Universitat Autònoma de Barcelona

Insights into the specificity and function of M14 metallocarboxypeptidases from structural and degradomic studies

Sebastian Martin Tanco

Universitat Autònoma de Barcelona

Insights into the specificity and function of M14 metallocarboxypeptidases from structural and degradomic studies

Doctoral thesis presented by Sebastian Martin Tanco
for the degree of PhD in Structure and Function of Proteins
from the Universitat Autònoma de Barcelona

Institut de Biotecnologia i Biomedicina
Unitat d'Enginyeria de Proteïnes i Proteòmica

Thesis supervised by
Prof. Julia Lorenzo Rivera and Prof. Francesc Xavier Avilés Puigvert

Sebastian Martin Tanco

Prof. Julia Lorenzo Rivera Prof. Francesc Xavier Avilés Puigvert

Barcelona, September 2012

A mi familia,

y a todos aquellos que en estos años actuaron como si lo fueran

Table of Contents

Abbreviations

Abbreviations

ACE	Angiotensin-converting enzyme
ACE2	Angiotensin-converting enzyme 2
ACI	Ascaris carboxypeptidase inhibitor
Ang-I	Angiotensin I
BMMCs	Bone marrow-derived mast cells
CCP	Cytosolic carboxypeptidase
CD	Catalytic domain
COFRADIC	COmbined FRActional DIagonal Chromatography
CPs	Carboxypeptidases
CPAs	A-type carboxypeptidases
CPA1	Carboxypeptidase A1
CPA2	Carboxypeptidase A2
CPA3	Carboxypeptidase A3
CPA4	Carboxypeptidase A4
CPA6	Carboxypeptidase A6
CPB	Carboxypeptidase B
CPU	Carboxypeptidase U
DmCPD1Bs	Drosophila melanogaster carboxypeptidase D variant 1B short
ECM	Extracellular matrix

ER	Endoplasmic reticulum
GEMSA	2-guanidinoethyl-mercaptosuccinic acid
LC/MS	Liquid chromatography/mass spectrometry
LCI	Leech carboxypeptidase inhibitor
MALDI-TOF MS	Matrix-assisted laser desorption ionization time-of-flight mass spectrometry
MC-CPA	Mast cell carboxypeptidase
MCPs	Metallocarboxypeptidases
mMCP-5	Mouse mast cell proteinase-5
MS/MS	Tandem mass spectrometry
PBS	Phosphate-buffered saline
PCI	Potato carboxypeptidase inhibitor
PCPA4	Procarboxypeptidase A4
PDB	Protein Data Bank
PVDF	Polyvinylidene difluoride
RP-HPLC	Reversed-phase high performance liquid chromatography
SCPs	Serine carboxypeptidases
SNP	Single-nucleotide polymorphism
TAFI	Thrombin-activatable fibrinolysis inhibitor
TAFI(a)	(activated) thrombin-activatable fibrinolysis inhibitor
TCI	Tick carboxypeptidase inhibitor

TFA	Trifluoroacetic acid
TLD	Transthyretin-like domain
TMAB	4-trimethyl-ammoniumbutyrate

Preface

The present thesis consists of four independent research works that are situated in the field of metallocarboxypeptidases: the study of their structure and their functional characterization.

The first work presents the crystal structure of a short isoform of *Drosophila melanogaster* carboxypeptidase D (CPD), which corresponds with the first repeat of the mammalian CPD homologues. This work was carried out in Barcelona, as part of a collaboration with the Proteolysis Lab of the Molecular Biology Institute of Barcelona, under the supervision of Dr. Julia Lorenzo Rivera, Prof. Francesc Xavier Aviles, Dr. Joan Lopez Arolas and Prof. Xavier Gomis-Ruth. Tibisay Guevara also participated in the realization of the work.

The second work comprises the biochemical and functional characterization of a cancer-associated human carboxypeptidase A4. This work was performed in collaboration with Fricker's lab at the Albert Einstein College of Medicine in New York, under the supervision of Dr. Julia Lorenzo Rivera, Prof. Francesc Xavier Aviles and Prof. Lloyd Fricker. In addition, Xin Zhang and Cain Morano contributed to the realization of this study.

The third and fourth works are the result of a collaboration with Proteome Analysis and Bioinformatics Unit at the University of Ghent in Belgium. As a result, Dr. Julia Lorenzo Rivera, Prof. Francesc Xavier Aviles, Prof. Petra Van Damme and Prof. Kris Gevaert participated in the supervision of both studies. The third work comprises the development of a novel proteome-derived peptide library to study the substrate specificity profiles of carboxypeptidases. Javier Garcia-Pardo, Sven Degroeve and Lennart Martens also experimentally contributes to this work.

The fourth work applies the C-terminal COFRADIC technology to screen for natural substrates of CCP1, a member of the novel subfamily of cytosolic carboxypeptidases, in a cellular context. This enzyme is considered to be implicated processes like degeneration and regeneration, thereby the study of its mechanism of action is of significant importance for a variety of neurological disorders. Olivia Tort also participated in the realization of this work.

The four research works presented in this thesis are the basis of the following scientific papers:

1. Tanco S, Zhang X, Morano C, Aviles FX, Lorenzo J and Fricker LD. "*Characterization of the substrate specificity of human carboxypeptidase A4 and implications for a role in extracellular peptide processing*". J Biol Chem (2010) 285: 18385-18396.
2. Tanco S, Arolas JL, Guevara T, Lorenzo J, Aviles FX and Gomis-Ruth FX. "*Structure-function analysis of the short splicing variant carboxypeptidase encoded by Drosophila melanogaster silver"*. J Mol Biol (2010) 401: 465-477.
3. Tanco S, Lorenzo J, Garcia-Pardo J, Degroeve S, Martens L, Aviles FX, Gevaert K and Van Damme P. "*Proteome-derived peptide libraries to study the substrate specificity profiles of carboxypeptidases".* Submitted to Molecular and Cellular Proteomics.
4. Tanco S, Tort O., Aviles FX, Gevaert K, Van Damme P and Lorenzo J. "*Positional proteomics identification of the natural substrates of cytosolic carboxypeptidases".* Manuscript in preparation.

Note about the format

This thesis is divided in four chapters that correspond to four different publications; two already published, one submitted and one still in the manuscript form. These publications have been adapted to provide them a unified format. Each chapter contain its own specific Introduction and Experimental section, although a general Background section provides the basic information for a better understanding of the thesis and its conclusions as a whole.

Background

B.1. Proteases

Proteases are enzymes that irreversibly cleave proteins by the catalysis of peptide-bond hydrolysis. The human genome encodes over 569 proteolytic enzymes or homologues, which represents 1.7% of human genes, indicating the importance of this class of enzyme. Proteases and inactiveprotease homologues form the largest enzyme family in man, with 566 members it is larger than the kinases (456 members) and second only to transcription factors in family size (Puente et al., 2003).

Initially proteases were considered unspecific protein degraders, a role that nonetheless is essential for living organisms (Barrett, 2000). These are composed largely of protein, and commonly to synthesize them, they need to obtain aminoacid units by breaking down preexisting protein molecules. In addition, processes like growth and remodeling of cells or tissues require the breakdown of old proteins in concert with the synthesis of new ones.

Recently more and more evidence well established that these enzymes, by the highly specific and limited substrate cleavage, can act as precise posttranslational modifiers of many protein molecules (auf dem Keller et al., 2007). For example, many newly synthesized proteins require a proteolytic processing to convert them to their biologically active forms. This is the case of the polyproteins produced by viruses, which need proteases to segment them. Notably, in both prokaryotic and eukaryotic cells, secreted proteins require the removal of the signal peptide in the secretory pathway. Many proteases are synthesized as catalytically inactive proenzymes, and are subsequently proteolytically activated at a biologically appropriate time and place (Barrett, 2000).

But proteases are not only important to initiate protein function but also to terminate it. A clear example of this is how caspases mediate apoptosis by the surgical cleavage of enzymes that are vital to the cell. In addition, the hydrolysis of proteins and polypeptides with signaling functions serves to keep the biological signals local in time and space (Barrett, 2000).

This way proteases regulate protein function by controlling appropriate intra- or extracellular localization; shedding from cell surfaces; activation or inactivation of proteases and other enzymes (like cytokines, hormones or growth factors); conversion of receptor agonists to antagonists; and exposure of cryptic neoproteins (product of a proteolytic cleavage that has a role distinct from the parent molecules) (Lopez-Otin and Overall, 2002).

With all proteins undergoing proteolysis — either during synthesis and maturation, to modify activity, or for degradation and turnover — proteases regulate virtually every biological process such as immunity, development, blood clotting and wound healing (Doucet and Overall, 2008). Proteases have been described to control processes such as DNA replication, cell-cycle progression, cell proliferation, differentiation and migration, morphogenesis and tissue remodeling, neuronal outgrowth, haemostasis, wound healing, angiogenesis and apoptosis (Lopez-Otin and Overall, 2002).

Accordingly, it is easy to see that deregulated proteolysis is a pathognomic feature of many diseases (Doucet and Overall, 2008). A deficiency, or a misdirected temporal and spatial activity, of these enzymes underlies several diseases such as cancer, arthritis, neurodegenerative and cardiovascular diseases. Moreover, many infectious microorganisms, viruses and parasites use proteases as virulence factors, and animal venom commonly contains proteases to effect tissue destruction or to evade host responses. Accordingly, many proteases or their substrates are an important focus of attention for the pharmaceutical industry as potential drug targets [4].

Different terms are used to define these enzymes. *Protease* is the most commonly used term. However, *peptidase* is also a general term for proteolytic enzymes, which can be thought of as a short form of “peptide bond hydrolase”. The *peptidase* root has been used to create terms for subtypes of proteolytic activities (i.e. endopeptidase, exopeptidase, aminopeptidase, and carboxypeptidase).

B.2 Classification of proteases

Any field of science needs a system of classification for the objects with which it deals, especially when they are numerous. For the classification of proteases there are three criteria that have worked well for that particular group of proteins (Barrett, 2000).

a) Catalytic type

Through evolution, proteases have adapted to the wide range of conditions found in complex organisms (e.g., variations in pH, reductive environment) and use different catalytic mechanisms for substrate hydrolysis. Proteases are thus generally classified according to the chemical nature of the catalytic site or, similarly, catalytic mechanism. Accordingly, there are six classes of proteases: aspartic, glutamic, cysteine, serine, threonine and metalloproteinase as well as a group of proteases of as yet unknown mechanism (Barrett et al., 2004). The threonine class, which is represented by the proteosome, that recently joined the original four classes proposed by Hartley (serine, cysteine, aspartic, and metallopeptidases). Glutamic proteases were discovered later and they constitute the only subtype not found in mammals so far. There are still a significant number of proteases for which the catalytic mechanism remains to be determined.

In the serine-, cysteine- and threonine-type of proteases the –OH or –SH group on the side chain of the amino acid at the heart of the catalytic site acts as a nucleophile in catalysis, forming an acyl enzyme intermediate. In contrast, the aspartic, glutamic and metallopeptidases seem to activate a bound water molecule that will serve as the nucleophile.

Table 1. Catalytic type of proteases and examples

Type	Examples
Aspartic	Pepsin, cathepsin E
Cysteine	Papain, cathepsin K
Metallo	Carboxypeptidase A, Thermolysin
Serine	Trypsin, prolyl oligopeptidase
Threonine	Proteasome
Unknown	gpr endopeptidase, prepilin type IV peptidase

b) The Type of Proteolytic Reaction Catalyzed

Enzymes are proteins that catalyze chemical reactions and so, it is logical to classify them based on the reactions they catalyze (Barrett, 2000; Barrett, 2001). The internationally recognized scheme for the classification and nomenclature of all enzymes, induding peptidases, is that of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) that assigns index numbers to enzymes (these numbers are known as "EC" numbers). In this system all enzymes are divided between six classes: (1) Oxido-reductases (2) Transferases (3) Hydrolases (4) Lyases (5) Isomerases (6) Ligases.

The hydrolases of peptide bonds, the peptidases, form subclass 3.4. Proteases are primarily divided into exopeptidases and endopeptidases. An exopeptidase acts only on one end of the polypeptide chain. Some act at a free N-terminus and release a single amino acid residue, a dipeptide, or a tripeptide (i.e., aminopeptidases, dipeptidylpeptidases, and tripeptidyl-peptidases respectively). Those acting at a free C-terminus may release a single residue (carboxypeptidases) or a dipeptide (peptidyl-dipeptidases). Other exopeptidases are specific for hydrolysis of dipeptides (dipeptidases) or remove terminal residues that are substituted, cyclized, or linked by

isopeptide bonds (i.e., peptide linkages other than those of α-carboxyl to α-amino groups; omega peptidases).

Instead, endopeptidases cleave internal bonds in polypeptide chains. Further, classification is based on the catalytic mechanism. Some endopeptidases are restricted to act on substrates smaller than proteins and these are described as "oligopeptidases".

The complete NC-IUBMB reccomendations can be found in http://www.chem.qmul.ac.uk/iubmb/enzyme/EC3/4/

Table 2. NC-IUBMB classification of peptidases

Sub-subclass	Peptidase type
3.4.11	Aminopeptidases
3.4.13	Dipeptidases
3.4.14	Dipeptidyl-peptidases
3.4.15	Peptidyl-di peptidases
3.4.16	Serine-type carboxypeptidases
3.4.17	Metallocarboxypeptidases
3.4.18	Cysteine-type carboxypeptidases
3.4.19	Omega peptidases
3.4.21	Serine endopeptidases
3.4.22	Cysteine endopeptidases
3.4.23	Aspartic endopeptidases
3.4.24	Metalloendopeptidases
3.4.25	Threonine endopeptidases
3.4.99	Endopeptidases of unknown type

c) Structural Relationships

In 1993 Rawlings and Barrett proposed the MEROPS system to classify proteases based on structural features that are believed to reflect evolutionary relationships among them (Rawlings and Barrett, 1993; Rawlings et al., 2012). This classification scheme deals with the limitations of the previous classification strategies, which were not able to account for the ample variability in structural and evolutionary

relationships (both in terms of primary and tertiary structure), catalytic mechanism and physiological function.

First, closely related proteases are grouped into families based on similarities in amino acid sequence. Each family of proteases is formed around a founding member, such as human carboxypeptidase A1 in family M14. The family is then built up by adding proteases that show statistically significant similarity in amino acid sequence to an existing member of the family. This analysis considers only the part of the protein that is responsible for the proteolytic activity (named “peptidase unit”). As such, for proteases that contain additional nonpeptidase domains, which might be shared with other groups of proteins, these additional domains are not considered for the classification. When families contain two or more distinct groups of proteases that differ greatly in primary structure, they are subdivided into subfamilies. Groups of families that show evidence of having had a common origin are then grouped together in clans.

An identifier is assigned to each family; it starts with a letter denoting the catalytic type of the peptidases (i.e., M for metalloproteases) plus a number. Subfamily names are derived the family name with the addition of a letter (e.g., M14A, M14B, M14C or M14D for subdivisions in family M14).

The complete list of families and clans recognized in the MEROPS can be found in the Web version of the MEROPS database (http://www.merops.co.uk).

B.3. Specificity-subsite terminology

The specificity of a protease is described using the conceptual model of Schechter & Berger (Schechter and Berger, 1967) that considers that each specificity subsite in the enzyme is able to accommodate the side-chain of a single amino acid residue in the substrate. Commonly, the active site of a peptidase is located on the surface of the enzyme and the substrate specificity is dictated by the properties of binding sites arranged along the groove on one or both sides of the catalytic site responsible for the peptide bond hydrolysis. Correspondingly, the enzyme binding sites are numbered

from the catalytic site S1, S2,…, Sn towards the N-terminus of the substrate, and S1', S2',..., Sn' towards the C-terminus. Similarly, the substrate amino acids that these binding sites accommodate are numbered P1, P2…Pn, and P1', P2'…Pn'

Substrate: P3 – P2 – P1 ↓ P1' – P2' – P3'

Protease : S3 – S2 – S1 * S1' – S2' – S3'

Here, the catalytic site of the enzyme is marked with an asterisc and the scissile bond is indicated by an arrow. In exopeptidases the cleft is likely to be "blind" on one side, not extending beyond S1 or S1' for an aminopeptidase or a carboxypeptidase, respectively.

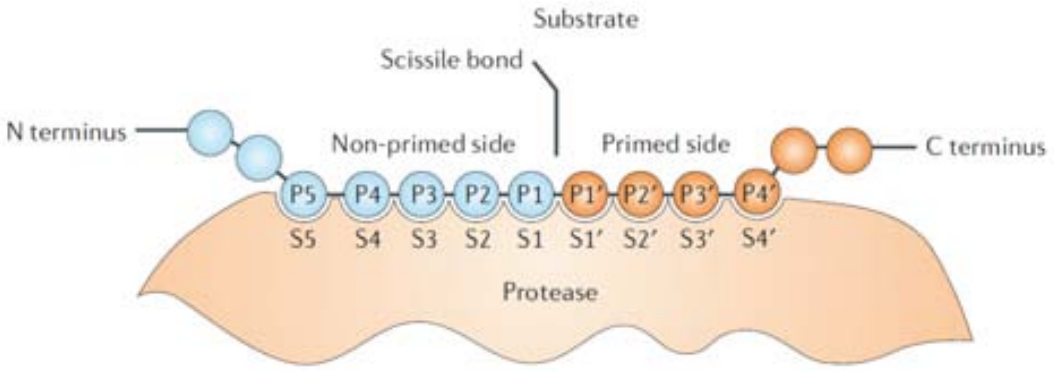

Figure 1. Representation of the conceptual model of Schechter & Berger

B.4 Carboxypeptidases

The main focus of this thesis will be on a type of exopeptidases named carboxypeptidases. Carboxypeptidases are proteolytic enzymes that catalyze the hydrolisis of peptidic bonds at the C-terminus of peptides and proteins. This hydrolysis may be a step in the degradation of substrate molecules or may result in the maturation of others. As for every type of proteolytic event the physiological effect of the hydrolytic action is thus varied and also site- and organism-dependent (Vendrell and Avilés, 1999)

The classification of carboxypeptidases follows the same rules that those described for proteolytic enzymes. Based on their catalytic mechanism, there are three types of carboxypeptidases: metallocarboxypeptidases (MCPs), serine carboxypeptidases (SCPs)

and cysteine carboxypeptidases. The classification of carboxypeptidases presented in Table 3 is based on that of the MEROPS database.

Table 3. Classification of carboxypeptidases. Adapted from (Vendrell and Avilés, 1999).

Catalytic type	Family	Enzymes	Catalytic feature
Metallo	**M2**	ACE2	**HEXXH…H**
	M14 subfamily A	CPA1 CPA2 CPA3 CPA4 CPA5 CPA6 CPB CPU/TAFI CPO	
	M14 subfamily B	CPE CPD CPM CPN CPZ CPX1 CPX2 AEBP1	**HXXE…H**
	M14 subfamily C	gamma-D-glutamyl-(L)-meso-diaminopimelate peptidase I	
	M14 subfamily D (proposed subfamily)	CCP1 CCP2 CCP3 CCP4 CCP5 CCP6	
	M15	zinc D-Ala-D-Ala carboxypeptidase	**HXXXXXXD…H**
	M20 subfamily A	glutamate carboxypeptidase	**HXD…D…EE…E…H**
	M20 subfamily D	carboxypeptidase Ss1	**DXD…D…EE…H**
	M28 subfamily B	glutamate carboxypeptidase II NAALADASE L peptidase	**HXD…D…EE…D…H**
	M32	carboxypeptidase Taq TcCP1 TcCP2	**HEXXH…H**
Serine	**S10**	Serine carboxypeptidase A Vitellogenic carboxypeptidase-like protein RISC peptidase	**catalytic triad S/D/H**
	S28	Lysosomal Pro-X carboxypeptidase	
Cysteine	**C1 subfamily A**	cathepsin X	**catalytic dyad C/H**

Whenever possible mammalian members of the family where shown.

The most crowded group of carboxypeptidases is the M14, to which belong the enzymes here studied, one of the most distributed and varied groups of carboxypeptidases. The second major group of carboxypeptidases is constituted by serine carboxypeptidases. The only representative of a cysteine carboxypeptidase is cathepsin X, a member of the cathepsin family.

Carboxypeptidases of the metallo-type are heterogeneous regarding their zinc-binding motifs, and are classified into distinct groups based on these motifs. These motifs comprise three metal binding residues: three histidines or two histidines plus an glutamate/aspartate. The extended zinc-binding consensus sequence may include a general base/acid glutamate. Some of these metallocarboxypeptidases display a catalytic site that is able to bind two metal ions, and thereby there are five metal binding residues. This is the case of the carboxypeptidases of the M20 and M28 families.

B.5. Metallocarboxypeptidases

The carboxypeptidases here studied are metallocarboxypeptidases that belong to the M14 family in the MEROPS classification and hereafter will be denoted as metallocarboxypeptidases (MCPs).

There are two well studied subfamilies of carboxypeptidases M14A and M14B according to the MEROPS database (Arolas et al., 2007; Gomis-Ruth, 2008; Vendrell et al., 2006). These two subfamilies are also described as A/B-type and N/E type respectively (according to their first identified members), or pancreatic and regulatory respectively. Recently, our group described a new subfamily of MCPs, the cytosolic carboxypeptidases or CCPs, which was proposed to constitute a new subfamily - M14D - in the MEROPS database (Kalinina et al., 2007; Rodriguez de la Vega et al., 2007). M14C is composed by microbial enzymes that participate in sporulation, such as gamma-D-glutamyl-(L)-meso-diaminopimelate peptidase I, and will not be consider in this description.

MCPs can also be classified, regarding their substrate specificity, into: CPs that prefer hydrophobic C-terminal amino acids (A-type enzymes), those that cleave C-terminal basic residues (B-type enzymes), those that recognize substrates with C-terminal aspartate or glutamate residues (O-type enzymes), and other CPs that display a broad substrate specificity (Lyons and Fricker, 2011). A-type MCPs are further classified into A1-type forms (that prefer both small aliphatic as well as bulky aromatic amino acids) and A2-type forms (showing a strong preference for large aromatic residues) (Gardell et al., 1988; Gomis-Ruth, 2008).

B.6. M14A metallocarboxypeptidases

To this subfamily belong the first studied carboxypeptidases, which were those enzymes produced by exocrine pancreas and secreted into the digestive tract to function in the breakdown of peptides (generated by digestive endopeptidases) in food (Waldschmidt-Leitz and Purr, 1929). These are the pancreatic carboxypeptidases A1 (CPA1), A2 (CPA2) and B (CPB).

Altogether, there are nine members of this subfamily in mammals. In addition to the pancreatic enzymes, this subfamily includes the carboxypeptidases A3 (CPA3 or mast cell carboxypeptidase), A4 (CPA4), A5 (CPA5), A6 (CPA6), O (CPO) and U (CPU or thrombin-activable fibrinolysis inhibitor (TAFI)). Orthologs can also be found in archaea, bacteria, plants, fungi, nematodes, amphibians or insect, among others.

They are structurally very uniform, since they are all secreted as inactive precursors (or zymogens) containing a proceeding signal peptide of 15-22 residues. They are known as procarboxypeptidases (PCPs), due to the presence of a 90-95 residues long pro-region that is placed at the N-terminus of the catalytic domain (CD) of 305-309 residues long (Fig. 2). This region is usually called “activation segment” and folds in a globular independent unit. The purpose of this pro-region is clear in the case of the pancreatic enzymes: the pancreas produce and store them as stable zymogens in granules. It blocks the access to the active site of the enzyme, which is already in an

active conformation in the zymogen. After secretion to the digestive tract, there is limited proteolysis by trypsin and the active enzymes are generated (Arolas et al., 2007). In summary, the role of this region is to maintain the enzymes in latency, until they are activated at the appropriate time and place.

Fig 2. Schematic representation of the structure of M14A MCPs. In green, the peptidase_M14 domain represents the carboxypeptidase domain. In red, Pro represents the pro-region or activation segment.

Two mammalian M14A MCPs escape to this description: CPA3 is stored in mast cell granules as an active enzyme, and its activation might occur within the granular compartment by a yet undefined protease (Pejler et al., 2009); CPA6 is activated by furin-like endoproteases in the secretory pathway and secreted as an active enzyme (Lyons et al., 2008).

It has been proposed that the pro-region would have a second role, contributing to the folding of the active domain during synthesis (Vendrell et al., 2000). CPO it is a curious case in this respect: CPO orthologs either do not contain any prodomain or contain a 20-28 residues long N-terminal domain (Lyons and Fricker, 2011). CPO does not require the presence of a pro-region to fold and render an active enzyme.

Although this subfamily is described as pancreatic/digestive carboxypeptidases, most of its members are involved in more selective processes. CPA3 is found in the secretory granules of mast cells and is suggested to have a role for CPA3 in regulating innate immune responses and mast cell granule homeostasis (Pejler et al., 2007). CPA4 presumably functions in neuropeptide processing and was linked to prostate cancer aggressiveness (Tanco et al., 2010). CPA6 is linked to the Duane syndrome, a congenital eye-movement disorder, and has been proposed to participate in the regulation of neuropeptides in the extracellular matrix within the olfactory bulb and other parts of the brain (Lyons et al., 2008). CPU/TAFI is a MCP of pharmacological interest for the prevention of thrombotic diseases. TAFI removes the C-terminal lysines

from partially degraded fibrin, and is one of the main intermediates in the cross regulation between coagulation and fibrinolysis (Declerck, 2011). CPO cleaves acidic amino acids from dietary proteins and peptides, and thus would complement the actions of well known digestive carboxypeptidases CPA1, CPA2 and CPB (Lyons and Fricker, 2011).

In terms of substrate specificity this subfamily it is rather variable, since there are examples of all kinds of substrate specificity: CPA1 to CPA6 are A-type MCPs, CPB and CPU/TAFI are examples of B-type MCPs, CPO is the classical example of a O-type MCP and CPT, a MCP from *Thermoactinomyces vulgaris*, displays an broad substrate specificity (Grishin et al., 2008).

B.7. M14B metallocarboxypeptidases

This subfamily of carboxypeptidases is composed of eight mammalian members. Five of them are considered to be enzymatically active: CPN, CPE, CPM, CPD and CPZ. Three members (CPX1, CPX2 and ACLP/AEBP1) lack critical active site residues and do not show activity against standard MCP substrates. The function of these inactive forms is not known: either cleave other kind of substrates or act as binding proteins (Arolas et al., 2007; Gomis-Ruth, 2008).

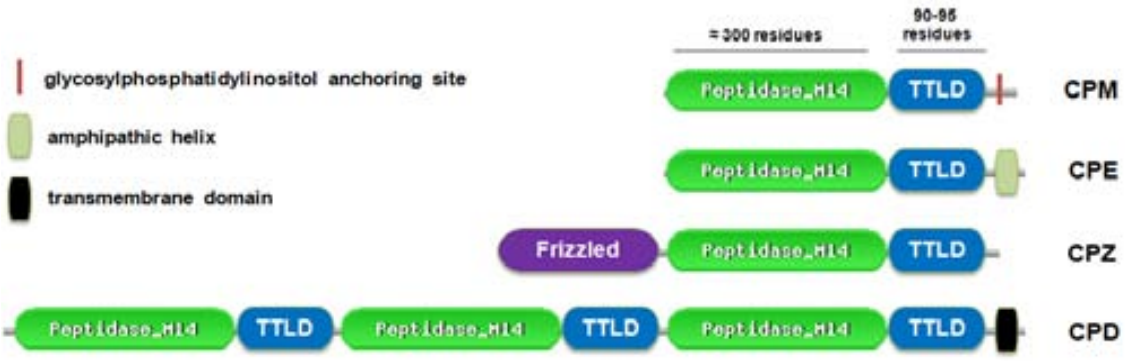

Fig 3. Schematic representation of the structure of M14B MCPs. In green, peptidase_M14 represents the catalytic carboxypeptidase domain. In blue, TLD represents the transthyretin-like domain. In purple, Frizzled represents the cysteine-rich region homologue to the frizzled receptors.

This group of MCPs is more variable structurally than M14A enzymes (Fig. 3). They are secreted as active enzymes, without a pro-region to regulate their activity. M14B MCPs contain a 80-90 residues long domain following the 300 residues CD, which shows structural homology to transthyretin (transthyretin-like domain or TLD), and similarly to the pro-region of M14A enzymes has been proposed to function as folding domain (Gomis-Ruth et al., 1999). All members of this subfamily contain additional domains. In the case of CPE, CPM and CPN this domain is small. CPE contains an amphipathic helix that allows the protein to bind peripherally to membranes of the secretory pathway. CPN, is found as a 280 kDa heterotetrameric complex, composed of two catalytic subunits of 48-55 kDa, and two carbohydrate-containing regulatory/noncatalytic subunits of 83 kDa each. These regulatory subunits have the function to carry and stabilize the catalytic subunits in the circulation (Skidgel, 1995). CPM contains a sequence for the addition of glycosylphosphatidylinositol that anchors the protein to the extracellular membrane, although soluble forms have been found in human body fluids (Skidgel et al., 1996). In CPZ this extra domain is larger, and is a cysteine-rich region homologue to the frizzled receptors and other Wnt-binding proteins. CPD contains multiple MCP domains. It is composed of three tandem repeats of approximately 390 residues, each comprising an active MCP domain followed by a C-terminal transthyretin-like domain (CD+TLD repeats). In addition it contains a transmembrane domain and a cytosolic tail 60 residues long. While the first two repeats are catalytically active the third repeat corresponds to an inactive MCP.

In contrast to the structural variability, they all show a similar B-type substrate specificity for basic arginines and lysines residues, although with differences in the preference for one or other residue. Most of the members of this subfamily function in the processing of peptide hormones and neurotransmitters. CPE is primarily found in the regulatory secretory vesicles and is mainly involved in the biosynthesis of numerous peptide hormones and neurotransmitters (Fricker, 2004a; Fricker and Snyder, 1983). CPN cleaves the C-terminal arginine residue of bradykinin, kallidin, complement anaphylatoxins (C3a, C4a, and C5a) and inflammatory fibrinopeptides, and has been implicated in inflammation (Skidgel and Erdos, 2007). The function of

CPM is not fully understood, but was proposed to be involved in the process of inflammation. In addition, CPM may act as binding partner in cell-surface protein–protein interactions (Deiteren et al., 2009). CPZ is a carboxypeptiddase that is present in the extracellular matrix and was shown to interact with Wnt4, and has thus been proposed to function in embryonic development (Moeller et al., 2003). CPD is found in the *trans* Golgi network and the plasma membrane. This CP is thought to further process proteins previously processed by the pro-protein convertases in the *trans* Golgi network (Fricker, 2004b).

B.8. Cytosolic metallocarboxypeptidases

This group of enzymes constitutes the most recently discovered subfamily of metallocarboxypeptidases (Kalinina et al., 2007; Rodriguez de la Vega et al., 2007). This subfamily was proposed to form a new subfamily (M14D) in the MEROPS database, provided that phylogenetic studies support the divergence of these enzymes from the previous MCP subfamilies in the MEROPS database (Rodriguez de la Vega et al., 2007). In contrast to the other subfamilies, this MCPs show cytosolic/nuclear localization (Kalinina et al., 2007), and for this reason were named cytosolic carboxypeptidases or CCPs. These enzymes present a wide taxonomic distribution: homologues of this subfamily can be found in Bacteria, Protista, Animalia, Fungi, or Plantae; but not in Archaea. The number of paralogs varies between species; in Bacteria there is usually only one gene, meanwhile in the ciliated protozoan *Paramecium tetraurelia* up to 32 genes can be found. In mammals there are 6 members, named CCP1 to CCP6. There is variability in terms of protein length and domain composition, as can be observed for the human members of this subfamily (Fig. 4)

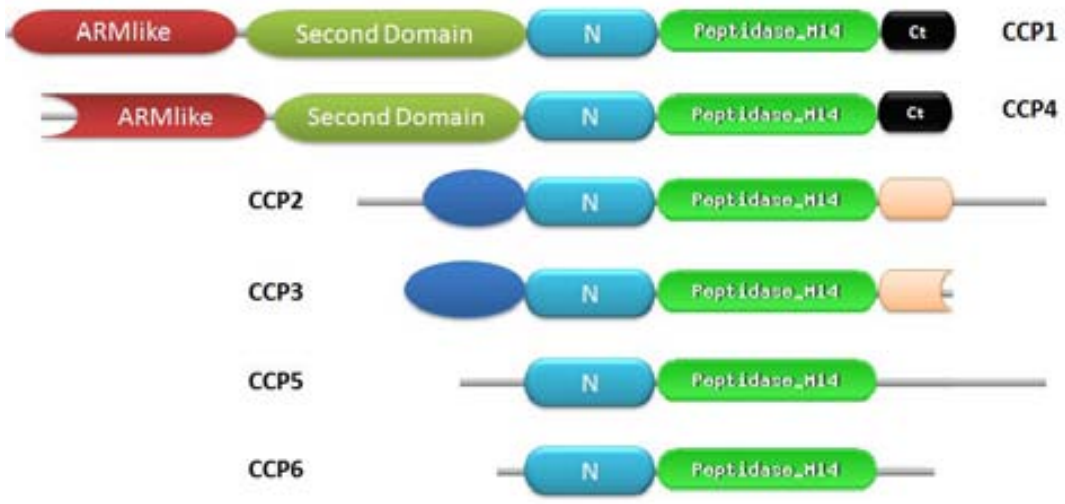

Fig 4. Schematic representation of the structure of CCPs. In green, peptidase_M14 represents the carboxypeptidase catalytic domain. In light blue, N represents the conserved N-terminal domain in CCPs. In red, Arm represents an Arm repeat domain. The other represented domains are poorly defined and show no homology with other known domains.

All the members of the CCP subfamily share a CD domain of approximately 300 residues and, in addition, an N-terminal conserved domain of 150 residues. Recently, the structure of a CCP from *Pseudomonas aeruginosa* was solved (Otero et al., 2012), showing that this N-terminal domain comprises a β-sandwich structure. This domain is different from the activation region of M14A MCPs and transthyretin-like domain of M14B MCPs. This N-terminal displays a previously undescribed fold and partially blocks the entrance to the active site, suggesting a role in the control of the active site accessibility. Of note is the presence of an armadillo repeat domain in the N-terminus of CCP1 and CCP4 (Fig. 4). These domains are often implicated in protein/protein or protein/nucleic acid binding.

Although they are quite variable structurally, they are quite similar in terms of substrate specificity and function. For mouse CCP1, CCP4, CCP5 and CCP6 it has been suggested a role as deglutamylases, shortening post-translationally added polyglutamate side chains in tubulin (Rogowski et al., 2010). These enzymes would trim gene-encoded C-terminal glutamate residues from tubulin (previously detyrosinated at its C-terminus), MLCK and telokin (Rogowski et al., 2010). Work performed in our laboratory suggests that CCP3 would perform a similar function (Tort

et al., unpublished results). Although it has been suggested a different function for CCP2 (Sahab et al., 2011), this could not be replicated in our group. Similarly, they would all have similar O-type substrate specificity for C-terminal glutamate residues.

B.9. Structure of metallocarboxypeptidases

The enzymes of the M14 family of MCPs have a compact globular shape, which resembles the volume obtained when a cone is extracted from a sphere. They present a funnel-like opening at the top, and for this, they are described as funnelins when are analyzed from a structural point of view (Fig. 5) (Gomis-Ruth, 2008).

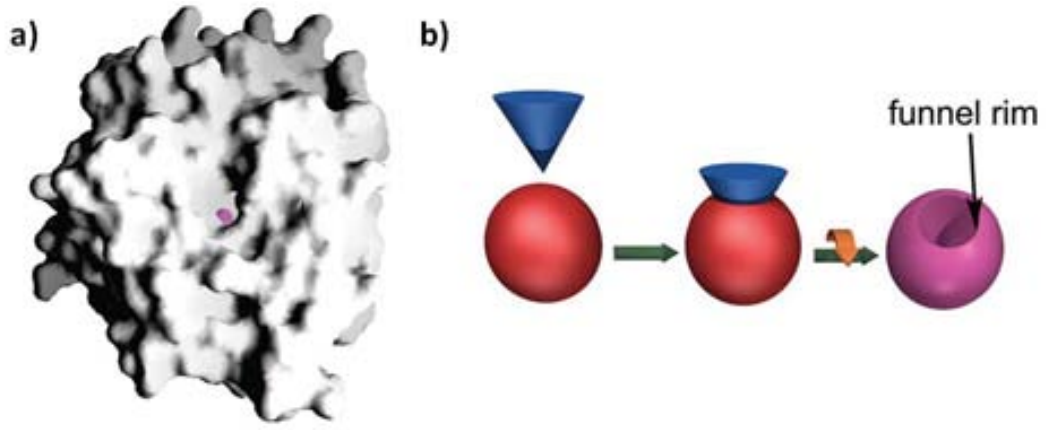

Fig. 5. Carboxypeptidase funnel-like structure. a) Molecular surface representation of human CPA4 showing the funnel-like aperture that leads to the catalytic zinc cation at the bottom (magenta sphere). b) Scheme illustrating the generation of an inverted spherical cone (magenta) through the intersection of a regular cone (blue) and a sphere (red). The resulting border is like a funnel rim. This figure is adapted from (Gomis-Ruth, 2008)

Structures derived from members of the three main subfamilies of MCPs show that all these enzymes share a similar conformational fold and conserve a similar architecture at the active site. MCP/Funnelin CD show a consensus α_8/β_8-topology conforming to an α/β-hydrolase or PLEES fold (Ollis et al., 1992; Puente and López-

Otín, 1997). They contain a central doubly-wound eight-stranded β-sheet (β1–β8) of mixed parallel/antiparallel topology with strand connectivity +1, +2, −1x, −2x, −2, +1x, -2. The core of the sheet consists of four parallel coplanar central strands (β3–β5 and β8) and the catalytic site is located at the C-terminal end of these strands. These four strands are flanked by two parallel strands on the top (β6–β7) and a β-ribbon (β1–β2) at the bottom (Fig. 6). This leads to a concave front side of the sheet, which accommodates helices α5 and α7, as well as the active-site cleft. All funnelins studied except DCPD2, HSCPM, and TVCPT contain an additional short helix inserted after the upper strand β7 (Figure 6b). At the back of the molecule, the convex side of the sheet embraces helices α1–α4, α6 and α8 and the surface N- and C-termini of the molecule.

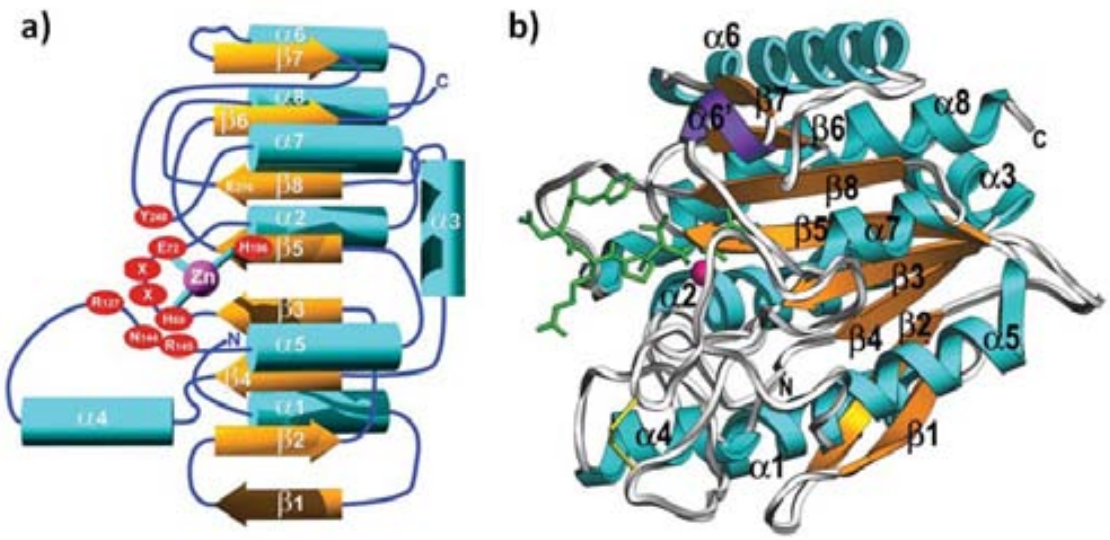

FIG. 6. Structural features of MCPs/funnelins. a)Topology scheme illustrating the consensus secondary-structure elements (α-helices in cyan and β-strands in orange) found in all known MCPs/funnelins structures. The highlighted amino acids are contained in the characteristic set of conserved residues HXXE+R+NR+H+Y+E, and are numbered using the numbering system of bovine CPA. The metal Zn^{+2} ion is shown as a magenta sphere. B) Richardson diagram of human CPA4. The consensus secondary-structure elements and the metal site are colored and labeled as in a). An additional element to the consensus α8/β8-structure is helix α6' (purple ribbon), integrated into loop Lβ7α7. A bound and subsequently cleaved hexapeptide (green stick model) is also displayed. The N- and C-termini are labeled for reference. This figure is adapted from (Gomis-Ruth, 2008).

The funnel-like access to the active site is shaped by a series of irregular segments of varying length, namely Lβ8α8, Lβ5β6, Lβ7α7, Lβ3α2, and, in particular, the residue

loop Lα4α5 in the lower left of the molecule, which delimits the active site cleft (Fig. 6). These segments are required for interactions of the CD with the pro-region in the zymogen and with the proteininhibitors. The most significant deviations from the canonical structure of bovine CPA are found in M14B MCPs structures and localized in this loop segments that shape the active site.

The catalytic zinc ion resides at the bottom of the active site cleft and is coordinated by His69 Nδ1 and by Glu72 atoms Oε2 and Oε1 in a bidentate manner. These two residues are provided by Lβ7α5 and belong to the characteristic HXXE consensus sequence of M14 MCPs. A third protein zinc ligand is His196 Nδ1, which together with a metal-bound solvent molecule gives rise to an overall distorted tetrahedral coordination sphere (this gives rise to a distorted tetrahedral coordination sphere, in which one of the tetrahedron corners bifurcates into the two carboxylate oxygen atoms of Glu72) (Fig. 6a).

Objectives

The laboratory of Professor Francesc Xavier Aviles at the Institut de Biotecnologia i Biomedicina at the Universitat Autònoma de Barcelona, has for many years focused its research in the study of the structure and function of metallocarboxypeptidases.

In this context, the general objective of this thesis has been to characterize different metallocarboxypeptidases of biomedical interest and to develop new tools for the study of these enzymes.

Specific objectives were proposed for each chapter:

CHAPTER I

- To optimize the recombinant expression and purification in *Pichia pastoris* of the short isoform (DmCPD1Bs) of *Drosophila melanogaster* carboxypeptidase D.
- To characterize in detail its enzymatic properties and the susceptibility to inhibitors
- To determine the three-dimensional structure of this protein.

CHAPTER 2

- To characterize the enzymatic properties of human carboxypeptidase A4 in term of substrate specificity and pH optimum
- To study the subcellular distribution and secretion of this enzyme.
- To develop a quantitative peptidomic approach to search for peptidic substrates of carboxypeptidases.

CHAPTER 3

- To develop a COFRADIC-based proteome-derived peptide library for in-depth profiling of the substrates preferences of carboxypeptidases.
- To apply this approach to human carboxypeptidases with known substrate preferences.
- To use this approach to delineate the poorly characterized substrate specificity profile of mast cell carboxypeptidase.

CHAPTER 4

- To apply C-terminal COFRADIC to search for natural substrates of cytosolic carboxypeptidase 1 (CCP1) in a cellular context.
- To determine the substrate preferences of this enzyme.
- To gain insights into the function of cytosolic carboxypeptidases in the cell.

Chapter 1

Structure-function analysis of the short splicing-variant carboxypeptidase encoded by *Drosophila melanogaster silver* (DmCPD1Bs).

1.1 INTRODUCTION

The *silver* (*svr*) gene was discovered in *Drosophila melanogaster* by Bridges in the early 1920s and it maps near the distal end of chromosome X (Lindsley and Zimm, 1992; Morgan et al., 1925; Settle et al., 1995; Tweedie et al., 2009). Adult fruitflies with mutations in this gene display cuticles that are pale and silvery in color due to reduced melanization, a finding that gave rise to the name of the gene. *Silver* includes nine exons (1A, 1B, and 2-8) which encode for eight different mRNA splicing variants. These result in two short protein variants of 433 (protein svr-PE *alias* "1A short") and 435 (svr-PF *alias* "1B short" and DmCPD1Bs) residues and in six long forms spanning between 1,259 and 1,439 residues (svr-PB, svr-PC, svr-PD, svr-PG, svr-PH, and svr-PI; see UniProt sequence database entry P42787 and Fig. 1) (Settle et al., 1995; Sidyelyeva and Fricker, 2002; Tweedie et al., 2009). The long variants possess the overall modular structure and sequence features of carboxypeptidase D (CPD), a glycosylated 180-kDa enzyme studied since the middle 1990s in fruitfly, mouse, rat, duck, bovine, chicken, and humans (Arolas et al., 2007; Song and Fricker, 1995, 1996). CPD was localized to the *trans*-Golgi network, the secretory and reuptake pathways, and transiently on the cell surface (Fricker, 2002). It was hypothesized to remove basic C-terminal residues from proteins and peptides supporting the endoproteolytic action of furin-like serine proprotein convertases of the subtilisin/kexin type. Thus, CPD would contribute to processing of hormones (e.g., adipokinetic hormone, bradykinin), growth factors, neurotransmitters, and other bioactive peptides (e.g., dynorphin B-14) that follow the secretory and endocytic pathways (Fricker, 2002; Fricker, 2004; Sidyelyeva et al., 2010; Skidgel and Erdos, 1998). It has been suggested that the Arg released by CPD from peptide substrates may stimulate production of NO, an important regulator of cellular processes like neurotransmission, vasodilation or tumor progression and survival (Abdelmagid and Too, 2008; Hadkar and Skidgel, 2001).

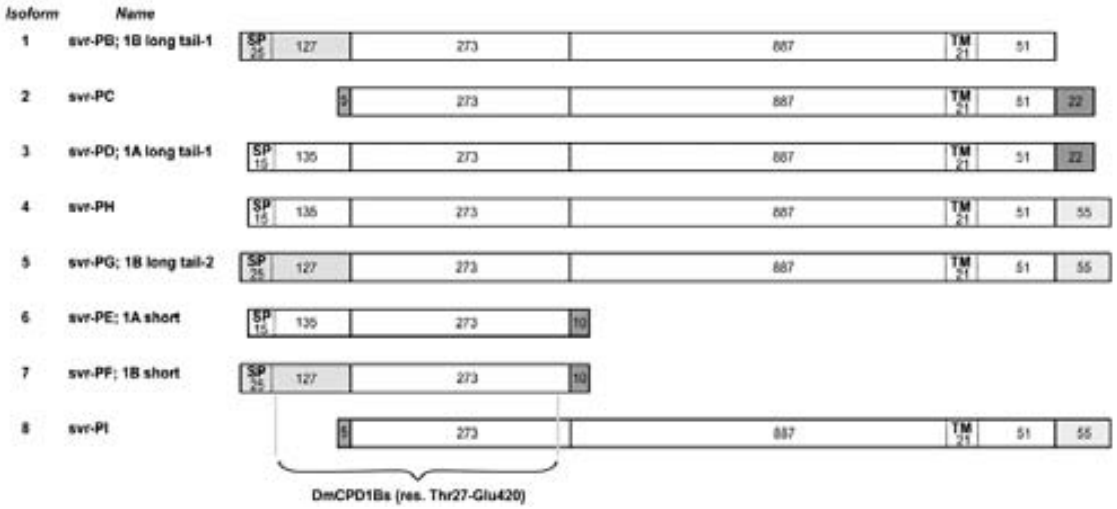

Figure 1. Protein isoforms described for the *silver* gene. These isoforms are the result of differential splicing (UniProt database entry P42787). Each exon product is shown as a distinct bar with the residues it spans. The N-terminal fragments of 150 (encoded by exon 1A) and 152 (encoded by exon 1B) residues show ~39 % sequence identity, the extent of the signal peptide (SP) of the former was estimated with program SignalP at www.expasy.ch. The short fragments of 5 and 10 residues, as well as the C-terminal fragments of 22 and 55, are unrelated to any other sequence of the protein. The extent of the transmembrane region (TM) is according to the UniProt entry. Isoform 7 *alias* svr-PF, 1B short, herein DmCPD1Bs, spans 435 residues, of which 27-420 correspond to the first CD+TLD repeat studied herein. The flanking peptides protrude from the surface and are unstructured. The last residue encoded by exon 1B is Gln152. See also (Sidyelyeva et al., 2006).

CPD belongs to the M14B subfamily of MCPs (Gomis-Rüth, 2008; Rawlings et al., 2010). The members of this subfamily are characterized for containing, in addition to its ~300-residue α/β-hydrolase domain (catalytic domain or CD), an additional ~80-residue C-terminal transthyretin-like domain (TLD) (Fricker, 2002; Fricker, 2004). In addition to CPD, seven further M14B MCPs are found in humans and mice, of which five show enzymatic activity (CPD, CPE, CPM, CPN, and CPZ) and three do not (AEBP1/ACLP, CPX1, and CPX2) (Arolas et al., 2007). In contrast, Drosophila comprises only CPD and CPM orthologs (Sidyelyeva et al., 2006; Sidyelyeva and Fricker, 2002). In vertebrates, CPM binds to cell membranes, but it also circulates in body fluids. It is assumed to participate in the extracellular processing of peptides and thus to have an activity comparable to CPD (Skidgel, 1988, 1996; Skidgel, 2004).

CPD is unique in having three CD+TLD repeats followed by a transmembrane anchor and a small cytosolic domain, which may be required for intracellular routing and export to nascent secretory vesicles (Arolas et al., 2007; Fricker, 2002; Gomis-

Rüth, 2008; Gomis-Rüth et al., 1999; Reznik and Fricker, 2001). Only the first two MCP domains are catalytically competent in mammals and duck. This also happens in *Drosophila* variants encoded by exon 1B (see (Sidyelyeva et al., 2006; Sidyelyeva and Fricker, 2002) and Fig. 1). When isolated, the two MCP repeats show different pH optima and specificity. The N-terminal repeat is more active at neutral pH and prefers arginine over lysine at the C-terminus of peptide substrates. In contrast, the second repeat prefers slightly acidic environments (pH 5-6) and excises lysines slightly more efficiently than arginines (Sidyelyeva and Fricker, 2002). Variants encompassing both repeats have broader pH optimum and substrate specificity than single-repeat variants. This is consistent with the conditions found throughout the secretory pathway, in which pH ranges from 5 to 7 (Sidyelyeva et al., 2006; Sidyelyeva and Fricker, 2002). The third MCP domain of CPD is catalytically incompetent as it lacks critical residues, but it may maintain the overall α/β-hydrolase scaffold (Fricker, 2002). In the duck, it was reported to be recognized by the pre-S region of the large envelope protein of hepatitis B virus during infection (Eng et al., 1998; Kuroki et al., 1994). This is reminiscent of ACE2, a CP belonging to another structural group, the cowrins (Gomis-Rüth, 2009), which is the receptor for human SARS coronavirus (Li et al., 2005). In contrast to the long, three-repeat variants, the two *Drosophila* short forms have only the first repeat, and they are secreted and soluble (Figs. 1 and 2). Among them, only DmCPD1Bs is catalytically active (Sidyelyeva et al., 2006; Sidyelyeva and Fricker, 2002). Although both short forms are present in the Golgi and may participate in physiological processes, the lack of membrane anchors makes them transient. Moreover, the endocytic vesicles contain only the long forms (Sidyelyeva et al., 2006). In addition, while the membrane anchors can regulate the long forms through localization/compartmentalization, the soluble DmCPD1Bs may be regulated in a different way along the secretory pathway.

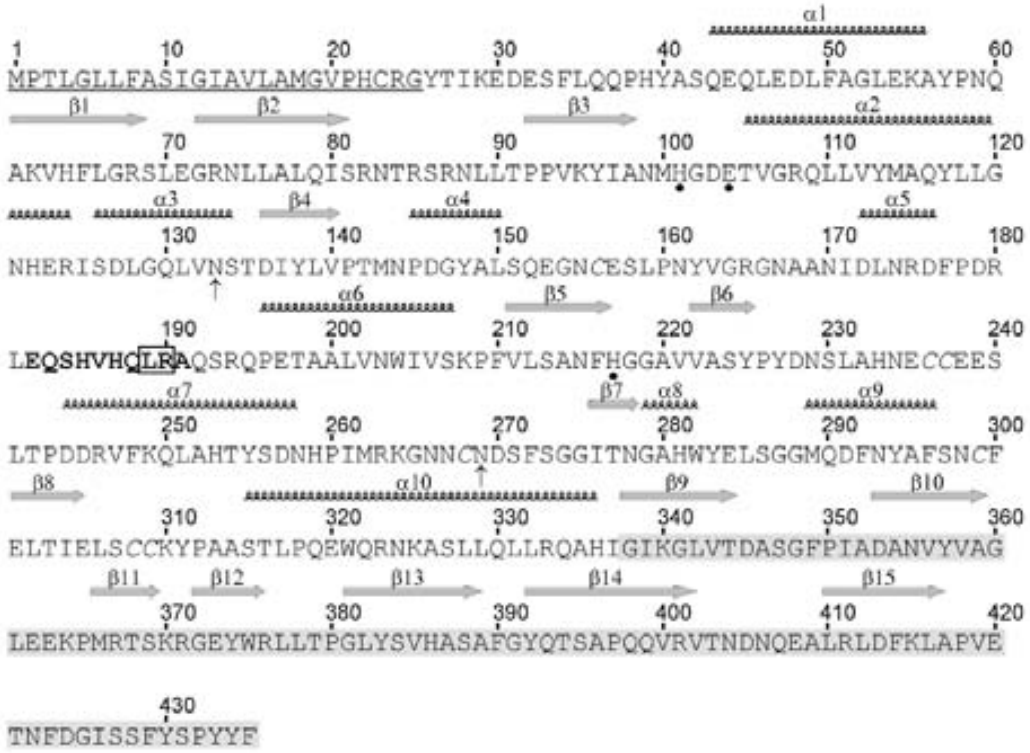

Fig. 2. Full-length amino-acid sequence of DmCPD1Bs in single-letter code. The 25-residue signal peptide is underlined, the residues involved in zinc coordination are labeled with full circles, and the two glycosylation sites are indicated by arrows. The cysteine pairing observed in the structure is Cys156-Cys309, Cys236-Cys237, and Cys268-Cys308. The sequential numbering employed throughout the text and the regular secondary structure elements are shown above each sequence block. The transthyretin-like domain is hallmarked by a light gray background. The segment of the regulatory loop, which is disordered in the structure, is shown in bold and the experimentally verified serine protease cleavage sites are framed.

Drosophila silver *alias* CPD is essential for life as mutants that do not express functional CPD are not viable and die early in the larval stage. The enzyme plays a major role in melanization and sclerotization of the insect cuticle, wing morphogenesis, catecholamine metabolism, phagocytosis, engulfment, memory, and sensitivity to cold and ethanol (Lindsley and Zimm, 1992; Morgan et al., 1925; Settle et al., 1995; Sidyelyeva et al., 2006; Sidyelyeva and Fricker, 2002; Sidyelyeva et al.; Tweedie et al., 2009; Wright, 1987). Interestingly, mutants that contain a functional DmCPD1Bs but no three-repeat variant are viable, but they have a silvery body and minor defects in the wings. Therefore, this short form is necessary and sufficient for viability. In order to investigate the detailed molecular basis for its activity and

regulation, we studied the structure and function of DmCPD1Bs, and deduced a plausible hypothesis for its regulation *in vivo*.

1.2 EXPERIMENTAL SECTION

Protein expression and purification

A clone encoding protein svr-PB of *Drosophila* was kindly provided by Dr. Lloyd D. Fricker (Albert Einstein College of Medicine, NY) and was used as template to generate a PCR-construct encoding DmCPD1Bs without the signal peptide, i.e. encompassing amino acids Tyr26 to Phe435 (UniProt P42787, isoform 7). This construct was cloned into the pPICZαA vector with a C-terminal His_6-tag, and transformed into *Pichia pastoris* strain KM71H following the manufacturer's instructions (EasySelect *Pichia* Expression Kit, Invitrogen). Selected transformant colonies were used to inoculate 1-L shake-flask cultures, which were grown at 28°C for 36h in buffered glycerol-complex medium (BMGY) until OD_{600} reached 20-30. Cells were collected by centrifugation at 3,000 *g*, gently resuspended in 100mL fresh BMMY medium (BMGY medium supplemented with 1 % of methanol instead of 1 % of glycerol), and cultured at 28°C for another 24h for protein expression. PMSF (0.3 mM) was added to the culture every 3-4h to prevent cleavage of the recombinant protein. For protein purification, the culture supernatant was equilibrated with 30 % ammonium sulfate, bound to a hydrophobic chromatography column (butyl-Toyopearl 650M, Tosoh Bioscience) connected to an ÄKTA Purifier system (GE Healthcare), and eluted with a decreasing gradient of ammonium sulfate. The eluted sample was subsequently dialyzed against 20 mM Tris·HCl, pH 8.0, and purified by anion-exchange chromatography (TSK-DEAE 5PW, Tosoh Bioscience) by using a linear gradient of 0.4 M ammonium acetate (0 to 30 %). DmCPD1Bs was then loaded onto a HiLoad Superdex 75 26/60 column (GE Healthcare) previously equilibrated with 50mM Tris·HCl, pH 7.5, 250 mM NaCl. Eluted fractions were analyzed by SDS-PAGE, and the purest samples containing the active enzyme were pooled. The protein was buffer-exchanged to 10mM Tris·HCl, pH 7.5, 50 mM NaCl, and concentrated by using an Amicon Centricon centrifugal device (10-kDa cutoff, Millipore). The typical final yield was ~ 0.3 mg of protein per liter of growth medium. Western blot using an anti-His monoclonal antibody (GE Healthcare) showed this protein lacked the C-terminal His_6-tag, possibly due to autolysis during production/purification, as described for other MCP s (Lin et al., 1996; Lyons et al.,

2008). Point mutants N133Q, N269Q, and N133Q/N269Q were obtained by site-directed mutagenesis by using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) and produced and purified in the same way as the wild-type protein. Human pancreatic procarboxypeptidase B was overexpressed and purified as described for DmCPD1Bs but without addition of PMSF. The enzyme was activated by controlled trypsin digestion as described (Barbosa Pereira et al., 2002).

Proteolytic activity assays

Activity was assayed by monitoring the rate of hydrolysis of hippuryl-L-Arg (Sigma) at λ =254 nm on a Cary 400 UV-Vis spectrophotometer (Varian) with 50 mM Tris·HCl, pH 7.5, 0.15 M NaCl as buffer at 37ºC. Initial turnover rates were determined from the first 5 %-10 % of the time-trace of each reaction for substrate concentrations close to the K_m value whenever possible. The kinetic parameters, k_{cat} and K_m, were calculated from at least six experimental points by direct fit to a Michaelis-Menten curve applying a nonlinear least-squares regression analysis using the Grafit program. Three independent experiments were performed.

Inhibitory assays

The apparent K_i of GEMSA (Calbiochem) against DmCPD1Bs was determined by measuring the inhibition of the hydrolysis of the chromogenic substrate *N*-(4-methoxyphenylazoformyl)-Arg-OH (Bachem) at λ=350nm by considering linear competitive kinetics as previously described (Fricker and Snyder, 1983). Assays were performed with 100 µM of substrate in 50 mM Tris·HCl, pH 7.5. Similar inhibitory experiments were carried out by using the protein carboxypeptidase inhibitors from potato (PCI), leech (LCI), tick (TCI), and human (latexin). These protein inhibitors were produced by heterologous overexpression in *Escherichia coli* and purified as published elsewhere (Pallarès et al., 2005; Sanglas et al., 2009).

Proteolytic susceptibility assays

1.3 μg of DmCPD1Bs was incubated with either 30 ng of bovine trypsin (Sigma) or 70 ng of subtilisin Carlsberg (Sigma) for 30 min at 37°C in 50 mM Tris·HCl, pH 7.5. Equivalent samples without serine protease were used as controls. Cleavage products were analyzed by Coomassie-stained 10 %-SDS-PAGE, transferred to polyvinylidene difluoride membranes (Millipore), and subjected to automated Edman degradation analysis in a Procise 492 protein sequencer (Applied Biosystems) at the inhouse proteomics service at the Universitat Autònoma de Barcelona. Digested samples and controls were tested for CP activity as described for the inhibitory assays. Each experiment was performed six times.

Dynamic light scattering and cross-linking experiments

Dynamic light scattering measurements were performed at six DmCPD1Bs concentrations (ranging from 0.5 mg/mL to 18 mg/mL) in 10 mM Tris·HCl, pH 7.5, 50 mM NaCl in a Zetasizer Nanoseries apparatus (Malvern Instruments). Each sample was measured 30 times and the results were analyzed with the Dispersion Technology Software v5.1. For the cross-linking experiments, DmCPD1Bs was prepared at six protein concentrations as above and treated with 0.8 mM glutaraldehyde for 2 min at room temperature in a final volume of 20 μL. The reaction was thereafter quenched with 12 μL of 50 mM Tris·HCl, pH 7.5. Samples were analyzed by Coomassie-stained 10 %-SDS-PAGE.

Crystallization and X-ray diffraction data collection

Crystallization assays followed the sitting-drop vapor diffusion method. Reservoir solutions were prepared by a Tecan robot and 100-nL crystallization drops were dispensed on 96x2-well MRC plates (Innovadyne) by a Cartesian (Genomic Solutions) or a Phoenix (Art Robbins/Rigaku) nanodrop robot at the High-Throughput Crystallography Platform (PAC) at the Barcelona Science Park. The best crystals appeared in a Bruker steady-temperature crystal farm at 4°C with protein solution (18 mg/mL in 10 mM Tris·HCl, pH 7.5, 50 mM NaCl) and 8 % PEG 4000, 0.2 M KSCN, 0.1 M

sodium cacodylate, pH 6.5 as reservoir solution. These conditions were efficiently scaled up to the microliter range with 24-well Cryschem crystallization dishes (Hampton Research). Complex crystals with GEMSA were obtained by soaking (10 mM inhibitor in reservoir solution) for 5 days. Crystals were cryo-protected with 16.5 % PEG 4000, 15 % glycerol, 0.2 M KSCN, 0.1 M sodium cacodylate, pH 6.5. A complete diffraction dataset at 2.7Å resolution was collected at 100K (Oxford Cryosystems 700 series cryostream) from a single liquid-N_2 flash-cryo-cooled GEMSA-complexed crystal on an ADSC Q315R CCD-detector at beam line ID23-1 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) within the Block Allocation Group "BAG Barcelona". The crystal was primitive orthorhombic, with four molecules per asymmetric unit. Diffraction data were integrated, scaled, merged, and reduced with programs XDS (Kabsch, 2001) and SCALA (Evans, 2006) within the CCP4 suite of programs (see Table 1). The presence of pseudomerohedral twinning following twin law “-h,l,k” was identified by using XTRIAGE within the PHENIX suite of programs (Adams et al., 2002).

Structure solution and refinement

The structure of DmCPD1Bs was solved by Patterson-search methods with program PHASER (McCoy et al., 2007) by using the co-ordinates of duck CPD domain II (Protein Data Bank (PDB) access code 1H8L; (Aloy et al., 2001; Gomis-Rüth et al., 1999)) as searching model. Subsequently, manual model building on a Silicon-Graphics workstation with program TURBO-Frodo (Carranza et al., 1999) alternated with crystallographic refinement (including TLS refinement) with program REFMAC5 within the CCP4 suite at initial stages and with program suite PHENIX under consideration of twinning (refined to a fraction of α=0.079) at the final stages until the model was completed (see Table 1). This model contained protein residues Lys29-Val419 (numbering according Fig. 2) for chain A, Ile28-Glu420 for chain B, Glu30-Val419 for chain C, and Thr27-Glu420 for chain D *plus* a zinc ion, an N-acetyl-glucosamine sugar moiety attached to Asn133, and a GEMSA molecule for each protomer. Although there was evidence that Asn270 was also glycosylated, the electron density maps were too weak to enable reliable modeling. Segments 183-190, 185-191, 184-188, and 182-190

were disordered in each of the four chains, respectively. SDS-PAGE analysis of washed crystals revealed only intact protein.

Table 1. Crystallographic data.

Space group / cell constants (a,b, and c, in Å)	$P2_12_12_1$ / 97.0, 135.7, 141.7
Wavelength (Å)	0.9724
No. of measurements / unique reflections	419,491 / 52,035
Resolution range (Å) (outermost shell)	49.0 – 2.70 (2.85 – 2.70)
Completeness (%)	99.9 (100.0)
R_{merge} [a]	0.099 (0.590)
$R_{r.i.m.}$ (= R_{meas}) [a] / $R_{p.i.m.}$ [a]	0.106 (0.629) / 0.037 (0.217)
Average intensity over stand. dev. (<[<I> / σ(<I>)]>)	22.5 (4.6)
B-Factor (Wilson) (Å^2) / Average multiplicity	54.0 / 8.1 (8.2)
Resolution range used for refinement (Å)	∞ – 2.70
No. of reflections in working set / in test set	51,270 / 765
Crystallographic R_{factor} (free R_{factor}) [a]	0.217 (0.283)
No. of protein atoms / solvent molecules / ions	12,113 / 211 / 4 Zn^{2+}
ligands	4 glycerol, 4 N-acetyl-glucosamine, 4 GEMSA
Rmsd from target values	
bonds (Å) / angles (°)	0.006 / 1.028
bonded B-factors (Å^2)	8.49
Average B-factors for protein atoms (Å^2)	55.2
Main-chain conformational angle analysis [b]	
Residues in favored regions / outliers / all residues	1,460 / 4 / 1,523

Values in parentheses refer to the outermost resolution shell.

[a] For definitions, see Table 1 in (Bayés et al., 2007).

[b] According to MOLPROBITY (Davis et al., 2007).

Miscellaneous

Figures were prepared with programs SETOR (Evans, 1993) and TURBO-Frodo. Interface analysis was performed with the PISA server at www.ebi.ac.uk. Model validation was performed with MOLPROBITY (Davis et al., 2007) and the WHATCHECK routine of program WHATIF (Vriend, 1990). The model of variant svr-PE was constructed with TURBO-Frodo based on the structure of DmCPD1Bs and a sequence alignment calculated with CLUSTAL-W (Thompson et al., 1994), which was modified to

include structural restraints. The dimer interface was calculated as half of the total surface area buried by the complex. The final coordinates of DmCPD1Bs are available from the PDB at www.pdb.org (access code 3MN8).

1.3 RESULTS AND DISCUSSION

Production and proteolytic susceptibility of DmCPD1Bs

Initial heterologous recombinant overexpression trials of DmCPD1Bs in the methylotrophic yeast, *Pichia pastoris*, rendered sufficient protein for structural studies. However, although we assayed several purification strategies, the presence of several bands in SDS-PAGE could not be avoided. N-terminal sequencing of these bands revealed they were cleavage products of the target protein occurring after Leu189 that co-purified with the intact enzyme. Subsequent assays with different broad-spectrum inhibitors against the major classes of endopeptidases revealed that addition of phenylmethanesulfonyl fluoride (PMSF) during protein expression significantly limited degradation, giving rise to mainly two major bands (Fig. 3). In contrast, inhibitors of metallo-, cysteine, and aspartyl proteinases had no appreciable effect (data not shown). As a control, intact protein obtained and purified in the presence of PMSF was assayed for susceptibility to proteolysis by commercial trypsin and subtilisin. These experiments revealed similar degradation patterns to those observed during production in the absence of PMSF. In the case of subtilisin, cleavage occurred after Leu189 and for trypsin, cleavage occurred after Arg190 (Figs. 2 and 3). According to the provisional genome sequence (see http://bioinformatics.psb.ugent.be/webtools/bogas and (De Schutter et al., 2009)), *Pichia pastoris* apparently encodes four soluble serine endopeptidases: a putative Kex2 subtilisin-like proprotein convertase; a sequence relative of the mammalian Omi/HtrA2 family of trypsin-like serine proteases; and two putative orthologs of *Saccharomyces cerevisiae* YscB *alias* cerevisin (UniProt P09232). The latter enzyme belongs to the subtilisin/kexin family of serine peptidases and preferentially cleaves after arginine, tyrosine, or leucine according to MEROPS database (Rawlings et al., 2008). Therefore, either of the cerevisin orthologs in *P. pastoris* may be responsible for the cleavage after Leu189 observed during recombinant overexpression of DmCPD1Bs.

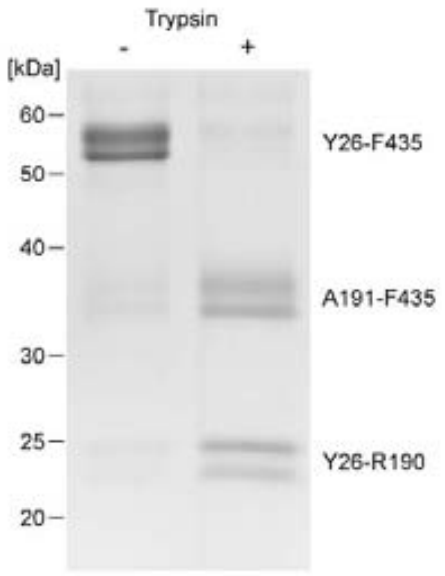

Fig. 3. Coomassie-stained SDS-PAGE analysis of DmCPD1Bs cleavage. Mature DmCPD1Bs (residues Tyr26-Phe435) reveals two main bands at ~52-58 kDa, which indicates heterogeneity in the glycosylation pattern. Incubation with trypsin produced a single cleavage within the regulatory loop at Arg190 and gave rise to two cleavage-product pairs of ~35 and ~23 kDa. The cleavage products are already present in the non-treated enzyme due to incomplete inhibition of the *Pichia pastoris* serine protease during protein expression and purification.

The two variant bands obtained corresponded to species of between 52 and 58 kDa and did thus not coincide with the theoretical molecular mass of the protein (47 kDa). Treatment with endoglycosidase F and N-terminal sequencing revealed they were glycosylation variants of the intact target protein (Fig. 3). There are two potential glycosylation sites in the enzyme at Asn133 and Asn269. We tried to reduce heterogeneity by constructing two mutants, N133Q and N269Q. However, the expression yield of both single mutants was dramatically lower than that of the wild type. In addition, both mutants still displayed heterogeneous glycosylation (data not shown). The double mutant N133Q/N269Q could not be expressed. These results underpinned the relevance of sugars for this protein, in good agreement with the importance of protein glycosylation in the secretory pathway (Ungar, 2009).

Intact DmCPD1Bs has proteolytic activity

The catalytic efficiency of the purified intact protein against the standard chromogenic type-B CP substrate, hippuryl-L-Arg (Table 2), was comparable to that of human CPE, human CPM, and human and bovine CPU (also known as activated thrombin-activatable fibrinolysis inhibitor, TAFIa). This is consistent with previous reports on DmCPD1Bs produced in baculovirus, which was also active (Sidyelyeva et al., 2006; Sidyelyeva and Fricker, 2002). Reported values for human CPN evince slightly lower activity for this substrate due to its preference for C-terminal lysine residues (Table 2). We further produced and purified human CPB to provide a standard for an M14A MCP for comparison and it had about ten-fold higher efficiency. This is in accordance with previous studies, which showed that "digestive" MCPs are more efficient CPs than "regulatory" forms against small substrates (Arolas et al., 2007; McKay et al., 1979). Moreover, intact DmCPD1Bs previously incubated with commercial subtilisin or trypsin (see above) had only 30 % activity, which is attributed to traces of intact enzyme that were clearly visible in a Coomassie-stained SDS-PAGE. We conclude that cleavage compromises the proteolytic activity of DmCPD1Bs.

Table 2. Kinetic constants for hydrolysis of hippuryl-L-Arg by different carboxypeptidases

Carboxypeptidase	K_{cat} (s^{-1})	K_m (μM)	K_{cat}/K_m (s^{-1} mM^{-1})
DmCPD1Bs	38.8±4.1	1112±167	34.9±8.9
hCPB	253.7±11.9	485.3±62.5	522.8±89.8
hTAFIa [a]	73.1±5.7	3350±1050	21.8±8.6
bTAFIa [a]	219.8±26.7	7310±2710	30.1±14.8
hCPE [b]	8.5	400	21.3
hCPN [c]	4	650	6.2
hCPM [d]	6.1±0.1	290±20	21±1.0

Data taken from [a] (Valnickova et al., 2009), [b] (Fricker and Devi, 1990), [c] (McKay et al., 1979), and [d] (Deiteren et al., 2007). h, human; b, bovine.

Inhibition assays against DmCPD1B

DmCPD1Bs is inhibited by the small-molecule inhibitor 2-guanidinoethyl-mercaptosuccinic acid (GEMSA) with an apparent inhibition constant (K_i) value of 0.67±0.08 μM. This fits in the range reported for other MCPs of B-type specificity, which show K_i values in the order of mM to nM for this inhibitor (Arolas et al., 2007). In contrast, DmCPD1Bs is not inhibited by protein inhibitors from potato, leech, tick, or mammal, even at micromolar concentrations. These inhibitors display high affinity for M14A MCPs, with low nanomolar inhibition constants, but they are inert against M14B-type MCPs (Arolas et al., 2005; Gomis-Rüth, 2008).

Structure of DmCPD1Bs in complex with GEMSA

DmCPD1Bs was crystallized in the presence of GEMSA, which significantly improved the crystal diffraction possibly by providing rigidity, as found in the case of TAFI (Marx et al., 2008). Overall, the structure subdivides into an N-terminal catalytic MCP/funnelin domain (Ile28-His336) and a C-terminal TLD (Ile337-Glu420). The catalytic domain (CD) has a compact globular shape, which resembles the volume obtained when a cone is extracted from a sphere (Fig. 4). There is a funnel-like opening to the left, within which lies the active-site cleft, at the base of the opening. The cleft is rather shallow, which is compatible with the capacity of MCPs to cleave a large variety of substrates, as only few contacts between the enzyme and the C-terminal stretch of a substrate are required to fix the latter to the active site. The DmCPD1Bs CD evinces an α_{8+2}/β_8-topology conforming to an α/β-hydrolase or PLEES fold (Ollis et al., 1992; Puente and López-Otín, 1997) (Fig. 4). It contains a central doubly-wound eight-stranded β-sheet (β1- β8), which is strongly twisted and spans the molecule vertically bottom to top. The sheet has a mixed parallel/antiparallel topology and a strand connectivity +1,+2,-1x,-2x,-2,+1x,-2, and its core consists of four parallel coplanar central strands (β3- β5 and β8). The catalytic site is located at the C-terminal end of these strands, as observed in other MCPs (Gomis-Rüth, 2008). These four strands are flanked by two parallel strands at the top (β6-β7) and a β-ribbon (β1- β2) at the bottom (Fig. 4). The sheet curvature leads to a concave front side, which shelters

helices α5, α6, α8 and α9, as well as the active-site cleft. At the back of the molecule, the convex side of the sheet accommodates the surface N- and C-termini of the molecule, as well as helices α1- α4, α7, α8, and the C-terminal helix α10 of the CD, which runs left to right across the molecule along the convex face of the β-sheet and is bent by ~40°. The funnel-like access to the active site is shaped by irregular segments of disparate length: the loop connecting strand β3 with helix α2 (Lβ3α2), Lβ5β6, Lβ6α7, Lα7β7, and, in particular, the 55-residue segment connecting strand β4 with helix α6 (segment β4α6 hereafter). This element contributes to the lower front of the molecule, shapes the funnel rim, and includes two short helices, α4 and α5. Between five and nine residues are missing in each of the four molecules present in the asymmetric unit within the surface-located Lα5α6 (hereafter referred to as "regulatory loop") within segment β4α6. Two disulfide bonds also contribute to the funnel rim by cross-linking loop Lα4α5 with Lβ8α10 (Cys156-Cys309) and Lα7β7 with Lβ8α10 (Cys268-Cys308). In addition, a conspicuous disulfide bond between neighboring residues is found within Lβ6α7 (Cys236-Cys237; see Fig. 4).

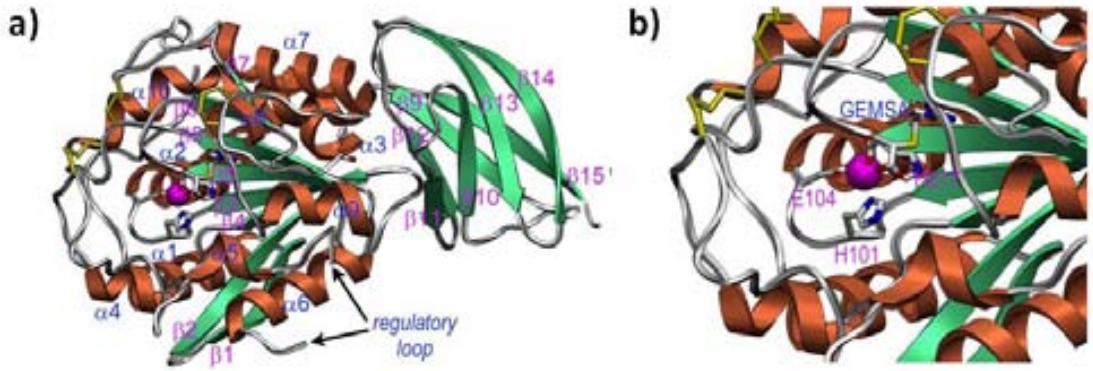

Fig. 4. Structure of DmCPD1Bs. a) Richardson plot of DmCPB1Bs in standard orientation (Gomis-Rüth, 2008) showing the catalytic domain (left) and the transthyretin-like domain (right). The repetitive secondary structure elements are shown as green arrows (strands β1-β8 in the catalytic domain and β9-β15 in the transthyretin-like domain) and coral ribbons (helices α1-α10) and labeled. The catalytic zinc ion is shown as a magenta sphere. The position of the regulatory loop is pinpointed by arrows. b) Close-up view of a) displaying the active-site cleft and the funnel-like border that gives access to it. The zinc-binding residues are presented as sticks and labeled, as is the bound GEMSA molecule.

The catalytic zinc ion resides at the bottom of the active-site cleft and is coordinated by the $N^{\delta 1}$ atoms of His101 and His217 and, bidentately, by Glu104 atoms

$O^{\varepsilon 2}$ and $O^{\varepsilon 1}$ (Fig. 4b). These residues are provided by Lβ3α2 and by the end of strand β5. The protein was co-crystallized in the presence of the small-molecule inhibitor GEMSA (Fig. 5), which mimics a product complex after substrate cleavage in which the C-terminal residue is still trapped in the S1' pocket. The carboxymethylene group of the inhibitor binds the catalytic zinc ion asymmetrically through two carboxylate oxygens, thus giving rise to an overall six-fold metal coordination. The other carboxylate group of GEMSA is anchored to Arg165 $N^{\eta 2}$, Asn174 $N^{\delta 2}$, and, bidentately, to Arg175 $N^{\eta 1}$ and $N^{\eta 2}$. The guanidinoethylmercapto group, which imitates an arginine side chain, occupies the S1' pocket and is surrounded by Ser224, Gly279, Trp282, Tyr283, Leu285, Gln290, and Thr303. The characteristic specificity for basic residues of M14B-type MCPs is provided in DmCPD1Bs by the side chain of Asp228, which binds one of the terminal guanidine nitrogen atoms of GEMSA. The role of the general acid/base, essential for catalysis in MCPs and other metal-dependent peptidases (Bayés et al., 2007; Gomis-Rüth, 2008), is fulfilled here by Glu305. The residues that give rise to the characteristic MCP motif, HXXE+R+NR+H+Y+E (Gomis-Rüth, 2008), are found in DmCPD1Bs at positions 101, 104, 165, 174, 175, 217, 283, and 305, respectively.

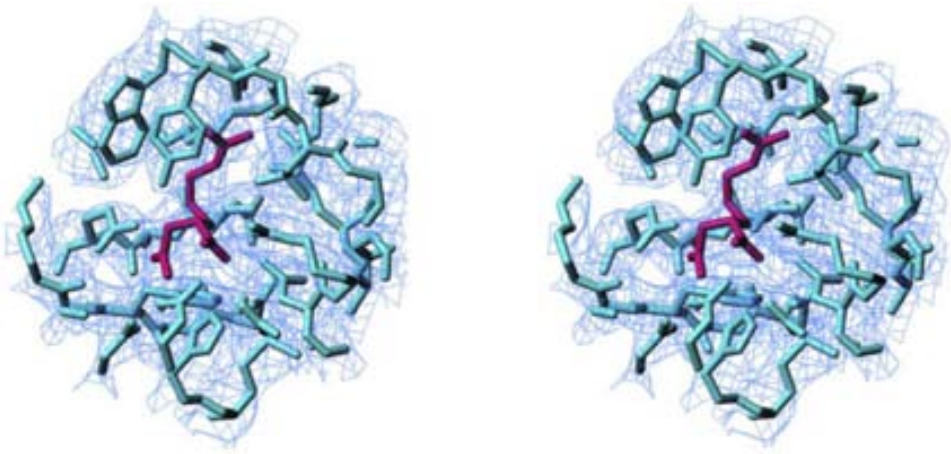

Fig. 5. Presence of GEMSA in the structure of DmCPD1Bs. Initial σ_A-weighted ($2mF_{obs}-DF_{calc}$)-type electron density map centered on the GEMSA-binding site. This map was calculated without the inhibitor molecule (omit map) and is shown contoured at 0.75 σ. The final refined model is superimposed (GEMSA, magenta sticks; DmCPD1Bs, cyan sticks).

The C-terminal TLD spans 84 residues in DmCPD1Bs and is rod-shaped (Fig. 4a). Its N- and C-termini are on opposite sides of the rod, which folds into an all-β seven-

stranded β-barrel or β-sandwich, with two layers of three mixed (β1, β4, and β7) and four antiparallel strands (β2, β3, β5, and β6), respectively, which are glued by a hydrophobic core. These strands are arranged as two subsequent Greek-key-like elements related by a two-fold axis perpendicular to the sandwich surface (Gomis-Rüth et al., 1999; Keil et al., 2007; Reverter et al., 2004). This domain has topological similarity with transthyretin, a conserved plasma protein that tends to aggregate, thus giving rise to several forms of amyloidosis (Prapunpoj and Leelawatwattana, 2009). However, transthyretin contains an eighth C-terminal β-strand that is absent in MCPs (Gomis-Rüth et al., 1999). The TLD is laterally attached to the CD and interacts with Lβ2β3, Lα6β5, Lα9β8, and Lβ6α7 of the latter through hydrophobic residues. It also interacts through a (bidentate) salt bridge formed by an aspartate (Asp244) at the beginning of helix α7 of the CD, which is conserved among M14B MCPs, and an arginine (Arg376) within the TLD. As in the case of related M14B MCPs, the function of TLD is uncertain. It may assist in folding, regulation of enzyme activity or protein-protein interactions (Arolas et al., 2005; Gomis-Rüth, 2008; Gomis-Rüth et al., 1999; Reznik and Fricker, 2001).

Hypothetical quaternary arrangement

Dynamic light-scattering and cross-linking experiments using glutaraldehyde showed that protein concentrations below 1 mg/mL rendered a monomer, whereas concentrations above 8 mg/mL gave rise to results that are compatible with a monomer-dimer equilibrium. Size-exclusion chromatography indicated that the protein is monomeric throughout the concentration range assayed (0.5 to 18 mg/mL, as injected into the FPLC system). In the crystals, which resulted from highly concentrated protein solution (18 mg/mL), two pairs of dimers were found in the asymmetric unit (Fig. 6). Inspection of the dimeric interface in the structure reveals that it measures ~950 $Å^2$, which is larger than the range described for experimentally validated “weak” transient homodimeric proteins (740±140 $Å^2$; (Nooren and Thornton, 2003)). Such proteins form both monomers and dimers at physiological concentration. In DmCPD1Bs, the dimer is symmetric, so that the same structural segments of each protomer are involved in complex formation: helix α6, strands β1 and β2, and loops

Lα5α6 and Lβ2β3 of each CD; and loops Lβ11β12 and Lβ9β10 of each TLD. Two symmetric bidentate salt-bridges (Asp353-Arg85) and a total of nine hydrogen bonds are observed at the interface. In addition, the two active-site clefts are located on opposite surfaces of the dimer, i.e. there would be no steric hindrance for substrate binding (Fig. 6a). However, there is no concluding evidence that the enzyme is a functional dimer *in vivo* since the protein seems to be monomeric in solution and the dimer could be regarded as an artefact of the crystallization process. Moreover, we do not know whether the formation of a dimer could have functional implications. To date, no higher oligomerization states than a monomer have been reported for MCPs. Only CPN is believed to function *in vivo* as a tetramer comprising two M14B MCP subunits linked to two non-catalytic regulatory domains (Skidgel and Erdös, 2007), but it is not known how these subunits are arranged within the heterotetramer.

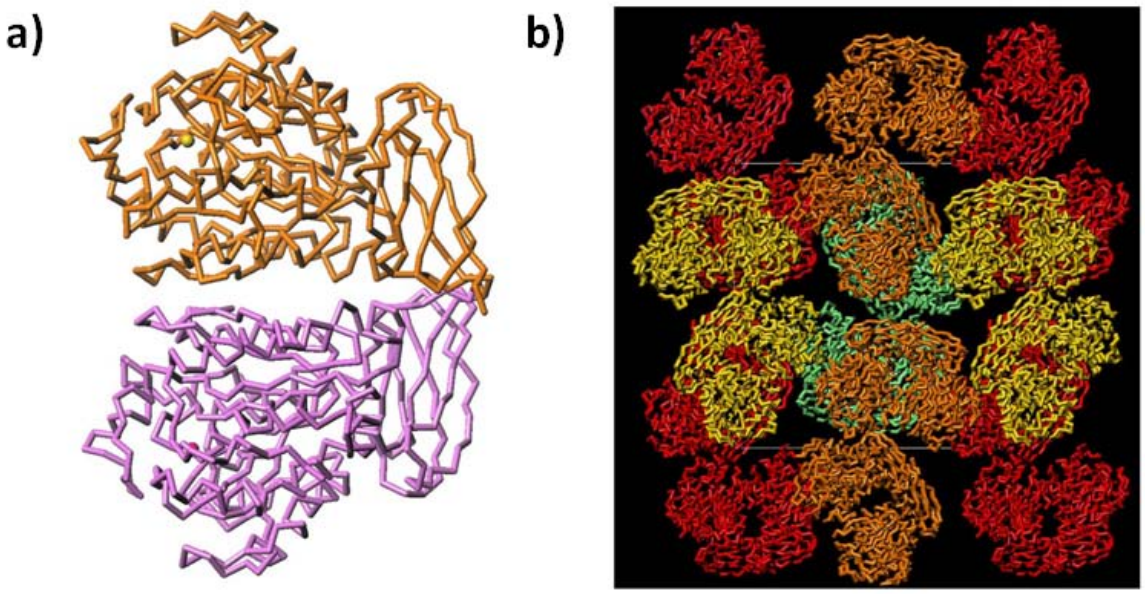

Fig 6. Quaternary structure arrangement. a) The quaternary structure of DmCPD1Bs in the crystal is a homodimer. b) The crystal is made up of such dimers, shown in red, orange, yellow and cyan. Two dimers are found in the crystallographic asymmetric unit. The unit-cell box is shown for reference.

Implications for the other splicing variants of the silver gene

The splicing variants of *silver* give rise to two types of N-terminal repeat (Fig. 1). These differ in the first 152 (encoded by exon 1B in DmCPD1Bs, svr-PB, and svr-PG) or 150 (encoded by exon 1A in svr-PD, svr-PE, and svr-PH) residues, which share ~39 %

sequence identity (Sidyelyeva and Fricker, 2002). These stretches give rise to the lower half of the CD until the middle of segment β4α6. The second half of the first repeat until the end at Glu420 (numbering hereafter is according to DmCPD1Bs if not otherwise stated) is identical in all proteins. Svr-PC and svr-PI do not possess a ~150-residue N-terminal segment but have a short five-residue tail. As they lack a signal peptide, it is questionable whether they are secreted. They may play an intracellular role instead.

A model of variant svr-PE, which had been shown to be catalytically incompetent by Fricker and co-workers (Sidyelyeva et al., 2006; Sidyelyeva and Fricker, 2002), based on the coordinates of DmCPD1Bs and a structure-based sequence alignment, reveals that the differences in sequence are compatible with maintenance of the overall structure upon minor readjustment of the main chain. Most changes are conservative or are compensated for by differences in size of the side chains in neighboring positions. However, there is a slight trend towards smaller side chains at positions contributing to the central hydrophobic core of the MCP/funnelin domain, which might lower thermal stability: Leu78Ala, Tyr96Leu, Ile97Val, Ile80Ala, Met100Ile, Leu111Val, Leu119Ala, Leu128Val, Met143Cys, and Ser151Ala. In addition, the presence of a histidine replacing an asparagine at position 121 would entail a minor rearrangement of the N-terminal stretch up to Phe34. In turn, the conservation of the N-glycoslyation site motif around Asn133 (Asn-Leu-Thr in exon 1A instead of Asn-Ser-Thr in exon 1B) indicates that the aspargine is also attached to a sugar moiety. However, the presence of a glutamine instead of a glycine at position 129 may entail a rotation of the sugar moiety around the linking bond, $N^{\delta 2}$-C1, which provides space for the glutamine side chain. In addition, only a four-residue insertion is found within Lβ2β3, which should not disturb the interaction of the catalytic domain with the TLD (see above) due to the maintenance of leucine residues at positions 89 and 90.

The main difference between the exon 1A- and the exon 1B-encoded fragments is the substitution of glutamine (Gln99 in UniProt entry P42787 isoform 6) for the zinc-binding histidine at position 101 in DmCPD1Bs (Fig. 1). Although glutamine is a rare zinc ligand in proteins (Auld, 2001; McCall et al., 2000), in the case of MCPs it may lead to metal-containing but catalytically impaired or even inert zinc sites, as reported for

zinc-ligand mutants of human glyoxalase I (Ridderstrom et al., 1998). A glutamine at the position equivalent to residue 101 in DmCPD1Bs is also found in a series of (putative) proteins from related *Drosophila* species, namely *D. ananassae* (UniProt B3N0D7), *D. willistoni* (B4NPC2), *D. mojavensis* (B4L2T3), *D. sechellia* (B4IEW7), *D. grimshawi* (B4JX73), *D. erecta* (B3NWB3), and *D. virilis* (B4M2C3). These entries span between 1302 and 1495 residues and could represent three-repeat orthologs of the *silver* gene product, also encoded by an exon that gives rise to a non-functional MCP/funnelin domain. In addition to fruitfly species, a glutamine was found in a potential three-repeat ortholog of the starlet sea anemone *Nematostella vectensis* (A7S4K6). In all these cases, it remains to be ascertained whether the glutamine-containing MCP-like domains are functional.

The six long constructs all have identical sequences for the putative second and third repeats, the transmembrane helix, and the first 51 residues of the cytosolic domain (Fig. 1). This means, they all have the second catalytically competent and the third silent repeats. They differ only in the very last segments of the cytosolic domains: svr-PB stops after the common stretch; svr-PC and svr-PD have additional 22 residues; and svr-PG, svr-PH, and svr-PI have a further 55 residues, which are unrelated to the others. These differences allow differential trafficking of the protein between the *trans*-Golgi network and the cell surface for the six membrane-anchored forms (Sidyelyeva et al., 2006; Sidyelyeva and Fricker, 2002).

Comparison with duck CPD domain II

DmCPD1Bs shows 40 % sequence identity with the second repeat of the duck ortholog, which is the only other CPD structure reported (Aloy et al., 2001; Gomis-Rüth et al., 1999; Sidyelyeva and Fricker, 2002). This sequence similarity redounds to close structural similarity, as reflected by an *rmsd* of 0.97Å for the 368 C^{α}-atoms deviating less than 3Å (see Fig. 7a). Significant deviations are found, however, within Lβ2β3, Lβ6α7, Lα7β7, and the regulatory loop, Lα5α6 (Fig. 7b). These loops contribute to shaping the funnel rim that allows access to the active site (see above). This rim is responsible for interaction with large protein substrates in MCPs (Gomis-Rüth, 2008).

In addition, significant differences are found in two loops of the TLD: Lβ9β10 and, in particular, Lβ14β15, which is on the back surface of the molecule. This region might be important for function in M14B MCP TLDs (see above) and structural differences might lead to disparate binding partners in duck CPD domain II and DmCPD1Bs. Overall, the most noteworthy difference is found in the regulatory loop, which is partially undefined in DmCPD1Bs (see above). As the protein found in the crystals is intact, this flexibility does not arise from cleavage but from intrinsic disorder, which, in turn, might have functional implications (see below).

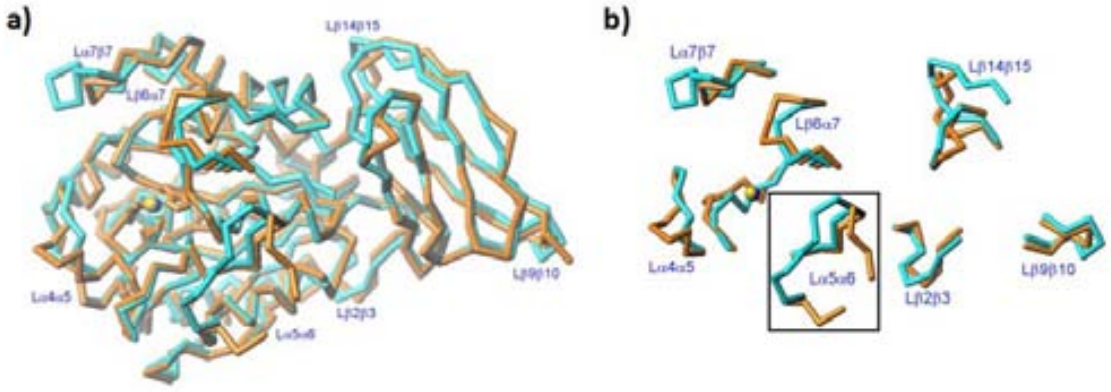

Fig. 7. Structural comparison of DmCPD1Bs. a) Superimposition of the C^{α}-traces of DmCPD1Bs (orange; yellow zinc ion) and duck CDP domain II (PDB 1QMU; cyan; blue zinc ion). b) Detail of a) showing only those loop segments that evince major conformational differences between the two structures. The regulatory loop (Lα5α6) is framed.

Proposed mechanisms of regulation

M14A MCPs are secreted as zymogens with an N-terminal 90-95-residue pro-domain in charge of latency maintenance until the environmental and temporal conditions for activity are given. This ensures that no undesired proteolytic activity occurs in the transit between the locus of biosynthesis and the final destination (Arolas et al., 2007; Avilés et al., 1993; Gomis-Rüth, 2008). In contrast, M14B MCPs are not secreted as latent zymogens, and the regulation of the catalytically active forms is believed to rely on their localization: most of these enzymes are bound to membrane or to the extracellular matrix (e.g., long CPD forms, CPM, and CPZ). However, soluble variants such as DmCPD1Bs may be regulated by other mechanisms.

CPD is palmitoylated at cysteine residues of its transmembrane domain in the duck, with implications for its localization and biological function (Kalinina and Fricker, 2003). Covalent attachment of fatty acids enhances the hydrophobicity of proteins and contributes to their membrane association, thus providing a means of regulation through localization. The structure of DmCPD1Bs has several exposed cysteine residues at the protein surface shaping the funnel rim, which could be potentially palmitoylated. In particular, adjacent residues Cys136 and Cys137, which were disulfide-linked in the crystal structure (see above), could be present as free thiols in physiological conditions and become palmitolylated. This could be a way to anchor DmCPD1Bs to membranes and thus regulate its intracellular trafficking.

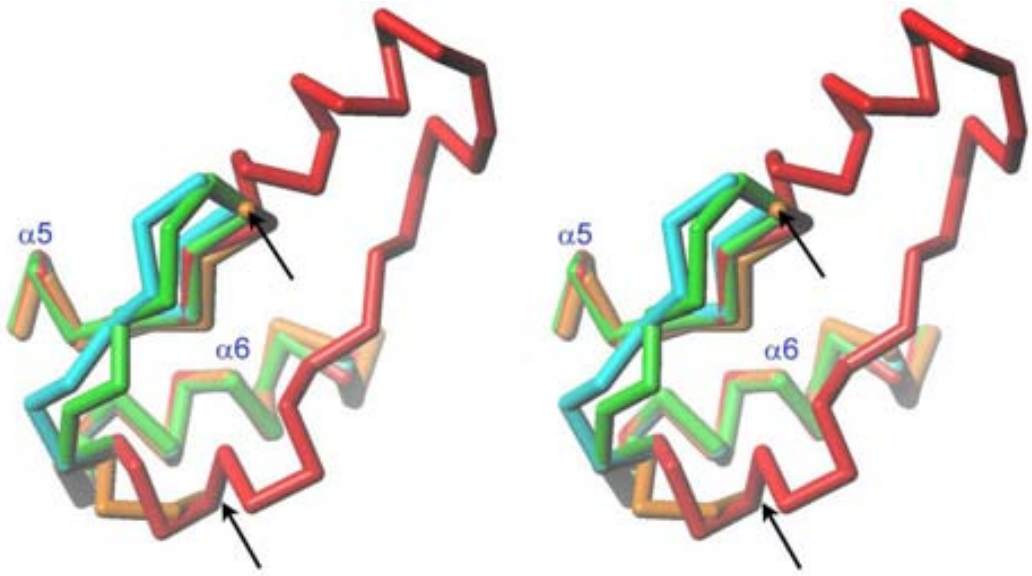

Fig. 8. Regulatory loop of DmCPD1Bs. Superimposition of the C^{α}-traces of DmCPD1Bs (orange); duck CDP domain II (PDB 1QMU) displayed in cyan (Aloy et al., 2001; Gomis-Rüth et al., 1999); human CPM (PDB 1UWY) displayed in green (Reverter et al., 2004); and human CPN (PDB 2NSM) displayed in red (Keil et al., 2007) corresponding to segments equivalent to α5, Lα5α6, and α6 of the fruitfly enzyme. The arrows point to the residues that flank the flexible region of the regulatory loop in DmCPD1Bs: up to nine residues are missing in each of the protomers.

Another, more likely mechanism of regulation could be limited proteolysis within the regulatory loop. Intrinsically flexible regions on the molecular surface of a protein directly correlate with instability, conformational changes and lability, and proteolytic susceptibility (Zappacosta et al., 1996), in particular in a protease-rich medium like that of the secretory pathway. This would explain why the regulatory loop is a hotspot for

cleavage by host proteases during recombinant heterologous overexpression (see above). The equivalent regions of all other M14B MCPs thus far structurally analyzed, i.e. duck CPD domain II, human CPN and human CPM, are well ordered and rigid, and they significantly deviate from the conformation found in DmCPD1Bs, thus pointing to potential differences in the function of this loop (Fig. 8). Single cleavage at Leu189 or Arg190 was found to inactivate the protein, possibly due to disruption of the regulatory loop itself and the preceding structural elements α5 and Lα4α5. Therefore, a similar cleavage *in vivo* by peptidases along the secretory pathway would provide a means of selective regulation of DmCPD1Bs activity.

Final considerations

In contrast to mammals, the fruitfly has only two regulatory/M14B MCPs, one CPD and one CPM ortholog, which is consistent with the fact that *Drosophila* contains fewer genes. Hence, the presence of eight splicing variants of CPD, which give rise to both single-repeat soluble and two- and three-repeat membrane-anchored forms, may be conceived as a strategy to compensate for this shortage (Sidyelyeva et al., 2006). This is reminiscent of the finding that only two matrix metalloproteinase paralogs are found in the fruitfly, while 23 have been described in humans (Llano et al., 2002). Of the five active M14B MCPs in mammals, CPN is secreted into plasma and forms a dimer of heterodimers. CPM is membrane-anchored and functions extracellularly. CPZ is partially targeted to the extracellular matrix and has a frizz9led-like cysteine-rich domain that may make it a Wnt-binding protein (Reznik and Fricker, 2001). DmCPD1Bs could have a homologous function to CPE, which is likewise a single-repeat MCP and is found in secretory vesicles (Reznik and Fricker, 2001). The fruitfly enzyme is active and cleaves synthetic substrates with similar efficiency to human CPE, although the former has a different pH optimum (Sidyelyeva and Fricker, 2002). DmCPD1Bs could function as a CPE-like enzyme in mildly acidic regions of the secretory pathway like the trans-Golgi network. However, further studies are required to validate this hypothesis. Finally, the structure of DmCPD1Bs conforms to the overall fold of N/E-type funnelins, as may be the case for CPE, but evinces two unique features that may correlate with its regulation: (i) two contiguous surface-located cysteines that may become

palmitoylated, thus targeting the enzyme to membranes; and (ii) a surface hotspot for proteolytic inactivation. In this way, DmCPD1Bs could participate in *Drosophila* development through proteolytic processing of substrates previously targeted by furin-like proprotein convertases and would itself be inactivated at later stages by limited proteolysis.

Chapter 2

Human carboxypeptidase A4:

Characterization of the substrate specificity and implications for a role in extracellular peptide processing

2.1. INTRODUCTION

Carboxypeptidases (CPs) hydrolyze a single amino acid from the C-terminus of peptides and proteins. MetalloCPs use a catalytic mechanism in which nucleophilic attack on a peptide bond is mediated by a Zn^{2+}-activated water molecule (Gomis-Ruth, 2008). Most metalloCPs are classified based on amino acid sequence, structure and function into subfamilies within clan MC, Family M14 (Rawlings et al., 2006). These subfamilies include; M14A, also known as the CPA/CPB subfamily; M14B, also known as the CPN/CPE subfamily; M14C, that includes gamma-D-glutamyl-(L)-meso-diaminopimelate peptidase I from Bacillus sphaericus, and a proposed fourth subfamily including cytosolic CPs (CCPs) (Kalinina et al., 2007; Rodriguez de la Vega et al., 2007). MetalloCPs have also been divided into subgroups based on substrate specificity: A-like enzymes with preference for hydrophobic C-terminal amino acids; B-like enzymes which cleave C-terminal basic residues; CPs with preference for glutamate in the C-terminal position; and CPs with a broad substrate specificity.

Carboxypeptidase A4 (CPA4) belongs to the M14A subfamily of carboxypeptidases. The pancreatic members of this subfamily (CPA1, CPA2 and CPB) act in the degradation of dietary proteins in the digestive tract. Other members of this subfamily display a wide spectrum of physiological roles. Mast cell carboxypeptidase (recently renamed CPA3) is found in the secretory granules of mast cells and plays a role in defense (Pejler et al., 2007). CPA6 is linked to Duane syndrome, a congenital eye-movement disorder, and has been proposed to participate in the regulation of neuropeptides in the extracellular matrix within the olfactory bulb and other parts of the brain (Lyons et al., 2008). CPU, also known as CPB2 and thrombin-activatable fibrinolysis inhibitor (TAFI), participates in fibrinolysis inhibition and possesses anti-inflammatory properties (Leung et al., 2008), and for these reasons is of interest to the pharmaceutical industry (Arolas et al., 2007).

CPA4 was originally referred to as CPA3 (Huang et al., 1999), but was renamed CPA4 to reflect the order in which the CPAs were discovered. Huang et al identified CPA4 in a search for mRNAs induced by sodium butyrate in androgen-independent prostate cancer cells (Huang et al., 1999). The histone deacetylase inhibitor

trichostatin A also induced the expression of CPA4 mRNA in PC-3, DU145 and BPH1 human prostate cancer cell lines (Huang et al., 1999). Huang *et al.* also reported that CPA4 mRNA expression is associated with hormone-regulated tissues, suggesting that it may have a role in cell growth and differentiation. The human CPA4 gene is located on chromosome 7q32, which is a region in the genome that might contain genes for prostate cancer aggressiveness (Witte et al., 2000). In addition, the CPA4 gene was found to be maternally imprinted in a tissue-specific manner (Bentley et al., 2003; Kayashima et al., 2003). Furthermore, imprinting in adult benign hypertrophic prostate tissue suggests that mutations or aberrant imprinting in CPA4 may be related to prostate cancer aggressiveness (Kayashima et al., 2003). Ross *et al.* studied the association of single-nucleotide polymorphisms (SNPs) on the CPA4 gene and its relation to prostate cancer and found that the nonsynonymous coding SNP (Gly303Cys) rs2171492 was associated with an increased risk of aggressive disease in younger men (Ross et al., 2009).

Despite the potential importance of CPA4 towards prostate and other cancers, no previous studies have examined the substrate specificity of CPA4 in detail. The three-dimensional structure of CPA4 has been determined for the zymogen state (Garcia-Castellanos et al., 2005) and for the active form in complex with latexin, a brain-derived CP inhibitor (Pallares et al., 2005), and in complex with a hexapeptide (Bayes et al., 2007). This latter study also compared the cleavage of a handful of synthetic peptides by CPA4, but no kinetic details were provided. To gain a better understanding of the enzymatic properties of CPA4, we used a variety of approaches to characterize the substrate specificity. A series of di-peptide chromogenic substrates were synthesized and tested with purified CPA4. A number of synthetic peptides were individually tested with purified CPA4; many of these peptides correspond to biologically active peptides. A third approach involved incubating CPA4 with a mixture of over 100 endogenous peptides extracted from mouse brain, using a quantitative peptidomics approach to identify substrates and products. The subcellular distribution, secretion, and pH optimum of CPA4 were also examined. Taken together, these studies provide a complementary and thorough analysis of the cleavage specificity of CPA4

and suggest that this enzyme functions as a soluble extracellular peptidase that removes C-terminal hydrophobic/aliphatic residues from secreted peptides.

2.2. EXPERIMENTAL SECTION

Protein Production and Purification.

Human proCPA4 (PCPA4) was produced using the vector pPIC9 and the methylotrophic yeast *Pichia pastoris* as an expression host and purified as described elsewhere (Pallares et al., 2005). The active enzyme was obtained through tryptic activation (at a 1/10 w/w ratio) for 60 min at room temperature, and the resulting product was subsequently purified by anion-exchange chromatography (TSK-DEAE 5PW) using a FPLC-Äkta system with a linear salt gradient from 0 to 30% of 0.4 M ammonium acetate in 20 mM Tris-HCl (pH 10.5). Eluted fractions were analyzed by SDS−PAGE, and the purest samples containing the active enzyme were pooled, desalted, and concentrated to 1 mg/mL by Amicon centrifugal filter devices.

Cell culture and Transfection

HeLa (human cervix adenocarcinoma), PC-3, DU145 and LNCaP (all three cell lines are human prostate carcinoma), 1BR3.G (human skin fibroblasts) and HEK 293T (human embryo kidney) cells were purchased from the American Type Culture Collection (Manassas, VA), and were cultured in the recommended growth medium: Eagle's Minimum Essential Medium or Dulbecco's Modified Eagle's Medium with 4500 mg/L glucose, GlutaMAX–I and pyruvate supplemented with 10% fetal calf serum (Invitrogen, Carlsbad, CA). The full-length cDNA sequence of human proCPA4 was cloned into pcDNA3.1 vector (Invitrogen) and the hemagglutinin (HA) epitope was introduced on the C-terminus of CPA4. The plasmid pDsRed2-ER from Clontech was used for endoplasmic reticulum (ER) localization. Cells were grown in 35 mm dishes and transfected with Lipofectamine 2000 according to the manufacturer's instructions.

Cell fractionation and processing

Forty eight hours after transfection, medium was removed, centrifuged and collected on ice. The dishes were gently washed with phosphate-buffered saline (PBS), and cells were removed by pipetting up and down in a mild extraction buffer consisting

of 50 mM Tris-HCl pH 8.0. The cell suspension was sonicated and centrifuged at 16,000 g for 10 min at 4°C. The pellet was resuspended in an equivalent amount of 50 mM Tris-HCl pH 8.0, containing 500 mM NaCl and 1% Nonidet P-40 and centrifuged as above to produce a detergent/high salt extract. Following the removal of cells, the plate was washed several times with PBS and the extracellular matrix (ECM) was extracted by adding hot SDS-PAGE sample buffer directly to the plate. All homogenates were frozen and stored at −70°C until analysis. For trypsin activation assays, 200 ng of trypsin (Sigma) were added to 20 µl of the extracellular medium of transfected HEK 293T cells and incubated for 30 min at 37°C.

Immunoblot

For immunoblotting, equivalent amounts of each lysate or fraction were analyzed by SDS/10% PAGE, and the separated proteins were electrophoretically transferred onto polyvinylidene difluoride (PVDF) filters (Millipore Corp.). Non-specific binding sites on the PVDF filters were blocked by incubation with 5% skim milk for 12 h. The PVDF filters were then incubated with anti-HA antibody (dilution 1:3,000; Sigma-Aldrich) for 1 h at 25°C. After washing with PBS containing 0.05% Tween 20, the filters were incubated with goat anti-mouse IgG-peroxidase (dilution 1:10,000; Pierce) for 1 h. After three rinses, immunoreactive bands were visualized with a chemiluminescence detection kit (ECL; Millipore).

Immunofluorescence

HeLa, HEK 293T, PC-3, DU145 or LNCaP cells were co-transfected with HA-tagged PCPA4 and pDsRed2-ER using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. After 48 h incubation cells were fixed in methanol, permeabilized with 0.1% Triton X-100, and incubated with anti-HA antibody (clone HA-7, Sigma, 1:2,000 dilution), followed by conjugated Alexa-488 anti-mouse-IgG (Molecular Probes, 1:500 dilution) and DAPI (4',6'-diamidino-2-phenylidole). Coverslips were mounted with Fluoprep mounting medium (Biomerieux), and analyzed with a confocal laser scanning microscope (Leica Microsystems, Bannockburn, IL).

Substrate synthesis

Chromogenic substrates were synthesized according to Blumberg et al (Blumberg and Vallee, 1975). 3-(2-furyl)acryloyl-Phe-N-hydroxysuccinimide (FA-Phe-Osu) was prepared according to the following procedure: a solution of 15 mmol FA-Phe and 15 mmol N-hydroxysuccinimide in 60 ml of dioxane-DMF, 1:1, was cooled to 4ºC and 15 mmol dicyclohexylcarbodiimide was added. The solution was stirred at 4ºC for 16 hr. The dicyclohexyl urea was filtered and the solution evaporated in vacuo. Furylacryloyl dipeptides were prepared by reacting FA-Phe-Osu with the appropriate amino acid according to the following procedure: the N-hydroxysuccinimide ester (5 mmol) in 25 ml of dioxane was added at room temperature to a solution of the amino acid (6 mmol) and sodium bicarbonate (12 mmol) in 25 ml of water. The mixture was kept at room temperature for 6 hr, the dioxane evaporated, and the residue acidified with 1N HCI. After the mixture was cooled overnight at 4ºC, the precipitate was collected by filtration and dried.

Kinetic measurements

The rate of hydrolysis of different substrates was continuously measured at 25ºC in 50 mM Tris 0.15 M NaCl at the optimal pH for each enzyme. The wavelengths varied slightly among substrates: 336 nm for FA-Phe-Phe, FA-Phe-Ile and FA-Phe-Trp; 338 nm for FA-Phe-Ala, FA-Phe-His and FA-Phe-Leu; and 340 nm for FA-Phe-Val and FA-Phe-Met. Initial rates determined from the first 5–10% of the time-trace of each reaction were obtained at substrate concentrations close to the K_m value whenever possible. The kinetic parameters, k_{cat} and K_m, were obtained using 6 experimental points by direct fit to a Michaelis–Menten curve using a nonlinear least-squares regression analysis. For those substrates for which the solubility of the substrates limited the measurements to concentrations well below K_m, the k_{cat}/K_m values were estimated from the slope of the linear portion of saturation curve using the simplified equation $v=(k_{cat}/K_m).[E_0].[S]$.

Measurement of equilibrium dissociation constants (K_i)

The method for reversible tight-binding inhibitors described by Bieth (Bieth, 1995) was used for K_i determination. Recombinant purified CPA4 at a concentration of 8 nM was assayed against increasing concentrations of inhibitor. Initial rates, determined from the first 5–10% of the time-trace of each reaction, were obtained for each inhibitor concentration using the N-(4-Methoxyphenylazoformyl)-L-phenylalanine (AAFP) substrate (Bachem). K_i values were obtained by direct fitting of fraction velocity versus inhibitor concentration data to the Morrison equation using the Grafit program.

Substrate specificity and kinetic measurements using peptides

CPA4 substrate specificity was analyzed using synthetic peptide substrates. Reactions were performed in 50 mM Tris-HCl 100 mM NaCl pH 7.5 at 37ºC using a peptide and enzyme concentrations of 10 µmol and 1 nmol, respectively. Peptide cleavage reactions were stopped at different times (30 minutes, 1h, 2h and 16h) by dropping the pH to 4 using 1% TFA. The reaction products were desalted using a C18 ZipTip (Millipore) following the manufacturer's protocol. The peptides were eluted with 1 µl of 70% acetonitrile, 0.1% trifluoroacetic acid, mixed with an equal volume of a saturated solution of α-hydroxycinnamic acid in the same solvent, and analyzed in a Bruker Daltonics Ultraflex matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometer.

For the determination of kinetic constants for the hydrolysis of peptides, rates of substrate hydrolysis were determined by reversed-phase high performance liquid chromatography (RP-HPLC) in a Waters Alliance apparatus using a Grace Vydac 218TP^{TM} C18 column (3 µm, 100 Å, 4.6 mm i.d. × 250 mm, Hesperia, CA) with a linear gradient of 5 to 35% acetonitrile and following the absorbance at either 214 or 280 nm. Reactions were performed at 37 °C for 30 minutes in 50mM Tris-HCl 100 mM NaCl pH 7.5 and stopped by the addition of 0.1% TFA. Substrate concentrations ranged from 0.2 to 5 times the K_m whenever possible and the enzyme concentration was chosen so that hydrolysis was less than 10% product formed. The kinetic parameters, k_{cat} and K_m, were obtained by direct fit to a Michaelis–Menten curve using the Grafit program

Peptidomics

Peptides were extracted from whole mouse brain (minus the olfactory bulb and cerebellum) as described (Che et al., 2007; Morano et al., 2008). In brief, four mice (C57B6/J, 2-3 months old) were sacrificed by decapitation and the head was immediately irradiated in a conventional microwave oven for 8 seconds to raise the brain temperature to 80°C and inactivate proteases that can rapidly degrade cellular proteins during postmortem processing of the tissue (Che et al., 2005). The mouse brain was isolated, the olfactory bulb and cerebellum were discarded, and the remainder of the brain was frozen in dry ice and stored at -70°C until peptide extraction. Peptides were extracted by sonication of each brain in 1.0 mL ice-cold water followed by incubation at 70°C for 20 min, cooling in an ice bath, acidification with 120 µL of 0.1 M HCl to a final concentration of 10 mM HCl, and centrifugation at 13,000 *g* for 30 min at 4°C. The supernatants (peptide extracts) from each of the four mice were pooled, mixed, and 60 µl aliquots neutralized to pH 8 by 75 mM borate buffer and incubated in the presence of purified CPA4 (20 ng, 670 pg, or 23 pg), or in the absence of enzyme, for 90 minutes. The reactions were stopped by the addition of 2.5 mM benzyl-succinic acid and heat-inactivated at 80°C for 10 min. Labeling with 4-trimethyl-ammoniumbutyrate (TMAB) isotopic tags containing either no (D0), 3 atoms (D3), 6 atoms (D6), or 9 atoms (D9) of deuterium was performed as described (Morano et al., 2008). After labeling, the TMAB reagent was quenched with glycine, the 4 samples were combined, and peptides were isolated by passage through a Microcon® YM-10 unit (Millipore) to remove proteins >10 kDa. The samples were treated with hydroxylamine to remove TMAB labels from Tyr residues, desalted with a PepClean™ C-18 spin column (Pierce), and concentrated to ~20 µl in a vacuum centrifuge.

Mass spectrometry was performed as previously described (Morano et al., 2008; Zhang et al., 2008). In brief, the peptide mixture was trapped and washed on a PepMap™ C18 trapping column (5 µm, 100 Å, 300 µm i.d. × 5 mm, LC Packing, Marlton, NJ) and separated on a Grace Vydac MS C18 capillary column (3 µm, 100 Å, 75 µm i.d. × 150 mm, Hesperia, CA) at 4 µL/min with an acetonitrile gradient in 0.1% formic acid. The column eluate was analyzed on an API Q-Star Pulsar-i™ quadrupole time-of-flight mass spectrometer (Applied Biosystems/MDS Sciex, Foster City, CA) in

information-dependent acquisition mode, with a collision energy for MS/MS (20-45 eV) based on the *m/z* value and charge state of the ion selected.

Data was analyzed as described (Morano et al., 2008). In brief, Mascot (http://www.matrix-science.com) was used followed by manual verification; identifications were rejected unless 80% or more of the major fragments observed in MS/MS matched predicted b- or y-series fragments (minimum five matches). Additional criteria included all of the following: a parent mass within 40 ppm (preferably 20 ppm) of the theoretical mass; the observed number of TMAB tags on the peptide matched the predicted number of free amines available (i.e., Lys residue and N-terminus); and the observed charge state(s) of the peptide was consistent with the expected number of positive charges. The intensity of the TMAB-labeled isotopic peaks were determined by measuring peak heights of the monoisotopic peak and the additional peaks containing 1 and 2 atoms of ^{13}C. The peak intensities of the samples incubated with enzyme were compared to the peak intensity of the sample incubated without enzyme; previous studies examining 4 aliquots of brain extract with the 4 different isotopic tags showed average ratios of 1.00 (Morano et al., 2008).

2.3 RESULTS

The sequence analysis of PCPA4 protein reveals the presence of a 16 amino acid-long signal peptide consensus sequence and therefore the protein is predicted to be translocated to the lumen of the ER upon synthesis. From the ER, many options are possible, including retention in the secretory pathway, targeting to lysosomes, or secretion. Because CPA4 lacks any transmembrane domains and other members of this subfamily are secreted, PCPA4 is predicted to also be secreted from cells. Some CPs such as CPA1 remain as soluble proteins in the extracellular environment (Hamstra and Rehemtulla, 1999) while others like CPA6 (Lyons et al., 2008) or mouse CPZ (Novikova et al., 2000) are retained within the ECM. In relation to proprotein processing, some proCPs are activated before or upon secretion by proteases present in the secretory pathway (Lyons et al., 2008) while others are activated after secretion by extracellular proteases (Hamstra and Rehemtulla, 1999). To study whether PCPA4 is secreted and activated, HEK 293T cells were transiently transfected with HA-tagged PCPA4 cDNA. After 48 h of transfection, cells, media and ECM were collected, fractionated and analysed by Western blot. A band of an apparent molecular mass of 50 kDa corresponding to the full-length PCPA4 was detected predominantly in the cell culture medium (Fig. 1 lane 5). PCPA4 was also detected in two different cell extracts: the zymogen appears in a low salt extract but also at low levels in a detergent/high salt extract (Fig. 1 lanes 1 and 3). The absence of any PCPA4 in the ECM shows that the protein remains as a soluble protein in the extracellular medium. In all cases PCPA4 is found in the full length form indicating that its activating protease is not present or active in sufficient levels in the system. To investigate if this behaviour would be similar in other cell lines, we repeated the transfection of PCPA4 in five different cell lines: DU145 and PC-3, which are prostate cancer cells previously found to express PCPA4 in relatively high levels (Huang et al., 1999); LNCaP, a prostate cancer cell line that does not express PCPA4; HeLa cells, which arise from a cervical adenocarcinoma; and 1BR.3.G, a human skin fibroblast cell line. In all cases, PCPA4 was predominantly found soluble in the extracellular medium in the zymogen form (data not shown).

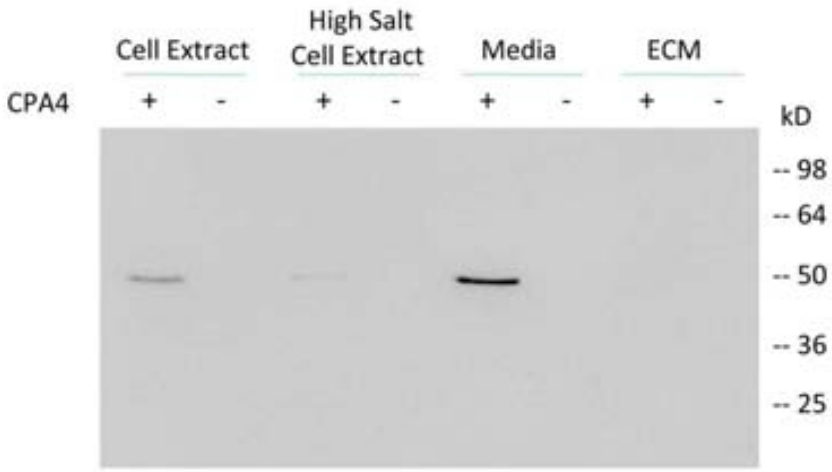

Figure 1. PCPA4 is a soluble secreted protein. PCPA4 transfection in mammalian cells. HEK 293T cells were transfected with empty vector (-) or plasmid expressing HA-tagged PCPA4 (+) and different fractions were collected and prepared. Equal amounts of each fraction were analyzed by Western blot using an HA antibody

To demonstrate that PCPA4 expressed in HEK 293T cells was properly folded and able to produce the enzymatically active form, extracellular medium collected after 48 h of transfection was treated with trypsin. Western blot analysis of these samples shows that trypsin causes the disappearance of the 50 kDa band corresponding to PCPA4 and the appearance of a 35 kDa band corresponding to the expected molecular weight of the active enzyme (Fig. 2). Enzymatic analysis of these samples shows no detectable CP activity in the extracellular medium of PCPA4 transfected cells, but incubation with trypsin resulted in the appearance of A-type CP enzymatic activity against the AAFP substrate (data not shown). The finding that PCPA4 can be activated with trypsin indicates that it is properly folded and that the absence of detectable zymogen processing in the HEK 293T cells, as well as in various other cell lines, is due to insufficient levels of the activating enzyme and not due to the improper folding of PCPA4 in these cell lines.

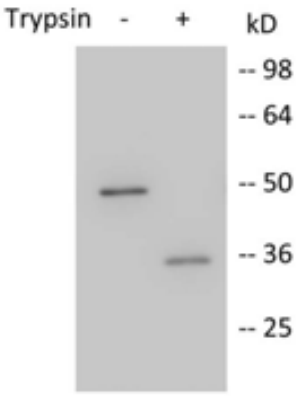

Figure 2. Trypsin processing of PCPA4. Extracellular media was treated with (+) or without (-) trypsin and analyzed by Western blot

The presence of PCPA4 in the cells was further investigated by immunocytochemistry. PCPA4 was found to show some overlap with an ER marker (Fig. 3), which is consistent with the finding that intracellular PCPA4 is present in the secretory pathway of the cell.

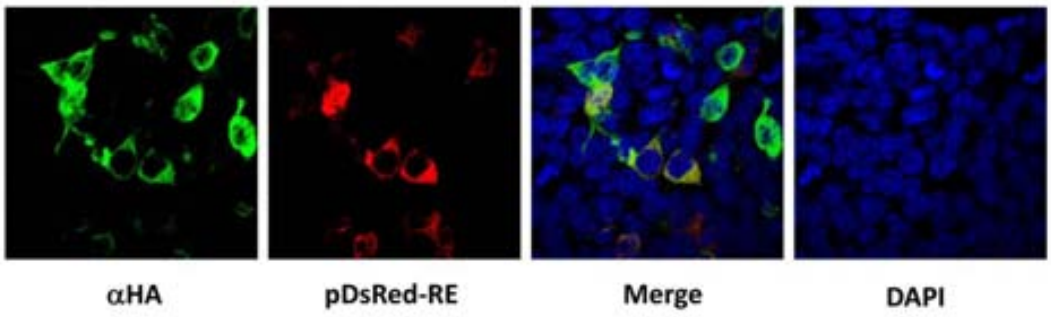

Figure 3. Subcellular distribution of PCPA4. HEK 293T cells were transfected with HA-tagged PCPA4 and pDsRed-ER cDNAs. After fixation, transfected cells were incubated with mouse anti-HA antibody followed by anti-mouse IgG Alexa 488 (green) and DAPI to detect cell nuclei.

Although there are several reports about CPA4 structure (Garcia-Castellanos et al., 2005; Pallares et al., 2005), expression (Huang et al., 1999) and relation to prostate cancer aggressiveness (Bentley et al., 2003; Kayashima et al., 2003), a detailed analysis of its enzymatic properties has not been previously reported. Recombinant protein obtained in the *Pichia pastoris* system as previously described (Pallares et al., 2005) was used for this purpose. Mature CPA4 was obtained after treatment of the

proenzyme with trypsin and subsequent purification. As expected, the active enzyme shows the ability to remove C-terminal residues from both FA-Phe-Phe and FA-Phe-Ala. No detectable activity towards substrates with basic (FA-Ala-Arg) or acidic (hippuryl-L-glutamic acid) C-terminal residues was found. This is in accordance with predictions based on its sequence and structure that includes CPA4 in the group of A-type CPs. The *Pichia*-expressed proenzyme shows 3% of residual activity against the AAFP substrate when compared to the active enzyme. Although the proenzyme shows residual activity, it could not be detected in the conditioned media of PCPA4 transfected HEK 293T cells because this activity falls below the limit of detection of the technique.

The influence of pH on CPA4 activity was analyzed using both FA-Phe-Phe and FA-Phe-Ala. The optimal pH for cleavage of either substrate is in the range of 8.5-9 with very low activity below pH 6 (Fig. 4). This suggests that CPA4 is maximally active in the extracellular environment after secretion, and has little activity within the acidic environment of the secretory pathway.

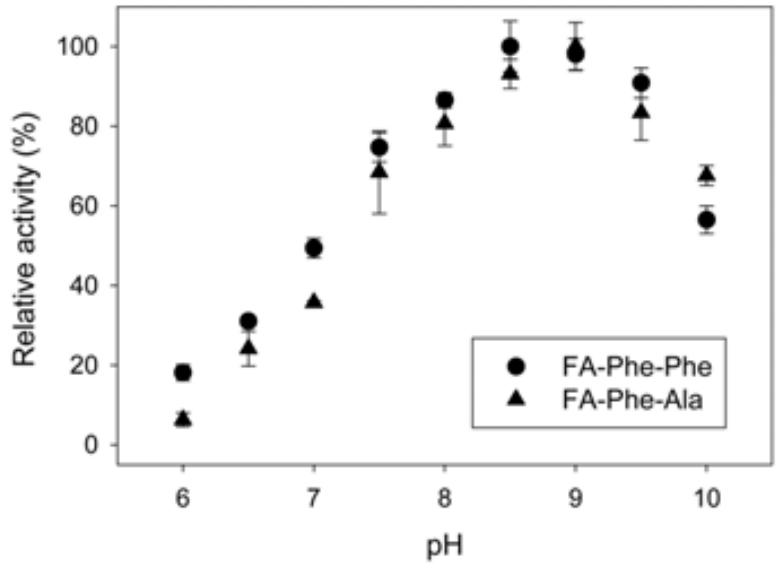

Figure 4. Effect of pH on CPA4 activity. CPA4 activity was determined by measuring initial rates using 100 µM FA-Phe-Phe or FA-Phe-Ala in 50 mM Tris-acetate buffer at the indicated pH at 25ºC. Error bars represent the standard error of the mean for three independent determinations.

The effect of several different CP inhibitors on CPA4 activity was assayed. Benzylsuccinic acid (Byers and Wolfenden, 1973), a potent synthetic inhibitor of the A-type CP subfamily, at a concentration of 100 µM produced 30% of inhibition on CPA4 activity and completely inhibited the enzyme when 1 mM concentration of the inhibitor was used. The Zn^{2+}-chelating agent 1,10-phenanthroline (Felber et al., 1962) showed 70% inhibition of the CP activity after 30 minutes of incubation at 1 mM concentration. CPA4 showed resistance against the action of EDTA, another chelating inhibitor of CPs: complete inhibition of CPA4 activity was only achieved using a 100 mM concentration of the inhibitor in an overnight incubation, while the same inhibitor concentration caused only 60% inhibition after 5 hours of treatment. Recombinant forms of potato CP inhibitor (PCI) (Ryan et al., 1974), leech CP inhibitor (LCI) (Reverter et al., 1998) and tick CP inhibitor (TCI) (Arolas et al., 2005) were tested with CPA4 and found to exhibit equilibrium dissociation constant (K_i) values in the low nanomolar range (Table 1). The protein inhibitor with the highest affinity for CPA4 was found to be that from tick *Rhipicephalus bursa*, for which a constant of 0.8 ± 0.3 nM was obtained (Table 1). The K_i values for CPA4 determined in the present study are generally comparable to values previously published for CPA4 with latexin (Pallares et al., 2005) and *Ascaris* carboxypeptidase inhibitor (ACI) (Sanglas et al., 2009); these results are included in Table 1 for comparison with the present data.

Table 1. Ki values for CPA4 against different proteinaceous carboxypeptidase inhibitors.

Inhibitor	K_i (nM)
ACI	23.9 ± 3.9[a]
Latexin	3.0 ± 0.3[b]
PCI	1.3 ± 0.1
LCI	7.3 ± 0.4
TCI	0.8 ± 0.3

Taken from [a] (30) and [b] (15).

To investigate the substrate specificity of CPA4 a new series of substrates of the type FA-Phe-Xaa were synthesized. These substrates incorporated the furylacryloyl moiety as a chromophore, allowing for continuous monitoring of CP activity at a wavelength that has little interference from proteins or aromatic derivatives (Peterson et al., 1982). This new series of substrates provided a broad range of

hydrophobic/aliphatic residues in the C-terminal position so that substrate specificities of A-type CPs could be evaluated. The different amino acids included in the C-terminal position were His, Leu, Ile, Phe, Trp, Ala, Val and Met. All substrates contain Phe in the penultimate position and are of the same length, allowing a direct comparison of the substrates' C-terminal residues. Kinetic constants were determined for CPA4 and also for human CPA1 and CPA2 to directly compare these related enzymes with the same substrates. CPA1 and CPA2 were obtained as recombinant proteins in the *Pichia pastoris* system, purified in their zymogen form, activated with trypsin and further purified as previously described (Pallares et al., 2008; Reverter et al., 1997). k_{cat}, K_m and k_{cat}/K_m values are shown in Table 1, and indicate that although all three enzymes have overlapping specificities they differ markedly. While CPA1 and CPA4 display similar and broader substrate specificities, CPA2 shows a substrate specificity that is essentially restricted to large aromatic residues in the C-terminal position. CPA4 shows a preference for hydrophobic C-terminal amino acids like Phe, Leu, Met, Val, and Ile, as judged by k_{cat}/K_m values. Of all the dipeptide substrates tested with CPA4, FA-Phe-Trp exhibited the highest k_{cat}, but the high K_m resulted in a lower k_{cat}/K_m for this substrate than for FA-Phe-Phe. Overall, the k_{cat} varied only 5-fold among the substrates that showed measureable k_{cat} values, while the K_m varied 32-fold and the k_{cat}/K_m varied 12-fold for these substrates (Table 2).

Table 2. Kinetic constants for peptide substrate hydrolysis by CPA1, CPA2 and CPA4.

	CPA4				CPA1				CPA2			
Substrate	k_{cat} (1/s)	K_m (μM)	k_{cat}/K_m ($M^{-1}.s^{-1}.10^{-6}$)	k_{cat}/K_m relative to FAPP	k_{cat} (1/s)	K_m (μM)	k_{cat}/K_m ($M^{-1}.s^{-1}.10^{-6}$)	k_{cat}/K_m relative to FAPP	k_{cat} (1/s)	K_m (μM)	k_{cat}/K_m ($M^{-1}.s^{-1}.10^{-6}$)	k_{cat}/K_m relative to FAPP
FA-Phe-Phe	44.3	55.6	7.97	1.00	76.6	20.4	37.6	1.00	121	36.4	33.3	1.00
FA-Phe-Trp	57.3	615	0.93	0.12	81.3	166	4.90	0.13	65.3	16.3	40.2	1.21
FA-Phe-Leu	13.4	19.4	6.93	0.87	34.0	15.7	21.7	0.58	18.6	393	0.47	0.014
FA-Phe-Ile	12.4	23.3	5.32	0.67	9.77	9.33	10.5	0.28	5.37	460	0.12	0.0035
FA-Phe-Met	23.9	40.0	5.81	0.73	51.8	128	4.03	0.11	NM	NM	0.040	0.0012
FA-Phe-Ala	24.3	372	0.65	0.082	78.0	300	2.58	0.069	NM	NM	NM	NM
FA-Phe-Val	19.4	57.3	3.38	0.42	12.3	8.95	14.8	0.39	NM	NM	0.030	0.0010
FA-Phe-His	NM	NM	0.27	0.034	NM	NM	1.04	0.028	NM	NM	NM	NM

NM: Not measurable, below the limit of detection.

To further characterize the catalytic activity of CPA4, its ability to cleave a series of peptides was studied. Some of the peptides chosen for this analysis are biologically active and are known to be secreted into the extracellular fluid of tissues where CPA4 is expressed. Others are not likely to be biological substrates of CPA4 but provide further information of the ability of this enzyme to cleave peptides. All of these peptides were previously tested against CPA6 (Lyons et al., 2008) so they are useful for comparison purposes. The enzyme, at a concentration of 1 nM, was incubated at 37ºC with the peptides in a final concentration of 10 µM. In order to obtain semi-quantitative information of the cleavage, different times of incubation were analysed by MALDI-TOF MS. While some peptides were extensively cleaved, other peptides were only partially cleaved, and others were not cleaved under the conditions tested (Table 3). CPA4 was found to completely cleave PEN and Met-enkephalin-Arg-Phe after 30 minutes of incubation. Big SAAS and neurotensin were also extensively cleaved after 1-2 h of incubation. Partial cleavage of approximately 50% of angiotensin was found after 30 minutes of incubation. Des-Asp^{1}-angiotensin and Leu-enkephalin were partially cleaved in 2 h; and Met-Enkephalin and CPA5 C-terminus needed an overnight incubation for a partial digestion by CPA4. Only a small percentage of cleavage was detected for human little SAAS. No digestion of the aldolase C-terminal peptide was detected, although it was expected to be cleaved because of the C-terminal Tyr. When a 10-fold higher concentration of enzyme (10 nM) was used, substantial cleavage of the C-terminal Tyr was observed after 2 hours of incubation. Neither bradykinin nor des-Arg^{9}-bradykinin were substrates for CPA4 (Table 3). Although des-Arg^{9}-bradykinin has a C-terminal Phe, the presence of a Pro in the penultimate position may explain why CPA4 is not able to cleave this substrate under the conditions tested.

By using higher concentrations of enzyme (20 nM) and substrate (50 µM) it was possible to examine the processivity of CPA4 cleavage. The sequential hydrolysis of amino acids from the C-terminus of neurotensin was observed (Fig. 5). The product of the cleavage of the C-terminal Leu was detected at 10 minutes of incubation, while hydrolysis of the penultimate Ile was observed at 1 hour of incubation (Fig. 5). Further removal of the Tyr was not detected, even upon prolonged incubation, presumably because of the negative influence of the Pro in the P1 position of this peptide (Fig. 5).

The action of CPA4 (20 nM) with angiotensin (200 µM) also showed processivity: the removal of C-terminal Leu, penultimate His and antepenultimate Phe is sequentially observed at 10 minutes, 30 minutes and 4 hours of incubation, respectively (data not shown).

Table 3. Analysis of peptide cleavage by CPA4 using MALDI-TOF MS

Peptides[a]	**Sequence and site of cleavage**[b]
Peptides extensively cleaved (> 70% cleavage)	
Neurotensin	pyroELYENKPRRPYI ↓ L
PEN (rat)	AVDQDLGPEVPPENVLGALLR ↓ V
Big SAAS (rat)	ARPVKEPRSLSAASAPLAETSTPLR ↓ L
Met-Enkephalin-Arg-Phe	YGGFMR ↓ F
Peptides partially cleaved (< 70% cleavage)	
Angiotensin I	DRVYIHPFH ↓ L
Des-Asp^{1} - Angiotensin I	RVYIHPFH ↓ L
Little SAAS (human)	GLSAASPPLAETGAPRR ↓ F
CPA5 C-terminus	CQTIMKHTLNHP ↓ Y
Leu-Enkephalin	YGGF ↓ L
Met-Enkephalin	YGGF ↓ M
Peptides not cleaved	
Aldolase C-terminus (mouse)[c]	LYIANHAY
Big LEN (rat)	LENSSPQAPARRLLPP
Bradykinin	RPPGFSPFR
Des-Arg^{9} – Bradykinin	RPPGFSPF
YE-17 (rat)	YPPENVLGALLRVKRLE
CPE-mid (bovine)	YDPDRPPCRKNDD

a) Peptides were incubated at a 10 µM concentration with CPA4 at a 1 nM concentration in 50 mM Tris-HCl 100 mM NaCl pH 7.5 at 37ºC and analyzed for cleavage at different times (0.5, 1, 2, and 16 h). Peptide names are indicated; note that PEN, SAAS, and LEN are names, not abbreviations.

b) Cleavage sites are indicated by arrow.

c) Although no cleavage was detected under the conditions used for all other peptides in this table (1 nM CPA4), extensive cleavage for this peptide was observed when enzyme concentration was increased to 10 nM.

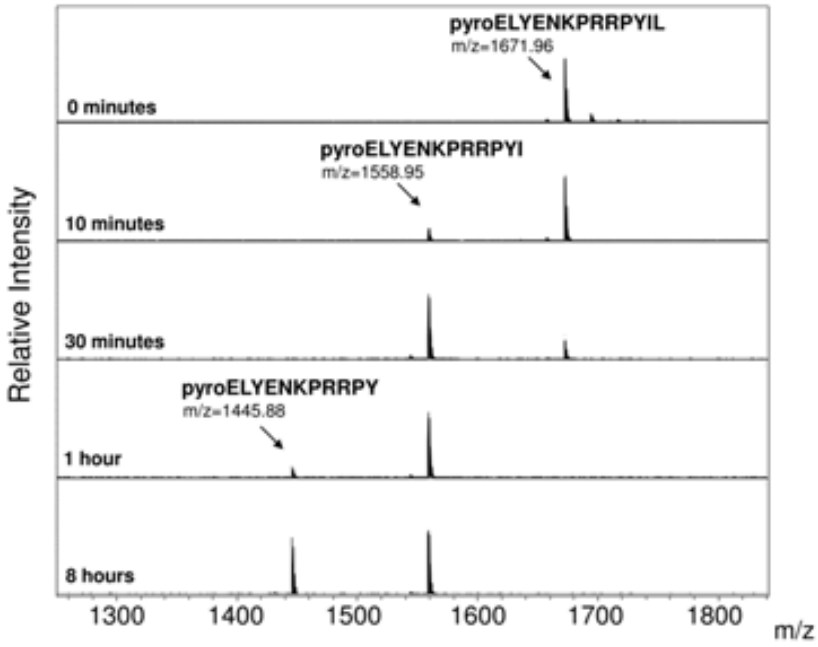

Figure 5. MALDI-TOF MS analysis of neurotensin cleavage by CPA4. MALDI-TOF MS spectra of 50 µM neurotensin incubated for different times at 37°C with 20 nM CPA4. Numbers above the major peaks indicate the monoisotopic masses of the MH+ ion (m/z). The peak with mass of 1671.96 corresponds to neurotensin (theoretical monoisotopic mass of neurotensin = 1671.91) and a first cleavage product with a mass of 1558.95 formed in the presence of CPA4 corresponds to the peptide lacking the C-terminal Leu (theoretical mass = 1558.84) and a second cleavage product with mass 1445.88 corresponds to the peptide lacking both C-terminal Leu and Ile (theoretical mass = 1445.75).

Five of the peptides identified as CPA4 substrates were further investigated by determining the kinetic constants for CPA4 catalyzed hydrolysis. The data is summarized in Table 4 and shows a broad variation in K_m and k_{cat} values for the different peptides. CPA4 cleaves neurotensin and angiotensin with high efficiency (43.8 $\mu M^{-1}.min^{-1}$ and 9.43 $\mu M^{-1}.min^{-1}$ respectively) and low K_m (0.329 µM and 9.23 µM respectively).

Table 4. Kinetic constants for the hydrolysis of biological peptides by CPA4

Peptide	k_{cat} (s^{-1})	K_m (µM)	k_{cat}/K_m ($\mu M^{-1}.min^{-1}$)
Neurotensin	0.240 ± 0.017	0.329 ± 0.049	43.8 ± 9.5
Met-enkephalin-Arg-Phe	1.45 ± 0.03	9.23 ± 0.64	9.43 ± 0.86
Angiotensin I	8.23 ± 0.34	227 ± 25	2.18 ± 0.31
Leu-enkephalin	2.33 ± 0.11	165 ± 19	0.85 ± 0.13
Met-enkephalin	3.43 ± 0.16	550 ± 52	0.374 ± 0.053

To evaluate the ability of CPA4 to cleave a larger number of peptides, a quantitative peptidomics approach using differential isotopic-tagging and MS was used. For this, peptides were extracted from heat-inactivated mouse brains and purified away from proteins by microfiltration (10 kDa filter cut-off), and then incubated with varying amounts of purified CPA4, or in the absence of enzyme, for 1.5 h (Fig. 6). After heat inactivation of CPA4, the various samples were labeled with one of four isotopic tags, combined, and analyzed by liquid chromatography / mass spectrometry (LC/MS). Over 100 peptides were detected in the LC/MS analysis, including many of the brain peptides tested in Table 3 (neurotensin, PEN, big SAAS, Met-enkephalin-Arg-Phe (i.e. the proenkephalin heptapeptide), little SAAS, Leu-enkephalin, Met-enkephalin, and big LEN). Angiotensin and bradykinin were not detected in the brain peptidome, reflecting the low abundance of these peptides in brain. Quantification of the relative peptides levels in the enzyme-treated and control incubations was performed by determining peak intensity; representative data is indicated in Fig. 6. Many of the observed peptides showed roughly equal peak intensities; these are neither substrates nor products of CPA4 under the incubation conditions employed. Some peptides showed a decrease in peak intensity only for the sample incubated with the highest amount of enzyme; these are weak substrates of CPA4. Other peptides showed a large decrease in peak intensity for both the highest and the medium amount of enzyme; these are good substrates of CPA4. Finally, some peptides showed a large increase in peak intensity when incubated with enzyme; these are products of CPA4. Control studies in which aliquots of mouse brain peptides were isotopically tagged and analyzed by LC/MS, without incubation with enzyme, showed 4 equal peaks for all of the peptides observed in the present study (Morano et al., 2008). Approximately half of the detected peptides could be identified by tandem mass spectrometry (MS/MS) sequence analysis. In addition to these sequenced peptides, several other peptides were tentatively identified based on matches to known or predicted peptides; the criteria for these matches included an observed monoisotopic mass that was within 0.004% of the theoretical mass, a correct charge (equal to the number of basic residues plus the N-terminus), and a correct number of isotopic tags incorporated (based on the number of free amines in the peptide). These tentatively identified peptides are included in Tables 5-8 with the sequence surrounded by

parentheses; all other sequences in these tables were confirmed by MS/MS sequencing (in addition to fitting the above criteria).

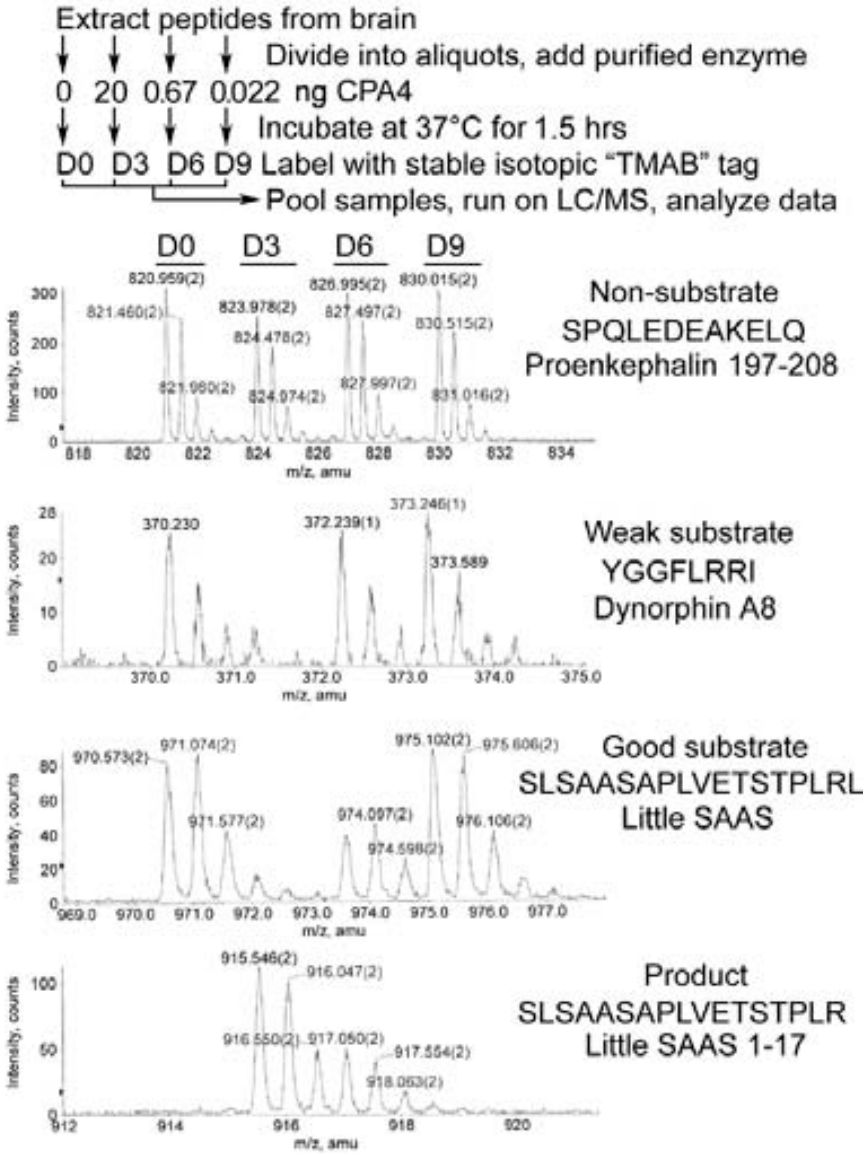

Figure 6. Quantitative peptidomics scheme and representative results. Peptides were extracted from microwave-irradiated mouse brain and separated from proteins by microfiltration with a 10 kDa filter. Aliquots were incubated with purified CPA4 (20 ng or 30-fold dilutions) or in the absence of CPA4. After incubation, the samples were labeled with one of 4 stable isotopic TMAB tags that differ in the number of hydrogens and deuteriums. Samples were then pooled and analyzed by liquid chromatography/mass spectrometry (LC/MS). Representative data are shown. Peptides that showed less than a 25% decrease in peak intensity for the sample incubated with the highest concentration of enzyme (the D3 TMAB tag) were considered non-substrates, as shown for the proenkephalin fragment SPGLEDEAKELQ. Peptides which showed a major decrease with the highest concentration of enzyme (>50% decrease) but less than a 25% decrease with the medium concentration of enzyme were considered weak substrates. An example of a weak substrate is the neuropeptide dynorphin A8. Peptides which showed a substantial decrease in peak intensity for the highest concentration of enzyme, and at least a 25% decrease for the medium concentration of enzyme were considered good substrates. An example is little SAAS, an 18-residue peptide with a C-terminal Leu. Peptides that showed an increase in peak intensity for the samples incubated with enzyme were considered products of CPA4. An example is little SAAS 1-17, which lacks the C-terminal Leu of little SAAS.

Table 5. Good substrates of CPA4. These are defined as those peptides affected (>25% decrease) by the medium concentration of enzyme, and with the high concentration of enzyme greatly reducing the level of peptide (>80% decrease).

Protein name	Peptide name	Sequence	z	T	Obs M	Theor M	ppm	Ratio, CPA4/no enzyme		
								Low	Med	High
Peptidylprolyl isomerase A	118-129	KTEWLDGKHVVF	4	3	1457.74	1457.767	-21	0.81	<0.10	<0.10
Chromogranin A	374-388	LEGEDDPDRSMKLSF	3	2	1737.82	1737.788	17	1.13	<0.10	<0.10
Proneurotensin	Neurotensin	pELYENKPRRPYIL	3	1	1671.94	1671.910	18	1.27	<0.20	<0.20
Proneurotensin	Neuromedin N	KIPYIL	2	2	745.47	745.474	1	1.03	0.21	<0.10
Proenkephalin	Heptapeptide	YGGFMRF	2	1	876.41	876.395	12	0.94	0.37	<0.10
Proenkephalin	Heptapeptide, Met-oxidized	YGGF-Mox-RF	2	1	892.40	892.398	8	1.03	0.40	<0.20
Peroxiredoxin V	N-terminus	APIKVGDAIPSVEVF	2	2	1540.87	1540.850	14	0.95	0.40	<0.20
Synaptosomal-associated protein (Snap91)	C-terminus	SPSPTPATQSPKKPPAKDPLADLNIKDFL	5	5	3072.68	3072.649	10	0.96	0.48	<0.10
ProSAAS	Little SAAS	SLSAASAPLVETSTPLRL	2,3	1	1812.02	1812.007	9	1.04	0.57	<0.10
Chromogranin B	64-86	SGKEVKGEEKGENQNSKFEVRLL	5,6	5	2604.38	2604.351	11	1.13	0.66	<0.10
Chromogranin B	600-613	QYDGVAELDQLLHY	2	1	1662.84	1662.789	28	1.04	0.67	<0.10
Propeptidyl-amidating-monooxygenase	Cleaved pro peptide	FRSPLSVF	2	1	951.53	951.527	-1	1.08	0.67	<0.20
ProSAAS	PEN	SVDQDLGPEVPPENVLGALLRV	2,3	1	2316.27	2316.232	17	1.04	0.71	<0.10
ProSAAS	PEN-20	SVDQDLGPEVPPENVLGALL	2	1	2061.12	2061.063	26	1.53	0.72	<0.20
Elongation factor 1 beta 2	N-terminus	GFGDLKTPAGLQVL	2	2	1414.78	1414.782	2	0.98	0.74	<0.20
ProSAAS	Big SAAS	ARPVKEPRSLSAASAPLVETSTPLRL	5	2	2745.61	2745.550	21	1.14	0.75	<0.10

Sequences were determined by tandem mass spectrometry (MS/MS) analysis, as described in Experimental Procedures. Abbreviations: Mox, oxidized methionine; pE, pyroGlu; z, charge; T, number of isotopic tags incorporated into peptide; Obs M, observed monoisotopic mass; Theor M, theoretical monoisotopic mass; ppm, difference between Obs M and Theor M in parts per million; Ratio CPA4/no enzyme, the ratio in peak intensity between the sample incubated with enzyme and the sample incubated without enzyme. When the peak intensity was below the background, the ratio is expressed as <0.10 or <0.20, depending on the signal strength and background levels.

Table 6. Weak substrates of CPA4. These peptides are defined as those peptides which were not substantially affected by the medium concentration of enzyme (less than 25% change), but which showed a decrease with the high concentration of enzyme.

Protein name	Peptide name	Sequence	z	T	Obs M	Theor M	ppm	Ratio, CPA4/no enzyme		
								Low	Med	High
Provasopressin	151-end	VQLAGTRESVDSAKPRVY	4	2	1975.08	1975.049	15	0.66	0.80	<0.10
Actin (beta or gamma)	169-177 of beta	ALPHAILRL	3	1	1002.64	1002.641	0	1.15	0.82	<0.10
Cathepsin D	138-155	YTVFDRDNNRVGFANAVV	3	1	2056.07	2056.013	26	0.91	0.83	<0.10
Secretogranin II	287-316	(SGQLGLPDEENRRESKDQLSEDASKVITYL)	5	3	3376.71	3376.675	10	0.91	0.85	<0.10
Peptidylprolyl isomerase A	26-39	ADKVPKTAENFRAL	4	3	1558.86	1558.847	7	1.11	0.92	0.21
Proteasome subunit, beta type, 6	34-41	(TTIMAVQF)	2	1	909.47	909.453	8	0.97	0.92	<0.20
Thioredoxin 1	N-terminus	VKLIESKEAFQEAL	3	3	1603.91	1603.882	15	1.01	0.93	0.38
Vacuolar ATP synthase subunit 2	C-terminus	EVRPQVHPNYRVTV	4	1	1692.91	1692.913	-2	0.87	0.93	<0.10
Secretogranin II	300-316	ESKDQLSEDASKVITYL	3	3	1924.98	1924.960	13	1.09	0.95	<0.10
Prodynorphin	Dynorphin A8	YGGFLRRI	3	1	980.56	980.556	-1	0.97	0.96	<0.10
Ribosomal protein S21	C-terminus	AKADGIVSKNF	3	3	1148.63	1148.619	7	0.90	0.99	<0.10
Proenkephalin	Octapeptide	YGGFMRSL	2	1	929.45	929.451	-3	1.03	1.00	0.17
Triose-phosphate isomerase	149-163	KVIADNVKDWSKVVL	4	4	1712.99	1712.982	5	0.92	1.00	<0.10
Chromogranin A	392-402	AYGFRDPGPQL	2	1	1219.61	1219.600	12	1.02	1.01	0.45
Proenkephalin	Leu-Enkephalin	YGGFL	1	1	555.27	555.269	10	1.10	1.01	<0.10
Clathrin light chain A	C-terminus	SVLISLKQAPLVH	3	2	1403.85	1403.850	0	1.08	1.02	<0.12
Chromogranin A	374-390	LEGEDDPDRSMKLSFRT	4	2	1994.97	1994.937	16	1.06	1.03	<0.10
Voltage-dependent anion channel protein 1	C-terminus	AGGHKLGLGLEFQA	3	2	1396.76	1396.746	10	1.08	1.03	0.23
Peptidylprolyl isomerase A	23-39	ELFADKVPKTAENFRAL	4	3	1948.06	1948.042	10	1.00	1.04	<0.10
Peptidylprolyl isomerase A	84-92	EDENFILKH	3	2	1143.56	1143.556	6	1.26	1.08	<0.20
Peptidylprolyl isomerase A	84-100	EDENFILKHTGPGILSM	3	2	1899.96	1899.948	5	1.10	1.10	<0.20
Procholecystokinin	46-63	AVLRTDGEPRARLGALLA	4	1	1878.11	1878.080	14	1.07	1.13	<0.10
Procholecystokinin	46-62	AVLRTDGEPRARLGALL	4	1	1807.05	1807.043	3	1.19	1.16	<0.10
Proenkephalin	Met-Enkephalin	YGGFM	1	1	573.23	573.226	10	1.15	1.20	0.18

Abbreviations are described in Table 5. Peptide sequences in parentheses are tentative identifications, based on similarities to known peptides, without MS/MS sequence information. These similarities include the monoisotopic mass, the charge (based on the number of basic residues in the peptide), and the number of isotopic tags incorporated into the peptide (based on the number of free amines). See Experimental Procedures for further details

Table 7. Products of CPA4. These peptides are defined as those peptides that increased >100% with one or more of the concentrations of CPA4.

Protein name	Peptide name	Sequence	Cleaved	z	T	Obs M	Theor M	ppm	Ratio, CPA4/no enzyme		
									Low	Med	High
Procholecystokinin	43-60	(QLRAVLRTDGEPRARLGA)	L	5	1	1978.14	1978.118	13	1.37	2.33	1.52
Procholecystokinin	46-60	AVLRTDGEPRARLGA	L	4	1	1580.88	1580.875	6	1.45	1.64	5.05
Procholecystokinin	46-61	AVLRTDGEPRARLGAL	L	4	1	1693.97	1693.970	2	1.00	3.60	12
Proneurotensin	Neurotensin 1-12	pELYENKPRRPYI	L	3	1	1558.86	1558.836	15	>5.6	>19	>7.2
Proneurotensin	Neurotensin 1-11	pELYENKPRRPY	I	3	1	1445.77	1445.752	12	nd	nd	>10
ProSAAS	Little SAAS 1-17	SLSAASAPLVETSTPLR	L	2	1	1698.94	1698.915	16	nd	>6.6	>11
ProSAAS	PEN 1-21	SVDQDLGPEVPPENVLGALLR	V	2,3	1	2217.19	2217.164	13	>1.4	>9.7	>20
ProSAAS	Big SAAS 1-25	ARPVKEPRSLSAASAPLVETSTPLR	L	5	2	2632.50	2632.466	13	nd	>8.8	>19
Chromogranin B	600-612	QYDGVAELDQLLH	Y	2	1	1499.75	1499.726	14	1.00	>10	>10
Chromogranin B	438-453	LLDEGHYPVRESPIDT	A	3	1	1839.92	1839.900	10	0.94	1.09	2.14

Abbreviations are defined in Tables 5 and 6. nd, not detected. For those peptides not detected in the "no enzyme" control, the ratio is expressed using the > sign; relative levels between the different concentrations of enzyme are reflected in the different lower limits for the increase. The amino acid cleaved by CPA4 to generate the observed peptide is indicated in column 4.

Table 8. Brain peptides that are not substrates or products of CPA4.

Protein name	Peptide name	Sequence	z	T	Obs M	Theor M	ppm	Ratio, CPA4/no enzyme		
								Low	Med	High
Phosphatidylethanolamine-binding protein 1	30-46	AGVTVDELGKVLTPTQV	3	2	1725.99	1725.951	23	0.80	1.03	0.74
Voltage-dependent anion channel protein 2	C-terminus	AGGHKLGLALELEA	3	2	1377.77	1377.762	5	0.91	0.88	0.79
Proopiomelanocortin	Phosphorylated CLIP	RPVKVYPNVAENE-phosphoS-AEAFPLEF	3	2	2585.27	2585.230	14	1.09	0.95	0.79
ProVasoactive Intestinal Peptide	111-122	ISSSISEDPVPI	2	1	1242.66	1242.634	25	1.32	1.28	0.80
Proenkephalin	197-208	SPQLEDEAKELQ	2,3	2	1385.68	1385.667	11	0.94	1.10	0.89
VGF	487-507	KKNAPPEPVPPPRAAPAPTHV	5	3	2170.23	2170.201	11	1.00	1.00	0.86
Chromogranin B	438-446	LLDEGHYPV	2	1	1041.52	1041.521	2	0.97	1.11	0.88
Prohormone convertase 2	94-104	IKMALQQEGFD	2	2	1278.64	1278.636	6	0.83	0.67	0.88
Chromogranin B	438-454	LLDEGHYPVRESPIDTA	3	1	1910.94	1910.930	7	0.82	0.95	0.90
Ubiquinol-cytochrome C reductase hinge protein	N-terminus	GDPKEEEEEELVDPLTTVREH	4	2	2450.20	2450.145	21	0.83	1.14	0.92
Ubiquinol-cytochrome c reductase complex 11 kDa protein	C-terminus	Cyano-CVAHKLFKNLK	3,4	3	1324.73	1324.749	-14	1.12	0.86	0.86
Prothyrotropin Releasing Hormone	83-106	EEKEEDVEAEERGDLGEVGAWRPH	5	2	2765.30	2765.253	15	1.09	1.09	0.92
Tubulin beta, isoforms 2 - 6	N-terminus	MREIVH(I/L)QA	3	1	1095.61	1095.586	20	1.00	1.00	0.92
Prothyrotropin Releasing Hormone	95-106	GDLGEVGAWRPH	3	1	1292.67	1292.626	33	1.08	1.15	0.93
Prodynorphin	Alpha-neoendorphin	YGGFLRKYPK	4	3	1227.66	1227.676	-17	1.09	0.93	0.93
Prohormone convertase 2	C-terminus	SLQSILRKN	3	2	1057.63	1057.624	6	1.05	0.92	0.93
Fibrinogen, alpha polypeptide	N-terminus	TDTEDKGEFLSEGGGVR	3	2	1795.85	1795.822	15	1.00	0.90	0.93
Prothyrotropin Releasing Hormone	25-50	LLEAAQEEGAVTPDLPGLEKVQVRPE	3,4	2	2787.50	2787.483	6	1.10	1.09	po
Proopiomelanocortin	CLIP	RPVKVYPNVAENESAEAFPLEF	3	2	2505.30	2505.260	14	1.10	1.12	0.94
Proenkephalin	Phosphorylated 238-261	FAESLP-phosphoS-DEEGENYSKEVPEIE	3	2	2577.11	2577.060	18	0.95	1.05	0.97
Protachykinin A	72-95	DADSSVEKQVALLKALYGHGQISH	4,5	3	2565.34	2565.319	8	0.86	1.10	1.00
Chromogranin B	588-597	SFARAPQLDL	2	1	1116.60	1116.590	11	1.03	0.88	0.98
Procholecystokinin	72-94	APSGRMSVLKNLQSLDPSHRISD	5	2	2507.31	2507.291	7	0.90	1.26	0.99
VGF	489-507	NAPPEPVPPPRAAPAPTHV	3	1	1914.04	1914.011	13	0.97	0.73	1.00
Cerebellin 4	64-77	ANSKVAFSAVRSTN	3	2	1450.77	1450.753	15	1.05	0.80	1.00

Thymosin beta-10	Entire protein	Ac-ADKPDMGEIASFDKAKLKKTETQEKNTLPTKETIEQEKRSEIS	9	8	*4936.37*	*4936.490*	-25	1.21	1.16	1.00
Proopiomelanocortin	J-peptide	AEEEAVWGDGSPEPSPRE-amide	3	1	1939.88	1939.855	12	1.03	1.16	1.01
Proenkephalin	218-228	VGRPEWWMDYQ	2	1	1465.68	1465.645	25	1.23	1.09	1.01
VGF	348-372	(DLGGRELQETQQERENEREEEAEQE)	4	1	3030.43	3030.340	29	0.98	1.00	0.83
Thymosin beta-4	Entire protein	Ac-SDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKQAGES	7	9	*4963.51*	*4963.540*	-5	1.15	1.10	1.02
Macrophage migration inhibitory factor	28-47	AQATGKPAQYIAVHVVPDQL	3	2	2105.15	2105.127	10	0.88	1.08	1.06
Chromogranin B	313-330	PSPKESKEADVATVRLGE	4	3	1911.98	1911.990	-5	1.00	1.04	1.07
ProSAAS	Little SAAS 1-16	SLSAASAPLVETSTPL	2	1	1542.84	1542.814	18	1.00	0.89	1.07
Cerebellin 1 precursor protein	Cerebellin	SGSAKVAFSAIRSTNH	3,4	2	1631.86	1631.838	13	0.92	1.26	1.10
Pyruvate kinase isozymes M1/M2	C-terminus	FTNTMRVVPVP	2	1	1259.69	1259.670	15	0.89	0.85	0.89
Procholecystokinin	71-94	KAPSGRMSVLKNLQSLDPSHRISD	6	3	2635.36	2635.386	-11	0.93	1.32	1.10
Tubulin beta, isoforms 2-6	N-terminus	MREIVH(I/L)QAGQ	3	1	1280.70	1280.674	17	1.02	1.15	1.04
Promelanin concentrating hormone	Neuropeptide EI	EIGDEENSAKFPI-amide	3	2	1446.71	1446.683	21	1.11	1.24	1.12
ProSAAS	Big LEN	LENPSPQAPARRLLPP	3	1	1754.99	1754.979	8	1.06	1.08	1.15
Proopiomelanocortin	Alpha-MSH	Ac-SYSMEHFRWGKPV-amide	3	1	1663.82	1663.790	16	1.04	1.31	1.29
Protachykinin A	72-94	(DADSSVEKQVALLKALYGHGQIS)	4	3	2428.27	2428.260	5	0.89	1.19	0.96
Hemoglobin alpha	2-13	VLSGEDKSNIKA	3	3	1259.67	1259.672	0	1.15	1.05	1.05
Proenkephalin	114-133	MDELYPMEPEEEANGGEILA	2	1	2236.00	2235.955	19	0.99	0.80	0.69
Mast cell protease -3 or -4 (MMCP-3 or MMCP-4)	N-terminus	IIGGVESRPHSRPY	4	1	1566.84	1566.827	9	0.97	1.24	1.15
Procorticotropin Releasing Factor	127-142	GAEDALGGHQGALERE	3	1	1608.78	1608.749	20	1.02	1.04	0.87
Synuclein, alpha or beta	N-terminus	Ac-MDVFMKGL	2	1	981.47	981.466	1	0.82	1.09	0.84
Hemoglobin beta	2-13	VHLTDAEKAAVS	3	2	1239.61	1239.646	-30	1.17	1.24	1.13
Procorticotropin Releasing Factor	127-142	GAEDALGGHQGALERE	3	1	1608.78	1608.749	20	1.25	1.28	1.08
Thymosin beta-4	C-terminus	PLPSKETIEQEKQAGES	3	2,3	1869.94	1869.932	6	1.03	1.12	0.85
ATP synthase subunit e	N-terminus	VPPVQVSPLIKFGRYS	3	2	1786.04	1786.014	14	1.13	1.04	1.02

Abbreviations are defined in Tables 5 and 6. po, Peak overlap with a contaminating peptide (can't quantify).

Analysis of the C-terminal residue (i.e. the P1' position) of the 16 best substrates of CPA4 identified from the brain peptidome showed a predominance of Leu and Phe, with 14 of the 16 peptides containing one of these two amino acids (Table 5 and Figure 7). The other two peptides in this group had C-termini of Tyr or Val. Similar analysis of the 24 weak substrates of CPA4 showed these 4 amino acids to be present in the C-terminal position, in addition to Ile, Met, Ala, Thr, and His (Table 6 and Figure 7). Analysis of the products of CPA4 focused on the downstream residue that was removed; of the 10 products that were identified, 6 required removal of a Leu, and the other four required removal of an Ile, Val, Tyr, or Ala (Table 7). Analysis was also performed on the peptides that were neither substrates nor products of CPA4 (Table 8 and Figure 7). Of the 50 peptides in this category, only a small number had C-terminal Leu or Phe residues; the majority of these peptides contained C-terminal residues that were basic (Lys, Arg), acidic (Asp, Glu), polar (Ser, Asn, Gln), or otherwise not represented in the best substrate group (Ala, His, Pro).

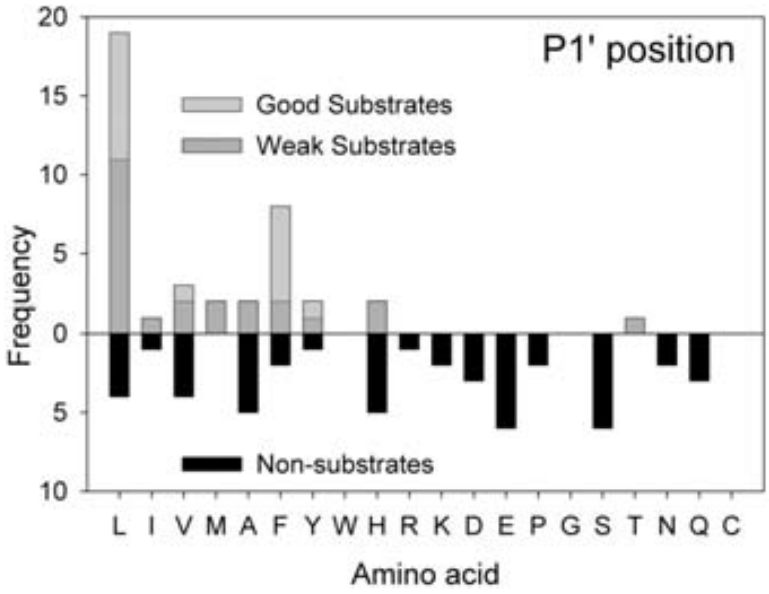

Figure 7. Analysis of the C-terminal residue (P1') of peptides detected using the quantitative peptidomics technique. The number of times each amino acid was present in the C-terminal position was determined for good substrates, weak substrates, and non-substrates, as defined in the Fig. 6 legend.

The finding that some, but not all of the peptides with C-terminal aliphatic and aromatic residues could be cleaved by CPA4 under the conditions tested suggested that additional residues contributed to the preference of this enzyme for particular

peptides. To analyze this, we examined the P1 residue in all peptides with a C-terminal hydrophobic residue. Altogether, there were 62 peptides with a C-terminal Leu, Ile, Val, Met, Ala, Phe, Tyr, His, or Thr; of these, 40 were found to be substrates and 22 were not cleaved under the conditions tested. The majority of the substrates contained an aliphatic residue (Leu, Ile, Val, Met, Ala), aromatic (Phe, Tyr), or basic (Arg, Lys) residue in the P1 position (Fig. 8). Furthermore, none of the substrates contained an acidic residue (Asp, Glu), a Pro, or a Gly in the P1 position. In contrast, only 2 of the non-cleaved peptides had P1 aliphatic, aromatic, or basic residues (Fig. 8). Furthermore, many of these non-cleaved peptides had acidic or alpha-helix-breaking residues in the P1 position. Thus, it appears that while the C-terminal residue plays a critical role in determining specificity of CPA4, the P1 residue also contributes to the cleavage specificity.

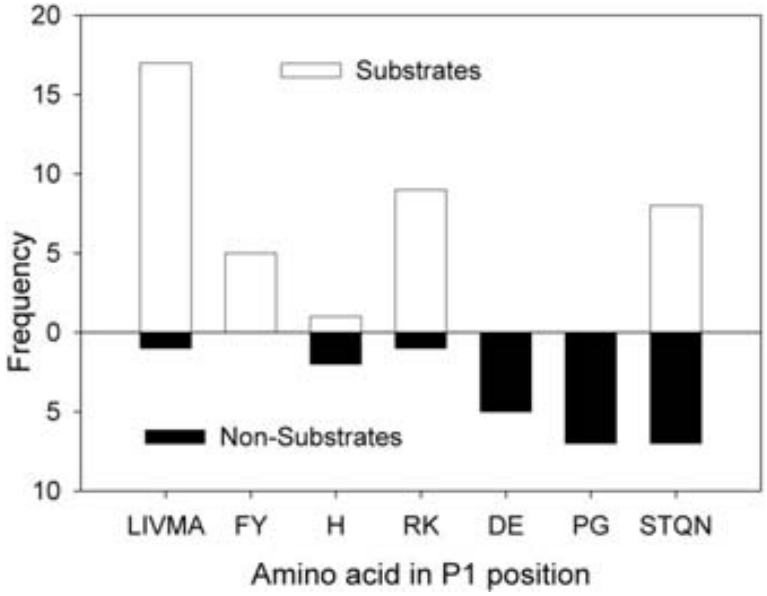

Figure 8. Analysis of the P1 residue of peptide containing hydrophobic C-terminal residues. For this analysis, only peptides that were substrates or potential substrates were considered, based on the analysis shown in Fig. 7. Those peptides with acidic, basic, or polar C-terminal residues were excluded. This analysis included 40 substrates (good and weak substrates were combined) and 22 non-substrates with similar C-terminal residues as the substrates (i.e. L, I, V, M, A, F, Y, H). The frequency of the various amino acids in the P1 position of each group were determined and combined into groups based on amino acid type; aliphatic (L, I, V, M, A), aromatic (F, Y), basic (R, K), acidic (D, E), alpha-helix breakers (P, G), and polar (S, T, Q, N). His was placed into its own group, and there were no data for Cys or Trp as these amino acids were not present in the P1 position in any of the peptides analyzed

2.4. DISCUSSION

Summary of cleavage specificity of CPA4

The major finding of the present study is the cleavage specificity of CPA4, which has not been previously examined. Because of the potential link between CPA4 and some types of cancer (described in Introduction), it is important to determine the substrate specificity of this enzyme so that predictions can be made regarding its biological role(s). Based on the present results, CPA4 is clearly a CPA-like enzyme with specificity for C-terminal aliphatic and aromatic residues, consistent with the substrate binding pocket predicted from analysis of the crystal structure of CPA4 (Bayes et al., 2007; Garcia-Castellanos et al., 2005; Pallares et al., 2005). The three types of substrates used in the present study (small chromogenic substrates, selected synthetic peptides, and a mixture of brain peptides) provide complementary information. While the small chromogenic substrates are ideal for kinetic analysis of k_{cat} and K_m values, these are not natural substrates and the chromogenic group can potentially affect the cleavage specificity. Synthetic peptides are useful to study precise substrate/product relationships, but are limited by the expense and the time to analyze each peptide. The quantitative peptidomics approach provides a great deal of information regarding which peptides are substrates, which are products, and which are not cleaved by CPA4. However, this approach does not provide kinetic information, and it is not possible to follow individual reactions as can be done with the synthetic peptides. Therefore, these three approaches have overlapping strengths and collectively provide a more complete picture than a single approach.

Taken together, the optimal C-terminal residue for CPA4 cleavage is Leu, Ile, Val, Met, Tyr, or Phe (Fig. 9). Other acceptable amino acids in this position include Ala, His, Thr, or Trp. CPA4 does not cleave C-terminal Pro, basic residues (Arg, Lys), acidic residues (Asp, Glu), or polar residues such as Ser, Asn, or Gln. Although Thr can be considered a polar residue, it is less polar than Ser. No information on Gly or Cys in the C-terminal position is available. The P1 position also has an influence on cleavage specificity, with aliphatic, aromatic, and basic residues favourable in this position (Fig. 9). Certain amino acids have a negative influence in the P1 position, such as acidic

residues, Pro, and Gly (Fig. 9). No information on Trp or Cys is available for the P1 position.

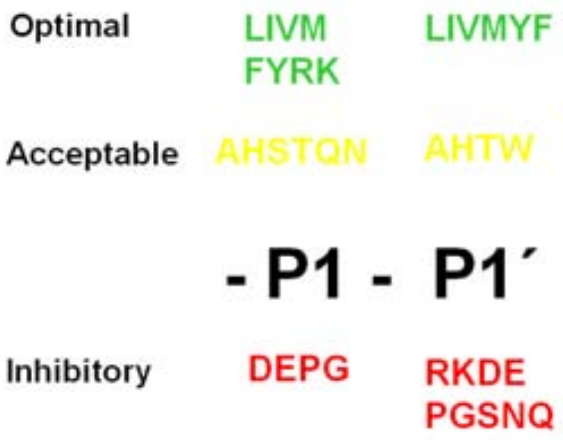

Figure 9. Summary of cleavage specificity of CPA4, based on results from synthetic chromogenic substrates, synthetic peptides, and endogenous mouse brain peptides. Those amino acids considered most favorable for cleavage by CPA4 are listed in the top row, and include L, I, V, M, Y, and F in the P1' position and these same residues plus R and K in the P1 position. Residues that are acceptable in these positions are listed in the middle row, and those residues that appear to be inhibitory for cleavage are listed in the bottom row.

Comparison among CPAs

The FA-Phe-Xaa series of substrates offer higher values of k_{cat}/K_m values than other small synthetic substrates, allowing measurements of CPA2 towards substrates with small hydrophobic C-terminal amino acids, which has not previously been reported. In most cases, the improvement of catalytic efficiency for CPA2 towards FA-Phe-Xaa substrates versus Cbz-$(Gly)_n$-Xaa substrates is due to lower K_m values but in some cases an increase in k_{cat} values also contributes to the difference (Reverter et al., 1997). This effect may be explained by the different amino acids in the penultimate position of these substrates (Phe versus Gly); if CPA2 is similar to CPA4 as described above from the peptidomics analysis, Phe is much preferred over Gly in the P1 position. In addition, the FA-blocking group is known to enhance the CP activity (Peterson et al., 1982). The kinetic data for CPA2 fully agree with that reported by Reverter *et al*., showing a preference for FA-Phe-Trp and FA-Phe-Phe (Reverter et al., 1997). Compared to FA-Phe-Trp, the k_{cat}/K_m values are 85-fold and 335-fold lower for FA-Phe-Leu and FA-Phe-Ile, respectively. In contrast with these previous reports we were able to estimate k_{cat}/K_m values for FA-Phe-Met and FA-Phe-Val that are approximately 1000 times lower than the one for FA-Phe-Trp.

Although CPA1 has been well studied over the past 50 years, particularly bovine CPA1 (originally referred to as CPA in the literature), the kinetic constants for human CPA1 have not been previously reported. The preference of CPA1 for Phe in the C-terminal position is more marked than that of CPA4, when the k_{cat}/K_m values for different substrates are compared. The k_{cat}/K_m ratio of CPA1 for FA-Phe-Phe is approximately two-fold greater than that of FA-Phe-Leu and FA-Phe-Val, four times higher than that of FA-Phe-Ile and ten-fold higher than the corresponding values for FA-Phe-Trp, FA-Phe-Met and FA-Phe-Ala. When k_{cat}/K_m values are compared between the three enzymes, CPA1 and CPA2 appear to be more efficient than CPA4 in the removal of C-terminal residues from peptides. For example, k_{cat}/K_m values for FA-Phe-Phe are approximately the same for CPA1 and CPA2, and four times lower for CPA4.

A comparison between CPA4 and CPA6 activities towards the same group of peptides show several differences. CPA6 cleaves both Leu and His from the C-terminus of angiotensin I and des-Asp^1 angiotensin I, whereas CPA4 only removes the Leu and not the His under the conditions described in Table 3. However, CPA4 can cleave both the Leu and His in angiotensin when higher enzyme and peptide concentrations are used (data not shown). Furthermore, FA-Phe-His is cleaved by CPA4 at a low rate and in the peptidomics study two peptides which contain C-terminal His residues were found to be weak substrates of CPA4. Therefore, CPA4 is able to cleave His from the C-terminus of some substrates. In addition to CPA4 and CPA6, previously it was found that CPA1 (Garcia et al., 1997) and CPA3 (Lundequist et al., 2004) can cleave C-terminal His residues. Although His is commonly considered to be a basic residue, at neutral pH values it will be mainly unprotonated and resemble an aromatic/aliphatic residue. Thus, it is not surprising that this residue is cleaved by CPA-like enzymes.

The substrate specificity of CPA4 determined in the present study fits very well with predictions made from structural analysis of this enzyme in complex with a hexapeptide (Bayes et al., 2007), and with results from structural analysis of many CPAs and Bs (Gomis-Ruth, 2008). The specificity of CPA4 for C-terminal hydrophobic residues such as Leu, Ile, Val, Met, Tyr, and Phe, and weaker hydrolysis of Trp, matches more closely the specificity of CPA1 rather than CPA2. The S1' specificity pocket is delimited by the side chains in positions 194, 203, 207, 243, 247, 250, 253-255, and

268 (using the numbering system of bovine CPA1). The residue in position 255 (Ile for CPA4) is considered to be the major determinant of substrate specificity for carboxypeptidases of the CPA/CPB subfamily (Grishin et al., 2008). In CPA4 the residues in positions 203, 247, and 250 (numbering system of CPA1), which are Met, Val, and Ala respectively, also contribute to the hydrophobicity of the S1' binding pocket. A comparison of the amino acids that shape the S1' pocket between CPA1, CPA2 and CPA4 explains the strong preference of hCPA2 for bulkier hydrophobic amino acids in comparison to CPA1 and CPA4 (Table 9). The presence of Ala in position 268 in CPA2, in contrast to the Thr present in CPA1 and CPA4, makes the specificity pocket of CPA2 larger and more hydrophobic, explaining the specificity of CPA2 for large amino acids like Trp (Faming et al., 1991; Garcia-Saez et al., 1997). The aminoacid in position 253 may also be responsible for this difference. The preference of CPA1 for C-terminal Phe is more marked than that of CPA4; this preference can be explained by the presence of a more hydrophobic environment in the S1' pocket of CPA1 provided by Ile243 and Ile247, compared to the Thr and Val in similar positions in CPA4

Table 9. Comparison of amino acids that shape the S1' pocket in different CPs.

	Homologous amino acid residues								
Position in bovine CPA (active form)	194	203	207	243	247	250	253	255	268
Bovine CPA	Ser 304	Leu 313	Gly 317	Ile 353	Ile 357	Ala 360	Gly 363	Ile 365	Thr 378
Human CPA4	Asp 306	Met 315	Gly 319	Thr 355	Val 359	Ala 362	Ser 365	Ile 367	Thr 380
Rat CPA4	Asp 306	Met 315	Gly 319	Thr 355	Val 359	Ala 362	Ser 365	Val 367	Thr 380
Human CPA1	Ser 304	Met 313	Gly 317	Ile 353	Ile 357	Ala 360	Ser 363	Ile 365	Thr 378
Rat CPA1	Ser 304	Leu 313	Gly 317	Ile 353	Ile 357	Ala 360	Ser 363	Ile 365	Thr 378
Human CPA2	Thr 302	Met 311	Gly 315	Ile 351	Ile 355	Ala 358	Gly 361	Ile 363	Ala 376
Rat CPA2	Thr 302	Met 311	Gly 315	Ile 351	Ile 355	Ala 358	Gly 361	Ile 363	Ala 376

The substrate binding pocket residues are indicated, which by convention in the field are numbered according to their position within the active form of bovine CPA, after removal of the signal peptide and the pro peptide. The corresponding residues in human and rat CPA1, CPA2, and CPA4 are indicated, along with their position within the precursor protein (preproCPA).

In the present study, the S1 subsite was also found to contribute to the specificity of CPA4 (Fig. 9); critical residues in this subsite are the Tyr in position 198 and Ser in position 199 (Bayes et al., 2007). Other residues known to contribute to the shape of the S1 subsite of CPAs include Val 247, Tyr248, Phe279 and Glu270, which form a mainly hydrophobic pocket that may explain the preference of CPA4 for substrates with hydrophobic amino acids in the P1 position. The preference of CPA4 for positively charged amino acids over negatively charged amino acids in the P1 position could be explained by interactions between the side chain of the P1 residue with the polar portion of Tyr198 and Ser199 in the S1 binding site (Gomis-Ruth, 2008).

Potential function of CPA4

The finding that PCPA4 is secreted from cells in a soluble form and, once activated, has a neutral pH optimum suggests a function in the extracellular environment rather than within the secretory pathway. Although trypsin was used to activate PCPA4 in the present study, this is not likely to represent the physiological activating enzyme because trypsin is pancreatic while CPA4 has a broader distribution. There are a number of trypsin-like enzymes expressed in brain and other tissues where PCPA4 is found (Wang et al., 2008). Based on modelling and structural considerations, the conversion of PCPA4 into active CPA4 requires cleavage at a single Arg in the sequence EGQER-SS. It is possible that the activation of PCPA4 may be an important control point that regulates the active form of the enzyme.

The distribution of CPA4 provides important clues as to the function of this enzyme. Huang et al (Huang et al., 1999) found CPA4 mRNA to be expressed at high levels in different prostate cancer cell lines and at extremely low levels in some normal human tissues including prostate and pancreas. CPA4 mRNA was induced in prostate cancer cell lines upon treatment with 10 mM sodium butyrate indicating that its expression is associated with *in vitro* differentiation. It was also reported that CPA4 expression is associated with hormone-regulated tissues. Through an imprinting analysis, Kayashima et al (Kayashima et al., 2003) found that CPA4 mRNA was also highly expressed in fetal tissues that are growing or differentiating with the transcriptionally active state of chromatin. Using RT-PCR Bentley et al (Bentley et al.,

2003) detected CPA4 expression in a wide variety of tissues including placenta, testis, uterus, heart, brain, intestine, kidney, and muscle. In mouse brain, *in situ* hybridization showed relatively high expression of CPA4 mRNA in the olfactory bulb, specifically in the granular and mitral layers (Lein, 2007), in an expression pattern similar to CPA6 mRNA (Fontenele-Neto et al., 2005). A US patent using TaqMan analysis of human tissues reported expression of CPA4 in kidney, spinal cord, and brain, with moderate levels in cortex and considerably higher levels in hypothalamus, especially in the arcuate and paraventricular nuclei (White, 2004). Taken together, it is clear that CPA4 mRNA is present in many tissues.

The association of CPA4 mRNA levels with differentiation and/or cancerous cells and tissues raises the possibility that this enzyme plays an important role in cleaving a factor or factors that are critical to the differentiated state. Many secreted factors are known to function as growth factors or to otherwise enable tissues to grow and metastasize (such as secreted proteases) and CPA4 may contribute to these processes. Some of the peptides found to be substrates in the present study include neurotensin, granins like chromogranin A and B and secretogranin II, and opioid peptides such as Met- and Leu-enkephalin. The cleavage performed by CPA4 would eliminate the binding of these peptides to their receptors. While enkephalin is a well-known neuropeptide that functions in analgesia and the reward pathway, additional roles for enkephalin include synaptogenesis and neurite outgrowth (Davila-Garcia and Azmitia, 1990; Escobar et al., 1997). Some granin-derived peptides have been postulated to participate as regulators of tissue repair, inflammatory reaction and innate immunity (Helle et al., 2007). Additional roles for some of these products include the stimulation of proliferation in tumor cells (Mosca et al., 2005) and invasive potential of prostate cancer cells partially through enhancement of cell mobility (Nagakawa et al., 2001). Similarly, neurotensin is also a well-studied neuropeptide that has recently been proposed to contribute to the progression of several types of cancer including prostate, breast, and pancreatic cancer (Dupouy et al., 2009; Myers et al., 2009; Vias et al., 2007). Because the enkephalins, secretogranins and neurotensin are secreted from some of the same tissues where CPA4 is also secreted, it is likely that this enzyme plays a role in the termination of the action of these peptides, as well as many others. A role

for CPA4 in the inactivation of peptides that function in cell proliferation would potentially provide a mechanism to account for the previous studies linking CPA4 with aggressiveness of some forms of cancer. Additional research efforts are needed to characterize and understand the biological effects of this enzyme and its SNP variant(s) in the tumoral process.

Chapter 3

Proteome-derived peptide libraries to study the substrate specificity profiles of carboxypeptidases

3.1. INTRODUCTION

Carboxypeptidases (CPs) catalyze the release of C-terminal amino acids from proteins and peptides (Arolas et al., 2007; Vendrell and Avilés, 1999), and are grouped according to the chemical nature of their catalytic site. Accordingly, there are three types of carboxypeptidases: metallocarboxypeptidases (MCPs), serine carboxypeptidases (SCPs) and cysteine carboxypeptidases. CPs can also be classified based on their substrate specificity; CPs that prefer hydrophobic C-terminal amino acids (A-like MCPs or C-type SCPs), those that cleave C-terminal basic residues (B-like MCPs or D-type SCPs), those that recognize substrates with C-terminal aspartate or glutamate residues, and other CPs that display a broad substrate specificity (Lyons and Fricker, 2011a; Mortensen et al., 1999).

CPs were initially considered as degrading enzymes associated with protein catabolism. However, accumulating evidence demonstrates that some CPs are (more) selective and play key roles in controlling various biological processes (Arolas et al., 2007; Rodriguez de la Vega et al., 2007). Angiotensin-converting enzyme 2 (ACE2), a MCP homologue of angiotensin-converting enzyme (ACE) that belongs to the M2 family of proteolytic enzymes according to the MEROPS classification, is a potent negative regulator of the renin-angiotensin system and plays a key role in maintaining blood pressure homeostasis. ACE2 cleaves off a C-terminal phenylalanine thereby converting angiotensin II to the heptapeptide angiotensin-(1-7), a peptide hormone that opposes the vasoconstrictor and proliferative actions of angiotensin II (Kuba et al., 2010). Cathepsin A, a lysosomal SCP, is also believed to function in blood pressure regulation, in this case through its action against vasoactive peptides like endothelin-1 or angiotensin I (Pshezhetsky and Hinek, 2009). Human carboxypeptidase A4 (hCPA4), a MCP from the M14 family, presumably functions in neuropeptide processing and was linked to prostate cancer aggressiveness (Tanco et al., 2010).

Besides their biological importance, CPs are also exploited in biotechnological and biomedical applications. Carboxypeptidase B (CPB) for instance, is a M14 MCP used for manufacturing recombinant human insulin. Recombinant preproinsulin is enzymatically

processed *in vitro* by pancreatic trypsin and carboxypeptidase B to generate the active insulin form (Son et al., 2009). Further, carboxypeptidase digestion has been used for determining the C-terminal sequence of purified proteins or peptides. The most popular CPs being the SCPs C, P and Y (Bergman et al., 2003). In addition, the food industry uses different SCPs to process protein products in order to reduce their bitter taste (Arai et al., 1970; Raksakulthai and Haard, 2001; Umetsu et al., 1983).

Identifying a protease's specificity and its natural substrates provides key information to understanding the molecular role of proteases (Barrios and Craik, 2002; Gosalia et al., 2005). Moreover, determination of a protease's specificity also provides a framework for the design of selective probes and potent and selective inhibitors (Diamond, 2007). While several factors impact on substrate selection, a key factor is the complementarity of a protease binding site with specific substrate side-chains.

Several approaches for determining protease substrate specificity based on peptide libraries have been developed, including substrate phage/bacterial display libraries, peptide microarrays, positional-scanning peptide libraries, mixture-based peptide libraries and proteome-derived peptide libraries (auf dem Keller and Schilling, 2010). The latter were more recently introduced by Schilling *et al.* (Schilling and Overall, 2008) and make use of natural peptide libraries generated by proteolysis of a model proteome using a specific protease (e.g., trypsin, chymotrypsin). Such peptide libraries are subsequently digested by a protease of interest and the resulting neo-N-terminal products are enriched and identified following LC–MS/MS analyses. This technology allows profiling of the substrate specificity of endoproteases and aminopeptidases. However, viewing the fact that only C-terminal cleavage products are isolated by this method, it cannot be used to study CPs since their resulting primed site cleavage products are typically only a single amino acid and thus are not compatible for subsequent LC-MS/MS based identification.

We here exploited the COFRADIC technology (Gevaert et al., 2007) and developed a carboxypeptidase peptide library assay that was used to determine the substrate specificity profile of 4 selected human carboxypeptidases. Given that MCPs are the most studied and thus a highly relevant group of CPs, the human carboxypeptidases A4

(hCPA4), A2 (hCPA2) and A1 (hCPA1) were used as model CPs. Two different peptide libraries, created using chymotrypsin or metalloendopeptidase Lys-N as peptide library generating proteases, were used to extensively profile the proteolytic substrate specificities of these MCPs. In addition, we profiled the substrate preferences for the yet uncharacterized mast cell carboxypeptidase (MC-CPA or mCPA3).

3.2. EXPERIMENTAL SECTION

Cell Culture

Human K-562 cells were obtained from the American Type Culture Collection (CCL-243, ATCC, Manassas, VA, USA), and cultured in GlutaMAX™ containing RPMI 1640 medium (Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal calf serum (Invitrogen), 100 units/ml penicillin (Invitrogen) and 100 µg/ml streptomycin (Invitrogen). Cells were maintained at 37°C in a 5% CO_2 humidified incubator.

Protein Production and Purification

The human carboxypeptidases A1 (hCPA1), A2 (hCPA2) and A4 (hCPA4) were obtained as recombinant proteins using the pPIC9 expression vector and the methylotrophic yeast *Pichia pastoris* as an expression host. Enzyme purifications were performed as described previously (Pallares et al., 2005; Pallares et al., 2008; Reverter et al., 1997). These enzymes were purified in their zymogen form and the active enzymes were obtained through tryptic activation (at a 1/50 (w/w) ratio) for 1 h at room temperature. The resulting mature and activated enzymes were subsequently purified by anion-exchange chromatography (TSK-DEAE 5PW) on a FPLC-Äkta system using a linear salt gradient (ranging from 0 to 30% of 0.4 M ammonium acetate in 20 mM Tris-HCl, pH 9.0). Eluted fractions were analyzed by SDS-PAGE, and the purest fractions containing the enzyme were pooled, desalted, and concentrated to 1 mg/ml by Amicon centrifugal filter devices (Ultra 0.5 ml 10 kDa MWCO columns (Millipore, Billerica, MA, USA)).

Mouse CPA3 (mCPA3) was purified from mouse bone marrow-derived mast cells (BMMCs) (kindly provided by Dr Gunnar Pejler, Swedish University of Agricultural Sciences) using a two-step purification procedure. Mast cells were lysed in 10 mM Tris-HCl, pH 7.4, 4 M NaCl, 0.1% PEG 3350 supplemented with a Complete EDTA-free Protease Inhibitor Cocktail Tablet (Roche Diagnostics) (buffer A) for 30 minutes at 4°C. The lysate

was subsequently centrifuged and the supernatant was diluted twenty-fold in 50 mM Tris-HCl, pH 7.4 and 100 mM NaCl, supplemented with a Complete EDTA-free Protease Inhibitor Cocktail Tablet (buffer B). The diluted extract was loaded on a Heparin HyperD® M column (Pall Biosepra, Cergy-Saint-Christophe, France) equilibrated with buffer B. mCPA3 was eluted using 5 volumes of buffer A. The eluate was concentrated on Amicon centrifugal filters devices, and fractionated on a Superdex 75 column in 25 mM Tris-HCl, pH 7.4, 1 M NaCl and 0.025% PEG 3350. The purest eluted fractions (analyzed by SDS-PAGE) showing the highest activity towards N-(4-methoxyphenylazoformyl)-Phe-OH (Bachem, Bubendorf, Switzerland) were pooled and concentrated by Amicon centrifugal filters devices (Ultra 0.5 ml 10 kDa MWCO columns, Millipore).

Preparation of Proteome-derived Peptide Libraries

Proteome-derived peptide libraries were generated from human K-562 cell extracts. Cells were repeatedly (3x) washed in digestion buffer (50 mM NH_4CO_3, pH 7.9) and re-suspended in this buffer at $2x10^7$ cells per ml. Then, these cell suspensions were subjected to three rounds of freeze-thaw lysis and the lysate was cleared by centrifugation for 10 min at 16,000 *g* at 4°C. To prepare the chymotryptic and metalloendopeptidase Lys-N proteome-derived peptide libraries, the lysates were respectively digested for 4 h at 37ºC using sequencing-grade chymotrypsin (Promega, Madison, WI, USA) at an enzyme/substrate ratio of 1/200 (w/w) or recombinant Lys-N (1/85, w/w) for two hours at 37ºC (U-Protein Express BV, Utrecht, The Netherlands). To stop proteolytic digestion, acetic acid was added to a 4% final concentration.

Peptide Library-based Carboxypeptidase Assay

For the chymotryptic peptide library and to prevent oxidation of methionines between the primary and secondary RP-HPLC runs, methionines were oxidized before the primary run. The methionine oxidation reaction proceeded in the injector compartment by transferring 20 µl of a freshly prepared aqueous 3% H_2O_2 solution to a vial containing 90

µl of the acidified peptide mixture (final concentration of 0.54% H_2O_2). This reaction proceeded for 30 min at 30°C after which the sample was immediately injected onto the RP-HPLC column. No prior methionine oxidation was performed when the Lys-N peptide libraries were assayed (see Results). From these mixtures, 100 µl (equivalent to ~350 µg of digested proteins) was injected onto a RP-column (Zorbax 300SB-C18 Narrow Bore, 2.1 mm internal diameter (I.D.) x 150 mm length, 5 µm particles; Agilent Technologies) for the first RP-HPLC run. Following 10 min isocratic pumping with solvent A (10 mM ammonium acetate in water/acetonitrile (98/2, v/v), pH 5.5), a gradient was started of 1% solvent B (10 mM ammonium acetate in water/acetonitrile (30/70, v/v), pH 5.5) increase per minute. The column was then run at 100% solvent B for 5 min, switched to 100% solvent A and re-equilibrated for 20 min with solvent A. The flow was kept constant at 80 µl/min using Agilent's 1100 series capillary pump with an 800 µl/min flow controller. Twenty fractions of 2 min intervals (from 20 to 60 min after sample injection) and twenty-six fractions (from 20 to 72 min after sample injection) were collected for the chymotryptic library and the Lys-N library respectively.

Each peptide fraction was dried and re-dissolved in 20 µl of CP-supplemented assay buffer (50 mM Tris-HCl, pH 8.0 and 100 mM NaCl prepared with 90% $H_2{}^{18}O$ water (Cambridge Isotope Labs, Andover, MA, USA) containing 7.3 units of CP per ml (which is approximately equivalent to a 10 nM CP concentration)). Note that one unit of CP activity was defined as the amount of enzyme hydrolyzing 1 nmol of N-(4-methoxyphenylazoformyl)-L-phenylalanine substrate per min at 37°C. CP hydrolysis was allowed to proceed for 2 h at 37°C and stopped by addition of 33 µl of 4% acetic acid in solvent A. All 20 (chymotryptic library) or 26 (Lys-N library) samples were reloaded on the same RP-column and separated using identical conditions. Per sample, twenty secondary fractions of 1 minute wide were collected in a time interval ranging from 21 to 1 minute before the fraction collection interval (but eluting the earliest at 10 min (start of the gradient)) used for the primary fraction. These peptide fractions were dried and, secondary fractions eluting 4 min apart were pooled by re-dissolving these in a final volume of 20 µl of 2 mM TCEP and 2% acetonitrile, similar to a pooling strategy described

previously (Staes et al., 2011). In total, 40 peptide fractions per setup were subjected to LC-MS/MS analysis.

LC-MS/MS Analysis

LC-MS/MS analysis was performed using an Ultimate 3000 RSLC nano LC-MS/MS system (Dionex, Amsterdam, The Netherlands) in-line connected to a LTQ Orbitrap Velos (Thermo Fisher, Bremen, Germany). 2 µl of the sample mixture was first loaded on a trapping column (made in-house, 100 µm I.D. x 20 mm length, 5 µm Reprosil–Pur Basic-C18-HD beads, Dr. Maisch, Ammerbuch-Entringen, Germany). After back-flushing from the trapping column, the sample was loaded on a reverse-phase column (made in-house, 75 µm I.D. x 150 mm length, 3 µm C18 Reprosil–Pur Basic-C18-HD beads). Peptides were loaded with solvent A' (0.1% trifluoroacetic acid in 2% acetonitrile) and were separated with a linear gradient from 98% of solvent A'' (0.1% formic acid in 2% acetonitrile) to 50% of solvent B' (0.1% formic acid in 80% acetonitrile) with a linear gradient of a 1.8% solvent B' increase per minute at a flow rate of 300 nl/min followed by a steep increase to 100% of solvent B'. The Orbitrap Velos mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS acquisition for the ten most abundant peaks in a MS spectrum. Full scan MS spectra were acquired in the Orbitrap at a target value of 1E6 with a resolution of 60,000. The ten most intense ions were then isolated for fragmentation in the linear ion trap, with a dynamic exclusion of 20 s. Peptides were fragmented after filling the ion trap at a target value of 1E4 ion counts. From the MS/MS data in each LC run, Mascot Generic Files were created using the Mascot Distiller software (version 2.3.2.0, Matrix Science, www.matrixscience.com/Distiller.html). While generating these peak lists, grouping of spectra was allowed with maximum intermediate retention time of 30 s and maximum intermediate scan count of 5. Grouping was done with a 0.005 Da precursor tolerance. A peak list was only generated when the MS/MS spectrum contained more than 10 peaks. There was no de-isotoping and the relative signal-to-noise limit was set at 2. The generated MS/MS peak lists were then searched with Mascot using the Mascot Daemon interface (version 2.3, Matrix Science). The Mascot search parameters were set as follows. Searches were performed in the Swiss-Prot

database with taxonomy set to human (either the 2011_05 or 2011_06 UniProtKB/Swiss-Prot database containing respectively 20,286 and 20,312 human protein sequence entries were used). Single ^{18}O modification of peptide C-termini, acetylation of protein N-termini and pyroglutamate formation of N-terminal glutamine were set as variable modifications. Methionine oxidation to methionine-sulfoxide was set as fixed modification for the chymotryptic library assays and as a variable modification for the Lys-N library assays. For the chymotryptic library the enzyme was set to "none" (since up to 3 missed cleavages could be detected), whereas for the Lys-N library a semi-Lys-N/P (semi Lys-N specificity with lysine-proline cleavage allowed) was set as enzyme allowing for one missed cleavage. The mass tolerance on the precursor ion was set to 10 ppm and on fragment ions to 0.5 Da. The peptide charge was set to 1+, 2+ or 3+ and instrument setting was put on ESI-TRAP. Only peptides that were ranked one and scored above the threshold score, set at 95% confidence, were withheld. According to the method described by Käll *et al.* (Kall et al., 2008), the false discovery rate was estimated and was found <2% at the spectrum level and <3% at the peptide level. Identified MS/MS spectra are available on-line in the PRoteomics IDEntifications database (PRIDE) (Martens et al., 2005) with accession codes 26851 to 26859, 27075 and 27076.

Comparative Modeling of mCPA3

The protein structure modeling of the active domain of mCPA3 was obtained from computational prediction using the I-TASSER server (Roy et al., 2010; Zhang, 2007). The confidence level of the models was evaluated by the C-score (1.64 for the model) and the TM-score (0.94±0.05) (Roy et al., 2010; Zhang, 2007). The model with a hexapeptide in the active site was derived from the structure of hCPA4 with PDB ID code 2PCU (Bayes et al., 2007) in which the amino acid in the P1' position was mutated to Trp. Molecular visualizations were performed using PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

3.3. RESULTS

Proteome-derived Peptide Library Generation and Carboxypeptidase Assay Workflow

To profile the substrate specificity of CPs, we opted for the use of proteome-derived peptide libraries which should allow comprehensive analyses of protease preferences (Schilling and Overall, 2008). Fig. 1 and 2 show the peptide library generation strategy and the assay workflow respectively. In this approach, proteome preparations from human K-562 cells were used as starting material, and peptide libraries were generated by proteolysis using either chymotrypsin (cleaves C-terminal of Tyr, Phe, Leu, Trp and Met) or metalloendopeptidase Lys-N (cleaves N-terminal of Lys). The resulting peptide mixtures were subsequently used to assay the substrate specificity profiles of carboxypeptidases (Fig.1).

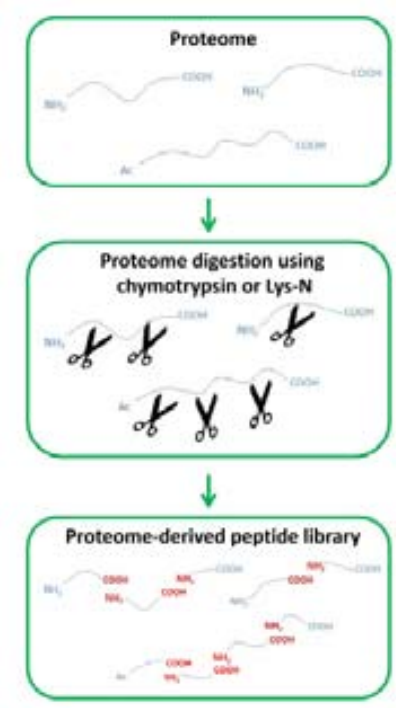

Fig. 1. Proteome-derived peptide library generation. Peptide library generation workflow. Peptide libraries are generated from whole proteomes by digestion using a peptide library-generating protease (e.g., chymotrysin or Lys-N). Ac- is indicative of N-terminal protein acetylation in vivo

Combined fractional diagonal chromatography (COFRADIC) relies on the principle of diagonal reversed-phase chromatography to select a representative set of peptides (Gevaert et al., 2007). A modification step (chemical or enzymatic reaction) is performed between two consecutive, identical chromatographic separations to infer different chromatographic properties to modified peptides, thereby allowing for their isolation. Here, CPs act as peptide modifiers, reducing the length of their respective substrates (by cleaving off a C-terminal residue) and thus altering their chromatographic properties.

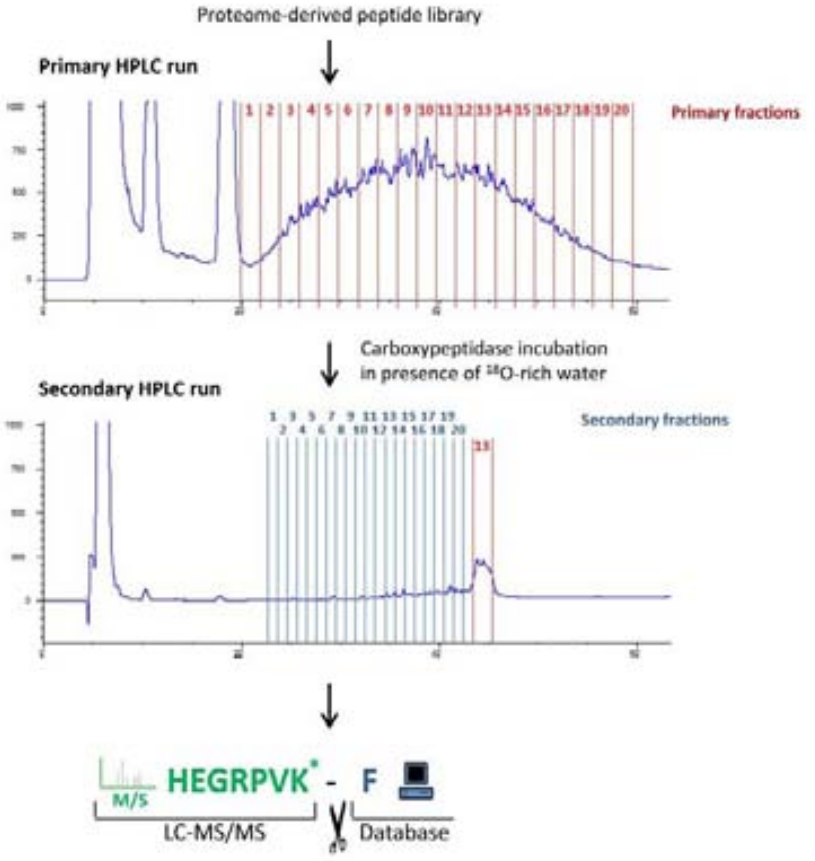

Fig. 2. Chymotryptic proteome-derived peptide library-based carboxypeptidase substrate screen. The peptide library is separated by reverse-phase HPLC (primary COFRADIC run) and collected in 20 primary fractions. Each dried fraction is reconstituted in a CP compatible buffer prepared in 95% 18O-rich water and incubated with the CP of interest (i.e. hCPA4 in this case). These altered primary fractions are then re-separated on the same column using identical conditions (secondary COFRADIC run). The shifted carboxypeptidase products are collected in 1 min fractions (secondary fractions) starting from 21 to 1 min before the elution time of the primary collection interval containing the unaltered peptides. The CP products contain a C-terminal 18O-label (here indicated with an asterisk) and can therefore be distinguished from their C-terminal unmodified counterparts following LC-MS/MS analyses and database searching. The cleaved C-terminal amino acid is subsequently inferred from the database.

The generated peptide library was first fractionated by RP-HPLC, after which each primary fraction was subjected to CP treatment in assay buffers prepared with ^{18}O-rich water. Then, each CP-modified primary fraction was re-separated using identical RP-HPLC conditions. Fig. 2 shows the chromatograms of typical primary and secondary COFRADIC runs (here the primary fraction eluting between 44 and 46 min was treated with hCPA4). Judged from initial tests with hCPA4, a hydrophilic shift was expected for CP substrates (Fig. 3). Consequently, CP peptide products were collected in 20 fractions of 1 minute wide intervals, each time starting 21 min before the elution of CP-unmodified peptides. Given that proteolysis is the hydrolytic cleavage of peptide bonds, we exploited carboxypeptidase-induced ^{18}O-labeling at the C-terminus of CP-generated peptides. Following LC-MS/MS analysis and database searching, this ^{18}O-label helps in identifying genuine CP oligopeptide substrates. Indeed, ^{18}O-labeling indicates that at least one amino acid was released from the C-terminus of the identified peptide. In the following sections, we only report those substrate peptides from which only one amino acid was cleaved off. In fact, such peptides represent the overall majority of peptides identified (typically more than 80%), although in several cases CPs were able to sequentially remove several amino acids from peptide C-termini (data not shown).

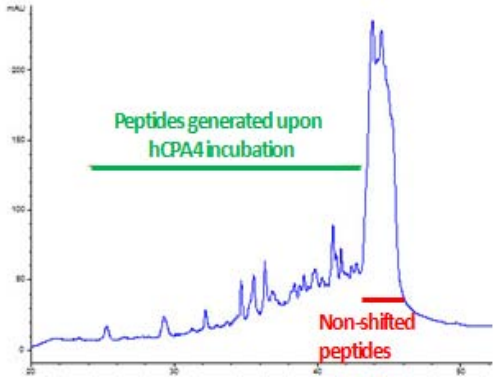

Fig. 3. Detailed zoom of a secondary COFRADIC run, showing that hydrophilic shifts were observed upon hCPA4 mediated C-terminal amino acid release.

The Substrate Specificity Profile of Human Carboxypeptidase A4

Although in principle the developed strategy is of general use to profile any kind of CPs, in this study we have focused on the metallo-forms of the M14 family of carboxypeptidases, one of the most distributed and varied groups of CPs. Using proteome-derived chymotrytic peptide libraries, we first assayed the previously characterized human carboxypeptidase A4 (Tanco et al., 2010; Van Damme et al., 2010). We identified 9729 ^{18}O-labeled hCPA4 products, representing 93% of all isolated peptides, thus pointing to a very efficient COFRADIC-based enrichment. From these peptides, the C-terminal amino acid cleaved off was inferred by database searching, which allowed determining the substrate preference of hCPA4. Note that the specificity determinants of MCPs are here described according to the model of Schechter & Berger (Schechter and Berger, 1967) in which each specificity subsite in the protease is able to accommodate the side-chain of a single amino acid residue in the substrate. The amino acids surrounding a MCP cleavage site are indicated as ...-P3-P2-P1↓P1´, with the residue C-terminal to the scissile bond being the primed side residue (P1'). The hCPA4 substrate specificity profile obtained by analyzing the hCPA4-treated proteome-derived chymotryptic peptide library, is shown as an iceLogo (Colaert et al., 2009) and as a heat map (Figs. 4a and 4b). For generating such iceLogos and heat maps, the frequency of amino acid occurrence at each position in hCPA4 substrates was compared with the amino acid occurrence found in the chymotryptic peptide library reference set. LC-MS/MS analysis of a representative set of the original peptide library input, identified 909 peptides and revealed that chymotrypsin generates a library composed of peptides mainly ending with Trp, Tyr, Leu, Phe or Met, although up to three missed cleavages could be observed (Figs. 5 and 6). In each case, to create the enzyme substrate specificity profiles using iceLogo, the reference set was used to correct for the bias introduced by the amino acid composition of the peptide library. The hCPA4-derived specificity profile reveals that the P1' position is the major specificity determinant for this M14 CP, and that amino acids at the P1 position also contribute to the substrate specificity. hCPA4 shows a preference for aliphatic and aromatic hydrophobic amino acids at P1' (Fig. 5). At the P1 position, basic and hydrophobic amino

acids are preferred, whereas acidic amino acids, Thr, and Gly are disfavored and Pro is even absent at this position (Fig. 7).

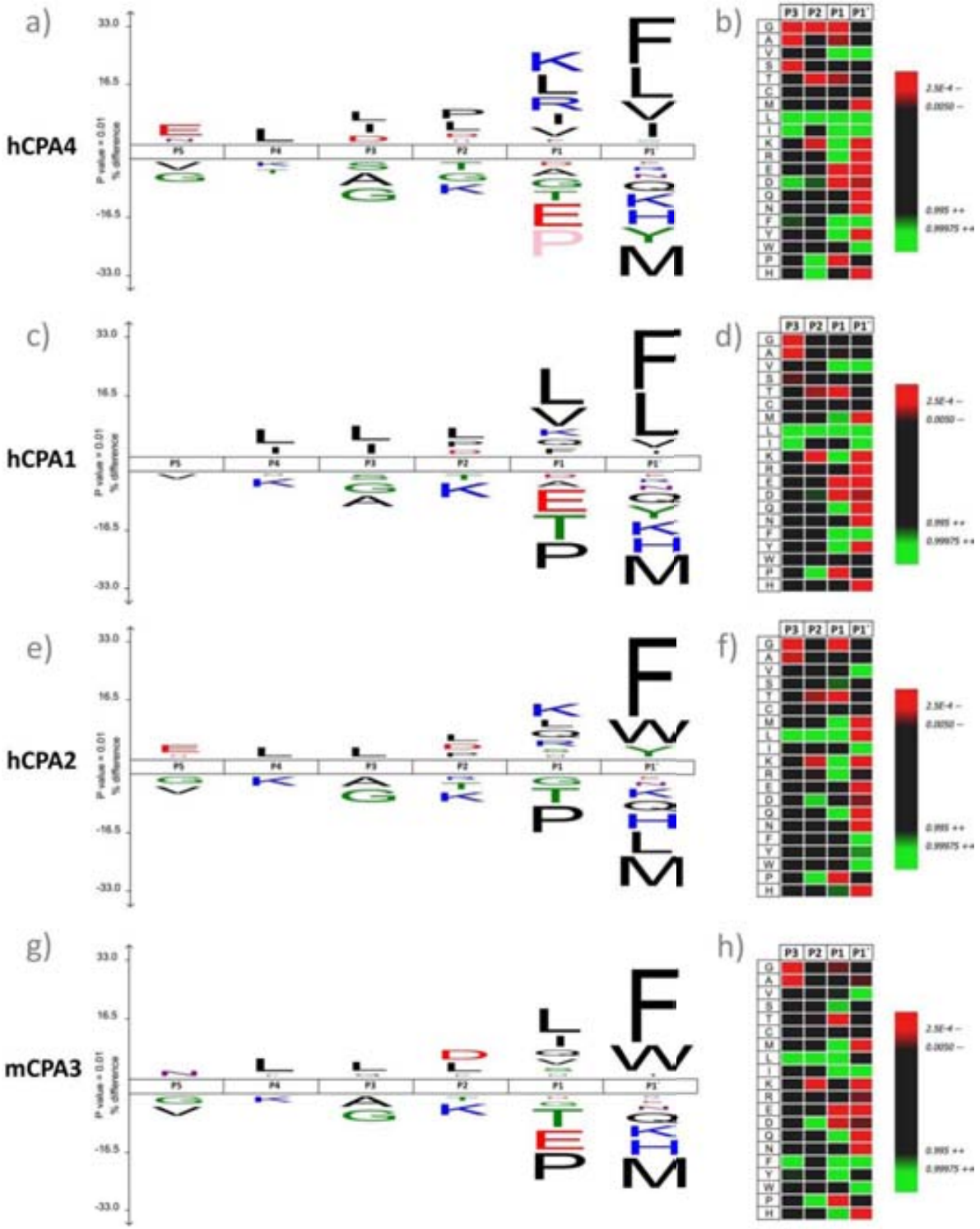

Fig. 4. Substrate specificity profiles of hCPA4, hCPA1, hCPA2 and mCPA3 using a chymotryptic proteome-derived peptide library. IceLogo and heat map representations of the substrate peptide sequences for hCPA4 (a and b respectively), hCPA1 (c, d), hCPA2 (e, f) and mCPA3 (g,h). These representations show the enriched and depleted residues present at the different identified MCP substrate positions as compared to the reference set. 9729 unique hCPA4 substrate peptides were identified, 4467 for hCPA1, 4207 for hCPA2 and 9680 for mCPA3. In all representations the substrate residues are depicted according to Schechter & Berger nomenclature (Schechter and Berger, 1967). The frequency of the amino acid occurrence at each position in the sequence set was compared with the occurrence in the peptide library reference set. This comparison corrects for the bias introduced by the amino acid composition of the peptide library. Only statistically significant residues with a p-value □ 0.01 are plotted in the iceLogo or colored in the heatmap. Amino acids height (iceLogo) or color (heat map) shows the degree of difference in the frequency of an amino acid in the experimental set as compared to the frequency in the reference set. In the iceLogo, residues that are statistically over- or underrepresented in the experimental set are respectively shown in the upper or lower part of the iceLogo. Residues colored in pink never occurred at a given position.

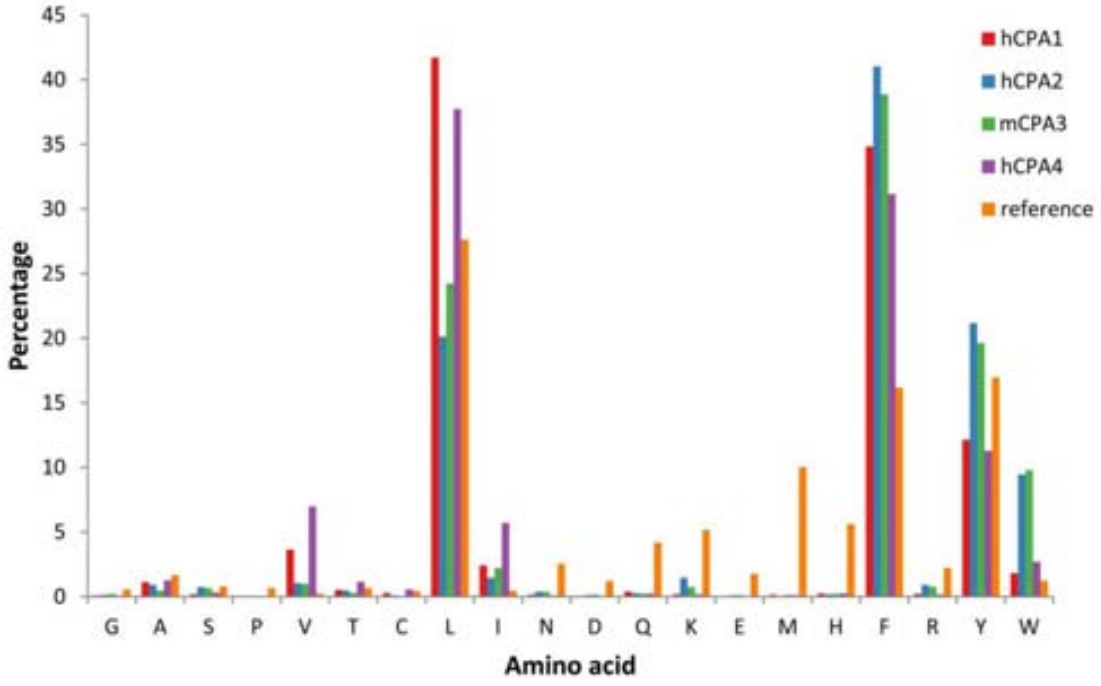

Fig. 5. Amino acid occurrence at P1' of the identified carboxypeptidase substrates identified using chymotryptic proteome-derived peptide libraries. Percentage of amino acid occurrence at the C-terminus of the identified hCPA1 (red), hCPA2 (blue), mCPA3 (green) and hCPA4 (purple) oligopeptide substrates using chymotryptic- peptide libraries are plotted and compared to the occurrence at that position for the reference chymotryptic peptide library (orange).

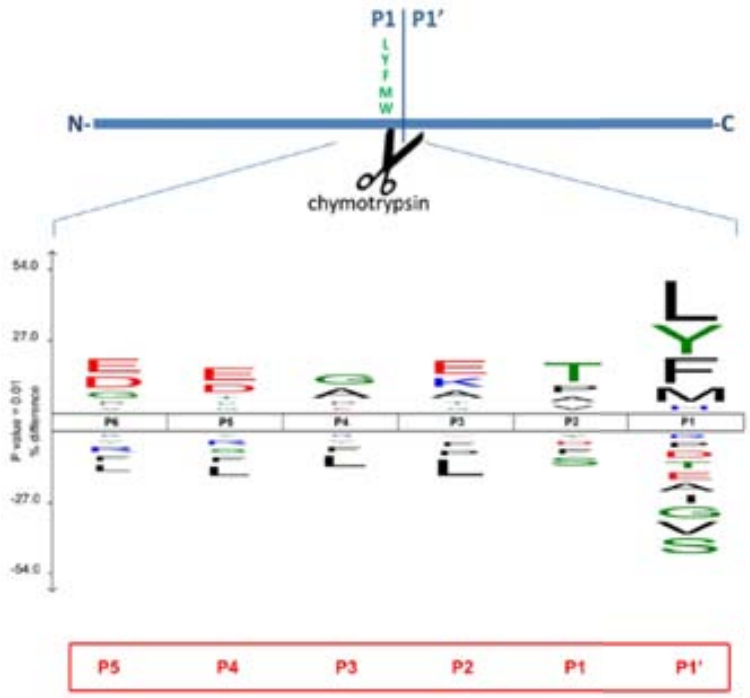

Fig. 6. Amino acid composition of the proteome-derived chymotryptic peptide library. Chymotrypsin cleaves C-terminal of Tyr, Phe, Leu, Trp and Met. The iceLogo representation shows the amino acid composition of the C-terminus according to the Schechter & Berger nomenclature (Schechter and Berger, 1967) of 910 chymotryptic library peptides. The iceLogo was created using the human Swiss-Prot database as reference set. The chymotryptic Schechter & Berger positions are depicted in the iceLogo and, the corresponding positions for the CPs substrates are indicated in a red box.

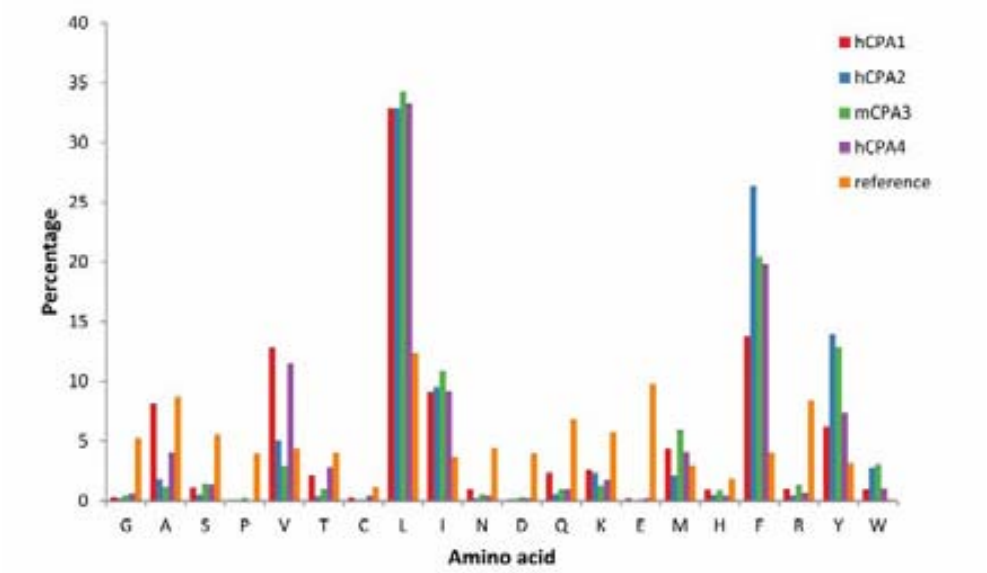

Fig. 7. Amino acid occurrence at the P1 position of the carboxypeptidase substrates identified using chymotryptic proteome-derived peptide libraries. Percentage of amino acid occurrence at the P1 position of the identified hCPA1 (red), hCPA2 (blue), mCPA3 (green) and hCPA4 (purple) oligopeptide substrates using chymotrypticsin-derived peptide libraries are plotted and compared to the occurrence at that position of the reference chymotryptic peptide library (orange).

This data is in good agreement with previous analyses using synthetic substrates, positional proteomics and peptidomics (Tanco et al., 2010; Van Damme et al., 2010), and thus validates our approach. However, a major difference is the nearly complete absence of Met at the P1' position, which was previously considered to serve as a good P1' substrate residue for this enzyme. Considering that Met oxidation might alter its cleavage susceptibility, a slightly modified assay was performed. Here, the methionine oxidation step preceding the primary HPLC run was omitted and as a result, the frequency of Met-ending peptides cleaved by hCPA4 increased 36-fold, from 0.11 % to 4.11% at P1' in the oligopeptide substrates (Fig. 8).

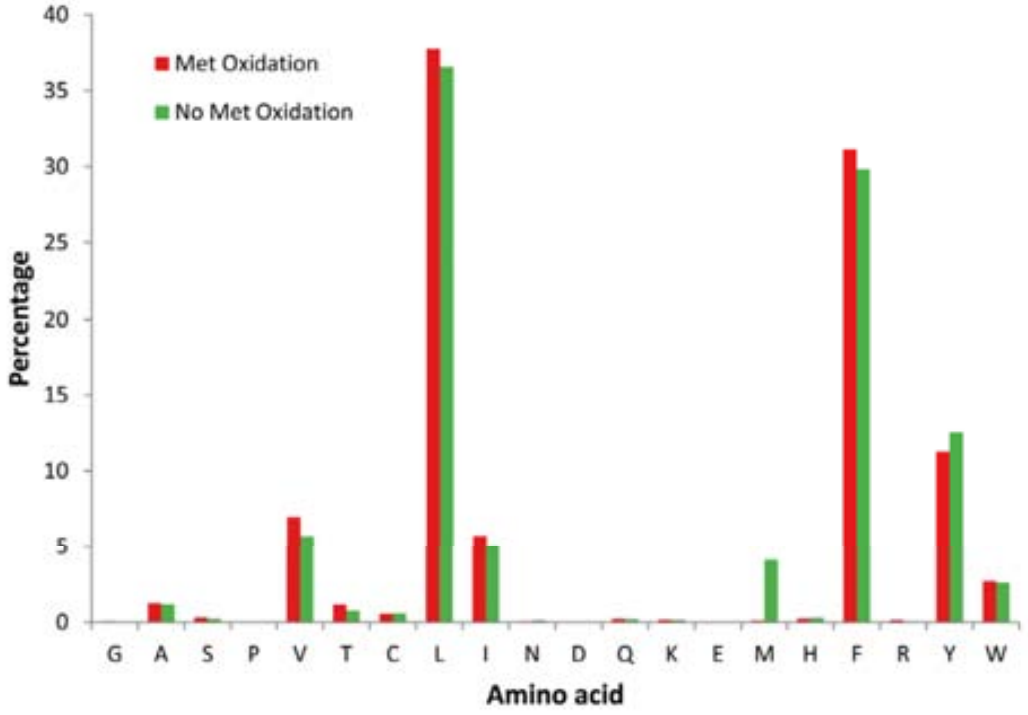

Fig. 8. Influence of Met oxidation on the P1' substrate specificity profile of hCPA4 using a chymotryptic proteome-derived peptide library. Percentage of amino acid occurrence at P1' of the identified hCPA4 oligopeptide substrates using chymotryptic proteome-derived peptides libraries with (red) and without (green) Met oxidation are plotted.

The Substrate Specificity Profile of the Human Carboxypeptidases A1 (hCPA1) and A2 (hCPA2)

We further validated our approach by assessing the specificity profiles of two previously characterized MCPs. A chymotryptic peptide library was used to assay hCPA1, yielding 4467 ^{18}O-labeled peptides (84% of all identified peptides), and hCPA2, yielding 4207 ^{18}O-labeled peptides (92% of all identified peptides). The hence derived specificity profiles are shown in Figs. 4c and 4d for hCPA1 and Figs. 4e and 4f for hCPA2. The amino acid occurrences used to build these profiles are further shown in Fig. 5 for the P1' position and in Fig. 7 for the P1 position. The P1' substrate preferences of hCPA1 are similar to those of hCPA4, though a decreased preference for basic amino acids at P1 is observed for hCPA1. Our data further confirm the previously described strong preference of hCPA2 for aromatic residues at the C-terminal position. Lys, Arg, Gln and Leu show an increased frequency of occurrence at P1 for hCPA2 as compared to the reference set. An inhibitory effect of Pro and a decreased occurrence of Thr and Gly when compared to the reference set at P1 can also be observed for hCPA2.

Substrate Specificity Characterization of Mast Cell Carboxypeptidase

Mast cell carboxypeptidase, or CPA3, is found in mast cell granules. Recent reports suggested a role for CPA3 in regulating innate immune responses and mast cell granule homeostasis (Pejler et al., 2009). Although CPA3 is known for many years, its functional properties remain poorly characterized. The difficulty to obtain large amounts of purified CPA3 from natural sources, its instability and the failures in establishing recombinant production of this MCP help to explain the lack of functional CPA3 studies. Previously, CPA3 was shown to hydrolyze the typical substrates of pancreatic CPA and peptides carrying a C-terminal hydrophobic residue, indicative of A-like MCP activity. Taking advantage of the fact that our approach requires only small amounts of CP for substrate specificity characterization, we used purified mCPA3 from mouse bone marrow-derived mast cells to characterize its substrate preferences (mouse and human CPA3 display 81.3% identity and 94.2% similarity). A two-step purification process based on heparin affinity column and gel filtration chromatography yielded a few micrograms of active

mCPA3. The purest fractions were incubated with a chymotryptic peptide library in the presence of a cocktail of protease inhibitors not targeting mCPA3. LC-MS/MS analysis of the COFRADIC-isolated peptides identified 9680 mCPA3 cleaved peptides (85% of all identified peptides). The substrate preferences of mCPA3 reveal a typical MCP specificity profile, as P1' mainly determines substrate selection additionally influenced by the amino acid at P1 (Figs. 4g and 4h). Interestingly, preferred substrates of mCPA3 contained large aromatic amino acids at their C-termini (Phe, Tyr or Trp), similar to hCPA2 (Fig. 5). Hydrophobic amino acids (Leu and Ile), are the main representatives at the penultimate position in the oligopeptide substrates identified, whereas Thr, Gly, Pro and acidic amino acids were again poorly represented (Fig. 7).

Carboxypeptidase Substrate Specificity Profiling Using Proteome-derived Lys-N Peptide Libraries

Protease-derived peptide libraries generated with endoproteases are composed of peptides with a characteristic amino acid preference at one of the peptide termini. Chymotrypsin generates peptides mainly ending with a C-terminal Trp, Tyr, Leu, Phe or Met residue (Fig. 5 and Fig. 6), thereby generating a vast number of potential substrates for A-like MCPs but possibly introducing a substrate bias. Therefore, we evaluated a peptide library created using the metalloendopeptidase Lys-N. This protease selectively cleaves the peptide bond N-terminal of Lys, and as a result it generates a C-terminal unbiased peptide library (i.e., C-termini of Lys-N peptides do not display any amino acid preference (Fig. 9)) (Hohmann et al., 2009; Raijmakers et al., 2010). This library typically contains longer peptides and thus on average more hydrophobic peptides that elute over a larger interval as compared to chymotryptic peptides. As a consequence, in the first RP-HPLC separation primary fractions were collected in a broader time-interval when compared to the chymotryptic library. The substrate specificities of hCPA1, hCPA2, mCPA3 and hCPA4 were then profiled using these peptide libraries. 3275 unique substrates for hCPA1 were identified (46% of all identified peptides), 2064 unique peptides for hCPA2 (36%), 2440 unique substrates for mCPA3 (51%) and 3503 unique peptides for hCPA4 (52%).

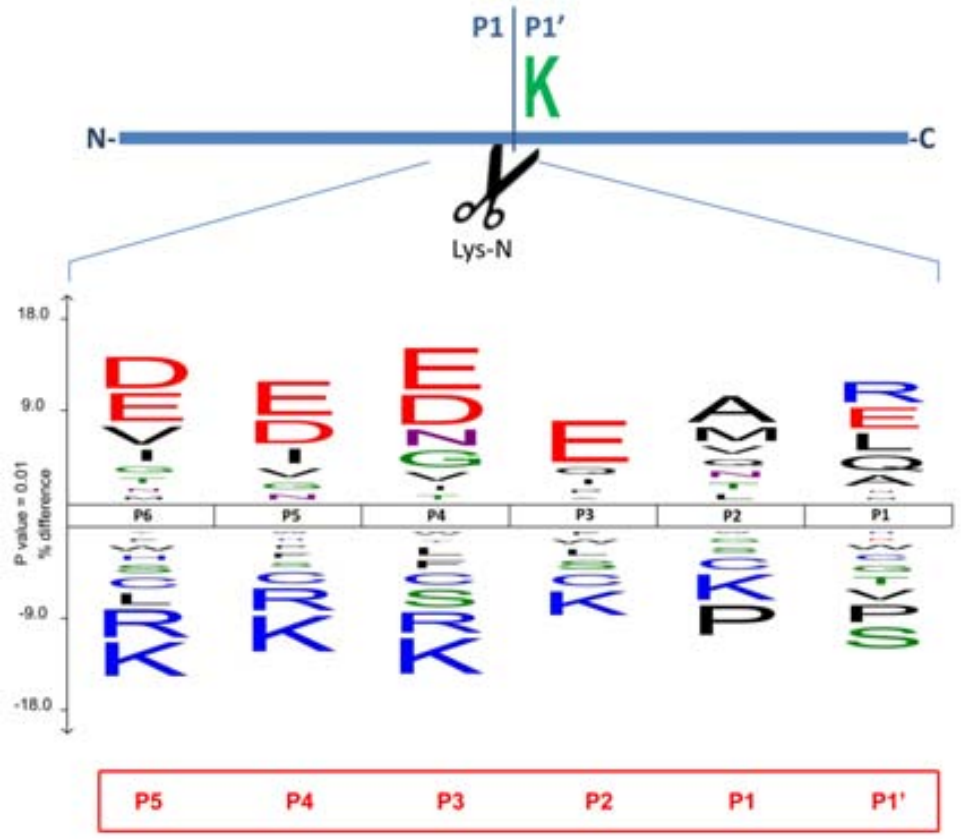

Fig. 9. Amino acid composition of the proteome-derived Lys-N peptide library. Lys-N cleaves N-terminal of Lys (Lys at P1'). The iceLogo representation shows the amino acid composition of the C-terminus according to the Schechter & Berger nomenclature (Schechter and Berger, 1967) of 7269 unique library reference peptides generated using Lys-N. The iceLogo was created using the human Swiss-Prot database as reference set. The Lys-N Schechter & Berger positions are depicted in the iceLogo and the corresponding positions for the CPs substrates are indicated in a red box.

The derived substrate preferences were further visualized as iceLogos (Fig. 10) and heat maps (Fig. 11). In addition, the amino acid occurrences used to build these profiles are shown in Fig. 12 for the P1' position and in Fig. 13 for P1. Of note here is that, in contrast to the chymotryptic peptide library, given the unbiased nature of the Lys-N generated peptide C-termini, the iceLogos show an overall higher contribution of P1' to the substrate specificity profile. Correspondingly, P1 is shown to contribute moderately to the MCP specificity.

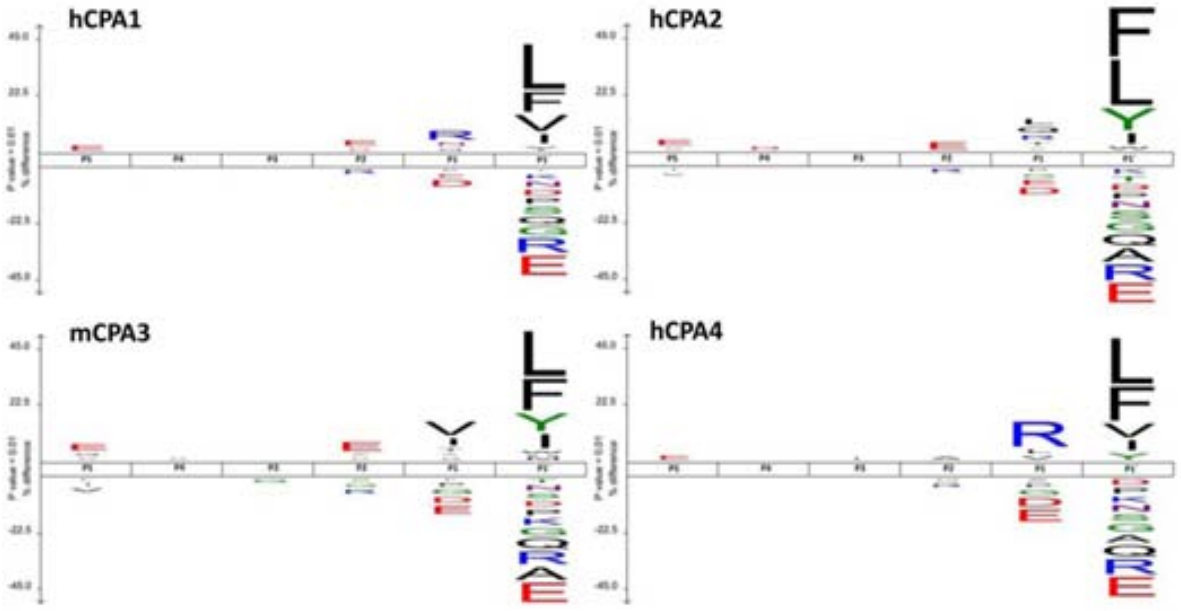

Fig. 10. Lys-N proteome-derived peptide library assay. IceLogo representations of the substrate preferences for hCPA1 (upper left), hCPA2 (upper right), mCPA3 (lower left) and hCPA4 (lower right) derived from Lys-N peptide libraries. These representations show the enriched and depleted residues present at the different identified MCP substrate positions as compared to the Lys-N peptide library reference set. 3275 hCPA1 substrate peptides were identified, 2064 for hCPA2, 2440 for mCPA3 and 3503 for hCPA4. In all representations, the substrate residues are depicted according to the Schechter & Berger nomenclature (Schechter and Berger, 1967).

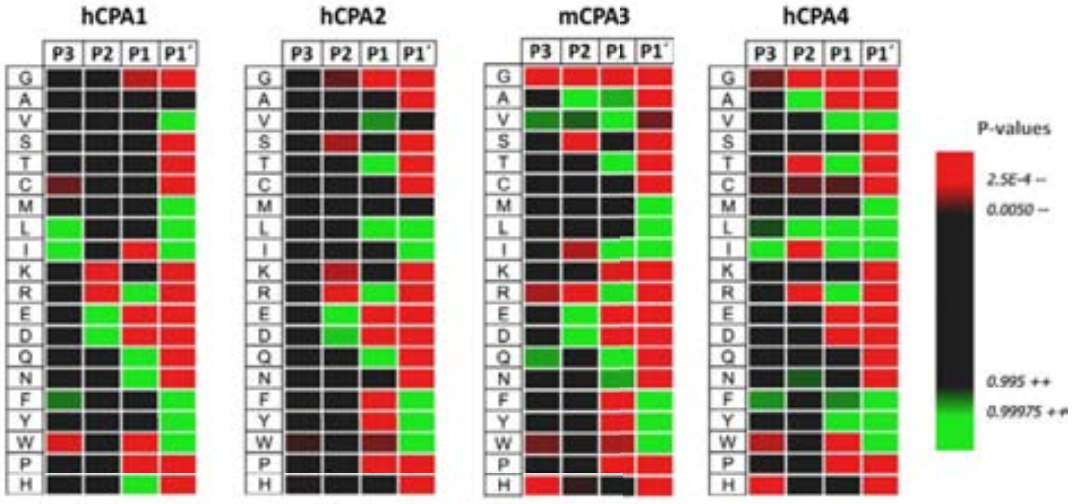

Fig. 11. **Heat maps depicting the substrate preferences of hCPA1, hCPA2, mCPA3 and hCPA4 using a Lys-N proteome-derived peptide library.** These representations show the statistically enriched and depleted residues present at the different identified MCP substrate positions as compared to the reference set. 3275 hCPA1 substrate peptides were identified, 2064 for hCPA2, 2440 for mCPA3 and 3503 for hCPA4.The substrate residues are depicted according to Schechter & Berger nomenclature (Schechter and Berger, 1967).

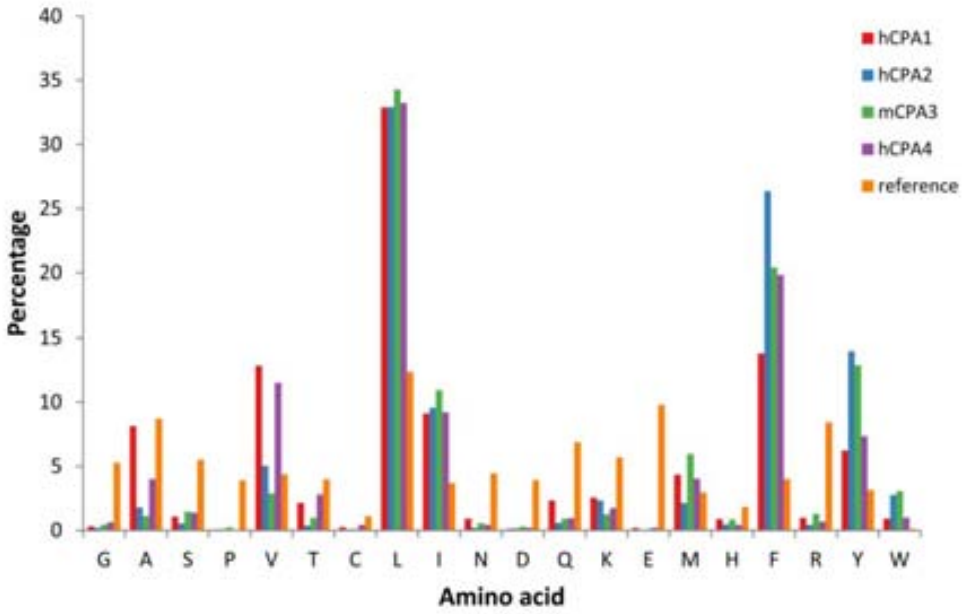

Fig. 12. P1' aminoacid occurrences for the Lys-N proteome-derived peptide library assay. C-terminal amino acid occurrences for the substrates of hCPA1 (red), hCPA2 (blue), mCPA3 (green) and hCPA4 (purple) using Lys-N peptide libraries are plotted and compared to the occurrence at that position for the reference Lys-N peptide library (orange).

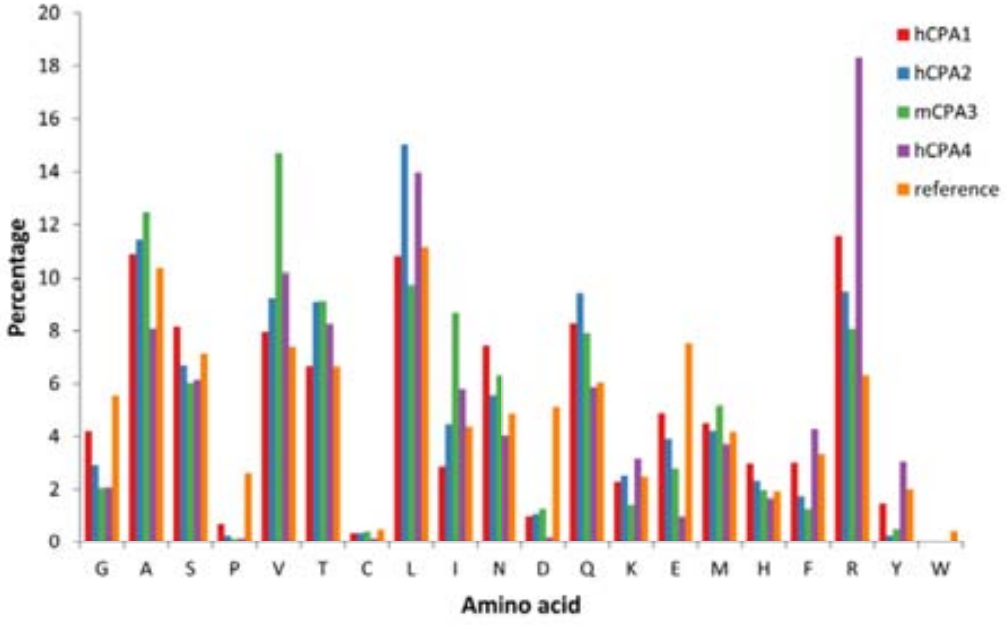

Fig. 13. P1 aminoacid occurrences for the Lys-N proteome-derived peptide library assay. Amino acid occurrences at the penultimate for the substrates of hCPA1 (red), hCPA2 (blue), mCPA3 (green) and hCPA4 (purple) using Lys-N peptide libraries are plotted and compared to the occurrence at that position for the reference Lys-N peptide library (orange).

A general analysis of the displayed specificity profiles reveals strong similarities with those derived using the chymotryptic library, in which hCPA2 and mCPA3 are more efficient in removing large hydrophobic amino acids like Trp or Tyr, when compared to hCPA1 and hCPA4 (Fig. 12). For the Lys-N library, no methionine oxidation step was performed, enabling for a better evaluation of the occurrence of Met in the overall substrate specificity profiles. As a consequence, Met appears even among the preferred P1' residues (Fig. 11). In relation to the penultimate amino acid preferences, a rather large discrepancy can be observed between the Lys-N and chymotryptic peptide library: threonine was generally found to be under-represented in this position in the chymotryptic library, but it seems to be tolerated for all the CPAs in the Lys-N library. Most probably, Thr has in fact a neutral to positive effect in this position, but the effect observed in the iceLogos and heat maps for the chymotryptic library might be an artifact caused by the relatively high levels of Thr in this position in the reference set for this library.

Classification of A-like Metallocarboxypeptidases According to Their Substrate Specificity Profiles

In mammals there are 6 genes of A-like M14 metallocarboxypeptidases (CPA1 to CPA6) and all of them hydrolyze C-terminal hydrophobic amino acids (Goldstein et al., 1989; Lyons and Fricker, 2010; Tanco et al., 2010; Wei et al., 2002). Given that CPA1 and CPA2 were the first characterized members showing characteristic specificities, A-like enzymes were further classified in A1-like forms (that prefer both small aliphatic as well as bulky aromatic amino acids) and A2-like forms (showing a strong preference for large aromatic residues). Figs. 5 and 12 show the frequencies of the amino acid occurrence at P1' for each of the MCPs here assayed (CPA1 to 4) and allow for a direct comparison of their substrate preferences. The substrate specificities of these enzymes can similarly be compared by plotting differential iceLogos (Fig. 14). These comparisons reveal hCPA4 as an A1-like and mCPA3 as an A2-like enzyme according to the established classification. Further, hierarchical cluster analysis of the P1' substrate specificity profile of the MCPs

analyzed using chymotryptic or Lys-N proteome-derived peptide libraries revealed that for both types of libraries, hCPA1 groups close to hCPA4 in terms of its P1' substrate specificity, while hCPA2 and mCPA3 cluster separately (Fig. 15), and thus enables to cluster enzymes according their differences in the observed substrate preferences.

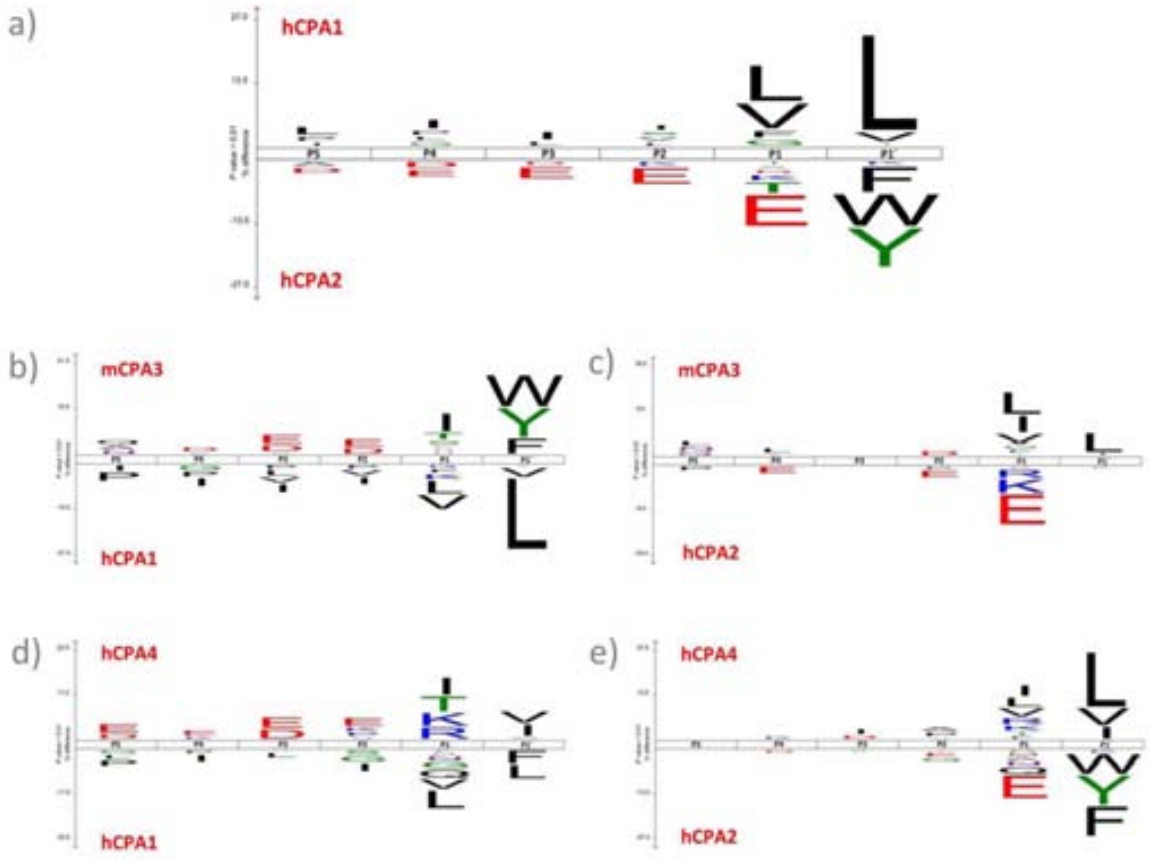

Fig. 14. Differential iceLogo representations of hCPA1, hCPA2, mCPA3 and hCPA4 cleavage motifs using a chymotryptic proteome-derived peptide library. Differential iceLogo representations of the substrate peptide sequences of hCPA1 vs. hCPA2 (a), mCPA3 vs. hCPA1 (b), mCPA3 vs. hCPA2 (c), hCPA4 vs. hCPA1 (d), and hCPA4 vs. hCPA2 (e). The differential iceLogos represent the enriched and depleted residues in the substrate peptides for the enzyme labeled in the upper part of the iceLogo using as reference set the substrates found for the enzyme labeled in the lower part of the iceLogo. In all representations the substrate residues are depicted according to Schechter & Berger nomenclature (Schechter and Berger, 1967).

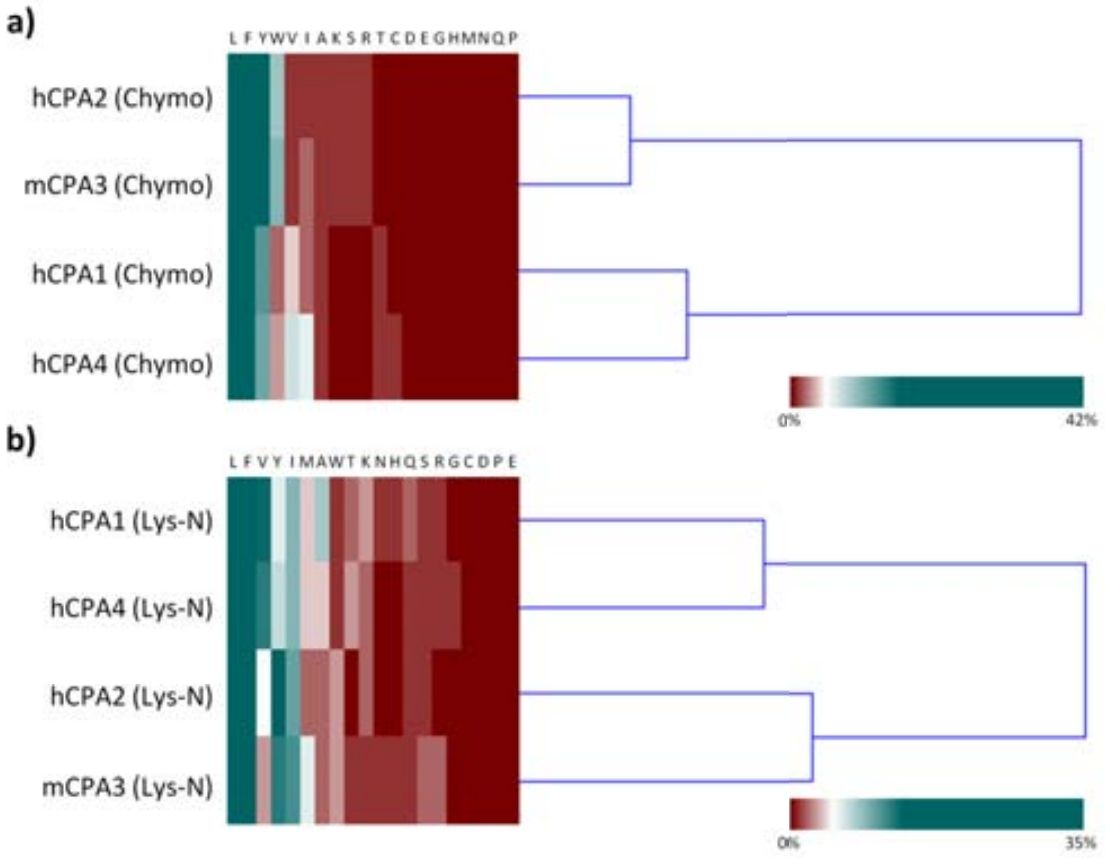

Fig. 15. Cluster analysis of the P1' substrate specificity profile of the MCPs analyzed using chymotryptic or Lys-N proteome-derived peptide libraries. a) Cluster analysis of the P1' substrate specificity data of hCPA1, hCPA2, mCPA3 and hCPA4 using chymotryptic proteome-derived peptide libraries. b) Cluster analysis of the P1' substrate specificity data of hCPA1, hCPA2, mCPA3 and hCPA4 using Lys-N proteome-derived peptide libraries. Each dendrogram represents the hierarchical Euclidian distance structure between the substrate specificity profiles that was computed using hierarchical clustering with the complete linkage algorithm and indicates the degree of similarity in the P1' substrate specificity profiles of the 4 MCPs analyzed. For both types of libraries, hCPA1 groups close to hCPA4 in terms of its P1' substrate specificity, while hCPA2 and mCPA3 cluster separately.

3.4. DISCUSSION

Carboxypeptidases participate in C-terminal peptide and protein processing and are thereby implicated in the regulation of a great variety of biological processes including blood coagulation/fibrinolysis (Declerck, 2011), blood pressure regulation (Kuba et al., 2010; Pshezhetsky and Hinek, 2009), pro-hormone and neuropeptide processing (Cawley et al., 2012; Lyons et al., 2008; Tanco et al., 2010; Zhang et al., 2008), and are also implicated in various pathological conditions such as cancer (Vovchuk and Petrov, 2008). Despite the importance of these processes, the lack of tools enabling the study of C-terminal processing has hampered identification of CP substrates. The inherently low chemical reactivity of the carboxylic acid group and the difficulty to discriminate peptide/protein C-terminal carboxyl functions from carboxyl functions of the acidic amino acid side chains have long been important obstacles for developing specific methodologies for enriching C-termini (Nakazawa et al., 2008; Toyo'oka, 1998). In addition, C-terminal sequencing has generally been less accessible and reliable than N-terminal sequencing, hampering the study of C-terminal proteolysis (Hardeman et al., 1998; Nakazawa et al., 2008; Samyn et al., 2000). Finally, C-terminal modifications usually affect the ability of CPs to recognize and cleave these peptides, thereby limiting the strategies for the design of probes.

In the present study we took advantage of the altered chromatographic behavior of intact peptides as compared to their proteolytic fragments to develop a sensitive and comprehensive method to profile carboxypeptidase substrate specificities *in vitro*. We presented a proteome-derived peptide library approach that allows for the MS-based identification of CP cleavage products, concomitantly allowing for cleavage site identification (both prime and nonprime side specificities). Further, ^{18}O-labeling of neo-C-termini permits to discriminate CP products from co-purifying peptides and we showed that up to a 90% of all COFRADIC-sorted and identified peptides report CP activity. Additionally, our approach can provide information about subsite cooperativity, although

no such cooperativity was observed for the here analyzed MCPs (data not shown). Our approach was optimized using CPs of the M14 family that show a strong preference for hydrophobic amino acids and as a result mainly hydrophilic shifts were observed upon C-terminal amino acid release (Fig. 3). We evaluated different members of the A-like subclass of MCPs and presented results that agree with previous data obtained using synthetic substrates, positional proteomics and peptidomics (Tanco et al., 2010; Van Damme et al., 2010). All of this suggests that our technique is of general use for assaying CPs that prefer hydrophobic amino acids (A-like in MCPs or C-type in SCPs). Further optimization might be required to extend this approach to other CPs. In addition, our approach might enable assaying the specificities of other C-terminal exopeptidases such as peptidyl dipeptidases, which include the angiotensin-converting enzyme (ACE).

Further, to the best of our knowledge this is the first statistically sound approach, profiling the preferred substrate sites for MCPs that extends analysis up to the P5 position. Using chymotrytic and Lys-N peptide libraries, we confirmed that the P1' position is the key determinant of substrate specificity for MCPs. It is clearly shown that P1 also influences substrate selection but to a lower extent. Our analysis suggests that for the MCPs under study, P2 to P5 have in general a very low influence on substrate selection.

Previous peptidomic studies typically made use of natural peptides isolates from cells and tissues as substrate pools for CPs (Lyons and Fricker, 2010, 2011b; Tanco et al., 2010). In principle our technology could also make use of such a natural peptide library; however one of the advantages of our proteome-derived approach is the ability to identify much larger numbers of substrates, which ensures a statistical more sound analysis of specificity profiles. As an example, a peptidomic analysis of hCPA4 (Tanco et al., 2010) identified 44 oligopeptide substrates, as compared to the 9729 and 3503 peptide substrates identified in this study using respectively a chymotryptic and a Lys-N peptide library. This is clearly explained by the fact that proteome-derived libraries offer a much greater number of possible substrates with broader sequence diversities. In addition and in contrast to peptidomic approaches, our approach enriches for CP products, which improves substrate identification. Another innovative aspect of our approach is the use of ^{18}O-labeling to

discriminate CP products from all other co-isolated peptides. Previously, peptidomic approaches reported of quantitative approaches by which efficient substrates can be distinguished from less efficient ones (Lyons and Fricker, 2011b). In principle, alike strategies to assess the efficiency of CP cleavage could be integrated when making use of proteome-derived peptide libraries. Further, peptidomics offers the possibility to evaluate naturally occurring and therefore biological relevant peptides, however some might hold stabilized C-termini, in which C-terminal processing have already occurred *in vivo*.

It was proposed that evolutionary traits allowed CPA1 and CPA2 to diverge from one another with respect to their substrate specificities, resulting in overlapping as well as complementary preferences (Gardell et al., 1988). hCPA2 displays a stronger preference for bulkier hydrophobic amino acids in comparison to hCPA1, which is able to more efficiently cleave small aliphatic residues (Tanco et al., 2010). This in contrast with bovine CPA, which is the only pancreatic CPA gene in *Bos taurus* and displays a broad substrate specificity. Our analysis confirms the characteristic preference of hCPA2 for large hydrophobic amino acids like Phe, Tyr or Trp. As for hCPA1, it confirms its less restricted specificity profile, since this enzyme performs better against amino acids like Leu, Val or Ala. Profiling the optimal peptide substrates of hCPA4 identified in this study, demonstrates that this enzyme is essentially an A1-like enzyme, as suggested previously (Tanco et al., 2010). The performance of the chymotryptic library seems to be better when describing the P1' substrate specificity of MCPs. The chymotryptic library-derived specificity profile for hCPA2 displays a more restricted preference for large aromatic hydrophobic amino acids at P1', in analogy with previously published kinetic data (Tanco et al., 2010). The Lys-N library seemingly is not able to discriminate between the differences in affinity for Leu or Ile that exist between hCPA1 and hCPA2 (Fig. 12). It is important to note that the use of chymotrypsin or Lys-N generates libraries with different peptide composition and length. In the former and although up to 3 missed cleavages could be observed, there is a lower content of internal hydrophobic residues, while the latter lack internal Lys residues and thus the presence of this amino acid cannot be profiled. Some of these particularities might explain why both libraries perform differently

when assessing certain aspects of MCP substrate preferences. The use of two different peptide libraries should minimize the bias introduced by each of them, and as a result, generates a better overall profile of enzyme preferences.

All the analyzed MCPs display overlapping P1 specificity profiles at the P1 position, although some subtle differences can be observed (Figs. 4, 7, 10, 11 and 13). It is striking to note the generally strong negative effect of a P1 Pro on substrate selection. Another common feature at P1 is the presence of Gly as a disfavored amino acid, although this effect it is less pronounced for hCPA1. This negative effect of Pro and Gly in the penultimate position seems to be a general feature for MCPs as it has also been observed for CPM (Deiteren et al., 2007), CPB (Deiteren et al., 2007), CPU/TAFI (Willemse et al., 2008), CPN (Willemse et al., 2008), CPE (Chen et al., 2001), CPD (Novikova et al., 1999) and CPA6 (Lyons and Fricker, 2010). Through molecular modeling and docking studies of CPM with different substrates, Deiteren *et al.* (Deiteren et al., 2007) suggest that a Gly at the P1 position side-chain permits an excessive flexibility of the substrate, explaining the unfavorable catalytic parameters for such substrates. In addition, the here studied enzymes in general disfavor acidic amino acids at P1, an observation which is more pronounced for hCPA4. Conversely, basic amino acids are favored for most of them, although again this effect is clearer for hCPA4, being most probably the most discriminating carboxypeptidase in the case of acid or basic P1 residues. In all cases, hydrophobic amino acids are among the preferred amino acids at P1 and their occurrence is even underrepresented at this position, viewing the fact that hydrophobic amino acids at the P1 position favor the CP-mediated release of a second C-terminal amino acid, and as a result only the products of the ultimate cleavage are purified and evaluated.

One of the advantages of our approach is that it allows, in a single experiment and using small quantities of a carboxypeptidase, a complete elucidation of CP cleavage site preferences. For instance, using very limited amounts of enzyme we here characterized the substrate specificity of mast cell carboxypeptidase (MC-CPA or CPA3), an enzyme that was discovered 25 years ago (Goldstein et al., 1987) but for which no in-depth analysis of its enzymatic properties was available. CPA3 was found to be implicated in the protection

against certain snake venom toxins in mice, and as a result this enzyme would play a role in regulating innate immune responses (Metz et al., 2006; Rivera, 2006). Recently, a mouse strain lacking CPA3 expression (MC-CPA$^{-/-}$ strain) showed altered mast cell staining, compatible with an immature phenotype, and displayed a strongly impaired storage of one of the MC chymases -mouse mast cell proteinase-5 (mMCP-5)- in the granules, indicating that CPA3 may participate in regulating secretory granule homeostasis (Feyerabend et al., 2005). Although, *in vitro* assays using synthetic CP substrates and some bioactive peptides have previously provided some insights into the cleavage properties and potential physiological substrates of this mast cell CP, we here generated a comprehensive peptide substrate specificity profile of mCPA3 (Figs. 4, 10 and 11). Mast cell CP was found to represent an A2-like MCP, with a marked preference for large hydrophobic amino acids like Trp, Tyr and Phe. This contradicts predictions based on structural models suggesting a preference of CPA3 for intermediate sized hydrophobic residues (Lyons and Fricker, 2010). A hierarchical cluster analysis, further confirms this classification, since it groups mCPA3 close to hCPA2 in terms of substrate specificity, while hCPA1 and hCPA4 cluster separately (Fig. 15). Although clearly being an A2-like MCP, it shows a relative higher preference for amino acids like Leu, Ile or Met, and a lower affinity for Phe in P1' when compared to hCPA2 (Figs. 5, 12 and 14). Despite the prediction that Ser 253 of mCPA3 (using the numbering system of bovine CPA) would interfere with the binding of very large amino acids like Trp (Lyons and Fricker, 2010), the here obtained specificity profile suggests that this position would not be critical in substrate specificity determination. As previously suggested (Faming et al., 1991; Garcia-Saez et al., 1997; Smith et al., 1997; Tanco et al., 2010), the amino acid in position 268 would be the key determinant of an A1-like or an A2-like specificity. In the case of mast cell carboxypeptidase, the presence of an Ala in position 268 (Ala-376 in mCPA3), in contrast to the Thr present in hCPA1 and hCPA4, might explain the specificity for bulkier amino acids like Tyr, Trp or Phe (Table 1). We built a 3D model of mCPA3 accommodating a hexapeptide in its active site (Fig. 16a), which shows that the amino acid at position 268 restricts the size of the S1' binding pocket. Figs. 16b and 16c illustrate how the

replacement of an Ala for a Thr at this position decreases the pocket dimensions and limits the ability of the enzyme to accommodate bulky C-terminal amino acids.

Table 1. Comparison of the amino acids that shape the S1' pocket in different CPs. The substrate S1' binding pocket residues are indicated, and by convention are numbered according to their position within the active form of bovine CPA, i.e., after removal of the signal peptide and the pro peptide. The corresponding residues in human and rat CPA1, CPA2, and CPA4; human and mouse CPA3 are indicated, along with their position within the precursor protein (preproCPA).

	Homologous amino acid residues								
Position in bovine CPA (active form)	**194**	**203**	**207**	**243**	**247**	**250**	**253**	**255**	**268**
Bovine CPA	Ser 304	Leu 313	Gly 317	Ile 353	Ile 357	Ala 360	Gly 363	Ile 365	Thr 378
Human CPA1	Ser 304	Met 313	Gly 317	Ile 353	Ile 357	Ala 360	Ser 363	Ile 365	Thr 378
Rat CPA1	Ser 304	Leu 313	Gly 317	Ile 353	Ile 357	Ala 360	Ser 363	Ile 365	Thr 378
Human CPA2	Thr 302	Met 311	Gly 315	Ile 351	Ile 355	Ala 358	Gly 361	Ile 363	Ala 376
Rat CPA2	Thr 302	Met 311	Gly 315	Ile 351	Ile 355	Ala 358	Gly 361	Ile 363	Ala 376
Human CPA3	Thr 302	Leu 311	Gly 315	Ile 351	Ile 355	Ile 358	Ser 361	Leu 363	Ala 376
Mouse CPA3	Thr 302	Leu 311	Gly 315	Ile 351	Ile 355	Thr 358	Ser 361	Leu 363	Ala 376
Human CPA4	Asp 306	Met 315	Gly 319	Thr 355	Val 359	Ala 362	Ser 365	Ile 367	Thr 380
Rat CPA4	Asp 306	Met 315	Gly 319	Thr 355	Val 359	Ala 362	Ser 365	Val 367	Thr 380

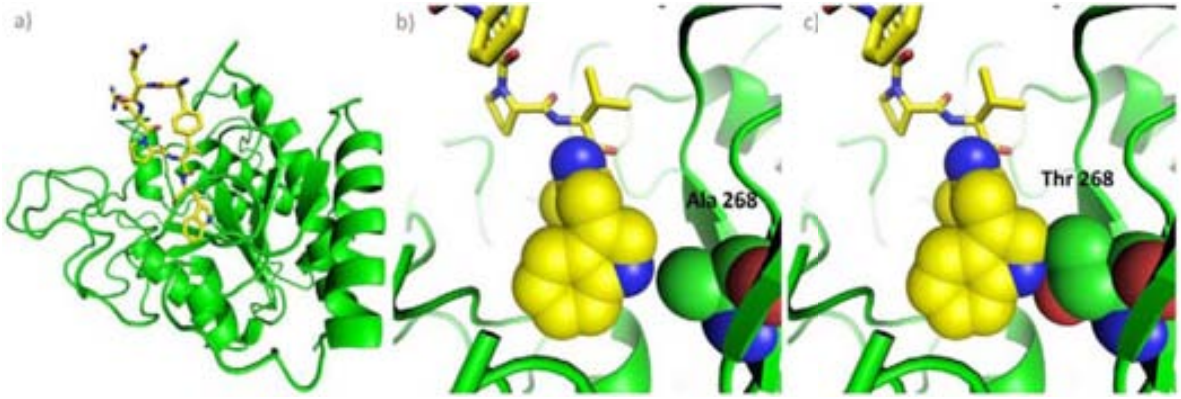

Fig 16. Structural based model of mouse CPA3 in complex with a hexapeptide. a) Cartoon representation of mCPA3 (shown in green) in complex with a Trp-ending hexapeptide (shown as a yellow stick model). b) Close-up view of the S1' specificity pocket. The hexapeptide is shown as a yellow stick model, with the exception of the C-terminal Trp of which the atoms (except of hydrogens) are shown as spheres. The Ala in position 268 is also indicated and shown as spheres. c) Detailed view of the S1' pocket of mCPA3 in which the Ala in position 268 has been mutated to Thr (Ala268>Thr). The atom color coding is as follows: green for C (enzyme), yellow for C (peptide), cyan for N, and red for O.

The description of mast cell carboxypeptidase as an A2-like MCP has implications on the products that this enzyme generates upon digestion of angiotensin I (Ang-I), one of the putative physiological substrates of CPA3. Recently, Pereira *et al.* (Pereira et al., 2012) compared the action of rat CPA1 and CPA2 on Ang-I. While an A2-like enzyme mainly produces Ang-(1-9) upon Ang-I digestion, an A1-like enzyme is more efficient in further processing Ang-(1-9) into Ang-(1-7) (Pereira et al., 2012). This observation can be explained based on the more restricted substrate specificity displayed by A2-like enzymes. Knowledge of the optimal cleavage sequences for mast cell carboxypeptidase will be very useful to corroborate *in vivo* cleavage events and might enable prediction of novel biological mast cell carboxypeptidase substrates, which are key to gain insight into the biological function of this enzyme.

Chapter 4

Positional proteomics identification of the natural substrates of cytosolic carboxypeptidases.

4.1 INTRODUCTION

Cytosolic carboxypeptidases (CCPs) constitute a recently described subfamily of M14 metallocarboxypeptidases (MCP), which clusters non-secretory forms essential for important functions of biological and biomedical interest (Rodriguez de la Vega et al., 2007). The human and mouse CCP subfamily is composed of six members (CCP1 to CCP6), all related to Nna1 (afterwards renamed CCP1) which is the first reported and best characterized member of this subfamily. In 2000, a screen for mRNAs up-regulated by axonal regeneration identified a gene product with homology to the MCP family; this gene product was named Nna1 (nervous system nuclear protein induced by axotomy) (Harris et al., 2000). Gene disruption of ccp1/nna1 leads to an ataxic phenotype in Purkinje Cell Degeneration (*pcd*) mice (Fernandez-Gonzalez et al., 2002). These mice display multiple defects including postnatal degeneration of several types of neuronal tissues and infertility, supporting a role of CCP1 in the survival neurons and other kind of cells. Consequently mouse Nna1/CCP1 is thought to function on axogenesis, dendritogenesis, fotoreceptor development and spermatozoid wellness (Fernandez-Gonzalez et al., 2002; Harris et al., 2000; Li et al., 2010; Wang and Morgan, 2007). The degeneration observed in the *pcd* mice can be rescued by wild-type (WT) CCP1 but not with an inactive CCP1 mutant, showing that all these biological functions depend on its carboxypeptidase activity (Chakrabarti et al., 2008; Wang et al., 2006)

Although the molecular pathways through which the loss of CCP1 function leads to neuronal degeneration and death in *pcd* mice remain undefined, different cellular mechanisms have been proposed: endoplasmic reticulum stress (Kyuhou et al., 2006), decreased peptide turnover downstream from the proteasome (Berezniuk and Fricker, 2010), mitochondrial dysfunction and altered proteolytic processing of Nna1-interacting proteins (Chakrabarti et al., 2009), a dysfunction of microtubule stability (Li et al., 2010), and a progressive transcriptional silencing due to the accumulation of DNA lesions (Baltanas et al., 2011a; Baltanas et al., 2011b; Valero et al., 2006). Particularly, it has been described that CCP1 contains nuclear localization signals and is localized in both the nucleus and cytoplasm (19). But the primary function of Nna1 in

nuclear processes remains undefined, although it has been suggested that it could play a role in chromatin remodeling, in particular during the motoneuron response to axotomy (Harris et al., 2000). In addition, it has been observed that *pcd* mutation induces the repression of protein-coding genes in extensive domains of the genome of PCs. This transcriptional repression involves a profound reorganization of the nuclear architecture of chromatin, that leads to the accumulation of unrepaired DNA and, ultimately, to neuronal cell death (Baltanas et al., 2011a).

In recent years, a primary function for CCPs in tubulin processing has been proposed. CCP2 is thought to participate in the tyrosination cycle of tubulin, showing the ability to cleave C-terminal tyrosine of tubulin (Sahab et al., 2011). Four members of the CCP subfamily (CCP1, CCP4, CCP5 and CCP6) are known to specifically trim the gene encoded C-terminal glutamate residue of previously detyrosinated tubulin (Rogowski et al., 2010). Moreover CCP1, CCP4 and CCP6 catalyze the shortening of polyglutamate chains, a posttranslational modification that generates glutamate side chains on tubulins and other proteins (Rogowski et al., 2010). CCP5 perform a similar function, but in addition it specifically removes the branching point of a polyglutamylation event (Kimura et al., 2010; Rogowski et al., 2010). The only member of the human/mouse CCP subfamily that remains without an assigned specificity is CCP3. Furthermore, considering that polyglutamylation is a general posttranslational modification that not only affects tubulin, it could be expected for CCPs have a role in the regulation of polyglutamylation of different proteins.

The posttranslational modifications in which CCPs seem to be involved have been proved to play a role in the regulation of many cellular functions. Notably, analysis of the *pcd* mice phenotype provides direct evidence that abnormally high polyglutamylation levels lead to neuronal degeneration (Rogowski et al., 2010). In addition, microtubule polyglutamylation has recently been shown to regulate the activity of ciliary dynein and of the microtubule-severing protein spastin, most probably through altering the binding of these proteins to microtubules.

In reality, the knowledge on the CCP actions is very limited and their natural substrates are largely unknown, besides tubulin. However, it was demonstrated that CCP1, CCP4 and CCP6 can shorten the gene encoded C-terminal glutamate stretch of

two important regulators of myosin functions; MLCK and telokin (Rogowski et al., 2010). This raised the possibility that CCPs could shorten gene-encoded glutamates stretches on the C-terminus of other cytosolic and nuclear proteins. In this report, we performed a degradomic screen for primary sequence CCP1 substrates, applying C-terminal COFRADIC, a recently developed terminal proteomics approach. As a result, we identified 6 new putative CCP1 substrates, all of them containing acidic stretches in their C-terminus.

4.2. EXPERIMENTAL PROCEDURES

Cell culture

HEK 293T cells (American Type Culture Collection (ATCC), Manassas, VA, USA) were cultured in Dulbecco's Modified Eagle's Medium with 4500 mg/L glucose, GlutaMAX–I and pyruvate supplemented with 10% fetal calf serum (Invitrogen, Carlsbad, CA, USA). Cells were grown in a constant temperature incubator set at 37 °C and 10% CO2. HEK 293F cells (Invitrogen) were cultured in FreeStyle 293 expression medium (Invitrogen). Cells were grown in flasks on a rotary shaker (120 rpm) at 37 °C in a 8% CO2 and 70% relative humidity atmosphere.

Molecular cloning and transfection

The full-length cDNA sequence of human CCP1 (Q9UPW5-1, Uniprot) was cloned into the pOPINFS vector (Berrow et al., 2007). Both, a *Strep*-tag II (Schmidt and Skerra, 2007) and the hemagglutinin (HA) epitope were introduced on the C-terminus of CCP1. The CCP1 (E270Q) mutant was generated by QuikChange site-directed mutagenesis according to the manufacturer's protocol (Stratagene, La Jolla, CA, USA). The full-length cDNA sequences of TRAF-type zinc finger domain-containing protein (TRAD1, Uniprot O14545) and human high mobility group protein B3 (HMGB3, Uniprot O15347) were cloned into the pTriEx-6 vector and as a result a Strep-tag II epitope was introduced on the protein N-terminus.

DNA transfections were carried out using polyethylenimine (PEI), linear, 25 kDa (PolySciences, Warrington, PA, USA). Briefly, DNA was mixed with PEI in a ratio of 1:3 and incubated in the presence of serum-free medium for 15 minutes at room temperature. HEK 293T, at a 40-50% confluency, were exposed to the transfection complex at 1.4 ug DNA per ml medium concentration for 60 h.

In vitro substrate validation assay

Affinity purified CCP1 was incubated at 25 ng/µl for 2h at 37ºC with affinity purified TRAD1 or HMGB3 at 25ng/µl in the activity buffer (80mM PIPES pH6.8, 1mM MgCl2) complemented with 1/500 EDTA-free protease inhibitor cocktail Set III (Calbiochem).

Immunoblot

For immunoblotting, equivalent amounts of each lysate or fraction were analyzed by SDS/10% PAGE, and the separated proteins were electrophoretically transferred onto polyvinylidene difluoride (PVDF) filters (Millipore Corp., Billerica, MA, USA). Non-specific binding sites on the PVDF filters were blocked by incubation with 5% skim milk for 2 h. The PVDF filters were then incubated with the primary antibody for 1 h at 25 °C. As primary antibodies were used: rabbit antiAGTPBP1 (dilution 1:1,000; ProteinTech Group, Chicago, IL, USA), rabbit anti-detyrosinated tubulin (1:2,000; Chemicon, Pittsburgh, PA, USA), rabbit anti-Δ2-tubulin (1:1,000; Millipore), rabbit anti-polyE (anti-polyglutamilation, 1:2,000; gift of M.A. Gorovsky), mouse anti-HA (HA-7; 1:3,000; Sigma), mouse anti-*Streptag* II (StrepMAB-Classic; 1:1,000; IBA), anti-tyrosinated tubulin (TUB-1A2; 1:5,000; Sigma) or anti-β-actin (1:100,000; Sigma) antibodies. After washing with PBS containing 0.05% Tween 20, the filters were incubated with HRP-labeled goat donkey anti-rabbit (1:5,000; GE Healthcare), goat anti-mouse (1:5,000; Biorad)anti-mouse IgG-peroxidase (dilution 1:10,000; Pierce) for 1 h. After three rinses, immunoreactive bands were visualized with Immobilon Western Chemiluminiscent HRP substrate (Millipore).

Isolation of C-terminal terminal peptides

HEK 293T cells were washed in and resuspended at 2×10^7 cells per ml in 50 mM sodium phosphate buffer, pH 8.0, 100 mM NaCl and 0.5 mM EDTA supplemented with Complete Protease Inhibitor Cocktail (Roche Diagnostics GmbH, Mannheim, Germany). Cells were then subjected to three rounds of freeze-thaw lysis and cleared by centrifugation. Guanidinium hydrochloride was added to the cell lysates to a final concentration of 4 M to denature and inactivate proteases. Primary free amines were

acetylated for 2 h at 30 °C by sulfo-N-hydroxysuccinimide acetate to a final concentration of 10 mM. O-acetylation was reversed by adding hydroxylamine to a final concentration of 40 mM and incubation for 10 min at 30 °C. The acetylation reagent was further quenched for 10 min at 30 °C by glycine (final concentration 20 mM). Excess reagents were removed by desalting over 1 ml Illustra NAP^{TM}-10 columns (and proteins were eluted in 1.5 mL of 20 mM NH_4HCO_3 (pH 7.9)).

Protein concentrations were determined with BioRad's Protein Assay (BioRad Laboratories, Munich, Germany) according to the manufacturer's instructions. After heating the proteins for 5 min at 95 °C followed by placing them on ice for 5 min, acetylated proteins were digested overnight at 37 °C with sequencing-grade, modified trypsin (Promega, Madison, WI, USA), at an enzyme/substrate ratio of 1:100 w/w.

The so-generated peptide mixtures were vacuum dried. Peptides starting with pyroglutamate were unblocked before Strong Cation Exchange (SCX) prefractionation: 25 µl of pGAPase (pyroglutamyl aminopeptidase, 25 U/ml) (TAGZyme kit, Qiagen Inc., Valencia, CA) was activated for 10 min at 37 °C by addition of 1 µl of 50 mM EDTA (pH 8.0), 1.25 µl of 800 mM NaCl and 11 µl of freshly prepared 50 mM cysteamine-HCl. To the pGAPase solution, 25 µl of glutamine cyclotransferase (Qcyclase, 50 units/ml, TAGZyme kit) was added. The dried peptides were redissolved in 212 µl of buffer containing 16 mM NaCl, 0.5 mM EDTA, 3 mM cysteamine and 50 µM aprotinin. The activated pGAPase and Qcyclase mixture was added to the sample, and the mixture was incubated for 1 h at 37 °C (total volume of 275 µl). Then, 275 µl acetonitrile was added, and the sample was acidified to pH 3 using a stock solution of 1% trifluoroacetic acid (TFA) in 50% acetonitrile. The sample was diluted with buffer A (10 mM sodium phosphate in 50% acetonitrile) to a final volume of 1 ml. This peptide mixture was loaded onto a SampliQ Silica Strong Cation Exchange (Si-SCX, 100mg, 1mL, Agilent Technologies, Waldbronn, Germany) preequilibrated with buffer A using an Agilent vacuum manifold according to the manufacturer's instructions. Sample loading was followed by washing the SCX cartridge with 5 mL of buffer A and 2 mL of buffer A supplemented with 5 mM NaCl.

Peptides recovered in the flow-through and wash fractions, containing the N- and C-terminal peptides were vacuum dried and re-dissolved in 100 µl of HPLC solvent A

(10 mM ammonium acetate in water/acetonitrile (98/2, v/v), pH 5.5). To prevent oxidation of methionines between the primary and secondary RP-HPLC runs, methionines were oxidized before the primary run. The methionine oxidation reaction proceeded in the injector compartment by transferring 20 µl of a freshly prepared aqueous 3% H_2O_2 solution to a vial containing 90 µl of the acidified peptide mixture (final concentration of 0.54% H_2O_2). This reaction proceeded for 30 min at 30 °C after which 100 µl of this peptide mixture (an equivalent of 300 µg of digested proteins) were immediately injected onto the RP-HPLC column. (Zorbax 300SB-C18 Narrow bore, 2.1 mm internal diameter (I.D.) × 150 mm length, 5 µm particles; Agilent Technologies).

RP-HPLC separation was performed as follows. Following 10 min isocratic pumping with HPLC solvent A, a gradient was started of 1% HPLC solvent B (10 mM ammonium acetate in water/acetonitrile (30/70, v/v), pH 5.5) increase per minute. The column was then run at 100% solvent B for 5 min, switched to 100% solvent A and re-equilibrated for 20 min with solvent A. The flow was kept constant at 80 µl/min using Agilent's 1100 series capillary pump with an 800 µl/min flow controller. Fractions of 4 min were collected from 20 to 80 min after sample injection.

These primary fractions were dried and redissolved in 20 µl of 100 mM sodium phosphate (pH 8.0). Then 20 µl of a freshly made solution of NHS-butyric acid (60 mM in 100 mM sodium phosphate; pH 8.0) was added ($^{12}C_4$-butyrylation was used for the control sample and $^{13}C_4$-butyrylation for the CCP1 overexpression sample). Acylation of primary amines was allowed for 2 h at 30 °C. The acetylation reagent was further quenched for 10 min at 30 °C by glycine (final concentration 20 mM). O-acetylation was reversed by heating the peptide mixture at 95 °C for 60 min.

Corresponding primary fractions from the two parallel primary RP-HPLC runs (the control and CCP1 overexpression samples) were pooled and acidified with 4 µl of acetic acid. The butyrylated fractions were reloaded on the same RP-HPLC column and separated using identical conditions. Twelve secondary fractions, highly enriched for N-terminal peptides, were collected during a 6 minutes time interval starting 2 minutes before primary fraction interval (30 s collection intervals), and the secondary fractions highly enriched for C-terminal peptides were collected up to 10 min after the primary

collection interval, again using 30 s collection intervals. This resulted in 32 collected fractions per secondary run and 480 fractions in total. To reduce LC-MS/MS analysis time, secondary fractions eluting 12 min apart were pooled, dried and redissolved in 20 µl of 20 µl of 2 mM TCEP and 2% acetonitrile. In total, 96 samples were analyzed by LC-MS/MS.

Shotgun proteomics

Guanidinium hydrochloride was added to the HEK 293T cells lysates to a final concentration of 4 M to denature and inactivate proteases. 1 mg of each protein sample were brought to 500 µl using 4 M guanidinium hydrochloride. Protein mixtures were desalted using 1 ml Illustra NAP^{TM}-5 columns (and proteins were eluted in 1 mL of 20 mM Triethylammoniumbicarbonate (pH 7.0)). Prior to digestion, protein concentrations were measured again with BioRad's Protein Assay. Proteins were digested overnight at 37 °C with endoproteinase Lys-C (endoLys-C, Roche Diagnostics Gmbh). Differential labeling of was performed by propionylation (Ghesquiere et al., 2009) on primary α-amines (peptide N-termini) and ε-amines (lysine side-chains), with isotopic variants of N-hydroxysuccinimide propionate. Control sample was labeled with $^{12}C_3$-propionate and the CCP1 overexpression sample was labeled with $^{13}C_3$-propionate, during 2 hours at 30 °C. Excess reagent was neutralized by adding 28 µl of a 1 M stock glycine to the sample and heated for 60 min at 95°C (in order to reverse unwanted O-propionylation of the side-chains of serines, threonines and tyrosines). Samples were acidified with formic acid to obtain a pH 2-3. 250 µg of each labeled sample was mixed in a 1:1 ratio, were vacuum dried and re-dissolved in 100 µl of HPLC solvent A.

RP-HPLC separation of the sample peptides was performed with previous Met oxidation, as already described. Then, following 10 min isocratic pumping with HPLC solvent A, a gradient was started of 1% HPLC solvent B (10 mM ammonium acetate in water/acetonitrile (30/70, v/v), pH 5.5) increase per minute. The column was then run at 100% solvent B for 5 min, switched to 100% solvent A and re-equilibrated for 20 min with solvent A. The flow was kept constant at 80 µl/min using Agilent's 1100 series capillary pump with an 800 µl/min flow controller. Fractions of 2 min were collected

from 20 to 80 min after sample injection, vacuum dried and redissolved in 20 µl of 20 µl of 2 mM TCEP and 2% acetonitrile. In total, 30 samples were analyzed by LC-MS/MS.

LC-MS/MS analysis.

LC-MS/MS analysis was performed using an Ultimate 3000 RSLC nano LC-MS/MS system (Dionex, Amsterdam, The Netherlands) in-line connected to a LTQ Orbitrap Velos (Thermo Fisher, Bremen, Germany). 2 µl of the sample mixture was first loaded on a trapping column (made in-house, 100 µm I.D. x 20 mm length, 5 µm Reprosil–Pur Basic-C18-HD beads, Dr. Maisch, Ammerbuch-Entringen, Germany). After back-flushing from the trapping column, the sample was loaded on a reverse-phase column (made in-house, 75 µm I.D. x 150 mm length, 3 µm C18 Reprosil–Pur Basic-C18-HD beads). Peptides were loaded with solvent A' (0.1% trifluoroacetic acid in 2% acetonitrile) and were separated with a linear gradient from 98% of solvent A'' (0.1% formic acid in 2% acetonitrile) to 50% of solvent B' (0.1% formic acid in 80% acetonitrile) with a linear gradient of a 1.8% solvent B' increase per minute at a flow rate of 300 nl/min followed by a steep increase to 100% of solvent B'. The Orbitrap Velos mass spectrometer was operated in data-dependent mode, automatically switching between MS and MS/MS acquisition for the ten most abundant peaks in a MS spectrum. Full scan MS spectra were acquired in the Orbitrap at a target value of 1E6 with a resolution of 60,000. The ten most intense ions were then isolated for fragmentation in the linear ion trap, with a dynamic exclusion of 20 s. Peptides were fragmented after filling the ion trap at a target value of 1E4 ion counts. From the MS/MS data in each LC run, Mascot Generic Files were created using the Mascot Distiller software (version 2.3.2.0, Matrix Science, www.matrixscience.com/Distiller.html). While generating these peak lists, grouping of spectra was allowed with maximum intermediate retention time of 30 s and maximum intermediate scan count of 5. Grouping was done with a 0.005 Da precursor tolerance. A peak list was only generated when the MS/MS spectrum contained more than 10 peaks. There was no de-isotoping and the relative signal-to-noise limit was set at 2. The generated MS/MS peak lists were then searched with Mascot using the Mascot Daemon interface (version 2.3, Matrix Science). The Mascot search parameters were set as follows. Searches were performed in the Swiss-Prot database with taxonomy set

to human (either the 2011_05 or 2011_06 UniProtKB/Swiss-Prot database containing respectively 20,286 and 20,312 human protein sequence entries were used). The Mascot search parameters were set as follows. For the C-terminal COFRADIC analysis, searches were performed in the Swiss-Prot database with taxonomy set to human (database version 56.4; 20,328 human protein sequences). Acetylation at lysines and protein N-termini and methionine oxidation to methionine-sulfoxide were set as fixed modifications. Pyroglutamate formation of N-terminal glutamine were set as variable modifications Arg-C (for identification of database-annotated termini) or semi-ArgC (for identification of neo-termini) were set as the used protease (one missed cleavage was allowed). For the shotgun assay, searches were performed in the Swiss-Prot database with taxonomy set to human. Enzyme was set to endoLys-C. Variable modifications were set to pyroglutamate formation of N-terminal glutamine and acetylation of the protein's N-terminus. Methionine oxidation to methionine-sulfoxide was set as fixed modification. Mass tolerance on the precursor ion was set to 10 p.p.m. and on fragment ions to 0.5 Da. In addition, Mascot's C13 setting was set to 1. The peptide charge was set to 1+, 2+ or 3+ and instrument setting was put on ESI-TRAP. Only peptides that were ranked one and scored above the threshold score, set at 95% confidence, were withheld. According to the method described by Käll et al. (Kall et al., 2008), the false discovery rate was estimated and was found <2% at the spectrum level and <3% at the peptide level.

For the C-terminal COFRADIC analysis, butyrylation quantification ($^{12}C_4$-butyrylated versus $^{13}C_4$-butyrylated) was carried out using the Mascot Distiller Quantitation tool (version 2.2.1). For the shotgun experiment, the Mascot Distiller Quantification tool was used to perform the propionylation quantification ($^{12}C_3$-N-propionylated versus $^{13}C_3$-N-propionylated). The quantification method details were as follows: constrain search, yes; protein ratio type, average; report detail, yes; minimum peptides, 1; protocol, precursor; allow mass time match, yes; allow elution shift, no; all charge states, yes; and fixed modifications, mass values. Ratios for the proteins were calculated by comparing the extracted ion chromatogram peak areas of all matched light peptides with those of the heavy peptides, and the results were all verified by visual inspection of mass spectrometry spectra.

3.4. RESULTS

It was demonstrated that CCP1, CCP4 and CCP6 are active carboxypeptidases and are able, not only to shorten the posttranslationally added polyglutamate side chains of tubulin, but also to act over gene-encoded C-terminal glutamate stretch of two important regulators of myosin functions; MLCK and telokin (Rogowski et al., 2010). They proposed that additional proteins, with multiple glutamate residues at their C-termini, might also be substrates of these enzymes.

To test this hypothesis, we made use of the recently developed C-terminal COmbined FRActional DIagonal Chromatography (COFRADIC) technology (Van Damme et al., 2010). The study of the substrates of a protease has generally been hypothesis-driven and, thus biased and limited in terms of throughput and sensitivity. In addition, biochemical substrate characterization might render substrates that might not be cleaved *in vivo (Schilling and Overall, 2007).* The development of proteomic tools has helped to deal with the complexity present in biological samples, allowing for protease substrate discovery at the biochemical, cellular, tissue or organism level. Particularly, terminal proteomics focuses in the study of protein termini and, given that proteolysis generates neo protein termini, is able to follow proteolytic events in a complex proteome. C-terminal COFRADIC technology, a terminal proteomics approach, enriches for C-terminal peptides from whole-proteome digests, allowing for the analysis of C-terminal protein modifications and C-terminal protein processing events (Van Damme et al., 2010). Thereby, it enables substrate determination of endo- as well as carboxy-peptidases (the latter substrate-category was previously inaccessible by means of N-terminal terminal proteomics strategies). Indeed, this technique was recently proven to be able to specifically identify hCPA4 protein substrates in a PC3 cell lysate *in vitro* (Van Damme et al., 2010).

We applied the C-terminal COFRADIC technology to analyze the C-terminal processing at the protein level, as a consequence of CCP1 overexpression in a human cell line. Berezniuk *et al.* showed that HEK 293T cells are a good cellular system to study the consequences of CCP1 actions (Berezniuk et al., 2012). This cell line, showed

the highest relative level of CCP1 mRNA, among the cell lines tested. In addition, the other CCPs are either present at low levels or absent. (Berezniuk et al., 2012). As a result this cell line, normally expressing CCP1, should contain any co-factors and/or binding partners that may be necessary for CCP1 function. Besides, when considering a possible redundant function of the different CCPs, the low levels of other CCPs mRNAs make HEK 293T cells suitable to study the consequences of CCP1 knock-down.

We transiently transfected full length CCP1 and analyzed the effect generated over the levels of the different tubulin forms after 72 h of transfection. Western-blot analyses showed that CCP1 generated increased levels of Δ2-tubulin (Fig. 1), as has been previously shown (Berezniuk et al., 2012; Rogowski et al., 2010). No effect on the levels of other tubulin forms (i.e., Tyr-tubulin, deTyr-tubulin, polyglutamylated tubulin, polyglycylated tubulin) could be observed (data not shown).

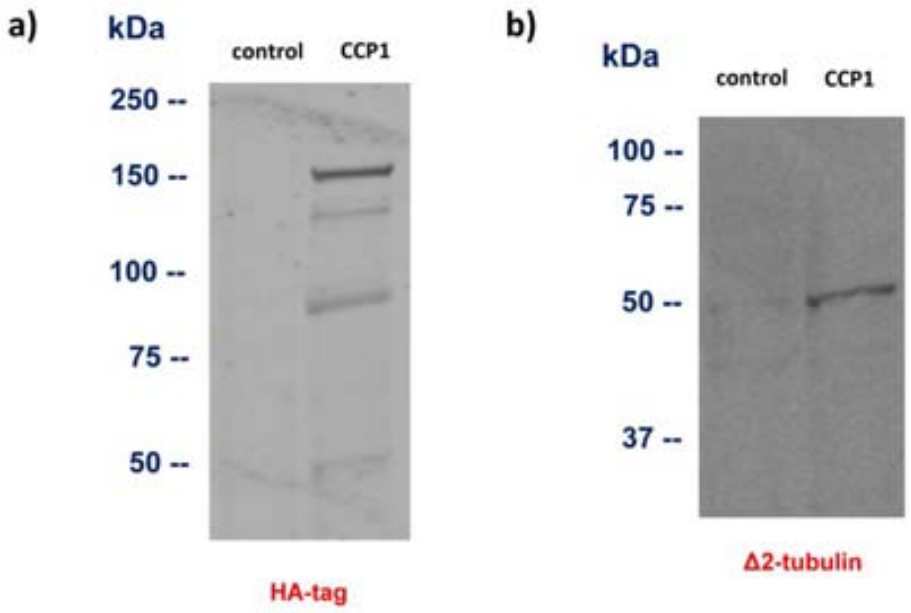

Fig. 1. CCP1 overexpression generates increased levels of Δ2-tubulin. a) CCP1 transfection in HEK 293T cells. Extracts from control cells or cells expressing HA-tagged CCP1 were prepared. Equal amounts of each extracts were analyzed by western blot using an HA antibody. b) Generation of Δ2-tubulin. Extracts from control cells or cells expressing HA-tagged CCP1 were were analyzed by western blot using a Δ2-tubulin antibody.

Of note is that when CCP1 is over-expressed degradation bands of approximately 120-130 kDa and 80-90 kDa can be observed. Western blots are performed using an

antibody against the HA-tag, which is located at the C-terminus of the protein, thereby the here observed proteolytic events might take place close to the N-terminus of the protein. When HEK 293T cells extracts were tested against a commercial CCP1 polyclonal antibody, a single major immunoreactive band at 140-150 kDa is observed (Fig. 2). Here again, equivalent lower molecular mass bands immunoreactive to CCP1 antibody can be observed. Similarly, Wu *et al.* reported the presence of lower immunoreactive bands for mouse CCP1, both in recombinant protein preparations and in *pcd* mouse protein extracts (in this case using a made in-house mouse CCP1-specific polyclonal antiserum) (Wu et al., 2012). Moreover, Rogowski *et al.* overexpressed EYFP-tagged murine CCP1 and degradation bands can also be seen in this case (Rogowski et al., 2010). Here, western blots recognize N-terminal EYFP, so the high molecular weight bands shown recognize proteolytic events at the C-terminus of CCP1.

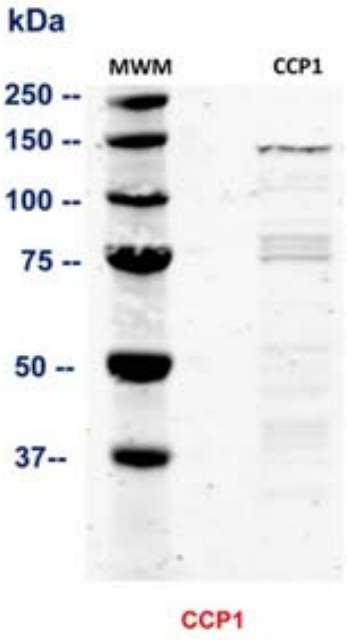

Fig. 2. Western-blot analysis of HEK 293T cell extracts using a CCP1polyclonal antibody.

Provided that the CCP1 overexpressed in HEK 293T cells is able to degrade tubulin C-terminus, we further proceeded with the C-terminal COFRADIC-based proteomics analysis (Van Damme et al., 2010) to study whether CCP1 is able to proteolytically process other proteins. Control HEK 293T cell extracts and HEK 293T cells overexpressing CCP1 extracts were processed in parallel. Briefly, following acetylation

of primary amines, the proteome is digested with trypsin, which now only cleaves after arginine residues. Then, a SCX step enriches for amino-blocked (i.e., acetylated) and α-amine-free C-terminal peptides from other peptides present in whole-proteome digests (Dormeyer et al., 2007; Staes et al., 2008). These peptides are not retained in the SCX column because they lack a net positive charge at low pH (Staes et al., 2008), and are thus collected in the flow-through.

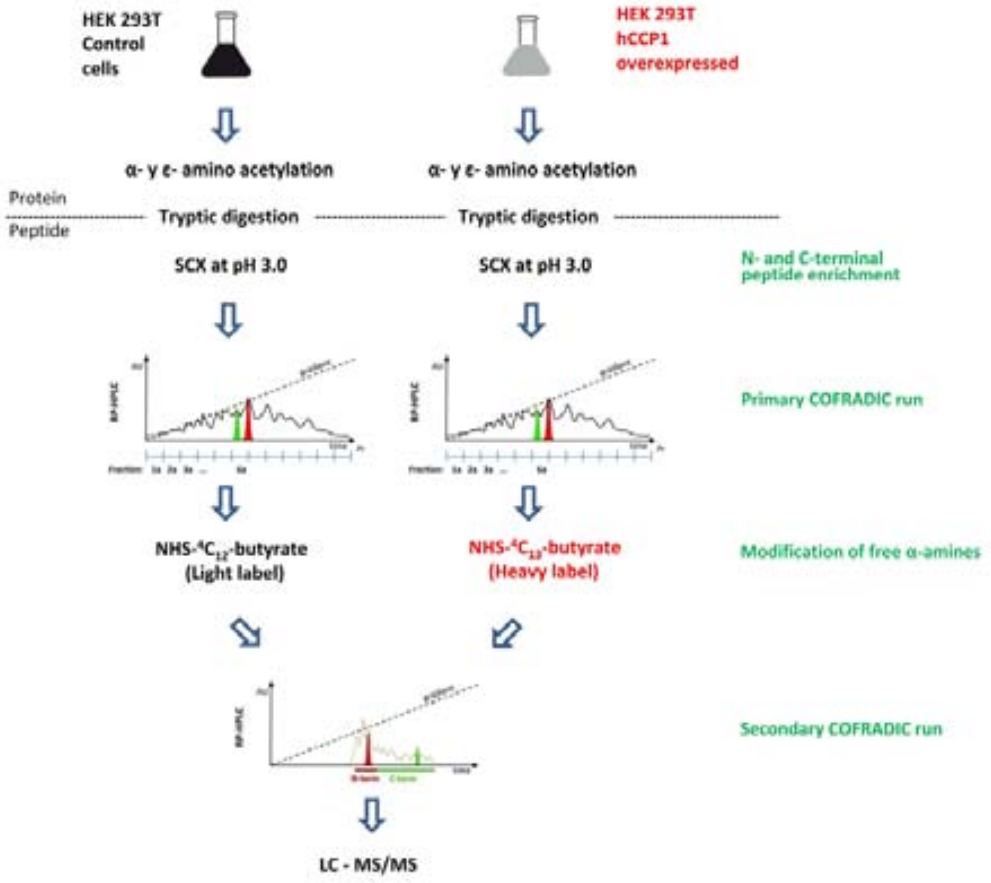

Fig. 3. Schematic representation of the C-terminal COFRADIC terminal proteomics procedure. See text for detail. In the RP-HPLC chromatograms, C-terminal tryptic peptides are represented in green and N-terminal peptides in red.

Further, C-terminal peptides were separated from N-terminal peptides by performing a specific COFRADIC method. Enriched protein termini were first fractionated in two parallel RP-HPLC runs, using equal amounts of peptide material from the control and CCP1 overexpression proteome. After this, each primary fraction were reacted with an N-hydroxysuccinimide ester of butyric acid, butyrylating only the C-terminal peptides (they differ from N-terminal peptides by the presence of a primary α-amine). At this point, C-terminal peptides were differentially labeled by using N-

hydroxysuccinimide (NHS) esters of $^{12}C_4$ or $^{13}C_4$ butyric acid, for the control proteome and CCP1 overexpressing proteome respectively. Corresponding fractions in time, from the two different proteomes, are pooled and re-separated using identical RP-HPLC conditions in a unique RP-HPLC run. While N-terminal peptides elute in the same time interval as the original primary fraction interval, butyrylated C-terminal peptides elute in an interval 4–12 min after the original interval. Thereby, N- and C-terminal peptides are fractionated in distinct secondary fractions and are analyzed by LC-MS/MS. Stable isotope-labeling of both proteomes with isotopic variants the NHS-butyrate allows for MS-based relative quantitation of the isolated C-terminal peptides.

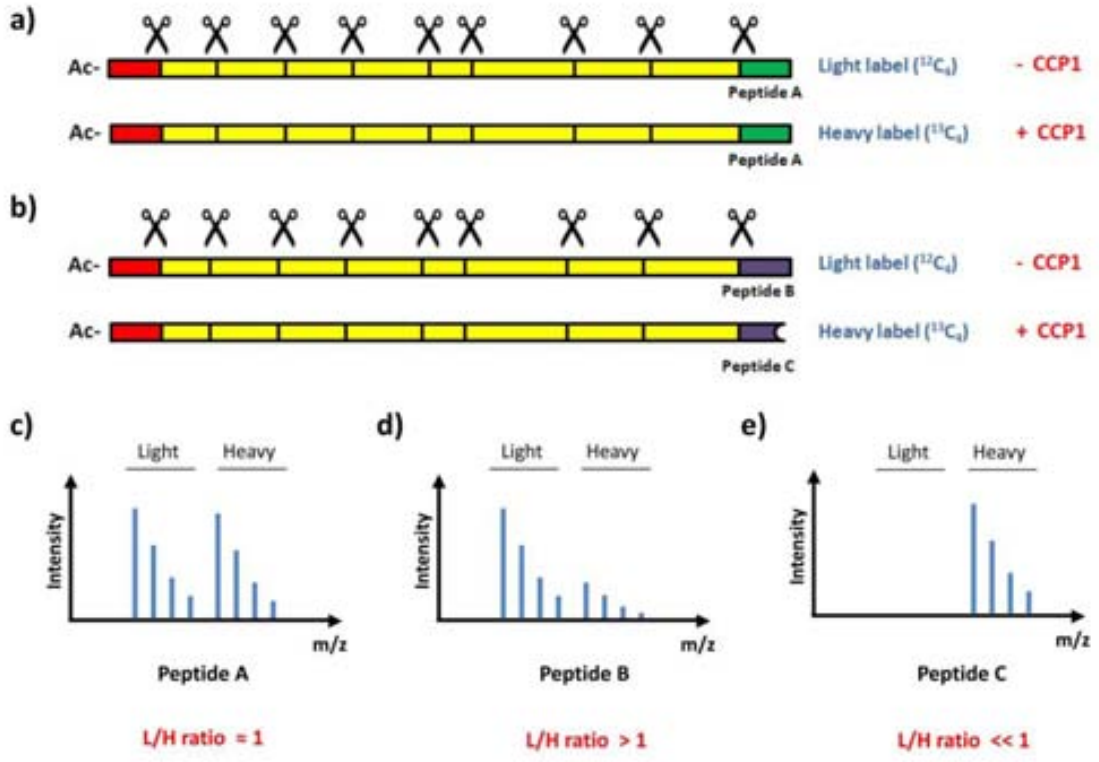

Fig. 4. Identification of C-terminal proteolytic processing in the C-terminal COFRADIC approach. a) Representation of a protein that is not processed in both proteomes. b) Representation pf a protein that is C-terminally processed at a higher extent in the sample where CCP1 is overexpressed. Scissors indicate trypsin cleavage sites. N-terminal tryptic peptides are colored in red. C-terminal tryptic peptides are colored in green/blue. Internal tryptic peptides are colored in yellow. c) Mass spectrum representation of a C-terminal peptide that is not processed in both proteomes. This MS would equally represent a C-terminal peptide showing basal degradation in both proteomes by a protease different from CCP1. d) Mass spectrum representation of the intact C-terminal peptide of a protein that is processed by CCP1 e) Mass spectrum representation of the neo-C-terminal peptide of a protein that is processed by CCP1.

Different possibilities can be considered for protein C-termini when both proteomes are compared. The majority of protein C-termini will not be processed by CCP1 (Fig. 4a) and, as a consequence, the corresponded isolated C-terminal peptides will be present at highly comparable concentrations in both proteomes (isotopic couples) (Fig. 4c). Similarly proteins with basal C-terminal degradation (by a different carboxypeptide or protease) will show of equal intensity in MS-spectra (although this C-termini will differ from database annotated protein C-termini). Those proteins that are substrates of CCP1 will be processed at a higher extent in the proteome where CCP1 is overexpressed (Fig. 4b). As a result, the corresponding intact C-terminal peptide (database annotated protein C-terminal peptide) will be found at higher levels in the control sample (Fig. 4d). Additionally, the corresponding neo-C-terminal peptide (C-terminal peptide generated by proteolytic cleavage of the protease of study) will be present at higher levels in the proteome where CCP1 is overexpressed. In some cases, processing products are not present at detectable levels in the control proteome, and so these neo-C-termini are only identified in the CCP1 overexpressing sample, as "singletons" (Fig. 4e). Neo-C-termini will be the only peptides considered in the search for CCP1 substrates, given that they lead to more confident results.

Isotopic butyrate labels are quantified during MS analysis, giving the relative amount of peptide derived from each proteome. While most of the peptides give L/H ratios that are closely distributed around 1 (Fig. 4c), intact C-termini from CCP1 protein substrates will display L/H ratios significantly >1 (Fig 4d) and neo-C-termini will have L/H ratios significantly <1 (Fig. 4e).

After LC-MS/MS analysis, we identified 1259 unique database-annotated N-termini and 886 unique database-annotated protein C-termini. In addition, we identified 55 processed protein C-termini. Most of them are found at comparable levels in both proteomes, and would correspond either to basal *in vivo* degradation products (derived from the protease activity distinct to CCP1) or to cleavage products generated during the sample analysis procedure. Nine processed C-termini can be considered as neo-C-termini, given that they were present at comparatively much higher levels in the CCP1 overexpressing proteome, and so they lead to putative CCP1 substrates.

As expected, two of these neo-C-termini, correspond to processed α-tubulin C-termini. As previously shown in Fig. 1b and further described in literature (Berezniuk et al., 2012; Rogowski et al., 2010; Wu et al., 2012), CCP1 is known to process the C-terminus of α-tubulin. C-terminal COFRADIC identified the processed C-terminus of α-tubulin 1B (EDMAALEKDYEEVGVDSVEGEGEEEG) with a L/H ratio of 0.012 (Fig. 5a). Moreover, C-terminal COFRADIC identified a neo-C-terminus (EDMAALEKDYEEVGADSADGEDEG) from a different α-tubulin isotype (α-tubulin 1C) with a L/H ratio of 0.015. Both peptides correspond to the C-terminus of Δ3-tubulin, a tubulin form that lacks the last three residues of α-tubulin. The C-terminus of α-tubulin is known to participate in a tyrosination cycle, which consists in the cyclic removal of the C-terminal tyrosine residue of the α-tubulin by tubulinyl-Tyr carboxypeptidase (TCP) and the readdition of this tyrosine by tubulin-tyrosine ligase (Fig. 5b). It has been suggested that CCP1 would not participate in tubulin detyrosination (Rogowski et al., 2010), and it was proposed that a different CCP (CCP2) would perform this function (Sahab et al., 2011). CCP1 would act over Glu-tubulin, removing the penultimate Glu residue of α-tubulin and generating Δ2-tubulin (Rogowski et al., 2010). Our results suggest that CCP1 is able to further process α-tubulin, by generating Δ3-tubulin out of Δ2-tubulin (Fig. 5b), as previously reported (Berezniuk et al., 2012). Intact α-tubulin peptides are found at comparable levels in both proteomes, correlating with our WB analysis showing equal levels of Tyr-tubulin.

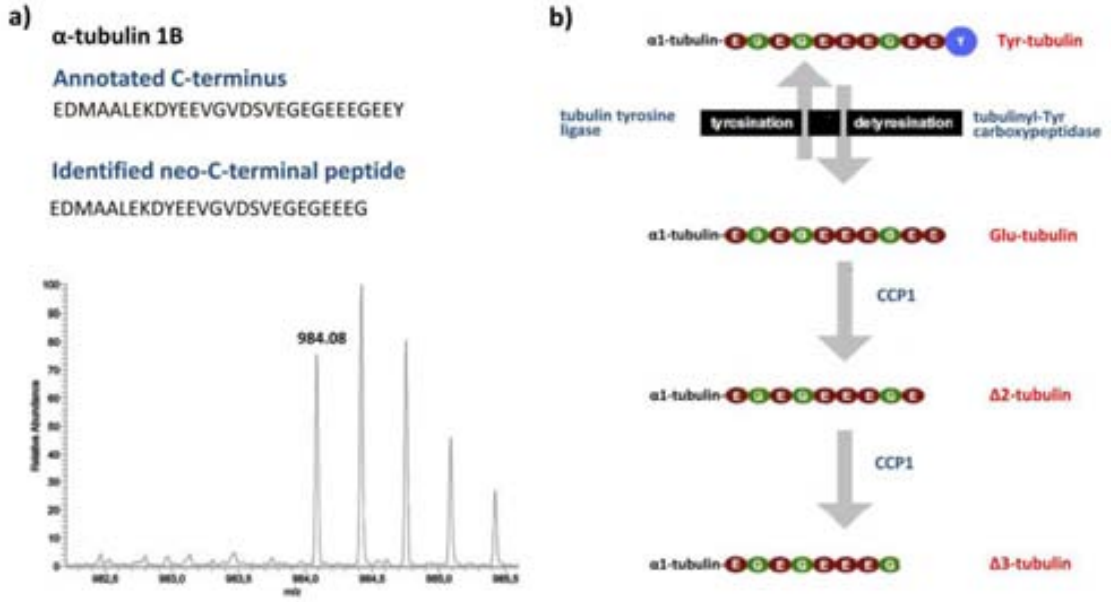

Fig. 5. Identification of α-tubulin as CCP1 substrate. a) Mass spectrum of the neo-C-terminus of α-tubulin 1B, But$^{13}C_4$-EDMAALEKDYEEVGVDSVEGEGEEEG (residues 423-448; But$^{13}C_4$, α-amino group modified $^{13}C_4$ butiric acid) b) C-terminal processing of α-tubulin.

Table 1. Neo-C-termini putative products of CCP1 carboxypeptidase activity

Protein	Accesion number	Identified neo-C-terminus	Cleaved amino acids	L/H ratio
Eukaryotic translation initiation factor 4H	Q15056	EEVVQKEQ	E	0.012
Stathmin	P16949	KNKESKDPADETEA	D	0.029
40S ribosomal protein S9	P46781	KNAKKGQGGAGAGDDE	EED	0.0025
		KNAKKGQGGAGAGDD	EEED	0.0090
TRAF-type zinc finger domain-containing protein	O14545	TAKAKPSKQQGAGDA	EEEEEE	0.14
High mobility group protein B3	O15347	KKVEEEDE	EEEEEEEEEEEEEDE	0.083
		KKVEEED	EEEEEEEEEEEEEDE	0.0039

A list of 7 additional neo-C-termini corresponding to 5 different putative CCP1 protein substrates is shown in Table 1. Notably, in all cases the amino acids released by CCP1 action would correspond to acidic residues, a substrate specificity that partially agrees with previous reports. We identified two peptides indicative of CCP1 mediated removal of only one amino acid from protein C-terminus. One of them shows the release of a Glu residue from the C-terminus of the eukaryotic translation initiation factor 4H (eIF4H). In the case stathmin, an aspartic acid would have been cleaved. Previous reports have shown the action of CCP1 over C-terminal Glu residues (Berezniuk et al., 2012; Rogowski et al., 2010; Wu et al., 2012), although it has been suggested that CCP1 it is not able to able to act over Asp residues (Rogowski et al., 2010; Wu et al., 2012).

Two peptides corresponding to processed C-termini of 40S ribosomal protein S9 (RS9) were identified (Table 1). Here, CCP1 would have cleaved 3 and 4 amino acids (i.e, C-terminal Asp and two or three subsequent Glu residues). Given that M14 CPs release one unique amino acid at a time, it is expected that CCP1 would release these amino acids in a sequential mode.

TRAF-type zinc finger domain-containing protein (TRAD1) is a putative CCP1 protein substrate. It contains a C-terminal stretch of six Glu residues and CCP1 seems to be able to remove all of them. Remarkably, we identified two neo-C-termini of the

high mobility group protein B3 (HMGB3), which show that up to 15 or 16 aminoacids have been processed by CCP1.

Table 2. Intact C-termini of putative CCP1 protein substrates

Protein	Identified C-terminal peptide	L/H ratio
Eukaryotic translation initiation factor 4H	EEVVQKEQE	0.95
Stathmin	KNKESKDPADETEAD	0.56
40S ribosomal protein S9	KNAKKGQGGAGAGDDEEED	0.83
High mobility group protein B3	KKVEEEDEEEEEEEEEEEEDE	0.76

Contrary to expected, the corresponding intact C-termini of the putative CCP1 substrates are not in any case significantly altered, when both proteomes are compared (Table 2). These were expected to give L/H ratios >1, with a ratio inversely proportional to the extent of proteolysis. Conversely, this would indicate that the extent of proteolysis of CCP1 substrates is very low, similarly to what we saw in the case of tubulin.

In parallel, we performed a shotgun proteomics experiment with the purpose getting more accurate relative quantification of protein between both proteomes. This approach, as before, uses stable isotope labeling for quantitative comparison of the different samples. Here, after endoproteinase Lys-C protein digestion, we employed isotopic variants of N-hydroxysuccinimide propionate (propionylation) to differentially label primary α-amines (peptide N-termini) and ε-amines (lysine side-chains) (Ghesquiere et al., 2009). Then labeled samples were mixed together and fractionated via RP-HPLC for subsequent LC-MS/MS analysis.

In the shotgun proteomics approach, the analyzed peptide mixture contains endoproteinase Lys-C digestion peptides: N-terminal and C-terminal peptides (together with neo-C-terminal peptides in the CCP1 experiment) plus an excess of internal digestion peptides. As a result of the LC-MS/MS analysis, we identified 3853 unique peptides. From these 127 corresponded to database-annotated C-termini and 18 to processed C-termini. From the latter, only one peptide can be considered neo-C-termini: the peptide RKRETDDEG from the acidic leucine-rich nuclear phosphoprotein

32 family member B (Uniprot Q92688) that displays an L/H ratio of 0.016. Consequently, CCP1 would cleave this protein releasing 3 residues: two last Asp residues and the penultimate Glu residues (-EDD-COOH).

Given the complex and unpredictable downstream effects of the increase of CCP1 activity in vivo, these 6 new proteins can only be considered putative CCP1 substrates that require careful validation. We chose two of these putative substrates to further work on their validation: TRAD1 and HMGB3.

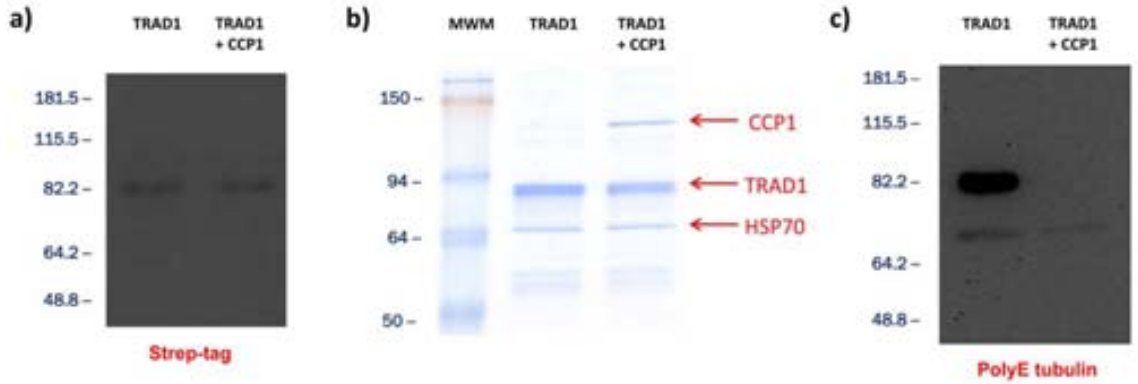

Fig. 6. CCP1 co-expression generates trimming of TRAD1 C-terminus. a) TRAD1 transfection in mammalian cells. HEK 293T cells were transfected with a plasmid expressing *strep*-tagged TRAD1 alone or co-expressed with CCP1. Equal amounts of each cell extract were analyzed by western blot using a *strep*-tag antibody. b) Analysis of affinity-purified TRAD1 by SDS-PAGE. Purification of TRAD1 from HEK 293T cells transfected with a plasmid expressing strep-tagged TRAD1 alone or co-expressed with CCP1. The identity of TRAD1 and contaminating bands (indicated in red) were confirmed by peptide mass fingerprinting analyses of protein bands directly excised from the gel. c) Analysis of the C-terminal integrity of affinity-purified TRAD1. Equal amounts of TRAD1 were analyzed by western blot using a polyE antibody.

For TRAD1 validation, an N-terminally *strep*-tagged version of the protein was expressed in HEK 293T cells in the presence or absence of CCP1 co-expression. Western-blot analysis of the derived protein extracts shows no effect in the electrophoretic mobility of TRAD1 protein (Fig. 6a). The polyE polyclonal antibody was designed to recognize glutamate side chains of at least three residues in tubulin (Shang et al., 2002), but Rogowsky *et al.* demonstrated that this antibody additionally recognizes gene-encoded C-termini with three or more glutamates (Rogowski et al., 2010). We used this antibody to evaluate the integrity of TRAD1 C-terminus, but due to sensibility problems we partially purified TRAD1 by means of its N-terminal *strep*-tag to perform this evaluation. Analysis of the affinity-purified TRAD1 shows that this protein

is efficiently purified in both proteomes, with HSP70 as a co-purifying protein (Fig. 6b). When CCP1 was co-expressed, CCP1 also appears as a co-purifying protein because this protein contains a *strep*-tag in its C-terminus (Fig. 6b). The polyE antibody recognizes purified TRAD1 when expressed alone but does not recognize TRAD1 when co-expressed with CCP1 (Fig. 6c). This is compatible with the ability of CCP1 to trim the C-terminal glutamate stretch of TRAD1.

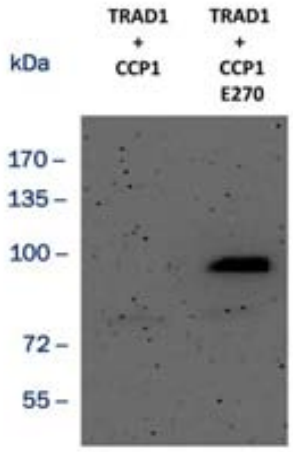

Fig. 7. *In vitro* validation of TRAD1 C-terminal processing by CCP1. TRAD1 was incubated 2 h at 37 °C with active and inactive CCP1, and by western blot using a polyE antibody.

However, C-terminus processing may be a result of either direct or indirect effects of CCP1 expression. For this reason, validation of true substrates is generally carried out by *in vitro* assays using purified proteins. Hence, affinity-purified TRAD1 was incubated at 37 °C for 2 hours with affinity-purified CCP1. As a control, TRAD1 was incubated with affinity-purified CCP1-E270. CCP1-E270 is an inactive mutant in a key catalytic residue for M14 metallocarboxypeptidases (E270Q using the numbering system of bovine CPA, E1102Q in human CCP1), a residue that polarizes the metal-bound water molecule responsible of the scissile peptide bond attack (Wu et al., 2012). TRAD1 is no longer recognized by the polyE antibody after CCP1 incubation, but not with inactive CCP1 (Fig. 7), suggesting again that CCP1 is able to act over TRAD1 C-terminus.

Similarly, for HMGB3 validation, an N-terminally *strep*-tagged version of the protein was co-expressed in HEK 293T cells with either active or inactive CCP1.

Western-blot analysis of the derived protein extracts shows that in addition to the HMGB3 band a second band with different electrophoretic mobility can be observed when HMGB3 is co-expressed with CCP1 (Fig. 8a). This band is not present in the case of the inactive CCP1, and is consistent with the observed processing of up to 16 acidic aminoacids in HMGB3 C-terminus (the *strep*-tag is present in the N-terminus of the protein). No *in vitro* validation of HMGB3 C-terminal processing could be observed when HGMB3 was assayed in similar conditions as those described for TRAD1 in vitro assay.

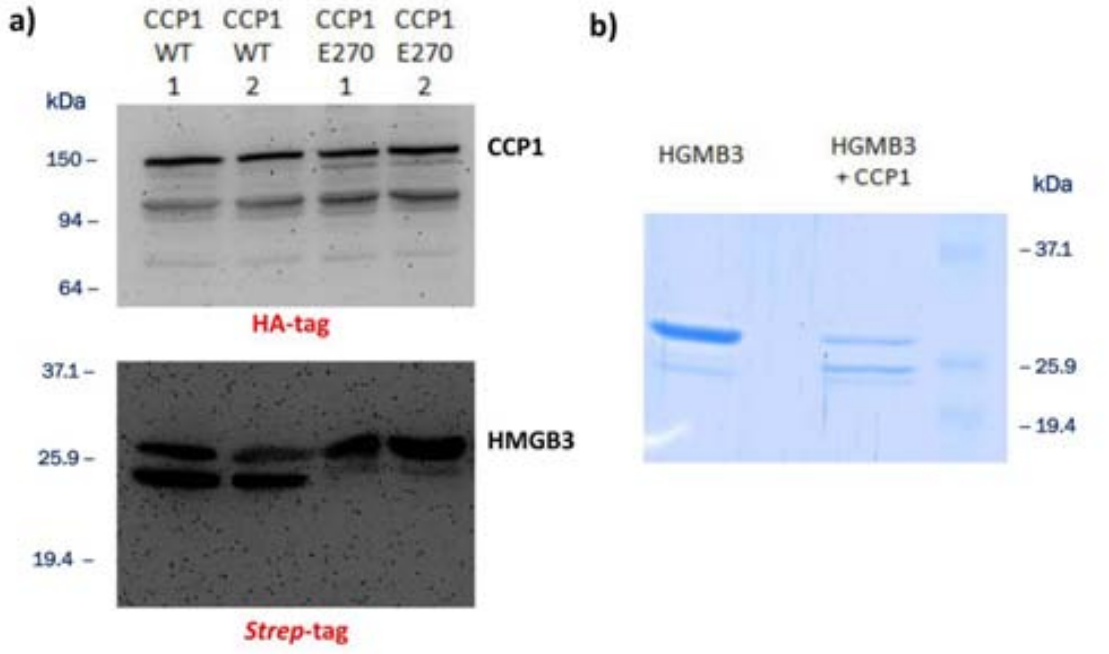

Fig. 8. CCP1 co-expression generates trimming of HMGB3 C-terminus. a) HMGB3 transfection in HEK 293T cells. A plasmid expressing *strep*-tagged HMGB3 was co-transfected with active (CCP1 WT) or inactive (CCP1 E270) CCP1. Equal amounts of each cell extract were analyzed by western blot using a HA-tag antibody, to show equal expression levels of both CCP1 variants. Western blotting using a strep-tag antibody shows HMGB3 processing when co-espressed with active CCP1. b) Analysis of affinity-purified TRAD1 by SDS-PAGE. Purification of HMGB3 from HEK 293T cells transfected with a plasmid expressing strep-tagged HMGB3 alone or co-expressed with CCP1. The identity of HMGB3 in all bands was confirmed by peptide mass fingerprinting analyses of protein bands directly excised from the gel.

Given that some partial processing was observed when HMGB3 was expressed with inactive CCP1, HMGB3 was affinity purified by means of its N-terminal *strep*-tag and analyzed by SDS-PAGE (Fig. 8b). When HMGB3 is co-expressed with CCP1, a band of lower molecular weight appears, similarly to what the western blot analysis has

shown. Interestingly, when HMGB3 is expressed alone, a degradation band can be observed (also present when CCP1 is co-expressed). In this case, processing in HMGB3 C-terminus seems to go further, since this degradation band shows a slightly lower molecular weight than that generated upon CCP1 co-expression. The identity of HMGB3 in all bands was confirmed by peptide mass fingerprinting analyses of protein bands directly excised from the gel.

4.4. DISCUSSION

It was previously hypothesized that the action of CCPs over gene-encoded C-termini could be a general feature for proteins ending in glutamates stretches (Rogowski et al., 2010). In this report, we identified six putative CCP1 substrates, all of them containing acidic stretches on their C-termini. This expands the substrate specificity of CCP1 to all acidic residues, since CCP1 would not only be able to act over glutamate residues, but also over C-terminal aspartate residues.

To the best of our knowledge, this is the first proteomic study aiming at identifying carboxypeptidase substrates in a cellular context. Degradomics discovery of carboxypeptidase substrates in complex proteomes was challenging until the development of approaches for protein C-terminal peptide enrichment (Schilling et al., 2010; Van Damme et al., 2010; Xu et al., 2011). The C-tails approach was used to investigate constitutive C-terminal proteolysis in *E. coli* (Schilling et al., 2010). The C-terminal COFRADIC technology proved to be useful to screen for protein substrates of human carboxypeptidase A4 in human PC3 cell lysates, when assayed *in vitro*. Here, C-terminal COFRADIC successfully identified neo-C-termini generated by CCP1 when expressed in HEK 293T cells.

Two different proteomic approaches were used in this study: a shotgun approach and a terminal proteomics approach (C-terminal COFRADIC). C-terminal COFRADIC, by enriching in C-terminal peptides from other non-informative tryptic peptides, increases the number of identified peptides and focuses the analysis in those peptides that directly point to CCP1 proteolytic action. Here, C-terminal COFRADIC was able to identify 9 neo-C-termini, while the shotgun approach provided information about only one C-terminus. However, the latter was not present among the 9 neo-C-termini identified by C-terminal COFRADIC. This can be explained by the use of a different protease (endoproteinase Lys-C vs trypsin) that generates different C-termini. In addition, the sample analysis procedure differs in both approaches, and then may favor the identification of peptides with certain characteristics.

Proteomic based screening of in vivo substrates performed by modulating the activity of the protease of interest in vivo, have unpredictable downstream effects as a consequence of the activity of the protease (Doucet et al., 2008; Huesgen and Overall, 2012). In our experiment, an additional level of complexity is added by the effects that CCP1 overexpression may produce in the cell. Correspondingly, the discovered neo-C-termini could be the result of either direct or indirect effects. For this reason, substrate candidates need to be validated by different approaches. A first level of analysis to discriminate indirect effects comes from the known properties of our protease (i.e., subcellular localization, substrate specificity). Table 3 summarizes the here identified putative CCP1 substrates and their properties. To enrich the analysis, MLCK1/telokin a previously identified substrate (Rogowski et al., 2010), not identified in this study was included in this table. When analyzing the subcellular localization of these candidates, all of them display either a cytoplasmic or nuclear localization, consistent with the described subcellular localization of CCP1 (Harris et al., 2000; Kalinina et al., 2007).

Table 3. Putative CCP1 substrates

Protein	Subcellular localization	Cleaved aminoacids	Function	*In vivo* Evidence
Tubulin	Cytoplasm	EEY	Microtubules	Yes (Paturle-Lafanechere et al., 1994)
Eukaryotic translation initiation factor 4H	Cytoplasm	E	Translation initiation	Yes (Richter-Cook et al., 1998)
Stathmin	Cytoplasm	D	Microtubules regulation	No
Protein 40S ribosomal S9	Cytoplasm	EEED	Translation initiation	No
TRAF-type zinc finger domain-containing protein	Cytoplasm	EEEEEE	Regulation TLR4 y RLH pathways	No
High mobility group protein B3	Nucleus	EEEEEEEEEEEEED	Transcription regulation, chromatin	Yes
Acidic leucine-rich nuclear phosphoprotein 32 family member B	Nucleus	EDD	Histone chaperone. Regulation of cell cycle progression.	No
Myosin light chain kinase 1/ Telokin	Cytoplasm	EEEEEE	Myosin regulation (microtubules)	Yes (Rusconi et al., 1997)

The derived substrate specificity shows that CCP1 restricts its preferences to acidic residues, which partially corresponds to previous reports. In this analysis, we consider that in the case of tubulin a different enzyme is responsible of the release of the C-terminal Tyr (Rogowski et al., 2010; Sahab et al., 2011). Substrate preference can also be visualized in the form of iceLogos (Colaert et al., 2009) (Fig. 9). For the iceLogo praparation we took into consideration that MCPs sequentially release aminoacids. As a result, when a few aminoacids are released to get to the final identified product, intermediate products of digestion act as new substrate and are further digested by CCP1. Consequently, intermediate products of digestion were considered as different individual substrates when preparing the iceLogo. The derived substrate specificity shows a strict preference for acidic aminoacids (Fig. 9). The iceLogo displays, in addition, an extended substrate specificity, with a critical role of P1 to P5 in substrate selection. This result agrees with a recent report that describes that for mouse CCP1 biotin-3E is a better substrate than biotin-2E, thereby showing its preference for longer acidic tails (Wu et al., 2012).

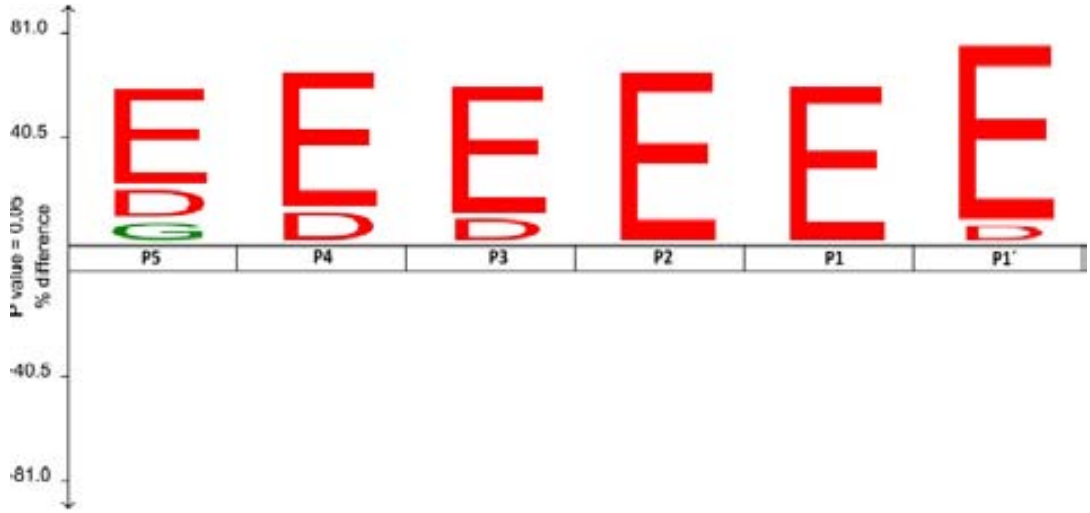

Fig. 6. Substrate specificity profile derived from putative CCP1 substrates. This representation shows the enriched residues present at the different identified CCP1 putative substrate positions as compared to the reference set. 37 CCP1 substrates were considered (when intermediate products of digestion were included as new substrates) and the iceLogo was created using the human Swiss-Prot database as reference set. In the representation the substrate residues are depicted according to Schechter & Berger nomenclature (Schechter and Berger, 1967). The frequency of the amino acid occurrence at each position in the sequence set was compared with the occurrence in the reference set. Only statistically significant residues with a p-value □ 0.05are plotted. Amino acids height shows the degree of difference in the frequency of an amino acid in the experimental set as compared to the frequency in the reference set. Residues that are statistically over- or underrepresented in the experimental set are respectively shown in the upper or lower part of the iceLogo.

The iceLogo shows that although both glutamate and aspartate residues seem to be cleaved, CCP1 displays a marked preference of glutamates over aspartates (Fig. 9). This probably helps to explain the discrepancy of these results with previous reports, which stated that CCP1 was not capable of releasing C-terminal aspartate residues (Rogowski et al., 2010; Wu et al., 2012). We hypothesized that CCP1 cleaves aspartate residues but much less efficiently than glutamate residues. This is partially supported by the iceLogo representation of CCP1 substrate specificity. In addition, although CCP1 seems to cleave aspartate residues of RS9 and HMGB3, their identified final products also show that CCP1 strongly slows down its proteolytic activity or may even stop it when it encounters an aspartate residue in P1 or P1' (Table 1).

In general, M14 MCPs with preference for acidic aminoacids, are able to cleave both glutamate and aspartate residues, although with different catalytic efficiency. The HaCA42 protein, a MCP from *Helicoverpa armigera,* is highly specific for C-terminal glutamate residues from peptides but was shown to slowly hydrolyze the C-terminal aspartate residue of the angiotensin 1 converting enzyme inhibitor peptide (Bown and Gatehouse, 2004). The recently described human and zebrafish CPO enzymes, are considered to cleave both aspartate and glutamate residues, based on the similar inhibitory effect that both amino acids exert on their activity (Lyons and Fricker, 2011). In all these cases, substrate specificity for acidic amino acids is mainly determined by an Arg at the position equivalent to residue 255 in bovine CPA. An Arg at this position is proposed to be determinant for CCP1 preference for acidic aminoacids (Wu et al., 2012).

As part of the validation process, it could be verified that the generation of the neo-C-termini and the ratios provided by C-terminal CORADIC are correct. This is typically done by Western blotting analysis, but in our case no antibody was available for this purpose. So, we co-transfected recombinant forms of the putative substrates with CCP1, to verify if in a cellular context the identified C-terminal trimming were true events. However, these might be true proteolytic events but the protein might still not be a CCP1 substrate, and be processed by a different protease as a result of downstream effects. To test this, purified substrates and CCP1 were assayed *in vitro*.

Analysis of the co-trasfection experiments provides results that are compatible with the identified C-terminal processing of both TRAD1 and HMGB3. This suggests that C-terminal COFRADIC provide real neo-C-termini. This is further supported by the identification as a substrate of tubulin, a previously described substrate of CCP1.

In vitro experiments validated TRAD1 (with a C-terminal stretch composed only by glutamate residues) as a CCP1 substrate (Fig. 7). This suggests that the identified neo-C-termini are a result of the direct action of CCP1 over these proteins. The previously discussed analysis of subcellular localization and substrate specificity is also in favor of this possibility. However, preliminary results failed to validate HGMB3 (that contains both aspartates and glutamates in its C-terminus) *in vitro.* This could be explained by the proposed lower efficiency of aspartate cleavage, which hampers to see *in vitro* cleavage. Another plausible explanation is that a different protein is responsible for aspartate release in the cell, helping CCP1 in the cleavage of HMGB3. Then CCP1 would only be able to cleave glutamate residues and thus *in vitro* validation of HMGB3 fails. The identified neo-C-terminus of stathmin, in which only one aminoacid is released (and corresponds to an aspartate), is not supportive of this hypothesis.

The determination of the cleavage site specificity derived from the putative CCP1 substrates allows prediction of novel substrates for this enzyme. The UniProt protein database was searched for proteins that contain an acidic strecht in its C-terminus. Table 4 displays 37 proteins that contain acidic C-terminal tails of at least five glutamates/aspartate residues. This number is higher than expected: among the 70000 human proteins in the UniProt database, less than one protein is expected to have these characteristics. These 37 proteins represent a number almost 53 times higher than that expected by probability. What is more, finding proteins with more than 10 C-terminal acidic residues is highly unlikely; however 9 human proteins have acidic stretches of at least 10 residues. Of note is the case of high mobility group protein B1 (UniProt P09429) that contains 30 acidic residues in its C-terminus.

Table 4. Human proteins with long acidic C-terminal tails

Number	Name	UniProt	C-terminus
1	High mobility group protein B1	P09429	SKKKKEEEEDEEDEEDEEEEEDEEDEDEEEDDDDE
2	Putative high mobility group protein B1-like 1	B2RPK0	SKKKKEEEEDEEDEEDEEEEDEEDEEDDDDE
3	High mobility group protein B2	P26583	SKKKNEPEDEEEEEEEEDEDEEEEDEDEE
4	High mobility group protein B3	O15347	ARKKVEEEDEEEEEEEEEEEEEDE
5	Homeobox protein Hox-A7	P31268	TAAADKADEEDDDEEEEDEEE
6	Cytosolic purine 5'-nucleotidase	P49902	PQEITHCHDEDDDEEEEEEEE
7	Translocation protein SEC63 homolog	Q9UGP8	QEDSEGFEDSFEEEEEEEEDDD
8	mTERF domain-containing protein	Q7Z6M4	DEDDNDEDDDDEDDDEAEDNDEDEDDDEEE
9	60S ribosomal protein L22	P35268	INQDEEEEEDED
10	Calsequestrin-1	P31415	LEDVLEGEINTEDDDDDDDD
11	Calcium-dependent secretion activator 1	Q9ULU8	LQGISMKDSDEEDEEDD
12	Putative high mobility group protein B3-like	P0C6E5	KKVEEEDEDEE
13	RNA polymerase-associated protein LEO1	Q8WVC0	VISDEEEEDDD
14	LIM domain and actin-binding protein 1	Q9UHB6	KRNRYYDEDEDEE
15	UPF0466 protein C22orf32	Q9H4I9	LLEEHDIFVPEDDDDDD
16	Calsequestrin-2	O14958	KINTEDDDEDDDDDNSDEEDNDDSDDDDDE
17	Cell cycle control protein 50B	Q3MIR4	VYIRYQDQDDDDEE
18	Dipeptidyl aminopeptidase-like protein 6	P42658	RIQDKLLTVTAKEDEEED
19	Homeobox protein Hox-B7	P09629	PGTTGQDRAEAEEEEEE
20	Metaxin-1	Q13505	LGMAEEDEEE
21	40S ribosomal protein S9	P46781	GQGGAGAGDDEEED
22	Protein timeless homolog	Q9UNS1	KKRYQIEDDEDD
23	TRAF-type zinc finger domain-containing protein 1	O14545	SKQQGAGDAEEEEEE
24	Ubiquitin thioesterase ZRANB1	Q9UGI0	SLSDGEEDEDDEDE
26	Acidic leucine-rich nuclear phosphoprotein 32 family member E	Q9BTT0	EEEEEEEGGLRGEKRKRDAEDDGEEEDD
27	Interferon alpha-inducible protein 6	P09912	KYLDSEEDEE
28	Myosin light chain kinase, smooth muscle	Q15746	AELIVETMEEGEGEGEEEEE
29	Choline-phosphate cytidylyltransferase A	P49585	KAAAYDISEDEED
30	Protein SSX1	Q16384	VIYEEISDPEEDDE
31	Protein SSX2	Q16385	VIYEEISDPEEDDE
32	Protein SSX3	Q99909	VIYEEISDPEEDDE
33	Protein SSX4	O60224	VVYEEISDPEEDDE
34	Protein SSX7	Q7RTT5	VIYEEISDPEEDDE
35	Protein SSX8	Q7RTT4	VIYEEIRDPEEDDE
36	Protein SSX9	Q7RTT3	VIYEEISDPEEDDE
37	Transmembrane protein 190	Q8WZ59	KESRDVEGGTEGEGTEEGEETEGEEEED

A validation that the here described putative CCP1 are true *in vivo* C-terminal modifications generated by CCP1 is lacking. There is still the possibility that these proteolytic events are a consequence of CCP1 overexpression, and are not biological relevant posttranslational modifications. Different strategies can be considered for this purpose (auf dem Keller and Schilling, 2010; Doucet et al., 2008; Huesgen and Overall, 2012): determination of temporal and spatial colocalization of CCP1 and its substrates in the cell; silencing of CCP1 in the cell and visualization of opposite effects on the putative substrates; *in vivo* detection of these proteolytic processes, possibly using antibodies that recognize these neo-C-termini; and, finally, mouse knockout/knock-in and transgenic models that would provide the ultimate validation of these substrates. Further work is required to completely validate these putative substrates.

We searched in literature for reports describing the primary sequence C-terminal modifications proposed for CCP1 (Table 3). Richter-Cook *et al.* described the purification of rabbit eIF4H, and found a form of the protein with a relative more basic pI, suggesting that it may be the result of limited proteolysis of its C-terminal glutamate (Richter-Cook et al., 1998). Rusconi *et al.*analyzed by MALDI-TOF the C-terminus of chicken telokin and found forms of the protein to which one to six C-terminal glutamate residues had been removed (Rusconi et al., 1997). The here described transfection experiments with HMGB3, show proteolyzed forms compatible with C-terminal trimming (Fig. 8). In the case of tubulin, C-terminal processing is well known (Janke and Bulinski, 2011; Paturle-Lafanechere et al., 1994; Sahab et al., 2012). However, the here identified neo-C-terminus corresponds to Δ3-tubulin a tubulin form that has only been described once (Berezniuk et al., 2012), generated after in vitro incubation of tubulin with CCP1.

The putative substrate HMGB3 provides clues of which could be the functional implications of CCP1 C-terminal processing. Lee *et al.* studied the role of acidic tails in HMG proteins, by comparing different members of the family with different tail lengths and by studying forms to which these tails were recombinantly removed (Lee and Thomas, 2000). HMG forms with longer acidic tails show lower affinity for linear DNA and four-way DNA junctions, and are less effective than in supercoiling and bending DNA. In addition, removing the C-terminal tail increases the DNA binding

affinity of these proteins (Lee and Thomas, 2000). Thereby the acidic tail of HMG protein would modulate DNA binding. In addition, their acidic tail has been associated with nuclear retention (Shirakawa et al., 1997) and in the stimulation of transcription (Aizawa et al., 1994). Marintcheva *et al*., proposed a similar role for the acidic tail of the ssDNA-binding protein of bacteriophage T7, and proposed a model that can be extended to other proteins like ribosomal proteins, translation factors or HMG proteins (Marintcheva et al., 2008). The acidic tail of these protein would regulate the affinity to their binding partners (DNA or other proteins), by binding and thereby shedding their basic binding clefts. In this context, CCP1 by shortening the length of these acidic tails would regulate protein-protein and DNA-protein interactions. Indeed, this is the proposed role for the posttranslational modifications that affect tubulin C-terminus: modulate the binding of motor proteins and MAPs to the C-terminal region of tubulin (Janke and Bulinski, 2011). Furthermore, Rusconi *et al.* proposed that the observed heterogeneity in telokin C-terminus might regulate its interaction with other proteins.

General discussion

Proteases are thought to regulate virtually every biological process; however for most proteases there is limited knowledge about their *in vivo* biological roles. This is particularly true for MCPs, which in general have proposed functions but usually there is no certainty about their real biological roles. The present thesis was aimed to gain insights into the understanding of the biological functions of different MCPs. Notably, we had the opportunity to work with members of the three main subfamilies of MCPs (M14A, M14B and M14D). For this purpose, we applied different approaches that included kinetic studies, cellular biology and structural characterization. In addition, we participate in the development of different proteomic techniques for carboxypeptidase substrate determination. This included a peptidomic and a proteome-derived peptide library approaches for assaying the action of MCPs *in vitro.* In addition, we applied the recently developed terminal proteomics approach (C-terminal COFRADIC) to search for natural substrates of cytosolic carboxypeptidases in a cellular context.

In Chapter 1, we analyzed MCPs from a structural point of view. The elucidation of the three dimensional structures has long been an aspect that provided many advances to the field of MCPs. The early studies on bovine CPA were useful to determinate the catalytic mechanism of these enzymes. It has been important also to define the key residues essential for the MCP function, to investigate determinants of substrate specificity, to analyze the interaction with its inhibitors, or to get insights into their regulation mechanisms, among other structure-function relationships. These studies are in addition important for structure-based approaches for drug discovery or the design of activity-probes. In this chapter we described the structure of the first repeat (catalytic domain plus TLD domain) of the *silver* gene, a CPD homolog in the fruit fly *Drosophila melanogaster.*

This gene was discovered in *Drosophila melanogaster*, as the gene responsible for the *silver* mutation, characterized for adult flies that display cuticles that are pale and silvery in color (due to reduced melanization) and pointed wings. There are in addition, several lethal *svr* mutations in this gene (Wright, 1987). CPD localizes in the trans-Golgi network, and thereby has been proposed to process neuropeptides. Mutation in the *silver* gene would generate impaired neuropeptide processing leading to the observed phenotype (Sidyelyeva et al., 2010). Eight different isoforms of the *silver* protein are

generated due to alternative splicing of the *silver* gene. While the long isoforms are membrane-bound tri-repeats similar to mammalian CPDs, there are two short soluble isoforms that correspond only to the first repeat of CPD (Sidyelyeva and Fricker, 2002). One of them (DmCPD1As) is an inactive form, whereas the other short form (DmCPD1Bs) is active and was here studied.

The only CPD structure previously reported corresponds to the second repeat of duck CPD (Gomis-Ruth et al., 1999). Thus we produced and purified DmCPD1Bs as a recombinant protein in *Pichia pastoris.* This short CPD form was active against typical CP B-type substrates, although with a lower efficiency than the M14A MCP human CPB. The obtained inhibition profile was in accordance with the profile obtained by other M14B enzymes: is inhibited by GEMSA with a *K*i in the µM range and is not inhibited by the reported MCP protein inhibitors. The protein was co-crystallized in the presence of the MCP inhibitor GEMSA. The solved structure of DmCPD1Bs matches to the structure of other M14B metallo-carboxypeptidases. The CD of DmCPD1Bs displays a α8+2/β8-topology typical of MCPs/funnelins, which forms a compact globular shape. There is a funnel-like opening to the left, within which lies the active-site cleft. The observed specificity for basic residues is provided by the side chain of Asp228, located at the bottom of the active-site cleft. The C-terminal TLD displays a rod-shape, with 7 β-sheets forming an all-β seven-stranded β-barrel or β-sandwich.

DmCPD1Bs appeared as a dimer in the crystal (generated from a very concentrated solution of the protein, 18mg/ml); indeed as a pair of dimers in the asymmetric unit. The dimer showed a dimeric interface of ~950 Å2, compatible with the interface of experimentally validated “weak” transient homodimeric proteins. However, the dimer could not be confirmed in solution. The protein seems to be a monomer in size-exclusion chromatography, while dynamic light-scattering and cross-linking experiments provide results compatible with a monomer-dimer equilibrium in solution (with the dimer favored in concentrated solutions), but that are not conclusive. It is worth mentioning that dimer formation is a hypothesized mechanism of regulation for CPD.

A differential structural element of DmCPD1Bs structure is the flexible loop Lα5α6. This is among the loops that shape the entrance to the funnel cavity. This loop appears as a flexible loop, which is consistent with the high susceptibility of this loop to be

proteolyzed by contaminanting proteases in the recombinant protein. There are not known mechanism of regulation for M14B MCPs, since they are synthesized as active enzymes and there is no known protein inhibitor for these enzymes. The regulation of the catalytically active forms is believed to rely on their localization: most of these enzymes are bound to membrane or to the extracellular matrix. DmCPD1Bs is a soluble isoform and other mechanisms have to be taken into consideration. The high susceptibility to proteolysis shown by this protein let us hypothesize that proteolysis could be a mechanism of proteolysis for this enzyme and similar ones.

In Chapter 2, we studied the enzymatic properties of human of a MCP form a different subfamily. CPA4 is an interesting MCP because of its potential link with prostate cancer aggressiveness, but lacked in-depth enzymatic characterization. The subcellular distribution and secretion of CPA4 was examined in transient transfection assays in human cell lines, showing that this enzyme is predominantly found soluble in the extracellular medium. Its neutral pH optimum is compatible with a function in the extracellular environment rather than in the secretory pathway. M14A MCPs are synthesized as inactive precursors, and CPA4 was found in the extracellular medium in the zymogen form, indicating that activating protease is not produced in sufficient amounts by the cell lines tested. This seems to be the general behaviour for most M14A MCPs, and CPA1 and TAFI are clear examples of this. CPA1 and TAFI are found in the extracelluar medium in the zymogen form when they are expressed in COS-1 (Zhao et al., 2003) and HEK 293 cells (Hamstra and Rehemtulla, 1999) respectively. Procarboxypeptidases are proteolytically activated at the appropriate moment by proteases present in the biological fluids where they perfom their functions. CPA1 is produced is produced in exocrine pancreas, and after secretion is activated by trypsin in the duodenum. TAFI instead is synthesized by the liver, and can be activated by thrombin, thrombin/thrombomodulin or plasmin in circulation (Colucci and Semeraro, 2012). For CPA4 it is not clear yet where it performs its functions (although it is present in many tissues) or its physiological activating enzyme (but it most probably corresponds to a trypsin-like protease).

We were able to profile the previously unknown CPA4 substrate preferences, by using three different approaches to characterize carboxypeptidase substrate specificities: a series of new chromogenic substrates, in vitro assays using biologically

relevant synthetic peptides and a quantitative peptidomics mass spectrometry-based approach. This analysis shows that CPA4 is clearly an A-type carboxypeptidase with specificity for C-terminal aliphatic and aromatic residues. Kinetic studies with chromogenic substrates allowed a detail comparison of its substrates preferences with those of human CPA1 and human CPA2. The specificity of CPA4 corresponds more closely with the specificity of CPA1 rather than CPA2. Nevertheless, when compared with CPA1, CPA4 shows a less marked preference for Phe in P1' position and a slightly better performance against intermediate hydrophobic aminoacids like Leu or Ile. The peptidomics analysis showed that the P1 position is also important for specificity. CPA4 prefers hydrophobic amino acids and positively charged amino acids in the P1 position. Instead negatively charged amino acids, proline and glycine are disfavoured in this position.

Several reports suggest a link of CPA4 with prostate cancer aggressiveness (Kayashima et al., 2003; Ross et al., 2009; Witte et al., 2000). Particularly, the nonsynonymous coding SNP (Gly303Cys) rs2171492 was associated with an increased risk of aggressive disease in younger men (Ross et al., 2009). In addition, CPA4 mRNA levels are increased in tissues that are growing or differentiating, suggesting a role of CPA4 in the differentiation process. In this study, several biologically relevant peptides where shown to be cleaved by CPA4 (i.e., enkephalins, secretogranins and neurotensin). Some of these peptides have roles in the progression and invasive potential of tumor cells (Dupouy et al., 2009; Mosca et al., 2005; Myers et al., 2009; Nagakawa et al., 2001; Vias et al., 2007). Altogether, we hypothesized a role for CPA4 in the extracellular environment in neuropeptide processing and regulation, providing a possible explanation for the link of CPA4 with prostate cancer aggressiveness.

In reality, additional research efforts are needed to completely understand which the biological role of CPA4 is. For example, several other A-type MCPs have been proposed to function in neuropeptide processing, including CPA3 (Bunnett et al., 1992), CPA5 (Wei et al., 2003) and CPA6 (Lyons et al., 2008). It is sometimes difficult to predict which of these CP will actually modify some of these substrates in situ. The analysis of the distribution pattern of each enzyme and the site of the neuropeptide production or action may help to understand this. Another point to consider is the functional consequence of the truncation of a C-terminal residue in the neuropeptide,

which may vary from peptide to peptide. In some cases this cleavage might abolish and in some other promote the biological activity, whereas it is also possible that sometimes the truncation is not coupled to a change in the biological function.

For the determination of CPA4 substrate specificity we made use of three different tools for substrate analysis. The small chromogenic substrates are ideal for kinetic analysis of k_{cat} and K_m values, which is the best way to compare the preferences of different enzymes. However, the number of substrates available for evaluation is sometimes limited, and the chromogenic group can potentially affect the activity of the enzyme. The analysis of selected synthetic peptides represents a more real situation, but the analysis might be expensive and biased by the peptide selection. Both these approaches are time consuming and might require important amounts of the enzyme under study. The peptidomic approach has the advantage that allows the analysis of hundreds of naturally occurring peptides in a single experiment, although it does not provide kinetic information. The developed peptidomics approach has been further used to evaluate the activity of other MCPs and endoproteases (Berti et al., 2009; Lyons and Fricker, 2010, 2011).

This peptidomic approach can be classified into the group of proteomic techniques for the *in vitro* mapping of protease active site specificity. In Chapter 3 we developed a second of these proteomic techniques, in this case a proteome-derive peptide library to study substrate specificity of carboxypeptidases. Substrate specificity profiling is an important step in the enzymatic characterization, because while several factors impact on substrate selection, a key factor is the complementarity of a protease binding site with specific substrate side-chains. Importantly, knowledge of the *in vitro* substrate specificity serves to corroborate *in vivo* cleavage events. Second, it allows to predict activity of the protease/carboxypeptidase toward novel substrates. Finally this knowledge provides a framework for the design of selective probes and potent and selective inhibitors (van den Berg and Tholey, 2012). The major advantages of MS-based approaches are their high sensitivity and selectivity, their ability to identify and characterize the primary sequence of the cleavage products (in some cases for both primed and non-primed sides), and their ability to perform quantitative analysis.

This peptide library is a COFRADIC-based approach, in which the MCP acts as a modifying step. The substrates of the MCP under study are modified by the action of

the enzyme allowing for their isolation and MS/MS analysis. The advantage of this approach is that it isolates and analyzes the N-terminal products of the protease action, allowing for the analysis of the MCP substrate preferences. Most peptides libraries analyze only the C-terminal products of the protease action (Schilling and Overall, 2008), but in the case of MCP these are only single amino acids and consequently difficult to analyze. The enrichment of CPs product was very efficient in some cases, reaching up to a 90 percent of all analyzed peptides. The principle behind the enrichment procedure is only valid for A-type carboxypeptidases, but the approach has to be adapted so it can be used to profile carboxypeptidases with different substrate specificity.

This approach is in addition a proteome-derived technique, and as a result once the non-primed side of the cleavage is identified the cleaved primed side can bioinformatically be identified. In other words, this approach allows for the simultaneous profiling of prime- and non-prime specificity. Proteome derived peptide libraries are generated by proteolysis of a model proteome (here a cell extract) using a specific protease (in our case chymotrypsin or Lys-N). The selection of a certain peptide library-generating protease determines the peptide composition and length of the resulting peptide library. The use of different peptide libraries minimizes the bias introduced by peptide library-generating protease, rendering a better overall profile of the MCP preferences. Furthermore this approach introduces the use of MCP-induced ^{18}O-labeling to discriminate MCP products from co-purifying peptides. ^{18}O-labeling is a strategy that introduces a label in the product of a proteolytic event when the protease catalyzes the hydrolysis in the presence of $H^{18}O_2$ (Ye et al., 2009). This is the first report of the use of ^{18}O-labeling as a valid approach to label the products of the carboxypeptidase action.

Using proteome-derived peptide libraries we determined the substrate preferences of human CPA1, CPA2 and CPA4, and mouse CPA3. This is the first report of the substrate preferences for this mast cell carboxypeptidase. This demonstrates one of the great advantages of this approach: by using only tiny amounts of purified enzyme it allows for a complete substrate specificity profiling (in a single experiment). The derived substrate specificity of human CPA1, CPA2 and CPA4 fully agrees with the results presented in Chapter 2, thereby validating this approach. With this approach

the extended substrate specificity profile of these MCPs was evaluated, showing that substrate selection is mainly dictated by the P1' and, at a lower extent, P1 positions. The analysis of the derived substrate preferences show that the substrates preferences of A-type MCPs can be fitted in two big groups. CPA2 and CPA3 can be described as A2-type MCPs displaying a strong preference for bulkier hydrophobic amino acids (i.e., Trp, Tyr and Phe). Instead CPA1 and CPA4 behave as A1-type MCPs, by being able to efficiently cleave both small aliphatic as well as bulky aromatic amino acids.

The number of identified CPA4 substrates is much larger in the peptide library than in the peptidomics analysis presented in Chapter 2: in the peptidomics analysis we identified 44 peptide substrates for CPA4; on the other hand 9729 and 3503 peptide substrates identified using the peptide a chymotryptic and a Lys-N peptide library, respectively. One possible explanation for this is that proteome-derived peptide libraries offer a much greater number of possible substrates with broader sequence diversities. In addition, in our peptide library there is an enrichment of the MCP products, which improves substrate identification. However, one of the advantages of the peptidomic approach is that it is a quantitative approach allowing distinguishing efficient substrates from less efficient ones. In addition, it offers the possibility to evaluate biological relevant peptides. These results can be compared with those presented by Van Damme *et al.* that searched for CPA4 substrates using the C-terminal COFRADIC approach, when the recombinant enzyme was added to PC3 cell extracts (Van Damme et al., 2010). Here only 7 substrates were identified and the substrate preferences corresponded to the specificity profiles presented in Chapters 2 and 3. The low number of CPA4 substrates can be explained due to the fact that C-terminal COFRADIC searches for protein substrates and not peptide substrates. Limited solvent exposure of protein C-termini and limited accessibility of these to the MCP active site might possibly explain why CPA4 seem to cleave more efficiently peptide substrates than protein substrates. The presence of the endogenous MCP inhibitor (latexin) in the PC3 cells extract might also account for the low CPA4 activity.

C-terminal COFRADIC was recently developed in the laboratory of Professor Gevaert, in collaboration with our laboratory (Van Damme et al., 2010). This is a protein-centric MS-based approach (because proteolysis by the protease/carboxypeptidase under study occurs at the protein level), and is suitable for

the investigation of proteolytic events at the cellular level. Protein-centric strategies provide a closer representation of the *in vivo* function for some proteases, since they take into account influences of protein secondary and tertiary structure. There is a need for these system-wide MS-based approaches for the study of protease substrates at the protein level, since the study of substrate specificity is usually insufficient to fully identify natural protease substrates. Just because a protease can cleave a substrate *in vitro* it does not mean it does *in vivo*. For example, compartmentalization is usually not maintained in *in vitro* assays and the test protease is allowed to cleave substrates that it cannot access *in vivo*. Occasionally, the inhibitory mechanisms present in the cell may be surpassed by the amount of added test protease.

C-terminal COFRADIC enriches for C-termini from whole-proteome digests, allowing for the analysis of C-terminal protein modifications and C-terminal protein processing events. Some of the characteristics of CCPs make them a suitable and relevant biological model to apply this technique. First, they are cytosolic/nuclear enzymes and so suitable for analysis, since protein cell extracts are relatively simpler proteomes to be analyzed (when compared to the extracellular medium for instance). Furthermore, they are expected to act over protein C-termini, and this is ideal since this is a protein-centric approach. The biological importance of these enzymes is demonstrated in the ataxic Purkinje cell degeneration (*pcd*) mouse, in which gene disruption of *Ccp1* leads to a phenotype characterized by neuronal degeneration (Wang and Morgan, 2007). Cytosolic carboxypeptidases are thought to post-translationally modify proteins at two levels. They have been shown to trim both gene-encoded acidic C-termini and post-translationally added glutamate side chains in tubulin (Janke and Bulinski, 2011). C-terminal COFRADIC is able to search for the first kind of substrates but not for the latter.

We searched for CCP1 substrates in HEK 293T by modulating (through CCP1 over-expression) its activity. C-terminal COFRADIC is a quantitative proteomic approach, and allows comparing two cellular conditions, allowing for the detection of neo-C-termini in the proteome where CCP1 activity is increased. As a result, seven new substrates were identified, for which from 1 up to 16 acidic amino acids were released. Here, C-terminal COFRADIC proved to be able to detect real C-terminal proteolytic events. This is supported by the fact that two of the identified substrates are known CCP1

substrates (α-tubulin 1B and 1C). In addition, some of the identified new substrates were validated in cellular experiments and *in vitro*. The derived substrate specificity, only partially corresponds to previous reports. It has been previously stated that CCP1 is not able to cleave aspartate residues. Here, CCP1 was shown to be able to cleave both glutamate and aspartate residues. Nevertheless our results suggest that CCP1 would cleave glutamate residues more efficiently that aspartate residues. This could be a possible explanation for the discrepancy with previous results. Finally, it is still possible that some of these substrates are cleaved as a consequence of CCP1 over-expression. Further experiments are needed to confirm that these are physiologically relevant CCP1 substrates.

Some important questions remain open in relation to the regulation of the activity of CCP1 inside the cell. Many putative substrates have been described for these enzymes. CCP1 is able to shorten the length of the polyglutamylation side chains in tubulin, but polyglutamylation is also present in a broad range of proteins (van Dijk et al., 2008). In addition, CCPs were proven to act over the gene-encoded C-terminal glutamate residues of tubulin, MLCK and telokin (Rogowski et al., 2010). We here extend this list, by showing that CCP1 is able to act over the acidic C-terminus of seven other proteins. As a result, a few questions remain unanswered. Are any only few of this substrate the physiologically relevant? Or is CCP1 able to process all of them in a physiological context? Are the other CCPs also able to modify them? Different regulation mechanism can be considered. Rodriguez de la Vega et al, showed that different CCPs display specific cellular distribution in the cell, suggesting that their regulation might rely on their localization (Rodriguez de la Vega et al, 2012). In addition, each CCP contain additional domain that may regulate in addition to their localization, their binding with substrates. Different additional mechanisms of regulation can be hypothesized for CCP1 (and other CCPs): The level of expression might constitute a mechanism of regulation, given that CCPs seem to be found at low levels in the cell. Proteolysis could be another mechanism, taking into account that degradation bands are usually observed in CCP1 western blots. Phosphorilation have been hypothesized as a regulation mechanisms; indeed some of the tyrosines in the PYTY motif (Rodriguez de la Vega et al., 2007) are highly conserved in the whole

subfamily. Further work will needed to completely elucidate the function of these enzymes and their mechanisms of regulation.

Concluding remarks

Chapter 1

1. The structure of the first repeat of *Drosophila melanogaster* carboxypeptidase D (DmCPD1Bs) in complex with GEMSA was solved at 2.70 Å resolution, and overall conforms to the structure of M14B MCPs
2. This enzyme was found as a homodimer in the crystal structure, but there is no concluding evidence that the enzyme behaves as a dimer in solution.
3. When such structure is compared with the structure of the second repeat of duck CPD, the most significant deviations are found in the loops that contribute to shape the access to the active site
4. A unique structural element in DmCPD1Bs structure is a surface hotspot targetable by peptidases, suggesting a regulatory mechanism through proteolytic inactivation.

Chapter 2

1. Human CPA4 is secreted from cells as a soluble proenzyme (PCPA4), and would function in neuropeptide processing in the extracellular environment
2. The optimal C-terminal residues for CPA4 cleavage are: Phe, Leu, Met, Ile, Tyr and Val. Residues like Ala, His, Thr, or Trp are cleaved less efficiently. The P1 position also has an influence on cleavage specificity.
3. A quantitative peptidomics approach was developed, which allows testing hundreds of biologically relevant peptides in a single experiment.
4. This approach identified neurotensin, granins and opioid peptides such as Met- and Leu-enkephalin as putative CPA4 substrates. These peptides are involved in the proliferation and differentiation of prostate cancer cells, potentially explaining the link between CPA4 and cancer aggressiveness.

Chapter 3

1. A COFRADIC-based proteome-derived peptide library approach was developed to study the substrate specificity profiles of carboxypeptidases.
2. Two different peptide libraries, generated either by chymotrypsin or by Lys-N, provided substrate specificities for human CPA1, CPA2 and CPA4, which are in agreement with previous reports.
3. This approach served to delineate the previously little studied substrate specificity profile of mouse CPA3, which behaves as an A2-type MCP.
4. The substrates preferences of A-type MCPs can be fitted in two groups; while CPA1 and CPA4 behave as A1-type MCPs, CPA2 and CPA3 can be described as A2-type MCPs.

Chapter 4

1. C-terminal COFRADIC was able to identify seven new putative CCP1 protein substrates in a cellular context.
2. Validation experiments demonstrated that C-terminal COFRADIC identified real CCP1-mediated proteolytic events, although further research is needed to fully confirm that these are physiologically relevant substrates.
3. The derived substrate specificity shows that CCP1 is able to trim both glutamate and aspartate C-terminal residues.
4. Long acidic C-terminal tails participate in protein binding, suggesting that CCP1, by shortening these tails, might function in the regulation of protein-protein interactions.

Bibliography

Abdelmagid, S.A., and Too, C.K. (2008). Prolactin and estrogen up-regulate carboxypeptidase-d to promote nitric oxide production and survival of MCF-7 breast cancer cells. Endocrinology 149, 4821-4828.

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr sect D 58, 1948-1954.

Aizawa, S., Nishino, H., Saito, K., Kimura, K., Shirakawa, H., and Yoshida, M. (1994). Stimulation of transcription in cultured cells by high mobility group protein 1: essential role of the acidic carboxyl-terminal region. Biochemistry 33, 14690-14695.

Aloy, P., Companys, V., Vendrell, J., Avilés, F.X., Fricker, L.D., Coll, M., and Gomis-Rüth, F.X. (2001). The crystal structure of the inhibitor-complexed carboxypeptidase D domain II and the modeling of regulatory carboxypeptidases. J Biol Chem 276, 16177-16184.

Arai, S., Noguchi, M., Kurosawa, S., Kato, H., and Fujimaki, M. (1970). APPLYING PROTEOLYTIC ENZYMES ON SOYBEAN. 6. Deodorization Effect of Aspergillopeptidase A and Debittering Effect of Aspergillus Acid Carboxypeptidase. Journal of Food Science 35, 392-395.

Arolas, J.L., Lorenzo, J., Rovira, A., Castella, J., Aviles, F.X., and Sommerhoff, C.P. (2005a). A carboxypeptidase inhibitor from the tick Rhipicephalus bursa: isolation, cDNA cloning, recombinant expression, and characterization. J Biol Chem 280, 3441-3448.

Arolas, J.L., Vendrell, J., Aviles, F.X., and Fricker, L.D. (2007). Metallocarboxypeptidases: emerging drug targets in biomedicine. Curr Pharm Des 13, 349-366.

auf dem Keller, U., Doucet, A., and Overall, C.M. (2007). Protease research in the era of systems biology. Biol Chem 388, 1159-1162.

auf dem Keller, U., and Schilling, O. (2010). Proteomic techniques and activity-based probes for the system-wide study of proteolysis. Biochimie 92, 1705-1714.

Auld, D.S. (2001). Zinc coordination sphere in biochemical zinc sites. Biometals 14, 271-313.

Avilés, F.X., Vendrell, J., Guasch, A., Coll, M., and Huber, R. (1993). Advances in metallo-procarboxypeptidases. Emerging details on the inhibition mechanism and on the activation process. European Journal of Biochemistry 211, 381-389.

Baltanas, F.C., Casafont, I., Lafarga, V., Weruaga, E., Alonso, J.R., Berciano, M.T., and Lafarga, M. (2011a). Purkinje cell degeneration in pcd mice reveals large scale chromatin reorganization and gene silencing linked to defective DNA repair. J Biol Chem 286, 28287-28302.

Baltanas, F.C., Casafont, I., Weruaga, E., Alonso, J.R., Berciano, M.T., and Lafarga, M. (2011b). Nucleolar disruption and cajal body disassembly are nuclear hallmarks of DNA damage-induced neurodegeneration in purkinje cells. Brain Pathol 21, 374-388.

Barbosa Pereira, P.J., Segura-Martin, S., Oliva, B., Ferrer-Orta, C., Avilés, F.X., Coll, M., Gomis-Rüth, F.X., and Vendrell, J. (2002). Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). J Mol Biol 321, 537-547.

Barrett, A.J. (2000). Proteases In Current Protocols in Protein Science, pp. 21.21.21 - 21.21.12.

Barrett, A.J. (2001). Proteolytic enzymes: nomenclature and classification. In Proteolytic enzymes: a practical approach (Oxford University Press).

Barrett, A.J., Rawlings, N.D., and Woessner, J.F. (2004). Handbook of Proteolytic Enzymes, 2nd edn (London, Elsevier Academic Press).

Barrios, A.M., and Craik, C.S. (2002). Scanning the prime-site substrate specificity of proteolytic enzymes: a novel assay based on ligand-enhanced lanthanide ion fluorescence. Bioorg Med Chem Lett 12, 3619-3623.

Bayes, A., Fernandez, D., Sola, M., Marrero, A., Garcia-Pique, S., Aviles, F.X., Vendrell, J., and Gomis-Ruth, F.X. (2007). Caught after the Act: a human A-type metallocarboxypeptidase in a product complex with a cleaved hexapeptide. Biochemistry 46, 6921-6930.

Bentley, L., Nakabayashi, K., Monk, D., Beechey, C., Peters, J., Birjandi, Z., Khayat, F.E., Patel, M., Preece, M.A., Stanier, P., et al. (2003). The imprinted region on human chromosome 7q32 extends to the carboxypeptidase A gene cluster: an imprinted candidate for Silver-Russell syndrome. J Med Genet 40, 249-256.

Berezniuk, I., and Fricker, L.D. (2010). A defect in cytosolic carboxypeptidase 1 (Nna1) causes autophagy in Purkinje cell degeneration mouse brain. Autophagy 6.

Berezniuk, I., Vu, H.T., Lyons, P.J., Sironi, J.J., Xiao, H., Burd, B., Setou, M., Angeletti, R.H., Ikegami, K., and Fricker, L.D. (2012). Cytosolic carboxypeptidase 1 is involved in processing alpha- and beta-tubulin. J Biol Chem 287, 6503-6517.

Bergman, T., Cederlund, E., Jornvall, H., and Fowler, E. (2003). C-terminal sequence analysis. Curr Protoc Protein Sci Chapter 11, Unit 11 18.

Berrow, N.S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., and Owens, R.J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. Nucleic Acids Res 35, e45.

Berti, D.A., Morano, C., Russo, L.C., Castro, L.M., Cunha, F.M., Zhang, X., Sironi, J., Klitzke, C.F., Ferro, E.S., and Fricker, L.D. (2009). Analysis of intracellular substrates and products of thimet oligopeptidase in human embryonic kidney 293 cells. J Biol Chem 284, 14105-14116.

Bieth, J.G. (1995). Theoretical and practical aspects of proteinase inhibition kinetics. Methods Enzymol 248, 59-84.

Blumberg, S., and Vallee, B.L. (1975). Superactivation of thermolysin by acylation with amino acid N-hydroxysuccinimide esters. Biochemistry 14, 2410-2419.

Bown, D.P., and Gatehouse, J.A. (2004). Characterization of a digestive carboxypeptidase from the insect pest corn earworm (Helicoverpa armigera) with novel specificity towards C-terminal glutamate residues. Eur J Biochem 271, 2000-2011.

Bunnett, N.W., Goldstein, S.M., and Nakazato, P. (1992). Isolation of a neuropeptide-degrading carboxypeptidase from the human stomach. Gastroenterology 102, 76-87.

Byers, L.D., and Wolfenden, R. (1973). Binding of the by-product analog benzylsuccinic acid by carboxypeptidase A. Biochemistry 12, 2070-2078.

Carranza, C., Inisan, A.-G., Mouthuy-Knoops, E., Cambillau, C., and Roussel, A. (1999). Turbo-Frodo. In AFMB Activity Report 1996-1999 (Marseille, CNRS-UPR 9039), pp. 89-90.

Cawley, N.X., Wetsel, W.C., Murthy, S.R., Park, J.J., Pacak, K., and Loh, Y.P. (2012). New roles of carboxypeptidase E in endocrine and neural function and cancer. Endocr Rev 33, 216-253.

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. Nat Methods 6, 786-787.

Colucci, M., and Semeraro, N. (2012). Thrombin activatable fibrinolysis inhibitor: at the nexus of fibrinolysis and inflammation. Thromb Res 129, 314-319.

Chakrabarti, L., Eng, J., Ivanov, N., Garden, G.A., and La Spada, A.R. (2009). Autophagy activation and enhanced mitophagy characterize the Purkinje cells of pcd mice prior to neuronal death. Mol Brain 2, 24.

Chakrabarti, L., Eng, J., Martinez, R.A., Jackson, S., Huang, J., Possin, D.E., Sopher, B.L., and La Spada, A.R. (2008). The zinc-binding domain of Nna1 is required to prevent retinal photoreceptor loss and cerebellar ataxia in Purkinje cell degeneration (pcd) mice. Vision Res 48, 1999-2005.

Che, F.Y., Lim, J., Pan, H., Biswas, R., and Fricker, L.D. (2005). Quantitative neuropeptidomics of microwave-irradiated mouse brain and pituitary. Mol Cell Proteomics 4, 1391-1405.

Che, F.Y., Zhang, X., Berezniuk, I., Callaway, M., Lim, J., and Fricker, L.D. (2007). Optimization of neuropeptide extraction from the mouse hypothalamus. J Proteome Res 6, 4667-4676.

Chen, H., Jawahar, S., Qian, Y., Duong, Q., Chan, G., Parker, A., Meyer, J.M., Moore, K.J., Chayen, S., Gross, D.J., et al. (2001). Missense polymorphism in the human carboxypeptidase E gene alters enzymatic activity. Hum Mutat 18, 120-131.

Davila-Garcia, M.I., and Azmitia, E.C. (1990). Neuropeptides as positive or negative neuronal growth regulatory factors: effects of ACTH and leu-enkephalin on cultured serotonergic neurons. Adv Exp Med Biol 265, 75-92.

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Bryan Arendall, W., 3rd, Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucl Acids Res 35 (Web Server issue), W375-W383.

De Schutter, K., Lin, Y.C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouze, P., Van de Peer, Y., and Callewaert, N. (2009). Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol 27, 561-566.

Declerck, P.J. (2011). Thrombin activatable fibrinolysis inhibitor. Hamostaseologie 31, 165-166, 168-173.

Deiteren, K., Hendriks, D., Scharpe, S., and Lambeir, A.M. (2009). Carboxypeptidase M: Multiple alliances and unknown partners. Clin Chim Acta 399, 24-39.

Deiteren, K., Surpateanu, G., Gilany, K., Willemse, J.L., Hendriks, D.F., Augustyns, K., Laroche, Y., Scharpe, S., and Lambeir, A.M. (2007a). The role of the S1 binding site of carboxypeptidase M in substrate specificity and turn-over. Biochim Biophys Acta 1774, 267-277.

Deiteren, K., Surpateanu, G., Gilany, K., Willemse, J.L., Hendriks, D.F., Augustyns, K., Laroche, Y., Scharpe, S., and Lambeir, A.M. (2007b). The role of the S1 binding site of carboxypeptidase M in substrate specificity and turn-over. Biochim Biophys Acta 1774, 267-277.

Diamond, S.L. (2007). Methods for mapping protease specificity. Curr Opin Chem Biol 11, 46-51.

Dormeyer, W., Mohammed, S., Breukelen, B., Krijgsveld, J., and Heck, A.J. (2007). Targeted analysis of protein termini. J Proteome Res 6, 4634-4645.

Doucet, A., Butler, G.S., Rodriguez, D., Prudova, A., and Overall, C.M. (2008). Metadegradomics: toward in vivo quantitative degradomics of proteolytic post-translational modifications of the cancer proteome. Mol Cell Proteomics 7, 1925-1951.

Doucet, A., and Overall, C.M. (2008). Protease proteomics: revealing protease in vivo functions using systems biology approaches. Mol Aspects Med 29, 339-358.

Dupouy, S., Viardot-Foucault, V., Alifano, M., Souaze, F., Plu-Bureau, G., Chaouat, M., Lavaur, A., Hugol, D., Gespach, C., Gompel, A., et al. (2009). The neurotensin receptor-1 pathway contributes to human ductal breast cancer progression. PLoS One 4, e4223.

Eng, F.J., Novikova, E.G., Kuroki, K., Ganem, D., and Fricker, L.D. (1998). gp180, a protein that binds duck hepatitis B virus particles, has metallocarboxypeptidase D-like enzymatic activity. J Biol Chem 273, 8382-8388.

Escobar, M.L., Barea-Rodriguez, E.J., Derrick, B.E., Reyes, J.A., and Martinez, J.L., Jr. (1997). Opioid receptor modulation of mossy fiber synaptogenesis: independence from long-term potentiation. Brain Res 751, 330-335.

Evans, P. (2006). Scaling and assessment of data quality. Acta Crystallogr sect D 62, 72-82.

Evans, S.V. (1993). SETOR : hardware lighted three-dimensional solid model representations of macromolecules. J Mol Graphics 11, 134-138.

Faming, Z., Kobe, B., Stewart, C.B., Rutter, W.J., and Goldsmith, E.J. (1991). Structural evolution of an enzyme specificity. The structure of rat carboxypeptidase A2 at 1.9-A resolution. J Biol Chem 266, 24606-24612.

Felber, J.P., Coombs, T.L., and Vallee, B.L. (1962). The mechanism of inhibition of carboxypeptidase A by 1,10-phenanthroline. Biochemistry 1, 231-238.

Fernandez-Gonzalez, A., La Spada, A.R., Treadaway, J., Higdon, J.C., Harris, B.S., Sidman, R.L., Morgan, J.I., and Zuo, J. (2002). Purkinje cell degeneration (pcd) phenotypes caused by mutations in the axotomy-induced gene, Nna1. Science 295, 1904-1906.

Feyerabend, T.B., Hausser, H., Tietz, A., Blum, C., Hellman, L., Straus, A.H., Takahashi, H.K., Morgan, E.S., Dvorak, A.M., Fehling, H.J., et al. (2005). Loss of histochemical identity in mast cells lacking carboxypeptidase A. Mol Cell Biol 25, 6199-6210.

Fontenele-Neto, J.D., Kalinina, E., Feng, Y., and Fricker, L.D. (2005). Identification and distribution of mouse carboxypeptidase A-6. Brain Res Mol Brain Res 137, 132-142.

Fricker, L.D. (2002). Carboxypeptidases D and E. In The Enzymes Co- and posttranslational proteolysis of proteins, R.E. Dalbey, and D.S. Sigman, eds. (London, Academic Press, Inc.), pp. 421-452.

Fricker, L.D. (2004a). 251. Metallocarboxypeptidase D. In Handbook of proteolytic enzymes, A.J. Barrett, N.D. Rawlings, and J.F. Woessner Jr., eds. (London, Elsevier), pp. 848-851.

Fricker, L.D. (2004b). Carboxypeptidase E. In Handbook of Proteolytic Enzymes (London, Elsevier Academic Press).

Fricker, L.D., and Devi, L. (1990). Comparison of a spectrophotometric, a fluorometric, and a novel radiometric assay for carboxypeptidase E (EC 3.4.17.10) and other carboxypeptidase B-like enzymes. Anal Biochem 184, 21-27.

Fricker, L.D., and Snyder, S.H. (1983). Purification and characterization of enkephalin convertase, an enkephalin-synthesizing carboxypeptidase. J Biol Chem 258, 10950-10955.

Garcia-Castellanos, R., Bonet-Figueredo, R., Pallares, I., Ventura, S., Aviles, F.X., Vendrell, J., and Gomis-Ruth, F.X. (2005). Detailed molecular comparison between the inhibition mode of A/B-type carboxypeptidases in the zymogen state and by the endogenous inhibitor latexin. Cell Mol Life Sci 62, 1996-2014.

Garcia-Saez, I., Reverter, D., Vendrell, J., Aviles, F.X., and Coll, M. (1997). The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. EMBO J 16, 6906-6913.

Garcia, K.C., Tallquist, M.D., Pease, L.R., Brunmark, A., Scott, C.A., Degano, M., Stura, E.A., Peterson, P.A., Wilson, I.A., and Teyton, L. (1997). Alphabeta T cell receptor interactions with syngeneic and allogeneic ligands: affinity measurements and crystallization. Proc Natl Acad Sci U S A 94, 13838-13843.

Gardell, S.J., Craik, C.S., Clauser, E., Goldsmith, E.J., Stewart, C.B., Graf, M., and Rutter, W.J. (1988). A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family. J Biol Chem 263, 17828-17836.

Gevaert, K., Impens, F., Van Damme, P., Ghesquiere, B., Hanoulle, X., and Vandekerckhove, J. (2007). Applications of diagonal chromatography for proteome-wide characterization of protein modifications and activity-based analyses. FEBS J 274, 6277-6289.

Ghesquiere, B., Colaert, N., Helsens, K., Dejager, L., Vanhaute, C., Verleysen, K., Kas, K., Timmerman, E., Goethals, M., Libert, C., et al. (2009). In vitro and in vivo protein-bound tyrosine nitration characterized by diagonal chromatography. Mol Cell Proteomics 8, 2642-2652.

Goldstein, S.M., Kaempfer, C.E., Kealey, J.T., and Wintroub, B.U. (1989). Human mast cell carboxypeptidase. Purification and characterization. J Clin Invest 83, 1630-1636.

Goldstein, S.M., Kaempfer, C.E., Proud, D., Schwartz, L.B., Irani, A.M., and Wintroub, B.U. (1987). Detection and partial characterization of a human mast cell carboxypeptidase. J Immunol 139, 2724-2729.

Gomis-Ruth, F.X. (2008). Structure and mechanism of metallocarboxypeptidases. Crit Rev Biochem Mol Biol 43, 319-345.

Gomis-Rüth, F.X. (2009). Catalytic domain architecture of metzincin metalloproteases. J Biol Chem, 15353-15357.

Gomis-Ruth, F.X., Companys, V., Qian, Y., Fricker, L.D., Vendrell, J., Aviles, F.X., and Coll, M. (1999). Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallocarboxypeptidase subfamily. EMBO J 18, 5817-5826.

Gosalia, D.N., Salisbury, C.M., Ellman, J.A., and Diamond, S.L. (2005). High throughput substrate specificity profiling of serine and cysteine proteases using solution-phase fluorogenic peptide microarrays. Mol Cell Proteomics 4, 626-636.

Grishin, A.M., Akparov, V., and Chestukhina, G.G. (2008). Structural principles of the broad substrate specificity of Thermoactinomyces vulgaris carboxypeptidase T--role of amino acid residues at positions 260 and 262. Protein Eng Des Sel 21, 545-551.

Hadkar, V., and Skidgel, R.A. (2001). Carboxypeptidase D is up-regulated in raw 264.7 macrophages and stimulates nitric oxide synthesis by cells in arginine-free medium. Mol Pharmacol 59, 1324-1332.

Hamstra, D.A., and Rehemtulla, A. (1999). Toward an enzyme/prodrug strategy for cancer gene therapy: endogenous activation of carboxypeptidase A mutants by the PACE/Furin family of propeptidases. Hum Gene Ther 10, 235-248.

Hardeman, K., Samyn, B., Van der Eycken, J., and Van Beeumen, J. (1998). An improved chemical approach toward the C-terminal sequence analysis of proteins containing all natural amino acids. Protein Sci 7, 1593-1602.

Harris, A., Morgan, J.I., Pecot, M., Soumare, A., Osborne, A., and Soares, H.D. (2000). Regenerating motor neurons express Nna1, a novel ATP/GTP-binding protein related to zinc carboxypeptidases. Mol Cell Neurosci 16, 578-596.

Helle, K.B., Corti, A., Metz-Boutigue, M.H., and Tota, B. (2007). The endocrine role for chromogranin A: a prohormone for peptides with regulatory properties. Cell Mol Life Sci 64, 2863-2886.

Hohmann, L., Sherwood, C., Eastham, A., Peterson, A., Eng, J.K., Eddes, J.S., Shteynberg, D., and Martin, D.B. (2009). Proteomic analyses using Grifola frondosa metalloendoprotease Lys-N. J Proteome Res 8, 1415-1422.

Huang, H., Reed, C.P., Zhang, J.S., Shridhar, V., Wang, L., and Smith, D.I. (1999). Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone deacetylase inhibitors during differentiation of prostate epithelial cancer cells. Cancer Res 59, 2981-2988.

Huesgen, P.F., and Overall, C.M. (2012). N- and C-terminal degradomics: new approaches to reveal biological roles for plant proteases from substrate identification. Physiol Plant 145, 5-17.

Janke, C., and Bulinski, J.C. (2011). Post-translational regulation of the microtubule cytoskeleton: mechanisms and functions. Nat Rev Mol Cell Biol 12, 773-786.

Kabsch, W. (2001). Chapter 25.2.9: XDS. In International Tables for Crystallography Volume F: Crystallography of Biological Macromolecules, M.G. Rossmann, and E. Arnold, eds. (Dordrecht (The Netherlands), Kluwer Academic Publishers (for The International Union of Crystallography)), pp. 730-734.

Kalinina, E., Biswas, R., Berezniuk, I., Hermoso, A., Aviles, F.X., and Fricker, L.D. (2007). A novel subfamily of mouse cytosolic carboxypeptidases. FASEB J 21, 836-850.

Kalinina, E., and Fricker, L.D. (2003). Palmitoylation of carboxypeptidase D. Implications for intracellular trafficking. J Biol Chem 278, 9244-9249.

Kall, L., Storey, J.D., MacCoss, M.J., and Noble, W.S. (2008). Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. J Proteome Res 7, 29-34.

Kayashima, T., Yamasaki, K., Yamada, T., Sakai, H., Miwa, N., Ohta, T., Yoshiura, K., Matsumoto, N., Nakane, Y., Kanetake, H., et al. (2003). The novel imprinted carboxypeptidase A4 gene (CPA4) in the 7q32 imprinting domain. Hum Genet 112, 220-226.

Keil, C., Maskos, K., Than, M., Hoopes, J.T., Huber, R., Tan, F., Deddish, P.A., Erdos, E.G., Skidgel, R.A., and Bode, W. (2007). Crystal structure of the human carboxypeptidase N (kininase I) catalytic domain. J Mol Biol 366, 504-516.

Kimura, Y., Kurabe, N., Ikegami, K., Tsutsumi, K., Konishi, Y., Kaplan, O.I., Kunitomo, H., Iino, Y., Blacque, O.E., and Setou, M. (2010). Identification of tubulin deglutamylase among Caenorhabditis elegans and mammalian cytosolic carboxypeptidases (CCPs). J Biol Chem 285, 22936-22941.

Kuba, K., Imai, Y., Ohto-Nakanishi, T., and Penninger, J.M. (2010). Trilogy of ACE2: a peptidase in the renin-angiotensin system, a SARS receptor, and a partner for amino acid transporters. Pharmacol Ther 128, 119-128.

Kuroki, K., Cheung, R., Marion, P.L., and Ganem, D. (1994). A cell surface protein that binds avian hepatitis B virus particles. J Virol 68, 2091-2096.

Kyuhou, S., Kato, N., and Gemba, H. (2006). Emergence of endoplasmic reticulum stress and activated microglia in Purkinje cell degeneration mice. Neurosci Lett 396, 91-96.

Lee, K.B., and Thomas, J.O. (2000). The effect of the acidic tail on the DNA-binding properties of the HMG1,2 class of proteins: insights from tail switching and tail removal. J Mol Biol 304, 135-149.

Lein, E.S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., et al. (2007). Genome-wide atlas of gene expression in the adult mouse brain. Nature 445, 168-176.

Leung, L.L., Myles, T., Nishimura, T., Song, J.J., and Robinson, W.H. (2008). Regulation of tissue inflammation by thrombin-activatable carboxypeptidase B (or TAFI). Mol Immunol 45, 4080-4083.

Li, F., Li, W., Farzan, M., and Harrison, S.C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309, 1864-1868.

Li, J., Gu, X., Ma, Y., Calicchio, M.L., Kong, D., Teng, Y.D., Yu, L., Crain, A.M., Vartanian, T.K., Pasqualini, R., et al. (2010). Nna1 mediates Purkinje cell dendritic development via lysyl oxidase propeptide and NF-kappaB signaling. Neuron 68, 45-60.

Lin, Q., Yu, N.J., and Spremulli, L.L. (1996). Expression and functional analysis of Euglena gracilis chloroplast initiation factor 3. Plant Mol Biol 32, 937-945.

Lindsley, D.L., and Zimm, G.G. (1992). The genome of Drosophila melanogaster. (San Diego, CA, Academic Press, Inc.).

Lopez-Otin, C., and Overall, C.M. (2002). Protease degradomics: a new challenge for proteomics. Nat Rev Mol Cell Biol 3, 509-519.

Lundequist, A., Tchougounova, E., Abrink, M., and Pejler, G. (2004). Cooperation between mast cell carboxypeptidase A and the chymase mouse mast cell protease 4 in the formation and degradation of angiotensin II. J Biol Chem 279, 32339-32344.

Lyons, P.J., Callaway, M.B., and Fricker, L.D. (2008). Characterization of carboxypeptidase A6, an extracellular matrix peptidase. J Biol Chem 283, 7054-7063.

Lyons, P.J., and Fricker, L.D. (2010). Substrate specificity of human carboxypeptidase A6. J Biol Chem 285, 38234-38242.

Lyons, P.J., and Fricker, L.D. (2011a). Carboxypeptidase O is a glycosylphosphatidylinositol-anchored intestinal peptidase with acidic amino acid specificity. J Biol Chem 286, 39023-39032.

Lyons, P.J., and Fricker, L.D. (2011b). Peptidomic approaches to study proteolytic activity. Curr Protoc Protein Sci Chapter 18, Unit18 13.

Llano, E., Adam, G., Pendas, A.M., Quesada, V., Sánchez, L.M., Santamaria, I., Noselli, S., and López-Otín, C. (2002). Structural and enzymatic characterization of Drosophila Dm2-MMP, a membrane-bound matrix metalloproteinase with tissue-specific expression. J Biol Chem 277, 23321-23329.

Marintcheva, B., Marintchev, A., Wagner, G., and Richardson, C.C. (2008). Acidic C-terminal tail of the ssDNA-binding protein of bacteriophage T7 and ssDNA compete for the same binding surface. Proc Natl Acad Sci U S A 105, 1855-1860.

Martens, L., Hermjakob, H., Jones, P., Adamski, M., Taylor, C., States, D., Gevaert, K., Vandekerckhove, J., and Apweiler, R. (2005). PRIDE: the proteomics identifications database. Proteomics 5, 3537-3545.

Marx, P.F., Brondijk, T.H., Plug, T., Romijn, R.A., Hemrika, W., Meijers, J.C., and Huizinga, E.G. (2008). Crystal structures of TAFI elucidate the inactivation mechanism of activated TAFI: a novel mechanism for enzyme autoregulation. Blood 112, 2803-2809.

McCall, K.A., Huang, C., and Fierke, C.A. (2000). Function and mechanism of zinc metalloenzymes. J Nutr 130, 1437S-1446S.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr 40, 658-674.

McKay, T.J., Phelan, A.W., and Plummer, T.H., Jr. (1979). Comparative studies on human carboxypeptidases B and N. Arch Biochem Biophys 197, 487-492.

Metz, M., Piliponsky, A.M., Chen, C.C., Lammel, V., Abrink, M., Pejler, G., Tsai, M., and Galli, S.J. (2006). Mast cells can enhance resistance to snake and honeybee venoms. Science 313, 526-530.

Moeller, C., Swindell, E.C., Kispert, A., and Eichele, G. (2003). Carboxypeptidase Z (CPZ) modulates Wnt signaling and regulates the development of skeletal elements in the chicken. Development 130, 5103-5111.

Morano, C., Zhang, X., and Fricker, L.D. (2008). Multiple Isotopic Labels for Quantitative Mass Spectrometry. Anal Chem 80, 9238-9309.

Morgan, T.H., Bridges, C.B., and Sturtevant, A.H. (1925). The genetics of Drosophila. Bibliogr Genet 2, 1-262.

Mortensen, U.H., Olesen, K., and Breddam, K. (1999). Carboxypeptidase C including carboxypeptidase Y. In Handbook of proteolytic enzymes, A.J. Barrett, N.D. Rawlings, and J.F. Woessner, eds. (London, Academic Press), pp. 389-393.

Mosca, A., Berruti, A., Russo, L., Torta, M., and Dogliotti, L. (2005). The neuroendocrine phenotype in prostate cancer: basic and clinical aspects. J Endocrinol Invest 28, 141-145.

Myers, R.M., Shearman, J.W., Kitching, M.O., Ramos-Montoya, A., Neal, D.E., and Ley, S.V. (2009). Cancer, Chemistry, and the Cell: Molecules that Interact with the Neurotensin Receptors. ACS Chem Biol 4, 503-525.

Nagakawa, O., Ogasawara, M., Murata, J., Fuse, H., and Saiki, I. (2001). Effect of prostatic neuropeptides on migration of prostate cancer cell lines. Int J Urol 8, 65-70.

Nakazawa, T., Yamaguchi, M., Okamura, T.A., Ando, E., Nishimura, O., and Tsunasawa, S. (2008). Terminal proteomics: N- and C-terminal analyses for high-fidelity identification of proteins using MS. Proteomics 8, 673-685.

Nooren, I.M.A., and Thornton, J.M. (2003). Structural characterisation and functional significance of transient protein-protein interactions. J Mol Biol 325, 991-1018.

Novikova, E.G., Eng, F.J., Yan, L., Qian, Y., and Fricker, L.D. (1999). Characterization of the enzymatic properties of the first and second domains of metallocarboxypeptidase D. J Biol Chem 274, 28887-28892.

Novikova, E.G., Reznik, S.E., Varlamov, O., and Fricker, L.D. (2000). Carboxypeptidase Z is present in the regulated secretory pathway and extracellular matrix in cultured cells and in human tissues. J Biol Chem 275, 4865-4870.

Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., et al. (1992). The a/b hydrolase fold. Prot Engng 5, 197-211.

Otero, A., Rodriguez de la Vega, M., Tanco, S., Lorenzo, J., Aviles, F.X., and Reverter, D. (2012). The novel structure of a cytosolic M14 metallocarboxypeptidase (CCP) from Pseudomonas aeruginosa: a model for mammalian CCPs. FASEB J 26, 3754-3764.

Pallares, I., Bonet, R., Garcia-Castellanos, R., Ventura, S., Aviles, F.X., Vendrell, J., and Gomis-Ruth, F.X. (2005). Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin. Proc Natl Acad Sci U S A 102, 3978-3983.

Pallares, I., Fernandez, D., Comellas-Bigler, M., Fernandez-Recio, J., Ventura, S., Aviles, F.X., Bode, W., and Vendrell, J. (2008). Direct interaction between a human digestive protease and the mucoadhesive poly(acrylic acid). Acta Crystallogr D Biol Crystallogr D64, 784-791.

Paturle-Lafanechere, L., Manier, M., Trigault, N., Pirollet, F., Mazarguil, H., and Job, D. (1994). Accumulation of delta 2-tubulin, a major tubulin variant that cannot be tyrosinated, in neuronal tissues and in stable microtubule assemblies. J Cell Sci 107 (Pt 6), 1529-1543.

Pejler, G., Abrink, M., Ringvall, M., and Wernersson, S. (2007). Mast cell proteases. Adv Immunol 95, 167-255.

Pejler, G., Knight, S.D., Henningsson, F., and Wernersson, S. (2009). Novel insights into the biological function of mast cell carboxypeptidase A. Trends Immunol 30, 401-408.

Pereira, H.J., Souza, L.L., Costa-Neto, C.M., Salgado, M.C., and Oliveira, E.B. (2012). Carboxypeptidases A1 and A2 from the perfusate of rat mesenteric arterial bed differentially process angiotensin peptides. Peptides 33, 67-76.

Peterson, L.M., Holmquist, B., and Bethune, J.L. (1982). A unique activity assay for carboxypeptidase A in human serum. Anal Biochem 125, 420-426.

Prapunpoj, P., and Leelawatwattana, L. (2009). Evolutionary changes to transthyretin: structure-function relationships. FEBS J 276, 5330-5341.

Pshezhetsky, A.V., and Hinek, A. (2009). Serine carboxypeptidases in regulation of vasoconstriction and elastogenesis. Trends Cardiovasc Med 19, 11-17.

Puente, X.S., and López-Otín, C. (1997). The PLEES-proteins - A family of structurally related enzymes widely distributed from bacteria to humans. Biochem J 332, 947-949.

Puente, X.S., Sanchez, L.M., Overall, C.M., and Lopez-Otin, C. (2003). Human and mouse proteases: a comparative genomic approach. Nat Rev Genet 4, 544-558.

Raijmakers, R., Neerincx, P., Mohammed, S., and Heck, A.J. (2010). Cleavage specificities of the brother and sister proteases Lys-C and Lys-N. Chem Commun (Camb) 46, 8827-8829.

Raksakulthai, R., and Haard, N.F. (2001). Purification and characterization of a carboxypeptidase from squid hepatopancreas (Illex illecebrosus). J Agric Food Chem 49, 5019-5030.

Rawlings, N.D., and Barrett, A.J. (1993). Evolutionary families of peptidases. Biochem J 290 (Pt 1), 205-218.

Rawlings, N.D., Barrett, A.J., and Bateman, A. (2010). MEROPS: the peptidase database. Nucleic Acids Res 38, D227-D233.

Rawlings, N.D., Barrett, A.J., and Bateman, A. (2012). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 40, D343-350.

Rawlings, N.D., Morton, F.R., and Barrett, A.J. (2006). MEROPS: the peptidase database. Nucleic Acids Res 34, D270-272.

Rawlings, N.D., Morton, F.R., Kok, C.Y., Kong, J., and Barrett, A.J. (2008). MEROPS: the peptidase database. Nucleic Acids Res 36 (Database issue), D320-D325.

Reverter, D., Garcia-Saez, I., Catasus, L., Vendrell, J., Coll, M., and Aviles, F.X. (1997). Characterisation and preliminary X-ray diffraction analysis of human pancreatic procarboxypeptidase A2. FEBS Lett 420, 7-10.

Reverter, D., Maskos, K., Tan, F., Skidgel, R.A., and Bode, W. (2004). Crystal structure of human carboxypeptidase M, a membrane-bound enzyme that regulates peptide hormone activity. J Mol Biol 338, 257-269.

Reverter, D., Vendrell, J., Canals, F., Horstmann, J., Aviles, F.X., Fritz, H., and Sommerhoff, C.P. (1998). A carboxypeptidase inhibitor from the medical leech Hirudo medicinalis. Isolation, sequence analysis, cDNA cloning, recombinant expression, and characterization. J Biol Chem 273, 32927-32933.

Reznik, S.E., and Fricker, L.D. (2001). Carboxypeptidases from A to Z : implications in embryonic development and Wnt binding. Cell Mol Life Sci 58, 1790-1804.

Richter-Cook, N.J., Dever, T.E., Hensold, J.O., and Merrick, W.C. (1998). Purification and characterization of a new eukaryotic protein translation factor. Eukaryotic initiation factor 4H. J Biol Chem 273, 7579-7587.

Ridderstrom, M., Cameron, A.D., Jones, T.A., and Mannervik, B. (1998). Involvement of an active-site Zn2+ ligand in the catalytic mechanism of human glyoxalase I. J Biol Chem 273, 21623-21628.

Rivera, J. (2006). Snake bites and bee stings: the mast cell strikes back. Nat Med 12, 999-1000.

Rodriguez de la Vega, M., Sevilla, R.G., Hermoso, A., Lorenzo, J., Tanco, S., Diez, A., Fricker, L.D., Bautista, J.M., and Aviles, F.X. (2007). Nna1-like proteins are active metallocarboxypeptidases of a new and diverse M14 subfamily. FASEB J 21, 851-865.

Rogowski, K., van Dijk, J., Magiera, M.M., Bosc, C., Deloulme, J.C., Bosson, A., Peris, L., Gold, N.D., Lacroix, B., Grau, M.B., et al. (2010). A family of protein-deglutamylating enzymes associated with neurodegeneration. Cell 143, 564-578.

Ross, P.L., Cheng, I., Liu, X., Cicek, M.S., Carroll, P.R., Casey, G., and Witte, J.S. (2009). Carboxypeptidase 4 gene variants and early-onset intermediate-to-high risk prostate cancer. BMC Cancer 9, 69.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5, 725-738.

Rusconi, F., Potier, M.C., Le Caer, J.P., Schmitter, J.M., and Rossier, J. (1997). Characterization of the chicken telokin heterogeneity by time-of-flight mass spectrometry. Biochemistry 36, 11021-11026.

Ryan, C.A., Hass, G.M., and Kuhn, R.W. (1974). Purification and properties of a carboxypeptidase inhibitor from potatoes. J Biol Chem 249, 5495-5499.

Sahab, Z.J., Hall, M.D., Me Sung, Y., Dakshanamurthy, S., Ji, Y., Kumar, D., and Byers, S.W. (2011). Tumor suppressor RARRES1 interacts with cytoplasmic carboxypeptidase AGBL2 to regulate the alpha-tubulin tyrosination cycle. Cancer Res 71, 1219-1228.

Sahab, Z.J., Kirilyuk, A., Zhang, L., Khamis, Z.I., Pompach, P., Sung, Y., and Byers, S.W. (2012). Analysis of tubulin alpha-1A/1B C-terminal tail post-translational poly-glutamylation reveals novel modification sites. J Proteome Res 11, 1913-1923.

Samyn, B., Hardeman, K., Van der Eycken, J., and Van Beeumen, J. (2000). Applicability of the alkylation chemistry for chemical C-terminal protein sequence analysis. Anal Chem 72, 1389-1399.

Sanglas, L., Aviles, F.X., Huber, R., Gomis-Ruth, F.X., and Arolas, J.L. (2009). Mammalian metallopeptidase inhibition at the defense barrier of Ascaris parasite. Proc Natl Acad Sci U S A 106, 1743-1747.

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

Schilling, O., Barre, O., Huesgen, P.F., and Overall, C.M. (2010). Proteome-wide analysis of protein carboxy termini: C terminomics. Nat Methods 7, 508-511.

Schilling, O., and Overall, C.M. (2007). Proteomic discovery of protease substrates. Curr Opin Chem Biol 11, 36-45.

Schilling, O., and Overall, C.M. (2008). Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nat Biotechnol 26, 685-694.

Schmidt, T.G., and Skerra, A. (2007). The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins. Nat Protoc 2, 1528-1535.

Settle, S.H., Jr., Green, M.M., and Burtis, K.C. (1995). The silver gene of Drosophila melanogaster encodes multiple carboxypeptidases similar to mammalian prohormone-processing enzymes. Proceedings of the National Academy of Sciences of the United States of America 92, 9470-9474.

Shang, Y., Li, B., and Gorovsky, M.A. (2002). Tetrahymena thermophila contains a conventional gamma-tubulin that is differentially required for the maintenance of different microtubule-organizing centers. J Cell Biol 158, 1195-1206.

Shirakawa, H., Tanigawa, T., Sugiyama, S., Kobayashi, M., Terashima, T., Yoshida, K., Arai, T., and Yoshida, M. (1997). Nuclear accumulation of HMG2 protein is mediated by

basic regions interspaced with a long DNA-binding sequence, and retention within the nucleus requires the acidic carboxyl terminus. Biochemistry 36, 5992-5999.

Sidyelyeva, G., Baker, N.E., and Fricker, L.D. (2006). Characterization of the molecular basis of the Drosophila mutations in carboxypeptidase D. Effect on enzyme activity and expression. J Biol Chem 281, 13844-13852.

Sidyelyeva, G., and Fricker, L.D. (2002). Characterization of Drosophila carboxypeptidase D. J Biol Chem 277, 49613-49620.

Sidyelyeva, G., Wegener, C., Schoenfeld, B.P., Bell, A.J., Baker, N.E., McBride, S.M., and Fricker, L.D. (2010). Individual carboxypeptidase D domains have both redundant and unique functions in Drosophila development and behavior. Cell Mol Life Sci 67, 2991-3004.

Skidgel, R.A. (1988). Basic carboxypeptidases : regulators of peptide hormone activity. Trends Pharmacol Sci 9, 299-304.

Skidgel, R.A. (1995). Human carboxypeptidase N: lysine carboxypeptidase. Methods Enzymol 248, 653-663.

Skidgel, R.A. (1996). Structure and function of mammalian zinc carboxypeptidases. In Zinc metalloproteases in health and disease, N.M. Hooper, ed. (London, Taylor and Francis Ltd.), pp. 241-283.

Skidgel, R.A. (2004). 252. Carboxypeptidase M. In Handbook of proteolytic enzymes, A.J. Barrett, N.D. Rawlings, and J.F. Woessner Jr., eds. (London, Elsevier), pp. 851-854.

Skidgel, R.A., and Erdos, E.G. (1998). Cellular carboxypeptidases. Immunol Rev 161, 129-141.

Skidgel, R.A., and Erdos, E.G. (2007). Structure and function of human plasma carboxypeptidase N, the anaphylatoxin inactivator. Int Immunopharmacol 7, 1888-1899.

Skidgel, R.A., McGwire, G.B., and Li, X.Y. (1996). Membrane anchoring and release of carboxypeptidase M: implications for extracellular hydrolysis of peptide hormones. Immunopharmacology 32, 48-52.

Smith, G.K., Banks, S., Blumenkopf, T.A., Cory, M., Humphreys, J., Laethem, R.M., Miller, J., Moxham, C.P., Mullin, R., Ray, P.H., et al. (1997). Toward antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1 and novel in vivo stable prodrugs of methotrexate. J Biol Chem 272, 15804-15816.

Son, Y.J., Kim, C.K., Kim, Y.B., Kweon, D.H., Park, Y.C., and Seo, J.H. (2009). Effects of citraconylation on enzymatic modification of human proinsulin using trypsin and carboxypeptidase B. Biotechnol Prog 25, 1064-1070.

Song, L., and Fricker, L.D. (1995). Purification and characterization of carboxypeptidase D, a novel carboxypeptidase E-like enzyme, from bovine pituitary. Journal of Biological Chemistry 270, 25007-25013.

Song, L., and Fricker, L.D. (1996). Tissue distribution and characterization of soluble and membrane-bound forms of metallocarboxypeptidase D. J Biol Chem 271, 28884-28889.

Staes, A., Impens, F., Van Damme, P., Ruttens, B., Goethals, M., Demol, H., Timmerman, E., Vandekerckhove, J., and Gevaert, K. (2011). Selecting protein N-terminal peptides by combined fractional diagonal chromatography. Nat Protoc 6, 1130-1141.

Staes, A., Van Damme, P., Helsens, K., Demol, H., Vandekerckhove, J., and Gevaert, K. (2008). Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). Proteomics 8, 1362-1370.

Tanco, S., Zhang, X., Morano, C., Aviles, F.X., Lorenzo, J., and 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, 18385-18396.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix.choice. Nucl Acids Res 22, 4673-4680.

Toyo'oka, T. (1998). Modern Derivatization Methods for Separation Science (New York, Wiley).

Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R., et al. (2009). FlyBase: enhancing Drosophila Gene Ontology annotations. Nucl Acids Res 37 (Database issue), D555-D559.

Umetsu, H., Matsuoka, H., and Ichishima, E. (1983). Debittering mechanism of bitter peptides from milk casein by wheat carboxypeptidase. Journal of Agricultural and Food Chemistry 31, 50-53.

Ungar, D. (2009). Golgi linked protein glycosylation and associated diseases. Semin Cell Dev Biol 20, 762-769.

Valero, J., Berciano, M.T., Weruaga, E., Lafarga, M., and Alonso, J.R. (2006). Pre-neurodegeneration of mitral cells in the pcd mutant mouse is associated with DNA damage, transcriptional repression, and reorganization of nuclear speckles and Cajal bodies. Mol Cell Neurosci 33, 283-295.

Valnickova, Z., Thaysen-Andersen, M., Hojrup, P., Christensen, T., Sanggaard, K.W., Kristensen, T., and Enghild, J.J. (2009). Biochemical characterization of bovine plasma thrombin-activatable fibrinolysis inhibitor (TAFI). BMC Biochem 10, 13.

Van Damme, P., Staes, A., Bronsoms, S., Helsens, K., Colaert, N., Timmerman, E., Aviles, F.X., Vandekerckhove, J., and Gevaert, K. (2010). Complementary positional proteomics for screening substrates of endo- and exoproteases. Nat Methods 7, 512-515.

van den Berg, B.H., and Tholey, A. (2012). Mass spectrometry-based proteomics strategies for protease cleavage site identification. Proteomics 12, 516-529.

van Dijk, J., Miro, J., Strub, J.M., Lacroix, B., van Dorsselaer, A., Edde, B., and Janke, C. (2008). Polyglutamylation is a post-translational modification with a broad range of substrates. J Biol Chem 283, 3915-3922.

Vendrell, J., and Avilés, F.X. (1999). Carboxypeptidases. In Proteases: new perspectives, V. Turk, ed. (Basel, Birkhauser), pp. 13-34.

Vendrell, J., Aviles, F.X., and Fricker, L.D. (2006). Metallocarboxypeptidases. In Handbook of Metalloproteins (John Wiley & Sons, Ltd).

Vendrell, J., Querol, E., and Aviles, F.X. (2000). Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. Biochim Biophys Acta 1477, 284-298.

Vias, M., Burtt, G., Culig, Z., Veerakumarasivam, A., Neal, D.E., and Mills, I.G. (2007). A role for neurotensin in bicalutamide resistant prostate cancer cells. Prostate 67, 190-202.

Vovchuk, I.L., and Petrov, S.A. (2008). [Role of carboxypeptidases in carcinogenesis]. Biomed Khim 54, 167-178.

Vriend, G. (1990). WHAT IF : a molecular modelling and drug design program. J Mol Graph 8, 52-56.

Waldschmidt-Leitz, E., and Purr, A. (1929). Über Proteinase und Carboxy-Polypeptidase aus Pankreas (XVII. Mitteilung zur Spezifität tierischer Proteasen). Berichte der deutschen chemischen Gesellschaft (A and B Series) 62, 2217-2226.

Wang, T., and Morgan, J.I. (2007). The Purkinje cell degeneration (pcd) mouse: an unexpected molecular link between neuronal degeneration and regeneration. Brain Res 1140, 26-40.

Wang, T., Parris, J., Li, L., and Morgan, J.I. (2006). The carboxypeptidase-like substrate-binding site in Nna1 is essential for the rescue of the Purkinje cell degeneration (pcd) phenotype. Mol Cell Neurosci 33, 200-213.

Wang, Y., Luo, W., and Reiser, G. (2008). Trypsin and trypsin-like proteases in the brain: proteolysis and cellular functions. Cell Mol Life Sci 65, 237-252.

Wei, S., Feng, Y., Kalinina, E., and Fricker, L.D. (2003). Neuropeptide-processing carboxypeptidases. Life Sci 73, 655-662.

Wei, S., Segura, S., Vendrell, J., Aviles, F.X., Lanoue, E., Day, R., Feng, Y., and Fricker, L.D. (2002). Identification and characterization of three members of the human metallocarboxypeptidase gene family. J Biol Chem 277, 14954-14964.

White, D. (2004). Use for Carboxypeptidase-A4 in the diagnosis and treatment of metabolic disorders In US Patent 2004/0077001 A1 by Millennium Pharmaceuticals Inc.

Willemse, J.L., Polla, M., Olsson, T., and Hendriks, D.F. (2008). Comparative substrate specificity study of carboxypeptidase U (TAFIa) and carboxypeptidase N: development of highly selective CPU substrates as useful tools for assay development. Clin Chim Acta 387, 158-160.

Witte, J.S., Goddard, K.A., Conti, D.V., Elston, R.C., Lin, J., Suarez, B.K., Broman, K.W., Burmester, J.K., Weber, J.L., and Catalona, W.J. (2000). Genomewide scan for prostate cancer-aggressiveness loci. Am J Hum Genet 67, 92-99.

Wright, T.R. (1987a). The genetics of biogenic amine metabolism, sclerotization, and melanization in Drosophila melanogaster. Adv Genet 24, 127-222.

Wright, T.R.F. (1987b). The genetics of biogenic amine metabolism, sclerotization, and melanization in Drosophila melanogaster. Adv Genet 24, 127-222.

Wu, H.Y., Wang, T., Li, L., Correia, K., and Morgan, J.I. (2012). A structural and functional analysis of Nna1 in Purkinje cell degeneration (pcd) mice. FASEB J.

Xu, G., Shin, S.B., and Jaffrey, S.R. (2011). Chemoenzymatic labeling of protein C-termini for positive selection of C-terminal peptides. ACS Chem Biol 6, 1015-1020.

Ye, X., Luke, B., Andresson, T., and Blonder, J. (2009). 18O stable isotope labeling in MS-based proteomics. Brief Funct Genomic Proteomic 8, 136-144.

Zappacosta, F., Pessi, A., Bianchi, E., Venturini, S., Sollazzo, M., Tramontano, A., Marino, G., and Pucci, P. (1996). Probing the tertiary structure of proteins by limited proteolysis and mass spectrometry: the case of Minibody. Prot Sci 5, 802-813.

Zhang, X., Che, F.Y., Berezniuk, I., Sonmez, K., Toll, L., and Fricker, L.D. (2008). Peptidomics of Cpe(fat/fat) mouse brain regions: implications for neuropeptide processing. J Neurochem 107, 1596-1613.

Zhang, Y. (2007). Template-based modeling and free modeling by I-TASSER in CASP7. Proteins 69 Suppl 8, 108-117.

Zhao, L., Buckman, B., Seto, M., Morser, J., and Nagashima, M. (2003). Mutations in the substrate binding site of thrombin-activatable fibrinolysis inhibitor (TAFI) alter its substrate specificity. J Biol Chem 278, 32359-32366.