Universitat Autònoma
de Barcelona

Structural and functional characterization of regulatory metallocarboxypeptidases: Studies on human carboxypeptidases D and Z, and the transthyretin-like domain

Javier Garcia Pardo

2015

Universitat Autònoma de Barcelona

Structural and functional characterization of regulatory metallocarboxypeptidases: Studies on human carboxypeptidases D and Z, and the transthyretin-like domain

Doctoral thesis presented by Javier Garcia Pardo for the degree of PhD in Biochemistry, Molecular Biology and Biomedicine 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 i Puigvert

Javier Garcia Pardo

Prof. Julia Lorenzo Rivera

Prof. Francesc Xavier Avilés Puigvert

Bellaterra, Setember 2015

Table of contents

TABLE OF CONTENTS

List of abbreviations

LIST OF ABBREVIATIONS

ACE	Angiotensin-converting enzyme
APR	Aggregation prone region
CCP	Cytosolic carboxypeptidase
CD	Catalytic domain
COFRADIC	COmbined FRActional DIagonal Chromatography
CP	Carboxypeptidase
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
ECM	Extracellular matrix
ER	Endoplasmic reticulum
FALS	Familial amyotrophic lateral sclerosis
GEMSA	2-guanidinoethyl-mercaptosuccinic acid
H-TTL	Transthyretin-like domain from the first catalytic domain of human Carboxypeptidase D
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

MS/MS	Tandem mass spectrometry
PBS	Phosphate-buffered saline
PCPA4	Procarboxypeptidase A4
PDB	Protein Data Bank
PVDF	Polyvinylidene difluoride
RP-HPLC	Reversed-phase high performance liquid chromatography
SCPs	Serine carboxypeptidases
SOD1	Superoxid dismutase 1
TAFI	Thrombin-activatable fibrinolysis inhibitor
TAFI(a)	(activated) thrombin-activatable fibrinolysis inhibitor
TFA	Trifluoroacetic acid
TLL	Transthyretin-like domain
TMAB	4-trimethyl-ammoniumbutyrate
TTR	Transthyretin

Preface

PREFACE

The present thesis consists of five independent research works situated in the field of metallocarboxypeptidases, particularly focusing in two members of the M14B subfamily: Carboxypeptidase D and carboxypeptidase Z.

The first chapter describes, for the first time, the biochemical characterization of the aggregational properties of a transthyretin-like domain from the first catalytic repeat of human Carboxypeptidase D, termed here as h-TTL. This work was performed in collaboration with Prof. Salvador Ventura (Institut de Biotecnologia i Biomedicina, Bellaterra, Spain). Dr. Ricardo Graña also contributed to this work.

The second chapter presents the crystal structure solved at ultra-high resolution of the h-TTL described in the first chapter. The information derived in the present study might facilitate the understanding of the biological roles of the transthyretin-like domains found in M14B subfamily members and would be an interesting tool to analyze in detail the structural properties, folding and physiological roles of these domains. The crystallographic studies performed in this chapter were done in collaboration with Prof. David Reverter (Institut de Biotecnologia i Biomedicina, Bellaterra, Spain). Dr. Pablo Gallego generously contributed experimentally to this work.

The third chapter comprises the characterization of the substrate specificity of human carboxypeptidase D by using a combination of quantitative peptidomics approaches. This unique enzyme with multiple catalytic sites might be implicated in the processing of neuropeptides and growth factors, thereby the study of its mechanism of action is of significant importance for biomedicine. This work was perfomed in collaboration with Prof. Lloyd D. Fricker (Albert Einstein College of Medicine, New York, USA) through a stay in his laboratory. Dr. Sebastian Tanco also contributed both experimentally and in the supervision of this work.

The fourth chapter describes de development of a simple and inexpensive method to produce heparin-affinity carboxypeptidases using mammalian cells, taking as

example the case of carboxypeptidase Z. Further, this method can be used to produce other metallocarboxypeptidases of interest for both biotechnological and biomedical applications. Dr. Sebastian Tanco was also implicated in the development of this methodology.

The fifth chapter applies several quantitative peptidomics approaches, in a similar manner with the third chapter, to characterize the substrate specificity of the human carboxypeptidase Z. Furthermore, this work present the modelling of the frizzled-like domain found in this enzyme in order to analyze their role in Wnt binding. This work was performed in collaboration with Prof. Lloyd D. Fricker (Albert Einstein College of Medicine, New York, USA) through a stay in his laboratory. Sebastian Tanco greatly contributed to this work. We thank Prof. Mireia Duñach and Dr. Beatriz del Valle for its advice on the field of Wnt signalling.

The five research works presented in this thesis are the basis of the following scientific publications:

I. J. Garcia-Pardo, R. Graña-Montes, M. Fernandez-Mendez, A. Ruyra, N. Roher, F. X. Aviles, J. Lorenzo, and S. Ventura. "*Amyloid Formation by Human Carboxypeptidase D Transthyretin-like Domain Under Physiological Conditions*". J. Biol. Chem. (2014) 289:33783-33796.

II. J. Garcia-Pardo, S. Tanco, S. Dasgupta, F. X. Avilés, J. Lorenzo and L.D. Fricker. "*Substrate specificity of human Metallocarboxypeptidase D: Comparison of the two active carboxypeptidase domains*". Submitted to the J. Biol. Chem. (2015)

III. J. Garcia-Pardo, S. Tanco, S. Dasgupta, F. X. Avilés, J. Lorenzo and L.D. Fricker. "*Substrate specificity and structural modeling of human carboxypeptidase Z: The unique protease with a Frizzled-like domain*". In preparation. (2015)

IV. J. Garcia-Pardo, S. Tanco, R. Fernández-Alvarez, F. X. Avilés and J. Lorenzo. "*A simple method to improve production of heparin-affinity metallocarboxypeptidases using mammalian cells*". In preparation. (2015)

Other articles co-authored which are not part of this thesis:

V. L. García-Fernández, J. Garcia-Pardo, O. Tort, I. Prior, M. Brust, J. Lorenzo and V.F. Puntes. "*Conserved effects and altered trafficking of Cetuximab antibodies conjugated to gold nanoparticles with control of their number and orientation*". Submitted to ACS Nano. (2015).

VI. D. Lufrano, J. Cotabarren, J. Garcia-Pardo, R. Fernandez-Alvarez, O. Tort, S. Tanco. F.X. Avilés, L. Julia and W.D. Obregón. "*Biochemical characterization of a new carboxypeptidase inhibitor from a variety of Andean potatoes*". Subbmited to Phytochemistry. (2015).

VII. C. Núñez, E. Oliveira, J. García-Pardo, M. Diniz, J. Lorenzo, J. L. Capelo and C. Lodeiro. "*A novel quinoline molecular probe and the derived functionalized gold nanoparticles: Sensing properties and cytotoxicity studies in MCF-7 human breast cancer cells*". J. Inorg. Biochem. (2014) 137:115–122.

VIII. E. Oliveira, H. M. Santos, J. Garcia-Pardo, M. Diniz, J. Lorenzo, B. Rodríguez-González, J. L. Capelo, and C. Lodeiro. "*Synthesis of functionalized fluorescent silver nanoparticles and their toxicological effect in aquatic environments (Goldfish) and HEPG2 cells*". Front. Chem. (2013) 1:1–11.

IX. S. Tanco, J. Lorenzo, J. Garcia-Pardo, S. Degroeve, L. Martens, F. X. Aviles, K. Gevaert, and P. Van Damme. "*Proteome-derived peptide libraries to study the substrate specificity profiles of carboxypeptidases*". Mol. Cell. Proteomics. (2013) 12:2096–110.

Note about the format

The present thesis is composed of five different works, all of them situated in the field of metallocarboxypeptidases. For a better comprehension, this thesis has been divided into five independent chapters corresponding to one already published work, one submitted work and others two still in the manuscript form. Furthermore, other interesting data not considered yet for publication, have been included due to its relevance. Each chapter includes a specific Introduction, Experimental section, Results and Discussion. Despite, a general Background was provided for a better understanding of the following chapters. The thesis ends with a general discussion and concluding remarks with the main findings from the present work.

Background

BACKGROUND

B.1 PROTEASES

B.1.1 BRIEF INTRODUCTION

Proteases (also termed as proteinases, peptidases or proteolytic enzymes) are enzymes that catalyze the hydrolysis of peptide bonds in proteins and peptides. Enzymes are large biological molecules, generally of proteinaceous nature, that greatly accelerate the rate of nearly all of the chemical reactions that occur in biological systems.

Proteases are ubiquitous throughout all kingdoms of living organisms, as well as in viruses (N.D. Rawlings et al. 2014). The human genome encodes over 560 proteases or homologues which represent about 2 % of the known genes, constituting one of the largest protein families (Puente et al. 2003). The complete repertoire of proteases expressed by an organism is defined as degradome. In the recent years, the development of degradomic approaches allowed the study of degradomes from different organism and at different physiopathological states. These studies have provided important information about the biological roles of proteases *in vivo* (Quesada et al. 2009).

Initially, proteases were characterized as nonspecific degradative enzymes associated with protein catabolism. However, it is becoming increasingly recognized that proteases can act as precise modifiers of many protein molecules to achieve a precise cellular control of biological processes in all living organisms (López-Otín & Overall 2002). Proteolytic enzymes are involved in multiple physiological and pathological processes. They exert a high-order of posttranslational control over metabolic reactions that sustain life such as cell-cycle progression, cell proliferation, cell death, DNA replication, tissue remodeling, wound healing and immune response. Furthermore, dysfunction of protease activity can lead to several pathologies such as cardiovascular and inflammatory diseases, cancer, osteoporosis and neurological disorders, as well as infectious diseases (Turk 2006).

B.1.2 CLASSIFICATION OF PROTEASES

Proteases are of great relevance to biology, medicine and biotechnology. Due to its importance, proteases have been traditionally classified according to different properties. However, there are three criteria that are more commonly used for that large family of proteins (Rawlings & Barrett 1999).

B.1.2.1 Catalytic mechanism

Through evolution, proteases have adapted to a wide range of conditions found in complex organisms (variations in pH, reductive environment, temperature and so on) and use different catalytic mechanisms for substrate hydrolysis. According to the chemical nature of the catalytic site or catalytic mechanism used (Turk 2006) proteases can classified in six major catalytic classes as serine, threonine, cysteine, aspartic, glutamic and metalloproteases, as well as a group of proteases which mechanism of action is yet unknown (see **Table B.1**).

Table B.1 Catalytic type of proteases and selected examples

Mechanism	Catalytic Type	Examples
Covalent catalysis	Serine	Trypsin, prolyl oligopeptidase
	Cysteine	Papain, Cathepsin K
	Threonine	Proteasomal subunits
Acid-base catalysis	Metallo	Caboxypeptidase A, Thermolysin
	Aspartic	Pepsin, Cathepsin E
	Glutamic	Aspergilloglutamic peptidase
Unknown	Unknown	Collagenase, yabG protein

In serine, cysteine and threonine proteases the nucleophile is part of an amino acid. This mechanism of action is known as covalent catalysis. For cysteine proteases, a –SH group on the side chain of a Cys amino acid located at the active site, whereas in

threonine and serine proteases, the nucleophile is a –OH group from a Thr and Ser residues, respectively. In this type of proteases, His residues normally function as a base. Instead, in metalloproteases, aspartic and glutamic proteases the nucleophile is an activated water molecule that interacts directly with the Zinc atom or with Asp or Glu residues, respectively, and acts as acid and base. This mechanism is known as non-covalent catalysis.

B.1.2.2 Type of reaction catalyzed and location of the cleavage site

One of the most commonly used systems for enzyme classification and nomenclature is the EC number system (Enzyme Commission number) that is recommended by the NC-IUBMB (Nomenclature Committee of the International Union of Biochemistry and Molecular Biology). This system of classification is based on the reactions they catalyze. Following this system, enzymes are divided in six classes: (1) Oxido-reductases, (2) Transferases, (3) Hydrolases, (4) Lyases, (5) Isomerases and (6) Ligases.

Proteases are enzymes that hydrolyze peptide bonds. For this, are classified into the subclass 3.4, according to the EC classification. Some proteases specifically cleave substrates either from the N- or C-termini (both known as exopeptidases) and/or in the middle of proteins and peptides (endopeptidases). Accordingly, proteases are primarily divided into aminopeptidases, carboxypeptidases or endopeptidases, respectively (see **Figure B.1**). Further, some exopeptidases act at a free N-terminus and release a single amino acid residue (aminopeptidases), dipeptides (dipeptidyl-peptidases) or tripeptides (dipeptidyl-peptidases). In a similar manner, other exopeptidases are specific for cleave a free C-terminus and release single amino acid residues (carboxypeptidases), dipeptides (peptidyl-dipeptidases) or tripeptides (peptidyl-tripeptidases).

Other exopeptidases catalyze the hydrolysis of specific N- or C-terminal residues that are modified with different post-translational mofifications (i.e residues that are substituted, cyclized, or linked to isopeptide bonds). This is de case of omega peptidases.

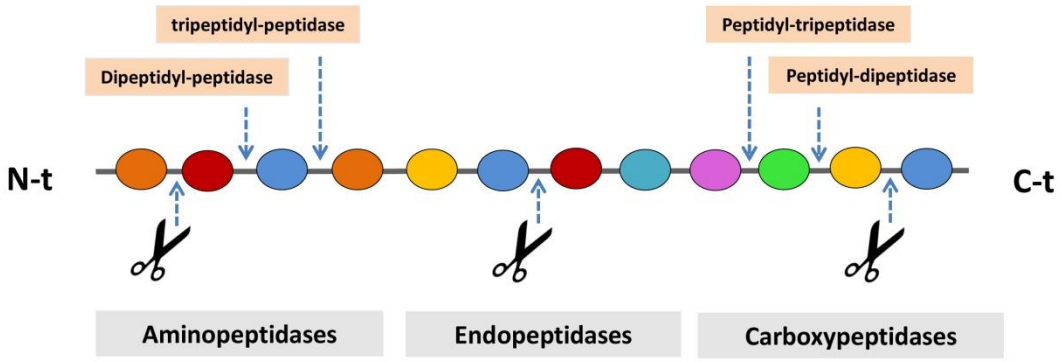


Figure B.1 Schematic representation of principal exopeptidases and endopeptidases cleavage sites. Scheme of proteases classification, according to the localization of its site of cleavage in a polypeptide substrate. Different single amino acids of the polypeptide chain are indicated with different colours. The N-termini (N-t) and C-termini (C-t) are indicated. The different sites of cleavage are indicated with arrows.

A difference with exopeptidases, endopeptidases hydrolyzes internal bonds in polypeptide chains and can be classified according to its catalytic mechanism (**Figure B.1**). The complete numerical classification according to the NC-IUBMB can be found in the following link: http://www.chem.qmul.ac.uk/iubmb/enzyme/EC3/4/. Despite, the most relevant subfamilies of proteases are summarized in **Table B.2**.

Table B.2 Classification of peptidases according to the NC-IUBMB

EC number	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	Metallocarboxypeptidases
3.4.25	Threonine endopeptidases
3.4.99	Endopeptidases of unknown type

B.1.2.3 Structural relationships

The MEROPS database was proposed 22 years ago by Rawlings and Barrett as a useful and systematic method to classify proteases based on sequential and structural features (Rawlings & Barrett 1993). This system classifies proteases in families, which are constituted by homologues with significant primary structure similarities and consequently, common ancestry. Protease families are further grouped into different clans based on its similar tertiary structure and catalytic mechanism. At the time of writing, there are nearly 250 families of proteases in the MEROPS and approximately 50 clans of proteases (N.D. Rawlings et al. 2014). An identifier is assigned to each family of proteases, corresponding with a letter denoting the catalytic type of peptidases (Aspartic (A), Cysteine (C), Glutamic (G), Metallo (M), Asparagine (N), Mixed (P), Serine (S), Threonine (T) and Unknown (U)) plus a number. Some families are divided into subfamilies because there is evidence of a very ancient divergence within the family and are represented with the addition of a letter (i.e. the family M14 of metallocarboxypeptidases is divided into four subdivisions M14A, M14B, M14C and M14D). The complete list of protease families and clans can be found in the web version of the MEROPS database (http://merops.sanger.ac.uk/).

B.1.3 SUBSTRATE SPECIFICITY OF PROTEASES AND ITS TERMINOLGY

To date, the most widely accepted conceptual model to explain the substrate specificities of proteases is the model of Schneider & Berger (Schechter & Berger 1967). This model considers that each specificity pocket in the catalytic site of the enzyme is able to accommodate the side-chain of a single amino acid residue in the substrate. Therefore, the structure of the active site of the protease determines which substrate residues are able to bind specific substrate binding sites of the protease.

In the substrate, residues are indicated as shown in **Figure B.2** (-P5-P4-P3-P2-P1-P1'-P2'-P3'-P4'-P5'). The exact site of protease cleavage is the peptide bond located between residues P1 and P1'. The corresponding surface of the protease that is able to accommodate a single side chain of the substrate is the specificity pocket (or specificity subsite).

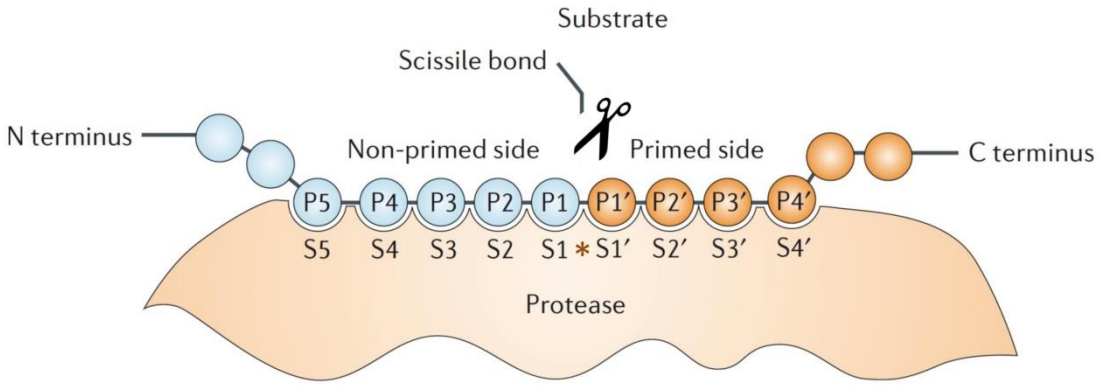


Figure B.2 Schematic representation of the model of Schneider & Berger. Schematic diagram of a protease binding a peptide substrate. The diagram shows the active site of a protease, in which are represented the locations of its specificity pockets (called S1 to Sn and S1' to Sn'). The enzyme accommodate several residues from the substrate numbered from P1 to Pn and P1' to Pn '. The primed sites correspond to the C-terminal region sequence from the scissile bond and the non-primed side to the N-terminal.

According to the substrate nomenclature, the specificity pockets located in the surface of the protease are named as S5, S4, S3, S2, S1 and S1', S2', S3', S4' and S5' (**Figure B.2**). In the figure above, the catalytic site of the enzyme is shown with and asterisk and the scissile bond by scissors. For exopeptidases the cleft is likely to be "blind" on the side, not extending beyond S1 or S1' for an aminopeptidase or carboxypeptidase, respectively.

B.1.4 CARBOXYPEPTIDASES

Concerning the cleavage site, carboxypeptidases (CPs) can be defined as proteolytic enzymes that catalyze the hydrolysis of peptide bonds at the C-terminus of peptides and proteins. The carboxypeptidase function may be carried out by three different types of exopeptidases: serine, cysteine or metallo carboxypeptidases (see **Table B.1** and **Table B.3**). They are widely distributed throughout all living organisms and perform a wide range of physiological functions, ranging from food digestion to the fine control of cell signaling.

Table B.3 Classification of carboxypeptidases.

Catalytic type	MEROPS family	Name	Catalytic signature
Metallo	M2	ACE2	HEXXH…H
	M14 subfamily A	CPA1 CPA2 CPA3 CPA4 CPA5 CPA6 CPB CPU or TAFI CPO	HXXE…H
	M14 subfamily B	CPE CPD CPM CPN CPZ CPX1 CPX2 AEBP1	
	M14 subfamily C	Gamma-D-glutammyl-(L)-meso-diaminopimelate peptidase I	
	M14 subfamily D	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 RISC peptidase	Catalytic triad S/D/H
	S28	Lysosomal Pro-X carboxypeptidase	
Cysteine	C1 subfamily A	Cathepsin X	Catalytic diad C/H

Adapted from (Vendrell & Avilés 1999), (N.D. Rawlings et al. 2014) and (Petrera et al. 2014).

The largest family of carboxypeptidases is the M14 family, which belongs to the group metallo-type carboxypeptidases (Vendrell & Avilés 1999). Metallo-type carboxypeptidases are not homogenous regarding their zinc binding motif and they can be classified into different groups based upon their zing binding site (see **Table B.3**). The proteolytic enzymes focus of this thesis, the M14 family members, contain a HXXE...H binding motif, where the three ligands of the metal ion are a two histidines and a glutamate residue. Other families within metallo-type carboxypeptidases have also three metal binding residues but, with three histidines or two histidines plus and glutamate/aspartate. Interestingly, the metal binding sequence can include a general base/acid glutamate and/or an additional zinc metal ion, thereby containing five ligand residues. There are only two families of carboxypeptidases with two zinc ions in their catalytic site; the M20 and M28 (Rowsell et al. 1997; Mesters et al. 2006).

Serine-type carboxypeptidases are the second crowded group of carboxypeptidases. This group is formed by the families S10 and S28 which share the same catalytic motif composed by a triad of a serine, an aspartic acid and a histidine residue (Jung et al. 1998; Soisson et al. 2010). A difference, cysteine-type carboxypeptidases is the most reduced group and its members contain a catalytic dyad with a cysteine and a histidine in their catalytic site. The most representative member of this last group is cathepsin X, an enzyme involved in lysosomal degradation of proteins and peptides, cell signaling and tumor progression (Kos et al. 2015).

B.1.5 METALLOCARBOXYPEPTIDASES

The main focus of this thesis is the study of metallocarboxypeptidases (MCPs). Metallocarboxypeptidases are a large and diverse group of carboxypeptidases, all of which utilize a coordinated divalent metal ion to catalyze the hydrolysis of peptide bonds. To date, MCPs are distributed along six families of proteases according to the MEROPS database (see **Table B.3**). Despite, here we studied a subgroup within the M14 family and hereafter will be denoted as MCPs.

B.1.6 FAMILY M14 OF METALLOCARBOXYPEPTIDASES

Based on this binding motifs and structural relationships, the M14 family of MCPs can be divided into four subfamilies: M14A, M14B, M14C and M14D (Neil D Rawlings et al. 2014). Within each subfamily, the degree of amino acid sequence identity is 25-63%, but it decreases to only 15-25% when the two subfamilies are compared (Arolas et al. 2007). Furthermore, MCPs can be classified, according to its substrate specificity, into MCPs with preference for hydrophobic C-terminal amino acids (A-type), for basic C-terminal amino acids (B-type), for acidic C-terminal amino acids (O-type) and MCPs with broad substrate specificity (Lyons & Fricker 2011). The A-like enzymes can be further divided into A1-type or A2-type depending on its preference for both small aliphatic and bulky aromatic amino acids or with a strong preference for large aromatic residues, respectively (Gardell et al. 1988; Tanco et al. 2013).

B.1.6.1 Subfamily M14A

The subfamily M14A of MCPs (also described as A/B type or procarboxypepidases (PCPs)) is one of the best studied subfamilies. In mammals, contains nine members: CPA1, CPA2, CPA3 (or mast cell carboxypeptidase), CPA4, CPA5, CPA6, CPB, CPU (also known as thrombin-activatable fibrinolysis inhibitor or TAFI) and CPO. Among them, pancreatic carboxypeptidases CPA1, CPA2 and CPB were the first studied carboxypeptidases, which are produced and secreted by the exocrine pancreas to the digestive tract to function in the breakdown of proteins and peptides (generated by digestive endopeptidases).

All members of the M14A subfamily are structurally very uniform, since they appear to be produced as inactive precursors containing a preceding signal peptide of 15-22 residues, a N-terminal pro-domain of 90-95 residues (also known as pro-region) and a catalytic domain (CD) of 305-309 residues. The N-terminal region is usually called “activation segment” and folds in a globular independent unit. In pancreatic MCPs, this domain blocks the access to the active site of the enzyme in order to produce and store them as stable zymogens into pancreatic secretory granules. After secretion to the digestive tract, trypsin-promoted limited proteolysis takes place and generate the

active enzyme that participate in the degradation of dietary proteins (Arolas et al. 2007) (**Figure B.3**).

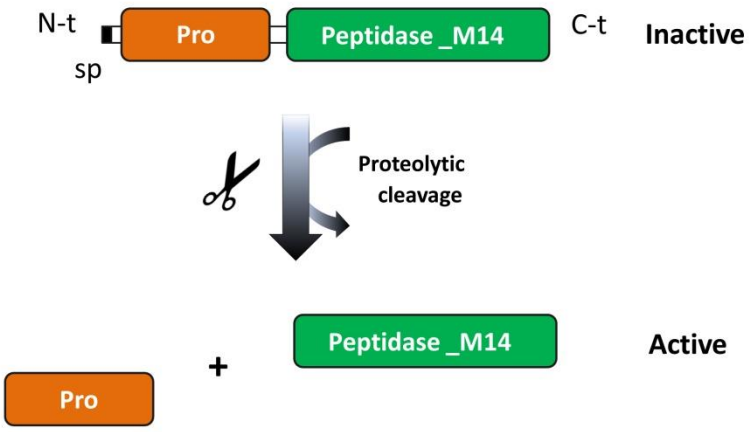


Figure B.3 Schematic representation of the structure and activation mechanism of M14 MCPs. Metallocarboxypeptidases within the subfamily M14A are produced typically as inactive precursors (zymogens) containing a N-terminal pro-domain (Pro, shown in orange) and the carboxypeptidase M14 catalytic domain (Peptidase_M14, shown in green). After secretion, the Pro-domain, also called "activation segment", is removed by limited proteolysis to obtain a fully active enzyme. The N-termini (N-t), C-termini (C-t) and the signal peptide sequence (sp) are indicated.

A similar activation mechanism is proposed for the rest of M14A MCPs, which are activated following secretion. Despite, there are three mammalian enzymes that scape to this general rule; CPA3, CPA6 and CPO. CPA3 is found in the secretory granules of mast cells, mainly in this active form, in complex with proteoglycans (Goldstein et al. 1989; Tanco et al. 2013). In a similar manner, CPA6 is activated by furin-like endoproteases in the secretory pathway and secreted as a fully active enzyme (Lyons et al. 2008).

Several studies have found that pro-domains present in M14A MCPs, besides the inactivation function, have an important contribution to the folding of the active domain during its synthesis (Phillips & Rutter 1996; Vendrell et al. 2000). CPO is a recently characterized member of the M14A subfamily of MCPs and is a unique metallocaroboxypeptidase within the M14A subfamily, since is translated and folded as an active enzyme without the need for a typical pro-domain. This intriguing enzyme

is a glycosylphosphatidylinositol-anchored intestinal peptidase with acidic amino acid specificity, which is probably involved in the cleavage of acidic amino acids from dietary proteins at the intestine and complements the actions of the pancreatic MCPs during food digestion (Lyons & Fricker 2011).

Although this subfamily was described originally as pancreatic/digestive carboxypeptidases involved primarily in general protein digestion, the majority of its members are involved in more selective processes. CPU (also known as CPB2, CPR, and plasma CPB) is synthetized and secreted by the liver as the inactive precursor named as PCPU. This circulating precursor form is also commonly referred to as thrombin-activatable fibrinolysis inhibitor or TAFI. Upon activation by thrombin alone or by the thrombin-thrombomodulin complex, the active form (TAFIa) cleaves C-terminal lysine residues from the fibrin surface, thereby decreasing its cofactor activity and regulating the rate of fibrinolysis. Due to its function on the edge of fibrinolysis and coagulation, TAFI raised the interest of the pharmacological industry to be used as drug target for prevention of thrombotic diseases (Arolas et al. 2007; Foley et al. 2013). CPA3 is stored in the secretory granules of mast cells, which are important effectors of the immune system. The exact role of this enzyme is not clear yet, although it might function in the regulation of the inflammatory response and/or protection against exotoxins (e.g. venom sarafotoxins) (Wernersson & Pejler 2014). CPA4 is widely expressed in tissues and its expression is greatly upregulated in some types of carcinomas. Furthermore, CPA4 has been proposed to participate in the inactivation of peptides that function in cell proliferation and in aggressiveness of prostate cancer (Sebastian Tanco et al. 2010) . CPA5 cleaves aliphatic C-terminal residues and is present in discrete regions of rodent pituitary and other tissues. Although, additional research efforts are needed to elucidates the biological effects of this unexplored enzyme (Wei et al. 2002). CPA6 is activated in the secretory pathway. After secretion, CPA6 is bound to the extracellular matrix where retains its enzymatic activity. CPA6 mutations were found to be associated with epilepsy and Duane syndrome (a congenital eye-movement disorder). However, further studies must be carried out to test if loss of CPA6 is sufficient to cause these pathologies (Lyons & Fricker 2010; Sapio & D. 2014).

B.1.6.2 Subfamily M14B

The subfamily M14B of MCPs (also named as N/E subfamily or regulatory carboxypeptidases) is composed of eight mammalian members (**Figure B.4**).

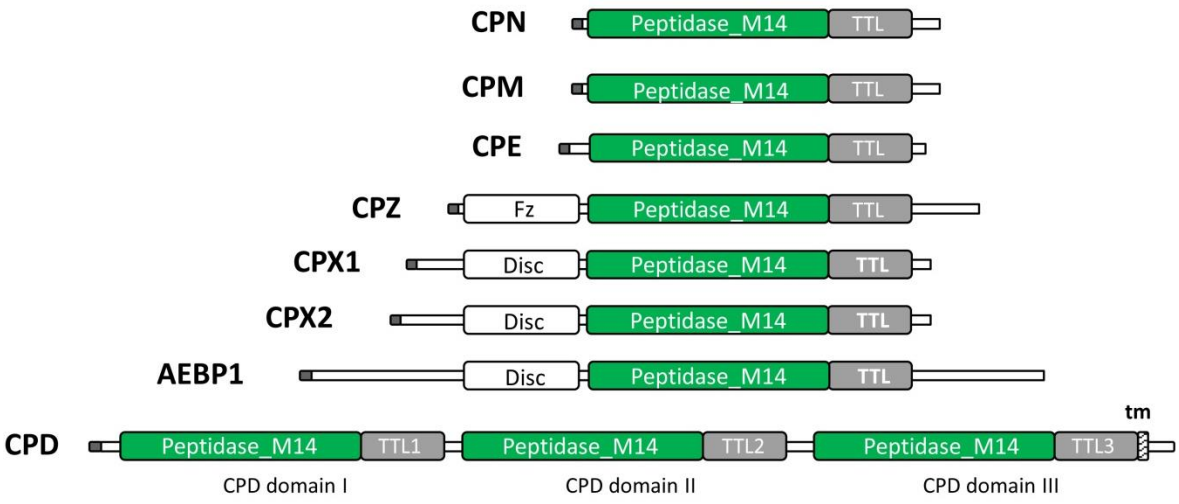


Figure B.4 Comparison of the domain structures of M14B subfamily MCPs members present in humans. All members of the M14B subfamily of MCPs contain an N-terminal signal peptide, a carboxypeptidase M14 catalytic domain (domain (Peptidase_M14, shown in green), an additional C-terminal domain with homology to transthyretin (TTL, shown in grey), and then additional sequences at the N- or C- termini (e.g. the frizzled-like domain present in CPZ (Fz) or the additional domains with amino acid similarity to discoidin-1 found in CPX1, CPX2 and AEBP1 (Disc)). Unlike the rest of MCPs, carboxypeptidase D contains three tandem repeats of the M14B catalytic unit (each one of them containing both the peptidase_M14 domain plus the TTL domain, named as CPD domains I, II and III) followed by a short transmembrane domain (tm) and a cytoplasmic tail.

Five members (CPE, CPN, CPM, CPD and CPZ) are catalytically active proteins with an strict B-like substrate specificity for cleaving only C-terminal basic residues. They are produced and secreted as constitutively active enzymes, without a pro-domain to regulate their activity. Nonetheless, some members within the M14B subfamily are catalytically inactive (AEBP1, CPX1 and CPX2), since they lack some relevant amino acids essential for the catalytic mechanism. The function of these inactive members remains unknown: either might cleave other substrates or act as binding proteins (Reznik & Fricker 2001; Arolas et al. 2007).

This subfamily of MCPs is structurally heterogeneous, in comparison with the M14A subfamily (see **Figure B.4**). Despite, all of them share a common structure with a carboxypeptidase M14 catalytic domain and a C-terminal domain of about 80 residues with structural homology to transthyretin (known as transthyretin-like domain or TTL). The role of TTL domains is yet unknown. Due to its similarity with pro-domains found in the M14B members, it has been proposed to function as folding domain or, alternatively, was proposed a probable function as binding element necessary for oligomerization and/or binding of the enzyme to other proteins or membranes (Reznik & Fricker 2001). Besides, all members of this subfamily contain N- and/or C- terminal domains of a variable length. In the case of CPN, CPM and CPE this additional domains are small. In CPZ, this domain is a large N-terminal sequence homologe to the frizzled receptors and other wnt-binding proteins (termed as frizzled-like domain). Likewise, CPX1, CPX2 and AEBP1 contain additional domains in the n-terminal region with amino acid similarity to discoidin-1 (named as disc domains). A difference, CPD contains three catalytic domains in the same polypeptide chain and a transmembrane domain, as well as a cytosolic tail of 60 residues long.

Of the five active members of the M14B subfamily, CPE and CPN were the first to be discovered and are the best characterized as neuropeptide processing enzymes. CPE is involved in the biosynthesis of neuropeptides and growth factors by removing C-terminal basic residues from peptide processing intermediates. This step occurs within the secretory pathway of neuroendocrine cells, following the action of a type of endoproteases known as pro-hormone convertases (Xin Zhang et al. 2008; Sapio & D. 2014). CPN and CPM have also been proposed to function in the processing of peptide hormones, but in contrast to CPE, these other carboxypeptidases perform its functions in the extracellular space. CPN is produced by the liver and circulates in plasma as a 280 kDa heterotetrameric complex that includes two catalytic and two regulatory/noncatalytic subunits of 48-55 kDa and of 83 kDa each, respetively. CPN has been proposed to removes C-terminal Arg amino acids from a variety of neuropeptides including bradykinin, kallidin and from anaphylotoxins C3a, C4a and C5a (Skidgel & Erdos 2007). CPM is mainly found in lung and placenta as a membrane-bound glycosyl-phosphatidylinositol-linked protein. In addition, soluble forms of CPM are found in

various body fluids. Although the function/s of this enzyme are not fully understood, it has been proposed that CPM might have a possible function in inflammation and, besides its peptidase activity, it could be act as a binding partner in cell-surface protein-protein interactions (Deiteren et al. 2009).

CPD has the broadest tissue distribution of all MCPs of both subfamilies M14A and m14B, being present in all tissues. This enzyme with three catalytic domains is located mainly in the trans-Golgi network and immature secretory vesicles (but not in mature vesicles) and in the plasma membrane. While the first two catalytic domains I and II are active enzymes only differing slightly in their enzymatic properties, the domain III correspond to an inactive MCP, more similar to CPX1/2 and AEBP1 in that it lacks the residues considered critical for the catalytic mechanism. Due to its localization and substrate specificity, Reznik and Fricker (Reznik & Fricker 2001) proposed that CPD primarily might function following the action of the trans-Golgi network endopeptidases, therefore likely to be involved in the production of receptors and growth factors which are processed by furin and related enzymes within the trans-Golgi network. Despite, its function remains unclear.

A difference with CPE and CPN, CPZ is one of the less-studied members of this subfamily. This enzyme is produced dynamically during embryonic development and is secreted to the extracellular space, where bounds to the ECM due to its heparin-affinity properties. Although the exact function of CPZ is not yet known, it is likely that this protein can play a role in development by removing C-terminal amino acids from neuropeptides and growth factors and/or through its interaction with Wnt proteins (Reznik & Fricker 2001).

B.1.6.3 Subfamily M14C

The M14C subfamily is composed by only one known member, the bacterial enzyme named γ-D-glutamyl-(L)-meso-diaminopimelate peptidase I. In terms of substrate specificity, this enzyme is unique since cleave γ-D-glutamil bonds to the L-terminus of meso-diapimelic acid during bacterial sporulation. Further work will be done in order to gain some understanding of the molecular mechanisms of this enzyme.

B.1.6.4 Subfamily M14D

The subfamily M14D comprises a newly discovered subfamily of MCPs, also known as cytosolic carboxypeptidases or CCPs. This subfamily was described eight years ago simultaneously by two research groups (Rodriguez de la Vega et al. 2007a; Kalinina et al. 2007), which provided phylogenetic studies supporting the divergence of these enzymes from the rest of M14 subfamilies.

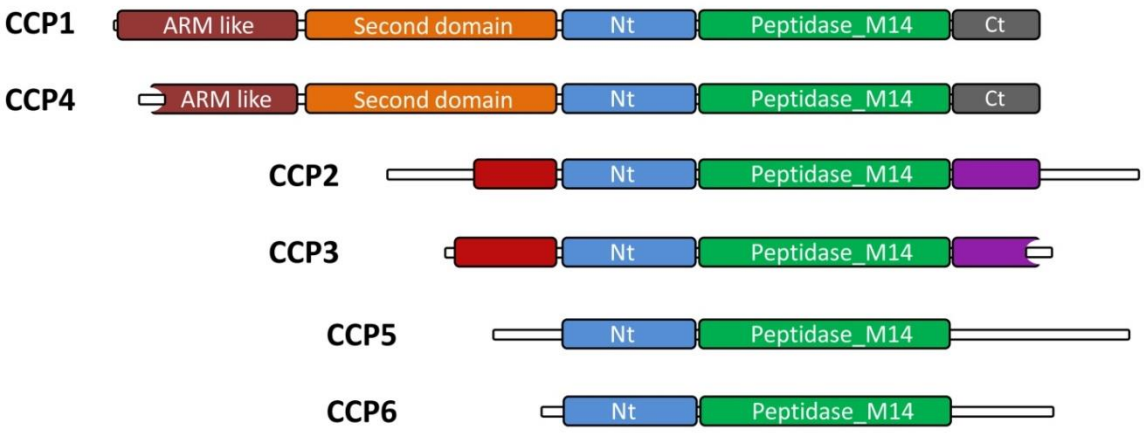


Figure B.5 Schematic representation of the domain structures of M14D subfamily members identified in humans. All members of the M14 subfamily of MCPs contain a carboxypeptidase M14 catalytic domain (Peptidase_M14, shown in green) and an N-terminal domain (Nt, shown in blue). Furthermore, all members have additional domains in the N-terminal and/or C-terminal region.

A difference with other M14 subfamilies, all of these enzymes show cytosolic/nuclear localization. In humans the M14D subfamily is composed by six members, which share a common catalytic domain and an N-terminal conserved domain of 150 residues with a new β-sandwich fold, known as N-terminal domain (Otero et al. 2012), as well as other additional N- or C-terminal domains (**Figure B.5**).

Although they are highly variable in terms of substrate specificity, all of them have similar O-like substrate specificities. The functions of all of them have been recently studied in detail (Rogowski et al. 2010; Berezniuk et al. 2012; Berezniuk et al. 2013; Tort et al. 2014). It has suggested that these enzymes remove glutamate residues from

the C-terminus of posttranslationally added polyglutamate side-chains in tubulin, as well as in other polyglutamilated proteins (Tanco et al. 2015)

B.1.7 STRUCTURE OF METALLOCARBOXYPEPTIDASES

B.1.7.1 Structure of the catalytic domain

Analysis of the increasing number of available three-dimensional (3-D) structures for the main subfamilies of MCPs shows that these enzymes share a common catalytic domain with a similar conformational fold. All of them also conserve a similar architecture at the active site, showing identical catalytic residues according to its substrate specificity (see **Figure B.6**).

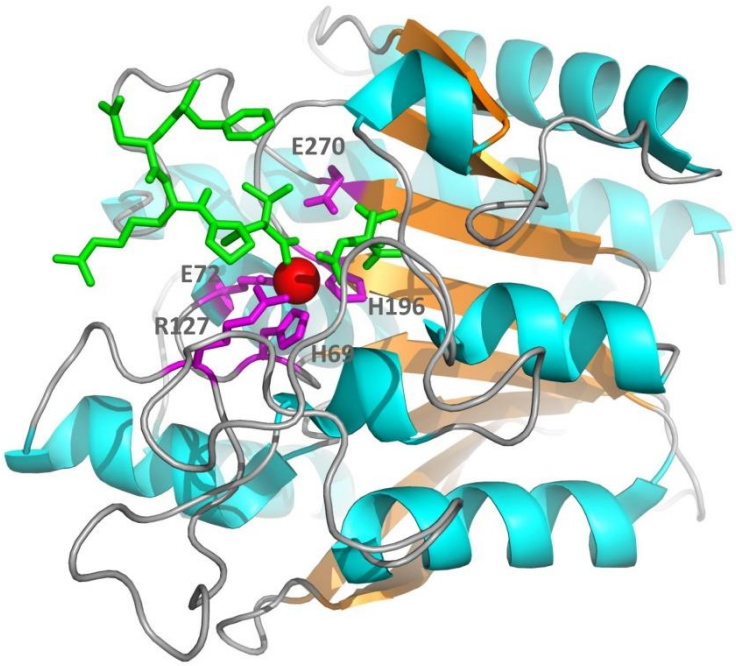


Figure B.6 Representative structure of the catalytic domain and active site residues of MCPs. Ribbon representation of the three-dimensional structure of the catalytic domain of hCPA4 in complex with a cleaved hexapeptide (PDB 2PCU). The α-helices and β-strands are shown in cyan and orange, respectively. The cleaved hexapeptide is shown as a green stick model. The sides chains of the main residues involved in zinc binding (H69, E72 and H196) and in the catalytic mechanism (E270 and R127) are depicted in magenta. The metal Zn^{2+} ion is shown as a red sphere. Image generated with PyMOL (DeLano 2002).

The catalytic domain of MCPs corresponds to a α/β hydrolase fold, formed by a central eight stranded β-sheet (β1- β8) with a twist of 120º between the first and the last strand, over which eight α-helices pack on both sides to form a globular protein which have a compact globular shape that, on its overall structure, resembles the volume obtained when a cone is extracted from a sphere (**Figure B.7-A**). They have a funnel-like opening at the top, and for this, Gomis-Rüth (Gomis-Rüth 2008) et al proposed the term funnelins to describe this family of proteases (**Figure B.7-B** and **C**).

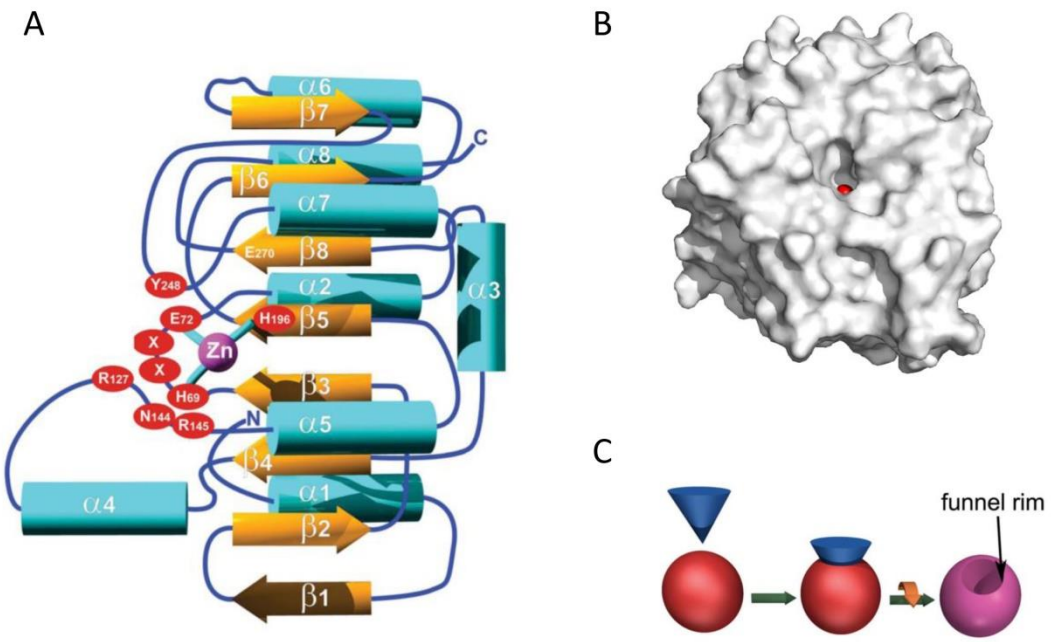


Figure B.7 Topology scheme and overall structure of funnelins. (A) Topology scheme illustrating the consensus secondary-structure elements (α-helices in cyan and β-strands in orane) shown in all MCPs, taking as example the hCPA structure. The highlighted amino acids are contained in the characteristic set of conserved residues, HXXE+R+NR+H+Y+E. Residues are numbered according to the bCPA1 numeration. The metal ion is shown as a magenta sphere. (B) Surface representation of hCPA4 showing the funnel-like aperture that leads to the catalytic zinc cation at the bottom (red sphere). (C) Schematic representation showing the conceptual model of funnelins proposed to describe MCPs, which illustrates the generation of a major (or inverted) spherical cone (magenta) through the intersection of a regular cone (blue) and a sphere (red). Figure adapted from (Gomis-Rüth 2008).

The β-strands of the catalytic domain have connectivity +1, +2, -1x, +2x, +2, +1x, -2. The core of the sheet is composed by four parallel copalanar central strands β3, β5 and β8. The catalytic site is located at the C-terminal end of these strands, which are

flanked by two parallel strands (β6- β7) and a β-ribbon (β1- β2). The concave front side of the sheet accommodates the helices α5 and α7, while α1- α4, α6 and α8 helices and the surface N- and C-termini of the molecule are located at the convex site of the sheets. The catalytic site access is shaped by a series of irregular segments of different length. This segments are Lβ8α8, Lβ5β6, Lβ7α7, Lβ3α2 and in particular, the loop Lα4α5 that contains a disulfide bridge between Cys138 and Cys161, according to bCPA1 reference, found in most M14A members (Arolas et al. 2007; Gomis-Rüth 2008) (**Figure B.7-A**).

Crystal structures of fruit fly CPD domain I, duck CPD domain II and human CPM showed that the major differences between the catalytic domain of M14B and M14A members are within the length of the loops between the β-sheets and α-helices that further close the aperture to the catalytic site, and limited the access of the typical M14A proteinaceous inhibitors (Aloy et al. 2001; Reverter et al. 2004; Keil et al. 2007).

The catalytic zinc cation is located at the bottom of the funnel where is coordinated with two histidine residues (H69 Nδ1, and H196 Nδ1), one glutamate residue (E72 atoms Oε2 and O ε1) and a water molecule in a mostly asymmetric, bidentate manner (Gomis-Rüth 2008). This triad of residues constitutes the characteristic consensus motif for M14 MCPs (**Figure B.6**).

B.1.7.2 Structure of the pro-domain

All members of the M14A subfamily of MCPs, with the exception of CPO, contain a pro-domain with a similar structure. The structure of this domain consist of 80 residues that form a globular part composed by an open sandwich with a twisted, four stranded antiparallel β-sheet with two antiparallel α-helices on top of them, defining a hydrophobic core (García-Castellanos et al. 2005; Gomis-Rüth 2008). This globular part is connected to the N-terminus of the enzyme moiety through a C-terminal connecting segment formed by a large α-helix that includes the activation scissile bond (**Figure B.8**).

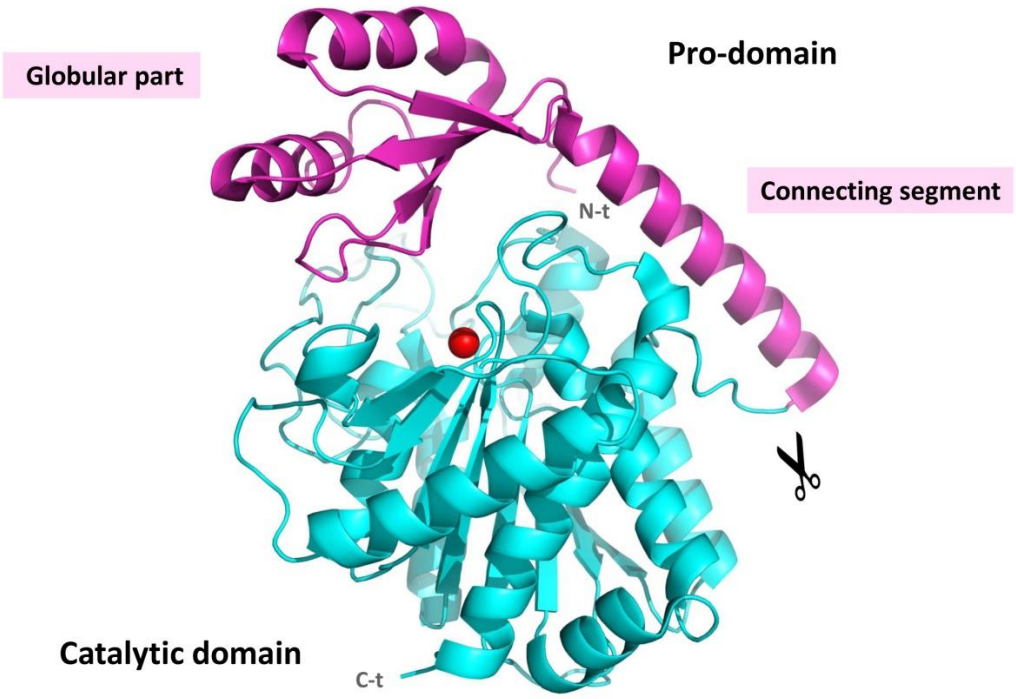


Figure B.8 Structure of hCP4 on its zymogenic form. Ribbon representation of hCPA (PDB 2BOA) with the Pro-domain (shown in magenta) and the mature enzyme moiety (shown in cyan). The pro-domain is composed by two different regions, a globular part and the connecting segment which includes the activation site (denoted with scissors). The metal ion is shown as a red sphere and N- and C-termini are indicated as N-t and C-t, respectively. Image generated with PyMOL (DeLano 2002).

The pro-domain can fold independently from the catalytic moiety, suggesting that might act as a chaperone *in vivo*. The interaction of this domain with the catalytic domain takes place at the side of the β-sheet, between β2 and β3, and the proximal residues Arg71 and Phe279/Tyr198 of the catalytic moiety. The interaction between the pro-domain and the catalytic domain blocks the access to the active site cleft and accounts for the almost completely loss of activity in the majority of the zymogenic forms of M14 MCPs, acting similarly to as non-competitive inhibitor (García-Castellanos et al. 2005). Although this is a general rule, it has also demonstrated that the zymogen form of TAFI displays continuous and stable carboxypeptidase activity against large peptides substrates (Valnickova et al. 2007).

B.1.7.3 Structure of the Transthyretin-like domain

All members of the M14B subfamily of MCPs share similar domain architecture with a typical M14 catalytic domain and an additional C-terminal domain related with transthyretin, known as TTL domains. The first three-dimensional structure of a MCPs TTL domain was solved in 1999 in complex with its catalytic moiety (Gomis-Rüth et al. 1999). This study revealed that this C-terminal domain shows a rod-like shape with dimensions of 25 x 25 x 40 Å and is formed by a β-sandwich containing two layers of three mixed strands and four antiparallel strands, which are held together by a hydrophobic core. This domain is connected to the catalytic part via helix I (connecting segment), of the later and is connected to the N-edge of the central β-sheet (**Figure B.9**).

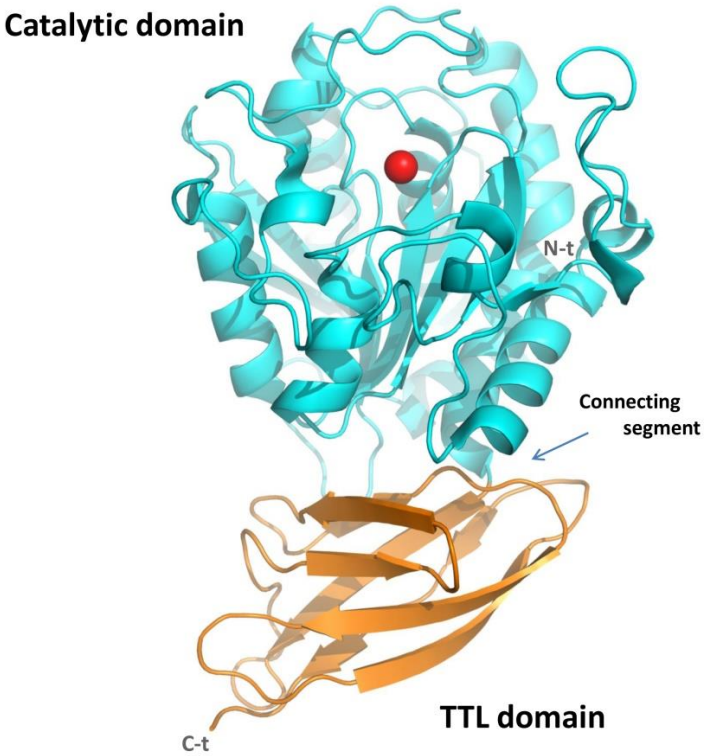


Figure B.9 Structure of the duck CPD domain II. Ribbon representation of the domain II of dCPD (PDB 1H8L), showing the location of the catalytic domain on top (shown in cyan) and the transthyretin-like domain (TTL domain) at the bottom (shown in orange). The metal ion is shown as a red sphere. The protein C- and N- termini are indicated (C-t and N-t, respectivley). Image generated with PyMOL (DeLano 2002).

The interactions between the TTL domain and the catalytic moiety are stablished mainly through hydrophobic interactions. Furthermore, at least one salt bridge was observed between the residues Asp216 of the catalytic domain and Arg343 within the TTL domain. Additionally, about 48 van der Waals interactions and 8 hydrogen bonds are stablished to maintain the structure.

B.1.8 CATALYTIC MECHANISM OF METALLOCARBOXYPEPTIDASES

B.1.8.1 Catalytic mechanism

The three-dimensional structure of bovine CPA (bCPA) was one of the first MCPs structures to be reported (Lipscomb et al. 1969) and is one of the most thoroughly enzymes studied in biochemistry. For this reason, the numbering system widely used corresponds to that of the archetypal bovine CPA and it will be used hereafter for the description of all key residues involved in substrate binding and catalysis. In MCPs, the active site is commonly located in a grove on the surface of the enzyme. The active site contains four groups of residues: (1) the three protein ligands of the catalytic zinc ion (His69, E72 and H196); (2) residues involved directly in the catalytic mechanism (Arg127 and Glu270); (3) residues involved in the anchoring and neutralization of the COOH group of the C-terminal residue of the substrate (Asn144 and Arg145), (3) residues responsible to shape the dead end pocket (shown as R in **Figure B.10**), which is complementary to the C-terminal side chain of the substrate and define the S1' subsite (**Figure B.10**).

Although residues involved in the catalytic mechanism have been extensively characterized, the catalytic mechanism of MCPs still remains controversial. Two mayor mechanisms have been proposed: the promoted-water mechanism and the nucleophilic mechanism (Breslow & Wernick 1977; Alvarez-Santos et al. 1987; Gomis-Rüth 2008). Even though, the most widely accepted is the promoted-water mechanism, in which a polarized zinc-bound water molecule (due to the attack of the zinc ion and the deprotonated carboxylate group of Glu270) acts as a $Zn\text{-}OH^-$ nucleophile attacking the carbonyl carbon of the scissile peptide bond of the substrate, leading a tetrahedral intermediate/transition state.

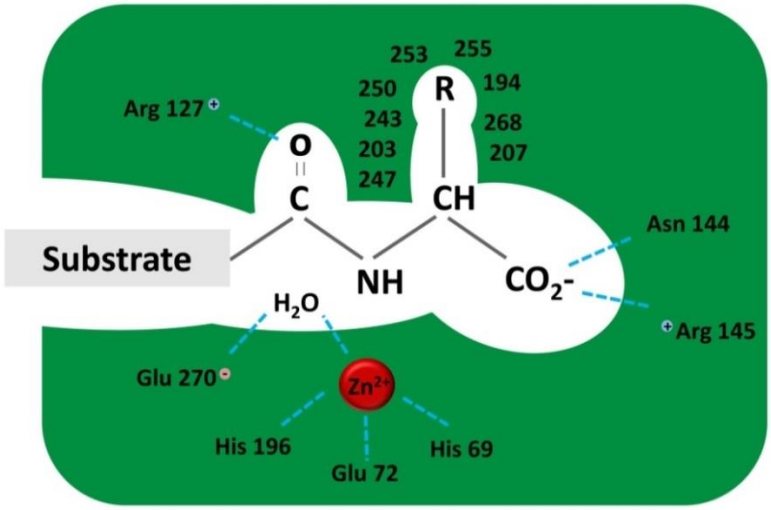


Figure B.10 Active site and major substrate-binding residues of A-like MCPs. Schematic representation of the metal-binding and substrate binding residues with shape the S1' pocket in the active site of metallocarboxypeptidases of the M14 family. Residues are numbered according to the bCPA1 numeration. Adapted from (Sebastian Tanco et al. 2010)

A hydrogen bond of the carbonyl oxygen to the guanidinium moiety of Arg127 polarized the carbonyl peptide bond prior to hydrolysis, and stabilizes the transition state of the tetrahedral intermediate together with the zinc ion and the Glu270. The tetrahedral intermediate subsequently collapses into products upon protonation of the amide nitrogen, left by the side chain of Glu270 and restore the catalytic machinery for a subsequent catalytic process. Thus, the carboxylate group of Glu270 serves as a general acid/general base in this mechanism.

B.1.8.2 Structural determinants of the substrate specificity

Several crystallographic studies have delimited the residues involved in substrate binding (García-Sáez et al. 1997; Gomis-Rüth et al. 1999; Guasch et al. 1992; Bayés et al. 2007; Arolas et al. 2007; Sebastian Tanco et al. 2010). However, the bCPA structure is still the most extensively used model to define the substrate specificity S1', S1, S2, S3 and S4 subsites (see section B.1.3). These subsites are located in the surface of the enzyme and accommodate the side chains of the C-terminal residues of the substrates (**Table B.4**). Among these subsites, the S1' and S1 are the most important to define the substrate specificity in the great majority of MCPs.

The S1' subsite accommodates the side chain of the C-terminal residue. This subsite is delimited in M14A MCPs by the side chains of residues located in positions 194, 203, 207, 243, 247, 250, 253, 254, 255 and 268 (**Figure B.10 and Table B.4**). The difference in specificity between CPA-like and CPB-like MCPs is primarily attributed to the Ser255 residue (Arolas et al. 2007; Sebastian Tanco et al. 2010). In CPB this position is occupied by a Asp amino acid. Moreover, Ile243 is replaced by a Gly243 in order to promote a polar environment. In CPA2, Ser194, Leu203 and Thr268 are changed by Ile, Met and Ala residues, respectively, facilitating the binding of bulky amino acids. By contrast, CPO contains an Ile amino acid in a homologous position, which leads to a substrate preference for acidic amino acids.

Table B.4 Summary of amino acid residues involved in the substrate specificity determination.

Subsite	Residues involved
S1'	Ser194, Leu203, Gly207, Ile243, Ile247, Ala250, Gly253, Ser254, Ile255, and Thr268
S1	Arg127, Tyr198, Ser199, Ile247, Tyr248, Glu270 and Phe279
S2	Arg71, Arg127, Asp142, Ser197, Tyr198 and Ser199
S3	Phe279
S4	Glu122, Arg124 and Lys128

According to (Arolas et al. 2007; Sebastian Tanco et al. 2010). Numbering according to bCPA reference

For M14B MCPs some of the important substrate-binding residues are different from those found in M14A MCPs. The most important difference is found in the residue in position 255 located at the bottom of the S1' binding pocket. This position is occupied by a Gln or Ser residue in CPE, CPD, CPN and CPM. Other difference is located a position equivalent to Ser207. This position is occupied by an Asp amino acid in the active members of the M14B subfamily and is a key determinant for its substrate preference for C-terminal basic residues (Arolas et al. 2007). The S1 subsite is shaped mainly by residues Tyr198 and Ser199. Additionally, other residues, such as Arg127, Val247, Tyr248, Phe279 and Glu270 also contribute to the shape and substrate specificity of this subsite (Arolas et al. 2007; Sebastian Tanco et al. 2010) (**Table B.4**).

B.2 PROTEIN FOLDING AND AGGREGATION

B.2.1 PROTEIN FOLDING

Protein folding can be defined as the process by which a polypeptide chain acquires its three dimensional native structure. Only correctly folded proteins have long-term stability in crowded biological environments and are able to interact selectively with their natural partners. The native conformation is constituted by a large number of dynamic states, which display different but energetically proximal minimums. The fluctuations between these species provide flexibility, indispensable for protein functional activity under physiological conditions. Uncovering the mechanisms through which such processes take place is one of the grand challenges of modern science, since the failure of proteins to fold correctly, or to remain correctly folded, is the origin of a wide variety of pathological conditions. Although the mechanism by how polypeptide chains self-assemble into highly structured states is currently better understood thanks to the advance of physical and chemical techniques, as well as computational methods, still remains shrouded in mystery (Dobson 2003; Chiti & Dobson 2006).

B.2.1.1 Basis of protein folding: Form the Anfinsen's postulate to the "New view"

In the early 1960s, Anfinsen and co-workers performed the first pioneering experiments of protein folding. His work was based on the *in vitro* study of the folding mechanism of ribonuclease A (Anfinsen et al. 1961). With this work, Anfinsen demonstrated that the amino acid sequence suffices to encode its three-dimensional structure and he concluded that driving force that guides protein folding is the search for the minimum free energy that corresponds with the native state. Based on *in vitro* experiments, he proposed that the folding process starts spontaneously after the newly synthetized amino acidic chain leave the ribosome.

Almost a decade later, Levinthal published a paradox based on a simple calculation, illustrating how the acquisition of a particular native structure by random search would take longer than the age of the universe, considering the extremely large

number of possible conformations. Thus, he concluded that protein folding must follow preferential folding pathways, adopting a limited number of conformations before reaching the native state (Levinthal 1968). This finding leaded the search of folding pathways in the forthcoming years. Several studies were performed to circumvent the Levinthal paradox, since it was envisaged that proteins could fold by defined pathways and mechanisms that removed the need to search all possible conformations. From such works different classical models have been proposed:

The sequential protein folding model, also known as framework or hierarchic model, was proposed in 1973 by Ptitsyn. This model suggests that secondary structures are the first formed elements, and the native form may be achieved then by docking of the secondary structures (Ptitsyn 1973). The diffusion-collision model was suggested in 1976 by karplus and Weaver. This model postulates the formation of secondary structures, followed by their diffusion, collision and coalescence to form tertiary structures (Karplus & Weaver 1976). The hydrophobic-collapse model is a hypothesis proposed to explain protein folding based on the observation that proteins usually contains a hydrophobic core in the protein interior. This model proposed in 1984 by Go, suggest that the initial steps in folding involve hydrophobic collapse. Acquisition of secondary structure and the correct packing interactions are then formed in a confined volume (Go 1984). The nucleation-condensation or nucleation-collapse model is a mix of the two previous models (hydrophobic collapse and framework mechanism). This model suggests that the folding starts with the formation of a more diffused nucleus, in comparison with the classical model in which a strong localized nucleus is formed (Wetlaufer 1973; Fersht 1997).

The field of protein folding has seen tremendous advances over the past 20 years. These technical advances have allowed the postulation of new models. The energy landscape theory was first proposed by Bryngelson and Onuchic 20 years ago (Bryngelson et al. 1995; Onuchic et al. 1997). This theory proposes that proteins do not fold following a restricted mechanism, instead postulate that folding is produced due to a stochastic search of the many conformations accessible to a specific polypeptide chain. This landscape describes the dependence of the free energy on all the coordinates determining the protein conformation; where the y-axis of the landscape

represents the internal free energy of a given polypeptide configuration whereas lateral axes represent the conformational coordinates. Internal free energy includes the energies of the three main forces that lead to the attainment of the secondary and tertiary structure: hydrogen bonds, electrostatic interactions and hydrophobic forces, as well as the torsion angle energy and the solvation free energy (**Figure B.11**) (Ahluwalia et al. 2013).

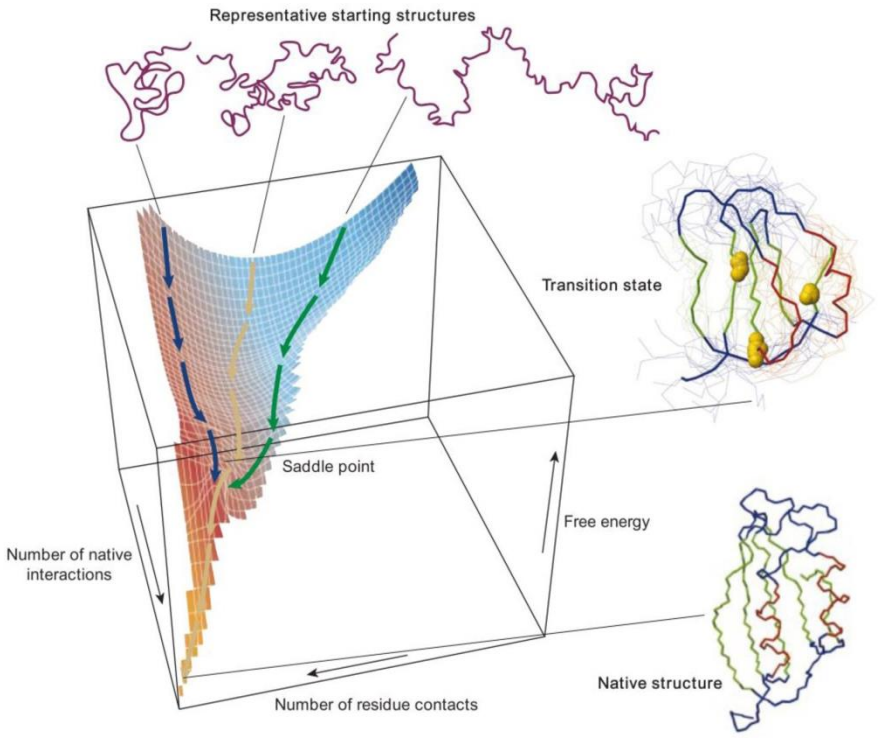


Figure B.11 Schematic representation of the energy landscape for protein folding. During protein folding, an ensemble of partially folded intermediates is formed. In the diagram the multiple computer-simulated denatured conformations are funnelled towards the acquisition of the native structure. Simplified trajectories for the folding of individual molecules are indicated with arrows.The transition state is the barrier that all proteins must cross if they are to fold to the native state. Adapted from (Dobson 2003).

The energy landscapes depend on the polypeptide sequence and environmental factors. For small proteins, different folding pathways are possible along the funnel with local minima within the funnel that give rise to folding intermediates. These intermediates can be on the pathway of the native state or also be off-pathways acting as folding traps (Dobson 2003).

B.2.2 PROTEIN MISFOLDING, AGGREGATION AND AMYLOID FORMATION

B.2.2.1 Protein misfolding

Misfolded proteins result when a protein follows the wrong folding pathway or energy-minimizing funnel. Mostly, only the native conformations are produced by the cell machinery. But, sometimes failures in the acquisition or maintenance of the native structure occur, and have important consequences, in many cases leading to cellular toxicity. These pathological conditions are generally named protein misfolding diseases or conformational diseases (Chiti & Dobson 2006).

Partially folded species expose to the solvent regions that are buried in their native states, which allow the establishment of non-native intermolecular interactions leading to protein deposition. Under normal circumstances, the cell has quality control mechanisms to prevent proteins from folding incorrectly, as well as to get rid of misfolded proteins. This mechanism involves the mediation of chaperones to assist protein folding or degradation through the ubiquitin-proteasome pathway (Figure B.12). Failures of these cellular machineries can lead to catastrophic effects for living organisms, such as the impairment of cell function or eventually the kill of the cell.

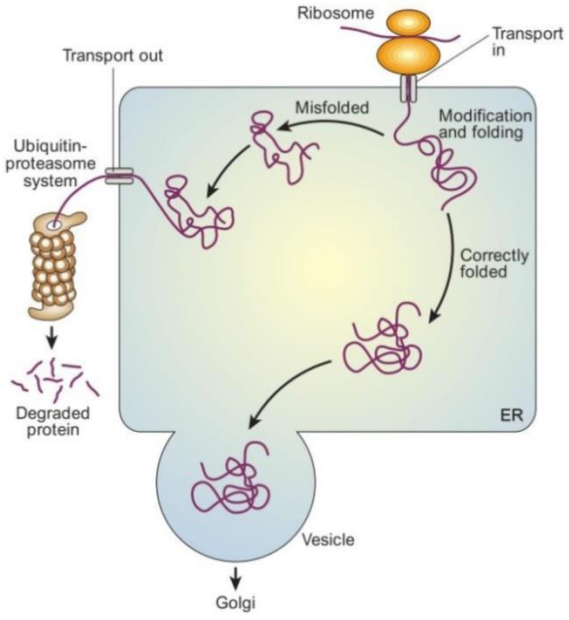


Figure B.12 Mechanism of regulation of protein folding. Molecular chaperones help proteins to fold properly after its synthesis by ribosomes in the Endoplasmic Reticulum (ER). Incorrect folded proteins are detected and transported to the cytoplasmic proteasome machinery to be degraded. Adapted from (Chiti & Dobson 2006)

B.2.2.2 Protein aggregation and amyloid formation

The Failures of the quality control mechanisms that redress protein misfolding in cells often lead to protein aggregation, protein deposition and disease. Aggregating proteins and peptides involved in conformational diseases display different structural states in solution, ranging from globular to totally unstructured conformations. However, the aggregated forms have many characteristics in common, such as birefringence or affinity for some specific dyes, and have striking similarities on its aggregation behavior. Usually, unfolded or partially unfolded proteins self-assemble to form small, soluble aggregates (or amyloids) that undergo further assembly into protofibrils or protofilaments and, subsequently into the mature fibrils (see **Figure B.13**) in a process known as amyloid formation (Chiti & Dobson 2006).

To date, more than forty human conformational diseases are related with the formation of intracellular and/or extracellular amyloid-like aggregates. Nonetheless, the conversion of some polypeptide chains into fibrillar species not only has deleterious consequences for cells. Several studies have reported that this mechanism could be further exploited by living organisms for functional purposes (Claessen et al. 2003; Barnhart & Chapman 2010; True & Lindquist 2000; Coustou et al. 1997) .

To start the amyloid formation process in a two-state folding protein model, the native ensemble has to cross energy barriers to be converted into non-native species. The most important of these barriers is the unfolding activation barrier (k_U), which is a rate-limiting step in amyloidegenic diseases. This aggregation process can be explained with the following equation 1, derived from the Lumry-Eyring model (Sanchez-Ruiz 1992).

Equation 1:

$$F \underset{k_F}{\overset{k_U}{\longleftrightarrow}} U \overset{k_A}{\longleftrightarrow} A$$

Where k_F, k_U and k_A are the folding, unfolding and aggregation rates.

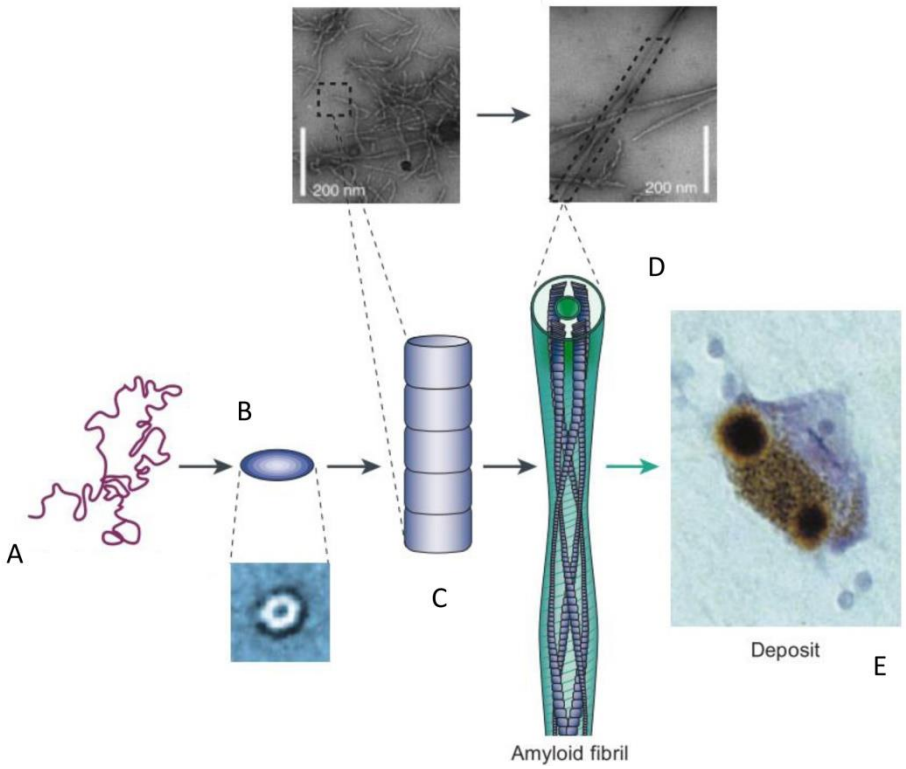


Figure B.13 Scheme of the general mechanism of protein aggregation and amyloid formation. Unfolded or partially folded protein species (A) self-associate to form soluble oligomers (B) that further assemble protofibrils (or protofilaments) (C). Finally, this protofilaments form mature amyloid fibrils (D) that often are accumulated inside or outside (E, deposit) the cells, leading to cell toxicity. Adapted from (Chiti & Dobson 2006).

Typically, the aggregation process follows a kinetic mechanism typical of nucleated processes such as crystallization (**Figure B.14**), in which the Initial structurally diverse precursors promote the slowly nucleation process in a concentration-dependent manner. In this step known as lag phase, protein monomers must transform their conformation into β-sheet structures to achieve a critical size. In this process thermodynamically unfavorable, the lag phase can be eliminated by the addition of preformed aggregates to fresh solutions (known as seeding). After the achievement of a critical number of oligomer species that form the aggregation nucleus, monomers are incorporated rapidly and efficiently during the fibril elongation process. This thermodynamically favorable process is called growth fase and takes place as a single-exponential curve.

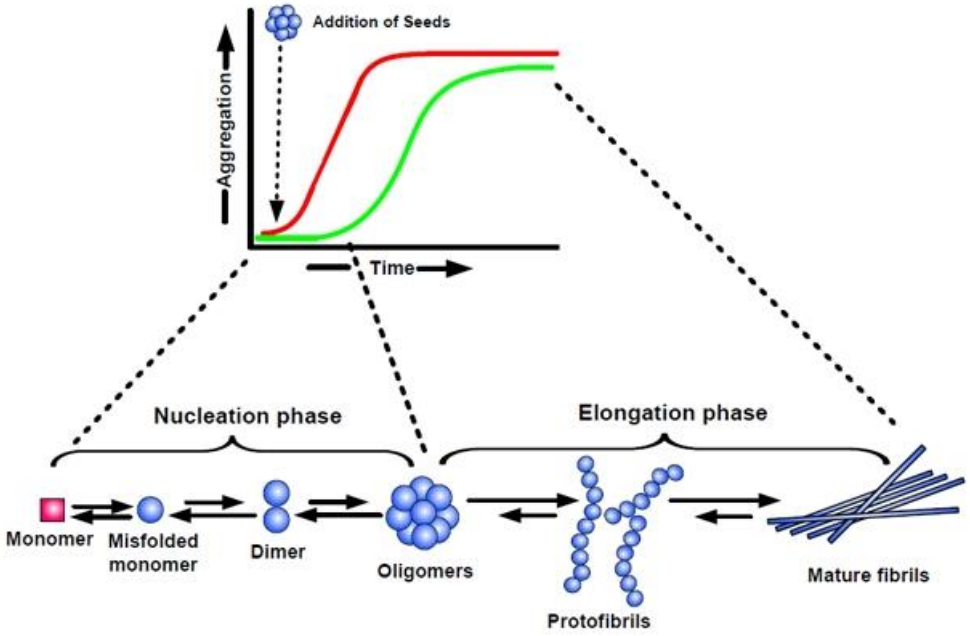


Figure B.14 Scheme of the nucleation-dependent model of amyloid aggregation. The amyloid formation kinetics contains two differentiated steps, the lag (or nucleation) phase and the growth (or elongation) phase. In the first phase, monomers undergo structural changes and unfolding, followed by its association into the oligomeric nuclei. The second phase is thermodynamically more favourable and is characterized by a rapid growth. Thus, the kinetics of the amyloid formation can be well represented by a sigmoidal curve with a lag phase followed by a growth phase (green curve). The addition of preformed seeds reduces the lag time and accelerates the aggregate formation. Adapted from (Kumar & Walter 2011)

B.2.2 AMYLOID FORMATION UNDER NATIVE-LIKE CONDITIONS

Recent *in vitro* studies have been demonstrated that a limited number of globular proteins have the ability to undergo amyloid fibril formation under solution conditions that promote their partial unfolding (Chiti & Dobson 2009b). Examples include Sso AcP acylphosphatase-like protein, lysozyme, superoxide dismutase 1, transthyretin, immunoglobulin light chains, β2-microglobulin, various forms of ataxins and prolactin (Zhuravlev et al. 2014; Dumoulin et al. 2006; Bemporad et al. 2008; Nordlund & Oliveberg 2006; Hörnberg et al. 2004). In these highly evolved states, the propensity of the proteins to form amyloid aggregates is generally very low.

Nevertheless, such folded proteins have been shown *in vitro* to undergo amyloid fibril formation readily under native-like conformations (N*), without the requirement to cross the major energy barrier for unfolding (**Figure B.15**). In certain cases, these proteins were able to initiate the self-assemble reaction in solution under conditions

that promote their partial unfolding, such as at low pH, high temperature, high pressure, and in the presence of co-solvents. These native-like conditions are thermodynamically different from the native state, and can be accessed directly from the native state through slight fluctuations or changes on the physiological conditions (i.e. pH and temperature) (**Figure B.15-A** and **B**). These fluctuations can be further facilitated by mutations that cause local structural fluctuations on its tertiary and/or quaternary structure (Chiti & Dobson 2009b).

One of the most-studied protein models for amyloid formation under native-like conditions is Sso AcP. Sso AcP is a small, α/β enzyme belonging to the acylphosphatase-like structural family with a α/β structure and an unstructured N-terminal segment. Many evidences have demonstrated that Sso AcP adopts *native-like* conformations under specific aggregating conditions (50 mM acetate buffer, pH 5.5, 25 ºC, in the presence of 15–25% (v/v) 2,2,2-trifluoroethanol) as revealed by a set of biochemical methods, such as far- and near-UV CD, enzymatic activity assays and stopped-flow measurements of folding and unfolding rates (Bemporad & Chiti 2009). During aggregation, the N-terminal segment of Sso AcP stablish interactions with the peripheral β-strand of the globular domain. In a first step, these interactions promote the formation of oligomeric species with a native-like conformation that evolve to the formation of amyloid-like structures.

Several variants of human lysozyme (containing the single substitutions I56T, F57I, W64R, D67H, as well as the double mutations F57I T70N or W112R T70N) associated with familial forms of system amyloidosis exhibit local cooperative unfolding under physiological conditions. *In vitro* experiments showed that in some of these mutants the partially unfolded state (N* in Figure **B.15-C**) is formed through a locally cooperative process and can be considered to be a conformational state thermodynamically distinct from the native state (N). This transition does not involve crossing the energy barrier for unfolding, since partially unfolded species are accessed at least five orders of magnitude faster than global unfolding, and must therefore result from inherent fluctuations associated with the native state.

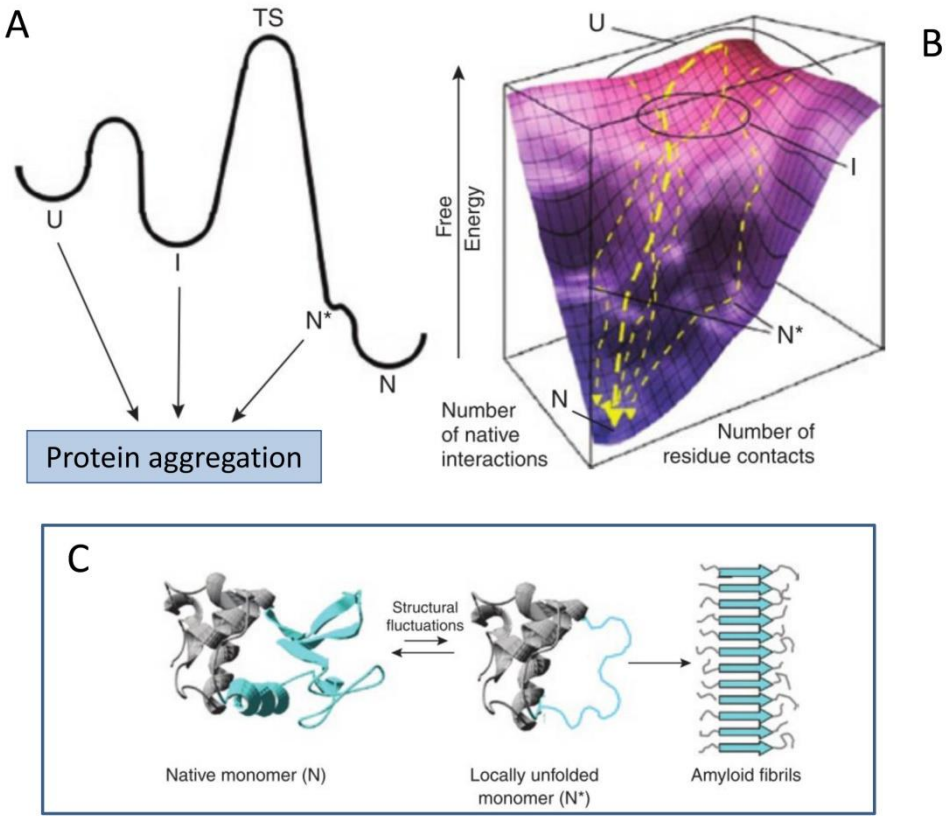


Figure B.15 Protein folding process and the principal pathways for protein aggregation. (A) Energy diagram for protein folding, according to the classic thermodynamic view. (B) Free energy landscape. (C) Schematic representation of the proposed process to explain the process of fibril formation under *native-like* conditions, taking as example the case of human lysozyme. Unfolded species (U) consists in a large ensemble of unstructured conformations. These unfolded species can collapse to form partially folded intermediaries (I) and then across the energy barrier for folding to reach the native state (N). Thermal fluctuations can lead into locally unfolded states, or also termed native-like states (N*). Such conformational ensembles represent high energy states with respect to N under physiological conditions, but are typically separated from N by a low energy barrier. Polypeptide chains adopting U, I and N* states are able to self-assemble and trigger amyloid formation, being N* more thermodynamically and kinetically more readily accessible from N than are I and U. Therefore, N* represent a key precursor to protein aggregation. Adapted from (Chiti & Dobson 2009b).

In superoxid dismutase 1 (SOD1) variant associated with familial amyotrophic lateral sclerosis (FALS), a difference with Sso AcP and lysozyme, locally destabilizing mutations cause the persistent, rather than transient, formation of conformations with locally unstructured regions in the absence of global unfolding (N*). This region is

mainly located in the loop VII, a SOD1 region highly flexible and structurally heterogeneous. One of such SOD1 mutants (S134N) displays a native-like dimeric structure, but temporary soluble oligomers can be formed through interactions involving this unstructured loop. Other mutations associated with FALS have been proposed to exert their pro-aggregating potential through a great variety of mechanisms; e.g. by decreasing the conformational stability of the native fold, by increasing the intrinsic aggregation propensity of the sequence, by causing metal loss, post-translational modifications, dimer dissociation and so on (Elam et al. 2003; Nordlund & Oliveberg 2006).

A similar behavior was found in other proteins such as transthyretin (TTR). The conversion of the will-type, as well as mutant form of TTR into amyloid fibrils is associated with a wide range of amyloidosis (see section B.2.3 below). It is widely accepted that the native TTR tetramer needs to dissociate into partially unfolded monomeric states to initiate the amyloid formation. The aggregation process is here thermodynamically favored by mutations associated with these group of familial diseases. However, again this conversion to the amyloid-competent form of the protein does not involve full unfolding of the monomeric TTR protein, indeed following tetramer dissociation, the TTR monomer undergoes only a local unfolding transition to reach the native-like state. A difference with other amyloidegenic proteins (such as lysozyme or SOD1), the fibril formation is not guided by the newly unfolded regions, but rather these unstructured regions remain exposed to the solvent in the fibrils (Hörnberg et al. 2004; Olofsson et al. 2004; Chiti & Dobson 2009b).

B.2.3 HUMAN DISEASES ASSOCIATED WITH PROTEIN AGGREGATION

A wide range of human diseases are consequence from the failure of a specific peptide or protein to adopt, or remain in, its native functional conformational state. These pathological conditions, known as protein misfolding or protein conformational diseases can be grouped into three categories, according to its anatomopathologycal features; neurodegenerative conditions, non-neuropathic localized amyloidosis and non-neuropathic systemic amyloidosis. All of the human conformational diseases

associated with formation of extracellular amyloid deposits or intracellular inclusions with amyloid-like characteristics are summarized in **Table B.5**.

It is well-known that some of these pathological conditions have a mainly sporadic origin (85%) and less frequently hereditary (10%). However, it is known that 5 % of these diseases can be transmissible in humans, as well as in other mammals, as spongiform encephalopathies.

Table B.5 Classification of human conformational diseases associated with amyloid deposition

	Disease	Aggregation protein/peptide	Native structure features
Neurodegenerative diseases	Alzheimer's disease	Amyloid β peptide	Natively unfolded
	Spongiform encephalopathies	Prion protein or fragments	Natively unfolded (residues 1–120) and α-helical (residues 121–230)
	Parkinson's disease	α-Synuclein	Natively unfolded
	Parkinson's disease	α-Synuclein	Natively unfolded
	Frontotemporal dementia with Parkinsonism	Tau	Natively unfolded
	Amyotrophic lateral sclerosis	Superoxide dismutase 1	All-β, Ig like
	Huntington's disease	Huntingtin with polyQ expansion Ataxins	Largely natively unfolded All-β,
	Spinocerebellar ataxias	Ataxins with polyQ expansion	All-β, AXH domain (residues 562–694); the rest are unknown
	Spinocerebellar ataxia	TATA box-binding protein with polyQ expansion	α+β, TBP like (residues 159–339); unknown (residues 1–158)
	Spinal and bulbar muscular atrophy	Androgen receptor with polyQ expansion	All-α, nuclear receptor ligand-binding domain (residues 669–919); the rest are unknown
	Hereditary dentatorubral-pallidoluysian atrophy	Atrophin-1 with polyQ expansion	Unknown
	Familial British dementia	ABri	Natively unfolded
	Familial Danish dementia	ADan	Natively unfolded
	AL amyloidosis	Immunoglobulin light chains or fragments	All-β, Ig like
	AA amyloidosis	Fragments of serum amyloid A protein	All-α, unknown fold

Non-neuropathic systemic amyloidoses	Familial Mediterranean fever	Fragments of serum amyloid A protein	All-α, unknown fold
	Senile systemic amyloidosis	Wild-type transthyretin	All-β, prealbumin like
	Familial amyloidotic polyneuropathy	Mutants of transthyretin	All-β, prealbumin like
	Hemodialysis-related amyloidosis	β2-microglobulin	All-β, Ig like
	ApoAI amyloidosis	N-terminal fragments of apolipoprotein AI	Natively unfolded
	ApoAII amyloidosis	N-terminal fragment of apolipoprotein AII	Unknown
	ApoAIV amyloidosis	N-terminal fragment of apolipoprotein AIV	Unknown
	Finnish hereditary amyloidosis	Fragments of gelsolin mutants	Natively unfolded
	Lysozyme amyloidosis	Variants of fibrinogen α-chain	α+β, lysozyme fold
	Fibrinogen amyloidosis	Variants of fibrinogen α-chain	Unknown
	Icelandic hereditary cerebral amyloid angiopathy	Mutant of cystatin C	α+β, cystatin like
Non-neuropathic localized diseases	Type II diabetes	Amylin, also called islet amyloid polypeptide (IAPP)	Natively unfolded
	Medullary carcinoma of the thyroid	Calcitonin	Natively unfolded
	Atrial amyloidosis	Atrial natriuretic factor	Natively unfolded
	Hereditary cerebral haemorrhage with amyloidosis	Mutants of amyloid β peptide	Natively unfolded
	Pituitary prolactinoma	Prolactin	All-α, 4-helical cytokines
	Injection-localized amyloidosis	Insulin	All-α, insulin like
	Aortic medial amyloidosis	Medin	Unknown
	Hereditary lattice corneal dystrophy	Mainly C-terminal fragments of kerato-epithelin	Unknown
	Corneal amylodosis associated with trichiasis	Lactoferrin	α+β, periplasmic-binding protein like
	Cataract	γ-Crystallins	All-β, γ-crystallin like
	Calcifying epithelial odontogenic tumors	Unknown	Unknown
	Pulmonary alveolar proteinosis	Lung surfactant protein C	Unknown
	Inclusion-body myositis	Amyloid βpeptide	Natively unfolded
	Cutaneous lichen amyloidosis	Keratins	Unknown

Adapted from (Chiti & Dobson 2006)

B.3 C-TERMINAL PROCESSING OF PEPTIDES AND GROWTH FACTORS

Neuropeptides and growth factors are essential molecules that serve many important roles in communication between cells. These essential molecules are under the fine control of proteases which usually regulate its biological activity (see selected examples at **Table B.6**). Among them, carboxypeptidases have demonstrated to play important roles by cleaving C-terminal amino acids in such substrates, leading to a complex modulation of its biological activities. These latter proteins perform a variety of important biological functions mainly in non-digestive tissues and fluids, acting in pro-hormone and neuropeptide processing, blood coagulation/fibrinolysis, inflammation, local anaphylaxis, cellular response and so on. Therefore, is essential to identify the role of a given protease in a given biological process and to understand protease signaling in health and disease (Turk 2006; Fricker 2005; Arolas et al. 2007).

Table B.6 Selected examples of neuropeptide and its precursors gene families.

Gene family	Examples of biologically active peptides
Opioid	Pro-enkephalyn (PENK), pro-opiomelanocortin (POMC)
Vasopresin/Oxitocin	Vasopresin (AVP), Oxytocin (OXT)
CCK/Gastrin	Gastrin (GAST), Cholecystokinin (CCK)
Oxitocin	Somatostatin (SST), Cortistatin (CST)
F and Y amide	Neuropetide FF (NPFF), Neuropeptide Y (NPY)
Calcitonin	Calcitonin (CALCA), Adrenomedullin (ADM)
Natriuretic Factor	Atrial natriuretic factor (NPPA), Brain natriuretic factor (NPPB)
Bombesin like	Gastrin releasing peptide (GRP), Neuromedin B (NMB)
Endotelin	Endotelin 1-3 (END 1-3)
Glucagon/Secretin	Glucagon (CGC), Secretin (SCT), Vasoactive intestinal peptide (VIP)
CRH	Corticotropin releasing hormone (CRH), Urocortin (UCN), Urotensin (VTS)
Kinin	Pre-protachykinin A and B (TAC 1 y TAC 3)
Neuromedin	Neuromedin S (NMS), Neuromedin U (NMU)
Tensins / Kinins	Angiotensin (AGT), Neurotensin (NTS)
Granins	Chromogranin A (CHGA), Chromogranin B (CHGB), Secretogranin (SCG)
Motilin	Motilin (MLN), Ghrelin (GHRL)

Galanin	Galanin (GAL), Galanin like (GALP)
Insulin	Insulin (INS), IGF-1 (IGF1), Relaxin 1 (RLN1)
GnRH	Gonadotropin releasing hormone (GnRH)
Cerebellins	Cerebellin 1 (CBLN1)
Neurexophilins	Neurexophillin 1 (NXPH1)

B.3.1 PRO-HORMONE AND NEUROPEPTIDE PROCESSING

Neuroendocrine peptides are involved in a large number of physiological processes, including feeding and body weight regulation, fluid intake and retention, pain, anxiety, memory, circadian rhythms and sleep/wake cycles, and reward pathways. Neuropeptides are an important group of "biologically active peptides," but the two terms are not interchangeable; peptides play many important biological roles and only some of these are produced in the brain. Thus, the term "neuropeptide" refers to a peptide that conveys information from one cell to another. Other types of biologically active peptides that are not neuropeptides include those that function as antibiotics in humans and many other organisms and peptide toxins such as those present in snakes, spiders, and other species. Up to date, hundreds of neuropeptides have been identified. Nonetheless, only a fraction of them are known to have biological functions (Fricker 2005; Fricker 2012).

B.3.1.1 Intracellular neuropeptide processing

The vast majority of mature neuropeptides are produced from larger protein precursors that are converted to the mature peptide forms thought the combined action of endo and exopeptidases (**Figure B.16**). These protein precursors (termed as pre-pro-peptides) are synthetized in the ER, immediately translocated to the ER lumen, where the signal peptide is removed by the proteolytic action of signal peptidases.

According to the classical pathway, pro-peptides are cleaved in the ER lumen by serine proteases belonging to the subtilisin family, named prohormone convertases 1 and 2, or PC 1 and 2. Basically, these enzymes are endopeptidases that cleaves at sites containing basic residues (e.g. at Lys-Arg, Arg-Arg or Arg-Xaa_n-Arg, where n is 2, 4, or 6) (see **Figure B.16**). Both enzymes PC 1 and 2 are strongly activated by the combination of a decreased pH and increased Ca^{2+} levels. Both enzymes are able to cleave many of the same substrates, although there are some sites cleaved preferentially by each (Zhou et al. 1999).

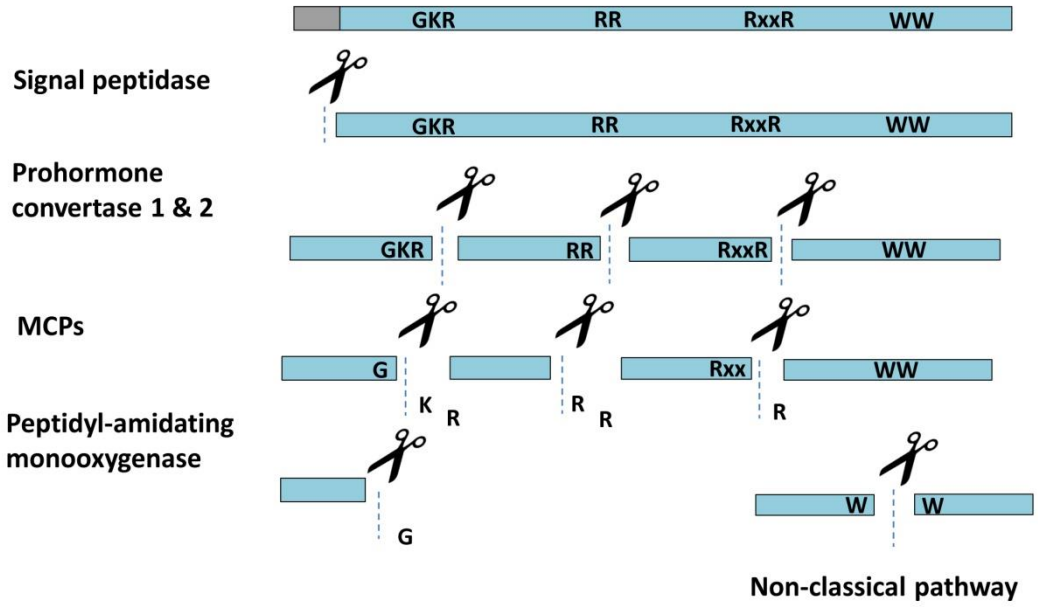


Figure B.16 Scheme of the classical and non-classical pathways for neuropeptide processing. After its synthesis in the ribosomes, the N-terminal signal drives translocation of the pre-pro-peptide into the lumen of the endoplasmic reticulum (ER), where the signal peptide is removed by signal peptidases. Then, following the classical scheme, the pro-peptide is processed at the C-terminal edge of sites containing basic residues (Arg or Lys) through the proteolytic action of prohormone convertases 1 and 2 (PC 1 and 2). Following, the resultant C-terminal basic residues are fully removed by MCPs (typically CPE, and probably others) to obtain the mature neuropeptides. Other enzymes, such as an amidating enzyme contributes to the scission of the remaining C-terminal Gly aminoacids. Alternative processing pathways (non-classical) have been proposed to explain the relative abundance of peptides cleaved at non-basic residues (i.e. the cleavage between two adjacent Trp residues). Nonetheless, the responsible enzymes for these non-classical processing are not known yet.

The resultant pro-peptides contain remaining C-terminal basic residues that are typically further cleaved by MCPs. The major peptide-processing carboxypeptidase is CPE. This enzyme, initially named as enkephalin convertase, is broadly expressed in the neuroendocrine system and processes many peptides containing C-terminal Arg, Lys or His residues, including those containing a Pro residue in P1 position (**Figure B.16**). This modification carried out by CPE is an important modification that is essential for the biological activity of many peptides (Xin Zhang et al. 2008). For many years, this enzyme was considered the unique MCPs involved in neuropeptide processing. However, recently it was suggested that others MCP like CPD or CPZ might be involved in this process (Fricker 2012).

Other alternative processing pathways (non-classical) have been proposed to explain the relative abundance of peptides cleaved at non-basic residues. Furthermore, it have been identified additional posttranslational processing events including acetylation, sulfation, phosphorylation, glycosylation, and additional proteolytic cleavages, as well as other less frequent modifications such as the n-octanoylation found within the ghrelin peptide (**Figure B.16**).

B.3.1.2 Extracellular neuropeptide processing and its biological implications

In some cases the peptides may be processed after secretion by extracellular proteases. The extracellular cleavage not always can lead to inactivation of the peptide. Sometimes the resultant product has a different affinity by its receptor or, even and increased biological activity. To date over 100 distinct peptide-binding receptors have been identified and characterized and other 100 peptide receptors remain still poorly characterized (Fricker 2012).

There are several examples where extracellular peptidases modulate the biological activity of the peptide toward one particular receptor. One of the best-characterized examples is bradykinin. This neuropeptide is synthetized mainly in plasma (produced by the proteolytic cleavage of kininogen), as a 9-residue peptide that stimulate the dilation of blood vessels. Circulating bradykinin can be metabolized by several enzymes, including two M14B MCPs; CPM and CPN. The scission of this C-

terminal residue modulate the biological activity of bradykinin, converting the peptide from a B2 agonist into a B1 agonist (Xianming Zhang et al. 2008) .

Although the biological roles of several peptides and its proteolytic products have been extensively characterized, much additional work is needed to decipher the biological activities of the majority of neuropeptides and to understand the contributions of MCPs to these complex neuropeptide signaling networks.

B.3.2 C-TERMINAL PROCESSING OF GROWTH FACTORS

Growth factors are secreted signaling molecules that control the activities of cells through intercellular communication. Typically, these heterogeneous groups of molecules bind to specific receptors on the surface of their target cells and induce a wide variety of intracellular signaling pathways that regulate an array of biological processes such as cell proliferation, activation, differentiation, and migration. Similarly to that described above for neuropeptides, several growth factors contain C-terminal relevant residues (such as the C-terminal basic amino acids found in EGF, BDNF, VEGF and several wnt family proteins), which often have a role for the modulation of its biological activity. In addition, the extracellular proteolysis of these c-terminal residues by MCPs (alone or in combination with endoproteases) can have different consequences, such as activation, inactivation and switch of function (Nakayama 1997; Reznik & Fricker 2001; Vincan 2009).

B.3.1 The EGF case

In humans, EGF is produced and secreted in tissues of epithelial, mesenchymal and neuronal origin (mainly within the digestive tract) as a pro-form, which is proteolytically processed from a high molecular-weight precursor into a biologically active peptide encompassing 53 amino acid residues (EGF1-53). The binding of the mature EGF1-53 to the EGF receptor (EGFR) leads to cell proliferation and wound healing *in vitro* and *in vivo* (Berlanga-Acosta et al. 2009). Although EGF1-53 is considered the mature form in humans, it has been found that the predominant circulating forms in the digestive tract, as well as in other body fluids and tissues are C-

terminal truncated forms (EGF1-49 and EGF1-48), probably due to its C-terminal proteolytic cleavage (Playford et al. 1995).

EGF structure is tightly related to its function (Ogiso et al. 2002). There are three well-characterized contact sites described between the ligand and the EGF receptor (EGFR) molecule (which is structured into four domains, named as domain I, II, III and IV). Within the EGF-EGFR complex, the B loop of EGF stablishes interactions with site 1 in domain I, the A loop of EGF interacts with site 2 in domain III, and the C- terminal region of EGF interacts with site 3 in domain III (**Figure B.17-A**).

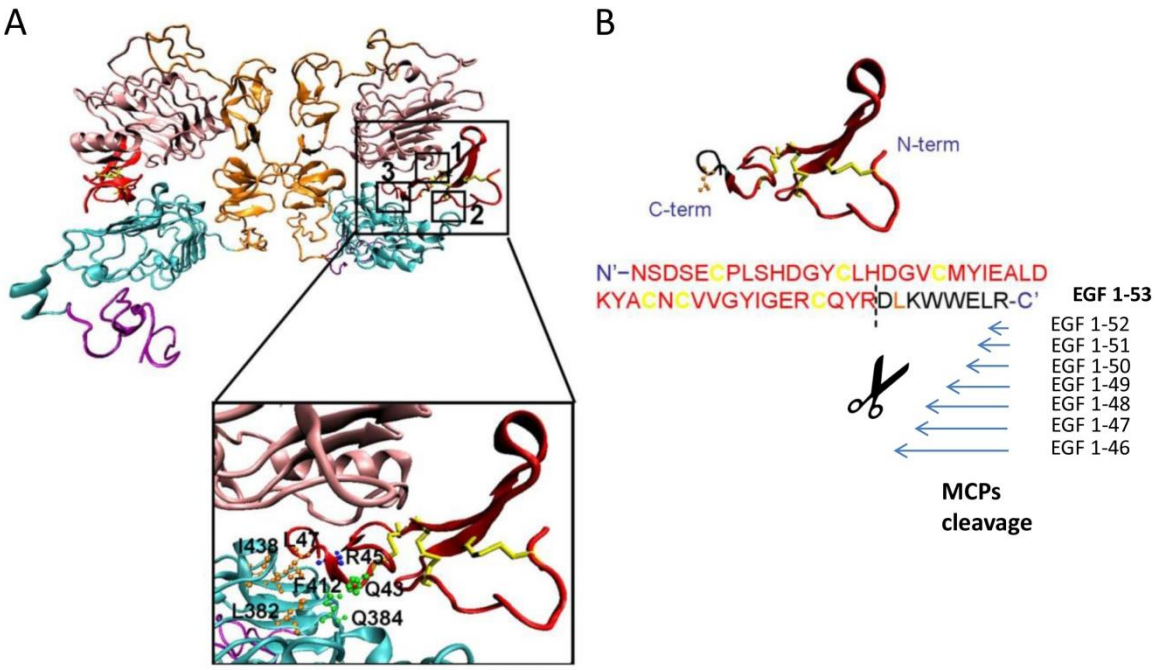


Figure B.17 Receptor binding and C-terminal proteolysis of human EGF. (A). Three-dimensional structure of the EGF-EGFR complex, showing the localization of the three binding sites. The magnification shows the interaction site 3 between C-terminal part of EGF and domain III, in which some of the most important residues are indicated. These residues interact with the EGF molecule through hydrophobic interactions (between Leu47 of EGF and Leu382, Phe412, and Ile438 of EGFR) and hydrogen bonds (between Gln384 side chain of EGFR and carbonyl and amide groups of Gln43 and Arg45 in the EGF molecule). (B) Ribbon representation EGF_{1-48} and EGF sequence showing the C-terminal truncated forms generated by C-terminal proteolysis by MCPs. Adapted from (Panosa et al. 2013).

The sequential lack of C-terminal amino acids at the EGF molecule by proteolysis (EGF1-52, EGF1-51, EGF1-50 and son on) reduce progressively its biological activity, due to the contribution of this C-terminal region for EGFR binding. Several studies have reported that EGF1-48 form is 50% less potent than the native form (Gregory et al. 1988; Goodlad et al. 1996) and even, the scission of the last 8 amino acids can transform the EGF1-45 molecule into a potent EGFR inhibitor (Panosa et al. 2013).

Nonetheless, the enzymes that cleave the c-terminal tail *in vivo* are not well-characterized. Recently, our research group has studied the contribution of several digestive carboxypeptidases, as well as other endopeptidases to the C-terminal human EGF cleavage (Lufrano & Garcia-Pardo, under preparation). Preliminary studies showed that the combined action of CPA-type, CPB-type and CPO-type MCPs suffice to remove these residues and consequently, lead to the reduction of its biological activity (**Figure B.17-B**).

B.3.3 QUANTITATIVE PEPTIDOMIC APPROACHES TO STUDY PROTEOLYTIC ACTIVITY OF METALLOCARBOXYPEPTIDASES

In recent years, a variety of Mass Spectrometry-based approaches have been developed and applied to study the substrate specificities *in vitro* and *in vivo* of a great number of proteases. Among them, peptidomics can be defined as the analysis of the peptide content of a biological sample (with peptides usually defined as amino acid polymers less than 10000 Da). Most of these peptides are generated by the action of proteases. Therefore, analysis of the peptide content of a sample provides information on the proteolytic activities required to generate the observed peptides. By using quantitative peptidomic approaches, additional information can be obtained. Using this approach, relative levels of peptides can be compared among two or more different samples, which enables studies examining the effects of increasing proteolytic activity (e.g. overexpression of proteases in a cell line) or reducing proteolytic activity (e.g. inhibition of a protease, RNA knock-down, or gene knock-out approaches). This methodology can provides information about the proteolytic activities that occur within a whole cell or a tissue (Lyons & Fricker 2012)..

Alternatively, quantitative peptidomic approaches also can be used to study the proteolytic activity of purified enzymes. In this case, the purified protease can be incubated *in vitro* with a complex mixture of peptides obtained from different sources (e.g. peptides extracted from a tissue, cell line or synthetic peptide library, among others) and then subjected to mass spectrometry analysis. By modifying the experimental conditions (times and/or concentrations of enzyme tested), the resulting data provides relevant information about the substrate specificity of the proteases analyzed.

Several studies have employed quantitative peptidomics to study proteolytic activities of M14A subfamily MCPs (Sebastian Tanco et al. 2010; Lyons & Fricker 2010). A key feature of these peptidomic approaches is the quantitative analysis of relative levels of peptides. To achieve this goal, trimethylammonium butyrate (TMAB) tags (see **Figure B.18-A**) are used to label the peptides of each independent sample, typically containing different amounts of enzyme. Following, these individual reactions are then pooled and analyzed by Liquid Chromatography and Mass Spectrometry (LC-MS) (**Figure B.18-B**).

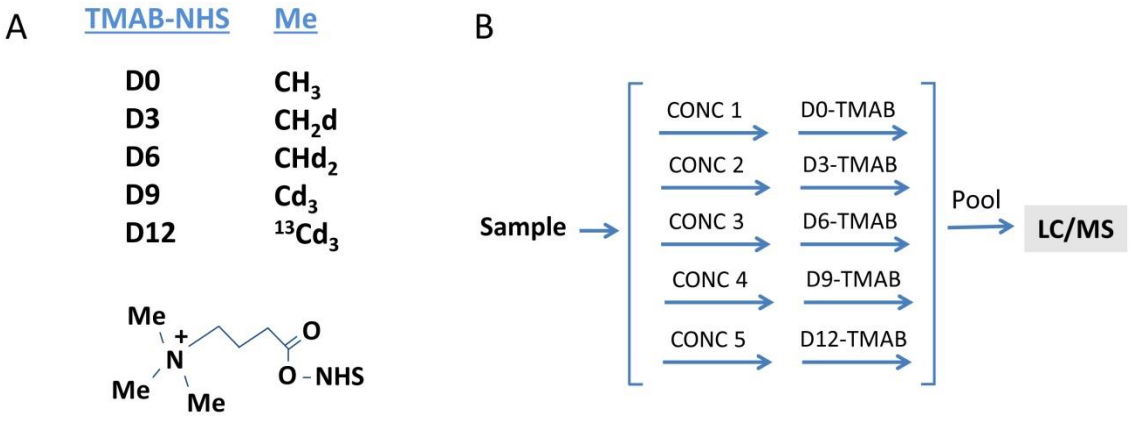


Figure B.18. Chemical reagents and typical scheme of the quantitative peptidomic approaches employed to study proteolytic activity of several metallocarboxypeptidases. (A). Chemical structure of the TMAB-NHS tags used, named as D0, D3, D6, D9 and D12. (B) Representative quantitative peptidomic experiment to examine the substrate specificities of MCPs using peptides extracted from a biological sample. To perform the experiment, this original sample is divided into aliquots, threated with different enzyme concentrations and then labelled (using a single TMAB tag for each reaction), pooled and finally subjected to LC-MS. Adapted from (Lyons & Fricker 2012).

In addition to this approach, other quantitative peptidomic techniques have been recently developed to enable the characterization of the substrate specificity of MCPS. (Tanco et al. 2015). One of the most promising approaches is C-terminal COFRADIC. The use of this technique has allowed the characterization of substrate specificities of several M14A MCPs (Tanco et al. 2013), as well as the identification of *in vivo* substrates for CCPs (Tanco et al. 2015).

Objectives

OBJECTIVES

The laboratory of Protein Engineering and Proteomics 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 main objective of this thesis has been to characterize different metallocarboxypeptidases of biotechnological and biomedical interest and to develop new tools for the production and characterization of these enzymes.

Specific objectives were prosed for each chapter:

Chapter I

- To optimize the recombinant expression and purification in *Escherichia coli* of the h-TTL domain.
- To characterize the amyloid formation under physiological conditions by the h-TTL domain.
- To study the interactions of the h-TTL domain with liposomes as a model membranes.

Chapter II

- To determine the three-dimensional structure of the h-TTL domain.
- To investigate the crystal structure of h-TTL to gain insights into the structure of the aggregation prone region.

Chapter III

- To optimize the production methodology to obtain soluble and active human CPD using mammalian cells.
- To obtain catalytically inactive single-point mutants for CPD domain I and II, as well as a double mutant for both domains.
- To characterize the substrate specificity of human carboxypeptidase D by using quantitative peptidomic approaches.

- To compare the substrate specificity of each catalytically active CPD domain and determine their optimum pH.
- To structurally model all three carboxypeptidase domains of human CPZ in order to identify the active site residues.

Chapter IV

- To develop an easy an inexpensive production system to obtain high yields of pure and active heparin-affinity metallocarboxypeptidases.

Chapter V

- To optimize the expression of human Carboxypeptidase Z with and without its frizzled-domain.
- To study the contribution of the frizzled-like domain to the enzymatic activity of human carboxypeptidase Z.
- To characterize the enzymatic properties of human carboxypeptidase Z in terms of substrate specificity and pH optimum.
- To structurally model the catalytic domain of human CPZ in order to identify the residues of the active site that participate in the catalysis, as well in ECM binding.
- To structurally model the frizzled-like domain to identify the structural feature potentially involved in Wnt binding.

Chapter I

Amyloid formation by human Carboxypeptidase D transthyretin-like domain under physiological conditions

CHAPTER I: AMYLOID FORMATION BY HUMAN CARBOXYPEPTIDASE D TRANSTHYRETIN-LIKE DOMAIN UNDER PHYSIOLOGICAL CONDITIONS

1.1 INTRODUCTION

Carboxypeptidases (CPs) perform many diverse functions in the body by removing amino acids from the C-termini of proteins and peptides. Four subfamilies of CPs can be defined based on their sequential and structural homology: M14A, M14B, M14C and M14D (Fernández et al. 2010; Rodriguez de la Vega et al. 2007b). Among them, the M14B subfamily is composed of five catalytically active members that display a stringent specificity for cleaving C-terminal basic residues only. The other members of this subfamily are inactive, lacking essential catalytic residues (Arolas et al. 2007; Reznik & Fricker 2001). All the members of this subfamily share a common structural architecture composed by a CP domain followed by a β-sandwich transthyretin-like (TTL) domain (Aloy et al. 2001; Gomis-Rüth et al. 1999; Keil et al. 2007; Reverter et al. 2004; Tanco et al. 2010).

Carboxypeptidase D (CPD) is a member of M14B subfamily that has a broad tissue distribution and functions in the processing of proteins and peptides in the secretory pathway (Sidyelyeva et al. 2006). CPD was first discovered as a 180 kDa protein from duck which binds hepatitis B viral particles (Eng 1998; Kuroki et al. 1995). Unlike all other members of the CP family, CPD contains three repeats followed by a transmembrane domain and a cytosolic tail. All three repeats contain a CP domain and a TTL domain. The function of these TTL domains is unknown, although it has been proposed that they could be involved in the regulation or oligomerization of the enzyme and/or in membrane binding (Arolas et al. 2007).

TTL domains receive their names due to their structural, but not sequential, similitude to transthyretin (TTR), a transport protein that distributes the two thyroid hormones T3 and T4 and retinol. TTR is a homotetrameric protein associated with senile systemic amyloidosis (SSA) (Cornwell, G. G. et al. 1988; Westermark et al. 1990) and familial amyloid polyneuropathy (FAP) (Saraiva et al. 1983; Hou et al. 2007). Dissociation of the TTR tetramer is a prerequisite for the development of these

disorders (Hammarström et al. 2002; Sekijima et al. 2005; Ferrão-Gonzales et al. 2003). TTR tetrameric structure dissociates into dimers, which are unstable in the absence of additional quaternary interactions, explaining why TTR exists in primarily tetramer-monomer equilibrium (Foss et al. 2005; Foguel 2005). The monomers constitute the building blocks for amyloid fibril formation (Ferreira et al. 2013; Lai et al. 1996; Jiang et al. 2001; Quintas et al. 2001).

TTR illustrates the generic overlap between interfaces and aggregation-prone regions in protein complexes, since many of the interactions promoting the formation of functional complexes, including hydrophobic and electrostatic forces, can potentially favour abnormal intermolecular association as well (Castillo & Ventura 2009; Pechmann et al. 2009). Avoidance of non-functional interactions significantly influences the evolution of the physico-chemical properties of proteins (Monsellier & Chiti 2007) and, in general, soluble monomeric proteins tend to minimize the presence of hydrophobic, potentially dangerous patches at their surfaces (Levy et al. 2012). TTR and human TTL share the same fold but differ in their quaternary structure, providing a privileged model system to dissect the determinants of protein solubility from native states.

The conversion of folded proteins into amyloid assemblies generally requires non-physiological conditions, such as extreme pHs (Guijarro et al. 1998; Castillo et al. 2013), organic co-solvents (Pallaràs et al. 2004; Chatani et al. 2012) or high temperatures (Sabate et al. 2012; Fändrich et al. 2001). These destabilizing environments cause proteins to partially or fully unfold leading to the exposure of aggregation prone regions, which are able to form intermolecular interactions, thus triggering aggregation (Invernizzi et al. 2012). However, increasing evidence supports the existence of an alternative pathway in which the aggregation of globular proteins into amyloids can depart from conformational states directly accessible from the native state without the requirement of a large unfolding (Chiti & Dobson 2009a; Soldi et al. 2005). The population of these nearly native or native-like states (N*) accounts for an increased aggregation propensity of the protein leading to formation of amyloid assemblies. They usually correspond to metastable conformers accessible through fluctuations of the native state (Chiti & Dobson 2009a; Zhuravlev et al. 2014). In these

cases, amyloid aggregates are formed without transitions across the major energy barrier for unfolding. To date, only a reduced number of proteins have been shown to form amyloids under conditions that are close to the physiological environment (Zhuravlev et al. 2014; Dumoulin et al. 2006; Bemporad et al. 2008; Nordlund & Oliveberg 2006; Hörnberg et al. 2004). Here we address the conformational and aggregational properties of a TTL domain from the first catalytic domain of human CPD (h-TTL) under close to native conditions and their functional implications, providing new insights on the interplay between the establishment of functional and deleterious protein interactions.

1.2 EXPERIMENTAL SECTION

1.2.1 RECOMBINANT h-TTL EXPRESSION AND PURIFICATION

The Transthyretin-like domain belonging to the first catalytic domain of human metallocarboxypeptidase D (residues 386-460) named here as h-TTL, was cloned into pET-22B vector to encode a C-ter hexahistidine fusion protein. For protein production, the plasmid was transformed into Escherichia coli BL21 (DE3) cells, which were then grown in 1 L of lysogeny broth (LB) medium with 50 µg.mL-1 ampicillin, at 37 ºC and 250 rpm to an OD600nm of 0.5 to 0.6. Once reached this cell density, protein expression was induced with 0.1 mM isopropyl-1-thio-β-Dgalactopyranoside (IPTG) for 16 h at 18 ºC. Then, the culture was centrifuged and the cell pellet was frozen at -20 ºC. After cell lysis by sonication in 1/50 the initial culture volume of Tris100 mM, NaCl 150 mM buffer at pH 8.0, TTL protein was purified under native conditions by affinity chromatography on a Chelating SepharoseTM Fast Flow (GE Healthcare) resin. The column was equilibrated and washed gently with a 100 mM Tris·HCl, 0.5 M NaCl buffer at pH 8.0 and the TTL protein was eluted in 3 column volumes (c.v.) of a 50 mM Tris·HCl, 0.15 M NaCl at pH 8.0 buffer containing 500 mM imidazole. The recombinant protein was further purified on a Superdex 75 HR 10/30 column (GE, Healthcare) and the protein buffer exchanged on a Sephadex G-25 column (GE, Healthcare) to a 20 mM phosphate, NaCl 100 mM buffer at pH 8.0. The purified TTL protein was flash frozen at approximately 2.2 mg·ml-1 and stored at -80 ºC.

1.2.2 INTRINSIC FLUORESCENCE

h-TTL intrinsic fluorescence was registered after equilibration at 25, 37, 42, 45 and 75 ºC in a Jasco FP-8200 spectrofluorimeter by measuring Tyr emission spectra between 280 and 400 nm upon excitation at 268 nm Slit widths were typically 5 nm for excitation and 5 nm for emission and the spectra were acquired at 0.5 nm intervals, 1000 nm·min–1 rate, and 0.1 s averaging time of a 25 µM protein concentrations in 20 mM phosphate, pH 8.0, NaCl 100 mM buffer.

1.2.3 SECONDARY STRUCTURE ANALYSIS BY CIRCULAR DICHROISM (CD)

h-TTL far-UV and near-UV CD spectra were recorded between 205 and 250 nm, and 250 and 320 nm, respectively, at 25, 37, 42, 45 and 75 ºC with a spectral resolution of 0.5 nm, using a Jasco 810 spectropolarimeter. Each spectrum was obtained by accumulating 20 scans of 25 µM protein sample in a 20 mM phosphate, NaCl 100 mM buffer at pH 8.0, in a 0.1 path-length quartz cell.

1.2.4 INTRINSIC FLUORESCENCE QUENCHING ASSAYS

Quenching of h-TTL intrinsic fluorescence was analysed by monitoring Tyr emission in the presence of acrylamide. Tyr fluorescent emission was recorded between 280 and 400 nm upon excitation at 268 nm and after equilibration at 25, 37, 42, 45 and 75 ºC using 10 µM protein samples with final quencher concentrations ranging from 0 to 0.25 M in a Jasco FP-8200 spectrofluorimeter. Data were fitted to the following equation:

$$\frac{I_0}{I} = (1 + K_{SV}[Q])e^{V[Q]}$$

Where I0 and I are the fluorescence intensities in the absence and presence of a concentration of quencher [Q]. KSV is the Stern-Volmer constant and V is the static quenching constant.

1.2.5 THERMAL AND CHEMICAL DENATURATION

h-TTL thermal denaturation was monitored by following the changes in Tyr intrinsic fluorescence at 303 nm upon excitation at 268 nm, in bis-ANS binding at 485 nm upon excitation at 370 nm and in CD ellipticity at 235 nm. Signal change was recorded using a 1 ºC·min-1 gradient and protein concentrations ranging from 5 to 25 µM.

Chemical denaturation of h-TTL was followed by motoring change in Tyr intrinsic fluorescence at 303 nm of 25 µM protein samples at different urea concentrations after equilibration at 25, 37, 42 or 45 ºC. Samples fluorescence was recorded in the 280 to 400 nm range after excitation at 268 nm, using a Jasco FP-8200 spectrofluorimeter.

Experimental data of thermal and chemical denaturation were fitted to a two-state unfolding model where the signals of the folded and the unfolded state are linearly dependent on temperature or denaturant concentration, equations 1 and 2, respectively using a non-linear least-squares algorithm provided with KaleidaGraph (Synergy Software):

(eq. 1)

$$y = \frac{(\alpha_F + \beta_F \cdot T + (\alpha_U + \beta_U \cdot T) \cdot e^{\frac{\Delta H_{VH}}{R} \cdot (\frac{1}{T_m} - \frac{1}{T})}}{1 + e^{\frac{\Delta H_{VH}}{R} \cdot (\frac{1}{T_m} - \frac{1}{T})}}$$

(eq. 2)

$$y = \frac{(\alpha_F + \beta_F \cdot D + (\alpha_U + \beta_U \cdot D) \cdot e^{m_{U-F} \cdot \frac{D - D_{50}}{R \cdot T}}}{1 + e^{m_{U-F} \cdot \frac{D - D_{50}}{R \cdot T}}}$$

Where y is the observed signal; T and D, experimental temperature and denaturant concentration, respectively; αF and αU are, respectively, the spectroscopic signals of the folded and the unfolded states either at standard temperature or in the absence of denaturant; βF and βU, the dependences of the signal on the change of temperature or denaturant concentration for the folded and the unfolded states, respectively; ΔHVH is the Van't Hoff's enthalpy of unfolding; Tm, the melting

temperature; mU-F, the proportionality constant between the free energy of unfolding and D; D50 is the denaturant concentration where the free energy of unfolding equals 0 and R is the gas constant.

1.2.6 STRUCTURE ALIGNMENT, 3D MODELLING AND PREDICTION OF AGGREGATION-PRONE REGIONS

The amino acid sequence of h-TTL and TTR were obtained from the UniProt database (http://www.uniprot.org). An structural alignment between h-TTL and TTR was generated by the Flexible structure AlignmenT by Chaining AFPs (Aligned Fragment Pairs) with Twists (FATCAT) algorithm (Ye & Godzik 2003) using the Protein Comparison Tool of the RCSB Protein Data Bank (PDB) (www.rcsb.org). Three dimensional (3D) structures of human TTR (3W3B) as well as the transthyretin-like domains of human carboxypeptidase M (CPM) (1UWY) and human carboxypeptidase N (2NSM) were obtained from the Protein Data Bank (www.rcsb.org). Structural models of TTL and transthyretin-like domains of human carboxypeptidase Z (CPZ), human carboxypeptidase E (CPE), human Adipocyte enhancer-binding protein 1 (AEBP1), human carboxypeptidase X1 (CPX1) and human carboxypeptidase X2 (CPX2) were constructed by using the automated I-TASSER online server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Zhang 2008a). Models with the best C-Score, based on the significance of threading template alignments and the convergence parameters, were selected. After I-TASSER models were built automatically, manual intervention was required to redefine secondary structure limits, based on the predictions of Jpred 3 (Cole et al. 2008), expert knowledge and experimental information. The primary sequence of h-TTL was used as input to predict its aggregation-prone regions (APRs) using the WALTZ algorithm (Maurer-Stroh et al. 2010). PyMOL (DeLano 2002) was used for figures generation and visual inspection of models.

1.2.7 IN VITRO PROTEIN AGGREGATION ASSAYS

h-TTL aggregation from soluble monomers was monitored by following the evolution over time of Th-T binding to 100 µM protein samples in 20 mM phosphate, NaCl 100 mM buffer at pH 8.0 and incubated under agitation at 25, 37, 42 or 45 ºC. Th-T binding was evaluated for samples obtained at different times by recording dye fluorescence as described below. h-TTL aggregation kinetics at different temperatures were represented as normalized Th-T fluorescence intensity at 480 nm against time and were fitted to an autocatalytic model, whenever possible, as described previously (Sabate et al. 2012). The effect of seeding on h-TTL aggregation kinetics was evaluated by adding at t=0 a 10% (w/w) of amyloid fibrils preformed at 42 ºC.

1.2.8 BINDING TO AMYLOID DYES

Thioflavin-T (Th-T) binding to h-TTL protein aggregates was measured by recording Th-T fluorescence using a Jasco FP-8200 spectrofluorimeter, with an excitation wavelength of 445 nm and an emission range between 460 and 600 nm. Spectra were registered of dilutions with a 25 µM final Th-T concentration in the absence or presence of 25 µM h-TTL aggregates at different temperatures in a 20 mM phosphate, NaCl 100 mM buffer at pH 8.0. For optical microscopy analysis, h-TTL protein aggregates were incubated for 1 h in the presence of 25 µM of Thioflavin-T (Th–T). After centrifugation at 14000xg for 5 min, the precipitated fraction was placed on a microscope slide and sealed. Th–T fluorescence images were obtained under UV light with a fluorescence microscope (Leica Microsystems).

Binding of 4,4'-bis[1-anilinonaphthalene 8-sulfonate] (bis-ANS) to h-TTL was evaluated by registering bis-ANS fluorescence between 400 and 650 nm after excitation at 370 nm in a Jasco FP-8200 spectrofluorimeter. Spectra were recorded at 25, 37, 42, 45 and 75 ºC after diluting native h-TTL in a 20 mM phosphate, NaCl 100 mM buffer at pH 8.0 with bis-ANS. Final protein and dye concentrations were 25 and 2.5 µM, respectively.

1.2.9 FT-IR SPECTROSCOPY

Attenuated Total Reflectance Fourier-transformed Infrared Spectroscopy (ATR-FTIR) analysis of h-TTL aggregates after 14 days incubation was performed using a Bruker Tensor 27 FTIR spectrometer (Bruker Optics) with a Golden Gate MKII ATR accessory. Each spectrumwas measured at a spectral resolution of 2 cm^{-1}. Infrared spectra between 1725 and 1575 cm^{-1} were fitted through overlapping Gaussian curves, and the amplitude, mass center, bandwidth at half of the maximum amplitude, and area for each Gaussian function were calculated employing the nonlinear peak-fitting program PeakFit (Systat Software).

1.2.10 TRANSMISSION ELECTRON MICROSCOPY

Samples of h-TTL incubated at 37 ºC and 42 ºC for 14 days were diluted tenfold and placed onto carbon-coated grids. After 5 minutes, grids were washed with distilled water and then, negatively stained with 2% (w/v) uranyl acetate for 2 min. Micro-graphs were recorded in a Hitachi H-7000 transmission electron microscope (TEM) operated at 75 kV accelerating voltage.

1.2.11 TTL AGGREGATION-PRONE PEPTIDE PREPARATION

A peptide with the sequence GTYNLTVVLTGYM, corresponding to the aggregation-prone region of h-TTL predicted using the WALTZ (Maurer-Stroh et al. 2010) algorithm, was purchased from EZBiolab, Inc. with a purity of 95.4%. Stock solutions were prepared at 5 mM in 1, 1, 1, 3, 3, 3-hexafluoro-2-propanol (hexafluoroisopropanol, HFIP), centrifuged at 15000 g at 4 ºC for 15 minutes and filtrated through millex-GV 0.22 mm filters to remove pre-aggregated species. After removing HFIP by evaporation, samples were stored at -80 ºC. For analysis, the peptide was resuspended in DMSO, further diluted to 100 µM in 20 mM phosphate, NaCl 100 mM buffer at pH 8.0 (with a maximum DMSO concentration of 5%) and finally bathsonicated for 10 min. Peptide aggregation was carried out without agitation at 25

ºC for 72 h and the amyloid properties of peptide aggregates were analysed as described above.

1.2.12 LIPOSOME PREPARATION AND LIPOSOME BINDING ASSAYS

Liposomes were prepared by the thin film hydratation method. Briefly, 1,2-didodecanoyl-sn-glycero-3-phosphocholine (DLPC), 1,2-dioleoyl-sn-glycero-3-phosphoric acid monosodium salt (DOPA), and Cholesterol were dissolved in chloroform solutions and mixed in a round-bottom flask at the desired molar ratios for DOPA (0.0:0.5:0.5), DOPA:DLPC (0.25:0.25:0.5) and DLPC (0.5:0.0:0.5) liposomes. The organic solvent was removed by rotary evaporation to obtain a dry lipid film that was then hydrated with 20 mM phosphate pH 8.0, 100 mM NaCl, buffer to give a lipid concentration of 10 mM. Multilamellar liposomes (MLV) were formed by constant vortexing followed by extrusion in an Extruder (Lipex Biomembranes) through polycarbonate membranes (Avanti Polar Lipids, USA) of variable pore size under nitrogen pressure. Liposomes were extruded in three steps: first through a 0.8 µm pore diameter filter, then through a 0.4 µm filter and finally through a 0.2 µm filters) until the obtention of large unilamellar liposomes (LUV). The particle size distribution and zeta potential (ζ) of the final liposomal formulations were determined by dynamic light scattering at 25 ºC using a Zetasizer NanoZS (Malvern Instruments).

The intrinsic fluorescence of 25 µM h-TTL in 20 mM phosphate, NaCl 100 mM buffer at pH 8.0 was monitored after equilibration at 25ºC in presence of DOPA, DOPA/DLPC or DLPC liposomes at 0.25, 0.5 and 1.0 mM final concentration as described above. The aggregation of h-TTL in the presence of DOPA, DOPA/DLPC, and DLPC liposomes was evaluated at 1 mM liposome concentration. Reactions were incubated at 42 ºC for up to 3 days and h-TTL aggregation was monitored as described above.

1.3 RESULTS

1.3.1 STRUCTURAL SIMILITUDE BETWEEN HUMAN TTL AND THE TTR MONOMER

In the crystal structure of TTR each monomer (A, B, C and D) is composed of two four-stranded β-sheets (with a DAGH and CBEF arrangement), which are connected by loops with a short α-helix located between β-strands E and F **(Figure 1)** (Hörnberg et al. 2000; Hamilton & Benson 2001). Duck CPD TTL shares topological similarity and connectivity with TTR over 78 Cα atoms (with a rmsd of 2.4 Å) despite their low sequence similarity (18% identity) and TTL lacking strand eight present in TTR.

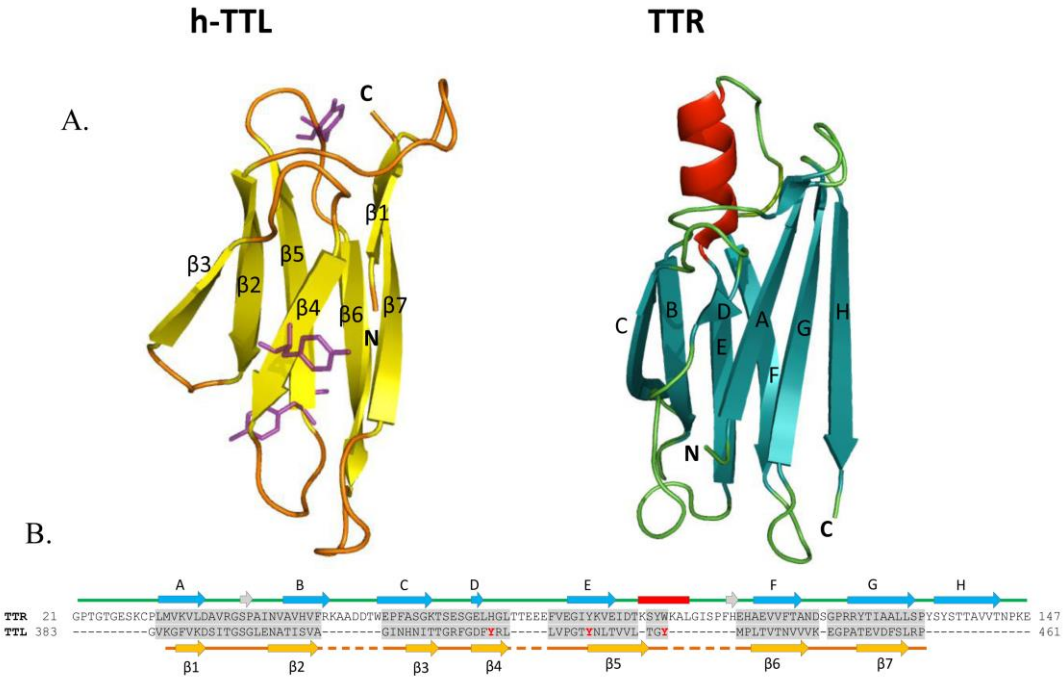


Figure 1. Structural comparison of the human carboxypeptidase D transthyretin-like domain (h-TTL) and human transthyretin (TTR). (A) Ribbon representation of h-TTL model (left) and TTR (right) 3D structures. The labels represent the secondary structure elements and the N- and C-termini. The side chains of Tyr 38, Tyr 46 and Tyr 55 in h-TTL (Tyr 420, Tyr 428 and Tyr 437 according to CPD numeration) are shown in purple. (B) Structure-based sequence alignment of of h-TTL and TTR sequences. Residues highlighted in grey correspond to regions with structural similarity and residues in red correspond to tyrosines.

The structure of the duck CPD TTL domain consists of a β-barrel made up by seven strands (β-1 to β-7) arranged in a antiparallel four-stranded (β-3, β-2, β-5 and β-6) and a mixed three-stranded β-sheet (β-1, β-4 and β-7), both folded together enclosing a hydrophobic core (Aloy et al. 2001; Gomis-Rüth et al. 1999; Keil et al. 2007). The same topology was found in the crystal structure of a short splicing variant of *Drospophila melanogaster* CPD (Sebastián Tanco et al. 2010). A similar transthyretin-like fold topology has been found in several, diverse protein families including dioxygenases, glucotransferases and glucoamylases (Harata et al. 1996; Orville et al. 1997; Sorimachi et al. 1997). Because the 3D-structure of the TTL domain of human CPD (h-TTL) is not available yet, we modelled it on top of solved TTL structures using I-TASSER. The resulting model and a structural alignment with human TTR are shown in **Figure 1**.

1.3.2 SPECTRAL PROPERTIES OF HUMAN TTL

We expressed and purified h-TTL. In gel filtration chromatography the protein elutes as a single peak, with an apparent Mw of 8.7 KDa, fairly close to the theoretical and mass spectrometry determined Mw of 9.3 KDa, indicating that the domain remains as a monomer in solution (data not shown). We monitored its conformational properties in the native state by far- and near-UV circular dichroism (CD) and intrinsic fluorescence at pH 8.0 and 25 °C at a 25 μM protein concentration. The far-UV CD spectrum of h-TTL exhibits a single minimum at 217 nm consistent with an all β-sheet structure **(Figure 2A)**. The spectrum exhibits a maximum at 236 nm (**Figure 2A**), which is usually attributed to the contribution of Trp and Tyr side chains since the amide contributions in this region are generally negative. Because h-TTL lacks Trp residues, this signature can be univocally assigned to tyrosine (Tyr) side chains. Accordingly, the near-UV CD spectrum exhibits a single negative band in the 270-280 nm interval **(Figure 2B).**

Human TTL contains three Tyr residues at positions 38, 46, and 55. Tyr38 is exposed to solvent whereas Tyr46 and Tyr55 are buried in the h-TTL structure. We monitored the intrinsic fluorescence of Tyr residues by exciting the protein at 268 nm and recording fluorescence between 280 and 400 nm. The fluorescence spectrum of h-TTL exhibits a characteristic Tyr emission maximum at 303 nm **(Figure 2C).**

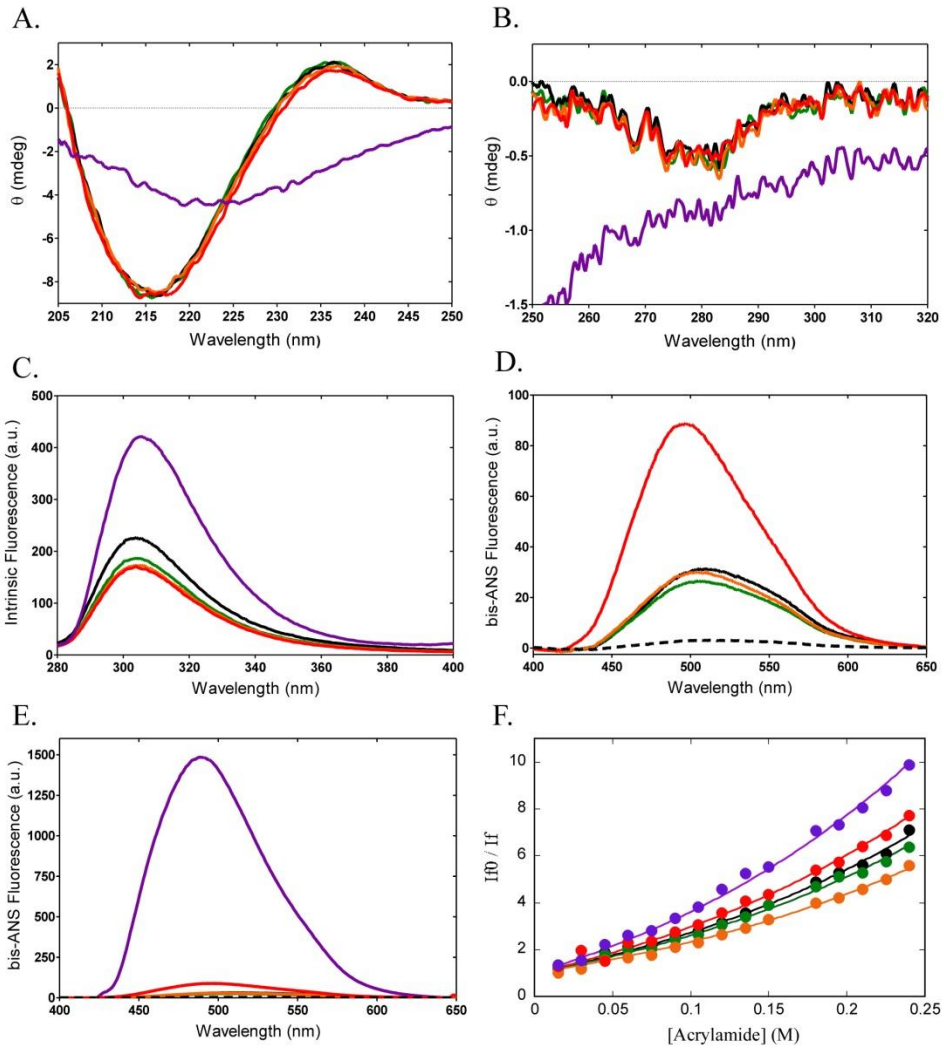


Figure 2. Conformational analysis of h-TTL. (A) Far UV-CD, (B) Near UV-CD, (C) intrinsic fluorescence, (D and E) bis-ANS binding and (F) Stern-Volmer plots of the acrylamide quenching of tyrosine intrinsic fluorescence at 25 (black), 37 (green), 42 (orange) 45 (red) and 75 (purple) ºC. In D and E, dashed lines represent free bis-ANS emission spectra.

1.3.3 THERMAL AND CHEMICAL UNFOLDING OF HUMAN TTL

The thermal stability of h-TTL was analyzed by monitoring the changes with temperature of Tyr intrinsic fluorescence at 303 nm, in CD ellipticity at 235 nm and in bis-ANS extrinsic fluorescence at 485 nm. The analysis were performed in the 15-75 °C range at pH 8.0 and 10 µM protein concentration **(Figure 3A-C)**. A single cooperative transition was observed in all cases, and the data could be fitted to a two-state temperature-induced unfolding model (R > 0.99). Melting temperatures (Tm) of 55.0 ±

0.1, 55.2 ± 0.1 and 53.7 ± 0.1 °C, were calculated from intrinsic fluorescence, CD and bis-ANS fluorescence data, respectively, indicating that all the probes report on the same global unfolding process. No detectable protein aggregation occurred during melting in the 5-25 µM concentration range (data not show).

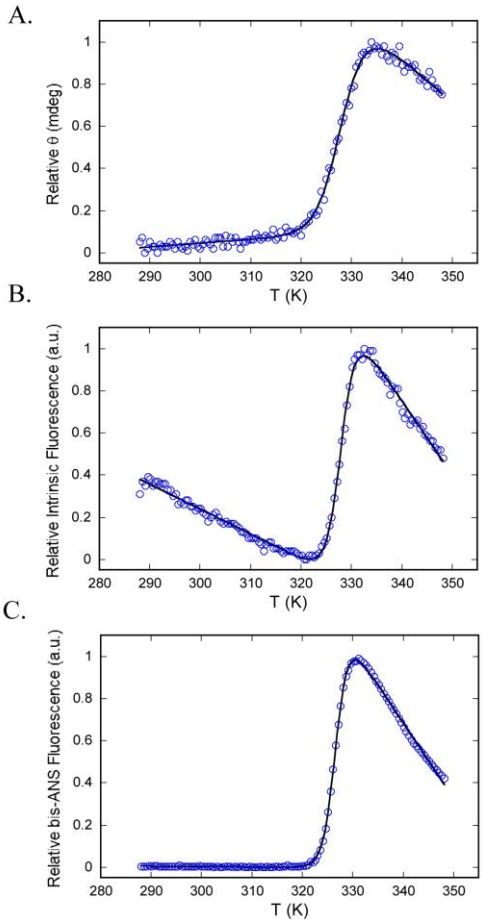


Figure 3. Thermal stability of h-TTL. (A) Thermal unfolding was monitored by following changes in (A) CD signal at 235 nm, (B) intrinsic fluorescence at 305 nm and (C) bis-ANS fluorescence at 485 nm, in the 15-75 ºC range.

Thermal unfolding data indicate that the protein is essentially folded in the 25-45°C temperature range. To further confirm this extent we analysed the far-UV, near-UV and intrinsic fluorescence of h-TTL at 25, 37, 42 and 45 °C and compared them with that of the denatured protein at 75 °C. The fluorescence spectra of h-TTL in the 37-45 °C range are identical and slightly less intense than that at 25 °C **(Figure 2C)**; all them being different from the spectra exhibited by the unfolded protein, where a large

increase of the fluorescence maximum was observed, indicating the exposure to solvent of previously protected Tyr residues (**Figure 2C**). We monitored the presence of exposed hydrophobic clusters in the structure of h-TTL at these temperatures by measuring their binding to 4,4'-bis[1-anilinonaphthalene 8-sulfonate] (bis-ANS), a dye that increases its fluorescence emission upon interaction with these nonpolar regions. The spectra in the 25-42 °C range were essentially identical, with an increase in fluorescence emission of bis-ANS at 45 °C (**Figure 2D**). Still, this change was much lower than the one observed for the unfolded protein at 75 °C (**Figure 2E**). To further confirm the native-like conformation of h-TTL in the 25-45 °C range we recorded the near- and far-UV CD spectra of the protein at 25, 37, 42 and 45 °C at pH 8.0. The spectra obtained at all these temperatures can be superposed in both the near and far-UV regions and are clearly different from that of the unfolded state at 75 °C (**Figure 2A and 2B**). Finally, the relative accessibility of Tyr residues was analyzed by acrylamide quenching. Upward curving in Stern-Volmer plots is consistent with h-TTL Tyr residues being located in different environments, with both static and dynamic quenching contributions (**Figure 2F**). The Stern-Volmer constant for h-TTL (K_{sv}) at 25°C is 10.8±0.1 M^{-1}. Partial or total unfolding is usually associated with an increase in K_{sv} values. This effect is observed at 45°C (K_{sv} = 12.6±0.2 M^{-1}) and, especially, at 75°C, (K_{sv} = 17.5±0.2 M^{-1}). In contrast, K_{sv} values of 9.9±0.1 M^{-1} and 7.7±0.2 M^{-1} were calculated for the protein at 37°C and 42°C, respectively. Despite K_{sv} constants obtained at different temperatures should be interpreted with caution, quenching experiments suggest that the protein remains in a compact conformation in the 25-42°C temperature range.

We later analysed the dependence of the conformational stability of h-TTL on the temperature by monitoring the resistance of the protein against chemical denaturation with urea at pH 8.0 and 25, 37, 42 or 45 °C at a 25 µM concentration, by following intrinsic fluoresce changes at 303 nm at equilibrium (**Figure 4**). A single detectable transition was observed in all cases, indicating that the protein unfolds cooperatively from an initially highly packed state. The main thermodynamic parameters of the unfolding reaction were calculated from the equilibrium curves assuming a two-state model (R > 0.99 in all cases). The thermodynamic stability of the domain exhibited a high dependence on the temperature (**Table 1**).

Table 1. Thermodynamic characterization of h-TTL at different temperatures by intrinsic florescence

Temperature (K) / (ºC)	$\Delta G_{U\rightleftharpoons F}^{H_2O}$ (kcal mol^{-1})	m (kcal mol^{-1} m^{-1})	$[Urea]_{50\%}$ (M)
298 / 25	6.12 ± 0.04	1.05 ± 0.01	5.83
310 / 37	4.69 ± 0.04	1.03 ± 0.01	4.55
315 / 42	3.61 ± 0.08	0.86 ± 0.02	4.20
318 / 45	1.74 ± 0.13	0.64 ± 0.03	2.72

At 25 °C h-TTL exhibits a Gibbs free energy difference, extrapolated to absence of denaturant, (ΔG^{H2O}) of 6.12± 0.04 kcal.mol^{-1}, with a half-transition denaturant concentration ($[Urea]_{50\%}$) of 5.83 M.

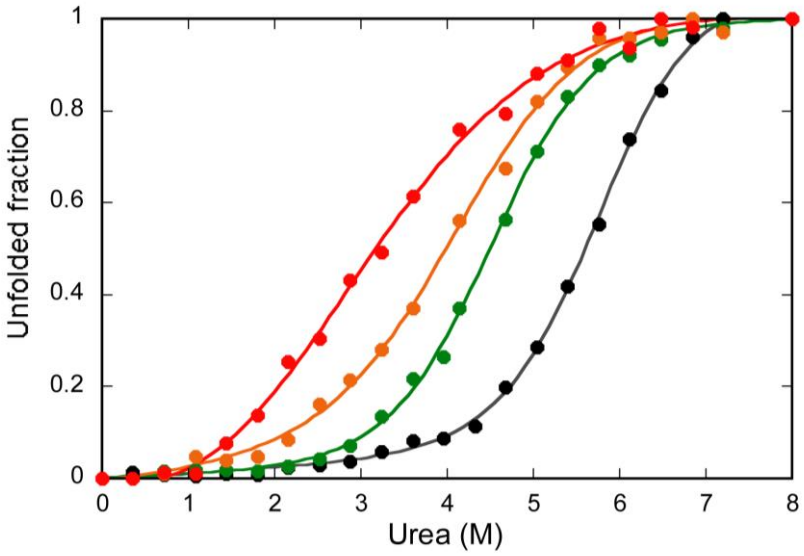


Figure 4. Stability of h-TTL against chemical denaturation. Equilibrium unfolding in urea at 25 (black), 37 (green), 42 (orange) and 45 ºC (red).

At 37, 42 and 45 °C the protein is destabilized by 1.43, 2.41 and 4.38 kcal·mol^{-1}, respectively (Table 1). In a similar manner, the m-value decreases as temperature increases, indicating lower cooperativity of the unfolding reaction at higher temperatures (Table 1).

1.3.4 AGGREGATION OF HUMAN TTL INTO AMYLOID-LIKE STRUCTURES

Since dissociated TTR monomers have been shown to readily aggregate into amyloid-like assemblies (Hammarström et al. 2002), it resulted interesting to evaluate the aggregative properties of a conformationally similar, yet monomeric, domain. Human TTL was incubated at 100 µM for 14 days at pH 8.0 and 25, 37 or 42 °C to assess whether this domain aggregates into amyloid-like structures under close to physiological conditions and if this process is dependent on the thermodynamic stability of the native state.

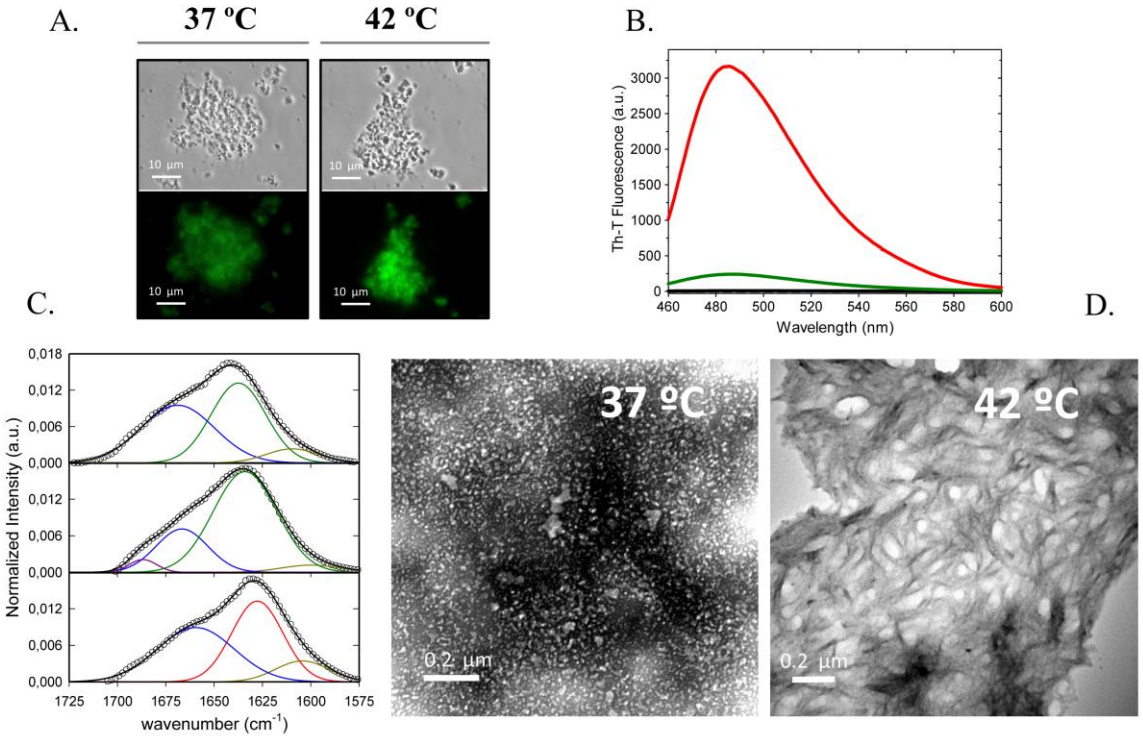


Figure 5. Morphological and conformational analysis of h-TTL aggregates. (A) Bright field (upper panels) and fluorescence (lower panels) imaging of h-TTL aggregates stained with Th-T after 14 days incubation at 37 and 42 ºC. (B) Th-T binding to native h-TTL (solid black lines) and h-TTL aggregates incubated at 37 (green) and 42 ºC (red). Free Th-T emission is represented as dashed lines. (C) Normalized IR spectra of native h-TTL (upper panel) and h-TTL aggregates incubated at 37 (middle panel) and 42 ºC (lower panel). Coloured lines represent different secondary structure elements arising from Gaussian deconvolution. (D) Representative TEM micrographs of h-TTL aggregates after 14 days incubation.

No apparent aggregation was observed at 25 °C (data not shown). In contrast, we did observe the formation of protein aggregates that were stained with the amyloid specific dye Thioflavin-T (Th-T), yielding bright green–yellow fluorescence against a dark background when observed by fluorescence microscopy upon incubation, at both 37 and 42 °C (**Figure 5A**).

Th-T fluorescence emission is enhanced in the presence of amyloid fibrils and the same behaviour is observed upon incubation of aggregated h-TTL with Th-T; especially for samples incubated at 42 °C, where the Th-T fluorescence at the 480 nm spectral maximum increases by 100-fold (**Figure 5B**). No Th-T binding to soluble h-TTL was detected at any of the assayed temperatures prior to incubation (data not shown).

Attenuated Total Reflectance Fourier-transformed Infrared Spectroscopy (ATR-FTIR) allowed addressing the structural features of both native and aggregated h-TTL. The absorbance spectra of native h-TTL in the amide I region is dominated by a band at 1637 $cm^{-1,}$ in agreement with its all-β fold (**Figure 5C** and **Table 2**).

Decovolution of the spectra demonstrates differences in the secondary structure content of the protein incubated at 37 and 42 °C (**Figure 5C** and **Table 2**). In this way, whereas the IR spectrum at 37 °C is dominated by a signal at 1634 cm^{-1} typically attributed to β-sheet structure, the main signal at 42 °C is shifted to 1627 cm^{-1}, indicating shorter hydrogen bonding and, therefore, more densely packed β-sheet structures, compatible with intermolecular contacts in an amyloid fold (**Table 2**).

The morphological features of h-TTL samples were analysed after incubation using transmission electron microscopy (TEM). As shown in **Figure 5D**, we detected the presence of protein aggregates at both 37 and 42 °C. Nevertheless, the size and morphology of the aggregates formed at the two temperatures were significantly different. In good agreement with the Th-T binding and secondary structure data, small oligomer-like aggregates, which tended to clump together, were observed at 37 ºC, whereas at 42 °C the protein assembled into dense amyloid-like fibrillar meshes.

1.3.5 DEPENDENCE OF THE AGGREGATION KINETICS OF HUMAN TTL ON THE TEMPERATURE

We analysed the kinetics of h-TTL aggregation and its dependence on the temperature by monitoring its Th-T binding over time. As shown in **Figure 6A,** subtle changes in temperature result in dramatic effects on the aggregation regime. The kinetics of amyloid fibril formation usually follows a sigmoidal behaviour characterized by three kinetic stages: (1) lag phase, (2) exponential growth phase, and (3) plateau phase. These phases, characteristic of most amyloid processes, reflect a nucleation-polymerization mechanism. The h-TTL aggregation kinetics at 42 °C can be rationalized according to this mechanism and fitted to an autocatalytic equation, with nucleation and elongation constants of 0.11 and 33.8 (10^6 min^{-1}), respectively and a lag time of 1410 min. Lowering the temperature to 37 °C makes the aggregation reaction exceedingly slow, reaching only 10% of the maximal Th-T binding recorded at 42 °C after 14 days. In contrast, increasing the temperature just by 3°C, to 45 °C, exacerbates aggregation, abrogating the lag phase of the reaction.

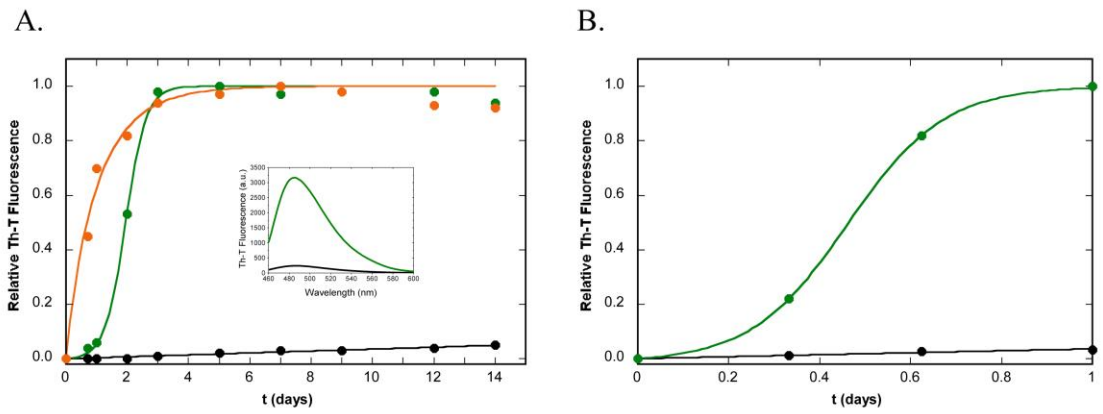


Figure 6. h-TTL aggregation kinetics. (A) Fraction of aggregated h-TTL, measured as Th-T binding, as function of time at 37 (black), 42 (green) and 45 ºC (orange). The inset shows Th-T emission spectrum of h-TTL aggregates after 14 days incubation at 37 (black) and 42 ºC (green). (B) Aggregation reaction performed at 25 ºC in the absence (black) or in the presence (green) of 10% of fibrils (w/w) preformed at 42 ºC.

We compared Th-T signal during aggregation at 42 °C with both intrinsic and bis-ANS fluorescence at each of the reaction stages (**Figure 7**). The low Th-T signal during the lag phase correlates with small changes in protein intrinsic fluorescence, suggesting that h-TTL does not suffer extensive unfolding at this stage. In contrast, bis-ANS fluorescence increases significantly during nucleation, indicating the formation of new hydrophobic clusters that likely reflect the assembly of native-like h-TTL into small oligomers (Bolognesi et al. 2010). Like Th-T signal, intrinsic fluorescence progressively increases at the growth and plateau stages, indicating that the h-TTL Tyr residues are in a different protein environment in the native and fibrillar states. The increase in intrinsic protein fluorescence is common to many anyloidogenic processes and seems to respond to the more hydrophobic environment that the amyloid fibril provides to aromatic residues as well as to their stacking (Ramírez-Alvarado et al. 2003), which is in agreement with the observed concomitant increase in bis-ANS binding.

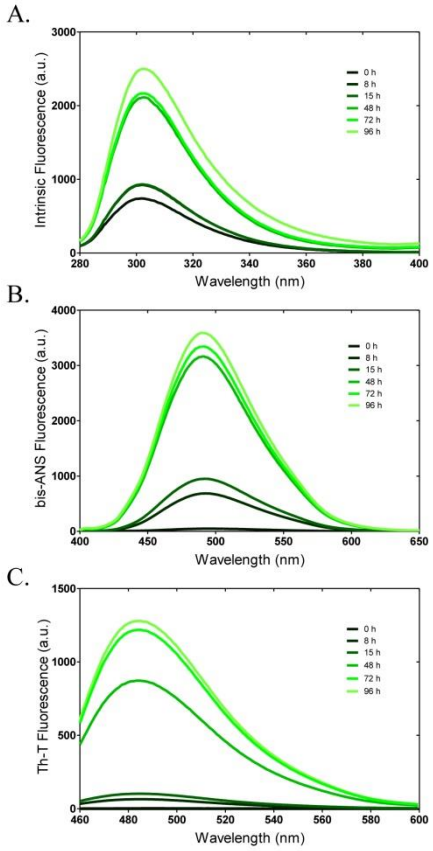


Figure 7. h-TTL conformational changes during its aggregation reaction. Changes monitored by (A) Intrinsic fluorescence, (B) bis-ANS binding and (C) Th-T binding at the different stages of the aggregation process: initial (0h), lag phase (8h and 15h), growth phase (48h) and plateau (96h). The reaction was performed at 42 ºC.

We further analysed whether the amyloid fibrils formed at 42 °C were able to seed the aggregation of soluble h-TTL at 25 °C, where no apparent aggregation occurs spontaneously. Interestingly enough, the presence of 10% of preformed fibrils promotes the formation of Th-T positive aggregates following a classical sigmoidal curve (**Figure 6B**). This suggests that the fibrillar state is able to recognize one or more regions in the structure of the folded protein and subsequently promote the incorporation of the complete soluble protein into the fibril.

1.3.6 HUMAN TTL DISPLAYS A HIGHLY AMYLOIDOGENIC SHORT-SEQUENCE STRETCH

It is now accepted that specific and continuous protein segments nucleate amyloid-like reactions and participate in the formation of the β-core of the mature fibrils (Ventura et al. 2004). Different computational methods have been developed to predict those sequential stretches (Castillo et al. 2011). Here we used WALTZ, which exploits a position-specific scoring matrix deduced from the biophysical and structural analysis of the amyloid properties of a large set of hexapeptides, to identify amyloid aggregating sequences (Maurer-Stroh et al. 2010). We identified a single amyloidogenic sequence stretch in h-TTL, encompassing residues G44-M56 (**Figure 8A**), which overlaps with β-sheet 5 and includes part of the loops at its N-ter and C-ter sides (**Figure 8B**).

To assess if this region has the ability to self-assemble and act as possible nucleation element in the aggregation process of h-TTL, we synthesized and characterized the amyloidogenic properties of the corresponding peptide (GTYNLTVVLTGYM). The peptide was incubated at 100 µM, pH 8.0 and 25 °C. In these conditions, the solution becomes slightly cloudy after 1 min. The formation of fibrillar structures with size and morphology compatible with an amyloid nature could be observed by TEM upon 3 days incubation (**Figure 8C**). We analyzed the secondary structure content of the fibrils by ATR-FTIR in the amide I region of the spectrum (**Figure 8D**). The absorbance spectrum in this region was dominated by a peak at 1626 cm^{-1} that indicates the presence of intermolecular β-sheet. Finally, Th-T binding (**Figure**

8E) confirms the amyloidogenic properties of the fibrillar structures formed by the most aggregation-prone sequence of h-TTR.

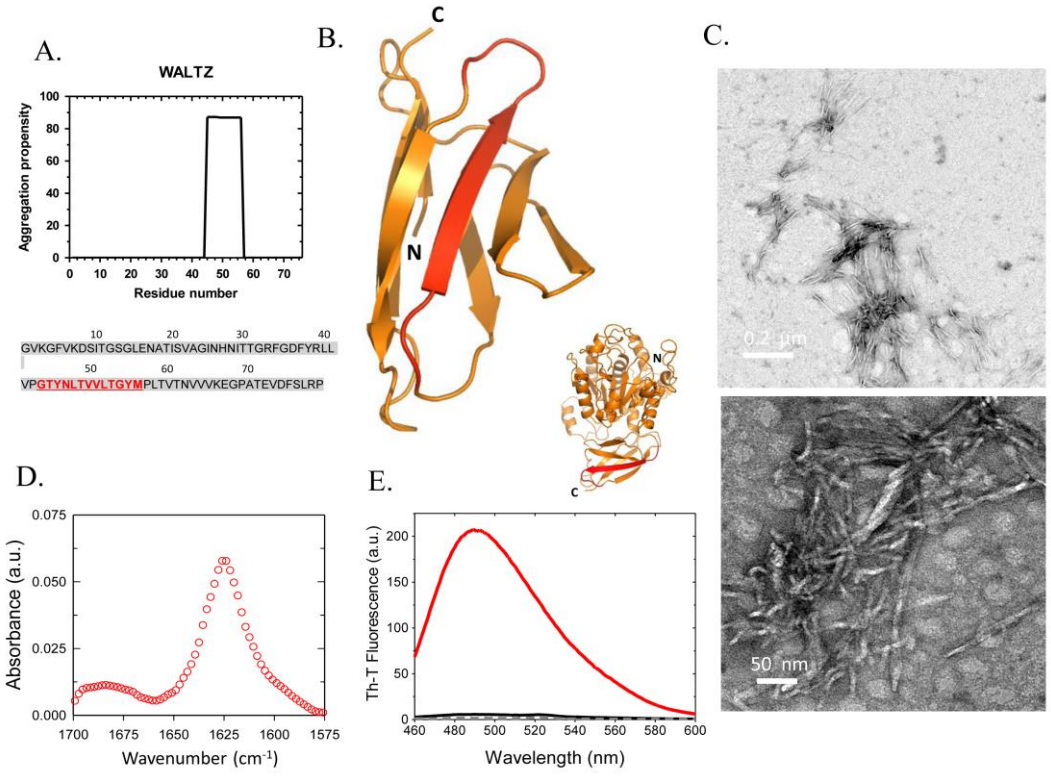


Figure 8. Aggregation properties of h-TTL. (A) Aggregation profile of h-TTL computed using WALTZ. The only detected aggregation-prone region (APR) corresponds to the peptide sequence GTYNLTVVLTGYM. (B) Structural location of the APR detected with Waltz (in red) in the h-TTL and the human carboxypeptidase D domain I (showing its catalytic domain at the top and the h-TTL domain at the bottom) structural models. (C) Representative TEM micrographs showing the fibrillar morphology of the h-TTL-derived peptide aggregates, the observed average fibril diameter is ~10 nm. (D) IR spectrum of the aggregated peptide, clearly dominated by a characteristic intermolecular β-sheet signal. (E) Th-T binding to peptide aggregated (in red), free Th-T emission spectrum is shown in black. The dashed grey line corresponds to the buffer alone.

1.3.7 BINDING TO MEMBRANES MODULATES HUMAN TTL AGGREGATION

It has been been proposed that h-TTL might be involved in binding of CPD to the cell membrane, and in fact other TTR-like proteins have been shown to serve this function (Kang et al. 2012).

The presence of biological membranes may strongly affect the aggregation of amyloidogenic proteins (Sabaté, Espargaró, et al. 2012). We addressed if this is the case for h-TTL using liposomes as mimics of natural cell membranes. The effect of negatively charged liposomes (1,2 dioleoyl-sn-glycero-3 phosphate sodium salt) (DOPA), neutral liposomes (1,2-didodecanoyl-sn-glycero-3-phosphocholine) (DLPC) and liposomes containing an equimolar mixture of charged and neutral lipids (DOPA:DLPC) was assessed.

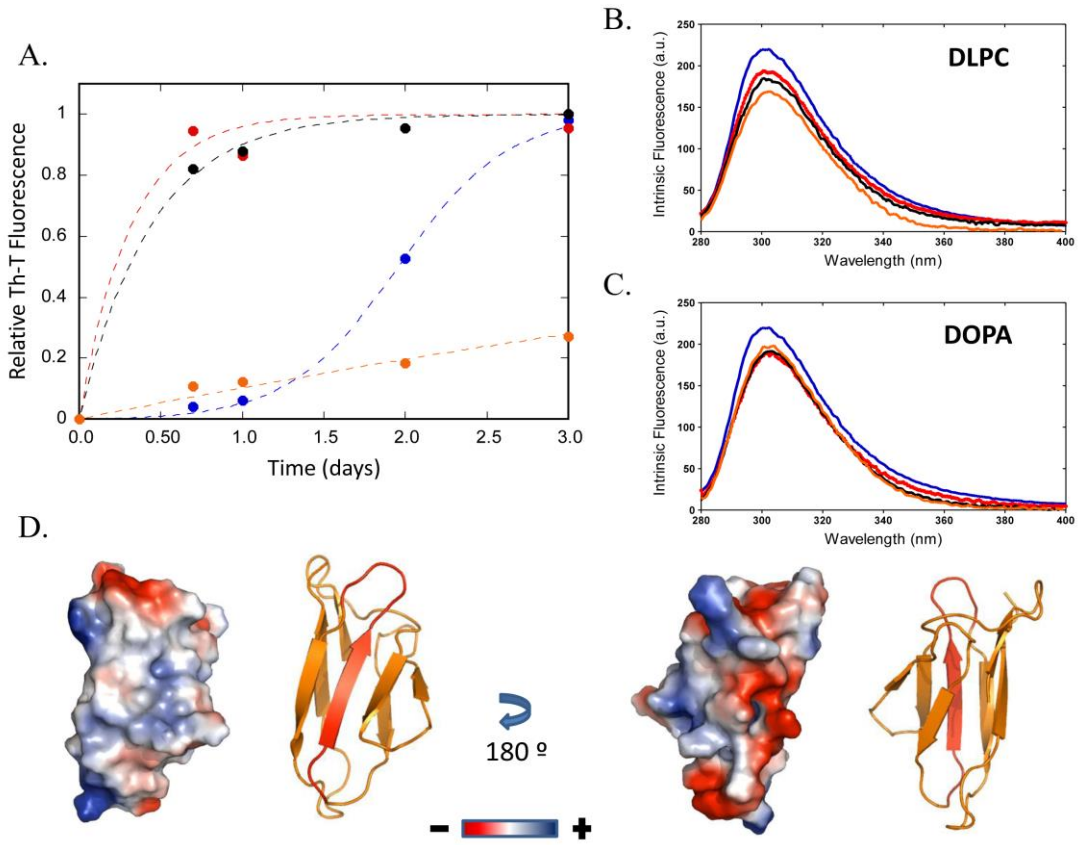


Figure 9. Effect of liposome composition on h-TTL aggregation. (A) Aggregation kinetics at 42 ºC, followed by Th-T binding, of h-TTL alone (blue) and in the presence of 1 mM DOPA (red), DOPA/DLPC (1:1) (black) or DLPC liposomes (orange). Intrinsic fluorescence spectra of h-TTL alone (blue) and in the presence of 0.1 mM (red), 0.5 mM (black) or 1.0 mM (orange) of (B) neutral DLPC and (C) highly negatively charged DOPA liposomes. (D) Electrostatic surface potential distribution and ribbon representation of h-TTL (in the same orientation) showing in red the location of the APR detected by Waltz. Two views are shown related by a 180º rotation around the z-axis. Blue indicates positive and red indicates negative charge potential.

As shown in **Figure 9A**, the effect of liposomes on the kinetics of h-TTL aggregation at 42 °C is dramatically dependent on their charge. The presence of DPLC liposomes exerts a strong inhibitory effect on h-TTL aggregation. DPLC liposomes with a z potential (ζ) of -1.6 quench the fluorescence of Tyr residues in h-TTL in a concentration dependent manner (**Figure 9B**), suggesting that this inhibitory effect might be mediated by liposome interaction with aromatic/hydrophobic residues in the native h-TTL state. The presence of DOPA containing liposomes with a z potential (ζ) of -46 strongly accelerates the reaction completely abrogating the lag phase. h-TTL is an acidic protein with a pI of 5.6 and, therefore, a generic adsorption to liposomes facilitated by the presence of a negative surface potential in the membrane is not expected. However, an inspection of the charge distribution of the h-TTL surface indicates an asymmetric localization of the negative and positive residues, which are placed in opposed faces of the protein (**Figure 9D**). The amyloidogenic β-strand 5 is located in the positively charged side. Therefore it is tempting to hypothesize that the strongly pro-aggregational effect exerted by negatively charged liposomes might be related to a preferential orientation of h-TTL relative to the lipidic surface, affecting either directly or indirectly the microenvironment of the most amyloidogenic region in h-TTL. In this orientation, the negatively charged side where the exposed Tyr38 resides will remain apart from the membrane, explaining the lower fluorescence quenching exerted by DOPA containing liposomes (**Figure 9C**).

1.4 DISCUSSION

All the members of the CP M14B subfamily share a β-sandwich TTL domain at the C-terminus, which function remains uncertain. Here we show that, the globular h-TTL domain displays an intrinsic propensity to aggregate into amyloid-like conformations, a property shared with the structurally homologous TTR monomer. The spectroscopic probes indicate that, at 42 °C, h-TTL has structural properties that are very similar to those of the native state at 25 °C. However, at 25 °C the protein remains soluble whereas at 42 °C the protein aggregates into amyloid-like structures following characteristic sigmoidal kinetics. Before aggregation the protein appears to retain a secondary structure identical to this of the native state and Tyr side-chains report on the existence of a compact conformation. Moreover, the binding to bis-ANS remains unaltered relative to the native structure, ruling out the emergence of additional hydrophobic patches on the surface at 42 °C.

The protein is destabilized at 42 °C, with a ΔG^{H2O} of 3.61 kcal·mol^{-1}, significantly smaller than the ΔG^{H2O} of 6.12 kcal·mol^{-1} measured at 25 °C. Despite h-TTL still exhibits an unfolding cooperativity characteristic of folded states at 42 ºC and spectroscopic properties indistinguishable of that of the protein at 25 °C, the m-value suggests an overall lower cooperativity of the protein at 42 °C, which might result from transient fluctuations around the native conformation at this temperature. In a similar manner to what has been proposed for lysozyme, a hyperthermophilic acylphosphatase, superoxide dismutase 1, TTR and β2-microglobulin (Chiti & Dobson 2009), they will likely be these conformers, in dynamic equilibrium with the native state, that would trigger the aggregation reaction. Accordingly, the aggregation rate is strongly dependent on the temperature, being very slow at 37°C, where the protein is 1.08 kcal·mol^{-1} more stable than at 42 °C, and extremely fast at 45 °C, where it is 1.87 kcal·mol^{-1} less stable. In fact, despite the secondary structure content of h-TTL at 45 °C is equivalent to that of the protein in the 25-42 °C range, both Tyr intrinsic fluorescence and ANS are higher, indicating partial exposure of previously hidden hydrophobic residues, a conformational feature that abrogates the lag phase of the aggregation, promoting the immediate incorporation of the protein into Th-T positive aggregates. Despite, the increase in aggregation rates of initially unfolded

amyloidogenic proteins with the temperature is a well-know phenomenon (Sabaté, Villar-Piqué, et al. 2012); the dependence of the reaction on the temperature observed in these cases is much lower than the one reported here for h-TTL. This responds to the fact that, for this domain, temperature sharply tunes the conformational stability and, therefore, controls the degree of structural fluctuations leading to the transient population of aggregation susceptible conformers. Therefore, as recently proposed for the src SH3 domain, it is likely that the structures sampled by the monomer under native conditions encode not only the structures in the fibril state but also the rate of fibril formation (Zhuravlev et al. 2014). Alternatively, it can be that a fraction of the low-populated unfolded species in equilibrium with the native ensemble might constitute a reservoir of conformations responsible for fibrillation, whose concentration would depend on the thermodynamic stability of the protein. In any case, our results indicate no global unfolding of the protein population occurs in the lag phase of the reaction, where the formation of oligomeric species is already taking place.

Most of the information we have on the aggregation of globular proteins has been obtained under strong denaturing conditions, in which polypeptide chains are significantly unfolded or populate molten globules. Despite the assays performed under these conditions have provided important mechanistic insights into amyloid fibril formation (Dobson 2004), it is also true that the protein repertoire in living organisms would hardly face these environments. Our results are, in line with recent data, obtained from structurally and sequentially unrelated protein models, which suggests that protein aggregation and subsequent amyloid formation can occur under conditions where a protein populates conformations close to its native state, but slightly destabilized (Chiti & Dobson 2009; Knowles et al. 2014). Cellular stress can suffice to promote such destabilization, increasing conformational fluctuations and/or the transient population of partially unfolded conformers, which could trigger aggregation if they exceeded their critical concentration for nucleation. At 25°C the h-TTL is stable enough to skip aggregation, but the protein already aggregates at 37°C, indicating that this side reaction can occur under mild physiological conditions. This is consistent with h-TTL having a highly amyloidogenic preformed β-strand. Local

fluctuations around this structural element would likely allow anomalous intermolecular interactions between h-TTL monomers, leading to the formation of an aggregated β-sheet structure without extensive unfolding. This is the mechanism driving the formation of amyloids by TTR, in which dissociation of the tetrameric structure results in the direct exposure to the solvent of preformed β-strands previously involved in inter-subunit contacts at the interface of the complex. It has been suggested that, upon tetramer dissociation, the TTR monomer experiences only a local unfolding transition, mainly involving the external C and D strands, with β-strands AGH and BEF retaining largely a native-like folded conformation (Lai et al. 1996). Moreover, the core structure of the fibrils have been shown to involve the AGH and BEF sheets of the monomer in a native-like conformations, in such a way that interactions between B and A strands as well as between F and H strands from adjacent molecules account for the cross-β structure of TTR fibrils (Olofsson et al. 2004). The fact that in h-TTL aggregation occurs without requiring and extensive unfolding suggests that the docking of preformed β-sheets might also account for the formation of amyloid-like structures in this structurally homologous domain. Despite its intrinsic aggregation-propensity, the h-TTL domain remains soluble at 25 °C, where the protein is energetically more stable. This suggests that for pathological proteins sharing the same mechanism for amyloid formation, small compounds able to stabilize the native state and therefore reduce the population of aggregation competent transient conformations might find therapeutic application (Bulawa et al. 2012). However, it is also true that even when they are in energetically stable conformations, these proteins are aggregation-susceptible, since, as shown here for h-TTL, the presence of small concentrations of preformed aggregates suffices to trigger their self-assembly reactions.

As shown in **Figure 10**, we predict that all TTL domains of the CPs in the M14B subfamily exhibit at least one, preformed, highly amyloidogenic β-strand, as predicted with Waltz. According to our data, this puts the respective domains at risk of aggregation. Because the formation of intracellular deposits reduces cell fitness, during the course of the evolution proteins have adopted sequential and structural strategies to escape from protein aggregation (Monsellier & Chiti 2007; de Baets et al. 2011;

Beerten et al. 2012). However, in certain cases, especially for all-β-sheet proteins, the presence of preformed amyloidogenic structural elements cannot be completely avoided since they are needed for functional purposes, like the formation of native intermolecular interfaces, as in TTR. This argues that the aggregation-prone face of h-TTL might serve to contact other molecules.

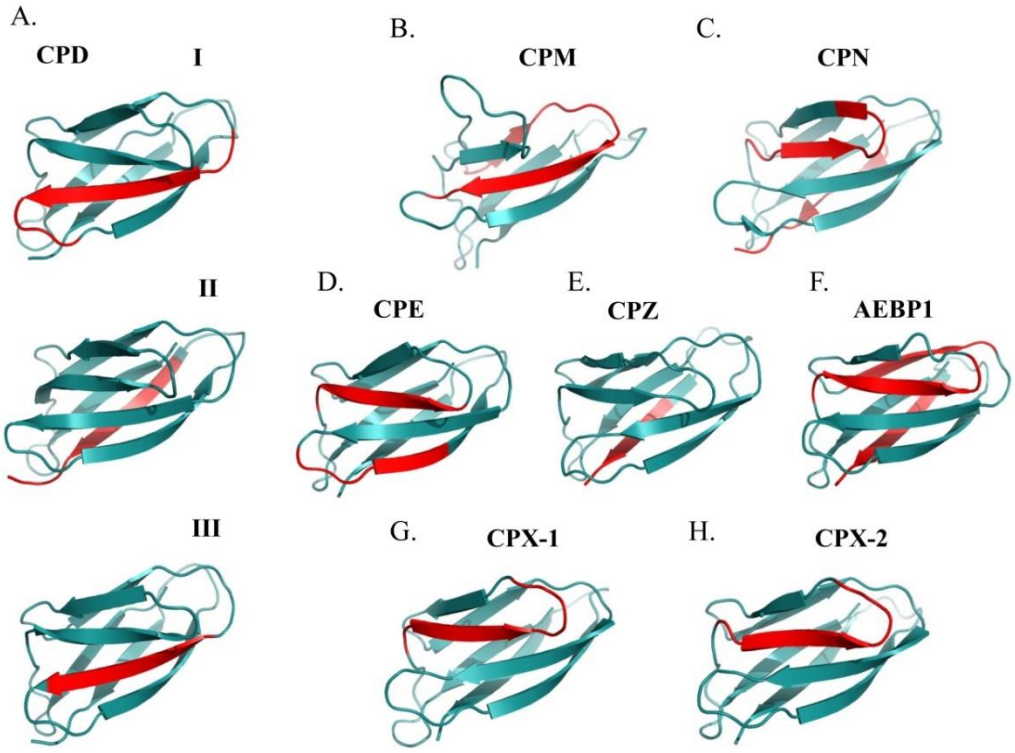


Figure 10. Aggregation-prone regions (APRs) detected in the transthyretin-like domains of the M14B subfamily members. APRs (in red) detected employing Waltz in the TTL domains of (A) CPD (CP domains I, II and III), (B) CPM, (C) CPN, (D) CPE, (E) CPZ, (F) AEBP1, (G) CPX-1 and (E) CPX-2. CPM and CPN structures correspond to PDB atomic coordinates deposited under accession codes 1UWY and 2NSM, respectively. The remaining ribbon representations correspond to models generated using the I-TASSER server.

It has been suggested that TTL domains might play a role in the binding the M14B subfamily CPs to membranes. Whereas it is true that h-TTL might bind to neutral membranes and that this interactions abrogate its aggregation, it is also true that most biological membranes display a negative charge which, due to the asymmetric distribution of the charges at the h-TTL surface would likely force the preferential orientation of the aggregation-prone face of the protein towards the membrane,

resulting in a high aggregation propensity, as shown here for negatively charged liposomes. This questions the putative membrane-binding role of TTL domains in CPs, which should serve therefore for alternative functions, yet to discover.

The aggregation mechanism described here, for the first time for a natively monomeric TTR-like protein, is being found also in a number of initially soluble globular proteins associated with protein deposition diseases and might be in fact quite generic for folds displaying preformed amyolidogenic elements in their structures, essentially β-sheets. This emphasizes the crucial role played by the protein quality machinery to preclude the aggregation of globular proteins under stress conditions that might decrease their conformational stability (Kim et al. 2013).

Chapter II

Crystal structure of the human carboxypeptidase D transthyretin-like domain solved at ultra-high resolution

CHAPTER II: CRYSTAL STRUCTURE OF THE HUMAN CARBOXYPEPTIDASE D TRANSTHYRETIN-LIKE DOMAIN SOLVED AT ULTRA-HIGH RESOLUTION

2.1 INTRODUCTION

All the members of the M14B subfamily of MCPs share a β-sandwich transthytretin-like (TTL) domain at the C-terminus, which function remains uncertain (Reznik & Fricker 2001). This domain with structural similarity to transthyretin, is also found in many other protein families, including dioxygenases, the N-terminus of prophage tail fibre proteins, Staphylococcus aureus collagen-binding surface proteins, C-terminus of cell surface antigens and polysaccharide lyases, among others (Knoot et al. 2015; Jensen et al. 2010; Hall et al. 2014; Young et al. 2014).

Recently, it has been demonstrated that one of the TTL domains belonging to the first catalytic domain of human carboxypeptidase D (h-TTL) has intrinsic propensity to aggregate into amyloid-like conformations under close to physiological conditions, rendering this protein to a privileged model for study of protein aggregation (Garcia-Pardo et al. 2014). Even though, all the structural information we have about TTL domains is based on the three-dimensional crystal structures of TTL domains produced together with its catalytic moiety, and solved with resolutions higher than 2.10 Å (Gomis-Rüth et al. 1999; Aloy et al. 2001; Keil et al. 2007; Tanco et al. 2010).

The constant advances in protein crystallography and the improvements in synchrotron radiation sources have brought biocrystallography to a context favourable to obtain structures at subatomic resolution. To date, 500 macromolecular X-ray crystal structures with resolutions higher than 1.0 Å have been deposited in the Protein Data Bank, and only 332 structures were solved below 0.95 Å.

In this study, we report a three-dimensional structure of the h-TTL domain, produced and crystallized independently from its catalytic domain with an overall resolution of 0.94 Å, which reveals its structure at subatomic resolution, and provides important information for future functional and structural insights.

2.2 EXPERIMENTAL SECTIION

2.2.1 PROTEIN EXPRESSION AND PURIFICATION

The Transthyretin-like domain (h-TTL) of human metallocarboxypeptidase D domain (residues 383-461) was produced as described previously (Garcia-Pardo et al. 2014) (see Chapter I). Briefly, the sequence of the h-TTL domain with and additional Ser residue was cloned into the pET-22B vector and then transformed into *Escherichia coli* BL21 (DE3) cells, which were then grown in 1 L of lysogeny broth (LB) medium with 50 µg.mL-1 ampicillin, at 37 ºC and 250 rpm to an OD600nm of 0.5 to 0.6. Once reached this cell density, protein expression was induced with 0.1 mM isopropyl-1-thio-β-Dgalactopyranoside (IPTG) for 16 h at 18 ºC. Then, the culture was centrifuged and the cell pellet lysed by sonication in 1/50 the initial culture volume of Tris100 mM, NaCl 150 mM buffer at pH 8.0. The h-TTL protein was purified by affinity chromatography on a Chelating SepharoseTM Fast Flow (GE Healthcare) resin. The column was equilibrated and washed gently with 100 mM Tris·HCl, 0.5 M NaCl buffer at pH 8.0 and the TTL protein was eluted in 3 column volumes (c.v.) of a 50 mM Tris·HCl, 0.15 M NaCl at pH 8.0 buffer containing 500 mM imidazole. The recombinant protein was further purified on a Superdex 75 HR 10/30 column (GE, Healthcare) and the protein buffer exchanged on a Sephadex G-25 column (GE, Healthcare) to a 20 mM phosphate, NaCl 100 mM buffer at pH 8.0. The purified protein was concentrated up to approximately 50 mg/ml-1 and stored at 4ºC until use.

2.2.2 MASS SPECTROMETRY ANALYSIS

After 50-fold dilution in milli-Q water, the purified h-TTL protein was analysed by Matrix-Assisted Laser Desorption and Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF), in linear positive mode. Samples were prepared by mixing equal volumes of a saturated solution of α-cyano-4-hydroxycinnamic acid. This protein-matrix mixture was spotted onto a MP 384 Polished Steel MALDI sample support (Bruker), evaporated to dryness and analysed in an Ultraflex MALDI-TOF mass spectrometer (Bruker Daltonics, Germany), previously calibrated with a mixture of standard peptides.

2.2.3 CRYSTALLIZATION AND DATA COLLECTION

Crystals of the h-TTL domain were obtained at 18 °C by hanging drop vapor diffusion methods. The reservoir solution contained 100 mM Bis-Tris 0, 25% PEG3350, pH 6.0. Single crystals appeared after 1 week from equal volumes of protein solution (about 50 mg/ml in 5 mM Tris-HCl pH8.0, 100 Mm NaCl) and reservoir solution. All crystals were cryo-protected in reservoir buffer containing 15% Glycerol and flash-frozen in liquid nitrogen prior to diffraction analysis (**Figure 1-A**). Diffraction data were recorded from cryo-cooled crystals (100 K) at the ALBA synchrotron in Barcelona (BL13-XALOC beamline). Data were integrated and merged using XDS (Kabsch 2010) (**Figure 1-B**), scaled, reduced, and further analysed using CCP4 (see **Table 1**).

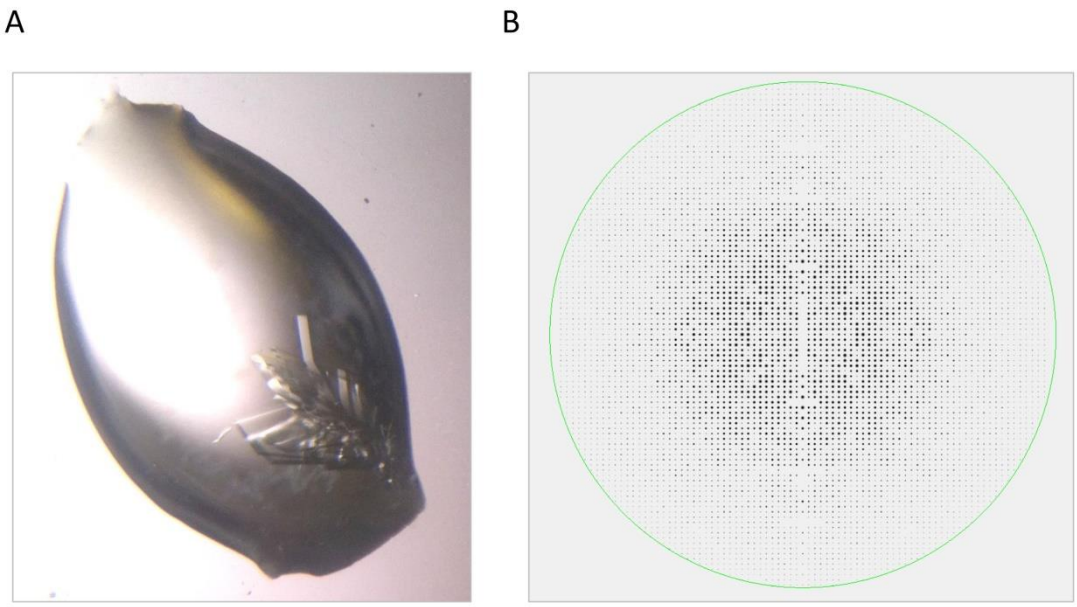


Figure 1. Crystallization of h-TTL and data collection. (A) (A) Image of h-TTL crystals appeared after one week incubation at 18 ºC. (B) Diffraction data of h-TTL, resultant from the diffraction of the crystals shown in (A). The green circle indicates the resolution limit of 0.95 Å.

TABLE 1. Data collection and refinement statistics

Data collection	**h-TTL**
Space group	P 1 21 1
Cell dimensions	
A, b, c (Å)	39.7, 46.0, 42.7
A, β, γ, (°)	90.0, 90.2, 90.0
Resolution (Å)	43.73 - 0.949
R_{merge} [a]	0.030
I/σ_1	13.7
Completeness(%)	95.5
Redundancy	2.5
Refinement	
Resolution (Å)	0.949
Nº Reflections	92883
R_{work}/R_{free}	0.153/0.175
Nº Atoms	1319
Nº aa protein	81
Water	249
R.m.s desviations	
Bond lengths (Å)	0.029
Bond angles	2.364
PDB code	4TST

[a] $R_{merge} = \sum|I_i - \langle I\rangle| / \sum I_i$, where I_i is the *i*th measurement of the intensity of an individual reflection or its symmetry-equivalent reflections and $\langle I\rangle$ is the average intensity of that reflection and its symmetry-equivalent reflections.
[b] $R_{work} = \sum||F_o| - |F_C|| / \sum|F_o|$ for all reflections and $R_{free} = \sum||F_o| - |F_C|| / \sum|F_o|$, calculated based on the 5% of data excluded from refinement.
[C] r.m.s root mean square deviation

2.2.4 STRUCTURE DETERMINATION AND REFINEMENT

The structure of h-TTL was determined from the x-ray data by molecular replacement using a previous solved structure from *Drosophyla melanogaster* CPD variant 1B short (Protein Data Bank code 3MN8) as a model. The initial electron density maps produced from molecular replacement programs were manually improved to build up complete models for h-TTL using the program COOT (Emsley et al. 2010). Model refinement was performed with Refmac (Winn et al. 2011) and Phenix (Adams et al. 2010). h-TTL showed assembling of two molecules in the asymmetric unit, and the Ramachandran analysis shows 96.85% of residues (123 residues) are in preferred regions, 2.36% of residues (3) are in allowed regions, and 0.79% of residues

(1) are in outlier regions. Refinement and data statistics are provided in **Table 1**. Structural representations were prepared with PyMOL (DeLano 2002).

2.2.5 ACCESION CODES

Coordinates and structure factors from the h-TTL structure were deposited in the Protein Data Bank with the accession code 4TST.

2.3 RESULTS

2.3.1 PROTEIN PRODUCTION AND PURIFICATION

In order to produce large amounts of protein for structural studies, a recombinant form of the h-TTL domain (corresponding to residues 383-461 in the full-length human CPD) was expressed in *Escherichia coli*. To facilitate the purification of this domain, six His residues were added to the C-terminus. The expressed protein was purified to homogeneity by using a two-step purification protocol (see above). Using this production and purification system, we produced more than 50 mg of pure h-TTL protein for each litter of cell culture (**Figure 2-A**). The recombinant h-TTL protein obtained has a mass of 9.3 kDa, determined by mass spectrometry analysis (**Figure 2-B**), consistent with the predicted theoretical mass of h-TTL.

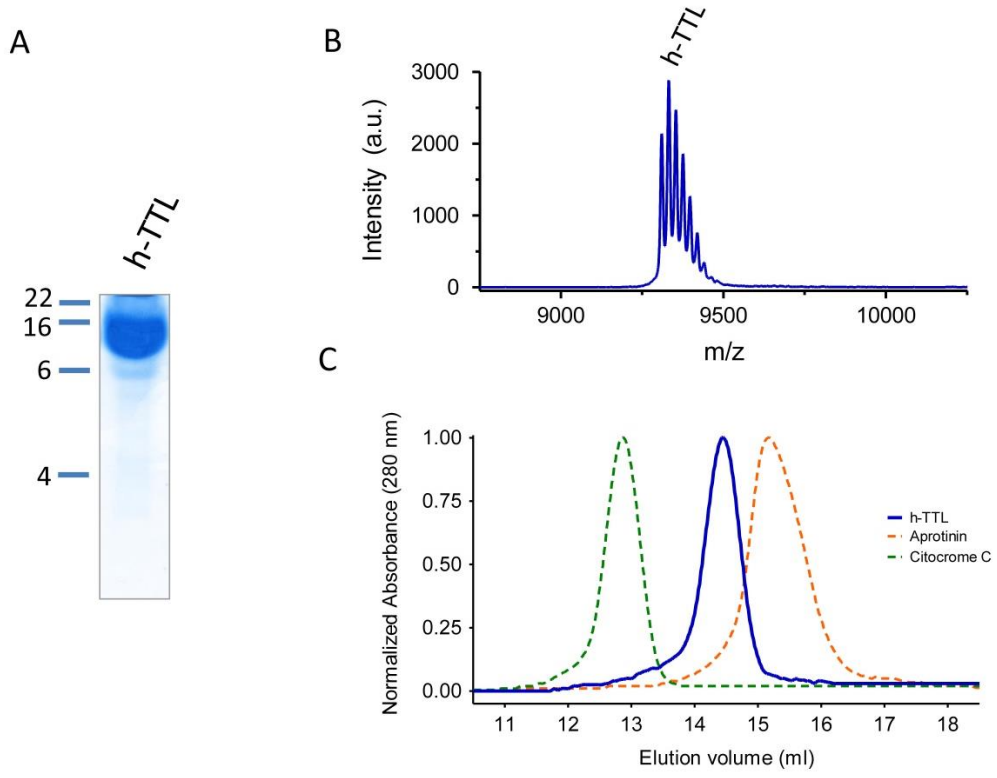


Figure 2. Recombinant production of h-TTL. (A) The purified h-TTL was visualized on a native-PAGE by coomassie staining. (B) MALDI-TOF spectrum of purified h-TTL, showing an average mass of 9.309 KDa. (C) Gel-filtration chromatography of human TTL (solid blue line). The recombinant protein elutes as a single peak with an apparent mass of 8.7 kDa, between the molecular weight markers Aprotinin (dashed orange line) and Citocrome C (dashed green line), with masses of 6.5 and 12.4 kDa, respectively.

In addition, in gel filtration chromatography, the protein elutes as a single peak, with an apparent mass of 8.7 kDa, fairly close to the theoretical and mass spectrometry determined masses, indicating that the domain remains as a monomer in solution (**Figure 2-C**). Is interesting to mention that the purified protein only can be visualized with a polyacrylamide gel electrophoresis under native conditions, since the aggressive denaturing conditions (such as the presence of SDS, or temperatures higher than 40 ºC), induce its non-reversible aggregation.

2.3.2 OVERALL CRYSTAL STRUCTURE OF THE H-TTL DOMAIN

The polypeptide chain of human h-TTL can be traced clearly in the final electron density map from Gly1 to Pro79 (corresponding to the segment Gly383-Pro461 in full-length human CPD). Due to its high-resolution different types of non-hydrogen atoms (i.e., sulfur, carbon, nitrogen and oxygen atoms) can be discriminated directly by the electron density map (see **Figure 3**).

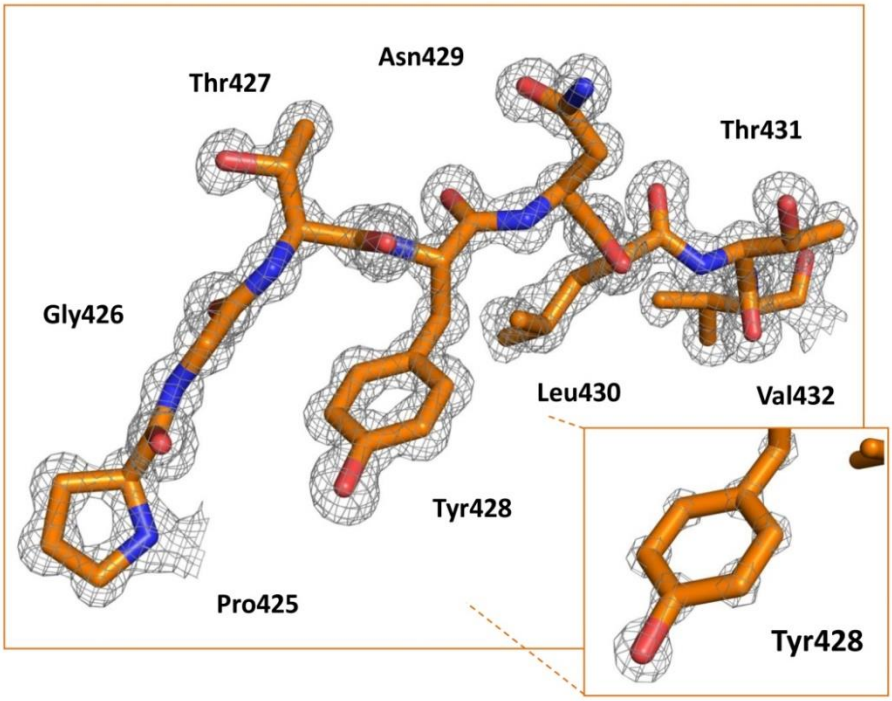


Figure 3. Representative electron density map contoured at 1.5 sigma around residues Pro425-Val432 of h-TTL. The inset shows the magnification of Tyr428 showing the electron density map contoured at 4.0 sigma. The figure was generated using PyMOL (DeLano 2002).

In the crystals, which resulted from highly concentrated protein solution (~50 mg/ml), h-TTL dimers were found in the asymmetric unit (**Figure 4**). On the contrary, under gelfiltration chromatography h-TTL at concentrations ranging from 0.5 to ~5 mg/ml elutes in a single peak corresponding to a monomeric protein (see **Figure 2-B**)

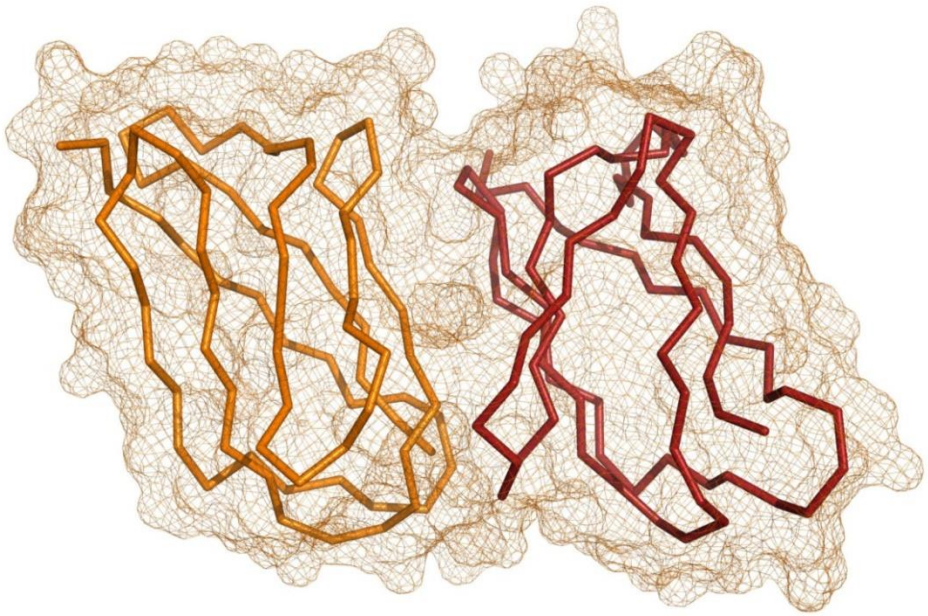


Figure 4. Structure of h-TTL dimers found in the asymmetric unit. The P1 21 1 crystal of h-TTL is made up of such dimers. Each crystalline monomer of h-TTL is shown with a different colour (orange or red) in the ribbon representation. The h-TTL surface is presented as an orange mesh.

Overall, the structure of h-TTL is rod-shaped, with the N-t and C-t located 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, β7) and four antiparallel strands (β2, β3, β5, and β6), respectively, which are glued by a hydrophobic core (**Figure 4**), similarly as described for other TTL domains of MCPs (Gomis-Rüth et al. 1999; Keil et al. 2007; Sebastián Tanco et al. 2010). These strands are arranged as two subsequent Greek-key-like elements related by a 2-fold axis perpendicular to the sandwich surface (**Figure 4**).

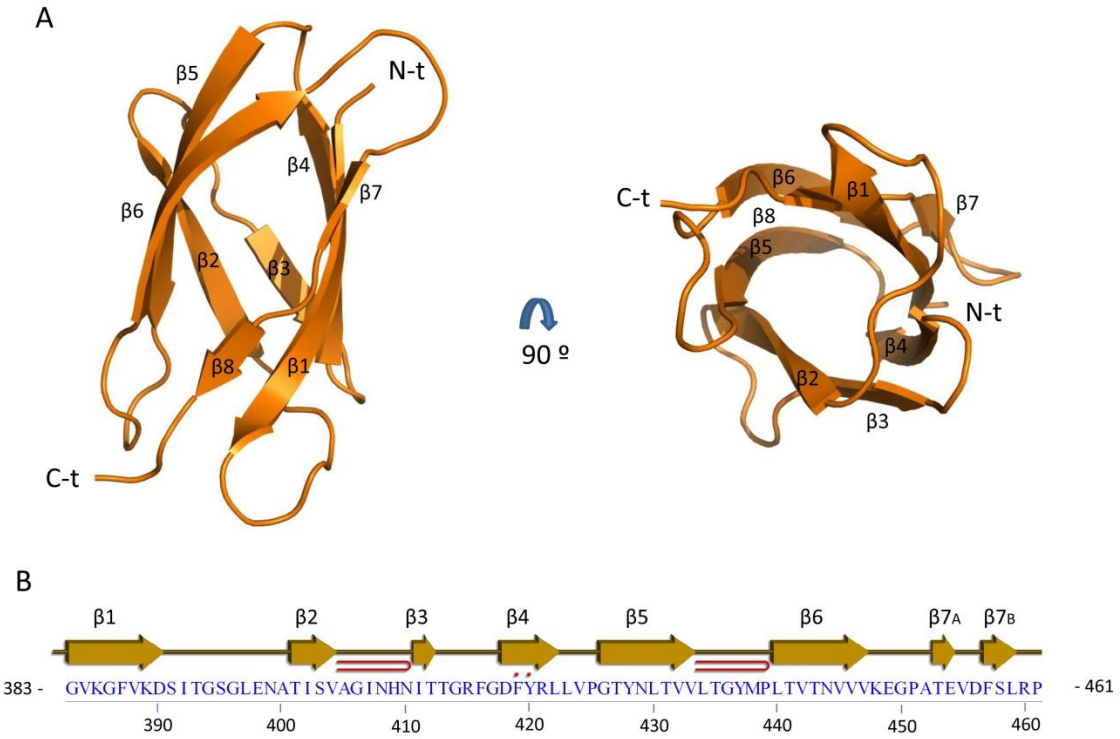


Figure 5. Structure of h-TTL (A) Ribbon representation of h-TTL (left), and the same structure rotated 90º, showing a top view (right). (B) Sequence and topology scheme of h-TTL, with β-strands indicated as arrows. The location of the two beta hairpins is indicated. Numeration corresponds to residues 383-461, according to full-length human CPD numeration (Uniprot accession code O75976).

.

The structure of this domain is maintained by a buried hydrophobic cluster running across the whole domain, made up by the side chains of Val384, Val388, Leu397, Ala400, Ile402, Ile407, Thr413, Gly414, Phe419, Leu423, Tyr428, Leu430, Val432, Leu434, Tyr437, Leu440, Val442, Val445, Thr453, Phe457, Ser458, Leu459 and Leu459.

Additionally, one glycerol molecule has been assigned in the structure based on the electron density (not shown). Their presence is chemically reasonable due to the excess of this compound in the cryoprotectant solution.

Due to the topology and strand connectivity, the h-TTL resembles the plasma protein transthyretin (also known as prealbumin). However, transthyretin contains an additional C-terminal β-strand that is absent in h-TTL (Damas & Saraiva 2000).

2.3.3 INSIGHTS INTO THE STRUCTURE OF THE AGGREGATION-PRONE REGION

In a previous study, it was suggested that human TTL displays a highly amyloidogenic short sequence stretch, encompassing residues Gly426 to Met438 (according to full-length human CPD numeration), which overlaps with the β-sheet 5 (**Figure 6-A)**, based on WALTZ predictions (see chapter I (Garcia-Pardo et al. 2014)). Local fluctuations around this structural element would likely allow anomalous intermolecular interactions between h-TTL monomers leading to amyloid formation without an extensive unfolding.

In the crystal h-TTL structure, this aggregation prone region (APR) is located within the fifth β-sheet, similarly as described previously using a modelled structure (Garcia-Pardo et al. 2014). A detailed analysis shows that this sequence comprises the peptide GTYNLTVVLTGYM. The first eight residues within the APR sequence are involved in the β-sheet structure formation, while the last five shape the second β-hairpin. Some of these residues have its side chains buried in the hydrophobic core (Tyr428, Leu430, Val432, Leu434, Tyr437), whereas the rest have its side chains exposed to the solvent (Gly426, Thr427, Asn429, Thr431, Val433, Thr435, Gly436) (**Figure6-B**) .

Gel filtration chromatography indicated that the protein is monomeric throughout the concentration range assayed (0.5 to ~5 mg/mL, as injected into the FPLC system). In the crystals, which resulted from highly concentrated protein solution, a dimeric structure was found in the asymmetric unit (see above). Inspection of the dimeric interface in the structure reveals that it measures ~ 350 Å^2. In h-TTL, the dimer is not symmetric, so that the different structural segments of each monomer are involved in complex formation. Both structures are stabilized by three hydrogen bonds (Asn408-Asn408, Arg421-Asn408 and Asp418-Thr435) and up to 26 non-bonded contacts, mainly stablished between hydrophobic residues. Some of these contacts are stablished in the interface between residues within the APR sequence (form one of the monomers) and residues from the adjacent h-TTL structure (**Figure 6-C**). Thus, the residue Thr431 stablishes a hydrogen bond with an Asp418 located in the opposite molecule. Similarly, Val433 and Thr435 contact a Tyr420 residue, within the proximal monomer. An additional contact was observed, it is stablished between Thr435 and the Leu422 residue belonging to the proximal h-TTL structure.

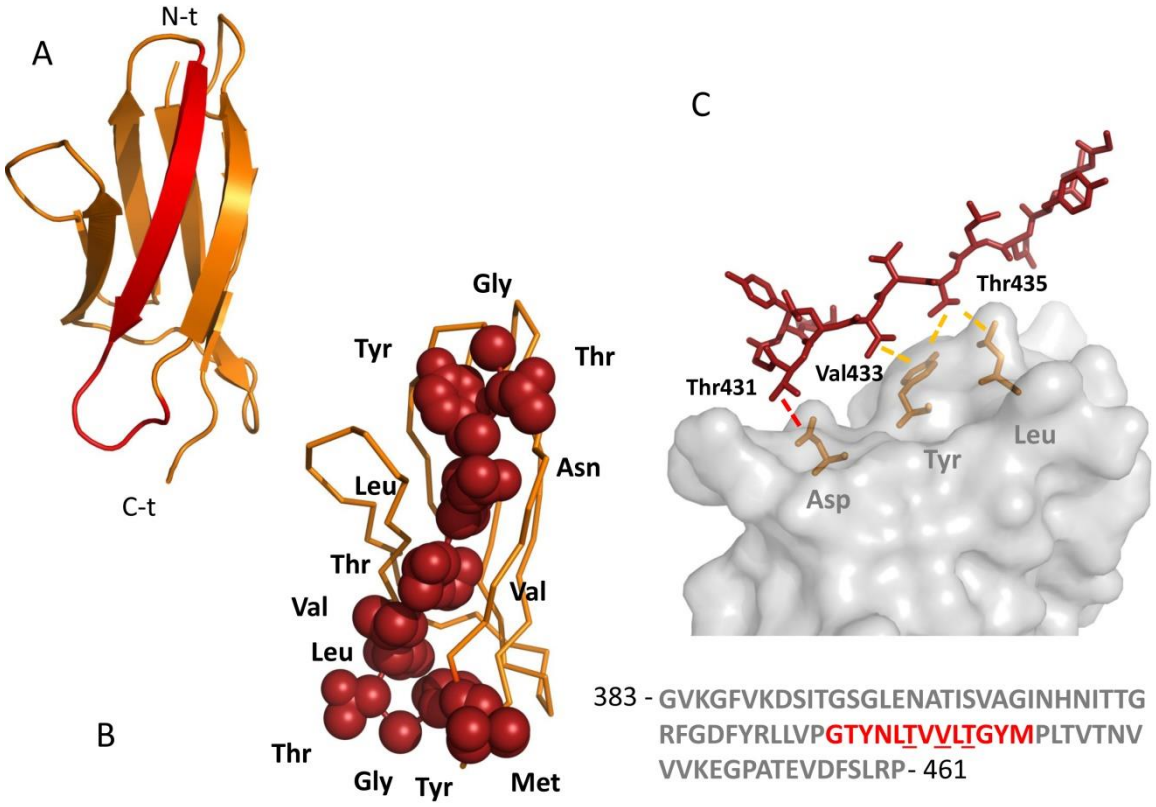


Figure 6. Structure of the aggregation prone region of h-TTL (A) Structural location of the APR found in h-TTL. (B) Ribbon diagram of the h-TTL, showing the side chains of the thirteen residues that conforms the APR peptide (side chains are shown as red spheres). Some of the residues within the APR region have its side chains buried in the hydrophobic core and some others expose its side chains to the h-TTL surface. (C) Stick model of the h-TTL APR peptide. The Hypothetic dimer formation occurs through hydrogen bonds (dashed red line) and non-bonded interactions (dashed yellow lines). In the figure are shown some of these interactions that involve residues within the APR sequence and the adjacent h-TTL monomer. Below, the residues that conform the APR are shown in red within the h-TTL sequence.

However, there is no evidence that the h-TTL forms a dimeric structure in vivo, protein seems to be monomeric in solution and the dimer could be an artifact induced by the crystallization process. Moreover, we do not know whether the formation of a dimer could have significance, since h-TTL has the ability to form amyloid structures, probably starting from a population of oligomeric species similar to that dimeric form observed in the crystal structure.

2.4 DISCUSSION

In this work, we report for the first time the ultra-high resolution crystal structure of a TTL domain belonging to the first catalytic domain of human CPD, produced and crystallized independently from its catalytic moiety. Such structure clearly shows the location of all the structural features that shape the h-TTL fold.

For crystallization, the protein was concentrated over 50 mg/ml without visible aggregation, suggesting that h-TTL is extremely soluble at temperatures between 4 and 18 ºC. This finding is in agreement with previous reports, which found that h-TTL is stable in solution at temperatures below 37 ºC (Garcia-Pardo et al. 2014).

The final resolution of the h-TTL structure was 0.949 Å. It is interesting to mention that this is crystal structure solved with higher resolution, among all the structures of MCPs (or parts of MCPs) known. Furthermore, to date is the crystal structure with higher resolution solved using the light from the ALBA synchrotron in Barcelona. It is known that some structural features in proteins are very difficult to determine accurately except in those refined high-resolution structures (e.g. dihedral angles of disulfides). For this, the crystal structure reported here represents a unique tool for future studies regarding the h-TTL domain (such as the study of its folding and its mechanism of aggregation). As example, based on this structural information, we have identified different residues that might form the hydrophobic core of h-TTL, stablishing the bases to characterize the kinetics of folding and unfolding of this domain (unpublished data). Understanding of the rules governing protein core evolution and the possible restrictions imposed by folding requirements is not only an intellectual challenge in the field of protein design but also of fundamental importance for understanding the mechanism that underlies protein aggregation.

Another goal of the present study is the structural analysis of the aggregation prone region sequence in the h-TTL crystal structure. This segment made up by thirteen residues is located in the fifth β-sheet and some of its residues have its side chains buried in the hydrophobic core of the protein. This suggests that some of these residues might server for folding or for the maintenance of the h-TTL structure. This finding is in agreement with previous reports, which suggested that the presence of

preformed structural elements cannot be completely avoided during protein evolution, because they are needed for functional purposes (Castillo & Ventura 2009). Moreover, some residues within the APR sequence have its side chains exposed to the solvent. Three of these residues stablish contacts with the adjacent h-TTL monomer, suggesting a possible role in the h-TTL oligomerization during the aggregation process.

The information collected or derived in the present study might facilitate the understanding of the differential biological roles of the TTL domains found in MCPs, as well as the study of its mechanism of folding and aggregation. This would be an interesting structure model to experimentally analyse the properties and roles of these domains, and to expand our knowledge about the intrinsic mechanism that drive protein aggregation in living organisms. In addition, the structure described here provides an important tool for the computational-based modelling of other TTL domains from other M14B MCPs.

Chapter III

Substrate specificity of human metallocarboxypeptidase D: comparison of the two active carboxypeptidase domains

CHAPTER III: SUSBSTRATE SPECIFICITY OF HUMAN METALLOCARBOXYPEPTIDASE D: COMPARISON OF THE TWO ACTIVE CARBOXYPEPTIDASE DOMAINS

3.1 INTRODUCTION

Metallocarboxypeptidases (MCPs) are zinc-dependent enzymes that cleave single amino acids from the C termini of peptides and proteins (Arolas et al. 2007). The first MCP to be identified was carboxypeptidase A1 (CPA1), a pancreatic enzyme that removes C-terminal hydrophobic residues. In the ensuing decades since the discovery of CPA1, dozens of additional MCPs have been identified based on sequence and/or structural similarity; these are classified as M14 family enzymes by MEROPS (N.D. Rawlings et al. 2014). All mammalian MCPs show marked preference for the residue in the P1' site, and are selective for either hydrophobic residues, basic residues, or acidic residues (Arolas et al. 2007; Lyons & Fricker 2011). Several M14 gene family members are not active towards conventional substrates, although it remains possible that they hydrolyze atypical substrates (Reznik & Fricker 2001; Hourdou et al. 1993). Most of the MCPs show differential patterns of expression. The combination of substrate specificity and distribution are responsible for the diverse functions of these proteins which ranges from digestion of food (for carboxypeptidases A1, A2, and B1) to the production of neuropeptides and peptide hormones (carboxypeptidase E) and the selective processing of tubulin (cytosolic carboxypeptidases) (Arolas et al. 2007; Berezniuk et al. 2012).

Carboxypeptidase D (CPD) belongs to the M14B subfamily of MCPs and was originally discovered in duck hepatocytes, as a 180-kDa membrane-bound glycoprotein (named as gp180) that has the ability to bind duck hepatitis B virus particles (Eng 1998; Glebe & Urban 2007; Kuroki et al. 1994). In most mammals, birds and in *Drosophila melanogaster,* CPD has a broad tissue distribution and has a similar structural architecture. CPD contains three copies of a ~400 amino acid-long segment with sequence homology to carboxypeptidase E and other members of the M14B subfamily.

The critical catalytic and substrate-binding residues found in all active members of the M14 gene family are conserved in CPD domains I and II but not in domain III (Figure 1). CPD also contains a transmembrane region of about 20 residues and a cytosolic tail of 60 residues (**Figure 1**). CPD is primarily localized in the trans-Golgi network (TGN), and cycles between the TGN and the cell surface. Sequences within the C-terminal tail bind cytosolic proteins and mediate TGN retention and intracellular trafficking (Kalinina et al. 2002).

The three CP domains of CPD are localized to the lumen of the secretory pathway, the cell surface, and the interior of endocytic vesicles. Based on the broad tissue distribution of CPD, its subcellular localization, and ability to cleave basic residues from several peptides that have been tested, CPD is thought to play a role in the further processing of proteins and peptides that are initially cleaved by furin and furin-like enzymes, within the secretory pathway and/or on the cell surface and within the endocytic pathway. The physiological role of the three CPD domains remains unclear. Previous studies testing a small number of synthetic substrates found that duck CPD domains I and II are enzymatically active, but the third domain of CPD is catalytically inactive toward traditional MCPs substrates. Also, CPD domain I is optimally active at neutral pH and cleaves a peptide with C-terminal arginine more efficiently than a similar peptide with C-terminal lysine. By contrast, the CPD domain II is optimally active at more acidic pH and cleaves a peptide with C-terminal lysine more efficiently than a peptide with C-terminal arginine (Eng 1998; Novikova et al. 1999). Sidyelyeva *et al.* investigated the function of the various CP domains through the creation of flies expressing specific forms of CPD in the svrPG33 mutant (Sidyelyeva et al. 2006). All mutants containing an active CP domain rescued the lethality with varying degrees, in all cases requiring the presence of inactive CPD domain III for full viability. Transgenic flies expressing active CPD domain I or domain II showed similar behaviours to each other and to the viable *svr* mutants, suggesting redundant functions in terms of processing peptides involved in viability and in behaviours like cold and ethanol sensitivity, as well as long-term memory.

The main objective of the present study was to investigate the substrate specificity of the full active human CPD and characterize its individual catalytic domains I and II. To gain a better understanding of the enzymatic properties of human CPD and its individual domains we used a variety of approaches. For this purpose we generated catalytically inactive single point mutants for the individual active domains I and II (E350Q, E762Q, respectively) as well as a double mutant (**Figure 1**). We characterized the enzymatic properties of the full active human CPD and the individual catalytically active domains, using a wide range of substrates and quantitative peptidomics approaches. These studies found that CPD domain III is inactive towards all peptides examined, while domains I and II have complimentary activities that provide for a wider ability to cleave peptides/proteins with basic C-terminal residues within the secretory pathway, on the cell surface, and within the endocytic pathway.

3.2 EXPERIMENTAL SECTION

3.2.1 RECOMBINANT PROTEIN PRODUCTION AND PURIFICATION

Human carboxypeptidase D (residues 32-1298) was cloned into the pTriEx^{TM}-7 expression vector (Merck Millipore) encoding for mouse IgM secretion signal sequence and an N-terminal Strep-Tag$^{®}$II fusion protein. CPD E270Q mutants (according to bovine CPA numbering) named here as E350Q, E762Q (for single mutants) and E350Q/E762Q (for the double mutant) were generated by PCR-driven overlap extension (Heckman & Pease 2007). For protein production in mammalian cells, DNA transfections of CPD and the CPD E270Q mutants were carried out using 25-kDa polyethylenimine (PEI) (Polysciences) in a ratio of 1:3 (µg DNA/µg PEI), as described previously (Tanco, Tort, et al. 2015). Briefly, HEK293F cells were diluted to a cell density of about 0.5x10^{6} cells/ml, grown for 24 h and then transfected with 1 µg DNA per ml of culture for 7 days. For protein purification, the culture supernatant was equilibrated with 30% ammonium sulphate, centrifuged, filtered through 0.2 µm filters and bound to a hydrophobic chromatography column (Toyopearl Butyl-650M, Tosoh Bioscience). Protein elution was carried out with a decreasing gradient of ammonium sulphate (30 to 0%) in a 100 mM Tris-HCl, pH 7.5 buffer. The eluted fractions containing CPD (as analysed by SDS-PAGE) were pooled and loaded onto a Strep-Tactin affinity column (IBA GmbH) equilibrated with binding buffer (100 mM Tris-HCl, pH 8.0, 150 mM NaCl), washed with 2 column volumes of binding buffer and eluted with the elution buffer (*i.e,,*binding buffer containing 2.5 mM *d*-desthiobiotin; IBA GmbH). Eluted fractions were analysed by SDS-PAGE, and the purest fractions were pooled and loaded onto a HiLoad Superdex 75 26/60 column (GE Healthcare) previously equilibrated with a 50 mM Tris-HCL, pH 7.5, 150 mM NaCl buffer. The purified proteins were flash frozen at a concentration of approximately 0.5 mg/ml and stored at -80 ºC.

3.2.2 CELL CULTURE

HEK293T cells (ATCC CRL-3216) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with GlutaMAX and 10% (v/v) fetal bovine serum (Invitrogen, Inc.) at 37°C, 10% CO_2 and 95% humidity. HEK293F cells (ATCC CRL-3216) were grown in FreeStyle 293 expression medium (Invitrogen, Inc.) in flasks on a rotary shaker (120 rpm) at 37°C, 8% CO_2 and 70% humidity.

3.2.3 HEK293T BORTEZOMIB TREATMENT AND PEPTIDE EXTRACTION

HEK293T cells were seeded in 24 150-mm cell culture plates and after growing up to 70% confluence, cells were treated with fresh media containing 0.5 µM bortezomib for 1 h. Following incubation, cells were washed three times with cold Dulbecco's phosphate-buffered saline (DPBS, Invitrogen), immediately scraped and centrifuged at 8000xg for 5 min. For peptide extraction, the cell pellet was resuspended in 1 ml of 80ºC water and the mixture was incubated for 20 min in an 80ºC water bath. Samples were cooled, transferred to 2 ml low retention microfuge tubes and centrifuged at 13,000xg for 20 min. Soluble fractions containing HEK293T cell peptides were stored at –70ºC overnight. Samples were centrifuged again and the supernatant of each tube was collected and concentrated in a vacuum centrifuge to a volume of 1.5 ml. Finally, samples were cooled and acidified with 0.1 M HCl to get a final concentration of 10 mM HCl. After 15 minutes of incubation, samples were centrifuged at 13000xg for 40 min at 4ºC and the supernatants stored at -80ºC until labelling.

3.2.4 PREPARATION OF TRYPTIC PEPTIDE LIBRARIES

Tryptic peptide libraries were generated by digesting five different proteins with trypsin. Trypsin is an endoprotease that cleaves C-terminal to arginine or lysine residues, producing peptides containing basic residues on their C-termini (Olsen et al. 2004). The proteins bovine serum albumin (BSA, Sigma-Aldrich), bovine thryroglobulin (Sigma-Aldrich), bovine α-lactalbumin (Sigma-Aldrich), human α-hemoglobin (Sigma-Aldrich) and human β-hemoglobin (Sigma-Aldrich) were digested separately and then

pooled in order to have a tryptic peptide library with a final peptide concentration of about 500 µM. To prepare the tryptic peptide libraries, 5 nmoles of each protein were digested for 16 h at 37ºC in a 20 mM borate, pH 7.6 buffer using sequencing-grade trypsin (Promega) at an enzyme/substrate ratio of 1/100 (w/w). The efficiency of protein digestion was checked by SDS-PAGE. To stop the proteolytic digestion, 0.1 M HCl was added to get a final concentration of…. All the reactions were combined and centrifuged at 13,000xg for 45 min at 4ºC. The supernatant containing tryptic peptides was filtered through a 10-kDa centrifugal filter device (Amicon, Merck Millipore), aliquoted and stored at -80ºC until use.

3.2.5 KINETIC MEASUREMENTS USING FLUORESCENT SUBSTRATES

Carboxypeptidase activity was assayed with the fluorescent substrates dansyl-Phe-Ala-Arg, dansyl-Phe-Gly-Arg or dansyl-Phe-Pro-Arg. To perform these assays, 0.2 mM of each substrate was incubated in a 100 µl reaction with variable amounts of enzyme in a 0.1 M Tris-acetate, 0.1 M NaCl buffer (pH 6.5) for 60 min at 37ºC. Reactions were stopped by adding 50 µl of 0.5 M HCl, 1 ml of chloroform was added to each reaction, tubes were mixed gently and centrifuged for 2 min at 300 x g. After centrifugation, 0.5 ml of the chloroform phase was transferred to new tubes and dried over night at 25 ºC in a fume hood. Dried samples were resuspended with 200 µl of PBS containing 0.1 % of Triton X-100. The amount of product generated was determined by measuring the fluorescence emission of samples at 500 nm upon excitation at 350 nm using a 96-well plate spectrofluorometer. For kinetic analysis, purified enzymes were assayed at different substrate concentrations (6.25, 12.5, 25, 50, 100, 200, 300 and 600 µM). In all cases, a maximum of 20% of the substrate was hydrolyzed. Kinetic parameters were determined by fitting the obtained data for each enzyme to the Michaelis-Menten equation (y= (V_{max} x X) /(K_m + X)) using GraphPad Prism software (Motulsky HJ n.d.), where X is the substrate concentration, V_{max} is the maximum enzyme velocity and K_m is the Michaelis-Menten constant. The pH optimum of purified enzymes was determined with 0.2 mM dansyl-Phe-Ala-Arg in 0.1 mM Tris-acetate, 150 mM NaCl buffer at the indicated pH.

3.2.6 PEPTIDOMICS

Quantitative peptidomics experiments were performed as described previously (Sebastian Tanco et al. 2010; Lyons & Fricker 2010), with slight modifications. We used the peptide library generated using trypsin or by peptide extraction from HEK293T cells, as described above. The peptide mix was incubated for 16 h at 37ºC in 100 mM borate (pH 6.5), 100 mM NaCl buffer with different amounts (0, 1, 10 and 100 nM) of purified rhCPD. In a second round of experiments, peptides from the peptide library generated with trypsin were incubated with rhCPD and with the CPD single point mutants (E350Q, E762Q and E350Q/E762Q) at a concentration of 100 nM for 16 hours at 37 ºC. After incubation, reactions were quenched and peptides were labeled using 4-trimethylammoniumbutyrate isotopic tags activated with N-hydroxysuccinimide (TMAB-NHS) containing either all hydrogen (D0), 3 deuteriums (D3), 6 deuteriums (D6), 9 deuteriums (D9), or 9 deuteriums and three atoms of ^{13}C (D12). The labels were dissolved in DMSO to a concentration of 0.4 mg/µl and 5 mg of label was used per sample. At the start of the experiment, pH of the sample was adjusted to 9.5 with 1M NaOH. Labeling was performed over 8 rounds; 1.6 µl of the label was added to the extract every 20 min. pH was measured between each round and if necessary, brought back to 9.5, only for the first five rounds. After labeling, 30 µl of 2.5 M glycine was added to quench any unreacted label. Labeled samples for a single experiment were pooled, filtered through a10-kDa centrifugal filter device and 30 µl of 2M hydroxylamine was added to hydrolyze any labeled tyrosines. This was done to ensure that only N-terminal amines and lysine side-chain amines of peptides are TMAB-labeled and not tyrosines. Samples were desalted through C-18 spin columns (Thermo-Scientific) by following the manufacturer's instructions. Peptides were eluted using 160 µl of 0.5% TFA and 70% acetonitrile, freeze-dried in a vacuum centrifuge, and subjected to liquid chromatography and mass spectrometry as described (Sebastian Tanco et al. 2010; Lyons & Fricker 2010). Identifications were performed using Mascot software (Matrix Science). The MS spectra were manually examined for peak sets reflecting peptides with the various isotopic forms of the TMAB-NHS tags. Identifications were rejected unless 80% or more of the major fragments from the MS/MS matched predicted b- or y-series fragments (with a minimum of five matches).

Additional criteria followed to consider the matches includes the coincidence with the parent mass within 50 ppm of the theoretical mass, an expected charge equal to basic residues plus the N-terminus and a correct number of isotopic tags incorporated considering the number of free amines present in the peptide. Peptidomics analyses were performed at least in triplicate.

3.2.7 SEQUENCE ALIGNMENT AND THREE-DIMENSIONAL MODELING

The amino acid sequences of human CPM and CPD were obtained from the UniProt database (accession P14384 and O75976, respectively). A sequence alignment between CPM and CPD domains I, II and III was generated using ClustalW2 (Larkin et al. 2007) from the EMBL-EBI. Three-dimensional structures of CPD domain I, II and III were constructed by using the automated I-TASSER on-line server (Zhang 2008a). Models with the best C-score, based on the significance of threading template alignments and the convergence parameters, were selected. After I-TASSER models were built automatically, PyMOL (DeLano 2002) was used for generation of figures and visual inspection of the models.

3.3 RESULTS

3.3.1 RECOMBINANT PROTEIN PRODUCTION AND PURIFICATION

Several expression systems were explored to produce active recombinant CPD. *Pichia pastoris* was previously used to produce CPD domain I and CPD domain II, as well as other members of the M14 MCP family (Ventura et al. 1999; Sebastian Tanco et al. 2010; Sebastián Tanco et al. 2010; Gomis-Rüth et al. 1999). However, human CPD containing all three domains did not express well in this system, possibly due to structural complexity and/or the numerous post-translational modifications like N-linked glycosylation. Previous studies obtained high levels of full-length CPD using the baculovirus expression system (Eng 1998; Novikova et al. 1999). Nonetheless, there are fundamental differences in the glycoprotein processing pathways of insects and higher eukaryotes, which might lead to structural and functional differences between native and recombinantly produced proteins. Recently, mammalian-based expression systems have emerged as promising systems to obtain high levels of complex mammalian proteins by transient or stable transfection (Vink et al. 2014; Portolano et al. 2014). For this reason, here we cloned the human carboxypeptidase D without its C-terminal transmembrane domain and cytosolic tail (residues 32-1298, named here as rhCPD) into the pTriExTM-7 expression vector that encodes an IgM secretion signal sequence and an N-terminal Strep-Tag II fusion protein (**Figure 1**). The CPD-pTrieX-7 construct was transfected into mammalian HEK293F cells for transient expression, by using polyethylenimine (PEI) as transfecting agent. The secreted recombinant protein showed a major band with a molecular weight of about 170 kDa by western blot analysis using an anti-strep tag II antibody (data not shown). Maximum levels of soluble rhCPD in the medium were detected 7 days post-transfection.

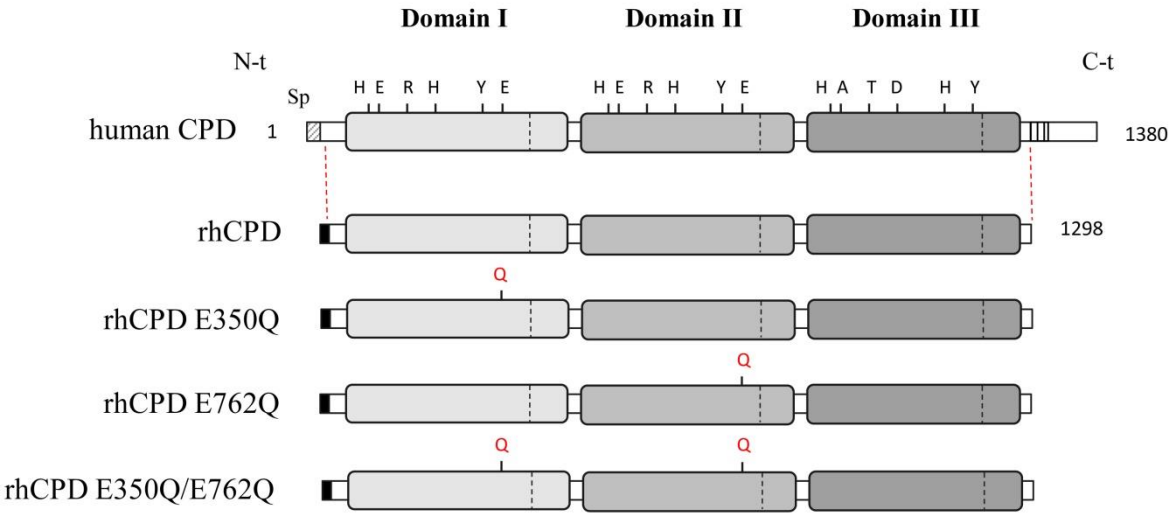


Figure 1. **Linear representation of full-length human CPD and recombinant forms showing the location of single point mutations.** The positions indicated in human CPD correspond to key residues essential for the catalytic mechanism: His69, Glu72, Arg145, His198, Tyr248, and Glu270 (according to the bCPA numbering). Recombinant proteins correspond to C-terminal truncated forms of human CPD, which lack the C-terminal transmembrane anchor and the highly conserved cytoplasmic tail. The mutations (i.e., Glu to Gln) performed to generate single point mutants for the CPD domain I (E350Q), domain II (E762Q) and a double mutant (E350Q/E372Q) are indicated.

A variety of conditions were tested to purify rhCPD from the medium, and optimal results were obtained following a three-step chromatography protocol. The first chromatographic step was a hydrophobic interaction chromatography. The eluates containing the rhCPD protein from the first purification step were pooled and purified over an anti-Strep tag affinity resin and, further fractionated on a size exclusion column (**Figure 2-A**). The purified rhCPD protein showed a major band of about 170 kDa when analyzed on SDS–PAGE by Coomassie blue staining (**Figure 2-B**), consistent with the predicted size. To characterize the substrate specificity of the individual catalytic domains we generated constructs with single point mutations in the domains I and II, as well as a double mutant (named here as E350Q, E762Q and E350Q/E762Q, respectively). All mutants were expressed and purified following the same protocol optimized for rhCPD (**Figure 2-B**). In gel filtration chromatography, all mutants eluted in a single peak with comparable apparent masses (data not shown), which indicates that the single point mutations do not have a major impact on its protein structure.

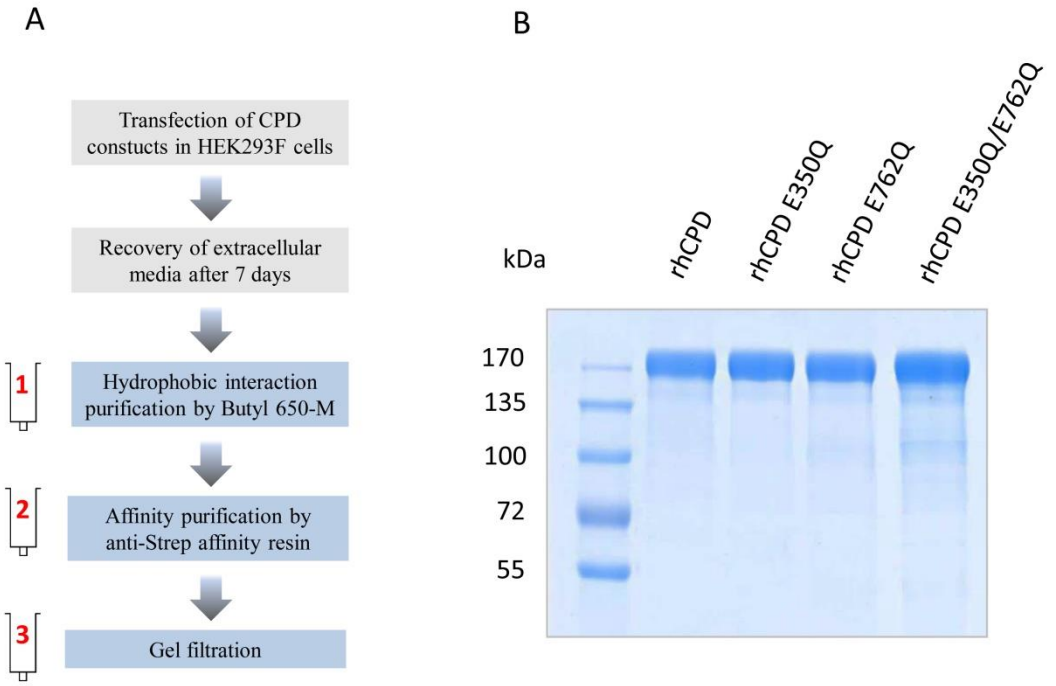


Figure 2. Expression and purification of CPD. (A) Schematic diagram of the strategy for expression and purification of rhCPD and mutants. Protein expression was performed by transient transfection of suspension-growing HEK293F cells. Extracellular medium was collected after 7 days incubation, followed by purification of the recombinant proteins in three steps; (1) hydrophobic interaction chromatography using a Butyl 650-M, (2) affinity chromatography using anti-strep tag resin, and (3) gel filtration chromatography. (B) Coomassie- stained SDS–PAGE showing the purity of recombinant CPD proteins.

3.3.2 EFFECT OF PH ON CPD ACTIVITY

The influence of pH on rhCPD and rhCPD single point mutants were examined using the fluorescent substrate dansyl-Phe-Ala-Arg. The construct E350Q shows optimal activity at pH 6.5, with activity >50% of the maximum over the range 5.0–7.0, while the construct E762Q has optimal activity at pH 7.0 and with >50% maximal activity over the range 6.5–7.5 (**Figure 3-B**). The pH optimum of the human protein with both domains I and II active is 6.5, with >50% maximal activity from 5.0–7.5 (**Figure 3-A**). The double mutant E350Q/ E762Q had no detectable enzyme activity at any of the pH values examined (**Figure 3-A**). These results with human CPD are similar to previous studies with duck CPD domains I and II, indicating that this property has been conserved through evolution (Novikova et al. 1999; Eng 1998).

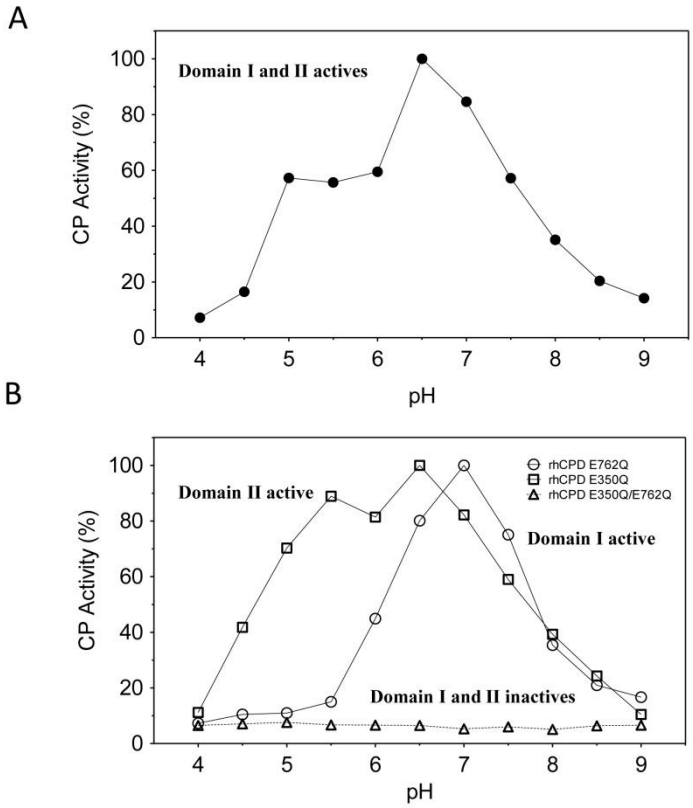


Figure 3. Effect of pH on the activity of different recombinant CPD forms. (A) Effect of pH on rhCPD and (B) single point mutants using 200 µM dansyl-Phe-Ala-Arg in a Tris-acetate buffer at the indicated pH for 60 min at 37ºC. The activity represented is the average of three independent measures with less than a 10% of variation. CP activity is represented as a percentage of the maximal activity, observed at optimal pH.

3.3.3 ACTIVITY OF CPD AND CPD SINGLE POINT MUTANTS AGAINST FLUORESCENT SUBSTRATES

The enzymatic activities of the rhCPD were tested against three fluorescent synthetic substrates. To compare substrates, different amounts of the purified enzyme were incubated with 200 µM of each dansylated peptide and the relative amount of product determined. Among the substrates evaluated, dansyl-Phe-Ala-Arg was the peptide most rapidly cleaved by rhCPD (**Supplemental Figure 1**); at enzyme concentrations where 70% of dansyl-Phe-Ala-Arg was cleaved, less than 5% of dansyl-Phe-Gly-Arg was hydrolyzed by the enzyme. No activity was detected towards the substrate with a Pro residue in penultimate position (*i.e.,* dansyl-Phe-Pro-Arg). We determined the kinetic parameters for dansyl-Phe-Ala-Arg for the full active rhCPD and

the single point mutants E350Q and E762Q (**Table 1)**. These studies were conducted at pH 6.5 because all three proteins showed ≥80% maximal activity at this pH value. The k_{cat} value of the E350Q mutant with only domain II active (7.0 ± 0.9 s^{-1}) is comparable to the k_{cat} of the E762Q mutant with domain I active (8.5 ± 0.5 s^{-1}) and both are smaller than the k_{cat} for rhCPD with both active domains (12.5 ± 0.5 s^{-1}). The lowest K_m value was obtained for the full active rhCPD, which showed a K_m of 152.7± 15.9 µM, while single point mutants E762Q and E350Q showed higher K_m values of 319.3 ± 37.1 µM and 843.7 ± 37.1 µM, respectively. The k_{cat}/k_m value for rhCPD was approximately 3 times and 10 times higher in comparison with values obtained for the E762Q and E350Q single point mutants, respectively (**Table 1**).

Table 1. Kinetic constants for hydrolysis of dansyl-Phe-Ala-Arg by rhCPD and rhCPD single point mutants

Substrate	**rhCPD**	**Active domain I** rhCPD (E762Q)	**Active domain II** rhCPD (E350Q)	**Double mutant** rhCPD (E762Q /E372Q)
Dansyl-Phe-Ala-Arg				
K_m (µM)	152.7 ± 15.9	319.3 ± 37.1	843.7 ± 139.1	ND
K_{cat} (s^{-1})	12.5 ± 0.5	8.5 ± 0.5	7.0 ± 0.9	ND
K_{cat}/K_M ($\mu M^{-1} S^{-1}$)	0.082 ± 0.012	0.027 ± 0.003	0.008 ± 0.001	ND

ND, not detectable

3.3.4 CHARACTERIZATION OF THE SUBSTRATE SPECIFICITY OF FULL ACTIVE HUMAN CPD BY QUANTITATIVE PEPTIDOMICS

The results obtained using synthetic substrates is useful, but is limited by the avaibility of substrates. To gain more information about the substrate specificity with a wide range of peptides, we applied several quantitative peptidomics approaches to address the substrate specificity of human CPD and unravel the contributions of each individual active domain. Quantitative peptidomics allows the study of proteolytic activity by analysing a large number of peptides substrates in a single experiment, thereby providing a more complete understanding of MCP substrate specificities (Sebastian Tanco et al. 2010; Lyons & Fricker 2010; Lyons & Fricker 2012; Tanco et al. 2013). First, to characterize the substrate specificity of the full active protein, we performed a peptidomics experiment in which different amounts of purified rhCPD

were incubated with a peptide mixture obtained by digestion of selected proteins with trypsin. This tryptic peptide library contains mainly peptides with basic residues (Arg or Lys) on their C-termini, and only a small number of peptides lacking C-terminal basic residues which arose from either the C-terminus of the proteins or from non-tryptic cleavages. After incubation with CPD, the individual reactions were each labeled with a different isotopic TMAB tag, combined and analysed by liquid chromatography/mass spectrometry (LC-MS) (see experimental scheme in **Figure 4-A**).

After LC-MS, over 55 peptides were identified through tandem mass spectrometry (MS/MS) and/or close matches with the theoretical mass of the peptides generated with trypsin. The criteria followed to consider the matches include coincidence of the observed mass with the monoisotopic theoretical mass (within 50 ppm); an expected charge equal to the number of TMAB-labeled lysines plus the N-terminus along with arginines and occasionally, histidines; and the correct number of isotopic tags incorporated (based on the number of free amines present in the peptide). Some of these peptides exhibited a peak set with roughly equal peak heights, revealing that these peptides were not substrates or products of rhCPD under the experimental conditions used (**Figure 4-B** and **Supplemental Table 2**). Some peptides were extensively cleaved, showing a complete or almost complete decrease in the peak intensity in the sample incubated with the highest concentration of enzyme (*i.e.,* 100 nM) and a partial decrease in the peak intensity of the sample incubated with a lower concentration of rhCPD (*i.e.,* 10 nM); these are considered as good substrates of rhCPD (**Figure 4-C** and **Table 2**). In addition, some peptides were only partially cleaved, exhibiting a small decrease in intensity with the highest concentration of enzyme assayed, and no major decrease in the peak intensity with the concentration of 10 nM; these are considered as weak substrates of rhCPD (**Figure 4-D** and **Table 3**). Some peptides showed an increase in peak intensities that correlated with the amount of rhCPD; these are considered to be products of rhCPD cleavage (**Figure 4-D** and **Table 4**).

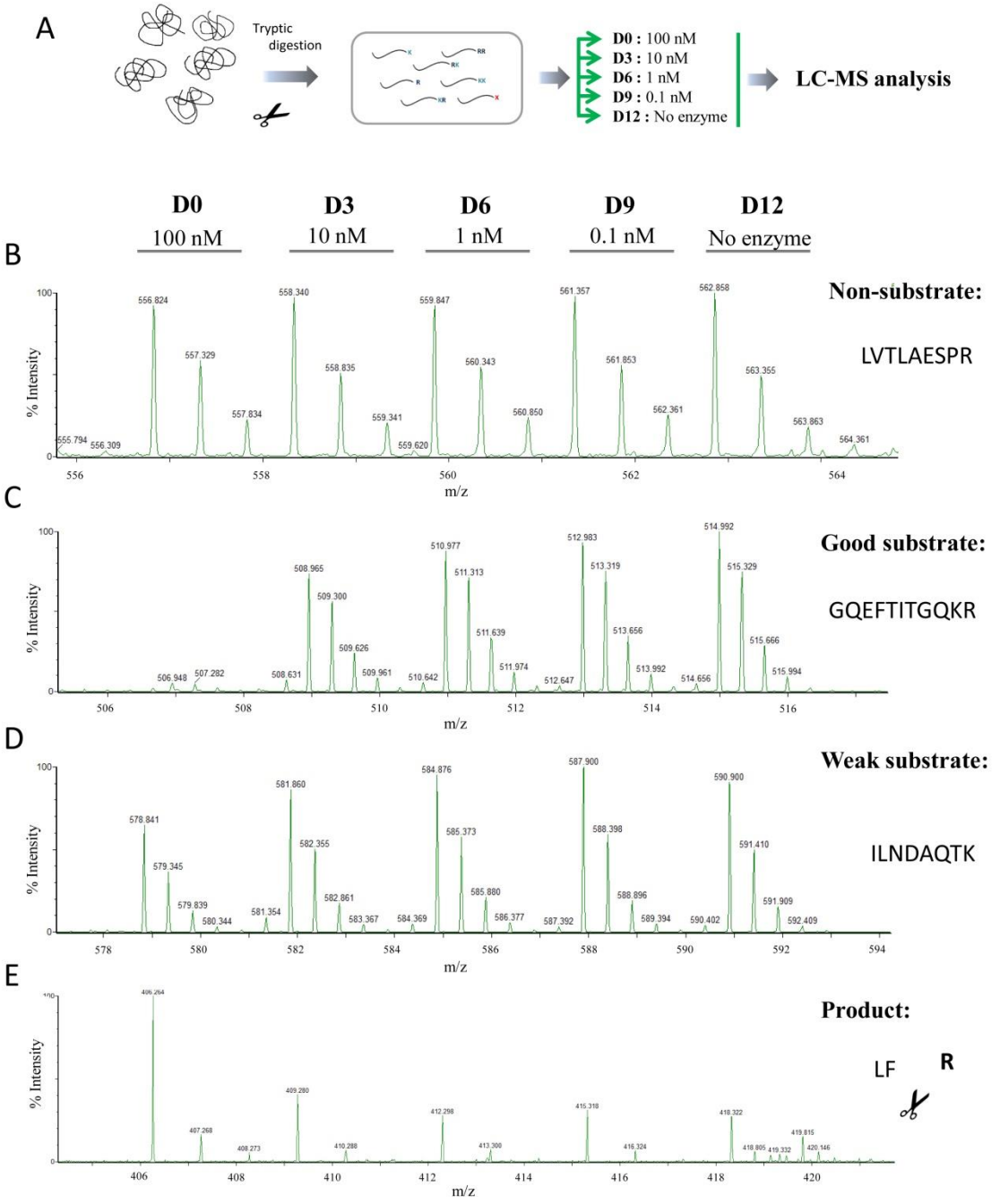


FIGURE 4. (A) Quantitative peptidomics scheme for the characterization of rhCPD substrate specificity using the tryptic peptide library and (B-E) representative spectra. Tryptic peptides were obtained from digestion of five selected proteins (BSA, bovine thryroglobulin, bovine α-lactalbumin and human α and β–hemoglobin) with trypsin. The resultant peptide library was aliquoted and digested with no enzyme or different rhCPD concentrations of 0.1, 1, 10, and 100 nM for 16 h at 37ºC. After incubation samples were labeled with one of five stable isotopic TMAB-NHS tags (D0= 100 nM; D3=10 nM; D6=1 nM; D9=0.1 nM; D12= No enzyme) Then, samples were pooled and analysed byLC-MS). Examples of representative data are shown for (B) non-substrates, (C) good substrates, (D) weak substrates and (E) products.

Table 2. Good substrates of rhCPD identified using the tryptic peptide library

Protein precursor	Peptide sequence	Z	T	Obs M	Theor M	ppm	Ratio rhCPD / No enzyme			
							100 nM	10 nM	1 nM	0.1 nM
Thyroglobulin	QQAAALAK	2	2	799.46	799.46	-4	<0.10	0.68	0.95	1.08
Thyroglobulin	FPLGESFLAAK	2	2	1178.64	1178.63	12	<0.10	0.94	1.12	1.09
α-Hemoglobin	VDPVNFK	2	2	817.43	817.43	2	<0.10	0.87	1.03	1.06
Thyroglobulin	GQEFTITGQKR	3	2	1263.67	1263.66	11	<0.10	0.75	0.96	0.95
Bovine serum albumin	ADLAK	2	2	516.28	516.29	-14	0.11	0.83	1.0	1.11
Thyroglobulin	KFEK	2	3	550.29	550.31	-38	0.11	0.67	0.85	1.07
Thyroglobulin	SLSLK	2	2	546.33	546.34	-15	0.13	0.89	0.98	1.13
Thyroglobulin	LPESK	2	2	572.32	572.32	-8	0.14	0.64	0.49	1.07
Thyroglobulin	KFEKLPESK	2	4	1104.61	1104.62	-5	0.15	0.66	0.96	1.19
Thyroglobulin	LTDEELAFPPLSPSR	2	1	1670.88	1670.85	20	0.16	0.78	1.10	1.07
Bovine serum albumin	LVNELTEFAK	2	2	1162.63	1162.62	12	0.20	0.94	1.09	1.09
α-Hemoglobin	LRVDPVNFK	3	2	1086.63	1086.62	13	0.21	0.88	1.09	1.06
Bovine serum albumin	LVTDLTK	2	2	788.47	788.46	9	0.40	0.84	1.05	1.08

Good substrates; peptides affected with a decrease ≥60% by the highest concentration of enzyme. Z, charge; T, number of isotopic tags incorporated into each peptide; Obs M, observed monoisotopic mass; Theor M, theoretical monoisotopic mass; ppm, difference between Obs M and Theor M (in parts per million); Ratio rhCPD/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.

Table 3. Weak substrates of rhCPD identified using the tryptic peptide library

Protein precursor	Peptide sequence	Z	T	Obs M	Theor M	ppm	Ratio rhCPD / No enzyme			
							100 nM	10 nM	1 nM	0.1 nM
Thyroglobulin	FEKLPESK	3	3	976.53	976.52	8	0.42	1.00	0.97	1.15
Thyroglobulin	KGQEFTITGQK	3	3	1235.66	1235.65	7	0.53	0.67	0.86	1.11
Thyroglobulin	ALEQATR	2	1	787.43	787.42	19	0.55	0.65	0.81	0.90
Thyroglobulin	FVAPESLK	2	2	889.50	889.49	10	0.61	0.79	1.00	1.03
Thyroglobulin	ILNDAQTK	2	2	901.49	901.49	5	0.67	0.90	1.10	1.10
Bovine serum albumin	AEFVEVTK	1	2	921.50	921.48	18	0.67	0.91	0.97	0.99
Bovine serum albumin	KVPQVSTPTLVEVSR	3	2	1638.95	1638.93	14	0.68	0.92	1.00	1.08
Bovine serum albumin	KQTALVELLK	3	3	1141.70	1141.71	-6	0.73	1.00	1.14	1.00
α-Hemoglobin	VLSPADKTNVK	3	3	1170.67	1170.66	5	0.74	0.86	1.00	1.08
Thyroglobulin	AVKQFEESQGR	2	2	1277.66	1277.64	17	0.76	0.78	0.86	1.12
Bovine serum albumin	VPQVSTPTLVEVSR	3	1	1510.87	1510.84	22	0.77	1.03	1.08	1.08
Thyroglobulin	ELSVLLPNR	2	1	1039.62	1039.60	2	0.79	0.97	1.06	1.12
β-Hemoglobin	VNVDEVGGEALGR	2	1	1313.69	1313.66	19	0.79	0.91	1.03	1.06

Weak substrates; peptides affected with a decrease ≥20% and <60% by the highest concentration of enzyme. See Table 2 for abbreviation definitions.

Table 4. Products of rhCPD identified using the tryptic peptide library

Protein precursor	Peptide sequence	Cleaved aa	Z	T	Obs M	Theor M	ppm	Ratio rhCPD / No enzyme			
								100 nM	10 nM	1 nM	0.1 nM
Thyroglobulin	LF	R	1	1	278.16	278.15	30	3.63	1.75	0.88	0.88
Bovine serum albumin	AEFVEVT	K	2	1	793.40	793.39	13	>5.00	ND	ND	ND
Bovine serum albumin	LVNELTEFA	K	1	1	1034.54	1034.53	12	>5.00	ND	ND	ND

Products; peptides with an increase >120% with one or more concentrations of enzyme; Cleaved aa, the amino acid cleaved by rhCPD to generate the observed peptide; ND, not detectable. See Table 2 for the rest of abbreviation definitions.

The majority of peptides with C-terminal (P1') Lys were either good or weak substrates of rhCPD, while fewer of the peptides with C-terminal Arg were good substrates and the majority of these peptides were non-substrates (**Figure 5-A** and **Supplemental Table 1**). The influence of the penultimate (P1) residue on rhCPD substrates and non-substrates was examined (**Figure 5-B**). Good substrates of rhCPD contained Ala, Ser, Phe, Leu, Lys, Thr, or Glu in the P1 position **(Figure 5B** and **Table 2)**. Weak substrates contained C-terminal Thr, Ser, Leu, Gly, Val, Glu, Asn or Gln (**Table 3**). No peptide with Pro, Asp and Ile in the P1 position was identified as a rhCPD substrate. The three products were generated by cleavage of peptides containing Arg or Lys at the C-termini and containing Ala, Phe or Thr amino acids at P1 position (**Table 4**).

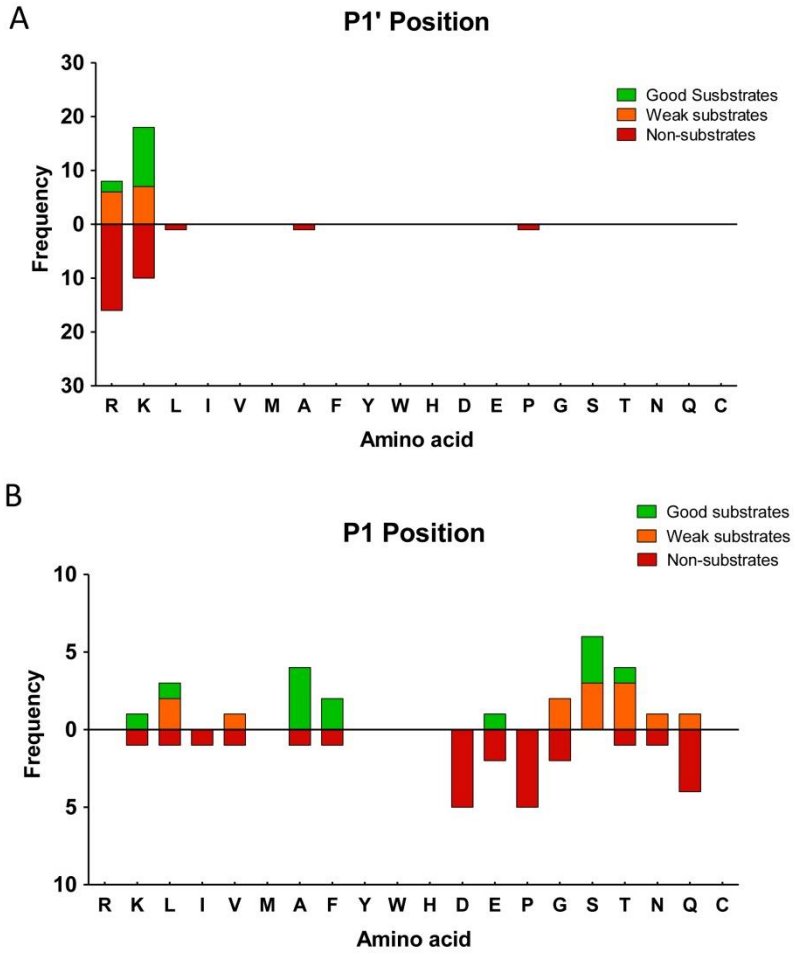


Figure 5. Analysis of the substrate preferences of rhCPD using the tryptic peptide library (A) Substrate preferences of rhCPD at C-terminal (P1') and (B) P1 positions. The number of times each amino acid was present in P1 or P1' was determined for good substrates, weak substrates and non-substrates and represented. For P1 analysis, only substrates with permissive P1' residues were considered.

As a side point, peptides containing Cys, Tyr, or His showed variable peak intensities that did not correlate with the amount of CPD. In addition, peptides containing cyanoCys, mono-iodoTyr, di-iodoTyr, mono-iodoHis, and di-iodoHis were identified from MS/MS analysis of the data; these peptides result from chemical reactions due to impurities in the TMAB labels (Fricker 2015). This finding might explain the absence and/or relatively low abundance of peptides with Cys, Tyr, or His residues in some previous peptidomics studies which using TMAB tags (Sebastian Tanco et al. 2010; Lyons & Fricker 2010; Morano et al. 2008). Consequently, peptides with Tyr, His, or Cys were not considered for the present work.

3.3.5 CHARACTERIZATION OF THE SUBSTRATE SPECIFICITY OF HUMAN CPD DOMAINS I AND II BY QUANTITATIVE PEPTIDOMICS

To gain insight into the substrate specificity of the domains I and II of human CPD, we repeated the quantitative peptidomics and compared rhCPD with the various mutant forms in a single LC/MS run. A schematic of one experiment and labeling scheme is shown in **Figure 6A**; we performed sufficient replicates so that at least three biological replicates of each enzyme form was tested. For this analysis, a single concentration (100 nM) of purified rhCPD, of single point mutants E350Q and E762Q, and the double mutant E350Q/E762Q were incubated with the peptide mixture obtained by digestion of selected proteins with trypsin. After incubation with the enzyme, the individual reactions were differentially labeled with isotopic TMAB tags, combined and subjected to LC-MS analysis (**Figure 6-A**). Over 48 peptides were identified through tandem mass spectrometry (MS/MS) and/or by close matching with the theoretical mass of the peptides generated with trypsin, following the same criteria as described above. After the analysis, the results revealed some peptides which showed roughly equal peak heights; these peptides were considered to be neither substrates nor products under the assay conditions used (**Supplemental Table 2**). Some peptides were cleaved by the enzyme with both domains I and II active, showing a complete or almost complete decrease in the peak intensity upon incubation with rhCPD; these are considered as substrates (which include good and weak substrates) of rhCPD (**Figure 6-B** and **Table 5**).

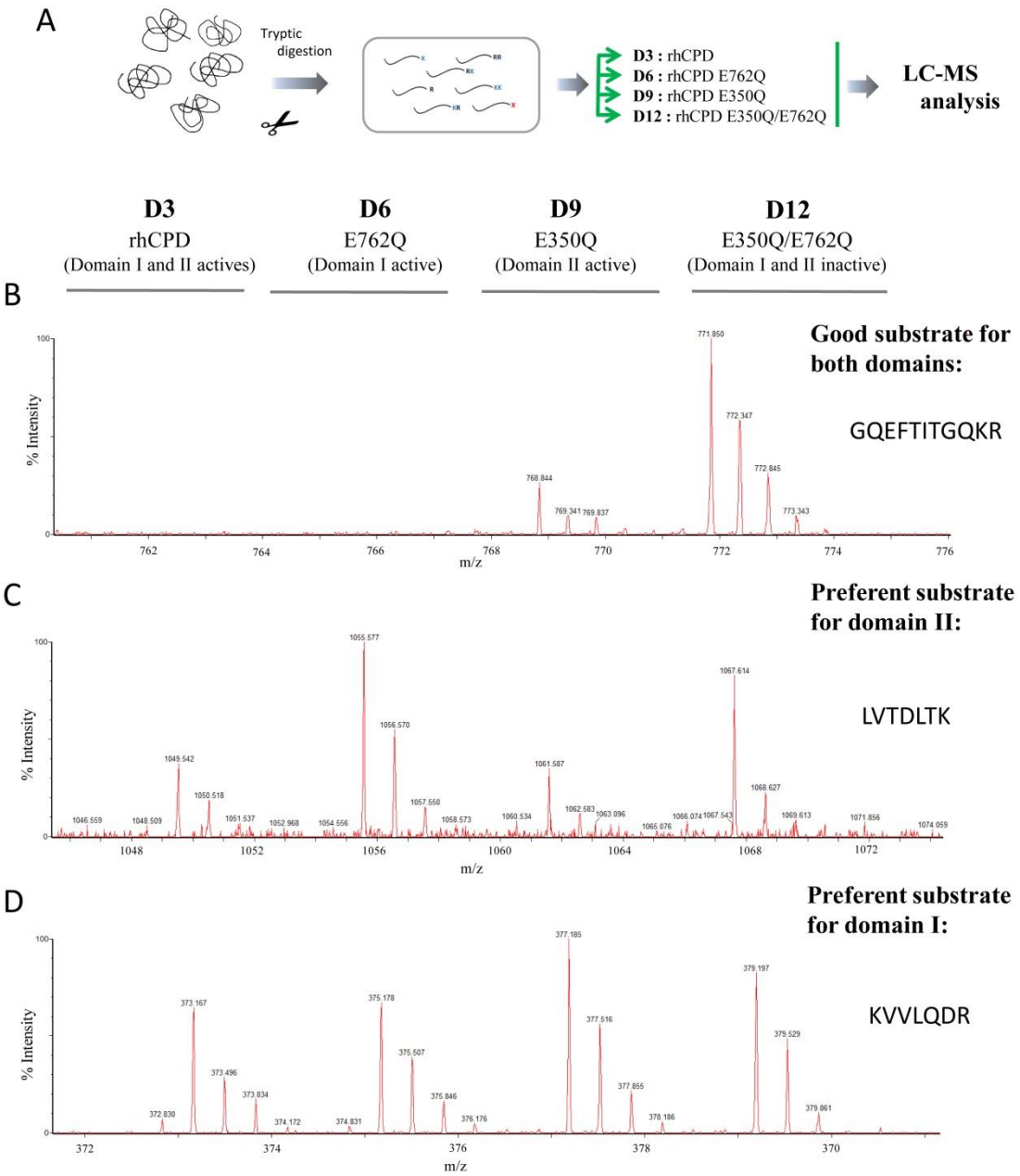


Figure 6. (A) Quantitative peptidomics scheme for the substrate characterization of CPD domains I and II using a tryptic peptide library and (B-E) representative spectra. Tryptic peptides were obtained as described above. The resultant peptide library was aliquoted and digested with 100 nM rhCPD, 100 nM rhCPD E350Q, 100 nM rhCPD E762Q or 100 nM rhCPD E350Q/E762Q for 16 h at 37ºC. Then, samples were labeled with one of the isotopic TMAB-NHS tags (D3=rhCPD; D6=rhCPD E762Q; D9=rhCPD E350Q; D12=rhCPD E350Q/E762Q). Finally, samples were pooled and analysed by LC-MS. Examples of representative data are shown for (B) good substrates of both domains I and II, (C) preferential substrates of domain I and (D) preferential substrates of domain II.

Table 5. Good and weak substrates identified within substrate characterization of CPD domains I and II using the tryptic peptide library

Protein precursor	Peptide sequence	Z	T	Obs M	Theor M	Ppm	Ratio enzyme / Control			Ratio dI active / d II active
							rhCPD	Domain I active	Domain II active	
Thyroglobulin	GQEFTITGQKR	2	2	1263.66	1263.66	4	<0.10	**<0.10**	**0.16**	<0.10
α-Hemoglobin	VDPVNFK	2	2	817.42	817.43	-20	<0.10	1.05	**<0.10**	>5.00
Thyroglobulin	FAATSFR	2	1	798.40	798.40	-3	0.12	**<0.10**	**0.24**	0.10
Bovine serum albumin	LVNELTEFAK	2	2	1162.62	1162.62	1	0.27	0.65	**<0.10**	6.88
Bovine serum albumin	LVTDLTK	2	2	788.45	788.46	-16	0.33	0.83	**0.29**	3.53
α-Hemoglobin	LRVDPVNFK	3	2	1086.57	1086.62	-44	0.34	1.09	**<0.10**	>5.00
Bovine serum albumin	KQTALVELLK	3	3	1141.66	1141.71	-41	0.42	1.09	0.42	2.60
Bovine serum albumin	AEFVEVTK	2	2	921.48	921.48	0	0.43	0.75	0.41	1.84
Thyroglobulin	KVVLQDR	2	2	856.49	856.51	-18	0.54	0.52	0.66	0.86
Thyroglobulin	AFLGTVR	2	1	762.44	762.44	5	0.59	0.56	0.53	1.07
Bovine serum albumin	KVPQVSTPTLVEVSR	3	2	1639.00	1638.93	41	0.59	0.96	0.51	1.86
Thyroglobulin	VVLQDR	2	1	728.42	728.42	0	0.64	0.79	0.79	1.00
Thyroglobulin	AVKQFEESQGR	3	2	1277.64	1277.64	3	0.64	0.75	0.75	1.00
Thyroglobulin	AISVPEDIAR	2	1	1069.63	1069.58	46	0.65	0.69	0.59	1.18
Thyroglobulin	ASGLGAAAGQR	2	1	957.52	957.50	17	0.72	0.56	0.83	0.67
Thyroglobulin	GQEIPGTR	2	1	856.47	856.44	33	0.72	0.59	0.59	1.00
Bovine serum albumin	IETMR	2	1	648.31	648.33	-26	0.74	0.59	0.74	0.80
Thyroglobulin	ELSVLLPNR	2	1	1039.64	1039.60	43	0.74	0.78	0.78	1.00
Thyroglobulin	KGQEFTITGQK	3	3	1235.63	1235.65	-15	0.78	0.99	0.74	1.33

Good substrates; peptides affected with a decrease ≥60% by rhCPD. Weak substrates; peptides affected with a decrease ≥20% and <60% by rhCPD; Ratio enzyme/control, the ratio in peak intensity between the sample incubated with rhCPD, E762Q or E350Q (as indicated for rhCPD, domain I active or domain II active, respectively) and the sample incubated with the double mutant (E350Q/E762Q); Ratio d I active /d II active, the ratio in peak intensity between the sample incubated with E762Q and the sample incubated with E350Q. Numbers underlined and highlighted in bold in columns 9 and 10 indicate peptides (within rhCPD substrates) considered as non-substrates and good substrates of domain I and II, respectively. See Table 2 for the rest of abbreviation definitions.

Among them, some peptides were cleaved by both domains showing a complete or almost complete decrease in the peak intensities after incubation with the single point mutants E350Q and E762Q (**Figure 6-B** and **Table 5**); these are considered as substrates of both domains. However, some other peptides substrates of rhCPD revealed a complete or almost complete decrease in the peak intensity upon incubation with the single point mutant E350Q and a slight or no decrease in the peak intensity upon incubation with E762Q; these are considered as preferential substrates of domain II (**Figure 6-C** and **Table 5**). Another group of peptides revealed a moderate decrease in the peak intensity upon incubation with the single point mutant E762Q and a slight or no decrease in the peak intensity upon incubation with the E350Q mutant; these are considered as preferential substrates of domain I (**Figure 6-D** and **Table 5**).

Analysis of the P1' residue of substrates of domain I showed a predominance of Arg, with 11 substrates containing C-terminal Arg and only 2 substrates containing C-terminal Lys (**Figure 7-A**). In contrast, substrates of domain II were more evenly divided between C-terminal Lys versus Arg (**Figure 7-C**). Several peptides with C-terminal Arg and Lys were also identified as non-substrates of the mutant CPD with single catalytic domains. For these peptides, the composition of the penultimate (P1) residue was further analysed (**Figure 7-B and 7-D**). Good substrates of the CPD domain I contained only Lys or Phe amino acids in this position. Weak substrates of CPD domain I contained Ala, Thr, Asp, Val, Met, Asn, Gln or Gly in the P1 position **(Figure 7-B)**. Good substrates of the CPD domain II contained Phe, Lys, Ala or Thr in the P1 position. Weak substrates of the CPD domain II contained Asp, Thr, Ala, Leu, Thr, Met, Val, Asn, Gln or Gly in this position **(Figure 7-D)**. No peptides with Pro, Ile, or Glu amino acids in the P1 position were identified as substrates of either domain under the experimental conditions performed.

Six peptides were identified as products of rhCPD and domain II; these resulted from cleavage of Arg or Lys from tryptic peptides, with Ser, Phe or Val at the P1 position of the cleavage site (**Table 6**). Three of these peptides required removal of C-terminal Arg, and these were also products of domain I.

Table 6. Products identified within substrate characterization of CPD domains I and II using the tryptic peptide library

Protein precursor	Peptide sequence		Z	T	Obs M	Theor M	Ppm	Ratio enzyme[1] / Control[2]			Ratio dI active / d II active
								rhCPD	Domain I active	Domain II active	
Thyroglobulin	GLFPS	R	1	1	519.26	519.26	12	>5.00	>5.00	>5.00	1.08
Thyroglobulin	AFLGTV	R	1	1	606.34	606.33	29	>5.00	>5.00	>5.00	0.89
Thyroglobulin	FAATSF	R	1	1	642.31	642.29	36	>5.00	>5.00	>5.00	0.88
α-Hemoglobin	VLSPADKTNV	K	2	2	1042.59	1042.57	23	>5.00	0.59	>5.00	<0.10
α-Hemoglobin	LRVDPVNF	K	2	1	958.56	958.52	40	>5.00	2.35	>5.00	<0.10
Thyroglobulin	FEKLPES	K	2	2	848.42	848.43	-5	>5.00	0.59	>5.00	<0.10

Products; peptides with an increase >120% with one or more concentrations of enzyme. Cleaved aa. The amino acid cleaved by rhCPD to generate the observed peptide. See Table 2 and 5 for the rest of abbreviation definitions.

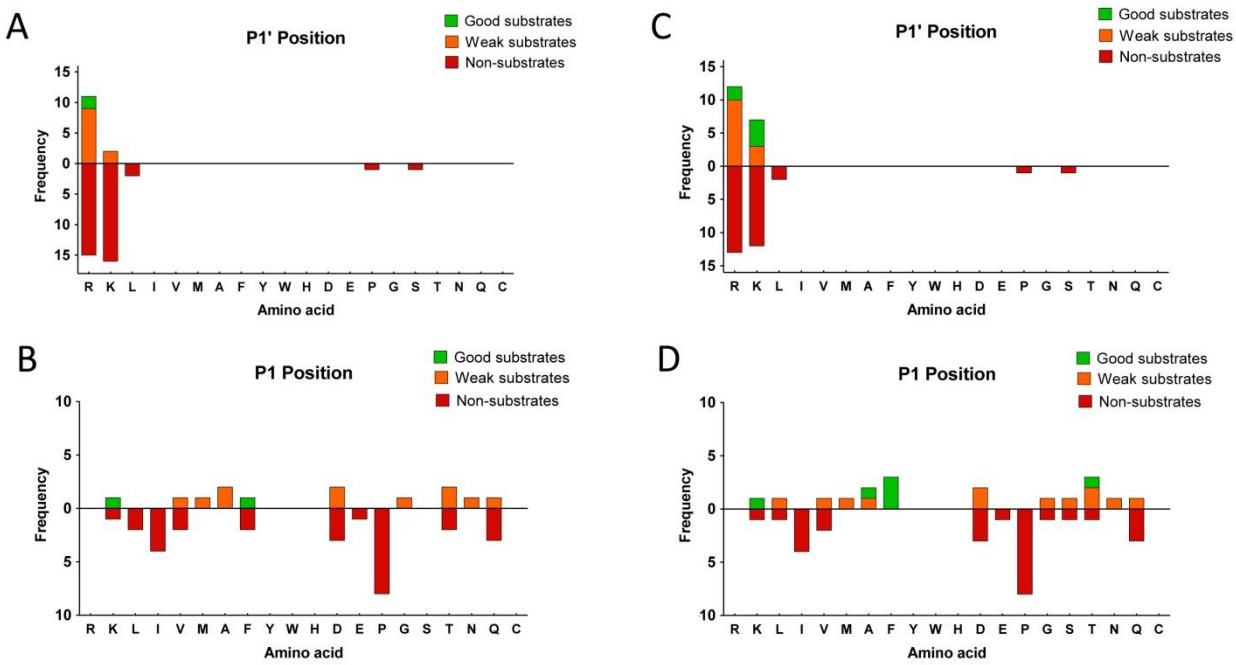


Figure 7. Analysis of the substrate preferences of CPD domains I and II using a tryptic peptide library. Substrate preferences of CPD domain I at (A) C-terminal (P1') and (B) P1 positions. Substrates preference of CPD domain II at (C) C-terminal (P1') and (D) P1 positions. The number of times each amino acid was present in P1 or P1' was determined for good substrates, weak substrates and non-substrates and represented. For P1 analysis, only substrates with permissive P1' for each domain were considered.

The other three peptides required removal of C-terminal Lys, and two of these peptides were only products of domain II and not domain I. The peptide LRVDPVNF was greatly elevated upon incubation with CPD with active domain II, and slightly elevated by incubation with CPD with active domain I (**Table 6**). Because the corresponding tryptic peptide contains a Lys on the C-terminus, this finding shows that CPD domain I is capable of cleaving C-terminal Lys, although not as efficient as CPD domain II.

3.3.6 HUMAN CPD CLEAVES EXCLUSIVELY C-TERMINAL BASIC RESIDUES

The tryptic peptide library provided a large number of peptides with C-terminal basic residues, but few peptides without C-terminal basic residues. To determine if rhCPD was able to cleave peptides lacking C-terminal basic residues, we performed a third quantitative peptidomics study using a HEK293T peptide extract as substrate. Different amounts of purified rhCPD were incubated with a peptide mixture extracted from bortezomib-treated HEK293T cells. Bortezomib, an antitumor drug that competitively inhibits the proteasomal beta-1 and beta-5 subunits, has been shown to paradoxically cause an increase in the levels of many intracellular peptides (Gelman et al. 2013). After incubation, the individual reactions were differentially labelled with isotopic TMAB tags, combined and analysed by liquid chromatography/mass spectrometry (LC-MS) (see experimental scheme in **Figure 8-A**). After LC-MS, more than 60 peptides were identified through tandem mass spectrometry (MS/MS). The majority of peptides that were identified did not contain C-terminal basic residues and were not affected by the treatment with rhCPD (**Figure 8B** and **Supplemental Table 3**). Of the peptides that contained C-terminal basic residues, some were extensively cleaved, showing a complete or almost complete decrease in the peak intensity upon incubation with the highest concentration of enzyme of 100 nM and a partial decrease in the peak intensity with a lower concentration of 10 nM; these are considered as good substrates of rhCPD (**Figure 8-C** and **Table 7**). Some peptides were only partially cleaved, exhibiting little decrease in intensity with the highest concentration of enzyme assayed and no or slight decrease in the peak intensity with the concentration of 10 nM; these are considered as weak substrates of rhCPD (**Figure 8-D** and **Table 8**). A small number of peptides showed an increase in the peak intensities related with increasing amounts of rhCPD; these are considered as products resultant after the rhCPD cleavage (**Figure 8-D** and **Table 9**).

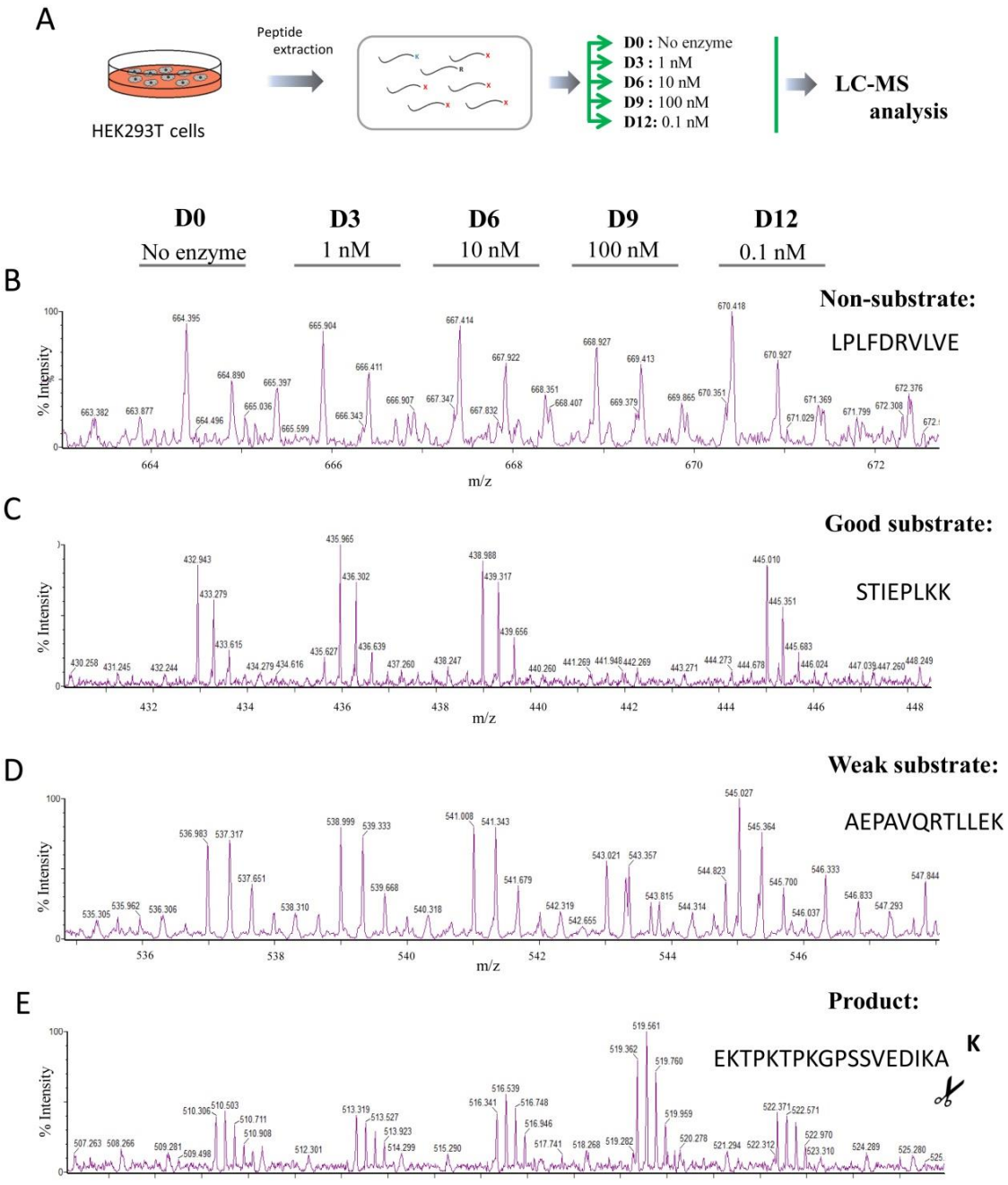


Figure 8. (A) Quantitative peptidomics scheme for the study of rhCPD substrate specificity using HEK293T derived peptide library and (B-E) representative spectra. Peptides were extracted from HEK293T cell cultures treated for 1 h at 37ºC with 0.5 µM bortezomib. The resultant peptide extract (i.e., HEK293T peptidome) was aliquoted and digested with no enzyme or different rhCPD concentrations of 0.1, 1, 10, and 100 nM at 37ºC for 16 h. After incubation samples were labeled with one of five stable isotopic TMAB-NHS tags (D0=100 nM; D3=10 nM; D6=1 nM; D9=0.1 nM; D12= No enzyme). Then, samples were pooled and analysed by LC-MS. Examples of representative data are shown for (B) non-substrates, (C) good substrates, (D) weak substrates and (E) products.

Table 7. Good substrates of rhCPD identified using HEK 293T peptides

Precursor	Sequence	Z	T	Obs M	Theor M	ppm	Ratio rhCPD / No enzyme			
							100 nM	10 nM	1 nM	0.1 nM
Eukaryotic translation initiation factor 5A	SAMTEEAAVAIKAMAK	3	3	1620.84	1620.821	11	<0.10	0.98	1.02	0.90
Acidic nuclear phosphoprotein pp32	STIEPLKK	3	3	914.53	914.54	-20	<0.10	1.12	1.15	1.00
40S Ribosomal protein S21	KADGIVSK	3	3	816.45	816.470	-20	<0.10	0.83	0.89	0.96
Histidine triad nucleotide-binding protein 1	Ac-ADEIAKAQVAR	2	1	1212.67	1212.646	21	0.15	1.03	1.03	1.14
Vimentin	AELEQLKGQGKSR	3	3	1442.78	1442.78	0	0.23	0.95	0.97	1.11
Eukaryotic translation initiation factor 5A	SAMoxTEEAAVAIKAMAK	3	3	1636.83	1636.821	3	0.32	0.98	1.00	0.94
Eukaryotic translation initiation factor 5A	NMDVPNIKR	3	2	1085.57	1085.565	4	0.38	0.95	1.05	0.72
Hematological and neurol. exp. 1 prot.	Ac-TTTTTFKGVDPNSRNSSR	3	1	2010.02	2009.977	20	0.40	1.13	1.07	0.97

See Table 2 and 5 for abbreviation definitions

Table 8.Weak substrates of rhCPD identified using HEK 293T peptides

Precursor	Sequence	Z	T	Obs M	Theor M	ppm	Ratio rhCPD / No enzyme			
							100 nM	10 nM	1 nM	0.1 nM
Hematological and neurol. exp. 1 prot.	Ac-TTTTTFKGVDPNSRNSSR	3	1	2010.02	2009.977	20	0.40	1.13	1.07	0.97
Cathepsin D	GPIPEVLK	2	2	851.51	851.512	1	0.58	0.99	0.95	0.94
CD99 antigen	AEPAVQRTLLEK	3	2	1353.77	1353.762	10	0.61	0.95	0.99	0.97
Ubiquitin-60S ribosomal protein L40	IIEPSLR	2	1	826.51	826.491	18	0.73	1.03	1.00	1.08

Weak substrates; peptides affected with a decrease ≥20% and <60% by the highest concentration of enzyme. See Table 2 for abbreviation definitions.

Table 9. Products of rhCPD identified using HEK 293T peptides

Precursor	Sequence	Cleaved aa	Z	T	Obs M	Theor M	ppm	Ratio rhCPD / No enzyme			
								100 nM	10 nM	1 nM	0.1 nM
40S Ribosomal protein S28	Ac-MoxDTSRVQPIKLA	R	3	1	1415.78	1415.749	25	1.56	1.12	1.21	1.02
Nucleophosmin	EKTPKTPKGPSSVEDIKA	K	5	5	1911.03	1911.03	-2	1.86	1.39	1.15	1.11

Products; peptides with an increase >120% with one or more concentrations of enzyme. Cleaved aa. The amino acid cleaved by rhCPD to generate the observed peptide. See Table 2 for the rest of abbreviation definitions.

Analysis of the P1' residue showed an exclusive preference for basic residues at this position (**Figure 9**).

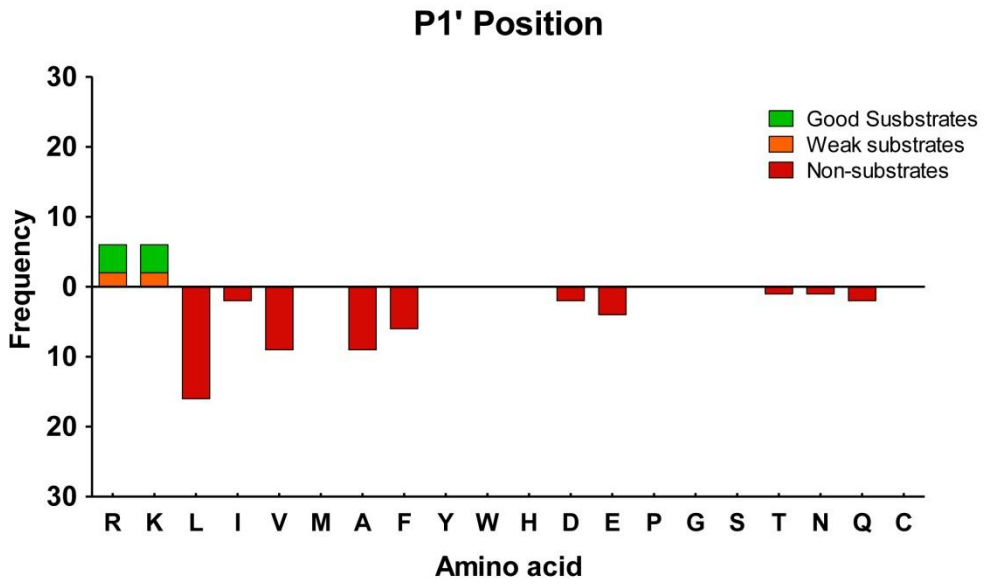


Figure 9. Analysis of the substrate preferences of rhCPD using a HEK293T derived peptidomics library. (A) Substrate preference of rhCPD at C-terminal (P1'). The number of times each amino acid was present in P1' was determined for good substrates, weak substrates and non-substrates.

3.3.7 COMPARATIVE MODELING OF THE ACTIVE SITES OF HUMAN CPD DOMAINS I, II AND III.

To date, a limited number of crystallographic studies have delimited the active site residues that are involved in the catalytic mechanism of the M14B subfamily of MCPs (Reverter et al. 2004; Keil et al. 2007; Sebastián Tanco et al. 2010; Gomis-Rüth et al. 1999). Because three-dimensional structures of human CPD domains are not available yet, we used I-TASSER (Zhang 2008a) to model all three human domains based on the previous structures from duck and *Drosophila melanogaster* solved by our groups (Gomis-Rüth et al. 1999; Sebastián Tanco et al. 2010).

For all human CPD domains, the basis of the active-site groove is formed by strands β5, β6, β8 and the N-terminal part of helix α2. The wall of the groove funnel is essentially defined by two projections of the α4–α5 multiple turn and by the α7–β7 hairpin loop, as described previously (Reverter et al. 2004). Surface representations of the catalytic sites of CPD domains I, II and III are shown in **Figure 10**. Like other MCPs, for human CPD domains I and II the active site clefts are located at the edge of the active-sites grooves within the β-sheet of the catalytic moiety. As shown in **Figure 10** and **Supplemental figure 2**, both domains I and II conserve the three protein ligands of the catalytic zinc ion (His69, Glu72 and His196, based on the conventional numbering system for the active form of bovine CPA1) and the residues directly involved in catalysis (Arg127, Arg145 and Glu270, according to the bovine CPA1 numbering). Further, both CPD domains I and II share the same residues involved in the main subsite, S1', for substrate binding (with Asp and Gln residues equivalent to positions 207 and 255 in bovine CPA1, respectively) that determine the substrate specificity for basic residues (**Figure 10**). Nonetheless, CPD domains I and II differ between them in some residues within such the S1' specificity pocket, as reported previously for the duck homologs (Reverter et al. 2004; Aloy et al. 2001). In particular, human CPD domain I contains Ser264 in a position comparable to Leu203 of bovine CPA and Ser198 of CPM (**Figure 10-B** and **Supplemental Figure 2**), whereas CPD domain II contain an Asn678 in the equivalent position.

We also performed the analysis of the residues that shape the S1 pocket in CPD domains I and II. In CPA1, the S1 subsite is shaped by Phe279, Tyr198, and Ser199 residues. Interestingly, and unlike CPA1, the CPD domain I and II as well as in CPM such hydrophobic presence at S1 subsite is provided by a Tyr, equivalent to the highly flexible Tyr248 in canonical CPs, which further closes the specificity pocket at the S1 subsite (**Figure 10**). Additionally for CPM, residues Lys245 and Met246 conforms a hydrophobic channel that can accommodate the side chain of P1 residues. In human CPD domains I and II, equivalent positions are occupied by Gly311 and Asp312 or Met723 and Tyr724 residues, respectively.

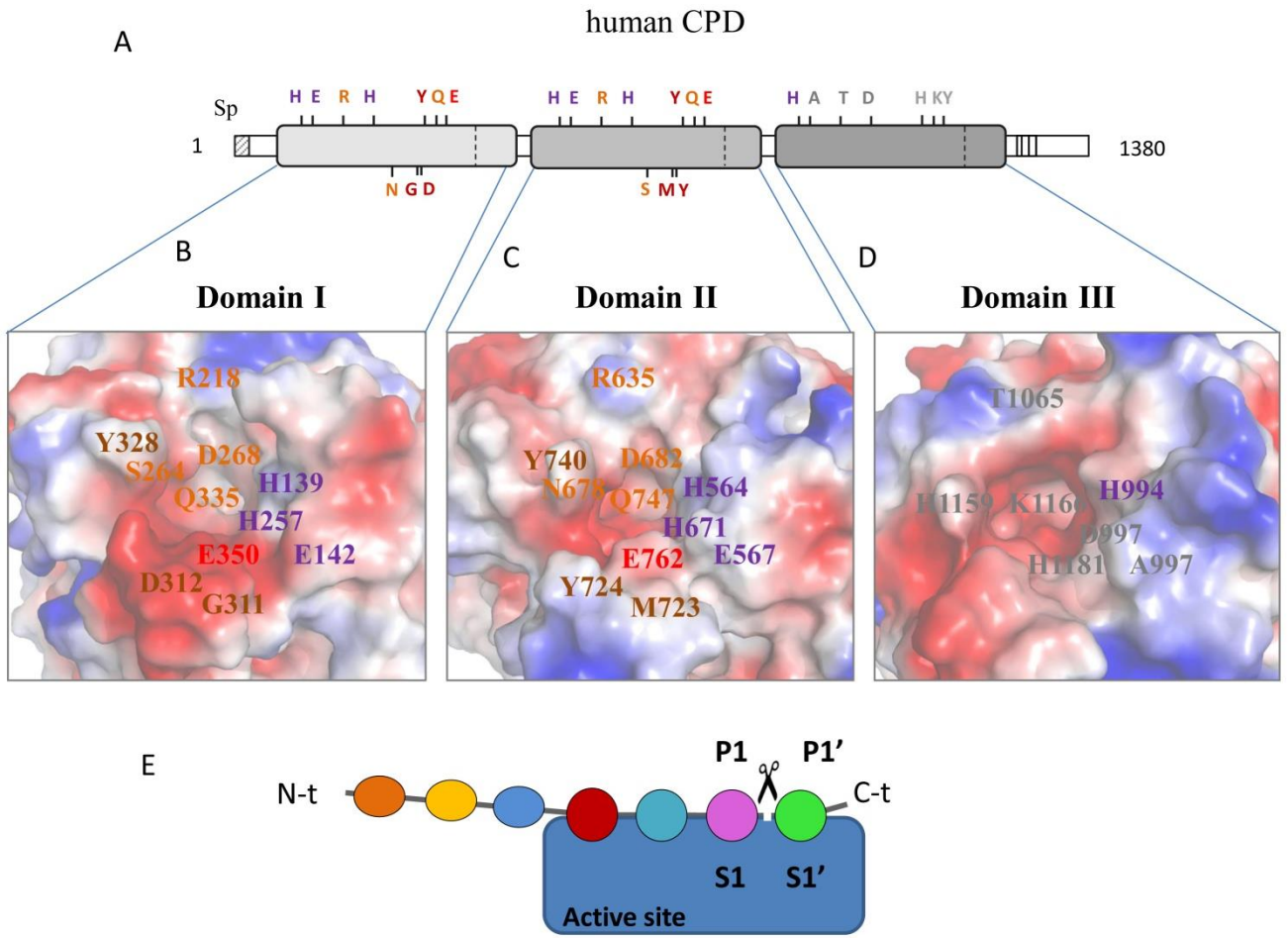


FIGURE 10. Structural modeling of the active sites of CPD domains I, II and III. (A) Linear representation of the full-length human CPD, showing the location of relevant amino acids involved in the catalytic mechanism. In violet, the residues directly involved in the zinc binding (*i.e.,* His69, Glu72 and His196) are shown. The catalytic residue Glu270 is indicated in red. In orange and brown, putative residues determining respectively the S1' and S1 specificity pockets are shown. Non-conserved residues, between inactive domain III and active domains I and II, are displayed in grey. Residues located below the linear representation correspond to non-conserved residues between CPD domains I and II. (B-D) Surface representation of the substrate binding sites, showing the location of the residues described above for (B) domain I, (C) domain II and (D) domain III. (E) Schematic representation of substrate binding sites, according to the model of Schechter and Berger (Schechter & Berger 1967).

Interestingly, inspection of the surface potential at the S1 pocket in CPD domain I shows a marked negative surface electrostatic potential in the region surrounding the mentioned Gly311 and Asp312 residues, while in the case of the domain II, a more hydrophobic region is observed due to the presence of Met723 and Tyr724 (**Figure 10-B and 10-C** and **Supplemental Figure 2**). Thus, these residue differences identified in the S1 pocket subsite between domains I and II might play an important role for substrate recognition and specificity.

In contrast to domains I and II, human CPD domain III lacks the majority of key residues for CP catalysis (See **Figure 10-C**). In CPD domain III, only the first His residue (*i.e.,* His69) involved in the zinc binding is conserved, while the other residues involved in metal binding (*i.e.,* Glu72 and His196) are replaced by Ala and Asp residues, which are not adequate to do such task, as viewed in the duck homolog (Aloy et al. 2001). Further, residues positions involved in the catalytic mechanism Glu272 and Arg135 are occupied in domain III by Tyr and His, respectively. Additionally, residues Asn144 and Arg145, responsible for the anchoring of the carboxylate group of the substrate, are replaced by a triad of Asp, Thr and Asp in the CPD domain III. Curiously, a similar triad was also found in homologous positions for the bacterial enzyme γ-D-glutamyl-(L)-meso-diaminopimelate peptidase I, which is a member of the M14 MCP gene family and a distant relative of CPD (Aloy et al. 2001; Garnier et al. 1985).

3.4 DISCUSSION

The primary goal of the present study was the detailed comparison of the substrate specificity of the first and second catalytic domains of CPD, as well as testing the third carboxypeptidase-like domain for catalytic activity using a broad array of peptides. CPD is the only member of the MCP family that contains multiple carboxypeptidase domains. Proteolytic enzymes with multiple catalytic centres are rare in nature, especially when such centres perform similar roles such as domains I and II of CPD. Another peptidase with two active domains (in most species and tissues) is angiotensin-converting enzyme (ACE); except for the form produced in sperm (which contains a single catalytic domain), ACE has two similar catalytic activities that differ mainly in their sensitivity to chloride and substrate preferences (Harrison & Acharya 2014; Bernstein et al. 2013). Another group of enzymes with multiple domains that perform related catalytic functions are the polyserases. As with CPD, the polyserases contain enzymatically active and inactive domains of unknown function. Polyserase-1 is post-translationally cleaved into distinct subunits that contain a single catalytic activity, but polyserase-2 and -3 remain intact as multi-domain-containing proteins (Cal et al. 2003; Cal et al. 2005; Cal et al. 2006). The presence of multiple catalytically active domains can increase the diversity of substrates cleaved, or the efficiency of the enzyme under different cellular conditions. Both appear to be the case for the two active domains of CPD.

One important finding of the present study is that human CPD domain I and II have distinct pH optima, with domain I working best at neutral pH while domain II is optimal at mildly acidic pH values. This observation is consistent with the properties found for CPD variants from duck and *Drosophila* observed in previous studies (Novikova et al. 1999; Sidyelyeva et al. 2006), suggesting that this feature is highly conserved through hundreds of millions of years of evolution. Although CPD is primarily detected in the trans Golgi network, its presence there is not static. CPD is present in vesicles that bud from the trans Golgi, but is retrieved from these vesicles and sorted back to the trans Golgi (unlike CPE, which moves into mature secretory granules). CPD is also transiently found on the cell surface, where it is internalized and a fraction is transported through the endocytic pathway back to the trans Golgi network. The pH of each of these compartments is different, ranging from neutral (the cell surface) to slightly acidic (the trans Golgi network) to acidic (endosomes). The broad pH range of domain II implies that this activity is the predominant one in the various cellular compartments, while domain I is like to be the major activity on the cell surface where the pH is neutral.

Another important finding of the present study is that human CPD domain I and II have differences in their substrate specificities. Both are specific for C-terminal basic residues, with no detectable cleavage of non-basic residues on the C-terminus. The present study tested dozens of peptides in a peptidomic assay, and extended previous studies done on duck and *Drosophila* CPD with a limited number of substrates. The preference of CPD domain I for Arg over Lys, and the broad ability of domain II to cleave both Arg and Lys with comparable efficiency was previously noted in a study that tested duck CPD with a single pair of peptides that differed only in the C-terminal residue. The present study tested dozens of peptides, and while each domain was able to cleave either Lys or Arg from some peptides, there was a clear preference for domain I to cleave Arg and not Lys. Taken together with the pH optima, it appears that when present on the cell surface, CPD domain I will be primarily active and cleave peptides/proteins with C-terminal Arg, while CPD present in the trans Golgi and endocytic pathways will have domain II active and cleave both C-terminal Lys and Arg. Although the previous study on duck CPD used only a single substrate containing C-terminal Lys, these previous results are in agreement with the large number of peptide substrates found in the present study for human CPD, implying that this feature has also been conserved through evolution.

Previous studies using small numbers of substrates have found that some members of the M14B subfamily of MCPs are more efficient at cleaving C-terminal Arg than Lys. For example, CPM cleaves Met^5-Arg^6-enkephalin with a k_{cat} of 934 min^{-1} and K_m of 46 µM, whereas Met^5-Lys^6-enkephalin is cleaved with K_m of 375 µM and k_{cat} of 663 min^{-1} (Skidgel et al. 1989). Other studies evaluated the specificity of MCPs using peptides with C-terminal Arg or Lys residues as competitive inhibitors. For example, CPE is inhibited by Leu^5-Lys^6-enkephalin with a k_i of 174 µM and by Leu^5-Arg^6-enkephalin with a k_i of 83 µM (Fricker & Snyder 1982). CPZ is inhibited by hippuryl-Arg but not by hippuryl-Lys (Novikova & Fricker 1999a). Taken together, CPZ and CPM are like CPD domain I with a strong preference for C-terminal Arg over Lys, while CPE and CPN are more like CPD domain II which does not have a marked preference for one basic residue over another. One residue within the substrate binding pocket that correlates with selectivity is the residue in position equivalent to Leu203 in CPA1. In CPD domain I a Ser is present in this position, while in CPD domain II an Asn is present. Similarly, CPM and CPZ contain a Ser in this position, and CPE and CPN contain Asn. Perhaps the superior capability of Ser of the binding site (versus Asn) to fit and stablish links with the guanidine group of Arg from substrates could play a differential positive factor in such a behaviour.

Another finding of the present study is that the third carboxypeptidase-like domain of human CPD is inactive as a carboxypeptidase when tested both along pH or in a wide peptidomic screening. In no case was carboxypeptidase activity detected for the form of CPD with mutations in a key catalytic residue in domain I and II. This observation is also consistent with the lack of a conserved active site, substrate-binding, and metal-binding residues in the third domain of CPD. For example, the critical active site Glu (Glu270 using bovine CPA1 numbering) is a Tyr in the third domain of human and duck CPD. In fact, most of the changes in active site residues found in the third domain of CPD are conserved between human and duck, suggesting that these differences are not simply random events but are important for the function of this domain. One possibility is that the third domain operates in peptide binding. While Glu270 is essential for catalytic activity, it is not required for substrate binding, and substitution of Gln for Glu permits the protein to bind peptide "substrates" but not hydrolyze them. However, other evolutionary changes in substrate binding residues in the third domain of CPD (conserved between human and duck) suggest that this domain is not likely to bind peptides that are substrates of domains I and II. One important residue for peptide binding is Arg145 (bovine CPA1 numbering)—this residue anchors the C-terminal carboxyl group of the substrate. However, in domain III the residue in a comparable position is Thr. This along with other changes suggests that if domain III functions in binding, it will have a distinct binding profile compared to domains I and II.

Three other members of the M14B subfamily of MCPs are also inactive towards standard carboxypeptidase substrates. Two of these, CPX2 (Xin et al. 1998) and AEBP1/ACLP (Layne et al. 1998) also contain Tyr in place of the critical active site Glu270, suggesting that this Tyr is important. The third inactive member of this subfamily, CPX1, does contain Glu in the comparable position but is missing other critical active site and substrate-binding residues. Other families of enzymes contain members that are considered catalytically inactive, and the functions of most of these proteins are not yet known (Freeman 2014). Recently, AEBP1/ACLP was found to induce phosphorylation and nuclear translocation of Smad3, and this was dependent on TGFβ receptor binding and kinase activity (Tumelty et al. 2014). Some of the active carboxypeptidases appear to have functions that are independent of their catalytic activity. For example, CPE was proposed to function as an extracellular trophic factor that protected neurons from hydrogen peroxide-, staurosporine- and glutamate-induced cell death, and this effect did not require CPE enzyme activity (Cheng et al. 2014). CPM was reported to be a positive allosteric modulator of the kinin B1 receptor, and this action is independent of CPM enzyme activity (Zhang et al. 2013). The emerging concept is that some

carboxypeptidases have multiple functions, including both catalytic and non-catalytic roles, while the inactive members of this gene family like CPD domain III would only have non-catalytic roles.

3.5 SUPPLEMENTAL INFORMATION

Supplementary table 1. Non-substrates of rhCPD identified using the tryptic peptide library

Protein precursor	Peptide sequence	Z	T	Obs M	Theor M	ppm	Ratio rhCPD / No enzyme			
							100 nM	10 nM	1 nM	0.1 nM
Thyroglobulin	LPFQK	2	2	631.37	631.37	-3	0.82	0.84	0.98	1.07
Thyroglobulin	LTDEELAFPPLSPSRETFLEK	3	2	2418.26	2418.23	12	0.82	0.85	1.15	1.09
Thyroglobulin	LQLVDAPPASLPDLQDVEEALAGK	3	2	2488.36	2488.31	19	0.82	0.83	0.85	0.97
Thyroglobulin	GQEIPGTR	2	1	856.45	856.44	13	0.83	0.83	0.9	0.93
Thyroglobulin	LQLVDAPPASLPDLQDVEEALAGK	3	2	2488.36	2488.31	19	0.82	0.83	0.85	0.97
Thyroglobulin	ALADLAKP	2	2	797.46	797.46	-1	0.91	0.88	0.97	1.03
Thyroglobulin	VDLLIGSSQDDGLINR	2	1	1713.93	1713.89	20	0.91	1.06	1.02	1.10
Thyroglobulin	AISVPEDIAR	2	1	1069.60	1069.58	17	0.93	1.05	1.07	1.02
Thyroglobulin	SALGEPKK	2	1	828.49	828.47	18	0.93	0.93	0.97	1.00
α-Hemoglobin	Ac-VLSPADKTNVK	2	2	1212.68	1212.67	9	0.94	1.03	1.00	0.94
Thyroglobulin	IDVALR	2	1	685.43	685.41	28	0.97	1.00	1.00	1.00
Thyroglobulin	LGGQEIR	2	1	771.44	771.42	22	1.00	1.00	1.07	1.00
Bovine serum albumin	DAIPENLPPLTADFAEDK	2	2	1954.97	1954.95	11	1.00	1.00	1.08	1.08
Bovine serum albumin	DDSPDLPK	2	2	885.41	885.41	6	1.00	0.93	1.00	1.03
Thyroglobulin	SLLLAPEEGPVSQR	3	1	1494.83	1494.80	23	1.02	1.15	1.15	1.24
Thyroglobulin	ALADLAKPL	2	2	910.55	910.55	3	1.06	1.03	1.06	1.12
Thyroglobulin	KVVLQDR	2	2	856.52	856.51	7	1.06	1.03	1.03	1.06
Thyroglobulin	LVTLAESPR	2	1	984.58	984.56	24	1.07	1.00	1.07	1.07
Thyroglobulin	VVLQDR	1	1	728.43	728.42	17	1.08	1.10	1.02	1.04
Thyroglobulin	RSLLLAPEEGPVSQR	3	1	1650.94	1650.91	17	1.08	1.10	1.14	1.12
Thyroglobulin	QAGVQAEPSPK	3	2	1110.57	1110.57	3	1.09	0.99	0.98	1.11
Bovine serum albumin	LVVSTQTALA	2	1	1001.60	1001.58	16	1.10	1.10	1.10	1.00
Thyroglobulin	ASGLGAAAGQR	2	1	957.52	957.50	18	1.10	1.07	1.10	1.07
Thyroglobulin	FLQGDR	1	1	734.39	734.37	22	1.11	1.01	1.07	1.07
Thyroglobulin	LNSNPASEAPK	2	2	1126.57	1126.56	5	1.12	0.97	1.06	1.03
Thyroglobulin	LQQNLFGGR	2	1	1031.57	1031.55	19	1.12	1.12	1.04	1.12
Thyroglobulin	VTLAADR	1	1	744.43	744.41	30	1.14	1.16	1.16	1.09
Trypsin[1]	VATVSLPR	2	1	841.52	841.50	19	1.14	1.25	1.18	1.07
Thyroglobulin	ETFLEK	2	2	765.39	765.39	3	1.14	1.07	1.18	1.14
Thyroglobulin	FAATSFR	2	1	798.42	798.40	24	1.15	1.15	1.11	1.07

[1]Fragment originated from trypsin autolysis. See Table 2 for the abbreviation definitions.

Suplemental table 2. Non-substrates identified within substrate characterization of CPD domains I and II using the tryptic peptide library

Protein precursor	Peptide sequence	Z	T	Obs M	Theor M	ppm	Ratio enzyme / Control		
							rhCPD	Domain I active	Domain II active
Thyroglobulin	LILPR	1	1	610.44	610.42	37	0.88	0.85	0.91
Thyroglobulin	ALADLAKPL	2	2	910.55	910.55	-3	0.88	0.88	0.92
Thyroglobulin	ALADLAKPLS	2	2	997.59	997.58	7	0.88	0.90	0.82
Thyroglobulin	RLVTLAESPR	3	1	1140.69	1140.66	22	0.89	0.98	1.00
Thyroglobulin	VTLAADR	2	1	744.42	744.41	8	0.90	0.95	1.05
Bovine serum albumin	LKPDPNTL	2	2	896.49	896.50	-3	0.92	1.20	1.09
Thyroglobulin	RVTLAADR	3	1	900.51	900.51	-5	0.92	1.01	0.84
Thyroglobulin	LVTLAESPR	2	1	984.59	984.56	35	0.92	0.96	0.92
Thyroglobulin	ALADLAKP	2	2	797.44	797.46	-32	0.92	0.92	0.87
Thyroglobulin	ILNDAQTK	2	2	901.49	901.49	1	0.94	1.20	0.65
Thyroglobulin	FVAPESLK	2	2	889.48	889.49	-11	0.94	1.18	0.82
Thyroglobulin	FARFTASCPPSIK	2	2	1423.78	1423.73	38	0.95	0.99	0.92
α-Hemoglobin	VLSPADKTNVK	2	3	1170.68	1170.66	13	0.96	1.13	0.78
Thyroglobulin	IDVALR	2	1	685.40	685.41	-9	0.98	0.84	0.90
Bovine serum albumin	DDSPDLPK	2	2	885.40	885.41	-15	1.03	1.18	1.32
Thyroglobulin	VLQFIR	2	1	774.48	774.48	12	1.04	0.98	0.94
α-Hemoglobin	Ac-VLSPADKTNVK	2	2	1212.65	1212.67	-14	1.05	1.19	0.83
Thyroglobulin	ETFLEK	2	2	765.37	765.39	-25	1.05	1.26	0.95
Thyroglobulin	LGGQEIR	2	1	771.43	771.42	13	1.06	1.06	1.11
Thyroglobulin	SALGEPKK	2	1	828.48	828.47	14	1.08	1.04	0.94
Thyroglobulin	GQEFTITGQK	2	2	1107.59	1107.56	24	1.08	1.18	0.98
Trypsin[1]	VATVSLPR	2	1	841.52	841.50	21	1.09	0.97	0.97
Thyroglobulin	LQQNLFGGR	2	1	1031.60	1031.55	46	1.09	0.65	1.09
Thyroglobulin	ILQR	1	1	528.34	528.34	-4	1.10	0.74	1.18
Thyroglobulin	FLQGDR	2	1	734.38	734.37	7	1.10	1.03	1.05
Thyroglobulin	LNSNPASEAPK	2	2	1126.58	1126.56	20	1.11	0.94	1.11
Thyroglobulin	QAGVQAEPSPK	2	2	1110.60	1110.57	25	1.11	1.05	1.05
Thyroglobulin	GLFPSR	2	1	675.36	675.37	-12	1.12	1.02	1.15
Thyroglobulin	LTGISIR	2	1	758.47	758.47	8	1.13	1.02	1.06

[1]Fragment originated from trypsin autolysis. See Table 2 for the abbreviation definitions.

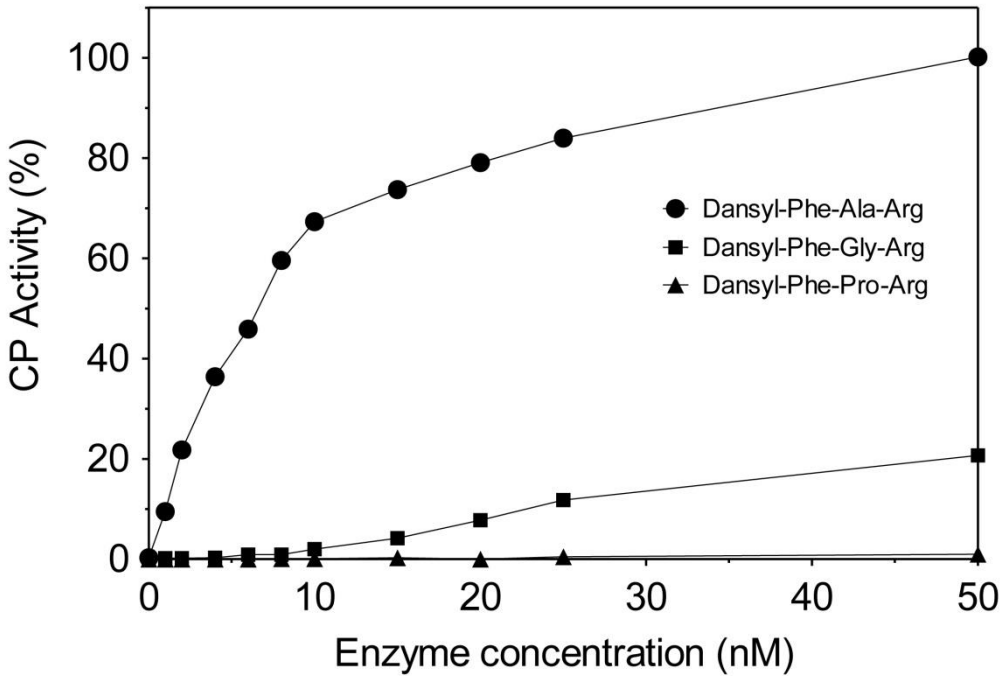
Supplemental table 3. Non-substrates of rhCPD identified using HEK 293T peptides

Precursor	Sequence	Z	T	Obs M	Theor M	ppm	Ratio rhCPD / No enzyme			
							100 nM	10 nM	1 nM	0.1 nM
40S Ribosomal protein S28	Ac-MDTSRVQPIKLA	2	1	1399.79	1399.749	30	0.82	1.08	1.00	0.97
40S Ribosomal protein S21	KADGIVSKNF	3	3	1077.59	1077.582	4	0.84	0.95	0.95	0.70
60S acidic ribosomal protein P2 polymorphism	DGKNIEDVIAQGIGKL	3	3	1668.92	1668.905	10	0.84	0.89	0.92	0.74
Heat shock 10kDa protein	AAETVTKGGIMLPEKSQGKVLQA	4	4	2355.32	2355.283	15	0.86	1.10	1.04	1.08
Heat shock 10kDa protein	GSGSKGKGGEIQPVSV	3	3	1485.79	1485.779	7	0.91	0.98	0.94	0.91
Superoxide dismutase 1	KGDGPVQGIINF	2	2	1243.67	1243.656	14	0.92	0.96	0.94	0.89
Prefoldin subunit 1	Ac-AAPVDLELKKAFTEL	2	2	1685.96	1685.924	22	0.92	0.92	0.95	0.99
Heterogeneous nuclear ribonucleoprotein D-like	KDAASVDKVLEL	3	3	1286.72	1286.708	7	0.93	0.89	0.97	0.95
Triosephosphate isomerase 1	SLGELIGTLNA	2	1	1086.62	1086.592	22	0.94	0.90	0.94	0.81
Heat shock 10kDa protein 1	LPLFDRVLVE	2	1	1199.72	1199.691	25	0.94	1.04	1.00	1.13
Triosephosphate isomerase 1	LDPKIAVA	2	2	825.50	825.496	3	0.94	1.06	0.94	0.94
Ubiquitin-60S ribosomal protein L40	IIEPSLRQL	2	1	1067.66	1067.634	21	0.94	0.97	0.94	0.89
60S Ribosomal protein L31	KNLQTVNVDEN	2	2	1272.64	1272.631	8	0.94	1.00	0.94	0.94
Elongation factor 1 beta	GFGDLKSPAGLQVL	2	2	1400.79	1400.770	12	0.95	1.05	0.89	1.03
Vimentin	LIKTVETRDGQVINETSQ	3	2	2030.11	2030.064	23	0.95	1.10	0.95	0.90
FK506 Binding Protein	VFDVELL	2	1	833.47	833.453	21	0.95	0.95	0.95	0.92
Nucleophosmin	ASIEKGGSLPKVEA	2	3	1384.76	1384.756	0	0.97	0.97	0.91	0.86
Cytochrome c oxidase subunit 5a	GISTPEELGLDKV	2	2	1356.73	1356.714	14	0.97	1.06	0.95	0.95
Peptidylprolyl isomerase A	ADKVPKTAENFRAL	3	3	1558.86	1558.847	8	0.97	1.09	0.99	0.99
Complement component 1 Q subcomponent-binding protein, mitochondrial	DRGVDNTFADELVELSTA	2	1	1950.96	1950.917	20	0.98	1.07	1.07	0.86
Heterogeneous nuclear ribonucleoprotein A/B isoform 1	FGEFGEIEAIEL	2	1	1352.69	1352.650	29	1.00	1.15	1.13	0.88
Nucleophosmin	GGFEITPPVVL	2	1	1127.65	1127.623	23	1.00	1.06	1.03	0.94
Peptidylprolyl isomerase A	ELFADKVPKTA	3	3	1217.67	1217.666	4	1.00	1.10	1.05	0.95
Heat shock 10kDa protein	TVVAVGSGSKGKGGEIQPV	3	3	1768.99	1768.968	13	1.00	1.09	0.85	0.96
FK506 Binding Protein	VFDVELLKLE	2	2	1203.70	1203.675	17	1.00	0.97	0.97	0.96
Peptidylprolyl isomerase A	ADKVPKTAENF	3	3	1218.62	1218.624	-3	1.01	0.99	0.99	0.89
Nucleophosmin	GGSLPKVEA	2	2	856.47	856.465	9	1.03	1.11	0.97	0.87

40S Ribosomal protein S21	AKADGIVSKNF	2	3	1148.62	1148.619	-1	1.04	1.05	0.95	0.93
40S Ribosomal protein S21	ADGIVSKNF	2	2	949.49	949.487	8	1.04	1.04	1.00	0.96
Elongation factor 1 beta	GFGDLKSPAGL	2	2	1060.57	1060.555	11	1.04	1.10	1.00	1.12
FK506 Binding Protein	GVQVETISPGDGRTFPKRGQ	4	2	2128.16	2128.102	25	1.05	1.07	1.03	1.02
Heterogeneous nuclear ribonucleoprotein D0	FGGFGEVESIEL	2	1	1282.64	1282.608	27	1.06	1.03	0.89	0.89
RNA binding motif protein 3	Ac-SSEEGKLFVGGLNF	2	1	1524.78	1524.746	25	1.06	1.06	1.06	0.94
Peptidylprolyl isomerase A	ELFADKVPKTAENFRAL	4	3	1948.05	1948.042	5	1.06	1.09	1.06	1.12
FK506 Binding Protein	VELLKLE	2	2	842.52	842.511	6	1.06	1.04	1.00	1.00
Eukaryotic translation initiation factor 4H	ATPLNQVANPNSAIFGGARPREEVVQ KEQE	4	2	3248.73	3248.654	24	1.07	1.13	1.00	0.93
Protein DJ-1 (Parkinson disease protein 7)	APLVLKD	2	2	754.46	754.459	-3	1.07	1.00	0.93	1.07
60S acidic ribosomal protein P2	VGIEADDDRLNKV	3	2	1442.76	1442.736	13	1.07	1.10	1.04	1.10
Elongation factor 1 beta	GFGDLKSPAGLQV	3	2	1287.70	1287.682	14	1.07	1.12	1.06	1.13
Peptidylprolyl isomerase A	VNPTVFFDI	2	1	1050.57	1050.539	26	1.08	1.06	1.00	1.00
40S Ribosomal protein S29	AKDIGFIKLD	3	3	1118.64	1118.634	3	1.08	1.06	1.00	1.00
Protein SET (Phosphatase 2A inhibitor I2PP2A) - isoform 2	Ac-SAPAAKVSKKEL	2	3	1269.73	1269.729	3	1.09	1.13	1.06	1.15
Heat shock 10kDa protein 1	AVGSGSKGKGGEIQPVSV	3	3	1655.89	1655.884	2	1.09	1.03	0.99	0.92
Heterogeneous nuclear ribonucleoprotein D-like	ASVDKVLEL	2	2	972.56	972.549	9	1.10	1.14	1.00	1.14
Heat shock 10kDa protein	VAVGSGSKGKGGEIQPVSV	3	3	1754.97	1754.953	11	1.11	1.06	0.83	0.98
Nucleophosmin	EKGGSLPKVEA	3	3	1113.60	1113.603	-5	1.12	1.12	1.12	0.86
Heat shock 10kDa protein 1	TVVAVGSGSKGKGGEIQPVSV	4	3	1955.07	1955.069	2	1.12	1.08	0.98	0.88
Complement component 1 Q subcomponent-binding protein, mitochondrial	ADRGVDNTFADELVEL	2	1	1762.88	1762.837	24	1.14	1.08	0.97	1.00
Heat shock 10kDa protein 1	VGSGSKGKGGEIQPVSV	3	3	1584.85	1584.847	2	1.14	1.09	1.00	1.09
Protein SET (Phosphatase 2A inhibitor I2PP2A) - isoform 1	SELIAKI	2	2	772.47	772.469	3	1.14	1.00	1.00	0.86
40S ribosomal protein S12	Ac-AEEGIAAGGVMDVNTALQEVLKT	3	1	2357.26	2357.178	35	1.15	1.13	1.15	1.06
40S Ribosomal protein S28	Ac-MoxDTSRVQPIKL	2	1	1344.74	1344.712	19	1.16	1.04	0.98	0.98

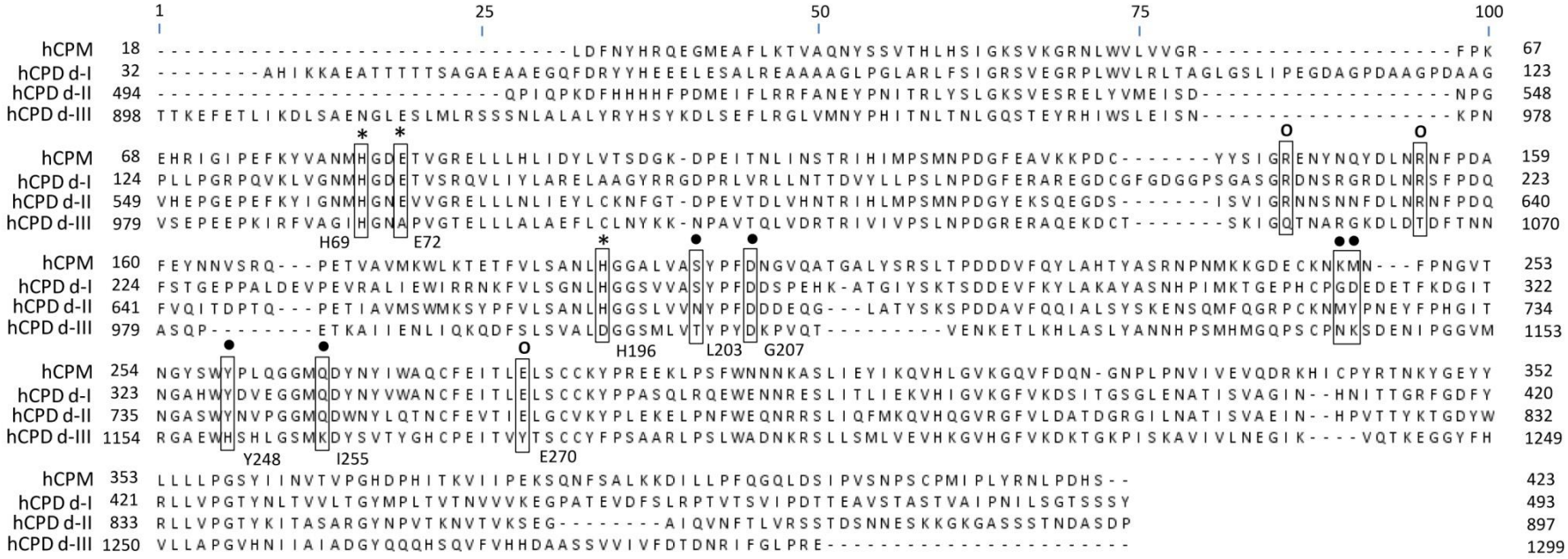
See Table 2 for the abbreviation definitions.

Supplemental Figure 1



Suplemental Figure 1. Relative amount of product formed by three different dansylated tripeptides incubated with various amount of purified rhCPD. Reactions containing 200 µM dansyl-Phe-Ala-Arg (filled circles), dansyl-Phe-Gly-Arg (filled squares) or dansyl-Phe-Pro-Arg (triangles/solid line) were incubated with different amounts of enzyme in a 100 mM Tris-acetate, pH 6.5, 150 mM NaCl buffer for 60 min at 37ºC.

Supplemental Figure 1



Supplemental Figure 2. Alignment of human CPD domains I, II and III, and human CPM. Stars were used to indicate residues involved in zinc binding. Closed circles show the location of essential catalytic residues. Open circles mark important residues involved in substrate binding and substrate specificity determination. Sequences of human CPD domains I, II and III correspond to residues 32-493, 494-897 and 898-1299, respectively. Sequence of human CPM corresponds to residues 18-423. Uniprot accession codes for both proteins sequences are O75976 for human CPD and P14384 for human CPM. The position numbers below the sequence indicate equivalent positions in the standard numbering system for bovine CPA. (Larkin et al. 2007).

Chapter IV

A simple method to improve recombinant protein production of Heparin-Affinity carboxypeptidases using mammalian cells

CHAPTER IV: A SIMPLE METHOD TO IMPROVE PROTEIN PRODUCTION OF HEPARIN-AFFINITY CARBOXYPEPTIDASES USING MAMMALIAN CELLS

4.1 INTRODUCTION

The rapid progress of biotechnology and biomedicine has created a constant need to produce a widely variety of structurally complex recombinant proteins. Metallocarboxypeptidases (MCPs) are exopeptidases that cleave C-terminal residues from its substrates which in the recent years have been emerged as promising drug targets in biomedicine (Arolas et al. 2007). Production of these enzymes can often require particular post-translational modifications, molecular chaperones and co-factors to support their elaborate folding and enzymatic activity. To solve these limitations the use of baculovirus or mammalian expression systems has been increasing because they are able to produce complex eukaryotic proteins that are otherwise problematic to express in other systems. Some of MCPs are proteins that have post-translational modifications and heparin-affinity properties which can lead in lack of protein production following classic procedures.

Here we describe a simple, fast, inexpensive and highly efficient method for expressing heparin-affinity MCPs in suspension grown mammalian cells. The approach combines transient expression in Human Embryonic Kidney 293F (HEK 293F) cells with the addition of sodium heparin after transfection to improve protein solubility of heparin-affinity MCPs. The cell line used has been derived from HEK 293 cells which have been adapted to grow in suspension using a serum-free media (Vink et al. 2014). To perform protein transfection we used polyethylenimine (PEI), an inexpensive polymeric reagent that has been reported to form complexes with DNA that can enter in the host cell and drive to protein expression (Tom et al. 2008). The addition of heparin 48 hours post-transfection allows the accumulation and consequently production of those heparin affinity MCPs.

This method is suitable for both small scale (~25 ml) and large-scale (up to 2000 ml) experiments and can produce high levels of purified MCPs. It is particularly useful for studying MCPs that require complex folding machineries, particular post-

translational modifications and present heparin-affinity properties that cannot be produced by bacteria, yeast or insect cells.

In the protocol described here, we present the high level expression of three human heparin affinity MCPs (carboxypeptidase Z (CPZ), carboxypeptidase A6 (CPA6) and Trombin-Actibable Fibrinolysis Inibitor (TAFI)). Further, we performed as representative example the purification of the human Carboxypeptidase Z form the extracellular medium. The purified protein is enzymatically active and can be used for high-throughput functional and structural studies.

4.2 EXPERIMENTAL SECTION

4.2.1 CELL CULTURE

HEK 293F cells (ATCC CRL-3216) were grown in FreeStyle 293 expression medium (Invitrogen, Inc.) in flasks on a rotary shaker (120 rpm) at 37°C, 8% CO_2 and 70% humidity. For maintenance, the cell culture was diluted each 48-72 hours, to maintain the cells at a density between $0.2x10^6$ and $3.0 x10^6$ cells/ml.

4.2.2 OPTIMIZED PROTOCOL FOR LARGE-SCALE CELL CULTURE TRANSFECTION

1.1 Seed HEK 293F cells at 0.5×10^6 cells/ml into a final volume of 450 ml in each 2-L shaker flask (See note 1 and 2 below).

1.2 Incubate for 24 hours in an orbital shaker incubator at 37 °C, 120 rpm, and 5% CO_2 until cells reach a density of 1.0×10^6 cells/ml.

1.3 Pipette a total of 500 µg of DNA (see Note 3) into 50 ml of FBS-free culture medium and vortex vigorously for 30 seconds (1 µg of DNA per millilitre of cell culture).

1.4. Add 1.5 ml of 1.0 mg/ml filter-sterilized PEI to the medium/DNA solution and vortex vigorously for 30 seconds.

1.5 Incubate the mix at room temperature 15-20 minutes.

1.6 Add the DNA/PEI mix to the cells

1.7 Following transfection, incubate the cells in an orbital shaker incubator at 37 °C, 120 rpm, and 5% CO2

1.8 After 48 hours post-transfection, add 0.5 ml of sodium heparin solution to the cells (see note 4).

1.9 Incubate the cells in an orbital shaker for additional 5-10 days at 37 ºC, 120 rpm, and 5% CO^2 (see note 5)

1.10 Harvest the cells by centrifuging at 3,000 x g for 5 min. Process immediately the extracellular conditioned medium or store at -80 °C, until purification.

Note 1: The protocol is suitable for any scale of expression, therefore volumes and quantities of reagents should be scaled proportionally. A suitable mammalian expression vector must be used for this protocol. Here we used both pTriEx-7 and pCDNa-3 expression vectors, that conveniently allow the extracellular expression of heparin-affinity proteins.

Note 2: Typically, cell culture flasks can accommodate a minimum of 1/10 of its nominal volume to a maximum of 1/4 of its nominal volume of suspension culture. For larger scale transfections use multiple culture flasks.

Note 3: Use plasmidic DNA with an appropriate quality, suitable for cell culture. Typically, DNA should be sterile and endotoxin-free, as well as free from other major contaminants.

Note 4: Use a sterile sodium heparin solution with 5000 IU/ml or 50 mg/ml, suitable for cell culture.

Note 5: The optimal incubation time for maximum protein expression should be determined for each recombinant protein.

4.2.3 OPTIMIZED PROTOCOL FOR PROTEIN PURIFICATION

2.1. Defrost the medium and add a cocktail of EDTA-free protease inhibitors (if is necessary, depending on our protein)(See note 6 and 7).

2.2 Equilibrate 10.0 ml of Heparin-affinity resin per liter L of culture medium by washing three times with resin equilibration buffer (100 mM tris-HCl, 100 mM NaCl, pH 7.4).

2.3 Pass through the equilibrated column the conditioned medium containing the recombinant protein, discard the flow-through.

2.4 Wash the resin with 20 ml of equilibration buffer 1 (100 mM tris-HCl, 100 mM NaCl, pH 7.4).

2.5 Elute the recombinant protein applying an increasing gradient of NaCl up to 1.5 M using the same equilibration buffer. Eluting buffer could be increased stepwise and

applied to the column (e.g. 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0 M NaCl). Alternatively, a linear NaCl gradient could be applied using an automated FPLC system. Collect a 10 µl sample of each eluate and add protein loading buffer for analysis.

2.7 Equilibrate 5.0 ml of anti strep-tag affinity resin per liter L of culture medium by washing three times with resin equilibration buffer (100 mM tris-HCl, 150 mM NaCl, pH 8.0).

2.8 Wash the resin with 25 ml of equilibration buffer 1 (100 mM tris-HCl, 150 mM NaCl, pH 8.0).

2.9 Elute the recombinant protein with 60 ml (12 column volumes) of elution buffer (100 mM tris-HCl, 150 mM NaCl, pH 8.0 and 2.5 mM d-*desthiobiotin*). Collect a 10 µl sample of each eluate and add protein loading buffer for analysis.

2.10 Equilibrate the size exclusion chromatography column with gel filtration buffer (25 mM tris-HCl, 150 mM NaCl, pH 8.0).

2.11 Filter the protein through a 0.22 µm filter and load the sample into the column and collect 10 µl of each eluted fraction

2.12 Run samples from steps 2.6, 2.9 and 2.11 and the gel filtration fractions from step on an SDS-PAGE and Coomassie stain for analysis.

Note 6: The protocol of purification described here has been optimized for the purification of human carboxypeptidase Z, therefore for other heparin-affinity proteins, some steps can be optimized to assure a maximum protein recovery.

Note 7: It is advisable to run an SDS-PAGE of samples from the initial conditioned medium prior to purification to confirm expression of the target protein.

4.2.4 CITOTOXITICY ASSAYS

HEK-293 F cells (ATCC CRL-3216) were seeded into 96-well plates at cell densities of 3.0x103 cells/well and incubated for 24 hours before sodium heparin (Hospira Prod. Farm. y S.L.) was added at a concentration from 0 to 2000 UI/ml. Growth inhibitory effect was measured after 24 and 72 hours treatment by the XTT assay (Núñez et al.

2014). Briefly, aliquots of 20 µl of XTT solution (2,3-bis-(2-methoxy-4-nitro-5-sulfophenyl)-2Htetrazolium-5-carboxanilide) were added to each well. After 4 hours the color formed was quantified by a spectrophotometric plate reader (Perkin Elmer Victor3 V) at 490 nm. Cell cytotoxicity was evaluated in terms of cell grown inhibition in treated cultures and expressed as % of the control condition.

4.2.5 CARBOXYPEPTIDASE ACTIVITY

Enzyme activity was typically assayed with the fluorescent substrate dansyl-Phe-Ala-Arg. To perform the experiments, reactions of 100 µl containing 0.2 mM of dansyl-Phe-Ala-Arg in 100 mM tris-acetate, pH 7.5, 100 mM NaCl buffer were incubated with a variable amount of enzyme for 60 min at 37 ºC. After incubation reactions were stopped by adding 50 µl of 0.5 M HCl. Then, 1 ml of chloroform was added to each reaction and tubes were mixed gently and centrifuged for 2 min at 300 x g. After centrifugation, 0.5 ml of the chloroform phase were transferred to new tubes and completely dried for overnight at 25 ºC. Finally, dried samples containing mainly the product generated in the enzymatic reaction were resuspended with 200 µl of PBS containing 0.1 % of Triton X-100. The amount of product generated was determined by measuring the fluorescence of samples at 395 nm upon excitation at 350 nm using a 96-well plate spectrofluorometer.

4.3 RESULTS

4.3.1 HEPARIN AFFINITY METALLOCARBOXYPEPTIDASES

Numerous proteins, including cytokines and chemokines, enzymes and enzyme inhibitors, extracellular matrix proteins, and membrane receptors, bind heparin. Although they are traditionally classified as heparin binding proteins, under normal physiological conditions these proteins actually interact with the heparan sulfate (HS) chains of one or more membrane or extracellular proteoglycans. Both Heparin and heparin sulfate (HS) are negatively charged, polydisperse linear polysaccharides. They are composed of α1-4 linked disaccharide repeating units containing a uronic acid and an amino sugar. They share common biosynthetic routes, as well as overall biochemical properties (**Figure 1-A**).

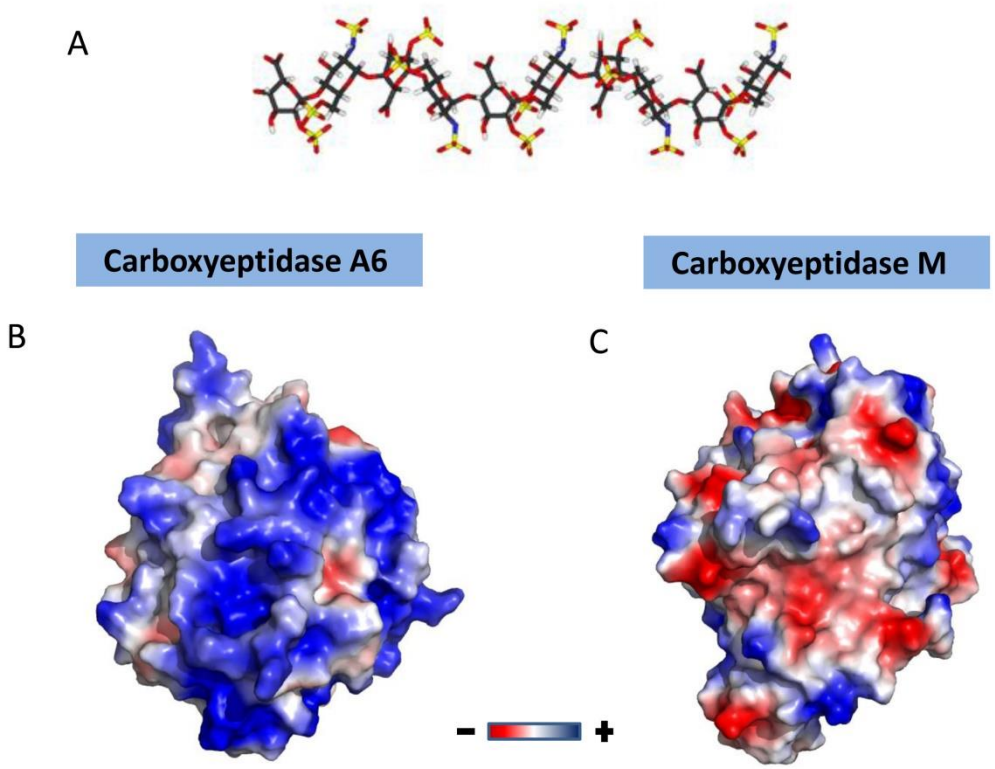


Figure 1. Heparin/HS affinity metallocarboxypeptidases (A) Stereo view of a heparin-derived duodecasaccharide in stick representation. (B) Electrostatic surface potential distribution of the catalytic domain of human CPA6 (modelled structure using I-TASER). (C) Electrostatic surface potential distribution of the catalytic domain of human CPM (based on protein data bank identifier 1UWY). Blue indicates positive and red indicates negative charge potential.

Although previous studies on CPA6, TAFI and CPZ demonstrated that these three carboxypeptidases have affinity by heparin (Lyons et al. 2008; Novikova et al. 2000; Foley et al. 2013; Sanglas et al. 2008; Goldstein et al. 1989) , this property has never been extensively studied within the MCPs protein family. For this, here we analysed the biochemical properties of all members of the M14 subfamily of proteases (see **Supplementary Table 1**).

Because generally Heparin/HS perform its biological functions at the cell surface and in the extracellular matrix, the protein should be present in a relevant biological context. The analysis of the biochemical properties of the majority of MCPs found in humans, show that among them, only four secreted MCPs display basic isoelectric points (pIs); CPA3, CPA6, TAFI and CPZ. The rest of extracellular MCPs have neutral to acidic pIs. It is intriguing that many proteins cytosolic carboxypeptidases (CCP2, CCP3 and CCP5) with its localization in the cytoplasm or the nucleus although display basic pIs, probably due to its involvement in the processing of the C-terminal acidic tails of tubulin and other nuclear proteins (Tanco et al. 2015)

It is widely accepted that best method to identify a heparin binding protein is to calculate the surface electrostatic potential. In the absence of a three-dimensional structure, homology modeling can be performed. Electrostatics clearly plays a major role in heparin/HS–protein interactions. As example, in **Figure 1-B** is shown the modeled structure of the catalytic domain of human CPA6, showing its electrostatic surface potential distribution. In the CPA6 model, is clearly visible the large number of basic residues (mainly Arg and Lys) located at the surface. Similar electrostatic surface potential distribution patterns were observed for CPA3, TAFI and CPZ (data not shown). In contrast, the electrostatic surface potential distribution in CPM, as example for MCP without a basic pI, shows a random distribution of the surface charges (see **Figure 1-C**). Thus, we can consider that only CPA3, CPA6, TAFI and CPZ are potential heparin/HS binding MCPs.

4.3.2 IMPROVED EXPRESSION OF HEPARIN-AFFINITY MCPS

There are numerous failed attempts to produce recombinant secreted heparin-affinity MCPs, following conventional expression systems (such as bacteria, yeast or insect cells). In general, overexpression of the genes in cultured cells resulted in a slight increase in the secreted protein, which could not be purified. Only, for CPA6, a recombinant His tagged form was previously stably expressed in mammalian HEK293 cells, obtaining relative good levels of active protein (Lyons et al. 2008).

To develop a new system to produce these enzymes, the plasmid pTriExTM-7 containing a recombinant form of human CPZ, was used to optimize transfection conditions for HEK 293F cell line. The CPZ levels in the extracellular medium were determined by western blot analysis. A wide variety of expression conditions were tested (including DNA amount, transfecting agent, their ratios and diluents, medium, cell density during transfection, as well as the supplementation with additives), and the best results were obtained following the optimized protocol described in the experimental section and **Figure 2**.

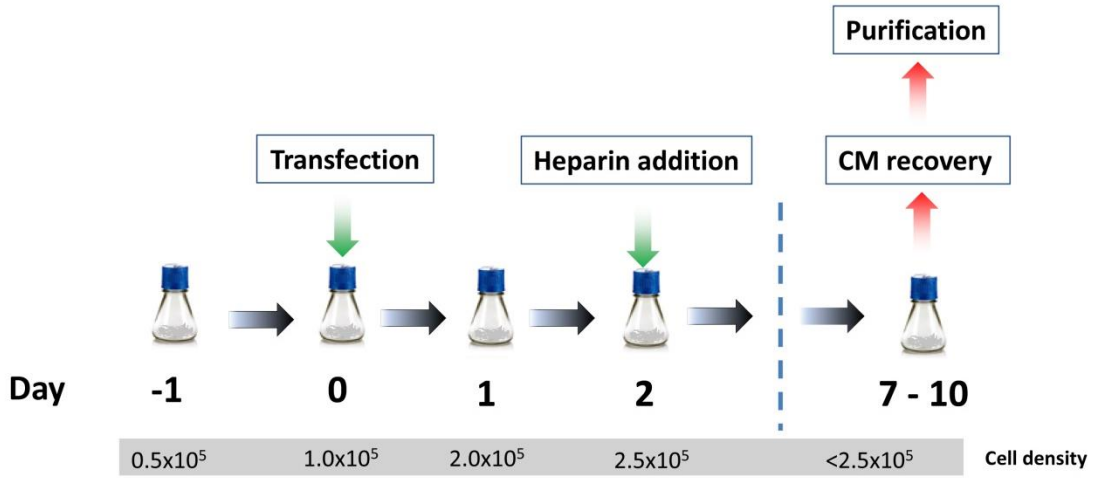


Figure 2. Schematic representation of the optimized procedure for production of Heparin-affinity MCPs. HEK 293F cells were seeded at a cell density of $0.5x10^5$ cells/ml. After 24 hours, cells were transfected with the DNA/PEI mix and after additional 48 hours, the cell culture was supplemented with 50 UI/ml of sodium heparin to improve protein production. After the appropriate incubation time (typically between 7 and 10 days), the cell culture can be recovered, centrifuged and the conditioned medium stored until protein purification.

The most important goal of the present procedure was the addition of sodium heparin to the culture after 48 hours post transfection, probably due to its solubilizing and protective effect for the recombinant protein. For this, we evaluated the cytotoxicity of this compound against HEK 293F cells, in order to determine the optimal concentration to be employed in the assay (**Figure 3**). We decided that the best concentration was 50 IU/ml, since it does not cause detectable toxicity to HEK 293F cells,at any of the incubation times assayed.

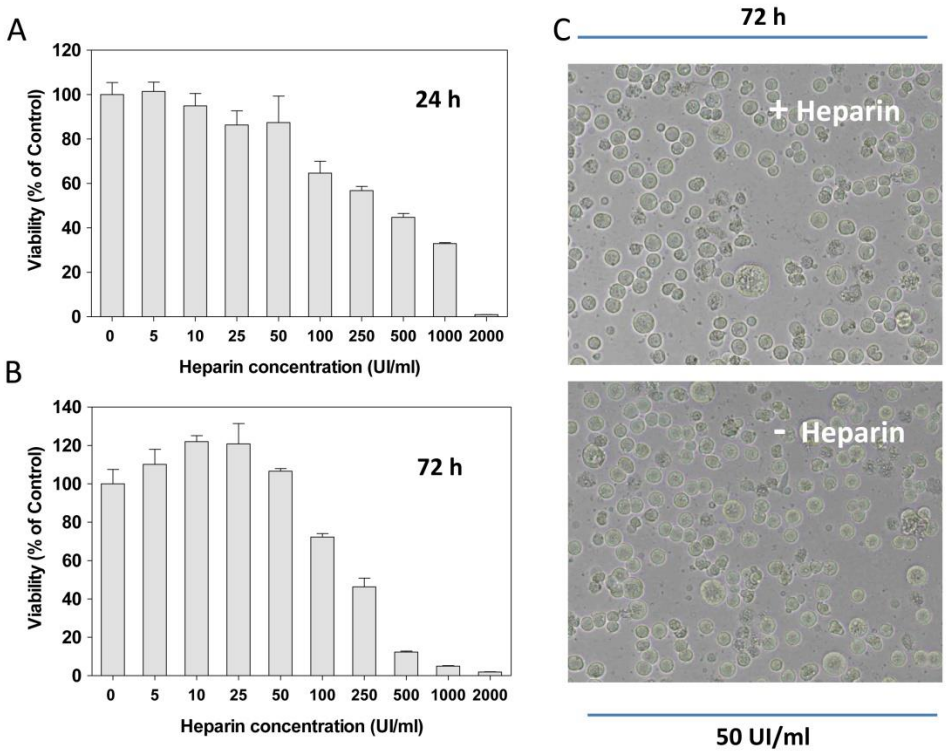


Figure 3. Citotoxicity of sodium Heparin against HEK 293F cells. The cytotoxicity of sodium heparin against HEK 293F cells was evaluated after 24 (A) and 72 (B) hours treatment with heparin at a concentration up to 2000 IU/ml. 50 IU/ml was the highest concentration of sodium heparin that did not caused notable cytotoxicity against HEK 293F. (C) Representative images of HEK 293F cells after 72 hours incubation, with or without 50 IU/ml of sodium heparin.

We also quantified the amounts of protein expression by densitometric analysis of the bands, it was estimated that the addition of 50 IU/ml of sodium heparin to the culture (after 48 hours pot-transfection) resulted in an increase of about ~8 fold in the amount of CPZ present in the extracellular medium of HEK 293F cells after 5-8 days

post-transfection, in comparison with the same protocol performed without the addition of this compound (**Figure 4**).

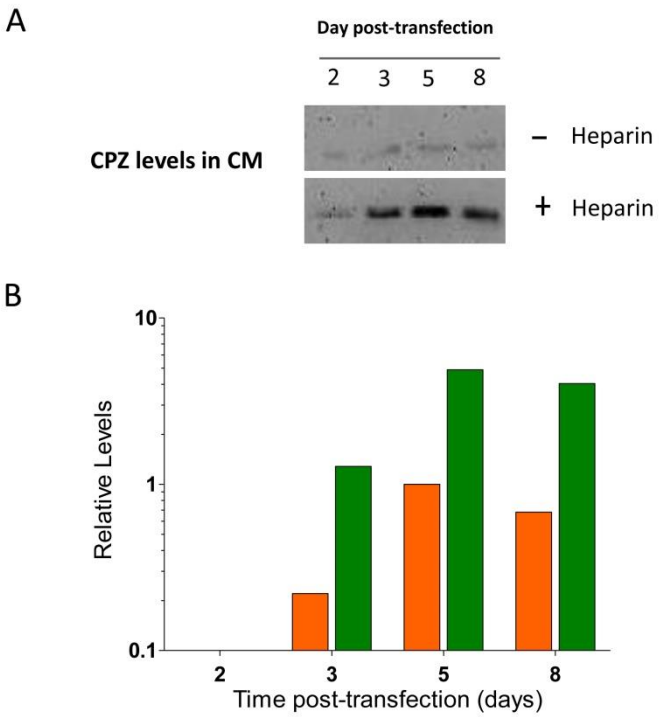


Figure 4. Effect of addition of sodium heparin on human CPZ production by HEK 293F cells. (A) Immunoblots of the human CPZ expression over time in HEK 293F cells in presence or absence of 50 IU/ml sodium heparin. Expression of CPZ was analysed with anti-Strep-tag. (B) Representative relative expression levels of human CPZ over time in HEK 293F cells. Values where normalized to the maximum CPZ signal detected in the conditioned medium in absence of heparin.

By using the same production protocol described for human CPZ, we evaluated the expression of other MCPs predicted to have heparin/HS binding properties, such as human TAFI or CPA6 (**Figure 5**). For TAFI, protein levels in the conditioned medium were increased in 10-fold after 5 days post-transfection, in comparison without heparin addition. In the case of CPA6 proteins levels raised about 90-fold after 6 days post-transfection in presence of heparin. The same protocol did not increased levels of other MCPs or proteins without heparin affinity properties, such as human CPD or human α-galactosidase (see **Figure5**).

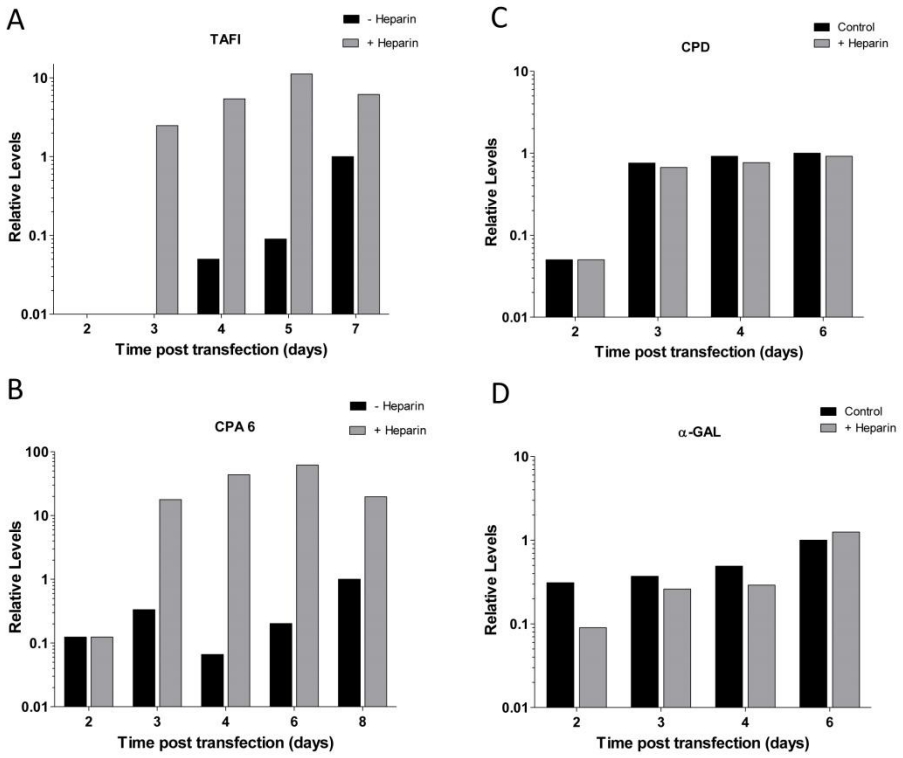


Figure 5. Effect of addition of sodium heparin on the production of TAFI, CPA6, CPD and α-galactosidase. Relative expression levels of human TAFI (A), CPA6 (B), CPD (C) and α-galactosidase (D). Expression of TAFI, CPD and α-galactosidase were analysed using an anti-Strep-tag antibody. Expression of CPA6 was detected using and anti HA-antibody. Values where normalized to the maximum protein signal detected in the conditioned medium in absence of heparin, for each case.

4.3.3 REPRESENTATIVE RESULTS FOR THE PURIFICATION OF THE HUMAN CARBOXYPEPTIDASE Z

Once best expression conditions were stablished, next step consisted in produce large amounts of conditioned medium for its purification. For that, 2-L shaker flasks were used to grow cell cultures of 450 ml. Such volumes were transfected with 50 ml of medium containing pTriExTM-7-CPZ and PEI complexes. After 48 hous post-transfection, the cell culture was supplemented with 50 IU/ml of sodium heparin. Then, at 9 days post-transfection, the conditioned medium from the culture was recovered and the protein purified through and optimized three-step purification protocol (**Figure 6**).

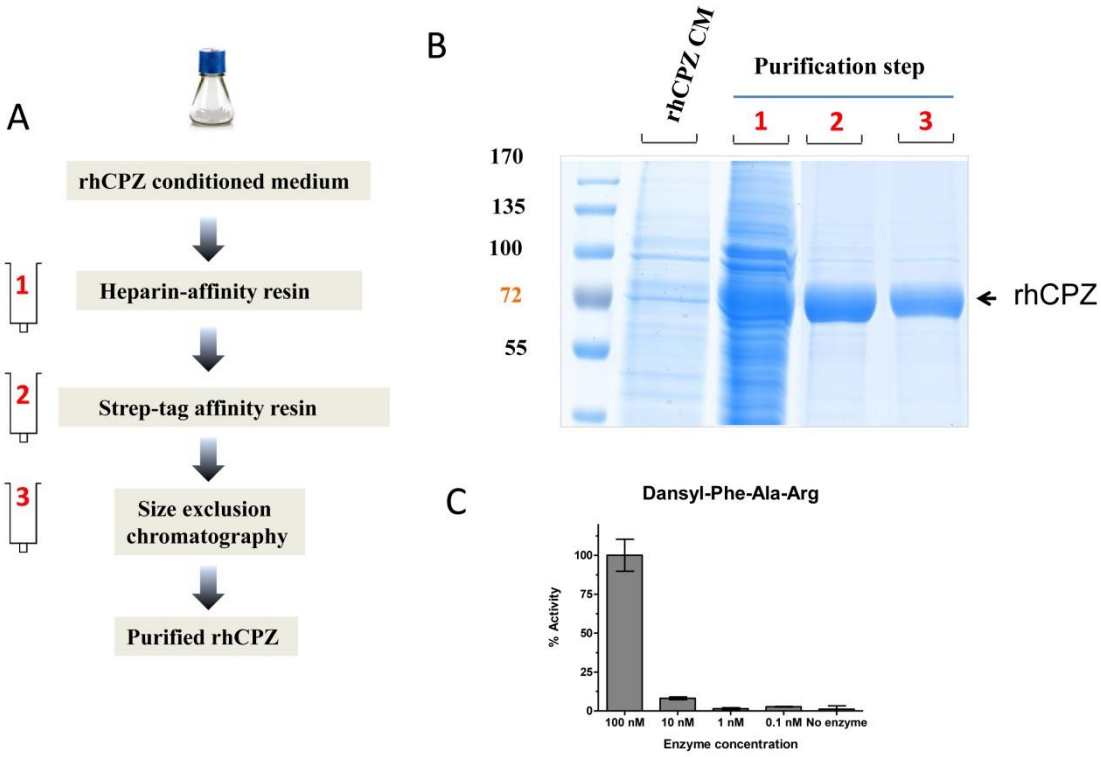


Figure 6. Purification of recombinant human CPZ. (A) Schematic diagram of the protocol followed for the purification of rhCPZ. For protein purification, the extracellular medium was collected after 9 days incubation, and recombinant proteins purified in three purification steps by (1) heparin-affinity chromatography, (2) affinity chromatography using anti-strep tag resin and by (3) gel-filtration chromatography. (B) An aliquot of the initial conditioned medium, as well as from the elution of each purification step, were visualized on SDS–PAGE by coomassie staining. (C) The purified protein shows a concentration-dependent activity against typical CPB-like substrates.

For protein purification, the clarified conditioned medium was applied into a heparin-affinity resin (Heparin HyperD®, PALL life sciences). Although we added sodium-heparin to the cell culture, this additive did not affect the ability of CPZ to bind the heparin-affinity resin. Eluted fractions containing CPZ from the first purification step were pooled and loaded into an anti-Strep tag affinity resin and further purified using a size exclusion chromatography (**Figure 6-A and 6-B**).

The resultant purified protein could be easily visualized on SDS–PAGE by Coomassie blue staining, free of any major contaminating proteins, suggesting that human CPZ can be efficiently purified directly from the conditioned medium, using this methodology, to a high degree of purity. The purified recombinant protein was enzymatically active, showing a concentration-dependent activity against typical substrates (see **Figure 6-C**).

4.4 DISCUSSION

We have developed a straightforward and efficient method for producing large amounts of recombinant heparin-affinity MCPs from mammalian cells. To achieve this, we used pure plasmid DNA of a high quality (ratio A_{260}/A_{280} between 1.8 and 2.0) which was mixed with PEI as transfecting agent and added to the cell culture as described in the protocol section. The HEK 293F cells used must be cultured in serum- and antibiotic-free media, therefore, sterile technique is strictly required for passaging and transfecting cells in order to avoid costly and time-consuming contaminations. For a successful transfection, cell viability should be over 90% and cultures should only contain single or dividing cells, with a cell density during culture maintenance between $0.5x10^6$ and 3.0×10^6 cells/ml, to keep the cells in optimal conditions. The ratio of DNA/PEI in the transfection and the time of expression can be optimized for each protein in order to have the best expression efficiency. Based on the cytotoxicity of sodium heparin against HEK 293F cells, we determined a fixed amount of heparin to be added in the culture that does not cause cell dead. However, the amount of sodium heparin added and the supplementation time can be adjusted for each protein and or experimental conditions to reach the best expression efficiency. Also, before scaling-up transfections, small scale transfections are particularly useful for testing different vectors and/or constructs containing different tags.

For protein purification, the choice of tags and purification buffers is critical, since they may interfere with structural elements or active sites of certain proteins, often resulting in reduced solubility and/or loss of enzymatic activity. Furthermore, is desirable to perform all the steps, keeping the protein sample cold, to reduce the risk of unwanted proteolytic degradation. Following the methodology described above, we purified recombinant human CPZ to a high degree of purity (see **Figure 6-B**), obtaining about 1-2 mg of pure protein for each L of cell culture. Previous trials showed that this enzyme is expressed but does not fold in *E. coli*. Other previous attempts to produces this protein, using baculovirus cells or mammalian cells without the addition of heparin, resulted in low amounts of recombinant protein production.

4.5 SUPPLEMENTAL INFORMATION

Supplemental Table 1. Properties of the majority of human MCPs

	Uniprot Code	Name	Localization	pI*	ECM/Heparin binding
M14A Subfamily	P15085	Carboxypeptidase A1	Extracellular	5.5 / 5.9	Unknown
	P48052	Carboxypeptidase A2	Extracellular	5.5 / 6.3	Unknown
	P15086	Carboxypeptidase B	Extracellular	6.2 / 6.7	Unknown
	P15088	Mast cell carboxypeptidase A (CPA3)	Extracellular	**9.2 / 9.5**	Unknown
	Q9UI42	Carboxypeptidase A4	Extracellular	6.1 / 7.1	Unknown
	Q8WXQ2	Carboxypeptidase A5	Unknown	6.0 / 5.8	Unknown
	Q8N4T0	Carboxypeptidase A6	Extracellular	**9.5 / 9.5**	**Yes**
	Q96IY4	Carboxypeptidase B2 (TAFI)	Extracellular	**7.7 / 8.1**	**Yes**
	Q8IVL8	Carboxypeptidase O	Extracellular	6.6	Unknown
M14B Subfamily	O75976	Carboxypeptidase D	TGN / Extracellular	5.6	Unknown
	P14384	Carboxypeptidase M	Cell membrane / Extracellular	6.7	Unknown
	P15169	Carboxypeptidase N catalytic chain	Extracellular	6.9	Unknown
	P16870	Carboxypeptidase E	Secretory vesicles	4.9	Unknown
	Q66K79	Carboxypeptidase Z	Extracellular	**8.3 / 9.3**	**Yes**
	Q8IUX7	Adipocyte enhancer-binding protein 1	Cytosolic	5.0	Unknown
	Q96SM3	Carboxypeptidase X1	Cytosolic	6.2	Unknown
	Q8N436	Carboxypeptidase-like protein X2	Cytosolic	6.4	Unknown
M14B Subfamily members	Q9UPW5	Cytosolic carboxypeptidase 1	Cytosolic / Nuclear	5.8	Unknown
	Q5U5Z8	Cytosolic carboxypeptidase 2	Cytosolic / Nuclear	9.1	Unknown
	Q8NEM8	Cytosolic carboxypeptidase 3	Cytosolic / Nuclear	9.0	Unknown
	Q96MI9	Cytosolic carboxypeptidase 4	Cytosolic / Nuclear	6.9	Unknown
	Q8NDL9	Cytosolic carboxypeptidase 5	Cytosolic / Nuclear	9.3	Unknown
	Q5VU57	Cytosolic carboxypeptidase 6	Cytosolic / Nuclear	5.8	Unknown

*pI values where calculated with the ProtParam tool from Expasy (http://web.expasy.org/protparam/).
The first value indicates the pI of the proform, while the second value indicates the pI of the active form.
In the case of CPZ the second pI value correspond to the enzyme without the frizzled like domain.

Chapter V

Substrate specificity and structural modeling of human carboxypeptidase Z: The unique protease with a frizzled-like domain

CHAPTER V: SUBSTRATE SPECIFICITY AND STRUCTURAL MODELING OF HUMAN CARBOXYPEPTIDASE Z: THE UNIQUE PROTEASE WITH A FRIZZLED-LIKE DOMAIN

5.1 INTRODUCTION

Metallocarboxypeptidases (MCPs) are zinc-containing exopeptidases that cleave single C-terminal amino acids from proteins and peptides (Arolas et al. 2007). The M14B subfamily of MCPs, also known as CPE/N, is composed in humans by eight members thought to play important roles in the processing of neuropeptides and growth factors (Arolas et al. 2007; Reznik & Fricker 2001). All members of this subfamily share a similar structural architecture with a conserved carboxypeptidase (CP) domain and a C-terminal transthyretin-like domain (TTL domain) (Bayés et al. 2007; Reverter et al. 2004; Keil et al. 2007; Sebastián Tanco et al. 2010). The latter possess a β-sandwich folding related to transthyretin (TTR), and its function is yet uncertain; although, it might be involved in the folding, regulation of the enzyme and/or in membrane/protein binding (Gomis-Rüth et al. 1999). Some members within the M14B subfamily are catalytically inactive (AEBP1, CPX1 and CPX2) since they lack some relevant amino acids essential for the catalytic mechanism. Nonetheless, five members (CPE, CPN, CPM, CPD and CPZ) are catalytically active proteins with stringent B-like substrate specificity (*i.e.,* cleaving only C-terminal basic residues) (Reznik & Fricker 2001). The most studied members of this subfamily of MCPs are carboxypeptidase E (CPE), carboxypeptidase N (CPN) and carboxypeptidase M (CPM). CPE has been involved mainly in the removal of C-terminal lysine and arginine residues from neuroendocrine peptide precursors in the secretory pathway (Xin Zhang et al. 2008). CPN circulates in plasma and removes C-terminal arginine residues from bradykinin and anaphylatoxins C3a, C4a and C5a altering the affinity of these neuropeptides by its receptors (Keil et al. 2007). In a similar manner, CPM has been proposed to remove C-terminal basic residues from proteins and peptides in the extracellular space (Deiteren et al. 2009; Zhang et al. 2013).

Carboxypeptidase Z (CPZ) belongs to the M14B subfamily of MCPs. In humans, CPZ is synthesized as a constitutively active enzyme of about 72 kDa, which is secreted through the regulated pathway to the extracellular medium. Part of the secreted CPZ appears to be bound to the extracellular matrix (ECM), due to its heparin binding properties. Although previous studies suggested that the presence of a highly conserved C-terminal stretch of ~35 amino acids might contribute to the ECM binding, the exact mechanism by which CPZ is attached to the ECM remains still unclear (Novikova & Fricker 1999b). Currently, limited information is available about the substrate specificity, function, structure and/or interactors of CPZ (Reznik & Fricker 2001).

Distinct from other MCPs, CPZ contains an N-terminal domain with homology with Frizzled receptors (known as Frizzled-like domain or motifs) (Song & Fricker 1997). This Frizzled-like domain (Fz) is a cysteine-rich sequence exclusively present in a reduced number of protein families including seven-span transmembrane receptors, type XVIII collagen, secreted Frizzled-like related proteins (sFRP), receptor protein tyrosine kinases, LDL-receptor class-A domain and in MCPs. Typically, Frizzled domains function as receptors for members of the *Wingless* (WG), or its mammalian homolog, the Wnt family proteins (Diekmann 1998). The activation of these receptors by WG/Wnt proteins leads to the initiation of intracellular signaling pathways (commonly known as Wnt signaling pathways) that mediate vertebrate and invertebrate development and tissue homeostasis, thanks to their influence on cell proliferation, differentiation, and migration (Reya & Clevers 2005; Clevers & Nusse 2012). Dysregulation of the Wnt signaling has been associated with a variety of human pathologies such as several hereditary diseases or cancer (Anastas & Moon 2013).

The physiological role of the Frizzled-like domain present in CPZ remains a mystery. However, the presence of this domain, together with the localization of CPZ in the ECM, as well as its dynamic expression during development, suggest that CPZ might interact with Wnt proteins. In addition, current evidence indicates that several proteins containing Frizzled-like domains, such as sFRP can act as Wnt modulators, inhibiting and/or enhancing its biological activities and influencing gene expression (Leimeister et al. 1998; Bovolenta et al. 2008). In a similar manner, CPZ was proposed

to have a role in Wnt signaling, probably by inhibiting its function (Reznik & Fricker 2001). Only two studies have demonstrated that CPZ have the ability to bind Wnt-4 and modulate its signaling in skeletal elements in chicken (Moeller et al. 2003) or regulating the terminal differentiation of growth plate chondrocytes *in vitro* (Wang et al. 2009). Moreover, several human Wnts have encoded C-terminal Arg or Lys residues immediately after the last Cys residue (with the exception of Wnt-2, Wnt-8, and Wnt-9), susceptible to be removed by the action of CPZ. Nevertheless, it is not yet clear how the presence (or absence) of these C-terminal residues can modulate Wnt function.

In the work presented here, we have used a combination of fluorescent substrates and quantitative peptidomics approaches to gain insights into the substrate specificity of human CPZ, and to elucidate the contribution of the N-terminal Frizzled-domain to its catalytic activity. Furthermore, we performed a structural modeling of the catalytic domain, as well as the Frizzled-like domain of human CPZ to dissect its structural features.

Using enzymatic assays, quantitative peptidomics, and structural modeling, we find that CPZ is a carboxypeptidase that specifically cleaves substrates and peptides with C-terminal basic amino acids, especially those containing C-terminal Arg residues. Moreover, we have characterized the structure and functions of the Frizzled-like domain, through the study of the activity of a CPZ N-terminal truncated form, and through the study of its structural characteristics. Our work on human CPZ completes previous works with an extensive description of its substrate preferences, suggesting an important functional role of the Frizzled domain into Wnt binding.

5.2 EXPERIMENTAL SECTION

5.2.1 RECOMBINANT PROTEIN PRODUCTION AND PURIFICATION

Two forms of human carboxypeptidase Z were cloned for protein expression into the pTriEx™-7 expression vector (Merck Millipore) to encode a mouse IgM secretion signal sequence and a N-ter Strep-Tag® II fusion protein. The first form, containing both the catalytic and Frizzled- like domain (residues 19-623), named here as rhCPZ, and the second CPZ form without the N-terminal Frizzled-like domain (residues 186-623), named as CPZΔFz. Protein production was performed in mammalian cells as recently described (Garcia-Pardo, Tanco, Fernandez-Alvarez, et al. 2015) (see Chapter IV). In brief, DNA transfections of rhCPZ and CPZΔFz were carried out using 25 kDa polyethylenimine (PEI) (Polysciences), in a ratio of 1:3 (µg DNA/ µg PEI). HEK293F cells were diluted to a cell density of about $0.5 \cdot 10^6$ cells/ml, grown for 24 h and then transfected with 1 µg DNA per ml of culture. After 48 transfection sodium heparin was added to the culture, and then cells incubated for additional 8 days. For protein purification, the culture supernatant was centrifuged, filtered through 0.2 µm filters and bound to a heparin-affinity chromatography (Heparin HyperD®, PALL life sciences). Protein elution was carried out with an increasing gradient of NaCl (from 0 to 1M) in a 100 mM Tris-HCl, ph 8.0 buffer. The eluted fractions containing CPZ (as analyzed by SDS-PAGE) were pooled and loaded onto a Strep-Tag affinity column (IBA-life technologies) equilibrated with binding buffer (100 mM Tris-HCL, pH 8.0, 150 mM NaCl buffer), washed with 5 column volumes of binding buffer and eluted with the same buffer containing 2.5 mM d-desthiobiotin (IBA-life technologies). Eluted fractions were analyzed by SDS-PAGE, and the purest samples were pooled and loaded onto a size exclusion chromatography (HiLoad Superdex 75 26/60 column (GE healthcare)) previously equilibrated with a 50 mM Tris-HCL, pH 7.5, 150 mM NaCl buffer. Purified proteins were flash frozen at a concentration of approximately 0.3 mg/ml and then stored at -80 ºC until use. rhCPD was produced as described previously (Garcia-Pardo, Tanco, Dasgupta, et al. 2015)

5.2.2 CELL CULTURE

HEK293T cells (ATCC CRL-3216) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with GlutaMAX and with 10% (v/v) fetal bovine serum (Invitrogen, Inc.) at 37°C, 10% CO_2 and 95% humidity. HEK293F cells (ATCC CRL-3216) were grown in FreeStyle 293 expression medium (Invitrogen, Inc.) in flasks on a rotary shaker (120 rpm) at 37°C, 8% CO_2 and 70% humidity.

5.2.3 HEK293T BORTEZOMIB TREATMENT AND PEPTIDE EXTRACTION

Peptides form HEK 293T were obtained as described previously (Garcia-Pardo, Tanco, Dasgupta, et al. 2015), (see Chapter III). Briefly, HEK293T cells were grown to 70% confluence in 150 mm cell culture plates in a DMEM Medium supplemented with 10% fetal bovine serum and containing a mixture of penicillin and streptomycin antibiotics. After growing, HEK293T cell plates were treated with fresh media containing 0.5 µM bortezomib for 1 hour at 37 ºC. For peptide extraction, cells were washed three times with cold Dulbecco's phosphate-buffered saline (DPBS, Invitrogen), immediately scraped and centrifuged at 8000xg for 5 min. The cell pellet was resuspended in 1 ml of 80 ºC water and the mixture was incubated for 20 min in an 80 ºC water bath. Then, samples were cooled, transferred to a 2 ml low retention microfuge tubes and centrifuged at 13,000xg for 20 minutes. The soluble fractions containing HEK293T cell peptides were stored at –70 ºC for overnight. Samples were then centrifuged again and the supernatant of each tube was collected and concentrated in a vacuum centrifuge to a volume of 1.5 ml. Finally, samples were cooled, acidified with HCl, centrifuged at 13,000xg for 40 min at 4ºC, and the supernatants stored at -70ºC until labeling.

5.2.4 PREPARATION OF TRYTIC PEPTIDE LIBRARIES

Tryptic peptide libraries were generated by digesting five different proteins with trypsin, similarly as described before (Garcia-Pardo, Tanco, Dasgupta, et al. 2015) (see

Chapter III). To generate the peptide library, five different proteins (bovine serum albumin (BSA, Sigma-Aldrich), bovine thryroglobulin (Sigma-Aldrich), bovine α-lactalbumin (Sigma-Aldrich), human α-hemoblin (Sigma-Aldrich) and human β-hemoglobin (Sigma-Aldrich)) were completely digested with trypsin to obtain final peptide mix with a concentration of about 500 µM. The efficiency of protein digestions (and consequently the successful tryptic peptides generation) was confirmed by SDS-PAGE. After digestion reactions were stopped by adding 0.1 M HCl, and independent protein reactions were combined and centrifuged at 13,000xg for 45 minutes at 4 ºC. Finally, the supernatant containing tryptic peptides was filtered through a 10 kDa centrifugal filter device (Amicon), aliquoted and stored at -80ºC.

5.2.5 KINETIC EXPERIMENTS USING FLUORESCENT SUBSTRATES

Enzyme activity was typically assayed with the fluorescent substrate dansyl-Phe-Ala-Arg. To perform the experiments, reactions of 100 µl containing 0.2 mM of dansyl-Phe-Ala-Arg in 100 mM tris-acetate, pH 7.5, 100 mM NaCl buffer were incubated with a variable amount of enzyme for 60 min at 37 ºC. After incubation, reactions were stopped by adding 50 µl of 0.5 M HCl. Then, 1 ml of chloroform was added to each reaction and tubes were mixed gently and centrifuged for 2 min at 300 x g. After centrifugation, 0.5 ml of the chloroform phase were transferred to new tubes and completely dried for overnight at 25 ºC. Finally, dried samples containing mainly the product generated in the enzymatic reaction were resuspended with 200 µl of PBS containing 0.1 % of Triton X-100. The amount of product generated was determined by measuring the fluorescence of samples at 395 nm upon excitation at 350 nm using a 96-well plate spectrofluorometer. The enzymatic activity of the enzymes was also evaluated against other dansylated tripeptides (such as dansyl-Phe-Gly-Arg and dansyl-Phe-Pro-Arg). For this, reactions with different amounts of purified rhCPZ and rhCPZΔFz were incubated as described above for Dansyl-Phe-Ala-Arg. For kinetic analysis, purified enzymes were incubated with different substrate concentrations (typically 0, 66, 125, 250, 500, 1000, 1500, and 250 µM). The amount of enzyme used in reactions was that which hydrolyze a maximum of 20% of the substrate. Kinetic parameters were determined by fitting the obtained data for each enzyme to the

Michaelis-Menten equation (y= (Vmax x X) /(Km + X)), by using GraphPad Prism software (Motulsky HJ n.d.), were X is the substrate concentration, Vmax the maximum enzyme velocity and Km the Michaelis-Menten constant. The pH optimum of purified enzymes were determined with 0.2 mM Dansyl-Phe-Ala-Arg in 0.1 mM Tris-Acetate with 150 mM NaCl buffer at the indicated pH.

5.2.6 PEPTIDOMICS

Quantitative peptidomics experiments were performed similarly as described previously for other MCPs and, recently, for human CPD (Sebastian Tanco et al. 2010; Lyons & Fricker 2010), (Garcia-Pardo, Tanco, Dasgupta, et al. 2015) (see Chapter III), with the following modifications. For this, peptides extracted from HEK293T cells or from the tryptic peptide library were incubated for 16 h at 37 ºC in 100 mM borate, pH 7.5, 100 mM NaCl buffer with different amounts (0, 1, 10 and 100 nM) of purified rhCPZ. After incubation, reactions were quenched with 2.5 M glycine and peptides labeled using standard labeling procedures with 4-trimethylammoniumbutyrate (TMAB) isotopic tags (Morano et al. 2008). Following, samples were pooled and subjected to liquid chromatography and mass spectrometry as described (Sebastian Tanco et al. 2010; Lyons & Fricker 2010) (Garcia-Pardo, Tanco, Dasgupta, et al. 2015). Data were analyzed using Mascot software (Available from the Matrix Science Website). Manual intervention was required for the verification of the obtained data. Thus, identifications were rejected unless 80% or more of the mayor fragments from the MS/MS matched predicted b- or γ-series fragments (with a minimum of five matches). Additional criteria followed to consider the matches includes the coincidence with the parent mass within 50 ppm of the theoretical mass, an expected charge equal to basic residues plus the N-terminus and a correct number of isotopic tags incorporated considering the number of free amines present in the peptide. Peptidomics analyses were performed at least by duplicate.

5.2.7 SEQUENCE ALIGNMENT AND STRUCTURAL MODELING

The amino acid sequence of human CPZ and mouse Frizzled-8 were obtained from the UniProt database (accession codes K66K79-2 and Q61091, respectively). A structural alignment between the Frizzled-like domain of CPZ and the Frizzled domain of Fz was generated by the flexible structure alignment by chaining aligned fragment pairs with twist (FATCAT) algorithm (Ye & Godzik 2003) using the protein comparison tool of the RCSB Protein Data Bank. Three-dimensional structures of mouse Fz8-CRD and Wnt-8 (4F0A) were obtained from the Protein Data Bank. Structural models of the catalytic domain of CPZ and the CPZ Frizzled-like domain were constructed by using the automated I-TASSER on-line server. Models with the best C-score, based on the significance of threading template alignments and the convergence parameters, were selected. After I-TASSER models were built automatically, manual intervention was required to redefine secondary structure limits, based on the predictions of Jpred 4 (Drozdetskiy et al. 2015), expert knowledge, and experimental information. PyMOL (DeLano 2002) was used for generation of figures and visual inspection of models.

5.3 RESULTS

5.3.1 PROTEIN PRODUCTION AND PURIFICATION

Recently, a mammalian-based expression system has been developed to produce high amounts of CPZ and other MCPs with affinity for heparin (Garcia-Pardo, Tanco, Fernandez-Alvarez, et al. 2015) (see Chapter IV above). To large amounts of protein for substrate specificity determination, recombinant human carboxypeptidase Z was produced following this methodology. To facilitate the expression and the following purification, the CPZ was truncated at the 3' end to delete the highly basic C-terminal region. The scission of this region increases the amount of protein expressed without altering its activity, secretion or ECM affinity properties (data not shown). Accordingly, two forms of human CPZ lacking the C-terminal region with or without the Frizzled-like domain (named here as rhCPZ and rhCPZΔFz) were cloned into the pTriExTM-7 expression vector to encode a IgM secretion signal sequence and an N-terminal Strep-Tag II fusion protein (**Figure 1**).

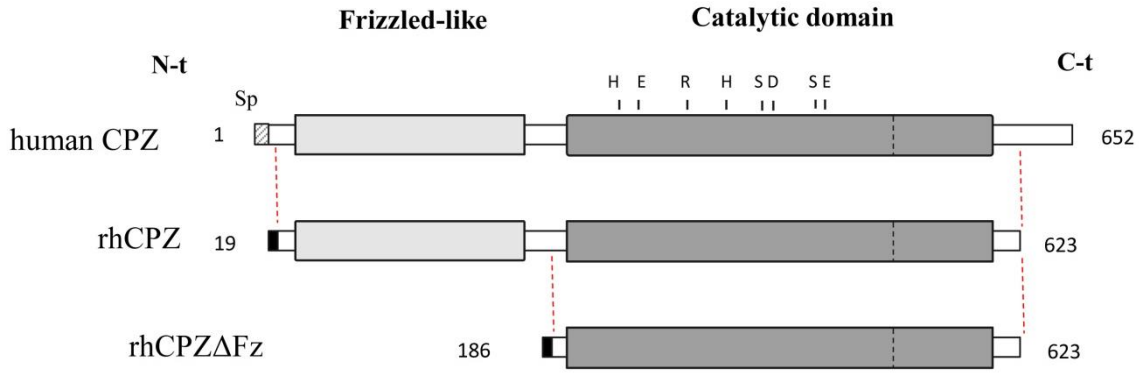


Figure 1. Linear representation of the full-length human CPZ and recombinant C-terminal truncated forms. The amino acids present in human CPZ correspond to key residues essential for the catalytic mechanism: His69, Glu72, Arg145, His198, Tyr248, and Glu270 (according to the bovine CPA numeration). Both C-terminal truncated forms of human CPZ were constructed lacking the c-terminal region. The recombinant form rhCPZΔFz also lacks the N-terminal part of 167 residues, corresponding to the Frizzled-like domain.

Both rhCPZ-pTrieX-7 and rhCPZΔFz-pTrieX-7 constructs were transfected into mammalian HEK293F cells for its transient expression. After 48 h post-transfection, recombinant proteins were detected on the extracellular media, showing majoritarian bands with molecular weights of ~70 kDa and ~55 kDa, respectively. Maximum levels of soluble purified rhCPD at the extracellular medium were detected after 10 days post-transfection, with protein levels up to 2-3 mg/L.

For protein purification of rhCPZ and rhCPZΔFz from the extracellular conditioned medium, the best results were obtained following a three-step chromatography protocol, as previously described (Garcia-Pardo, Tanco, Fernandez-Alvarez, et al. 2015) (**Figure 2**).

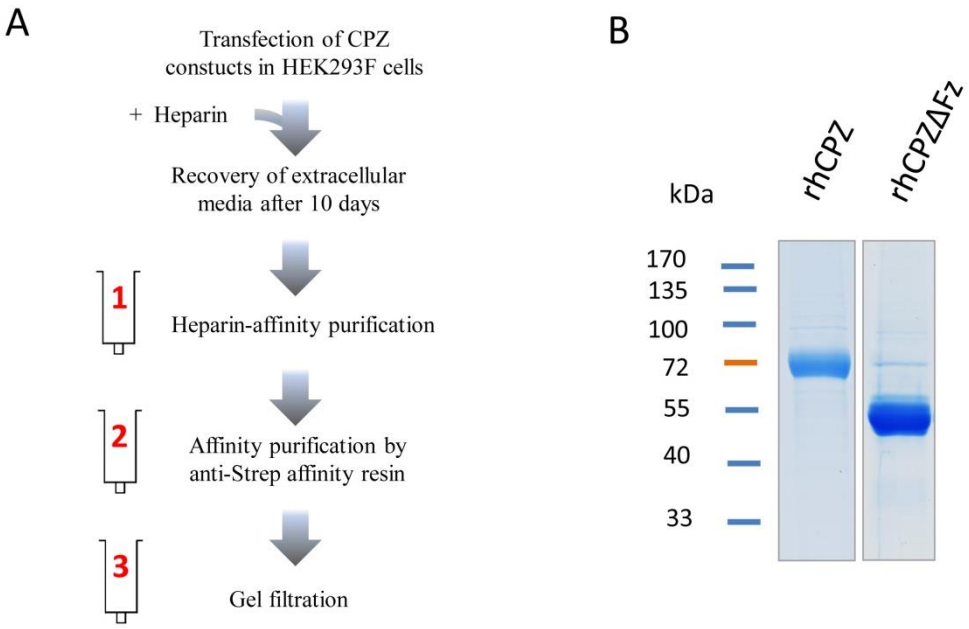


Figure 2. Expression and purification of CPZ. (A) Schematic diagram of the protocol followed for the expression and purification of rhCPZ and rhCPZΔFz. Protein expression was performed by high-level transient transfection in suspension-growing HEK (Human Embrionic Kidney) 283F cells, followed by the addition of heparin 24 hours post-transfection. For protein purification, the extracellular medium was collected after 10 days incubation, and recombinant proteins purified in three purification steps by (1) heparin-affinity chromatography, (2) affinity chromatography using anti-strep tag affinity resin and by (3) gel-filtration chromatography. (B) The purified proteins were visualized on SDS–PAGE by Coomassie staining.

The first chromatographic step was a heparin-affinity chromatography. The eluates containing both rhCPZ and rhCPZΔFz from the first purification step were pooled and purified over an anti-Strep tag affinity resin and, further fractionated on a size exclusion column (**Figure 2-A**). After purification, the obtained recombinant proteins could be easily visualized on SDS–PAGE by Coomassie blue staining, free of any major contaminating proteins (**Figure 2-B**). In gel filtration chromatography, both purified proteins elute mainly as single peaks, showed apparent masses, in agreement with its monomeric molecular weights in solution.

5.3.2 ENZYMATIC CHARACTERIZATION USING FLUORESCENT SUBSTRATES

The influence of pH on rhCPZ and rhCPZΔFd were evaluated using the fluorescent substrate dansyl-Phe-Ala-Arg. The pH optimum of the full-length protein was 6.0–8.5. At pH 5.5 the enzyme activity decreases to approximately 30% of the optimal activity observed at pH 7.5. The CPZ form without the Frizzled-like domain (rhCPZΔFd) showed a similar behavior at the different pH, reaching the pH optimum at pH 6.0–8.5 (Supplemental Figure 1).

The Enzymatic activities of the rhCPZ were tested against three different dansylated tripeptides. To compare substrates, different amounts of the purified enzyme were incubated with 200 µM of each fluorescent substrate in a 100 mM Tris-acetate, pH 7.5, 100 mM NaCl buffer, and the relative amount of product determined as described in the experimental section. Of the substrates examined, only dansyl-Phe-Ala-Arg was cleaved by CPZ under the experimental conditions assayed. When analysed in the same conditions, while 20% of dansyl-Phe-Ala-Arg was cleaved by CPZ, rhCPD was able to cleave three times more substrate (**Supplemental Figure 2-A**). No activity was detected towards dansyl-Phe-Gly-Arg or dansyl-Phe-Pro-Arg whe assayed to up to 50 nM of rhCPZ (**Supplemental Figure 2-B and C**). In contrast, 50 nM rhCPD cleaved approximately 20% of the dansyl-Phe-Gly-Arg, when tested in similar conditions. Further, we determined the kinetic parameters for rhCPZ and rhCPZΔFd using dansyl-Phe-Ala-Arg as substrate (**Table 1)**. The k_{cat} value obtained for rhCPZ with the Frizzled-like domain was 5.3 ± 0.6 s^{-1} fairly close to the k_{cat} of the rhCPZΔFd without the Frizzled-like domain which showed a value of 6.2 ± 0.8 s^{-1}.

Similarly, rhCPZ showed a K_M of 1905 ± 360 µM and a k_{cat}/K_{mM} of 0.0028 ± 0.0008 comparable to the enzyme without the Frizzled-like domain rhCPZΔFd, with K_m and k_{cat}/K_M values of 1667± 385 μM^{-1} s^{-1} and 0.0039 ± 0.0014 μM^{-1} s^{-1}, respectively.

Table 1. Kinetic constants for hydrolysis of dansyl-Phe-Ala-Arg by rhCPZ and rhCPZΔFz

Substrate	rhCPZ	rhCPZΔFz
Dansyl-Phe-Ala-Arg		
K_M (μM)	1905 ± 360	1667 ± 385
K_{cat} (s^{-1})	5.3 ± 0.6	6.2 ± 0.8
K_{cat}/K_M (μM^{-1} S^{-1})	0.0028 ± 0.00084	0.0039 ± 0.00140

5.3.3 CHARACTERIZATION OF THE SUBSTRATE SPECIFICITY OF HUMAN CPZ BY QUANTITATIVE PEPTIDOMICS

We applied a combination of quantitative peptidomics approaches to study the substrate specificity of human CPZ. In the first experiment, different amounts of purified rhCPZ were incubated with a peptide mixture extracted from bortezomib-treated HEK293T cells. This peptide mix contains hundreds of different peptides with a wide variety of residues in the C-terminal position. After incubation, the individual reactions were differentially labeled with isotopic TMAB tags and then combined and analyzed by LC-MS (see experimental scheme in **Figure 3-A**). After LC-MS, over 60 peptides were identified through tandem mass spectrometry (MS/MS) and/or close matches with the theoretical mass of the peptides previously identified. The criteria followed to consider the matches includes the coincidence with the monoisotopic mass of peptides identified in previous experiments with an error less than 50 ppm, a charge equal to the number of basic residues plus the N-terminus, and a correct number of isotopic tags incorporated. The analysis of the LC-MS profilesallowed us to identify some peptides which in the peak set exhibited roughly equal peak heights, revealing that these peptides were not substrates or products of rhCPZ, under the experimental conditions used (**Figure 3-B** and **Supplemental Table 1**).

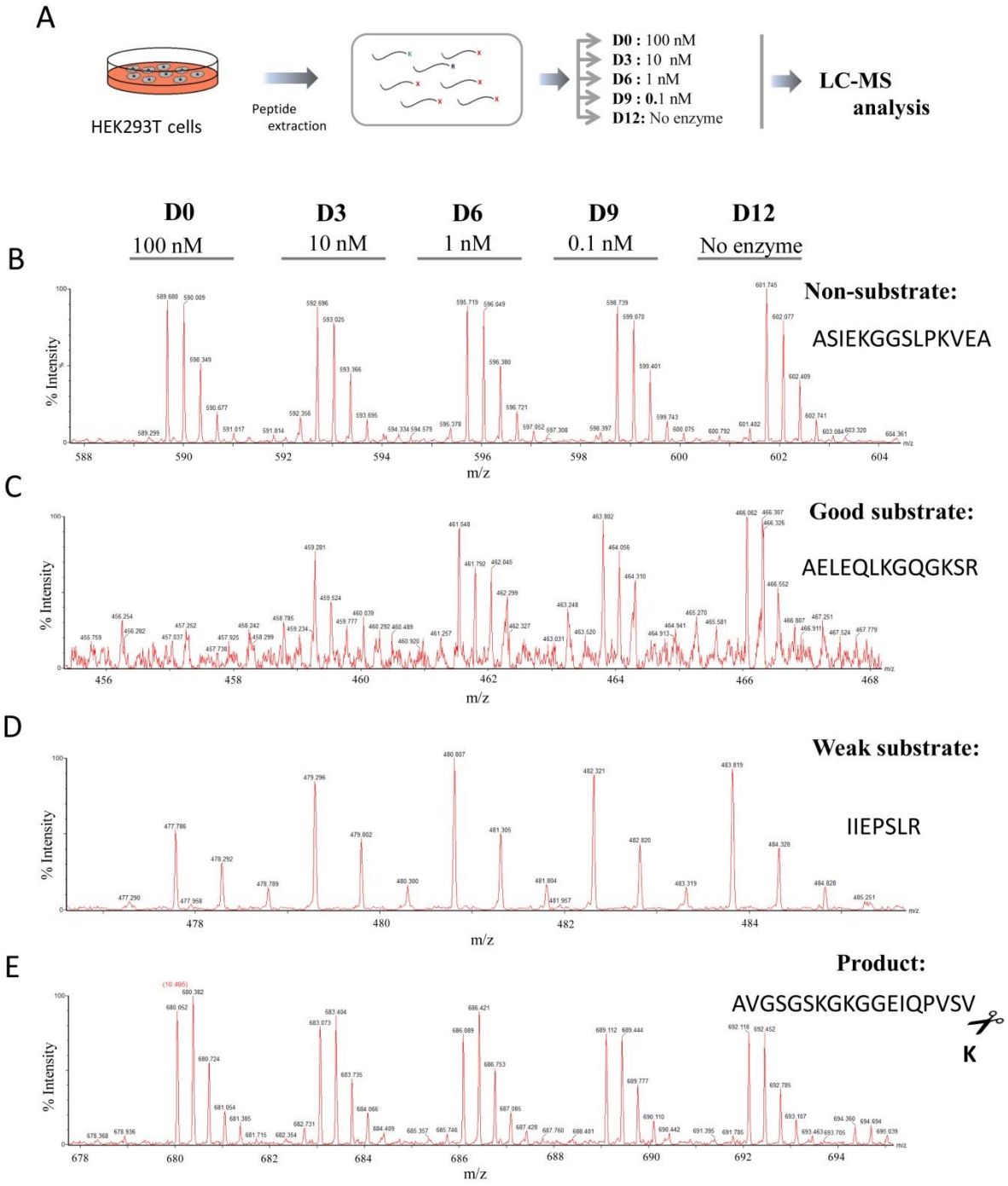


Figure 3. (A) Quantitative peptidomics scheme for the rhCPZ substrate characterization using HEK 293T peptides and (B-E) representative results. HEK 293T peptides were extracted from HEK 293 cell cultures treated for 1 h at 37 ºC with 0.5 µM bortezomib. The resultant peptide extract was aliquoted and digested with no enzyme or different rhCPZ concentrations of 0.1. 1. 10 and 100 nM at 37 ºC for 16 h. After incubation samples were labeled with one of five stable isotopic TMAB tags (D0= 100 nM; D3=10 nM; D6=1 nM; D9=0.1 nM; D12= No enzyme) Then, samples were pooled and analyzed by Liquid chromatography and Mass Spectrometry (LC-MS). Examples of representative data are shown for (B) non-substrates, (C) good substrates, (D) weak substrates and (E) products.

Some peptides were extensively cleaved, showing a complete or almost complete decrease in the peak intensity upon incubation with the highest enzyme concentration (*i.e.,* of 100 nM) and a partial decrease in the peak intensity with a lower enzyme concentration (*i.e.,* 10 nM); these are considered as good substrates of rhCPZ (**Figure 3-C** and **Table 2**). Some peptides were only partially cleaved, exhibiting little decrease in intensity with the highest concentration of enzyme assayed and no or slight decrease in the peak intensity with the concentration of 10 nM; these are considered as weak substrates for rhCPZ (**Figure 3-D** and **Table 2**). A reduced number of peptides showed an increase in the peak intensities related with increasing amounts of rhCPD; these are considered as products of rhCPZ action (**Figure 3-E** and **Table 3**).

Analysis of the C-terminal (P1') residue of rhCPZ substrates and non substrates, showed that rhCPZ only cleaves Lys and Arg residues (**Figure 4**).

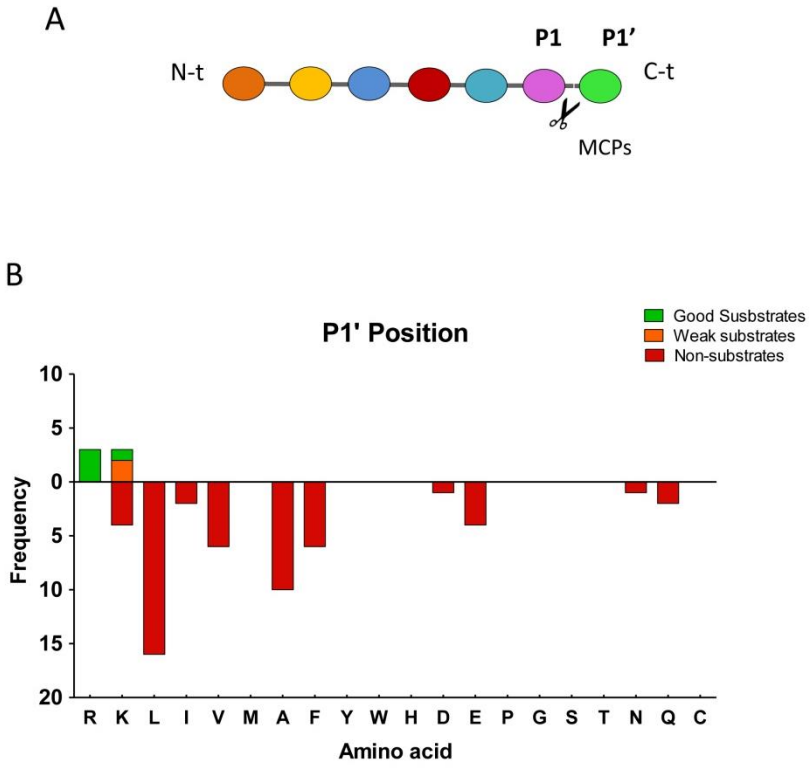


Figure 4. Analysis of the substrate preferences of rhCPZ using HEK 293T peptides. (A) schematic representation of the most important residues involved in a typical MCPs cleavage, according to model proposed by Schechter and Berger (Schechter & Berger 1967). (B) Substrate preference of rhCPZ at C-terminal (P1') position. The number of times each amino acid was present in P1' was determined for good substrates, weak substrates and non-substrates and represented.

This P1' residue is the most important residue to determine the substrate specificity of MCPs, although other residues in positions P1 or even farther from the cleavage site (P2, P3 and so on) can contribute to the substrate specificity, according to the proposed model of Schechter and Berger (**Figure 4-A)** (Schechter & Berger 1967). Thus, almost all the good and weak substrates identified showed Arg at C-terminal position, and only one peptide showed a Lys amino acid (**Figure 4-B)**. All of those peptides identified as substrates for this enzyme contained hydrophobic residues with small side chains or residues with polar uncharged side chains such as Ala, Ser, Lys or Thr amino acids in penultimate (P1) position (**Table 2**). Moreover, a widely variety of C-terminal residues (Leu, Ile, Val, Ala, Phe, Asp, Glu, Gln, Asn amino acids) were identified as non-substrates for rhCPZ (**Supplemental Table 1**).

Four peptides were identified as products of rhCPZ. These resulted from cleavage of Lys or Arg from tryptic peptide precursors, with Ala, Phe, Val or Leu at the P1 position of the cleavage site (**Table 3**).

Table 2. Good and weak substrates of rhCPZ identified using HEK 293T peptides

Type	Precursor	Peptide sequence	Z	T	Obs M	Theor M	ppm	Ratio rhCPZ / No enzyme			
								100 nM	10 nM	1 nM	0.1 nM
Good	Histidine triad nucleotide-binding protein 1	Ac-ADEIAKAQVAR	2	1	1212.66	1212.65	14	0.02	0.94	1.08	1.00
Good	Eukaryotic translation initiation factor 5A	SAMoxTEEAAVAIKAMAK	3	3	1636.81	1636.82	-5	0.07	0.15	0.95	0.95
Good	Vimentin	AELEQLKGQGKSR	4	3	1442.78	1442.78	-3	0.18	1.00	1.03	0.95
Good	Hematological and neurological expressed 1 protein	Ac-TTTTTFKGVDPNSRNSSR	3	1	2010.00	2009.98	13	0.31	0.89	1.24	1.07
Weak	Eukaryotic translation initiation factor 5A	NMDVPNIKR	3	2	1085.56	1085.57	-6	0.45	n.d.	n.d.	0.82
Weak	Ubiquitin-60S ribosomal protein L40	IIEPSLR	2	1	826.49	826.49	0	0.60	1.00	1.13	1.04

Good substrates; peptides affected with a decrease ≥60% by the highest concentration of enzyme; weak substrates; peptides affected with a decrease ≥20% and <60% by the highest concentration of enzyme; Z, charge; T, number of isotopic tags incorporated into each peptide; Obs M, observed monoisotopic mass; Theor M, theoretical monoisotopic mass; ppm, difference between Obs M and Theor M (in parts per million); Ratio rhCPZ/no enzyme, the ratio in peak intensity between the sample incubated with enzyme and the sample incubated without enzyme.

Table 3. Products of rhCPZ identified using HEK 293T peptides

Precursor	Sequence	Cleaved aa	Z	T	Obs M	Theor M	ppm	Ratio rhCPZ / No enzyme			
								100 nM	10 nM	1 nM	0.1 nM
Heat shock 10kDa protein 1 (chaperonin 10)	AVGSGSKGKGGEIQPVSV	K	3	3	1655.89	1655.89	2	1.30	1.16	1.28	0.93
Peptidylprolyl isomerase A	ADKVPKTAENF	R	3	3	1218.62	1218.62	-3	1.36	1.09	1.13	1.09
Nucleophosmin	EKTPKTPKGPSSVEDIKA	K	5	5	1911.00	1911.03	-18	1.46	1.00	1.23	1.15
FK506 Binding Protein	VFDVELL	K	1	1	833.45	833.45	2	1.50	1.28	ND	1.25

Products; peptides with an increase >120% with one or more concentrations of enzyme. Cleaved aa. The amino acid cleaved by rhCPZ to generate the observed peptide. See Table 2 for the rest of abbreviation definitions.

5.3.4 CHARACTERIZATION OF THE SUBSTRATE SPECIFICITY OF HUMAN CPZ BY QUANTITATIVE PEPTIDOMICS USING A TRYPTIC PEPTIDE LIBRARY

The HEK 293T peptide mix used in the first peptidomics experiment as substrate gave important information about the substrate preferences of CPZ for C-terminal basic residues versus other non-basic amino acids. Nonetheless, this peptide mix contains a reduced number (less than 10%) of peptides with Arg or Lys amino acids in C-terminal position, which are potential substrates for CPZ. For this reason, this approach offer limited information about its substrate preferences, especially for other positions different than P1'. To address an extensive characterization of the substrate specificity of human CPZ, we performed a second peptidomics experiment in which different amounts of purified rhCPz were incubated with a peptide mixture obtained by digestion of selected proteins with trypsin (see Experimental section above). This tryptic peptide library, used before to characterize the substrate specificities of the first and second domains of human CPD, contains mainly peptides with basic residues (Arg or Lys) at C-terminal position (P1`), having a great variety of amino acids at P1 position (Garcia-Pardo, Tanco, Dasgupta, et al. 2015). After incubation with the enzyme, individual reactions were differentially labeled with isotopic TMAB tags, combined and analyzed by liquid chromatography/mass spectrometry (LC-MS) (see experimental scheme in **Figure 5-A**). After LC-MS, more than 50 peptides were identified through tandem mass spectrometry (MS/MS) and/or close matches with the theoretical mass of the peptides generated with trypsin. The criteria followed to consider the matches were the same described above for the first peptidomics experiment (see above). Some of these peptides identified exhibited a peak set with roughly equal peak heights, revealing that these peptides were not substrates or products of rhCPZ, under the experimental conditions used (**Figure 5-B** and **Supplemental Table 2**). Further, some peptides were extensively cleaved, showing a complete or almost complete decrease in the peak intensity upon incubation with the highest concentration of enzyme (*i.e.,* 100 nM) and a partial decrease in the peak intensity with a lower concentration (*i.e.,* 10 nM); these are considered as good substrates of rhCPZ (**Figure 5-B** and **Table 4**).

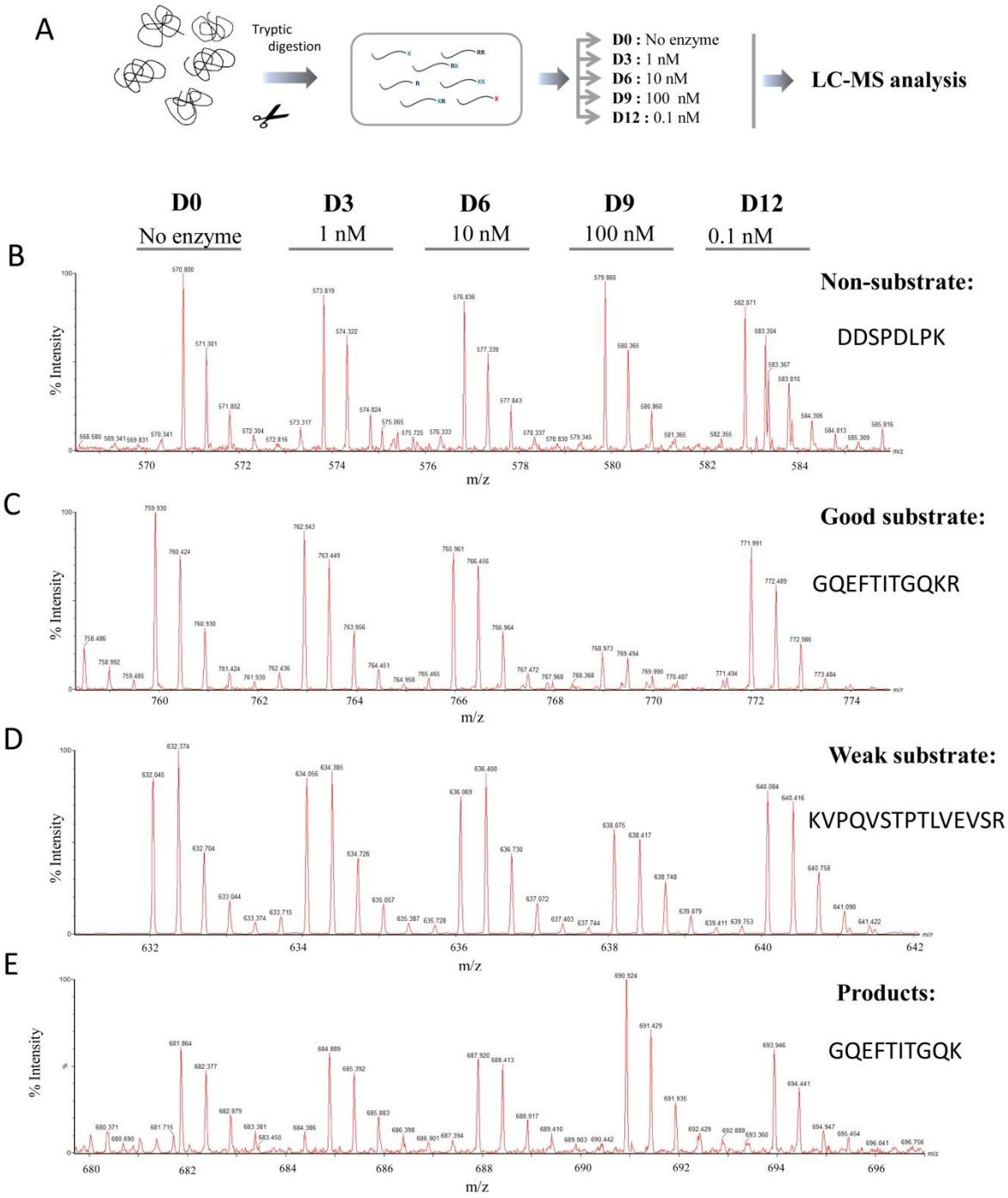


Figure 5. (A) Quantitative peptidomics scheme for the rhCPZ substrate characterization using the tryptic peptide library and (B-E) representative results. Tryptic peptides were obtained from digestion of five selected proteins (BSA, bovine thryroglobulin, bovine α-lactalbumin and human α and β–hemoblin) with trypsin. The resultant peptide library was aliquoted and digested with no enzyme or different rhCPZ concentrations of 0.1. 1. 10 and 100 nM for 16 h at 37ºC. After incubation samples were labeled with one of five stable isotopic TMAB tags (D0= No wnzyme; D3=1 nM; D6=10 nM; D9=100 nM; D12= 0.1 nM) Then, samples were pooled and analyzed by Liquid chromatography and Mass Spectrometry (LC-MS). Examples of representative data are shown for (B) non-substrates, (C) good substrates, (D) weak substrates and (E) products.

In addition, some peptides were only partially cleaved, exhibiting little decrease in intensity with the highest concentration of enzyme assayed, and no or slight decrease in the peak intensity with the concentration of 10 nM; these are considered as weak substrates of rhCPZ (**Figure 5-D** and **Table 4**). On the other hand, a reduced number of peptides showed an increase in peak intensities that correlates with the amount of rhCPZ; these are considered as products of rhCPZ cleavage (**Figure 5-D** and **Table 4**).

Analysis of the C-terminal (P1') residue of rhCPZ substrates versus non-substrates, showed again preference for Lys and Arg amino acids (see **Figure 6-A**).

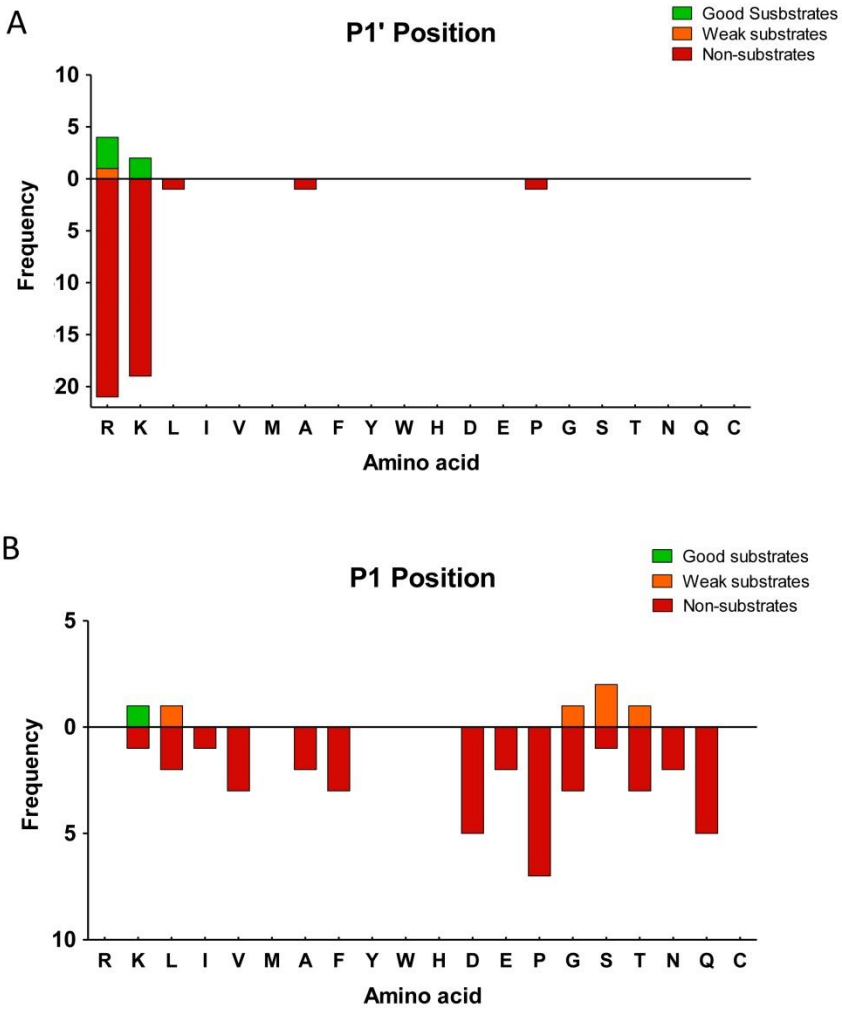


Figure 6. Analysis of the substrate preferences of rhCPZ determined using the tryptic peptide library. (A) Substrate preferences of rhCPZ at C-terminal (P1') and (B) P1 positions. The number of times each amino acid was present in P1 or P1' was determined for good substrates, weak substrates and non-substrates and represented. For P1 analysis, only substrates with permissive P1' residues according to (A) were considered.

Nevertheless, a large number of peptides with C-terminal Arg and Lys were identified as non-substrates, probably as a result of low activity of this enzyme, as well as the influence of positions other than P1' (*i.e.,* P1 or even farther from the C-terminal residue) (**Supplemental Table 2**). Therefore, we analyzed the influence of the penultimate (P1) residue on rhCPZ substrates and non-substrates (see **Figure 6-B**). We observed that the only one good substrate identified for rhCPZ contain a Lys amino acid in this position, and weak substrates contained C-terminal Ser, Leu, Gly or Thr amino acids (**Table 4**). No peptides with Pro, Asp, Gln, Phe, Val, Ala, Asn, Glu or Ile amino acids at P1 position were identified as a rhCPZ substrates (see **Table 4**).

In addition, we identified two peptides as products for rhCPZ. Both peptides are obtained from the proteolytic cleavage of tryptic peptide precursors with C-terminal Arg residues, with Phe or Lys residues in the P1 position (**Figure 5-E** and **Table 5**).

Table 4. Good and weak substrates of rhCPZ identified using the tryptic peptide library

Type	Protein precursor	Peptide sequence	Z	T	Obs M	Theor M	ppm	Ratio rhCPZ / No enzyme			
								100 nM	10 nM	1 nM	0.1 nM
Good	Thyroglobulin	GQEFTITGQKR	2	2	1263.66	1263.6	-2	0.38	1.13	0.97	1.16
Weak	Thyroglobulin	ALEQATR	2	1	787.42	787.42	3	0.57	1.04	1.00	1.09
Weak	Thyroglobulin	AVKQFEESQGR	3	2	1277.64	1277.64	0	0.68	0.86	0.95	0.89
Weak	Bovine serum albumin	KVPQVSTPTLVEVSR	3	2	1638.93	1638.93	-2	0.70	0.90	0.95	0.85
Weak	Bovine serum albumin	KQTALVELLK	3	3	1141.69	1141.71	-22	0.73	0.80	0.89	0.87
Weak	Thyroglobulin	LPESK	2	2	572.31	572.32	-24	0.77	0.83	0.83	0.90

Good substrates; peptides affected with a decrease ≥60% by the highest concentration of enzyme; weak substrates; peptides affected with a decrease ≥20% and <60% by the highest concentration of enzyme; Z, charge; T, number of isotopic tags incorporated into each peptide; Obs M, observed monoisotopic mass; Theor M, theoretical monoisotopic mass; ppm, difference between Obs M and Theor M (in parts per million); Ratio rhCPDZno enzyme, the ratio in peak intensity between the sample incubated with enzyme and the sample incubated without enzyme.

Table 5. Products of rhCPZ identified using the tryptic peptide library

Protein precursor	Peptide sequence	Cleaved aa	Z	T	Obs M	Theor M	ppm	Ratio rhCPZ / No enzyme			
								100 nM	10 nM	1 nM	0.1 nM
Thyroglobulin	GQEFTITGQK	R	2	2	1107.56	1107.56	-4	1.52	0.78	1.04	1.00
Thyroglobulin	LF	R	1	1	278.16	278.15	14	1.56	0.88	0.88	1.00

Products; peptides with an increase >120% with one or more concentrations of enzyme; Cleaved aa, the amino acid cleaved by rhCPZ to generate the observed peptide; ND, not detectable. See Table 2 for the rest of abbreviation definitions.

5.3.5 STRUCTURAL MODELING OF THE CATALYTIC DOMAIN OF HUMAN CPZ

To gain more insights into the structure and function of human CPZ, we modeled the catalytic domain of this enzyme based on the previous solved crystal structure of human CPN (PDB 2NSM), human CPM (PDB 1UWY) and the Drosophila melanogaster and duck CPD domains I (PDB 3MN8) and II (PDB 1H8L), respectively, by using I-TASSER (Zhang 2008b). The resultant model shown in **Figure 7-A**, shows both sequential and topological similitude with the rest of the carboxypetidase domains of M14B MCPs with solved structures. Among them, CPN has the higher sequential (51%) identity, as well as topological similitude (with a root mean square desviation or RMSD of 0.48 Å) to human CPZ. In the CPZ model, two independent domains are clearly visible, the CP domain and the typical C-terminal transthyretin-like domain found in all members of the M14B subfamily of MCPs, which shares topological similarity and connectivity with transthyretin (**Figure 7-A**).

The predicted pI of full-length human CPZ (including both the carboxypeptidase and the transthyretin-like domain) is ~8.3. This pI value rises to ~9.3 when the enzyme lacks its N-terminal transthyretin-like domain and is the responsible of its ECM, as well as heparin affinity properties. The representation of the electrostatic potential into the CPZ surface shows that majority of basic residues are located on the face of the molecule opposite that of the active site (**Figure 7-B**), suggesting that this may orient CPZ with respect to the ECM and facilitate access of substrates to the catalytic cleft.

A detailed analysis of the catalytic site of human CPZ (**Figure 7-C** and **7-D**) shows that the three protein ligands of the catalytic zinc ion His69, Glu72 and His196 (according to bCPA numeration) are conserved. Similarly, those residues directly involved in catalysis Arg145 and Glu270, are also conserved. Nonetheless, some of residues involved in the substrate binding and specificity (Leu203, Gly 207 and Ile255 in bCPA) are substituted by a Ser, Asp and Ser amino acids, respectively (**Figure 7-C** and **7-D**). These residue substitutions are also found in other MCPs like the first domain of human CPD, which contains a Ser and Asp residues homologous to Leu203 and Gly207 of bCPA, whereas the second domain of CPD contains Asn in an equivalent position to

Leu203. A difference, CPZ has a Ser residue in a position equivalent to I255 in bCPA, while both CPD domains have an Asn amino acid (**Figure 7-C** and **7-D**).

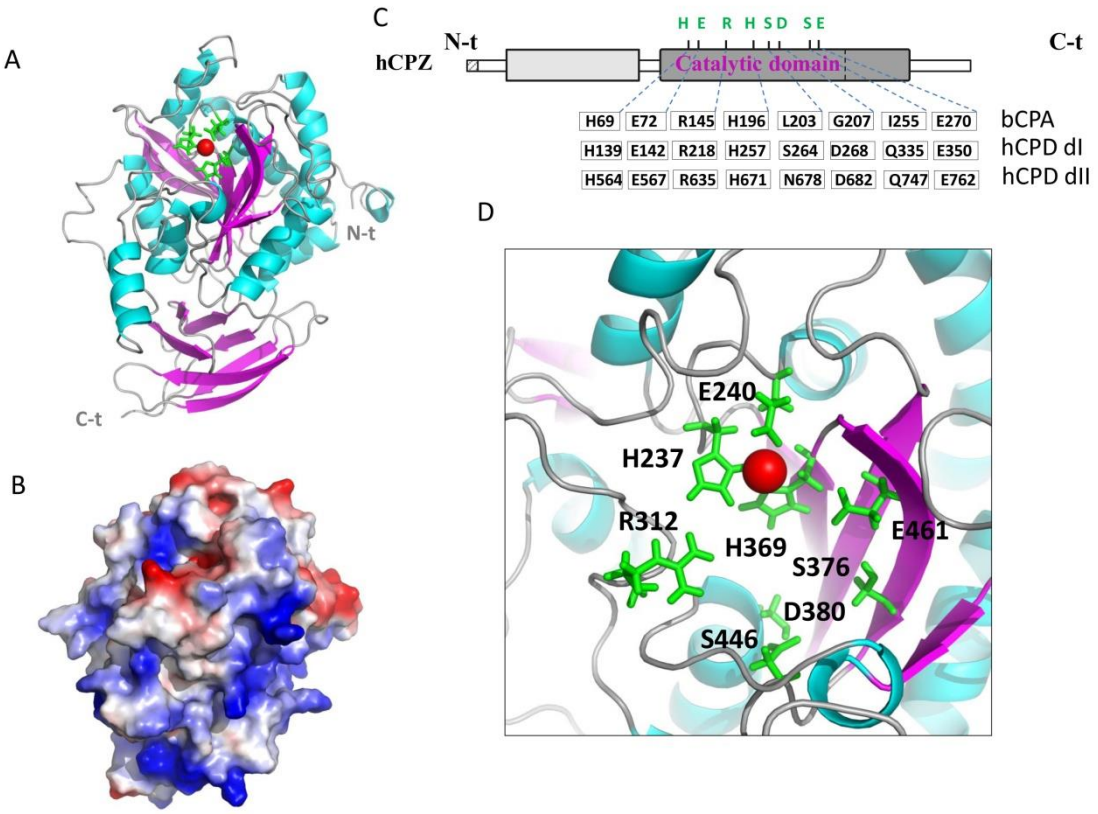


Figure 7. Structural modeling of the catalytic domain of human CPZ. A) Ribbon representation of human CPZ structure showing the catalytic moiety at the top, and the transthyrety-like domain at the bottom. The side chains of the three residues involved in the Zn^{2+} binding are indicated in green. B) Electrostatic surface potential distribution of the catalytic domain of human CPZ (in the same orientation than A). Blue indicates positive and red indicates negative charge potential. (C) Linear representation of the full-length human CPZ, showing the location of relevant amino acids involved in the catalytic mechanism and substrate binding (Arg312, Ser376, Asp380, Ser446 and Glu461), as well as in zinc binding (His267, Glu240, His369). Residues found in equivalent positions in bovine CPA (bCPA, as reference), in domain I of human CPD (hCPDdI) and in domain II of human CPD (hCPDdII) are indicated. (D) Magnification of the active site of human CPZ, showing the location of these residues important for the catalytic mechanism, and substrate specificity determination. The Zn^{2+} metal ion is shown in all representations as a red sphere.

5.3.6 STRUCTURAL MODELING OF THE CPA FRIZZLED-LIKE DOMAIN: INSIGHTS INTO THEIR STRUCTURE AND WNT RECOGNITION

To date, a reduced number of crystal structures of Frizzled domains have been solved (Dann et al. 2001; Janda et al. 2012). Because the three-dimensional structure of the transthyretin-like domain of human CPZ is still known, here we modeled it based on previous solved structures from the cysteine rich domains (or Frizzled-like domains) of mouse Fz8 (PDB 4F0A) and sFRP3 (PDB 1IJX) by using I-TASSER (Zhang 2008b). The derived three-dimensional model is shown in **Figure 8-A**.

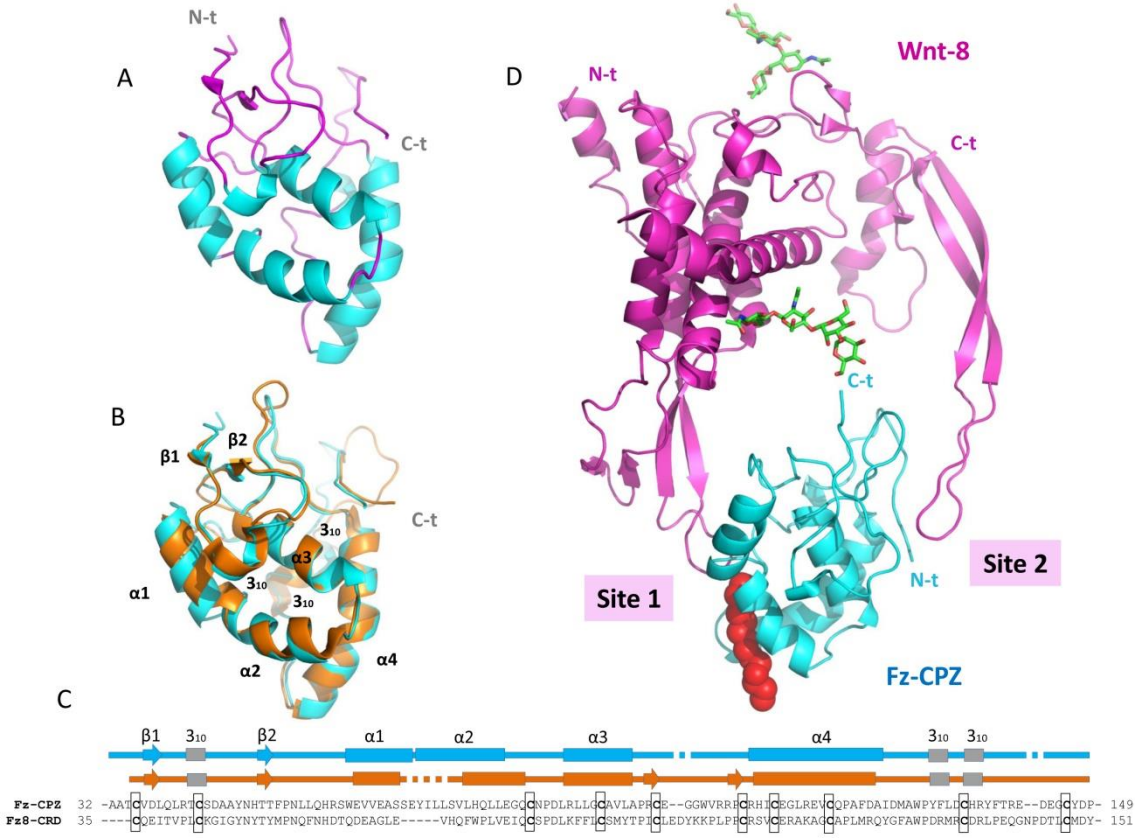


Figure 8. Structural modeling of the Frizzled-like domain of human CPZ and of their interaction with Wnt-8. A) Ribbon representation of the highly conserved Frizzled-like domain of human CPZ. (B) Structural comparison of the Frizzled-like domain of human CPZ (Fz-CPZ) and the cysteine rich domain of mouse Fz8 (Fz8-CRD). (C) Structure-based sequence alignment of Fz-CPZ and Fz8-CRD. (D) Ribbon representation of Wnt-8 in complex with the Frizzled-like domain of CPZ. The extended palmitoleic acid (PAM) group is represented with red spheres, and the most important sites of interaction between Fz-CPZ and Fz8-CRD are indicated (Site 1 and Site 2). For all proteins, the N-terminus and C-terminus are indicated (as N-t and C-t, respectively).

The Frizzled-like domains of mouse Fz8 and mouse sFRP3 share topological similarity and connectivity with the Frizzled-like domain of human CPZ. The RMSD calculation between the cysteine rich domain of mouse Fz8 and our modeled structure gave a value of 1.66, despite they share only a 28% of sequential identity (**Figure 8-B** and **8-C**). On its overall structure, the CPZ Frizzled-like domain is composed mainly by four α-helices stabilized through all ten conserved cysteines forming disulphide bonds. In addition to helical regions, two short β-strands at its N-terminus form a minimal β-sheet with β2 passing through a knot created by disulphide bonds (**Figure 8-A**). No major differences were observed between the three-dimensional structures of the cysteine rich domain of mouse Fz8 and the Frizzled-like domain of human CPZ. These slight differences arise mainly in the length and limits of the first and second α-helix (**Figure 8-B** and **8-C**).

Due to the structural similitude between the Frizzled-like domain of human CPZ and the extracellular domain of mouse Fz8, we used the recently solved structure of the complex between mouse Fz8-CRD and Xenopus Wnt-8 (XWnt-8) to model the interaction of our modeled structure with Wnt-8 (**Figure 8-D**). To gain insights into Wnt recognition by the CPZ Frizzled-like domain, we analyzed the most important structural elements of this domain for Wnt binding. In the resultant complex, XWnt-8 appears to grasp the Frizzled-like domain of CPZ at two opposing sites using extended lipid modified thumb and index fingers projecting from a central “palm” domain. The two contact sites are well defined; the “site 1” and “site 2 (see **Figure 8-**D). A detailed analysis of site 1 shows that the surface clef that accommodates the 16-C palmitoleic acid extended from the XWnt-8 Ser197, is conserved in the Frizzled-like domain of human CPZ. This surface cleft is shaped by the made up of helix 1 and 2, helix 4 and the fist loop immediately after helix 4. This surface cleft is lined by hydrophobic amino acids, similarly as described for Fz8-CRD and other Frizzled-like domains (Janda et al. 2012). The site 2 is located at the opposite side of site 1 in the CPZ Frizzled-like molecule and is also conserved. This second interaction site is formed by depression between inter-helical loops on the Frizzled-like domain and is the responsible for the accommodation of the XWnt-8 finger loop.

The C-terminal tail of the Frizzled-like domain is connected with the N-terminal tail of the CPZ CP domain through a ~10 amino acids segment. In the Fz-CPZ-XWnt-8 complex, the C-terminal tail of the Frizzled-domain is extended to the side of the main plain of the structure, and oriented to the C-terminal tail of XWnt-8. Although the structural arrangement between both domains (Frizzled-like domain and the catalytic domain of CPZ) is known, this finding allows the possibility that the catalytic domain remain close to the C-terminal residue of Wnt proteins, facilitating the cleavage of its C-terminal basic amino acids, following Wnt recognition (**Figure 8-D**).

5.4 DISCUSSION

The major finding of the present study was the determination of the cleavage specificity of human Carboxypeptidase Z, which has not been previously examined. It was previously found that CPZ was a secreted and active enzyme with a substrate preference for C-terminal Arg residues, based exclusively on a limited number of synthetic substrates (Novikova & Fricker 1999a). In this study, we used synthetic fluorescent substrates and quantitative peptidomics approaches to characterize the substrate specificity of this enzyme using a broad spectrum of peptides. The synthetic substrate dansyl-Phe-Ala-Arg, enabled us to determine that human CPZ has maximum activity at neutral pH (at pH 7.5). This observation is in agreement with previous studies on human CPZ, which tested the activity of this enzyme against a couple of dansylted tripeptides differing one from another only in the P3 position (Novikova & Fricker 1999a). One major limitation of this previous work was the obtainment of enough protein to calculate all the kinetic parameters. We solved this limitation through the use of a recently described production methodology, which allowed us to obtain mg of pure and active protein (Garcia-Pardo, Tanco, Fernandez-Alvarez, et al. 2015). Human CPZ has a k_{cat}/k_m value of 0.0028 ± 0.0008 μM^{-1} S^{-1} and a k_m of 1905 ± 360 µM. This result fits with previous results, which showed a k_m value for human CPZ of ~ 2 µM. When the k_{cat}/k_m values are compared with other enzymes of the same subfamily previously described, human CPD appear to be ~30 fold more efficient than human CPZ in the cleavage of the same synthetic substrate under a similar experimental conditions (Garcia-Pardo, Tanco, Dasgupta, et al. 2015). A detailed comparison between the k_{cat}/k_m values of both CPD domains and CPZ, show that while the first domain of human CPD is ~10 fold more active, the second domain only has ~3 fold more activity than CPZ. In a similar manner, CPE also appear to be ~3600 fold more efficient than human CPZ in the cleavage of this substrate (Fricker & Snyder 1982) . Synthetic substrates are useful to study precise substrate/product relationships and kinetic parameters, but are limited by the expense and the time to analyse each substrate. In the present study, we tested dozens of peptides using a combination of quantitative peptidomic approaches, in order to perform an in deep characterization of the substrate specificity of CPZ. Results showed a clear specificity for both Arg or Lys

C-terminal residues, with a slight preference to cleave Arg versus Lys amino acids. Previous studies have demonstrated that CPZ is synthetized and immediately secreted to the extracellular medium, where is attached to the ECM (Novikova et al. 2000). Moreover, CPZ is expressed dynamically during mouse development (Novikova et al. 2001). These findings, taken together with the pH optimum and the substrate preference of this enzyme suggest that carboxypeptidase Z might play a mayor function in the extracellular space, primarily by cleaving peptides and/or proteins with C-terminal Arg/Lys residues.

Carboxypeptidase Z is a striking and unique enzyme, since besides the catalytic domain, contains an additional N-terminal cysteine rich domain of ~120 amino acids with a ~30 % sequential homology to the Frizzled protein family of Wnt receptors (Xiaonan et al. 1998). Only a reduced number of protein families contain Frizzled motifs (Diekmann 1998), and the majority of these proteins are presumed to bind Wnt proteins in a conserved fashion. This is the case of the Frizzled family of seven-pass transmembrane receptors that following Wnt binding, activate intracellular signaling events. Another group of proteins with cysteine rich domains are the secreted Frizzled Related Proteins (sFRPs). This family of secreted molecules act typically as Wnt inhibitors. Nonetheless, recent observations have offered a new perspective on their functions and mechanisms of action (Bovolenta et al. 2008). From the study of the structural model of the Frizzled-like domain, we found that the Frizzled-like domain of human CPZ has structural similarity and connectivity with the cysteine rich domain of Frizzled receptors (taking as example the mouse Fz8-CRD), and conserves the most important elements for Wnt recognition. Recently, it is emerging the idea that some carboxypeptidases can be involved in multiple functions independent from its catalytic activities. For example, CPE was proposed to be a positive/negative modulator of the Wnt signaling pathway, through its binding to the Wnt-3A-Frizzled receptor complex (Skalka et al. 2013). Accordingly, the Frizzled-like domain of CPZ might have independent functions, without the requirement of its catalytic activity or alternatively, may have related functions with the catalytic domain.

Another important goal of the present study is the production and enzymatic characterization of human CPZ without its N-terminal Frizzled-like domain. We found

that the lack of the Frizzled-like domain in human CPZ does not affect substantively its enzymatic activity against dansyl-Phe-Ala-Arg, since both proteins showed comparable kinetic parameters. This observation is also consistent with the conservation of all the residues in the active site important for the catalytic mechanism of human CPZ. For example, this enzyme conserves the three protein ligands of the catalytic zinc ion (His69, E72 and His196 using bovine CPA1 numbering), as well as the critical Glu (Glu270 in bovine CPA1) and Arg (Arg145 in bovine CPA1) residues, required for catalytic activity and for the anchoring of the C-terminal carboxyl group of the substrate, respectively. Moreover, residues Leu203 and Gly207 found in bovine CPA and responsible of the substrate specificity determination are substituted in CPZ by Ser and Asp amino acids. Similar residue substitutions were found in other MCPs of the same subfamily. This is the case of the first domain of human CPD, which has also preference for substrates with C-terminal Arg residues and optimum activity at neutral pH (Garcia-Pardo, Tanco, Dasgupta, et al. 2015). This finding is consistent with previous reports that related the presence of these two residues in the specificity pocket, with a substrate preference for C-terminal Arg versus Lys (Reverter et al. 2004; Garcia-Pardo, Tanco, Dasgupta, et al. 2015). Despite these similitudes, one residue within the substrate binding pocket of CPZ equivalent to I255 in bovine CPA1 is replaced in CPZ by a Ser amino acid. A difference with CPZ, in other members of the same subfamily with a similar substrate specificity for basic residues (CPE, CPD, CPN and CPM), this residue is substituted by Gln. One possibility is that the presence of this residue in the catalytic site of CPZ might explain the low activity of CPZ against the majority of substrates, in comparison with the rest of MCPs. The presence of this Ser in CPZ might provide a favourable environment in the catalytic pocket for the binding, and subsequent cleavage of other non-conventional substrates.

Previous studies using a low number of small synthetic substrates have found that CPZ is only able to cleave substrates with C-terminal Arg residues and among them, only was efficient at cleaving those containing Ala in P1 position (Novikova & Fricker 1999a). In the present study, in which we used dozens of peptides, we found that this enzyme prefer substrates with small hydrophobic side chains or with polar uncharged side chains like Ser, Thr, Leu or Gly in the P1 position. One residue with Arg in P1

position was also cleaved efficiently by CPZ. The presence of the Frizzled-like domain in CPZ together with its dynamic expression patter during development and its localization in the ECM can indicate a special affinity for Wnt molecules. It is intriguing that the majority of human Wnt proteins are predicted to contain C-terminal basic residues, immediately after the last conserved Cys amino acid (**Figure 9**). However, it is not known whether the major forms of these proteins that exist *in vivo* contain the C-terminal basic residue, or whether it has been removed by proteases. In previous studies, it was hypothesised that CPZ may cleave these C-terminal residues (Reznik & Fricker 2001; Moeller et al. 2003; Wang et al. 2009). As discussed above, in the present study we found that CPZ can remove C-terminal basic residues from a wide variety of peptides, especially those containing amino acids with uncharged side chains at the P1 position. This finding fits well with the presence of a Cys residue (with similar chemical properties to Ser/Thr) found immediately before to the C-terminal amino in the majority of human Wnt proteins, implying that this molecules are potential substrates for CPZ. If processing does occur at this site, it is not known whether this affects the biological activity. It is possible that removal of the C-terminal residue activates or inactivates Wnt proteins, render them susceptible to further degradation or altering the targeting of Wnts within the extracellular matrix. For example, in the case of human Epidermal Growth Factor (EGF) , the cleavage of three or more C-terminal residues lead to the loss of its biological activity, since these C-terminal residues are important for receptor biding (Calnan et al. 2000; Panosa et al. 2013). We have recently found that the combination of several digestive MCPs suffices to cleave its C-terminal tail and remove the biological activity of human EGF *in vitro* (Unpublished data). However, the C-terminal tail of Wnt proteins does not contact Fzd8 in the XWnt-8-Fzd8 structure. Nonetheless, it is interesting to speculate that the C–terminal residue/s in Wnt proteins might serve for recognition and/or stabilization of the interactions with co-receptors (e.g. LRP binding), or serve for other structural/functional purposes. Thus, it is possible that removal of the C-terminal residue activates or inactivates the Wnt, renders it susceptible to further degradation or alters the targeting of the Wnt within the extracellular matrix. In support of this, the polarized nature of the model for CPZ suggests that this enzyme may orient itself with respect to the negatively charged glycoproteins of the ECM, remaining its active site

accessible for potential substrates. It is interesting that a similar charge polarity has also been found for other carboxypeptidases with similar heparin-binding properties, such as mast cell carboxypeptidase or human carboxypeptidase A6 (Lyons et al. 2008; Wernersson & Pejler 2014).

Taken together, our results establish that CPZ is able to cleave peptides and protein substrates with C-terminal basic residues, having preference for those substrates containing basic or polar uncharged side chains in the P1 position. These findings, together with the structural analysis of the Frizzled-like domain and the catalytic domain of CPZ revealed that this enzyme is likely to play an important role in the activation and/or inactivation of Wnt signaling molecules. Nonetheless, further functional studies are needed to test these hypotheses, and unravel the exact biological functions of this intriguing enzyme in Wnt signaling.

5.5 SUPPLEMENTAL INFORMATION

Supplemental table 1. Non-substrates of rhCPZ identified using HEK 293T peptides

Precursor	Sequence	Z	T	Obs M	Theor M	ppm	Ratio rhCPZ / No enzyme			
							100 nM	10 nM	1 nM	0.1 nM
Eukaryotic translation initiation factor 4H	ATPLNQVANPNSAIFGGARPREEVVQKEQE	4	2	3248.69	3248.65	12	0.83	1.07	1.07	1.07
Acidic nuclear phosphoprotein pp32	STIEPLKK	3	3	914.53	914.54	-20	0.85	0.90	1.15	1.07
Elongation factor 1 beta	GFGDLKSPAGLQV	2	2	1287.68	1287.68	0	0.87	0.87	1.00	0.95
Protein SET (Phosphatase 2A inhibitor I2PP2A) - isoform 2	Ac-SAPAAKVSKKEL	2	3	1269.72	1269.73	-9	0.90	0.93	0.95	0.96
RNA binding motif protein 3	Ac-SSEEGKLFVGGLNF	2	1	1524.77	1524.75	15	0.90	1.00	0.85	0.85
Vimentin	LIKTVETRDGQVINETSQ	3	2	2030.07	2030.04	2	0.90	0.95	0.90	0.95
FK506 Binding Protein	GVQVETISPGDGRTFPKRGQ	4	2	2128.12	2128.12	9	0.94	1.00	1.05	0.91
Ubiquitin-60S ribosomal protein L40	IIEPSLRQL	2	1	1067.64	1067.64	9	0.94	0.90	0.88	0.81
40S Ribosomal protein S28	Ac-MoxDTSRVQPIKLA	2	1	1415.76	1415.75	6	0.94	1.04	1.04	1.19
Complement component 1 Q subcomponent-binding protein, mitochondrial	ADRGVDNTFADELVEL	2	1	1762.86	1762.84	14	0.94	1.06	1.29	1.06
Peptidylprolyl isomerase A	ADKVPKTAENFRAL	4	3	1558.84	1558.850	-7	0.96	0.99	1.02	1.11
Heat shock 10kDa protein 1 (chaperonin 10)	GGIMLPEKSQGKVLQA	3	3	1654.89	1654.90	-9	0.97	0.97	1.07	0.50
Nucleophosmin	ASIEKGGSLPKVEA	3	3	1384.75	1384.76	-5	0.99	0.99	0.95	0.97

40S Ribosomal protein S21	KADGIVSK	3	3	816.44	816.47	-33	1.00	0.94	1.09	1.12
40S Ribosomal protein S28	Ac-MoxDTSRVQPIKL	2	1	1344.72	1344.71	5	1.00	1.07	1.07	1.19
Heat shock 10kDa protein 1 (chaperonin 10)	GGIMLPEKSQGKVL	3	3	1455.79	1455.81	-14	1.00	0.98	1.05	1.00
Heterogeneous nuclear ribonucleoprotein D0	FGGFGEVESIEL	2	1	1282.61	1282.61	2	1.00	1.00	1.13	0.94
Heat shock 10kDa protein 1 (chaperonin 10)	GSGSKGKGGEIQPVSV	3	3	1485.78	1485.78	-2	1.03	0.93	1.13	0.95
Peptidylprolyl isomerase A	ELFADKVPKTA	3	3	1217.65	1217.67	-16	1.05	1.00	1.16	1.05
Cytochrome c oxidase subunit 5a	GISTPEELGLDKV	2	2	1356.72	1356.71	3	1.06	1.00	1.09	1.09
Nucleophosmin	GGFEITPPVVL	2	1	1127.63	1127.62	4	1.06	1.06	1.13	1.06
Peptidylprolyl isomerase A	VNPTVFFDI	2	1	1050.54	1050.54	5	1.07	1.07	0.93	1.07
FK506 Binding Protein	VFDVELL	2	1	833.46	833.45	3	1.07	1.00	1.07	1.14
Heat shock 10kDa protein 1 (chaperonin 10)	TVVAVGSGSKGKGGEIQPVSV	4	3	1955.07	1955.07	1	1.08	1.10	1.14	1.17
FK506 Binding Protein	VFDVELLKLE	2	2	1203.68	1203.68	2	1.08	1.07	1.11	1.08
40S Ribosomal protein S29	AKDIGFIKLD	3	3	1118.61	1118.63	-18	1.08	0.93	1.03	0.98
Cathepsin D	GPIPEVLK	2	2	851.51	851.51	-4	1.08	1.00	1.00	0.83
Protein SET (Phosphatase 2A inhibitor I2PP2A) - isoform 1,2, or 3	SELIAKI	2	2	772.46	772.47	-17	1.11	1.07	1.15	1.11
Heterogeneous nuclear ribonucleoprotein A/B isoform 1, 2 or 3	FGEFGEIEAIEL	2	1	1352.67	1352.65	15	1.11	1.00	1.11	1.06

Heat shock 10kDa protein 1 (chaperonin 10)	VGSGSKGKGGEIQPVSV	3	3	1584.84	1584.85	-3	1.12	1.10	1.13	1.18
Elongation factor 1 beta	GFGDLKSPAGL	2	2	1060.54	1060.56	-10	1.13	1.06	1.06	1.03
CD99 antigen	AEPAVQRTLLEK	3	2	1353.76	1353.76	1	1.13	1.08	1.11	0.91
Nucleophosmin	EKGGSLPKVEA	3	3	1113.57	1113.60	-26	1.13	1.07	1.24	1.00
Peptidylprolyl isomerase A	ELFADKVPKTAENFRAL	4	3	1948.03	1948.04	-4	1.13	0.93	1.03	1.20
60S acidic ribosomal protein P2 polymorphism	DGKNIEDVIAQGIGKL	3	3	1668.88	1668.91	-15	1.14	1.05	1.19	1.09
Superoxide dismutase 1	KGDGPVQGIINF	2	2	1243.65	1243.66	-6	1.14	1.06	1.14	0.94
40S Ribosomal protein S21	KADGIVSKNF	3	3	1077.57	1077.58	-13	1.14	1.03	1.08	1.08
40S Ribosomal protein S21	AKADGIVSKNF	2	3	1148.61	1148.62	-9	1.15	0.86	1.15	0.98
Heterogeneous nuclear ribonucleoprotein D-like	KDAASVDKVLEL	3	3	1286.69	1286.71	-12	1.15	1.09	1.15	1.09
Triosephosphate isomerase 1	LDPKIAVA	2	2	825.48	825.50	-21	1.15	1.06	1.06	1.00
Heat shock 10kDa protein 1 (chaperonin 10)	LPLFDRVLVE	2	1	1199.70	1199.69	9	1.15	1.19	1.19	1.03
60S acidic ribosomal protein P2	VGIEADDDRLNKV	3	2	1442.74	1442.74	0	1.16	1.02	1.05	1.05
Triosephosphate isomerase 1	SLGELIGTLNA	2	1	1086.60	1086.59	3	1.16	1.16	1.19	1.09
40S Ribosomal protein S21	ADGIVSKNF	2	2	949.48	949.48	-6	1.17	1.10	1.10	1.04
60S Ribosomal protein L31	KNLQTVNVDEN	2	2	1272.64	1272.63	4	1.17	0.97	1.10	0.91
Nucleophosmin	TPKTPKGPSSVEDIKA	4	4	1653.87	1653.89	-12	1.17	1.08	1.17	1.04

Heterogeneous nuclear ribonucleoprotein D-like	ASVDKVLEL	2	2	972.54	972.55	-8	1.17	1.06	1.14	1.07
40S Ribosomal protein S21	AKADGIVSKNF	3	3	1148.60	1148.62	-17	1.18	1.06	1.06	0.94
Elongation factor 1 beta	GFGDLKSPAGLQVL	2	2	1400.76	1400.77	-6	1.18	0.94	0.94	0.82
Nucleophosmin	GGSLPKVEA	2	2	856.46	856.47	-10	1.18	0.99	1.13	0.99
FK506 Binding Protein	VELLKLE	2	2	842.50	842.51	-14	1.19	1.00	1.14	1.04
Complement component 1 Q subcomponent-binding protein, mitochondrial	DRGVDNTFADELVELSTA	2	1	1950.95	1950.92	17	1.19	1.13	0.88	1.13

Non-substrates; peptides no affected (with a decrease ≤20% and an increase ≤120% by the highest concentration of enzyme). See Table 2 for the abbreviation definitions.

Supplemental table 2. Non-substrates of rhCPZ identified using the tryptic peptide library

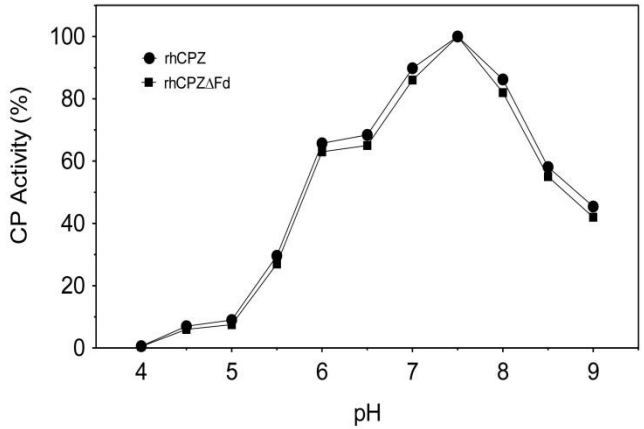
Protein precursor	Sequence	Z	T	Obs M	Theor M	ppm	Ratio rhCPZ / No enzyme			
							100 nM	10 nM	1 nM	0.1 nM
α-Hemoglobin	VDPVNFK	2	2	817.42	817.43	-14	0.82	0.76	0.91	0.61
α-Hemoglobin	LRVDPVNFK	3	2	1086.62	1086.62	-3	0.83	0.80	0.75	0.70
Thyroglobulin	ILNDAQTK	2	2	901.48	901.49	-11	0.85	0.82	0.88	0.82
Thyroglobulin	VTLAADR	2	1	744.42	744.41	12	0.86	0.86	0.97	0.86
Thyroglobulin	ETFLEK	2	2	765.38	765.39	-8	0.87	0.90	1.00	0.83
Thyroglobulin	SALGEPKK	2	1	828.48	828.47	7	0.89	0.86	0.97	0.83
Thyroglobulin	KFEKLPESK	4	4	1104.59	1104.62	-27	0.89	0.90	0.97	0.73
Bovine serum albumin	LVNELTEFAK	2	2	1162.62	1162.62	-4	0.89	0.89	0.92	0.87
Thyroglobulin	IDVALR	2	1	685.42	685.41	12	0.90	0.90	0.93	0.93

Thyroglobulin	QAGVQAEPSPK	2	2	1110.57	1110.57	0	0.91	0.83	1.00	0.83
Thyroglobulin	RSLLLAPEEGPVSQR	3	1	1650.93	1650.91	13	0.93	0.93	1.00	0.83
Thyroglobulin	ALADLAKP	2	2	797.45	797.46	-17	0.93	0.83	0.90	0.87
Thyroglobulin	AVKQFEESQGR	2	2	1277.64	1277.64	1	0.94	0.91	0.97	0.85
Thyroglobulin	ELSVLLPNR	2	1	1039.61	1039.60	8	0.94	1.00	1.00	0.91
Thyroglobulin	LTDEELAFPPLSPSR	3	1	1670.86	1670.85	7	0.95	1.08	1.14	1.02
Bovine serum albumin	DDSPDLPK	2	2	885.39	885.41	-19	0.95	0.92	0.97	0.90
α-Hemoglobin	VLSPADKTNVK	3	3	1170.65	1170.66	-5	0.96	0.91	0.88	0.90
Bovine serum albumin	LVTDLTK	2	2	788.45	788.46	-7	0.97	0.92	0.90	0.88
Thyroglobulin	LGGQEIR	2	1	771.43	771.42	12	0.98	0.95	0.96	0.87
Thyroglobulin	SLSLK	2	2	546.33	546.34	-24	1.00	0.83	0.83	0.83
Thyroglobulin	KVVLQDR	2	2	856.51	856.51	0	1.00	0.97	0.97	0.91
Thyroglobulin	FVAPESLK	2	2	889.48	889.49	-6	1.00	0.93	0.93	0.83
Thyroglobulin	ALADLAKPL	2	2	910.55	910.55	1	1.00	0.91	0.94	0.94
Thyroglobulin	ASGLGAAAGQR	2	1	957.50	957.50	2	1.00	0.90	0.97	0.94
Thyroglobulin	LVTLAESPR	2	1	984.57	984.56	8	1.00	0.93	0.86	0.93
Thyroglobulin	LNSNPASEAPK	2	2	1126.56	1126.56	1	1.00	0.83	0.86	0.83
Thyroglobulin	KGQEFTITGQK	2	3	1235.63	1235.65	-13	1.00	0.87	0.89	0.84
Thyroglobulin	VVLQDR	2	1	728.42	728.42	-1	1.01	0.94	1.05	0.94
Thyroglobulin	SLLLAPEEGPVSQR	3	1	1494.81	1494.80	7	1.02	0.86	1.09	0.85
Bovine serum albumin	DAIPENLPPLTADFAEDK	3	2	1954.97	1954.95	8	1.03	0.90	1.05	0.89
Bovine serum albumin	LVVSTQTALA	2	1	1001.58	1001.58	0	1.03	1.06	1.07	1.07
α-Hemoglobin	Ac-VLSPADKTNVK	2	2	1212.67	1212.67	1	1.03	1.08	1.02	1.00
Thyroglobulin	QQAAALAK	2	2	799.44	799.46	-20	1.03	0.87	0.93	0.83

Bovine serum albumin	AEFVEVTK	2	2	921.47	921.48	-7	1.05	0.97	1.01	0.90
Trypsin	VATVSLPR	2	1	841.51	841.50	12	1.06	0.91	1.03	0.84
Thyroglobulin	FAATSFR	2	1	798.41	798.40	8	1.07	1.00	1.03	1.00
Thyroglobulin	GQEIPGTR	2	1	856.45	856.44	11	1.07	1.00	1.07	1.07
Thyroglobulin	LQQNLFGGR	2	1	1031.54	1031.55	-9	1.07	1.00	0.97	1.04
Thyroglobulin	FLQGDR	2	1	734.37	734.37	6	1.08	0.94	1.00	0.86
Thyroglobulin	RLVTLAESPR	3	1	1140.67	1140.66	11	1.09	0.97	0.98	1.06
Thyroglobulin	LTDEELAFPPLSPSRETFLEK	3	2	2418.25	2418.23	9	1.09	0.97	1.00	0.91
β-Hemoglobin	VNVDEVGGEALGR	2	1	1313.67	1313.66	10	1.17	1.00	1.17	1.00
Thyroglobulin	VDLLIGSSQDDGLINR	2	1	1713.91	1713.89	14	1.18	1.09	1.18	1.09

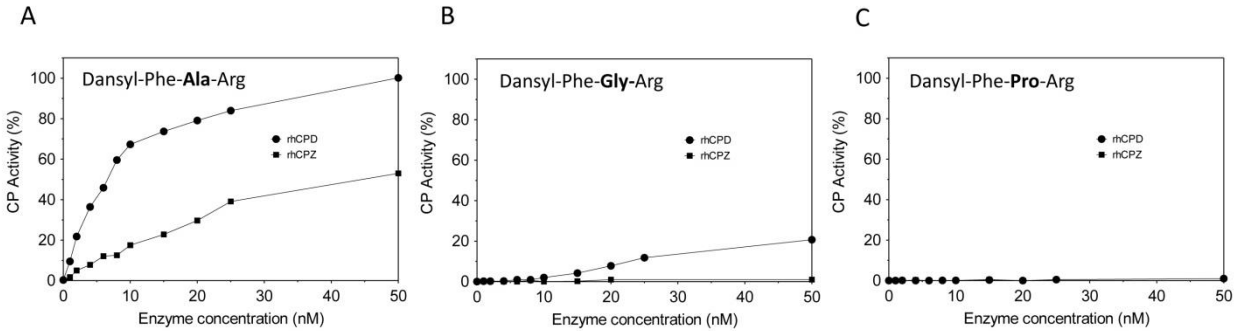
[1]Fragment originated from trypsin autolysis; Non-substrates; peptides no affected (with a decrease ≤20% and an increase ≤120% by the highest concentration of enzyme). See Table 2 for the abbreviation definitions.

Supplemental Figure 1



Supplemental figure 1. Effect of pH on the activity of CPZ purified proteins. Effect of pH on rhCPZ and rhCPZΔFz activity using 200 µM Dansyl-Phe-Ala-Arg in a tris-acetate buffer at the indicated pH for 60 min at 37 ºC, as described in the experimental section. Activity was represented to the maximal activity at optimal pH and represents the average of three independent measures with less than a 10% of variation.

Supplemental Figure 2



Supplemental figure 2. Relative amount of product formed by three different dansylated tripeptides incubated with various amount of purified rhCPZ and rhCPD as control enzyme. A) Reactions containing 200 µM Dansyl-Phe-Ala-Arg, (B) Dansyl-Phe-Gly-Arg (filled squares/solid line) or (C) Dansyl-Phe-Pro-Arg were incubated with different amounts of rhCPZ (triangles/solid line) or rhCPD (filled circles/solid line) in a 100 mM Tris acetate pH 7.5 or in 100 mM Tris acetate pH 6.9, 150 mM NaCl buffer, respectively, for 60 minutes at 37 ºC. Samples were analyzed as described in the in the Experimental section.

General discussion

GENERAL DISCUSSION

Metallocarboxypeptidases are important enzymes that actively participate in the digestion of proteins and peptides. In the recent years, novel members of this family of enzymes have been found in different extra-pancreatic tissues and fluids, comprising a wide range of physiological roles. These recent findings have expanded the implications of MCPs through different areas, including biochemistry, biotechnology and biomedicine and have increased the potential biomedical applications of this large family of proteolytic enzymes.

The present thesis has the aim to gain insights into the knowledge of the structure and biological roles of different regulatory MCPs. We applied a wide range of biochemical approaches to elucidate the biological role of two of these enzymes; human carboxypeptidase D and human carboxypeptidase Z. In addition, we also decided to extent the study to the structure and biological roles of the transthyretin-like domains that are found in all members of this subfamily of proteases. We took as an example the first TTL domain belonging to the first catalytic domain of human CPD.

In the first part of this thesis, we demonstrated for the first time the ability of the h-TTL domain to form amyloid aggregates under native conditions. For this, we provide biophysical and biochemical data that clearly show that at 42 °C, h-TTL has structural properties that are very similar to those of the native state at 25 °C. While at 25 °C the protein remains soluble, at 42 °C the protein aggregates into amyloid-like structures following characteristic sigmoidal kinetics. Before aggregation the protein appears to retain a secondary structure identical to this of the native state and Tyr side-chains report on the existence of a globular compact conformation. Generally, the conversion of folded proteins into amyloid assemblies generally requires non-physiological conditions, such as extreme pHs, organic co-solvents or high temperatures. These destabilizing environments cause proteins to partially or fully unfold leading to the exposure of aggregation prone regions, which are able to form intermolecular interactions, thus triggering aggregation. However, a reduced number of proteins (such as lysozyme, a hyperthermophilic acylphosphatase, superoxide dismutase 1 and

TTR and β2-microglobulin) have been reported to form amyloid aggregates without transitions across the major energy barrier for unfolding, under conditions that are close to the physiological environment. Similarly, in h-TTL amyloid formation occurs under conditions where a protein populates conformations close to its native state (N*), but slightly destabilized, probably due to local fluctuations around a highly amyloidogenic preformed β-strand (see **Figure 1**).

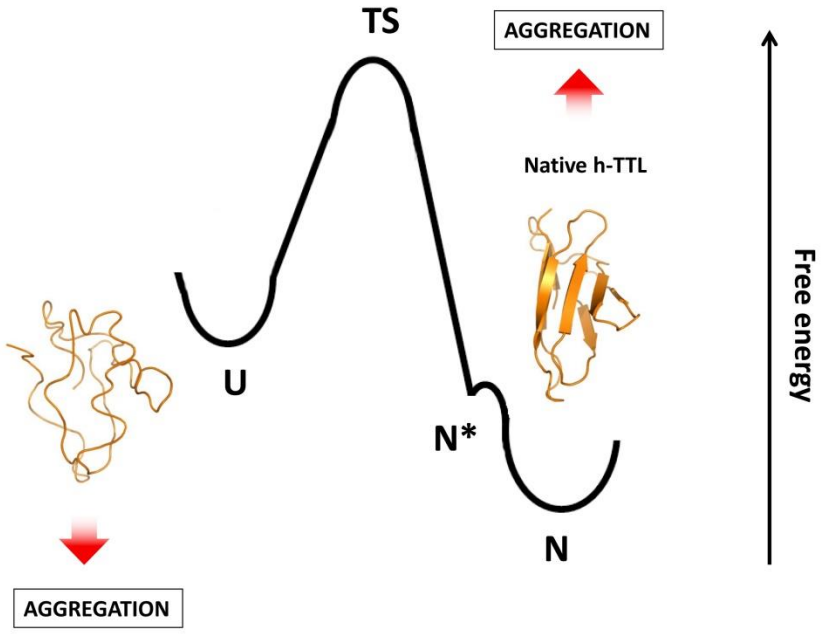


Figure 1. Proposed mechanism for the aggregation of h-TTL under native conditions. Scheme of the free energy landscape proposed for h-TTL. In the scheme is shown the unfolded state (U), consisting of a large ensemble of unstructured conformations. These unstructured conformations can cross the mayor free energy barrier for folding to the native state (N). In the case of h-TTL one or more locally unfolded states (N*) may become accessible from N via thermal fluctuations, probably around a highly amyloidogenic preformed β-strand. Protein molecules adopting the U or N* states conformations can all self-assemble and consequently trigger amyloid formation.

Cellular stress can suffice to promote such destabilization, increasing conformational fluctuations and/or the transient population of partially unfolded conformers, which could trigger aggregation if they exceeded their critical concentration for nucleation. At 25°C the h-TTL is stable enough to skip aggregation, but the protein already aggregates at 37°C, indicating that this side reaction can occur under mild physiological conditions.

A similar mechanism of aggregation was observed in TTR, in which dissociation of the tetrameric structure results in the direct exposure to the solvent of preformed β-strands previously involved in inter-subunit contacts at the interface of the complex. It has been suggested that, upon tetramer dissociation, the TTR monomer experiences only a local unfolding transition, mainly involving the external C and D strands, with β-strands AGH and BEF retaining largely a native-like folded conformation.

In chapter II, we analysed h-TTL from a structural point of view. Here, we provide the crystal structure of the h-TTL domain at 0.95 Å, which represent the highest resolution model of all the structures solved within the MCPs field, as well as structure function implications. Moreover, this is the first report in which is produced and crystallized a TTL domain independently from the catalytic moiety. This domain alone has the ability to fold in a β-sandwich structure with a folding rate (kf) of 13.4 S^{-1} in water (unpublished results). This folding rate is relatively faster in comparison with other domains, such as the SH3 domain with a kf of 3.9 S^{-1} (Ventura et al. 2002). This finding lead to the possibility that this domain might serve for the correct and efficient folding of the carboxypeptidase domain in homology to the prodomains found in other subfamilies of MCPs (Phillips & Rutter 1996). Although, additional studies are needed to confirm this hypothesis. On its overall structure h-TTL is rod-shaped, with the N-t and C-t located 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 and four antiparallel strands, respectively, which are glued by a hydrophobic core. The aggregation prone region found in h-TTL consists on thirteen residues and it is located in the fifth β-sheet, and has some of its residues buried in the hydrophobic core of the protein and the rest exposed to the solvent. This suggests that some of these residues might serve for folding or for structure maintenance. In the crystal structure h-TTL appeared as a dimer (generated from a very concentrated solution of protein used in the crystallization experiment) with a dimer interface of ~350 $Å^2$. However, the dimer could not be confirmed in solution, which suggests that this dimeric structure is an artefact generated during the crystallization process or, alternatively, this could be real oligomeric species formed in the initial steps required for amyloid formation. The crystal dimeric structure is stabilized by three hydrogen bonds and up to 26 non-

bonded contacts mainly stablished between hydrophobic residues. Some of these contacts are found in the interface between residues within the aggregation prone region sequence (form one of the monomers) and residues from the adjacent h-TTL structure.

In chapter III, we studied the enzymatic and functional properties of human carboxypeptidase D using a combination of quantitative peptidomic approaches. Carboxypeptidase D is a membrane-bound multicatalytic enzyme (with three tandem carboxypeptidae domains) enriched in the trans Golgi network that cycles to the cell surface through exocytic and endocytic pathways. After production and purification in mammalian cells we found that this enzyme behaves as a trimer in gel filtration chromatography (with an apparent molecular weight of 450 kDa). This finding is consistent with results obtained from dynamic-light scattering and single particle electronic microscopy analysis (unpublished results) that suggested that this protein is arranged in a trimeric structure in solution. This finding would have important biological implications for the CPD function, since the complete trimeric structure of CPD will form a multi-enzymatic machinery with 9 catalytic sites (with 3 of them probably inactive) working together in the cell membrane. Moreover, we observed that this complex is stable at high salt concentrations (up to 1 M NaCl), suggesting that the structure is stabilized by non-electrostatic interactions (unpublished results). Nonetheless, more structural studies are needed to unravel the conformational arrangement of this intriguing enzyme.

To investigate the enzymatic properties of each domain in human CPD, a critical active site Glu in domain I and II was mutated to Gln and the proteins expressed, purified and assayed with a range of substrates. One of the findings of this study was that human CPD domain I and II have distinct pH optima, with domain I working better at neutral pH while domain II works optimal at mildly acidic pH values. This observation is consistent with the properties of duck and Drosophila enzymes found in previous studies (Novikova et al. 1999; Sidyelyeva et al. 2006), suggesting that this feature is highly conserved through hundreds of millions of years of evolution. Other finding was that human CPD domain I and II have differences in their substrate specificities. Both are specific for C-terminal basic residues, with no detectable

cleavage of non-basic amino acids at the C-terminus. To evaluate the activity of the two active CPD domains, we tested dozens of peptides in a peptidomic assay, and extended previous studies done on duck and Drosophila CPD using a limited number of substrates. We found that each domain was able to cleave either Lys or Arg from some peptides. Since domain I displays a clear preference to cleave Arg and not Lys, and also poses neutral optimum pH, it appear that when CPD is present on the cell surface domain I will be the primarily active and cleaves peptides/proteins with C-terminal Arg. By contrast, the CPD present in the trans Golgi and endocytic pathways will have domain II active and cleave both C-terminal Lys and Arg. A difference with the first two domains, the third carboxypeptidase-like domain of human CPD is inactive as a carboxypeptidase. The structural analysis of the models for the individual CPD domains I, II and III, revealed that the third human CPD domain lacks some of conserved active site, substrate-binding, and metal-binding residues. Most of these changes are conserved between human and duck, providing evidence that these differences are not simply random events but are important for the function of this domain and suggesting that this domain might function in peptide binding, having other non-catalytic roles. This is in agreement with the emerging concept that some carboxypeptidases have multiple functions, including both catalytic and non-catalytic purposes. With this work we stablish the basis for the characterization of the substrate specificity of each active domain of human CPD, however additional research efforts are needed to completely understand the biological roles of this enzyme and to decipher the exact implications of each catalytic domain.

In chapter IV, we described an optimized methodology to produce large amounts of soluble and active heparin-affinity metallocarboxypeptidases. Production of these enzymes can often require particular post-translational modifications, molecular chaperones and co-factors to support their elaborate folding and enzymatic activity. To solve these limitations we used a mammalian-based expression system. The major goal of the present optimized procedure is the addition of sodium heparin after 48 hours post-transfection, in order to improve the amount of protein in the extracellular medium. Using this production system, we improved the expression of three human

heparin-affinity MCPs (carboxypeptidase Z (CPZ), carboxypeptidase A6 (CPA6) and Trombin-Actibable Fibrinolysis Inibitor (TAFI)).

Furthermore, we performed as representative example, the purification of human Carboxypeptidase Z form the extracellular medium. The purified protein is enzymatically active and can be used for functional and structural studies.

In chapter V, we describe the functional characterization of human carboxypeptidase Z by using quantitative peptidomics. This enzyme is unique among the proteases, since is the only proteolytic enzyme with an N-terminal frizzled-like domain. In humans, CPZ is synthetized as a constitutively active enzyme of about 72 kDa, which is secreted through the regulated pathway to the extracellular medium. After secretion an important fraction is bound to the extracellular matrix due to its heparin-binding properties. The analysis of the electrostatic potential into the CPZ surface using a modelled structure showed that the positive charges are located on the face of the molecule that of the active site, suggesting that this may orient CPZ with respect to the ECM. Perhaps the most fundamental benefit of the association of CPZ with the ECM is the capacity to tether and present the enzyme at specific locations in tissues, facilitating the access of substrates to the catalytic cleft. The heparan sulfate molecules that conforms the ECM structure are ideally suited for this role. Its chains on cells can provide ~10^6 binding sites for ligands, a number that far exceeds the number of other types of receptors on the plasma membrane (Xu & Esko 2014). Moreover, the binding of this enzyme to the ECM might have regulatory effects on the CPZ function, such as to act as scaffold to modulate protein-protein interactions, or have an allosteric regulatory effect.

Using the protocol described in chapter IV, we expressed and purified two recombinant forms of human CPZ with or without the frizzled-like domain. Using synthetic substrates, as well as quantitative peptidomics we determined the substrate preferences of human CPZ. This is the first extensive report of the substrate preferences for this enzyme. In previous reports only a couple of fluorescent substrates were tested (Novikova & Fricker 1999a). Here, we tested the activity of CPZ against dozens of peptides by using a similar methodology to that described in chapter III for the characterization of the substrate specificity of the two active

carboxypeptidase D domains. The derived substrate specificity for CPZ shows that this enzyme cleaves exclusively substrates with Arg or Lys C-terminal residues, having a slight preference to hydrolyse Arg versus Lys amino acids. Also, we demonstrated that this enzyme has a k_{cat}/k_m value of 0.0028 ± 0.0008 μM^{-1} S^{-1} against Dansl-Phe-Ala-Arg, being one of the less actives enzymes of this subfamily against conventional substrates. Moreover, we found that lack of the frizzled-like domain in human CPZ does not affect substantively its enzymatic activity against dansyl-Phe-Ala-Arg, since both proteins showed comparable kinetic parameters.

In the present study, in which we used dozens of peptides, we found that this enzyme prefer substrates with small hydrophobic side chains or with polar uncharged side chains like Ser, Thr, Leu, Gly or Arg in the P1 position. The presence of the frizzled domain in CPZ, taken together with its dynamic expression patter during development and localization in the ECM can indicate a special affinity for Wnt molecules. Furthermore, the presence of a Cys residue found immediately before to the C-terminal basic amino acid (Arg or Lys) in the majority of human Wnt proteins fits well with the substrate specificity found here for CPZ, since Cys residues share similar chemical properties with Ser amino acids. If processing does occur at this site in wnt proteins, it is not known whether this affects the biological activity. It is possible that removal of the C-terminal residue activates or inactivates the Wnt, renders it susceptible to further degradation or alters the targeting of the Wnt within the extracellular matrix. In support of this hypothesis, we found that the modelled structure for the frizzled-like domain of human CPZ conserves the most important elements for Wnt binding. Despite our findings, some important questions remain open in relation to the involvement of this enzyme with Wnt molecules. Further work will needed to completely elucidate the function of this enzyme and their possible contribution to the regulation of the Wnt signalling.

Concluding remarks

CONCLUDING REMARKS

Chapter I

1. The h-TTL domain of human carboxypeptidase D could be expressed and purified from *E. coli* as a monomeric protein.
2. The h-TTL domain of human carboxypeptidase D forms amyloid aggregates without extensive unfolding under physiological conditions (at temperatures higher than 37 ºC) .
3. The monomeric transthyretin fold has an inherent propensity to aggregate due to the presence an aggregation prone region localized within the fifth β-sheet.
4. The interaction of h-TTL with membranes modulates its aggregation kinetics. While neutral liposomes abrogate its aggregation, negatively charged model membranes accelerate the aggregation reaction.

Chapter II

5. The structure of the h-TTL domain was solved by X-ray crystallography at ultra-high resolution (0.95 Å), and overall conforms the structure of TTL domains found in M14B MCPs.
6. This domain was found as a homodimer in the crystal structure, but there is no evidence that might behave as a dimer in solution.
7. The crystal dimeric structure is stabilized by three hydrogen bonds and up to 26 non-bonded contacts found in the interface between both monomers, and some of them involve residues within the aggregation prone region.

Chapter III

8. Three catalytically inactive single point mutants for the individual active domains I and II, as well as a double mutant were expressed and purified using a mammalian cell-based expression system.
9. Using quantitative peptidomic approaches we determined that CPD with no mutations or with mutations in just one of the two domains were active towards a subset of peptides with C-terminal Lys or Arg residues.
10. The optimal C-terminal residues for human CPD domain I are Arg (only a couple substrates had C-terminal Lys from dozens of peptide substrates), whereas CPD domain II cleaved equally peptides with C-terminal Lys and Arg.

Chapter IV

11. An easy and inexpensive protocol was developed to improve the production of heparin-affinity carboxypeptidases based on mammalian cells, by supplementing the cell culture medium with sodium heparin.
12. Using this protocol we were able to express and purify high amounts of soluble and active human carboxypeptidase Z.

Chapter V

13. Two recombinant forms of human CPZ with and without the frizzled-like domain were expressed and purified using the protocol described in chapter IV.
14. Using quantitative peptidomic approaches we determined that human CPZ cleaves exclusively peptides with C-terminal basic residues.

15. The lack of the frizzled-like domain had not influence on the catalytic activity of human CPZ against a typical carboxypeptidase substrate.

16. A structural model of the catalytic domain of human CPZ was built, predicting that this enzyme has all the residues essential for the catalytic mechanism.

17. The structural model of the frizzled-like domain of human CPZ has revealed structural similarity and connectivity with the cysteine rich domain of Frizzled receptors, retaining all the major structural determinants needed for Wnt binding.

Bibliography

BIBLIOGRAPHY

Adams, P.D. et al., 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp.213–221.

Ahluwalia, U., Katyal, N. & Deep, S., 2013. Models of Protein Folding. *Journal of Proteins & Proteomics*, 3(December), pp.85–93. Available at: http://www.jpp.org.in/index.php/jpp/article/view/17.

Aloy, P. et al., 2001. The crystal structure of the inhibitor-complexed carboxypeptidase D domain II and the modeling of regulatory carboxypeptidases. *The Journal of biological chemistry*, 276(19), pp.16177–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11278909 [Accessed March 7, 2014].

Alvarez-Santos, S., González-Lafont, A. & Lluch, J.M., 1987. On the water-promoted mechanism of peptide cleavage by carboxypeptidase A. A theoretical study. *Chemistry Letters*, (1), pp.1–4.

Anastas, J.N. & Moon, R.T., 2013. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer*, 13(1), pp.11–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23258168.

Anfinsen, C.B. et al., 1961. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, 47(9), pp.1309–1314.

Arolas, J.L. et al., 2007. Metallocarboxypeptidases: emerging drug targets in biomedicine. *Current pharmaceutical design*, 13(4), pp.349–66. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17311554.

De Baets, G. et al., 2011. An evolutionary trade-off between protein turnover rate and protein aggregation favors a higher aggregation propensity in fast degrading proteins. *PLoS Computational Biology*, 7(6), pp.1–8.

Barnhart, M.M. & Chapman, M.R., 2010. Curli Biogenesis and Function. , pp.131–147.

Bayés, À. et al., 2007. Caught after the act: A human A-type metallocarboxypeptidase in a product complex with a cleaved hexapeptide. *Biochemistry*, 46(23), pp.6921–6930.

Beerten, J., Schymkowitz, J. & Rousseau, F., 2012. Aggregation prone regions and gatekeeping residues in protein sequences. *Current topics in medicinal chemistry*, 12(22), pp.2470–8.

Bemporad, F. et al., 2008. A model for the aggregation of the acylphosphatase from Sulfolobus solfataricus in its native-like state. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1784(12), pp.1986–1996. Available at: http://dx.doi.org/10.1016/j.bbapap.2008.08.021.

Bemporad, F. & Chiti, F., 2009. "Native-like aggregation" of the acylphosphatase from Sulfolobus solfataricus and its biological implications. *FEBS Letters*, 583(16), pp.2630–2638. Available at: http://dx.doi.org/10.1016/j.febslet.2009.07.013.

Berezniuk, I. et al., 2012. Cytosolic carboxypeptidase 1 is involved in processing α- and β-tubulin. *Journal of Biological Chemistry*, 287(9), pp.6503–6517.

Berezniuk, I. et al., 2013. Cytosolic carboxypeptidase 5 removes α- And γ-linked glutamates from tubulin. *Journal of Biological Chemistry*, 288(42), pp.30445–30453.

Berlanga-Acosta, J. et al., 2009. Epidermal growth factor in clinical practice – a review of its biological actions, clinical indications and safety implications. *International Wound Journal*, 6(5), pp.331–346.

Bernstein, K.E. et al., 2013. A modern understanding of the traditional and nontraditional biological functions of angiotensin-converting enzyme. *Pharmacological reviews*, 65(1), pp.1–46. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3565918&tool=pmcentrez&rendertype=abstract.

Bolognesi, B. et al., 2010. ANS binding reveals common features of cytotoxic amyloid species. *ACS Chemical Biology*, 5(8), pp.735–740.

Bovolenta, P. et al., 2008. Beyond Wnt inhibition: new functions of secreted Frizzled-related proteins in development and disease. *Journal of cell science*, 121(Pt 6), pp.737–746.

Breslow, R. & Wernick, D.L., 1977. Unified picture of mechanisms of catalysis by carboxypeptidase A. *Proceedings of the National Academy of Sciences of the United States of America*, 74(4), pp.1303–1307.

Bryngelson, J.D. et al., 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins: Structure, Function and Genetics*, 21(3), pp.167–195.

Bulawa, C.E. et al., 2012. Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade. *Proceedings of the National Academy of Sciences*, 109(24), pp.9629–9634.

Cal, S. et al., 2005. Human polyserase-2, a novel enzyme with three tandem serine protease domains in a single polypeptide chain. *Journal of Biological Chemistry*, 280(3), pp.1953–1961.

Cal, S. et al., 2006. Identification and characterization of human polyserase-3, a novel protein with tandem serine-protease domains in the same polypeptide chain. *BMC biochemistry*, 7, p.9.

Cal, S. et al., 2003. Polyserase-I, a human polyprotease with the ability to generate independent serine protease domains from a single translation product. *Proceedings of the National Academy of Sciences of the United States of America*, 100(16), pp.9185–9190.

Calnan, D.P. et al., 2000. Potency and stability of C terminal truncated human epidermal growth factor. *Gut*, 47(5), pp.622–627.

Castillo, V. et al., 2011. Prediction of the aggregation propensity of proteins from the primary sequence: Aggregation properties of proteomes. *Biotechnology Journal*, 6(6), pp.674–685.

Castillo, V., Chiti, F. & Ventura, S., 2013. The N-terminal helix controls the transition between the soluble and amyloid states of an FF domain. *PloS one*, 8(3), p.e58297. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3591442&tool=pmcentrez&rendertype=abstract [Accessed May 7, 2014].

Castillo, V. & Ventura, S., 2009. Amyloidogenic regions and interaction surfaces overlap in globular proteins related to conformational diseases. *PLoS computational biology*, 5(8), p.e1000476. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2719061&tool=pmcentrez&rendertype=abstract [Accessed February 20, 2014].

Chatani, E. et al., 2012. Polymorphism of beta2-microglobulin amyloid fibrils manifested by ultrasonication-enhanced fibril formation in trifluoroethanol. *Journal of Biological Chemistry*, 287(27), pp.22827–22837.

Cheng, Y., Cawley, N.X. & Loh, Y.P., 2014. Carboxypeptidase E (NF-α1): A new trophic factor in neuroprotection. *Neuroscience Bulletin*, 30(4), pp.692–696.

Chiti, F. & Dobson, C.M., 2009a. Amyloid formation by globular proteins under native conditions. , 5(1), pp.15–23.

Chiti, F. & Dobson, C.M., 2009b. Amyloid formation by globular proteins under native conditions. *Nature chemical biology*, 5(1), pp.15–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19088715 [Accessed May 2, 2014].

Chiti, F. & Dobson, C.M., 2006. Protein misfolding, functional amyloid, and human disease. *Annual review of biochemistry*, 75, pp.333–366.

Claessen, D. et al., 2003. A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in Streptomyces coelicolor by forming amyloid-like fibrils. *Genes and Development*, 17(14), pp.1714–1726.

Clevers, H. & Nusse, R., 2012. Wnt/β-catenin signaling and disease. *Cell*, 149(6), pp.1192–1205.

Cole, C., Barber, J.D. & Barton, G.J., 2008. The Jpred 3 secondary structure prediction server. *Nucleic acids research*, 36(Web Server issue), pp.W197–201. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2447793&tool=pmcentrez&rendertype=abstract [Accessed April 29, 2014].

Cornwell, G. G., 3rd et al., 1988. Evidence that the amyloid fibril protein in senile systemic amyloidosis is derived from normal prealbumin. *Biochem. Biophys. Res. Commun.*, 154, pp.648–653.

Coustou, V. et al., 1997. The protein product of the het-s heterokaryon incompatibility gene of the fungus Podospora anserina behaves as a prion analog. *Proceedings of the National Academy of Sciences of the United States of America*, 94(18), pp.9773–9778.

Damas, a M. & Saraiva, M.J., 2000. Review: TTR amyloidosis-structural features leading to protein aggregation and their implications on therapeutic strategies. *Journal of structural biology*, 130(2-3), pp.290–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10940233 [Accessed March 5, 2014].

Dann, C.E. et al., 2001. Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains. *Nature*, 412(6842), pp.86–90.

Deiteren, K. et al., 2009. Carboxypeptidase M: Multiple alliances and unknown partners. *Clinica Chimica Acta*, 399(1-2), pp.24–39. Available at: http://dx.doi.org/10.1016/j.cca.2008.10.003.

DeLano, W.L., 2002. The PyMOL Molecular Graphics System. Available at: http://www.pymol.org.

Diekmann, H., 1998. The frizzled motif : in how many different protein families does it occur? , (November), pp.415–417.

Dobson, C.M., 2004. Principles of protein folding, misfolding and aggregation. *Seminars in Cell and Developmental Biology*, 15(1), pp.3–16.

Dobson, C.M., 2003. Protein folding and misfolding. *American Scientist*, 426(5), pp.884–890.

Drozdetskiy, a. et al., 2015. JPred4: a protein secondary structure prediction server. *Nucleic Acids Research*, pp.1–6. Available at: http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkv332.

Dumoulin, M., Kumita, J.R. & Dobson, C.M., 2006. Normal and aberrant biological self-assembly: Insights from studies of human lyspzyme and its amyloidogenic variants. *Accounts of Chemical Research*, 39(9), pp.603–610.

Elam, J.S. et al., 2003. Amyloid-like filaments and water-filled nanotubes formed by SOD1 mutant proteins linked to familial ALS. *Nature structural biology*, 10(6), pp.461–467.

Emsley, P. et al., 2010. Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, 66(4), pp.486–501.

Eng, F.J., 1998. gp180, a Protein That Binds Duck Hepatitis B Virus Particles, Has Metallocarboxypeptidase D-like Enzymatic Activity. *Journal of Biological Chemistry*, 273(14), pp.8382–8388. Available at: http://www.jbc.org/cgi/doi/10.1074/jbc.273.14.8382 [Accessed March 7, 2014].

Fändrich, M., Fletcher, M. a & Dobson, C.M., 2001. Amyloid fibrils from muscle myoglobin. *Nature*, 410(6825), pp.165–166.

Fernández, D. et al., 2010. Progress in metallocarboxypeptidases and their small molecular weight inhibitors. *Biochimie*, 92(11), pp.1484–500. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20466032 [Accessed March 7, 2014].

Ferrão-Gonzales, A.D. et al., 2003. Hydration and packing are crucial to amyloidogenesis as revealed by pressure studies on transthyretin variants that either protect or worsen amyloid disease. *Journal of Molecular Biology*, 328(4), pp.963–974.

Ferreira, P. et al., 2013. Structure-based analysis of A19D, a variant of transthyretin involved in familial amyloid cardiomyopathy. *PloS one*, 8(12), p.e82484.

Fersht, a R., 1997. Nucleation mechanisms in protein folding. *Current opinion in structural biology*, 7(1), pp.3–9.

Foguel, D., 2005. High pressure studies on transthyretin. *Protein and peptide letters*, 12(3), pp.245–249.

Foley, J.H. et al., 2013. Insights into thrombin activatable fibrinolysis inhibitor function and regulation. *Journal of Thrombosis and Haemostasis*, 11(SUPPL.1), pp.306–315.

Foss, T.R., Wiseman, R.L. & Kelly, J.W., 2005. The pathway by which the tetrameric protein transthyretin dissociates. *Biochemistry*, 44(47), pp.15525–33.

Freeman, M., 2014. The Rhomboid-Like Superfamily : Molecular Mechanisms and Biological Roles. , (June), pp.1–20.

Fricker, L.D., 2015. Limitations of Mass Spectrometry-Based Peptidomic Approaches. *Journal of The American Society for Mass Spectrometry*. Available at: http://link.springer.com/10.1007/s13361-015-1231-x.

Fricker, L.D., 2005. Neuropeptide-processing enzymes: applications for drug discovery. *The AAPS journal*, 7(2), pp.E449–55. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2750981&tool=pmcentrez&rendertype=abstract.

Fricker, L.D., 2012. *Neuropeptides and Other Bioactive Peptides: From Discovery to Function* L. D. Fricker, ed., Morgan & Claypool Publishers.

Fricker, L.D. & Snyder, S.H., 1982. Enkephalin convertase: purification and characterization of a specific enkephalin-synthesizing carboxypeptidase localized to adrenal chromaffin granules. *Proceedings of the National Academy of Sciences of the United States of America*, 79(12), pp.3886–3890.

García-Castellanos, R. et al., 2005. Detailed molecular comparison between the inhibition mode of A/B-type carboxypeptidases in the zymogen state and by the endogenous inhibitor latexin. *Cellular and Molecular Life Sciences*, 62(17), pp.1996–2014.

Garcia-Pardo, J., Tanco, S., Fernandez-Alvarez, R., et al., 2015. A simple method to improve protein production of heparin-affinity carboxypeptidases using mammalian cells. *In preparation*.

Garcia-Pardo, J. et al., 2014. Amyloid Formation by Human Carboxypeptidase D Transthyretin-like Domain Under Physiological Conditions. *The Journal of biological chemistry*.

Garcia-Pardo, J., Tanco, S., Dasgupta, S., et al., 2015. Substrate specificity of human metallocarboxypeptidase D: Comparison of the two active carboxypeptidase domains. *In preparation*.

García-Sáez, I. et al., 1997. The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. *The EMBO journal*, 16(23), pp.6906–6913.

Gardell, S.J. et al., 1988. A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family. *Journal of Biological Chemistry*, 263(33), pp.17828–17836.

Garnier, M. et al., 1985. Purification and partial characterization of the extracellular gamma-D-glutamyl-(L)meso-diaminopimelate endopeptidase I, from Bacillus sphaericus NCTC 9602. *European journal of biochemistry / FEBS*, 148(3), pp.539–543.

Gelman, J.S. et al., 2013. Alterations of the Intracellular Peptidome in Response to the Proteasome Inhibitor Bortezomib. *PLoS ONE*, 8(1).

Glebe, D. & Urban, S., 2007. Viral and cellular determinants involved in hepadnaviral entry ENVELOPE. , 13(1), pp.22–38.

Go, N., 1984. The consistency principle in protein structure and pathways of folding. *Advances in Biophysics*, 18, pp.149–164.

Goldstein, S.M. et al., 1989. Human mast cell carboxypeptidase. Purification and characterization. *The Journal of clinical investigation*, 83(5), pp.1630–1636.

Gomis-Rüth, F.X. et al., 1999. Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallocarboxypeptidase subfamily. *The EMBO journal*, 18(21), pp.5817–26. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1171647&tool=pmcentrez&rendertype=abstract.

Gomis-Rüth, F.X., 2008. Structure and mechanism of metallocarboxypeptidases. *Critical reviews in biochemistry and molecular biology*, 43(5), pp.319–345.

Goodlad, R. a, Boulton, R. & Playford, R.J., 1996. Comparison of the mitogenic activity of human epidermal growth factor I-53 and epidermal growth factor I-48 in vitro and in vivo. *Clinical science (London, England : 1979)*, 91(4), pp.503–507.

Gregory, H. et al., 1988. The contribution of the C-terminal undecapeptide sequence of urogastrone-epidermal growth factor to its biological action. *Regulatory peptides*, 22(3), pp.217–226.

Guasch, a et al., 1992. Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. *Journal of molecular biology*, 224(1), pp.141–157.

Guijarro, J.I. et al., 1998. Amyloid fibril formation by an SH3 domain. *Proceedings of the National Academy of Sciences of the United States of America*, 95(8), pp.4224–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=22470&tool=pmcentrez&rendertype=abstract.

Hall, M. et al., 2014. Structure of the C-terminal domain of AspA (antigen I/II-family) protein from Streptococcus pyogenes. *FEBS Open Bio*, 4, pp.283–289. Available at: http://dx.doi.org/10.1016/j.fob.2014.02.012.

Hamilton, J. a & Benson, M.D., 2001. Transthyretin : a review from a structural perspective. *Cellular and Molecular Life Sciences*, 58(10), pp.1491– 1521. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11693529.

Hammarström, P. et al., 2002. Sequence-dependent denaturation energetics: A major determinant in amyloid disease diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 99 Suppl 4, pp.16427–16432.

Harata, K. et al., 1996. X-ray structure of cyclodextrin glucanotransferase from alkalophilic Bacillus sp. 1011. Comparison of two independent molecules at 1.8 A resolution. *Acta crystallographica. Section D, Biological crystallography*, 52(Pt 6), pp.1136–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15299574 [Accessed March 12, 2014].

Harrison, C. & Acharya, K.R., 2014. ACE for all – a molecular perspective. *Journal of Cell Communication and Signaling*, 8(3), pp.195–210. Available at: http://link.springer.com/10.1007/s12079-014-0236-8.

Heckman, K.L. & Pease, L.R., 2007. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature protocols*, 2(4), pp.924–932.

Hörnberg, a et al., 2000. A comparative analysis of 23 structures of the amyloidogenic protein transthyretin. *Journal of molecular biology*, 302(3), pp.649–669.

Hörnberg, A. et al., 2004. The β-strand D of transthyretin trapped in two discrete conformations. *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1700(1), pp.93–104.

Hou, X., Aguilar, M.-I. & Small, D.H., 2007. Transthyretin and familial amyloidotic polyneuropathy. Recent progress in understanding the molecular mechanism of neurodegeneration. *The FEBS journal*, 274(7), pp.1637–50.

Hourdou, M.L. et al., 1993. Characterization of the sporulation-related gamma-D-glutamyl-(L)meso-diaminopimelic-acid-hydrolysing peptidase I of Bacillus sphaericus NCTC 9602 as a member of the metallo(zinc) carboxypeptidase A family. Modular design of the protein. *The Biochemical journal*, 292 Pt 2, pp.563–570.

Invernizzi, G. et al., 2012. Protein aggregation: Mechanisms and functional consequences. *The International Journal of Biochemistry & Cell Biology*, 44(9), pp.1541–1554. Available at: http://dx.doi.org/10.1016/j.biocel.2012.05.023.

Janda, C.Y. et al., 2012. Structural Basis of Wnt Recognition by Frizzled. *Science*, 337(6090), pp.59–64.

Jensen, M.H. et al., 2010. Structural and biochemical studies elucidate the mechanism of rhamnogalacturonan lyase from aspergillus aculeatus. *Journal of Molecular Biology*, 404(1), pp.100–111. Available at: http://dx.doi.org/10.1016/j.jmb.2010.09.013.

Jiang, X. et al., 2001. An engineered transthyretin monomer that is nonamyloidogenic, unless it is partially denatured. *Biochemistry*, 40(38), pp.11442–11452.

Jung, G., Ueno, H. & Hayashi, R., 1998. Proton-relay system of carboxypeptidase Y as a sole catalytic site: studies on mutagenic replacement of his 397. *Journal of biochemistry*, 124(2), pp.446–450.

Kabsch, W., 2010. Xds. *Acta Crystallographica Section D: Biological Crystallography*, 66(2), pp.125–132.

Kalinina, E. et al., 2007. A novel subfamily of mouse cytosolic carboxypeptidases. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 21(3), pp.836–850.

Kalinina, E., Varlamov, O. & Fricker, L.D., 2002. Analysis of the carboxypeptidase D cytoplasmic domain: Implications in intracellular trafficking. *Journal of Cellular Biochemistry*, 85(1), pp.101–111.

Kang, Y. et al., 2012. Structural study of TTR-52 reveals the mechanism by which a bridging molecule mediates apoptotic cell engulfment. *Genes and Development*, 26(12), pp.1339–1350.

Karplus, M. & Weaver, D.L., 1976. Protein-folding dynamics. *Group*, 260, pp.404–406. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1264226.

Keil, C. et al., 2007. Crystal structure of the human carboxypeptidase N (kininase I) catalytic domain. *Journal of molecular biology*, 366(2), pp.504–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17157876 [Accessed March 7, 2014].

Kim, Y.E. et al., 2013. *Molecular chaperone functions in protein folding and proteostasis.*, Available at: http://www.ncbi.nlm.nih.gov/pubmed/23746257.

Knoot, C.J., Purpero, V.M. & Lipscomb, J.D., 2015. Crystal structures of alkylperoxo and anhydride intermediates in an intradiol ring-cleaving dioxygenase. *Proceedings of the National Academy of Sciences*, 112(2), pp.388–393. Available at: http://www.pnas.org/lookup/doi/10.1073/pnas.1419118112.

Knowles, T.P.J., Vendruscolo, M. & Dobson, C.M., 2014. The amyloid state and its association with protein misfolding diseases. *Nature reviews. Molecular cell*

biology, 15(6), pp.384–96. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24854788.

Kos, J. et al., 2015. Seminars in Cancer Biology Intracellular signaling by cathepsin X : Molecular mechanisms and diagnostic and therapeutic opportunities in cancer. *Seminars in Cancer Biology*, 31, pp.76–83.

Kumar, S. & Walter, J., 2011. Phosphorylation of amyloid beta (Aβ) peptides - A trigger for formation of toxic aggregates in Alzheimer's disease. *Aging*, 3(8), pp.803–812.

Kuroki, K. et al., 1994. A cell surface protein that binds avian hepatitis B virus particles. *Journal of virology*, 68(4), pp.2091–2096.

Kuroki, K. et al., 1995. gp180, a host cell glycoprotein that binds duck hepatitis B virus particles, is encoded by a member of the carboxypeptidase gene family. *Journal of Biological Chemistry*, 270(25), pp.15022–15028. Available at: http://www.jbc.org/content/270/25/15022.short [Accessed March 7, 2014].

Lai, Z., Colón, W. & Kelly, J.W., 1996. The acid-mediated denaturation pathway of transthyretin yields a conformational intermediate that can self-assemble into amyloid. *Biochemistry*, 35(20), pp.6470–6482.

Larkin, M. a et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), pp.2947–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17846036 [Accessed May 23, 2014].

Layne, M.D. et al., 1998. Aortic carboxypeptidase-like protein, a novel protein with discoidin and carboxypeptidase-like domains, is up-regulated during vascular smooth muscle cell differentiation. *Journal of Biological Chemistry*, 273(25), pp.15654–15660.

Leimeister, C., Bach, A. & Gessler, M., 1998. Developmental expression patterns of mouse sFRP genes encoding members of the secreted frizzled related protein family. *Mechanisms of Development*, 75(1-2), pp.29–42.

Levinthal, C., 1968. Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique*, 65, pp.44–45. Available at: http://www.biochem.wisc.edu/courses/biochem704/Reading/Levinthal1968.pdf.

Levy, E.D., De, S. & Teichmann, S. a, 2012. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proceedings of the National Academy of Sciences of the United States of America*, 109(50), pp.20461–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3528536&tool=pmcentrez&rendertype=abstract.

Lipscomb, W.N. et al., 1969. THE STRUCTURE OF CARBOXYPEPTIDASE A, IX. THE X-RAY DIFFRACTION RESULTS IN THE LIGHT OF THE CHEMICAL SEQUENCE. *Proceedings of the National Academy of Sciences of the United States of America*, 64(August 1966), pp.28–35.

López-Otín, C. & Overall, C.M., 2002. Protease degradomics: a new challenge for proteomics. *Nature reviews. Molecular cell biology*, 3(7), pp.509–519.

Lyons, P.J., Callaway, M.B. & Fricker, L.D., 2008. Characterization of carboxypeptidase A6, an extracellular matrix peptidase. *Journal of Biological Chemistry*, 283(11), pp.7054–7063.

Lyons, P.J. & Fricker, L.D., 2011. Carboxypeptidase O is a glycosylphosphatidylinositol-anchored intestinal peptidase with acidic amino acid specificity. *Journal of Biological Chemistry*, 286(45), pp.39023–39032.

Lyons, P.J. & Fricker, L.D., 2012. Peptidomic approaches to study proteolytic activity. *Changes*, 29(6), pp.997–1003.

Lyons, P.J. & Fricker, L.D., 2010. Substrate specificity of human carboxypeptidase A6. *Journal of Biological Chemistry*, 285(49), pp.38234–38242.

Maurer-Stroh, S. et al., 2010. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nature methods*, 7, pp.237–242.

Mesters, J.R. et al., 2006. Structure of glutamate carboxypeptidase II, a drug target in neuronal damage and prostate cancer. *The EMBO journal*, 25(6), pp.1375–1384.

Moeller, C. et al., 2003. Carboxypeptidase Z (CPZ) modulates Wnt signaling and regulates the development of skeletal elements in the chicken. *Development (Cambridge, England)*, 130(21), pp.5103–11. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12944424 [Accessed March 4, 2014].

Monsellier, E. & Chiti, F., 2007. Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO reports*, 8(8), pp.737–742.

Morano, C., Zhang, X. & Fricker, L.D., 2008. Multiple isotopic labels for quantitative mass spectrometry. *Analytical Chemistry*, 80(23), pp.9298–9309.

Motulsky HJ, GraphPad Prism. Available at: www.graphpad.com.

Nakayama, K., 1997. Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *The Biochemical journal*, 327 (Pt 3, pp.625–635.

Nordlund, A. & Oliveberg, M., 2006. Folding of Cu/Zn superoxide dismutase suggests structural hotspots for gain of neurotoxic function in ALS: parallels to precursors in amyloid disease. *Proceedings of the National Academy of Sciences of the United States of America*, 103(27), pp.10218–10223.

Novikova, E., Fricker, L.D. & Reznik, S.E., 2001. Metallocarboxypeptidase Z is dynamically expressed in mouse development. *Mechanisms of Development*, 102(1-2), pp.259–262.

Novikova, E.G. et al., 2000. Carboxypeptidase Z is present in the regulated secretory pathway and extracellular matrix in cultured cells and in human tissues. *The Journal of biological chemistry*, 275(7), pp.4865–4870.

Novikova, E.G. et al., 1999. Characterization of the enzymatic properties of the first and second domains of metallocarboxypeptidase D. *Journal of Biological Chemistry*, 274(41), pp.28887–28892.

Novikova, E.G. & Fricker, L.D., 1999a. Purification and characterization of human metallocarboxypeptidase Z. *Biochemical and biophysical research communications*, 256(3), pp.564–568.

Novikova, E.G. & Fricker, L.D., 1999b. Purification and characterization of human metallocarboxypeptidase Z. *Biochemical and biophysical research communications*, 256(3), pp.564–568.

Núñez, C. et al., 2014. A novel quinoline molecular probe and the derived functionalized gold nanoparticles: Sensing properties and cytotoxicity studies in MCF-7 human breast cancer cells. *Journal of Inorganic Biochemistry*, 137, pp.115–122. Available at: http://dx.doi.org/10.1016/j.jinorgbio.2014.04.007.

Ogiso, H. et al., 2002. Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. *Cell*, 110(6), pp.775–787.

Olofsson, A. et al., 2004. Probing Solvent Accessibility of Transthyretin Amyloid by Solution NMR Spectroscopy. *Journal of Biological Chemistry*, 279(7), pp.5699–5707.

Olsen, J. V, Ong, S.-E. & Mann, M., 2004. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular & cellular proteomics : MCP*, 3(6), pp.608–614.

Onuchic, J.N., Luthey-Schulten, Z. & Wolynes, P.G., 1997. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annual Review of Physical Chemistry*, 48(1), pp.545–600.

Orville, a M. et al., 1997. Structures of competitive inhibitor complexes of protocatechuate 3,4-dioxygenase: multiple exogenous ligand binding orientations

within the active site. *Biochemistry*, 36(33), pp.10039–51. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9254599.

Otero, a. et al., 2012. The novel structure of a cytosolic M14 metallocarboxypeptidase (CCP) from Pseudomonas aeruginosa: a model for mammalian CCPs. *The FASEB Journal*, 26(9), pp.3754–3764.

Pallaŕs, I. et al., 2004. Amyloid fibril formation by a partially structured intermediate state of α-chymotrypsin. *Journal of Molecular Biology*, 342(1), pp.321–331.

Panosa, C. et al., 2013. Development of an Epidermal Growth Factor Derivative with EGFR Blocking Activity. *PLoS ONE*, 8(7).

Pechmann, S. et al., 2009. Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 106(25), pp.10159–10164.

Petrera, A., Lai, Z.W. & Schilling, O., 2014. Carboxyterminal Protein Processing in Health and Disease : Key Actors and Emerging Technologies.

Phillips, M. a. & Rutter, W.J., 1996. Role of the prodomain in folding and secretion of rat pancreatic carboxypeptidase A1. *Biochemistry*, 35(21), pp.6771–6776.

Playford, R.J. et al., 1995. Epidermal growth factor is digested to smaller, less active forms in acidic gastric juice. *Gastroenterology*, 108(1), pp.92–101.

Portolano, N. et al., 2014. Recombinant Protein Expression for Structural Biology in HEK 293F Suspension Cells: A Novel and Accessible Approach. *Journal of Visualized Experiments*, 1(92), pp.1–8. Available at: http://www.jove.com/video/51897/recombinant-protein-expression-for-structural-biology-hek-293f.

Ptitsyn, O., 1973. Stages in the mechanism of self-organization of protein molecules. *Dokl Akad Nauk SSSR*, 210(5), pp.1213–5.

Puente, X.S. et al., 2003. Human and mouse proteases: a comparative genomic approach. *Nature reviews. Genetics*, 4(7), pp.544–558.

Quesada, V. et al., 2009. The Degradome database: mammalian proteases and diseases of proteolysis. *Nucleic acids research*, 37(Database issue), pp.D239–D243.

Quintas, A. et al., 2001. Tetramer Dissociation and Monomer Partial Unfolding Precedes Protofibril Formation in Amyloidogenic Transthyretin Variants. *Journal of Biological Chemistry*, 276(29), pp.27207–27213.

Ramírez-Alvarado, M., Cocco, M.J. & Regan, L., 2003. Mutations in the B1 domain of protein G that delay the onset of amyloid fibril formation in vitro. *Protein science : a publication of the Protein Society*, 12(3), pp.567–576.

Rawlings, N.D. et al., 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res*, 42, pp.D503–D509.

Rawlings, N.D. et al., 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic acids research*, 42(Database issue), pp.D503–9.

Rawlings, N.D. & Barrett, A.J., 1993. Evolutionary families of metallopeptidases. *Methods in Enzymology*, 248, pp.183–228.

Rawlings, N.D. & Barrett, A.J., 1999. MEROPS: The peptidase database. *Nucleic Acids Research*, 27(1), pp.325–331.

Reverter, D. et al., 2004. Crystal structure of human carboxypeptidase M, a membrane-bound enzyme that regulates peptide hormone activity. *Journal of molecular biology*, 338(2), pp.257–69. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15066430 [Accessed March 7, 2014].

Reya, T. & Clevers, H., 2005. Wnt signalling in stem cells and cancer. *Nature*, 434(7035), pp.843–850.

Reznik, S.E. & Fricker, L.D., 2001. Carboxypeptidases from A to Z : implications in embryonic development and Wnt binding. *Cellular and Molecular Life Sciences*, 58, pp.1790–1804.

Rodriguez de la Vega, M. et al., 2007a. Nna1-like proteins are active metallocarboxypeptidases of a new and diverse M14 subfamily. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 21(3), pp.851–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17244817 [Accessed March 1, 2014].

Rodriguez de la Vega, M. et al., 2007b. Nna1-like proteins are active metallocarboxypeptidases of a new and diverse M14 subfamily. *The FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 21(3), pp.851–865.

Rogowski, K. et al., 2010. A family of protein-deglutamylating enzymes associated with neurodegeneration. *Cell*, 143(4), pp.564–578. Available at: http://dx.doi.org/10.1016/j.cell.2010.10.014.

Rowsell, S. et al., 1997. Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure (London, England : 1993)*, 5(3), pp.337–347.

Sabate, R. et al., 2012. Native structure protects SUMO proteins from aggregation into amyloid fibrils. *Biomacromolecules*, 13(6), pp.1916–26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22559198.

Sabaté, R., Espargaró, A., et al., 2012. Effect of the surface charge of artificial model membranes on the aggregation of amyloid β-peptide. *Biochimie*, 94(8), pp.1730–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22542639 [Accessed February 19, 2014].

Sabaté, R., Villar-Piqué, A., et al., 2012. Temperature dependence of the aggregation kinetics of Sup35 and Ure2p yeast prions. *Biomacromolecules*, 13(2), pp.474–483.

Sanchez-Ruiz, J.M., 1992. Theoretical analysis of Lumry-Eyring models in differential scanning calorimetry. *Biophysical journal*, 61(4), pp.921–935.

Sanglas, L. et al., 2008. Structure of Activated Thrombin-Activatable Fibrinolysis Inhibitor, a Molecular Link between Coagulation and Fibrinolysis. *Molecular Cell*, 31(4), pp.598–606.

Sapio, M. & D., F.L., 2014. Carboxypeptidases in disease: Insights from peptidomic studies. *Proteomics Clinical applications*, 29(6), pp.327–337.

Saraiva, M. et al., 1983. Presence of an abnormal transthyretin (prealbumin) in Portuguese patients with familial amyloidotic polyneuropathy. *Trans. Assoc. Am. Physicians*, 96, pp.261–270.

Schechter, I. & Berger, a, 1967. On the size of the active site in proteases. I. Papain. *Biochemical and biophysical research communications*, 27(2), pp.157–162.

Sekijima, Y. et al., 2005. The biological and chemical basis for tissue-selective amyloid disease. *Cell*, 121(1), pp.73–85.

Sidyelyeva, G., Baker, N.E. & Fricker, L.D., 2006. Characterization of the molecular basis of the Drosophila mutations in carboxypeptidase D. Effect on enzyme activity and expression. *The Journal of biological chemistry*, 281(19), pp.13844–52. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16556608 [Accessed April 18, 2014].

Skalka, N. et al., 2013. Carboxypeptidase E: a negative regulator of the canonical Wnt signaling pathway. *Oncogene*, 32(23), pp.2836–2847.

Skidgel, R. a, David, R.M. & Tan, F., 1989. Purification and characterization of a membrane-bound carboxypeptidase that cleaves peptide hormones. *The Journal of biological chemistry*, pp.2236–2241.

Skidgel, R. a & Erdos, E.G., 2007. Structure and function of human plasma carboxypeptidase N, the anaphylatoxin inhibitor. *In Immunopharmacol*, 7(14), pp.1888–1899.

Soisson, S.M. et al., 2010. Structural definition and substrate specificity of the S28 protease family: the crystal structure of human prolylcarboxypeptidase. *BMC structural biology*, 10, p.16.

Soldi, G. et al., 2005. Amyloid formation of a protein in the absence of initial unfolding and destabilization of the native state. *Biophysical journal*, 89(6), pp.4234–4244.

Song, L. & Fricker, L., 1997. Cloning and expression of human carboxypeptidase Z, a novel metallocarboxypeptidase. *Journal of Biological Chemistry*, 272(16), pp.10543–10550. Available at: http://www.jbc.org/content/272/16/10543.short [Accessed March 4, 2014].

Sorimachi, K. et al., 1997. Solution structure of the granular starch binding domain of Aspergillus niger glucoamylase bound to beta-cyclodextrin. *Structure (London, England : 1993)*, 5(5), pp.647–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9195884.

Tanco, S. et al., 2010. Characterization of the substrate specificity of human carboxypeptidase A4 and implications for a role in extracellular peptide processing. *Journal of Biological Chemistry*, 285(24), pp.18385–18396.

Tanco, S., Tort, O., et al., 2015. C-terminomics Screen for Natural Substrates of Cytosolic Carboxypeptidase 1 Reveals Processing of Acidic Protein C termini. *Molecular & Cellular Proteomics*, 14(1), pp.177–190. Available at: http://www.mcponline.org/lookup/doi/10.1074/mcp.M114.040360.

Tanco, S. et al., 2013. Proteome-derived peptide libraries to study the substrate specificity profiles of carboxypeptidases. *Molecular & cellular proteomics : MCP*, 12(8), pp.2096–110. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23620545.

Tanco, S. et al., 2010. Structure-function analysis of the short splicing variant carboxypeptidase encoded by Drosophila melanogaster silver. *Journal of molecular biology*, 401(3), pp.465–77. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20600119 [Accessed January 31, 2014].

Tanco, S., Gevaert, K. & Van Damme, P., 2015. C-terminomics: Targeted analysis of natural and posttranslationally modified protein and peptide C-termini. *Proteomics*, 15(5-6), pp.903–914. Available at: http://doi.wiley.com/10.1002/pmic.201400301.

Tom, R., Bisson, L. & Durocher, Y., 2008. Transfection of HEK293-EBNA1 cells in suspension with linear PEI for production of recombinant proteins. *Cold Spring Harbor Protocols*, 3(3), pp.1–5.

Tort, O. et al., 2014. The cytosolic carboxypeptidases CCP2 and CCP3 catalyze posttranslational removal of acidic amino acids. , pp.1–35.

True, H.L. & Lindquist, S.L., 2000. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature*, 407(6803), pp.477–483.

Tumelty, K.E. et al., 2014. Aortic carboxypeptidase-like protein (ACLP) enhances lung myofibroblast differentiation through transforming growth factor ?? receptor-dependent and -independent pathways. *Journal of Biological Chemistry*, 289(5), pp.2526–2536.

Turk, B., 2006. Targeting proteases: successes, failures and future prospects. *Nature reviews. Drug discovery*, 5(9), pp.785–799.

Valnickova, Z. et al., 2007. Thrombin-activable fibrinolysis inhibitor (TAFI) zymogen is an active carboxypeptidase. *Journal of Biological Chemistry*, 282(5), pp.3066–3076.

Vendrell, J. & Avilés, F.X., 1999. Carboxypeptidases. In V. Turk, ed. *Proteases: new perspectives*. Basel, Birkhauser, pp. 13–34.

Vendrell, J., Querol, E. & Avilés, F.X., 2000. Metallocarboxypeptidases and their protein inhibitors: Structure, function and biomedical properties. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1477(1-2), pp.284–298.

Ventura, S. et al., 2002. Conformational strain in the hydrophobic core and its implications for protein folding and design. *Nature structural biology*, 9(6), pp.485–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12006985 [Accessed April 28, 2014].

Ventura, S. et al., 2004. Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19), pp.7258–7263.

Ventura, S., Villegas, V. & Sterner, J., 1999. Mapping the pro-region of carboxypeptidase B by protein engineering. *Journal of Biological …*, 274(28), pp.19925–19933. Available at: http://www.jbc.org/content/274/28/19925.short.

Vincan, E., 2009. *Wnt Signaling* E. Vincan, ed., human Press.

Vink, T. et al., 2014. A simple, robust and highly efficient transient expression system for producing antibodies. *Methods*, 65(1), pp.5–10. Available at: http://dx.doi.org/10.1016/j.ymeth.2013.07.018.

Wang, L., Shao, Y.Y. & Ballock, R.T., 2009. Carboxypeptidase Z (CPZ) links thyroid hormone and Wnt signaling pathways in growth plate chondrocytes. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research*, 24(2), pp.265–273. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3276606&tool=pmcentrez&rendertype=abstract [Accessed March 4, 2014].

Wei, S. et al., 2002. *Identification and Characterization of Three Members of the Human Metallocarboxypeptidase Gene Family Identification and Characterization of Three Members of the Human Metallocarboxypeptidase Gene Family*,

Wernersson, S. & Pejler, G., 2014. Mast cell secretory granules: armed for battle. *Nature reviews. Immunology*, 14(7), pp.478–94. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24903914.

Westermark, P. et al., 1990. Fibril in senile systemic amyloidosis is derived from normal transthyretin. *Proceedings of the National Academy of Sciences of the United States of America*, 87(7), pp.2843–2845.

Wetlaufer, D.B., 1973. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 70(3), pp.697–701.

Winn, M.D. et al., 2011. Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), pp.235–242.

Xiaonan, X. et al., 1998. Cloning , Sequence Analysis , and Distribution of Rat Metallocarboxypeptidase Z. *DNA and cell biology*, 17(4), pp.311–319.

Xin, X. et al., 1998. Identification of mouse CPX-2, a novel member of the metallocarboxypeptidase gene family: cDNA cloning, mRNA distribution, and protein expression and characterization. *DNA and cell biology*, 17(10), pp.897–909.

Xu, D. & Esko, J.D., 2014. Demystifying Heparan Sulfate-Protein Interactions. *Annual review of biochemistry*, (February), pp.1–29. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24606135.

Ye, Y. & Godzik, a., 2003. Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics*, 19(Suppl 2), pp.ii246–ii255. Available at: http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btg1086 [Accessed May 7, 2014].

Young, P.G. et al., 2014. Structural Conservation, Variability, and Immunogenicity of the T6 Backbone Pilin of Serotype M6 Streptococcus pyogenes. *Infection and Immunity*, 82(7), pp.2949–2957. Available at: http://iai.asm.org/cgi/doi/10.1128/IAI.01706-14.

Zhang, X. et al., 2008. Carboxypeptidase M and kinin B1 receptors interact to facilitate efficient B1 signaling from B2 agonists. *Journal of Biological Chemistry*, 283(12), pp.7994–8004.

Zhang, X. et al., 2008. Peptidomics of Cpefat/fat mouse brain regions: Implications for neuropeptide processing. , 107(6), pp.1596–1613.

Zhang, X., Tan, F. & Skidgel, R. a., 2013. Carboxypeptidase M is a positive allosteric modulator of the kinin B1 receptor. *Journal of Biological Chemistry*, 288(46), pp.33226–33240.

Zhang, Y., 2008a. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, 9, p.40. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2245901&tool=pmcentrez&rendertype=abstract [Accessed April 29, 2014].

Zhang, Y., 2008b. I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, 9, p.40. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2245901&tool=pmcentrez&rendertype=abstract [Accessed January 20, 2014].

Zhou, a et al., 1999. Proteolytic processing in the secretory pathway. *The Journal of biological chemistry*, 274(30), pp.20745–20748.

Zhuravlev, P.I. et al., 2014. Propensity to form amyloid fibrils is encoded as excitations in the free energy landscape of monomeric proteins. *Journal of Molecular Biology*, 426(14), pp.2653–2666. Available at: http://dx.doi.org/10.1016/j.jmb.2014.05.007.