




Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

Obtention, Biochemical and Structural Characterization of a hALDH1A3 mutant and Ovocalyxin-32, and field improvements

Sergi Montané Bel

November 2020

Obtention, Biochemical and Structural characterization of a hALDH1A3 mutant and Ovocalyxin-32, and field improvements

Doctoral thesis presented by Sergi Montané Bel for the degree of PhD in Biochemistry,
Molecular Biology and Biomedicine from Universitat Autònoma de Barcelona

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

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

Sergi Montané Bel

Dr. Julia Lorenzo Rivera

Dr. Francesc Xavier Avilés i Puigvert

Bellaterra, November 2020

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS.....	5
TABLE OF ABBREVIATIONS	11
INTRODUCTION	14
THE WORLD OF ENZYMES, INHIBITORS AND REGULATORS	14
A1. OXIDOREDUCTASES.....	17
A1.1. Aldehyde dehydrogenases.....	18
A1.1. Members of ALDH family in humans.....	23
B1. PROTEASES	27
B1.1. Metalloproteases.....	29
B1.2. Substrate specificity and catalytic mechanism of MCPs.....	35
B2. PROTEASE INHIBITORS	39
B2.1. Metalloprotease inhibitors	40
B2.2. Avian egg.....	46
COMMENTS AND JUSTIFICATIONS ON THE SUBJECTS OF THIS THESIS.....	55
OBJECTIVES.....	59
CHAPTER I: CHARACTERIZATION OF A MUTANT VARIANT OF HUMAN ALDH1A3 (C314S MUTANT).....	63
1.1. INTRODUCTION	63
1.2. EXPERIMENTAL PROCEDURES.....	64
1.2.1. Bradford assay	64
1.2.2. Sodium Dodecyl Sulphate Polycrylamide Gel Electrophoresis (SDS-PAGE)	64
1.2.3. Coomassie SDS-PAGE staining.....	66
1.2.4. DNA mutagenesis	66
1.2.5. Protein expression and purification	68
1.2.6. Activity fluorescence ALDH1A3 C314S mutant activity assay.	69
1.2.7. Crystallization of protein and data collection	70

1.2.8. Structure solution	70
1.3. RESULTS	72
1.3.1. Protein production and purification	72
1.3.2. Dehydrogenase activity for ALDH1A3 C314S	73
1.3.3. Resolution Structural characteristics from hALDH1A3 C314S mutant	75
1.4. DISCUSSION.....	82
1.5. SUPPLEMENTARY DATA.....	84
CHAPTER II: EFFICIENT AND FACILE PRODUCTION OF RECOMBINANT CARBOXYPEPTIDASES IN MAMMALIAN CELLS.....	87
2.1. INTRODUCTION	87
2.2. EXPERIMENTAL PROCEDURES.....	89
2.2.1. Sodium Dodecyl Sulphate Polycrylamide Gel Electrophoresis (SDS-PAGE).....	89
2.2.2. Western blot	89
2.2.3. DNA Techniques	90
2.2.4. Cell Culture.....	91
2.2.5. Data evaluation	93
2.3. RESULTS	94
2.3.1. Heparin affinity proteins	94
2.3.2. Cell Viability in Heparin presence.....	97
2.3.3. Improved expression of Heparin affinity MCPs.....	98
2.4. DISCUSSION.....	102
CHAPTER III: EXTRACTION, PURIFICATION AND CHARACTERIZATION OF NATURAL OVOCALYXIN-32	105
3.1. INTRODUCTION	105
3.2. EXPERIMENTAL PROCEDURES.....	108
3.2.1. Bicinchoninic Acid Assay	108
3.2.2. Estimation total pure protein by absorbance	109
3.2.3. Sodium Dodecyl Sulphate Polycrylamide Gel Electrophoresis (SDS-PAGE):.....	110

3.2.4. Coomassie SDS-PAGE staining.....	110
3.2.5. Extraction and Purification of natural Ovocalyxin-32:.....	110
3.2.6. Enzyme Activity	111
3.3. RESULTS	116
3.3.1. Extraction, identification and purification of the protein	116
3.3.2. Analysis of the proteolytic capabilities of purified OCX32.....	122
3.3.3. Structural modelling of globular domain of OCX32.....	124
3.4. DISCUSSION.....	134
CONCLUDING REMARKS.....	139
BIBLIOGRAPHY	143

TABLE OF ABBREVIATIONS

TABLE OF ABBREVIATIONS

AAFP	N-(4-Methoxyphenylazoformyl)-L-Phenylalanine
a-Gal	Alpha-Galactosidase
ALDH	Aldehyde Dehydrogenase
ALDH1A1	Aldehyde Dehydrogenase 1A1
ALDH1A2	Aldehyde Dehydrogenase 1A2
ALDH1A3	Aldehyde Dehydrogenase 1A3
ALDH2	Aldehyde Dehydrogenase 2
bCPA1	bovine Carboxypeptidase A1
BSA	Bovine Serum Albumina
DEAB	N,N-diethylaminobenzaldehyde
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
<i>E.Coli</i>	<i>Escherichia coli</i>
ECI	Endogenous Carboxypeptidase Inhibitor
ECM	Extracellular matrix
hCPA1	human carboxypeptidase A1
hCPA2	human Carboxypeptidase A2
hCPA4	human Carboxypeptidase A4
hCPA6	human Carboxypeptidase 6
hCPD	human Carboxypeptidase D
hCPO	human Carboxypeptidase O
hCPZ	human Carboxypeptidase Z
HEK 293F	Human Embryonic Kidney 293F
HMGB3	High mobility group box 3
HRP	Horse-readish peroxidase
hTAFI	human Thrombin-Activatable fribrinolysi inhibitor
LB	Luria-Bertani
LXN	Latexin

MCPs	Metallo-carboxypeptidases
MLCK	Myosin light chain kinase
Mr	molecular mass
NAD ⁺	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide reduced
O/N	Over Night
OCX32	Ovocalyxin-32
PBS	Phosphate-Buffered Saline
pCPB	porcine Carboxypeptidase B
PEG	Polyethylene glycol
PEI	Polyethylenimine
pI	Isoelectrical point
PVDF	polyvinylidene difluoride
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
ST6	Beta-galactoside alpha-2,6-sialyltransferase 1
TRAF1	TRAF-type zinc finger domain containing protein
wt	wild type
XTT	2,3-bis-(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide

INTRODUCTION

INTRODUCTION

THE WORLD OF ENZYMES, INHIBITORS AND REGULATORS

Nowadays, the scientific fields of proteins and all related biomolecules are among the main topics in biology, biotechnology and medicine. Structural and functional variety and versatility of the proteins are at the root of their wide localizations and deep participation in a myriad of distinct biological structures and functions. Within them enzymes stand out because of the vital role they play in most biochemical reactions in organisms. Life depends of their presence and optimal function.

Enzymes are mostly proteins that perform catalytical functions in living organisms (there are exceptions: i.e. a small number of catalytic RNA molecules and chemical molecules). Enzymes act on chemical entities called substrates, with high specificity, converting them to products (Figure 1).

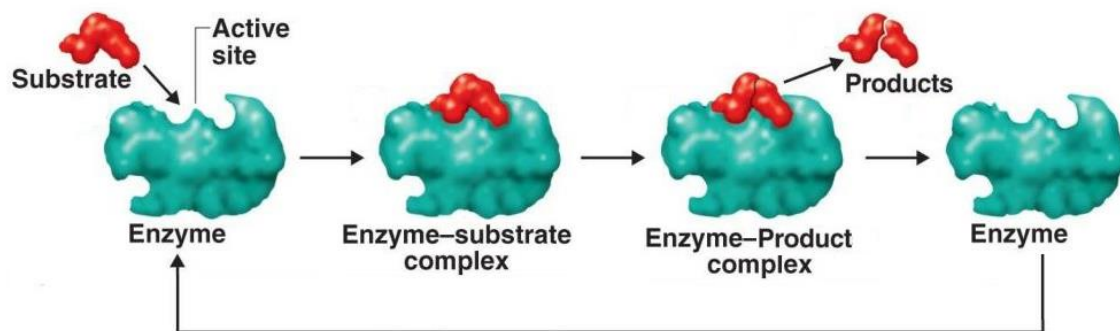


Figure 1. Graphic representation of simplest functional model of an enzyme. An enzyme generally interacts with molecules known as substrates by its active site, and enzyme-substrate complex is formed. After its formation, the complex helps to transform substrate in product, forming an enzyme-product complex. Finally, the enzyme releases the product, and it is ready for another catalytical cycle. *Image reproduced from web page <https://quizlet.com/158784011/principles-of-microbial-metabolism-iii-flash-cards/>.*

Enzymes increase the reaction rate by lowering its activation energy (Figure 2). This action means that enzymes can make the conversion of substrate to product many millions of times faster than the natural spontaneous reaction. Frequently, the presence of an enzyme makes possible a biochemical reaction which could not be feasible in physiological conditions and/or in the required time-scales (Radzicka and Wolfenden 1995).

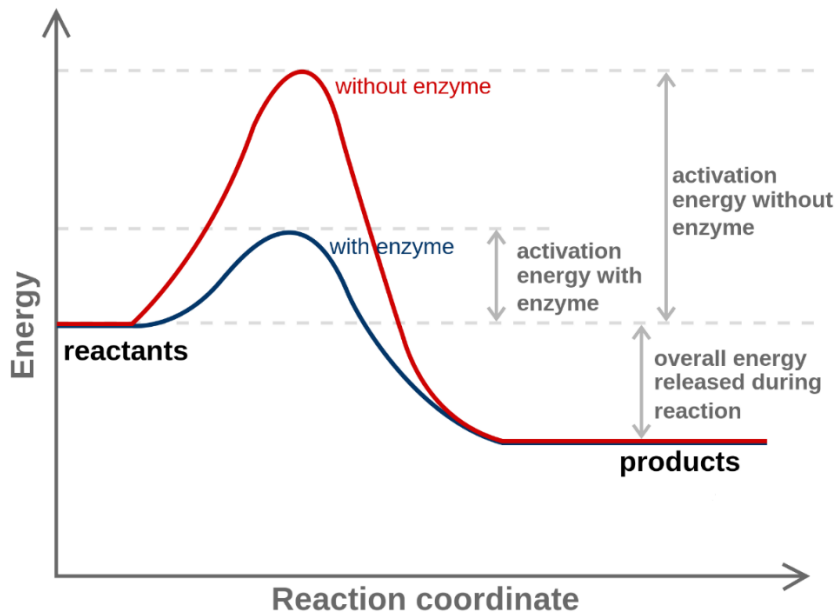


Figure 2. Graph of energy in biochemical reaction. As it is represented, activation energy of reaction decreases in presence of enzyme, this makes easier and faster the conversion from substrate to product. Red line represents activation energy without enzyme and blue line represents activation energy with enzyme. *Image reproduced from rkuhnke.eu, Enzymes.*

Some enzymes require the interaction with complementary molecules or ions to be functional, these elements are called coenzymes, or prosthetic group if the link between the protein part (apoenzyme) and the coenzyme is a tight or even a covalent bond. When apoenzyme and coenzyme interact, enzyme is active, and it is called holoenzyme. In addition, an enzyme can be modified by other processes in order to regulate its enzymatic activity, as phosphorylation, glycosylation or proteolytic cleavage. Moreover, other molecules can regulate enzyme activity, acting as activators or inhibitors. These molecules, mostly inhibitors, are widely distributed in nature, for example: many defence deterrents and natural poisons are enzyme inhibitors, and some of them can be used for therapeutic purposes,

In nature, a huge number of enzymes have been found and described, and generally have been called according to the name of their substrates or the type of reaction they catalyze. Other enzymes are named by the source in which they were found or by the broad function, before the specific activity has been defined. In order to standardize and classify all known enzymes, in 1955 the International Congress of Biochemistry in Brussels set up an Enzyme Commission (EC). Its classification of enzymes is based in which kind of activity the enzyme carries on. The enzyme list and classification have been updated among the last years, the last reorganization has been

INTRODUCTION

done by EC in August 2018 (Table 1) (Tipton 2018). The enzymes treated in this memory, Oxidoreductases and proteases, belong to EC classes 1 and 3.

TABLE 1: International Classification of Enzymes depending their function.

EC Class nº.	Class name	Type of reaction catalysed	Typical reaction
1	Oxidoreductases	Transfer of electrons (hydride ions or H atoms)	$AH_2 + B = A + BH_2$ $AH_2 + B^+ = A + BH + H^+$
2	Transferase	Group transfer reactions	$AX + B = A + BX$
3	Hydrolases	Hydrolysis reactions (transfer of functional groups to water)	$AB + H_2O = AH + BOH$
4	Lyases	Cleavage of C-C, C-O, C-N, or other bonds by elimination, leaving double bonds or rings, or addition of groups to double bonds	$A=B + X-Y = A-B + H_2O$ $\quad \quad \quad \quad $ $\quad \quad \quad X \quad Y$ $A=B + X-Y = A-B + CO_2$ $\quad \quad \quad \quad $ $\quad \quad \quad X \quad Y$
5	Isomerases	Transfer of groups within molecules to yield isomeric forms	$A = B$
6	Ligases	Formation of C-C, C-S, C-O, and C-N bonds by condensation reactions coupled to cleavage of ATP or similar cofactor.	$A + B + XTP = AB + XDP + P$ $A + B + XTP = AB + XMP + PP$
7	Translocases	Catalyse the movement of ions or molecules across membranes or their separation within membranes	

A1. OXIDOREDUCTASES

Oxidoreductases are enzymes that catalyse the transfer of electrons from one molecule, the reductant, also called the electron donor, to another, the oxidant, also called the electron acceptor, generally taking nicotinamide adenine dinucleotide phosphate (NADP⁺) or nicotinamide adenine dinucleotide (NAD⁺) as cofactors. (Price and Stevens 1999). Oxidoreductase enzymes play significant roles in different processes both at the aerobic and anaerobic metabolism, such as oxidative phosphorylation or degradation of retinaldehyde, keeping the redox balance for cellular homeostasis, among others.

Oxidoreductases is a very large family, which includes englobes different kinds of enzymes, but all have in common the electron transport between molecules. So, they can be classified according their functional mechanism of transport:

- Oxidases generally considered as such when molecular oxygen functions as an acceptor of hydrogen or electrons.
- Dehydrogenases, which work by oxidizing a substrate through transferring hydrogen to an acceptor that is either NAD/NADP or a flavin enzyme.
- Peroxidases are placed in peroxisomes and con catalyse the reduction of hydrogen peroxide.
- Hydroxylases are responsible of hydroxyl group aggregation into an organic compound (Huang and Groves 2017).
- Oxygenases are enzymes that oxidize a substrate by transferring the oxygen from molecular oxygen to it.
- Reductases can act like oxidases, but catalysing reductions.

It has long been an important goal in biotechnology to develop practical biocatalytic as applications of oxidoreductases. For instance, research on the construction of bioreactors for pollutants biodegradation and biomass processing, and the development of oxidoreductase-based approaches for synthesis of polymers and functionalized organic substrates achieved great progress. During the past few years, significant breakthroughs have been made in the development of oxidoreductase-based diagnostic tests and improved biosensors, and the design of innovative systems for the regeneration of essential coenzymes.

A1.1. Aldehyde dehydrogenases

Aldehydes are organic compounds widespread in nature. They can be produced endogenously during numerous physiological processes, including biotransformation and oxidative stress, and can behave both as beneficial or as toxic molecules, or generate toxic adducts with other biomolecules (Fritz and Petersen 2013). Thus, they are formed in the biotransformation of amino acids, neurotransmitters, carbohydrates, alcohols, biogenic amines, vitamins and steroids, and through processes such as lipid peroxidation, in which 4-hydroxy-nonenal and malondialdehyde are synthesized (O'Brien et al. 2005). Aldehydes are often generated during the metabolism of a number of drugs and environmental agents. Various aldehydes, including ethanol-derived, acetaldehyde, and the anticancer drugs cyclophosphamide and ifosfamide are important aldehyde precursors (Marchitti et al. 2008). The control of endogenous and exogenous aldehydes (i.e. xenobiotics) by natural aldehyde-transforming enzymes, as aldehyde dehydrogenases, constitute therefore important biological homeostatic and protective metabolic mechanisms.

Aldehyde dehydrogenases (ALDHs) include a cluster of evolutionarily related NAD(P)⁺-dependent enzymes catalysing the oxidation of a wide spectrum of aldehydic substrates, generated from various endogenous and exogenous precursors, to their corresponding carboxylic acids. Aldehydes are detoxified primarily through reductive and oxidative Phase I enzyme-catalysed reactions, including the enzyme systems for aldehyde reduction, alcohol dehydrogenase (ADH), aldo-keto reductase (AKR) and short-chain dehydrogenase/reductase (SDR), and those for aldehyde oxidation, xanthine oxidase (XO), aldehyde oxidase (AOX) and ALDH (Figure 3).

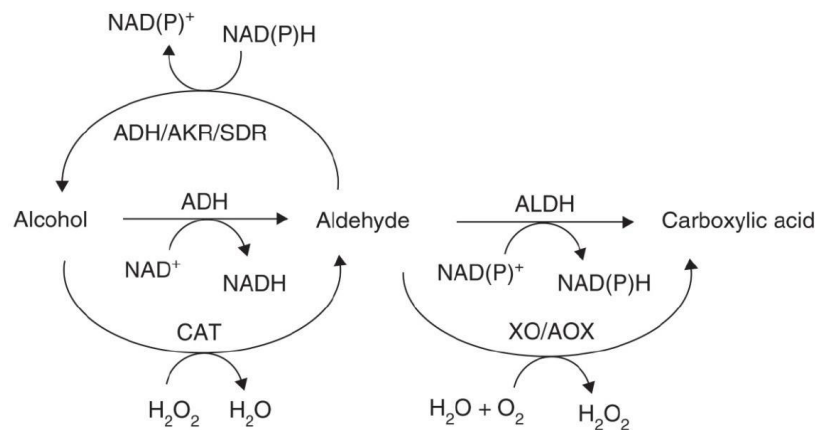


Figure 3. Enzymatic metabolism of aldehydes. ADH: Alcohol dehydrogenase; AKR: Aldo-keto reductase; ALDH: Aldehyde dehydrogenase; AO: Aldehyde oxidase; CAT: Catalase; SDR: Short chain dehydrogenase/reductase; XO: Xanthine oxidase. Reproduced from (Marchitti et al. 2008).

For the human species, there are four well defined members of Aldehyde Dehydrogenases: ALDH1A1, ALDH1A2, ALDH1A3 and ALDH2. Such enzyme variants belong the large ALDH superfamily, which includes tetrameric and dimeric oligomers formed by subunits of 55kDa approximately. At the structural level, the crystallographic structures of distinct ALDHs allows to delineate 3 regions or domains well differentiated:

- On the N-terminal, is the region for cofactor binding domain. On this family the only cofactor which could bind to protein is NAD⁺. This binding is not caused by the canonical NAD⁺ binding moiety, Gly-X-Gly-X-X-Gly residues, (Wierenga and Hol 1983). On ALDH2 (model), the linear sequence that most closely to the previous motif and localized on the region where the cofactor is bound is Gly-Ser-Thr-Glu-Val-Gly.
- Catalytic domain is localized on the middle of 55kDa subunit. On this region, substrate-binding pocket is also localized. The size of the substrate access channel is the determinant for the specificity on ALDH. ALDH1A possesses a wider opening leading to active site allowing the catalysis of big substrates, whereas ALDH2 is more constricted, allowing only the entrance a small aldehydes (Morgan and Hurley 2015).
- At the C-terminal, ALDHs, presents an oligomerization domain, which allows enzymes subdomains being organised on functional dimers and tetramers, as is observed on crystallized proteins.

Most of ALDH enzymes share a high sequence identity in their cofactor-binding site and in the catalytic centre. The principal residues conserved on the catalytic centre are a cysteine,

INTRODUCTION

which is the essential nucleophile, numbered as Cys302 on human ALDH2 numbering (ALDH model), for the reaction. The second conserved residue is a glutamic, numbered Glu268 (ALDH2 numbering), which acts as the proton acceptor from catalytic water to activate de nucleophilic cysteine. And finally, an asparagine, numbered as Asn169, which is the responsible to stabilize the transition state intermediate (Farrés et al. 1995).

At the functional level, most ALDHs are catalytically active proteins. The crystallographic structures of ALDHs have allowed to disclose their catalytic mechanism. The members of the ALDH1A subfamily are tetrameric cytosolic enzymes that catalyse the oxidation of the retinol metabolite, retinaldehyde, to retinoic acid (Wang et al. 1996, Zhao et al. 1996). In contrast, ALDH2 is the primary enzyme involved in the oxidation of acetaldehyde during ethanol metabolism (Klyosov et al. 1996). The kinetic mechanism seems to be similar in all ALDHs. In general, ALDHs exhibit a sequential ordered bi-bi mechanism, in which the oxidized cofactor first binds to the enzyme to allow the subsequent substrate attachment (Figure 4) (Bradbury and Jakoby 1971, Valenzuela-Soto and Muñoz-Clares 1993).

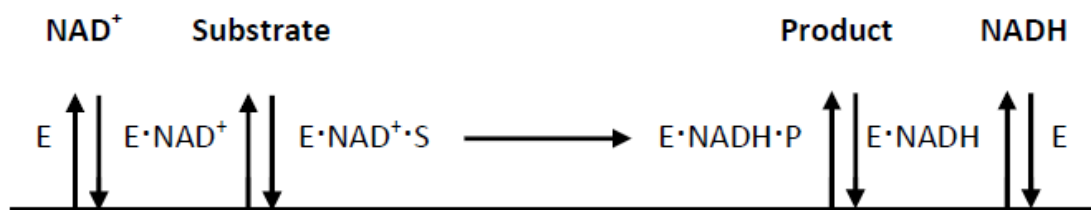


Figure 4. Sequential ordered bi-bi mechanism to the irreversible reaction catalysed by ALDHs. The oxidized cofactor binds first to the enzyme and allows substrate binding. When the oxidation reaction occurs, the product is released prior to the reduced cofactor. Reproduced from (Valenzuela-Soto and Muñoz-Clares 1993).

The variability of residues in the substrate-binding site clearly indicates evolutionary substrate specificities of individual ALDHs. However, farther to the interior of the active site strict conservations are compatible with a common catalytic mechanism (Liu et al. 1997). Site-directed mutagenesis experiments have revealed that catalysis occurs in sequential steps. For mechanism description we used human ALDH2 numbering, because is high studied as can be consider as a model. The mechanism is the following: on the first step the catalytic residue is activated by abstraction of a proton from a water molecule between Cys302 and Glu268. Once the catalytic residue is activated, its thiolate group made a nucleophilic attack on the electrophilic aldehyde (Figure 5.1), As result of previously step, the tetrahedral thiohemiacetal intermediate is formed (diacylation) with concomitant hydride transfer to the pyridine ring of

INTRODUCTION

NAD⁺ (Figure 5.2) After the intermediate is overcome, the catalytic cysteine presents a thioester, which suffers another nucleophilic attack from the remnant OH⁻ from water molecule forming a new intermediate (Figure 5.3) The subsequent step is the hydrolysis of the resulting thioester intermediate setting free the product (Figure 5.4). The final steps are the regeneration of the enzyme (Figure 5.5) and the dissociation of the reduced cofactor and the product (Figure 5.6), allowing the binding of new molecules of NAD⁺ and substrate (Hempel et al. 1999, Koppaka et al. 2012).

On cytoplasmatic and mitochondrial ALDHs from variety of mammalian sources has also been reported esterase activity. This alternative hydrolysis takes place on the same catalytic centre where dehydrogenase activity is produced. This esterase activity remains requiring Cysteine as nucleophilic element, and the Glutamic as the base to activate Cysteine (Farrés et al. 1995, Wang and Weiner 1995).

INTRODUCTION

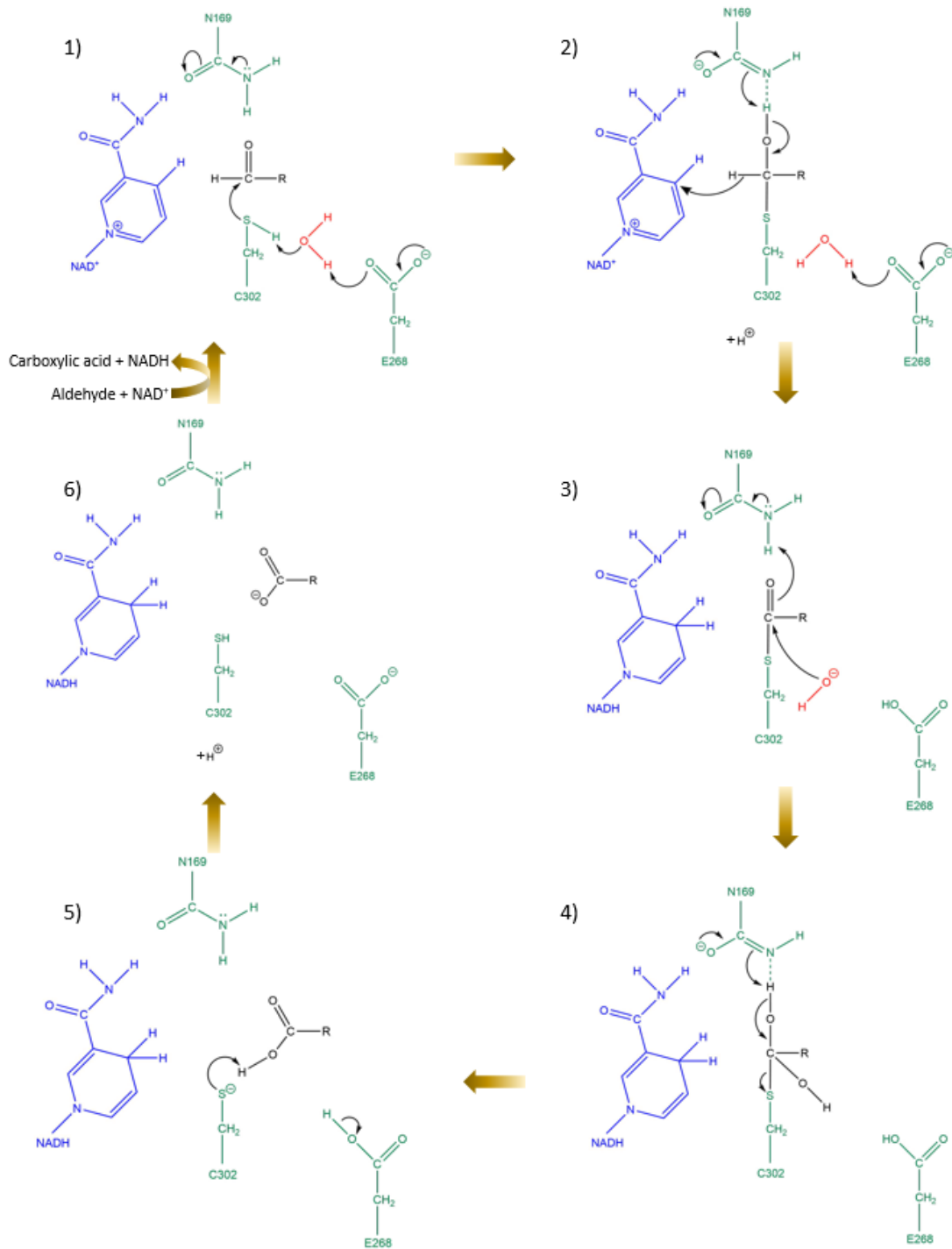


Figure 5. Catalytic mechanism of aldehyde oxidation based on ALDH2 (ALDH model). 1)The thiolate attacks the substrate carbonyl carbon. 2) A transition-state tetrahedral intermediate is formed with a hydride transfer.3) The thioester is poised for base catalysis initiated by abstraction of a proton from a water molecule between Cys302 and Glu268. 4) Now a second tetrahedral intermediate is generated as consequence of water deportation. 5) Collapse of this intermediate yield the product acid which may reprotonate the thiolate. 6) The enzyme can bind reactants again. Reproduced from(Hempel et al. 1999).

A1.1. Members of ALDH family in humans

On humans, ALDHs members, as we described, play a widespread physiological role on all organism. Numerous polymorphisms exist in the human ALDH genes, some of which cause inborn errors of metabolism and contribute to clinically relevant diseases (Vasilidou and Pappa 2000). The human ALDHs family is composed by 19 members, but the most studied are:

ALDH1A1 is ubiquitously distributed in the adult epithelium of various organs (King and Holmes 1997, Zhai et al. 2001). It is one of three highly conserved cytosolic isozymes that catalyse the oxidation of the retinal (retinaldehyde) to retinoic acid (RA) (Wang et al. 1996, Zhao et al. 1996). Furthermore, ALDH1A1 is one of the enzymes involved in the metabolism of the ethanol metabolite, acetaldehyde (Ueshima et al. 1993). The enzyme also plays a key role in the cellular defence against oxidative stress. Like the other ALDHs, ALDH1A1 may also play an important role in cancer therapeutics (Sládek et al. 2002, Moreb et al. 2007).

ALDH1A2 is a cytosolic enzyme expressed in various embryonic and adult tissues (Hsu et al. 2000). ALDH1A2, as ALDH1A1, catalyses the oxidation of retinal to RA (Niederreither et al. 1999). Furthermore, like ALDH1A1, ALDH1A2 also plays role on ethanol metabolism (toxicity defence) (Fatma et al. 2004) and protective role on oxidative stress (Wang et al. 1996). ALDH1A2 has been suggested as a candidate tumour suppressor gene in prostate cancer (Kim et al. 2005).

ALDH1A3 is a cytosolic homotetramer that participates in the synthesis of RA from the retinal and plays an important role in embryonic development (Graham et al. 2006). It participates in the development of the eye (Molotkov et al. 2006), nucleus accumbens and olfactory bulbs (Wagner et al. 2002), hair follicles (Everts et al. 2007), the forebrain (Molotkova et al. 2007) and cerebral cortex (Wagner et al. 2006). The enzyme also plays a role in mitigating oxidative stress (Kim et al. 2005). The deficiency of ALDH1A3 has been involved on different kinds of cancers.

ALDH2 expression is distributed in various tissues including liver, kidney, heart, lung and brain (Goedde and Agarwal 1990). It is the major enzyme involved on the acetaldehyde oxidation during ethanol metabolism (Klyosov et al. 1996). It also acts as a nitrate reductase for the activation of nitroglycerin used to treat angina and heart failure (Li et al. 2006). Mostly of the pathologies related to ALDH2 even its association to cancers, is due ALDH2*2 allele, and the consequent presence of protein on the tetramer, because this is enzymatically inactive (Chao et al. 1993).

The previously described ALDHs are not the only ones involved on human pathologies. Other humans ALDHs involved on pathologies are:

- ALDH1L1 displays a role in cancer by regulating cellular proliferation (Krupenko and Oleinik 2002, Oleinik and Krupenko 2003, Oleinik et al. 2005).
- The reduced expression levels of ALDH3A1 in humans, produce pathologies in corneas (Pei et al. 2006).
- Dysfunction on this ALDH3A2 results in Sjögren-Larsson syndrome (SLS), which is a neurocutaneous disorder characterized by congenital ichthyosis, mental retardation and spastic tetraplegia (Rizzo and Carney 2005).
- ALDH6A1 has been identified as a cardiac protein possibly involved on age-dependent degeneration (Kanski et al. 2005).
- The mutations on ALDH7A1 are the basis for a kind of epilepsy, characterized by the onset of intractable seizures during infancy and early childhood. It could be prevented by a high dose of Vitamin B6, every day (Mills et al. 2006).

A1.1.1. ALDH1A3

On this thesis we focus on human ALDH1A3 which is a cytosolic homotetramer (56kDa every subunit) (Figure 6). This dehydrogenase is integral dedicated to the oxidation of retinal to retinoic acid. In humans, high levels of ALDH1A3 expression has been observed in salivary gland, stomach, breast, kidney and fetal nasal mucosa (Hsu et al. 1994, Zhang et al. 2005). Furthermore, its deregulation on different tissues is a biomarker for cancer phenotype. On MCF-7 breast cancer cell lines, ALDH1A3 expression is downregulated (Rexer et al. 2001). Its gene is upregulated by induction of wild type p53 in cultured human colon cancer cells (Okamura et al. 1999). Additionally, this protein is highly expressed in pancreatic and ovarian cancer, even is not expressed on these tissues on normal conditions (Saw et al. 2012, Jia et al. 2013). On gastric cancer cells, its expression has been silenced by methylation (Yamashita et al. 2006). Furthermore, on human glioblastoma cells, its expression is upregulated by interleukin, which is described as anticancer cytotoxic agent (Han et al. 2005). To sum up, the ALDH1A3 role on the regulation of different kinds of cancer, is clearly evidenced.

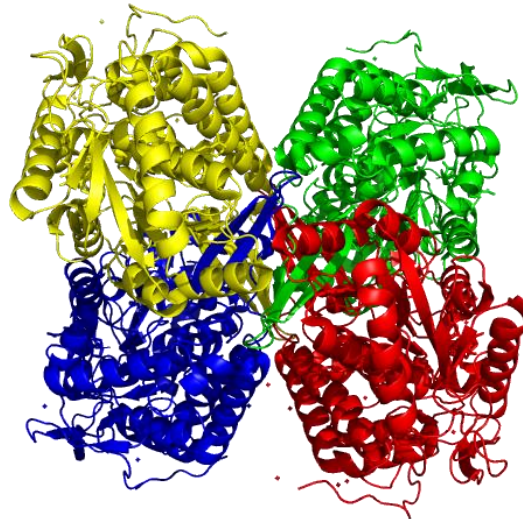


Figure 6. Cartoon representation of the tetramers of ALDH1A3 wt. Every monomer is colored with a different color. Pictures were created using PyMOL program. The ALDH1A3 wt structure was obtained from PDB ID 5FHZ.

This enzyme also displays an important role in embryogenesis, having a critical effect on this biological process (Graham et al. 2006). ALDH1A3 is involved on the central nervous system development and the morphogenesis of anterior head structures, playing an essential role on both biological processes (Suzuki et al. 2000, Molotkova et al. 2007, Kin Ting Kam et al. 2012). In the last years, it has been established a relation between ALDH1A3 and anophthalmia/microphthalmia in humans (Mory et al. 2014).

ALDHs have a wide tissue distribution, with different substrate specificities, and play distinct roles or type of functions. One of the most well-known functions is the conversion of redox metabolites that are essential in the organism, such as retinaldehyde. ALDHs oxidates, in an irreversible way, the retinaldehyde to retinoic acid. (Yoshida et al. 1993, Yoshida et al. 1998, Duester 2000), preventing its transformation in more reduced forms which can be toxic in excess. The particularly critical role of these enzymes in the cellular protection against toxic species, is evidenced by the fact that allelic variants of ALDH genes (leading to perturbations in aldehyde metabolism) are at the molecular basis of several disease states and metabolic anomalies (Vasiliou et al. 2004)

As we described at the beginning of the ALDHs section, they are enzymes widely distributed in distinct tissues, such as liver, stomach, seminal vesicle, brain and lung, localizing at cytosol and mitochondria. Several members of the ALDHs subfamily display important functions on the organism (Marchitti et al. 2008). Recently, they have been described as therapeutic target for

INTRODUCTION

the treatment of different types of cancer, particularly the members of the A1 subfamily (Pequerul et al. 2020), including through chemoresistance (Clark and Palle 2016). Furthermore, It has been observed that the use of ALDH inhibitors on cancer cells generates cell death, breaking new ground for the future of the cancer treatment (Jiménez et al. 2019). For this reason, we consider the study of ALDH as an important area of research.

In this thesis we have particularly focused on ALDH1A3, and more specifically on a mutant of this enzyme with kinetic and structural interest for our study and for the field.

B1. PROTEASES

Proteases, or proteolytic enzymes, a term equivalent but also somewhat distinct than peptidases, are enzymes which belong to the family of hydrolases that catalyse the hydrolysis of peptide bonds in proteins and peptides. For this reason, proteases are classified into subclass 3.4, according to the EC classification, within the 3.0 class of hydrolases. These enzymes account more than 2% of the genes in humans and other mammals (López-Otín and Matrisian 2007), and are also found in many other organism, including those involved in microbial, viral and other parasitic infections. That is, proteases constitute one of the largest functional groups of proteins (Barrett 1994) which are ubiquitous throughout all kingdoms of living organisms (Hedstrom 2002, Barrett et al. 2012).

Proteases are involved, directly or indirectly, in most of the biochemical and physiological processes in living organisms: various vitals processes for cell life (growth, nutrition, migration and invasion), fertilization, differentiation, apoptosis and dead, among others. The effects of such previous processes are present from the molecular and cellular level to organ and organism level to produce cascade systems, such as inflammation and many other complex processes at physiological and pathophysiological level (Abbernante and Fairlie 2005). Consequently, proteases are of great interest for agroalimentary, sanitary and pharmaceutical industries, i.e. as potential drugs targets.

Proteases can be classified according to different criteria:

- **Type of reaction and cleavage sites:** Some proteases cleave substrates in N- and C-terminal (both are exopeptidases), and other proteases cleaves in the middle of substrate (endopeptidases). According to this, proteases can be aminopeptidases, carboxypeptidases or endopeptidases. Further than these kinds of proteases also exist omega peptidases, which are exopeptidases that catalyse the hydrolysis of specific modified residues with different post-translational modifications.
- **Structural relationships:** This type of classification is based on sequential and structural features of the proteases. All these ones, here termed under the name of peptidases, are classified as described in the MEROPS database (Rawlings et al. 2018). This system classifies proteases/peptidases in families whose members have sequence homology and a common ancestor. The classification is hierarchical: sequences are assembled into families, and families are assembled into clans, the latter including members with related three-dimensional structure. Each peptidase, family and clan have a unique

identifier composed by a letter and numbers. The complete list of families can be found at the web version of the MEROPS database (<https://www.ebi.ac.uk/merops/>).

- **Functional group present in active site or catalytic mechanism:** Due to evolution, proteases have adapted to a wide range of conditions, and they use different catalytic mechanisms for hydrolysis. According to the functional essential residue present in the active site, proteases belong to different catalytic classes, as metallo-, serine-, cysteine-, or aspartic-proteases. Recently, this has been expanded to threonine-, asparagine-, and glutamic proteases, as shown in the MEROPS database, although the latter are still minority classes. Some of the most important features of the such main classes are summarised below.
 - The Metalloproteases have a metal ion, frequently zinc, coordinated by a binding motif in its active site; this metal ion use to act as a nucleophilic reagent for catalysis. Metalloproteases are not homogenous; their zinc binding motif differs between proteases and they can be classified into different families according to their zinc binding motif. The present thesis is focused in one of the largest family of metalloproteases, the M14, whose zinc binding motif is HXXE...H. The zinc binding motif is formed by two histidines and a glutamic acid residue. Other metalloproteases families differs in the number of histidines, presence of glutamic acid or aspartic acid, or even in the number of metal ions coordinated in the active site. (Vendrell and Avilés 1999) Is a class with a great structural and functional variety, and alternative representative members are aminopeptidases N/I/P, thermolysin, collagenases, MMPs, and ADAMs, as well as many peptidases from parasites, among others.
 - The Serine Proteases share the same catalytic motif in all their members. Such catalytic motif is composed by a triad of residues, serine/histidine/aspartic; in this sequence serine serves as nucleophilic residue at active site. Among the most representative members are trypsin, chymotrypsin and elastase-like variants, which play crucial roles in protein digestion and turnover, homeostasis, reproduction, immune response, and even in signal transductions (Wang et al. 2008)
 - The Cysteine proteases contain a very reactive cysteine residue at the active site, which acts as the nucleophile. These proteases share a catalytic motif, which is formed by cysteine/histidine/aspartic triad, although in some members a cysteine/histidine dyad is enough. Is also a class with a great internal variety, and some representative members are papain and pineapple proteases, cathepsins,

caspases and calpain, among others One particular example is cathepsin X, involved in cell signaling, lysosomal degradation of peptides and tumor progression (Kos et al. 2015).

- Aspartic proteases present one or two reactive aspartate residues in the active site for catalysis of their substrates. It is the class less populated among the previously mentioned. Pepsin is among the most well-known examples of aspartic protease, as one of the main proteases of mammalian digestive system (Fujinaga et al. 1995). Another well-known and interesting member is HIV protease, devoted to the processing of its viral proteins, initially translated as a long joint precursor requiring multiple cleavages to generate the mature proteins.

B1.1. Metalloproteases

Metalloproteases (MCPs) are a large and diverse groups of peptidases that catalyse the hydrolysis of peptide bonds at the C-terminal of peptides and proteins, and a metal ion are involved in its catalytic mechanism (Barrett et al. 2012, Rawlings et al. 2018). Frequently, their action, even if removing a single residue, causes strong effects in the biological activity of their peptide and protein substrates (Skidgel 1996). In this thesis we focused in one specific family of MCPs, the M14 family.

Metalloproteases from the family M14, have been divided into four subfamilies based on structural similarity and sequence homology: M14A (also termed as the A/B subfamily), M14B (the N/E subfamily), M14C, and the M14D subfamily, also known as cytosolic metalloproteases (CCPs) (Rodriguez de la Vega et al. 2007, Rawlings et al. 2018). MCPs could also be classified based on their substrate specificity (Lyons and Fricker 2011):

- **A-type:** This kind of MCPs has preference for C-terminal residues having aromatic or branched aliphatic side chains, and they can also be subdivided into two subtypes: A1 or A2. The A1 enzymes show cleavage preference for small aliphatic and aromatic residues, whilst the A2 enzymes prefer large aromatics residues (Gardell et al. 1988, Tanco et al. 2013).
- **B-type:** its preference is for basic residues at the C-terminal of substrates.
- **O-type:** this a rather novel variant, with preference for acidic residues at the C-terminal of substrates (Lyons and Fricker 2011).

M14A subfamily: Members of this subfamily are secreted as soluble proteins, which usually are synthesized as inactive precursors with a preceding signal peptide of 15-22 residues (Figure

INTRODUCTION

10). The pro-enzymes of M14A subfamily are known as procarboxypeptidases (PCPs) and contain a 90-95 residues segment at their N-terminal, which is called pro-region. Pro-region folds into a globular independent unit, suggesting that might act as chaperone *in vivo* (Phillips and Rutter 1996, Vendrell et al. 2000), and it is linked to the enzyme through a connecting part (Coll et al. 1991, Guasch et al. 1992). The structure of this pro-domain is similar on all the members of M14A. Furthermore, these carboxypeptidases contain a catalytic domain of approximately 300 amino acid residues. The interaction between the catalytic domain and pro-region blocks the access to the active site, acting as inhibitor (Garcia-Castellanos et al. 2005). To be active, such zymogens must be processed and set free of pro-region (Figure 7).

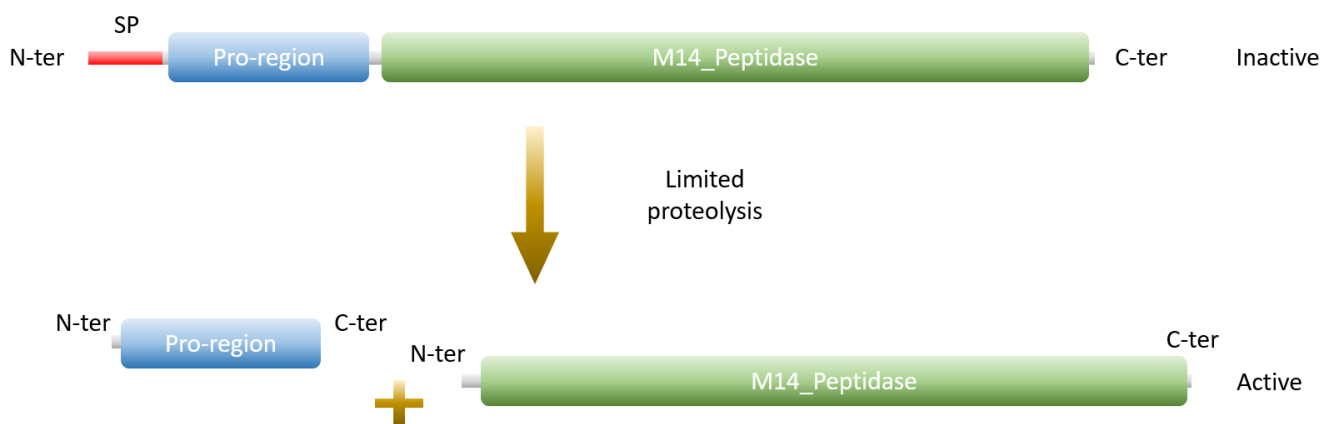


Figure 7. Schematic representation of M14A peptidase domain distribution and its activation mechanism on mammalian. PCPs, from M14A peptidase subfamily, have a pro-region or “activation segment” next to catalytic domain. Pro-region blocks active site and enzyme remains inactive. After PCP is secreted and reaches functional localization, “activation segment” is removed by limited proteolysis to obtain fully active enzyme. The N-terminal (N-ter), C-terminal (C-ter) and signal peptide (SP) are indicated. Reproduced from (Garcia-Pardo 2015).

The members of M14A subfamily are:

- **Carboxypeptidases A1 (CPA1), A2 (CPA2), and B (CPB)** are synthesized as inactive zymogens (PCPA1, PCPA2, and PCPB), mainly in the pancreatic acinar cells until they are secreted into the digestive tract and their pro-region is cleaved by limited tryptic digestion in the duodenum (Arolas et al. 2007). The difference between these enzymes is in their specificity of substrate: CPA1 displays cleavage specificity for aliphatic and smaller aromatic residues, meanwhile CPA2 shows preference for larger aromatic residues (as Trp). Instead CPB cleaves basic residues (Gardell et al. 1988, Pascual et al. 1989, Arolas et al. 2007). The difference in specificity between A-type and B-type proteases is caused by residue in position 255, a hydrophobic residue in A1 or A2 (Ile, Val, Leu or Met), whilst in

INTRODUCTION

B-type is an acidic residue (Asp) (Vendrell et al. 2004). Furthermore, CPA2 isoforms have also been reported in rat extra-pancreatic tissues such as brain, testis, and lung; these CPA2 isoforms are shorter and have a distinct role from the pancreatic isoform (Normant et al. 1995, Reverter et al. 1997).

- **Carboxypeptidase A3 (CPA3)**, also known as mast cell CPA, is found in the secretory granules of mast cells mainly in its active form, in complex with proteoglycans (Springman et al. 1995). The precise role of this carboxypeptidase still needs to be defined, although evidences have been collected favouring a role in the regulation of the local inflammatory response and/or protection against exotoxins (Springman et al. 1995, Lundequist et al. 2004, Metz et al. 2006, Wernersson and Pejler 2014).
- **Carboxypeptidase A4 (CPA4)** is widely expressed in tissues and its expression is upregulated in some types of carcinomas. A study of substrate specificity showed that this enzyme was able to hydrolyse hydrophobic C-terminal residues as Phe, Leu, Ile, Met Tyr and Val (Tanco et al. 2010). Some peptides such as neurotesin, granin and opioid peptides, described as involved in cell proliferation and differentiation, are transformed by the CPA4. It has been proposed that hCPA4 could participate in its inactivation, influencing cell proliferation and aggressiveness of prostate cancer (Tanco et al. 2010).
- **Carboxypeptidase A5 (CPA5)** has been suggested that plays a role in the processing of pro-opiomelanocortin-derived peptide, and possibly other peptides that undergo C-terminal removal of aliphatic or aromatic residues, which are specifically hydrolyzed (Wei et al. 2002, Wei et al. 2003).
- **Carboxypeptidase A6 (CPA6)**, in contrast to other M14 MCPs, is activated in the secretory pathway and, after its secretion, binds to extracellular matrix where retains its enzymatic activity. Substrate specificity studies indicate its preference for large hydrophobic C-terminal amino acids and only very weak activity toward small amino acids and histidine (Lyons and Fricker 2010). Mutations in CPA6 gene has been linked to Duane syndrome, a congenital disorder of the visual system (Pizzuti et al. 2002), as well as to epilepsy (Lyons and Fricker 2010, Sapio and Fricker 2014).
- **Carboxypeptidase O (CPO)** is membrane protein attached via a phosphatidylinositol glycan (GPI) anchor and is found on the apical surface of intestinal epithelial cells. It is the only member of this subfamily which does not present the inhibitory pro-region domain in the N-terminal, so it is synthesized as active proteases. This enzyme is able to cleave proteins and synthetic peptides with toward acidic C-terminal amino acids (O-type proteases).

- **Carboxypeptidase U (CPU)**, commonly named as TAFI (Thrombin-Activable Fibrinolysis Inhibitor), is one of the most cited M14 MCPs, which cleaves the basic amino acids Lys and Arg from the C terminal of a variety of peptides and proteins (B-type) (Skidgel 1996). Activated TAFI plays a key role in the regulation of fibrinolysis (Bouma and Mosnier 2004, Hendriks 2004, Marx 2004, Willemse and Hendriks 2007, Rijken and Lijnen 2009). It removes the C-terminal lysine residues from the fibrin clots, thus preventing the proteolytic action of plasmin on such clots and, therefore, fibrinolysis. This enzyme has become an important target for thrombotic diseases (Sanglas et al. 2008).

M14B subfamily: Proteases of M14B subfamily are produced as active enzymes, and its specific action relies on subcellular localization in order to prevent inappropriate cleavages that would otherwise damage the cell. Instead of a pro-segment, these proteins contain a transthyretin-like (TTL) domain (80-90 residues) at the C-terminal of the catalytic domain, of an unknown role. Due to its similarity with pro-domains found in the M14B members, it has been proposed to function as a folding domain or, alternatively, as a binding element necessary for oligomerization and/or interaction with other proteins or membranes (Reznik and Fricker 2001). The interface between the TTL domain and the catalytic domain is stabilised mainly through hydrophobic interactions. Members of this subfamily contain other domains and even repeats of the carboxypeptidase domain (Vendrell et al. 2004, Rodriguez de la Vega et al. 2007). Overall this subfamily of MCPs is structurally more heterogeneous, in comparison with the M14A subfamily (Figure 8).

The M14B subfamily is composed by eight members in mammals: five members are catalytically active (CPN, CPM, CPE, CPZ and CPD) and other 3 apparently non-active (CPX1, CPX2 and AEBP1) (Figure 8).

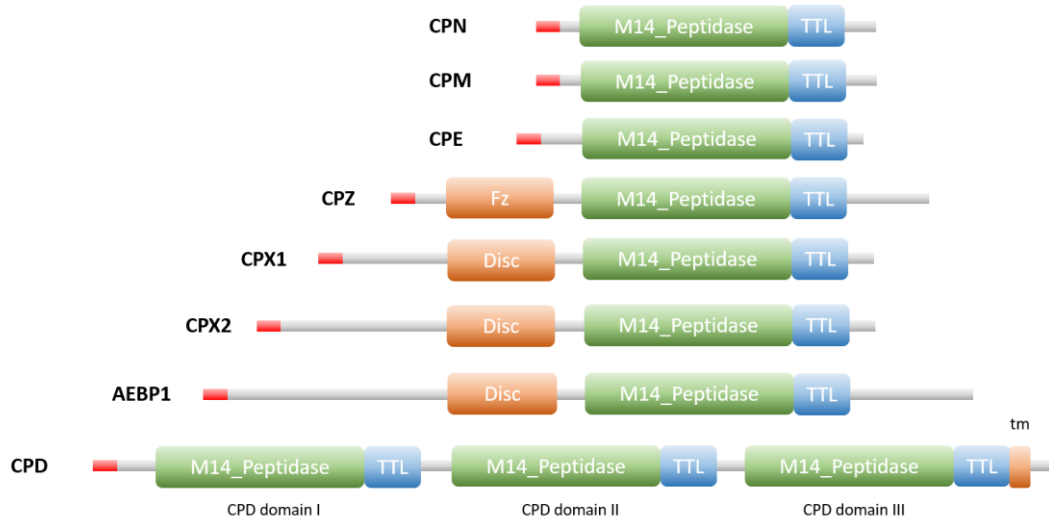


Figure 8: Schematic representation of 8 members from M14B subfamily in mammals. All members contain a signal peptide located in N-ter (represented as red fragment), a M14 carboxypeptidase catalytic domain (M14_Peptidase, in green) and a C-terminal domain homologous to transthyretin (TTL, in blue). In addition, some members have extra N-terminal domain (for example, CPZ has a frizzled-like domain (Fz) or CPX1, CPX2 and AEBP1 show and N-terminal domain with high similarity to discoidin-1). As it is shown, CPD domain structure differs from the rest of M14B subfamily members, it has three tandem repeats of M14 peptidase plus a TTL domain (labelled in the figure as CPD domain), followed by a short transmembrane domain (tm, in orange) and a cytoplasmic tail.

- **Carboxypeptidase E (CPE)** is one of the most active members of M14B MCPs, particularly regarding the biosynthesis of neuropeptides and peptide hormones, and as such is abundant in brain and endocrine organs. CPE removes C-terminal Lys and Arg residues from intermediates formed by the action of prohormone convertases on peptide precursors (Fricker 2018). The CPE monomer has a 53kDa size, with pH 5.6 as the optimum for its proteolytic activity, coinciding with the one inside the secretory granules. CPE has been proposed to have key functions in the metabolism and trafficking of peptide hormones and neurotransmitters. (Cool et al. 1997).
- **Carboxypeptidase D (CPD)** is a 180-kDa protein that contains three carboxypeptidase-like domains, a transmembrane domain, and a cytosolic tail (figure 5). CPD has the broadest tissue distribution of all MCPs, being present in all tissues. Of the three carboxypeptidase-like domains, only the first two have enzymatic activity toward standard substrates (Novikova et al. 1999). It has described as an important member in the processing of proteins that transit the secretory pathway (Sidelyeva and Fricker 2002). Putative substrates include neuropeptides, growth factors and receptors that are produced from larger precursors by cleavage at basic amino acid-containing sites (Fricker 2013), somewhat complementing the very active CPE in such tasks (Fricker 2018).

INTRODUCTION

- **Carboxypeptidase N (CPN)** is a tetrameric complex secreted from the liver into blood. The tetramer is composed by two dimers each formed by a carboxypeptidase domain of 55kDa and a regulatory domain of 83 kDa. CPN removes basic residues from the C-terminal of various blood proteins/peptides with regulatory activity, i.e. from bradykinin, which alters the affinity of this peptide for the two distinct bradykinin receptors. In addition, CPN has been proposed to cleave C-terminal basic amino acids from anaphylatoxin C3a, C4a, and C5a (Skidgel and Erdos 2004).
- **Carboxypeptidase M (CPM)** is a glycosylated monomer of 62 kDa consisting of a single amino acid chain, which is mainly bound on the cell surface through a GPI anchor to the plasma cell membrane (Deddish et al. 1990, Shimamori et al. 1990, Skidgel 1996, Tan et al. 2003). In such form is distributed in various tissues, with highest levels in lung and placenta. Enzymatic release of CPM from the membrane upregulates its synthesis, maintaining constant levels of CPM on the cell surface (McGwire et al. 1999). It is also present in soluble form in distinct body fluids. It cleaves Arg- or Lys- C-terminal residues from a variety of peptides, and as such it is expected its participation in the extracellular regulation of peptide hormone activities and processing, as in kinins, enkephalins and opioid peptides (Li and Skidgel 1999).
- **Carboxypeptidase Z (CPZ)** was discovered during a search for novel carboxypeptidases that could compensate for defective CPE in Cpe fat /Cpe fat mice (Song and Fricker 1997, Fricker and Leiter 1999). Its specificity and size is similar to CPE, but with lower levels of activity and more restricted localizations (Garcia-Pardo et al. 2020). CPZ contains a cysteine-rich domain at the N-terminus that has amino acid sequence similarity to Wnt-binding proteins (Novikova et al. 2000), and its is expected to be connected to the Wnt pathways. Its exact function is unknown, but it can play a role in development, cleaving residues from neuropeptides and growth factors, or/and through its interaction with Wnt proteins (Reznik and Fricker 2001).

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 and Fricker 2001, Arolas et al. 2007).

M14C subfamily: The M14C subfamily comprises the bacterial orthologs, which includes - D-glutamyl-(L)-meso-diaminopimelate peptidase I from *Bacillus sphaericus* (Hourdou et al. 1993). Relevant information about them can be found at the MEROPS database (Rawlings et al. 2018).

M14D subfamily: This subfamily was recently described by our research group (Kalinina et al. 2007, Rodriguez de la Vega et al. 2007), and their members are also known as cytosolic carboxypeptidases (CCPs), indicating its cytosolic or nuclear localization, a difference with the other M14 subfamilies. They tend to be of diverse sizes and larger than most MCPs (Fig. 6), with restricted specificity for C-terminal acidic residues and, besides a canonical carboxypeptidase domain, they contain others of expecting interacting nature. In humans, M14D subfamily is composed by six members, from CCP1 to CCP6. These enzymes notably act as glutaminases on tubulins at the microtubules. Also, they show activity on others proteins rich in acidic residues at the C-terminus, like telokin, MLCK, TRAD1 and HMGB3 (Tanco et al. 2015)(Figure 9).

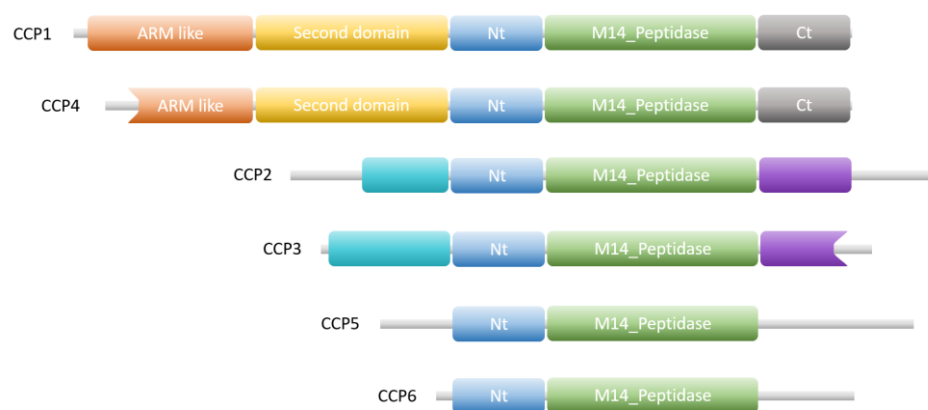


Figure 9: Schematic representation of the domain structures of M14D subfamily members in humans. All members of the M14D subfamily share the M14 peptidase catalytic domain (M14_Peptidase, in green) and an N-terminal domain (N-t, in blue). In addition, every member of the family has other domains in the N-terminal and/or C-terminal region.

B1.2. Substrate specificity and catalytic mechanism of MCPs

For the interaction between the protease and substrate, a conceptual model of how this interaction happens was proposed (Schechter and Berger 1967). Accordingly, the protease has several specificity pockets in the recognition centre, which are able to accommodate the side chain of residues from substrate, generally one residue by each counterpart. Following this premise, the structure of the recognition centre determines the substrate that is able to accommodate and be cleaved by the protease. A graphic representation of the model described is following (Figure 10)

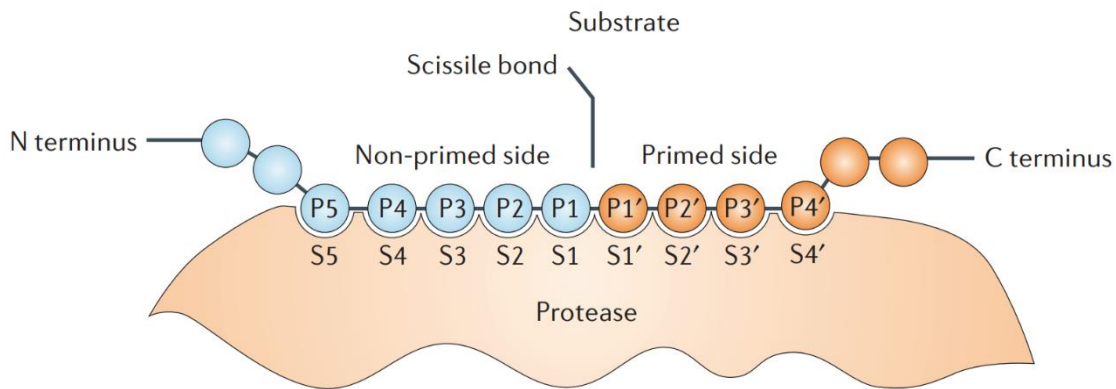


Figure 10. Schematic representation of the model from Schechter and Berger. The residues from the substrate are labelled as P plus a number (from P5 to P4'). The specificity pockets from a protease are labelled as S plus a number (from S5 to S4'). The initial (primary) cleavage site from a protease is located between position S1 and S1', and the peptide bond which will be cleaved is described as Scissile bond. Reproduced from (Turk 2006).

In order to describe the catalytic mechanism for M14 MCPs, the enzyme of reference is bovine CPA1 (bCPA1), the canonical model, which is one of the most studied enzymes in Biochemistry. Most crystallographic and kinetic data currently support the promoted-water model, where the zinc ion from CPA1 acts as electrophilic catalyst providing an electrostatic stabilization for the negatively charged intermediates generated during the hydrolysis. The water molecule act as a potent nucleophile to attack the scissile peptide bond of the substrate (Fife and Przystas 1986, Álvarez-Santos et al. 1998).

The nucleophilicity of the metal-bound water is enhanced by a hydrogen bond with Glu270 which acts as the general acid-base catalyst. While Glu270 abstracts a proton from the water molecule, Arg127, which polarizes the peptide bond prior to hydrolysis, also stabilizes the tetrahedral transition state (Figure 11A, 11B) (Christianson and Lipscomb 1989, Phillips et al. 1990). Subsequently the tetrahedral intermediate collapses into products due to protonation of amide nitrogen by the side chain of Glu270, and MCP restores its catalytic machinery (Figure 11C). The zinc binding residues are His69, Glu72 and His196, with the catalytic water molecule also affording a coordination bond. From crystallographic studies, amino acids Asn144, Arg145, Tyr248 in S1' and Arg127 and Glu270 in S1 have been defined as important for substrate binding and catalysis and are shared by all enzymatically active members of the MCPs family. The terminal carboxylate group of the peptide substrate is fixed by Asn144, Arg145 and Tyr248, while the carbonyl group of the scissile peptide bond becomes positioned near Glu270, Arg127 and zinc. Differences are restricted to the positions that define subsites S2, S3 and S4 and to amino acids defining specificity for a C-terminal sidechain. The amino acid residue at position

INTRODUCTION

The enzymatically active members of the M14B subfamily do not use the residue in the position equivalent to residue 255 of bCPA1 to provide the specificity for basic C-terminal residues (Gomis-Rüth et al. 1999). This residue is either a Gln (CPE, CPD domains 1 and 2, CPM, CPN) or Ser (CPZ). Based on the crystal structure, the role of binding to the positive charge of the substrate's side chain in CPD domain 2 is carried out by Asp192, at a position equivalent to Gly207 in CPA (Gomis-Rüth et al. 1999, Aloy et al. 2001).

MCPs represent potential drug targets due to their diversity of structures and functions, involvement in allergic and inflammatory responses, cardiovascular diseases, cancer, in neuropeptides/hormones processing, as well as in the case of CCPs, relationship with pathologies derived from microtubules processing. (Wei et al. 2003, Handa et al. 2019, van der Laan et al. 2019, Zhang et al. 2019). Any dysfunction on many MCPs could generate a pathology, for this reason it is necessary to know how carboxypeptidase activities are regulated in nature, in order to devise related therapeutic or biotechnological approaches. Thus, the activity of MCPs can be affected and regulated by environmental conditions: pH, temperature, presence of ions, salts and effectors. Compartmentalization and synthesis as zymogens are two more ways of regulating their activity and have to be taken into account. However, one of the most adaptive and applicative method, is the use of specific and high affinity regulatory molecules, generally inhibitors. This kind of molecules binds specifically to MCPs to turn down their catalytic action and can be used to specifically mitigate or chemically knock-out the activity of diverse MCPs.

B2. PROTEASE INHIBITORS

Natural protease or peptidase inhibitors are usually proteic polypeptides, but there are also organic molecules which could act as inhibitor too, both able to inhibit or stop the action of proteases. Proteinaceous protease inhibitors are generally genome-encoded and are widely distributed in various tissues of animals, plants, and microorganisms (Laskowski and Kato 1980, Covalada et al. 2012, Fernández et al. 2013), a natural set greatly populated by the appearance of powerful genomic-proteomic screening approaches. Regarding proteases, we shall preferentially deal with this class in the present memory. Besides, a considerable set of small synthetic inhibitors of proteases are increasingly available, either of organic or peptide nature.

Generally, protease inhibitors have features resembling substrates, being specially powerful when they behave as transition state analogues and act as competitive molecules. In the case of protein protease inhibitors, they frequently have a good class specificity, but sometimes they are useful when such specificity is broader. These broader inhibitors are able of inhibit proteases belonging to different mechanistic classes, but not all with the same efficiency (Rawlings et al. 2004). The presently most accepted classification of protease inhibitors consider them into 4 major classes: inhibitors of serine-, metallo-, cysteine- and aspartic proteases, ordered from the most to the less varied and populated (Hartley 1960). Besides, it could also be considered, but as very minor representatives, the inhibitors of threonine-, asparagine- and glutamic-proteases.

Protease inhibitors play important basic roles, such as to prevent proteolysis in sites where this activity must not occur and regulate limited proteolysis (Bode and Huber 2000). Even some inhibitors described could play an important role in defence against protease from predator (Mebis and Gebauer 1980, Shiomi et al. 1985, Sencic and Macek 1990, Castañeda et al. 1995, Covalada et al. 2012).

These functions generate an interest to study protease inhibitors, not only as valuable molecules to establish structure-function relationship, but also to their potential role as a biotechnological and biomedical tools.

B2.1. Metalloprotease inhibitors

Unlike endoproteases for which numerous examples of proteinaceous inhibitors have been reported (Rawlings et al. 2018), MCPs inhibitors are limited in number and in functional variety. Thus, up to date, no proteinaceous inhibitors for the N/E and CCPs variants of MCPs, that is for M14B and M14C subfamilies, have been found in nature (Arolas et al. 2007). Besides, most of the described MCPs inhibitors are from exogenous sources and addressed to exogenous needs, with the exception of the reported mammalian endogenous carboxypeptidase inhibitor (ECI or Latexin).

The MCPs inhibitors well described and characterized to the date are:

- **Potato carboxypeptidase inhibitor (PCI):** this polypeptide, having potato tuber as its richer source and consisting of 39 residues (molecular mass of 4.2 kDa), is the most extensively studied proteinaceous MCP inhibitor. It is a tight-binding, substrate-like, competitive and low nanomolar (K_i) inhibitor against A/B-type MCPs (Hass et al. 1975, Arolas et al. 2005). PCI is organized in a 27-residue globular core, with N- and C-terminal tails of 7 and 5 residues, respectively. The core is stabilized by three disulphide bonds, which confer high stability to the protein. Its secondary structure is simply formed by a short helix and a very small antiparallel β -sheet (Figure 12. A). PCI, as its name described, was initially isolated from potato (*Solanum tuberosum*) (Ryan et al. 1974) but nowadays is produced recombinantly at high yield (Marino-Buslje et al. 1994). Tomatoe (*Solanum lycopersicum*) contains a highly homologous inhibitor, TCI, with 70% sequence identity and strong immunological cross reactivity (Blanco-Aparicio et al. 1998). The function of potato and tomato inhibitors is probably related to plant defence mechanisms (Ryan 1989).
- **Carboxypeptidase inhibitor from *Ascaris suum* (ACI):** This inhibitor consists of 67 residues (molecular mass of 7.5 kDa) and did not show significant homology with any known protein, except at C-terminus. The crystal structure of ACI, in complex with hCPA1, showed a protein with a fold consisting of two tandem homologous domains, each containing a β -ribbon and two disulphide bonds (Figure 12. B). The inhibitor was isolated from pig roundworms (*Ascaris suum*) (Sanglas et al. 2009).
- **Tick carboxypeptidase inhibitor (TCI):** TCI is formed by 75 amino acid residues (molecular mass of 7.9 kDa). Its three-dimensional structure, derived either associated to CPA or to CPB, shows two different domains, one presents homology with PCI and is linked by small α -helix to a second domain which shows high structural homology to proteins of the β -defensin-fold family (Figure 12. C) (Arolas et al. 2006). The protein was isolated from the ixodid tick (*Rhipicephalus bursa*) (Arolas et al. 2005).

INTRODUCTION

- ***Sabellastarte magnifica* inhibitor of carboxypeptidases A and serine proteases (SmCI):** this is a 165 residues glycoprotein (molecular mass of 19.7 kDa) with three BPTI Kunitz domains. The X-ray crystal structures of the 3 domains are typical for members of this family (Figure 12. D). This molecule is able to inhibit several A-type MCPs as bCPA1, through its "N-terminus" in a non-canonical unexpected way, and also displays strong inhibitory activity against trypsin, chymotrypsin and pancreatic elastase (Alonso-del-Rivero et al. 2012). The inhibitor was found and isolated from fan worm (*Sabellastarte magnifica*).
- **Leech carboxypeptidase inhibitor (LCI):** LCI is composed by 66 amino acid residues. It does not show sequence similarity to any other protein, except at its C-terminal tail, by convergent evolution to the other MCPs inhibitors. According to its derived 3D structure, this region shares structural homology with the other inhibitors, suggesting a similar mechanism of inhibition (Reverter et al. 1998). In addition, the structure of LCI defines a new protein motif of five-stranded antiparallel β -sheet and one short α -helix (Figure 12. E). This inhibitor was isolated from leech (*Hirudo medicinalis*).
- ***Nerita versicolor* Carboxypeptidase Inhibitor (NvCI):** this is a monomeric protein of 53 residues (molecular mass of 5.9 kDa) which is described as the most efficient inhibitor of MCPs (Covalada et al. 2012), with picomolar K_s against MCPs. Its derived three-dimensional crystal structure, in complex with hCPA4, presents a simple globular fold stabilized by three disulphide bonds. In addition, its secondary structure presents two antiparallel β -sheets connected by 3 loops with amino- and carboxy-terminal tails (Figure 12. F). It is the first inhibitor of MCPs isolated and characterized in depth from marine organism, in this case from the marine snail (*Nerita versicolor*) (Covalada et al. 2012).

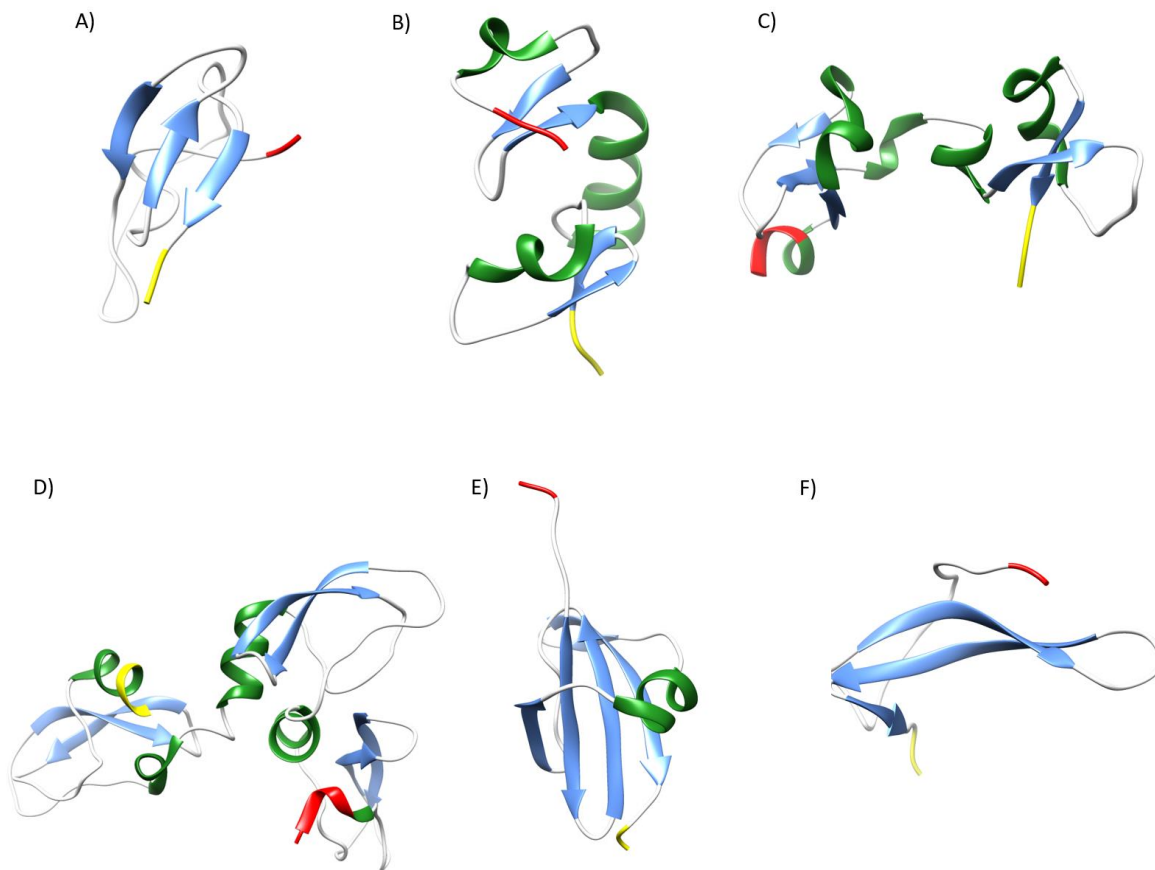


Figure 12. Three-dimensional structure of exogenous MCPs inhibitors. A) PCI. PDB code: 4SGB B) ACI. PDB code: 3FJU C) TCI. PDB code: 1ZLH D) SmCI. PDB code: 4BD9 E) LCI. PDB code: 1DTV and F) NvCI. PDB code: 4A94. Code color: red fragments depict N-terminals, yellow fragments depict C-terminals, the reactive site region. Painted in green are α -helix elements, in blue β -sheets elements and in white the unstructured regions and loops.

- **Endogenous carboxypeptidase inhibitor or Latexin (ECI or LXN):** this inhibitor, also known as tissue carboxypeptidase inhibitor or latexin (LXN) is the largest proteinaceous M14 carboxypeptidase inhibitor reported. It is a protein of 222 residues (25 kDa) whose sequence does not present significant homology with the rest of reported MCPs inhibitors. Latexin is a reversible and potent inhibitor of hCPA1, hCPA2, hCPB, hCPA4, among other A/B-type MCPs (Pallarès et al. 2005), with K_s in the nanomolar range. It lacks a signal peptide, a fact that suggests a cytosolic localization. Therefore, a rather general functional role, such as the control of cytosolic protein degradation was suggested (Normant et al. 1995). Three-dimensional structure of latexin shows two structurally related domains linked by a connecting helix. Each domain comprises an extended α -helix followed by a strongly twisted four-stranded antiparallel β -sheet of simple up-and-down connectivity that embraces the helix establishing hydrophobic contacts (Figure 13) (Pallarès et al. 2005).

Latexin is expressed in neural and non-neural tissues of rat but also a latexin homologue has been found in humans, both consisting of 222 residues (Liu et al. 2000).

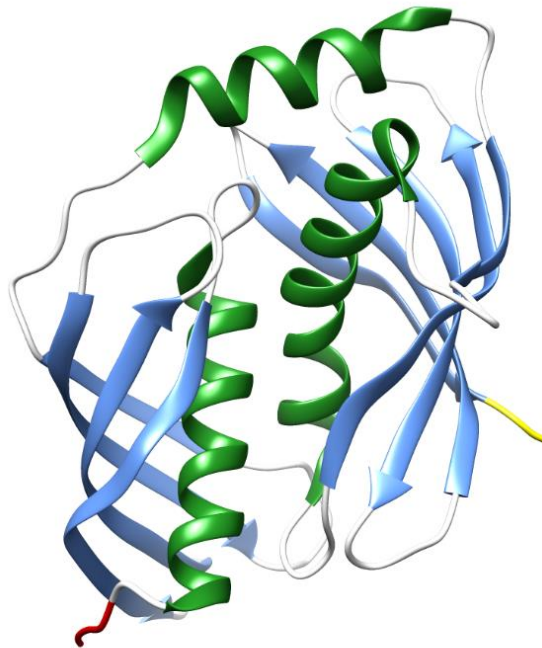


Figure 13: Three-dimensional structure of human endogenous carboxypeptidase inhibitor or latexin. Its structure is composed by two identical topological domains with similar cystatin structure. The both domains are faced by their α -helices, and they are connected by a third α -helix as a connection segment (Pallarès et al. 2005).

B2.1.1. Mechanism of MCPs inhibition

The mechanism of inhibition of the exogenous small inhibitors PCI, ACI, LCI and TCI relies on the interaction of their C terminal tail with the active site groove of the carboxypeptidase in a way that mimics substrate interaction (Vendrell et al. 2000). So, they are substrate-like inhibitors. Although there is no sequence similarity between all these inhibitors along most of such molecules (the N-terminal + central globular regions), there is a significant conservative sequential homology at the C-terminal ends (Figure 14. A). More important yet, comparisons of the crystal structures of such inhibitors in complex with MCPs indicate that the C-terminal tails of the inhibitors display a very close conformation and structural alignment, and that the last residue is coordinated with the zinc atom of the active-site of the enzyme (Figure 14. B). Thus, even though these exogenous carboxypeptidase inhibitors are isolated from evolutionary distant organisms, this is a good example of convergent evolution dictated by the architecture of the active site of the target enzymes.

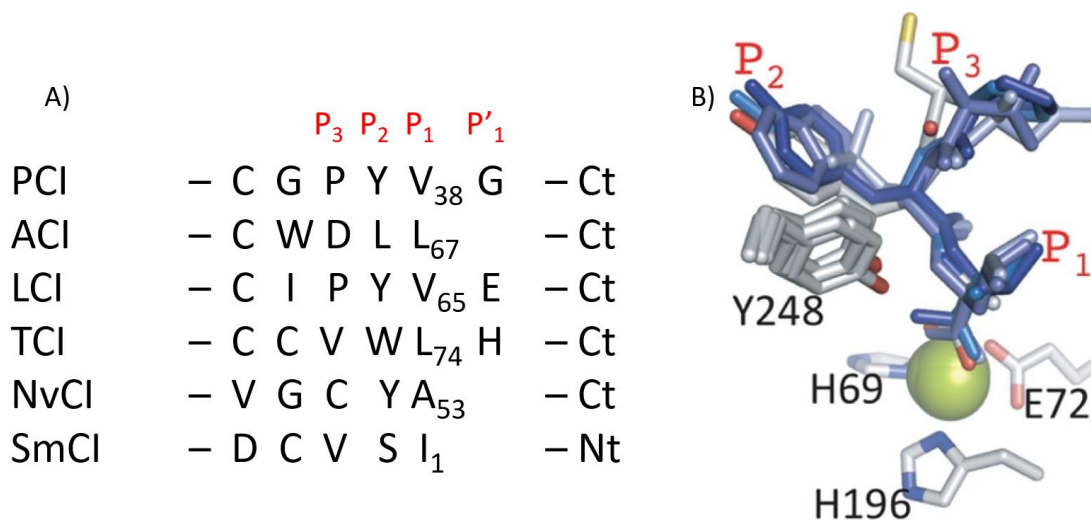


Figure 14. Sequential and structural alignment of C-terminal from exogenous M14A proteinaceous inhibitors. A) Aligned representation of the different C-terminus from exogenous proteinaceous MCPs inhibitors, except SmCI which, in a non-canonical way, uses its N-terminal to interact with the target. The residues involved in primary interaction are identified as P₃, P₂, P₁ and P'₁. B) Stick conformational representation of residues in positions P₃, P₂ and P₁, described in image A. The C-terminal tails used in this graphical representation are from PCI, ACI, LCI, TCI and NvCI inhibitors. As it is shown, the C-terminal of inhibitors interacts with zinc ion, mimicking a substrate behaviour. Zinc is represented as a yellow sphere. Reproduced from (Covaleda et al. 2012).

Additionally, TCI anchors to the surface of A/B-type MCPs in a double-headed manner not seen for the other inhibitors, where the last 3 residues of the C-terminal tail in the inhibitor interact with the active-site of the enzyme in a standard manner but its N-terminal domain binds to an exosite distinct from the active-site groove (Arolas et al. 2005).

Remarkably, the SmCI shows an atypical inhibition mechanism because it is conducted by the N-terminal of the protein, of which its first residue interacts with the zinc ion, in contrast with the rest of exogenous inhibitors described.

This mechanism previously described in which the C-terminal tail from inhibitor is essential, is very different from mechanism of ECI/Latexin. The C-terminal tail of this inhibitor does not seem to be a suitable interactor for carboxypeptidases (Reverter et al. 1998). ECI inhibitory mechanism imitates the pro-region domain of PCPs. This endogenous inhibitor interacts with the enzyme through a large surface. The main interaction area includes the lower barrel surface of latexin around the β -sheet located at its C-terminal domain. The active-site blocking area involves an inhibitory loop located in the central part of the mentioned β -sheet (Pallarès et al. 2005).

B2.1.2. Latexin family proteins

The latexin family is composed by proteins related to latexin that belong to MEROPS proteinase inhibitor family I47 (Rawlings et al. 2004).

In addition, it is also composed by two more members: the Retinoic Acid Receptor Responder 1 (RARRES1) or Tazarotene Induced Gene 1 (TIG1), and the Ovocalyxin-32 (OCX32). RARRES1 is a transmembrane protein whose expression is up regulated by retinoic acid. It is also described as a tumour suppressor whose diminished expression is involved in the malignant progression of prostate, breast cancer and hepatocellular carcinoma (Jing et al. 2002, Peng et al. 2012, Chen et al. 2014).

The other important member of the family is Ovocalyxin-32, a 32kDa eggshell matrix protein present at high levels in the uterine fluid during the terminal phase of eggshell formation and also predominantly localised in the outer eggshell. The timing of OCX32 secretion into the uterine fluid suggests that, at least, it may play a role in the termination of mineral deposition (Hincke et al. 2003). OCX32 protein possesses limited identity (32%) to latexin and RARRES1.

To go into detail on this protein, we focus in describing the structures where OCX32 is localized, the avian egg. Avian egg has thoroughly studied at nutritional level for food science and, among others, as a source of to obtain components for biotechnological and biomedical applications.

B2.2. Avian egg

Avian egg is a reproductive structure which evolved to protect the embryo of potential physical, microbial or thermal attacks from external environment. In addition, this structure also supplies functional help for the proper development of the embryo. This description is focused on chicken egg. The egg can be divided into two main parts: the vitellus or deutoplasm (egg yolk) and the albumen or glair (egg white) (Hincke et al. 2008).

Vitellus: This part is the nutrient-bearing portion of the egg whose primary function is to supply nourishment for the development of the embryo. Yolk makes up about 33% of the liquid of the egg and approximately is has three times more energy content than albumen, mostly due to its unsaturated fatty acids and saturated fatty acids composition (Hilditch and Williams 1964, Bitman et al. 1975).

Other elements present in vitellus are proteins, representing about 16% of all yolk composition. These proteins are involved in diferents biological roles: development of the embryo (Goulas et al. 1996, Sawaguchi et al. 2006), defense (Williams 1962, Kiosseoglou 2004), and on lipid and cation storage (Anderson et al. 1998, Thompson and Banaszak 2002).

Furthermore, vitellus has the highest percentage of minerals of all egg, which is rich in phosphorus, calcium and iron among others (Stadelman and Pratt 1989, Hallberg et al. 1997, Hallberg and Hulthén 2000). On the other side, deutoplasm is also rich in some vitamins (Booth et al. 1992, Nys and Sauveur 2004, Seuss-baum 2007).

At physical level, the vitellus can be divided in different parts (Figure 15):

- **Germinal disc:** this structure appears as white plaque at surface of the oocyte. This disc is either a blastoderm (fertilized) or blastodisc (unfertilized).
- **White yolk:** also known as latebra, its name is due to their images in nuclear magnetic resonance (NMR). When the yolk is analysed by NMR, the latebra space is represented as a clearer space than the rest of the yolk because its low lipidic composition. White yolk is localised in the centre of yolk, and its function is no described yet, but it could act as scaffold where other deutoplasm elements could be bound.
- **Nucleus of Pander:** small nucleus of white yolk. Its function is nourishment as the rest of the yolk, it does not show another important role for embryo development (Schmitt 2005).
- **Yellow yolk:** primarily vitamins and minerals source of the yolk. In addition, it is also composed by the half of yolk proteins and all lipids.

- **Vitelline membrane:** extracellular protein membrane covering the yolk. It is composed of glycoproteins. Recent studies suggested that this layer corresponds in fact to innermost portion of chalaziferous layer (Rahman et al. 2007).

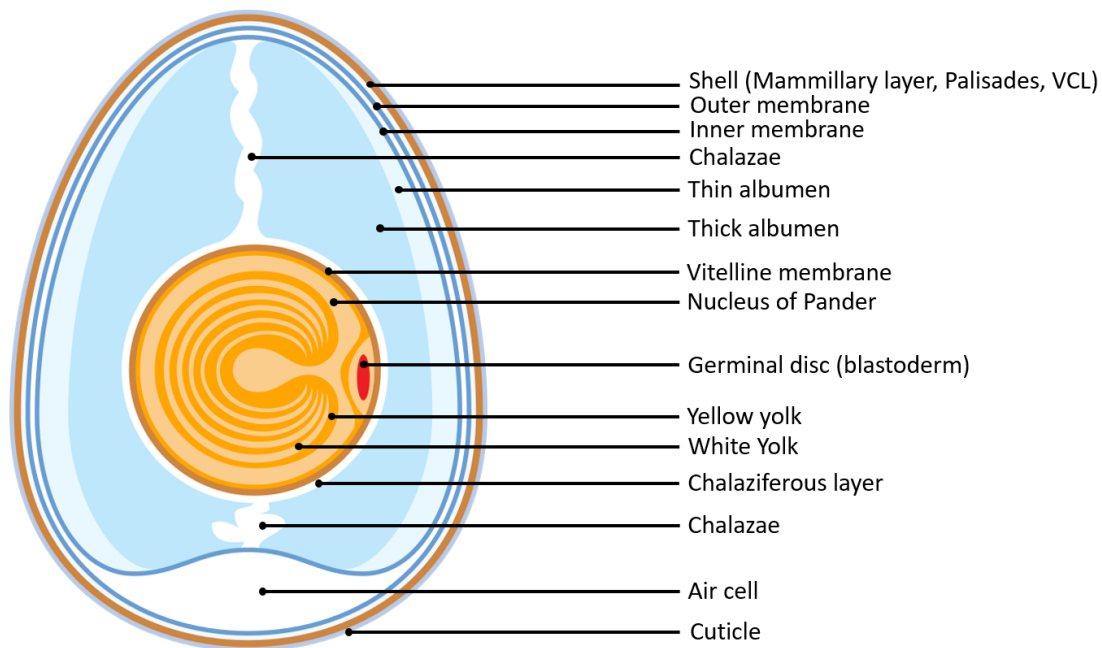


Figure 15: Schematic representation of physical parts of an egg. Every major morphological part is labelled.

Albumen: This part represents two-thirds of an avian egg by weight. It is formed by layers of various secreted elements, mostly proteins, from anterior section of oviduct (Gill 2007). The most abundant proteins on this region are ovalbumin (Mine 1995), ovotransferrin (Superti et al. 2007), ovomucoid (Lineweaver and Murray 1947), ovomucin (Watanabe et al. 1998) and globulins, together with lysozyme (Lesniewski et al. 2004, Vilcacundo et al. 2018). All these proteins play roles for embryo welfare: since support role until defence against microbiological threats.

Other elements present in albumen are lipids, but their concentration is negligible if compared with egg yolk. And carbohydrates are present too in albumen, but more than half are combined to proteins, and the rest is mostly glucose (Mine 1995, Ternes 2008).

At physical level, the albumen could be divided in different parts (Figure 15):

- **Chalaziferous layer:** It is a gelatinous layer which covers egg yolk. (Romanoff and Romanoff 1949). This layer extends to become chalazae, which hold the yolk.
- **Chalazae:** Also known as Chalazae cord, are two twisted clockwise bands at both sides of the yolk membrane, it allows the yolk being suspended in the centre of egg. The chalazae are slightly elastic and permit limited rotation of the yolk (Okubo et al. 1997).

INTRODUCTION

- **Thin albumen and thick albumen:** The elements which compose these layers are the same, the difference is their proportions. In thick albumen the concentration of ovomucin is four times higher than in thin albumen, and this difference produces an increase in thick albumen viscosity (Okubo et al. 1997).

All this set of elements are contained inside a structure called eggshell. It is not a compact structure, because it has between 7.000 - 17.000 small pores (Dennis et al. 1996), which cross eggshell until membranes; these structures give permeability to egg allowing the entrance of foreign elements. Eggshell is made up of inner and outer shell membrane, air cell, mammillary vesicles, palisades, vertical crystal layer and the cuticle (Figure 16).

- **Inner and Outer shell membrane:** Eggshell membranes contribute to shell strength, probably by serving as a reinforcement of the crystalline part of the shell. It acts as support
- **Air cell:** This air cell rests between the outer and inner membranes at the bigger end of the egg.
- **Mammillary layer:** Mammillary bodies or cones are the nucleation sites of the mineral portion of eggshell, and they are located in the surface of outer shell membrane. The individual fibres of outer eggshell membrane are embedded into mammillary cones (Dennis et al. 1996).
- **Palisades:** This layer consists of columns of calcite crystals disposed perpendicular to the eggshell surface and extend outwards from the mammillary cones. The palisade layer ends in a thin vertical crystal layer aligned perpendicular to the shell surface. Its thickness is around 200µm corresponding to two-thirds of eggshell. Its role is crucial in the determination of eggshell breaking strength.
- **Vertical crystal layer:** this layer gives rigidity to eggshell. Calcium carbonate (CaCO_3) is one of the principal elements which forms this layer. This layer is the final part of the calcified structure of eggshell.
- **Cuticle:** it is a thin organic layer (12µm) that coats vertical crystal surface. Fundamentally, it is composed by glycoproteins (90% of elements), but also by lipids and polysaccharides. Cuticle proteins regulate eggshell permeability, but besides these proteins has shown a strong antimicrobial activity (Muñoz et al. 2015).

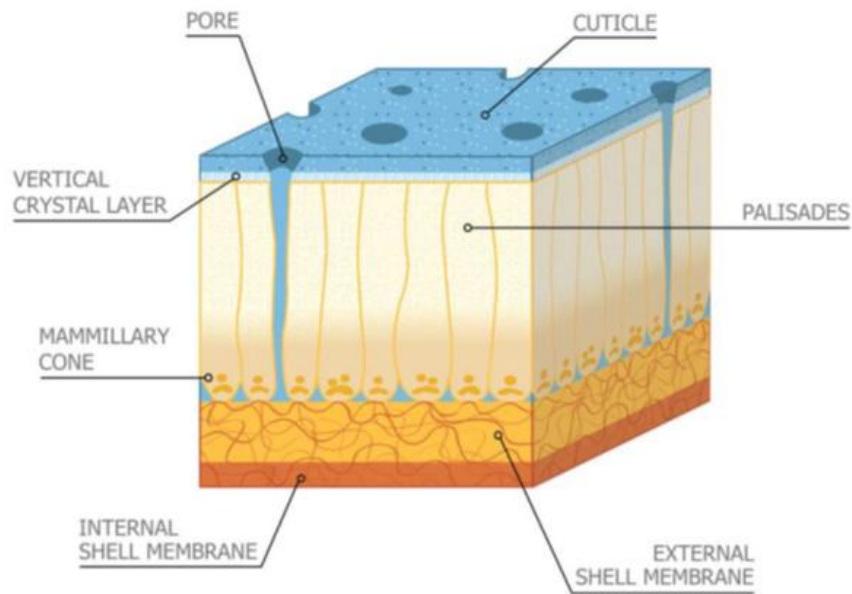


Figure 16: Graphic representation of cross-sectional view of chicken eggshell. Every major morphological part is labelled. Reproduced from (Hincke et al. 2012)

The main objective of role of it is protecting embryo from physical damage, microorganisms, and small predators, and this must be compatible with easy breakage from inside to allow hatching of embryo. In addition, eggshell components take part in embryo survival and development processes like regulatory role in metabolic gases and water exchange or acting as calcium source (Hincke et al. 2008, Kulshreshtha et al. 2018).

B2.2.1. Formation of the egg

Domestic hens can produce more than 300 eggs per year. This is possible due to their continuous ovulation. In the ovary of hens more than 12,000 oocytes are present, but only a small proportion (300-500) will later acquire yolk and develop into mature ovulatory follicles to give rise to egg formation (Nys and Guyot 2011).

The process of egg formation starts 15-60 minutes after oviposition, when the next mature follicle ovulates, and oocyte is released. The ovulation cycle takes between 24 to 27 hours, it depends on breed of the hen and its ratio of production (Nys and Guyot 2011, Johnson 2015).

The ovum released is captured by the infundibulum (Figure 17) and thus funnelled into the reproductive tract. During this part of the process, fertilization of the yolk and formation of the vitelline membrane happen. At the same time, the process of formation of chalazae also begins, and is finished in the magnum.

INTRODUCTION

Subsequent to the infundibulum, the vitellus moves through the magnum (Figure 17) where albumen formation process takes the next 3-4 hours. The principal component of albumen is water, with high concentration of proteins which are produced in the mucosal cells from magnum (Etches 1996).

Following the oviduct tract, the egg (it already has albumen and deutopasm) is transferred into isthmus (Figure 17) where the inner and outer shell membrane are formed. These are impermeable for egg white, but they allow the exchange of gas, minerals and water. On the surface of outer shell membrane, mamillary cores are formed, which are the responsables of eggshell formation.

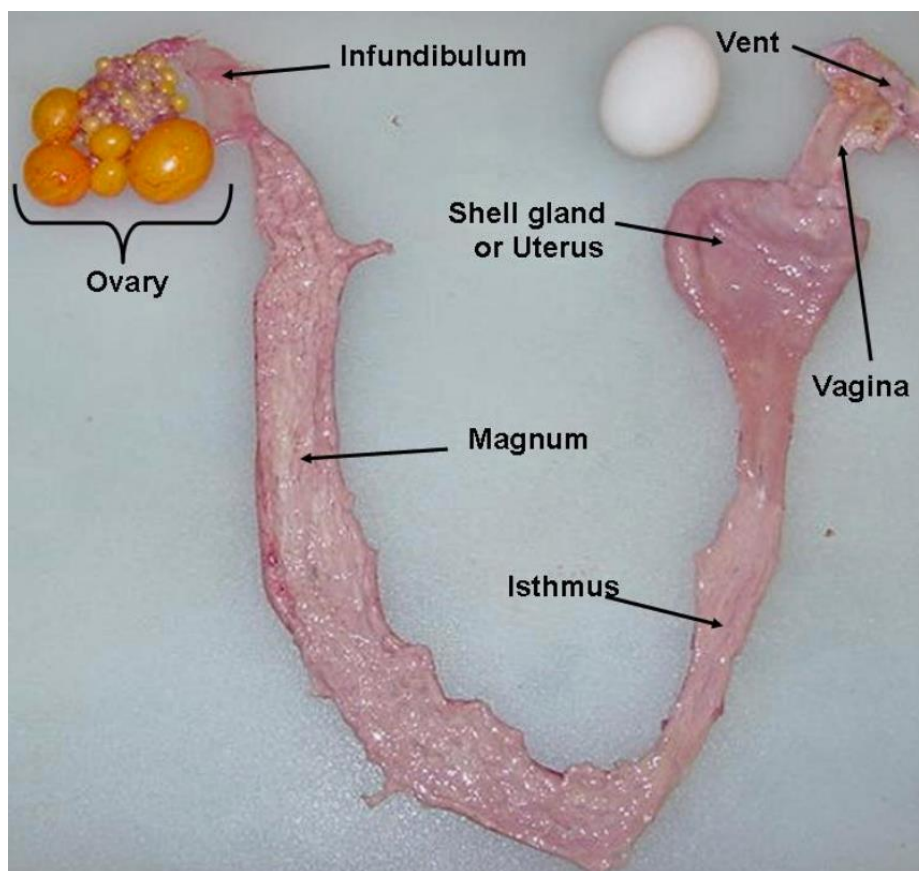


Figure 17: Picture of the reproductive tract of a female chicken (hen). Most birds have only one tract, the left ovary and oviduct (Pollock and Orosz 2002). Picture reproduced from Jacquie Jacob and Tony Pescatore; College of Agriculture, Food and Environment, University of Kentucky.

Once the membranes are formed, the egg is moved to uterus (Figure 17), where it will develop its characteristic ovoid form. The calcification process takes up to 19 hours, it is characterized by the secretion of calcium and bicarbonate (HCO_3^-) into uterine lumen in order to form calcite crystals. In order to help, some matrix glycoproteins and polysaccharides are secreted too, their function is acting as scaffold to become calcified (Nys et al. 1999).

Previously to oviposition, when the mineralization process is done, one last process remains: cuticle formation. Elements of the cuticle (mostly proteins and polysaccharides) are formed and deposited on the egg. Accordingly, its primary function is to prevent penetration of bacteria into the egg and to limit the loss of water from the egg (Schusser et al. 2013).

The oviposition is initiated by the relaxation of the uterine sphincter and the contraction of uterine smooth muscles, letting the egg pass through the vagina and the lent of the hen (Figure 12). Within the coming 60 minutes after oviposition the next ovulation will take place.

B2.2.2. Proteins in eggshell

As it is described above, the eggshell is a structure that not only contains the vitellus and albumen, but also plays a regulatory role in egg permeability and protective role from exogenous threats. These roles are facilitated due to presence of some specific proteins on eggshell. These proteins are:

Ovocleidin-17 (OC-17): It is an abundant matrix phosphoprotein with C-Type lectin domain. It is distributed throughout the shell matrix but is concentrated in the mammillary bodies (Hincke et al. 1995, Mann and Siedler 1999, Mann et al. 2007). In its surface, this protein presents a large number of Arg and Lys residues forming a cationic antimicrobial structure, conferring to OC-17 a high antimicrobial activity (Wellman-Labadie et al. 2008, Freeman et al. 2010, Freeman et al. 2012).

Ovocalyxin 36 (OCX-36): Its aminoacidic sequence displays high identity to lipopolysaccharide-binding proteins (LPS), bacteripermeability-increasing proteins (BPIP) and Plunc family proteins, proteins that are key members of the innate immune system (Bingle and Craven 2004, Tian et al. 2010). This fact suggests a potential role in egg defence binding to bacterial lipopolysaccharide from cell wall. It has been observed that OCX-36 also plays an important role during calcification (Gautron et al. 1997, Cordeiro et al. 2013, Kovacs-Nolan et al. 2014).

Ovocleidin-116 (OC-116): One of the most abundant in eggshell. It is mainly localized in palisade layer (Hincke et al. 1999) It is secreted by the granular cells of the surface epithelium on uterus in the growth phase of eggshell calcification (Nys et al. 2004). In some studies, OC-116 showed high affinity for CaCO₃ and because this is suggested a biological role as an important actor in eggshell calcification growth (Hernández-Hernández et al. 2008, Hernández-Hernández et al. 2008).

Ovocalyxin-32 (OCX32): The protein was identified as a 32-kDa uterine fluid protein which is abundant in the terminal phase of shell formation (Gautron et al. 2001, Hincke et al. 2003).

INTRODUCTION

OCX32 is expressed at high levels in the uterine and isthmus regions of the oviduct and is secreted by epithelial cells of these sections (Gautron et al. 2001). A proteomics study identified phosphorylation of OCX32 at serines and threonines between positions 257 and 268, but exact sites were not determined (Mann et al. 2007) In eggshell it localizes to outer palisade layer, vertical crystal layer and predominantly in the cuticle. (Figure 18) (Gautron et al. 2001)

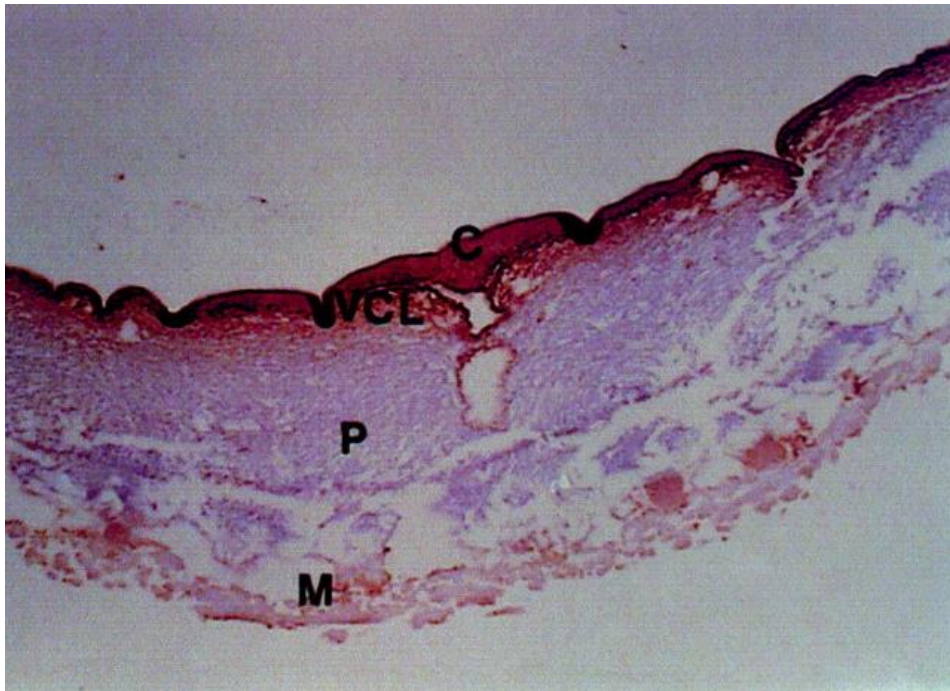


Figure 18: Immunohistochemistry to localize the 32-kDa protein (OCX32) in eggshell matrix. Histological image stained with anti 32-kDa protein serum from rabbit as primary antibody and an anti-rabbit antibody-colloidal gold conjugate as secondary antibody, *M*, membranes and mammillary layers; *P*, palisade layer; *VCL*, vertical crystal layer; *C*, cuticle. Reproduced from (Gautron, Hincke et al. 2001).

OCX32 has limited identity (32%) to two mammalian proteins: latexin (LXN), which is described as ECI exogenous inhibitor of MCPs inhibitors at B2.1.2. section, and Retinoic Acid Receptor Responder 1 (RARRES1), which is expressed in multiple tissues, even in skin where it was originally isolated (Nagpal et al. 1996). The expression of RARRES1 is upregulated by tazarotene and retinoic acid and downregulated by promoter and CpG methylation in human neoplastic cells (Shutoh et al. 2005). RARRES1 is implicated in the therapeutic treatment of psoriasis by retinoic acid (Nagpal et al. 1996) and it could act as tumour suppressor protein (Jing et al. 2002, Roy et al. 2017), but recently in some kinds of cancer has observed its role as promoter of tumour progression (Wang et al. 2013, Zimpfer et al. 2016).

Phylogenetic analysis suggests that the OCX32 is orthologous to the RARRES1 gene and paralogous to the LXN gene. As shown in the tree, the ancestral latexin-like gene might have duplicated and given formation to OCX32 (in birds)/RARRES1 (in mammals) and their paralog

latexin (LXN) (Figure 19.A). This hypothesis is supported by genomic analyses, in the chicken genome, OCX32 is localized next to LXN on chromosome 9, while RARRES1 is next to LXN on chromosome 3 in human genome. By comparison between chicken (*Gallus gallus*) and human (*Homo sapiens*) (Figure 19.B). RARRES1 is localized in the orthologous position of OCX32 in human (Tian et al. 2010).

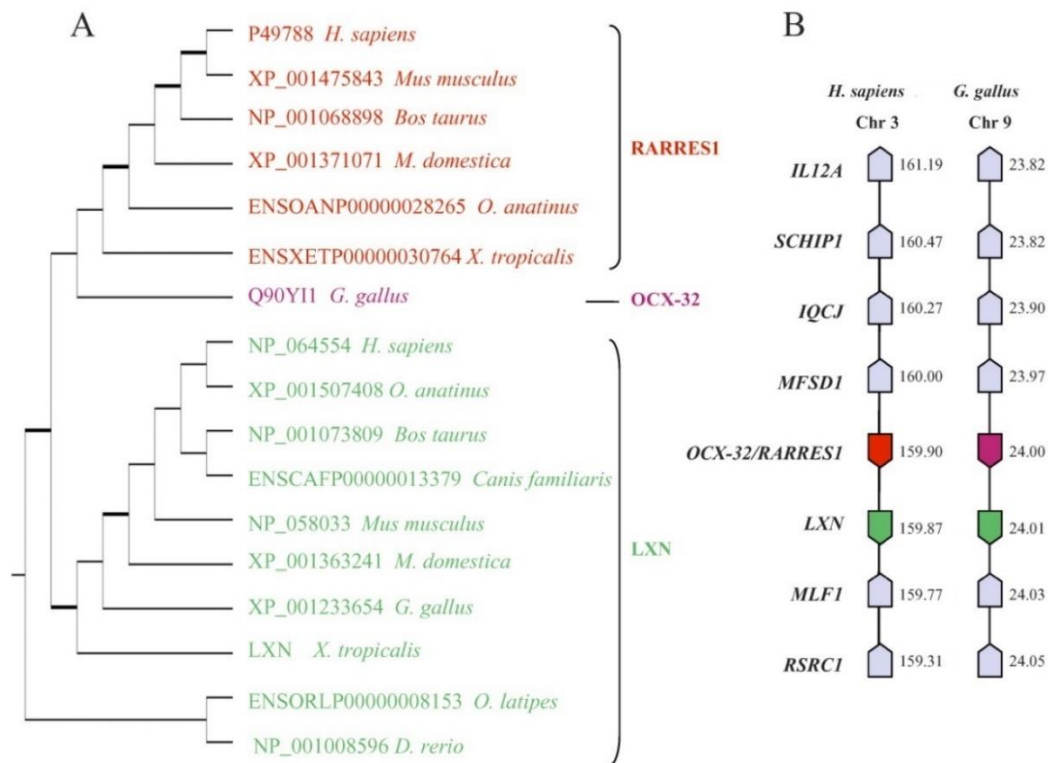


Figure 19: Gene divergence between the *G. gallus* eggshell-expressed ovocalyxin-32 with its ortholog, the RARRES1 gene in mammals and other species. A) Consensus phylogenetic tree of ovocalyxin-32 and its relatives from other species. **B)** The genomic localization of OCX32 in the *G. gallus* and syntenic comparison with *H. sapiens*. Numbers on the right of pentagons are gene locations expressed in Mb for different species. Reproduced from (Tian et al. 2010).

Due to time when OCX32 is secreted into the uterine fluid, it has been suggested that it plays a role in the termination of eggshell calcification. This role was hypothesized when morphological changes in calcite crystals of CaCO₃ after incubation with from fresh uterine were observed (Dominguez-Vera et al. 2000, Hernández-Hernández et al. 2008). Another potential role is as “first-responder” molecule against bacterial colonization of the eggshell due its localization on the cuticle. This role was confirmed in some antimicrobial activity studies (Xing et al. 2007, Wellman-Labadie et al. 2008, Wellman-Labadie et al. 2008). The last potential role described for OCX32 is carboxypeptidase inhibitor, due its homology with latexin, an already described carboxypeptidase inhibitor.

INTRODUCTION

Our research group at UAB has been focused on the study of carboxypeptidases and their inhibitors for long time. Most of the proteinaceous MCPs inhibitors described use to be small (low weight), with simple three-dimensional structure and disulphide rich, except latexin. As described before, latexin is representative of a protein family of which, OCX32 is considered a member in spite that its activity and potential role as inhibitor is not characterized. So, we consider the study of this protein and potential inhibitor as interesting for the field of carboxypeptidases, their inhibitors and protein structure function.

COMMENTS AND JUSTIFICATIONS ON THE SUBJECTS OF THIS THESIS

In this PhD thesis we have been working on two apparently, and mostly, different subjects: an aldehyde dehydrogenase enzyme, ALDH1A3, belonging to the field of oxidoreductases-dehydrogenases, and an enzyme inhibitor, OCX32, belonging to the field of metalloproteases and inhibitors. It could be argued that they both belong to classes of proteins with a generic function or tendency for "protection", as it happens with the mechanisms for the removal or mitigation of intracellular aldehydes (as retinal for ALDH1A3), and for the inactivation of proteases by protease inhibitors (as it might be interpreted for the strong carboxypeptidase inhibitory capability of OCX3, confirmed in this thesis). However, actually, little is known on the true extra/systemic functions of ALDH1A3, besides the particular transformation of retinal to retinol at cytosol or mitochondria, as exemplified by the nowadays envisaged drug-based knocking of such enzymes for cancer therapy. Also, little is known about the same on OCX3, besides its supposed role in facilitating the deposition of a mineral shell on top in avian eggs and in acting as a protective film against the penetration of microorganism inside the eggs.

What is more certain is that our research group at UAB, expert on proteases and inhibitors, in order to increase research capability and adapt to difficult times for scientific research financing, established a long-term collaboration with another neighbouring research group expert on oxidoreductases. That is, both groups experts on enzymology and on structure-function analysis are sharing grants and publication. This has been one of the main practical reasons behind, but we are also genuinely trying to establish bridges between both expertises and between both fields. This is an important reason for this PhD thesis.

OBJECTIVES

OBJECTIVES

This thesis has been carried out in the Protein Engineering and Proteomics group of Institut de Biotecnologia i Biomedicina at Universitat Autònoma de Barcelona. The group, headed by Drs. Francesc Xavier Avilés and Julia Lorenzo, is specialized on protein sciences in general, and more specifically on the study of metallo-carboxypeptidases and their inhibitors. It also works in collaboration with nearby and international groups in related synergic research. One of the strongest collaborations is with a group specialized on oxidoreductases (i.e. alcohol and aldehyde dehydrogenases), headed by Drs. Jaume Farrés and Xavier Parés, localized at the Departament de Bioquímica i Biologia Molecular at Universitat Autònoma de Barcelona.

In this context, the main objective of this thesis has been to characterize enzymes and inhibitors of biotechnological and biomedical interest in the fields of carboxypeptidases and oxidoreductases and develop new tools and approaches for the production and characterization of these biomolecules. It is expected that this would reinforce joint bridges and collaborative research between both groups.

Specific objectives were proposed for each chapter:

CHAPTER I

- To obtain by recombinant approaches the aldehyde dehydrogenase ALDH1A3-C314S mutant variant, supposed to be catalytical inactive or with decreased activity, to be used as a reference molecule.
- Characterize any remanent catalytic activity on the C314S mutant.
- Analyse by X-ray crystallography and/or computational approaches the three-dimensional structure of the C314S mutant.
- Investigate the interaction between the enzyme, the cofactor and ligands.

CHAPTER II

- To analyse the protein surface and the distribution along it of positively charged patches of some MCPs, as potential heparin binding regions.
- Evaluate the contribution of heparin to the recombinant production of our proteins of interest in mammalian cell cultures, as well as its cytotoxicity
- Develop and evaluate easy and inexpensive production systems to obtain high yields of MCPs by the heparin-affinity approach.

CHAPTER III

- To isolate in native, pure and soluble state a potential inhibitor of MCPs from a natural source (eggshell), the Ovocalyxin-32 (OCX32) protein.
- Characterize the purified protein as inhibitor against different proteases, and particularly against M14 metallo-carboxypeptidases.
- Determine the three-dimensional model structure of OCX32, by in silico methods, and desirably also by X-ray crystallography, if feasible, to elucidate the potential mechanism of carboxypeptidase inhibition.

CHAPTER I

CHARACTERIZATION OF A MUTANT VARIANT OF ALDH1A3 (MUTANT C314S)

CHAPTER I: CHARACTERIZATION OF A MUTANT VARIANT OF HUMAN ALDH1A3 (C314S MUTANT)

1.1. INTRODUCTION

Aldehydes are highly reactive molecules related with cell toxicity, and its metabolism involves several enzymes systems. One of the most important in human cells is the aldehyde dehydrogenase superfamily. Aldehyde dehydrogenase (ALDHs) are proteins responsible on the detoxification of specific aldehyde substrates from alcohol metabolism and lipid peroxidation (Vasiliou et al. 2004). They are ubiquitous abroad human tissues and display distinct substrate specificity (Sládek 2003, Vasiliou et al. 2004) Generally, ALDH serve to protect cells from damaging effects of aldehydes by their oxidation to carboxylic acids (Pappa et al. 2005). In the last years, ALDHs medical importance is evidenced when human pathological phenotypes are linked or influenced to mutations in different ALDH genes that cause misfunction, absence, inactivation or protein deficiency. ALDHs are related to different pathologies such as many cancers, neurological abnormalities, and metabolic diseases, among others (Vasiliou et al. 2004, Marchitti et al. 2008).

. From the large number of members that forms the ALDHs superfamily in humans, we focus on ALDH1A3. Human ALDH1A3 is a protein of 512 amino acid residues and is localized in the cytoplasm, nucleus, and mitochondria (<http://www.genecards.org/>) ALDH1A3 participates on the synthesis of retinoic acid from retinaldehyde, compound that is required for growth and development of different organs and also participates on the embryonic development (Graham et al. 2006, Wagner et al. 2006, Everts et al. 2007).

In recent studies, ALDH1A3 have been established as interesting target for the treatment of different types of cancers. It is highly expressed in pancreatic, ovarian and breast cancer and high-grade gliomas but it is not expressed or in very low levels in the corresponding para-neoplastic tissues (Marchitti et al. 2008, Duan et al. 2016). Accordingly, targeting these enzymes with suitably designed inhibitors represents a successful strategy for some cancer's treatment (Jiménez et al. 2019).

With the objective to obtain more information about this protein we designed and obtained a mutant variant where the catalytic centre is supposed to be inactive, because we substituted the Cysteine 314 (C314) by a Serine (C314S). The objective of this mutation is not only to realize a characterization with kinetical purposes if not to obtain a good three-dimensional structure of this protein alone or in complex with substrates, the cofactor and some inhibitors. We think that the structure of this mutant could become a good tool to be used for structure-based drug design inhibitors to mitigate the effects of ALDH1A3 in diseases.

1.2. EXPERIMENTAL PROCEDURES

1.2.1. Bradford assay

Bradford assay is a spectroscopic total protein determination method useful for its reproducibility, stability and rapidity. This method is based on the absorbance shift observed in acidic solution of dye Coomassie® Brilliant Blue G-250 (Bradford 1976). This dye binds to the protein, aromatic and basic aminoacids (Compton and Jones 1985), by an electrostatic attraction due to its sulfonic groups. The interaction between dye and protein results in a colour change from reddish brown to blue.

For total protein quantification, Pierce™ Coomassie (Bradford) Protein Assay Kit from Thermo Fisher Scientific was used when possible, following Microplate protocol detailed in manual of the manufacturer. Standards were freshly prepared with Bovine Serum Albumin (BSA) from 0.1 mg·mL⁻¹ stocks. The diluent used was the buffer from the samples. To proceed with protein quantification, 150 µL of samples and standards were mixed with 150 µL of kit reagent per well in a 96-well microplate. Every sample or standards were added in triplicates. After 5 minutes at room temperature incubation, absorbance was quantified on a spectrophotometric plate reader (Perkin Elmer Victor3 V) at 620nm.

1.2.2. Sodium Dodecyl Sulphate Polycrylamide Gel Electrophoresis (SDS-PAGE)

Electrophoresis is the motion of particles inside a fluid when a uniform electric field is applied. This technique is used in order to separate macromolecules based on size; the separation of this macromolecules is done with a gel made of polymer. Focus on proteins, there are many factors which intervene in protein mobility as protein size, charge type and density, and finally protein morphology.

In order to use the protein size as unique factor to conditioning the electrophoretic mobility, the rest of the factors must be equalised by different surfactants. First, one used is Sodium dodecyl sulphate (SDS), which is an anionic detergent with denaturing properties for proteins. Furthermore, SDS covers the natural charge of the protein and gives to it a net negative charge.

In addition, a reductant agent is also used for breaking covalent bonds which are involved in protein conformation (disulphide bonds), the agent used was β -mercaptoethanol. Proteins incubated at high temperature in combination with SDS and β -mercaptoethanol presence were successful denaturated (Shapiro et al. 1967, Laemmli 1970).

Samples were mixed with 0.2 volumes of Laemmli 6x (2% w/v SDS, 10% v/v Glycerol, 60 mM Tris-HCl pH6.8, 0.01% v/v Bromophenol Blue and 2% v/v β -mercaptoethanol), incubated for 5 minutes at 95°C and loaded into the gel. The gel is composed by two layers:

- **Stacking gel:** this layer has large sized pores that allow the proteins to migrate freely and get stacked at interface between gels. This layer makes sure that the proteins start migrating from same level. It is composed by 125 mM Tris-HCl pH 6.8 (Sigma-Aldrich), 0.1% w/v SDS (Sigma Aldrich), 0.05% w/v Ammonium Persulphate (Sigma-Aldrich), 4% v/v 40% Acrylamide/Bis Solution 37,5:1 (NBS Biological) and 0.025% v/v TEMED (GE Healthcare).
- **Resolving gel:** this layer has small sized pores due to higher acrylamide concentration. These smaller pores generate size restriction which allows the protein separation by size. It is composed by 375 mM Tris-HCl pH 8.8 (Sigma-Aldrich), 0.1% w/v SDS (Sigma Aldrich), 0.05% w/v Ammonium Persulphate (Sigma-Aldrich), 8-20% (variable) v/v 40% Acrylamide/Bis Solution 37,5:1 (NBS Biological) and 0.15% v/v TEMED (GE Healthcare).

Samples are run in an electrophoresis chamber with Running buffer (25 mM Tris-HCl, 192 mM Glycine and 0.1% SDS) for 20-25 minutes at constant current of 20mA for each gel which is running. When samples have overcome stacking gel after 25 minutes, current is change to 40 mA for each gel, as long as required.

SDS-PAGE protein separation is used to be combined with protein identification techniques as Coomassie staining for unspecific protein detection or western blot for specific protein detection.

1.2.3. Coomassie SDS-PAGE staining

After an SDS-PAGE is done, in the gel is impossible to determine the protein separation pattern, the gel needs a staining process. The most common used method is using Coomassie G-250 dye, as is described before, this dye binds aromatic and basic residues from proteins and this interaction makes a blue signal, in this case a blue band which represents the protein (Fazekas de St Groth et al. 1963).

Original staining process includes two principal steps: first one is staining, all the gel is stained in a blue organic solution of Coomassie on mild shaking, in order to staining proteins. This solution is composed by organic solvents as methanol and acetic acid with Coomassie dye. This process is incubated about an hour. The second step is unstaining process, with an organic solution composed by methanol and acetic acid too, where the gel is submerged. The unstaining process is composed by some washes until the unstaining solution gets clean, obtaining a clear background in the gel and blue bands which represents proteins (Meyer and Lamberts 1965).

The staining solution that we use in laboratory is a commercial one called BlueSafe® from NZYTech. BlueSafe® is safer than Coomassie staining because does not contain any methanol or acetic acid in its composition. Furthermore, it does not require the use of a destaining solution.

1.2.4. DNA mutagenesis

In vitro site-directed mutagenesis is an invaluable technique for characterizing the dynamic, complex relationships between protein structure and function, for studying gene expression elements, and for carrying out vector modification. The rapid three-step procedure generates mutants with high efficiency in a single reaction (Figure 20).

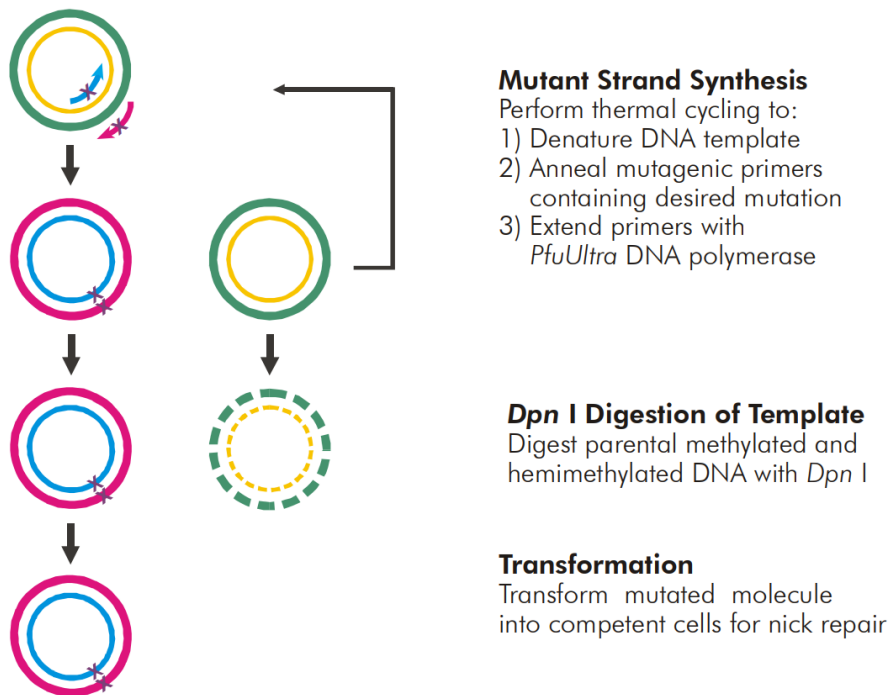


Figure 20. Overview of the Quick-change II site-directed mutagenesis method. Green-Yellow strands are paternal or original plasmid DNA. Magenta-Blue strands mutated primers and new synthesis mutated plasmid DNA.

Human ALDH1A3 was mutated using the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies). Quick-Change system is based on polymerase chain reaction (PCR), which is a common molecular biology technique that enables researchers to make multiple copies of a specific region of DNA. Once the DNA has been sufficiently amplified, the resulting product can be sequenced, analysed by gel electrophoresis, or cloned into a plasmid for experimental purposes. On this case the purpose is to generate a new plasmid with the mutation incorporate on the gene sequence.

For this purpose, the following primers were used: ALDH1A3 C314S forward primer (5'-CTTCAACCAAGGCCAGTGTAGCACGGCAGC -3') and ALDH1A3 C314S reverse primer (5'-GCTGCCGTGCTACACTGGCCTTGGTTGAAG -3'). The red coloured nucleotide corresponds to the mutated ones.

The PCR steps were:

- 1) DNA polymerase activation (95°C for 1 min),
- 2) denaturing (95°C for 30 s),
- 3) annealing (65°C for 30 s)
- 4) extension (72°C for 60 s/kb).

Then, the steps 2-4 were repeated for additional 18 cycles, and one final extension step at 72°C for 10 minutes to be sure that all sequences are cloned until the end. Then, the PCR products were digested with *DpnI* to eliminate the parental DNA. The resulting constructs were transformed into *E. coli* XL1Blue, and sequenced.

1.2.5. Protein expression and purification

Mutant ALDH1A3 was expressed from the pET-30 Xa/LIC constructs, which allow protein expression with an N-terminal (His)₆ tag under the control of T7 RNA polymerase promoter and operon *lac*. In *E. coli* BL21(DE3) pLys. The expression of T7 RNA polymerase is also under the control of the operon *lac* promoter. Consequently, the constitutive expression of the operon *lac* repressor (*lac I*), contained in both the *E. coli* genome and the expression vector, assures the expression of the protein only in presence of an inducer such as IPTG. For protein expression, transformed *E. coli* BL21(DE3) pLys cells were grown O/N in 25 mL LB medium in the presence of 40 µg/mL kanamycin and 34 µg/mL of chloramphenicol at 37°C. This culture was used to inoculate 1 L of 2XYT medium in the presence of kanamycin and chloramphenicol and cells were incubated at 37°C until an O.D.595 of 0.8 was reached. Protein expression was then induced by the addition of 1 mM IPTG and cells were grown O/N at 22°C.

After O/N incubation, cells were collected by centrifugation at 12400 x g and 4°C for 10 min. The resulting pellet was resuspended in 30 mL of Bind Buffer per liter of culture. Cell lysis was performed by freezing (O/N at -20°C) and thawing at room temperature. Then, 1 mg/mL lysozyme, 20 µg/mL DNase, 1% Triton-X 100, 1 mM protein inhibitor PMSF and 5 mM DTT were added, followed by two cycles of sonication. The cellular extract was centrifuged at 15000 x g and 4°C for 20 min. The soluble fraction was collected and filtered. Protein purification was performed by affinity chromatography on a nickel-charged chelating Sepharose™ Fast Flow 10-mL column (His Trap column), which specifically binds the protein due to the His tag, using an ÄKTA™ FPLC purification system. Protein was eluted by applying an increasing stepwise gradient (5, 60, 100 and 300 mM) of imidazole in 50 mM Tris/HCl and 0.5 M NaCl, pH 8.0. The enzyme is eluted at 300 mM imidazole. The imidazole present in the eluted protein fractions was removed through a PD-10 gel filtration-desalting column. Protein including the (His)₆ tag was stored frozen at -80°C in 50 mM Tris/HCl, 0.5 mM NaCl at pH 8.0.

1.2.6. Activity fluorescence ALDH1A3 C314S mutant activity assay.

Dehydrogenase activity with hexanal as substrate, was monitored using a fluorimeter (Cary Eclipse Varian) at 25°C, in order to determine if the mutant is active. ALDH1A3 C314S assays were performed in 50 mM HEPES, 50 mM MgCl₂, 5 mM DTT, pH 8.0. NAD⁺ concentration was 500 μM, at final volume of 1mL. The enzymatic reaction was initiated by the addition of substrate, dissolved in reaction buffer. Fluorescence of NADH was followed at 460 nm with excitation at 340 nm and spectral bandwidth of 10 nm. The reaction mixture also contained 5 μM NADH as an internal standard to obtain absolute reaction rates, which were calculated according to the equation:

$$v = \frac{dF}{dt} \cdot \frac{C_{st}}{F_{st}}$$

Where C_{st} is the standard NADH concentration (5μM), F_{st} the standard fluorescence and dF/dt the slope of the time dependent fluorescence obtained on the assays (Sołobodowska et al. 2012). Assays without enzyme were carried out as controls, and also to obtain the fluorescence standard.

Inhibition assays for this mutant enzyme were also realized in order to determine the effect of these inhibitors on ALDH1A3 *wild-type* (*wt*) and ALDH1A3 C314S. Single-point analyses of enzymatic activity at 10 μM of inhibitor were performed at similar conditions of ALDH1A3 assays, using a fluorimeter (Varian Cary Eclipse). The reaction was followed by fluorescence emission at 460 nm with the excitation from 340nm, corresponding to the NADH production. Enzymatic inhibition assays were performed in a final volume of 1 mL, 50 mM HEPES, 50 mM MgCl₂, on this case without DTT at pH 7.2, and with 500 μM NAD⁺. The compound tested as inhibitor was DEAB, which was dissolved in DMSO and assayed in a final concentration of 1% (v/v) DMSO using hexanal as a substrate at 250μM, at 25°C. DEAB is already described as inhibitor for ALDH and even it has been evaluated its efficiency for ALDH1A3 previously (Morgan et al. 2015). The reaction mixture was incubated for 20 min at room temperature (25°C), before adding the Substrate and standard of of NADH (5μM). The percentage of remaining activity was calculated from the ratio of activity at a given inhibitor concentration to the control activity with 1% DMSO without inhibitor compound added.

1.2.7. Crystallization of protein and data collection

For the crystallization process we produced the recombinant protein, purified and concentrated to 10mg/mL in 5mM Tris-HCl, pH 8.0, 150mM NaCl, 10mM DTT and 2mM NAD⁺. Crystals of ALDH1A3 C314S mutant were obtained at 18°C by hanging drop vapor diffusion methods. The reservoir solution contained 100mM BisTris pH5.5, 2.2M Ammonium sulphate, 5% PEG 400. The crystals appeared after 7 days of incubation (Figure 21).

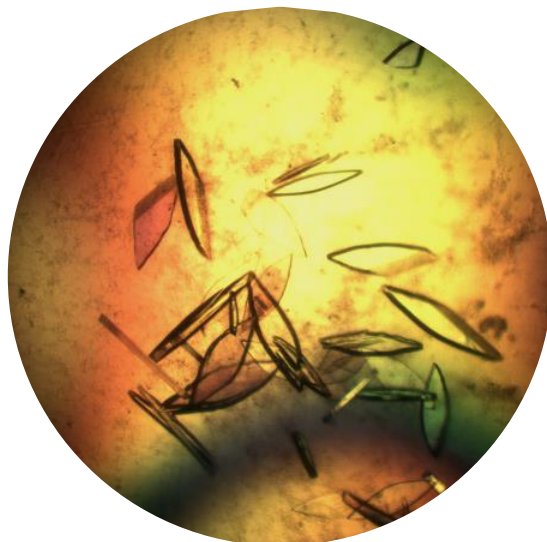


Figure 21. Picture of human ALDH1A3 C314S crystals appeared after 7 days of incubation at 18°C.

Diffraction data was collected from cryo-cooled crystals (-193°C) in a single sweep at the XALOC beamline at the ALBA synchrotron (Juanhuix et al. 2014). The crystal belonged to space group P2₂1₂1. Data was processed using the Global phasing AutoPROC program (Vornrhein et al. 2011). The resolution was cut-off at 1.759Å. The current structure was refined until an R-factor of 0.1782% and R-free of 0.2128%. The Ramachandran analysis shows 96.49% of residues on favoured regions, and 0.32% which are considered outliers residues, see Table X for further statistics. This model structure even could be refined, and these values are temporal.

1.2.8. Structure solution

The ALDH1A3 C314S structure was solved by molecular replacement with PHASER (McCoy et al. 2007) from the CCP4 package (Winn et al. 2011) using the wild-type structure of ALDH1A3 (PDB ID 5FHZ) as model. A near-complete initial model was obtained with Phenix autobuild (Adams et al. 2010). The structure was completed through alternate manual model building with Coot v.0.8.9 (Emsley et al. 2010) and refinement with PHENIX v.1.16_3549 (Adams et al. 2010). The model was validated and further adjusted and refined using MolProbity (Williams et al. 2018). The crystallographic and refinement parameters are given in Table 2.

TABLE 2. Statics of the current model of the ALDH1A3 C314S mutant.

DATA COLLECTION	
λ (Å)	0.9793
Space group	P2 ₂ 1 ₂ 1
Cell dimensions	
a, b, c (Å)	89.149, 157.871, 81.079
a, b, γ (°)	90, 90, 90
Resolution range(Å)	59.98 - 1.758 (1.957-1.758)
Ellipsoidal Completeness (%)	89.9 (53.7)
$\langle I / \sigma(I) \rangle$	7.5 (0.2)
Average multiplicity	6.7
Rsym (%) ^b	0.118
Rmeas (%) ^c	0.128
CC(1/2) (%)	0.998
REFINEMENT	
Resolution	1.758
Number of reflections	
Total	464337 (22605)
Unique	69352 (3468)
Rcryst ^d / Rfree ^e (%)	0.1782, 0.2128
R.m.s deviations	
Bond lengths (Å)	0.011
Bond angle (°)	1.295
Molprobrity score	
Clashscore ()	7.53
Poor rotamer (%)	0
Ramachandran statistics	
Favoured (%)	96.49
Outliers (%)	0.32
Overall score (%) ^{1.64}	1.64
Isotropic B factor analysis	
Average model B-factors (Å ²)	33
B-factor from Wilson plot (Å ²)	25.84

^a Throughout the table, the values in parentheses are for the outermost resolution shell.

^b $R_{\text{Sym}} = \sum_h | \hat{I}_h - I_{h,i} | / \sum_h \sum_i I_{h,i}$, where $\hat{I}_h = (1/n_h) \sum_i I_{h,i}$ and n_h is the number of times a reflection is measured.

^c $R_{\text{meas}} = [\sum_h (n_h / [n_h - 1])^{1/2} \sum_i | \hat{I}_h - I_{h,i} |] / \sum_h \sum_i I_{h,i}$, where $\hat{I}_h = (1/n_h) \sum_i I_{h,i}$ and n_h is the number of times a reflection is measured.

^d $R_{\text{cryst}} = \sum_{hkl} | |F_{\text{obs}}| - k |F_{\text{calc}}| | / \sum_{hkl} |F_{\text{obs}}|$

^e $R_{\text{free}} = \sum_{hkl \in T} | |F_{\text{obs}}| - k |F_{\text{calc}}| | / \sum_{hkl \in T} |F_{\text{obs}}|$ where T represents a test set comprising ~5% of all reflections excluded during refinement.

1.3. RESULTS

1.3.1. Protein production and purification

Once we obtained the mutant clone with the correct sequence for our interest, the plasmid is transformed on the *E.coli* BL21 (DE3) pLys, which is a specific strain for recombinant protein production. The protocol used is already described on Materials and Methods section. Once the protein is expressed, the cellular pellet was collected and lysed, and we only recovery the soluble fraction of the lysate, because we expected to find the produced protein on this fraction. The protein was purified from homogeneity by using one step purification protocol (described on Experimental procedures) (Figure 22.A). Using this production and purification system, we produced around 10 mg of pure protein for litter of cell culture (Figure 22.B).

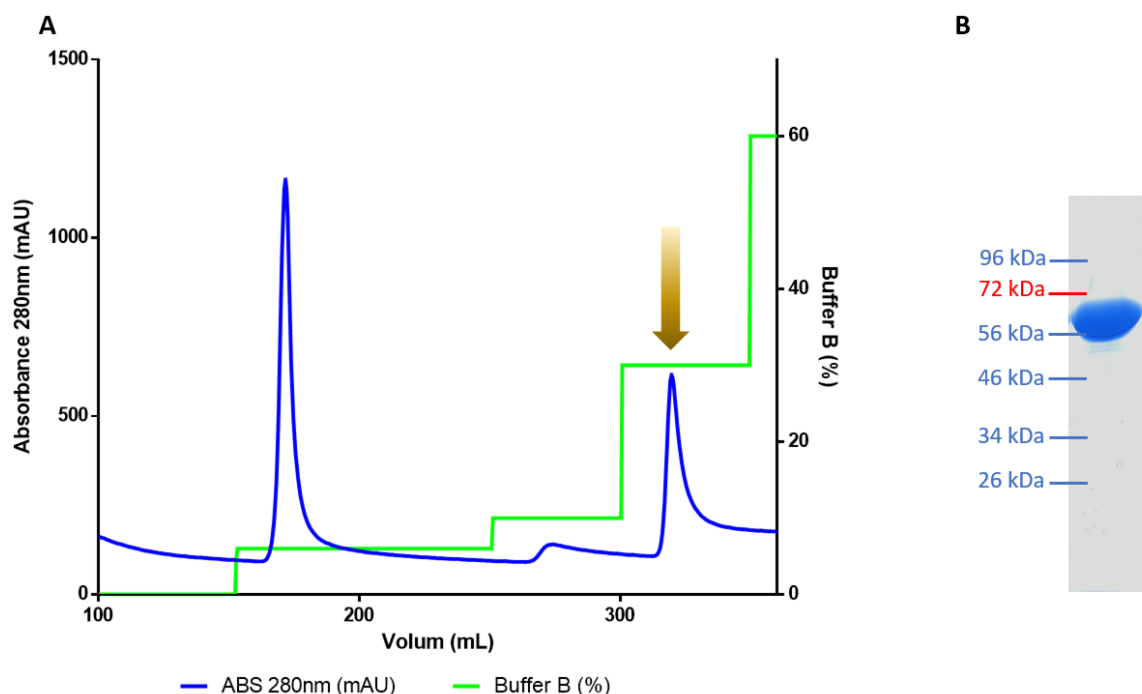


Figure 22. Recombinant production of human ALDH1A3 C314S. A) Elution profile of His-tag affinity chromatography. The peak where recombinant protein is eluted is marked by yellow arrow. B) The purified ALDH1A3 C314S was visualized on a SDS-PAGE by BlueSafe® (Comassie) staining.

On the bibliography, biologically functional ALDH1A3 *wt* is described as a homotetramer, 4 identical monomers intimately associated, and this quaternary structure was observed on the asymmetric unit of the crystal, when the ALDH1A3 structure was solved (Moretti et al. 2016). In order to confirm that our mutant also oligomerises as a tetramer, we run a Size Exclusion Chromatography (SEC), to confirm this quaternary structure. We used a Superdex 200 Increase (GE Healthcare), which allows us to separate proteins from 10.000 to 600.000 Da. The pure

protein was injected into the column, and the elution peak was on 12.1 mL which corresponds to 226.3 kDa as molecular weight (Figure 23) and the theoretical weight from the monomer of the expressed protein is 56 kDa. So, if we divide the Mr obtained from SEC between the theoretical Mr, we obtain the value of 4. This value indicates that the C314S mutant also oligomerises on a homotetramer.

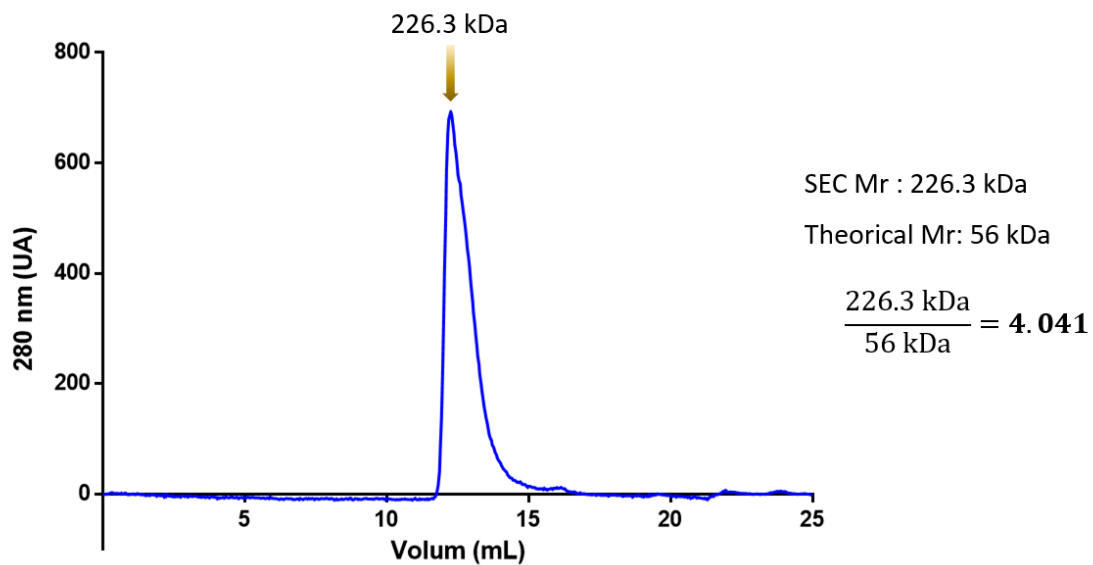


Figure 23. SEC chromatography of human ALDH1A3 C314S mutant. The recombinant protein elutes as a single peak with an apparent mass of 226.3 kDa. The expected mass of produced protein was around 56 kDa. So, the non-denaturalized recombinant protein is forming homotetramers as ALDH1A3 *wt* does too.

1.3.2. Dehydrogenase activity for ALDH1A3 C314S

In order to test the role of specific residues in determining the kinetic properties of ALDH1A3 enzyme, we performed site-directed mutagenesis of catalytical cysteine. ALDH1A3 *wt* kinetic features have already been described recently for different substrates (Pequerul et al. 2020). On this study, hexanal is described as the best substrate to evaluate its activity. The conditions to measure the activity of the enzyme are described on the section Materials and Methods. On our study of activity firstly we compared the activity of the ALDH1A3 *wt* against ALDH1A3 C314S mutant. The results shown a big difference on activity on both proteins. The standard activity of ALDH1A3 C314S mutant is only $\leq 1\%$ of the standard activity of ALDH1A3 *wt* (Figure 24). Even this protein is a new mutant, never described before, this loss of activity has already been observed on other members of ALDHs superfamily, for the same mutation (Farrés et al. 1995). The remnant activity observed by the mutant could be generated by the OH group from Serine, which could act as a nucleophile, as SH from Cysteine does on ALDH1A3 *wt*. Clearly, this possible nucleophile is lesser reactive.

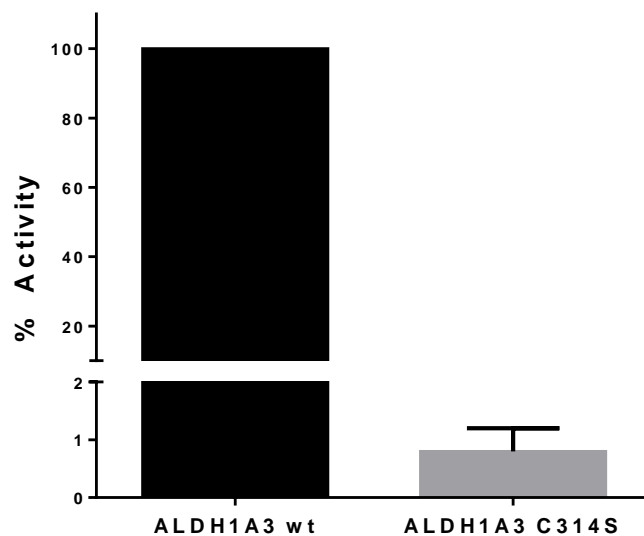


Figure 24. Graphical representation of ALDH1A3 wt and ALDH1A3 C314S mutant activities. Data is normalised on the top activity of ALDH1A3 wt.

We tested the activity on the presence of inhibitor already described for ALDH1A3. DEAB is an ALDHs inhibitor. More than ALDHs classical inhibitor, it is considered as a slow substrate. Furthermore, the ALDH inhibition by DEAB is the basis of Aldefluor™ assay, which is used to identify, evaluate, and isolate normal and cancer stem cells, that express high levels of ALDH (Moreb et al. 2012). The mechanism suggested for this inhibitor is the formation of a quinoid-like resonance state following hydride transfer that is stabilized by local structural features that exist in several of the ALDH enzymes (Morgan et al. 2015) (Supplemental figure 1). We tested the activity of the mutant, on the presence of 10 μ M of DEAB and compare its effect on wt too (Figure 25).

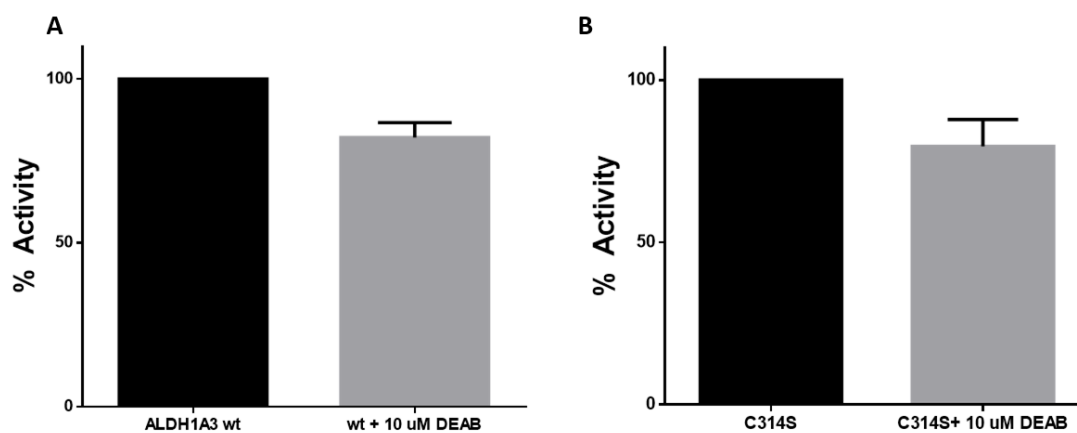


Figure 25. Graphical representation of ALDHs activities in presence or not the inhibitor DEAB. A) The enzyme used is ALDH1A3, and data normalised on its standard activity. B) The enzyme used is ALDH1A3 C314S mutant, and data is normalised on its standard activity.

We observed inhibition in both cases, on ALDH1A3 wt was expected the decrease of activity, because it was already observed and described (Morgan et al. 2015, Jiménez et al. 2019). But the inhibitory effect on the mutant was not expected. The most probably explanation is that inhibitor interacts with a cysteine adjacent to the mutated residue (C313), and by sterically effect avoids the interaction of substrate, with the nucleophile element.

In order to determine this effect on the mutant, a structural analysis is required to obtain more information. Hence, we decide to focus on the obtention of a high-resolution X-ray crystallographic structure, and proceed on soaking studies with some inhibitors, DEAB among them.

1.3.3. Resolution Structural characteristics from hALDH1A3 C314S mutant

The structure of ALDH1A3 wild type by X-ray crystallography has already been resolved at 2.9Å of resolution (Moretti et al. 2016). On our study we focused on a mutant variant of ALDH1A3, as we described previously, for a better understanding of ALDH1A3 catalytical mechanism and the resource of potential inhibitors, for their use as drugs. Our mutant protein has replaced the catalytical cysteine for a serine(C314S), which is already done on others ALDH but never before on this one (Farrés et al. 1995). We proceed to obtain the X-ray crystallographic structure, with objective to visualize how this mutation could alter the structure of the enzyme.

On the structure that we obtained, which was resolved at 1.758Å of resolution, the crystal asymmetric unit was formed by a dimer, and the homotetramer could be extrapolated by symmetry (Figure 26). We obtained a better resolution structure which allows us to compare the mutant with the structure already deposited on PDB.

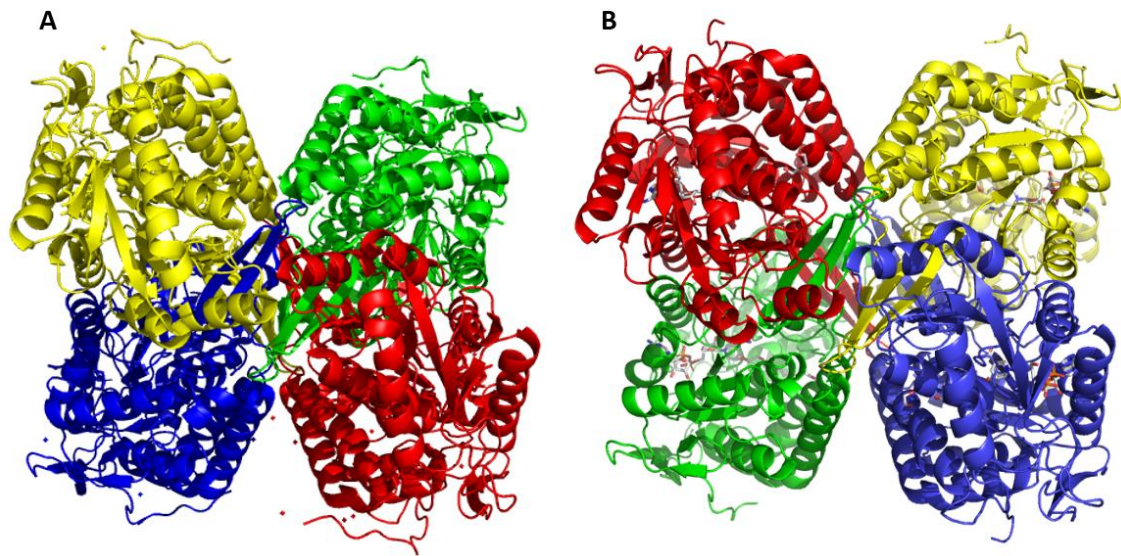


Figure 26. Cartoon representation of the tetramers of ALDH1A3 *wt* (A) and ALDH1A3 C314S mutant (B). Both are forming this homotetramer, and at quaternary level, there are not differences. Every monomer is colored with a different color. Pictures were created using PyMOL program. The ALDH1A3 *wt* structure was obtained from PDB ID 5FHZ.

Once we detected the quaternary structure of the ALDH1A3, we focus on the monomer. The monomer of ALDH1A3 mutant C314S maintains the structure described for ALDH1A3 wild type, and the rest of all ALDH classes. The structure of monomeric ALDH1A3 *wt* is composed by 13 α -helices, 19 β -strands and the connecting loops, distributed on three functional domains (Figure 27.A). The C314S mutant also shows the same elements and the three differentiated domains (Figure 27.B), the NAD binding domain localized on N-terminal, followed by the catalytic domain and the oligomerization domain, localized on the C-terminal.

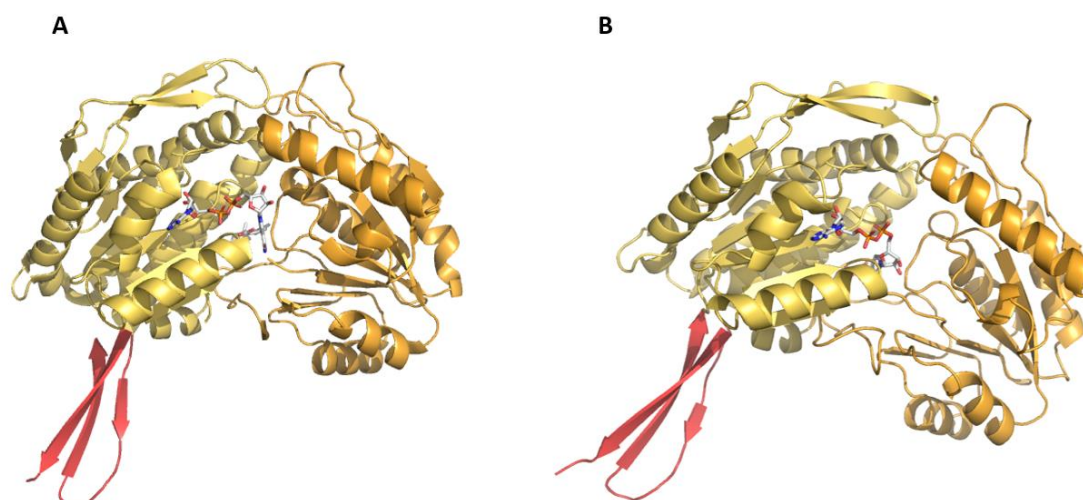


Figure 27. Stereoview comparing the monomer subunit of ALDH1A3 wt (A) and ALDH1A3 C314S mutant (B). In both structures, the NAD⁺ binding domain is coloured in light yellow; the catalytic domain is represented on bright orange and the oligomerization domain is coloured in red. Images were created by PyMOL program. ALDH1A3 wt structure was obtained from PDB ID 5FHZ

On the cofactor interaction region, our solved structure presents a better resolution than the deposited allowing us to define interactions with the cofactor, not described previously. The adenine, the adenine ribose and the two phosphate groups show the same conformation in both monomers, but the nicotinamide and the nicotinamide ribose are observed in two different conformations between monomers. Even their differences on conformations, the structure and disposition on both monomers are close, this happen for the disruption on electronic map. On this description we focus on monomer A, because we considered as it is better resolved.

As already described, the number of stabilizing interactions sustains the NAD⁺ recognition indicating that the cofactor binds strongly to human ALDH1A3 as also suggested by the increased enzyme stability when it is complexed with NAD⁺ (Moretti et al. 2016). In particular, the adenine base is held in place by van der Waals contacts with Gly 237, Pro 238, Gly241 and Ala242. The interaction between enzyme cofactor also are composed by hydrogen bonds. The adenine ribose presents hydrogen bond between its O2 β and the Lys204 N ζ atom (3.0 Å) and Glu207Oe2 (2.7Å). We also observed another hydrogen bond for the adenine ribose, between O3 β and the Thr178 O atom (3.0Å) (Figure 28).

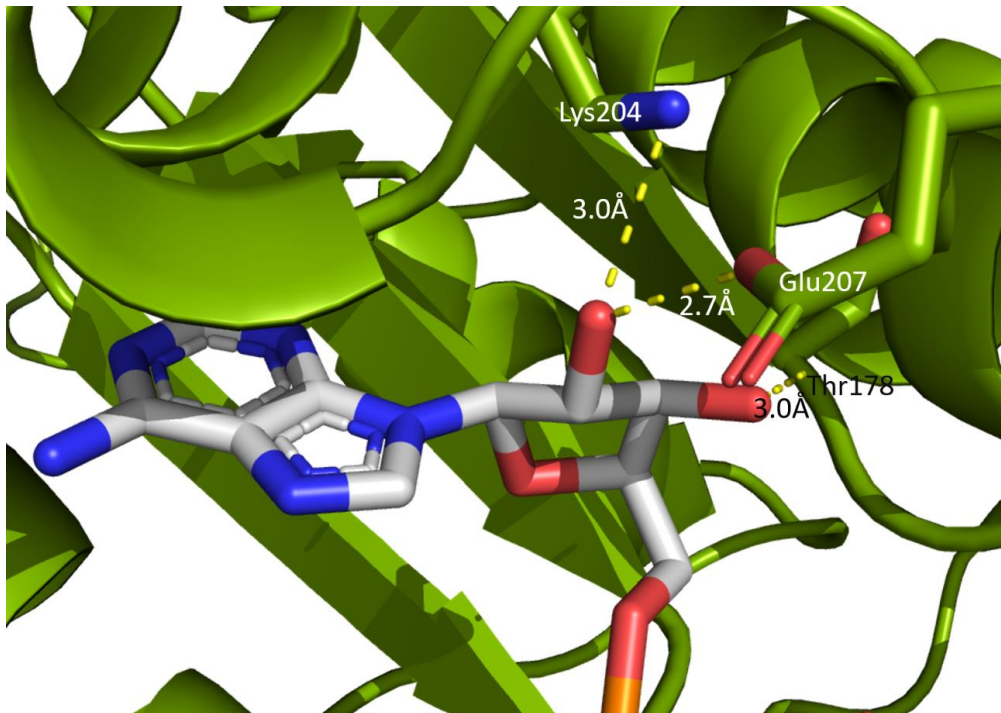


Figure 28. Graphic representations of hydrogen bonds of adenine ribose ring from NAD⁺. In green is represented the protein, in red the oxygen atoms, in white the carbon atoms and in blue the nitrogen atoms. The hydrogen bonds are represented by yellow discontinued lines, with their distances.

The first phosphate is stabilized by three hydrogen bonds, two are made by its O2 α oxygen with the O γ of Ser258 (2.7Å) and the N of Ser258 (2.8Å). On the deposited structure from the wt the interaction between Ser258 and NAD⁺ was already described, but it only presents one hydrogen bond. The third hydrogen bond is shared for both phosphates, because it implies the O3 from NAD⁺, and the interaction is with the N of Ser258 (3.3Å). The second phosphate from cofactor also presents a hydrogen bond to let the cofactor stabilized, the bond is formed by its O1 η and the N ϵ 1 of Trp180 (3.4Å) (Figure 29).

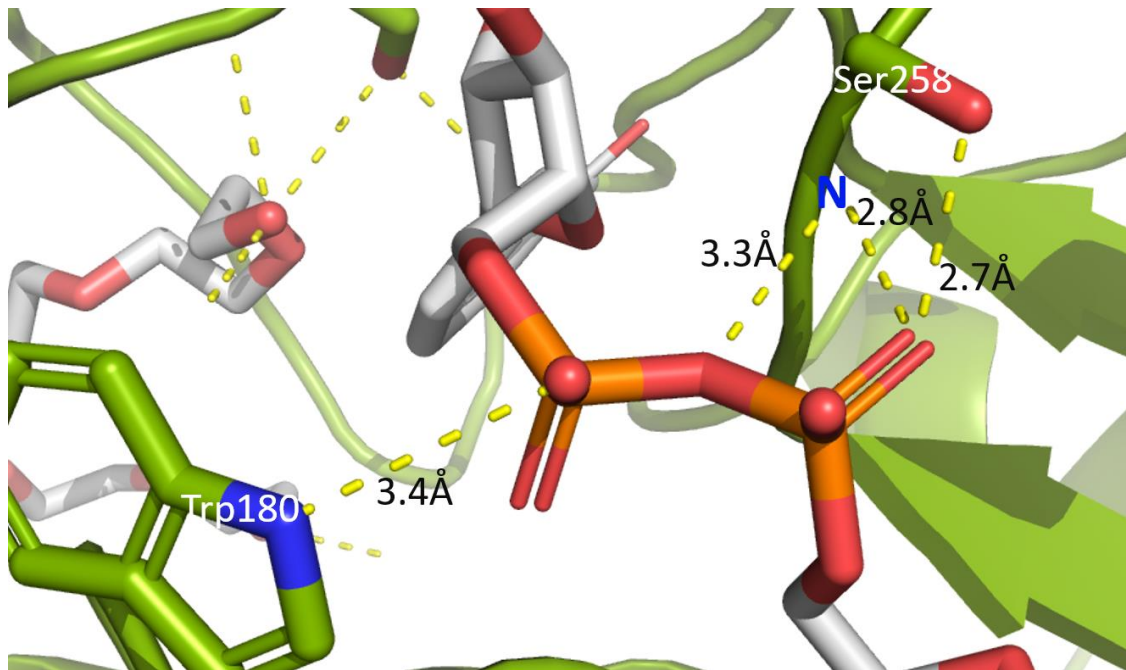


Figure 29. Graphic representations of hydrogen bonds of central phosphate groups from NAD⁺. In green is represented the protein, in red the oxygen atoms, in white the carbon atoms, in blue the nitrogen atoms and orange are phosphate atoms. The hydrogen bonds are represented by yellow discontinued lines, with their distances.

On our structure, we report the nicotinamide group from cofactor is oriented toward the catalytic site as is described in bibliography (Moretti et al. 2016), adopting the hydride transfer position. This position reveals that the C4 of the nicotinamide ring (hydrogen acceptor) has a distance of 3.4 Å from the O_γ of the Serine on 314 position, which substitutes the S_γ from the catalytic cysteine on ALDH1A3 *wt*. In this conformation, the nicotinamide ribose oxygens are stabilized by hydrogen bonds. In particular, the O2 δ contacts the O ϵ 2 of Glu411 (3.3 Å) and the O3 δ engages the O ϵ 1 (3.5 Å) of Glu411. Finally, the last interaction which is observed, and obviously it cannot be observed on the *wt*, is an hydrogen bond between N7 η from the NAD⁺ and the O δ from Ser314 of the mutant (Figure 30).

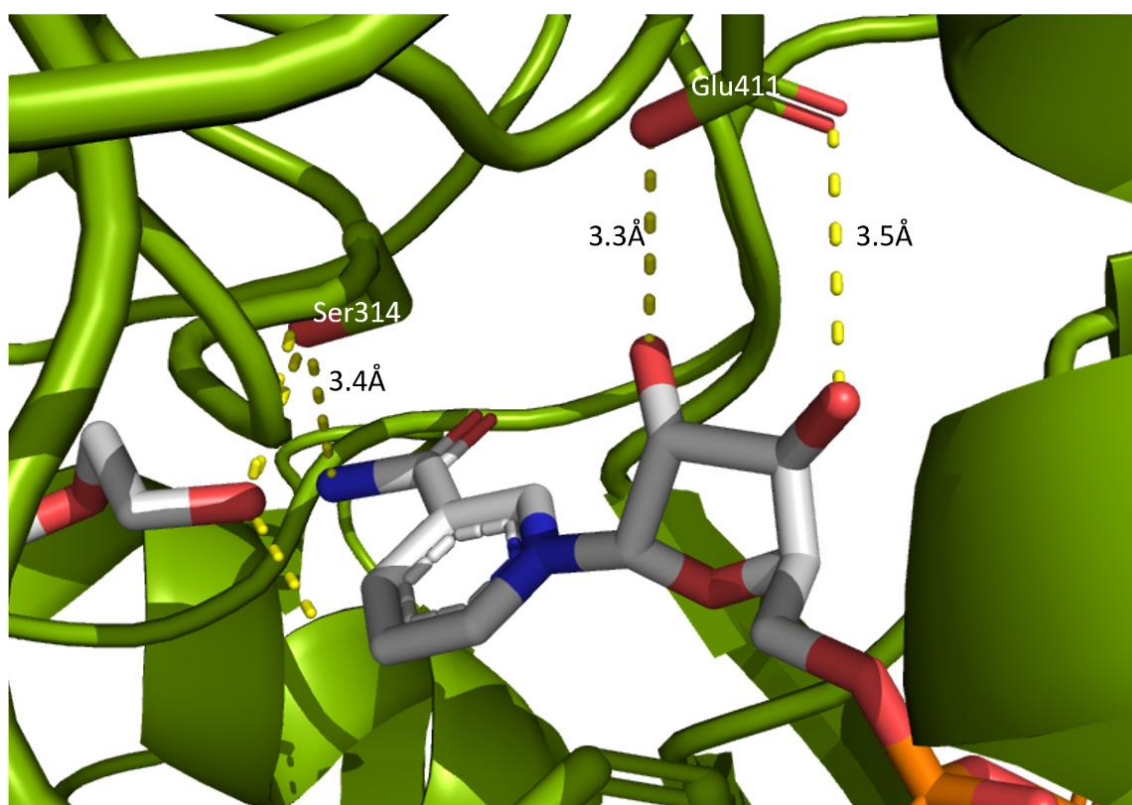


Figure 30. Graphic representations of hydrogen bonds of Nicotinamide ring from NAD⁺. In green is represented the protein, in red the oxygen atoms, in white the carbon atoms, in blue the nitrogen atoms and orange are phosphate atoms. The hydrogen bonds are represented by yellow discontinued lines, with their distances.

To sum up all previous described interactions, we present you the following table (Table 3).

TABLE 3. Resume interaction by hydrogen bonds (distances) (Å)

ALDH NAD ⁺	Thr178 O	Trp180 Nε1	Lys204N z	Glu207 Oε2	Ser258 N	Ser258 Oγ	Ser314 Oγ	Glu411 Oε1	Glu411 Oε2
O3					3.3				
O2α					2.8	2.7			
O1η		3.4							
O2δ									3.3
O3δ								3.5	
N7η							3.4		
O2β			3.0	2.7					
O3β	3.0								

Finally, we also have observed the substrate pocket and compared with the deposited ALDH1A3 structure. The natural substrate of this kind of enzyme, as is described on bibliography, is the all-trans-retinal. Its degradation to acid retinoic is vital for organism survival. In our study we crystallized the C314S mutant with cofactor only and not with substrate in order to simplify

the experiment. In the described and deposited structure the molecule fitted on the catalytical pocket is the product, retinoic acid (Figure 34.A), but in structure we did not dare to try to modelate the all trans retinal because the electronic map on the substrate space was not clear, and we only let two molecules of low weight PEG (Figure 34.B). The PEG modelled is even a lower than the used for crystallization, this could be happening because the PEG used is not pure or could even been degraded.

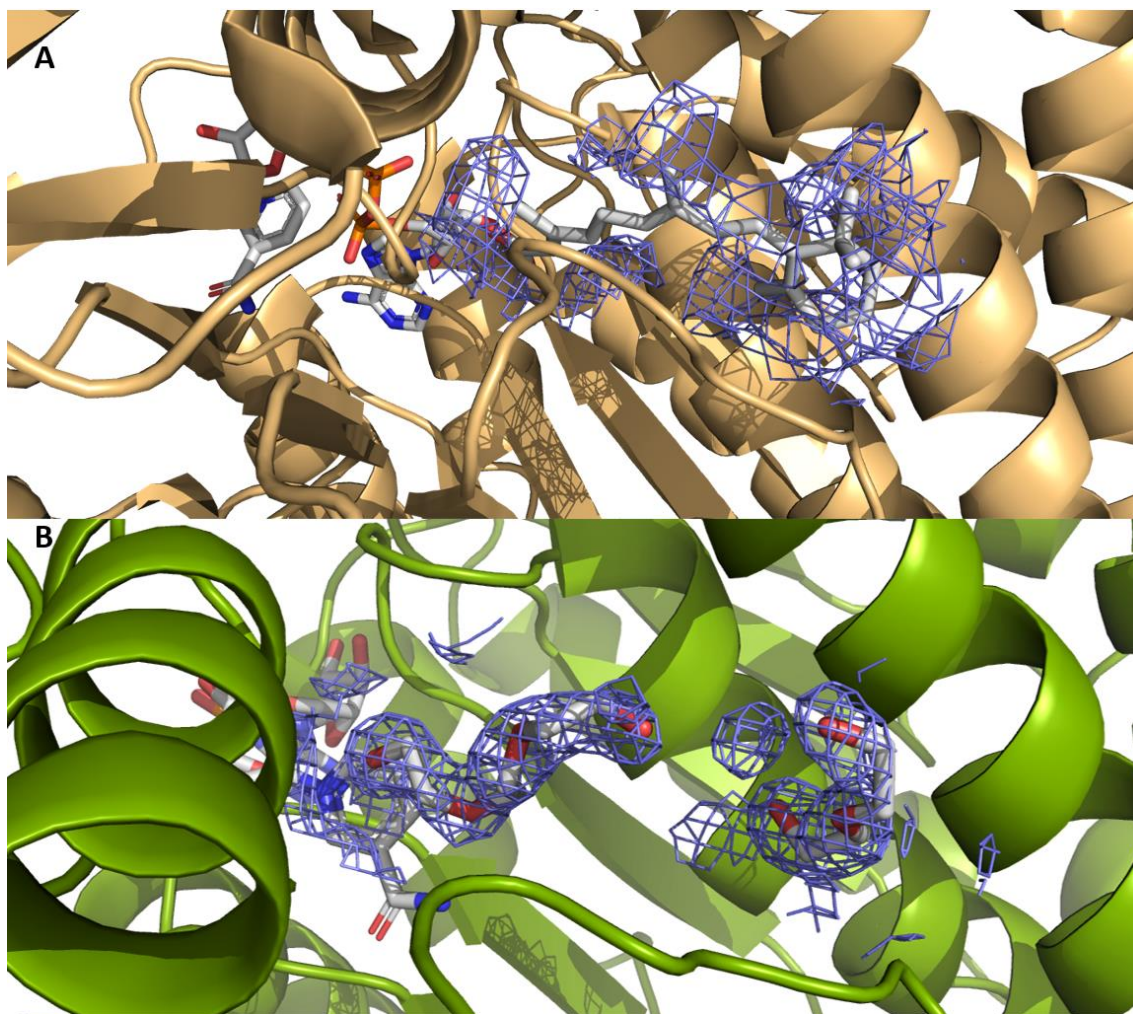


Figure 31. Electronic map of the molecules localized on the substrate pockets at 1.0 sigma. A) ALDH1A3 substrate is represented as the white molecule, and with blue colour is represented the electronic map of the substrate molecule. As it could be observed the electronic map of the substrate is not really adjusted to the molecule at all. B) ALDH1A3 C314S mutant substrate pocket, on white is represented the molecules which represents the substrate and with blue colour is represented the electronic map of these substrate molecule. As we described, we did not fit the retinoic acid or all-*trans*-retinal molecule because we only could fit two molecules of PEG.

1.4. DISCUSSION

The ALDH1A3 is an enzyme which is involved on different biological roles, some of them involved on cancer phenotypes. So, this marks the protein as target for study on its kinetics and structure, and how different inhibitors could act as potential drugs for the treatment of ALDHs related pathologies.

In the process to obtain the mutant, we found some troubles on the one-site directed mutagenesis because the catalytical region is codified by a huge number of guanines and cytosines on its gene. Fortunately, only one copy of the gene mutated is required to produce more quantity of the plasmid with the mutated gene.

We produced a low activity mutant variant of ALDH1A3. The substitution of Cysteine by a Serine on the catalytical residue position, causes a reduction of 99% of total activity if we compare with the *wt* activity. The remnant activity observed could be caused by the Serine and its OH group on the side chain. The oxygen can act as the nucleophilic element for the reaction, in a lesser level than the sulphur from Cysteine, on the *wt*, does. We tried to characterize the kinetic parameters of this mutants, but its low standard activity causes, that the measures of their activity at different concentrations of substrates by fluorescence requires higher concentrations of substrate. Furthermore, it exists technical limitations due the measures of fluorescence.

We also studied the activity of the mutant against a known inhibitor, DEAB. After we understood the suggested mechanism, we supposed that the inhibitor would not present any effect on the C314S mutant. Unfortunately, we observed a comparable inhibitory of DEAB on C314S mutant as *wt*. After we observed this inhibitory effect, we supposed that the inhibitor is binding, using the same mechanism reported (Morgan et al. 2015), but on the adjacent Cysteine (C313) to the catalytical residue. This potential bond should cause stereochemical problems for the interaction between the substrate and the nucleophile. This causes that we realized the need to obtain the structure of the enzyme bond to the inhibitor to characterizes what is happening.

The production of the mutant has allowed us not only to study their activity, also it could allow us to deepen on the ALDH1A3 monomer structure, concretely on the interaction of cofactor NAD⁺. We obtained a good resolution structure (1.758Å) which has allowed us to describe more hydrogen bonds between the NAD⁺ and the enzyme, than already are described by ALDH1A3 deposited structure. Some of these newfound hydrogen bonds coincides and gives support to

the already described interaction between NAD⁺ cofactor and NAD⁺ cofactor binding domain on others ALDH, as ALDH2 (model one) (Ni et al. 1997, Sheikh et al. 1997) (Table 4)

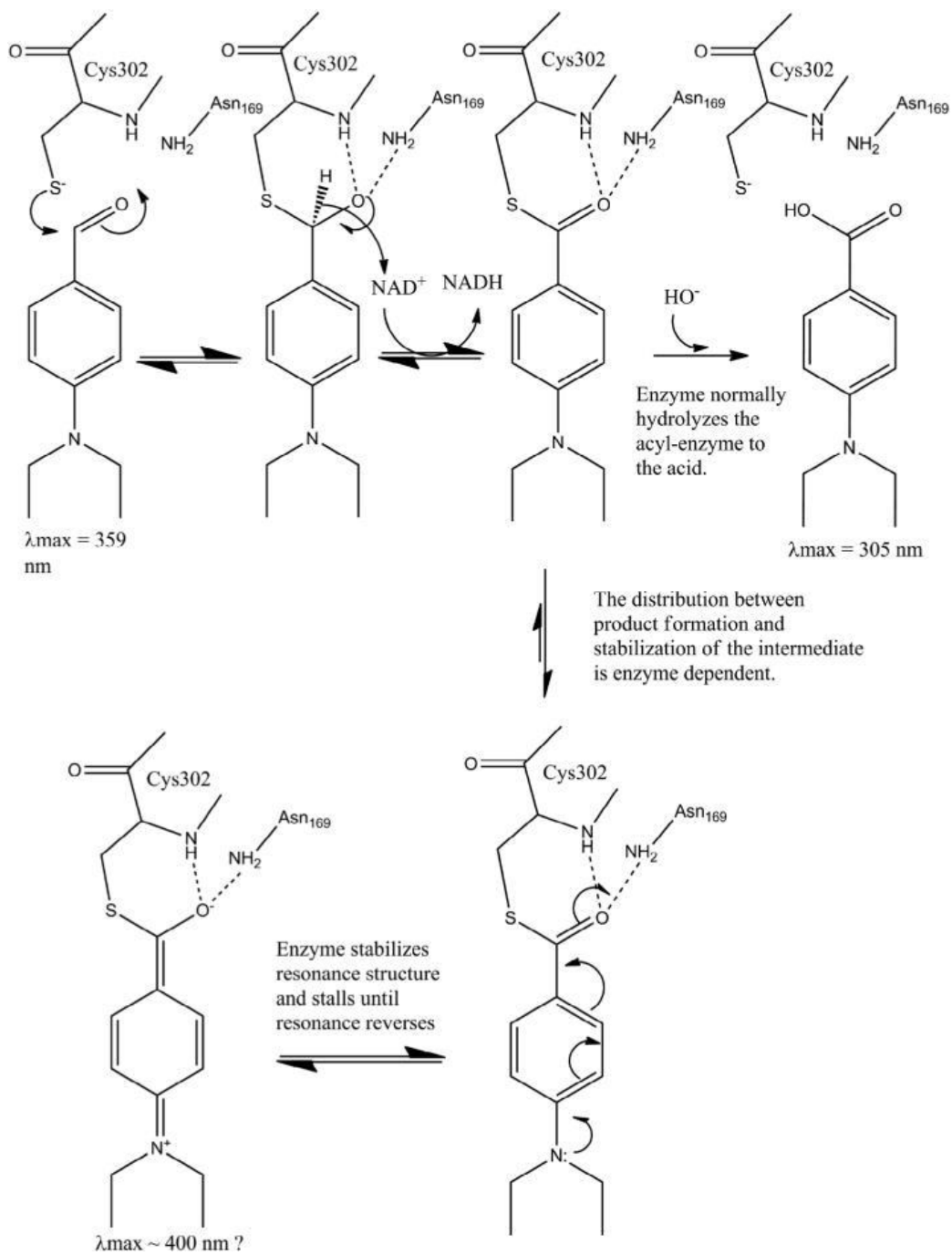
TABLE 4: Residues conserved on NAD⁺ hydrogen bounding between ALDH2 and the mutant C314S

ALDH2	ALDH1A3 C314S	Where interacts	Conserved
Lys192	Lys204	Adenine ribose	Yes
Glu195	Glu207	Adenine ribose	Yes
Ile166	Thr178	Adenine ribose	No
Glu399	Glu411	Nicotinamide ribose	Yes

One of the advantages, that we found on the resolution of this structure, is its high stability. Possibly, cysteine substitution by serine, making lesser reactive enzyme, has made that the structure be more stable and compacted. This both properties made that resolution of the structure became good. Furthermore, it is important to say that the obtained structure for us is not deposited on the PDB yet, which means that even can be refined, although the structure is not changing a lot.

The obtention of this structure is a previous work for future soaking studies, which are already carrying on about how different inhibitors, among them DEAB, interacts with ALDH1A3 and describe their mechanism of inhibition and describe its specificity.

1.5. SUPPLEMENTARY DATA



Supplementary Figure 1. Suggested mechanism of inhibition by DEAB to ALDH2 (ALDH model). Reproduced from (Morgan et al. 2015).

CHAPTER II

EFFICIENT AND FACILE PRODUCTION OF RECOMBINANT CARBOXYPEPTIDASES IN MAMMALIAN CELLS

CHAPTER II: EFFICIENT AND FACILE PRODUCTION OF RECOMBINANT CARBOXYPEPTIDASES IN MAMMALIAN CELLS

2.1. INTRODUCTION

The rapid progress of biotechnology and biomedicine has created a constant need to produce a variety of complex recombinant proteins for different purposes. Their production often is linked to post-translational modifications and the use of molecular chaperones and co-factors to obtain a completely functional protein. To overcome these limitations is well-known the use of mammalian expression systems because they are able to produce complex proteins with post-translational modifications which could be problematic for other recombinant protein production systems.

In the last years, transient expression in mammalian cells has been a useful tool to overcome folding and post-translational modifications problems from others expression systems and produce research quantities of proteins. The system is based on expression in the fast growing suspension cell line Freestyle™ 293-F, and transfection of the mammalian expression plasmid with polyethylenimine (PEI) (Longo et al. 2013, Vink et al. 2014). PEI binds DNA forming positively charged complexes, and the complexes interact to anionic cell surface. Subsequently, complex is endocytosed and the DNA released into the cytoplasm (Boussif et al. 1995, Sonawane et al. 2003).

Metalloproteases (MCPs) are structurally complex exopeptidases that cleave C-terminal residues from its substrates. In the last years, MCPs have been described as promising drug targets in biomedicine because members of this populated group of proteases have been associated with different biological roles, which range from food digestion to the precise control of neuropeptides activity. (Arolas et al. 2007). MCPs synthesis often requires post-translational modifications, molecular chaperones and co-factors to obtain a properly folding and enzymatic activity. Furthermore, some MCPs present regions that are rich in basic residues allowing them to bind to heparin by electrostatic interactions. We find an example in the case of the highly unstable MCP Thrombin-activable fibrinolysis inhibitor (TAFI) that the structural works demonstrated that possess an instability region that is clearly stabilized by heparin binding, and this region is rich on basic residues (Sanglas et al. 2008) or human carboxypeptidase A6 (hCPA6), which is MCP bonded to extracellular matrix (ECM) and whose attachment to ECM is blocked by presence of heparin, letting the hCPA6 soluble in cell culture medium. Modelling of hCPA6 shows that the majority of basic residues are located on the face of the molecule opposite that of the

active site, suggesting that this may orient CPA6 with respect to the ECM and the region where heparin might bind the MCP (Lyons et al. 2008).

Heparin is a highly sulphated polysaminoglycan that is used as anticoagulant reagent to maintain the blood flow in medicine among other applications. Structurally, heparin is composed by α 1-4 linked disaccharide repeating units (Figure 32). The heparin and its derived molecule Heparan Sulphate (HS) are used to be associated with cations, due to they are negatively charged, in solutions and could play a role in metal ion sequestration and transfer (Heidarieh et al. 2013). Heparin also interacts with huge number of proteins, from soluble proteins until extracellular signalling molecules as Epithelial Growth Factor (EGF), even membrane-binding proteins.

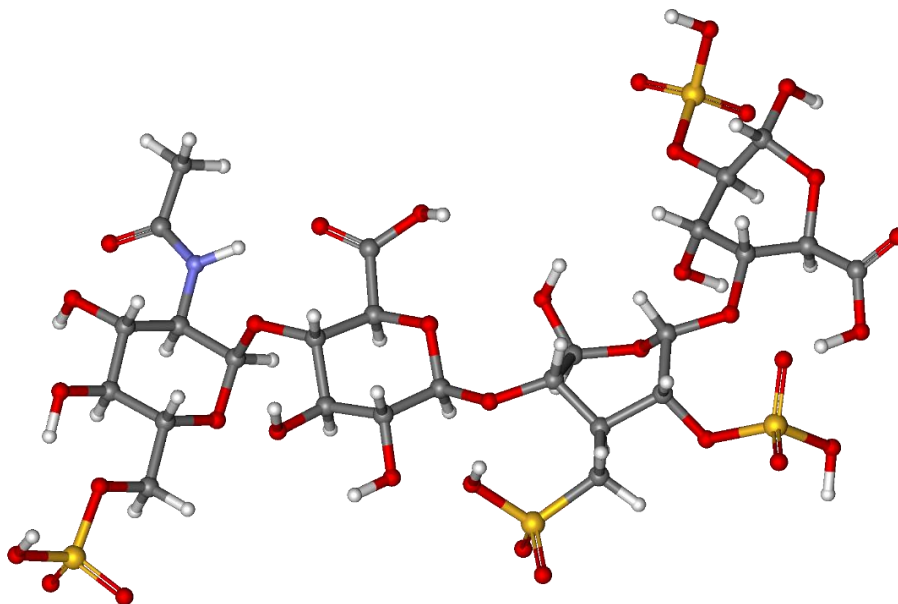


Figure 32: Balls and sticks representation of Heparin molecule. Negative charge of molecule comes from sulphate groups represented as gold and red balls in the picture.

In this chapter we describe a simple and non-expensive method to optimize heparin affinity MCPs production in suspension growing mammalian cells by addition of sodium heparin in the cell culture media. The heparin addition in the culture medium allows the stabilization of the heparin-affinity proteins and they are accumulated in the extracellular media in a bigger proportion than without heparin. In this work, we tested different MCPs which have heparin affinity regions, deduced from its pI. These MCPs are human carboxypeptidase Z (hCPZ) (Novikova et al. 2000), human carboxypeptidase A6 (hCPA6) (Lyons et al. 2008), human thrombin activation fibrinolysis inhibitor (TAFI) (Sanglas et al. 2008). In addition, we tested other

MCPs, which do not present these heparin affinity regions, because they do not show a basic pI. These MCPs are human carboxypeptidase O (hCPO) and human carboxypeptidase D (hCPD). Both proteins have a transmembrane domain, which was truncated in their expression. In addition, we tested two more proteins from other families: Beta-galactoside alpha-2,6-sialyltransferase 1 (ST6) and alpha galactosidase (α -Gal) as positive and negative controls respectively.

2.2. EXPERIMENTAL PROCEDURES

2.2.1. Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

The methodology used is the same process described in Experimental Procedures in Chapter I. Only differs on the percentage of polyacrylamide used.

2.2.2. Western blot

Western blot or protein immunoblot is an analytical technique in molecular biology used to detect and quantify specific proteins in a sample. This technique is based on the ability of antibodies to specifically attach to an antigen, or epitopes of proteins in that case.

Western blot technique is combined with SDS-PAGE because high density of acrylamide prevents from antibodies interacting electrophoretic separated proteins. In order to solve this problem, proteins are transferred from polyacrylamide gel into membranes with high affinity for proteins (nitrocellulose or polyvinylidene difluoride, PVDF). (Towbin et al. 1979, Eisenstein 2005). Transfers were run in an electrophoresis chamber with Transfer Buffer (25mM Tris-HCl, 192mM Glycine and 10% Methanol) at constant voltage of 100 V (perpendicular to the gel). Time of transference depends of protein size, electrophoretic gel thickness and polyacrylamide percentage.

After the proteins are transferred, membranes were equilibrated with Phosphate-buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄) with 0.05% v/v of Tween[®] 20 Sigma-Aldrich, Merck, under mild shaking. In order to avoid unspecific interaction of antibodies with PVDF membranes, they were blocked through an incubation with 5% solution of non-fat dry milk (Central Lechera Austuriana) in PBS-Tween[®] for 1 hour at room temperature under mild shaking.

Once are blocked, membranes were washed 4 times with PBS-Tween[®] for 10 minutes under shaking. Target proteins were detected using primary antibodies which recognize some protein

regions or recombinant added tags to protein, diluted in 2% solution of non-fat dry milk in PBS-Tween®. Incubations were realized 2 hours at room temperature or overnight at 4°C under mild shaking.

After primary antibodies incubation, they were recovered, and membranes were washed 4 times for 10 minutes with PBS-Tween®. HRP-labelled secondary antibodies (they recognize heavy chain from primary antibodies) were freshly diluted 1:5000 in a 2% of non-fat dry milk in PBS-Tween® and incubated 2 hours at room temperature. Membranes were washed 4 times with PBS-Tween® for 10 minutes under mild shaking. Proteins bands were revealed by chemiluminescence using as substrate Luminata Forte (EMD-Millipore). Images were obtained by Molecular Imager Versadoc MP4000 system (BioRad).

2.2.3. DNA Techniques

2.2.3.1. Plasmid production and extraction

In order to obtain more quantity of DNA plasmid, we use the stain XL1-Blue from Escherichia coli for preparation of high-quality DNA plasmid. XL1-Blue was transformed following the protocol:

1. 0.1 ng from the original DNA plasmid stock was mixed with one aliquot of 50 µL of XL1-Blue at 0.45 of OD600. The mix was incubated 30 minutes in ice.
2. After that, thermal shock was realized at 42°C for 1 minute and then the mixture is incubated on ice 10 minutes more.
3. Already incubation with ice was finished, 700 µL of liquid Luria-Bertani (LB) medium was added to DNA/cells mixture. This mixture was incubated at 37°C on a rotatory shaker at 220-250 rpm for 40 minutes.
4. Subsequently this incubation, aliquot was centrifuged and 700 µL of medium were discarded. The pellet was resuspended in 50 µL of medium. This cellular resuspension was seeded in a LB plate with an antibiotic to force a positive selection, and the plate was incubated in a heater at 37°C for 16 hours.

After this last incubation, some colonies with DNA plasmid were grown in the plate. One colony was recollected and seeded in 5 mL of LB with antibiotic to prepare a pre-inoculum for a bigger volume of culture. When the pre-inoculum was grown, it was seeded to a volume of LB with antibiotic (between 50 mL to 250 mL) and it was incubated at 37°C during necessary time to obtain optimal biological mass, for induction or extraction of plasmid DNA.

For DNA extraction using volumes of culture from 50mL up to 250mL, we use NZYMidiprep kit from NZYTech. The result is high-concentrated and high-quality DNA plasmid ready to transfect or transform.

2.2.4. Cell Culture

2.2.4.1. Cell lines and maintenance

The cell line used was Human Embryonic Kidney 293-F (HEK 293F), which is derived from HEK 293-T cell line but was adapted to grow in suspension for an easier growth to large scales and easier recovery of conditioned media. The recovery of conditioned media is important on our project cause the proteins expressed are cloned inside pTriex7 on phase to mouse IgM secretion signal sequence (Figure 33). This allows the protein produced inside cell being secreted on conditioned media (Kurosawa et al. 1994, Kim et al. 2003) and it can be easy recovered for being analysed on our study.

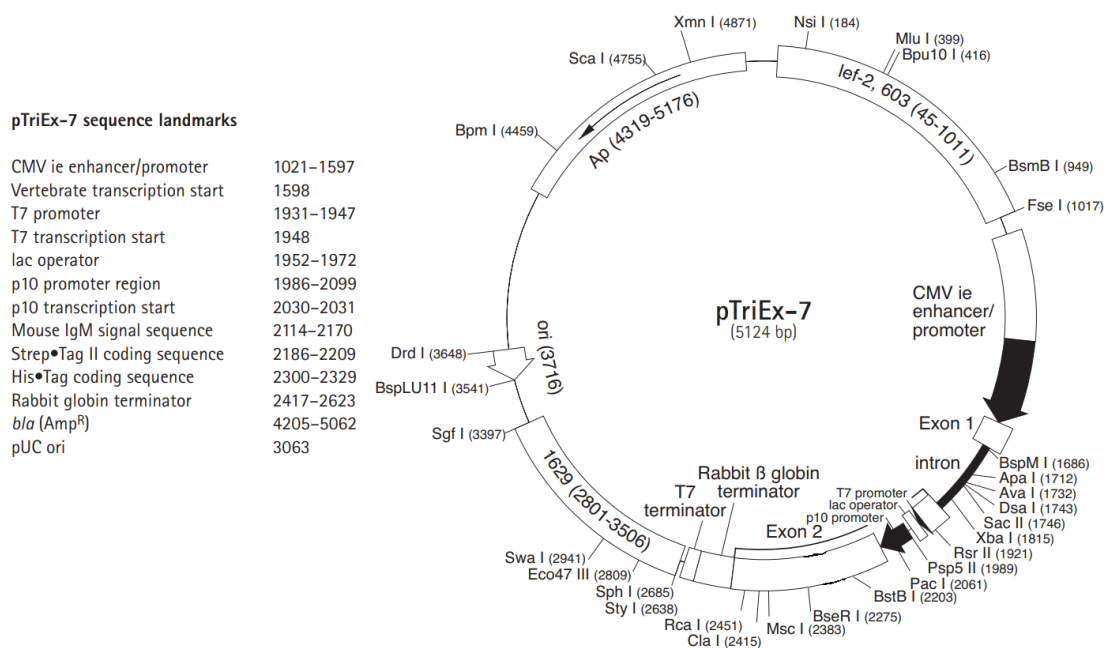


Figure 33. pTriex7 DNA plasmid map. This map shows the different elements that forms pTriex7 plasmid. Figure reproduced from Novagen, Merck Millipore, Inc.

The medium where cells were grown in Freestyle 293 expression medium from Life Technologies, Thermo Scientific. Cells were grown in flask on an orbital shaker at 37°C, 8% CO₂, 70% humidity and 135 rpm. Its maintenance is made every 48h by dilution of its density up to 0.2·10⁶ - 1.0·10⁶ cells/mL.

2.2.4.2. Cytotoxicity assay

Cytotoxicity assays were realized in order to determine HEK 293F cell viability and cell proliferation with the presence of sodium heparin in media. The molecule which is used for these assays is 2,3-Bis-(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide salt (XTT), whose reduction produces a colorimetric change in solution and this change is the signal measured. The reduction is produced by active reductases from living cells, so this process is a good marker of cell metabolic activity, and consequently cell viability (Mosmann 1983, Cory et al. 1991, Stevens and Olsen 1993).

Cells were seed at $3.0 \cdot 10^3$ cells/well of cell density in 96-well plates and incubated 24 hours at 37°C. After this incubation, sodium heparin is added at different concentration, from 0 to 2000 UI/mL. Growth inhibitory effect of sodium heparin was measured after 24 and 72 hours of incubation from heparin addition. Aliquots of 20 μ L of XTT solution were, added in each well and after 4 hours of incubation, the colorimetric change was quantified by a spectrophotometric plate reader (Perkin Elmer Victor3 V) at 490nm. Cell viability in treated cultures was evaluated and expressed as percentage of control condition (Núñez et al. 2014).

2.2.4.3. Transfection process

Transfection is known as the process to include a plasmid DNA inside mammalian cells. The transfection could be done as stable or as transient, they both differ in their time effects. The stable transition is for a long-term expression from transfected DNA in cell, and this DNA is passed on to daughter cells. Meanwhile, in transient transfection, the cell will express the DNA for short period of time and this DNA will not be passed to daughter cells. In this thesis, we focused on transient transfection to achieve the goals proposed.

2.2.4.3.1. Transient transfection protocol

1. One day prior to transfection, HEK 293F cells are seeded at $0.5 \cdot 10^6$ cells/mL into a final volume of 18 mL in a 125mL flask. It is recommended that cell culture should be accommodate as a 1/5 of total volume of culture flask. After 24h of incubation in an orbital shaker incubator at 37°C, 135 rpm, and 8% CO₂, the cell culture should reach a concentration of $1.0 \cdot 10^6$ cells/mL.

Prepare transfection complexes by mixing 20ug of DNA into 2 mL of Freestyle 293 fresh medium and vigorously vortexed for 30 seconds (1 μ g of DNA per millilitre of cell culture).

After, add to DNA/medium mix, 60ug of Transporter 5™ (commercial prepared PEI MAX at 1 mg/mL) and vortex vigorously for 30 seconds more.

2. Incubate the DNA/PEI/medium mix 20-25 minutes at room temperature, during this time DNA/PEI complexes are formed. After incubation of the mix, it is slowly added to the cells, and the cells are incubated in an orbital shaker incubator at 37°C, 135 rpm and 8% CO₂.
3. After 48h from transfection, 20µL of sterile sodium heparin solution at 5000 IU/mL are added to the cell culture medium.
4. After heparin sodium addition, cells are incubated as long as required (maximum 8 days) at 37°C, 135 rpm and 8% CO₂, in an orbital shaker, in order to let cells produce recombinant proteins.

A sample of the expression process is periodically collected, for its analysis by western blot (Figure 34).

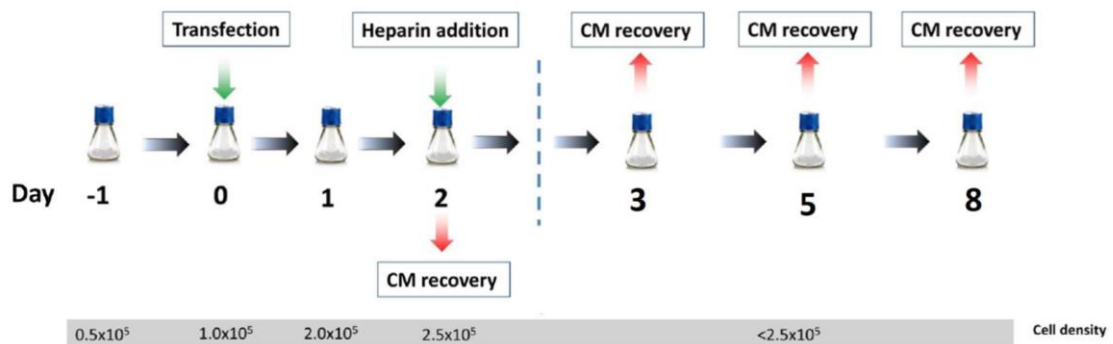


Figure 34. Schematic representation of the procedure used for production of Heparin affinity proteins. All the proteins were tested with the same protocol and the same days for no differences between proteins.

2.2.5. Data evaluation

All the data obtained was evaluated in two-way analysis of variance (ANOVA). For this case, the two independent variables are time post transfection and presence of sodium heparin in extracellular media, and the dependent variable is relative expression levels of each protein.

2.3. RESULTS

2.3.1. Heparin affinity proteins

Heparin could bind a large number of proteins, and this group englobes from cytokines until membrane receptors, including enzymes and extracellular matrix proteins among others. They all are secreted proteins. This kind of proteins interacts with the Heparan sulphate chains from membrane and extracellular proteoglycans.

To define if a protein can bind heparin or not, we focus on the isoelectrical points (pI) of the proteins. It is a widely known that the best method to identify a heparin binding protein is to calculate the surface electrostatic potential. In the heparin-protein interaction, electrostatics is perhaps one of the most important role, so the protein that interacts heparin must have a positively charged surface, and this causes that protein has a basic isoelectric point.

If we apply this concept to MCPs, we can suggest that MCPs with basic IP could bind to heparin. In fact, on our research group we already observed the ability of hTAFI to bind to heparin, and their interaction allowed to stabilize hTAFI (Sanglas et al. 2008). This fact has suggested us a new hypothesis, if we add heparin to the cell culture media during the recombinant production of basic IP carboxypeptidase, heparin will bind to a new synthesized protein, helping in its folding, solubility and stability. This facile and cheap strategy will be useful for our research on MCPs because these enzymes are complex and some of them are difficult to produce. Then, the addition of heparin in the cell culture medium, where is protein secreted due to the IgM exporting signal, should increase the levels of soluble protein in the extracellular media, if this one has a basic IP. The MCPs that have a basic IP are CPA3, CPA6, CPZ and TAFI. (Table 5).

In this study, we included two proteins as positive and negative controls. Beta-galactoside alpha-2,6-sialyltransferase 1 (ST6) a membrane protein, that is responsible to process sialic acid transference from CMP-sialic acid to galactose-containing substrates. ST6 is used as positive control due its capacity of interact with Heparin and Heparan sulphate molecules (Kuhn et al. 2013) and it also has a basic IP. As negative control we include alpha-galactosidase (α -Gal), it is a glycoside hydrolase enzyme which catalyses the removal of the terminal α -galactose from polysaccharides (Scriver et al. 2001). This protein is localized inside the lysosomes at intracellular level. This protein shows acidic IP and non-interaction with heparin is described in bibliography.

TABLE 5. Properties of the different proteins used in this work.

	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 (CPA3)	Extracellular	9.2/9.5	Unknown
	Q9UI42	Carboxypeptidase A4	Extracellular	6.1/7.1	Unknown
	Q8WXQ2	Carboxypeptidase A5		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	YES
	Q8IUX7	Adipocyte enhancer- binding protein1	Cytosolic	5.0	Unknown
	Q96SM3	Carboxypeptidase X1	Cytosolic	6.2	Unknown
	Q8N436	Carboxypeptidase- like protein X2	Cytosolic	6.4	Unknown
M14D Subfamily	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
Controls	P06280	Alpha-galactosidase (α -Gal)	Lisosomal (intracellular)	5.2	NO
	P15907	Beta-galactoside alpha-2,6- sialyltransferase 1 (ST6)	Golgi Aparatus (Lumen)	9.1	YES

*IPvalues were calculated with ProtoPara tool from ExPasy (<http://webexpasy.org/protparam/>).

The first value indicates the IP of proform, while the second value indicates the IP of the active form.

Adapted from (Garcia-Pardo 2015)

As it is reported for the different MCPs which we focused on this project, the heparin binding regions are surface areas rich on basic residues, what generates a basic IP on the MCPs. For this reason, we have studied the different protein surfaces and their charge distribution to deepen on the interaction between protein and heparin in each protein. Some of MCPs studied do not have a resolved structure so we must make a model to obtain the information. What we observed on this study is that as expected the protein reported as heparin affinity protein show a bigger number of positively charged patches than negatively charged in their surfaces. These positively charged patches are described as the regions where heparin might be bonded and are showed in the figure 35 in blue color (Lyons et al. 2008, Sanglas et al. 2008). We can observe these regions in hCPA6, TAFI, hCPZ and ST6 proteins (Figure 35). Otherwise, the other proteins involved in the project, do not have a basic IP and show a homogeneity in the distribution in their surface charges, and therefore we consider that they do not have heparin affinity.

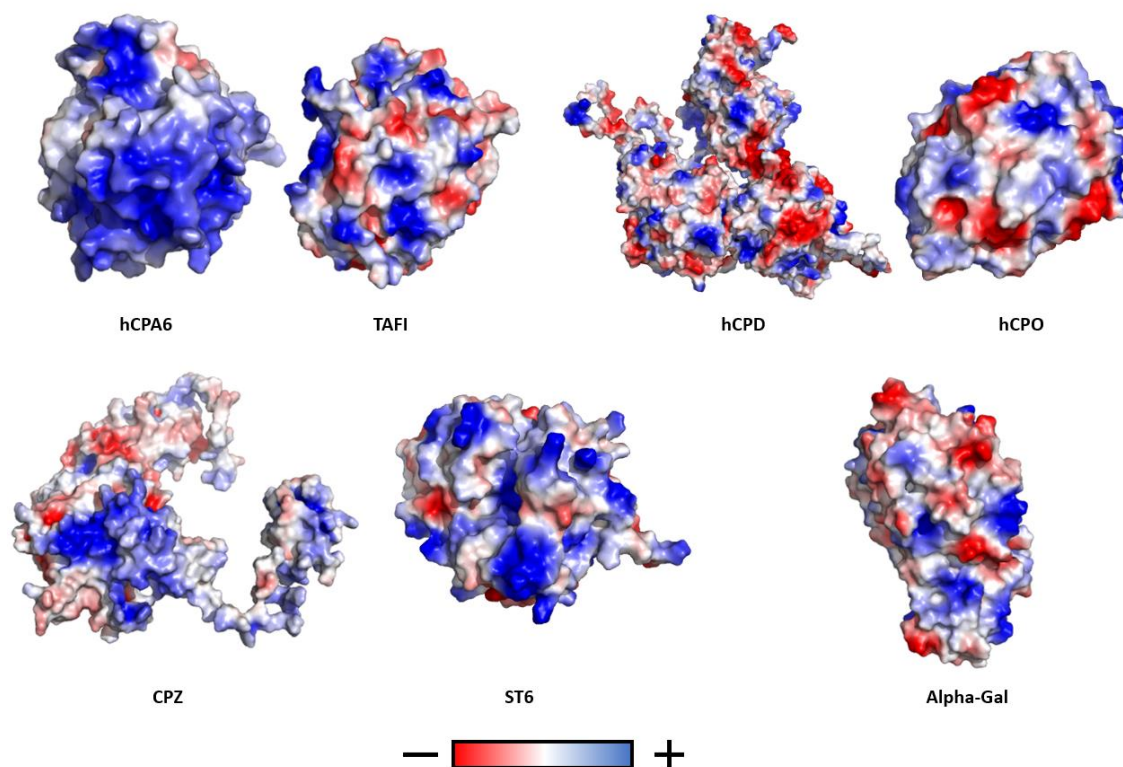


Figure 35. Heparin affinity proteins. Electrostatic surface distribution of different proteins employed for this project. Catalytic domain of human CPA6 (hCPA6), human CPD (hCPD) and human CPZ (hCPZ) are modelled structure using RaptorX. Human α -Gal (PDB code: 1R46). Human CPO (hCPO) (PDB code: 5MRV), Human TAFI (hTAFI) (PDB code: 1KWM), and Human ST6 (PDB code: 4JS1).

2.3.2. Cell Viability in Heparin presence

The first step to assure the success in our strategy is to test the effect of the addition of heparin during the expression process therefore heparin causes cell toxicity at high doses (Gurbuz et al. 2013). For that reason, heparin was evaluated as biocompatible component for cell viability, to underline the toxic effects and determine a threshold value for its toxicity, we searched for a dose–response testing different concentrations of commercial heparin solution in cultures of HEK 293F cells that were incubated at different times.

HEK 293F is a human embryo kidney cell line that is used commonly for mammalian cell recombinant protein production. The viability of cells was measured after 24 or 72 hours before the heparin addition (Figure 36).

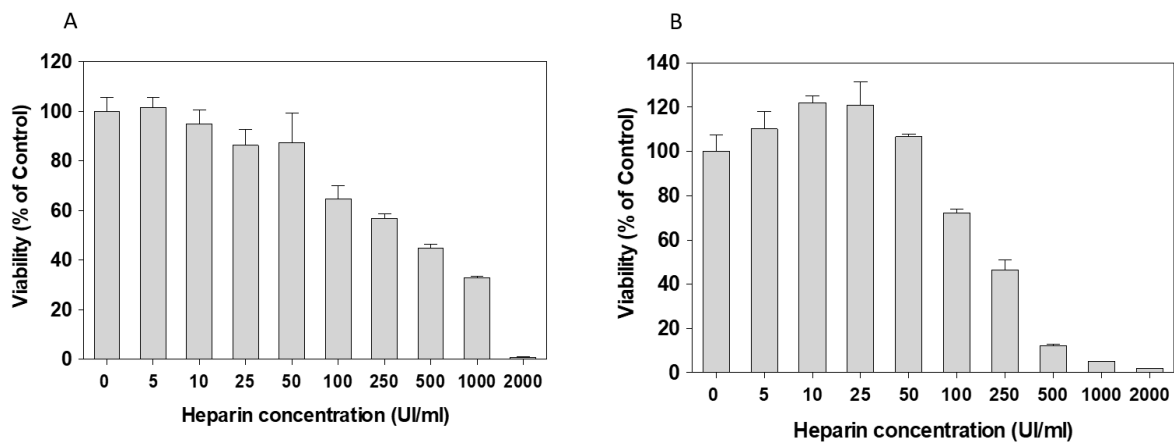


Figure 36 Graphic representations of cell viability. Cells have been incubated at different heparin concentrations at different times. A) 24 hours B) 72 hours. Reproduced of (Garcia-Pardo 2015)

At both times, we determined that the optimal Heparin concentration is 50UI/mL because is observed that this is the maximal concentration that preserves 100% of cell viability.

We also analysed if the heparin presence in the cell culture medium of HEK 293F cells provokes some morphological change in the cells which could modify its metabolism and interfere with the recombinant protein production. Thus, cells were incubated during 72h in the presence and absence of heparin at 50 UI/mL. After the incubation, we evaluated their state by microscopy and took some pictures, the cells were not treated by Trypan Blue (Figure 37).

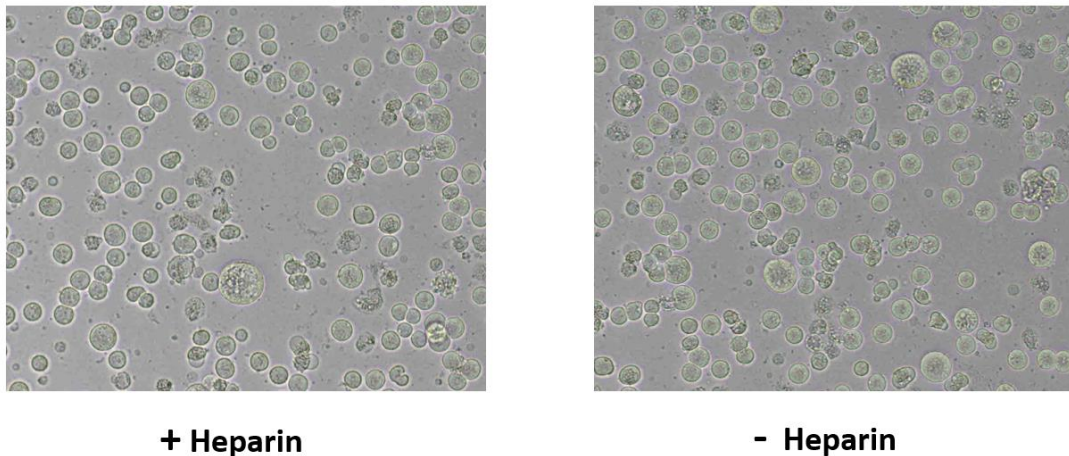


Figure 37. Images of HEK293F suspension cells obtained in presence and absence of 50UI/mL Heparin after 72h of incubation. No Trypan blue coloured. Reproduced of (Garcia-Pardo 2015)

HEK 293F cells are mammalian stem cells which grow in suspension, so they morphology used to be circular. But it has been reported, HEK 293 cell line could be differentiated under chemical stimulations.(Lu et al. 2004). We tested if the prolonged exposition to heparin provokes this differentiation. As it is observed, the presence of heparin at final concentration of 50 UI/mL does not affect cell viability and does not change the cell morphology. So, heparin presence in cell medium at 50UI/mL of concentration is not toxic for the cell.

2.3.3. Improved expression of Heparin affinity MCPs

The cDNA of each protein gene was cloned into the mammalian expression vector pTriex-7, for its transfection. This vector uses the IgM exporting domain for extracellular protein production. The optimized protocol is already described on Experimental procedures on this chapter.

The efficiency of the protocol proposed was evaluated by western blot using antibodies of the selected proteins (hCPZ, hCPA6, hTAFI, ST6, hCPO, hCPD and α -Gal) at different post transfection days in order to evaluate the potential amount increase in protein presence for the addition of sodium heparin. Values were normalized to maximum signal detected from protein studied in conditioned medium in absence of sodium heparin.

We divided the results into two groups: on the one hand, the MCPs and the negative control protein α -Gal that heparin has not effect on their expression, and on the other hand, the group

in which MCPs and the positive control protein ST6 are affected by heparin addition promoting their overexpression.

In the first block of results, we found the negative control protein, α -Gal, a protein located in lysosomes, with an acidic IP and. On the results, as expected, the control levels of proteins expressed over time increased, but there were no differences between the addition or not of heparin to the cell culture medium (Figure 38. A and B). Similar results were observed in hCPD truncated expression (Figure 38. C and D) and hCPO truncated expression (Figure 38. E and F) (Garcia-Guerrero et al. 2018). These observed results are the expected for these MCPs, which do not have basic IPs.

The second group of proteins comprises the proteins whose expression was increased by the addition of heparin in the extracellular media. In the case of ST6, that is the protein used as positive control because its interaction with heparin is already reported in bibliography (Kuhn et al. 2013), heparin presence in conditioned media increases the amount of protein around 2-fold (Figure 39. A and B). The first MCP used in the present study, hTAFI, heparin presence increases the amount of over expressed protein on extracellular media above 6-fold (Figure 39. C and D). In the case of hCPZ, an interesting protein of study due to its potential role in Wnt signalling pathway, the presence of heparin increases is around 4-fold the expression levels (Figure 39. E and F). The last MCP studied was hCPA6, which is the most benefited of this procedure. For this protein, the presence sodium heparin increased the overexpression levels above 50-fold (Figure 39. G and H). This is an important fact in our lab because the amount of hCPZ obtained before the implementation of this protocol was insufficient to allow its biochemical characterization (Garcia-Pardo et al. 2020)

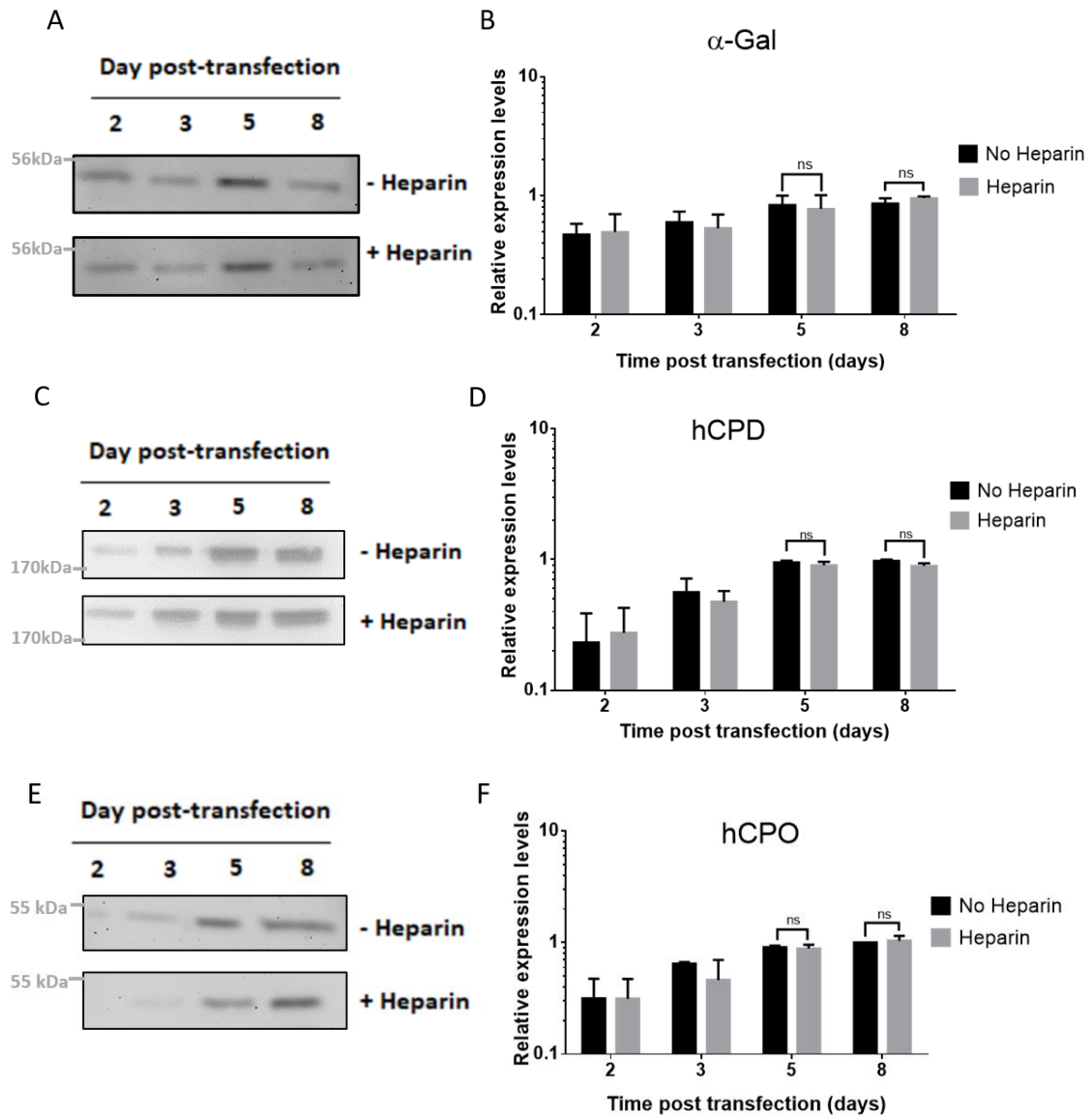


Figure 38. Effects of heparin addition in expression of α -Gal, hCPD and hCPO in mammalian cell culture of HEK 293F cells. Immunoblots of α -Gal (A), hCPD (C) and hCPO (E) expressed by HEK 293F cells over time in presence and absence of sodium heparin. Presence of protein was detected by mouse anti-Strep-tag II mAb (IBA technologies) as primary antibody and goat anti-mouse IgG (H+L) HRP conjugate (BioRad) as secondary antibody. Graphic representation of relative expression levels of human α -Gal (B), hCPD (D), and hCPO (F) over time in HEK 293F cells. Values were normalized to maximum signal detected in conditioned medium in absence of sodium heparin for each protein. (n=3, ns= no significance).

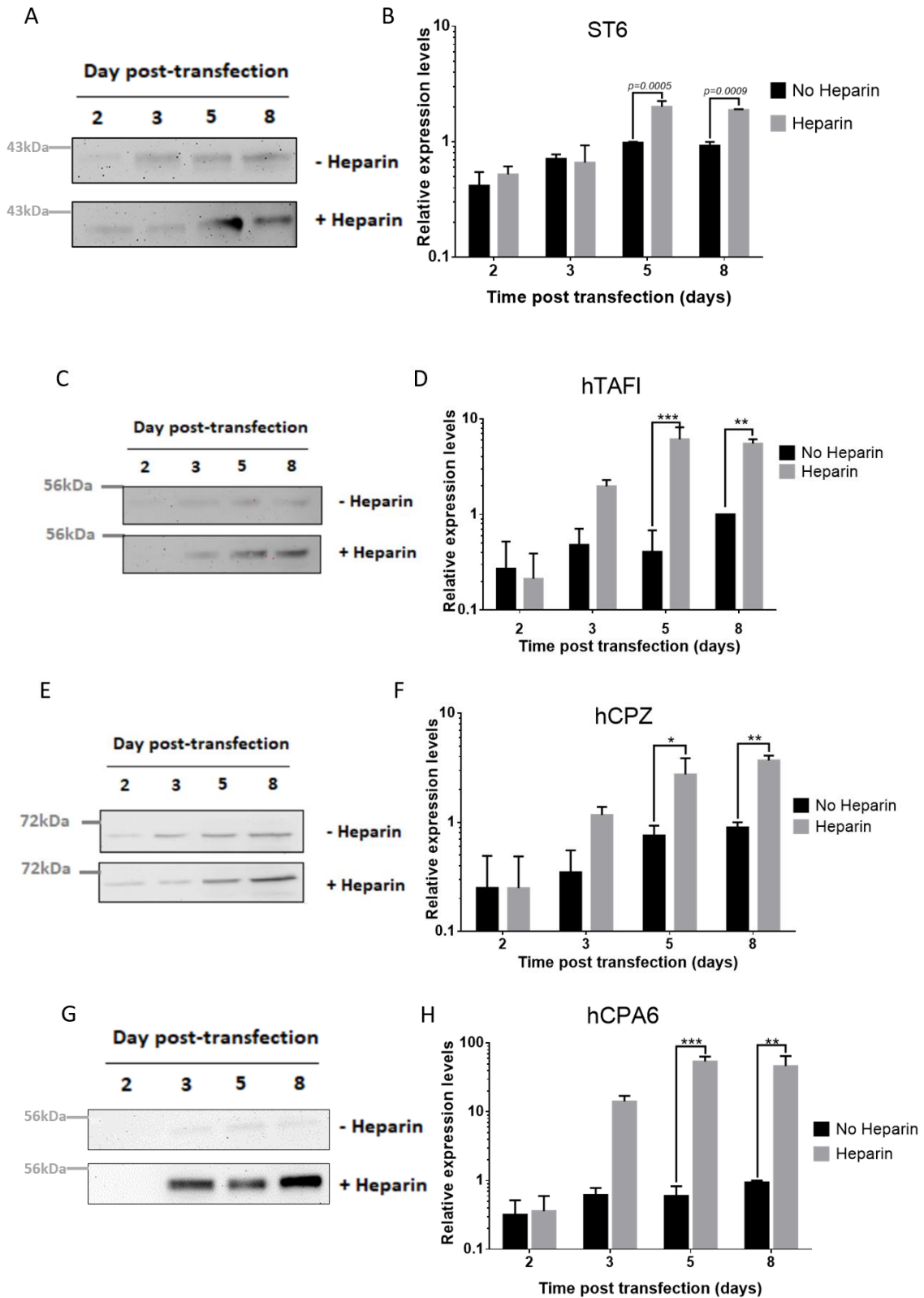


Figure 39. Effects of heparin addition in expression of ST6, hTAFI, hCPZ and hCPA6 in mammalian cell culture of HEK 293F cells. Immunoblots of ST6 (A), hTAFI (C) and hCPZ (E) and hCPA6 (G) expressed by HEK 293F cells over time in presence and absence of sodium heparin. Presence of protein was detected by mouse anti-Strep-tag II mAb (IBA technologies) as primary antibody for ST6, hTAFI, and hCPZ and mouse anti-HA mAb (Sigma Aldrich, Inc) for hCPA6. As secondary antibody for both primary antibodies a goat anti-mouse IgG (H+L) HRP conjugate (BioRad) is used. Graphic representation of relative expression levels of human ST6(B), hTAFI (D), hCPZ (F) and hCPA6 (H) over time in HEK 293F cells. Values were normalized to maximum signal detected in conditioned medium in absence of heparin for each protein. (n=3, (B) p-values in the graph, (D) *** $p=0.0004$ ** $p=0.0032$, (F) * $p=0.029$ ** $p=0.0022$, (H)*** $p=0.0004$ ** $p=0.0021$).

2.4. DISCUSSION

In this chapter is described a cheap, efficient, and straightforward protocol for producing large amounts of recombinant heparin-affinity MCPs from suspension HEK 293F cells. These cells must be cultured in serum-free and antibiotic-free medium, and the passaging cell also must be done with sterile technique in order to avoid contaminations which increase the cost and the time-consuming. As it is described before, DNA transfection is required for the protein expression. For a successful transfection, some points must be required. One of them is cell viability, it should be over 90% and cultures must contain single and dividing cells. Cell viability is evaluated during culture maintenance keeping cells in optimal conditions. Another point to take care is DNA, the plasmid used must be pure and high quality ($A_{260}/A_{280}=1.8-2.0$). The presence of organic elements or proteins from DNA purification process can interfere in transfection. The last point which intervenes in transfection is transfection agent. The ratio of DNA/transfection agent should be optimized, generally a ratio of 1:3 works for most of proteins to generate stable DNA/transfection agent complexes (Longo et al. 2013, Portolano et al. 2014, Zhang et al. 2017).

Following the protocol described, another key element is sodium heparin. Cytotoxicity of sodium heparin is evaluated to determine the top concentration which does not cause cell death on HEK 293F cells, and not produce morphological changes. So, a safety concentration of sodium heparin is found for its addition to extracellular media, making easier the interaction of molecules of heparin and protein expressed.

The results obtained in this study shown the positive effect of heparin on the expression of basic IPproteins, or heparin affinity proteins. The increase amount differs in each protein. The increase is not observed in negative control and both carboxypeptidases which are not described with basic pI. These results are beneficial for the chapter

This study completes a work started previously in our research group on also was evaluated the activity of carboxypeptidases expressed in both situations, the heparin affinity carboxypeptidases and heparin non-affinity carboxypeptidases (Garcia-Pardo 2015, Garcia-Guerrero et al. 2018).

CHAPTER III

EXTRACTION, PURIFICATION AND CHARACTERIZATION OF OVOCALYXIN-32

CHAPTER III: EXTRACTION, PURIFICATION AND CHARACTERIZATION OF NATURAL OVOCALYXIN-32

3.1. INTRODUCTION

Carboxypeptidases are proteolytic enzymes involved in a significant number of functions of living organisms, from food digestion to blood coagulation, inflammation, local anaphylaxis, hormone/neuropeptide processing, insect/vegetal attack defence strategies, among others (Reynolds et al. 1989, Bayés et al. 2005, Sanglas et al. 2008, Sanglas et al. 2010). In general, these proteases are secreted and their enzymatic action takes place normally in the extracellular space, except in the case of a novel subfamily of cytosolic carboxypeptidases, which are presumably involved in the processing of tubulin (Kalinina et al. 2007, Rodriguez de la Vega et al. 2007), as well as other proteins with repeated acidic residues at the C-terminus and strong biological action (Tanco et al. 2015). Knowledge of the control of interference with those mechanisms, by rational structure based or other drug design approaches, is great interest for biotechnical and biomedical industrial purposes.

The M14 carboxypeptidase inhibitors are mostly small proteins, between 5 and 9 kDa, which inhibits the activity of carboxypeptidase blocking the active site access by a substrate-like mechanism, using its C-terminal tail. But there are other kinds of proteinaceous inhibitor with a larger size.

One of the few large mass protein inhibitors known for metallo-carboxypeptidases (MCPs) is latexin, a 25-kDa protein discovered in the rat brain. Latexin, alias ECI, inhibits only members of the M14A metallo-carboxypeptidases subfamily. This subfamily, also known as pancreatic-like MCPs, comprise at least eight members: A1-6, B and B2 or TAFI. These M14A metallo-carboxypeptidases are synthesized as zymogens with, N-ter prodomain, which inhibits their activity by hindering substrate access to active site. For the activation-maturation of such carboxypeptidases a limited proteolysis that cleaves is required, that irreversibly releases the prodomain.

The structure of human latexin consists of two topologically equivalent subdomains, reminiscent of cystatins, folded in an α -helix enveloped by a curved β -sheet. They are packed against each other and linked by a connecting segment encompassing a third α -helix. Its mechanism of inhibition differs from the rest of inhibitors of M14 MCPs described because the C-terminal of this inhibitor does not interact with catalytic centre. Instead of the latexin sits on the top of the catalytic centre enzyme, mainly interacting by a central loop of its C-terminal

domain, emulating the prodomain of carboxypeptidases and avoiding the interaction of enzyme with its substrate.

Thus, complex formation with hCPA4 involves at least 10 residues of latexin (see Figure 40), which again is unlike other metalloprotease/inhibitor complexes, in which, characteristically, a smaller surface and lower number of inhibitor residues, about 4-5, interact with the enzyme, as previously described at introduction.

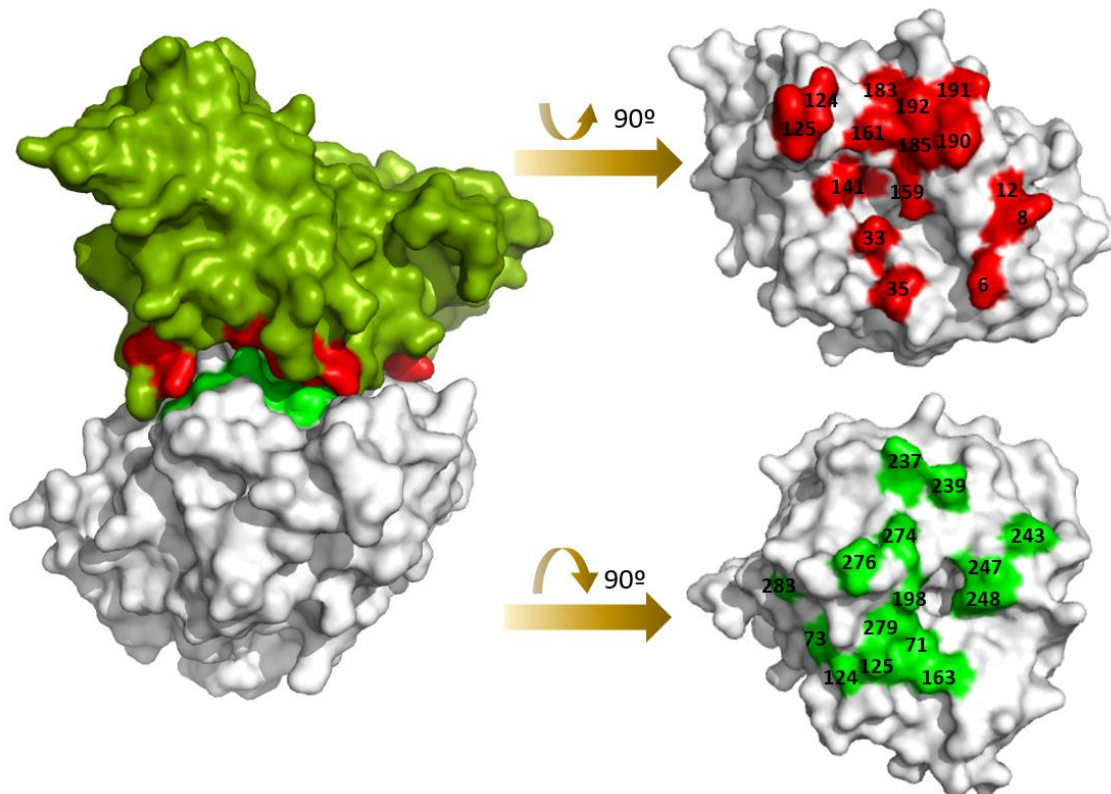


Figure 40. Graphic representation of the complex of LXN-hCPA4. The picture also shows the residues involved on interaction between inhibitor and carboxypeptidase. On the complex, the green molecule is LXN, and the grey is hCPA4. In the representation of residues involved, red one is LXN and green is hCPA4.

Rather, latexin inhibition is mainly caused by an inhibitory loop provided by the central part of the β -sheet, localized close to C-term, reminiscent of the inhibition through the prodomain in the M14A zymogens. This loop is shaped by the end of strand β 7, the beginning of β 8 and the non-structured residues which form the loop. The connecting loop β 7- β 8 (with the Gln190/Glu191 pair as the closer ones) penetrates the protease moiety, showing a behaviour similar to the one found in cystatins (inhibitors of cysteine proteases) (Figure 41). (Stubbs *et al.*, 1990).

Strikingly important for complex stability is the interaction of Gln190 (of the inhibitor) with Arg71 (of the enzyme). The latter is a basic residue present throughout M14A MCPs subfamily and is required for maintenance of the crucial salt bridge between Asp36 from the enzyme region (Asn36 in hCPA4) and the pro-domain of PCPs. This basic residue is absent in M14B forms, which are not secreted as proenzymes.

Interestingly, Gln190 from inhibitor, furthermore, interacts with Tyr248 of hCPA4, important for catalysis and which is involved in hydrogen bonding of the P1 amide nitrogen of substrates through its O_H group. The position of Gln190 of inhibitor within the inhibitory loop is maintained by an intramolecular interaction with the $N\epsilon 2$ atom of the preceding His185. This histidine further interacts via an edge-to-face Van-der-Waals contact with the side chains of Phe279 and Tyr198 of the protease moiety. (Figure 41) (Pallarès et al. 2005).

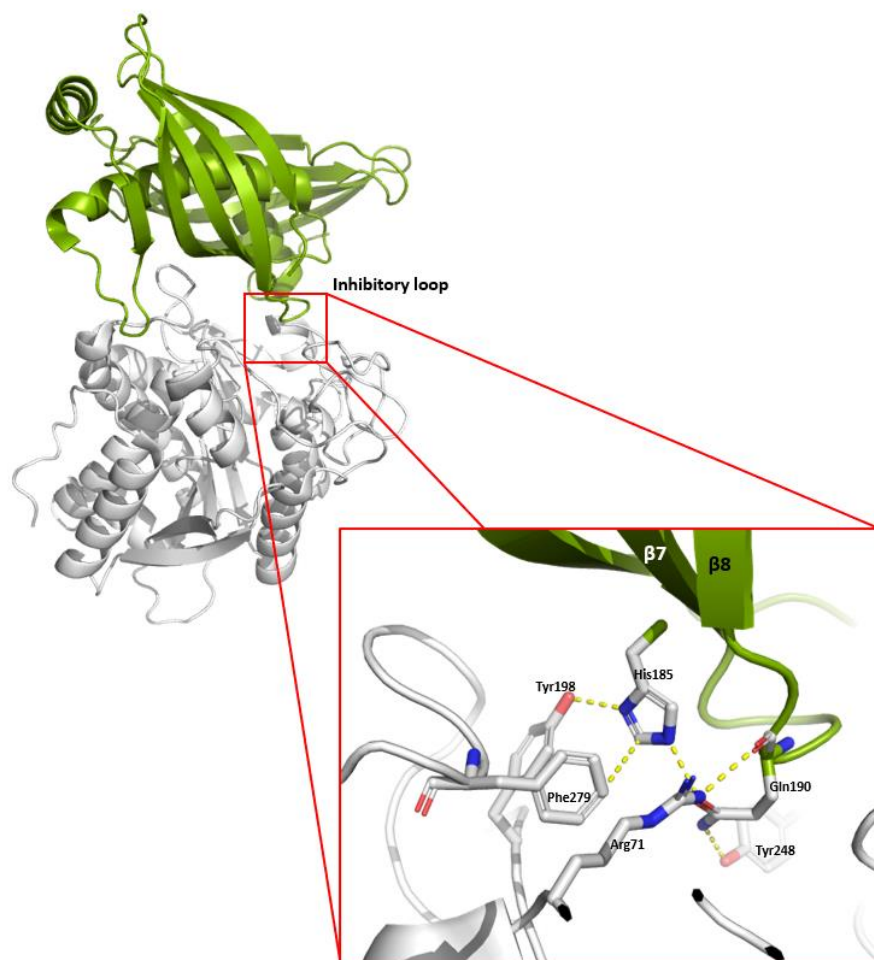


Figure 41. Graphic representation of key interaction between latexin and hCPA4. From the inhibitory loop located between the $\beta 7$ and $\beta 8$, Gln190 from latexin interacts to Arg71 from hCPA4; this interaction emulates the binding of prodomain to M14A carboxypeptidase by a salt bridge. This interaction is stabilized in latexin by intramolecular interaction between Gln190 and His185, also by salt bridge. Grey is hCPA4, green is latexin.

OCX32 is described as a paralog of latexin and member of the latexin family (I47 inhibitor in MEROPS), and its sequence has a 32% of homology with latexin sequence. When the sequences of both, OCX32 and latexin are aligned, it is revealed that several key residues for the interaction of latexin-carboxypeptidase are conserved in OCX32 (Figure 42); like His185/195 and Gln190/200 (besides 5 strict identities and 9 conservative fittings in the sequences around). The fact that the residues are conserved can give to OCX32 the skill to act as a potential metalloprotease inhibitor, as we shall show later on.

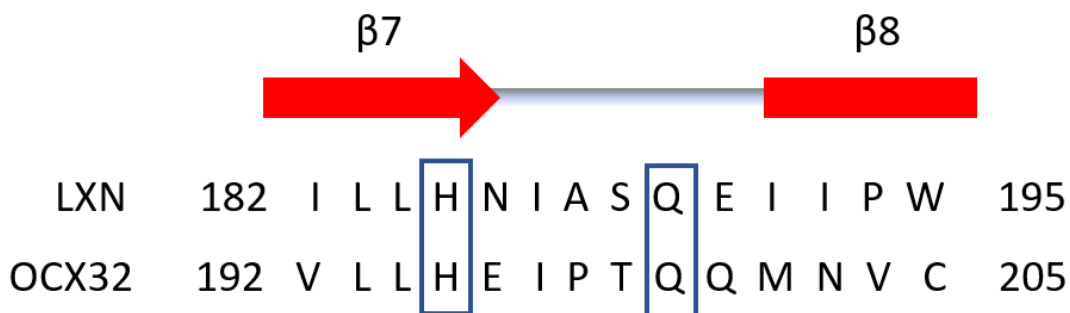


Figure 42. Partial alignment of latexin and OCX32. Gln190 on latexin corresponds to Gln200 on OCX32, and His185 corresponds to His195.

In this chapter we shall focus in the obtention of natural OCX32 from a natural source and its purification, for a its subsequent characterization as protease inhibitor.

3.2. EXPERIMENTAL PROCEDURES

3.2.1. Bicinchoninic Acid Assay

Bicinchoninic Acid (BCA) Assay is a biochemical detergent-compatible method for determining the total protein concentration in a solution, with high sensitivity (Smith et al. 1985). The assay is based in a two-reaction process which key reagent is the cation copper (II) (Cu^{2+}). Firstly Cu^{2+} ions are reduced from Cu^{2+} to Cu^+ , due the presence of peptide bonds at 37°C . So, the amount of generated Cu^+ is proportional to the amount of protein present in the solution. Afterward, each Cu^+ ion forms complexes with two molecules of bicinchoninic acid chelate, resulting as purple-colored solution (Wiechelman et al. 1988).

For total protein quantification, Pierce™ BCA Protein Assay Kit from Thermo Fisher Scientific was used when possible, following Microplate protocol detailed in the manual of the manufacturer. The standards were freshly prepared with Bovine Serum Albumin (BSA) from a $2 \text{ mg}\cdot\text{mL}^{-1}$ concentration, diluted with buffer from the samples.

To proceed with protein quantification, 25 μL of samples and standards were mixed with 200 μL of Working Reagent from kit per well in a 96-well microplate. Every sample or standards were added in triplicates. This method requires 30 minutes of incubation at 37°C, in order to let both reactions happen. Absorbance was quantified on a spectrophotometric plate reader at 550 nm.

3.2.2. Estimation total pure protein by absorbance

Proteins have a characteristic and specific absorbance spectrum. Its spectrum shows two peaks in UV region: first one at 214 nm, where peptide bonds absorb, and the second one at 280 nm, where Tryptophan and Tyrosine absorb, whilst Phenylalanine and disulphide bonds contributes in the latter too but in a much lower level (Figure 43) UV absorbance at 280 nm is routinely used to estimate protein concentration in laboratories due to its simplicity, ease of use and affordability. Measurements, in a given and appropriate buffer, are quick and highly reproducible since there is no need for incubation. In addition, this method also requires an extremely small sample volume since modern spectrophotometers use microcells or a sample retention system during the measurement.

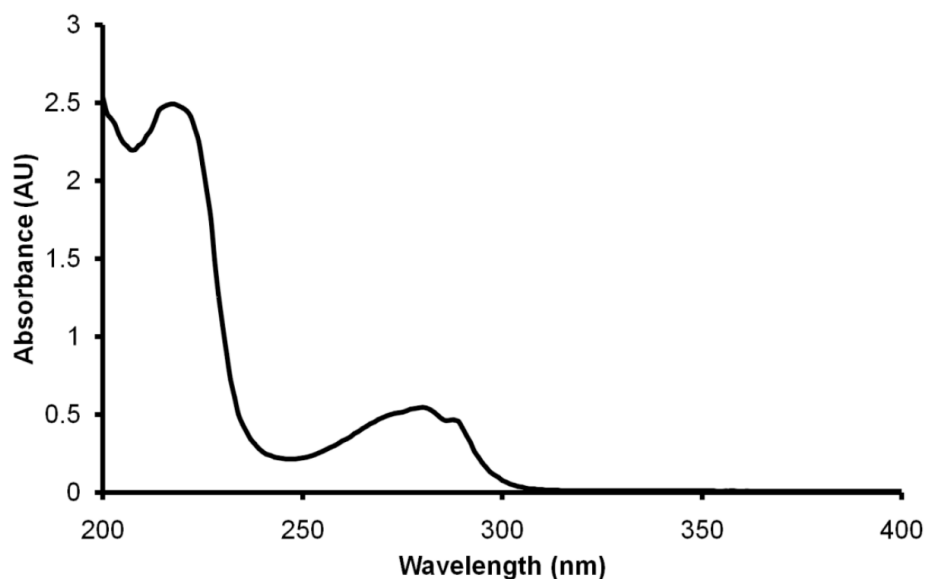


Figure 43. Example of protein absorbance spectrum in Ultra-Violet (UV) range. As is described a protein presents two peaks of absorbance at UV light region, one at 214nm (peptide bonds), and a second one at 280nm (aromatic residues and disulphide bonds).

The modern derivation of the Beer–Lambert law, necessary to follow for correct readings, correlates the absorbance, to the concentration of solution and the thickness of the material sample (path length). It is resumed in a simple formula:

$$\text{Abs}_{280} = \varepsilon \cdot l \cdot c$$

Where

- Abs_{280} is value of absorbance at 280 nm
- ϵ is molar extinction coefficient, this value means how strongly the component absorbs light in 280 nm. In proteins, this value could be calculated from aminoacidic sequence in ProtParam tool - ExPasy (<https://web.expasy.org/protparam/>). Its units are $M^{-1}\cdot cm^{-1}$
- l is the light pathlength in cm
- c is the molar concentration of the species present in solution in M ($mol\cdot L^{-1}$)

3.2.3. Sodium Dodecyl Sulphate Polycrylamide Gel Electrophoresis (SDS-PAGE):

The methodology used is the same process described in experimental section in Chapter I.

3.2.4. Coomassie SDS-PAGE staining

The methodology used is the same process described in experimental section in Chapter I.

3.2.5. Extraction and Purification of natural Ovocalyxin-32:

As described in bibliography, there are two main biologic sources to obtain natural Ovocalyxin-32: one is from uterine fluid III from the chicken, and the second is from the cuticle, the most external part of eggshell. Due to easy acquisition of eggs, the chosen resource was the last one.

The extraction protocol for protein from eggshell was adapted from (Gautron et al. 2001). The obtention of protein was performed incubating eggs inside 1 N HCl for 5 minutes, with gentle mechanic stirring to improve extraction yield. After extraction, sample was dialyzed and equilibrated in a 50mM $(NH_4)_2SO_4$ pH 5.24. Before its fractionation, the extract was centrifuged in order to separate the soluble fraction (which is used to purify) from an insoluble red fraction (Figure 44), and the former was filtered through a $0.22\mu m$ pore filter, in order to remove remaining aggregates from the sample.

The protein purification was performed by using a combination of two consecutive chromatographic steps: on a weak cation exchange- and on a size exclusion column. Thus, the dialyzed supernatant from the eggshell extraction, in 50mM $(NH_4)_2SO_4$ pH 5.24, was applied to a CM Sepharose FF (1.6cmX20cm, from GE), to ensure binding of the protein through its rich surface in basic residues, to the matrix. The bound proteins were eluted using a linear salt gradient from the equilibration buffer to the same + 2M of NaCl. Identification of rich OCX32 elution peak(s) was performed by SDS-PAGE in Tris-glycine buffer.

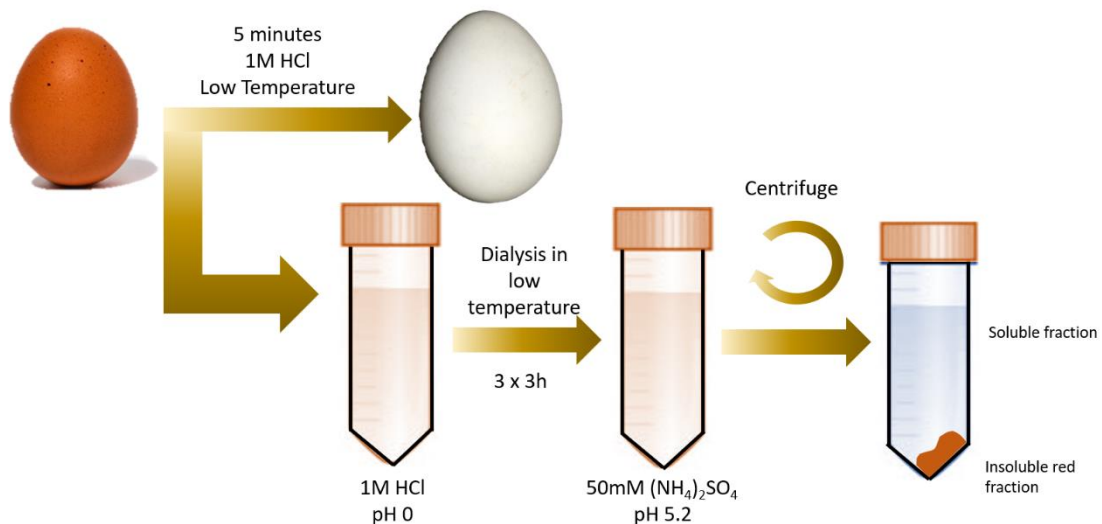


Figure 44. Schematic representation of protein extraction process from chicken eggshell. In order to optimize the extraction and maximize the quantity of protein extracted from the eggshell, the first steps of the process (extraction and start of dialysis) must be done fast and in low temperature.

Afterwards, the different elution peaks were analysed by SDS-PAGE, the protein of interest detected, and fractions which compose the peak of interest were mixed and concentrated for a next step of purification. The subsequent Size Exclusion Chromatography (SEC) used, after cation exchange purification was a Superdex 75 column HR/10/30 with 10mm of diameter and a height of packed bed of 30 cm, which fractionate proteins in the 60,000-3,000 Da range, very appropriate to purify OCX32. The column was equilibrated and eluted with 50mM (NH₄)₂SO₄ at pH 5.24.

3.2.6. Enzyme Activity

From enzyme assays the data obtained can be evaluated from a qualitative approach, a clear positive or negative result, or can be evaluated from a quantitative approach to determine the kinetics parameters for the enzyme. Other purposes of enzymatic assays are to identify specificity of enzyme, to prove its presence or absence in a distinct specimen (like an organism or a tissue) and to determine the amount of the enzyme in the sample. This kind of assays is also used in order to evaluate potential inhibitors and characterize them.

3.2.6.1. Determination of kinetics parameters

The protocol to study enzymatic kinetics starts with the determination of the enzymatic velocity at different concentration of substrate. Their graphical representation results in a hyperbolic curve (Figure 45.A) when the experimental data is represented by the substrate concentration at the X-axis and velocity at the Y-axis. The kinetic experimental data of most

enzymes, without presence of activator or inhibitor, could fit in the final derivation of Michaelis-Menten equation (Figure 45.B) (Michaelis and Menten 1913).

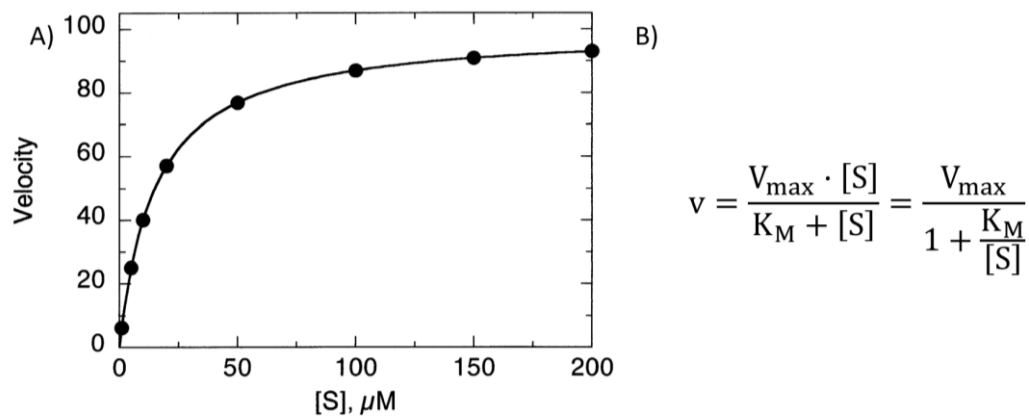


Figure 45. Graphic representation of velocities and Michaelis-Menten equation. A) Every velocity is calculated and represented in its correspondent substrate concentration. Its representation is an hyperbolic curve which is described mathematically as Michaelis-Menten equation (Copeland 2000). **B)** Michaelis-Menten equation where v is experimental velocity, V_{\max} is maximum velocity, K_M is Michaelis-Menten constant, which is the substrate concentration at which the reaction velocity is 50% of the V_{\max} and $[S]$ is initial concentration of substrate in reaction.

The activity of enzymes can be blocked by inhibitors and this process can be analysed in order to determine the effected kinetic parameters. Among the inhibitors we can differentiate two mechanistic kinds, reversible and irreversible inhibitors. Reversible inhibitors are a major research focus for both fundamental applied research (i.e. at pharmaceutical industry), so is an interesting field of study. There are two types of reversible inhibition, competitive, which substrate and inhibitor compete for a same interaction site in enzyme, or non-competitive, in which inhibitor interacts in different site of enzyme.

In addition, another type of inhibitor exists the one involving tight-binding inhibitors. This kind of inhibitors strongly interacts with enzyme, in a nearly stoichiometric ratio. A dose-response curve is done to determine if a protein or peptide is a tight-binding inhibitor, and the IC_{50} value derived from this data is close or similar to the concentration of active enzyme in the assay. In this case, "Morrison" is one of the favoured approaches to be used to determine inhibitor constants, i.e. for different peptidases and natural inhibitors, in which this inhibition is quite frequent. Residual enzymatic activity (V_i/V_0) is calculated for each inhibitor concentration and the obtained values are adjusted to the Morrison's quadratic equation (Morrison 1982). The following analytical equation applies:

$$\frac{V_i}{V_0} = 1 - \frac{([E_0] + [I_0] + K_i^{\text{app}}) - \sqrt{([E_0] + [I_0] + K_i^{\text{app}})^2 - 4[E_0][I_0]}}{2[E_0]}$$

In this equation: $\frac{V_i}{V_0}$ is residual enzymatic activity, $[E_0]$ is total active enzyme concentration in assay, $[I_0]$ is total active inhibitor concentration in assay and K_i^{app} is apparent inhibition constant.

However, the K_i value assessed is apparent and the real value can only be calculated, taking into account the effect of substrate concentration on the inhibitory activity, according to equation described below (Bieth 1995).

$$K_i = \frac{K_i^{\text{app}}}{\frac{S_0}{K_M} + 1}$$

In this equation: K_i is real inhibitory constant, K_i^{app} is apparent inhibitory constant, $[S_0]$ is initial substrate concentration and K_M is Michaelis-Menten constant

All experimental data is fitted to equation for tight-binding inhibitors by non-linear regression using the GraphPad Prism 6 software (GraphPad Software, Inc, USA).

In the present project, enzymatic assays have been performed for different proteases and each protease requires different optimal conditions (see next subsection). To collect the experimental data, the absorbance was measured at different times, with a microplate reader, to define a velocities of substrates hydrolysis. In the presence of inhibitors this velocity will decrease depending of the concentrations of them. In order to simplify the data collection, we used a microplate based protocol (Tellechea et al. 2016)

3.2.6.2. Metalloprotease

3.2.6.2.1. Metalloproteases 14 A (M14A)

CPA activity in the absence and presence of inhibitors.

Substrate: N-(4-Methoxyphenylazofonyl)-Phenylalanine-OH potassium salt (AAFP)

[CAS No396717-86-5] supplied by Bachem (Switzerland). **Wavelength (λ):** 340 nm

Activity buffer: 20 mM Tris-HCl, pH 7.5, containing 500 mM NaCl, 1% v/v DMSO and 0.05% w/v Brij-35

Concentration in the assay: 100 μ M

Enzyme: CPA1 from bovine pancreas supplied by Sigma-Aldrich (USA)

Concentration in the assay: 0.5 nM

Enzyme: CPA1 from human, recombinant production

Concentration in the assay: 1 nM

Enzyme: CPA2 from human, recombinant production
Concentration in the assay: 1 nM

Enzyme: CPA4 from human, recombinant production
Concentration in the assay: 10 nM

CPB activity in the absence and presence of inhibitors

Substrate: N-(4-Methoxyphenylazofornyl)-Arginine-OH potassium salt (AAFA)
 [CAS No442158-31-8] supplied by Bachem (Switzerland). **Wavelength (λ):** 340 nm
Activity buffer: 20 mM Tris-HCl, pH 7.5, containing 100 mM NaCl, 1% v/v DMSO and 0.05% w/v
 Brij[®]-35
Concentration in the assay: 100 μ M

Enzyme: CPB from porcine pancreas supplied by SIGMA., Inc. (USA)
Concentration in the assay: 3 nM

3.2.6.2.2. Metallocoarboxypeptidases 14 B (M14B)

Activity buffer: 100 mM Tris-Acetate, pH 6.5, containing 100 mM NaCl
Substrate: N-(4-Methoxyphenylazofornyl)-Arginine-OH potassium salt (AAFA)*
 [CAS No 442158-31-8] supplied by Bachem (Switzerland). **Wavelength (λ):** 340 nm
Concentration in the assay: 100 μ M

* This substrate is not the best but is observed that hCPD cleavage Arg and it works for hCPD.

Enzyme: Carboxypeptidase D from human, recombinant produced.
Concentration in the assay (M): 100 nM

3.2.6.3. Serine proteases:

3.2.6.3.1. Subtilisin

Activity buffer: 50 mM Tris-HCl, pH 8.5 containing 10% v/v DMSO
Substrate: Benzylcarbonyl-glycyl-glycyl-L-Leucine 4-nitroanilide (GGLPNA) [CAS No 53046-98-3] supplied by Bachem (Switzerland). **Wavelength (λ):** 414 nm
Concentration in the assay: 400 μ M

Enzyme: Subtilisin from *Bacillus licheniformis* supplied by Sigma-Aldrich (USA)
Concentration in the assay: 20nM

3.2.6.3.2. Trypsin

Activity buffer: 20 mM Tris-HCl, pH 8.0 containing 20 mM CaCl₂, 150 mM NaCl and 0.05% v/v
 Triton X-100
Substrate: Na-Benzoyl-L-arginine 4-nitroanilide hydrochloride (BAPNA) [CAS No 21653-40-7]
 supplied by Bachem (Switzerland). **Wavelength (λ):** 414 nm
Concentration in the assay: 1000 μ M

Enzyme: Trypsin from bovine pancreas supplied by Sigma-Aldrich (USA)
Concentration in the assay: 28nM

3.2.6.4. *Cysteine Protease*

3.2.6.4.1. Papain

Activity buffer: 100 mM phosphate buffer, pH 6.5 containing 100 mM KCl, 0,1 mM EDTA, 3 mM Dithioerythritol (DTT) and 0.05% w/v Brij®-35

Substrate: L-Pyroglutamyl-L-phenylalanyl-L-leucine-p-nitroanilide (PFLNA) [CAS No 85901-571] supplied by Bachem (Switzerland). **Wavelength (λ):** 414 nm

Concentration in the assay: 400 μ M

Enzyme: Papain from *Carica papaya* supplied by Roche (Switzerland)

Concentration in the assay: 45nM

3.3. RESULTS

3.3.1. Extraction, identification and purification of the protein

OCX32 is a protein deposited over eggshell surface. in significant quantities. His sequence is known since 2001 (Gautron et al. 2001), cDNA-derived from a mix of ESTs and automated Edman analysis of peptides, and his function, by now, seems to be related to both the eggshell calcification and microbial protection. However, its low solubility, the difficulties for the derivation of a proper recombinant expression system linked to GST and the indications on its inhibitory capability over M14 MCPs (Xing et al. 2007), encouraged us to improve its production and solve the related unknowns. To skip the mentioned difficulties, we focused on the natural protein because the OCX32 source is easy to obtain and the extraction is already described (Gautron et al. 2001).

In order to know which source or kind of hen eggs should be used, we tested two different origins of eggs: either from free-range hens or from commercial farms and distributors. Commercial eggs can be bought in any supermarket, but free-range hens' eggs are more difficult to obtain. Fortunately, we have contacted with farmers who kindly supplied all required eggs, in fresh and untreated state. The protein was extracted from the eggshell of both kind of eggs, using a modified protocol described in the Experimental procedures subsection, and the extracts analysed by SDS-PAGE (Figure 46).

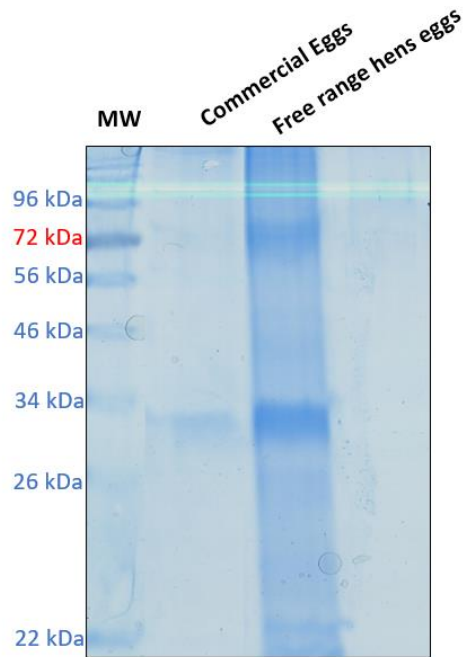


Figure 46. SDS-PAGE of commercial eggs and free-range hens' eggs extracts. *MW*: molecular weight ladder. *Commercial eggs*: soluble extract from commercial eggs. *Free range hens' eggs*: soluble extract from free-range hens' eggs. In order to have the closest conditions between both extracts, the commercial eggs used are labeled as free-ranged hens' eggs too.

After this gel, we can conclude that the best source is free-range hens' eggs, directly from the farm because they have a higher concentration of protein on the eggshell surface than commercial eggs. Besides, such "natural" origin allowed us to guarantee both the freshness and the lack of any chemical treatment of the eggs (i.e. to increase its conservation time).

Although an extraction protocol is already described and used from OCX32 bibliography (Gautron J et al, JBC, 2001), we must be sure that the process of extraction in 1N HCl works well and is respectful with protein integrity, solubility and activity, in our case. For this reason, we had always these properties in mind, and analysed-identified most of the used proteins of soluble extracts using the peptide mass fingerprint (PMF) approach by MALDI-TOF.MS (see Figure 47). The bands were scissed from SDS-PAGE gel and analysed at the Servei de Proteòmica i Bioinformàtica (SePBio) from Universitat Autònoma de Barcelona.

Hence, the maintenance of such properties for OCX32 in the extract was confirmed in all different batches produced. Furthermore were identified two more proteins in the extract: Ovocleidin-116 (OC116) and Ovocleidin-17 (OC17), so the protein should be purified from the rest of extract, before to be characterized (Figure 47). The protocol used is already described, but in this Thesis is adapted for our purposes.

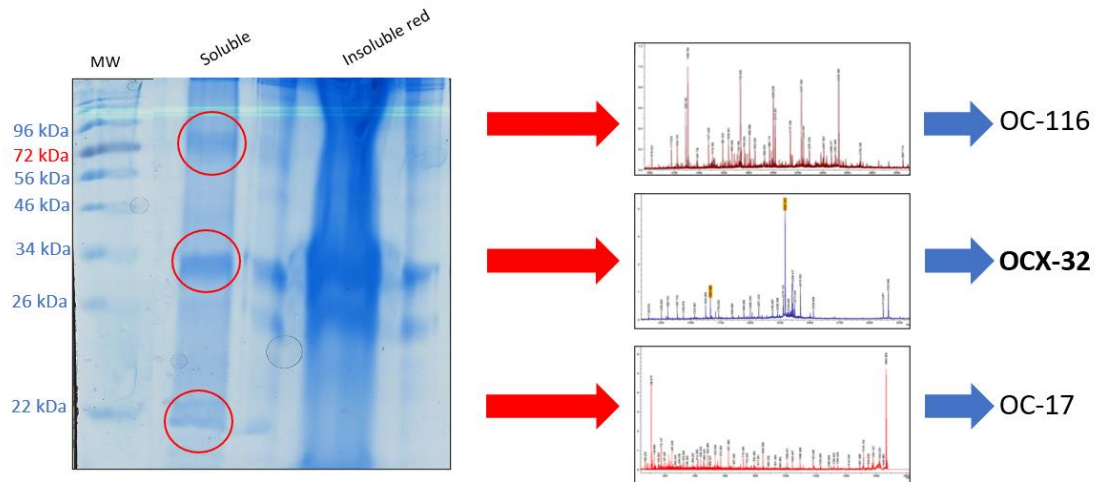


Figure 47. Graphic representation of how the bands have been processed and proteins identified. Three major protein bands are found in the soluble extract. The band closest to 34 kDa marker was identified as OCX32.

In the protocol described in bibliography, eggs' extract after dialysis is loaded in the Carboxymethyl Sepharose (CM-Sepharose) column. OCX32 has a basic pI (the value calculated by the ExPasy ProtoParam from aminoacid sequence is 8.99) which allows this protein to be purified by a weak cationic exchange chromatography, at a low acidic buffer. The presence of H^+ in the solution, at low pH, confers to the protein with a positive net charge, allowing it to interact with the carboxymethyl group ionized as $-CH_2OCH_2COO^-$ from CM-Sepharose. In the protocol from bibliography, the equilibration buffer used was 50mM Sodium Acetate at pH 4.0. However, this buffer interferes in metalloproteinase activity, and acts as an inhibitor. Other buffers, in combination with salts at distinct ionic strengths, were tested and finally 50mM $(NH_4)_2SO_4$ at pH 5.24 was selected to perform all purification process.

So, after the extraction process was completed, the extract was quickly dialysed to 50mM $(NH_4)_2SO_4$, at pH 5.24. After the dialysis, the extract is centrifuged to separate soluble and insoluble red part, and the soluble part is filtered through 0.22 μ m PVDF Protein Low-Binding filter. The filtered extract is loaded onto 20mL of CM Sepharose column bed packed inside an XK 16/20, and the height of the column is 10cm. The purification was at 4°C, in order to stabilize the protein loaded (Figure 48).

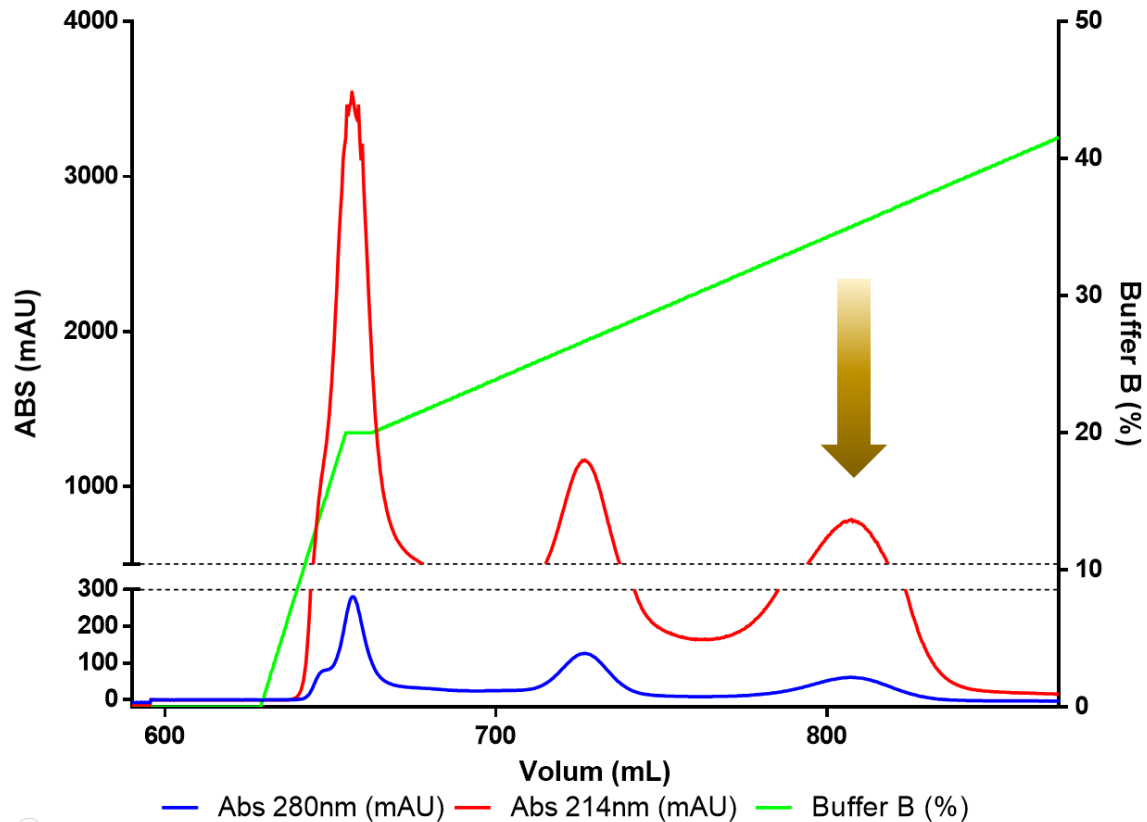


Figure 48. Graphic representation of CM Sepharose purification of the extract. The section of the chromatography here displayed is the elution of the protein, each peak was analysed by SDS-PAGE to identify where is the OCX32 rich fraction is localized. The elution peak of OCX32 is marked by yellow arrow. Buffer A or equilibration buffer is 50mM $(\text{NH}_4)_2\text{SO}_4$, at pH 5.24 and Buffer B is 50mM $(\text{NH}_4)_2\text{SO}_4$, 2M NaCl, at pH 5.24.

Once the fractions which compose the peak of OCX32 are mixed, they are concentrated using Amicon® Ultra-15 10K Centrifugal Filter Devices. Subsequently the protein purified and concentrated is loaded into a Size Exclusion Chromatography (SEC), Superdex 75 HR/10/30 with 10mm of diameter and a height of packed bed of 30 cm. The volume approximately of this SEC is 24mL. The chromatography was done at room temperature. This chromatography also will be used as a technique to evaluate de purity of the sample and desalting it (Figure 49).

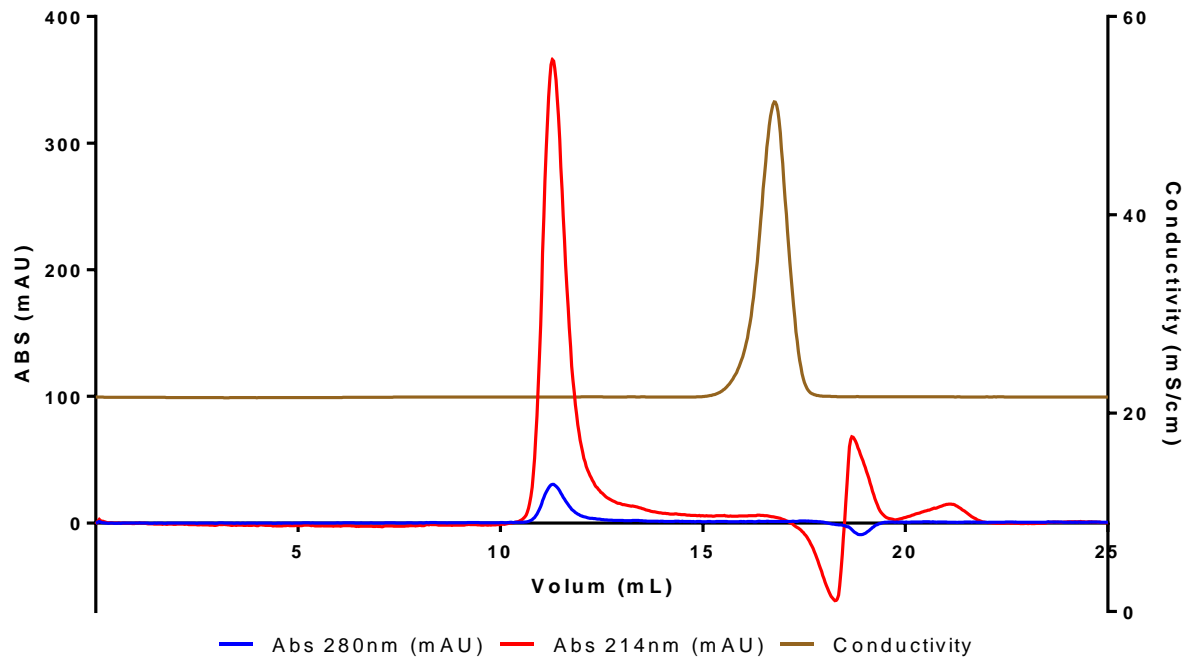


Figure 49. Graphic representation of Size Exclusion Chromatography from OCX32 peak purified. The graphic shows that the OCX32 is pure after CM Sepharose chromatography, moreover the desalting process is a success as the conductivity line shows.

To sum up and have an overall view of the process of purification, a sample on every step on purification process was taken and analysed by SDS-PAGE (figure 50).

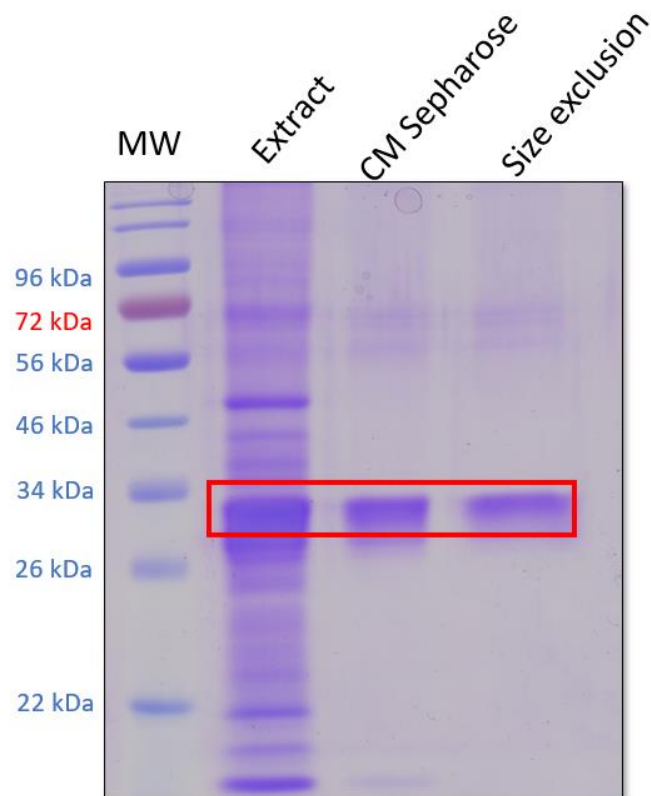


Figure 50. SDS-Electrophoretic gel evaluating the different steps of the purification process. The OXC32 band is highlighted by the red rectangle. As it is shown, after CM Sepharose chromatography the protein is almost pure, and size exclusion chromatography is done to confirm such state of the protein and bring it into the finally desired buffer.

After this combined isolation process, we confirmed that the obtained OXC32 is pure and, in 25mM $(\text{NH}_4)_2\text{SO}_4$, pH 5.2 buffer, ready to be further characterized (i.e. to study it as a proteinase inhibitor and derive the enzymatic parameters on distinct target enzymes). The productive yield of the purification process along the distinct steps, starting from 12 hen eggs has been (by measuring the protein content by the bicinchoninic assay) 52,086mg of protein in the extraction step, 2,012mg of protein after the cation-exchange chromatography and 1,2 mg of protein after the size exclusion chromatography. Overall, 1,205 mg of pure OXC32 were obtained starting from 12 hen eggs. All the data is recollected and represented on Table 6.

TABLE 6 Summary of the purification procedure of natural OCX32

<i>Steps</i>	Protein total (mg)	Concentration (mg/mL)	Inhibitory Activity (U)	Specific inhibitory Activity (U/mg)	Yield (%)	Purification (- fold)
<i>Extract</i>	52.086±7.23	0.08681±0.01205	3.26±0.533	0.0627±0.09	100±0.0	1.0±0.0
<i>Cation exch.</i>	2.0123±0.21	0.05678±0.00954	0.53±0.12	0.2646±0.81	16.3±1.4	4.2±0.7
<i>SEC</i>	1.2055±0.3195	0.02679±0.0071	0.43±0.07	0.3624±0.67	13.4±2.0	5.8±0.6

Data is the mean ± SD

3.3.2. Analysis of the proteolytic capabilities of purified OCX32

Proteases are enzymes which breaks peptide bond, and their activity can be blocked by inhibitors as we described before. This part of the thesis is focused on the inhibitory kinetics analysis of OCX32. As a member of I47 inhibitor family, the same as LXN, so we make the hypothesis that OCX32 could act as competitive reversible inhibitor, probable also tight-binding, as LXN.

Most kinetics studies with competitive reversible inhibitors are usually done with a higher inhibitor concentration than enzyme in the assays, in order to easily detect a significant inhibitory activity. In these conditions, the absolute or relative concentration of inhibitor bound to enzyme is constant and negligible, so free inhibitor concentration is virtually the total inhibitor concentration used in assay (Strelow et al. 2012). In this case, the formula of Michaelis-Menten is useful for K_i determination.

However, competitive reversible tight-binding inhibitors requires a lower concentration of inhibitor, that is inhibitory activity can be detected at equivalent concentration as enzyme, usually in nanomolar range. In this case, the free inhibitor concentration in the assay is no equal to total inhibitor concentration. For this reason, classical formulas are not valid for studying tight-binding inhibitor/enzyme kinetics, and particular analytical strategies, like Morrison's one, are necessary to determine K_i precisely (Morrison 1982).

To evaluate such possibility, the enzymatic study OCX32 as inhibitor must be analysed at different preincubation times of inhibitor with enzyme, i.e. at 37°C, to determine its IC_{50} for each time and, therefore, define whether this is a tight-binding case or not. In the OCX32 case, we selected 5 min and 20 min as the study times (Figure 51).

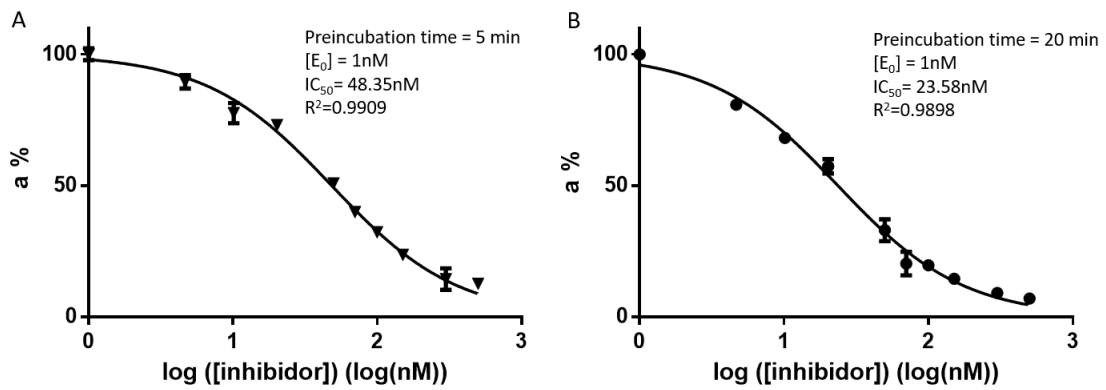


Figure 51. Representation of the of the percentage of activity against \log_{10} ([Inhibidor]) for OXC32. The enzyme used is bCPA1 at 1 nM of concentration. A) Representation of activity loss of bCPA1 at 5 minutes of preincubation at different inhibitor concentration. B) Representation of activity loss of bCPA1 at 20 minutes of preincubation at different inhibitor concentration.

The values IC₅₀ of OXC32 at both analysed times are in the same range of concentration than the active enzyme, so we can define OXC32 as tight-binding inhibitor and its inhibition can be calculated with formula described by Morrison (Morrison, 1982). For more complete analysis, this inhibitor was tested against different proteases available for our laboratory. Some of them were commercial enzymes, and other are recombinant forms expressed by us, so in both cases their titration and determination their active concentrations were important. Noteworthy, this titration was also done on OXC32 from a natural source, eggs, and care had to be taken to avoid that the extraction process could affect its capability as inhibitor. If the titration process is not followed, we only could derive an apparent K_i .

It appeared clearly that carboxypeptidase inhibitory activity of OXC32, with K_i s at nanomolar levels can be demonstrated for different carboxypeptidase variants (see Table 7), OXC32 was also tested against other peptidases in order to define its narrow or large specificity. The proteases tested in this assay were bCPA1, hCPA1, hCPA2, hCPA4, pCPB, hCPD, Subtilisin, Trypsin and Papain. Clearly, OXC32 only display inhibition against M14 MCPs.

TABLE 7 K_i of OCX 32 on different proteases.

Type of proteases	Protease	K_i (nM)
M14A Carboxypeptidase	bCPA1	5.427±0.2615
	hCPA1	2.118±0.2713
	hCPA2	16.28±1.31
	hCPA4	24.6±2.34
	pCPB	6.922±0.416
M14B Carboxypeptidase	hCPD	Not observed
Serine Protease	Subtilisin	Not observed
	Trypsin	Not observed
Cysteine Protease	Papain	Not observed

Data is the $K_i \pm SD$

As described in the table, all the values of K_i were determined for a series of proteases belonging to different catalytic types, being M14A carboxypeptidases the only ones affected, with derived K_i s in nanomolar range. OCX32 is therefore a very powerful inhibitor for these enzymes.

3.3.3. Structural modelling of globular domain of OCX32

As described previously at introduction, together with information on latexin like family inhibitors, OCX32 is an eggshell protein which presents a 32% of sequential homology with latexin. This sequential homology is sufficiently high to predict and model a potential detailed three-dimensional fold, as it can be observed using different online predictors of secondary and tertiary structure as RaptorX (<http://raptorx.uchicago.edu/>) or SWISS-MODEL (<http://swissmodel.expasy.org>).

The latexin (LXN) three-dimensional structure can be described as formed by two similar identical cystatin-like domains, each one with an α -helix over an antiparallel β -sheet, linked together by another α -helix, letting both α -helices buried inside the protein folded. Latexin structure and inhibitory mechanism has been described in detail by our group (Pallarès et al. 2005). Latexin inhibition is primarily caused by an inhibitory loop provided by the central part of the β -sheet of the C-terminal domain which interacts with a region from MCPs in a way reminiscent of the inhibition through prodomain in the zymogen. This loop, which is shaped by the end of strand $\beta 7$ and the beginning of $\beta 8$, is the closest part of the inhibitor that interacts with the protease moiety. This behavior is also found in cystatins, inhibitors of cysteine proteases which also employ a β ribbon structure for inhibition, and with a single domain of this type, as seen in the steffin/papain complex (Stubbs et al. 1990)

We work very hard to try to obtain the three-dimensional crystal structure of OCX32, we realized many attempts to crystallise OCX32 alone or in complex with recombinant hCPA4, but unfortunately, it was not possible. We attributed these failures to the low solubility displayed by OCX32 in standard buffers, even though we assayed many distinct buffers and addition of salts, with some partial success, but none allowed us to get crystals when the crystallisation agents were added, even in the presence of mild detergents.

To overcome this drawback, we derived and built an *in-silico* model of how OCX32 could fold and inhibit MCPs, based on the structures and mechanism already described for LXN and hCPA4.

In order to derive such OCX32 three-dimensional structure model, we followed the protocol described by the online server SWISS-MODEL (Figure 52) (Waterhouse et al. 2018). This is a well-known and trusted web-based integrated service dedicated to protein structure homology modelling. It guides the user in building protein homology models at different levels of complexity. For our model we started with the sequence OCX32 from *Gallus gallus* filed in UniprotKB with code Q90YI1. This sequence was processed on SWISS-MODEL for a blastp to determine proteins sequence with shared homology. All such proteins with shared homology are potential templates for the model.

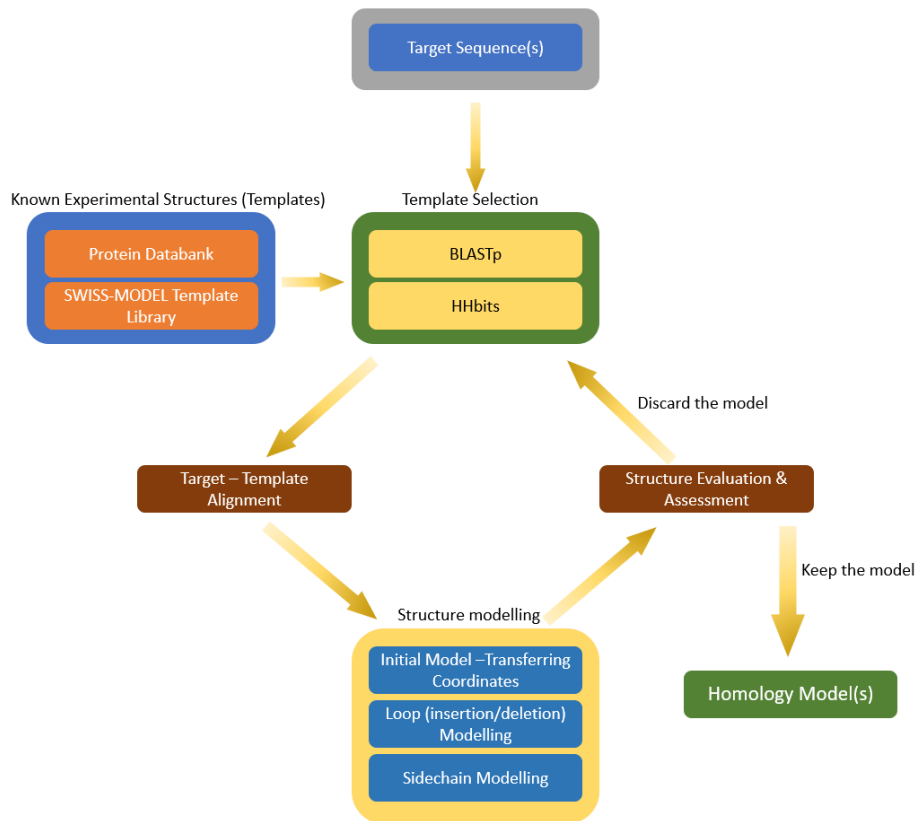


Figure 52: Workflow of homology modelling with the SWISS-MODEL expert system. Building a homology model comprises four main steps, first one is the identification and selection of structural template. The following step is alignment of target sequence and template structure with an extra module specialised on template comparison. The third step is the model building and the final step is the model quality evaluation. Each of the above steps can be repeated, until a satisfying modelling result is achieved. Image reproduced from (Waterhouse et al. 2018)

The results of the search of templates gave us the best score for human LXN (PDB code: 2bo9.1. B). As mentioned, the sequential homology between OCX32 and human LXN is around 32%. This search also shows the predicted secondary structure of both proteins and alignment between both sequences, showing conserved residues, conservative changes, and gaps between sequences. Furthermore, we took into account that the sequential sequential homology of such pair of proteins is particularly rich between residues 17 and 222 of OCX32, approximately, so its potential globular structure, similar to the LXN one, is localised in the centre of the protein (Figure 53).



Figure 53. Alignment of OCX32 sequence and LXN (template) sequence. Alignment was performed using blastp. The figure also shows a secondary structure prediction of OCX32 due its homology to LXN. The query is OCX32 sequence and Sbjct is human Latexin sequence. Blue arrows signal marks key residues on the interaction with hCPA4.

Once the template is selected, the model is built by transferring the atoms coordinates from the template to target protein. After this first step, the second one is the loop formation, on the sequence where no secondary structure is predicted. These regions could present troubles on model building due to their flexibility, or even if loop region sequence differs from template. The final step on model building is sidechain modelling; this process is done following the coordinates from the template. All stereochemical irregularities and clashes introduced in the modelling process are solved by energy minimization process in the refinement of the model.

The predictive model, obtained from the sequence of OCX32 is a three-dimensional structure very similar to LXN, with two cystatins-like domains linked by a-helix (Figure 54.A, B). The topological similitude of both structures is evaluated by root mean square deviation or RMSD, that has a value of 1.203 Å. The quality of this model can also be evaluated by different factors, the most used being the QMEAN (Qualitative Model Energy ANalysis), which is a scoring function describing the major geometrical aspects of protein structures. (Benkert et al. 2011, Studer et al. 2020). On general, for a good value of QMEAN, it should be close to 0. On OCX32 starting model, the QMEAN value is -2.40. (Supplementary Figure 2)

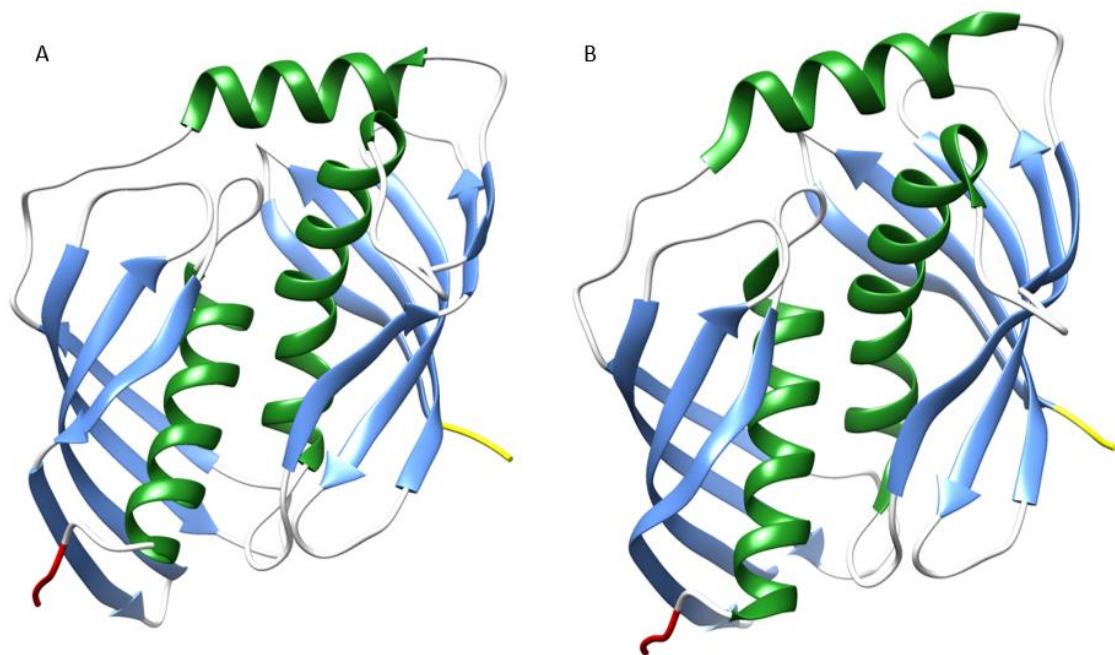


Figure 54. Comparison of three-dimensional structure of LXN and the model OCX32, using LXN as template. A) LXN structure from PDB file 2bo9. B) Modelled OCX32 structure.

3.3.3.1. Modelling the inhibitory mechanism of OCX32 using hCPA4-LXN complex as template

Once we got an initial version model of our inhibitor, we focused on designing how this protein could interact with hCPA4. For the modelling of OCX32-hCPA4 complex, the PDB file of OCX32 was aligned over the structure of LXN interacting hCPA4 in the crystalline structure of the complex (PDB code 2bo9). This is based on for the study of interactions already described for LXN-hCPA4 complex, trying to see whether they could appear for OCX32 theoretically bound to hCPA4 (Figure 55). For this work we use the PyMol software, which also allows us to visualize potential spatial problems on the localization of the side chain from residues.

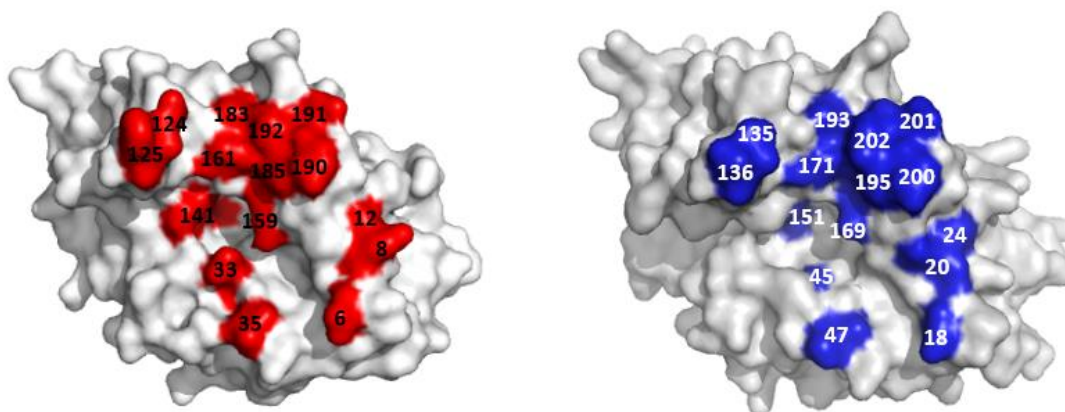


Figure 55. Surface representation of LXN and OCX32. Resalted residues are the responsible of interaction with hCPA4. Molecule with red residues is LXN, and the one with blue residues is OCX32. A) Latexin molecule displaying in red the residues that involved in hCPA4 interaction. B) OCX32 molecule displaying in blue the residues involved in hCPA4.

First, we focused on the most important interaction within LXN-hCPA4. As it is described on the introduction of this chapter, Gln190 from LXN plays a crucial role on the inhibitory mechanism due to its interaction with Arg71 from hCPA4. Interactions specially observed are between the atoms Gln190O and Gln190Oε1 from LXN to Arg71Nη2 of hCPA4. Gln190 also interacts with Tyr248 from hCPA4, which is involved in hydrogen bonding of the P1 amide nitrogen of substrate by its Oη. Gln190 homologous position on OCX32 is Gln200 which in our model also fits in the same position and could have the same interactions forming the crucial salt bridge for the interaction (Figure 56).

Another important residue for interaction is the His185 from LXN. As is described previously in the introduction of chapter III, His185 also plays a crucial role keeping the position of the Gln190 for its interaction with the basic residue of hCPA4 (Arg71). His185 also interacts with other residues but from the hCPA4: His185Cε1 establish a hydrogen bond to Phe279Cζ from hCPA4 and the His185Nδ1 interacts to Tyr198Oη of carboxypeptidase. On OCX32, His195 acts as His185 homologous, at the same distance from the Gln200 which establish the salt bridge with MCP allowing inhibition. In our model also fits for the same interactions (Figure 56).

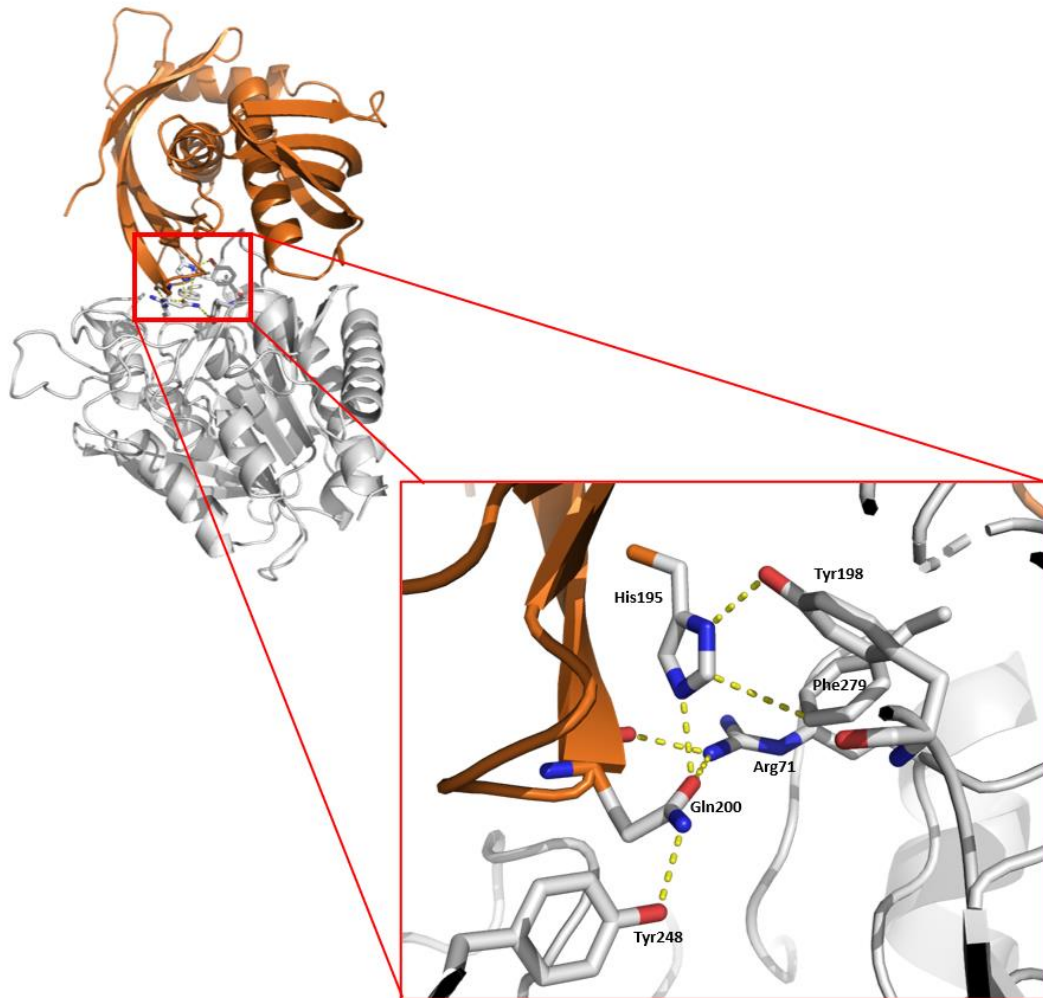


Figure 56. Graphical representation from OXC32 model interacting with hCPA4. Orange molecule is OXC32, and clear grey is hCPA4. On zoom square, there are represented the interaction between the side chains of both proteins on the predicted salt bridge described between.

Once these important interactions described on LXN-hCPA4 complex were also observed on OXC32-hCPA4 inhibition, we focused on the rest of interactions described for the inhibition of hCPA4 by LXN and analysed if they were available for the model of OXC32. (Figure 57) The interaction between Thr6 γ 2 from LXN and Gln239C β from hCPA4 changes. On the OXC32, the residue on homologous position of Thr6 is Ser18. Even being a conservative change (both are polar residues) Ser18 does not present a γ in its side chain, but it presents $O\gamma$ which could form a hydrogen bond with Gln239O ϵ 1 from hCPA4.

On the other hand, Tyr8N from LXN displays a hydrogen bond with Thr245O from hCPA4. On OXC32, the residue with the same coordinates is Arg20, which could do the same interaction with the carboxypeptidase, because the interaction is between the peptide bond on both residues.

The next interaction to be analysed is a hydrogen bond established between Arg12N η 1 and Val247O from hCPA4. The position of residue Arg12, on OCX32 is substituted by Trp24 in LXN. For this case, Trp24 does not fit into the model because it has not a N η 1 to do the hydrogen bond, and this interaction cannot be emulated by OCX32.

Glu33 from LXN has a long side chain which allows interact with two residues from hCPA4. The interactions observed are Glu33O ϵ 2 interacts to Thr274O γ 1 from carboxypeptidase and Glu33O ϵ 2 interacts to Thr276O γ 1 also from hCPA4. On the model of OCX32 interacting hCPA4, the homologous of Glu33 is Ala45. Ala has a smaller side chain than Glu, this causes that interactions previously described for this position, for OCX32 do not exist, making the interaction between inhibitor and carboxypeptidase weaker than LXN.

Also Gln35 of LXN shows an interaction with Glu237 of hCPA4. The specific interaction observed is Gln35N ϵ 2 interacting to Glu237O ϵ 2. On our model of OCX32, the Gln35 is replaced by His47. Even so, the interaction observed on LXN and hCPA4 could be conserved by His47N ϵ 2 from OCX32 to Glu237O ϵ 2.

LXN also interacts by a hydrogen bond between Asn125O and Arg124N ϵ from hCPA4. On the OCX32 Asn125 is replaced by Ser135, this replacement does not affect the interaction due its interaction involves the peptide bond from residue of inhibitor.

Phe126 is a residue involved on some interactions with hCPA4. These interactions are Phe126O interacting with Ala283N from hCPA4, and Phe126C ϵ 2 interacting with Trp73C ζ 3. On OCX32, the Phe126, is substituted by Asn136 and regarding intermolecular interactions, Asn136O interacting to Ala283N from hCPA4 is possible but Asn has a smaller side chain than Phe. Although, we found an interaction with an optimized rotamer of Asn136 from OCX32. So, the interaction is between Asn136O δ 1 from OCX32 and Trp73C ζ 3 from carboxypeptidase.

Trp141 is an LXN residue which interacts with Thr276 from hCPA4. The interaction observed is between their both side chains, Trp141C η 2 interacts with Thr276C γ 2 from hCPA4. On OCX32, the Trp141 position is replaced by Ile151. Its side chain is smaller than side chain of Trp. The interaction is due to a C atom from the two rings of Trp, but Ile does not have that C atom to do interaction. So, OCX32 does not present this interaction on the model.

Lys159 of LXN shows an intermolecular interaction with Thr274 from hCPA4. The interaction observed is Lys159N ζ interacting to Thr274O. The homologous of Lys159 on OCX32 is Leu169. On this case, Leu169 from OCX32 is not able to reproduce this interaction, because the Leu side chain is smaller than Lys side chain, and Leu side chain cannot interact with Thr274.

For Val161 presents an interaction with hCPA4, the interaction is Val161C γ 1 interacting with Leu125C δ 2 from carboxypeptidase. On OCX32, the residue of Val 161 is substituted by Ala171. The Ala171, is not able to maintain interaction because its side chain does not present two carbon on γ position to allow the interaction with Leu125 from hCPA4. Ala is a smaller residue than Val.

Leu183 side chain from LXN interacts with side chain of Leu125 from hCPA4, the interaction is Leu183C δ 1 from inhibitor to Leu125C δ 2 of carboxypeptidase. On OCX32, the residue on homologous position is Leu193. Even Leu193 from OCX32 fits on the coordinates from Leu183 from LXN, due its stereochemical limitations could alter the interaction. For this case, the interaction the carboxypeptidase could also be observed but this interaction is between Leu193C δ 2 from OCX32 and Leu125C δ 2 from hCPA4.

Another interaction is Glu191O ϵ 1 from LXN interacting with Glu163N from hCPA4. For OCX32, Gln201 is localized on the same coordinates than Glu191. The side chains of both residues are similar, so the interaction between Glu201O ϵ 1 from OCX32 and Glu163N from hCPA4 can happen. For Gln201, both interactions described could still being present because one is involved the peptide bond and, in the second one, side chains of both residues (Glu and Gln) are similar, and this allows the interaction OCX32 too.

Ile192 of LXN also interacts with Leu125 from hCPA4. The interaction is Ile192C γ 2 interacting Leu125C δ 2. On OCX32, the homologous of Ile192 is Met202. This change of residue in this position alters the interaction due to Met has a bigger side chain, and this makes more difficult the potential interaction between Met202 from OCX32 and Leu125 from hCPA4, in addition its size generates potential stereochemical problems for its accommodation. Anyway, we found a potential Van der Waals interaction between Met202S δ from OCX32 and Leu125C δ 2 from hCPA4.

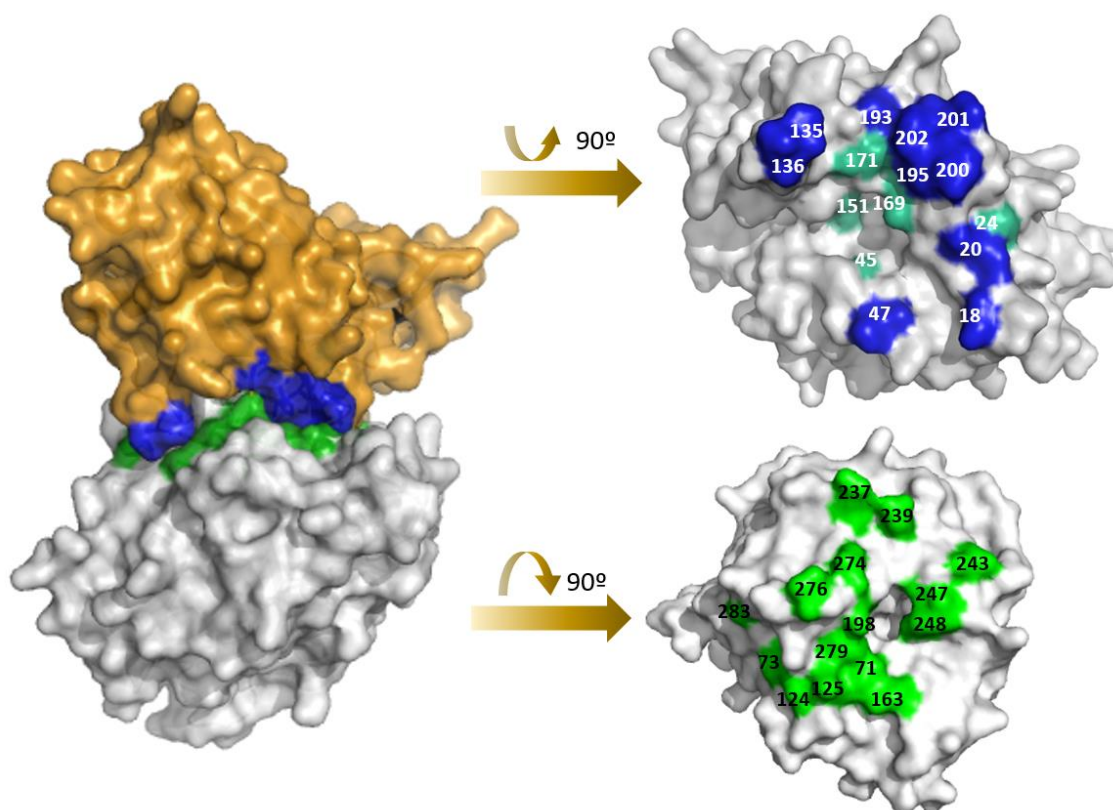


Figure 57. Model of OXC32 inhibition over hCPA4. The complex is the model of the interaction between OXC32 and hCPA4. On the proteins twisted, the one with blue and cyan residues is OXC32. And the protein with green residues is hCPA4. On OXC32, the residues coloured on cyan, are residues that lose their contact in comparison with LXN.

Once all the sidechain from the different residues of OXC32 and hCPA4 has been modelled according their possible interactions, using as template LXN-hCPA4, the model must be evaluated to determine quality. Now, after the model has been already adjusted to LXN and to its mechanism of inhibition, the RMSD value has decreased to 0.141Å. The pdb file obtained from the model adjusted has been submitted to Structural Assessment from the SWISS-MODEL web page, in order to evaluate the quality of this model of the complex of OXC32 and hCPA4 (Supplementary figure 3).

The complex that we modelled was evaluated by a QMEAN value of -0.66. After all the optimization process the value has been improved. Finally, we obtain a *in silico* model of how the inhibitor OXC32 is interacting and inhibiting the hCPA4 (Figure 57).

3.4. DISCUSSION

In the present study we focused on the process of characterization of a protein with capability for different interesting roles, since its extraction and purification from a natural source, followed by experimental functional =enzymological inhibitory= studies, until a *in silico* model for its three-dimensional structure and potential inhibitory interaction with a target enzyme, inhibiting the latter. We divided the results on three principal blocks to describe the work done on it.

The first block has been the extraction and purification of OCX32. In order to guarantee its native state, we tested various extraction procedures. All this work done was difficult because starting from the solid state in which the protein assembles in the cuticle of the egg, associated to the mineral phase of the shell, our goal was to transfer it to the aqueous solution, attaining a soluble state, whilst respecting its functional capabilities. The initially described protocol used 1N HCl for extraction from eggshell, without caring for protein integrity, given that the main goal was to get it in enough quantity for sequence analysis and antibody derivation. We had to tune it to avoid protein denaturation or degradation due to the HCl, in order to characterize, in respectful conditions, some previous indications on its enzymatic inhibitory properties. Also, to get it fully soluble in order to make feasible crystallographic analysis. A positive sign was the fact that when the extract was dialysed against low concentration of $(\text{NH}_4)_2\text{SO}_4$ (at pH 5.4), a salt known as a good folder (Wingfield 2001), both enzymatic inhibition and solubility appeared. Thus, in these conditions OCX32 seemed soluble, purifiable, active and well folded. Even in this situation, OCX32 in solution was still relatively unstable, easily forming aggregates at concentrations over 100ug/mL, particularly when suffering changes in temperature. With all this information, an after testing many distinct saline conditions and effectors (as arginine, heparin, mild detergents ...), we found how to maintain OCX32 reasonable soluble and stable for the activity studies.

In the second block, and once we got the protein more or less stable for biochemical study, we characterised it as an inhibitor of metallocarboxypeptidases (MCPs). Our group at UAB has a large and productive experience on the study of MCPs and their inhibitors, and particularly on tight-binding inhibitors of them. The protein that we study on this chapter, OCX32, is homologous to latexin, already described as an endogenous inhibitor of MCPs (Pallares et al, 2005) and a member of the I47 inhibitor family (LXN family), with which it keeps a significant level of sequence identity (32%). With this information we hypothesised that OCX32, could also play a role as strong, competitive, tight-binding inhibitor of MCPs, as is LXN, besides its primary

roles at the eggshell. Firstly, we analysed its potentially inhibitory character. We tested OCX32 against a good MCP model (bCPA1), both at nanomolar concentration too. This study was done at different preincubation times, 5 min and 20 min, to facilitate the formation of inhibitor-enzyme complex. On both cases we observed that OCX32 shows strong inhibition on bCPA1. Therefore, OCX32 shows inhibitory power against a MCP model, at nanomolar concentration and at the 5-20min range. Once, we know that it could act as inhibitor we focused to prove the potential role of OCX32 as inhibitor on other MCPs, as bCPA1, hCPA1, hCPA2, hCPA4, pCPB, confirming for all of them nanomolar K_i s. So, we can confirm that natural extracted OCX32 from eggshell is a tight-binding inhibitor. A functional question arises from this: could such inhibitory capability participate in the egg microbial defence by the eggshell, or could it be related to another unknown function?

In the third, and final, block of this chapter we focused on the obtention of three-dimensional structure of OCX32, by X-ray crystallography. We tried to obtain the OCX32 in complex with hCPA4, as it was done in our group for LXN (Pallarès et al. 2005). We made many attempts to crystallize both the complex and the isolated inhibitor, but we failed. As we described, OCX32 shows important instability and begins to aggregate when its concentration increases above 100ug/mL, so we tried to add effectors that could help in its solubilization and, therefore, subsequent crystallization (it is well known that highly soluble proteins tend to crystallize easily). Among this, the mild detergent LMNG (Lauryl Maltose Neopentyl Glycol), was one of the most efficient solubilizers for OCX32. Unfortunately, even in its presence we had no success in our crystallization attempts.

As alternative to obtain the crystal structure we designed a computational-based model of OCX32, that is *in silico*. Such model has been derived from the sequence, by alignment to a conformational template, that in our case is LXN. LXN is the one with which we got the best score for sequence homology and, along modelling, the coordinates of its atoms are transferred to the atoms of OCX32, to form the backbone of the protein and determine the secondary and tertiary structures. Once we got the OCX32 model, we focused on the conformational alignment of both LXN and OCX32 3D structures over hCPA4 structure. Identifying the main residues of LXN which interact with hCPA4, we can extrapolate some of the main specific interactions between residues on the homologous positions at OCX32 and hCPA4, taking care of the side chains rotamers of the involved residues. We observed that most of the interactions, could be conserved between OCX32 and hCPA4, from the LXN-hCPA4 complex, but others are lost, debilitating the interaction between OCX32 and hCPA4, or being substituted by alternative ones. The lost of contacts between residues could explain the difference between K_i on both proteins.

on the same target MCP (i.e. CPA4). The final modelled structure of OCX32 in complex with hCPA4 has been evaluated to determine its quality, with a good result.

Overall, the smooth modelling of OCX32 over LXN, as well as the good fitting of such model on CPA4, mimicking the experimental complex and crystal structure of LXN with such carboxypeptidase, gives confidence that actually the structure and inhibitory capability of OCX32 are prepared to act as such in a real scenario, that there are significant probabilities that such protein acts as a regular, and strong, inhibitor of MCPs, and uses such capability in a related real role in nature. Anyway, it justifies the interest of keeping investigating on this protein and on its potential natural role as a metallo-carboxypeptidase inhibitor, possibly regulating the activity of one of these enzymes, as it has been already proposed for the structurally related protein RARRES1, a potential tumour protector protein and natural binder of AGL2/CCP2, a cytosolic metallocarboxypeptidase (Sahab et al. 2011).

CONCLUDING REMARKS

CONCLUDING REMARKS

CHAPTER I

1. The ALDH1A3-C314S mutant, from human has been cloned, recombinantly overexpressed and purified from *Escherichia coli*.
2. This ALDH1A3 C314S mutant keeps about 1% of activity in comparison to the wild type ALDH1A3, as it has already been observed on other ALDH (Farrés et al. 1995).
3. The activity of the C314S mutant is still affected by the addition of the DEAB inhibitor, in a similar way as it happens with the ALDH1A3 *wt*.
4. The ALDH1A3 C314S mutant oligomerises in homotetramers as also ALDH1A3 *wt* does.
5. The structure of the monomer from ALDH1A3-C314S mutant, with NAD⁺ as a cofactor, has been solved by X-ray crystallography at high resolution (1.75Å)
6. The isolated protein has been found as a homodimer in the crystal structure, but there is evidence that also oligomerises on homotetramers.
7. The crystal structure of the monomer interacting with cofactor showed different well defined hydrogen bonds, some of them already reported on the bibliography, but there are new hydrogen bonds involving more interacting residues.

CHAPTER II

8. The study of charge distribution on the surface of proteins with basic pI, indicates the occurrence of positively charged patches on them. Such are potentially binding regions for the heparin.
9. The study of the cytotoxicity of heparin on HEK 293F culture cells has determined that the best concentration of heparin to be used in cell cultures for long times is 50UI/mL.
10. An easy and inexpensive protocol has been developed to improve the recombinant production of heparin-affinity carboxypeptidases in mammalian cell culture by the heparin-affinity based approach, by supplementing the culture medium with heparin.

CHAPTER III

11. The Ovocalyxin-32 (OCX32) protein has been identified by MALDI-TOF peptide mass fingerprinting (PMF), from eggshell extract.
12. OCX32, in native, pure and soluble state, has been isolated from a natural source, the eggshell, for its functional characterization.
13. OCX32 shows a strong inhibitory activity against M14A metalloproteases (MCPs). It acts as tight-binding inhibitor and all K_i determined for MCPs are in the nanomolar range.
14. The naturally extracted OCX32 is unstable at concentrations above to 100 μ g/mL, in standard buffers and conditions.
15. The three-dimensional structure of OCX32, alone and in complex with hCPA4, has been modelled *in silico* using latexin (LXN) and LXN-hCPA4 three-dimensional structures as templates.
16. The crucial residues of LXN for its interaction with hCPA4, are conserved in the homologous OCX32. Such residues are Glu 190/His185 in LXN, and Glu200/His195 in OCX32, respectively.
17. The rest of interactions described in the LXN-hCPA4 complex, also be observed in the modelled interaction between OCX32 and hCPA4, probably due to sequence and conformational evolutionary conservation among these proteins.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Abbernante, G. and D. Fairlie (2005). "Protease inhibitors in the clinic." Med Chem **1**(1): 71-104.
2. Adams, P., P. Afonine, G. Bunkóczi, V. Chen, I. Davis, N. Echols, J. Headd, L. Hung, G. Kapral, R. Grosse-Kunstleve, A. McCoy, N. Moriarty, R. Oeffner, R. Read, D. Richardson, J. Richardson, T. Terwilliger and P. Zwart (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." Acta Crystallogr D Biol Crystallogr **66**(Pt 2): 213-221.
3. Alonso-del-Rivero, M., S. Trejo, M. Reytor, M. Rodriguez-de-la-Vega, J. Delfin, J. Diaz, Y. Gonzalez-Gonzalez, F. Canals, M. Chavez and F. Aviles (2012). "Tri-domain bifunctional inhibitor of metallocarboxypeptidases A and serine proteases isolated from marine annelid *Sabellastarte magnifica*." J Biol Chem **287**(19): 15427-15438.
4. Aloy, P., V. Companys, J. Vendrell, F. Aviles, L. Fricker, M. Coll and F. Gomis-Ruth (2001). "The crystal structure of the inhibitor-complexed carboxypeptidase D domain II and the modeling of regulatory carboxypeptidases." J Biol Chem **276**(19): 16177-16184.
5. Álvarez-Santos, S., A. González-Lafont, J. Lluch, B. Oliva and F. Avilés (1998). "Theoretical study of the role of arginine 127 in the water-promoted mechanism of peptide cleavage by carboxypeptidase A." New J Chem **22**(4): 319-325.
6. Anderson, T. A., D. G. Levitt and L. J. Banaszak (1998). "The structural basis of lipid interactions in lipovitellin, a soluble lipoprotein." Structure **6**(7): 895-909.
7. Arolas, J., S. Bronsoms, S. Ventura, F. Aviles and J. Calvete (2006). "Characterizing the tick carboxypeptidase inhibitor: molecular basis for its two-domain nature." J Biol Chem **281**(32): 22906-22916.
8. Arolas, J., J. Lorenzo, A. Rovira, J. Castellà, F. Aviles and C. Sommerhoff (2005). "A carboxypeptidase inhibitor from the tick *Rhipicephalus bursa*: isolation, cDNA cloning, recombinant expression, and characterization." J Biol Chem **280**(5): 3441-3448.
9. Arolas, J., G. Popowicz, J. Lorenzo, C. Sommerhoff, R. Huber, F. Aviles and T. Holak (2005). "The three-dimensional structures of tick carboxypeptidase inhibitor in complex with A/B carboxypeptidases reveal a novel double-headed binding mode." J Mol Biol **350**(3): 489-498.
10. Arolas, J., J. Vendrell, F. Aviles and L. Fricker (2007). "Metallo-carboxypeptidases: Emerging Drug Targets in Biomedicine." Curr Pharm Des **13**(4): 349-366.
11. Barrett, A. (1994). "Classification of peptidases." Methods Enzymol **244**: 1-15.
12. Barrett, A., N. Rawlings and J. Woessner (2012). Handbook of of Proteolytic Enzymes. London, UK, Elsevier Academic Press.

BIBLIOGRAPHY

13. Bayés, A., M. Comellas-Bigler, M. de la Vega, K. Maskos, W. Bode, F. Aviles, M. Jongsma, J. Beekwilder and J. Vendrell (2005). "Structural basis of the resistance of an insect carboxypeptidase to plant protease inhibitors." *102*(46): 16602-16607.
14. Benkert, P., M. Biasini and T. Schwede (2011). "Toward the estimation of the absolute quality of individual protein structure models." *Bioinformatics* **27**(3): 343-350.
15. Bieth, J. (1995). *Theoretical and Practical Aspects of Proteinase Inhibition Kinetics. Methods in Enzymology: Proteolytic Enzymes: Aspartic and Metallo Peptidases*. A. J. Barrett, Academic Press. **248**: 59-84.
16. Bingle, C. and C. Craven (2004). "Meet the relatives: a family of BPI- and LBP-related proteins." *Trends Immunol* **25**(2): 53-55.
17. Bitman, J., T. Wrenn, D. Wood, G. Mustakas, E. Baker and W. Wolf (1975). "Effects of feeding formaldehyde treated, full fat soybean flours on milk fat polyunsaturated fatty acids." *J Am Oil Chem Soc* **52**(10): 415-418.
18. Blanco-Aparicio, C., M. Molina, E. Fernández-Salas, M. Frazier, J. Mas, E. Querol, F. Avilés and R. de Llorens (1998). "Potato carboxypeptidase inhibitor, a T-knot protein, is an epidermal growth factor antagonist that inhibits tumor cell growth." *J Biol Chem* **273**(20): 12370-12377.
19. Bode, W. and R. Huber (2000). "Structural basis of the endoproteinase protein inhibitor interaction." *Biochim Biophys Acta* **1477**(1-2): 241-252.
20. Booth, S. L., T. Johns and H. V. Kuhnlein (1992). "Natural Food Sources of Vitamin A and Provitamin A." *Food and Nutrition Bulletin* **14**(1): 1-15.
21. Bouma, B. and L. Mosnier (2004). "Thrombin Activatable Fibrinolysis Inhibitor (TAFI) at the Interface between Coagulation and Fibrinolysis." *Pathophysiol Haemost Thromb* **33**: 375-381.
22. Bousif, O., F. Lezoualc'h, M. Zanta, M. Mergny, D. Scherman, B. Demeneix and J. Behr (1995). "A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine." *Proc Natl Acad Sci U S A* **92**(16): 7297-7301.
23. Bradbury, S. and W. Jakoby (1971). "Ordered binding of substrates to yeast aldehyde dehydrogenase." *J Biol Chem* **246**(6): 1834-1840.
24. Bradford, M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." *Anal Biochem* **72**: 248-254.
25. Castañeda, O., V. Sotolongo, A. Amor, R. Stöcklin, A. Anderson, A. Harvey, A. Engström, C. Wernstedt and E. Karlsson (1995). "Characterization of a potassium channel toxin from the Caribbean Sea anemone *Stichodactyla helianthus*." *Toxicon* **33**(5): 603-613.

BIBLIOGRAPHY

26. Chao, Y., S. Liou, S. Tsai and S. Yin (1993). "Dominance of the mutant ALDH2(2) allele in the expression of human stomach aldehyde dehydrogenase-2 activity." Proc Natl Sci Counc Repub China B **17**(3): 98-102.
27. Chen, X., W. Wu and J. Ding (2014). "Aberrant TIG1 methylation associated with its decreased expression and clinicopathological significance in hepatocellular carcinoma." Tumour Biol **35**(2): 967-971.
28. Christianson, D. and W. Lipscomb (1989). "Carboxypeptidase A." Acc Chemi Res **22**(2): 62-69.
29. Clark, D. and K. Palle (2016). "Aldehyde dehydrogenases in cancer stem cells: potential as therapeutic targets." Annals of translational medicine **4**(24): 518-518.
30. Coll, M., A. Guasch, F. Avilés and R. Huber (1991). "Three-dimensional structure of porcine procarboxypeptidase B: a structural basis of its inactivity." Embo j **10**(1): 1-9.
31. Compton, S. and C. Jones (1985). "Mechanism of dye response and interference in the Bradford protein assay." Anal Biochem **151**(2): 369-374.
32. Cool, D., E. Normant, F. Shen, H. Chen, L. Pannell, Y. Zhang and Y. Loh (1997). "Carboxypeptidase E is a regulated secretory pathway sorting receptor: genetic obliteration leads to endocrine disorders in Cpe(fat) mice." Cell **88**(1): 73-83.
33. Copeland, R. (2000). Kinetics of Single-Substrate Enzyme Reactions. Enzymes: A practical Introduction to Structure, Mechanism, and Data Analysis, Wiley: 109-145.
34. Cordeiro, C., H. Esmaili, G. Ansah and M. Hincke (2013). "Ovocalyxin-36 is a pattern recognition protein in chicken eggshell membranes." PloS one **8**(12): e84112-e84112.
35. Cory, A., T. Owen, J. Barltrop and J. Cory (1991). "Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture." Cancer Commun **3**(7): 207-212.
36. Covalada, G., M. Alonso-del-Rivero, M. Chavez, F. Aviles and D. Reverter (2012). "Crystal structure of novel metallo-carboxypeptidase inhibitor from marine mollusk *Nerita versicolor* in complex with human carboxypeptidase A4." J Biol Chem **287**(12): 9250-9258.
37. Deddish, P., R. Skidgel, V. Kriho, X. Li, R. Becker and E. Erdős (1990). "Carboxypeptidase M in Madin-Darby canine kidney cells. Evidence that carboxypeptidase M has a phosphatidylinositol glycan anchor." J Biol Chem **265**(25): 15083-15089.
38. Dennis, J., S. Xiao, M. Agarwal, D. Fink, A. Heuer and A. Caplan (1996). "Microstructure of matrix and mineral components of eggshells from White Leghorn chickens (*Gallus gallus*)." Journal of Morphology **228**(3): 287-306.
39. Dominguez-Vera, J., J. Gautron, J. Garcia-Ruiz and Y. Nys (2000). "The effect of avian uterine fluid on the growth behavior of calcite crystals." Poult Sci **79**(6): 901-907.

BIBLIOGRAPHY

40. Duan, J., J. Cai, Y. Guo, X. Bian and S. Yu (2016). "ALDH1A3, a metabolic target for cancer diagnosis and therapy." *139*(5): 965-975.
41. Duester, G. (2000). "Families of retinoid dehydrogenases regulating vitamin A function." *European Journal of Biochemistry* **267**(14): 4315-4324.
42. Eisenstein, M. (2005). "Westward expansion." *Nature Methods* **2**(10): 796-796.
43. Emsley, P., B. Lohkamp, W. Scott and K. Cowtan (2010). "Features and development of Coot." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 4): 486-501.
44. Etches, R. (1996). *Reproduction in poultry*, Wallingford : CAB international.
45. Everts, H., J. Sundberg, L. King, Jr and D. Ong (2007). "Immunolocalization of enzymes, binding proteins, and receptors sufficient for retinoic acid synthesis and signaling during the hair cycle." *J Invest Dermatol* **127**(7): 1593-1604.
46. Farrés, J., T. Wang, S. Cunningham and H. Weiner (1995). "Investigation of the Active Site Cysteine Residue of Rat Liver Mitochondrial Aldehyde Dehydrogenase by Site-Directed Mutagenesis." *Biochemistry* **34**(8): 2592-2598.
47. Fatma, N., E. Kubo, L. Chylack, Jr., T. Shinohara, Y. Akagi and D. Singh (2004). "LEDGF regulation of alcohol and aldehyde dehydrogenases in lens epithelial cells: stimulation of retinoic acid production and protection from ethanol toxicity." *Am J Physiol Cell Physiol* **287**(2): C508-516.
48. Fazekas de St Groth, S., R. Webster and A. Datyner (1963). "Two new staining procedures for quantitative estimation of proteins on electrophoretic strips." *Biochim Biophys Acta* **71**: 377-391.
49. Fernández, D., I. Pallarès, G. Covalada, F. Avilés and J. Vendrell (2013). "Metalloprotease inhibitors and their inhibitors: recent developments in biomedically relevant protein and organic ligands." *Curr Med Chem* **20**(12): 1595-1608.
50. Fife, T. and T. Przystas (1986). "Divalent metal ion catalysis in amide hydrolysis. The hydrolysis of N-acylimidazoles." *J Am Chem Soc* **108**(15): 4631-4636.
51. Freeman, C., J. Harding, D. Quigley and P. Rodger (2010). "Structural Control of Crystal Nuclei by an Eggshell Protein." *Angew Chem Int Ed* **49**(30): 5135-5137.
52. Freeman, C., J. Harding, D. Quigley and P. Rodger (2012). "Protein binding on stepped calcite surfaces: Simulations of ovocleidin-17 on calcite {31.16} and {31.8}." *Physical chemistry chemical physics : PCCP* **14**: 7287-7295.
53. Fricker, L. (2013). Metalloprotease D. *Handbook of Proteolytic Enzymes*. N. Rawlings and G. Salvesen, Academic Press: 1353-1357.

BIBLIOGRAPHY

54. Fricker, L. (2018). Carboxypeptidase E and the Identification of Novel Neuropeptides as Potential Therapeutic Targets. Apprentices to Genius: A tribute to Solomon H. Snyder. S. Bryant. San Diego, Calif, Kruze, Z. **82**.
55. Fricker, L. and E. Leiter (1999). "Peptides, enzymes and obesity: new insights from a 'dead' enzyme." Trends Biochem Sci **24**(10): 390-393.
56. Fritz, K. and D. Petersen (2013). "An overview of the chemistry and biology of reactive aldehydes." Free Radic Biol Med **59**: 85-91.
57. Fujinaga, M., M. Chernaia, N. Tarasova, S. Mosimann and M. James (1995). "Crystal structure of human pepsin and its complex with pepstatin." Protein Sci **4**(5): 960-972.
58. Garcia-Castellanos, R., R. Bonet-Figueredo, I. Pallares, S. Ventura, F. Aviles, J. Vendrell and F. Gomis-Rutha (2005). "Detailed molecular comparison between the inhibition mode of A/B-type carboxypeptidases in the zymogen state and by the endogenous inhibitor latexin." Cell Mol Life Sci **62**(17): 1996-2014.
59. Garcia-Guerrero, M., J. Garcia-Pardo, E. Berenguer, R. Fernandez-Alvarez, G. Barfi, P. Lyons, F. Aviles, R. Huber, J. Lorenzo and D. Reverter (2018). "Crystal structure and mechanism of human carboxypeptidase O: Insights into its specific activity for acidic residues." Proc Natl Acad Sci U S A **115**(17): E3932-e3939.
60. Garcia-Pardo, J. (2015). Structural and functional characterization of regulatory metallocarboxypeptidases studies on human carboxypeptidases D and Z, and the transthyretin-like domain. PhD Doctoral Thesis, Universitat Autònoma de Barcelona.
61. Garcia-Pardo, J., S. Tanco, M. Garcia-Guerrero, S. Dasgupta, F. Avilés, J. Lorenzo and L. Fricker (2020). "Substrate Specificity and Structural Modeling of Human Carboxypeptidase Z: A Unique Protease with a Frizzled-Like Domain." International Journal of Molecular Sciences **21**(22).
62. Gardell, S., C. Craik, E. Clauser, E. Goldsmith, C. Stewart, M. Graf and W. Rutter (1988). "A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family." J Biol Chem **263**(33): 17828-17836.
63. Gautron, J., M. Hincke, K. Mann, M. Panheleux, M. Bain, M. McKee, S. Solomon and Y. Nys (2001). "Ovocalyxin-32, a novel chicken eggshell matrix protein. isolation, amino acid sequencing, cloning, and immunocytochemical localization." J Biol Chem **276**(42): 39243-39252.
64. Gautron, J., M. Hincke and Y. Nys (1997). "Precursor matrix proteins in the uterine fluid change with stages of eggshell formation in hens." Connect Tissue Res **36**(3): 195-210.
65. Gill, F. (2007). Ornithology. Ornithology. W. Freeman, Macmillan: 361.
66. Goedde, H. and D. Agarwal (1990). "Pharmacogenetics of aldehyde dehydrogenase (ALDH)." Pharmacol Ther **45**(3): 345-371.

BIBLIOGRAPHY

67. Gomis-Ruth, F. (2008). "Structure and mechanism of metallo-carboxypeptidases." Crit Rev Biochem Mol Biol **43**(5): 319-345.
68. Gomis-Rüth, F., V. Companys, Y. Qian, L. Fricker, J. Vendrell, F. Avilés and M. Coll (1999). "Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallo-carboxypeptidase subfamily." Embo j **18**(21): 5817-5826.
69. Goulas, A., E. L. Triplett and G. Taborsky (1996). "Oligophosphopeptides of varied structural complexity derived from the egg phosphoprotein, phosvitin." Journal of Protein Chemistry **15**(1): 1-9.
70. Graham, C., K. Brocklehurst, R. Pickersgill and M. Warren (2006). "Characterization of retinaldehyde dehydrogenase 3." Biochem J **394**(Pt 1): 67-75.
71. Guasch, A., M. Coll, F. Avilés and R. Huber (1992). "Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation." J Mol Biol **224**(1): 141-157.
72. Gurbuz, H., A. Durukan, H. Sevim, E. Ergin, A. Gurpınar and C. Yorgancıoğlu (2013). "Heparin toxicity in cell culture: a critical link in translation of basic science to clinical practice." Blood Coagul Fibrinolysis **24**(7): 742-745.
73. Hallberg, L., L. Hultén and E. Gramatkovski (1997). "Iron absorption from the whole diet in men: how effective is the regulation of iron absorption?" The American journal of clinical nutrition **66**(2): 347-356.
74. Hallberg, L. and L. Hulthén (2000). "Prediction of dietary iron absorption: an algorithm for calculating absorption and bioavailability of dietary iron." The American journal of clinical nutrition **71**(5): 1147-1160.
75. Han, J., L. Yang and R. Puri (2005). "Analysis of target genes induced by IL-13 cytotoxin in human glioblastoma cells." J Neurooncol **72**(1): 35-46.
76. Handa, T., A. Katayama, T. Yokobori, A. Yamane, T. Fujii, S. Obayashi, S. Kurozumi, R. Kawabata-Iwakawa, N. Gombodorj, M. Nishiyama, T. Asao, K. Shirabe, H. Kuwano and T. Oyama (2019). "Carboxypeptidase A4 accumulation is associated with an aggressive phenotype and poor prognosis in triple-negative breast cancer." Int J Oncol **54**(3): 833-844.
77. Hartley, B. (1960). "Proteolytic enzymes." Annu Rev Biochem **29**: 45-72.
78. Hass, G., H. Nau, K. Biemann, D. Grahn, L. Ericsson and H. Neurath (1975). "The amino acid sequence of a carboxypeptidase inhibitor from potatoes." Biochemistry **14**(6): 1334-1342.
79. Hedstrom, L. (2002). "Introduction: Proteases." Chemical Reviews **102**(12): 4429-4430.

BIBLIOGRAPHY

80. Heidarieh, M., M. Maragheh, M. Shamami, M. Behgar, F. Ziaei and Z. Akbari (2013). "Evaluate of heavy metal concentration in shrimp (*Penaeus semisulcatus*) and crab (*Portunus pelagicus*) with INAA method." Springerplus **2**(1): 72.
81. Hempel, J., J. Perozich, T. Chapman, J. Rose, J. Boesch, Z. Liu, R. Lindahl and B. Wang (1999). Aldehyde Dehydrogenase Catalytic Mechanism. Enzymology and Molecular Biology of Carbonyl Metabolism 7. H. Weiner, E. Maser, D. Crabb and R. Lindahl. Boston, MA, Springer US: 53-59.
82. Hendriks, D. (2004). Carboxypeptidase U. Handbook of Proteolytic Enzymes. A. Barrett, N. Rawlings and J. J. Woesner. London, Elsevier. **1**: 825-828.
83. Hernández-Hernández, A., J. Gómez-Morales, A. Rodríguez-Navarro, J. Gautron, Y. Nys and J. García-Ruiz (2008). "Identification of Some Active Proteins in the Process of Hen Eggshell Formation." Crystal Growth & Design **8**(12): 4330-4339.
84. Hernández-Hernández, A., M. Vidal, J. Gómez-Morales, A. Rodríguez-Navarro, V. Labas, J. Gautron, Y. Nys and J. García Ruiz (2008). "Influence of eggshell matrix proteins on the precipitation of calcium carbonate (CaCO₃)." Journal of Crystal Growth **310**(7-9): 1754-1759.
85. Hilditch, T. P. and P. N. Williams (1964). The chemical constitution of natural fats, London: Chapman & Hall.
86. Hincke, M., J. Gautron, K. Mann, M. Panhéleux, M. McKee, M. Bain, S. Solomon and Y. Nys (2003). "Purification of ovocalyxin-32, a novel chicken eggshell matrix protein." Connect Tissue Res **44 Suppl 1**: 16-19.
87. Hincke, M., J. Gautron, C. Tsang, M. McKee and Y. Nys (1999). "Molecular Cloning and Ultrastructural Localization of the Core Protein of an Eggshell Matrix Proteoglycan, Ovocleidin-116." **274**(46): 32915-32923.
88. Hincke, M., Y. Nys, J. Gautron, K. Mann, A. Rodriguez-Navarro and M. McKee (2012). "The eggshell: structure, composition and mineralization." Front Biosci (Landmark Ed) **17**(4): 1266-1280.
89. Hincke, M., C. Tsang, M. Courtney, V. Hill and R. Narbaitz (1995). "Purification and immunochemistry of a soluble matrix protein of the chicken eggshell (ovocleidin 17)." Calcified Tissue International **56**(6): 578-583.
90. Hincke, M., O. Wellman-Labadie, M. McKee, J. Gautron, Y. Nys and K. Mann (2008). Biosynthesis and structural assembly of eggshell components. Egg Bioscience and Biotechnology. Y. Mine. Chichester, John Wiley: 97-127.
91. Hourdou, M., M. Guinand, M. Vacheron, G. Michel, L. Denoroy, C. Duez, S. Englebert, B. Joris, G. Weber and J. Ghuysen (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." Biochem J **292 (Pt 2)**(Pt 2): 563-570.

BIBLIOGRAPHY

92. Hsu, L., W. Chang, L. Hiraoka and C. Hsieh (1994). "Molecular cloning, genomic organization, and chromosomal localization of an additional human aldehyde dehydrogenase gene, ALDH6." Genomics **24**(2): 333-341.
93. Hsu, L., W. Chang and A. Yoshida (2000). "Mouse type-2 retinaldehyde dehydrogenase (RALDH2): genomic organization, tissue-dependent expression, chromosome assignment and comparison to other types." Biochim Biophys Acta **1492**(1): 289-293.
94. Huang, X. and J. Groves (2017). "Beyond ferryl-mediated hydroxylation: 40 years of the rebound mechanism and C-H activation." Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry **22**(2-3): 185-207.
95. Jia, J., H. Parikh, W. Xiao, J. Hoskins, H. Pflücke, X. Liu, I. Collins, W. Zhou, Z. Wang, J. Powell, S. Thorgeirsson, U. Rudloff, G. Petersen and L. Amundadottir (2013). "An integrated transcriptome and epigenome analysis identifies a novel candidate gene for pancreatic cancer." BMC Med Genomics **6**: 33.
96. Jiménez, R., R. Pequerul, A. Amor, J. Lorenzo, K. Metwally, F. Avilés, X. Parés and J. Farrés (2019). "Inhibitors of aldehyde dehydrogenases of the 1A subfamily as putative anticancer agents: Kinetic characterization and effect on human cancer cells." Chemico-Biological Interactions **306**: 123-130.
97. Jing, C., M. El-Ghany, C. Beesley, C. Foster, P. Rudland, P. Smith and Y. Ke (2002). "Tazarotene-induced gene 1 (TIG1) expression in prostate carcinomas and its relationship to tumorigenicity." J Natl Cancer Inst **94**(7): 482-490.
98. Johnson, A. (2015). *Reproduction in the Female*. Sturkie's Avian Physiology: Sixth Edition, Elsevier Inc.: 635-665.
99. Juanhuix, J., F. Gil-Ortiz, G. Cuní, C. Colldelram, J. Nicolás, J. Lidón, E. Boter, C. Ruget, S. Ferrer and J. Benach (2014). "Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron." Journal of synchrotron radiation **21**(Pt 4): 679-689.
100. Kalinina, E., R. Biswas, I. Berezniuk, A. Hermoso, F. Aviles and L. Fricker (2007). "A novel subfamily of mouse cytosolic carboxypeptidases." FASEB J **21**(3): 836-850.
101. Kanski, J., A. Behring, J. Pelling and C. Schöneich (2005). "Proteomic identification of 3-nitrotyrosine-containing rat cardiac proteins: effects of biological aging." Am J Physiol Heart Circ Physiol **288**(1): H371-381.
102. Kim, H., J. Lapointe, G. Kaygusuz, D. Ong, C. Li, M. van de Rijn, J. Brooks and J. Pollack (2005). "The retinoic acid synthesis gene ALDH1a2 is a candidate tumor suppressor in prostate cancer." Cancer Res **65**(18): 8118-8124.
103. Kim, H., S. Yang, Y. Lee, S. Do, I. Chung and J. Yang (2003). "High-level expression of human glycosyltransferases in insect cells as biochemically active form." Biochemical and Biophysical Research Communications **305**(3): 488-493.

BIBLIOGRAPHY

104. Kin Ting Kam, R., Y. Deng, Y. Chen and H. Zhao (2012). "Retinoic acid synthesis and functions in early embryonic development." Cell & Bioscience **2**(1): 11.
105. King, G. and R. Holmes (1997). "Human corneal and lens aldehyde dehydrogenases. Purification and properties of human lens ALDH1 and differential expression as major soluble proteins in human lens (ALDH1) and cornea (ALDH3)." Adv Exp Med Biol **414**: 19-27.
106. Kiosseoglou, V. (2004). "Interactions and competitive adsorption effects in egg-based products." World's Poultry Science Journal **60**(3): 311-320.
107. Klyosov, A., L. Rashkovetsky, M. Tahir and W. Keung (1996). "Possible role of liver cytosolic and mitochondrial aldehyde dehydrogenases in acetaldehyde metabolism." Biochemistry **35**(14): 4445-4456.
108. Koppaka, V., D. Thompson, Y. Chen, M. Ellermann, K. Nicolaou, R. Juvonen, D. Petersen, R. Deitrich, T. Hurley and V. Vasiliou (2012). "Aldehyde dehydrogenase inhibitors: a comprehensive review of the pharmacology, mechanism of action, substrate specificity, and clinical application." Pharmacol Rev **64**(3): 520-539.
109. Kos, J., T. Vižin, U. Fonović and A. Pišlar (2015). "Intracellular signaling by cathepsin X: molecular mechanisms and diagnostic and therapeutic opportunities in cancer." Semin Cancer Biol **31**: 76-83.
110. Kovacs-Nolan, J., C. Cordeiro, D. Young, Y. Mine and M. Hincke (2014). "Ovocalyxin-36 is an effector protein modulating the production of proinflammatory mediators." Veterinary Immunology and Immunopathology **160**(1): 1-11.
111. Krupenko, S. and N. Oleinik (2002). "10-formyltetrahydrofolate dehydrogenase, one of the major folate enzymes, is down-regulated in tumor tissues and possesses suppressor effects on cancer cells." Cell Growth Differ **13**(5): 227-236.
112. Kuhn, B., J. Benz, M. Greif, A. Engel, H. Sobek and M. Rudolph (2013). "The structure of human α -2,6-sialyltransferase reveals the binding mode of complex glycans." Acta Crystallogr D Biol Crystallogr **69**(9): 1826-1838.
113. Kulshreshtha, G., A. Rodriguez-Navarro, E. Sanchez-Rodriguez, T. Diep and M. Hincke (2018). "Cuticle and pore plug properties in the table egg." Poult Sci **97**(4): 1382-1390.
114. Kurosawa, N., T. Hamamoto, Y. Lee, T. Nakaoka, N. Kojima and S. Tsuji (1994). "Molecular Cloning and Expression of Gal NAc α 2,6-Sialyltransferase." J Biol Chem **269**(2): 1402-1409.
115. Laemmli, U. (1970). "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4." Nature **227**(5259): 680-685.
116. Laskowski, M. J. and I. Kato (1980). "Protein inhibitors of proteinases." Annu Rev Biochem **49**: 593-626.

BIBLIOGRAPHY

117. Lesnierowski, G., R. Cegielska-Radziejewska and J. Kijowski (2004). "Thermally and chemical-thermally modified lysozyme and its bacteriostatic activity." World's Poultry Science Journal **60**(3): 303-310.
118. Li, X. and R. Skidgel (1999). "Release of glycosylphosphatidylinositol-anchored carboxypeptidase M by phosphatidylinositol-specific phospholipase C upregulates enzyme synthesis." Biochem Biophys Res Commun **258**(1): 204-210.
119. Li, Y., D. Zhang, W. Jin, C. Shao, P. Yan, C. Xu, H. Sheng, Y. Liu, J. Yu, Y. Xie, Y. Zhao, D. Lu, D. Nebert, D. Harrison, W. Huang and L. Jin (2006). "Mitochondrial aldehyde dehydrogenase-2 (ALDH2) Glu504Lys polymorphism contributes to the variation in efficacy of sublingual nitroglycerin." J Clin Invest **116**(2): 506-511.
120. Lineweaver, H. and C. Murray (1947). "Identification of the trypsin inhibitor of egg white with ovomucoid." J Biol Chem **171**(2): 565-581.
121. Liu, Q., L. Yu, J. Gao, Q. Fu, J. Zhang, P. Zhang, J. Chen and S. Zhao (2000). "Cloning, tissue expression pattern and genomic organization of latexin, a human homologue of rat carboxypeptidase A inhibitor." Mol Biol Rep **27**(4): 241-246.
122. Liu, Z., Y. Sun, J. Rose, Y. Chung, C. Hsiao, W. Chang, I. Kuo, J. Perozich, R. Lindahl, J. Hempel and B. Wang (1997). "The first structure of an aldehyde dehydrogenase reveals novel interactions between NAD and the Rossmann fold." Nat Struct Biol **4**(4): 317-326.
123. Longo, P., J. Kavran, M. Kim and D. Leahy (2013). "Transient mammalian cell transfection with polyethylenimine (PEI)." Methods Enzymol **529**: 227-240.
124. López-Otín, C. and L. Matrisian (2007). "Emerging roles of proteases in tumour suppression." Nat Rev Cancer **7**(10): 800-808.
125. Lu, P., A. Blesch and M. Tuszynski (2004). "Induction of bone marrow stromal cells to neurons: Differentiation, transdifferentiation, or artifact?" Journal of Neuroscience Research **77**(2): 174-191.
126. Lundequist, A., E. Tchougounova, M. Abrink and G. Pejler (2004). "Cooperation between mast cell carboxypeptidase A and the chymase mouse mast cell protease 4 in the formation and degradation of angiotensin II." J Biol Chem **279**(31): 32339-32344.
127. Lyons, P., M. Callaway and L. Fricker (2008). "Characterization of carboxypeptidase A6, an extracellular matrix peptidase." J Biol Chem **283**(11): 7054-7063.
128. Lyons, P. and L. Fricker (2010). "Substrate Specificity of Human Carboxypeptidase A6." The Journal of biological chemistry **285**: 38234-38242.
129. Lyons, P. and L. Fricker (2011). "Carboxypeptidase O is a glycosylphosphatidylinositol-anchored intestinal peptidase with acidic amino acid specificity." J Biol Chem **286**(45): 39023-39032.

BIBLIOGRAPHY

130. Mann, K., J. Olsen, B. Macek, F. Gnad and M. Mann (2007). "Phosphoproteins of the chicken eggshell calcified layer." Proteomics **7**(1): 106-115.
131. Mann, K. and F. Siedler (1999). "The amino acid sequence of ovocleidin 17, a major protein of the avian eggshell calcified layer." Biochem Mol Biol Int **47**(6): 997-1007.
132. Marchitti, S., C. Brocker, D. Stagos and V. Vasiliou (2008). "Non-P450 aldehyde oxidizing enzymes: the aldehyde dehydrogenase superfamily." Expert Opin Drug Metab Toxicol **4**(6): 697-720.
133. Marino-Buslje, C., M. Molina, F. Canals, F. Avilés and E. Querol (1994). "Overproduction of a recombinant carboxypeptidase inhibitor by optimization of fermentation conditions." Applied Microbiology and Biotechnology **41**(6): 632-637.
134. Marx, P. (2004). "Thrombin-activatable fibrinolysis inhibitor." Curr Med Chem **11**(17): 2335-2348.
135. McCoy, A., R. Grosse-Kunstleve, P. Adams, M. Winn, L. Storoni and R. Read (2007). "Phaser crystallographic software." Journal of Applied Crystallography **40**(4): 658--674.
136. McGwire, G., R. Becker and R. Skidgel (1999). "Carboxypeptidase M, a glycosylphosphatidylinositol-anchored protein, is localized on both the apical and basolateral domains of polarized Madin-Darby canine kidney cells." J Biol Chem **274**(44): 31632-31640.
137. Mebs, D. and E. Gebauer (1980). "Isolation of proteinase inhibitory, toxic and hemolytic polypeptides from a sea anemone. *Stoichactis* sp." Toxicon **18**(1): 97-106.
138. Metz, M., A. Piliponsky, C. Chen, V. Lammel, M. Abrink, G. Pejler, M. Tsai and S. Galli (2006). "Mast cells can enhance resistance to snake and honeybee venoms." Science **313**(5786): 526-530.
139. Meyer, T. and B. Lamberts (1965). "Use of coomassie brilliant blue R250 for the electrophoresis of microgram quantities of parotid saliva proteins on acrylamide-gel strips." Biochimica et Biophysica Acta **107**(1): 144-145.
140. Michaelis, L. and M. Menten (1913). "Die Kinetik der Invertinwirkung." Biochem Z **49**: 333-369.
141. Mills, P., E. Struys, C. Jakobs, B. Plecko, P. Baxter, M. Baumgartner, M. Willemsen, H. Omran, U. Tacke, B. Uhlenberg, B. Weschke and P. Clayton (2006). "Mutations in antiquitin in individuals with pyridoxine-dependent seizures." Nat Med **12**(3): 307-309.
142. Mine, Y. (1995). "Recent advances in the understanding of egg white protein functionality." Trends in Food Science & Technology **6**(7): 225-232.

BIBLIOGRAPHY

143. Molotkov, A., N. Molotkova and G. Duester (2006). "Retinoic acid guides eye morphogenetic movements via paracrine signaling but is unnecessary for retinal dorsoventral patterning." Development **133**(10): 1901-1910.
144. Molotkova, N., A. Molotkov and G. Duester (2007). "Role of retinoic acid during forebrain development begins late when Raldh3 generates retinoic acid in the ventral subventricular zone." Dev Biol **303**(2): 601-610.
145. Moreb, J., D. Mohuczy, B. Ostmark and J. Zucali (2007). "RNAi-mediated knockdown of aldehyde dehydrogenase class-1A1 and class-3A1 is specific and reveals that each contributes equally to the resistance against 4-hydroperoxycyclophosphamide." Cancer Chemother Pharmacol **59**(1): 127-136.
146. Moreb, J., D. Ucar, S. Han, J. Amory, A. Goldstein, B. Ostmark and L. Chang (2012). "The enzymatic activity of human aldehyde dehydrogenases 1A2 and 2 (ALDH1A2 and ALDH2) is detected by Aldefluor, inhibited by diethylaminobenzaldehyde and has significant effects on cell proliferation and drug resistance." Chem Biol Interact **195**(1): 52-60.
147. Moretti, A., J. Li, S. Donini, R. Sobol, M. Rizzi and S. Garavaglia (2016). "Crystal structure of human aldehyde dehydrogenase 1A3 complexed with NAD⁺ and retinoic acid." Scientific Reports **6**(1): 35710.
148. Morgan, C. and T. Hurley (2015). "Development of a high-throughput in vitro assay to identify selective inhibitors for human ALDH1A1." Chem Biol Interact **234**: 29-37.
149. Morgan, C., B. Parajuli, C. Buchman, K. Dria and T. Hurley (2015). "N,N-diethylaminobenzaldehyde (DEAB) as a substrate and mechanism-based inhibitor for human ALDH isoenzymes." Chem Biol Interact **234**: 18-28.
150. Morrison, J. (1982). "The Slow-Binding and Slow, Tight-Binding Inhibition of Enzyme-Catalyzed Reactions." Trends in Biochemical Sciences **7**(3): 102-105.
151. Mory, A., F. Ruiz, E. Dagan, E. Yakovtseva, A. Kurolap, X. Parés, J. Farrés and R. Gershoni-Baruch (2014). "A missense mutation in ALDH1A3 causes isolated microphthalmia/anophthalmia in nine individuals from an inbred Muslim kindred." European Journal of Human Genetics **22**(3): 419-422.
152. Mosmann, T. (1983). "Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays." J Immunol Methods **65**(1-2): 55-63.
153. Muñoz, A., N. Dominguez-Gasca, C. Jimenez-Lopez and A. Rodriguez-Navarro (2015). "Importance of eggshell cuticle composition and maturity for avoiding trans-shell Salmonella contamination in chicken eggs." Food Control **55**: 31-38.
154. Nagpal, S., S. Patel, A. Asano, A. Johnson, M. Duvic and R. Chandraratna (1996). "Tazarotene-Induced Gene 1 (TIG1), a Novel Retinoic Acid Receptor-Responsive Gene in Skin." J Invest Dermatol **106**(2): 269-274.

BIBLIOGRAPHY

155. Nagpal, S., S. Thacher, S. Patel, S. Friant, M. Malhotra, J. Shafer, G. Krasinski, A. Asano, M. Teng, M. Duvic and R. Chandraratna (1996). "Negative regulation of two hyperproliferative keratinocyte differentiation markers by a retinoic acid receptor-specific retinoid: insight into the mechanism of retinoid action in psoriasis." Cell Growth Differ **7**(12): 1783-1791.
156. Ni, L., S. Sheikh and H. Weiner (1997). "Involvement of glutamate 399 and lysine 192 in the mechanism of human liver mitochondrial aldehyde dehydrogenase." The Journal of biological chemistry **272**(30): 18823-18826.
157. Niederreither, K., V. Subbarayan, P. Dollé and P. Chambon (1999). "Embryonic retinoic acid synthesis is essential for early mouse post-implantation development." Nat Genet **21**(4): 444-448.
158. Normant, E., C. Gros and J. Schwartz (1995). "Carboxypeptidase A Isoforms Produced by Distinct Genes or Alternative Splicing in Brain and Other Extrapancreatic Tissues." **270**(35): 20543-20549.
159. Normant, E., M. Martres, J. Schwartz and C. Gros (1995). "Purification, cDNA cloning, functional expression, and characterization of a 26-kDa endogenous mammalian carboxypeptidase inhibitor." Proc Natl Acad Sci U S A **92**(26): 12225-12229.
160. Novikova, E., F. Eng, L. Yan, Y. Qian and L. Fricker (1999). "Characterization of the Enzymatic Properties of the First and Second Domains of Metallo-carboxypeptidase D." **274**(41): 28887-28892.
161. Novikova, E., S. Reznik, O. Varlamov and L. Fricker (2000). "Carboxypeptidase Z Is Present in the Regulated Secretory Pathway and Extracellular Matrix in Cultured Cells and in Human Tissues." Journal of Biological Chemistry **275**(7): 4865-4870.
162. Núñez, C., E. Oliveira, J. García-Pardo, M. Diniz, J. Lorenzo, J. Capelo and C. Lodeiro (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**: 115-122.
163. Nys, Y., J. Gautron, J. Garcia-Ruiz and M. Hincke (2004). "Avian eggshell mineralization: biochemical and functional characterization of matrix proteins." Comptes Rendus Palevol **3**(6-7): 549-562.
164. Nys, Y. and N. Guyot (2011). Egg formation and chemistry. Improving the Safety and Quality of Eggs and Egg Products. Y. Nys, M. Bain and F. Van Immerseel, Woodhead Publishing: 83-132.
165. Nys, Y., M. Hincke, J. Arias, J. Garcia-Ruiz and S. Solomon (1999). "Avian Eggshell Mineralization." Avian and Poultry Biology Reviews **10**: 143-166.
166. Nys, Y. and B. Sauveur (2004). "Valeur nutritionnelle des œufs." INRA Prod. Anim. **17**(5): 385-393.

BIBLIOGRAPHY

167. O'Brien, P., A. Siraki and N. Shangari (2005). "Aldehyde sources, metabolism, molecular toxicity mechanisms, and possible effects on human health." Crit Rev Toxicol **35**(7): 609-662.
168. Okamura, S., C. Ng, K. Koyama, Y. Takei, H. Arakawa, M. Monden and Y. Nakamura (1999). "Identification of seven genes regulated by wild-type p53 in a colon cancer cell line carrying a well-controlled wild-type p53 expression system." Oncol Res **11**(6): 281-285.
169. Okubo, T., S. Akachi and H. Hatta (1997). Structure of hen eggs and physiology of egg laying. Hen Eggs: Their Basic and Applied Science. T. Yamamoto, L. Juneja, H. Hatta and M. Kim. New York, CRC Press: 1-12.
170. Oleinik, N., N. Krupenko, D. Priest and S. Krupenko (2005). "Cancer cells activate p53 in response to 10-formyltetrahydrofolate dehydrogenase expression." Biochem J **391**(Pt 3): 503-511.
171. Oleinik, N. and S. Krupenko (2003). "Ectopic expression of 10-formyltetrahydrofolate dehydrogenase in A549 cells induces G1 cell cycle arrest and apoptosis." Mol Cancer Res **1**(8): 577-588.
172. Pallarès, I., R. Bonet, R. Garcia-Castellanos, S. Ventura, F. Aviles, J. Vendrell and F. Gomis-Ruth (2005). "Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin." Proc Natl Acad Sci U S A **102**(11): 3978-3983.
173. Pappa, A., D. Brown, Y. Koutalos, J. DeGregori, C. White and V. Vasiliou (2005). "Human aldehyde dehydrogenase 3A1 inhibits proliferation and promotes survival of human corneal epithelial cells." J Biol Chem **280**(30): 27998-28006.
174. Pascual, R., F. Burgos, M. Salva, F. Soriano, E. Mendez and F. Avilés (1989). "Purification and properties of five different forms of human procarboxypeptidases." Eur J Biochem **179**(3): 609-616.
175. Pei, Y., R. Reins and A. McDermott (2006). "Aldehyde dehydrogenase (ALDH) 3A1 expression by the human keratocyte and its repair phenotypes." Exp Eye Res **83**(5): 1063-1073.
176. Peng, Z., R. Shen, Y. Li, K. Teng, C. Shapiro and H. Lin (2012). "Epigenetic repression of RARRES1 is mediated by methylation of a proximal promoter and a loss of CTCF binding." PLoS One **7**(5): e36891.
177. Pequerul, R., J. Vera, J. Giménez-Dejóz, I. Crespo, J. Coines, S. Porté, C. Rovira, X. Parés and J. Farrés (2020). "Structural and kinetic features of aldehyde dehydrogenase 1A (ALDH1A) subfamily members, cancer stem cell markers active in retinoic acid biosynthesis." Arch Biochem Biophys **681**: 108256.
178. Phillips, M., R. Fletterick and W. Rutter (1990). "Arginine 127 stabilizes the transition state in carboxypeptidase." J Biol Chem **265**(33): 20692-20698.
179. Phillips, M. and W. Rutter (1996). "Role of the prodomain in folding and secretion of rat pancreatic carboxypeptidase A1." Biochemistry **35**(21): 6771-6776.

BIBLIOGRAPHY

180. Pizzuti, A., G. Calabrese, M. Bozzali, L. Telvi, E. Morizio, V. Guida, V. Gatta, L. Stuppia, A. Ion, G. Palka and B. Dallapiccola (2002). "A Peptidase Gene in Chromosome 8q Is Disrupted by A Balanced Translocation in a Duane Syndrome Patient." Invest Ophthalmol Vis Sci **43**(12): 3609-3612.
181. Pollock, C. and S. Orosz (2002). "Avian reproductive anatomy, physiology and endocrinology." Vet Clin North Am Exot Anim Pract **5**(3): 441-474.
182. Portolano, N., P. Watson, L. Fairall, C. Millard, C. Milano, Y. Song, S. Cowley and J. Schwabe (2014). "Recombinant protein expression for structural biology in HEK 293F suspension cells: a novel and accessible approach." J Vis Exp(92): e51897.
183. Price, N. and L. Stevens (1999). Fundamentals of Enzymology: The Cell and Molecular Biology of Catalytic Proteins. USA, Oxford University Press.
184. Radzicka, A. and R. Wolfenden (1995). "A proficient enzyme." Science **267**(5194): 90-93.
185. Rahman, M. A., Baoyindeliger, A. Iwasawa and N. Yoshizaki (2007). "Mechanism of chalaza formation in quail eggs." Cell and tissue research **330**(3): 535-543.
186. Rawlings, N., A. Barrett, P. Thomas, X. Huang, A. Bateman and R. Finn (2018). "The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database." Nucleic Acids Res **46**(D1): D624-d632.
187. Rawlings, N., D. Tolle and A. Barrett (2004). "Evolutionary families of peptidase inhibitors." Biochem J **378**(Pt 3): 705-716.
188. Rawlings, N. D., A. J. Barrett, P. D. Thomas, X. Huang, A. Bateman and R. D. Finn (2018). "The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database." Nucleic Acids Research **46**(D1): D624-D632.
189. Reverter, D., I. García-Sáez, L. Catasús, J. Vendrell, M. Coll and F. Avilés (1997). "Characterisation and preliminary X-ray diffraction analysis of human pancreatic procarboxypeptidase A2." FEBS Lett **420**(1): 7-10.
190. Reverter, D., J. Vendrell, F. Canals, J. Horstmann, F. Avilés, H. Fritz and C. Sommerhoff (1998). "A carboxypeptidase inhibitor from the medical leech *Hirudo medicinalis*. Isolation, sequence analysis, cDNA cloning, recombinant expression, and characterization." J Biol Chem **273**(49): 32927-32933.
191. Rexer, B., W. Zheng and D. Ong (2001). "Retinoic acid biosynthesis by normal human breast epithelium is via aldehyde dehydrogenase 6, absent in MCF-7 cells." Cancer research **61**(19): 7065-7070.
192. Reynolds, D., D. Gurley, R. Stevens, D. Sugarbaker, K. Austen and W. Serafin (1989). "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison of the

BIBLIOGRAPHY

protein with mouse mast cell carboxypeptidase A and rat pancreatic carboxypeptidases." Proc Natl Acad Sci U S A **86**(23): 9480-9484.

193. Reznik, S. and L. Fricker (2001). "Carboxypeptidases from A to Z: implications in embryonic development and Wnt binding." Cell Mol Life Sci **58**(12-13): 1790-1804.

194. Rijken, D. and H. Lijnen (2009). "New insights into the molecular mechanisms of the fibrinolytic system." J Thromb Haemost **7**(1): 4-13.

195. Rizzo, W. and G. Carney (2005). "Sjögren-Larsson syndrome: diversity of mutations and polymorphisms in the fatty aldehyde dehydrogenase gene (ALDH3A2)." Hum Mutat **26**(1): 1-10.

196. Rodriguez de la Vega, M., R. Sevilla, A. Hermoso, J. Lorenzo, S. Tanco, A. Diez, L. Fricker, J. Bautista and F. Avilés (2007). "Nna1-like proteins are active metalloproteinases of a new and diverse M14 subfamily." Faseb j **21**(3): 851-865.

197. Romanoff, A. and A. Romanoff (1949). The avian egg, New York : John Wiley & Sons, Inc. : London: Chapman & Hall, Ltd.

198. Roy, A., M. Ramalinga, O. Kim, J. Chijioke, S. Lynch, S. Byers and D. Kumar (2017). "Multiple roles of RARRES1 in prostate cancer: Autophagy induction and angiogenesis inhibition." PLoS One **12**(7): e0180344.

199. Ryan, C. (1989). "Proteinase inhibitor gene families: strategies for transformation to improve plant defenses against herbivores." Bioessays **10**(1): 20-24.

200. Ryan, C., G. Hass and R. Kuhn (1974). "Purification and properties of a carboxypeptidase inhibitor from potatoes." J Biol Chem **249**(17): 5495-5499.

201. Sahab, Z., M. Hall, Y. Me Sung, S. Dakshanamurthy, Y. Ji, D. Kumar and S. Byers (2011). "Tumor suppressor RARRES1 interacts with cytoplasmic carboxypeptidase AGLL2 to regulate the alpha-tubulin tyrosination cycle." Cancer Res **71**(4): 1219-1228.

202. Sanglas, L., J. Arolas, Z. Valnickova, F. Aviles, J. Enghild and F. Gomis-Rüth (2010). "Insights into the molecular inactivation mechanism of human activated thrombin-activatable fibrinolysis inhibitor." **8**(5): 1056-1065.

203. Sanglas, L., F. Aviles, R. Huber, F. Gomis-Ruth and J. Arolas (2009). "Mammalian metalloproteinase inhibition at the defense barrier of *Ascaris* parasite." Proc Natl Acad Sci U S A **106**(6): 1743-1747.

204. Sanglas, L., Z. Valnickova, J. Arolas, I. Pallarés, T. Guevara, M. Solà, T. Kristensen, J. Enghild, F. Avilés and F. Gomis-Rüth (2008). "Structure of Activated Thrombin-Activatable Fibrinolysis Inhibitor, a Molecular Link between Coagulation and Fibrinolysis." Mol Cells **31**(4): 598-606.

BIBLIOGRAPHY

205. Sapio, M. and L. Fricker (2014). "Carboxypeptidases in disease: insights from peptidomic studies." Proteomics Clin Appl **8**(5-6): 327-337.
206. Saw, Y., J. Yang, S. Ng, S. Liu, S. Singh, M. Singh, W. Welch, H. Tsuda, W. Fong, D. Thompson, V. Vasiliou, R. Berkowitz and S. Ng (2012). "Characterization of aldehyde dehydrogenase isozymes in ovarian cancer tissues and sphere cultures." BMC Cancer **12**: 329.
207. Sawaguchi, S., N. Ohkubo and T. Matsubara (2006). "Identification of Two Forms of Vitellogenin-derived Phosvitin and Elucidation of Their Fate and Roles During Oocyte Maturation in the Barfin Flounder." J Zoological Science(11): 1021-1029, 1029.
208. Schechter, I. and A. Berger (1967). "On the size of the active site in proteases. I. Papain." Biochemical and Biophysical Research Communications **27**(2): 157-162.
209. Schmitt, S. (2005). "From eggs to fossils: epigenesis and transformation of species in Pander's biology." The International Journal of Developmental Biology **49**(1): 1-8.
210. Schusser, B., E. Collarini, H. Yi, S. Izquierdo, J. Fesler, D. Pedersen, K. Klasing, B. Kaspers, W. Harriman, M. van de Lavoie, R. Etches and P. Leighton (2013). "Immunoglobulin knockout chickens via efficient homologous recombination in primordial germ cells." Proceedings of the National Academy of Sciences **110**(50): 20170.
211. Scriver, C., A. Beaudet, W. Sly, D. Valle, B. Childs, K. Kinzler and B. Vogelstein (2001). The Metabolic and Molecular Bases of Inherited Disease. Biochemistry (Moscow). New-York, McGraw-Hill. **67**.
212. Sencic, L. and P. Macek (1990). "New method for isolation of venom from the sea anemone *Actinia cari*. Purification and characterization of cytolytic toxins." Comp Biochem Physiol B **97**(4): 687-693.
213. Seuss-baum, I. (2007). Nutritional Evaluation of Egg Compounds. Bioactive Egg Compounds. R. Huopalahti, R. López-Fandiño, M. Anton and R. Schade. Berlin, Heidelberg, Springer Berlin Heidelberg: 117-144.
214. Shapiro, A., E. Vinuela and J. J. Maizel (1967). "Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels." Biochem Biophys Res Commun **28**(5): 815-820.
215. Sheikh, S., L. Ni, T. Hurley and H. Weiner (1997). "The potential roles of the conserved amino acids in human liver mitochondrial aldehyde dehydrogenase." The Journal of biological chemistry **272**(30): 18817-18822.
216. Shimamori, Y., Y. Kumagai, Y. Watanabe and Y. Fujimoto (1990). "Human placental carboxypeptidase M is anchored by a glycosyl-phosphatidylinositol moiety." Biochem Int **20**(3): 607-613.
217. Shiomi, K., E. Tanaka, H. Yamanaka and T. Kikuchi (1985). "Isolation and characterization of a lethal hemolysin in the sea anemone *Parasicyonis actinostoloides*." Toxicon **23**(5): 865-874.

BIBLIOGRAPHY

218. Shutoh, M., N. Oue, P. Aung, T. Noguchi, K. Kuraoka, H. Nakayama, K. Kawahara and W. Yasui (2005). "DNA methylation of genes linked with retinoid signaling in gastric carcinoma: expression of the retinoid acid receptor beta, cellular retinol-binding protein 1, and tazarotene-induced gene 1 genes is associated with DNA methylation." Cancer **104**(8): 1609-1619.
219. Sidyelyeva, G. and L. Fricker (2002). "Characterization of Drosophila carboxypeptidase D." J Biol Chem **277**(51): 49613-49620.
220. Skidgel, R. (1996). Zinc Metalloproteases In Health And Disease, CRC Press.
221. Skidgel, R. and E. Erdos (2004). Lysine carboxypeptidase. Handbook of Proteolytic Enzymes. A. Barrett, N. Rawlings and J. Woessner. London, Elsevier. **1**: 837-840.
222. Sládek, N. (2003). "Human aldehyde dehydrogenases: potential pathological, pharmacological, and toxicological impact." J Biochem Mol Toxicol **17**(1): 7-23.
223. Sládek, N., R. Kollander, L. Sreerama and D. Kiang (2002). "Cellular levels of aldehyde dehydrogenases (ALDH1A1 and ALDH3A1) as predictors of therapeutic responses to cyclophosphamide-based chemotherapy of breast cancer: a retrospective study. Rational individualization of oxazaphosphorine-based cancer chemotherapeutic regimens." Cancer Chemother Pharmacol **49**(4): 309-321.
224. Smith, P., R. Krohn, G. Hermanson, A. Mallia, F. Gartner, M. Provenzano, E. Fujimoto, N. Goeke, B. Olson and D. Klenk (1985). "Measurement of protein using bicinchoninic acid." Anal Biochem **150**(1): 76-85.
225. Sołobodowska, S., J. Giebułtowicz, R. Wolinowska and P. Wroczyński (2012). "Contribution of ALDH1A1 isozyme to detoxification of aldehydes present in food products." Acta Pol Pharm **69**(6): 1380-1383.
226. Sonawane, N., F. Szoka, Jr and A. Verkman (2003). "Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine-DNA polyplexes." J Biol Chem **278**(45): 44826-44831.
227. Song, L. and L. Fricker (1997). "Cloning and expression of human carboxypeptidase Z, a novel metalloproteinase." J Biol Chem **272**(16): 10543-10550.
228. Springman, E., M. Dikov and W. Serafin (1995). "Mast Cell Procarboxypeptidase A: Molecular modeling and biochemical characterization of its processing within secretory granules." **270**(3): 1300-1307.
229. Stadelman, W. and D. Pratt (1989). "Factors influencing composition of the hen's egg." World Poultry Sci J **45**(3): 247-266.
230. Stevens, M. and S. Olsen (1993). "Comparative analysis of using MTT and XTT in colorimetric assays for quantitating bovine neutrophil bactericidal activity." Journal of Immunological Methods **157**(1): 225-231.

BIBLIOGRAPHY

231. Strelow, J., W. Dewe, P. Iversen, H. Brooks, J. Radding, J. McGee and J. Weidner (2012). Mechanism of Action Assays for Enzymes. Assay Guidance Manual. S. Markossian, G. Sittampalam, A. Grossman et al. Bethesda (MD), Eli Lilly & Company and the National Center for Advancing Translational Sciences: 1-24.
232. Stubbs, M., B. Laber, W. Bode, R. Huber, R. Jerala, B. Lenarcic and V. Turk (1990). "The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction." EMBO J **9**(6): 1939-1947.
233. Studer, G., C. Rempfer, A. Waterhouse, R. Gumienny, J. Haas and T. Schwede (2020). "QMEANDisCo-distance constraints applied on model quality estimation." Bioinformatics **36**(6): 1765-1771.
234. Superti, F., M. Ammendolia, F. Berlutti and P. Valenti (2007). Ovotransferrin. Bioactive Egg Compounds. R. Huopalahti, R. López-Fandiño, M. Anton and R. Schade. Berlin, Springer.
235. Suzuki, R., T. Shintani, H. Sakuta, A. Kato, T. Ohkawara, N. Osumi and M. Noda (2000). "Identification of RALDH-3, a novel retinaldehyde dehydrogenase, expressed in the ventral region of the retina." Mech Dev **98**(1-2): 37-50.
236. Tan, F., S. Balsitis, J. Black, A. Blöchl, J. Mao, R. Becker, D. Schacht and R. Skidgel (2003). "Effect of mutation of two critical glutamic acid residues on the activity and stability of human carboxypeptidase M and characterization of its signal for glycosylphosphatidylinositol anchoring." Biochem J **370**(Pt 2): 567-578.
237. Tanco, S., J. Lorenzo, J. Garcia-Pardo, S. Degroeve, L. Martens, F. Aviles, K. Gevaert and P. Van Damme (2013). "Proteome-derived peptide libraries to study the substrate specificity profiles of carboxypeptidases." Mol Cell Proteomics **12**(8): 2096-2110.
238. Tanco, S., O. Tort, H. Demol, F. Aviles, K. Gevaert, P. Van Damme and J. Lorenzo (2015). "C-terminomics screen for natural substrates of cytosolic carboxypeptidase 1 reveals processing of acidic protein C termini." Mol Cell Proteomics **14**(1): 177-190.
239. Tanco, S., X. Zhang, C. Morano, F. Aviles, J. Lorenzo and L. Fricker (2010). "Characterization of the substrate specificity of human carboxypeptidase A4 and implications for a role in extracellular peptide processing." J Biol Chem **285**(24): 18385-18396.
240. Tellechea, M., J. Garcia-Pardo, J. Cotabarren, D. Lufrano, L. Bakas, F. Avilés, W. Obregon, J. Lorenzo and S. Tanco (2016). "Microplate Assay to Study Carboxypeptidase A Inhibition in Andean Potatoes." Bio-Protocol **6**.
241. Ternes, W. (2008). Naturwissenschaftliche Grundlagen der Lebensmittelzubereitung. Germany, Behrs Verlag.
242. Thompson, J. R. and L. J. Banaszak (2002). "Lipid-Protein Interactions in Lipovitellin." Biochemistry **41**(30): 9398-9409.

BIBLIOGRAPHY

243. Tian, X., J. Gautron, P. Monget and G. Pascal (2010). "What makes an egg unique? Clues from evolutionary scenarios of egg-specific genes." Biol Reprod **83**(6): 893-900.
244. Tipton, K. (2018) "Enzyme Nomenclature News: Translocases (EC 7): A new EC Class."
245. Tort, O., S. Tanco, C. Rocha, I. Bièche, C. Seixas, C. Bosc, A. Andrieux, M. Moutin, F. Avilés, J. Lorenzo and C. Janke (2014). "The cytosolic carboxypeptidases CCP2 and CCP3 catalyze posttranslational removal of acidic amino acids." Mol Biol Cell **25**(19): 3017-3027.
246. Towbin, H., T. Staehelin and J. Gordon (1979). "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications." **76**(9): 4350-4354.
247. Turk, B. (2006). "Targeting proteases: successes, failures and future prospects." Nat Rev Drug Discov **5**(9): 785-799.
248. Ueshima, Y., Y. Matsuda, M. Tsutsumi and A. Takada (1993). "Role of the aldehyde dehydrogenase-1 isozyme in the metabolism of acetaldehyde." Alcohol Alcohol Suppl **1b**: 15-19.
249. Valenzuela-Soto, E. and R. Muñoz-Clares (1993). "Betaine-aldehyde Dehydrogenase from Leaves of *Amaranthushypochondriacus* L. Exhibits an Iso Ordered Bi Bi Steady State Mechanism." J Biol Chem **268**: 23818-23824.
250. van der Laan, S., G. Dubra and K. Rogowski (2019). "Tubulin glutamylation: a skeleton key for neurodegenerative diseases." **14**(11): 1899-1900.
251. Vasiliou, V. and A. Pappa (2000). "Polymorphisms of Human Aldehyde Dehydrogenases." Pharmacology **61**(3): 192-198.
252. Vasiliou, V., A. Pappa and T. Estey (2004). "Role of human aldehyde dehydrogenases in endobiotic and xenobiotic metabolism." Drug Metab Rev **36**(2): 279-299.
253. Vendrell, J. and F. Avilés (1999). Carboxypeptidases. Proteases New Perspectives. V. Turk. Basel, Birkhäuser Basel: 13-34.
254. Vendrell, J., F. Avilés and L. Fricker (2004). Metallocoarboxypeptidases. Handbook of Metalloproteins. A. Messerschmidt, R. Huber, T. Poulas et al.
255. Vendrell, J., E. Querol and F. Avilés (2000). "Metallocoarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties." Biochim Biophys Acta **1477**(1-2): 284-298.
256. Vilcacundo, R., P. Méndez, W. Reyes, H. Romero, A. Pinto and W. Carrillo (2018). "Antibacterial Activity of Hen Egg White Lysozyme Denatured by Thermal and Chemical Treatments." Sci Pharm **86**(4).

BIBLIOGRAPHY

257. Vink, T., M. Oudshoorn-Dickmann, M. Roza, J. Reitsma and R. de Jong (2014). "A simple, robust and highly efficient transient expression system for producing antibodies." Methods **65**(1): 5-10.
258. Vonrhein, C., C. Flensburg, P. Keller, A. Sharff, O. Smart, W. Paciorek, T. Womack and G. Bricogne (2011). "Data processing and analysis with the autoPROC toolbox." Acta Crystallographica Section D **67**(4): 293--302.
259. Wagner, E., T. Luo and U. Dräger (2002). "Retinoic acid synthesis in the postnatal mouse brain marks distinct developmental stages and functional systems." Cereb Cortex **12**(12): 1244-1253.
260. Wagner, E., T. Luo, Y. Sakai, L. Parada and U. Dräger (2006). "Retinoic acid delineates the topography of neuronal plasticity in postnatal cerebral cortex." Eur J Neurosci **24**(2): 329-340.
261. Wang, X., P. Penzes and J. Napoli (1996). "Cloning of a cDNA encoding an aldehyde dehydrogenase and its expression in Escherichia coli. Recognition of retinal as substrate." J Biol Chem **271**(27): 16288-16293.
262. Wang, X., H. Saso, T. Iwamoto, W. Xia, Y. Gong, L. Pusztai, W. Woodward, J. Reuben, S. Warner, D. Bearss, G. Hortobagyi, M. Hung and N. Ueno (2013). "TIG1 promotes the development and progression of inflammatory breast cancer through activation of Axl kinase." Cancer Res **73**(21): 6516-6525.
263. Wang, X. and H. Weiner (1995). "Involvement of Glutamate 268 in the Active Site of Human Liver Mitochondrial (Class 2) Aldehyde Dehydrogenase As Probed by Site-Directed Mutagenesis." Biochemistry **34**(1): 237-243.
264. Wang, Y., W. Luo and G. Reiser (2008). "Trypsin and trypsin-like proteases in the brain: proteolysis and cellular functions." Cell Mol Life Sci **65**(2): 237-252.
265. Watanabe, K., Y. Tsuge, M. Shimoyamada, N. Ogama and T. Ebina (1998). "Antitumor Effects of Pronase-Treated Fragments, Glycopeptides, from Ovomucin in Hen Egg White in a Double Grafted Tumor System." Journal of Agricultural and Food Chemistry **46**(8): 3033-3038.
266. Waterhouse, A., M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. Heer, T. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede (2018). "SWISS-MODEL: homology modelling of protein structures and complexes." Nucleic Acids Research **46**(W1): W296-W303.
267. Wei, S., Y. Feng, E. Kalinina and L. Fricker (2003). "Neuropeptide-processing carboxypeptidases." Life Sci **73**(6): 655-662.
268. Wei, S., S. Segura, J. Vendrell, F. Aviles, E. Lanoue, R. Day, Y. Feng and L. Fricker (2002). "Identification and characterization of three members of the human metallocarboxypeptidase gene family." J Biol Chem **277**(17): 14954-14964.
269. Wellman-Labadie, O., R. Lakshminarayanan and M. Hincke (2008). "Antimicrobial properties of avian eggshell-specific C-type lectin-like proteins." FEBS Lett **582**(5): 699-704.

BIBLIOGRAPHY

270. Wellman-Labadie, O., J. Picman and M. Hincke (2008). "Antimicrobial activity of cuticle and outer eggshell protein extracts from three species of domestic birds." British Poultry Science **49**(2): 133-143.
271. Wernersson, S. and G. Pejler (2014). "Mast cell secretory granules: armed for battle." Nat Rev Immunol **14**(7): 478-494.
272. Wiechelman, K., R. Braun and J. Fitzpatrick (1988). "Investigation of the bicinchoninic acid protein assay: identification of the groups responsible for color formation." Anal Biochem **175**(1): 231-237.
273. Wierenga, R. and W. Hol (1983). "Predicted nucleotide-binding properties of p21 protein and its cancer-associated variant." Nature **302**(5911): 842-844.
274. Willemse, J. and D. Hendriks (2007). "A role for procarboxypeptidase U (TAFI) in thrombosis." Front Biosci **12**: 1973-1987.
275. Williams, C., J. Headd, N. Moriarty, M. Prisant, L. Videau, L. Deis, V. Verma, D. Keedy, B. Hintze, V. Chen, S. Jain, S. Lewis, W. Arendall, 3rd, J. Snoeyink, P. Adams, S. Lovell, J. Richardson and D. Richardson (2018). "MolProbity: More and better reference data for improved all-atom structure validation." Protein Sci **27**(1): 293-315.
276. Williams, J. (1962). "Serum proteins and the livetins of hen's-egg yolk." The Biochemical journal **83**(2): 346-355.
277. Wingfield, P. (2001). "Protein precipitation using ammonium sulfate." Curr Protoc Protein Sci **Appendix 3**: Appendix-3F.
278. Winn, M., C. Ballard, K. Cowtan, E. Dodson, P. Emsley, P. Evans, R. Keegan, E. Krissinel, A. Leslie, A. McCoy, S. McNicholas, G. Murshudov, N. Pannu, E. Potterton, H. Powell, R. Read, A. Vagin and K. Wilson (2011). "Overview of the CCP4 suite and current developments." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 235-242.
279. Xing, J., O. Wellman-Labadie, J. Gautron and M. Hincke (2007). "Recombinant eggshell ovocalyxin-32: expression, purification and biological activity of the glutathione S-transferase fusion protein." Comp Biochem Physiol B Biochem Mol Biol **147**(2): 172-177.
280. Yamashita, S., Y. Tsujino, K. Moriguchi, M. Tatematsu and T. Ushijima (2006). "Chemical genomic screening for methylation-silenced genes in gastric cancer cell lines using 5-aza-2'-deoxycytidine treatment and oligonucleotide microarray." Cancer Sci **97**(1): 64-71.
281. Yoshida, A., L. Hsu and Y. Yanagawa (1993). Biological Role of Human Cytosolic Aldehyde Dehydrogenase 1: Hormonal Response, Retinal Oxidation and Implication in Testicular Feminization. Enzymology and Molecular Biology of Carbonyl Metabolism 4. H. Weiner, D. W. Crabb and T. G. Flynn. Boston, MA, Springer US: 37-44.
282. Yoshida, A., A. Rzhetsky, L. Hsu and C. Chang (1998). "Human aldehyde dehydrogenase gene family." Eur J Biochem **251**(3): 549-557.

BIBLIOGRAPHY

283. Zhai, Y., Z. Sperkova and J. Napoli (2001). "Cellular expression of retinal dehydrogenase types 1 and 2: effects of vitamin A status on testis mRNA." J Cell Physiol **186**(2): 220-232.
284. Zhang, F., Y. Zhang, L. Sun, M. Chen, Y. Ran and L. Sun (2019). "Carboxypeptidase A4 promotes migration and invasion of lung cancer cells, and is closely associated with lymph node metastasis." **3**(2): 44-51.
285. Zhang, G., J. Liu, W. Fan, Q. Chen and B. Shi (2017). "An Efficient Transient Expression System for Enhancing the Generation of Monoclonal Antibodies in 293 Suspension Cells." Curr Pharm Biotechnol **18**(4): 351-357.
286. Zhang, X., Q. Zhang, D. Liu, T. Su, Y. Weng, G. Ling, Y. Chen, J. Gu, B. Schilling and X. Ding (2005). "Expression of cytochrome p450 and other biotransformation genes in fetal and adult human nasal mucosa." Drug Metab Dispos **33**(10): 1423-1428.
287. Zhao, D., P. McCaffery, K. Ivins, R. Neve, P. Hogan, W. Chin and U. Dräger (1996). "Molecular identification of a major retinoic-acid-synthesizing enzyme, a retinaldehyde-specific dehydrogenase." Eur J Biochem **240**(1): 15-22.
288. Zimpfer, A., F. Dammert, A. Glass, H. Zettl, E. Kilic, M. Maruschke, O. Hakenberg and A. Erbersdobler (2016). "Expression and clinicopathological correlations of retinoid acid receptor responder protein 1 in renal cell carcinomas." Biomark Med **10**(7): 721-732.