



Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular

***Discovery and characterization of  
small molecular weight  
metallocarboxypeptidase inhibitors***

**Daniel Fernández**

Doctoral Thesis

2009





Universitat Autònoma de Barcelona

Departament de Bioquímica i Biologia Molecular

***Discovery and characterization of  
small molecular weight  
metallocarboxypeptidase inhibitors***

Doctoral Thesis presented by Daniel Fernández  
to fulfill the Ph. D. degree in Biotechnology  
from the Universitat Autònoma de Barcelona

Daniel Fernández

Francesc X. Avilés Puigvert      Josep Vendrell Roca  
*Thesis co-directors*

Barcelona, 2009



*A mis padres*

*A Silvia*



# Index

Preface

List of published material

Abbreviations

Introduction

<b>1</b>	<b>The proteases</b>	<b>1</b>	
	<i>1.1</i>	<i>Types of proteases</i>	<i>1</i>
	<i>1.2</i>	<i>The metallocarboxypeptidases</i>	<i>2</i>
	<i>1.3</i>	<i>Inhibitors of CPs from natural sources</i>	<i>4</i>
	<i>1.4</i>	<i>CPA as a model metalloprotease</i>	<i>5</i>
<b>2</b>	<b>Biology - carboxypeptidases in health and disease</b>	<b>7</b>	

<i>2.1 CP activities in the gastrointestinal tract</i>	7
<i>2.2 CP roles in acute pancreatitis and cancer</i>	8
<i>2.3 Extra-pancreatic carboxypeptidases</i>	9
<i>2.4 A CPB-like antifibrinolytic</i>	10
<b>3 Biotechnology - from protein hydrolyzates to cancer therapy</b>	<b>13</b>
<b>4 Organic synthetic inhibitors of caboxypeptidases</b>	<b>15</b>
<i>4.1 Ground-state inhibitors</i>	16
<i>4.2 Transition-state inhibitors</i>	21
<b>Objectives</b>	<b>23</b>



## Results and Discussion

<b>5</b>	<b>Discovery of novel inhibitors</b>	<b>25</b>
<i>5.1</i>	<i>The arginine derivatives (family DF1)</i>	<i>26</i>
<i>5.2</i>	<i>Thioxophosphoranyl oxiranes (DF2) and cyclobutane <math>\beta</math>-peptides (DF3)</i>	<i>28</i>
<i>5.3</i>	<i>L-alanyl derivatives (DF4) and organosilicon compounds (DF5)</i>	<i>33</i>
<i>5.4</i>	<i>Peptoids (DF6)</i>	<i>37</i>
<i>5.5</i>	<i>Oxadiazoles (DF7)</i>	<i>40</i>
<b>6</b>	<b>Bioinformatics</b>	<b>43</b>
<i>6.1</i>	<i>The crystal structure pool</i>	<i>43</i>
<i>6.2</i>	<i>Interactions of the catalytic domain</i>	<i>44</i>
<i>6.3</i>	<i>Nature of the ODA patches</i>	<i>46</i>
<i>6.4</i>	<i>Predictive success of the method</i>	<i>47</i>
<i>6.5</i>	<i>Surface area of the ODA patches</i>	<i>49</i>
<i>6.6</i>	<i>Sequence-structure relationships</i>	<i>50</i>
<i>6.7</i>	<i>Protein-ligand relationships</i>	<i>54</i>
<i>6.8</i>	<i>Regionalization of the ODA patches</i>	<i>55</i>
<i>6.9</i>	<i>Accuracy of the predictions</i>	<i>57</i>
<i>6.10</i>	<i>Potential roles of the ODA patches</i>	<i>60</i>

<b>7 Crystallization and X-ray diffraction studies</b>	<b>65</b>
<i>7.1 Structure of a new CPB crystal form</i>	<i>66</i>
<i>7.2 Assignment of residues</i>	<i>68</i>
<i>7.3 Comparison to the known tetragonal crystal</i>	<i>70</i>
<i>7.4 The aromatic cluster</i>	<i>72</i>
<i>7.5 The role of Tyr248 and its flanking region</i>	<i>75</i>
<i>7.6 CPA crystallization</i>	<i>80</i>
<i>7.7 The complex of CPA and a mechanism-based inhibitor</i>	<i>81</i>
<i>7.8 Conformational changes</i>	<i>86</i>
<i>7.9 The zinc environment</i>	<i>87</i>
<i>7.10 The ligand structure</i>	<i>88</i>
<b>Overall discussion</b>	<b>91</b>
<b>Conclusions</b>	<b>97</b>

<b>8 Methods</b>	<b>99</b>
<i>8.1 General materials and media</i>	99
<i>8.2 Protein handling</i>	101
<i>8.3 Kinetic measurements and data treatment</i>	103
<i>8.4 X-ray crystallography - crystallogenesis</i>	105
<i>8.5 X-ray crystallography - data handling</i>	107
<i>8.6 X-ray crystallography - structure solution</i>	110
<i>8.7 Bioinformatics - the ODA method</i>	114
<i>8.8 Bioinformatics - virtual screening</i>	117
<i>8.9 Bioinformatics - resources</i>	120
<b>Bibliography</b>	<b>121</b>
<b>Acknowledgements</b>	<b>137</b>
<b>Publications</b>	<b>139</b>



# Preface

When I started working at our group at the UAB, I was not aware of the impact this endeavor would be operating upon me. Peptidases are proteins that have to be carefully faced up with because of their natural job of degrading other proteins. Therefore, finding a means to modulate their activity seemed both a bewildering and challenging task. And also a top-motivating one...

Five years later this feeling persists. Although some advances have been made, there seems to be no efficient means to modulate carboxypeptidase action by designed small molecular weight drugs, and therefore to shift the balance from disease to health. However, some things need time...

Peptidases let me introduce into molecular and structural biology, and bioinformatics. More importantly, they allowed me to meet a fascinating new world, paradoxically the "Old world", in a journey to a new Continent, country and nation. To a lovely city and a stimulating environment. As with peptidases, I've made countless people discoveries. The latter are of course not measurable by any scientific method, but constitute an important part of life.

However, as this document is not intended to be a biographical account, I'll focus on Science.

## *Notice about the format of this Thesis*

At the time of writing, parts of the work done during my Ph. D. have either been published, “lost” in the review process, or remain inedit. Thus, the experimental results are organized with a general sense in Chapters 5 to 7, to cover the different approaches employed. These Chapters are not intended as a compendium of publications. Published work is included in a separate section. The same was applied to Methods. Chapter 8 covers in a general way all aspects of the techniques used, while the peculiarities of each of them can be found in the attached papers.

I’ve tried to organize all parts in a well-balanced structure. Whenever possible, I’ve summarized all the information available. To begin with, Chapters 1 to 4 review background information. This is intended as a general Introduction. A brief introductory comment is included as a reference in each section of Chapters 5 to 7. Information on each specific topic can be found in the attached published material. The Discussion and Conclusions are also general.

# List of published material

## 1 Articles described in this Thesis

**Fernández D**, Testero S, Vendrell J, Avilés FX, Mobashery S. (2010). *Chem Biol Drug Des*, in press\*

*\*Our work was awarded the cover of volume 74 to appear in Feb/2010*

**Fernández D**, Boix E, Pallarès I, Avilés FX, Vendrell J. (2009). *Biopolymers*, in press

**Fernández D**, Avilés FX, Vendrell J. (2009). *Chem Biol Drug Des* 73:75-82

**Fernández D**, Torres E, Avilés FX, Ortuño RM, Vendrell J. (2009). *Bioorg Med Chem* 17:3824-3828

**Fernández D**, Avilés FX, Vendrell J. (2009). *Eur J Med Chem* 44:3266-3271

**Fernández D**, Illa O, Avilés FX, Branchadell V, Vendrell J, Ortuño RM. (2008). *Bioorg Med Chem* 14:4823-4828

**Fernández D**, Vendrell J, Avilés FX, Fernández-Recio J. (2007). *Proteins: Structure, Function, and Bioinformatics* 60:131-144

## 2 Inedit works

**Fernández D**, Montenegro M, Avilés FX, Vendrell J, García-Viloca M. (2009). *In preparation*

**Fernández D**, Avilés FX, Vendrell J, Pérez-Trujillo M, Parella T. (2009). *Submitted to J Inorg Biochem*

### 3 Crystal structures validated and deposited with major databases

Cambridge Crystallographic Data Centre (CCDC)

1. deposit number: 671796

Protein Data Bank (PDB)

1. ID: 2v77

2. ID: 3glj

3. ID: 3hlp

4. ID: 3huv

5. ID: 3i1u

### 4 Other articles

Pallarès I, **Fernández D**, Comellas-Bigler M, Fernández-Recio J, Ventura S, Avilés FX, Bode W, Vendrell J. (2008). *Acta Crystallogr D Biol Crystallogr* 64:784-791

Niemirowicz G, **Fernández D**, Solá M, Cazzulo JJ, Avilés FX, Gomis-Rüth FX. (2008). *Mol Microbiol* 70:853-866

Bayés A, **Fernández D**, Solá M, Marrero A, García-Piqué S, Avilés FX, Vendrell J, Gomis-Rüth FX. (2007). *Biochemistry* 46:6921-6930



# Abbreviations

ACE, angiotensin-converting enzyme  
ADEPT, antibody-directed enzyme prodrug therapy  
Ala, alanine  
Arg, arginine  
Asn, asparagine  
Asp, aspartic acid  
bCPA, bovine carboxypeptidase A (= CPA)  
bCPA1, bovine carboxypeptidase A1 (= bCPA)  
bPCPA1, bovine procarboxypeptidase A1  
BzISA, 2-benzylsuccinic acid  
CBP, cyclobutane  $\beta$ -peptide  
CCP, cytosolic carboxypeptidase  
CP, carboxypeptidase  
CPA, carboxypeptidase A  
CPB, pancreas carboxypeptidase B (= CPB1)  
CPD, carboxypeptidase D  
CPE, carboxypeptidase E  
CPM, carboxypeptidase M  
CPN, carboxypeptidase N  
CPU, human plasma CPB (= TAFI/CPR/CPB2)  
CPZ, carboxypeptidase Z  
CTGC, chymotrypsinogen C  
Cys, cysteine  
Da, dalton  
dCPD, domain II of duck carboxypeptidase D (= CPD)  
DMSO, dimethylsulfoxide  
ECI, endogenous carboxypeptidase inhibitor (= latexin)  
GEMSA, 2-(2-guanidinoethylthio)succinic acid  
Gln, glutamine  
Glu, glutamic acid

Gly, glycine  
HaCPA, CPA from cotton bollworm (*Helicoverpa armigera*)  
hCPA1, human carboxypeptidase A1  
hCPA2, human carboxypeptidase A2  
hCPA4, human carboxypeptidase A4  
hCPB, human pancreas CPB (= CPB/CPB1)  
hCPU, human plasma CPB (= TAFI/CPR/CPB2)  
hCPM, human carboxypeptidase M (= CPM)  
His, histidine  
hPCPA1, human procarboxypeptidase A1  
HzCPB, CPB from corn earworm (*Helicoverpa zea*)  
Ile, isoleucine  
iPr, *iso*-propyl  
LCI, leech carboxypeptidase inhibitor  
Leu, leucine  
Lys, lysine  
MD, molecular dynamics  
MES, 2-(*N*-morpholino)ethanesulfonic acid  
Met, methionine  
MMP, matrix metalloproteinase  
MR, molecular replacement  
MW, molecular weight  
ODA, Optimal Docking Area  
PCI, potato carboxypeptidase inhibitor  
PCP, procarboxypeptidase  
pCPB, porcine pancreas carboxypeptidase B  
PDB, Protein Data Bank  
Phe, phenylalanine  
PPE, proproteinase E  
Pro, proline  
SDS, sodium dodecyl sulfate  
Ser, serine  
Tris, 2-amino-2-(hydroxymethyl)propane-1,3-diol  
TAFI, thrombin-activable fibrinolysis inhibitor (=CPU/CPR/plasma CPB/CPB2)

TCI, tick carboxypeptidase inhibitor

TcMCP-1, metallocarboxypeptidase 1 from *Trypanosoma cruzi*

TcMCP-2, metallocarboxypeptidase 2 from *Trypanosoma cruzi*

Thr, threonine

TvCPT, carboxypeptidase T from *Thermoactinomyces vulgaris*

Trp, tryptophan

Tyr, tyrosine

Val, valine

Z, carbobenzoxy



# Introduction



# 1 The proteases

Proteases are proteins that catalyze the hydrolysis of the C-N bond, or peptide bond, of protein and peptide substrates and are therefore called peptidases as well. Proteases were recognized early in the history of biochemistry. In the nineteenth century, one primary focus of research was on digestive proteases like pepsin and trypsin. Another important digestive protease named carboxypeptidase was isolated and characterized in the last years of the second decade of the twentieth century. Digestive proteases are located in the gastro-intestinal tract and are mainly responsible for the digestion of food proteins (Erickson & Kim, 1990).

Peptidases located extracellularly in the blood or other extracellular compartments of the body often play regulatory roles in carefully controlled processes, such as blood clotting, fibrinolysis, complement activation, fertilization, and hormone production. Intracellularly located proteases exhibit a wide range of roles. They are found in the Endoplasmic Reticulum, the Golgi apparatus, or the lysosomes, among others. Their functions include the processing of peptide hormones, cytokines, and ubiquitin mediated proteolysis, among others (Fricker, 1988; Skidgel & Erdös, 1998; Fujiwara et al., 1999).

## *1.1 Types of proteases*

Several protease classification schemes have emerged. Proteases can be classified according to their mechanism of action, or to specific groups that are present in the active center. The following groups can be distinguished:

- i) serine peptidases (typified by trypsin and chymotrypsin)
- ii) cysteine peptidases (i.e., cathepsins, calpains and caspases)
- iii) aspartyl peptidases (i.e., pepsin and rennin)
- iv) metallo peptidases (i.e., angiotensin-converting enzyme, the extracellular matrix metalloproteinases and metallocarboxypeptidases)
- (v) peptidases with an as yet unclear reaction mechanism

The Enzyme Commission (EC), established after the 1955 international meeting, defined the EC classification scheme that codifies enzyme activities into six main classes, according to the global reaction catalyzed. Each enzyme activity was assigned a systematic name and a code number in such a way that, for instance, the peptidase with the trivial name chymotrypsin is coded EC 3.4.21.1, while carboxypeptidase A1 (CPA1) is systematized as EC 3.4.17.1. Evolutionary relationships are not fully considered by the current EC classification (García-Carreño & Navarrete del Toro, 1997).

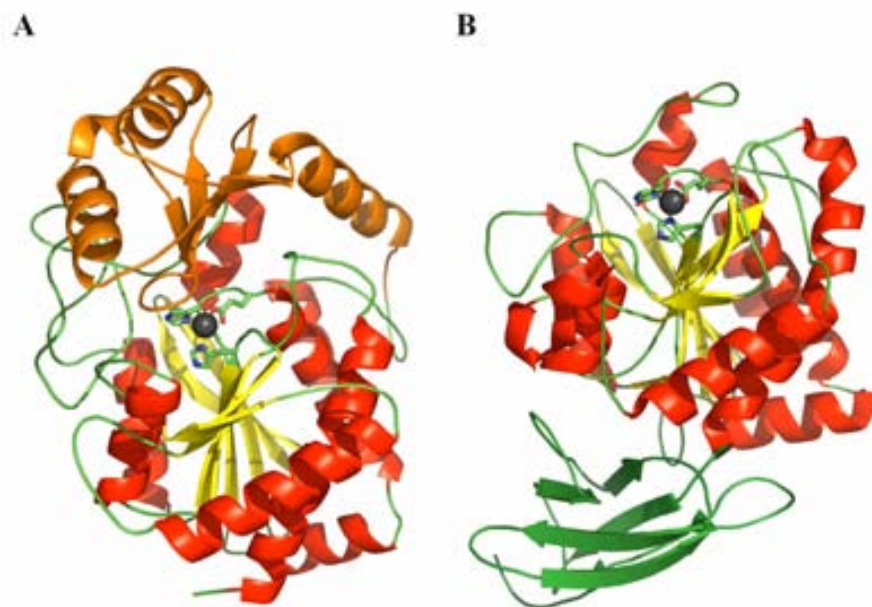
A further organization used aminoacid sequence and three-dimensional (3D) structural information to group proteases into clans and families (Rawlings et al., 2006). This scheme has been developed into the searchable online database MEROPS (<http://merops.sanger.ac.uk/>). The MEROPS accession code is formed by a three-character identifier of the family to which the peptidase belongs, and a three-figure number. For example, the identifier of chymotrypsin, the first peptidase in family S1, is S01.001. The letter “S” designates chymotrypsin to be a serine peptidase. CPA1, M14.001, belongs to clan MC, family M14, subfamily M14A. Within this classification scheme an additional subfamily is termed M14B and includes some CPB-like enzymes.

## *1.2 The metallo-carboxypeptidases*

Metallo-carboxypeptidases (CPs) are zinc-dependent exopeptidases, releasing the C-terminal residue from peptide and protein substrates. The CPs are grouped by sequence homology and overall structure in clan MC, family M14 of peptidases. At least, there exist four M14 subfamilies. The best characterized CPs belong to the M14A (i.e. CPA, CPB, and CPU) and M14B subfamilies (CPD, CPE, CPM, and CPN, among others). Many of these CPs have their 3D structure determined by X-ray diffraction. The overall structural features of CPs are as follows. The catalytic domain common to all CPs displays the  $\alpha/\beta$  hydrolase fold shown in Fig. 1, where the active center is located in a cavity that harbors the catalytic  $Zn^{2+}$  ion. M14A proteins are synthesized as zymogens with an N-terminal non-catalytic, 90-95 residues long subdomain termed pro domain (Fig. 1, panel A), which is not found in the M14B



proteins. The latter possess instead a 80-residues long C-terminal subdomain (Fig. 1, panel B) which might facilitate the anchoring onto the cellular membrane (Vendrell et al., 2004). The M14A peptidases acquire the full active state upon loss of the pro domain by proteolytic cleavage (Burgos et al., 1991; Avilés et al., 1993; Opezzo et al., 1994), although for CPU this is still a matter of debate (Valnickova et al., 2007; Anand et al., 2008).



**Figure 1.** View of the 3D structures of representative CPs. *Panel A)* An M14A subfamily member, bovine CPA1 zymogen (PDB code 1pyt) with the N-terminal pro domain shown in orange. *Panel B)* An M14B subfamily member, human CPM (1uwy) with the C-terminal domain depicted in green. In both structures, the catalytic domain has the  $\alpha$ -helices and the  $\beta$ -sheets in red and yellow, respectively. The HEH triad of residues coordinating the catalytic zinc ion (gray sphere) is shown as a stick model.

Following the bovine CPA1 numbering, key amino acid residues involved in the catalytic machinery and substrate binding include His69, Glu72 and His196 (for coordination of the zinc ion), Arg127 and Glu270 (directly involved in the catalysis; both residues define subsite S1), as well as Asn144, Arg145, and Tyr248 (subsites S1') which anchor and neutralize the C-terminal residue of the

substrate. Here, the carboxylate from the peptide substrate is bound by an ion pair with the guanidinium group of Arg145 and two additional hydrogen bonds from Asn144 and Tyr248. The enzyme-substrate interaction is even more facilitated by the overall hydrophobic character of the active site pocket which reduces competitive solvation of the binding sites by water molecules (Schmuck, 2006). One further feature is that the key determinant of substrate specificity among CPs hinges on the residue at position 255.

The catalytic mechanism of CPA has been one of the most investigated (Christianson & Lipscomb, 1989; Lipscomb & Sträter, 1996; Krämer, 1999). Two pathways, in which a tetrahedral transition state is formed, have been proposed to model it. In the anhydride mechanism, a carboxylate nucleophile (presumably Glu270) would attack the scissile amide bond, previously activated by zinc coordination of the carbonyl oxygen. The anhydride intermediate would then be hydrolyzed by a hydroxide ion. In the second mechanism, the water-promoted pathway, a zinc-hydroxyl nucleophile would attack directly the amide bond. The role of the carboxylate would be the deprotonation of the zinc-bound water.

### *1.3 Inhibitors of CPs from natural sources*

Protein inhibitors of the CPs are found in a number of tissues and organisms. They are small to medium range polypeptides, from 39 to 222 residues, that bind to the target macromolecules with nano- and subnanomolar affinities (Vendrell et al., 2000). The protein inhibitors of the CPs have a compact globular fold, generally sustained by disulphide bridges. The carboxypeptidase inhibitor from potato (PCI) is a small polypeptide (39 residues) with a 5-residue long C-terminal tail that makes interactions with the primary binding site of the catalytic domain and also coordinates to the zinc ion in a substrate-like manner (Rees & Lipscomb, 1982). Similar characteristics are found in other small polypeptides such as the leech carboxypeptidase inhibitor (LCI, 66 residues; Reverter et al., 2000), the tick carboxypeptidase inhibitor (TCI, 76 residues; Arolas et al., 2005) and the inhibitor from *Ascaris* (ACI, 67 residues; Sanglas et al., 2009). Human latexin, or endogenous carboxypeptidase inhibitor (ECI), is a 222-residue protein sequentially and

structurally unrelated to the latter forms and is devoid of disulphide bridges (Pallarès et al., 2005). The protein inhibitors identified until now affect the CPs of the M14A subfamily but are ineffective against the M14B carboxypeptidases. On the contrary, several different kinds of small molecular weight organic compounds are known as inhibitors of all CPs.

## *1.4 CPA as a model metalloprotease*

As a much studied zinc-dependent protease, CPA has served as a prototype for the development of therapeutic protocols that may be applied to other metalloproteases as well (Kim, 2001). The early determination of the crystal structure of bovine CPA allowed the creation of what was considered later as one of the first hits coming from a structure-based approach to drug discovery, although the true biological target for this drug was a CPA-related metallopeptidase, angiotensin-converting enzyme (ACE; family M2). ACE has been for long considered as a drug target for the control of hypertension, and therefore prompted the search by various pharmaceutical companies of small molecular weight synthetic compounds able to interfere with its activity. Researchers at Bristol-Myers Squibb (BMS) came up with the idea that CPA and ACE share a common mode of action and thereby would possess a similar active site architecture. Captopril (1-[(2-(S)-3-mercapto-2-methyl-1-oxopropyl]-L-proline), originated from a CPA-based crude model of ACE, was the first orally active ACE inhibitor to be launched into the market (Cushman et al., 1977; Ondetti, 1994; Cushman & Ondetti, 1999). Today, other groups of metallopeptidases such as the matrix metalloproteinases (family M10) and the metallocarboxypeptidases (family M14) are as ACE druggable targets and likewise exploited in the drug discovery process (Mittl & Grütter, 2006; Arolas et al., 2007).

The M14 family of peptidases, or CPs, will be the focus of our Thesis work. As will be reviewed in the following three Chapters, CPs have been a matter of analysis in diverse research areas.



## 2 Biology - carboxypeptidases in health and disease

### *2.1 CP activities in the gastrointestinal tract*

A normal healthy adult requires approximately 0.75 g/kg/day of good quality, highly digestible protein to maintain a positive nitrogen balance and to provide essential amino acids. The digestion of dietary protein begins in the stomach, where proteolysis is initiated by two types of pepsins, and continues through the action of several proteolytic enzymes released by the pancreatic acinar cells: trypsin, chymotrypsin, elastase, and carboxypeptidases A and B (Erickson & Kim, 1990). CPA was the first enzyme for which zinc has been shown to have a biological role. Maternal dietary deficiency of zinc during pregnancy may lead to serious embryonic and fetal development problems. Lower than normal levels of major proenzymes, like proCPA and chymotrypsinogen, have been found in pancreata of zinc-deficient fetuses (Robinson & Hurley, 1981). Some investigations linked the availability of zinc and the protein content from different meals to the modulation of pancreatic carboxypeptidase activity in rats (Berger and Schneeman, 1986; 1988), birds (Levey et al., 1999) and pigs (Martínez-Montemayor et al., 2008). The effect over digestion of food components such as melanoidins and bile acid conjugates, has been related to inhibition of CPA and CPB enzyme activity (Huijghebaert & Hofmann, 1986; Maeda & Takahashi, 1989; Ibarz et al., 2008). Besides, various different flavonoids, like the benzo- $\gamma$ -pyrone compounds quercetin and luteolin found in higher plants, and thus components of the human diet (Parellada et al., 1998), and the foodstuffs phenylethylaminoalanine (Friedman et al., 1986) and phytic acid (Martin & Evans, 1989), were shown to be CPA inhibitors. Further studies suggested a role of CPs in alleviating heavy metal poisoning (Kojima et al., 1994) as well as in detoxification of ochratoxin A, a carcinogenic food contaminant (Pohland et al., 1992; Stander et al., 2001).

On the other hand, the same basis of the beneficial effects of proteases, i.e., their hydrolytic action, may neutralize therapeutic success. A wide range of proteases in the gastrointestinal tract act together as a barrier for orally administered vaccines and drugs, limiting the usefulness of peptide and protein drugs. The quest for effective measures to counterbalance digestive protease action is an active research field (Friis et al., 1996; Langguth et al., 1997; Fredholt et al., 2000; Bernkop-Schnürch & Schmitz, 2007). Recent investigations stressed that the enzymatic activity of a digestive protease like CPA may be affected by using the cyclic oligomer cyclodextrin (Easton et al., 2001; Villalonga et al., 2007) or by direct binding of a drug-protecting mucoadhesive (Pallarès et al., 2008). Along with various extra-pancreatic family members (see paragraph 2.3, below), the pancreatic CPs have been implicated in tissue organogenesis and developmental processes (Nishioka et al., 2003; Guimarães et al., 2007; Zhou et al., 2007; Wang et al., 2007a; Marquez-Curtis et al., 2008).

## *2.2 CP roles in acute pancreatitis and cancer*

The activity of proteolytic enzymes is tightly controlled by various means, but they behave abnormally under certain circumstances. Significant levels of active CPA, CPA zymogen, and its pro domain, have been detected in serum of patients suffering from acute pancreatitis and pancreas cancer. A low acute pancreatitis incidence has been reported in the United Kingdom (10 cases / over 100,000) and in Germany (15/100,000) but in contrast the incidence is higher in the USA (40-80/100,000) (Kylänpää-Bäck et al., 2002). Pancreatic cancer is the fourth leading cause of death among men and fifth among women in the USA, and fifth in both men and women in Japan. The overall 5-year survival after a diagnosis of pancreatic carcinoma remains very low in contrast to most other gastrointestinal malignancies (Matsugi et al., 2007).

Therefore, measuring of CPA (or trypsin-activated CPA zymogen) activity in clinical assays would be a means to diagnose these diseases (Saruta et al., 1986; Brown et al., 1987; Kazmierczak & van Lente, 1989; Gilvarg, 1998; Stewart & Gilvarg, 1999). A similar use of CPA for the detection of pancreatic disorders in diabetes mellitus patients has been proposed (Fujii et al., 1977). Likewise, the early detection of acute pancreatitis and pancreas cancer have been linked to CPB and its pro domain (Geokas et al., 1974; Delk et al., 1985; Appelros et al., 1998; Müller et al., 2002; Borgström and Regnér, 2005; Sáez et al., 2005; Pearce et al., 2006; Regnér, et al., 2008). A variation of this procedure would be an immunological assay that includes the use of monoclonal antibodies raised against CPA1 and CPA2 (Hamada et al., 2006).

## *2.3 Extra-pancreatic carboxypeptidases*

CPs also perform a variety of tasks in other compartments in the body. These CPs resemble functionally CPB in that they process peptide and protein substrates with a basic C-terminal residue. CPD, a reference enzyme of this family, is primarily a membrane-bound enzyme in the *trans*-Golgi network where it processes polypeptides or prohormones that transit the constitutive secretory pathway and has also been shown to cleave extracellular substrates and generate arginine, resulting in enhanced nitric oxide production in macrophages (Hadkar and Skidgel, 2001). A novel CPD variant has been recently characterized in haematopoietic tumour cells (O'Malley et al., 2005). Mice lacking CPE develop multiple endocrine disorders, including obesity, diabetes, and neurological deficits (Wang et al., 1998; Zhu et al., 2005; Jeffrey et al., 2008). CPE may be used as a predictor of type II diabetes (Chen & Meyer, 2001). CPM binds to the cell surface through a glycosylphosphatidylinositol anchor (Reverter et al., 2004) and is involved in the process of inflammation, based on its potential to modulate the activity of the anaphylatoxins and kinins (Deiteren et al., 2009). CPN, a tetrameric protein synthesized in the liver and secreted in blood, inactivates complement anaphylatoxins (C3a, C4a, and C5a) by cleaving their carboxy-terminal Arg residue, is a regulator of inflammation

(Mueller-Ortiz et al., 2009). CPN has recently been linked to clinical cases of angio-oedema (Willemse et al., 2008). CPA3 (mast cell carboxypeptidase) resembling CPA1 functionally but CPB structurally, is released from mast cell secretory granules by immunologic response and plays a role in both host defense and in allergic reactions (Goldstein et al., 1989; Lützelshwab et al., 1997). CPA3 seems to be converted proteolytically from the zymogen in the secretory granules (Henningsson et al., 2005). Another CPB-like enzyme is CPZ which, unlike most non-pancreatic CPs, harbors a cysteine-rich-domain N-terminal to the catalytic domain, and is believed to be involved in developmental processes (Moeller et al., 2003). As recently reviewed by Vovchuk and Petrov (2008), many of the above mentioned enzymes play a role in carcinogenesis and cancer malignization processes.

A further, completely new, subfamily of M14 peptidases should also be mentioned. Nna1-like proteins, which preferentially (but not exclusively) localize in the cytoplasm of several different cell types, were assimilated to the classical metallo-carboxypeptidases, although they are ADP/ATP-dependent and have the potential to act on protein substrates, particularly in the processing of tubulin. These distinct cytosolic carboxypeptidases (CCPs) were grouped into the proposed M14D subfamily (de la Vega et al., 2007). A member of this subfamily, a CCP from *Plasmodium falciparum*, would be a potential target for the development of antimalarial agents (Avilés et al., 2008).

## ***2.4 A CPB-like antifibrinolytic***

Almost coincidentally four research groups reported on the discovery of a new CPB-like activity in plasma, distinct from pancreatic CPB and plasma circulating CPN. Not surprisingly, the newly discovered enzyme was given different names: TAFI (“thrombin-activable fibrinolysis inhibitor”; Bajzar et al., 1995), human plasma CPB (Eaton et al., 1991), CPU (“U” stands for unstable; Hendriks et al., 1990), and CPR (“R” denotes the preference to cleave substrates with an arginine at the C-terminus; Campbell & Okada, 1989). CPU/TAFI, which is synthesized in the liver, acts during coagulation or in



response to inflammation (Asai et al., 2004; Chen et al., 2005; Rooth et al., 2007; Leung et al., 2008). CPU is formed during coagulation and fibrinolysis from its zymogen precursor by the action of proteolytic enzymes, such as plasmin, thrombin, or the thrombin-thrombomodulin complex. Once activated, CPU cleaves basic amino acids at the C-terminus of fibrin fragments. The loss of C-terminal lysines and thereby of lysine binding sites for plasminogen then serves to inhibit fibrinolysis. By inhibition of the loss of lysine binding sites for plasminogen and thus increase the rate of plasmin formation, effective inhibitors of CPU are expected to facilitate fibrinolysis (Mosnier & Bouma, 2006). In addition, due to its ability to inhibit fibrinolysis, CPU itself may be useful as a therapeutic protein capable of acting as a pro-coagulant.

CPU is believed to play a central regulatory role in the fibrinolytic cascade, and the manipulation of CPU levels or activity in biological fluids may have important therapeutic applications with respect to hemorrhagic disorders including vascular and heart pathologies and stroke (Montaner et al., 2003; Leebeek et al., 2005; Tregouet et al., 2009). Thus, human CPU is a recognized valuable biomedical target for chemotherapy. This captured the interest of academic and industry research groups as well, which developed different means to quantitate CPU levels in biological fluids (Watanabe et al., 1998; Ferrer et al., 2003; Quentin, 2003; Tani et al., 2003; Greenfield & An, 2004; Cronet et al., 2005; Ziegler et al., 2006). Inhibitors of CPU would enhance fibrinolysis and have an anti-coagulant effect. Small molecular weight organic inhibitors developed to interfere with the enzyme's activity were since then increasingly found in the literature (as will be reviewed in Chapter 4).



### 3 Biotechnology - from protein hydrolyzates to cancer therapy

Probably the major applications entailed the use of CPs for the analysis of the primary structure of peptides and proteins (Gladner & Folk, 1958; Slobin & Carpenter, 1966; Oppenheim et al., 1986) and as biocatalysts in organic chemistry. Along with the serine protease CPY (from yeast), CPA is used for trimming the C-terminal polyhistidine tag of recombinantly expressed proteins. CPB has also been tested for the quality control (Yang et al., 1999) and the final maturation (Norgaard & Andersen, 2008) of the industrially produced recombinant insulin polypeptide. CPA was used to resolve isomers in the chemoenzymatic synthesis of the antitumor antibiotic deoxyspergualin (Umeda et al., 1987) and for the enantioselective synthesis of a tyrosine precursor (Hasegawa & Shinohara, 1998). The finding that a number of proteases, including CPA and CPB, are stable and operative in various organic media (Kamo & Tsugita, 1987; Simon et al., 1998) inspired further developments. Thus CPB, integrated in a microbiosensor in which the protease is immobilized on a solid support, was shown to have potential to be used as a transducer for the detection of peptide and protein content in liquid matrices (Biloivan et al., 2006).

Industrial peptides may be produced by enzyme-catalyzed reactions with some advantages over solid-phase synthesis such as stereoselectivity and the by-pass of unwanted protecting and deprotecting steps. The synthesis of Phe-Phe, among other dipeptides, was effected by tailoring the organic media such as to induce the reversal of the CPA-catalyzed reaction (Vertesi & Simon, 1998). Food products and pharmaceutical formulations make use of enzyme-catalyzed protein hydrolyzates. These protein hydrolyzates may frequently lead to an intense bitter taste which reduces the commercial value of the product. Moreover, a diet free of aromatic aminoacids like phenylalanine might be ingested by patients with phenylketonuria, preventing the outcome of brain injuries. Protein hydrolyzates may be prepared in a controlled protein hydrolysis condition using an immobilized CPA-agarose column, with enhanced release of

aromatic residues than the soluble protein, and under various operation conditions without affecting the CPA performance (Tardioli et al., 2003; Pinto et al., 2008; Galvão et al., 2009).

A number of CPs have found application as biotherapeutic drugs. So, CPA may be used to treat disorders like asthma and cancer (Zhu, 2004) while various proteases from mast cells, including CPA3, may be of utility in cases of envenomation from snake bites (Metz & Galli, 2007). An engineered more stable form of CPU, a CPB-related enzyme, has been proposed to be employed as a pro-coagulant or as an antifibrinolytic agent (Greenfield & An, 2004; Cronet et al., 2005; 2008). Probably, one of the most attractive potential biomedical uses of any CP is linked to cancer therapy. Carboxypeptidase G2 or glutamate carboxypeptidase (family M20) and CPA1 (family M14) have been the subject of considerable interest in experimental prodrug strategies for cancer therapy in the approach known as “antibody-directed enzyme prodrug therapy” or ADEPT (Senter et al., 1993; Huennekens, 1997; Cheng & Roffler, 2008). Engineered forms of CPA1 and CPB, fused or conjugated to a human or humanized tumor-selective antibody, and with altered substrate preference, have been widely investigated for ADEPT (Smith et al., 1997; Edge et al., 1998; Wolfe et al., 1999; Slater et al., 2002). In spite of dramatic advances in this therapeutic approach, some problems like the reactivity of the immunocomplexes, poor biodistribution and pharmacokinetics, and conjugate localization, remain a challenge. Recent studies on CPA1 linked to methoxypoly(ethylene glycol) chains demonstrated promising properties both *in vitro* and *in vivo* (Ton et al., 2005).

## 4 Organic synthetic inhibitors of carboxypeptidases

This last part of the introduction is devoted to a description of known small molecular weight organic CP inhibitors. In the light of the amount of information available on CP inhibitors it is wise to review former medicinal chemistry efforts to understand current developments in this field. However, this is not intended as an exhaustive account. The number and complexity of organic synthetic inhibitors directed towards CPs is growing with each passing day. Thus, we decided to include selected reference compounds rather than a comprehensive register of CP inhibitors.

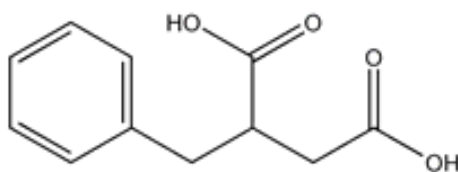
Key to the inhibitor's function are:

- i) a fragment able to dock into the specificity pocket,
- ii) a hydrolysis-resistant chemical bonding and especially,
- iii) a metal-coordinating group.

Different kinds of metal-coordinating groups have been widely investigated through time. The most popular ones are surveyed in more detail below. Compounds may be further classified as ground- or transition-state analogues. This means that some of them mimic the structure of the substrates and/or products (the ground-state inhibitors) while others bear analogy to the supposed transient entity that originates during the transition state. One further category includes the mechanism-based inactivators, some of which will be encountered later in Chapter 7.

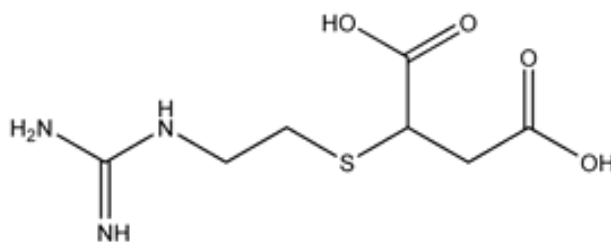
## 4.1 Ground-state inhibitors

Ground-state inhibitors may often take the form of the combination of the products of the enzymic reaction. Such an idea inspired Byers & Wolfenden the design of *DL*-BzISA (**1**), a bi-product analog of the CPA catalyzed reaction (Byers & Wolfenden, 1972; 1973). This concept was translated later to the design of *DL*-GEMSA (McKay & Plummer, 1978; McKay et al., 1979), an inhibitor to CPB and many CPB-like enzymes (**2**). Commercially available **1** and **2** serve today as standard synthetic inhibitors for the characterization of CP activity. The inspection of both chemical structures easily reveals the beauty of their design: the carboxylate functional groups serve as anchor points of the molecule while the aromatic ring (in **1**) and the guanidinium side chain (in **2**) provide a basis for docking into the specificity pocket of each enzyme.



**1**

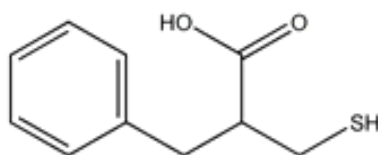
*DL*-BzISA, 2-benzylsuccinic acid  
 $K_i$  bCPA = 1.1  $\mu$ M



**2**

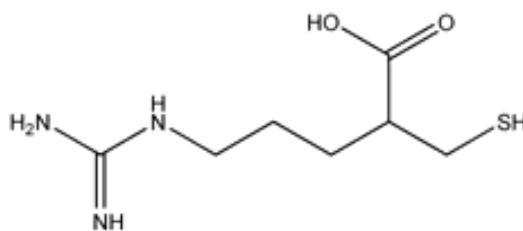
*DL*-GEMSA, 2-(2-guanidinoethylthio)succinic acid  
 $K_i$  pCPB = 1.0  $\mu$ M, hCPN = 1.0  $\mu$ M, hCPE = 0.9  $\mu$ M, hCPU = 4.4  $\mu$ M

Modification of the original molecules by researchers at Bristol-Myers Squibb gave compounds **3**, SQ14,603, and **4**, SQ24,798 (Ondetti et al., 1979). The striking difference relies in that these compounds have one carboxylate moiety replaced by a thiol group. The tremendous increase in inhibitory potency was evidently due to the powerful coordination of the catalytic zinc ion by the sulfur atom.



**3**

SQ14,603; 2-benzyl-3-mercaptopropanoic acid  
 $K_i$  bCPA = 11.0 nM

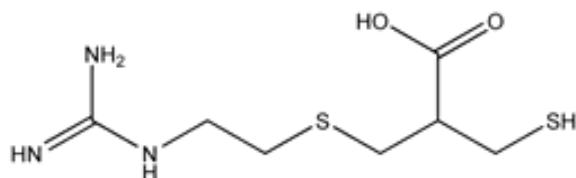


**4**

SQ24,798; 5-guanidino-2-(mercaptomethyl)pentanoic acid  
 $K_i$  pCPB = 0.42 nM

Holmquist & Vallee (1979) extended the analysis of mercaptans, i.e. thiol-based compounds, to various different zinc metallopeptidases including CPA and ACE. Later, Kim and associates prepared different aryl analogues of **3** to study the structural requirements of binding to the principal recognition subsite, S1', of CPA (Kim et al., 1991; Kim & Kim, 1991). These researchers also synthesized optically pure isomers as well as  $\alpha,\alpha$ -disubstituted variants of **3** to assess their inhibitory abilities against CPA (Kim & Kim, 1993; Lee & Kim, 2003). Further investigations on thiol-based compounds as CPA inhibitors were performed by Suh and coworkers (Suh et al., 1995) and Park & Kim (2002).

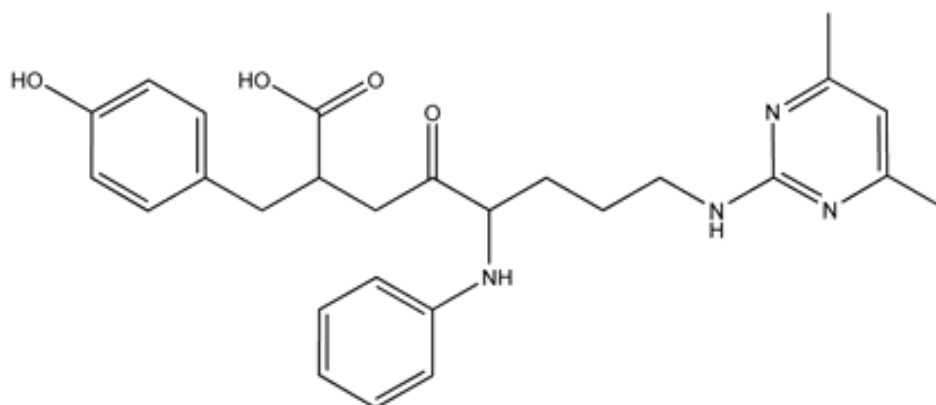
A closely related compound (**5**) demonstrated potent inhibition towards several human CPB-like enzymes (Plummer & Ryan, 1981; Fricker et al., 1983; Mao et al., 2003; Adler et al., 2005; Do et al., 2005).



**5**

Plummer's inhibitor, 3-(2-guanidinoethylthio)-2-(mercaptomethyl)propanoic acid  
 $K_i$  hCPN = 2 nM, hCPM 3 nM, hCPB 8.2 nM, hCPE 44 nM, hCPZ 50 nM, hCPU 200 nM

Chemical compounds isolated from natural sources had not escaped scrutiny as CP inhibitors (Pitout & Nel, 1969; Umezawa et al., 1984; Cannell et al., 1988; Moriguchi et al., 1988; Inamori et al., 1998; Kusano et al., 1998; Sakagami et al., 1999; Morita et al., 2001; 2002; 2003; Buchanan et al., 2008). Compound **6**, a derivative of arphamenine B, a product isolated from culture filtrates of *Chromobacterium violaceum*, was identified as a nanomolar inhibitor of CPA (Ohuchi et al., 1984). This compound is notable because of the presence of a tyrosyl side chain, which may confer selectivity toward A-type carboxypeptidases.

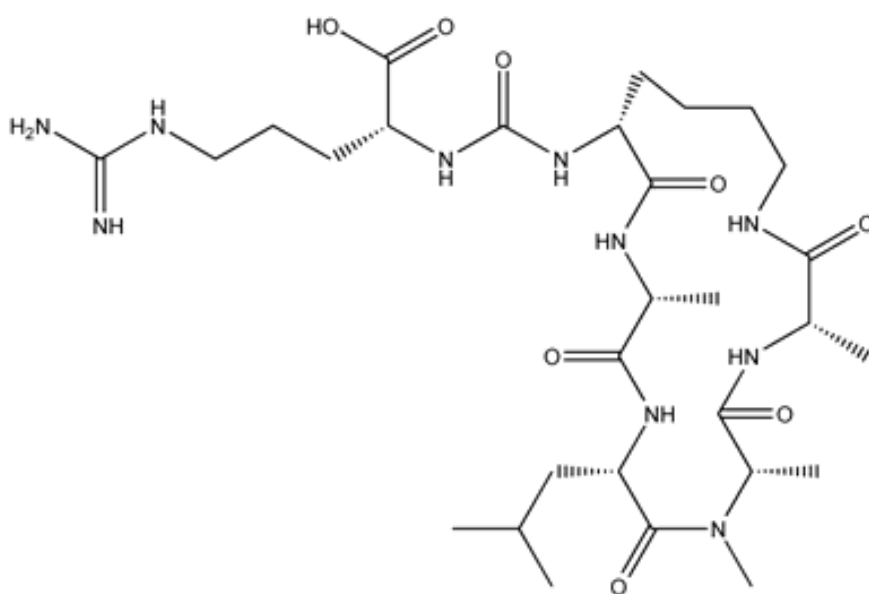


**6**

8-(4,6-dimethylpyrimidin-2-ylamino)-2-(4-hydroxybenzyl)-4-oxo-5-(phenylamino)octanoic acid  
 $K_i$  CPA = 19 nM, CPB 1.1  $\mu$ M



Anabaenopeptins G, I, J and T, isolated from cultured cyanobacteria *Oscillatoria agardhii* and *Aphanizomenon flos-aquae*, are cyclic peptides containing non-natural aminoacids and ureido-type bonds which exerted strong CPA inhibition (Itou et al., 1999; Kodani et al., 1999; Murakami et al., 2000). Cyclic anabaenopeptin-type compounds, like **7**, have been disclosed as CPU inhibitors and thus hold potential application for the treatment of blood-related diseases (Björquist et al., 2005; Astra Zeneca AB).



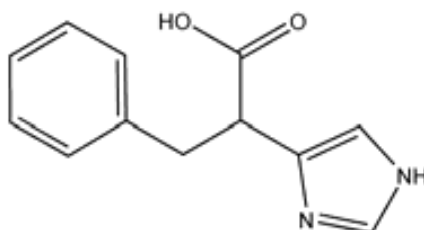
**7**

Anabaenopeptin-type cyclic peptide  
IC<sub>50</sub> CPU = 0.2 μM

Products from natural sources have also inspired the development of new CP inhibitors. In one example of design assisted by structural knowledge, a linear nonapeptide synthesized from the proteinaceous carboxypeptidase inhibitor from potato (PCI) was found to be a potent *in vitro* CPU inhibitor ( $K_i = 8.8$  nM) without affecting CPN (Lazoura et al., 2002).

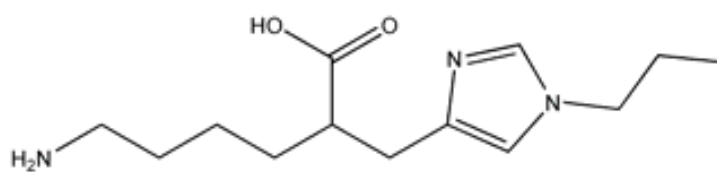
One further innovation that can be mentioned is the placement in the structure of a ring system as a zinc coordinating group. This led firstly to a series of CPA inhibitors developed by Dong H. Kim and coworkers (University of

Pohang, Seoul, Korea), that incorporate the imidazole ring as exemplified by **8** (Lee et al., 1997; Han & Kim, 2001).

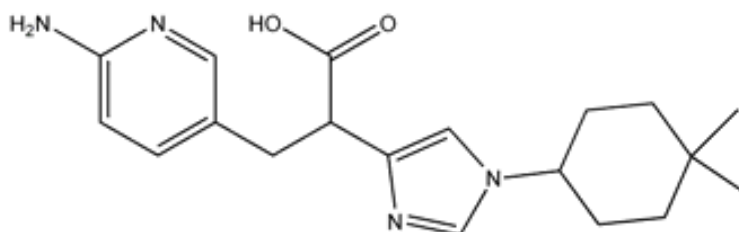


**8**  
2-(1*H*-imidazol-4-yl)-3-phenylpropanoic acid  
 $K_i$  bCPA = 0.8  $\mu$ M

This concept was exploited later for the development of human plasma CPB or CPU/TAFI inhibitors like **9** (Allerton et al., 2002; 2003a; 2003b [Pfizer Inc.]; Barrow et al., 2003; Nantermet et al., 2004 [Merck & Co.]), and the more advanced **10** (Kallus et al., 2005; 2007 [Sanofi-Aventis Deutschland GmbH]), which displays an arginine mimetic moiety.



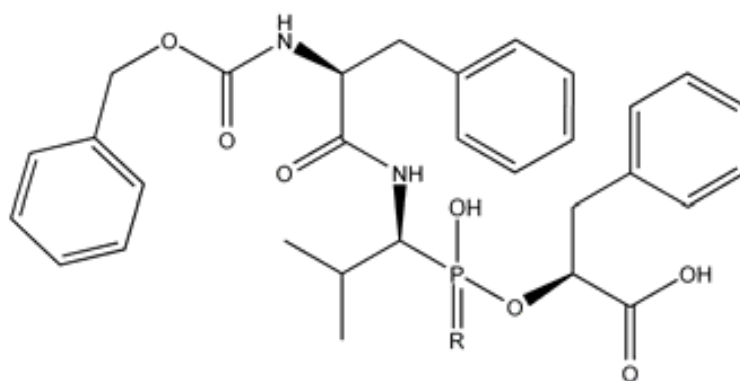
**9**  
6-amino-2-((1-propyl-1*H*-imidazol-4-yl)methyl)hexanoic acid  
 $K_i$  CPU = 0.31  $\mu$ M



**10**  
3-(6-aminopyridin-3-yl)-2-(1-(4,4-dimethylcyclohexyl)-1*H*-imidazol-4-yl)propanoic acid  
 $IC_{50}$  CPU = 3.0 nM

## 4.2 Transition-state inhibitors

The transition-state analogues may be recognized by the presence of a tetrahedrally-hybridized heteroatom, like phosphorus. The heteroatom usually replaces the last amide bond in a peptidic-like structure. Kinetic activity measurements and structural evidence suggested that the ionized P-O<sup>-</sup> group in the inhibitor with a tetrahedral phosphorus atom forms a coordinating bond with the zinc ion while the amino acid side chains are accommodated in diverse subsites along the active site cleft. Phosphorus-containing compounds have been widely accepted as CP inhibitors (Jacobsen & Bartlett, 1981; Grobelny et al., 1985; Yamauchi et al., 1985; Bartlett et al., 1987; Hanson et al., 1989; Bal et al., 1990; Phillips et al., 1992; Hill & Lowe, 1995; Payne et al., 1996; Fushihara et al., 2003; 2004; Suzuki et al., 2004). Some of them may be exceptionally potent. Such is the case of compounds Z-Phe-Val-po-Phe (Z = carbobenzyoxy), **11a**, the phosphonate ester analogue of the tripeptide Z-Phe-Val-Phe (Kaplan & Bartlett, 1991), and its thiophosphonate relative **11b**, Z-Phe-Val-thiopo-Phe (Fan et al., 1999).



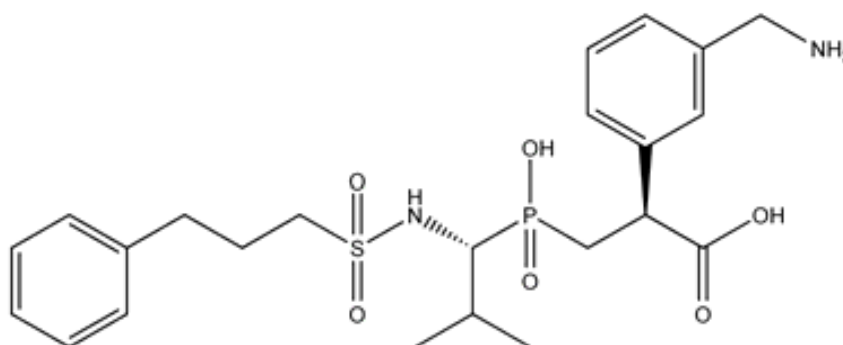
**11a**

R = O, Z-Phe-Val-po-Phe,  $K_i$  bCPA = 11.0 fM

**11b**

R = S, Z-Phe-Val-thiopo-Phe,  $K_i$  bCPA = 1.3 fM

The phosphonate and thiophosphonate ester tripeptides preserve one carboxylate group and the aromatic ring as **1**, but now the zinc coordination hinges on the phosphorus-containing linker between the Val-Phe group. Crystal structures of phosphorus-containing tripeptides in complex to CPA stressed their resemblance to the tetrahedral intermediate expected in the water-promoted mechanism, and not that expected in the anhydride mechanism involving nucleophilic attack by Glu270 (Kim & Lipscomb, 1990; 1991). Although being examples of the tightest macromolecule-ligand binding in Nature, these compounds did not share the fate of captopril, the ACE antihypertensive, and remain material for research purposes. Notwithstanding this BX 528, **12**, has showed promising properties in animal models of thrombolysis (Wang et al., 2007b; 2007c [Berlex Biosciences (a subsidiary of Schering AG)]).



**12**

BX528, (*S*)-2-[3-(aminomethyl)phenyl]-3-{hydroxy[*R*]-2-methyl-1-[[3-phenylpropyl)sulfonyl]aminopropyl]phosphorylpropanoic acid  
IC<sub>50</sub> CPU = 2.0 nM

Designed to mimic the tripeptide Phe-Val-Lys, **12** bears a sulfonamide and a phosphinate group replacing the corresponding amide bonds. A crystal structure of **12** in complex to porcine pancreas CPB revealed in atomic detail the bonding capabilities of this inhibitor together with an interesting account of its development (Adler et al., 2008).

# Objectives



# Objectives

In the introductory part of the Thesis we have witnessed the evolution of small molecular weight compounds from investigative purposes to potential applications like chemotherapy. However, in spite of dramatic progresses in this area, no compound targeted to a carboxypeptidase has reached the clinical practice yet. Wondering whether innovation is still feasible in this highly sought matter, we used different approaches to discover chemical entities useful as new scaffolds for the design of chemotherapeutic agents targeted to disease-linked metallo-carboxypeptidases.

Thus, the main objectives of this work are:

1 The exploration of structural space to identify and characterize protein-protein and protein-ligand interacting sites among M14 peptidases (or metallo-carboxypeptidases)

2 The exploration of chemical space to identify and characterize novel small molecular weight chemical scaffolds useful for obtaining new ligands for the above mentioned proteases

3 The structural characterization of protein-ligand complexes, i.e. metallo-carboxypeptidases bound to small molecular weight inhibitors, via modelling or X-ray diffraction analysis





# Results and Discussion



## 5 Discovery of novel inhibitors

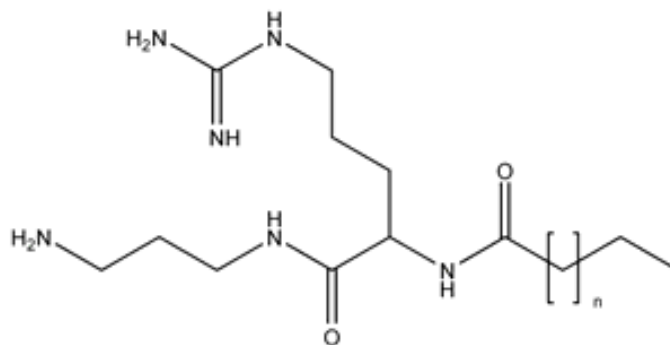
The goal of this part of the Thesis work was the identification of novel small-molecule synthetic CP inhibitor scaffolds. To this end, a variety of different classes of organic compound collections were screened as CP ligands. The screening collections can fall into two categories: i) “physical” compounds, real chemical entities either as unique compounds or as complex mixtures, and ii) “virtual” compounds, typically a list of entries in 1D ASCII format. An estimation of binding affinity can be obtained in a reasonable time from the *in silico* evaluation of all the records contained in such a database. To link both categories, a number of virtual compounds, judged by some qualifying criteria as candidates for further testing, were commercially acquired to become part of our physical compound inventory.

Overall, the performance of the screening procedure was good, showing a notable predictive power, but some failures were faced depending on the particular compounds. The peculiarities of the screening and hit selection for each family of compounds is fully detailed in each particular paper (see Publications Section). What matters here is the description of the features of the best representants of the different families screened. Table 1 lists in more detail the identity of the groups of small-molecules screened against M14 proteases.

Entry	Sample size and category	Supplier	Ref
DF1	16 (real)	Pere Clapés, IIQAB-CSIC, Barcelona, Spain	-
DF2 DF3	21 (real)	Rosa Ortuño, Departamento de Química Orgánica, UAB	Fernández et al., 2008 Fernández et al., 2009a
DF4 DF5	≈ 10,000 (virtual) 21 (real)	Paul J. Hergenrother, Univ. of Illinois (USA)	-
DF6	5,200 (real) 17 (real)	Àngel Messeguer, IIQAB-CSIC, Barcelona, Spain	-
DF7	≈ 300,000 (virtual) 34 (real)	Commercial (ASINEX)	Fernández et al., 2009b Fernández et al., 2009c

## 5.1 The arginine derivatives (entry DF1)

DF1 compounds were obtained through a collaboration with Dr. P. Clapès (IIQAB-CSIC, Barcelona). Compounds belonging to this group were demonstrated before to act as antimicrobials (Castillo et al., 2006) and surfactants (Piera et al., 1998; Morán et al., 2001). These compounds are, with some exceptions, derivatives of the amino acid arginine. As such, they were tested as inhibitors against B-type M14 proteases, enzymes with preference to cleave basic residue from the C-terminus of peptide and protein substrates. Some of them behaved as potent inhibitors against the lepidopteran CPB, or HzCPB, notably compound **13**, which displayed a  $K_i$  value below 10  $\mu\text{M}$ . This inhibitory potency is well below the value found for the vertebrate counterparts. Molecular modelling suggested that the hydrophilic part of **13** is inadequately positioned so as to coordinate the catalytic zinc ion in the active site of CPB, indicating that this compound owes much of its inhibitory activity to the large hydrophobic tail (Fig. 2).

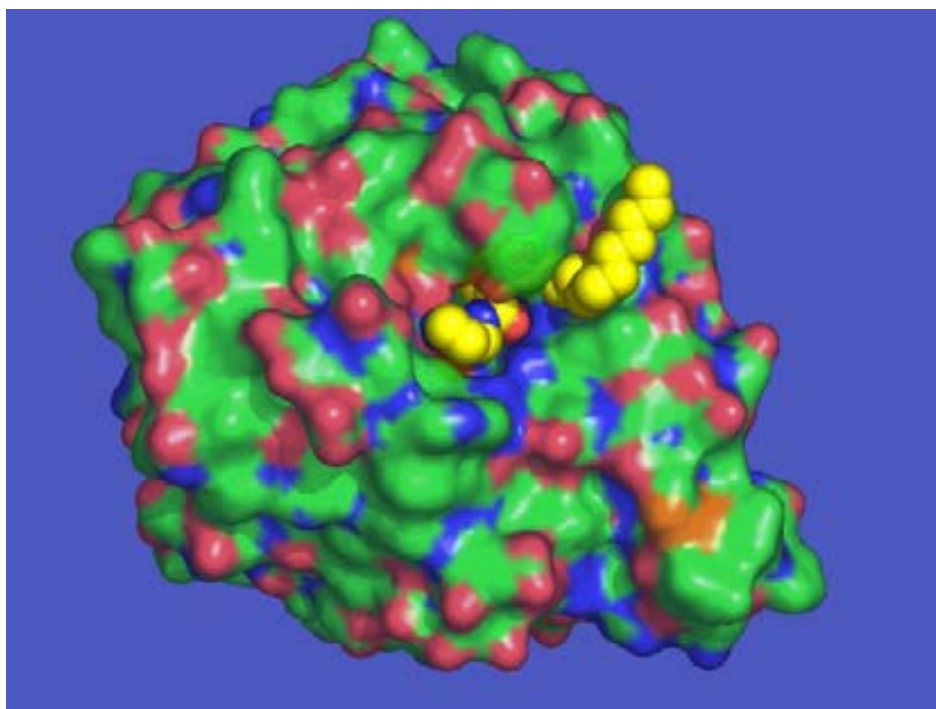


**13**

$n = 9$ ,  $K_i$  HzCPB = 7.9  $\mu\text{M}$

The fact that compound **13** inhibits the insect CPB more potently than its human counterpart, hCPB (the hCPB/HzCPB ratio is almost 10) may be related to what we found in a computational prediction of protein-protein docking sites. As will be fully discussed in Chapter 6, the analysis of M14 proteases using the

Optimal Docking Area (ODA) method revealed that the protein-protein docking surface displayed by the insect CPB is more than twice that of the human enzyme. Part of the protein-protein docking sites are occupied by the long hydrophobic tail of **13** as predicted by molecular docking (see Fig. 2). This configuration allows the hydrophilic polar part of the molecule to be buried in the active site cleft, while the hydrocarbon chain adopts an extended conformation. Because of its size, the latter would not fit into the active site cleft and therefore docks into an extended depression over the protein's surface.

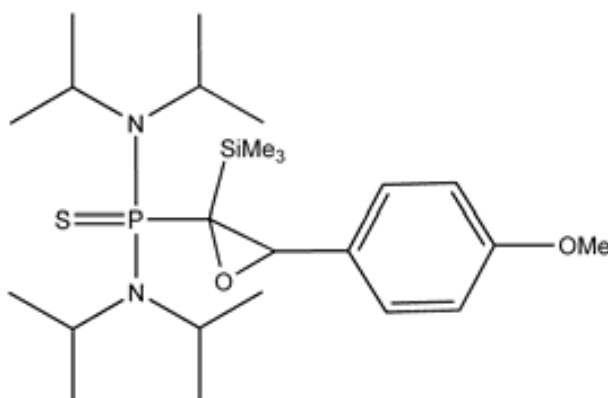


**Figure 2.** Docking of **13** in the crystal structure of lepidopteran HzCPB, 2c1c. The protein target is shown as a partially transparent surface colored by atom types, while the inhibitor is shown in CPK model with yellow carbons. The catalytically important residue Tyr248 is shown as a sticks model. Color codes: nitrogen (blue), oxygen (red), carbon (green), sulfur (orange).

## 5.2 Thioxophosphoranyl oxiranes (DF2) and cyclobutane $\beta$ -peptides (DF3)

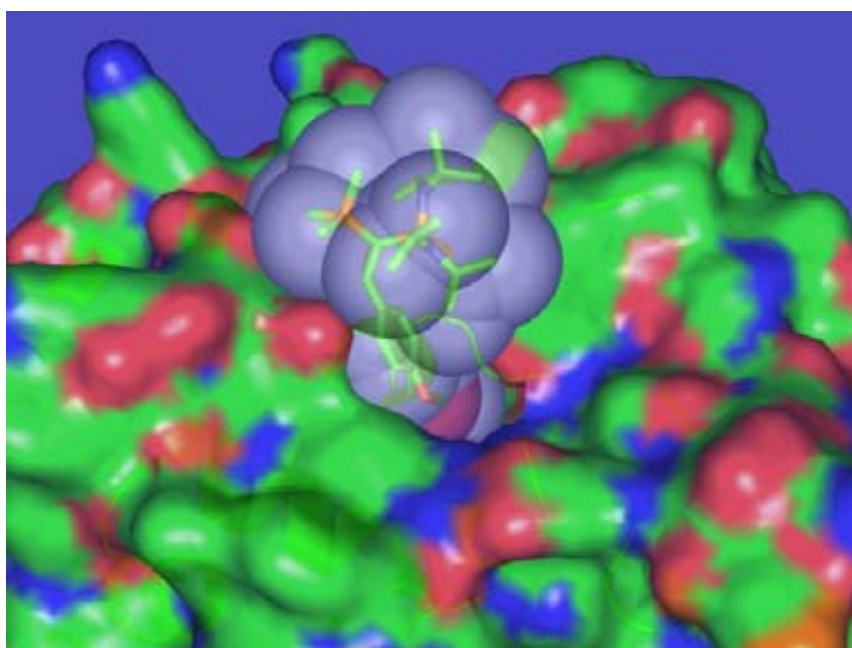
The thioxophosphoranyl oxiranes and cyclobutane  $\beta$ -peptides (CBPs) were kindly provided by Dr. Rosa M. Ortuño (Department of Organic Chemistry, Universitat Autònoma de Barcelona). These compounds were not originally intended as protease inhibitors, but their interest resides in their usefulness as templates and building blocks in organic synthesis.

The realization that DF2 compounds performed as potent inhibitors of an enzyme catalyzed reaction surpassed our expectations. Although the thioxophosphoranyl oxiranes have the ionizable groups fully blocked, the effect over the inhibitory potency was not deleterious, suggesting a distinct inhibitory activity mode, as will be discussed below. In spite of having the possible zinc coordinating atoms completely blocked with a pair of bulky *iPr* (*iPr* = *iso*-propyl) derivatives, these thioxo-compounds performed as low micromolar inhibitors of human CPB. Thus, compound **14** is a 1  $\mu$ M human CPB inhibitor.



**14**  
 $K_i$  hCPB = 1.0  $\mu$ M

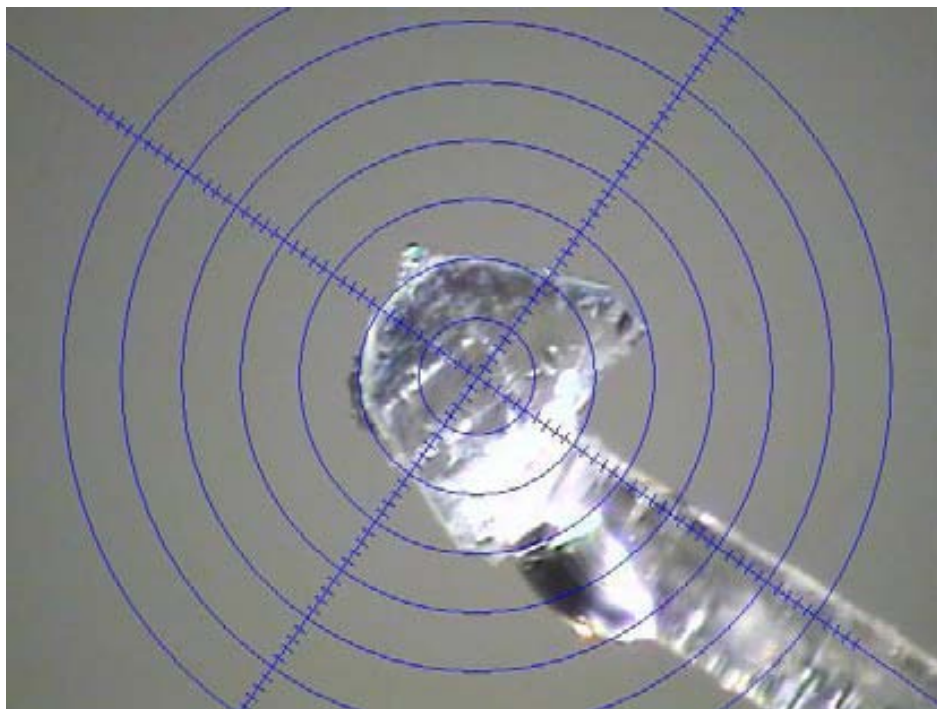
Docking of one such compound illustrated its probable mode of action (the modelling procedures are fully described in the Methods Section). As shown in Fig. 3, the bulky part of the inhibitor blocks the entrance of the enzyme's active site cleft, while the narrow end (the aryl moiety) enters it and contacts some residues inside the cavity. No atom from the inhibitor is in direct contact to the catalytic zinc ion, in sharp contrast to all known small-molecule CP inhibitors. This clearly suggests that binding may be mainly mediated by van der Waals contacts and hydrophobic interactions. On the whole, this mode of binding by capping the entrance of the active site cleft, resembles that of a proteinaceous inhibitor. Thus, thioxophosphoranyl oxiranes may be regarded as nature-inspired small-molecule CP inhibitors.



**Figure 3.** Docking of one thioxophosphoranyl oxirane compound (shown as a partially transparent surface colored light blue) into human CPB crystal structure, 1zli. The bulk of this small-molecule is facing the external solvent. The protein is also shown as a partially transparent surface. The atom color codes are as in the legend to Fig. 2.

A co-crystallization experiment to obtain a crystal structure of the complex between one thioxophosphoranyl oxirane and the target macromolecule furnished an adventitious outcome. A single-crystal grew up

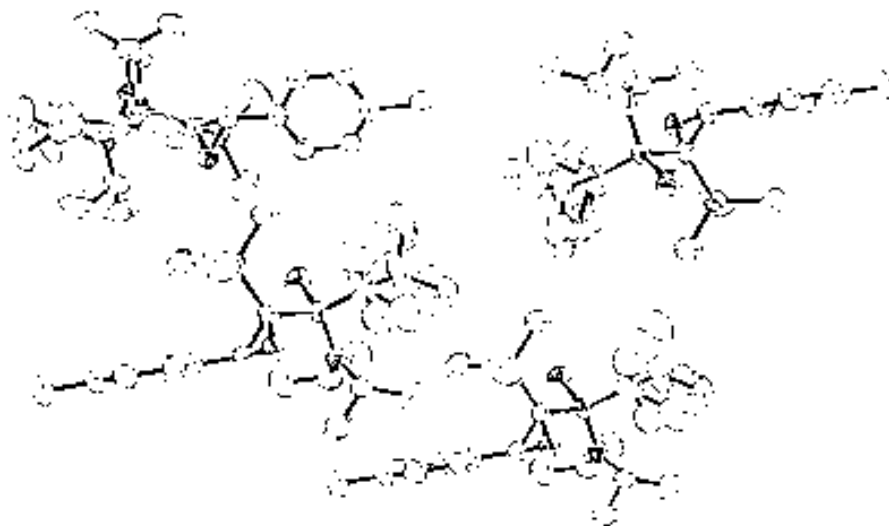
from a drop (Fig. 4) which, analyzed through X-ray diffraction, turned out to not be the desired protein-ligand complex, but rather the very inhibitor molecule.



**Figure 4.** A large crystal grew up from a co-crystallization experiment. The identity of the salt-like molecule was unambiguously determined by using single-crystal X-ray diffraction. The picture was taken at the X-ray generator. The crystal is mounted at the tip of a glass fiber using a popular fast glue. The separation between adjacent concentric lines is  $0.1 \mu\text{M}$ . The face opposite to the glued one shows an irregular shape due to the manual division of the crystal into small pieces.

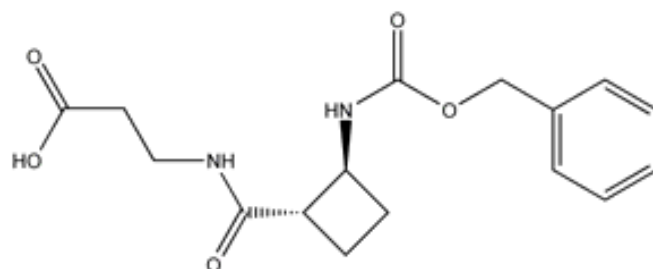
Interestingly, the packing interactions in the small-molecule unit cell are mediated solely by van der Waals contacts and hydrophobic interactions (Fig. 5). In addition, the  $(i\text{Pr})_2$  groups are loosely interacting with neighbouring molecules and show therefore positional disorder (Fernández et al., 2008).





**Figure 5.** An ORTEP (Farrugia, 1997) plot showing the packing in the crystal of one thioxophosphoranyl oxirane with atomic displacement ellipsoids drawn at the 50 % probability level. Heavy atom ellipsoids are shown as octant-shading style, while those from C are in full surface style. Note the disorder at the (iPr)<sub>2</sub> groups seen as the largest ellipsoids in the structure. Hydrogen atoms were omitted for clarity.

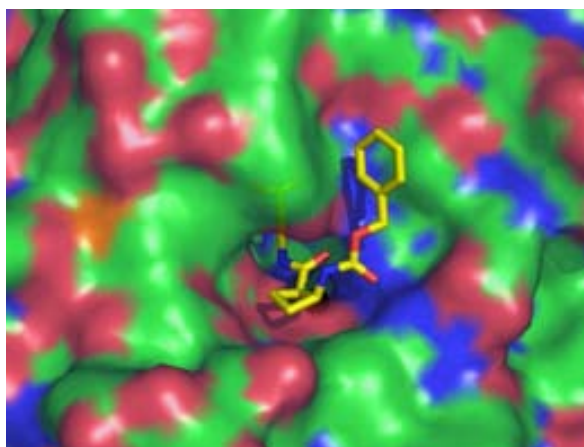
The cyclobutane  $\beta$ -peptides (CBPs), retain to some degree the character of the conventional  $\alpha$ -peptides, although with some marked contrasts: i) CBPs are structurally rigid entities due to the presence of the cyclobutane moiety, ii) they are synthetic building blocks able to form a wide diversity of arrangements, and most importantly, iii) CBPs are easily to produce by organic synthetic methods. As the previously commented compounds, the CBPs have not been designed as protease inhibitors. Their CBP inhibitory activity is comparatively weaker than that of other compounds in our work; however, they were fairly selective for a given macromolecule (Fernández et al., 2009a). The chemical structure of **15** is shown as an example of a CBP.



**15**

$K_i$  hCPB = 43.0  $\mu$ M

Binding of **15** to the protein target was *a priori* guessed as problematic because of its rigid structure as opposed to the more flexible one from the natural CP substrate, an  $\alpha$ -peptide. However, the docked structure indicated that the compound is able to fit in the active site of the enzymes, CPA and CPB, in an unprecedented manner. The cyclobutane  $\beta$ -peptide structure is bent at the position of the four-membered ring, and the shape of the inhibitor fits the bulge created by the phenol ring of the catalytically relevant residue Tyr248. One end of the inhibitor contacts the bottom of the active site cleft while the other end is exposed at the surface, where the *Z* moiety is immersed in a depression partially walled by Tyr248 phenolic ring. In Fig. 6, the interaction between the aromatic rings is edge-to-face; for other congeners, it may be face-to-face.



**Figure 6.** Docking of the cyclobutane  $\beta$ -peptide **15** (yellow carbons) into human CPB crystal structure, 1zli. The black sphere represents the catalytic zinc ion. Other details as in Fig. 2.

## 5.3 *L*-alanyl derivatives (DF4) and organosilicon compounds (DF5)

These compounds were obtained from a generous gift by Prof. Paul J. Hergenrother (Department of Chemistry, University of Illinois at Urbana-Champaign, IL, USA). We identified these compounds from: i) a high-throughput virtual screening procedure, and ii) a selection guided by a preference of a chemical element.

Class DF4 consists of a series of *Z*-*L*-Ala derivatives harboring a hydrocarbon chain of variable length ranging from zero to fourteen C atoms (the general scheme can be seen as an inset in Table 2). In spite of having a simple chemical design lacking an obvious metal-coordinating functionality, they behave as potent CP inhibitors against CPB type carboxypeptidases (Table 2).

Compound	R group	bCPA <sup>b</sup>	pCPB	HaCPA	H <sub>z</sub> CPB
<b>16a</b>	H	118	129	163	89.1
<b>16b</b>	CH <sub>3</sub>	95.4	126	79	135
<b>16c</b>	(CH <sub>2</sub> ) <sub>1</sub> CH <sub>3</sub>	117	30.7	37.5	67.3
<b>16d</b>	(CH <sub>2</sub> ) <sub>3</sub> CH <sub>3</sub>	124	12.3	12.5	3.3
<b>16e</b>	(CH <sub>2</sub> ) <sub>5</sub> CH <sub>3</sub>	58.2	8.5	90.6	3
<b>16f</b>	(CH <sub>2</sub> ) <sub>7</sub> CH <sub>3</sub>	70.2	11.5	37.5	2.2
<b>16g</b>	(CH <sub>2</sub> ) <sub>9</sub> CH <sub>3</sub>	38.5	12.1	109	0.76
<b>16h</b>	(CH <sub>2</sub> ) <sub>11</sub> CH <sub>3</sub>	10.8	5.75	2.6	2
<b>16i</b>	(CH <sub>2</sub> ) <sub>13</sub> CH <sub>3</sub>	84.2	32.8	9.4	4.6

Notes  
<sup>a</sup>*K<sub>i</sub>* in μM  
<sup>b</sup>bCPA, bovine carboxypeptidase A; pCPB, porcine carboxypeptidase B; HaCPA, carboxypeptidase A from cotton bollworm (*Helicoverpa armigera*); H<sub>z</sub>CPB, carboxypeptidase B from corn earworm (*Helicoverpa zea*)

The *Z-L*-Ala derivatives show inhibitory potencies in the lower micromolar range. Within the series, compounds with no or shorter hydrophobic tails, **16a-c**, are less potent than the variants with longer hydrocarbon chains. The inhibitory potency increased from compound **16d** on, to reach a value below 11  $\mu\text{M}$  against all target enzymes for compound **16h**. In particular, **16g** is an extremely potent inhibitor of insect CPB, HzCPB, displaying a  $K_i$  value of 0.76  $\mu\text{M}$ . On the whole, the *Z-L*-Ala derivatives behaved as weak inhibitors against CPA, especially the bovine counterpart. Besides, some compounds show selectivity in binding CPB enzymes, as **16d** and **16e** show a 10- and 7-fold CPB/CPA inhibition ratio for the mammalian and insect classes, respectively. As class DF4 compounds possess no conspicuous metal-coordinating functionality, we can conclude that binding is mainly governed by hydrophobic interactions, in a manner reminiscent of class DF1 compounds. In both cases, the preferential reactivity of organic compounds harboring long hydrocarbon chains against insect CPs may be in line with what was seen from the ODA analysis of protein-protein docking areas (see Chapter 6).

Class DF5 involve silicon-containing compounds. Naturally-occurring biological active compounds of silicon, and other semi-metals, are found in some prokaryotes, fungi and plants (Rezanka & Sigler, 2008). In organic chemistry, silicon is a versatile element that can afford, readily and with a relatively low toxicity, a rich variety of structures (Greenwood & Earnshaw, 1997; Miura & Hosomi, 2004). The synthesis of organosilicon compounds have captured the interest of medicinal chemists since silicon is an excellent tetrahedral isostere to carbon with similar bonding capabilities and distinct physicochemical properties. An added benefit is the possibility of generating new chemical entities exploiting established structure-activity relationships. Examples on the use of organosilicon compounds as protease ligands include some potent silicon diol inhibitors of the CPA-related metallopeptidase angiotensin-converting enzyme (ACE), and the HIV aspartyl protease (Showell & Mills, 2003). Silicon-bearing compounds used as CP inhibitors have been given little attention, and only a few examples can be found in the literature: 2,2-dimethyl-2-silapentane-5-sulfonic acid (Frye & Sebastian, 1990; Sebastian et

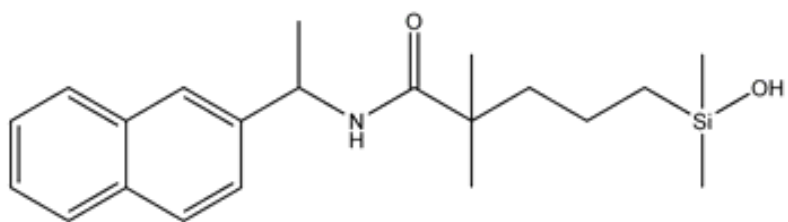
al., 1996), which harbors a trimethylsilyl moiety, and a derivative of carboxyethyl germanium sesquioxide (Komuro et al., 1986; Kakimoto et al., 1988).

11 silicon-containing compounds were screened against representative M14 proteases. As collected in Table 3, the organosilicon compounds behaved as potent CP inhibitors.

<b>Table 3</b>				
<b>Inhibitory constants for DF5 compounds<sup>a</sup></b>				
Compound	bCPA <sup>b</sup>	hCPB	HzCPB	pCPB
<b>17a</b>	4.4	0.87	0.31	7.7
<b>17b</b>	202	13.6	15.4	27.4
<b>17c</b>	157	63.7	138	59.7
<b>17d</b>	1,080	787	1,630	1,210
<b>17e</b>	68.4	9.3	0.49	96.9
<b>17f</b>	410	146	310	211
<b>17g</b>	100	8.0	0.42	3.0
<b>17h</b>	118	43.2	0.6	168
<b>17i</b>	48.7	577	1.04	1,054
<b>17j</b>	350	29.0	9.3	40.2
<b>17k</b>	1,180	27.1	1.48	950

Notes  
<sup>a</sup> $K_i$  in  $\mu\text{M}$   
<sup>b</sup>bCPA, bovine carboxypeptidase A; hCPB, human carboxypeptidase B; HzCPB, corn earworm (*Helicoverpa zea*) carboxypeptidase B; pCPB, porcine pancreatic carboxypeptidase B

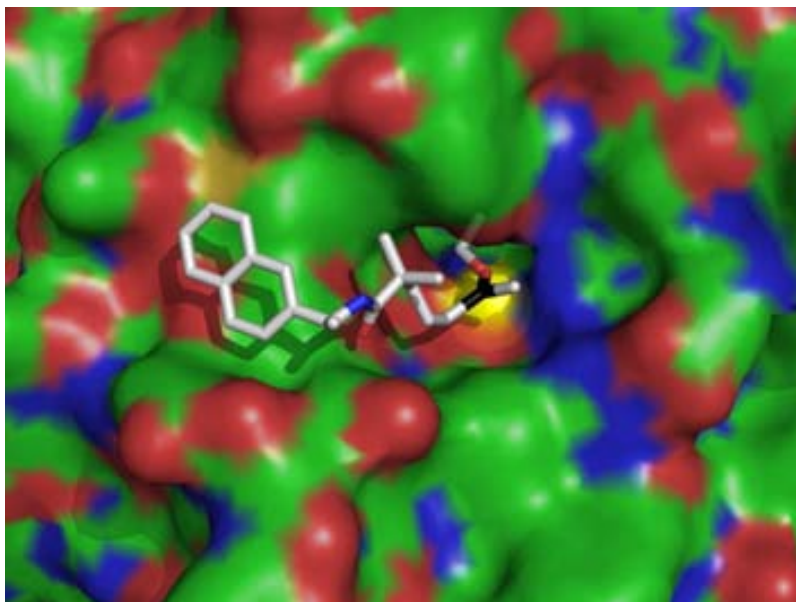
Compound **17a**, a large four-membered heterocyclic fused system, is exceptional since its low micromolar inhibition of both CPA and CPB. The predicted binding mode of **17a** indicated that it docks over the protein surface and locates at the entrance of the enzyme's active site cleft, far from the specificity pocket of the enzyme. Other compounds performed as specific CPB inhibitors. Some of these like **17g**, possess an extended backbone akin to an  $\alpha$ -peptide.



**17g**

5-(ethoxy(methylene)silyl)-2,2-dimethyl-*N*-(1-(naphthalen-2-yl)ethyl)pentanamide  
 $K_i$  hCPB = 8.0  $\mu$ M

Because of its features, we can regard **17g** as the most interesting lead compound of the group. Molecular docking of **17g** (Fig. 7), revealed that the silicon atom resides in the proximity of the catalytic zinc ion while the naphthyl moiety is placed over the protein's surface. The surface patch where the naphthyl moiety sits highlights an extended docking area useful to be exploited in the design of improved CPB inhibitors.



**Figure 7.** Docking of **17g** into human CPB crystal structure, 1zli. The inhibitor is shown in sticks model with white carbons and the Si atom colored black. The yellow sphere represents the catalytic zinc ion. Other details as in Fig. 2.

## 5.4 Peptoids (DF6)

Peptoids were obtained through a collaboration with Dr. Àngel Messeguer (Departament de Química Orgànica Biològica, IIQAB-CSIC, Barcelona). Peptoids are structurally related to natural peptides, and thus belong to the category of peptidomimetics. At variance with the naturally occurring peptides, a peptoid is an oligomer of *N*-substituted glycines (Anne et al., 1998). In fact, DF6 is a library made of three units of *N*-alkyl-glycine in which the three substituent positions R<sub>1</sub>, R<sub>2</sub> and R<sub>3</sub> are positionally varied. These three points of diversity originate the three sub-libraries R<sub>1</sub>-R<sub>3</sub>, which comprise more than 5,000 compounds in all. In the first step, the structurally diverse compounds are submitted to screening in a rapid search for lead compounds without isolation or purification of individual molecules. An ensuing step would be the deconvolution process which allows the identification of individual compounds from the mixtures and their subsequent use for testing their biological activity.

Results from the screening of the molecular library of peptoids against carboxypeptidases are shown in Table 4.

<b>Table 4</b>			
<b>First step screening: peptoid mixtures with biological activity</b>			
Target	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>
bCPA <sup>a</sup>	<b>13</b> (64.8) <sup>b</sup>	<b>28</b> (66.6)	<b>37</b> (15.0) <b>40</b> (15.0) <b>44</b> (20.4) <b>42</b> (21.7)
HaCPA	<b>13</b> (59.5)	<b>36</b> (29.6) <b>29</b> (33.3)	<b>47</b> (40.7) <b>48</b> (40.7) <b>52</b> (40.7)
hCPB	<b>9</b> (56.3) <b>13</b> (65.6)	<b>21</b> (31.9) <b>25</b> (43.2) <b>26</b> (32.8)	<b>37</b> (40.6) <b>42</b> (53.1)
HzCPB	<b>12</b> (46.0) <b>5</b> (59.0)	<b>34</b> (27.3)	<b>50</b> (56.3)
Notes			
<sup>a</sup> bCPA, bovine carboxypeptidase A; hCPB, human carboxypeptidase B; HaCPA, carboxypeptidase A from cotton bollworm ( <i>Helicoverpa armigera</i> ); HzCPB, carboxypeptidase B from corn earworm ( <i>H. zea</i> )			
<sup>b</sup> Numbers in <b>boldface</b> indicate the mixture label. Values in parenthesis refer to the residual activity with respect to the control (%)			

Some peptoid mixtures showed an inhibitory activity against CPs. Although R<sub>1</sub> affected only slightly the enzymatic activity, the effect of R<sub>2</sub> and R<sub>3</sub> is more pronounced. It should be mentioned that complex mixtures of compounds may actually result in combined, complex effects. The actual results show, in fact, an integration of partial activities: some peptoid mixtures may be activators while others may be inhibitors of the enzymatic activity. To obtain what is called a defined peptoid, the inhibitory functionalities must be concentrated into a single peptoid structure. The first screening suggested that such a second generation can be created because activity was detected across the three sub-libraries. As a result of the deconvolution process, 17 unique peptoids were synthesized and assayed as CP inhibitors. Some of them were effective as human CPB inhibitors, and showed an inhibitory activity against other human CPB-like enzymes (Table 5) but no activity was detected against CPA.

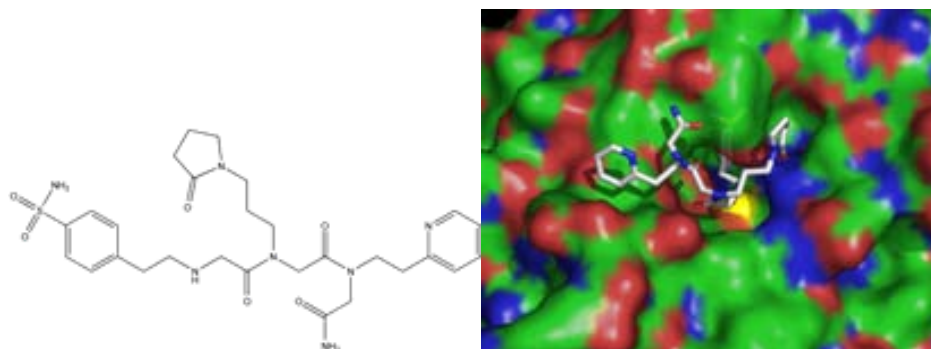
Compound	hCPB	hCPN	hCPU
<b>18a</b>		+ <sup>b</sup>	
<b>18b</b>		+	+
<b>18c</b>	40.5 <sup>a</sup>		
<b>18d</b>	8.9	+	
<b>18e</b>	9.0	+	
<b>18f</b>	13.6		
<b>18g</b>	3.4	+	
<b>18h</b>	2.6	+	+

Notes  
<sup>a</sup>K<sub>i</sub> value (μM)  
<sup>b</sup>the plus sign (+) marks that activity was detected at a fixed unique inhibitor concentration

As can be seen from inspection of Table 5, only half of the defined peptoids synthesized exerted an inhibitory effect against the target macromolecules. The most potent ones, **18g** and **18h**, show an inhibitory activity that falls in the lower micromolar range. **18g** cross-reacted with CPN, but did not inhibit CPU (CPN and CPU are two plasma-circulating carboxypeptidases). **18h** was found to inhibit CPN and CPU as well. To identify potential points for improvement, and to obtain a more discriminant inhibitor, compound **18h** was



further investigated by molecular docking. The chemical structure of the compound and the result from docking are shown below in Fig. 8.

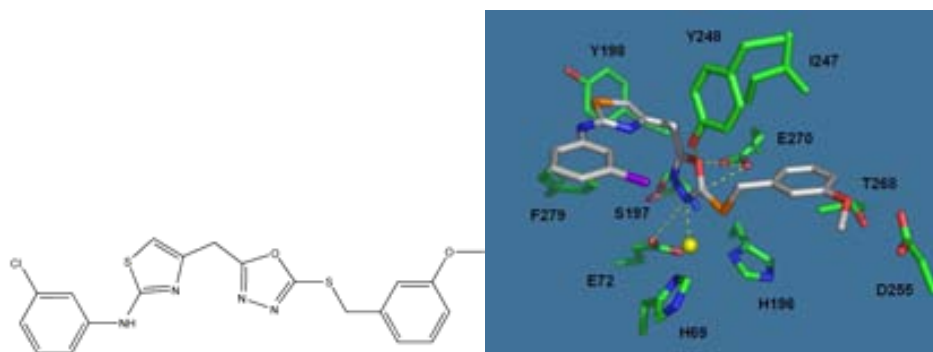


**Figure 8.** (*Left Panel*), chemical connectivity of **18h**,  $K_i$  hCPB = 2.6  $\mu$ M. (*Right Panel*), predicted binding of **18h** in the crystal structure of human CPB, 1zli. The inhibitor is shown with white carbons, while the protein is represented as a semi-transparent surface colored by atom types. The bulge at about 12 o'clock corresponds to the Tyr248 phenolic ring. The catalytic zinc ion is shown as a yellow sphere. Other details as in Fig. 2.

According to docking, inhibitor **18h** sits over three surface patches that surround the active site cleft entrance. Two of the peptoid's substituents wrap Tyr248 phenolic ring, while a third one sits over a less pronounced crevice. The aryl sulfonimine moiety points toward the bottom of the active site cleft. Only the carbonyl group in this branch is in the proximity of the catalytic zinc ion. Here, a potential anchor point for the molecule is not fully exploited by the peptoid and would constitute one point of improvement. The aryl sulfonimine moiety extends towards the bottom of the active site cleft contacting residues near the specificity pocket of the macromolecule. It is desirable to make convenient exploit of the peptoid's backbone to afford optimized third-generation inhibitors. These structures have the added benefit of their ability to overcome the problems of metabolic instability in the gastrointestinal tract.

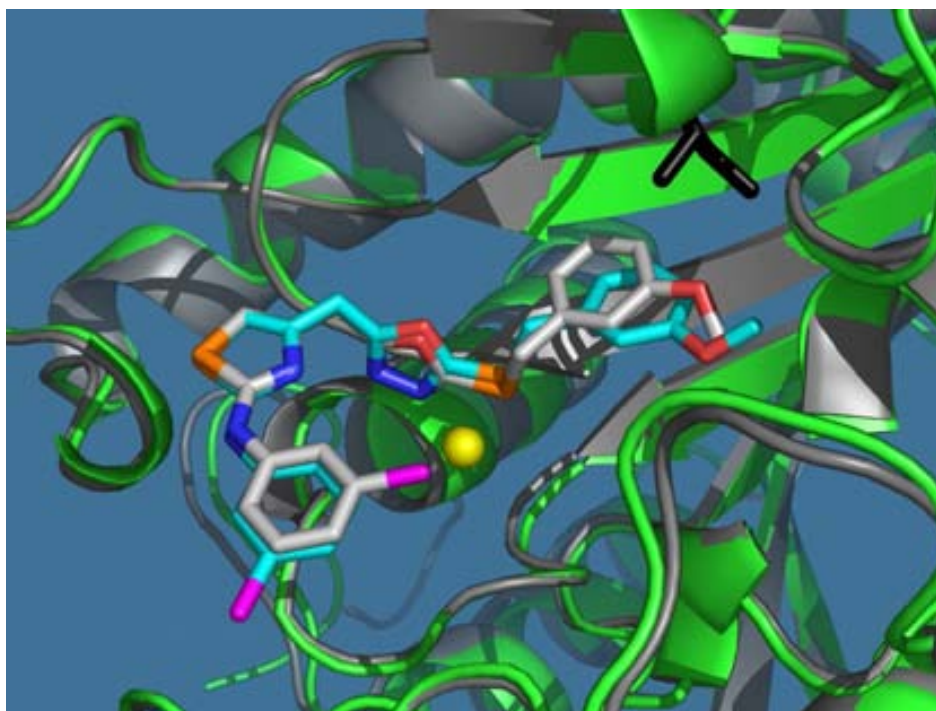
## 5.5 Oxadiazoles (DF7)

This scaffold was identified from database searches as published (Fernández et al., 2009b; 2009c). Compound **19**, *N*-(3-chlorophenyl)-4-((5-(3-methoxybenzylthio)-1,3,4-oxadiazol-2-yl)methyl)thiazol-2-amine, is an outstanding member of this group, which performed as a potent CPB inhibitor. **19** harbors a 1,3,4-oxadiazole moiety, which is a distinct feature among small-molecule CP inhibitors. Docking of **19** into human CPB illustrated its mode of action (Fig. 9). It involves electrostatic interactions with crucial residues for catalysis (Tyr248, Glu270). One nitrogen atom from the 1,3,4-oxadiazole moiety is coordinating the catalytic zinc ion while the oxygen atom is hydrogen bonded to the Tyr 248 phenol hydroxyl. The latter interaction may assure the firm binding of this part of the inhibitor molecule. Other parts of the 1,3,4-oxadiazole inhibitor seem less favourably engaged to residues at the active site of the enzyme, suggesting possible points of improvement.



**Figure 9.** (Left Panel), chemical connectivity of **19**,  $K_i$  hCPB = 1.2  $\mu$ M. (Right Panel), predicted binding of **19** into human CPB (1zli) active site cleft. Important residues are shown as the side chain in stick model (green carbons) while the ligand is shown with white carbons. The catalytic zinc ion is depicted as a sphere colored yellow. Hydrogen bond interactions are indicated by yellow dashed lines.

As shown in Fig. 10 below, the methoxyphenyl fragment is constrained by an Ile residue at position 243 in CPA (the equivalent residue in CPB is Gly). Therefore, the 1,3,4-oxadiazole scaffold of inhibitor **19** represents a new type of chemical bonding that can be considered as an interesting lead for further development in the design of novel inhibitors targeted to CPBs.



**Figure 10.** A comparison of **19** binding into human CPB (green cartoon) and bovine CPA (grey). Ile243 in CPA is highlighted black in contrast to a Gly in the equivalent position in CPB to mark its effect on **19** binding. The catalytic zinc ion is depicted as a sphere colored yellow.



# 6 Bioinformatics

This Chapter concerns our investigations aimed at the rationalization of functional and structural aspects among M14 peptidases from a bioinformatics approximation. This was done using the ODA (“Optimal Docking Area”) method. Chronologically, this work has been performed before others in this Thesis; as a consequence, certain results derived from bioinformatics relate to findings in the previous Chapter. Therefore, the analysis of protein-protein interactions among carboxypeptidase crystal structures with the ODA method will be described first, to continue with the application of this kind of analysis to the particular cases of protein-protein interactions in the activation process of mammalian and ruminant CPA zymogen forms as well as to the understanding of protein-ligand interactions with inhibitors identified by the methods in Chapter 5.

## *6.1 The crystal structure pool*

Crystal structures of M14 proteases with available 3D coordinates from the PDB (as of early 2007) were used throughout the study. Thus, this work lacked the analysis of some interesting CPs whose 3D structures were disclosed recently, like CPN, which physiologically is a tetramer, and the CPU zymogen from different sources. The crystal structures of unbound proteins, namely those coded 1m4l, 1z5r, 2c1c, and 1obr, were used to build models of complexes with different binding partners. In the cases of the structures of the bovine CPA1 (1m4l) and the porcine CPB (1z5r) it served the purpose of checking the accuracy of our predictions against the available structure of the complexes 1pyt and 1nsa, respectively. On the contrary, since there is no structural information available on complexes of TvCPT (1obr) and HzCPB (2c1c), the predictions must be evaluated with caution in these cases. We regard the catalytic domain as central to all CPs; therefore, the ligand molecule

in the complex is referred to as the binding or interacting partner. Overall, 24 polypeptide chains were used for our study.

## *6.2 Interactions of the catalytic domain*

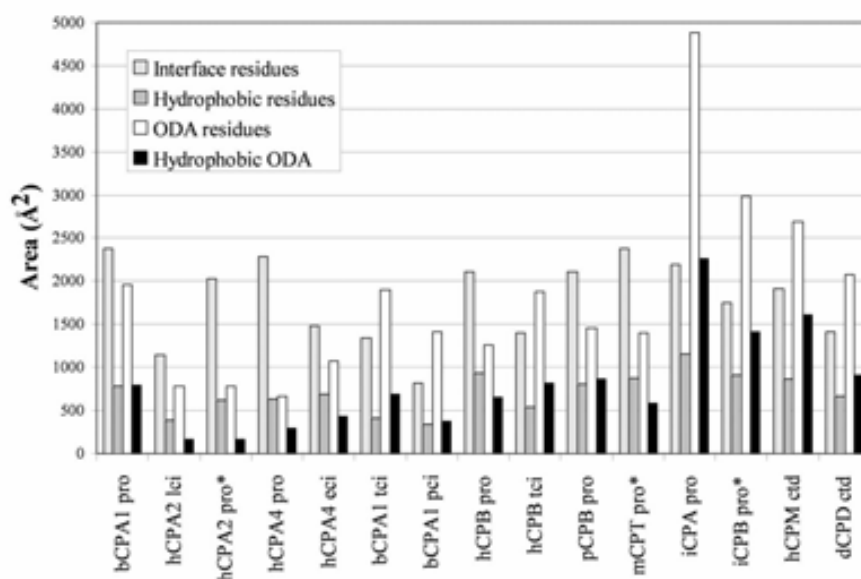
As can be seen from Table 6 (next page), the experimental interactions of the catalytic domain involve from 22 to 29 residues with the N-terminal pro domain (entries 1-7, and 10; column 7) and slightly fewer residues with the C-terminal domain (18 on average for entries 8, 9). In contrast, the number of residues found at the interface with the protein inhibitors is quite variable and ranges from 9 to 20. The results obtained for the modelled protein-protein complexes formed upon structural superimposition of the unbound to the bound catalytic domain are in good agreement with respect to the real complexes. This can be seen, for instance, by comparing the data of the real complexes of bCPA1 (entry 3 in Table 6) to the modelled ones obtained by structural superimposition of the unbound to the bound catalytic domain (marked by an asterisk). Besides, the values obtained for the modelled complexes of HzCPB (entry 7) and TvCPT (entry 10) are comparable to those obtained for the real complexes.

<b>Table 6</b> <b>Analysis of protein-protein interactions in the catalytic domain of CPs</b>								
	Catalytic domain <sup>a</sup>	Chain length <sup>b</sup>	ODA residue number <sup>c</sup>	Energy <sup>d</sup>	Partner mol <sup>e</sup>	Interface residues <sup>f</sup>	Predictive value <sup>g</sup>	Rank of ODA at interface <sup>h</sup>
1	hCPA2	303	10	-19.31	pro*	22	80	1
					lci	14	30	1
2	hCPA4	310	8	-17.12	pro	25	88	1
		307	13	-16.74	eci	19	60	2
3	bCPA1	309	21	-18.69	pro	27	50	2
		307	17	-18.21	pci	10	20	1
		301	22	-23.58	tci	15	60	1
		307	16	-17.98	pro*	29	80	1
						pci*	12	20
					tci*	17	70	2
4	HaCPA	317	54	-51.32	pro	25	90	1
5	hCPB	307	12	-16.75	pro	25	80	1
		306	23	-28.48	tci	19	90	1
6	pCPB	306	17	-16.66	pro	23	80	1
		304	12	-19.46	pro*	23	50	1
7	HzCPB	312	38	-30.33	pro*	23	90	1
					tci*	20	70	1
8	dCPD	300	21	-30.42	ctd	16	10	7
9	hCPM	295	28	-30.96	ctd	20	10	1
10	TvCPT	323	18	-18.57	pro*	27	80	1
					pci*	9	40	1

Notes  
<sup>a</sup>h, human; b, bovine; p, porcine; i, insect; d, duck; m, microbial.  
<sup>b</sup>Number of residues contributing to ODA calculation.  
<sup>c</sup>Number of significant ODAs.  
<sup>d</sup>Energy value of the lowest-energy ODA residue, kcal/mol  
<sup>e</sup>pro, pro domain; lci, leech carboxypeptidase inhibitor; eci, human endogenous carboxypeptidase inhibitor; pci, potato carboxypeptidase inhibitor; tci, tick carboxypeptidase inhibitor; ctd, C-terminal domain. An asterisk (\*) denotes that the analysis was performed on a modelled complex formed by structural superimposition of the unbound onto the bound molecule.  
<sup>f</sup>Number of residues of the catalytic domain that are at the interface.  
<sup>g</sup>Percentage of the ODA residues (among the significant top 10) of the catalytic domain that are at the interface, %  
<sup>h</sup>Rank of the lowest-energy ODA residue located at interface.

## 6.3 Nature of the ODA patches

The surface area formed by the residues at the known experimental interfaces ranges from 800 Å<sup>2</sup> (in the complexes of the catalytic domain with the protein inhibitors) to almost 2500 Å<sup>2</sup> (in the interaction between the catalytic domain and the N-terminal pro domain) (see Fig. 11). The average percentage of hydrophobic residues at the known experimental interfaces is close to 40% (ranging between 33 and 53%), which correlates well with an average of 42% of the total surface area being of hydrophobic nature. Except for the insect HaCPA, there are fewer hydrophobic residues at the interface with the binding partner for CPA1, CPA2, and CPA4 (close to 35%) than for hCPB, pCPB, dCPD, and TvCPT (between 45-50%). Notably, the interface in the insect CPA has an above average hydrophobic content (53%). A similar trend is observed for the insect HzCPB (52%) in the complex modelled by structural superimposition.



**Figure 11.** Surface area of the residues at the predicted ODA patches and at the experimental protein-protein interfaces. The values shown correspond to the total set of predicted ODA residues (open bars) and to the subset of hydrophobic ODA residues (filled bars). The complexes of human CPA2, microbial CPT and insect CPB modelled upon structural superimposition are marked with an asterisk (\*).



## 6.4 Predictive success of the method

The previously described set of structures of CPs was analyzed through the Optimal Docking Area (ODA) method in search of putative protein-binding areas. We evaluated whether the predicted protein interaction sites defined by the significant low-energy ODA residues were located in the experimentally determined protein-protein interfaces. Since in some cases more than ten ODA residues with  $< -10$  kcal/mol were identified that defined surface areas much larger than the average protein-protein interfaces, it was decided to use the ten lowest-energy ODAs for the predictions.

The ODA positive predictive value for the ODA patches in the central catalytic domain (Table 6), although being generally good among CPs, is quite variable and seems to be dependent on the kind of binding partner that is present in the complex. In 10 out of 10 cases (100% success rate), including the unbound structures for which the complex structure had to be modelled, the ODA residues were correctly located (positive predictive value  $\geq 50\%$ ) at the interface with the N-terminal pro domain. Similarly, in 4 out of 4 cases (100% success rate), the ODA residues correctly predicted the interface with tick carboxypeptidase (TCI) inhibitor. However, in none of the cases did the ODA residues fully predict the location of the interface with the leech (LCI), human endogenous (ECI), or potato (PCI) carboxypeptidase inhibitors. Interestingly, in 16 of 21 cases, the ODA residue ranked 1 (according to ODA energies) was at the interface with either the N-terminal pro domain or any inhibitor (including those inhibitors for which the overall ODA positive prediction values were below the 50% cutoff). Moreover, in 19 of 21 cases (90% success rate), an ODA residue with either rank 1 or 2 (according to energies) is located at the experimental interface. This indicates that the lowest-energy ODAs were indeed located at a surface region that might be important not only for binding to the N-terminal pro domain, but also to any of the inhibitors, in spite of the different atomic interactions that every type of inhibitor may be involved in. In the majority of the M14A cases (excluding insect and bacterial CPs), the residue with lowest-energy ODA is either aspartate or glutamate at position 163 (bovine CPA1 numbering). As it is discussed later, this position is strictly conserved in the

M14A subfamily, and is located in a loop that is one of the easiest points of cleavage by endoproteases. A summary of this analysis can be found in Table 7.

Subfamily	ODA range <sup>b</sup>	Energy <sup>c</sup>	Success rate <sup>d</sup>	Binding partner	ODA range <sup>e</sup>	Energy <sup>f</sup>	Success rate <sup>g</sup>
M14A <sup>h</sup>	8-54	-51.32	100 (5/5)	pro domain	1-24	-24.79	20 (1/5)
M14A <sup>i</sup>	10-23	-28.48	60 (3/5)	inhibitor	0-25	-38.43	60 (3/5)
M14B	21/28	-30.96	0 (0/2)	c-terminal domain	8/15	-25.41	100 (2/2)
Overall catalytic domain	8-54	-51.32	67 (8/12)	Overall binding partner	0-25	-38.43	50 (6/12)

Notes

<sup>a</sup>Complexes formed by structural superimposition are not summed up here.

<sup>b</sup>Maximum and minimum number of ODA residues of the catalytic domain.

<sup>c</sup>Energy value of the first ranked ODA residue of the catalytic domain, kcal/mol

<sup>d</sup>Percentage of cases where the ODA positive predictive value for predicting the interface is above the cutoff of 50%. Values in brackets indicate the number of cases above the minimum and the total number of cases, %

<sup>e</sup>Maximum and minimum number of ODA residues in the binding partner molecule.

<sup>f</sup>Energy value of the first ranked ODA residue of the binding partner molecule, kcal/mol

<sup>g</sup>Ratio of cases where the ODA positive predictive value is above the 50% level over the total number of binding partner molecules. Values in brackets indicate the number of cases above the minimum and the total number of cases, %

<sup>h</sup>CPs of the M14A subfamily where the pro domain structure is present.

<sup>i</sup>CPs of the M14A subfamily that are in complex with inhibitor.

When all the predictions (in catalytic domains and in their partners) are compared with only the experimentally determined interfaces (i.e. no complex models included), in a total of 14 over 24 cases, the predicted ODA patches (see the “Overall catalytic domain” and “Overall partner” entries in the Table 7) were successfully located (that is, more than 50% of the predicted residues were at the interface), which yields an overall success rate of 58% for the analyzed set of structures. The success rates of the ODA predictions for the catalytic domain were in general high, except in the case of its interactions with the C-terminal domain. Indeed, the ODA predictions in the catalytic domain of the members of the M14A subfamily (entries M14A<sup>h</sup> and M14A<sup>i</sup> in Table 7) were correctly located in 8 out of 10 cases (i.e. success rate of 80%). In spite of being sequentially and structurally unrelated, the success rate is similarly high for the CP protein inhibitors (there are no significant ODA residues in the TCI structure,

which is present in two complexes). On the contrary, the predictions for the N-terminal pro domains were poor: only in one case (the pro domain of the porcine CPB) the ODA positive predictive value is above the 50% minimum level.

## *6.5 Surface area of the ODA patches*

The overall average number of ODA residues in a catalytic domain is 16. The maximum ODA patch area is found in the insect HaCPA, covering almost 5000 Å<sup>2</sup> (Fig. 11) and involving 54 ODA residues, where 44 % of them are hydrophobic. This is followed by the other insect CP HzCPB (with an area of 3000 Å<sup>2</sup> and 38 ODA residues, where 50 % of them are hydrophobic). For the catalytic domain of the complete M14A subfamily (which comprises CPs where the N-terminal pro domain structure is present and also those that are in complex with an inhibitor), there are between 8-23 ODA residues (50 and 43 % of them hydrophobic, respectively) at the surface of the protein, forming an ODA patch area which extends from 800 to 2000 Å<sup>2</sup> (Fig. 11). The CPs of the M14B subfamily show intermediate values, with 21 and 28 ODA residues (52 and 54% of them hydrophobic, respectively) that display ODA patch areas of 2100 and 2700 Å<sup>2</sup> for the duck CPD and the human CPM, respectively. Thus, there seems to be an increasing trend in the number of ODA residues/covered surface area from CPA/Bs < CPN/Es < insect CPs.

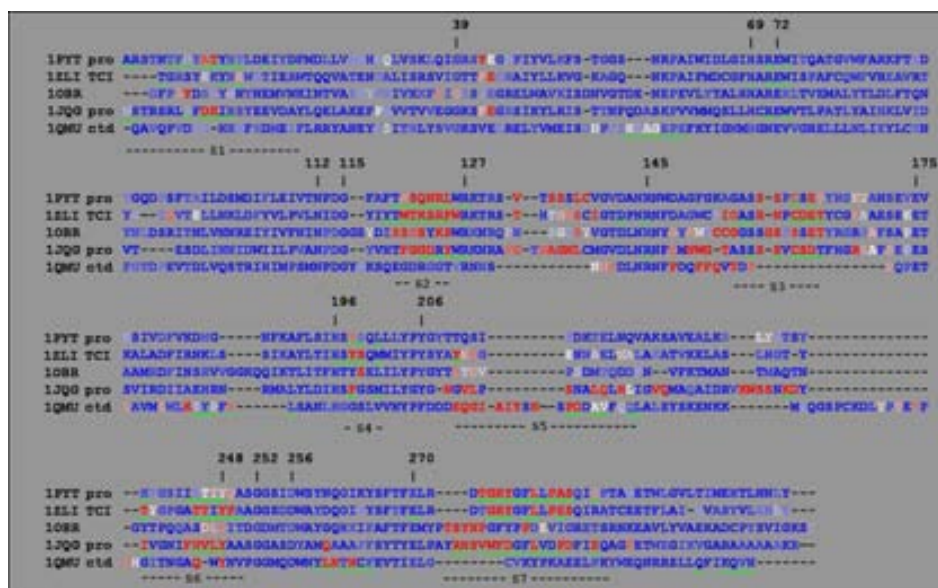
Previous tests indicate that the ODA method is not sensitive to small conformational changes in the binding interfaces. For instance, the ODA results do not significantly vary between the unbound and the complexed states, irrespectively of the conformational changes in the interface side-chains (Fernández-Recio et al., 2005). In addition, when several structures of the same protein were available, the ODA results were practically the same, in spite of the existence of small side-chain conformational changes. Finally, we have built alternative models modifying the side-chain conformations with SCWRL, and the ODA results do not vary when only a few side-chains are modified. The only exception is that, in some cases, a protein structure solved by NMR can have significant conformational changes as compared to the same protein structure

solved by X-ray crystallography, which may result in different ODA values (Fernández-Recio et al., 2005), but it is not the case in this study.

## *6.6 Sequence-structure relationships*

The alignment of the sequences of the CPs used in this study found very few residues fully conserved, and most of them were buried from solvent and therefore not relevant for ODA analysis. Actually, only Tyr248 (bovine CPA1 numbering) is fully conserved at an interface in all CPs, and happens to be predicted by ODA. Within the M14A subfamily, in addition to Tyr248, there are other residues quite conserved: position 163 can be either Glu or Asp, and is at the interface in all members (except in the bCPA/pro structure), and position 198 can be either Tyr or Phe and is at the interface in all members. All of these interface conserved residues in the M14A subfamily are predicted by ODA. However, although the conservation at the individual residue level is not high, our results show that the ODA residues in the catalytic domain are grouped in sequence segments that are quite conserved along the CPs (Fig. 12). The ODA residues identified are involved in nearly all the protein-protein interactions established by the catalytic domain (> 70% in the cases of the N-terminal pro domain, the C-terminal domain and the protein inhibitors). We analyzed the intermolecular interactions established by each of the sequence segments of ODA residues. The first low-energy ODA segment (labelled S1 in Fig. 12) is at the N-terminus of the catalytic domain, and all interactions found here are with the N-terminal pro domain. Probably because of the smaller size of the protein inhibitors, none of their residues are interacting with this region. On the contrary, and in the rest of the ODA-based residue segments, the protein inhibitors seem to interact with similar residues as the N-terminal pro domains. The sequence segment extending for more than 10 residues after Pro205 contains a number of significant ODA residues which do not seem to be involved in known protein-protein interactions, with the exception of a short segment of residues that are at the interface with the C-terminal domain in the members of the M14B subfamily.

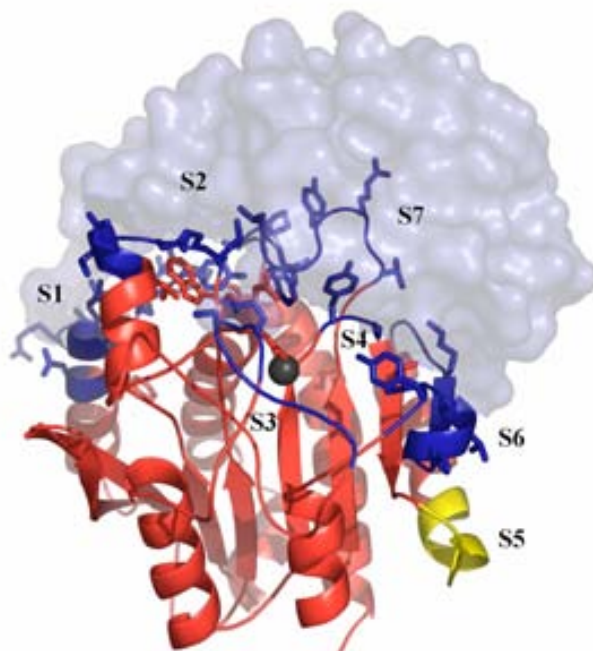
Thus, the 209-221 residue segment (labelled S5 in Fig. 21) was defined as a “putative binding region”.



**Figure 12.** Sequence alignment based on topological equivalence of the catalytic domain of selected representative CPs of the dataset. The numbers label important residues in the structure. The different sequence segments derived from the significant ODA residues are labelled from S1 to S7. The residues underlined in green are located at the known experimental interface (i.e. they have at least one atom within 5Å distance from any atom from the partner molecule). Residue segment number 5 (S5) is a putative binding region. The alignment was made with CLUSTALW ([www.ebi.ac.uk/clustalw](http://www.ebi.ac.uk/clustalw)) and was manually refined to account for experimental information (location of catalytic relevant residues, metal binding residues, and secondary structure elements). Sequence similarity values within A/B CPs (first four entries) ranged between 24% and 46%, whereas the average sequence similarity between A/B CPs and the only N/E CP in the figure (1qmu) was 18%. The alignment has been colored with an in-house script, according to ODA energy values from red (energy value < -10.0 kcal/mol) to blue (energy value = 0.0).

The analysis of the conservation of the ODA regions at the sequence level indicates that they are formed by segments of residues interspersed throughout the whole catalytic domain. However, they generally appear surrounding the active site cleft in the tertiary structure (Fig. 13). What follows is a brief description of the main structural characteristics of the ODA-based predicted binding region (the segments of residues numbered S1-S4,S6,S7 in

Fig. 13) that accounts for nearly all the known protein-protein interactions established by the catalytic domain, and of the features of the ODA-based putative binding region (segment S5 in Fig. 13). Residue segment S1 consists of the first 20 N-terminal residues of the catalytic domain, some of which establish non-covalent interactions to the N-terminal pro domain. Segment S2 contains residues 120-125, all of which are involved in protein-protein interactions either with the N-terminal pro domain or with the protein inhibitors. In the “reference” structure of bovine CPA1 this region spans the end of helix 4 and the first part of the following loop.



**Figure 13.** View of the binding regions in a representative CP, bovine CPA1 (1pyt in PDB). The structure of the CP domain is represented as a cartoon where the different residue segments, numbered according to their location in the sequence (see the text for a complete discussion), are shown in blue. The side-chains of the experimentally known interface residues (those with at least one atom within 5Å distance from any atom of the partner molecule) are shown in blue stick. The ODA-based putative binding region (segment S5) is in yellow. The zinc ion (gray sphere) and the N-terminal pro domain (depicted as a surface) are also shown.

Segment S3, containing residues 154-164, is entirely a loop. It has been described that this loop is one of the most susceptible points of cleavage of the CPs by endoproteases, a fact that might be related to both easy accessibility and regulatory processes. Thus, it has been reported that a specific cleavage of bovine CPA1 by subtilisin at the Ala154-Gly155 bond in this loop promoted catalytic and conformational changes to the catalytic domain (Solomon et al., 1990). Members of the M14A subfamily display different interactions with either the N-terminal pro domain or the protein inhibitor in this region. Remarkably, the residue at position 163 is an acidic one (aspartate/glutamate) in all CPs except for those in the M14B subfamily. In the M14A subfamily, this residue seems to change its conformation depending on the kind of binding partner molecule since it establishes a hydrogen bond to the LCI and TCI protein inhibitors (but not to PCI), although such interaction does not take place with the N-terminal pro domain.

Two residues in segment S4, Tyr198 and Ser199, are involved in four interactions with either the N-terminal pro domain or the protein inhibitors. Binding region 5 runs from residues 234 to 249 across the last part of helix 6, the loop connecting it with  $\beta$ -strand 7,  $\beta$ -strand 7 and a short segment in the next loop. The only strictly conserved residue is Tyr248, which makes interactions with residues of the N-terminal pro domain and the protein inhibitors. It is worth mentioning that the loop containing segment S6 is longer for the members of the M14B subfamily and might be a determinant for substrate or ligand selectivity. Residue segment S7, which includes residues 274-284, is found at the C-terminal part of the catalytic domain. It is essentially a loop containing residues that are to a greater extent involved in hydrophobic interactions with the binding partner molecules.

The residue segment S5 covers the loop between  $\beta$ -strand 6 and helix 6 of the catalytic domain. Neither the N-terminal pro domain nor the protein inhibitors seem to dock to this region. A superimposition with the CPs of the M14B subfamily reveals that this region overlaps with a limited number of residues that are interacting with the C-terminal domain. However, since no C-terminal domain has been described for any CP of the M14A subfamily, this ODA-based putative binding region may be involved in interactions with partners of different nature, yet to be discovered. Again, the loop is longer and

contains more ODA residues in the members of the M14B subfamily. One might hypothesize that this binding region performs an as yet unknown general function. A potential role is related to its susceptibility to aggregation. It is known that the CPs of the M14B subfamily might suffer an aggregation/solubilization transition along the sorting process at the secretory pathway (Song et al., 1995). Besides, binding of a monoclonal antibody recognizing the bovine CPA1 epitope located in the sequence 209-218 inside the loop was shown to protect the enzyme from aggregation (Song et al., 1995; Katzav-Gozansky et al., 1996). Taken together, these observations could suggest that the putative binding region might play a functional and structural role in the folding process of the CPs.

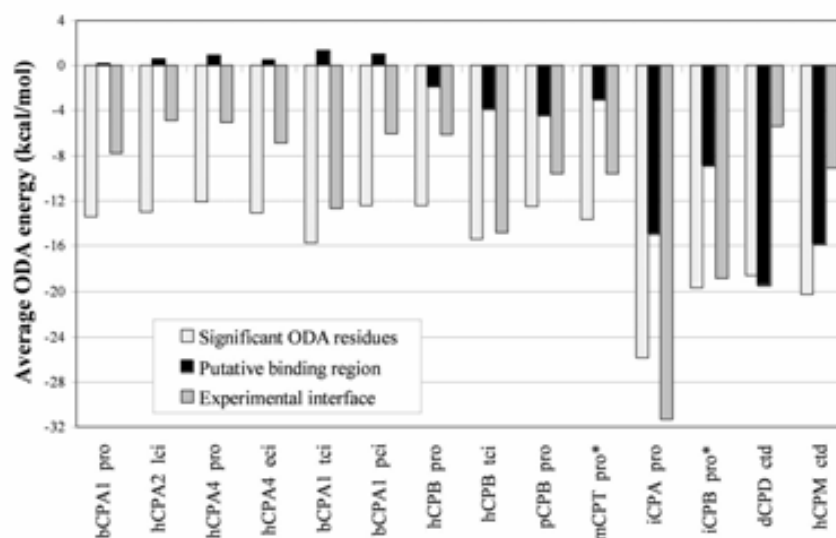
Previous molecular dynamics (MD) simulations of bCPA1 (Makinen et al., 1989), showed pronounced fluctuations in two regions: residues 120-126 (which is a region making interactions with extended oligopeptides), and residues 270-281 (which are lining the active site). Interestingly, these regions showing high mobility in MD calculations are two of our ODA-based predicted binding regions (segments S2 and S6). It would be interesting to include this kind of large protein fluctuations in the ODA analysis, but this is ongoing work that will be the object of a separate large systematic study.

## *6.7 Protein-ligand relationships*

The regions defined by the significant ODA residues, located either at the experimentally known interfaces or in putative binding regions, seem to contribute differently to the interaction with the corresponding partners, according to the particular type of CP (Fig. 14). The average ODA energy of the significant ODA residues (i.e. those with ODA energy < -10 kcal/mol) is around -12.0 kcal/mol for all catalytic domains of the M14A subfamily (all the A&B CPs) with the exception of the insect variants, whereas it shows slightly lower energy values (around -20 kcal/mol) for the CPs of the M14B subfamily (dCPD and hCPM species, here), and reaches a minimum value of -26.0 kcal/mol for the insect HaCPA. These results emphasize that the average energy profile of the



ODA patches is similar among the CPs, in spite of the changes at the primary sequence level. In contrast, the average ODA energy of the residues forming the putative binding region (as previously defined) is above zero for CPA1, CPA2, and CPA4, and reaches values below zero for the rest of MCPs, again being much lower (larger in absolute value) for the two insect CPs and for the members of the M14B subfamily.

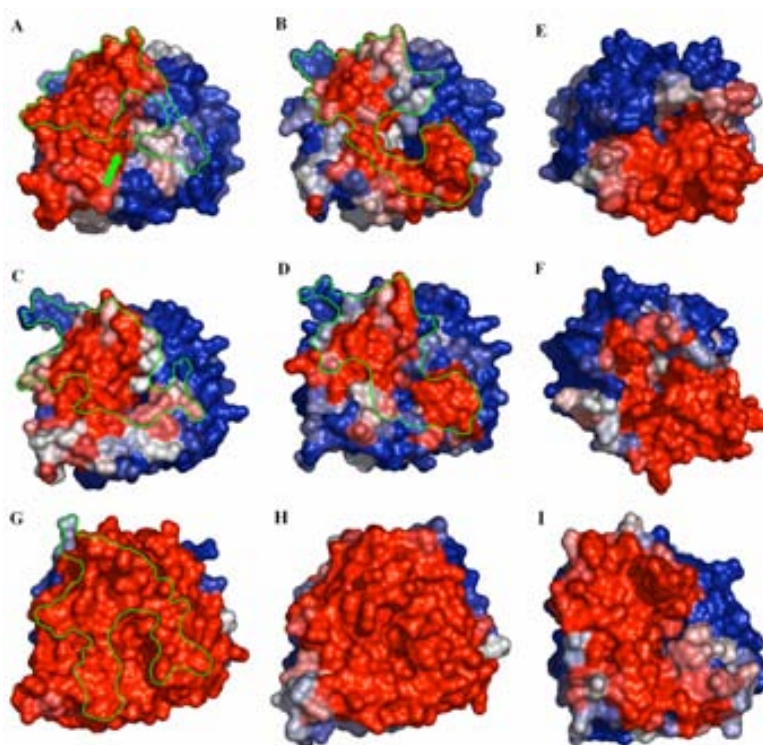


**Figure 14.** Average ODA energy values for different groups of residues of the catalytic domain: significant ODA residues (i.e. those with ODA energy < -10 kcal/mol), residues at the ODA-based putative binding regions, and residues at the experimental interfaces. The complexes of insect CPB and microbial CPT with a pro domain modelled upon structural superimposition are marked with an asterisk (\*).

## 6.8 Regionalization of the ODA patches

The energy-based colored ODA residues for selected CPs is shown in Fig. 15. The most prominent feature among all the CPs is the large ODA area displayed by the insect peptidases. Other differences are evident when comparing the pattern of the ODA patches between the CPs of the M14A (Fig. 15, panels A-D,G-I) and M14B (E,F) subfamilies. Firstly, the surroundings of the entrance to the active site cleft for all CPs of vertebrate origin (Fig. 15 A-F) appear to be covered by regions of ODAs of different energy levels. On the

contrary, for the insect CPs (Fig. 15 G,H) this face of the enzyme is a homogeneous low-energy ODA patch. Secondly, when comparing the ODA patches between the CPs of the M14A (Fig. 15 A-D) and M14B (E,F) subfamilies (with the exception of the insect and bacterial variants), we found that for the CPs of the M14A subfamily there are a number of regions of ODAs of different energy levels whereas those on the M14B CPs tend to be clustered.



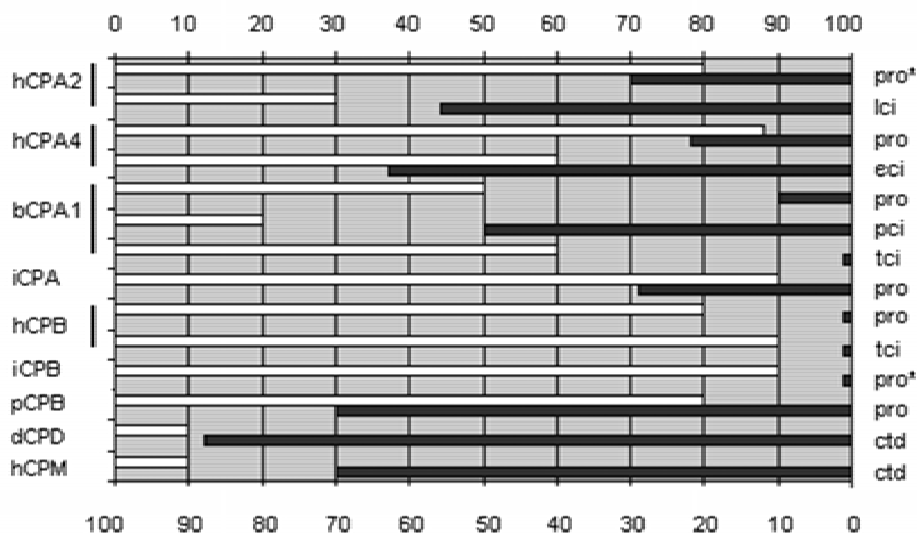
**Figure 15.** Representation of the surfaces of the ODA patches for the CPs. The view is with the active site cleft pointing towards the reader. In *panel A*, the zinc ion is shown in green at the bottom of the active site cleft, which is indicated by a green arrow. The structures have similar orientation, and are grouped into four classes. Upper left side and center, mammalian M14A CPs: A) bCPA1, bovine CPA1, B) pCPB, porcine CPB, C) hCPA4, human CPA4, D) hCPB, human CPB; upper right side, M14B CPs: E) dCPD, domain II of duck CPD, F) hCPM, human CPM; lower left side, insect M14A CPs: G) HaCPA, CPA from *H. armigera*, H) HzCPB, CPB from *H. zea*; lower right side, a bacterial M14A CP: I) TvCPT, CPT from *T. vulgaris*. The color ramp is from red (ODA energy value < -10.0 kcal/mol) to blue (ODA energy value = 0.0). The experimental interface is shown in green contour line for those cases where the structure of the complex is known (A-D,G).

As the experimental evidence is against the inhibition of any of the CPs of the M14B subfamily by the currently known natural protein inhibitors, and in favor that the CPs in this subfamily contain no N-terminal pro domain comparable to the M14A ones, we might hypothesize that the regionalization of ODA patches is an important determinant for the binding of a particular partner molecule. Together with the quantitative analysis in the first part of this section, it seems that the pattern of ODA patches distinguishes between the different subfamilies (e.g. in M14A the average value of ODA residues is -12 kcal/mol, whereas in M14B the average value is -20 kcal/mol; see Fig. 14) and, within them, among the type and origin of the different forms (e.g. within M14A, differences in average ODA values are found between vertebrate, insect, and bacterial variants, in which the average ODA values are around -12, -23 and -14 kcal/mol, respectively). This might be useful for predicting, besides the area of the studied molecule that is likely to be involved in putative interactions with ligands, the type and surface of those yet unknown partner molecules.

## *6.9 Accuracy of the predictions*

We then analyzed the contribution of the ODA residues to the protein-protein interactions in the CPs. The percentage of significant ODA residues that are located at the real interface (i.e. the positive predictive value) is a good indication of the involvement of the low-desolvation surface areas in the interaction. When we calculated this value for both the catalytic domain and the binding partners, we found that at least one, or in some instances both, of the interacting partners contribute to the binding energy with favorable desolvation. Besides, this effect may be related to the kind of binding partner (Fig. 16). In the subfamily M14A there is a large number of ODA residues belonging to the catalytic domain, but only a few of the top ten ODA residues of the N-terminal pro domain, at the interface (with the sole exception of the pro domain of the porcine CPB). The reverse situation was observed for the CPs of the M14B subfamily: only a few ODA residues belonging to the catalytic domain are at the interface with the C-terminal domain; however, the latter domain contributes

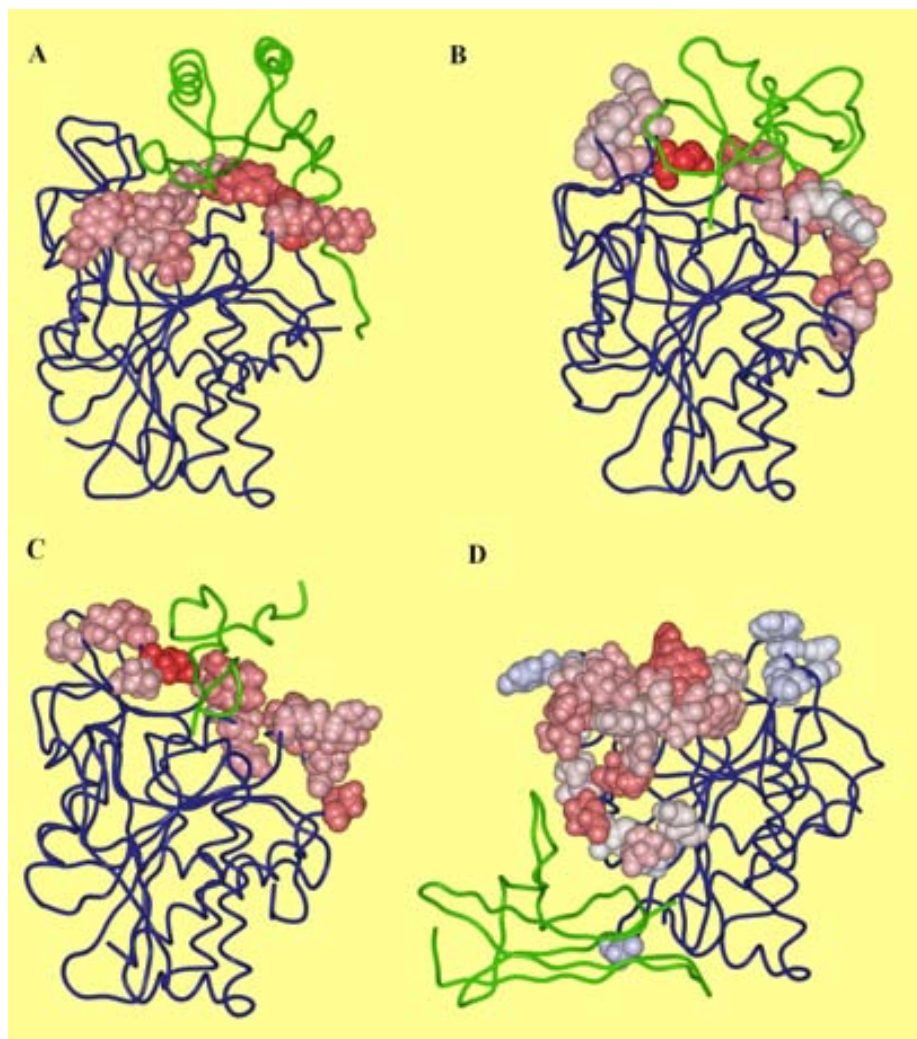
with at least seven out of ten ODA residues to the known interface (see entries dCPD and hCPM in Fig. 16).



**Figure 16.** Accuracy (positive predictive value) of ODAs in CPs on prediction of protein-protein interactions. The white boxes, from left to right, show the ODA positive predictive values for the catalytic domain. The dark grey boxes, from right to left, indicate the ODA positive predictive value levels of the corresponding interacting partners. The complexes of human CPA2 and insect CPB with a pro domain modelled by structural superimposition are marked with an asterisk (\*).

For all the CPs analyzed here it was possible to find at least one ODA hot-spot (i.e. residues with ODA value  $< -10$  kcal/mol) on the protein surface. These ODA hot spots defined a region on the protein surface that, remarkably, was located in or in the vicinity of the known binding site in most of the complexes. In some cases, the catalytic domain is known to be involved in interactions with different proteins using the same binding site, as for human CPB (Fig. 17, panel A), bovine CPA1 (Fig. 17 B), and human CPA4, where the predicted ODA points are correctly located in the interfaces with the N-terminal pro domain and a protein inhibitor. In other cases, the ODAs were correctly located (as compared with other CPs), but only a limited number of ODAs are at the real interface (e.g. with the small-size protein inhibitor PCI; Fig. 17 C) or they are found far from the known site of interaction as in the case of the CPs of the

M14B subfamily, where only a few ODAs are located at the interface with the C-terminal domain (Fig. 17 D).



**Figure 17.** View of ODA residues in the 3D structures of selected CPs. The ODA residues are located in the vicinity of the active site cleft of the catalytic domain (shown in blue) in the 3D structure of the CPs and are involved in protein-protein interactions with different partner molecules (green). The significant ODA residues are depicted as CPK models and colored from red (highest ODA energy value) to white (lowest ODA energy value). The structures are in a similar orientation and correspond to: A) human CPB with the N-terminal pro domain (PDB code 1kwm), B) bovine CPA1 in complex with TCI (1zlh), C) bovine CPA1 in complex with PCI (4cpa), and D) human CPM with the C-terminal domain (1uwy). The figure was made with the program ICM-Browser ([www.molsoft.com](http://www.molsoft.com)).

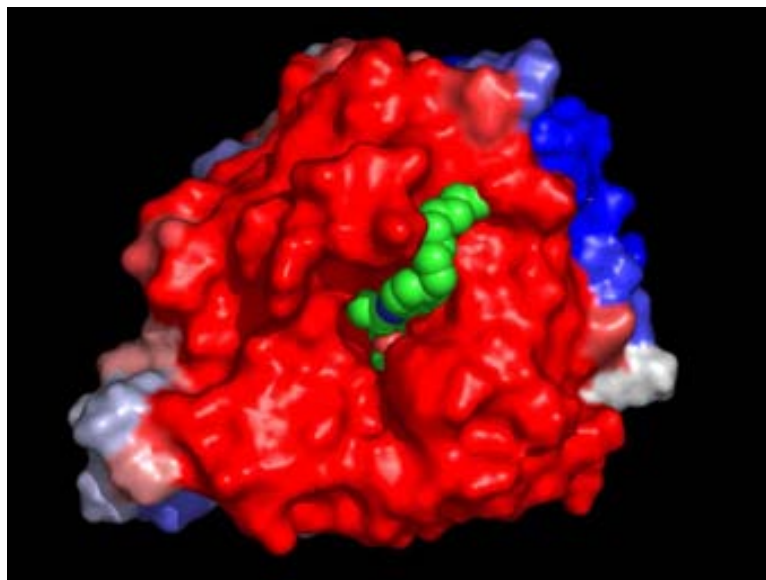
In some cases, an accurate prediction was not possible even though a conspicuous number of ODA residues were found on the protein surface. The complexes between the catalytic domain and the small-size protein inhibitors PCI and LCI are examples in which positive predictive values lower than 50% were obtained. This might be related to the shape or size of the ligand (for the same number of significant ODA residues, fewer of them could be statistically involved in an interaction with a smaller ligand as compared with that of a larger one). In these cases, the contribution to the binding energy provided by the coordination to the zinc ion may outweigh the other available electrostatic contributions. On the other hand, for the CPs of the M14B subfamily, the predicted low-energy docking surface is close to the active site, with only a few ODAs located at the interface with the C-terminal domain, suggesting that the catalytic activity is energetically favored with respect to the mechanical function, i.e. the attachment to the plasma membrane, for certain members. In a broader generalization, a low ODA positive predictive value could just be an artifact of the method, but we cannot disregard the existence of putative docking sites for potential ligands yet to be discovered.

## *6.10 Potential roles of the ODA patches*

The surface area of the residues of the catalytic domain at the experimental interfaces is below 2500 Å<sup>2</sup> (Fig. 11). The average ODA patch area (1700 Å<sup>2</sup>) agrees well with the peak value of area distribution found in a survey of protein-protein complexes of known 3D structure (Janin, 2000). In some instances, the area of the ODA patches parallels the surface area at the experimental interfaces, but there are exceptions, as is the case of the ODA patches in human CPA2 and CPA4, where the surface areas are under 1000 Å<sup>2</sup>, near the lower limit found in the above mentioned survey. On the other hand, the areas of the ODA patches in the CPs of insect origin are above the maximum value of the survey. One possible explanation for this feature is that such an enlarged area of the ODA patches may help the insect to “concentrate” the peptide and protein substrates from the diet in the midgut (depending on the

availability of food or other factors). One form of control mechanism, different from the interaction with the N-terminal pro domain or a protein inhibitor, might also be envisaged. Since the ODA low-energy surface area surrounds the active site cleft, a complex that may regulate the accessibility of the substrates to the protease could be built through the interaction of the catalytic domain with the membrane of the insect midgut (or with another protein receptor).

The ODA method was then applied to analyze protein-ligand interactions, involving small-molecule inhibitors. Some compounds disclosed in Chapter 5 were found as excellent inhibitors of the lepidopteran insect CPB, HzCPB. Although being inhibitors to other CPs as well, they show consistently weaker  $K_i$  values against the vertebrate CPBs. Thus, we used molecular docking and combined it to the ODA analysis. We docked compound **16g**, the *Z-L*-Ala derivative with a 10-C hydrocarbon chain, into the crystal structure of the lepidopteran insect CPB, HzCPB. **16g** is the most potent inhibitor of class DF4 compounds as shown by the  $K_i$  value of 0.76  $\mu\text{M}$  against HzCPB. A superimposition of the docked compound to the ODA-colored protein surface (Fig. 18) revealed the optimal docking of this inhibitor in favorable protein-protein interaction sites. The fact that **16g** is a weak inhibitor for the lepidopteran CPA, in spite of both proteins having equivalent ODA surface areas, may reflect the ability of the compound to better fit into micro-environments along the surface. A comparison of ODA surfaces among HaCPA and HzCPB in Fig. 15, panels G and H, respectively, suggests that the path for the hydrocarbon chain observed in the latter does not exist in the former protein. This might provide an inadequate docking for the long carbon tail, possibly resulting in a drop in inhibitory potency. **16g** shows an almost 20-fold more potent inhibitory profile against the insect versus the porcine CPB. As in the insect enzyme, in pCPB there is a path that the hydrocarbon tail can follow. However, the ODA protein-protein interaction surface areas are clearly dissimilar, as indicated by a comparison between pCPB and HzCPB surfaces (Fig. 15, panels B and H, respectively). The conclusions drawn from this analysis may hold true for a structurally similar compound as well: the arginine derivative **13** (Chapter 5; Fig. 2).



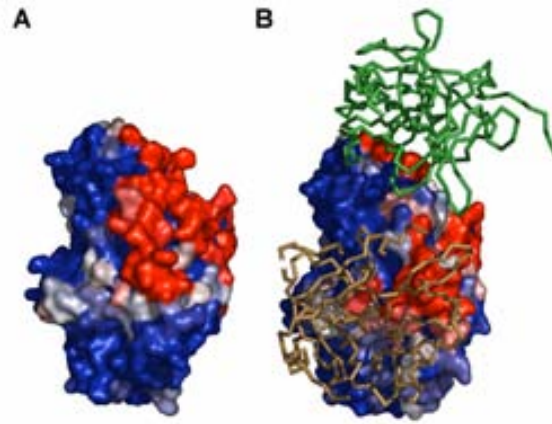
**Figure 18.** Ligand **16g** (CPK model in green) docked into insect CPB, HzCPB, 2c1c. Note the path followed by the hydrocarbon tail of the inhibitor. To aid for comparisons, the surface of the protein is colored according to ODA energy values and in the same orientation as in Fig. 15.

In a different scenario, we have applied the ODA method to analyze binding sites in the crystal structure of human CPA1 (hCPA1) in the context of the zymogen activation process (Pallarès et al., 2008). As commented upon above, the M14 family of CPs were the subject of a structural and functional characterization of binding sites following this method. In the mentioned study it was found that, although the set of CP enzyme domains were structurally very similar, the protein-protein docking surfaces of different sub-classes differed substantially and could be distinguished based on the analysis of ODA residues over the surface of the protein.

Alignment of the sequences of hCPA1 and bCPA1 indicates that, regardless of the conservation of individual residues, grouping of ODA residues in the catalytic domains is very similar in both proteins (Figure 19). All segments of clustered ODA residues in the catalytic domains correspond to







**Figure 20.** Representation of ODA patches for (A) a virtual hPCPA1 constructed using the newly determined hCPA1 structure, 2v77, and the pro domain of the porcine counterpart, and (B) the natural bovine ternary complex built by bPCPA1, PPE (green backbone representation) and CTGC (gold). Same color scale as in Figure 15.

A visual observation of the ODA-colored structures indicates that the virtual hPCPA1 displays a PPE binding surface with ODA energy patches extensive enough to accommodate a PPE subunit, as could be expected from the natural occurrence of the porcine zymogen as a binary complex with PPE (Ventura et al., 1997). On the other side, bPCPA1 is richer in ODA residues at the surface where CTGC sits, where residues contained in the long  $\alpha$ -helix (Val82-Arg99) connecting the domains display the highest ODA energy values (residues 90-99) and are of particular importance in bPCPA1-CTGC binding. Conceivably, as the long  $\alpha$ -helix contains the cleavable arginine residue, species-dependent binding of pro-proteinases to this region and the subsequent formation of oligomeric zymogen complexes modulates the kinetics of active enzyme generation. The computation of ODA energy patches for the known structures of CPs and PCPs thus confirms both the observed occurrence of complexes and the differences in activation behavior between ruminant and mammalian PCPs.

# 7 Crystallization and X-ray diffraction studies

In preceding Chapters it was advanced that the structural analysis of protein-ligand interactions might afford useful information for the design of new generations of ligands. For this purpose, one may resort to computational molecular docking (Krumrine et al., 2003), as we did previously, or use X-ray diffraction crystallography studies to obtain the spatial coordinates of the crystal content. X-ray crystallography is an experimental technique that may afford unique details at atomic level but demands the preparation of the protein samples in a very special way, as crystals. Crystallization experiments were carried out with several proteins and, as outlined in Table 8, the outcome was quite diverse.

Protein	Source	Crystallized	Diffacted	Structure
HaCPA <sup>a</sup>	Recombinant	No	No	No <sup>b</sup>
HzCPB	Recombinant	No	No	No <sup>c</sup>
bCPA	Commercial	Yes	Yes	Yes
pCPB	Recombinant	Yes	Yes	Yes
HaCPO	Recombinant	No	No	No
TcMCP1, TcMCP2	Recombinant	Yes	Yes	No <sup>d</sup>

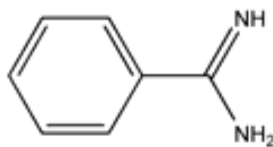
Notes  
<sup>a</sup>bCPA, bovine carboxypeptidase A; pCPB, porcine carboxypeptidase B; HaCPA, carboxypeptidase A from cotton bollworm (*Helicoverpa armigera*); HzCPB, carboxypeptidase B from corn earworm (*Helicoverpa zea*); HaCPO, carboxypeptidase O from cotton bollworm (*H. armigera*); TcMCP1 & TcMCP2, metallo-carboxypeptidases 1 & 2 (family M32 of peptidases) from *Trypanosoma cruzi*  
<sup>b</sup>HaCPA zymogen structure is known, PDB code 1jqg. We used the catalytic domain for screening of crystallization conditions  
<sup>c</sup>HzCPB catalytic domain is known, PDB code 2c1c; however it was crystallized as the zymogen. We used the catalytic domain for screening of crystallization conditions  
<sup>d</sup>refers to complexes with different ligands

CPA and CPB are benchmark metalloproteases, while those from lepidopteran insects and trypanosomes are relatively unknown. Therefore, a

wide range of crystallization conditions were screened, especially for the latter enzymes. Aside from obtaining crystals of model proteases, the other enzymes proved difficult to materialize. We will thus comment further on the results with CPA and CPB, since improvements on the crystallization process of the other enzymes are currently underway. Crystallization experiments were aimed at obtaining structures of these proteins bound to different organic synthetic ligands. It turned out to be a bewildering task in that, in spite of continuous efforts, only CPA was obtained in a complex to a small-molecule inhibitor. In the case of CPB, crystals showing a new packing arrangement were obtained. Thus, we will first describe the CPB crystal structure and then move to the CPA-bound crystal structure.

## 7.1 Structure of a new CPB crystal form

The first crystal structure determination of a metal-dependent proteolytic enzyme zymogen, and only third among different protease classes, corresponds to the porcine pancreatic CPB zymogen (Coll et al., 1991). This crystal structure (PDB code 1nsa) at 2.3 Å belongs to the tetragonal  $P4_3$  space group and shows the unique independent polypeptide chain bound to three copies of a small-molecule ligand, the arginine mimetic benzamidine (**20**).



**20**  
Benzamidine

Our newly determined CPB zymogen structure crystallized in the triclinic  $P1$  space group and diffracted to 1.88 Å. Further crystallographic and refinement data can be found in Table 9.

<b>Table 9 Statistics of data collection and refinement for CPB zymogen</b>	
Parameters	Value
Wavelength used during data collection	0.8123 Å
Unit Cell Constants	a = 49.59 Å, b = 49.97 Å, c = 50.63 Å $\alpha = 109.9^\circ$ , $\beta = 104.5^\circ$ , $\gamma = 96.2^\circ$
Resolution range	19.63 - 1.88 Å
Space group	P1 (1mol / asymmetric unit)
Number of measured reflections	129,122
Number of unique reflections	33,758
$R_{\text{merge}}^a$ (overall/outermost shell)	5.0 / 22.2 %
Completeness (overall/outermost shell)	96.4 / 88.3 %
$I / \sigma I$ (overall/outermost shell)	18.6 / 5.2
Reflections used for refinement	33,081 (total) / 676 (test set)
Crystallographic $R_{\text{factor}}^b$ / $R_{\text{free}}^c$	16.4 / 19.4 %
Deviation from ideality	
r.m.s.d. bond lengths	0.009 Å
r.m.s.d. bond angles	1.05°
Number of protein atoms/total	3,195 / 3,552
$B$ -factor statistics (Å <sup>2</sup> )	
$B$ -factor	17.1 (overall) / 23.5 (Wilson plot)
Catalytic domain, main/side chain	13.5 / 15.0
Pro domain, main/side chain	18.5 / 20.7
Zn <sup>2+</sup> (1 in total)	21.1
Glycerol atoms (30 in total / 5 mols per monomer)	32.4
Solvent atoms (326 in total)	30.7
Protein geometry <sup>d</sup>	
Ramachandran favored	97.5 % (383 of 393 residues)
Ramachandran allowed	99.7 % (392 of 393 residues)
Ramachandran outliers	0.3 % (1 of 393 residues, Ser199)
Residues with bad bonds/angles	0.00 / 0.00 %
Rotamer outliers	2.33 %
Notes	
<sup>a</sup> $R_{\text{merge}} = \sum_{\text{hkl}} \sum_{j=1}^N  I_{\text{hkl}} - I_{\text{hkl}}(j)  / \sum_{\text{hkl}} \sum_{j=1}^N I_{\text{hkl}}(j)$ , where $N$ is the redundancy of the data. The outermost shell is 1.93 - 1.88 Å	
<sup>b</sup> $R_{\text{factor}} = \sum_{\text{hkl}}   F_{\text{obs}}  -  F_{\text{calc}}   / \sum_{\text{hkl}}  F_{\text{obs}} $ , where $F_{\text{obs}}$ and $F_{\text{calc}}$ are the observed and calculated structure factor amplitudes of reflection hkl	
<sup>c</sup> $R_{\text{free}}$ = is equal to $R_{\text{factor}}$ for a randomly selected 2 % subset of reflections that were not used in refinement	
<sup>d</sup> According to Molprobit (Davis et al., 2004)	

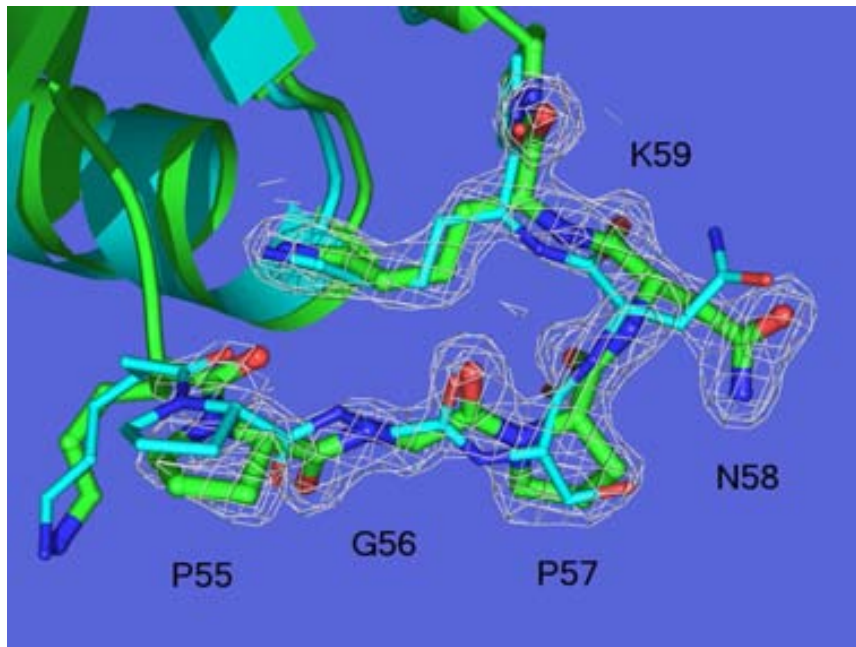
## 7.2 Assignment of residues

An alignment of the sequence retrieved from the earlier determination and the current genomic sequencing information data (Ventura et al., 1999) gave a sequence identity level of only 86%, highlighting the fact that only partial sequence information was available at the time the tetragonal crystal structure was solved (1nsa; Coll et al., 1991). In the present structure, all the residues were assigned according to the sequence information and carefully checked in the electron density maps. The CPB zymogen structure encompasses residues Phe7-Arg95 from the N-terminal pro domain and Thr4-Leu308 from the C-terminal catalytic domain. A sequence alignment of 393 residues in the 1nsa structure revealed 36 discrepancies (Table 10). Some of them might be undetectable in the electron density maps, i.e., Asp/Asn, Gln/Glu, Val/Thr, while others involve larger changes on the shape of the side chains. In the latter case, the electron density maps obtained in our structure allowed us to confirm the sequence data information.

This work/sequence <sup>a</sup>	1nsa	This work/sequence	1nsa
<i>Leu25<sup>b</sup></i>	<i>Glu</i>	Thr154	Val
<i>Arg93</i>	<i>Val</i>	Thr158	Val
<i>Val94</i>	<i>Ser</i>	Asp159	Asn
Lys23	Glu	Asp162	Asn
Glu28	Lys	Asn216	Asp
Thr36	Ser	Asn221	Ser
Leu43	Asp	Ala225	Gly
Pro57	Ser	Thr232	Ser
His76	Gln	Lys237	Ser
Glu85	Asp	Thr239	Ser
Leu88	Arg	Ala244	Ser
Ser94	Ala	Asp260	Asn
Asn101	Asp	Tyr277	Phe
Lys102	Asn	Ile280	Val
Thr136	Ser	Glu291	Gln
Thr137	Ser	Ile297	Val
Ile139	Thr	Val304	Thr
Asp148	Asn	Gly306	Glu

Notes  
<sup>a</sup>Genbank sequence accession code P09955  
<sup>b</sup>Residues in *italics* belong to the pro domain

The residue side chains that deviate most from sequence data information will be commented further. In the pro domain, the basic Arg93 had been interpreted as a valine in the tetragonal crystal structure. The inspection of a list of intermolecular interactions indicated that Arg93 side chain resides in an open space in such a way that no interactions are observed for the guanidinium N atoms. Leu25 had been assigned as a glutamate in the previous structure. However, the environment surrounding this residue is hydrophobic and less well suited for an acidic side chain. A number of residues belonging to the catalytic C-terminal domain were misassigned in 1nsa. Leu43, Pro57, Leu88, Ile139, and Ala244 all reside in hydrophobic environments which is against the presence of polar atoms as have been assigned before. Other residues, like Asn221, Lys237 and Tyr277 lie in less packed microenvironments. Notably, Tyr277 phenolic OH is only at 3.9 Å from a water molecule (it was assigned as a phenylalanine in 1nsa). As exemplified in Fig. 21, the current CPB structure at high resolution provides now direct evidence on the position and interactions of the corresponding side chains.



**Figure 21.** Partial structure of CPB zymogen in the triclinic (green carbons; this work) and tetragonal (cyan; 1nsa) forms. The region surrounding Pro57 is shown as sticks. The electron density (wire mesh) is contoured at a 1.5  $\sigma$  level.

Regarding the identity of residues that mediate in packing intermolecular interactions, there is no side chain other than Glu85 (originally assigned as an aspartate) in conflict with the sequence data information. This implies a structural stabilization effect that solvent exposed side chains may achieve when involved in packing intermolecular interactions. As the molecular packing in the triclinic and tetragonal crystals is completely shifted, we analyzed in depth the intermolecular contacts in both systems, as will be discussed next.

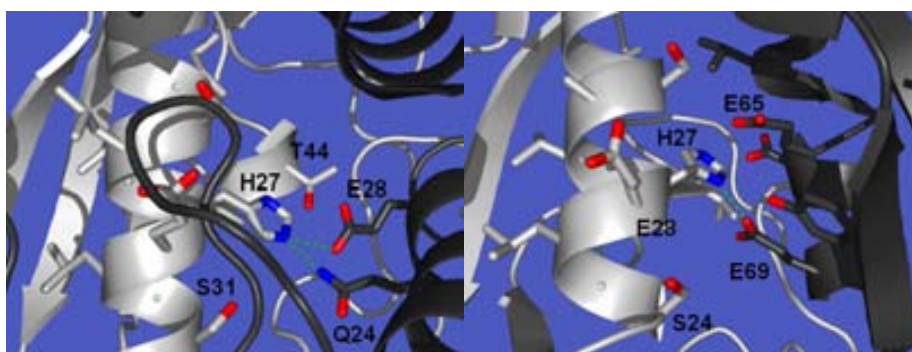
### *7.3 Comparison to the known tetragonal crystal*

The CPB zymogen structure presented herein harbors one independent polypeptide chain per asymmetric unit and unit cell constants  $\mathbf{a}=49.59 \text{ \AA}$ ,  $\mathbf{b}=49.97 \text{ \AA}$ ,  $\mathbf{c}=50.63 \text{ \AA}$ ,  $\alpha=109.9^\circ$ ,  $\beta=104.5^\circ$ ,  $\gamma=96.2^\circ$ . A comparison to the tetragonal  $P4_3$  unit cell (constants  $\mathbf{a}=\mathbf{b}=103.1 \text{ \AA}$ ,  $\mathbf{c}=46.6 \text{ \AA}$ ,  $\alpha=\beta=\gamma=90.0^\circ$ ) indicates that two cell dimensions are halved ( $\mathbf{a}/2$  and  $\mathbf{b}/2$ ) in the P1 form while the third one is conserved. The triclinic unit cell volume is roughly a quarter of the tetragonal one, indicating the operation of the four-fold axis. The orientation between interacting polypeptide chains is different in both crystal forms, as observed by generation of symmetry-related mates from superimposed parent molecules. This fact might be related to shifted intermolecular contacts established between neighbouring molecules (Table 11; next page). Only a few residues with conserved contact interactions could be detected, and will be briefly mentioned.

In the pro domain, His27 maintains strong hydrogen bonds (donor to acceptor distance around  $2.5 \text{ \AA}$ ) to neighbouring atoms. Interestingly, the acceptor atoms belong to different domains of the enzyme, residues Gln24 and Glu28 from the catalytic domain in the triclinic crystal, or Glu69 from a symmetry-related pro domain in the tetragonal crystal (Fig. 22, left and right panel, respectively). This is one example of different interfacial contacts across both crystals.

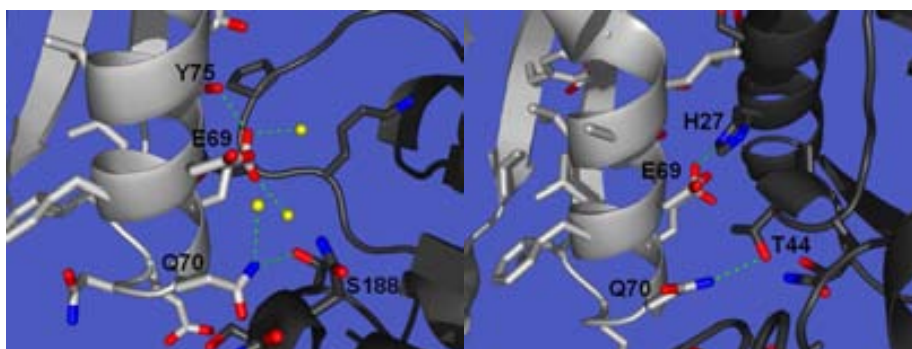


Table 11 List of residues participating in intermolecular contacts <sup>a</sup>	
Triclinic CPB	Tetragonal CPB
<i>Glu20<sup>b</sup></i>	<i>His27*</i>
<i>Asn21</i>	<i>Pro48</i>
<i>His27<sup>*c</sup></i>	<i>Glu69*</i>
<i>Glu28</i>	<i>Gln70*</i>
<i>Thr32</i>	<i>Glu72</i>
<i>Glu69*</i>	
<i>Gln70*</i>	
<i>Leu73</i>	
<i>Tyr75</i>	
Thr4	Tyr92
Ser8	Arg93
Gln24	Thr158
Asn58	Asp159
Arg84	Lys177
Glu85	Asp181
Asn133	Arg184
Tyr248*	Ser188
His307	Tyr248*
Notes	
<sup>a</sup> Contact distance cutoff 3.0 Å	
<sup>b</sup> Residues belonging to the pro domain are in <i>italics</i>	
<sup>c</sup> An asterisk (*) marks residues with conserved contact interactions	



**Figure 22.** Partial packing diagram showing His27 (pro domain) interactions. (*Left panel*), in the triclinic cell; (*right panel*), in the tetragonal cell. Symmetry-related residues within 5 Å from His27 are shown with white or grey carbons (symmetry-related chain) in sticks model. Green dashed lines indicate hydrogen bonds while polar atoms are depicted in red (oxygen) and blue (nitrogen). Only selected side chains are labelled.

Two consecutive residues, Glu69 and Gln70, sustain a number of intermolecular contacts with different acceptors depending on the crystal structure, making them the major contributors for the binding of the pro domain in both crystal forms (Fig. 23). In the tetragonal cell, these interactions result in the arrangement of symmetry-related pro domains while in the triclinic cell the arrangement is with a catalytic domain.

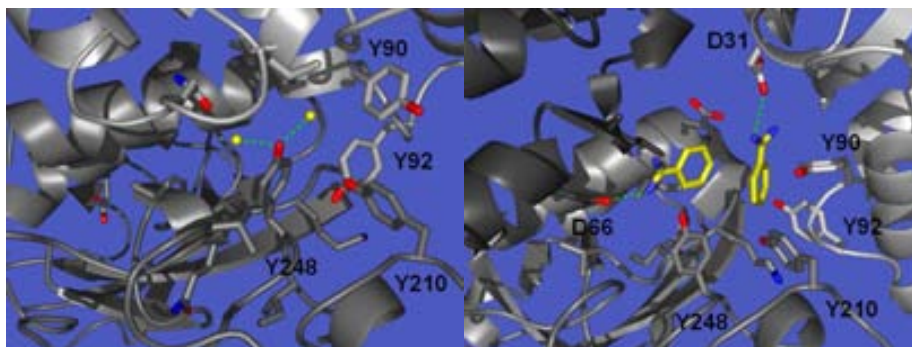


**Figure 23.** Partial packing diagram showing Glu69/Gln70 (pro domain) interactions. (*Left panel*), in the triclinic cell; (*right panel*), in the tetragonal cell. Water molecules are represented as yellow spheres. Other details as in the legend to Fig. 22.

## 7.4 The aromatic cluster

In the earlier CPB zymogen structure determination, benzamidine was generously added during protein preparation and crystallization to avoid activation by proteolytic cleavage. As a consequence, three independent benzamidine molecules are present in the tetragonal CPB zymogen structure. As can be seen from Fig. 24, a second benzamidine molecule, generated by symmetry, locates nearby Tyr248. A comparison to our triclinic CPB structure revealed interesting features of the most widely investigated residue in CPs. In the triclinic unit cell (Fig. 24; left panel), the phenolic hydroxyl oxygen engages two solvent water molecules while the tetragonal Tyr248 is bound to one benzamidine molecule in a dry micro-environment (Fig. 24; right panel). These

two ligand molecules become thus the core of an aromatic cluster with several neighboring symmetry-related aromatic side-chains.



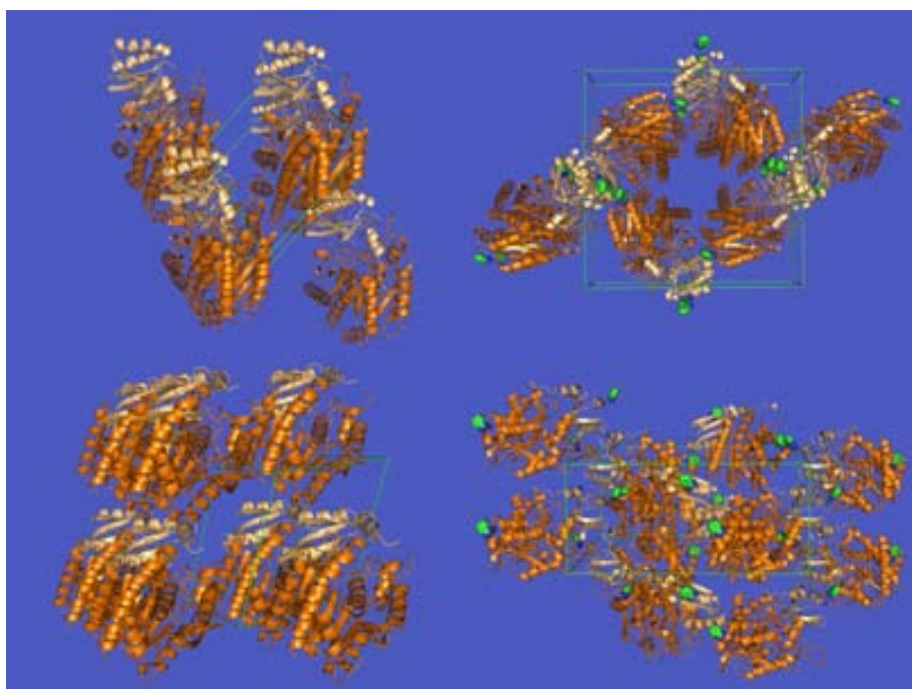
**Figure 24.** Packing diagram showing Tyr248 (catalytic domain) interactions. (*Left panel*), in the triclinic cell; (*right panel*), in the tetragonal cell. Benzamidine molecules are in sticks model with yellow carbons. Other details as in the legend to Fig. 22.

The analysis of the intermolecular interactions made up by the cluster of benzamidine-contacting residues show that water molecules are displaced in favor to interactions with larger ligands (Table 12).

<b>Table 12</b> <b>Intermolecular contacts made up by benzamidine-bound residues<sup>a</sup></b>	
Benzamidine-contacting residues	Interactions by the equivalent residue in the triclinic CPB crystal
<i>Glu20<sup>b</sup></i>	Glu85/Wat158
<i>Pro48</i>	-
<i>Leu62</i>	-
<i>Ala63</i>	Wat323
<i>Asp66</i>	Wat215/Wat251/Wat321
Asp31	Wat144
Tyr90	-
Gly91	-
Tyr92	Tyr210
Tyr210	Tyr92
Tyr248	Wat112/Wat189
Notes	
<sup>a</sup> Contact distance cutoff 3.0 Å; (-) no contact observed	
<sup>b</sup> Residues in italics belong to the pro domain	

One of the benzamidine molecules settled to the S1' specificity pocket, forming a salt bridge to Asp255, thus mimicking an arginine side chain (this is the small molecule found nearby GEMSA in the superimposition shown in Fig.

26; see next Section). The other two benzamidine molecules are placed over the surface, at opposite ends of the longest protein axis, i.e. one residing by the pro domain while the other is located near the catalytic domain. This arrangement allows each benzamidine molecule to interact with symmetry-related protein copies, in such a way that two of them form the core of a cluster of residues belonging to three protein molecules. In the present work, no such small-molecule was added to prevent proteolysis and the molecular packing was shifted (Fig. 25).



**Figure 25.** Packing diagram of CPB zymogen. *Upper row:* left panel, the triclinic cell; right panel, the tetragonal cell. The view is with the shortest unit cell vector pointing towards the reader. *Lower row:* the same, but the view is horizontally rotated 90° respect the upper figures. The pro domain is shown in beige, while the catalytic domain is in orange. The benzamidine molecules in the tetragonal cell (green carbons) are depicted as CPK model.

In the triclinic crystal, the CPB zymogen molecules show an alternate pattern of pro domains and catalytic domains (Fig. 25, upper row, left panel). At variance, the molecular packing in the tetragonal cell is mediated by stacking of

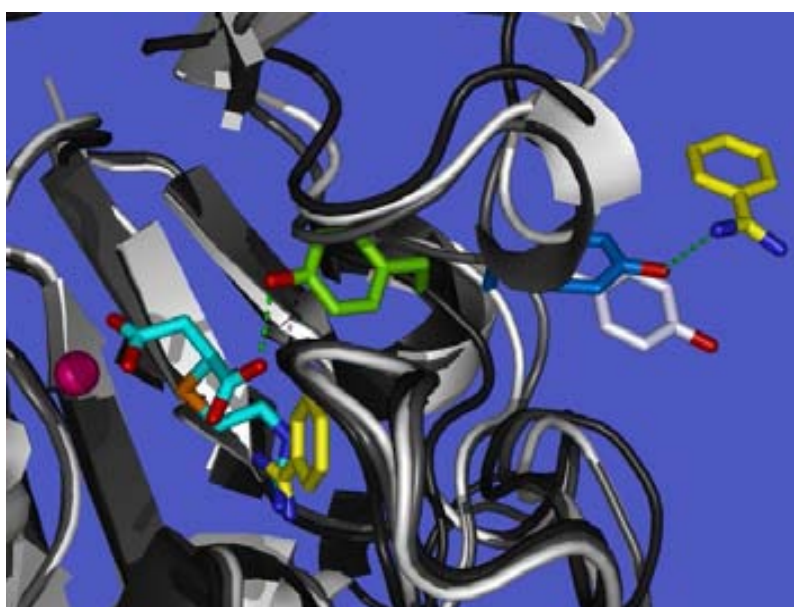
protein molecules along the shortest unit cell axis (Fig. 25, upper row, right panel) in such a way that columns of pro domains and catalytic domains are formed and the interface between adjacent polypeptide chains is completed with benzamidine molecules. The polar part of the neighbouring benzamidine molecules are engaged to a number of hydrogen bonds (with Asp31 and Tyr248 from the catalytic domain: 3.15 and 2.98 Å, respectively; with Asp66 from the pro domain: 2.84 Å) while the aromatic rings reside in a hydrophobic environment sustained by several aromatic and aliphatic residues. The benzyl rings of the core benzamidine molecules are separated by about 4 Å. Some tyrosyl residues seem to exchange partners upon formation of the benzamidine core. In the triclinic crystal Tyr92 and Tyr210, from the catalytic domain, are mutually engaged and form a hydrogen bond through their phenolic hydroxyl group, although the contact distance (3.2 Å) may reflect a somewhat weak interaction. These tyrosyl residues adopt different conformations in the tetragonal crystal, where the Tyr92 phenolic hydroxyl makes a strong hydrogen bond to the main-chain oxygen of Pro48 from the pro domain (contact distance 2.6 Å) while Tyr210 is in the proximity of one core benzamidine molecule. As a result, in the tetragonal crystal, the phenolic hydroxyl groups become separated by more than 8 Å. By this means, large residues exposed to the bulk solvent seem to change their conformation in order to accommodate an organic small-molecule ligand with distinct physicochemical characteristics. As a result, a new micro-environment is created, whereby the organic ligand molecules may be trapped and bonded, thus forming a stable core. The formation of this core would favor neighbouring molecules to pack and hence crystallize in a particular way.

## *7.5 The role of Tyr248 and its flanking region*

There are no conserved intermolecular contacts for residues belonging to the catalytic domain, with the sole exception of Tyr248. This is a strictly conserved residue for all the CP family members. The precise role of Tyr248 has been a matter of much debate, although the latest evidence proposed it to

be essential (Cho et al., 2001). Although partial details have become clearer, much of the action of Tyr248 during catalysis is barely understood. A two-state model seems to be operative comprising an open as well as a closed conformation, where Tyr248 is capping the entrance of the active cleft.

When we extended our analysis to a third CPB, which has Tyr248 in the closed conformation, further details emerged by comparison of the three superimposed crystal structures (Fig. 26).



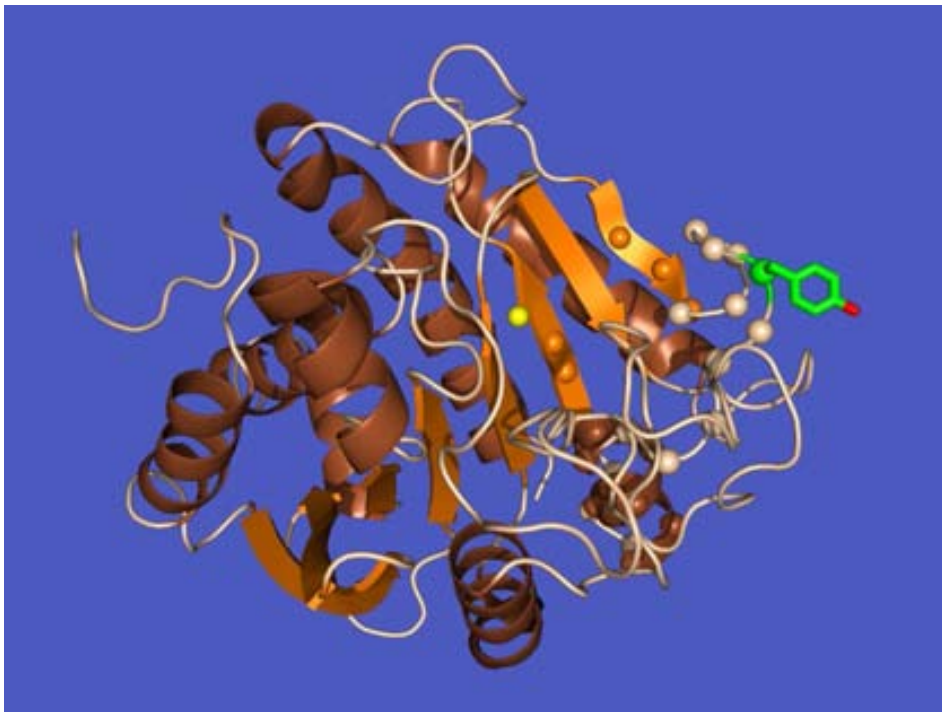
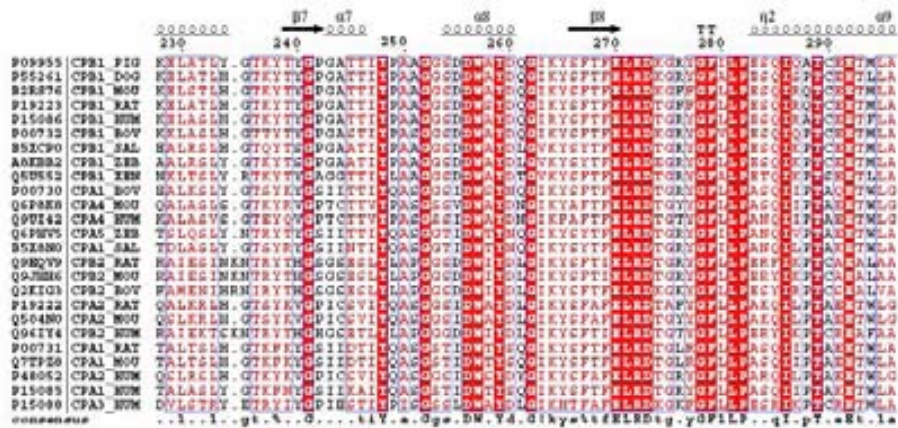
**Figure 26.** The cartoons represent: grey, unbound CPB, this work; white, benzamidine-bound CPB (1nsa); black, GEMSA-bound plasma CPB (3d67). Tyr248 side-chain, benzamidine (yellow), and GEMSA (cyan) are shown in sticks model. The catalytic zinc ion (far left) is a sphere in magenta. Tyr248 color codes: white (unbound CPB), blue (benzamidine-bound), green (GEMSA-bound CPB). Hydrogen bond contacts are marked by green dashed lines.

Triclinic CPB, determined in this work, shows the loop supporting Tyr248 fully exposed to the external solvent (Fig. 24, left panel; and Fig. 26, Tyr248 in white). In the tetragonal CPB it is instead in contact to the benzamidine molecule (Fig. 26, Tyr248 in blue). Now, we can see the Tyr248 loop in a different conformation, slightly displaced toward the entrance of the active site cleft. It might represent an intermediate step between the unbound and the bound state. In the latter, the phenolic ring points toward the catalytic zinc ion

and the hydroxyl group is hydrogen bonded to the carboxylate of GEMSA (Fig. 26, Tyr248 in green). In this structure the loop holding Tyr248 shifted its conformation again, closer to the catalytic site.

The structural superimposition of three CPBs suggests that Tyr248 would left solvent water to contact to a larger ligand, here benzamidine. Once bound, the Tyr248 aromatic ring undergoes a 120° twist to escort the attached ligand toward the catalytic site. Then, in the enzyme-bound complex, the phenolic hydroxyl holds up in the down position securing the C-terminal carboxylate from the ligand, until the catalytic reaction is over. We may thus conclude that Tyr248 hydroxyl group would be the attachment point of the ligand from the early events to the end of the enzymic reaction.

The movement of the loop holding Tyr248 prompted us to combine crystallographic and bioinformatic analyses which finally gave us a new picture of the possible Tyr248 function. First, to detect rigid regions or subdomains in the newly determined CPB crystal structure, a normal mode analysis was performed using HingeProt (Emekli et al., 2008). Two lowest modes with hinge residues 109, 185, 202, 220, 239 and 268 (mode 1), defining four rigid parts, and 14, 64, 82, and 192 (mode 2) yielding three rigid parts were predicted. According to this model, the enzyme elongates/shrinks in such a way that regions shaping the entrance to the active site cleft produce a mouth-like movement. Interestingly, one of the subdomains predicted by mode 2 is almost coincident with previous results obtained from CPA (Liebman et al., 1985; Pritchard et al., 2003). Besides, this subdomain harbors the most mobile side chain in the structure, Tyr248, whose flexibility in CPA has already been investigated (Firth-Clark et al., 2008; Zen et al., 2008). From our observations, hinge residues 239 and 268 in CPB would be the boundaries of a rigid region linked to the catalytic mechanism which contains the mobile Tyr248 side chain. The 239-268 region encompasses  $\beta$ -strand 7, the short  $\alpha$ -helix 7, the unstructured sequence 248-253,  $\alpha$ -helix 8, and part of the  $\beta$ -strand 8 (see Fig. 27). Sequence alignments indicate that this 29 AA sequence is uniquely present in M14 peptidases, with highest scores to CPB and also to various CPA isoforms, suggesting that this region of the carboxypeptidase molecule evolved to perform a function aside from merely holding the Tyr248 aromatic ring in place.

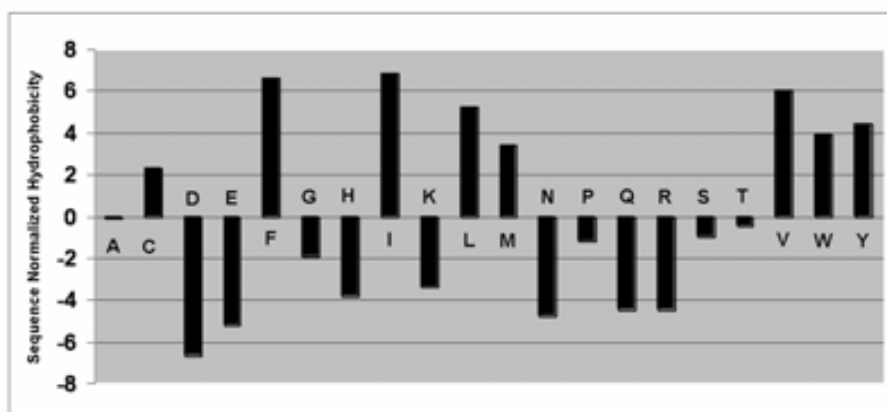


**Figure 27.** *Up*, sequence alignment of residues spanning the rigid subdomain in members of the M14 protease family. *Down*, cartoon representation of trimeric CPB structure highlighting the region limited by hinge residues 239 and 268 that contains Tyr248. The position of the C $\alpha$  atoms are indicated by a sphere and the side chain of Tyr248 is shown in stick model. The catalytic zinc ion is shown as a yellow sphere.

Next, the physicochemical nature of the wild type sequence and its mutants was computationally investigated using a well known amino acid aggregation propensity scale. As implemented in the AGGRESCAN software,



such a scale assigns high aggregation propensities to hydrophobic amino acids such as Ile, Phe, Val, Leu, Trp, Tyr and Val and low aggregation propensities to Asp, Glu, Asn, Arg, Gln, His and Lys (Sánchez de Groot et al., 2005). Considering that peptidases bind to peptide substrates during their normal mechanism of action, the aggregation propensity scale used here, regarded in the broader sense of protein-protein interactions, may be a valid measure of the ligand-protease association process. Figure 28 shows the predicted hydrophobicity value for the 239-268 sequences depending on the nature of the side chain at position 248. This single change suffices to introduce a dramatic shift in the overall properties of the sequence. Phe, Ile, Leu and Val confer the highest hydrophobicity to the whole sequence. The wild type Tyr248 sequence scores in a second subset, together with Trp and Met, which differs widely from the more hydrophilic sequences created by the polar side chains as well as by Ala, Gly and Thr.



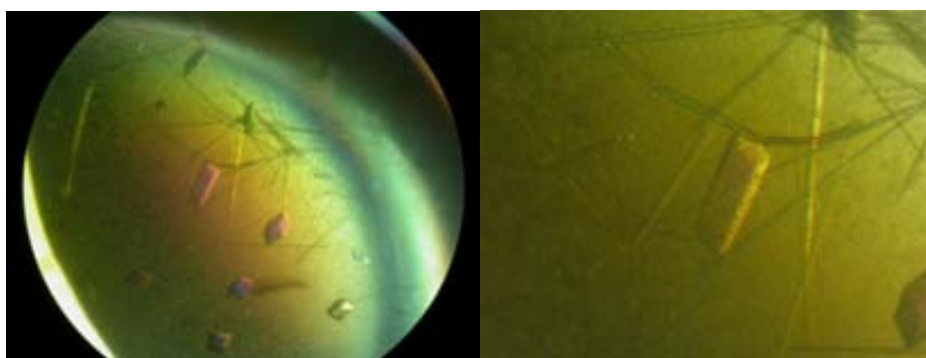
**Figure 28.** Sequence normalized hydrophobicity for the 239-268 sequence of CPB as computed by AGGRESCAN. The overall hydrophobicity value for each one of 20 mutants at position 248 was estimated from the experimental aggregation propensity calculated using a 5-residue sliding window. The data shown are absolute, normalized values that result from subtraction of the overall average value from the normalized hydrophobicity of each mutant.

Kinetic data using peptide substrates showed that the  $k_{cat}/K_M$  ratios are reduced about 50 and 300 fold for the Tyr248Phe and Tyr248Ala mutations, respectively (Cho et al., 2001), indicating the preference for a tyrosyl residue at

the entrance of the active site cleft. Aside from considerations about the involvement of the hydroxyl group in catalysis (see previous section), we might propose that part of the effects of the mutations over the enzymic activity may arise from ligand recognition and binding. The presence of Tyr at position 248 inside the sequence 239-268 suggests a choice over the more hydrophobic Phe-containing sequence, to which a peptide ligand may be more prone to stick and remain bound, and the less hydrophobic Ala-containing sequence which would be less suited to desolvate the peptide substrate. As a consequence, for Ala or Phe substitutions, the overall physicochemical nature of the sequence would impair the catalytic efficiency due to a poorer or stronger desolvation effect as compared to Tyr.

## 7.6 CPA crystallization

CPA has been a matter of intense research since its discovery in the second decade of the last century. Although a published procedure for producing CPA crystals using dialysis buttons is known (Kim & Lipscomb, 1991) we investigated further to obtain crystals of adequate size for X-ray diffraction. Several crystallization trials using CPA gave different outcomes, including X-ray diffraction quality crystals, as shown in Fig. 29. These crystals were obtained in the sitting drop setting from a new crystallization condition. They were used for soaking with different organic synthetic ligands.



**Figure 29.** *Left*, prism-shaped CPA crystals. *Right*, a magnified view.

It turned out from X-ray diffraction analysis that in most cases CPA is unbound. One important exception will be discussed next.

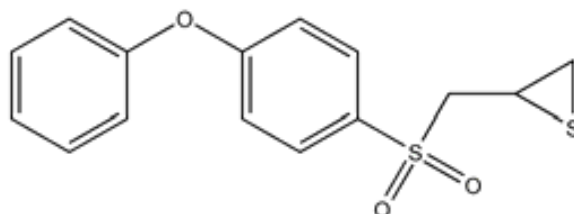
## *7.7 The complex of CPA and a mechanism-based inhibitor*

This part of the work involving mechanism-based inactivators was only made possible thanks to a collaboration to Prof. Shahriar Mobashery (Department of Chemistry and Biochemistry, University of Notre Dame, IN, USA). This allowed us to obtain by chemical synthesis one CPA mechanism-based inactivator which was then subjected to X-ray diffraction analysis.

A mechanism-based inactivator (“suicide substrate” or “ $k_{\text{cat}}$  inactivator”) is an innocuous agent that is converted from the latent to the active state only inside the target enzyme (Walpole & Wigglesworth, 1989). Once bound, the latent compound is converted into a powerful inhibitor thus avoiding contact to off-targets. The mechanism-based inactivators bear a reactive group that is chemically rearranged upon enzymic attack. In this way, the target macromolecule becomes irreversibly inhibited as the compound cannot be detached from the enzyme by dialysis. Compounds bearing different kinds of reactive groups have been assayed as CPA mechanism-based inactivators (Breckenridge & Suckling, 1986; Ner et al., 1987; Mobashery et al., 1990; Ghosh et al., 1991; Kemp et al., 1993; Husbands et al., 1994; Kim et al., 1994; Tanaka et al., 1994; Lee & Kim, 1996; Kim & Lee, 1997; Kim et al., 1998; Han et al., 2000; Chung et al., 2001; Park et al., 2001).

One kind of reactive group comprises the thiirane-based compounds. A thiirane ring is a three-membered heterocycle composed by two carbons and one sulfur atom, which may revert into a thiol group upon enzymic breakdown. An example of this may be SB-3CT, **21**, which is a selective inhibitor of gelatinases MMP2 and MMP9 (family M10 of proteases). The oxirane counterpart to **21** was found to be only a weak MMP2 inhibitory profile (three orders of magnitude less potent than the thiirane), supportive of the notion that

the thiirane ring is more reactive than the oxirane (Brown et al., 2000; Kleifeld et al., 2001; Ikejiri et al., 2005).

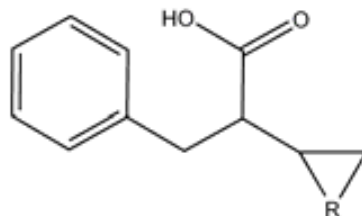


**21**

SB-3CT, 2-((4-phenoxyphenyl)sulfonyl)methylthiirane  
 $K_i$  MMP2 = 13.9 nM

Chemotherapeutic compounds that may be used as drugs or imaging agents targeted to M14 proteases are highly sought, and the complexity among CPs calls for the development of selective compounds. Mechanism-based inactivators are of special interest since their high selectivity for the target enzyme and long-lasting duration of action (Kim & Mobashery, 2001; Kim, 2004). Based on the notion that a covalently bound mechanism-based CP inactivator might be an ideal starting point for such an endeavour, a screening of a collection of thiirane-based MMP ligands (about 500 compounds) against representative M14 proteases was performed. Unfortunately, these molecules did not prove efficacious in inhibition of these metalloproteases. In an effort to elucidate the minimal structural motifs for inhibition of carboxypeptidases by thiirane-based inhibitors, we turned to two examples from the literature.

Compound **22b** (2-benzyl-3,4-epithiobutanoic acid), as well as its oxirane version, **22a** (2-benzyl)-3,4-epoxybutanoic acid), have been reported as mechanism-based inactivators of CPA (Kim & Chung, 1995; Lee et al., 1995; Kim & Kim, 1996; Lee & Kim, 2002). An earlier structural determination of CPA in complex with oxirane **22a** (Yun et al., 1992, Ryu et al., 1997) has been reported, although the atomic coordinates have not been deposited. Kinetic results showed that thiirane **22b** and oxirane **22a** displayed a similar inhibitory potency as CPA inactivators, contrarily to the example of gelatinase inhibition by **21**. Compounds **22** had to be synthesized *de novo* since they are no longer available either from academic or commercial sources.



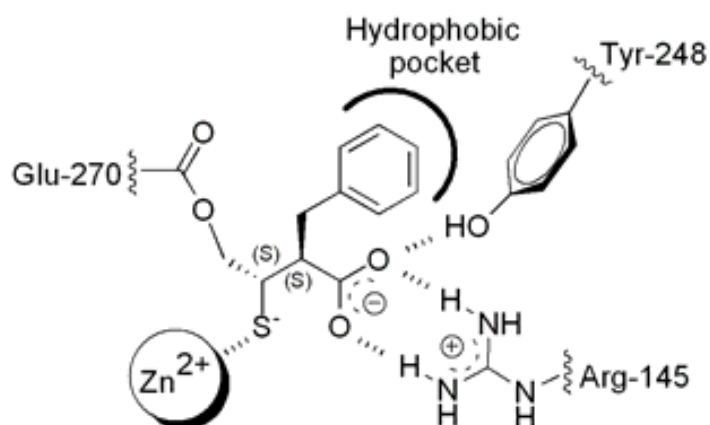
**22a**

R = O, 2-benzyl)-3,4-epoxybutanoic acid  
 $k_{\text{inact}}$  CPA 86  $\mu\text{M}$  (2*S*, 3*R*-**22a**) & 155  $\mu\text{M}$  (2*R*, 3*S*-**22a**)

**22b**

R = S, 2-benzyl-3,4-epithiobutanoic acid  
 $k_{\text{inact}}$  CPA 94  $\mu\text{M}$  (2*R*, 3*R*-**22b** & 2*S*, 3*S*-**22b** racemic mixture)

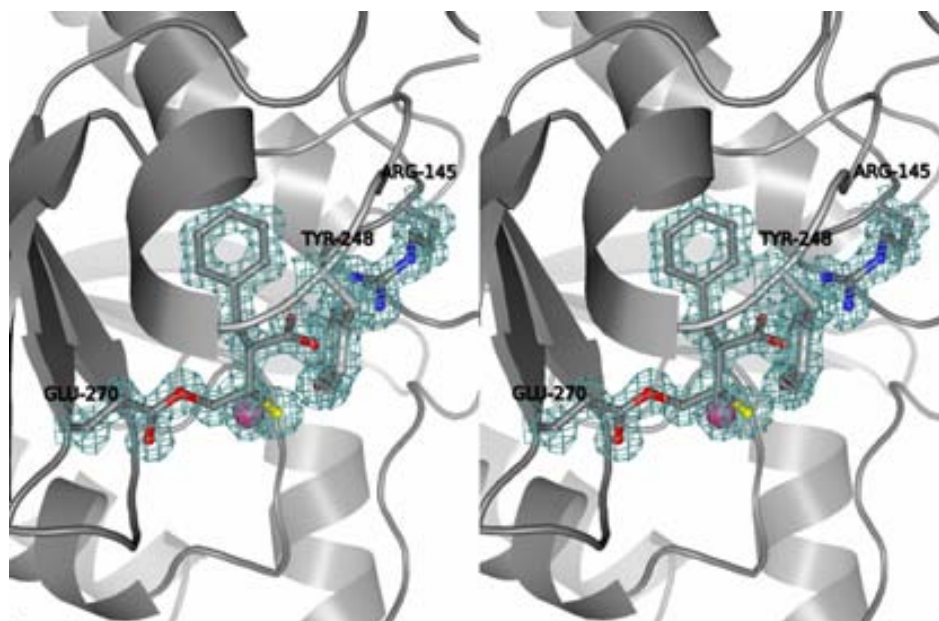
Soaking of CPA crystals with thiirane **22b** allowed us to form the complex. The crystals diffracted to near atomic resolution (1.38Å), and that of highest resolution was fully refined and is presented herein. A pictorial interpretation of the binding of thiirane **22b** is given in Fig. 30. Crystallographic and refinement details are given in Table 13 (next page) while a plot of the crystal structure of CPA in complex to **22b** is shown in Fig. 31 (page 81).



**Figure 30.** A schematic view of interactions between inhibitor **22b** and CPA.

<b>Table 13</b>	
<b>Statistics of data collection and refinement for thiirane-bound CPA</b>	
Parameters	Value
Wavelength used during data collection	0.8123 Å
Unit Cell Constants	a = 42.32 Å, b = 57.50 Å, c = 57.08 Å $\alpha = 90.0^\circ$ , $\beta = 99.1^\circ$ , $\gamma = 90.0^\circ$
Resolution range	31.27 - 1.38 Å
Space group	P2 <sub>1</sub> (1mol / asymmetric unit)
Number of reflections	488,546 (measured) / 54,345 (unique)
$R_{\text{merge}}^a$	6.8 (overall) / 37.9 % (outermost shell)
Completeness & multiplicity (overall/outermost shell)	96.6 / 77.8 % 6.1 / 5.3
I / $\sigma$ I (overall/outermost shell)	8.8 / 1.8
Reflections used for refinement	52,590 (total) / 1,102 (test set)
Crystallographic $R_{\text{factor}}^b$ / $R_{\text{free}}^c$	14.3 / 15.7 %
Deviation from ideality	
r.m.s.d. bond lengths	0.008 Å
r.m.s.d. bond angles	1.14°
Number of protein atoms/total atoms	2,438 / 2,930
<i>B</i> -factor statistics (Å <sup>2</sup> )	
<i>B</i> -factor	10.7 (overall) / 9.3 (Wilson plot)
Catalytic domain, main/side chain	7.3 / 8.4
Zn <sup>2+</sup> (1 in total / 1 mol per monomer)	5.3
Inhibitor atoms (14 in total / 1 mol per monomer)	5.8
Glycerol atoms (42 in total / 7 mols per monomer)	19.4
Solvent atoms (435 in total)	26.0
Protein geometry <sup>d</sup>	
Ramachandran favored	97.3 % (293 of 301 residues)
Ramachandran allowed	99.7 % (300 of 301 residues)
Ramachandran outliers	0.3 % (1 of 301 residues, Ser199)
Residues with bad bonds/angles	0.00 / 0.00 %
Rotamer outliers	0.77 %
Notes	
<sup>a</sup> $R_{\text{merge}} = \sum_{\text{hkl}} \sum_{j=1}^N  I_{\text{hkl}} - I_{\text{hkl}}(j)  / \sum_{\text{hkl}} \sum_{j=1}^N I_{\text{hkl}}(j)$ , where <i>N</i> is the redundancy of the data. The outermost shell is 1.46 -1.38 Å	
<sup>b</sup> $R_{\text{factor}} = \sum_{\text{hkl}}   F_{\text{obs}}  -  F_{\text{calc}}   / \sum_{\text{hkl}}  F_{\text{obs}} $ , where the <i>F</i> <sub>obs</sub> and <i>F</i> <sub>calc</sub> are the observed and calculated structure factor amplitudes of reflection hkl	
<sup>c</sup> <i>R</i> <sub>free</sub> = is equal to <i>R</i> <sub>factor</sub> for a randomly selected 2 % subset of reflections that were not used in refinement	
<sup>d</sup> According to Molprobit (Davis et al., 2004)	

From the initial maps, the extra electron density clearly indicated that the thiirane is bound to the enzyme's nucleophile Glu270 (the equivalent residue in MMP2 is Glu404). In addition, the thiol group of the inhibitor coordinates to the catalytic zinc ion in a monodentate manner to form the tetrahedral coordination sphere of the metal. This confirms that a breakdown of the thiirane ring was operated as part of the normal catalytic mechanism carried out by the metalloprotease. The compound was prepared as the racemic mixture. However, the 2*S*,3*S*-**22b** is the species that fitted into the electron density map implying that the enzyme is enantioselective in its interaction with the inhibitor. This configuration corresponds to the *D* series in accordance with has seen before for oxirane **22a**.



**Figure 31.** A stereo plot of a  $2F_{\text{obs}} - F_{\text{calc}}$  electron density map showing the active site of CPA with bound thiirane **22b** (sticks model). The inhibitor is covalently bound to Glu270 (also in sticks), as implied by the continuous electron density. Side-chains from residues important for binding are shown as sticks model and labelled. The catalytic zinc ion is shown as a sphere in magenta, while the sulfur from **22b** is in yellow. Oxygen (colored red) and nitrogen (blue) atoms important for binding interactions are marked. The map (wire mesh in green) is contoured at a  $1.5 \sigma$  level.

## 7.8 Conformational changes

As compared to native unbound CPA crystal structure, some conformational changes occurred upon inhibitor binding. The most mobile residue was the catalytically crucial Tyr248 followed by Arg145 (see Fig. 32, below). The phenol ring moved from the surface of the protein to the bottom of the active site cleft to make a strong hydrogen bond to **22b** carboxylate group (the bond distance is 2.59 Å). This carboxylate group is held in place by Arg145 through a salt bridge, and by a hydrogen bond interaction to Asn144 N $\delta$ 1. The benzyl ring of **22b** is accommodated in the S1' pocket, and is thus interacting with Ile255, the specificity-determinant residue. In this pocket, residues Leu203 and Ile243 make C<sub>H</sub>- $\pi$  interactions with the benzyl ring of **22b**. To make such apolar interactions, many of the water molecules present in the native CPA S1' pocket have been dispersed upon **22b** binding. A list of geometric details can be found in Table 14.

Interaction	Parameter (Å, °)	Comments
His69-Zn, His196-Zn, Glu72-Zn, S <sub>1</sub> -Zn	2.10, 2.11, 2.04, 2.33	Zinc coordination sphere
Glu270 O $\epsilon$ 2-C <sub>4</sub>	1.61	Covalent bond to nucleophile Glu270
Tyr248 OH-COOH	2.59	Hydrogen bond phenolic ring
Arg145 N $\eta$ 1-COOH	2.89	Salt bridge
Arg145 N $\eta$ 2-COOH	2.78	
Arg127 N $\eta$ 2-COOH	3.35	Anchoring of COOH
Asn144 N $\delta$ 1-COOH	2.97	
Leu203 C $\delta$ 1-C <sub>phenyl</sub>	3.85	Hydrophobic interactions of the benzyl ring (selected)
Ile243 C $\delta$ 1-C <sub>phenyl</sub>	3.42	
Thr268 C $\gamma$ 2-C <sub>phenyl</sub>	3.68	
C <sub>3</sub> -S <sub>1</sub> , C <sub>4</sub> -C <sub>3</sub> -S <sub>1</sub>	1.81, 100.9	Geometry around the sulfur

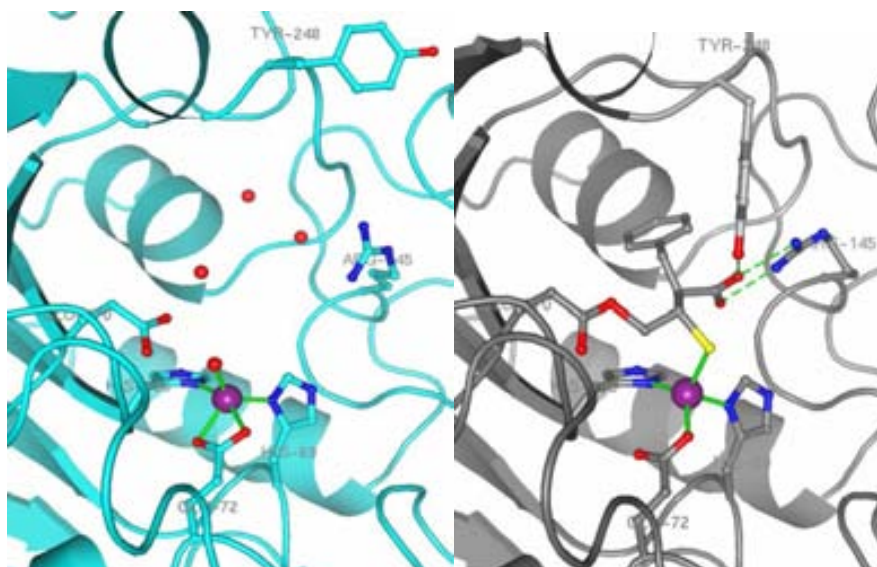
The Glu270 side chain conformation in the **22b**-bound CPA crystal structure is approximately *-gauche* (as measured by the C $\alpha$ -C $\beta$ -C $\gamma$ -C $\delta$  dihedral angle, -74.4°), close to the value adopted by the native counterpart. On the whole, our observations point to the fact that little changes in key micro-environments in the binding pocket of the enzyme would allow for an exquisite



recognition and binding of the ligand, shielding it from solvent contact, despite of it being conformationally dissimilar to a natural substrate.

## 7.9 The zinc environment

The environment around the catalytic zinc was perturbed by **22b** binding. A comparison to an unbound CPA can be made by inspection of Fig. 32 (see next page). In the bound structure, the coordination number of the zinc ion is four and the resulting geometry is tetrahedral, with three protein ligand atoms at the base and the sulfur of the inhibitor at the apex of a regular tetrahedron. The Zn-N distances are equal (2.10 Å and 2.11 Å for His69 and His196, respectively), while the Zn-S distance is the longest, 2.33 Å. The Zn-O distances are differing by almost 1 Å (2.04 *versus* 3.00 Å), clearly showing that Glu72 is a monodentate ligand. The conformation of the glutamate carboxylate is *syn* as the O-C-O-Zn dihedral angle is -2.56°. In native unbound CPA, the zinc coordination number is 5 owing to the presence of the second O atom from Glu72 (the Zn-O distances are 2.13 Å and 2.26 Å). A water molecule is instead at the apex of a distorted tetrahedron, but at a shorter distance from the metal ion (2.07 Å). Again, the conformation of the glutamate carboxylate is *syn* (O-C-O-Zn = -1.69°). In general, the coordination spheres are comparable in size, although subtle differences produce a more compact metal ion environment in the native with respect to the bound structure.



**Figure 32.** Comparison of the active site cavities of unbound (left panel) and thiirane-bound (right panel) CPA structures. In unbound CPA, the zinc (magenta sphere) coordination environment corresponds to a distorted tetrahedron which is held up by interactions (green solid lines) with three CPA residues and is completed by a fourth water ligand (red sphere). Notice the side chain conformations of Arg145 and Tyr248 and the water filled cavity. In thiirane-bound CPA, the zinc environment is tetrahedral owing to a slight change of Glu72 side chain, which now provides only one ligand atom. The sulfur of the thiirane has displaced the fourth water ligand. Now, Tyr248 and Arg145 have suffered conformational changes to contact the inhibitor via hydrogen bonds (green dashes). The orientation is similar in both structures.

## 7.10 The ligand structure

Compound **22b** adopted a *trans* conformation in binding to CPA (the main  $C_1-C_2-C_3-S_1$  backbone dihedral angle being  $-162.4^\circ$ ). The bond distance from  $C_4$  to Glu270  $O_{\epsilon 2}$ , 1.61 Å, is slightly larger than the reported  $C_4-O$  bond length of 1.4 Å for the oxirane (Yun et al., 1992; Ryu et al., 1997). The  $C_3-C_4-O_{\epsilon 2}$  bond angle compares favorably to that seen in CPA in complex with a iodopropanoate inactivator,  $114.3^\circ$  versus  $113.5^\circ$  (Massova et al., 1996). The bond angles around the  $C_3$  carbon atom, which bridges the sulfur coordinated to the zinc with  $C_4$  covalently bond to Glu270, average  $108.3^\circ$ . Remarkably, the bond angle of

100.9° (for C<sub>4</sub>-C<sub>3</sub>-S<sub>1</sub> linkage) is well below the ideal tetrahedral value of 109.5°, suggesting a structural distortion at the end of the molecule. This effect may be physically assimilated to a spring that is strained between the metal atom and Glu270 from the protein. The possible effects that this distortion might impose on the binding of the inhibitor to the target enzyme, and hence over its inhibitory potency, would only be assessed by analyzing compounds structurally related to the inhibitor analyzed here.



# Overall discussion



# Overall discussion

Proteolytic enzymes are fascinating proteins because of their power to modify other proteins permanently. Proteases are ubiquitous in Nature and have been estimated to make up approximately 2 % of all gene products, suggesting that the human genome encodes about 700 proteases. Proteases participate in a number of physiologically important processes, from digestion of food protein to proteolytic cascades, playing a central role in organ development, protein hormone maturation, fibrinolysis and blood coagulation, among others. Protease activity is highly regulated, but sometimes proteases escape control, leading either to increased or decreased activity and biological malfunctioning or diseases. Several different functional kinds of proteases exist. Among them, we used the M14 family of proteases, or metalloproteases (MCPs or CPs) as targets for our study.

The M14 family appears as a complex and populated array of proteases. Some of them may have an industrial application, such as in the processing of recombinant insulin. Other CPs have been disclosed as biopharmaceuticals in cancer treatment or as antifibrinolytics. CPs of pharmaceutical importance include CPA1, as a biomarker for early cancer detection, and its relatives CPA2 (with high expression levels in various disease states; Golz et al., 2006), CPA3 (a carboxypeptidase expressed in mast cells that acts in host defense), CPA4 (potentially linked to prostate cancer; Ross et al., 2009) and CPA6 (upregulated in oral squamous-cell carcinoma, a cancer of head and neck; Fialka et al., 2008). Further relevant CPs include the human CPB released from the pancreatic gland (involved in acute pancreatitis and pancreatic cancer) and various non-pancreatic CPB-like enzymes. Non-mammalian CPs, such as crop pest insect CPA and CPB, are potential targets for insecticide development (Bayés et al., 2005). TcMCP-1 and TcMCP-2 from *Trypanosoma cruzi*, a disease causing flagellate protozoan parasite, might be considered as novel targets belonging to the M32 family of peptidases (Niemirowicz et al., 2007). Along with human cytosolic carboxypeptidases (CCPs), the latter species have

begun to be better understood in terms of biological function and structural properties.

Structural and functional approaches have been combined in this Thesis in order to understand the determinants of MCP inhibition by small molecules. The work and the results that have been previously presented may be divided in three major areas which will be summarized and discussed below.

(i) With the aim of identifying potential functional differences between several M14 family members, and starting from their crystal structures, an analysis of protein-protein docking interfaces was carried out based on the use of the ODA (Optimal Docking Area) method. The crystal structure pool included a number of well-known mammalian CPs, one CP from bacterial origin, and two proteases from insects. The polypeptide chains of available M14 protease crystal structures were separated into four components: the catalytic domain common to all CPs, the N-terminal pro domain, the C-terminal extension, and finally the proteinaceous inhibitors. As the catalytic domain was arbitrarily considered the central part of the bound systems, the other pieces were designed as the “binding partner”. 24 polypeptide chains were thus analyzed. Significant ODA residues, i.e. those ODAs above a sensible threshold energy value, were calculated for all the pieces. When the original bound systems were reconstructed, the number of ODA residues at the interface was compared to the number of real interface residues. In 80% of cases, the ODA method rendered a correct prediction of known protein-protein binding sites (considering the catalytic domain alone). Only two proteins escaped this trend. However, the overall success of the method dropped to 58% due to bad predictions from the binding partner side, among other factors. The predictions were poor for the small proteinaceous inhibitors (especially PCI) and for CPD/CPM catalytic domains, which at variance with the M14A subfamily, do not possess pro-region. This may be due to an artifact of our procedure: at the time the ODA analysis was performed, some crystal structures, like that of CPN (which is a heterotetramer) and the CPU zymogen (from two different sources) were not available. Thus, the number of protein interfaces analyzed was over-represented for the M14A subfamily, as opposed to the only two M14B subfamily members, CPD and CPM.



The ODA patches locate close to and in the neighbourhood of the active site cleft, with no ODA residues residing in other regions over the protein's surface. This may be related to the biological function of the M14 proteases: the protein-protein binding surfaces might have evolved to recognize and channel the peptide substrate towards the active site cleft.

A plot of ODA residues over the protein's surface suggests that the M14 proteases show non-identical protein-protein interaction sites. This is also relevant because of close structural homology among CPs. It is possible to distinguish regions in the mapping of ODA points onto the protein's surface and from sequence alignment. For example, surface patches form differently sized regions for human CPA and CPM. In the case of the enzymes from lepidopteran insects, only one region - an entire ODA patch - is observed and encompasses the surroundings of the active site cleft entry. This represents the greatest surface area found for any CP in the family. Aside from its prodomain, there is a scarcity of structural and biological information about HaCPA, the insect enzyme, that may preclude any precise conclusions from these results. We speculated that such enlarged protein-protein docking area might be related to an as yet unknown function of these CPs. Of interest here is to mention that some small synthetic inhibitors displayed the tightest inhibitory effect on the lepidopteran insect CPs. A docking onto the ODA mapped structure reveals the fit of one inhibitor, a *Z*-L-Ala derivative harboring a long hydrophobic tail, into a narrow crevice that extends along the protein's surface from the entrance of the active site cleft. This last example illustrates the advantages of the combined use of computational tools in understanding the behaviour of some of the inhibitors, which were able to bind to the mostly featureless CP surface.

(ii) Related to the goal of obtaining new inhibitors for MCPs, the screening of large compound collections is one of the possible avenues able to provide valuable ligands for novel disease targets (Hergenrother, 2006). Such an approach allowed us to identify new small molecular weight CP inhibitors. Their inhibitory potency seems to be slim as compared to other known inhibitor molecules. However, what we wanted in this exercise was to discover novel scaffolds for the design of new CP inhibitors. We classify our findings in seven different chemical classes of potentially new small-molecule CP scaffolds. The

identified compounds have not been originally designed to perform a CP inhibitory activity, and thus they may be regarded as interesting leads for further developments. Importantly enough, various small molecular weight compounds performed well as human CPB inhibitors. The 1,3,4-oxadiazole-bearing compound showed potent inhibition against this enzyme (with a  $K_i$  value of about 1  $\mu\text{M}$ ), and harbors a new type of zinc chelator, distinct from known inhibitors. The strengths and weaknesses of the lead compounds have been highlighted in fitting the target macromolecule by molecular docking. Some points along the oxadiazole scaffold seem to fit and interact well with the macromolecule structure, as exemplified by the binding of Tyr248 to the oxazolyl oxygen atom, although other points are less well exploited. Other classes of chemicals, like the peptoids, the thioxophosphoranyl oxiranes and the cyclobutane  $\beta$ -peptides share with the oxadiazoles the ability to block the CP enzymatic activity, although with different potencies. The latter are also distinct because of their ease of obtention. The peptoids, which behaved as inhibitors of CPB and CPB-like enzymes, were synthesized from selected library mixtures. Those scaffolds might be used as a starting point in the design of novel inhibitors of CPU given that CPB is often used as a model for this purpose. Recent progresses in the structural characterization of this long time elusive 3D structure may allow a structure-guided approach. Inhibitors of CPU would enhance fibrinolysis and have an anti-coagulant effect. It is thus desired that new generations of organic synthetic inhibitors, based on the discovered scaffolds, may provide potent inhibition and selectivity for the target macromolecule.

(iii) The two experimental areas described before depend on the structural knowledge of the target enzymes. To this end, crystallization of several macromolecules was attempted. In the course of our investigations we obtained a new crystal form of the CPB zymogen, which has distinct features compared to the previously known structure. A comparison of the newly determined CPB in the triclinic unit cell with the previously known tetragonal unit cell, indicated a completely different packing arrangement. The presence in the tetragonal form of a small molecular weight compound between symmetry-related polypeptide chains would have facilitated the formation of a hydrophobic

nucleus in which some aromatic residues are clustered. On the contrary, the hydrophobic cluster is not observed in the triclinic crystal since the members of the cluster are now solvent-exposed. One of the aromatic residues member of the cluster is the catalytically relevant Tyr248. The fact that the structure determined herein presents Tyr248 fully solvated, as compared to the tetragonal form, and the computation of the physicochemical features of the Tyr248 flanking region, allowed us to propose a major role for this residue in ligand binding and recognition. According to our observations, the presence of a tyrosyl ring at position 248 (according to bovine CPA numbering) would provide a delicate balance between extreme hydrophobicity and hydrophilicity of the region guarding the entrance to the active site cleft. Added to this, mutants could show a different ability to desolvate or de-aggregate the ligand, and thus would impair the efficiency of the catalytic reaction.

Despite much efforts, the obtention of crystal structures of target CPs in complex to desired small-molecule ligands was difficult to attain. The notable exception was the obtention of the complex structure of a thiirane mechanism-based inactivator (**22b**) bound to CPA. A mechanism-based inactivator exploits the reactivity of part of the molecule in its interaction with the target enzyme. The nucleophilic attack by Glu270 to the thiirane moiety, with the attendant ring opening and formation of a covalent linkage, was verified as part of the normal catalytic mechanism of the enzyme. The complex was obtained at the highest resolution for a CPA bound crystal structure. Thus, finer details of atomic bonding could be observed. For example, the electron density clearly indicated the covalent linkage between the inhibitor and the nucleophile Glu270. The aromatic ring of the inhibitor docks into the primary specificity pocket while Tyr248 is in the closed conformation contacting the inhibitor's carboxylate group. The inhibitor is coordinating the zinc ion, and thus perturbs the zinc environment geometry, in a monodentate manner as (presumably) the thiolate. By analogy to a comparable gelatinase inhibitor, this would yield an extremely powerful species. However, this expectation did not materialize, suggesting that the molecule is not optimally engaged in interacting with the macromolecule. Our analysis of protein-ligand interactions showed that part of the molecule has a distorted geometry, and this would result in a suboptimal binding. This

information may be useful to understand the properties of other classes of metallopeptidases, such as gelatinases, due to the lack of structural data available on them (Lee et al., 2008). Aside from this, the concept of a covalent binder to CPs may be useful in the design of bioimaging agents to analyze the activity of known or barely known CPs, either an isolated species or rather a relatively crude cell extract. Compounds of this class may act as covalent tags to M14 proteases. It should be mentioned that the design of bioimaging agents to track CP activity may be compromised by the very function of the enzyme. As it performs as an exopeptidase, the room allowed for substrate binding is limited as compared to endoproteases. Then, the use of peptide substrates adorned with conventional chromophoric groups, which tend to be large aromatic moieties (Lavis & Raines, 2008), would be unsuited to be used for CPs. Therefore, one future direction of work might entail the chemical modification of compounds **22** to afford a bioimaging agent for sensible carboxypeptidase detection. The crystal structure of **22b** in complex to CPA would thus represent the starting point for a structure guided drug design effort.

# Conclusions



# Conclusions

**1** The characterization of protein-protein binding sites among M14 peptidases was used to gain a further insight into their association capabilities towards other macromolecules as well as against small molecular weight ligands

- The predictions were found to be correct in 80% of the cases when considering the central catalytic domain of MCPs
- MCPs from the M14A and M14B families can be distinguished by their pattern of ODA patches
- Insect MCPs display surface areas more than twice larger than their vertebrate counterparts
- Differences among the various enzyme classes can be exploited in the design of inhibitors
- The ODA analysis has been successfully applied to the analysis of interactions involving small molecule inhibitors and to the analysis of the activation processes of natural zymogen binary or ternary complexes

**2** Seven new classes of chemical scaffolds were identified as metalloprotease inhibitors and potential leads for further development

- Arginine derivatives, with a preference for lepidopteran enzymes
- The thioxophosphoranyl oxiranes, the cyclobutane  $\beta$ -peptides, the *L*-alanyl derivatives and the organosilicon compounds bear unusual functionalities distinct from known inhibitors
- Peptoids and oxadiazoles were obtained from screening, *in vitro* as well as computational, of large libraries. The best representatives are inhibitors in the low micromolar range

**3** A new CPB crystal form was obtained at higher resolution than previous structures

- It has allowed for an accurate assignment of the polypeptide chain in accordance to the genomic sequence information.
- The phenolic group of Tyr248, initially in the open position in the unbound enzyme, anchors the ligand and then serves as a pivot that moves inwards to the catalytic site to finally adopt the closed position.
- The region surrounding Tyr248, uniquely present in M14 peptidases, is flanked by hinge residues and tracks the movement of Tyr248

**4** The crystal structure of carboxypeptidase A in complex to a thiirane-based inactivator was determined at high resolution, representing the first example among metalloproteases of this kind of interaction

- The thiirane, a mechanism-based inactivator, covalently modified Glu270 at the active site
- This crystal structure confirms the ring opening at the active site of the enzyme with the attendant formation of a thiol group
- Structural information invites for further developments in this class of small molecular weight inhibitors



# Methods



# 8 Methods

The detailed description of each particular methodology can be found in each attached paper in the Publications Section. To avoid duplication, this Chapter will introduce the experimental procedures used in a broader sense.

## *8.1 General materials and media*

Either the solutions and the material was sterile autoclaved for 25 min at 120 °C and 1 atm, or in case of non-autocleavable solutions, they were sterile filtered under a laminar flow workbench with 0.22 µm Millex-GS filter units (Millipore, Billerica, MA, USA). Depending on the purpose, the solvent may be distilled water or deionized ultrapure water from the Milli-Q system (Millipore). Media for growth of bacteria (*E. coli*) or yeast (*P. pastoris*) were prepared according to the manufacturer's instructions or following standard laboratory procedures. For molecular biology procedures, *E. coli* subcloning-grade strains XL1-Blue (Stratagene, La Jolla, CA, USA) and MC1061, and *P. pastoris* strain KM71 (Invitrogen, San Diego, CA, USA) were used throughout. The extracellular expression of insect CPs was performed in *P. pastoris* cells harboring the pPIC9 expression vector (Invitrogen) according to known procedures (Bayés et al., 2003). The production, extraction and purification of plasmid DNA (with the GFX Micro Plasmid Prep Kit; GE Healthcare, Piscataway, NJ, USA) bearing the insect CP clone has been performed regularly to retransform fresh *P. pastoris* cells (using the Spheroplast Method Transformation; Invitrogen).

A list of appliances and equipment used is found below (Table 15).

<b>Table 15 Equipment and appliances</b>		
Name and Model	Supplier	Notes
Presoclave 75	Selecta (Spain)	Autoclave
AJ100, PJ300, H64	Mettler (Germany)	Balance
DC-SDS PowerPac 300	Bio-Rad (USA)	Vertical SDS Gel Chamber System
MiniProt II		
DC Power Supply PS 500XT		
Miniphor Submarine Electrophoresis Unit LKB 2013	Pharmacia (Sweden)	Horizontal Agarose gel Electrophoresis System
Maldi-TOF Biflex	Bruker (Germany)	
Äkta FPLC System	Amersham Pharmacia Biotech (USA)	Includes Fraction Collector and UNICORN Control System
TSK-Gel DEAE-5PW	TosoHaas (Germany)	FPLC Columns
Toyopearl HIC Butyl-650 M	TosoHaas (Germany)	HIC column matrix
CARY 100 Bio UV-Vis	Varian (Australia)	UV-Visible spectrophotometer
Freezers (-20°C, -80°C)	Sanyo (Japan)	
Max <sup>o</sup>	Barnstead (USA)	Horizontal Shaking Incubator with temperature control
J2-21, JA14, JA20	Beckman Coulter (Ireland)	Centrifuge/Fixed Angle Rotors
Incubator	Heraeus (Germany)	Fixed Temperature
Speed-Vac Concentrator	Savant (USA)	
Comfort Refrigerator	Liebherr (Germany)	Fridge +4°C and -20°C
Magnetic Stirrer A-03	SBS (Spain)	
Milli-Q	Millipore (USA)	Ultrafiltration Water System
Micromax	Micro-IEC (USA)	Benchttop Microcentrifuge
Micro pH 2001 Includes Ingold Electrode (Switzerland)	Crison Instruments SA (Spain)	pH meter
TWO-30	Faster (Italy)	Sterile Working Bench
Reax 2000	Heidolph (Germany)	Vortex
Megafuge 1.0 R	Heraeus Sepatech (Germany)	Refrigerated Ultracentrifuge
Tectron Bio Medic 60	Selecta (Spain)	Water Bath and Thermostatic Controller
Sonicator		Water Bath and Thermostatic Controller
96-well plates	Iwaki (Japan)	ELISA microtiter plates, flat bottom
Crystallization plates	Greiner (USA), Jena Bioscience (Germany), Axygen (USA), Cryschem (Hampton Research), Emerald Biosciences	Hanging and sitting drop
Rumed	Rubarth Apparate GmbH; Germany	Crystal Farm, steady temperature
Hampton Screen I and II	Hampton Research (USA)	Crystallization screens
Victor3	PerkinElmer (USA)	Microplate reader
Leica	Germany	Magnifier
General Reactives	Bio-Rad Laboratories (USA), Merck (Germany), Sigma Aldrich (USA), Acros Organics (USA), Bachem (Switzerland)	

## 8.2 Protein handling

Proteins for inhibition kinetic assays and crystallization were obtained from different sources. Mammalian model CPs, such as human and porcine CPB, were recombinantly expressed in the *P. pastoris* system, proteolytically activated, and purified as already described (Ventura et al., 1999). CPs from lepidopteran insects, viz. *Helicoverpa armigera* CPA and *H. zea* CPB, were also expressed in *P. pastoris* and prepared as reported (Estébanez-Perpiñá et al., 2001; Bayés et al., 2003). Recombinant *Trypanosoma cruzi* TcMCP-1 and TcMCP-2 were expressed as fusion proteins in *E. coli* as published (Niemirowicz et al., 2007). For some experiments we used commercially available bovine pancreas CPA (Sigma, St. Louis, MO, USA) and porcine CPB (Calbiochem, San Diego, CA, USA).

The expression and purification of protein was carried out following established protocols. For example, the large scale expression of lepidopteran insect CPs was performed in 2 liter flasks. First, the cell biomass was allowed to grow until  $OD_{600} = 2-6$  units, at which point the cells are harvested and resuspended in a fraction volume of the medium (usually one fifth or one tenth of the volume). Then, the expression period was induced by adding methanol (up to 2 % v/v) to the medium. Normally, the growth period and the protein expression process require 4-5 days to complete. The protein content was recovered from the supernatant of the centrifuged medium, and precipitated in 0.02 M Tris-HCl, pH 7.5, containing 30 % ammonium sulfate.

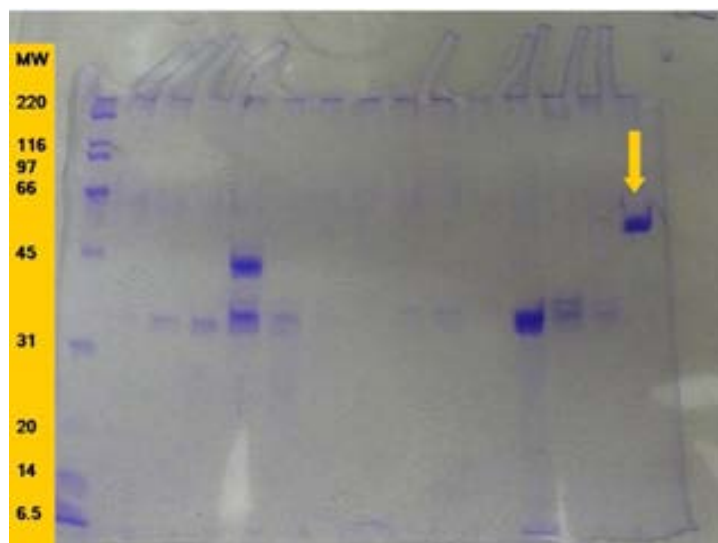
The purification of the insect CPs was carried out in two steps. First, the protein sample in ammonium sulfate is loaded into a hydrophobic interaction chromatography (HIC). Protein binds to the matrix because its hydrophobic groups became exposed due the high salt content. Protein is separated using a gradient of decreasing salt strength, using buffer 0.02 M Tris-HCl, pH 7.5, and pooled with a fraction collector. As the following step is a ion exchange chromatography (IEC), the sample needs to be dialyzed to eliminate salts that can disturb binding to the resin. The dialysis was performed in dialysis sacks with a cut-off of 12 or 25 kDa (Medicell International LTD, London, UK) in a large volume of buffer, usually 0.02 M Tris-HCl, pH 7.5. IEC chromatography was performed on an Äkta FPLC System (GE Healthcare) equipped with a Frac-

950 fraction collector (Figure 33). This chromatography exploits the reversible effect of ion exchange between covalent bonded groups of the column matrix and the protein surface ions. Here we use a DEAE anionic exchange chromatography with positively charged functional groups that exchange negatively charged ions.



**Figure 33.** Äkta-FPLC purifier system equipped with fraction collector Frac-950 and external pump P-960.

The proteins obtained recombinantly are the immature forms or zymogens. To be used in the inhibition assays they must be activated via proteolytic treatment. The activation is usually performed with trypsin, as published (Bayés et al., 2003). The purity of the protein samples was further evaluated by a denaturing SDS-polyacrylamide gel electrophoresis (Fig. 34).



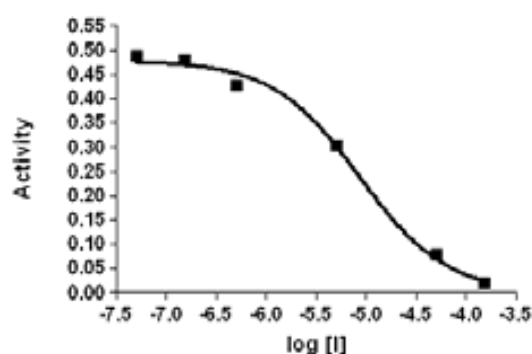
**Figure 34.** A 12 % SDS-polyacrylamide gel electrophoresis with protein bands stained with a dye, Coomassie brilliant blue. The arrow (lane 15) marks a band corresponding to the  $\approx 60$  KDa TcMCP-1. Other lanes correspond to different CPB batches indicating the high molecular weight zymogen of about 45 KDa and the low molecular weight mature enzyme of  $\approx 35$  KDa. Lane 1 corresponds to the molecular weight markers.

Active enzymes were kept in a freezer at  $-20^{\circ}$  for long term storage or at  $4^{\circ}$  for immediate use. Protein concentration was always estimated using the measurement of UV absorbance at 280 nm in a spectrophotometer. The concentration of the protein is then calculated with the Lambert-Beer equation using the extinction coefficient  $E_{(280)}$ , which represents the absorbance of a 1 g / l solution of the protein.

### 8.3 Kinetic measurements and data treatment

Enzyme inhibition assays were performed for a wide range of organic synthetic ligands. They are much smaller in size than the smallest known proteinaceous CP inhibitor, PCI. Typically, their MW range from 250 to about 550 Da. Whenever possible, the synthetic organic compounds were dissolved at the highest concentration possible, preferably in water or a suitable solvent. For example, the commercial screening compounds were kept as 10 mM stock

solutions in DMSO while the arginine derivatives were dissolved at 25 mM in 0.02 M Tris, pH 7.5. Heating, shaking or sonication were used to assure complete dissolution of the sample. The working solutions included various concentrations of each compound and one substrate concentration, usually at the  $K_m$  value. According to preliminary results, the final assayed concentrations were step-wise varied between 50 nM and 100  $\mu$ M (for commercial compounds) and between 200 nM and 1000  $\mu$ M (for the acyl arginine derivatives and other compounds). One blank tube was used as the control without inhibitor. The enzyme concentration was adjusted for to the kinetic measurements to take place within the initial velocity regime. The experimental setting involved the use of 1 ml polystyrene cuvettes or 96 well microtiter plates. In the former case, the changes in absorbance due to proteolytic activity were followed continuously in a Cary UV-Vis Spectrophotometer (Varian Inc., Palo Alto, CA, USA). A multimode plate reader Wallac 1420 Victor3 system (PerkinElmer Life Sciences, Waltham, MA, USA) with automatic dispensing of substrate to start the reaction was used in the microtiter plate procedure. In both cases, kinetic measurements were performed at room temperature. The kinetics data were then treated with the GraphPad Version 5.0 package (GraphPad Software, San Diego, CA, USA). After ensuring a good agreement between the experimental observations and the inhibition model, the  $K_i$  value was calculated from the  $IC_{50}$  following the procedure by Cheng & Prusoff (Cheng & Prusoff, 1973). An example semilog graph of inhibitory activity is shown in Fig. 35.



**Figure 35.** Plot showing the fit of an inhibitory model (continuous line) to the experimental measurements (full squares).



The experimental assay conditions are listed in Table 16.

<b>Table 16</b>		
<b>Assay conditions for carboxypeptidases</b>		
Enzyme	Substrate/Concentration	Buffer
HaCPA	<i>N</i> -(4-methoxyphenylazobenzoyl)-Phe-OH <sup>a</sup> 0.1 mM	50 mM Tris 0.5 M NaCl pH 7.5
H <sub>z</sub> CPB	<i>N</i> -(3-[2-furyl]acryloyl)-alanyl-lysine <sup>b</sup> 0.06 mM	20 mM Tris 0.1 M NaCl pH 7.5
bCPA	<i>N</i> -(4-methoxyphenylazobenzoyl)-Phe-OH 0.1 mM	50 mM Tris 0.5 M NaCl pH 7.5
pCPB hCPB	<i>N</i> -(4-methoxyphenylazobenzoyl)-Arg-OH <sup>c</sup> 0.06 mM/0.13/0.2 mM	20 mM Tris 0.1 M NaCl pH 7.5
TcMCP-1	<i>N</i> -(3-[2-furyl]acryloyl)-alanyl-lysine 0.16 mM	MES 0.1 M pH 6.2
TcMCP-2	<i>N</i> -(3-[2-furyl]acryloyl)-phenylalanyl- phenylalanine <sup>d</sup> 0.05 mM	Tris 0.05 M CoCl <sub>2</sub> 50 μM pH 7.5
Refs <sup>a</sup> Mock et al., 1996 <sup>b</sup> Plummer & Kimmel, 1980 <sup>c</sup> Mock & Stanford, 2002 <sup>d</sup> Peterson et al., 1982		

## 8.4 X-ray crystallography - crystallogenesis

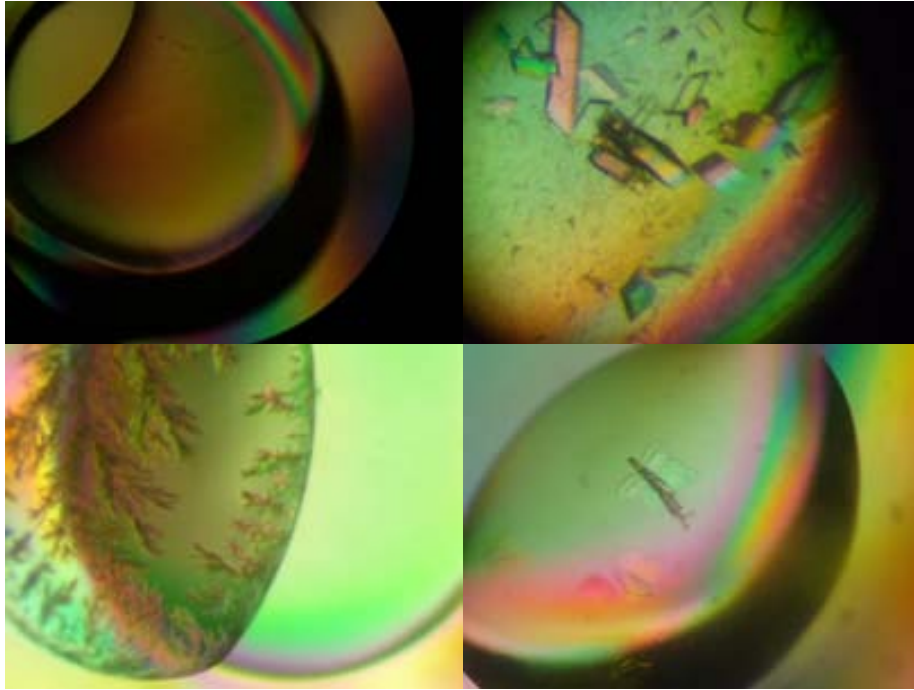
To perform an X-ray diffraction experiment it is necessary to prepare the protein sample in a very special way. The protein should form an ordered three-dimensional array, a crystal. Therefore, the protein needs to be removed from the solution and be kept within a region of the phase diagram known as the nucleation zone. Here is where crystallization takes place and results in the growth of crystals. Regions outside the nucleation zone led to the formation of material inadequate to X-ray diffraction analysis. Routinely, proteins are reluctant to be in the nucleation zone. Therefore, conditions need to be screened in order to find out the suitable one to let the protein form crystals. The

first approach involved the use of crystallization conditions for known CPs. More extended screening was performed with crystallization conditions either from commercial suppliers or made by hand following the most successful receipts from public structural genomics initiatives. In the latter case, the typical preparation involves mixing of stock solutions of salts, buffer or a precipitant. The solutions were freshly prepared and stored in a cold room either in 1.5 ml eppendorf microtubes or 1.2 ml deep well blocks. Besides varying the chemicals used for protein crystallization, other variables such as protein concentration, temperature, and the inclusion of additives were also screened. The experiment can be carried out in 24 or 96 well plates, either in a hanging or sitting drop setting (Fig. 36), and the plates are stored in a steady temperature incubator or crystal farm.



**Figure 36.** Crystallization plates. Left, a Cryschem 24 well sitting drop plate. Right, a Corning 96 well plate for nanodrops.

A drop is made up by a mixture of equal volumes of protein (usually at high concentration) and reservoir (the crystallization mixture). As the concentration of the crystallization mixture in the drop is half of that in the reservoir, liquid evaporates from the drop to reach equilibrium. This is called vapour diffusion, one of the most popular methods by which a protein can be crystallized. As a result of vapour diffusion, protein concentrates. The time course of protein crystallization may range from a few days to several months. Crystallization drops were observed with a magnifier. Some possible outcomes that a crystallization screening experiment may take are illustrated in Fig. 37.

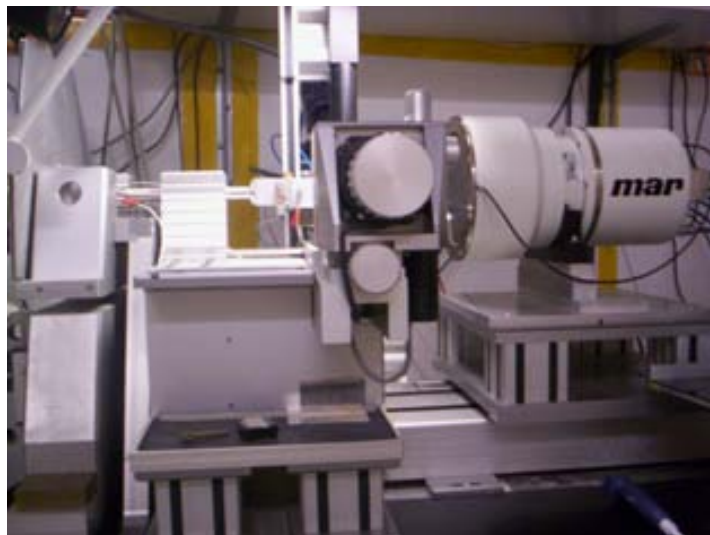


**Figure 37.** Photographs from some crystallization screening experiments. Upper row. *Left*, a clear drop without crystals; *right*, CPB crystals from macroseeding (protein precipitate is also seen). Lower row. *Left*, dendrites; *right*, plate-shaped crystals grew up from a contaminant fiber.

## 8.5 X-ray crystallography - data handling

Crystals are solvated to a great extent. This means that their handling is a delicate task because they are easy to destroy. The crystal is mounted into a tool in the form of a hook called loop. The loop, made up by a nylon fiber (to avoid its diffraction), is attached to a plastic base containing a small magnetized metal piece. The latter serves to fix the loop in the goniometer head of the beamline. The goniometer head serves to position the crystal in the X-ray beam, in such a way that its center remains in position as the crystal revolves around the  $\phi$  axis. This axis lies perpendicular to the X-ray beam and is shifted by a small angle, typically  $0.2 - 1.0^\circ$ , during diffraction. The movement around  $\phi$  gives

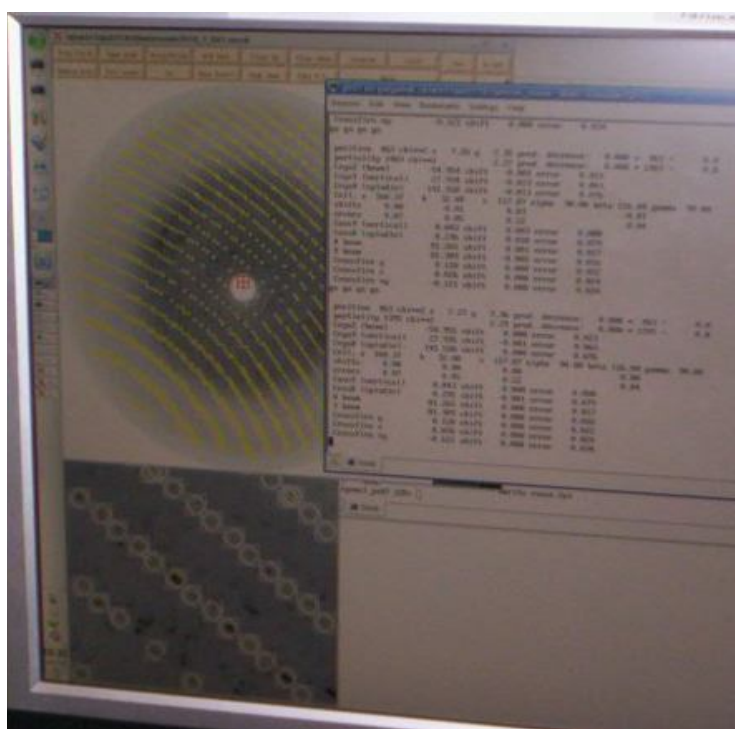
rise to successive diffraction images of slices of the crystal lattice. The  $\phi$  axis is positioned amid the X-ray beam and the detector (Fig. 38).



**Figure 38.** The working bench at an X-ray beamline. At the left is the X-ray beam source, at the right the detector. The large wheel at the front (at 12 o'clock) is the  $\phi$  axis control device.

The raw measurements are recorded as data collection progresses. The amount of data to be collected depends on the crystal symmetry, the quality of the crystal, and the resolution. The resolution is related to the extent at which diffraction is observed, that is, the spatial distribution at which spots are smeared in the image. This is related to Bragg's Law, as we will see next. The raw measurement data in each image contains intensity information (along with geometry data like crystal-to-detector distance, the position of the beam centre, and so on). Intensities are retrieved from raw data in a complex process called integration. First, data is reduced, that is, images are read by a program which guesses the symmetry operations and the dimension of the unit cell, and tries to adjust the image spot profile. The algorithm searches the best fit by adjusting some parameters like the crystal-to-detector distance, the mosaicity (the degree of crystal internal disorder), the position of the beam center, the size of the spot, among others. Once this has been done for every image, the reduced data has to be compiled into one unique set. Further steps like adjustments to account

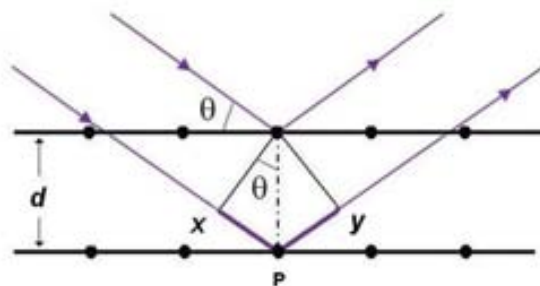
for possible intensity shifts between images (scaling), the combination of multiple readouts of a single reflection into one (merging), and the conversion of intensities to amplitudes (truncation) provide the list of reflections, part of the information necessary to solve the crystal structure. A look to a typical data reduction screen is shown in Fig. 39.



**Figure 39.** Processing of an X-ray diffraction image. The raw measurement data (black spots) are fitted by an algorithm that takes into account geometric and crystallographic information. Based on this, the program can retrieve the reflections (white circles enclosing spots). The window at the right shows statistics of the fitting process.

## 8.6 X-ray crystallography - structure solution

X-ray crystallography demands the use of special equipment, either an in-house X-ray generator or a huge facility, the X-ray light source termed synchrotron. X-rays are light characterized by a very short wavelength. In fact, X-rays are roughly of the same order of magnitude as a single C-C bond length, about  $1.5 \text{ \AA}$  ( $1 \text{ \AA}$  is  $100 \text{ pm}$  in standard CGS units). X-rays are a very penetrating radiation and only massive bodies, like the bones, can absorb them. X-rays interact with matter perturbing the electron density cloud, which turn thus into an excited energy state. As electrons regain basal state energy, they emit secondary waves of the same wavelength as the incident light. Waves emanated from different atoms interact in a way that they combine into a resultant, the diffracted wave, whose amplitude can be calculated but the phase relations between the components are lost. This constitutes the phase problem. The amplitude of the diffracted wave is a mathematical summation over all contributions from the diffracting atoms, but without knowledge of their relative phase relations the outcome is inviable. The more electrons possesses one atom, the more its contribution to the total amplitude. In a practical setting, the X-ray diffraction pattern records the intensities of the resultant, diffracted waves. The recorded intensities obey a precise geometric rule (Fig. 40).



**Figure 40.** Rays incide at an angle  $\theta$  on a crystal lattice at a point  $p$ . Atoms scatter the rays by the same angle, as in optical reflection. Two consecutive lattice planes separated by distance  $d$  produce emergent waves which have a phase difference. Geometrically this corresponds to the sum of the distances from  $p$  to  $x$  and  $y$ , so that the total phase difference of the emergent beams equals two times  $\sin \theta$ .

The mathematical derivation is known as Bragg's Law (Eq. 1).

$$d_{\min} = \lambda / 2 \sin \theta_{\max} \quad [\text{Eq. 1; Bragg's Law}]$$

where  $d$  is the distance between crystal lattice planes, and more generally is taken as the resolution (in Å),  $\lambda$  is the wavelength of the incident beam, and  $\theta$  is the scattering angle. Since  $\lambda$  is the constant magnitude, there exists an inverse correspondence between  $d$  and  $\theta$ . In practical terms, when resolution is good ( $d$  small) then  $\theta$  must be large. As mentioned above, this derives in largely smeared diffraction images. High-resolution intensities have magnitudes that fall off strongly with  $\theta$  because the atomic scattering factors are highly  $\theta$ -dependent.

Bragg's Law relates the observations, the measured intensities, to the objects present in the crystal lattice. The measured intensities are said to belong to the reciprocal space in opposition to the direct space, which is the domain of the electrons. The way round from one space to another is managed adequately by the mathematics of the Fourier transform.

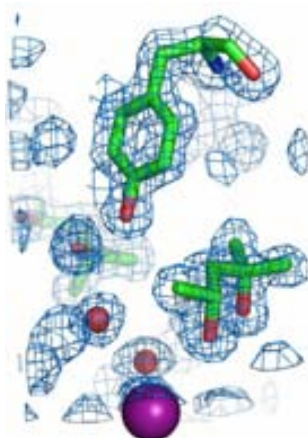
Thus, to calculate the electron density we use Eq. 2:

$$\rho(x,y,z) = (1 / V) \times \sum_h \sum_k \sum_l |F(h k l)| \exp[-2\pi i(hx + ky + lz) + i\alpha(h k l)] \quad [\text{Eq. 2}]$$

In Eq. 2,  $(x,y,z)$  is the position of every point in the unit cell,  $V$  is the unit cell volume, and  $(h k l)$  denominates every crystal lattice plane. The summation represents actually a vector, whose magnitude is given by  $F(h k l)$ , derived from the observed intensities, or  $F_{\text{obs}}$ , and whose phase angle is given by the expression inside the square brackets. The phase angle  $\alpha(h k l)$  is the unknown from any X-ray diffraction experiment.

The phase problem may be overcome with the aid of a number of methods. Phases can be obtained for an unknown macromolecule structure given that sufficient sequence homology exists to a protein whose crystal structure is already known, assuming that the folds of the sequence-related proteins will be similar. The method used to disclose the unknown structure is called the molecular replacement (MR) method. The problem is to translate the known protein structure from its original crystal arrangement to the one which is being determined. The placement of the known structure on the target unit cell

requires its proper orientation and precise position. It involves the two steps of rotation and translation. In the rotation step, the spatial orientation of the known and unknown molecule with respect to each other is determined, while in the translation step the now correctly oriented molecule is superimposed onto the other molecule. The rotation and translation searches are made with the Patterson function which uses intensities as coefficients and thus is a phaseless Fourier summation. Such a summation will only give interatomic vectors, each one arising from the origin of the Patterson function. In the rotation search, the Patterson function is convoluted with a rotated version of itself. When the rotation angles are correctly defined, then the convoluted Patterson function will be a maximum. The same mathematical principle applies to the translation function, although here the search is for vectors relating atoms from the problem and the model proteins. Once a solution of the search is encountered, we may proceed to refine the structure until geometry and stereochemistry achieve reasonable values. This may take the form of a minimization procedure (least-squares method), in which the parameters of a Fourier summation, comprising the problem protein amplitudes ( $F_{\text{obs}}$ ) and the calculated phases ( $\alpha_{\text{calc}}$ ), are changed iteratively until the difference between the  $F_{\text{obs}}$  and  $F_{\text{calc}}$  reaches a minimum value. A modern formulation uses a statistical probability method (maximum likelihood) for refinement (e.g. as in the program REFMAC). Visual inspection of electron density maps, given by the Fourier synthesis, is an important step during protein crystal structure refinement. Figure 41 shows a portion of an electron density map.





**Figure 41** (*Opposite page*). Electron density map (wire mesh in blue) contoured at 1.0  $\sigma$  level. The catalytically important Tyr248 residue (at 12 o'clock), one small molecule ligand (at 3 o'clock), water molecules (red spheres) and the zinc ion (magenta sphere) are shown.

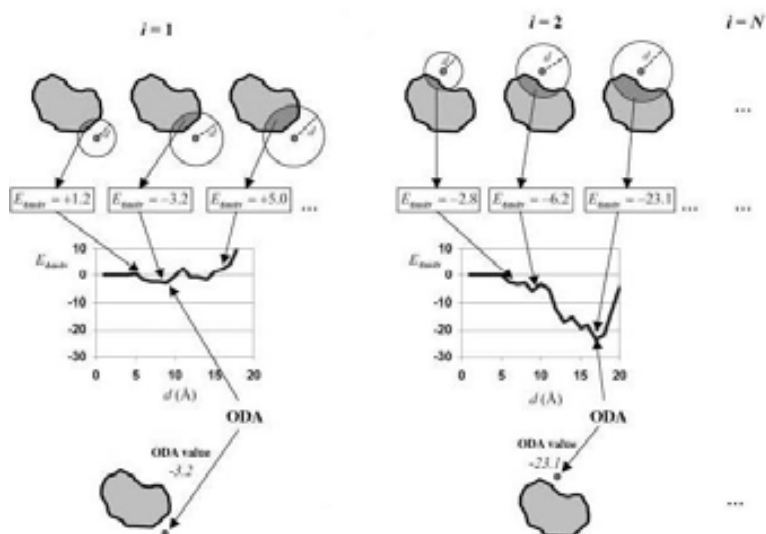
The phase problem for the small molecular weight compound crystal structure was solved by using direct methods, a completely different approach. Today, this *ab initio* technique is still not applicable in the practice for determining a full macromolecular structure. The method relies in the use of a starting subset of strong reflections that maintain a certain crystallographic relationship. One reflection can be calculated as the sum of the products of all pairs of reflections whose indices sum to it. Thus, for example:  $F_{321} = F_{100} \times F_{221} + F_{110} \times F_{211} + F_{111} \times F_{210} + F_{200} \times F_{121} + \dots$  Some reflections in this expression may be weak, and their product may not contribute significantly to the sum, so that it may be reduced to a three F-term equation:  $F_{321} \approx F_{100} \times F_{221}$ . This is known as the "triplet relationship". If both reflections in the pair are very strong and the reflection being calculated is very strong as well, then there is a high probability that this relationship alone will give a good indication of the phase of the reflection. If the reflections  $F_{100}$  and  $F_{221}$  both have the same sign (positive or negative) it is likely that the sign of  $F_{321}$  is positive (phase angle of  $0^\circ$ ). On the other hand, if  $F_{100}$  and  $F_{221}$  have opposite signs, it is probable  $F_{321}$  will have a negative sign (phase angle of  $180^\circ$ ). If the phases of two of the reflections are known, then the phase of the third reflection is determined by them. As the structure solution progresses, the phases from further strong reflections can be estimated with high probability from the known phases in the starting subset. The structure solution will then give the atomic coordinates for each atom in the asymmetric unit. One further step is structure refinement, the optimization of the values of the coordinates and other crystal parameters with the goal to minimize the differences between the calculated and the observed structure factors. This can be done by a least-squares iteration as before.

Small molecular weight compound X-ray diffraction measurements were carried out at the Unitat de Cristalografia de la Universitat Autònoma de Barcelona. The program SMART (<http://www.bruker.axs.nl>) was used for data reduction while the structure was solved and refined with SHELXS and SHELXL

(Sheldrick, 1997) implemented in the WinGX suite (Farrugia, 1999). Protein X-ray diffraction experiments were performed at the X11 and X13 beamlines of the EMBL (European Molecular Biology Laboratory)-Outstation at DESY (Deutsches Elektronen Synchrotron), Hamburg, Germany. The software used at different stages of protein structure solution include: MOSFLM/SCALA (Leslie, 1991), DENZO/SCALEPACK (Otwinowski & Minor, 1997), and Automar (Klein & Bartels, 2001). Programs of the CCP4 suite (Collaborative Computational Project, 1994) were used: PHASER (Read, 2001) for molecular replacement and REFMAC (Murshudov et al., 1997) for refinement. MolProbity (Davis et al., 2004) and PROCHECK (Laskowski et al., 1993) were used to check the protein coordinates against reference geometric values. COOT (Emsley & Cowtan, 2004) was employed for visualization of the electron density maps and manual adjustments of the model. Images were generated with Pymol (www.pymol.org). Reference books on the topics covered in this Section include: Massa W (2000). Crystal structure determination; Drenth J (2007). Principles of protein X-ray crystallography.

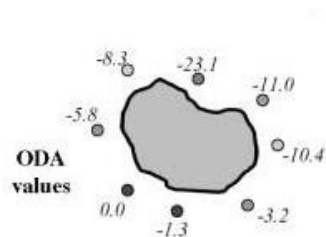
## *8.7 Bioinformatics - the ODA method*

The ODA (Optimal Docking Area) method was used in the search of areas with favorable energy change when buried in a protein-protein association. The ODA method identifies continuous surface patches with optimal desolvation energy based on atomic solvation parameters adjusted for protein-protein docking. We followed the published procedure (Fernández-Recio et al., 2005), with some modifications. First, the center of coordinates of each residue side-chain was used to define the starting points around the protein, instead of using the systematically distributed surface points generated in the original method. Then the docking surface energy of the different-sized patches was calculated at each point. Thus, the docking surface energy of the patch is dependent on the atomic solvation energy contribution of the residues that lie within a sphere of radius  $d$  ( $d = 1, 2, \dots, 20 \text{ \AA}$ ) from the starting point (Fig. 42).



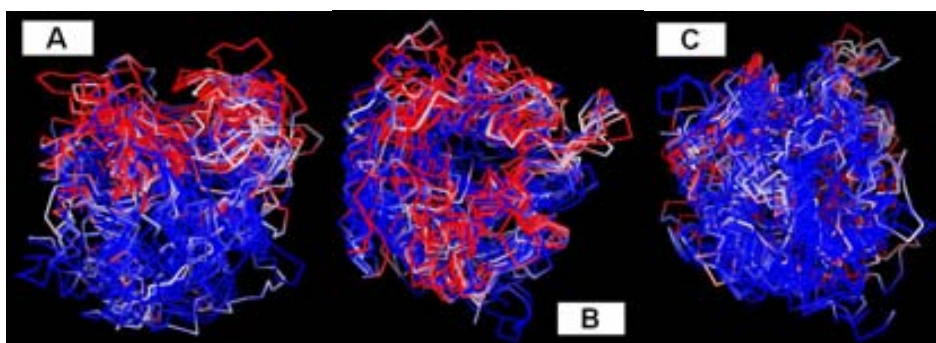
**Figure 42.** Flowchart of ODA analysis (further details in text)

Among all surface patches generated from each starting point, the patch with lowest energy value is called an ODA, and its optimal energy value is assigned to that point, and therefore to the residue from which that point was generated. ODAs were considered only for residues at the surface of the protein, i.e. when the relative solvent-accessible surface area (SASA) of the residue is greater than 0.25 (value obtained in previous internal tests in order to disregard poorly exposed surface residues, which do not significantly contribute to the desolvation effects). Zinc atoms were not considered in the calculations, since they were not significantly solvent-exposed (relative SASA less than 0.25). Finally, the ODA energy values per residue were used to define a region over the protein surface most likely to be involved in protein-protein binding to the partner molecule. An ODA was considered significant when its optimal docking energy value was  $< -10.0$  kcal/mol. Significant ODAs can be graphically displayed on the surface of the proteins and can be used to analyze if they are conserved in the family (Fig. 43).



**Figure 43.** Pictorial representation of ODA values over a generic protein surface

The ODA energy values are then stored in the protein's PDB file, in place of the *B*-factors. Thus, the coloring of the polypeptide chain by energy values is an easy way to assess the tendency of the ODA residues to be clustered around the active site cleft (Fig. 44).



**Figure 44.** The backbone of M14 protease crystal structures colored by ODA energy values, from red (ODA energy value < -10.0 kcal/mol) to blue (ODA energy value = 0.0). Carboxypeptidases were structurally superimposed. The view in panel B is rotated horizontally 90° with respect to panel A, while panel C is rotated vertically 180° with respect to panel B. The center figure shows the active site cleft pointing towards the reader.

The positive predictive value and the global success rate were the two parameters used to assess the performance of the ODA method. For a given unbound molecule, the positive predictive value was defined as the percentage of ODA hot spots (i.e. formed by up to a maximum of ten residues with ODA energy < -10 kcal/mol) that were correctly located in the real protein-protein interface (i.e. that at least one atom of the given residue is located at a distance

of  $< 5 \text{ \AA}$  from any heavy atom of the interacting partner after superimposing the structure of the unbound protein onto the equivalent molecule in the structure of the protein-protein complex). The success rate is the percentage of cases where the positive predictive value is  $\geq 50\%$ , meaning that at least half of the ODA residues are located at the real interface.

Protein interactions were analyzed for the available structures of bound and unbound CPs. The analysis of the protein interactions was performed using the program ICM-Browser, defining the real interface as all residues in one molecule that have at least one atom within a distance of  $5.0 \text{ \AA}$  from any non-hydrogen atom of the interacting partner. We have considered here domain-domain interactions, i.e. between the catalytic domain and its N- or C-terminal domain, as well as protein-protein interactions, i.e. in a complex with a heterologous protein. Whenever necessary, we used SCWRL Version 3.0 program (Canutescu et al., 2003) to reconstruct amino acid residues lacking side chain atoms.

## ***8.8 Bioinformatics - virtual screening***

In all cases, the ArgusDock algorithm was employed as the engine for the screening procedure. ArgusDock, implemented with the program ArgusLab (Planaria Software LLC, UK; [www.arguslab.com](http://www.arguslab.com)) is a freely available software package which has similarities to the commercial packages DOCK and Glide (Joy et al., 2006). A DOCK-type of calculation was used in all instances. The protein targets were either protein structures retrieved from the Brookhaven Protein Data Bank (PDB) or structures generated by homology to a sequence-related model protein. Protein crystal structures at (usually) high resolution were chosen as the target for the screening. The targets for docking are the crystal structures of CPA from *Helicoverpa armigera* (HaCPA) and CPB from *Helicoverpa zea* (HzCPB), which are deposited in the PDB under the codes 1jqg and 2c1c, respectively. Human CPB (1zli) and bovine CPA (2ctc) were the mammalian counterparts. Using the zinc ion at the active site, a docking box centred on it was defined after several trials in search for the optimal box

volume. The latter was found to have typical dimensions of 20Å x 15Å x 20Å. This encompasses all the known subsites for substrate binding, thus allowing a complete sampling of the macromolecule active site cleft. A flexible ligand docking was employed.

The algorithm virtually translates each compound from infinity (that means, at zero interaction energy) to the docking box. Once inside the box, the program translates and rotates the compound in order to find the best possible pose, i.e., the location that allows the interaction energy to reach the minimum value. In the process, the algorithm may rotate the torsional bonds so that an exploration of the conformational space of each compound is performed.

The output of the docking is a list of codenames arranged in ascendent energy values (i.e. from the highest to the lowest negative energy value). This allows for the selection of the most interesting compounds, which then can be “visually” inspected by the user. The frequent crashes of the docking process, that force to reset it from the start, are due to errors in the database records. Therefore, the latter had to be polished by using CORINA Version 3.2 (Gasteiger et al., 1990). This greatly helped in the identification and removal of problematic entries, and eased the whole screening with laboratory bench top PC computers. The screening was performed with ArgusLab and took several days of computation before completion. CORINA was also used to convert the 2D coordinates into 3D compounds (containing hydrogen atoms) in the adequate format recognized by the docking process.

The results obtained from the screening collections were carefully inspected to remove unwanted compounds. The 500 lowest-energy poses were first pre-filtered using the FAFDrugs server (Miteva et al., 2006). The properties monitored included molecular weight, number of hydrogen bond acceptors and donors, number of flexible and rigid bonds, number of carbon and hetero-atoms, estimation of the partition coefficient (log P), and calculation of polar surface area. This automatic procedure was followed by a careful, knowledge-based, manual filtering to yield a selected final set. The knowledge-based manual filtering was performed by visual inspection of the compounds, which were selected from the preference of CPs to bind to short peptides, applying the following general rules:

- i) compounds containing a linear and extended backbone having 3-5 functionalities (i.e. ring systems and/or trigonal planar  $sp^2$ -carbons, joined by single bonds) were preferred as the first choice;
- ii) compounds having bulky systems, like those containing three or more fused rings, and tribenzene-containing molecules, were discarded;
- iii) compounds having stereogenic centres were also disregarded; and
- iv) structures containing a chemical linkage susceptible to be hydrolyzed by the target enzyme were also disliked.

In this way family DF7 of compounds resulted from a careful selection among more than 500 best ranked poses. In the end, 34 compounds were purchased from a commercial supplier (ASINEX Europe BV, The Netherlands; [www.asinex.com](http://www.asinex.com)). The same methodology was employed to identify inhibitors of the metalloprotease 1 from *Trypanosoma cruzi* (TcMCP-1). For the final publication, some selected compounds were docked with AutoGrid implemented with AutoDock program (Goodsell & Olson, 1990). AutoDock allows for a robust and fast search of docking space in the prediction of the interaction energy between the substrate and the macromolecule. For energy evaluation, the macromolecule is first divided into a collection of small boxes (a grid) and then a probe atom is placed at each grid point. The energy of interaction between the probe atom with the protein is assigned to the grid point. An affinity grid is calculated for each type of atom in the substrate together with an electrostatic grid. The energy of the interaction is found by interpolation of the substrate affinity grid and the grid points surrounding each atom in the substrate.

## 8.9 Bioinformatics - resources

This is an account of some computer programs and world wide web based servers that were employed throughout (Table 17).

<b>Table 17</b>		
Name	Application	URL
ExpASy Proteomics Server	Sequence analysis tools, similarity searches, homology modelling, etc.	<a href="http://www.expasy.ch/">http://www.expasy.ch/</a>
Aggrescan	Prediction of "hot spots" of aggregation in polypeptides	<a href="http://bioinf.uab.es/aggrescan/">http://bioinf.uab.es/aggrescan/</a>
Molprobity	Evaluation of X-ray structure quality	<a href="http://molprobity.biochem.duke.edu/">http://molprobity.biochem.duke.edu/</a>
FAF-Drugs	ADME/Tox	<a href="http://bioserv.rpbs.univ-paris-diderot.fr/Help/FAFDrugs.html">http://bioserv.rpbs.univ-paris-diderot.fr/Help/FAFDrugs.html</a>
SABBAC version 1.2	Reconstruction of the complete protein structure from $\alpha$ -carbons	<a href="http://bioserv.rpbs.jussieu.fr/cgi-bin/SABBAC">http://bioserv.rpbs.jussieu.fr/cgi-bin/SABBAC</a>
HingeProt	Detection of hinge sites in proteins	<a href="http://www.prc.boun.edu.tr/appserv/prc/hingeprot/">http://www.prc.boun.edu.tr/appserv/prc/hingeprot/</a>
CCP4	A suite of programs to determine macromolecular 3D structures	<a href="http://www.ccp4.ac.uk/">http://www.ccp4.ac.uk/</a>
COOT	Model building, model completion and validation	<a href="http://www.yesbl.york.ac.uk/~emsey/cool/">http://www.yesbl.york.ac.uk/~emsey/cool/</a>
PyMol	Molecular visualization	<a href="http://www.pymol.org/">http://www.pymol.org/</a>
ArgusLab	Molecular modeling, graphics, and drug design	<a href="http://www.arguslab.com/">http://www.arguslab.com/</a>
CORINA	Generation of 3D structures of small molecules	<a href="http://www.mol-net.com">http://www.mol-net.com</a>
Automar	Image processing, data reduction	<a href="http://software.marresearch.de/automar/index.htm">http://software.marresearch.de/automar/index.htm</a>
AutoDOCK	Molecular docking	<a href="http://autodock.scripps.edu/">http://autodock.scripps.edu/</a>
ChemDraw	Creation of structures from chemical names	<a href="http://www.cambridgesoft.com/">http://www.cambridgesoft.com/</a>
GraphPad	Enzyme kinetics data analysis	<a href="http://www.graphpad.com/">http://www.graphpad.com/</a>
EdiSDF	Visualization, management and edition of SDF molecular files	<a href="http://infochim.u-strasbg.fr/">http://infochim.u-strasbg.fr/</a>
WinGX	Solving, refining and analysing single crystal X-ray diffraction data for small molecules	<a href="http://www.chem.gla.ac.uk/~louis/software/wingx/">http://www.chem.gla.ac.uk/~louis/software/wingx/</a>
MEROPS	Protease database	<a href="http://merops.sanger.ac.uk/">http://merops.sanger.ac.uk/</a>
PDB	Brookhaven Protein Data Bank	<a href="http://www.rcsb.org/pdb/">http://www.rcsb.org/pdb/</a>
ADIT	Protein 3D structure validation and deposition	<a href="http://www.rcsb.org/adit/">http://www.rcsb.org/adit/</a>



# Bibliography



# Bibliography

Adler M, Bryant J, Buckman B, Islam I, Larsen B, Finster S, Kent L, May K, Mohan R, Yuan S & Whitlow M. (2005). Crystal structures of potent thiol-based inhibitors bound to carboxypeptidase B. *Biochemistry* 44:9339-9347

Adler M, Buckman B, Bryant J, Chang Z, Chu K, Emayan K, Hrvatin P, Islam I, Morser J, Sukovich D, West C, Yuan S, Whitlow M. (2008). Structures of potent selective peptide mimetics bound to carboxypeptidase B. *Acta Crystallographica Section D: Biological Crystallography* 64:149-157

Allerton CMN, Blagg J, Bunnage ME, Steele J. (2002). Substituted imidazoles as TAFIA inhibitors. Patent WO 02/14285

Allerton CMN, Bull DJ, Bunnage ME, Maguire RJ, Steele J. (2003a). 3-(imidazolyl)-2-alkoxypropanoic acids as TAFIA inhibitors. Patent WO 03/061652

Allerton CMN, Bunnage ME, Steele J. (2003b). 3-(imidazolyl)-2-aminopropanoic acids for use as TAFI-A inhibitors for the treatment of thrombotic diseases. Patent WO 03/061653

Anand K, Pallarès I, Valnickova Z, Christensen T, Vendrell J, Wendt KU, Schreuder HA, Enghild JJ, Avilés FX. (2008). The crystal structure of thrombin-activable fibrinolysis inhibitor (TAFI) provides the structural basis for its intrinsic activity and the short half-life of TAFIa. *Journal of Biological Chemistry* 283:29416-29423

Anne C, Fournie-Zaluski M-C, Roques BP, Cornille F. (1998). Solid phase synthesis of peptoid derivatives containing a free C-terminal carboxylate. *Tetrahedron Letters* 39:8973-8974

Appelros S, Thim L, Borgström A. Activation peptide of carboxypeptidase B in serum and urine in acute pancreatitis. *Gut* 42:97-102

Arolas JL, Popowicz GM, Lorenzo J, Sommerhoff CP, Huber R, Avilés FX, Holak TA. (2005). The three-dimensional structures of tick carboxypeptidase inhibitor in complex with A/B carboxypeptidases reveal a novel double-headed binding mode. *Journal of Molecular Biology* 350:489-498

Arolas JL, Vendrell J, Avilés FX, Fricker LD. (2007). Metallo-carboxypeptidases: emerging drug targets in biomedicine. *Current Pharmaceutical Design* 13:349-366

Asai S, Sato T, Tada T, Miyamoto T, Kimbara N, Motoyama N, Okada H, Okada N. (2004). Absence of procarboxypeptidase R induces complement-mediated lethal inflammation in lipopolysaccharide-primed mice. *Journal of Immunology* 173:4669-4674

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

Avilés Puigvert FX, Lorenzo Rivera J, Rodríguez-Vera M, Querol Murillo E, Bautista Marugan M, Díez Martín A, Bautista Santa Cruz JM. (2008). Therapeutic agents for treatment of malaria. Patent WO 2008/077977

Bajzar L, Manuel R, Nesheim ME. (1995). Purification and characterization of TAFI, a thrombin-activable fibrinolysis inhibitor. *Journal of Biological Chemistry* 270:14477-14484

Bal W, Bertini I, Kozlowski H, Monnanni R, Scozzafava A, Siatecki GZ. (1990). The unusual behavior of the inhibitor S(+)(1-amino-2-phenylethyl)phosphonic acid towards carboxypeptidase A. *Journal of Inorganic Biochemistry* 40:227-235

Barrow JC, Nantermet PG, Stauffer SR, Ngo PL, Steinbeiser MA, Mao S-S, Carroll SS, Bailey C, Colussi D, Bosserman M, Burlein C, Cook JJ, Sitko G, Tiller PR, Miller-Stein CM, Rose M, McMasters DR, Vacca JP, Selnick HG. (2003). Synthesis and evaluation of Imidazole acetic acid inhibitors of activated thrombin-activatable

fibrinolysis inhibitor as novel antithrombotics. *Journal of Medicinal Chemistry* 46:5294-5297

Bartlett PA, Marlowe CK, Giannousis PP, Hanson JE. (1987). Phosphorus-containing peptide analogs as peptidase inhibitors. *Cold Spring Harbor Symposia on Quantitative Biology* 52:83-90

Bayés A, Sonnenschein A, Daura X, Vendrell J & Avilés FX (2003). Procarboxypeptidase A from the insect pest *Helicoverpa armigera* and its derived enzyme. Two forms with new functional properties. *European Journal of Biochemistry* 270:3026-3035

Bayés A, Comellas-Bigler M, Rodríguez de la Vega M, Maslos K, Bode W, Avilés FX, Jongsma MA, Beekwilder J & Vendrell J. (2005). Structural basis of the resistance of an insect carboxypeptidase to plant protease inhibitors. *Proceedings of the National Academy of Sciences of the USA* 102:16602-16607

Berger J, Schneeman BO. (1986). Stimulation of bile-pancreatic zinc, protein and carboxypeptidase secretion in response to various proteins in the rat. *Journal of Nutrition* 116:265-272

Berger J, Schneeman BO. (1988). Intestinal zinc and carboxypeptidase A and B activity in response to consumption of test meals containing various proteins by rats. *Journal of Nutrition* 118:723-728

Bernkop-Schnürch A, Schmitz T. (2007). Presystemic metabolism of orally administered peptide drugs and strategies to overcome it. *Current Drug Metabolism* 8:509-517

Biloivan OA, Verevka SV, Dzyadevych SV, Jaffrezic-Renault N, Zine N, Bausells J, Samitier J, Errachid A. (2006). Protein detection based on microelectrodes with the PPy[3,3-Co(1,2-C<sub>2</sub>B<sub>9</sub>H<sub>11</sub>)]<sub>2</sub> solid contact and immobilized proteinases: preliminary investigations. *Materials Science and Engineering C* 26:574-577

Björquist P, Buchanan M, Campitelli M, Carroll A, Hyde E, Neve J, Polla M, Quinn R. (2005). Use of cyclic anabaenopeptin-type peptides for the treatment of a condition wherein inhibition of carboxypeptidase U is beneficial, novel anabaenopeptin derivatives and intermediates thereof. Patent WO 2005/039617

Borgström A, Regnèr S. (2005). Active carboxypeptidase B is present in free form in serum from patients with acute pancreatitis. *Pancreatology* 5:530-536

Breckenridge RJ, Suckling CJ. (1986). Enzyme inhibition by electrophilic cyclopropane derivatives. *Tetrahedron* 42:5665-5677

Brown KS, Kingsbury WD, Hall NM, Dunn GL, Gilvarg C. (1987). Determination of carboxypeptidase A using *N*-acetyl-phenylalanyl-3-thiaphenylalanine as substrate: application to a direct serum assay. *Analytical Biochemistry* 161:219-225

Brown S, Bernardo MM, Li Z-H, Kotra LP, Tanaka Y, Fridman R, Mobashery S. (2000). Potent and Selective Mechanism-Based Inhibition of Gelatinases. *Journal of the American Chemical Society* 122:6799-6800

Buchanan MS, Carroll AR, Edser A, Sykes M, Fechner GA, Forster PI, Guymier GP, Quinn RJ. (2008). Lysianadioic acid, a carboxypeptidase B inhibitor from *Lysiana subfalcata*. *Bioorganic & Medicinal Chemistry Letters* 18:1495-1497

Burgos FJ, Salva M, Villegas V, Soriano F, Méndez E, Avilés FX. (1991). Analysis of the activation process of porcine procarboxypeptidase B and determination of the sequence of its activation segment. *Biochemistry* 30:4082-4089

Byers LD, Wolfenden R. (1972). A Potent Reversible Inhibitor of Carboxypeptidase A. *Journal of Biological Chemistry* 247:606-608

Byers LD, Wolfenden R. (1973). Binding of the By-product Analog Benzylsuccinic Acid by Carboxypeptidase A. *Biochemistry* 12:2070-2078

Campbell W, Okada H. (1989). An arginine specific carboxypeptidase generated in blood during coagulation or inflammation which is unrelated to carboxypeptidase N or its subunits. *Biochemical and Biophysical Research Communications* 162:933-939

- Cannell RJP, Kellam SJ, Owsianka AM, Walker JM. (1988). Results of a large scale screen of microalgae for the production of protease inhibitors. *Planta Medica* 54:10-14
- Canutescu AA, Shelenkov AA & Dunbrack RL Jr. (2003). A graph theory algorithm for protein side-chain prediction. *Protein Science* 12:2001-2010
- Castillo JA, Infante MR, Manresa A, Vinardell MP, Mitjans M, Clapés P. (2006). Chemoenzymatic synthesis and antimicrobial and haemolytic activities of amphiphilic bis(phenylacetylarginine) derivatives. *ChemMedChem* 1:1091-1098
- Chen H, Meyer J. (2001). Method of detecting risk of type II diabetes based on mutations found in carboxypeptidase E. Patent US 2001/0024792
- Chen C-C, Lee K-D, Gau J-P, Yu Y-B, You J-Y, Lee S-C, Hsu Y-B, Chau W-K, Ho C-H. (2005). Plasma antigen levels of thrombin-activatable fibrinolysis inhibitor did not differ in patients with or without disseminated intravascular coagulation. *Annals of Hematology* 84:675-680
- Cheng T-L, Roffler S. (2008). Membrane-tethered proteins for basic research, imaging, and therapy. *Medicinal Research Reviews* 28:885-928
- Cheng Y, Prusoff WH. (1973). Relationship between the inhibition constant (K<sub>1</sub>) and the concentration of inhibitor which causes 50 per cent inhibition (I<sub>50</sub>) of an enzymatic reaction. *Biochemical Pharmacology* 22:3099-3108
- Cho JH, Kim DH, Lee KJ, Kim DH, Choi KY. (2001). The role of Tyr248 probed by mutant bovine carboxypeptidase A: insight into the catalytic mechanism of carboxypeptidase A. *Biochemistry* 40:10197-10203
- Christianson DW, Lipscomb WN. (1989). Carboxypeptidase A. *Accounts on Chemical Research* 22:62-69
- Chung SJ, Chung S, Lee HS, Kim E-J, Oh KS, Choi HS, Kim KS, Kim YJ, Hahn JH, Kim DH. (2001). Mechanistic insight into the inactivation of carboxypeptidase A by  $\alpha$ -benzyl-2-oxo-1,3-oxazolidine-4-acetic acid, a novel type of irreversible inhibitor for carboxypeptidase A with no stereospecificity. *Journal of Organic Chemistry* 66:6462-6471
- Coll M, Guasch A, Avilés FX, Huber R. (1991). Three-dimensional structure of porcine procarboxypeptidase B: a structural basis of its inactivity. *EMBO Journal* 10:1-9
- Collaborative Computational Project, no. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D: Biological Crystallography* 50:760-776
- Cronet P, Knecht W, Malmberg-Hager CA-C, Andersson M, Furebring C. (2005). Carboxypeptidase U (CPU) mutants. Patent WO 2005/052149
- Cronet P, Knecht W, Malmberg-Hager CA-C, Andersson M, Furebring C. (2008). Carboxypeptidase U (CPU) mutants. Patent US 2008/0175834
- Cushman DW, Cheung HS, Sabo EF, Ondetti MA. (1977). Design of potent competitive inhibitors of angiotensin-converting enzyme. Carboxyalkanoyl and mercaptoalkanoyl amino acids. *Biochemistry* 16:5484-5491
- Cushman DW, Ondetti MA. (1999). Design of angiotensin converting enzyme inhibitors. *Nature Medicine* 5:1110-1112
- Davis IW, Murray LW, Richardson JS, Richardson DC. (2004) MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Research* 32:W615-W619
- Deiteren K, Hendriks D, Scharpe S, Lambeir AM. (2009). Carboxypeptidase M: multiple alliances and unknown partners. *Clinica Chimica Acta* 399:24-39
- Delk AS, Durie PR, Fletcher TS, Largman C. (1985). Radioimmunoassay of active pancreatic enzymes in sera from patients with acute pancreatitis. I. Active carboxypeptidase B. *Clinical Chemistry* 31:1294-1300
- Do YH, Gifford-Moore DS, Beight DW, Rathnachalam R, Klimkowski VJ, Warshawsky AM, Lu D. (2005). Inhibition of thrombin activatable fibrinolysis inhibitor by cysteine derivatives. *Thrombosis Research* 116:265-271
- Drenth J. (2007). Principles of protein X-ray crystallography. 3 rd Ed. Springer, New York, USA

- Eaton DL, Malloy BE, Tsai SP, Henzel W, Drayna D. (1991). Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase B from human plasma. *Journal of Biological Chemistry* 266:21833-21838
- Easton CJ, Harper JB, Head SJ, Lee K, Lincoln SF. (2001). Cyclodextrins to limit substrate inhibition and alter substrate selectivity displayed by enzymes. *Journal of the Chemical Society, Perkin Transactions 1*, pp. 584-587
- Edge M, Forder C, Hennam J, Lee I, Tonge D, Hardern I, Fitton J, Eckersley K, East S, Shufflebotham A, Blakey D, Slater A. (1998). Engineered human carboxypeptidase B enzymes that hydrolyse hippuryl-L-glutamic acid: reversed-polarity mutants. *Protein Engineering* 11:1229-1234
- Emekli U, Schneidman-Duhovny D, Wolfson HJ, Nussinov R, Haliloglu T. (2008). HingeProt: automated prediction of hinges in protein structures. *Proteins* 70:1219-1227
- Emsley P, Cowtan K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography* 60:2126-2132
- Erickson RH, Kim YS. (1990). Digestion and absorption of dietary protein. *Annual Review of Medicine* 41:133-139
- Estébanez-Perpiñá E, Bayés A, Vendrell J, Jongsma MA, Bown DP, Gatehouse JA, Huber R, Bode W, Avilés FX & Reverter D. (2001). Crystal structure of a novel mid-gut procarboxypeptidase from the cotton pest *Helicoverpa armigera*. *Journal of Molecular Biology* 313:629-638
- Fan H, Zhao Y, Byers L, Hammer RP. (1999). Synthesis of phosphonopeptide and thiophosphonopeptide analogs as inhibitors of carboxypeptidase A. *American Peptide Symposia, Peptides for the New Millennium*, pp. 91-93
- Farrugia LJ. (1997). In CK Johnson, MN Burnett, "ORTEP-3 for Windows" molecular plotting program. Based on ORTEP-III (Version 1.02). Department of Chemistry, University of Glasgow
- Farrugia LJ. (1999). WinGX suite for small-molecule single-crystal crystallography. *Journal of Applied Crystallography* 32:837-838
- Fernández D, Vendrell J, Avilés FX & Fernández-Recio J. (2007). Structural and Functional Characterization of Binding Sites in Metallo-carboxypeptidases Based on Optimal Docking Area Analysis. *Proteins: Structure, Function, and Bioinformatics* 60:131-144
- Fernández D, Illa O, Avilés FX, Branchadell V, Vendrell J, Ortuño RM. (2008). Thioxophosphoranyl aryl- and heteroaryloxiranes as the representants of a new class of metallo-carboxypeptidase inhibitors. *Bioorganic & Medicinal Chemistry* 14:4823-4828
- Fernández D, Torres E, Avilés FX, Ortuño RM, Vendrell J. (2009a). Cyclobutane-containing peptides: evaluation as novel metallo-carboxypeptidase inhibitors and modelling of their mode of action. *Bioorganic & Medicinal Chemistry* 17:3824-3828
- Fernández D, Avilés FX, Vendrell J. (2009b). Aromatic organic compounds as scaffolds for metallo-carboxypeptidase inhibitor design. *Chemical Biology and Drug Design* 73:75-82
- Fernández D, Avilés FX, Vendrell J. (2009c). A new type of five-membered heterocyclic inhibitors of basic metallo-carboxypeptidase. *European Journal of Medicinal Chemistry* 44:3266-3271
- Fernández-Recio J, Totrov M, Skorodumov C & Abagyan R. (2005). Optimal docking area: a new method for predicting protein-protein interaction sites. *Proteins: Structure, Function, and Bioinformatics* 58:134-43.
- Ferrer M, Zuck P, Kolodin G, Mao SS, Peltier RR, Bailey C, Gardell SJ, Strulovici B, Inglese J. (2003). Miniaturizable homogenous time-resolved fluorescence assay for carboxypeptidase B activity. *Analytical Biochemistry* 317:94-98
- Fialka F, Gruber RM, Hitt R, Opitz L, Brunner E, Schliephake H, Kramer F-J. (2008). CPA6, FMO2, LGI1, SIAT1 and TNC are differentially expressed in early- and late-stage oral squamous cell carcinoma - A pilot study. *Oral Oncology* 44:941-948

- Firth-Clark, S.; Kirton, S. B.; Willems, H. M.; Williams, A. (2008). De novo ligand design to partially flexible active sites: application of the ReFlex algorithm to carboxypeptidase A, acetylcholinesterase, and the estrogen receptor. *Journal of Chemical Information and Modeling* 48:296-305
- Fredholt K, Adrian C, Just L, Larsen DH, Weng S, Moss B, Friis GJ. (2000). Chemical and enzymatic stability as well as transport properties of a leu-enkephalin analogue and ester prodrugs thereof. *Journal of Controlled Release* 63:261-273
- Fricker LD, Plummer TH Jr, Snyder SH. (1983). Enkephalin convertase: potent, selective, and irreversible inhibitors. *Biochemical and Biophysical Research Communications* 111:994-1000
- Fricker LD. (1988). Carboxypeptidase E. *Annual Review of Physiology* 50:309-321
- Friedman M, Grosjean OK, Zahnley JC. (1986). Inactivation of metalloenzymes by food constituents. *Food Chemistry and Toxicology* 24:897-902
- Friis GJ, Bak A, Larsen BD, Frokjaer S. (1996). Prodrugs of peptides obtained by derivatization of the C-terminal peptide bond in order to effect protection against degradation by carboxypeptidases. *International Journal of Pharmaceutics* 136:61-69
- Frye JL, Sebastian JF. (1990). Effect of sulfonates on the esterase activity of carboxypeptidase A. *Biochemistry and Cellular Biology* 68:1062-1065
- Fujii S, Yamagata S, Tanaka K, Wada M, Akai T. (1977). Studies on the diagnostic significance of serum carboxypeptidase A activity in diabetes mellitus. *Japanese Journal of Medicine* 16:106-111
- Fujiwara H, Imai K, Inoue T, Maeda M, Fujii S. (1999). Membrane-bound cell surface peptidases in reproductive organs. *Endocrine Journal* 46:11-25
- Fushihara K, Kikuchi C, Matsushima T, Kanemoto K, Satoh E, Yamamoto T, Suzuki K. (2003). Phosphonic acid derivatives having carboxypeptidase B inhibitory activity. Patent US 6,576,627
- Fushihara K, Kikuchi C, Matsushima T, Kanemoto K, Satoh E, Yamamoto T, Suzuki K. (2004). Phosphonic acid derivative having inhibitory activity against carboxypeptidase B. Patent US 6,737,416
- Galvão CMA, Pinto GA, Jesus CDF, Giordano RC, Giordano RLC. (2009). Producing a phenylalanine-free pool of peptides after tailored enzymatic hydrolyses of cheese whey. *Journal of Food Engineering* 91:109-117
- García-Carreño FL & Navarrete del Toro MA. (1997). Classification of proteases without tears. *Biochemical Education* 25:161-167
- Gasteiger J, Rudolph C, Sadowski J. (1990). Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Computer Methodology* 3:537-547
- Geokas MC, Wollesen F, Rinderknecht H. (1974). Radioimmunoassay for pancreatic carboxypeptidase B in human serum. *Journal of Laboratory and Clinical Medicine* 84:574-583
- Ghosh SS, Wu Y-Q, Mobashery S. (1991). Peptidic mechanism-based inactivators for carboxypeptidase A. *Journal of Biological Chemistry* 266:8759-8764
- Gilvarg C. (1998). Method of detecting procarboxypeptidase A and carboxypeptidase A levels in biological fluids. Patent WO 98/45471
- Gladner JA, Folk JE. (1958). Carboxypeptidase B. II. Mode of action on protein substrates and its application to carboxyl terminal group analysis. *Journal of Biological Chemistry* 231:393-401
- Goldstein SM, Kaempfer CE, Kealey JT, Wintroub BU. (1989). Human mast cell carboxypeptidase. Purification and characterization. *Journal of Clinical Investigation* 83:1630-1636
- Golz S, Brüggemeier U, Geerts A. (2006). Diagnostics and therapeutics for diseases associated with carboxypeptidase A2 (CPA2). Patent WO 2006/010499
- Goodsell DS, Olson AJ. (1990). Automated docking of substrates to proteins by simulated annealing. *Proteins* 8:195-202

- Greenfield RS, An SSA. (2004). Method of preparation of stabilized thrombin-activatable fibrinolysis inhibitor (TAFI) and methods of use thereof. Patent US 6,808,927
- Greenwood NN & Earnshaw A. (1997). Chapter 9, Silicon, pp. 328-366, in *Chemistry of the Elements*, 2nd Ed. Butterworth-Heinemann, Oxford, UK
- Grobelny D, Goli UB, Galardy RE. (1985). 3-phosphonopropionic acids inhibit carboxypeptidase A as multisubstrate analogues or transition-state analogues. *Biochemical Journal* 232:15-19
- Guimarães AHC, Laurens N, Weijers EM, Koolwijk P, van Hinsbergh VWM, Rijken DC. (2007). TAFI and Pancreatic Carboxypeptidase B Modulate In Vitro Capillary Tube Formation by Human Microvascular Endothelial Cells. *Arteriosclerosis, Thrombosis, and Vascular Biology* 27:2157-2162
- Hadkar V, Skidgel RA. (2001). Carboxypeptidase D is up-regulated in RAW 264.7 macrophages and stimulates nitric oxide synthesis by cells in arginine-free medium. *Molecular Pharmacology* 59:1324-1332
- Hamada T, Matsugi S, Tanaka T. (2006). Antibodies, assay method by using them and judgment method for pancreas cancer. Patent US 2006/0073526
- Han MS, Kim DH. (2001). Effect of zinc ion on the inhibition of carboxypeptidase A by imidazole-bearing substrate analogues. *Bioorganic & Medicinal Chemistry Letters* 11:1425-1427
- Han MS, Ryu CH, Chung SJ, Kim DH. (2000). A novel strategy for designing irreversible inhibitors of metalloproteases: acetals as latent electrophiles that interact with catalytic nucleophile at the active site. *Organic Letters* 2:3149-3152
- Hanson JE, Kaplan AP, Bartlett PA. (1989). Phosphonate analogues of carboxypeptidase A substrates are potent transition-state analogue inhibitors. *Biochemistry* 28:6294-6305
- Hasegawa H, Shinohara Y. (1998). Synthesis of (2S)-2-amino-3-(2',6'-dibromo-4'-hydroxy)phenyl-propionic acid (2,6-dibromo-L-tyrosine). *Journal of the Chemical Society, Perkin Transactions 1*, pp. 243-247
- Hendriks D, Wang W, Scharpe S, Lommaert M-P, van Sande M. (1990). Purification and characterization of a new arginine carboxypeptidase in human serum. *Biochimica et Biophysica Acta* 1034:86-92
- Henningsson F, Yamamoto K, Saftig P, Reinheckel T, Peters C, Knight SD, Pejler G. (2005). A role for cathepsin E in the processing of mast-cell carboxypeptidase A. *Journal of Cell Science* 118:2035-2042
- Hergenrother PJ. (2006). Obtaining and screening compound collections: a user's guide and a call to chemists. *Current Opinion in Chemical Biology* 10:213-218
- Hill JM, Lowe G. (1995). Synthesis of phosphorothioate esters of L-phenyl-lactic acid as transition-state inhibitors of carboxypeptidase A. *Journal of the Chemical Society, Perkin Transactions 1*, pp. 2001-2007
- Holmquist B, Vallee BL. (1979). Metal-coordinating substrate analogs as inhibitors of metalloenzymes. *Proceedings of the National Academy of Sciences of the USA* 76:6216-6220
- Huennekens FM. (1997). Development of methotrexate  $\alpha$ -peptides as prodrugs for activation by enzyme-monoclonal antibody conjugates. *Advances in Enzyme Regulation* 37:77-92
- Huijghebaert SM, Hofmann AF. (1986). Pancreatic carboxypeptidase hydrolysis of bile acid-amino acid conjugates: selective resistance of glycine and taurine amidates. *Gastroenterology* 90:306-315
- Husbands S, Suckling CA, Suckling CJ. (1994). Latent inhibitors part 10. The inhibition of carboxypeptidase A by tetrapeptide analogues based on 1-aminocyclopropane carboxylic acid. *Tetrahedron* 50:9729-9742
- Ibarz A, Garza S, Pagan J. (2008). Inhibitory effect of melanoidins from glucose-asparagine on carboxypeptidase activity. *European Food Research and Technology* 226:1277-1282



- Ikejiri M, Bernardo MM, Meroueh SO, Brown S, Chang M, Fridman R, Mobashery S. (2005). Design, synthesis, and evaluation of a mechanism-based inhibitor for gelatinase A. *Journal of Organic Chemistry* 70:5709-5712
- Inamori Y, Shinohara S, Tsujibo H, Okabe T, Morita Y, Sakagami Y, Kumeda Y & Ishida N. (1998). Antimicrobial activity and metalloprotease inhibition of hinokitiol-related compounds, the constituents of *Thujopsis dolabrata* S. and *Z. Hondai* MAK. *Biological and Pharmaceutical Bulletin* 22:990-993
- Itou Y, Suzuki S, Ishida K & Murakami M. (1999). Anabaenopeptins G and H, Potent Carboxypeptidase A Inhibitors from the Cyanobacterium *Oscillatoria agardhii* (NIES-595). *Bioorganic & Medicinal Chemistry Letters* 9:1243-1246
- Jacobsen NE, Bartlett PA. (1981). A phosphoramidate dipeptide analogue as an inhibitor of carboxypeptidase A. *Journal of the American Chemical Society* 103:654-657
- Janin J. (2000). Kinetics and thermodynamics of protein-protein interactions. In: Kleanthous C, editor. Protein-Protein Recognition. Volume 31, Frontiers in Molecular Biology. New York: Oxford University Press, Inc.; pp. 1-32
- Jeffrey KD, Alejandro EU, Luciani DS, Kalynyak TB, Hu X, Li H, Lin Y, Townsend RR, Polonsky KS, Johnson JD. (2008). Carboxypeptidase E mediates palmitate-induced  $\beta$ -cell ER stress and apoptosis. *Proceedings of the National Academy of Sciences of the USA* 105:8452-8457
- Joy S, Nair PS, Hariharan R, Pillai MR. (2006). Detailed comparison of the protein-ligand docking efficiencies of GOLD, a commercial package and ArgusLab, a licensable freeware. *In Silico Biology* 6:601-605
- Kakimoto N, Katayama T, Hazato T, Ohnishi T. (1988). Opioid peptide-degrading enzyme inhibitor with a germanium compound. Patent US 4,748,187
- Kallus C, Heitsch H, Lindenschmidt A, Grueneberg S. (2005). Imidazole derivatives used as TAFIA inhibitors. Patent WO 2005/105781
- Kallus C, Heitsch H, Wehner V. (2007). Imidazole derivatives as inhibitors of TAFIA. Patent WO 2007/045339
- Kamo M, Tsugita A. (1987). Carboxypeptidase digestion in the presence of detergents. *Journal of Biochemistry* 102:243-246
- Kaplan AP, Bartlett PA. (1991). Synthesis and evaluation of an inhibitor of carboxypeptidase A with a  $K_i$  value in the femtomolar range. *Biochemistry* 30:8165-8170
- Katzav-Gozansky T, Hanan E, Solomon B. (1996). Effect of monoclonal antibodies in preventing carboxypeptidase A aggregation. *Biotechnology and Applied Biochemistry* 23:227-230
- Kazmierczak SC, van Lente F. (1989). Measuring carboxypeptidase A activity with a centrifugal analyzer: analytical and clinical considerations. *Clinical Chemistry* 35:251-255
- Kemp A, Ner SK, Flees L, Suckling CJ, Tedford MC, Bell AR, & Wrigglesworth R. (1993). Latent Inhibitors. Part 9. Substrate Activated Time-dependent Inhibition of Carboxypeptidase A by Aminocyclopropanecarboxylic Acid Derivatives and Analogues. *Journal of the Chemical Society, Perkin Transactions 2*, pp. 741-748
- Kim DH. (2001). Origin of chiral pharmacology: stereochemistry in metalloprotease inhibition. *Mini Reviews in Medicinal Chemistry* 1:155-161
- Kim DH. (2004). Chemistry-based design of inhibitors for carboxypeptidase A. *Current Topics in Medicinal Chemistry* 4:1217-1226
- Kim DH, Chung SJ. (1995). Inactivation of carboxypeptidase A by 2-benzyl-3,4-epithiobutanoic acid. *Bioorganic & Medicinal Chemistry Letters* 5:1667-1672
- Kim DH, Chung SJ, Kim E-J, Tian GR. (1998). Irreversible inhibition of zinc-containing protease by oxazolidinone derivatives. Novel inactivation chemistry. *Bioorganic & Medicinal Chemistry Letters* 8:859-864
- Kim DH, Kim KB. (1991). The function of S1' subsite pocket of carboxypeptidase A. *Bioorganic & Medicinal Chemistry Letters* 1:323-326

- Kim DH, Kim YJ. (1993). Synthesis and inhibitory activity of optically active 2-benzyl-3-mercaptopropanoic acid against carboxypeptidase A. *Bioorganic & Medicinal Chemistry Letters* 3:2681-2684
- Kim DH, Kim YM, Li Z-H, Kim KB, Choi SY. (1994). A new type of carboxypeptidase A inhibitor: design, synthesis, and mechanistic implication. *Pure & Applied Chemistry* 66:721-728
- Kim DH, Lee KJ. (1997). *O*-(hydroxyacetyl)-L- $\beta$ -phenyllactic acid as a new type of mechanism-based inactivator for carboxypeptidase A. *Bioorganic & Medicinal Chemistry Letters* 7:2607-2612
- Kim DH, Mobashery S. (2001). Mechanism-based inhibition of zinc proteases. *Current Medicinal Chemistry* 8:959-965
- Kim DH, Shin YS, Kim KB. (1991). The structural feature of S1' subsite of carboxypeptidase A. *Bioorganic & Medicinal Chemistry Letters* 1:317-322
- Kim H, Lipscomb WN. (1990) Crystal structure of the complex of carboxypeptidase A with a strongly bound phosphonate in a new crystalline form: comparison with structures of other complexes. *Biochemistry* 29:5546-5555
- Kim H, Lipscomb WN. (1991). Comparison of the structures of three carboxypeptidase A-phosphonate complexes determined by X-ray crystallography. *Biochemistry* 30:8171-8180
- Kim YM, Kim DH. (1996). Convenient Preparation of All Four Possible Stereoisomers of 2-Benzyl-3,4-epoxybutanoic Acid, Pseudomechanism-based inactivator for Carboxypeptidase A via  $\alpha$ -Chymotrypsin-Catalyzed Hydrolysis. *Bulletin of the Korean Chemical Society* 17:967-969
- Kleifeld O, Kotra LP, Gervasi DC, Brown S, Bernardo MM, Fridman R, Mobashery S, Sagi I. (2001). X-ray absorption studies of human matrix metalloproteinase-2 (MMP-2) bound to a highly selective mechanism-based inhibitor. comparison with the latent and active forms of the enzyme. *Journal of Biological Chemistry* 276:17125-17131
- Klein C, Bartels K. (2001). Automar - GUI for image processing. Version 2.0.1. Marresearch GmbH, Germany
- Kodani S, Suzuki S, Ishida K & Murakami M. (1999). Five new cyanobacterial peptides from water bloom materials of lake Teganuma (Japan). *FEMS Microbiology Letters* 178:343-348
- Kojima S, Makihira T, Funakoshi T, Shimada H. (1994). Effect of terbium on protease activity in pancreas of mice. *Research Communications in Molecular Pathology and Pharmacology* 85:227-235
- Komuro T, Kakimoto N, Katayama T, Hazato T. (1986). Inhibitory effects of Ge-132 (carboxyethyl germanium sesquioxide) derivatives on enkephalin-degrading enzymes. *Biotechnology and Applied Biochemistry* 8:379-386
- Krämer R. (1999). Bioinorganic models for the catalytic cooperation of metal ions and functional groups in nuclease and peptidase enzymes. *Coordination Chemistry Reviews* 182:243-261
- Krumrine J, Raubacher F, Brooijmans N, Kuntz I. (2003). Principles and methods of docking and ligand design, pp. 443-476, in *Structural Bioinformatics, Methods of Biochemical Analysis Vol. 44*. Bourne PE, Weissig H, Eds. Wiley-Liss, Inc., Hoboken, New Jersey
- Kusano G, Takahira M, Shibano M, Kusano A, Okamoto Y, Tsujibo H, Numata A & Inamori Y. (1998). Studies on inhibitory activities of fukiic acid esters on germination,  $\alpha$ -amylase, and carboxypeptidase A. *Biological and Pharmaceutical Bulletin* 21:997-999
- Kylänpää-Bäck M-L, Kempainen E, Puolakkainen P. (2002). Trypsin-based laboratory methods and carboxypeptidase activation peptide in acute pancreatitis. *Journal of Pancreas* (Online) 3:34-48
- Langguth P, Bohner V, Heizmann J, Merkle HP, Wolfram S, Amidon GL, Yamashita S. (1997). The challenge of proteolytic enzymes in the intestinal peptide delivery. *Journal of Controlled Release* 46:39-57

- Laskowski RA, MacArthur MW, Hutchinson E. & Thornton, JM. (1993). PROCHECK: a program to check the stereochemical quality of protein structures *Journal of Applied Crystallography* 26:283-291
- Lavis LD, Raines RT. (2008). Bright ideas for chemical biology. *ACS Chemical Biology* 3:142-155
- Lazoura E, Campbell W, Yamaguchi Y, Kato K, Okada N, Okada H. (2002). Rational structure-based design of a novel carboxypeptidase R inhibitor. *Chemistry & Biology* 9:1129-1139
- Lee HS, Kim DH. (2003). Synthesis and evaluation of  $\alpha,\alpha$ -disubstituted-3-mercaptopropanoic acids as inhibitors for carboxypeptidase A and implications with respect to enzyme inhibitor design. *Bioorganic & Medicinal Chemistry* 11:4685-4691
- Lee KJ, Joo KC, Kim E-J, Lee M, Kim DH. (1997). A new type of carboxypeptidase A inhibitors designed using an imidazole as a zinc coordinating ligand. *Bioorganic & Medicinal Chemistry* 5:1989-1998
- Lee KJ, Kim DH. (1996). Inactivation of a prototypic zinc-containing protease with (S)-2-benzyl-2-(2-oxo-2-isoxazolidinyl)acetic acid. *Bioorganic & Medicinal Chemistry Letters* 6:2431-2436
- Lee M, Heseck D, Shi Q, Noll BC, Fisher JF, Chang M, Mobashery S. (2008). Conformational analyses of thiirane-based gelatinase inhibitors. *Bioorganic & Medicinal Chemistry Letters* 18:3064-3067
- Lee M, Kim DH. (2002). Synthesis and kinetic evaluation of racemic and optically active 2-benzyl-2-methyl-3,4-epoxybutanoic acids as irreversible inactivators for carboxypeptidase. *Bioorganic & Medicinal Chemistry* 10:913-922
- Lee SS, Li Z-H, Lee DH, Kim DH. (1995). (2R,3S)- and (2S,3R)-2-benzyl-3,4-epoxybutanoic acid as highly efficient and fast acting pseudomechanism-based inactivators for carboxypeptidase A: design, asymmetric synthesis and inhibitory kinetics. *Journal of the Chemical Society, Perkin Transactions 1: Organic and Bio-Organic Chemistry*, pp. 2877-2282
- Leebeek FWG, van Goor MPJ, Guimarães AHC, Brouwers G-J, de Maat MPM, Rijken DC. (2005). High functional levels of thrombin-activatable fibrinolysis inhibitor are associated with an increased risk of first ischemic stroke. *Journal of Thrombosis and Haemostasis* 3:2211-2218
- Leslie AGW. (1991). Molecular data processing. In Moras D, Podjarny AD and Thierry JC. (eds), *Crystallographic Computing 5*. Oxford University Press, Oxford, pp. 50-61.
- Leung LL, Myles T, Nishimura T, Song JJ, Robinson WH. (2008). Regulation of tissue inflammation by thrombin-activatable carboxypeptidase B (or TAFI). *Molecular Immunology* 45:4080-4083
- Levey DJ, Place AR, Rey PJ, Martínez del Río C. (1999). An experimental test of dietary enzyme modulation in pine warblers *Dendroica pinus*. *Physiological and Biochemical Zoology* 72:576-587
- Liebman MN, Venanzi CA, Weinstein H. (1985). Structural analysis of carboxypeptidase A and its complexes with inhibitors as a basis for modeling enzyme recognition and specificity. *Biopolymers* 24:1721-1758
- Lipscomb WN, Sträter N. (1996). Recent advances in zinc enzymology. *Chemical Reviews* 96:2375-2433
- Lützel Schwab C, Pejler G, Aveskogh M, Hellman L. (1997). Secretory granule proteases in rat mast cells. Cloning of 10 different serine proteases and a carboxypeptidase A from various rat mast cell populations. *Journal of Experimental Medicine* 185:13-29
- Maeda Y, Takahashi M. (1989). Hydrolysis and absorption of a conjugate of ursodeoxycholic acid with *para*-aminobenzoic acid. *Journal of Pharmacobio-Dynamics* 12:744-753
- Makinen MW, Troyer JM, van der Werff H, Berendsen HJC, van Gunsteren WF. (1989). Dynamical Structure of Carboxypeptidase A. *Journal of Molecular Biology* 207:201-216

- Mao S-S, Colussi D, Bailey CM, Bosserman M, Burlein C, Gardell SJ, Carroll SS. (2003). Electrochemiluminescence assay for basic carboxypeptidases: inhibition of basic carboxypeptidases and activation of thrombin-activatable fibrinolysis inhibitor. *Analytical Biochemistry* 319:159-170
- Marquez-Curtis L, Jalili A, Deiteren K, Shirvaikar N, Lambeir AM, Janowska-Wieczorek A. (2008). Carboxypeptidase M expressed by human bone marrow cells cleaves the C-terminal lysine of stromal cell-derived factor-1alpha: another player in hematopoietic stem/progenitor cell mobilization?. *Stem Cells* 26:1211-1220
- Martin CJ, Evans WJ. (1989). Phytic acid-enhanced metal ion exchange reactions: the effect on carboxypeptidase A. *Journal of Inorganic Biochemistry* 35:267-288
- Martínez-Montemayor MM, Hill GM, Raney NE, Rilington VD, Tempelman RJ, Link JE, Wilkinson CP, Ramos AM, Ernst CW. (2008). Gene expression profiling in hepatic tissue of newly weaned pigs fed pharmacological zinc and phytase supplemented diets. *BMC Genomics* 9:421 (doi:10.1186/1471-2164-9-421)
- Massa W. (2000). Crystal structure determination. 1<sup>st</sup> English Version, Springer Verlag, Berlin, Germany
- Massova I, Martin P, de Mel S, Tanaka Y, Edwards B, Mobashery S. (1996). Crystallographic and Computational Insight on the Mechanism of Zinc-Ion-Dependent Inactivation of Carboxypeptidase A by 2-Benzyl-3-Iodopropanoate. *Journal of the American Chemical Society* 118:12479-12480
- Matsugi S, Hamada T, Shioi N, Tanaka T, Kumada T, Satomura S. (2007). Serum carboxypeptidase A activity as a biomarker for early-stage pancreatic carcinoma. *Clinica Chimica Acta* 378:147-153
- McKay TJ, Phelan AW, Plummer TH Jr. (1979). Comparative studies on human carboxypeptidases B and N. *Archives of Biochemistry and Biophysics* 197:487-492
- McKay TJ, Plummer TH Jr. (1978). By-product analogues for bovine carboxypeptidase B. *Biochemistry* 17:401-405
- Metz M, Galli SJ. (2007). Proteases for treatment of venomous bites. Patent WO 2007/149343
- Miteva MA, Violas S, Montes M, Gomez D, Tuffery P, Villoutreix BO. (2006). FAF-Drugs: free ADME/tox filtering of compound collections. *Nucleic Acids Research* 34(Web Server issue), W738-W744
- Mittl PR, Grütter MG. (2006). Opportunities for structure-based design of protease-directed drugs. *Current Opinion in Structural Biology* 16:769-775
- Miura K, Hosomi A. (2004). Silicon in organic synthesis. Chapter 10, Vol. 2, in Main Group Metals in Organic Synthesis, Yamamoto H & Oshima K, Eds. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, pp. 409-592
- Mobashery S, Ghosh SS, Tamura SY, Kaiser ET. (1990). Design of an effective mechanism-based inactivator for a zinc protease. *Proceedings of the National Academy of Sciences of the USA* 87:578-582
- Mock WL, Liu Y, Stanford DJ. (1996). Arazoformyl peptide surrogates as spectrophotometric kinetic assay substrates for carboxypeptidase A. *Analytical Biochemistry* 239:218-222
- Mock WL, Stanford DJ. (2002). Anisylazobenzylarginine: a superior assay substrate for carboxypeptidase B type enzymes. *Bioorganic & Medicinal Chemistry Letters* 12:1193-1194
- Moeller C, Swindell EC, Kispert A, Eichele G. (2003). Carboxypeptidase Z (CPZ) modulates Wnt signaling and regulates the development of skeletal elements in the chicken. *Development* 130:5103-5111
- Montaner J, Ribó M, Monasterio J, Molina CA, Alvarez-Sabín J. (2003). Thrombin-activatable fibrinolysis inhibitor levels in the acute phase of ischemic stroke. *Stroke* 34:1038-1040
- Morán C, Clapés P, Comelles F, García T, Pérez L, Vinardell P, Mitjans M, Infante MR. (2001). Chemical structure/property relationship in single-chain arginine surfactants. *Langmuir* 17:5071-5075

- Moriguchi M, Umeda Y, Miyazaki K, Nakamura T, Ogawa K, Kojima F, Inuma H, Aoyagi T. (1988). Synthesis of histargin and related compounds and their inhibition of enzymes. *Journal of Antibiotics* 41:1823-1827
- Morita Y, Matsumura E, Tsujibo H, Yasuda M, Sakagami Y, Okabe T, Ishida N, Inamori Y. (2001). Biological activity of  $\alpha$ -thujaplicin, the minor component of *Thujopsis dolabrata* SIEB. et ZUCC. var. *hondai* Makino. *Biological and Pharmaceutical Bulletin* 24:607-611
- Morita Y, Matsumura E, Tsujibo H, Yasuda M, Okabe T, Sakagami Y, Kumeda Y, Ishida N, Inamori Y. (2002). Biological activity of 4-acetyltropolone, the minor component of *Thujopsis dolabrata* SIEB. et ZUCC. var. *hondai* Mak. *Biological and Pharmaceutical Bulletin* 25:981-985
- Morita Y, Matsumura E, Okabe T, Shibata M, Sugiura M, Ohe T, Tsujibo H, Ishida N, Inamori Y. (2003). Biological activity of tropolone. *Biological and Pharmaceutical Bulletin* 26:1487-1490
- Mosnier LO, Bouma BN. (2006). Regulation of fibrinolysis by thrombin activatable fibrinolysis Inhibitor, an unstable carboxypeptidase B that unites the pathways of coagulation and fibrinolysis. *Arteriosclerosis, Thrombosis, and Vascular Biology* 26:2445-2453
- Mueller-Ortiz SL, Wang D, Morales JE, Li L, Chang J-Y, Wetsel RA. (2009). Targeted disruption of the gene encoding the murine small subunit of carboxypeptidase N (CPN1) causes susceptibility to C5a anaphylotoxin-mediated shock. *Journal of Immunology* 182:6533-6539
- Muller CA, Appelros S, Uhl W, Buchler MW, Borgstrom A. (2002). Serum levels of procarboxypeptidase B and its activation peptide in patients with acute pancreatitis and non-pancreatic diseases. *Gut* 51:229-235
- Murakami M, Suzuki S, Ito Y, Kodani S & Ishida K. (2000). New Anabaenopeptins, Potent Carboxypeptidase-A Inhibitors from the Cyanobacterium *Aphanizomenon flos-aquae*. *Journal of Natural Products* 63:1280-1282
- Murshudov GN, Vagin AA, Dodson EJ. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica Section D: Biological Crystallography* 53:240-255
- Nantermet PG, Barrow JC, Lindsley SR, Young M, Mao S-S, Carroll S, Bailey C, Bosserman M, Colussi D, McMasters DR, Vacca JP, Selnick HG. (2004). Imidazole acetic acid TAF1a inhibitors: SAR studies centered around the basic P<sub>1</sub> group. *Bioorganic & Medicinal Chemistry Letters* 14:2141-2145
- Ner SK, Suckling CJ, Bell AR, Wrigglesworth R. (1987). Inhibition of carboxypeptidase by cyclopropane-containing peptides. *Journal of the Chemical Society, Chemical Communications* pp. 480-482
- Niemirowicz G, Parussini F, Agüero F, Cazzulo JJ. (2007). Two metallo-carboxypeptidases from the protozoan *Trypanosoma cruzi* belong to the M32 family, found so far only in prokaryotes. *Biochemical Journal* 401:399-410
- Nishioka Y, Higuchi T, Sato Y, Yoshioka S, Tatsumi K, Fujiwara H, Fujii S. (2003). Human migrating extravillous trophoblasts express a cell surface peptidase, carboxypeptidase-M. *Molecular Human Reproduction* 9:799-806
- Norgaard P, Andersen AS. (2008). Method for making matured insulin polypeptides. Patent WO 2008/037735
- Ohuchi S, Suda H, Naganawa H, Kawamura K, Aoyagi T, Umezawa H. (1984). Structure-activity relationships among derivatives of arphamenines, inhibitors of aminopeptidase B. *Journal of Antibiotics* 37:1741-1743
- O'Malley PGP, Sangster SM, Abdelmagid SA, Bearne SL, Too CKL. (2005). Characterization of a novel, cytokine-inducible carboxypeptidase D isoform in haematopoietic tumour cells. *Biochemical Journal* 390:665-673
- Ondetti MA, Condon ME, Reid J, Sabo EF, Cheung HS, Cushman DW. (1979). Design of potent and specific inhibitors of carboxypeptidase A and B. *Biochemistry* 18:1427-1430

- Ondetti MA. (1994). From peptides to peptidases: a chronicle of drug discovery. *Annual Review of Pharmacology and Toxicology* 34:1-16
- Oppenheim FG, Yang Y-C, Diamond RD, Hyslop D, Offner GD, Troxler RF. (1986). The primary structure and functional characterization of the neutral histidine-rich polypeptide from human parotid secretion. *Journal of Biological Chemistry* 261:1177-1182
- Oppezzo O, Ventura S, Bergman T, Vendrell J, Joernvall H, Avilés FX. (1994). Procarboxypeptidases in rat pancreas. Overall characterization and comparison of the activation processes. *European Journal of Biochemistry* 222:55-63
- Otwinowski Z & Minor W. (1997). Processing of X-ray diffraction data collected in oscillating mode. *Methods in Enzymology* 276:307-326
- Pallarès I, Bonet R, García-Castellanos R, Ventura S, Avilés FX, Vendrell J, Gomis-Rüth FX. (2005). Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin. *Proceedings of the National Academy of Sciences of the USA* 102:3978-3983
- Pallarès I, Fernández D, Comellas-Bigler M, Fernández-Recio J, Ventura S, Avilés FX, Bode W, Vendrell J. (2008). Direct interaction between a human digestive protease and the mucoadhesive poly(acrylic acid). *Acta Crystallographica Section D: Biological Crystallography* 64:784-791
- Parellada J, Suarez G, Guinea M. (1998). Inhibition of zinc metallopeptidases by flavonoids and related phenolic compounds: structure-activity relationships. *Journal of Enzyme Inhibition* 13:347-359
- Park JD, Lee KJ, Kim DH. (2001). A new inhibitor design strategy for carboxypeptidase A as exemplified by *N*-(2-chloroethyl)-*N*-methylphenylalanine. *Bioorganic & Medicinal Chemistry* 9:237-243
- Park JD, Kim DH. (2002). Cysteine derivatives as inhibitors for carboxypeptidase A: synthesis and structure-activity relationships. *Journal of Medicinal Chemistry* 45:911-918
- Pascual R, Burgos FJ, Salva M, Soriano F, Méndez E, Avilés FX. (1989). Purification and properties of five different forms of human procarboxypeptidases. *European Journal of Biochemistry* 179:609-616
- Payne DJ, Bateson JH, Tolson D, Gasson B, Khushi T, Ledent P, Frere J-M. (1996). Phosphoramidate analogues of dipeptides with carboxypeptidase A and  $\beta$ -lactamase-inhibitory activity: elucidation of the mechanism of  $\beta$ -lactamase inhibition by electrospray mass spectrometry. *Biochemical Journal* 314:457-461
- Pearce CB, Sadek SA, Walters AM, Goggin PM, Somers SS, Toh SK, Johns T, Duncan HD. (2006). A double-blind, randomised, controlled trial to study the effects of an enteral feed supplemented with glutamine, arginine, and omega-3 fatty acid in predicted acute severe pancreatitis. *Journal of Pancreas (Online)* 7:361-371
- Peterson LM, Holmquist B, Bethune JL. (1982). A unique activity assay for carboxypeptidase A in human serum. *Analytical Biochemistry* 125:420-426
- Phillips MA, Kaplan AP, Rutter WJ, Bartlett PA. (1992). Transition-state characterization: a new approach combining inhibitor analogues and variation in enzyme structure. *Biochemistry* 31:959-963
- Piera E, Comelles F, Erra P, Infante MR. (1998). New alkyl amide type cationic surfactants from arginine. *Journal of the Chemical Society, Perkin Transactions 2*, pp. 335-341
- Pinto GA, Tardioli PW, Cabrera-Padilla RY, Galvão CMA, Giordano RC, Giordano RLC. (2008). Amino acids yields during proteolysis catalyzed by carboxypeptidase A are strongly dependent on substrate pre-hydrolysis. *Biochemical Engineering Journal* 39:328-337
- Pitout MJ, Nel W. (1969). The inhibitory effect of ochratoxin A on bovine carboxypeptidase A *in vitro*. *Biochemical Pharmacology* 18:1837-1843
- Pohland AE, Nesheim S, Friedman L. (1992). Ochratoxin A: a review. *Pure & Applied Chemistry* 64:1029-1046

- Plummer TH Jr., Kimmel MT. (1980). An improved spectrophotometric assay for human plasma carboxypeptidase N. *Analytical Biochemistry* 108:348-353
- Plummer TH Jr., Ryan TJ. (1981). A potent mercapto bi-product analogue inhibitor for human carboxypeptidase N. *Biochemical and Biophysical Research Communications* 98:448-454
- Pritchard L, Cardle L, Quinn S, Dufton M. (2003). Simple intrasequence difference (SID) analysis: an original method to highlight and rank sub-structural interfaces in protein folds. Application to the folds of bovine pancreatic trypsin inhibitor, phospholipase A2, chymotrypsin and carboxypeptidase A. *Protein Engineering* 16: 87-101
- Quentin G. (2003). Novel chromogenic substances and use thereof for the determination of carboxypeptidase activities. Patent WO 03/004516
- Rawlings ND, Morton FR & Barrett AJ. (2006). MEROPS: the peptidase database. *Nucleic Acids Research* 34:D270-274
- Read RJ. (2001). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallographica Section D: Biological Crystallography* 57:1373-1382
- Rees DC, Lipscomb WN. (1982). Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *Journal of Molecular Biology* 160:475-498
- Regnér S, Appelros S, Hjalmarsson C, Manjer J, Sadic J, Borgström A. (2008). Monocyte chemoattractant protein 1, active carboxypeptidase B and CAPAP at hospital admission are predictive markers for severe acute pancreatitis. *Pancreatology* 8:42-49
- Reverter D, Fernández-Catalán C, Baumgartner R, Pfänder R, Huber R, Bode W, Vendrell J, Holak TA, Avilés FX. (2000). Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nature Structural Biology* 7:322-328
- Reverter D, Maskos K, Tan F, Skidgel RA, Bode W. (2004). Crystal structure of human carboxypeptidase M, a membrane-bound enzyme that regulates peptide hormone activity. *Journal of Molecular Biology* 338:257-269
- Rezanka T, Sigler K. (2008). Biologically active compounds of semi-metals. *Phytochemistry* 69:585-606
- Robinson LK, Hurley LS. (1981). Effect of maternal zinc deficiency or food restriction on rat fetal pancreas. I. Procarboxypeptidase A and chymotrypsinogen. *Journal of Nutrition* 111:858-868
- Rooth E, Wallen H, Antovic A, von Arbin M, Kaponides G, Wahlgren N, Blombäck M, Antovic J (2007) Thrombin activatable fibrinolysis inhibitor and its relationship to fibrinolysis and inflammation during the acute and convalescent phase of ischemic stroke. *Blood Coagulation and Fibrinolysis* 18:365-370
- Ross PL, Cheng I, Liu X, Cicek MS, Carroll PR, Casey G, Witte JS. (2009). Carboxypeptidase 4 gene variants and early-onset intermediate-to-high risk prostate cancer. *BMC Cancer* 9:69 (doi:10.1186/1471-2407-9-69)
- Ryu S-E, Choi H-J, Kim DH. (1997). Stereochemistry in Inactivation of Carboxypeptidase A. Structural Analysis of the Inactivated Carboxypeptidase A by an Enantiomeric Pair of 2-Benzyl-3,4-epoxybutanoic Acids. *Journal of the American Chemical Society* 119:38-41
- Sáez J, Martínez J, Trigo C, Sánchez-Payá J, Compañy L, Laveda R, Griño P, García C, Pérez-Mateo M. (2005). Clinical value of rapid urine trypsinogen-2 test strip, urinary trypsinogen activation peptide, and serum and urinary activation peptide of carboxypeptidase B in acute pancreatitis. *World Journal of Gastroenterology* 11:7261-7265
- Sakagami Y, Tsujibo H, Hirai Y, Yamada T, Numata A, Inamori Y. (1999). Inhibitory activities of 2-pyridinecarboxylic acid analogs on phytogrowth and enzymes. *Biological and Pharmaceutical Bulletin* 22:1234-1236

- Sánchez de Groot N, Pallarès I, Avilés FX, Vendrell J, Ventura S. (2005). Prediction of "hot spots" of aggregation in disease-linked polypeptides. *BMC Structural Biology* 5:18
- Sanglas L, Avilés FX, Huber R, Gomis-Rüth FX, Arolas JL. (2009). Mammalian metallopeptidase inhibition at the defense barrier of *Ascaris* parasite. *Proceedings of the National Academy of Sciences of the USA* 106:1743-1747
- Saruta H, Ashihara Y, Sugiyama M, Roth M, Miyagawa E, Kido Y, Kasahara Y. (1986). Colorimetric determination of carboxypeptidase A activity in serum. *Clinical Chemistry* 32:748-751
- Schmuck C. (2006). How to improve guanidinium cations for oxoanion binding in aqueous solution? The design of artificial peptide receptors. *Coordination Chemistry Reviews* 250:3053-3067
- Sebastian JF, Liang G, Jabarin A, Thomas K, Wu HB. (1996). Effect of enzyme-substrate interactions away from the reaction site on carboxypeptidase A catalysis. *Bioorganic Chemistry* 24:290-303
- Senter PD, Wallace PM, Svensson HP, Vrudhula VM, Kerr DE, Hellstroem I, Hellstroem KE. (1993). Generation of cytotoxic agents by targeted enzymes. *Bioconjugate Chemistry* 4:3-9
- Sheldrick GM. (1997). SHELX97. Programs for Crystal Structure Analysis (Release 97-2). University of Göttingen, Germany.
- Showell GA, Mills JS. (2003). Chemistry challenges in lead optimization: silicon isosteres in drug discovery. *Drug Discovery Today* 8:551-556
- Simon LM, Laszlo K, Vertesi A, Bagi K, Szajani B. (1998). Stability of hydrolytic enzymes in water-organic solvent systems. *Journal of Molecular Catalysis B: Enzymatic* 4:41-45
- Skidgel RA, Erdös EG. (1998). Cellular carboxypeptidases. *Immunological Reviews* 161:129-141
- Slater AM, Blakey DC, Davies DH, Hennem JF, Hennequin LFA, Marsham PR, Dowell RI. (2002). Chemical compounds. Patent US 6,436,691
- Slobin LI, Carpenter FH. (1966). Kinetic studies on the action of carboxypeptidase A on bovine insulin and related model peptides. *Biochemistry* 5:499-508
- Smith GK, Banks S, Blumenkopf TA, Cory M, Humphreys J, Laethem RM, Miller J, Moxham CP, Mullin R, Ray PH, Walton LM, Wolfe LA III. (1997). Toward antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1 and novel in vivo stable prodrugs of methotrexate. *Journal of Biological Chemistry* 272:15804-15816
- Solomon BM, Larsen KS, Riordan JF. (1990). Catalytic and conformational changes induced by limited subtilisin cleavage of bovine carboxypeptidase A. *Biochemistry* 29:7303-7309
- Song L, Fricker LD. (1995). Calcium- and pH-dependent aggregation of carboxypeptidase E. *Journal of Biological Chemistry* 270:7963-7967
- Stander MA, Steyn PS, van der Westhuizen FH, Payne BE. (2001). A kinetic study into the hydrolysis of the ochratoxins and analogues by carboxypeptidase A. *Chemical Research and Toxicology* 14:302-304
- Stewart JD, Gilvarg C. (1999). Determination of the activity of carboxypeptidase A in the blood of healthy human adults. *Clinica Chimica Acta* 281:19-28
- Suh J, Lee SH, Uh JY. (1995). Zn(II)-chelating inhibitors of carboxypeptidase A. *Bioorganic & Medicinal Chemistry Letters* 5:585-588
- Suzuki K, Muto Y, Fushihara K, Kanemoto K, Iida H, Sato E, Kikuchi C, Matsushima T, Kato E, Nomoto M, Yoshioka S, Ishii H. (2004). Enhancement of fibrinolysis by EF6265 [(S)-7-amino-2-[[[(R)-2-methyl-1-(3-phenylpropanoylamino)propyl]hydroxyphosphinoyl]methyl]heptanoic acid], a specific inhibitor of plasma carboxypeptidase B. *Journal of Pharmacology and Experimental Therapeutics* 309:607-615



- Tanaka Y, Grapsas I, Dakoji S, Cho YJ, Mobashery S. (1994). Conscripting the active-site zinc ion in Carboxypeptidase A in Inactivation chemistry by a new type of irreversible enzyme inactivator. *Journal of the American Chemical Society* 116:7475-7480
- Tani S, Akatsu H, Ishikawa Y, Okada N, Okada H. (2003). Preferential detection of pro-carboxypeptidase R by enzyme-linked immunosorbent assay. *Microbiology and Immunology* 47:295-300
- Tardioli PW, Fernández-Lafuente R, Guisan JM, Giordano RLC. (2003). Design of new immobilized-stabilized carboxypeptidase A derivative for production of aromatic free hydrolysates of proteins. *Biotechnology Progress* 19:565-574
- Ton GN, Weichert JP, Longino MA, Fine JP, Kwon GS. (2005). Methoxypoly(ethylene glycol)-conjugated carboxypeptidase A for solid tumor targeting Part II: pharmacokinetics and biodistribution in normal and tumor-bearing rodents. *Journal of Controlled Release* 104:155-166
- Tregouet DA, Schnabel R, Alessi MC, Godefroy T, Declerck PJ, Nicaud V, Munzel T, Bickel C, Rupprecht HJ, Lubos E, Zeller T, Juhan-Vague I, Blankenberg S, Tiret L, Morange PE for the AtheroGene Investigators. (2009). Activated thrombin activatable fibrinolysis inhibitor levels are associated with the risk of cardiovascular death in patients with coronary artery disease: the AtheroGene study. *Journal of Thrombosis and Haemostasis* 7:49-57
- Umeda Y, Moriguchi M, Ikai K, Kuroda H, Nakamura T, Fujii A, Takeuchi T, Umezawa H. (1987). Synthesis and antitumor activity of spergualin analogues III. Novel method for synthesis of optically active 15-deoxyspergualin and 15-deoxy-1-O-methylspergualin. *Journal of Antibiotics* 40:1316-1324
- Umezawa H, Aoyagi T, Ogawa K, Iinuma H, Naganawa H, Hamada M, Takeuchi T. (1984). Histargin, a new inhibitor of carboxypeptidase B, produced by actinomycetes. *Journal of Antibiotics* 37:1088-1090
- Valnickova Z, Thogersen IB, Potempa J, Enghild JJ. (2007). Thrombin-activable fibrinolysis inhibitor (TAFI) zymogen is an active carboxypeptidase. *Journal of Biological Chemistry* 282:3066-3076
- de la Vega MR, Sevilla RG, Hermoso A, Lorenzo J, Tanco S, Diez A, Fricker LD, Bautista JM, Avilés FX. (2007). Nna1-like proteins are active metallo-carboxypeptidases of a new and diverse m14 subfamily. *FASEB Journal* 20:851-865
- Vendrell J, Querol E & Avilés FX. (2000). Metallo-carboxypeptidases and their protein inhibitors. Structure, functions and biomedical properties. *Biochimica et Biophysica Acta* 1477:284-298
- Vendrell J, Avilés FX & Fricker LD. (2004). Metallo-carboxypeptidases, in *Handbook of Metalloproteins* (Messerschmidt, A., Bode, W. & Cygler, M., eds) pp. 176-189, John Wiley & Sons, Ltd., Chichester
- Ventura S, Gomis-Rüth FX, Puigserver A, Avilés FX, Vendrell J. (1997). Pancreatic procarboxypeptidases: oligomeric structures and activation processes revisited. *Biology and Chemistry* 378:161-165
- Ventura S, Villegas V, Sterner J, Larson J, Vendrell J, Hershberger CL, Avilés FX. (1999). Mapping the pro-region of carboxypeptidase B by protein engineering. Cloning, overexpression, and mutagenesis of the porcine proenzyme. *Journal of Biological Chemistry* 274:19925-19933
- Vertesi A, Simon LM. (1998). Carboxypeptidase A-catalyzed dipeptide synthesis in organic media. *Journal of Biotechnology* 66:75-82
- Villalonga R, Cao R, Fragoso A. (2007). Supramolecular chemistry of cyclodextrins in enzyme technology. *Chemical Reviews* 107:3088-3116
- Vovchuk IL, Petrov SA. (2008). The role of carboxypeptidases in carcinogenesis. *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry* 2:267-274
- Walpole CS, Wrigglesworth R. (1989). Enzyme inhibitors in medicine. *Natural Products Reports* 6:311-346

- Wang H, Zhou Q, Kesinger JW, Norris C, Valdez C. (2007a). Heme regulates exocrine peptidase precursor genes in zebrafish. *Experimental and Biological Medicine* (Maywood) 232:1170-1180
- Wang W, Cain BM, Beinfeld MC. (1998). Adult carboxypeptidase E-deficient *fat/fat* mice have a near-total depletion of brain CCK 8 accompanied by a massive accumulation of glycine and arginine extended CCK. *Endocrine* 9:329-332
- Wang YX, Zhao L, Nagashima M, Vincelette J, Sukovich D, Li W, Subramanyam B, Yuan S, Emayan K, Islam I, Hrvatin P, Bryant J, Light DR, Vergona R, Morser J, Buckman BO. (2007b). A novel inhibitor of activated thrombin-activatable fibrinolysis inhibitor (TAFIa) - part I: pharmacological characterization. *Thrombosis and Haemostasis* 97:45-53
- Wang YX, da Cunha V, Vincelette J, Zhao L, Nagashima M, Kawai K, Yuan S, Emayan K, Islam I, Hosoya J, Sullivan ME, Dole WP, Morser J, Buckman BO, Vergona R. (2007c). A novel inhibitor of activated thrombin activatable fibrinolysis inhibitor (TAFIa) - part II: enhancement of both exogenous and endogenous fibrinolysis in animal models of thrombosis. *Thrombosis and Haemostasis* 97:54-61
- Watanabe M, Ishikawa Y, Campbell W, Okada H. (1998). Measurement of arginine carboxypeptidase-generating activity of adult plasma. *Microbiology and Immunology* 42:393-397
- Willemse JL, Hendriks DF, Chen D. (2008). Major carboxypeptidase N deficiency. *Clinica Chimica Acta* 389:181-182
- Wolfe LA, Mullin RJ, Laethem R, Blumenkopf TA, Cory M, Miller JF, Keith BR, Humphreys J, Smith GK. (1999). Antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1: in vitro and in vivo studies with prodrugs of methotrexate and the thymidylate synthase inhibitors GW1031 and GW1843. *Bioconjugate Chemistry* 10:38-48
- Yamauchi K, Ohtsuki S, Kinoshita M. (1985). Phosphonodipeptides containing (2-aminoethyl)phosphonic acid (ciliatine): transition state analogue inhibitors of carboxypeptidase A. *Biochimica et Biophysica Acta* 827:272-282
- Yang Z-H, Du X, Tang J-G. (1999). Separation and characterization of trypsin and carboxypeptidase B-digested products of met-lys-human proinsulin. *Applied Biochemistry and Biotechnology* 76:107-114
- Yun M, Park C, Kim S, Nam D, Kim SC, Kim DH. (1992). The x-ray crystallographic study of covalently modified carboxypeptidase A by 2-benzyl-3,4-epoxybutanoic acid, a pseudomechanism-based inactivator. *Journal of the American Chemical Society* 114:2281-2282
- Zen A, Carnevale V, Lesk AM, Micheletti C. (2008). Correspondences between low-energy modes in enzymes: dynamics-based alignment of enzymatic functional families. *Protein Science* 17:918-929
- Zhou Q, Law AC, Rajagopal J, Anderson WJ, Gray PA, Melton DA. (2007). A multipotent progenitor domain guides pancreatic organogenesis. *Developmental Cell* 13:103-114
- Zhu X, Wu K, Rife L, Cawley NX, Brown B, Adams T, Teofilo K, Lillo C, Williams DS, Loh YP, Craft CM. (2005). Carboxypeptidase E is required for normal synaptic transmission from photoreceptors to the inner retina. *Journal of Neurochemistry* 95:1351-1362
- Zhu Z. (2004). Regulation of human carboxypeptidase A. Patent US 2004/0115670
- Ziegler H, Prasa D, Sturzebrecher J, Wilkstroem P. (2006). Substrates for TAFI (a). Patent US 2006/00668457

# Acknowledgements



# Acknowledgements

I would like to express my sincere gratitude to Profs. F. Xavier Avilés and Josep Vendrell for making possible my stay in their labs and for giving me the opportunity to carry out my Ph. D. at the Universitat Autònoma de Barcelona. A Xavier y a Pep les estoy muy agradecido por la confianza depositada en mí, lo que me ha permitido desarrollar mi trabajo de una manera continua, sin prisas pero sin pausas. *Moltes gracies per tot!*

Quisiera agradecerles a Angel Ramírez-Ortiz (CBMSO, Madrid) y a Juan Fernández-Recio (IRB-CSIC, Barcelona) el haberme permitido acercarme al mundo bioinformático. Una muy grata experiencia! Asimismo a Xavier Gomis-Rüth (IRB-CSIC, Barcelona) el haberme abierto las puertas a la biología estructural.

I would like to give my thanks to Prof. Shahriar Mobashery for the enjoyable stay at his group in Notre Dame, IN, USA. It was a delightful experience for me! *Awesome!* A Sebas y Leti por haberme acogido y ayudado, mil gracias! I have also the nicest words for Pete, Takao, Major, Mayland, Mijoon...

Al Prof. Juan J. Cazzulo (Instituto de Investigaciones Biotecnológicas, Universidad Nacional de General San Martín, Buenos Aires) por todo su apoyo en el proyecto de proteasas de *Trypanosoma*. Gabriela, *suerte con los mutantes y con la Tesis!*

*Una pila de companys* a quienes agradecer. Comenzando en can PROs, quienes hacen del día a día algo especial: Natàlia, Virginia, Alba, MarthaLi, Raimon, Adri, Vero, Anna, Irantzu... Por supuesto que no me olvido de los enzimos: Sebas Tanco, Sebas Trejo, Laura, Montse, Mónica, Silvia, Joan, Julia...ni de algunos ex-PROs: Àlex, Roman, y Núria.

Tuve el placer de contar con excelentes estudiantes: Ricardo, Gemma, Ivette y Bernat, a todos ellos gracias por haberme “aguantado” y *suerte en el futuro!*

En el Departamento, quienes hacen que los experimentos funcionen (*por lo menos los equipos y reactivos!*): Salva (Bartolomé) y Helena...Y agradecer a Ester Boix toda su paciencia y entusiasmo para resolver todo tipo de situaciones (*y, por supuesto, estructuras!*)

A Mireia García-Viloca y Manuel Montenegro (IBB) y a Angel Alvarez-Larena (Unitat de Cristallografia) las constructivas discusiones computacionales y de difracción de rayos X. A Rosa Ortuño (Departament de Química Orgánica) por su entusiasmo en la obtención de nuevos inhibidores.

Last but not least I would like to thank the initial financial support from the Alβan Programme, and at a later stage, from CAMP project (VI EU Framework Programme).

# Publications





# Publications

## Order of papers:

**Fernández D**, Testero S, Vendrell J, Avilés FX, Mobashery S. (2010). *Chem Biol Drug Des*, author proofs

**Fernández D**, Boix E, Pallarès I, Avilés FX, Vendrell J. (2009). *Biopolymers*, author proofs

**Fernández D**, Avilés FX, Vendrell J. (2009). *Chem Biol Drug Des* 73:75-82

**Fernández D**, Torres E, Avilés FX, Ortuño RM, Vendrell J. (2009). *Bioorg Med Chem* 17:3824-3828

**Fernández D**, Avilés FX, Vendrell J. (2009). *Eur J Med Chem* 44:3266-3271

**Fernández D**, Illa O, Avilés FX, Branchadell V, Vendrell J, Ortuño RM. (2008). *Bioorg Med Chem* 14:4823-4828

**Fernández D**, Vendrell J, Avilés FX & Fernández-Recio J. (2007). *Proteins: Structure, Function, and Bioinformatics* 60:131-144



# The X-Ray Structure of Carboxypeptidase A Inhibited by a Thiirane Mechanism-Based Inhibitor

Daniel Fernández<sup>1</sup>, Sebastian Testero<sup>2</sup>,  
Josep Vendrell<sup>1</sup>, Francesc X. Avilés<sup>1,\*</sup> and  
Shahriar Mobashery<sup>2,\*</sup>

<sup>1</sup>Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, and Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain

<sup>2</sup>Departament of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA

\*Corresponding authors: Shahriar Mobashery, mobashery@nd.edu; Francesc X. Avilés, francesc Xavier.aviles@uab.es

**The three-dimensional X-ray crystal structure of carboxypeptidase A, a zinc-dependent hydrolase, covalently modified by a mechanism-based thiirane inactivator, 2-benzyl-3,4-epithiobutanoic acid, has been solved to 1.38 Å resolution. The interaction of the thiirane moiety of the inhibitor with the active site zinc ion promotes its covalent modification of Glu-270 with the attendant opening of the thiirane ring. The crystal structure determination at high resolution allowed for the clear visualization of the covalent ester bond to the glutamate side chain. The newly generated thiol from the inhibitor binds to the catalytic zinc ion in a monodentate manner, inducing a change in the zinc ion geometry and coordination, while its benzyl group fits into the S1' specificity pocket of the enzyme. The inhibitor molecule is distorted at the position of the carbon atom that is involved in the ester bond linkage on one side and the zinc coordination on the other. This particular type of thiirane-based metalloprotease inhibitor is for the first time analyzed in complex to the target protease at high resolution and may be used as a general model for zinc-dependent proteases.**

**Key words:** M14 family of proteases, mechanism-based inactivation, metalloprotease, thiirane, X-ray crystallography


Received 21 September 2009, revised xx Xxx 200x and accepted for publication xx Xxx 200x

Metalloproteases from the M14 family (carboxypeptidases, CPs) are involved in many physiologic and pathological processes such as tissue organogenesis (1–3), acute pancreatitis (4,5), diabetes (6,7), inflammation (8,9), fibrinolysis (10,11), neurologic diseases (12), and cancer (13–16). Circulating forms of CP may be used as prognostic tools for the early detection of cancer and other diseases (17). The

M14 family of proteases comprises an extended and complex array of zinc-dependent proteins with wide genomic distribution (18–20). They share with matrix metalloproteases (MMPs; e.g., gelatinases, collagenases, matrilysins, and stromelysins) a similar catalytic mechanism, whereby the catalytic zinc ion predisposes the substrate to turnover with the involvement of a conserved glutamate residue in the active site (21). Notwithstanding the shared features of their respective mechanisms, MMPs are endopeptidases, whereas CPs are exopeptidases that remove the C-terminal residues from peptide substrates.

Most known small-molecule inhibitors of CPs are chelators of the active site zinc ion. These chelators are often non-discriminant, as their function hinges on coordination with metals. The limited selectivity among potent zinc chelators can be exemplified by thiol-dependent CP inhibitors, such as **1** (2-mercaptomethyl-3-guanidinoethylthiopropanoic acid, Plummer's inhibitor (22)), which was still found to inhibit other M14 proteases (23). A similar concern has been encountered in the development of MMP inhibitors with sub-type selectivity (24). Mechanism-based inactivators ('suicide substrate' or ' $k_{cat}$  inactivator') exploit features of the catalytic mechanisms of the targeted enzymes (25). An innocuous agent is converted to the inhibitory species only within the active site of the target enzyme. Compound **2**, 2-((4-phenoxyphenylsulfonyl)methyl)thiirane, SB-3CT, is a prototypical example. It was designed as a selective gelatinase (MMP-2 and MMP-9) inhibitor. Its unique features include a biphenyl moiety that would fit the S1' pocket of gelatinases and a thiirane ring able to revert into a thiol group upon enzymic reaction (26,27). The oxirane counterpart to **2** is three orders of magnitude less potent than the thiirane (28).

Chemotherapeutic compounds that may be used as drugs or imaging agents targeted to M14 proteases are highly sought (18,29), and the complexity among CPs calls for the development of selective compounds. Based on the notion that a covalently bound mechanism-based CP inactivator might be an ideal starting point for such an endeavor, a screening of our collection of thiirane-based MMP ligands (~500 compounds) against representative M14 proteases was performed. Unfortunately, these molecules did not prove efficacious in inhibition of these carboxypeptidases. We hasten to add that thiirane-based inhibitors of MMPs have not been amenable to study of their structures in complex with the MMPs by X-ray analysis, hence structural information is currently lacking (30). In an effort to elucidate the minimal structural motifs for inhibition of carboxypeptidases by thiirane-based inhibitors, we turned to two examples from the literature. Compound **3** (2-benzyl-3,4-epithiobutanoic acid), as well as its oxirane version, **4** (2-benzyl)-3,4-epoxybutanoic acid),

	C	B	D	D			9	0	7	B	Dispatch: 14.10.09	Journal: CBDD	CE: Hemalatha
	Journal Name		Manuscript No.				Author Received:		No. of pages: 5		PE: Gayathri		

**3** has been reported as mechanism-based inactivators of CPA (31–33). An earlier structural determination of CPA in complex with oxirane **4** (34,35) has been reported, although the atomic coordinates have not been deposited. Kinetic results showed that thiirane **3** and oxirane **4** displayed a similar inhibitory potency as CPA inactivators, contrarily to the example of gelatinase inhibition by **2**.

We synthesized compound **3** for its study with CPA. Its binding to the active site would allow for the coordination of the thiirane sulfur with the zinc ion, making the thiirane more electrophilic and predisposing it to nucleophilic attack by the active site glutamate. Herein, we report the X-ray structure of the enzyme inhibited by covalent modification with this compound.

## Methods and Materials

CPA crystals were grown from a 2  $\mu$ L:2  $\mu$ L mixture of enzyme solution (14 mg/mL in 0.02 M Tris, pH 7.5) and precipitant (20% PEG

3350, 0.2 M NH<sub>4</sub>Cl, 0.02 M Tris pH 7.0) by the vapor diffusion method. The largest crystals were harvested and transferred to a 2  $\mu$ L drop containing the reservoir solution plus 10 mM of **3**. Soaking lasted for one week before data collection time at which no crystal cracking or damage was evident. Crystals were harvested, briefly bathed in cryoprotectant buffer (i.e., the reservoir solution with added 30% glycerol), flash frozen in nitrogen stream and diffracted. Some crystals gave excellent quality diffraction images and that of the highest resolution was used for structure determination by the molecular replacement procedure using native CPA (PDB: 2ctb) as the model. Programs from the CCP4 suite were used in different stages of structure determination (36). Extra electron density was clearly evident in the region near the catalytic zinc ion and was easily interpreted as the inhibitor molecule. Further, clear continuous electron density unequivocally revealed the covalent linkage between the inhibitor and Glu-270 side chain. The agreement between the model and the experimental observations is excellent, as indicated by the crystallographic and refinement statistics (Table 1). Geometric details of inhibitor–enzyme interactions are listed in Table 2.

Parameters	Value
Wavelength used during data collection	0.8123 Å
Unit cell constants	a = 42.32 Å, b = 57.50 Å, c = 57.08 Å $\alpha = 90.0^\circ$ , $\beta = 99.1^\circ$ , $\gamma = 90.0^\circ$
Resolution range	31.27–1.38 Å
Space group	P2 <sub>1</sub> (1 mol/asymmetric unit)
Number of measured reflections	488 546
Number of unique reflections	54 345
$R_{\text{merge}}^a$ (overall/outermost shell)	6.8/37.9%
Completeness & multiplicity (overall/outermost shell)	96.6/77.8%
$I/\sigma I$ (overall/outermost shell)	6.1/5.3
Reflections used for refinement (total/test set)	52 590/1102
Crystallographic $R_{\text{factor}}^b/R_{\text{free}}^c$	14.3/15.7%
Deviation from ideality	
r.m.s.d. bond lengths	0.008 Å
r.m.s.d. bond angles	1.14°
Number of protein atoms/total atoms	2438/2930
B-factor statistics (Å <sup>2</sup> )	
Overall B-factor/Wilson plot B-factor	10.7/9.3
Catalytic domain, main/side chain	7.3/8.4
$Zn^{2+}$ (1 in total/1 mol per monomer)	5.3
Inhibitor atoms	5.8
(14 in total/1 mol per monomer)	
Glycerol atoms	19.4
(42 in total/7 mols per monomer)	
Solvent atoms (435 in total)	26.0
Protein geometry <sup>d</sup>	
Ramachandran favored	97.3% (293 of 301 residues)
Ramachandran allowed	99.7% (300 of 301 residues)
Ramachandran outliers	0.3% (1 of 301 residues, Ser-199)
Residues with bad bonds/angles	0.00/0.00%
Rotamer outliers	0.77%

**Table 1:** Statistics of data collection and refinement for thiirane-bound CPA

<sup>a</sup> $R_{\text{merge}} = \sum_{\text{hkl}} \sum_{j=1}^N |I_{\text{hkl}} - h_{\text{kl}}(j)| / \sum_{\text{hkl}} \sum_{j=1}^N I_{\text{hkl}}(j)$ , where  $N$  is the redundancy of the data. The outermost shell is 1.46–1.38 Å.

<sup>b</sup> $R_{\text{factor}} = \sum_{\text{hkl}} \|F_{\text{obs}} - F_{\text{calc}}\| / \sum_{\text{hkl}} |F_{\text{obs}}|$  where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure factor amplitudes of reflection hkl.

<sup>c</sup> $R_{\text{free}} = R_{\text{factor}}$  for a randomly selected 2% subset of reflections that were not used in refinement.

<sup>d</sup>According to Molprobit (41).

## Mechanism-Based Thiirane Inhibitor of Carboxypeptidase

**Table 2:** Geometric details of thiirane-CPA interactions

Interaction	Parameter (Å, °)	Comments
Zn/His-69, Zn/His-196, Zn/Glu-72-Zn, Zn/S <sub>1</sub>	2.10, 2.11, 2.04, 2.33	Zinc coordination sphere
Glu-270 O <sub>ε2</sub> -C <sub>4</sub>	1.61	Covalent bond to nucleophile Glu-270
Tyr-248 OH-COOH	2.59	Hydrogen bond phenolic ring
Arg-145 N <sub>H1</sub> -COOH	2.89	Salt bridge
Arg-145 N <sub>H2</sub> -COOH	2.78	
Arg-127 N <sub>H2</sub> -COOH	3.35	Hydrogen bond of inhibitor COOH
Asn-144 N <sub>δ1</sub> -COOH	2.97	
Leu-203 C <sub>δ1</sub> -C <sub>phenyl</sub>	3.85	Hydrophobic interactions of the benzyl ring (selected)
Ile-243 C <sub>δ1</sub> -C <sub>phenyl</sub>	3.42	
Thr-268 C <sub>γ2</sub> -C <sub>phenyl</sub>	3.68	
C <sub>3</sub> -S <sub>1</sub> , C <sub>4</sub> -C <sub>3</sub> -S <sub>1</sub>	1.81, 100.9	Geometry around the sulfur

## Results and Discussion

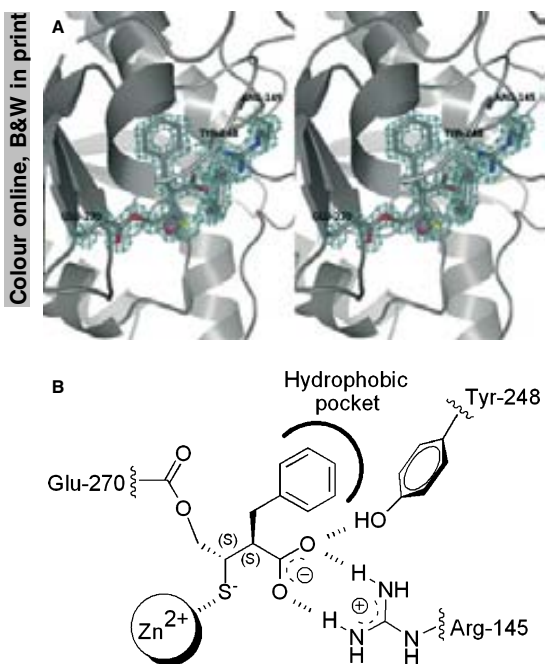
The three-dimensional structure of the complex between CPA, a model zinc-dependent protease, and a thiirane mechanism-based inactivator was solved to 1.38 Å resolution. The CPA crystal structure presented herein is the highest resolution structure for this enzyme, with a nearly 80% data completeness in the highest resolution shell. The structure has been deposited with the PDB with accession code 3i1u. As shown in Figure 1, the existence of continuous electron density between the bound inhibitor and the side chain of the glutamate indicates that Glu-270, corresponding to the conserved active site glutamate of the zinc-dependent proteases, is covalently modified by the inhibitor. The establishment of the covalent bond is accompanied by the opening of the thiirane ring with the attendant coordination of the newly formed thiol group with the catalytic zinc ion in a monodentate manner. This arrangement gives a tetrahedral coordination sphere for the zinc ion, with three ligands from the protein and the thiol from the inhibitor. Because thiiranes are quite stable at physiological pH, the involvement of the active site zinc ion is essential for the ring opening. The coordinated thiol (likely as a thiolate) is the remnant of the interaction of the thiirane sulfur, before the ring opening. Although inhibitor **3** was prepared as a racemic mixture, the X-ray structure reveals that it is only the 2*S*,3*S*-**3** configuration that fits into the electron density map, indicating that the enzyme is enantioselective in its interaction with the inhibitor. This configuration corresponds to that of the D-amino acid series, in accordance with what has been observed for oxirane **4** (34,35).

When compared to the structure of the native CPA, some conformational changes occur upon inhibitor binding. The catalytically important Tyr-248 (37) and Arg-145 have experienced motion. Tyr-248 has been observed in two conformational states in the several structures available for CPA. One brings it to a hydrogen bonding distance of the bound peptide substrate (the 'closed' position) and the other is away from it (the 'open' position) (38,39). In the X-ray structure of the inhibited complex, the Tyr-248 phenol group moves from the surface closer to the active site cleft to make a strong hydrogen bond to the carboxylate group of the inhibitor (the bond distance is 2.59 Å; the 'closed' position). This carboxylate group, which corresponds to the terminal carboxylate of the peptide substrate, is held in place by hydrogen bonds to the side chain of Arg-145 and Asn-144 N<sub>δ1</sub>. The benzyl ring of the inhibitor is buried in

the S1' pocket, thus interacting with Ile-255. This interaction is believed to impart CPA selectivity toward hydrolytic processing of the C-terminal hydrophobic amino acids. In this pocket, residues Leu-203 and Ile-243 make C<sub>H</sub>-π interactions with the benzyl ring of the inhibitor. The presence of this benzyl ring displaces the water molecules present in the native CPA S1' pocket. A superimposition with the related carboxypeptidase B (CPB) reveals that the benzyl moiety would clash with Asp-255 of CPB, which corresponds to Ile-255 of CPA, hence the origin of the likely selectivity for CPA. The Glu-270 side chain conformation, which is covalently tethered to the inhibitor, is approximately *gauche* (as measured by the C<sub>α</sub>-C<sub>β</sub>-C<sub>γ</sub>-C<sub>δ</sub> dihedral angle of -74.4°). These observations point to the fact that little changes in key microenvironments in the binding pocket of the enzyme would allow for an exquisite recognition and binding of the ligand, despite it being configurationally dissimilar to the natural peptide substrate.

The zinc environment is perturbed in the complex. The coordination number of the zinc ion is four in the complex with the inhibitor and the resulting geometry is tetrahedral, with three protein ligand atoms at the base and the thiol sulfur of the inhibitor at the apex of a regular tetrahedron. The Zn-N distances are equal (2.10 Å and 2.11 Å for His-69 and His-196, respectively), while the Zn-S distance is 2.33 Å. The two Zn-O distances from the side chain of Glu-72, the third coordinating amino acid, differ by almost 1 Å (2.04 versus 3.00 Å). This observation shows that Glu-72 is a monodentate ligand to the zinc ion. The conformation of the glutamate carboxylate is *syn* as the O-C-O-Zn dihedral angle is -2.56°. In the native CPA, the Zn coordination number is five because of the bidentate coordination by Glu-72 (the Zn-O distances are 2.13 Å and 2.26 Å). A water molecule is at the apex of the distorted tetrahedron, at a distance of 2.07 Å. Again, the conformation of the glutamate carboxylate is *syn* (O-C-O-Zn = -1.69°). On the whole, the metal ion environment is perturbed by the inhibitor in the CPA-bound structure.

The inhibitor adopted a *trans* conformation in binding to CPA (the main C<sub>1</sub>-C<sub>2</sub>-C<sub>3</sub>-S<sub>1</sub> backbone dihedral angle is -162.4°). The C<sub>3</sub>-C<sub>4</sub> and C<sub>3</sub>-S<sub>1</sub> bond lengths are normal, 1.49 and 1.81 Å, respectively. The distance from C<sub>4</sub> to Glu-270 O<sub>ε2</sub>, 1.61 Å, is longer than the C-O ester bond distances reported in the other CPA structures (34,35). The C<sub>3</sub>-C<sub>4</sub>-O<sub>ε2</sub> bond angle compares favorably to that seen



**Figure 1:** (A) The structure of the CPA-inhibitor complex. A  $2F_{\text{obs}} - F_{\text{calc}}$  electron density map showing the CPA active site with the product of the reaction of thirane **3** with the enzyme. The electron density map (chicken wire), calculated deleting the Glu-270 side chain and inhibitor coordinates, is contoured at a  $1.5\sigma$  level. The inhibitor and residues important for binding are shown in capped sticks and are labeled. The catalytic zinc ion is in magenta, while the sulfur atom from the inhibitor is in yellow. Other atoms are colored blue (nitrogen) and red (oxygen). A continuous electron density is clearly seen along the bond linking Glu-270 side chain and the inhibitor. (B) Schematic representation of the CPA-inhibitor complex.

in CPA inhibited by 2-benzyl-3-iodopropanoic acid, a halogen-based covalent inhibitor of CPA:  $114.3^\circ$  versus  $113.5^\circ$  (40). The bond angles to heavy atoms around  $C_3$  average  $108.3^\circ$ , with a minimum of  $100.9^\circ$  for  $C_4-C_3-S_1$ . This angle is below the ideal tetrahedral value of  $109.5^\circ$  indicating a distortion of the tetrahedral geometry around  $C_3$  carbon. This carbon atom resides amidst the ester bond to Glu-270 side chain and the zinc-coordinating sulfur,  $S_1$ . Whether this distortion would have an effect over the inhibitor binding to the enzyme, and hence to its inhibitory potency, might only be assessed by analyzing compounds structurally related to the inhibitor analyzed here.

In summary, we observed in atomic detail the binding mode of a mechanism-based thirane inhibitor for CPA. The bound inhibitor fits in the  $S1-S1'$  subsites of CPA, a model metallopeptidase. The high-resolution complex between a zinc-dependent protease and a thirane-based inactivator has been obtained for the first time, allowing for a complete analysis of the covalently modified protein. This

structure constitutes the foundation for future development of additional reagents targeting metalloproteases.

## Acknowledgments

The work in the USA was supported by a grant from the National Institutes of Health. Financial support from the Ministerio de Ciencia e Innovación, Spain (grant BIO2007-68046), Generalitat de Catalunya (grant 2005SGR-1037) and CAMP project 108830 (VI EU Framework Programme) is gratefully acknowledged. DF was recipient of a grant from the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya. We wish to express our gratitude to the local staff at the X13 beamline EMBL-DESY, Hamburg, Germany.

## References

- Marquez-Curtis L, Jalili A, Deiteren K, Shirvaikar N, Lambeir A.M., Janowska-Wieczorek A. (2008) Carboxypeptidase M expressed by human bone marrow cells cleaves the C-terminal lysine of stromal cell-derived factor-1alpha: another player in hematopoietic stem/progenitor cell mobilization? *Stem Cells*;26:1211–1220.
- Wang H., Zhou Q., Kesinger J.W., Norris C., Valdez C. (2007) Heme regulates exocrine peptidase precursor genes in zebrafish. *Exp Biol Med (Maywood)*;232:1170–1180.
- Zhou Q., Law A.C., Rajagopal J., Anderson W.J., Gray P.A., Melton D.A. (2007) A multipotent progenitor domain guides pancreatic organogenesis. *Dev Cell*;13:103–114.
- Borgstrom A., Regner S. (2005) Active carboxypeptidase B is present in free form in serum from patients with acute pancreatitis. *Pancreatology*;5:530–536.
- Saez J., Martinez J., Trigo C., Sanchez-Paya J., Company L., Lavada R. *et al.* (2005) Clinical value of rapid urine trypsinogen-2 test strip, urinary trypsinogen activation peptide, and serum and urinary activation peptide of carboxypeptidase B in acute pancreatitis. *World J Gastroenterol*;11:7261–7265.
- Jeffrey K.D., Alejandro E.U., Luciani D.S., Kalynyak T.B., Hu X., Li H. *et al.* (2008) Carboxypeptidase E mediates palmitate-induced beta-cell ER stress and apoptosis. *Proc Natl Acad Sci USA*;105:8452–8457.
- Kitagawa N., Yano Y., Gabazza E.C., Bruno N.E., Araki R., Matsumoto K. *et al.* (2006) Different metabolic correlations of thrombin-activatable fibrinolysis inhibitor and plasminogen activator inhibitor-1 in non-obese type 2 diabetic patients. *Diabetes Res Clin Pract*;73:150–157.
- Asai S., Sato T., Tada T., Miyamoto T., Kimbara N., Motoyama N. *et al.* (2004) Absence of procarboxypeptidase R induces complement-mediated lethal inflammation in lipopolysaccharide-primed mice. *J Immunol*;173:4669–4674.
- Deiteren K., Hendriks D., Scharpe S., Lambeir A.M. (2009) Carboxypeptidase M: multiple alliances and unknown partners. *Clin Chim Acta*;399:24–39.
- Rooth E., Wallen H., Antovic A., von Arbin M., Kaponides G., Wahlgren N. *et al.* (2007) Thrombin activatable fibrinolysis inhib-

## Mechanism-Based Thiirane Inhibitor of Carboxypeptidase

- itor and its relationship to fibrinolysis and inflammation during the acute and convalescent phase of ischemic stroke. *Blood Coagul Fibrinolysis*;18:365–370.
11. Willemse J.L., Hendriks D.F. (2007) A role for procarboxypeptidase U (TAFI) in thrombosis. *Front Biosci*;12:1973–1987.
  12. Zhu X., Wu K., Rife L., Cawley N.X., Brown B., Adams T. *et al.* (2005) Carboxypeptidase E is required for normal synaptic transmission from photoreceptors to the inner retina. *J Neurochem*;95:1351–1362.
  13. Fialka F., Gruber R.M., Hitt R., Opitz L., Brunner E., Schliephake H. *et al.* (2008) CPA6, FMO2, LGI1, SIAT1 and TNC are differentially expressed in early- and late-stage oral squamous cell carcinoma – a pilot study. *Oral Oncol*;44:941–948.
  14. O'Malley P.G., Sangster S.M., Abdelmagid S.A., Bearne S.L., Too C.K. (2005) Characterization of a novel, cytokine-inducible carboxypeptidase D isoform in haematopoietic tumour cells. *Biochem J*;390:665–673.
  15. Ross P.L., Cheng I., Liu X., Cicek M.S., Carroll P.R., Casey G. *et al.* (2009) Carboxypeptidase 4 gene variants and early-onset intermediate-to-high risk prostate cancer. *BMC Cancer*;9:69.
  16. Tsakiris I., Soos G., Nemes Z., Kiss S.S., Andras C., Szanto J. *et al.* (2008) The presence of carboxypeptidase-M in tumour cells signifies epidermal growth factor receptor expression in lung adenocarcinomas: the coexistence predicts a poor prognosis regardless of EGFR levels. *J Cancer Res Clin Oncol*;134:439–451.
  17. Matsugi S., Hamada T., Shioi N., Tanaka T., Kumada T., Satomura S. (2007) Serum carboxypeptidase A activity as a biomarker for early-stage pancreatic carcinoma. *Clin Chim Acta*;378:147–153.
  18. Arolas J.L., Vendrell J., Aviles F.X., Fricker L.D. (2007) Metallo-carboxypeptidases: emerging drug targets in biomedicine. *Curr Pharm Des*;13:349–366.
  19. Rodriguez de la Vega M., Sevilla R.G., Hermoso A., Lorenzo J., Tanco S., Diez A. *et al.* (2007) Nna1-like proteins are active metallo-carboxypeptidases of a new and diverse M14 subfamily. *FASEB J*;21:851–865.
  20. Vendrell J., Aviles F.X., Fricker L.D. (2004) Metallo-carboxypeptidases. In: Messerschmidt A., Bode W., Cygler M., editors. *Metallo-carboxypeptidases*. Chichester: John Wiley & Sons, Ltd; p. 176–189.
  21. Lipscomb W.N., Strater N. (1996) Recent advances in zinc enzymology. *Chem Rev*;96:2375–2434.
  22. Plummer T.H. Jr, Ryan T.J. (1981) A potent mercapto bi-product analogue inhibitor for human carboxypeptidase N. *Biochem Biophys Res Commun*;98:448–454.
  23. Mao S.S., Colussi D., Bailey C.M., Bosserman M., Burlein C., Gardell S.J. *et al.* (2003) Electrochemiluminescence assay for basic carboxypeptidases: inhibition of basic carboxypeptidases and activation of thrombin-activatable fibrinolysis inhibitor. *Anal Biochem*;319:159–170.
  24. Fisher J.F., Mobashery S. (2006) Recent advances in MMP inhibitor design. *Cancer Metastasis Rev*;25:115–136.
  25. Walpole C.S., Wrigglesworth R. (1989) Enzyme inhibitors in medicine. *Nat Prod Rep*;6:311–346.
  26. Ikejiri M., Bernardo M.M., Bonfil R.D., Toth M., Chang M., Fridman R. *et al.* (2005) Potent mechanism-based inhibitors for matrix metalloproteinases. *J Biol Chem*;280:33992–34002.
  27. Kleinfeld O., Kotra L.P., Gervasi D.C., Brown S., Bernardo M.M., Fridman R. *et al.* (2001) X-ray absorption studies of human matrix metalloproteinase-2 (MMP-2) bound to a highly selective mechanism-based inhibitor. comparison with the latent and active forms of the enzyme. *J Biol Chem*;276:17125–17131.
  28. Brown S., Bernardo M.M., Li Z.-H., Kotra L.P., Tanaka Y., Fridman R. *et al.* (2000) Potent and selective mechanism-based inhibition of gelatinases. *J Am Chem Soc*;122:6799–6800.
  29. Fernandez D., Aviles F.X., Vendrell J. (2009) Aromatic organic compounds as scaffolds for metallo-carboxypeptidase inhibitor design. *Chem Biol Drug Des*;73:75–82.
  30. Lee M., Heseck D., Shi Q., Noll B.C., Fisher J.F., Chang M. *et al.* (2008) Conformational analyses of thiirane-based gelatinase inhibitors. *Bioorg Med Chem Lett*;18:3064–3067.
  31. Kim D.H., Chung S.J. (1995) Inactivation of carboxypeptidase A by 2-benzyl-3,4-epithiobutanoic acid. *Bioorg Med Chem Lett*;5:1667–1672.
  32. Kim Y.M., Kim D.H. (1996) Convenient preparation of all four possible stereoisomers of 2-benzyl-3,4-epoxybutanoic acid, pseudomechanism-based inactivator for carboxypeptidase A via a-chymotrypsin-catalyzed hydrolysis. *Bull Korean Chem Soc*;17:967–969.
  33. Lee S.S., Li Z.-H., Lee D.H., Kim D.H. (1995) (2R,3S)- and (2S,3R)-2-benzyl-3,4-epoxybutanoic acid as highly efficient and fast acting pseudomechanism-based inactivators for carboxypeptidase A: design, asymmetric synthesis and inhibitory kinetics. *J Chem Soc Perkin 1*;?????:2877–2882.
  34. Ryu S.-E., Choi H.-J., Kim D.H. (1997) Stereochemistry in inactivation of carboxypeptidase A. Structural analysis of the inactivated carboxypeptidase A by an enantiomeric pair of 2-benzyl-3,4-epoxybutanoic acids. *J Am Chem Soc*;119:38–41.
  35. Yun M., Park C., Kim S., Nam D., Kim S.C., Kim D.H. (1992) The x-ray crystallographic study of covalently modified carboxypeptidase A by 2-benzyl-3,4-epoxybutanoic acid, a pseudomechanism-based inactivator. *J Am Chem Soc*;114:2281–2282.
  36. Collaborative Computational Project, number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*;50:760–776.
  37. Cho J.H., Kim D.H., Lee K.J., Choi K.Y. (2001) The role of Tyr248 probed by mutant bovine carboxypeptidase A: insight into the catalytic mechanism of carboxypeptidase A. *Biochemistry*;40:10197–10203.
  38. Firth-Clark S., Kirtan S.B., Willems H.M., Williams A. (2008) De novo ligand design to partially flexible active sites: application of the ReFlex algorithm to carboxypeptidase A, acetylcholinesterase, and the estrogen receptor. *J Chem Inf Model*;48:296–305.
  39. Pallares I., Fernandez D., Comellas-Bigler M., Fernandez-Recio J., Ventura S., Aviles F.X. *et al.* (2008) Direct interaction between a human digestive protease and the mucoadhesive poly(acrylic acid). *Acta Crystallogr D Biol Crystallogr*;D64:784–791.
  40. Massova I., Martin P., de Mel S., Tanaka Y., Edwards B., Mobashery S. (1996) Crystallographic and Computational Insight on the Mechanism of Zinc-Ion-Dependent Inactivation of Carboxypeptidase A by 2-Benzyl-3-Iodopropanoate. *J Am Chem Soc*;118:12479–12480.
  41. Davis I.W., Murray L.W., Richardson J.S., Richardson D.C. (2004) MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res*;32:W615–W619.





# Analysis of a New Crystal Form of Procarboxypeptidase B: Further Insights into the Catalytic Mechanism

AQ2

Daniel Fernández,<sup>1,2</sup> Ester Boix,<sup>2</sup> Irantzu Pallarès,<sup>1,2</sup> Francesc X. Avilés,<sup>1,2</sup> Josep Vendrell<sup>1,2</sup><sup>1</sup> *Departament de Bioquímica i Biologia Molecular, Facultat de Biociències and Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain*<sup>2</sup> *Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain*

Received 10 June 2009; revised 22 September 2009; accepted 24 September 2009

Published online 00 Month 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bip.21320

## ABSTRACT:

A new triclinic crystal structure form of porcine pancreatic procarboxypeptidase B (PCPB) was obtained at higher resolution than the previously known tetragonal crystal structure. This new crystal polymorph has allowed for a corrected, accurate assignment of residues along the polypeptide chain based on the currently available gene sequence information and crystallographic data. The present structure shows unbound PCPB in a distinct molecular packing as compared to the previous benzamidine complexed form. Its catalytically important Tyr248 residue is oriented and hydrogen-bonded to solvent water molecules, and locates the furthest away from the catalytic zinc ion as compared to previous structures. A relatively long stretch of residues flanking Tyr248 and guarding the access to the catalytic zinc ion was found to be sequentially unique to the M14 family of peptidases. Predictions from a normal mode analysis indicated that this stretch of residues belongs to a rigid subdomain in the protein structure. The specific presence

of a tyrosyl residue at the most exposed position in this region would allow for a delicate balance between extreme hydrophobicity and hydrophilicity, and affect substrate binding and the kinetic efficiency of the enzyme.

© 2009 Wiley Periodicals, Inc. *Biopolymers* 00: 000–000, 2009.

**Keywords:** X-ray crystallography; crystal polymorphism; carboxypeptidase B; sequence analysis; M14 peptidase family; enzyme mechanism

This article was originally published online as an accepted preprint. The “Published Online” date corresponds to the preprint version. You can request a copy of the preprint by emailing the *Biopolymers* editorial office at [biopolymers@wiley.com](mailto:biopolymers@wiley.com)

## INTRODUCTION

The zinc-dependent metalloproteases (CPs), or M14 peptidases, are a group of enzymes that hydrolyze peptide substrates from the C-terminus. The biological functions of CPs range from the digestion of diet protein to tightly regulated processes such as fibrinolysis and hormone maturation.<sup>1</sup> Human pancreatic CPB, secreted as a proenzyme by the acinar cells, is a good diagnostic and prognostic variable in acute pancreatitis and nonpancreatic diseases.<sup>2–6</sup> Increasing evidence is being collected regarding the involvement of different CPs in cancer malignization processes,<sup>7,8</sup> suggesting that CPB may be one potential biomarker for the detection of early stage carcinomas. CPB-related enzymes secreted by the liver, like carboxy-

AQ5

Correspondence to: Ester Boix; e-mail: [josep.vendrell@uab.es](mailto:josep.vendrell@uab.es) or Josep Vendrell; e-mail: [josep.vendrell@uab.es](mailto:josep.vendrell@uab.es)

Contract grant sponsor: Ministerio de Ciencia e Innovación, Spain

Contract grant number: BIO2007-68046

Contract grant sponsors: Generalitat de Catalunya

Contract grant number: 2005SGR-1037

© 2009 Wiley Periodicals, Inc.

AQ6

2 Fernández et al.

**Table I** Statistics of Data Collection, Processing, and Refinement

Wavelength used during data collection	0.8123 Å
Unit Cell Constants	$a = 49.59 \text{ \AA}$ , $b = 49.97 \text{ \AA}$ , $c = 50.63 \text{ \AA}$ $\alpha = 109.9^\circ$ , $\beta = 104.5^\circ$ , $\gamma = 96.2^\circ$
Resolution range	19.63–1.88 Å
Space group	P1 (1mol / asymmetric unit)
Number of measured reflections	129,122
Number of unique reflections	33,758
$R_{\text{merge}}^a$ (overall/outermost shell)	5.0/22.2%
Completeness (overall/outermost shell)	96.4/88.3%
$I/\sigma I$ (overall/outermost shell)	18.6/5.2
Reflections used for refinement (total/test set)	33,081/676
Wilson plot $B$ factor (Å <sup>2</sup> )	23.5
Crystallographic $R_{\text{factor}}^b/R_{\text{free}}^c$	16.4/19.4%
r.m.s.d. bond lengths	0.009 Å
r.m.s.d. bond angles	1.05°
Number of protein atoms/total atoms	3195/3552
$B$ -factor statistics (Å <sup>2</sup> )	
Overall structure	17.1
Catalytic domain, main/side chain	13.5/15.0
Prodomain, main/side chain	18.5/20.7
Zn <sup>2+</sup> (1 in total/1 mol per monomer)	21.1
Glycerol atoms (30 in total/5 mols per monomer)	32.4
Solvent atoms (326 in total)	30.7

<sup>a</sup>  $R_{\text{merge}} = \sum_{\text{hkl}} \sum_{j=1}^N |I_{\text{hkl}} - I_{\text{hkl}}(j)| / \sum_{\text{hkl}} \sum_{j=1}^N I_{\text{hkl}}(j)$ , where  $N$  is the redundancy of the data. The outermost shell is 1.93–1.88 Å.

<sup>b</sup>  $R_{\text{factor}} = \sum_{\text{hkl}} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{\text{hkl}} |F_{\text{obs}}|$ , where the  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure factor amplitudes of reflection hkl.

<sup>c</sup>  $R_{\text{free}} = R_{\text{factor}}$  for a randomly selected 2% subset of reflections that were not used in refinement.

<sup>d</sup> According to Molprobit.<sup>17</sup>

AQ7

peptidase N and human plasma CPB (alternatively termed thrombin-activatable fibrinolysis inhibitor, TAFI, CPU, or CPR), are involved in angioedema,<sup>9</sup> diabetes,<sup>10,11</sup> as well as in fibrinolysis<sup>12,13</sup> and inflammation.<sup>14,15</sup>

The 3D crystal structure of porcine of pancreatic procarboxypeptidase B (PCPB) was the first to be obtained for a precursor of a metal-dependent proteolytic enzyme.<sup>16</sup> This original structure (PDB entry: 1nsa) belonging to the tetragonal  $P4_3$  space group and solved at 2.3 Å, showed that a small-molecule ligand (benzamidine), which was added to prevent activation from contaminant proteases, can induce conformational changes in the structure. Since only partial genome sequence information was available at the time the 1nsa structure was determined, we decided to derive the porcine CPB zymogen crystal structure in the light of current sequence data. The protein crystallized at a higher resolution and in a different crystal form (a crystal polymorph). In the region that guards the entrance to the active site cleft, Tyr248 is at the furthest distance from the catalytic zinc ion and hydrogen bonded to water molecules, as opposed to what is observed in the tetragonal crystal structure. The precise role of this residue during catalysis has generated much debate over the years, and still deserves attention as a mechanistic

model for this class of peptidases. A strategy combining structure data analysis from both the newly determined one and previous ligand-bound CPB forms and bioinformatics tools was followed to gain further insights into its behavior. In summary, the novel CPB zymogen crystal structure is presented here and details regarding the catalytic mechanism of action are discussed.

## RESULTS AND DISCUSSION

### Analysis of Sequence and Structural Data

A crystal that diffracted beyond 1.88 Å and belonging to the triclinic space group P1 was characterized throughout this study. Crystallographic details are presented in Table I. The structure has been validated and deposited with the PDB with accession code 3glj.

The resolution now obtained is higher than that of the previously determined tetragonal crystal (PDB code: 1nsa),<sup>16</sup> for which only partial sequence information was available at the time the structure was solved. In the present structure, all residues were assigned according to the updated sequence information<sup>18</sup> and carefully checked in the electron density

T1

Biopolymers

**Table II** Amino Acid Side Chains that Have Been Assigned According to Sequence and Confirmed by Structural Information<sup>a</sup>

This work/sequence	Insa <sup>b</sup>	This work/sequence	Insa
<i>Leu25<sup>c</sup></i>	<i>Glu</i>	Thr154	Val
<i>Arg93</i>	<i>Val</i>	Thr158	Val
<i>Val94</i>	<i>Ser</i>	Asp159	Asn
Lys23	Glu	Asp162	Asn
Glu28	Lys	Asn216	Asp
Thr36	Ser	Asn221	Ser
Leu43	Asp	Ala225	Gly
Pro57	Ser	Thr232	Ser
His76	Gln	Lys237	Ser
Glu85	Asp	Thr239	Ser
Leu88	Arg	Ala244	Ser
Ser94	Ala	Asp260	Asn
Asn101	Asp	Tyr277	Phe
Lys102	Asn	Ile280	Val
Thr136	Ser	Glu291	Gln
Thr137	Ser	Ile297	Val
Ile139	Thr	Val304	Thr
Asp148	Asn	Gly306	Glu

<sup>a</sup> UniProt accession code P09955.<sup>b</sup> Previously determined structure; see Ref. 16.<sup>c</sup> Residues in *italics* belong to the pro-domain.

AQ3 maps. The alignment of 393 residues revealed 36 discrepancies (Table II), representing about a 10% of the total. At the resolution of the previous crystal structure determination some of them, that is, Asp/Asn, Gln/Glu, Val/Thr, might be undetectable in the electron density maps, while others involve larger changes on the side chains shape. Some examples of misassignments are Leu25, previously assigned as a glutamate despite being located in a hydrophobic environment, Arg93, previously interpreted as a valine, which resides in an open space where no interactions are observed for the guanidinium N atoms, or residues Leu43, Pro57, Leu88, Ile139, and Ala244 in the enzyme moiety that reside in hydrophobic environments, which is against the polar side chains that had been assigned before (for a complete description of the numbering system used with metalloproteases see García-Sáez et al.<sup>19</sup>). Overall, the quality of the electron density maps obtained in our structure allowed us to assign all residues in accordance to the sequence information. An example of this procedure is residue Pro57 in the catalytic domain (see Figure 1).

### Crystal Packing Interactions in the Two PCPB Polymorphic Forms

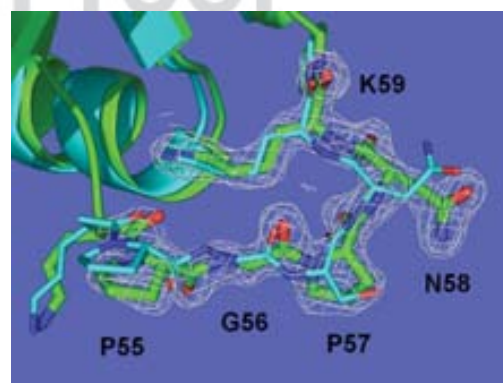
The newly determined porcine CPB zymogen structure crystallized in the triclinic P1 space group with one independent

Biopolymers

polypeptide chain in the asymmetric unit and cell constants  $a = 49.59 \text{ \AA}$ ,  $b = 49.97 \text{ \AA}$ ,  $c = 50.63 \text{ \AA}$ ,  $\alpha = 109.9^\circ$ ,  $\beta = 104.5^\circ$ ,  $\gamma = 96.2^\circ$ . The triclinic CPB zymogen structure encompasses residues Phe7-Arg95 from the N-terminal pro-domain and Thr4-Leu308 from the C-terminal catalytic domain and displays one catalytic zinc ion per polypeptide chain.

A comparison to the known tetragonal CPB zymogen structure (Insa) in the P4<sub>3</sub> space group and also one molecule in the asymmetric unit, revealed no gross changes in the overall structure. Some differences are local or involve the whole polypeptide chain, for example, when comparing the way molecules are packed in the crystal lattice. The unit cell constants of the tetragonal crystal, viz.  $a = b = 103.1 \text{ \AA}$ ,  $c = 46.6 \text{ \AA}$ ,  $\alpha = \beta = \gamma = 90.0^\circ$ , indicate that two cell dimensions are halved ( $\sim a/2$  and  $b/2$ ) in the P1 form while the third one is conserved. The triclinic unit cell volume is roughly a quarter of the tetragonal one. Besides, the orientation among symmetry-related equivalents is shifted in both crystal forms owing to different intermolecular packing interactions (see Figure 2).

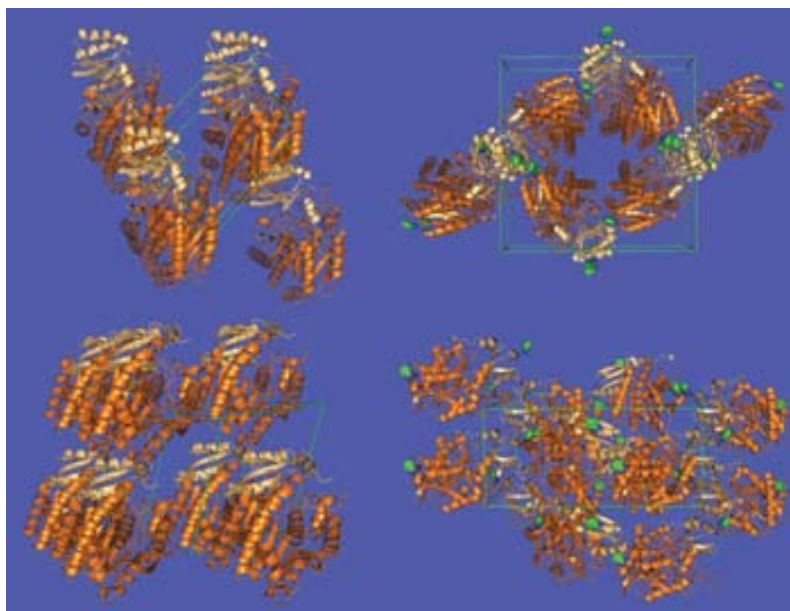
In the earlier CPB zymogen structure determination (Insa), benzamidine was added to avoid activation by proteolytic cleavage. Two of the three benzamidine molecules present in the tetragonal structure are placed over the surface, interacting with symmetry-related protein copies. In the triclinic crystal, no such small-molecule was added for crystallization and the arrangement of CPB zymogen molecules shows an alternate pattern of prodomains and catalytic domains (Figure 2, upper row, left panel). In contrast, the



**FIGURE 1** Partial view of CPB structure localized around residue Pro57. In this work (thick sticks model, green carbons), a proline residue fits into the electron density (gray mesh) as opposed to a serine assigned in the previously reported structure (Insa) (thin sticks, cyan carbons). The weighted 2Fo-Fc map is contoured at a 1.5 sigma level.

F2

C  
O  
L  
O  
R



**FIGURE 2** Comparison of the molecular packing of the PCPB zymogen in two space groups. Left panel, the triclinic cell (this work) and right panel, the tetragonal cell (Insa). Upper row: the view is with the shortest unit cell vector pointing toward the reader. Lower row: the view is horizontally rotated 90° with respect to the upper figures. The prodomain is shown in beige, while the catalytic domain is in orange. The benzamidinium molecules in the tetragonal cell (green carbons) are depicted as a CPK model.

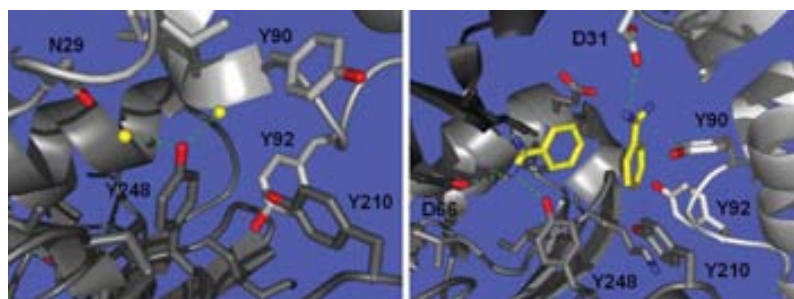
molecular packing in the tetragonal cell is mediated by stacking of protein molecules along the shortest unit cell axis (Figure 2, upper row, right panel), forming columns of prodomains and catalytic domains. A detailed view of one such intermolecular packing interaction involves the catalytically important residue Tyr248. In the triclinic unit cell (Figure 3; left panel), the phenolic hydroxyl oxygen engages two solvent water molecules while Tyr248 is bound to one benzamidinium molecule in a micro-environment devoid of water in the tetragonal crystal (Figure 3; right panel).

#### The Role of Tyr248 Side Chain and Its Flanking Region

Tyr248 is a strictly conserved residue for all M14 peptidases. The precise role of Tyr248 phenolic ring during catalysis has been a matter of debate over the years. Kinetic measurements have contributed to two different views about the role of Tyr248. Working with the canonical M14 peptidase CPA, some evidences pointed out that the phenolic hydroxyl would not play a crucial role during catalysis.<sup>20</sup> However, later investigations proposed that the phenolic hydroxyl is

crucial for catalysis by providing further activation of the zinc-bound water molecule and leading to the generation of the reactive species that attacks the scissile peptide bond, and is also important for ligand anchoring inside the active site cleft.<sup>21</sup> It should be mentioned that most members of the newly described cytosolic carboxypeptidases (CCPs) have a Phe in the position equivalent to Tyr248 after sequential alignment,<sup>22</sup> indirectly supporting the notion that the side chain at this position does not influence catalysis. However, structural and kinetic data on the members of what has been proposed to be a new M14D subfamily<sup>23</sup> are necessary to clarify this issue.

The kinetic aspects of peptidase mechanism of action commented upon above would entail various consecutive steps such as trapping the ligand from solution, binding, escorting it to the active site cleft, and anchoring it facing the catalytic zinc ion, to finally release the products into the solvent and prepare the enzyme for a new round of peptidase activity. Some of these events are poorly understood and may be technically difficult to assess. Structural data indicate that Tyr248 aromatic ring adopts the “closed” (or “down”) conformation when it acts as a cap facing a ligand bound to the



**FIGURE 3** Partial packing diagram showing the interactions of Tyr248. In the triclinic cell (left panel), the phenolic ring is bound to two crystallization water molecules. In the tetragonal cell (right panel), Tyr248 hydroxyl is hydrogen bonded to one cocrystallization benzamidine molecule. Symmetry-related polypeptide chains are shown in tones of gray. Benzamidine molecules are in sticks model with yellow carbons. Oxygen and nitrogen atoms are colored red and blue, respectively. Hydrogen bond interactions are shown as green dashed lines. Relevant side chains are shown as sticks model and labeled.

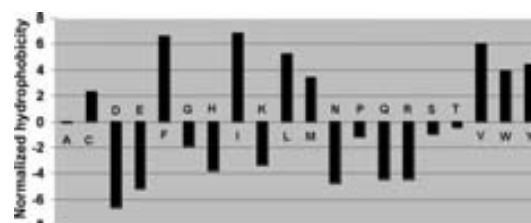
enzyme, or the “open” (or “up”) conformation when exposed to the solvent. The catalytic mechanism would be operating between these limit structures and, according to the latest kinetics data, Tyr248 would play a key role during it. Further details emerged as we investigated the triclinic PCPB by a combination of crystallographic and bioinformatic analyses.

First, to detect rigid regions or subdomains in the newly determined PCPB crystal structure, a normal mode analysis was performed using HingeProt.<sup>24</sup> Two lowest modes with hinge residues 109, 185, 202, 220, 239, and 268 (Mode 1), defining four rigid parts, and 14, 64, 82, and 192 (Mode 2) yielding three rigid parts were predicted. According to this model, the enzyme elongates/shrinks in such a way that regions shaping the entrance to the active site cleft produce a mouth-like movement. Interestingly, one of the subdomains predicted by Mode 2 is almost coincident with previous results obtained from CPA.<sup>25,26</sup> Besides, this subdomain harbors the most mobile side chain in the structure, Tyr248, whose flexibility in CPA has already been investigated.<sup>17,27</sup> From our observations, hinge residues 239 and 268 in CPB would be the boundaries of a rigid region linked to the catalytic mechanism that contains the mobile Tyr248 side chain. The 239–268 region encompasses  $\beta$ -strand 11, the short  $\alpha$ -helix 10, the unstructured sequence 248–253,  $\alpha$ -helix 10, and part of the  $\beta$ -strand 12 (see Figure 5a). Sequence alignments indicate that this 29 AA sequence is uniquely present in M14 peptidases, with highest scores to CPB and also to various CPA isoforms, suggesting that this region of the carboxypeptidase molecule evolved to perform a function aside from merely holding the Tyr248 aromatic ring in place.

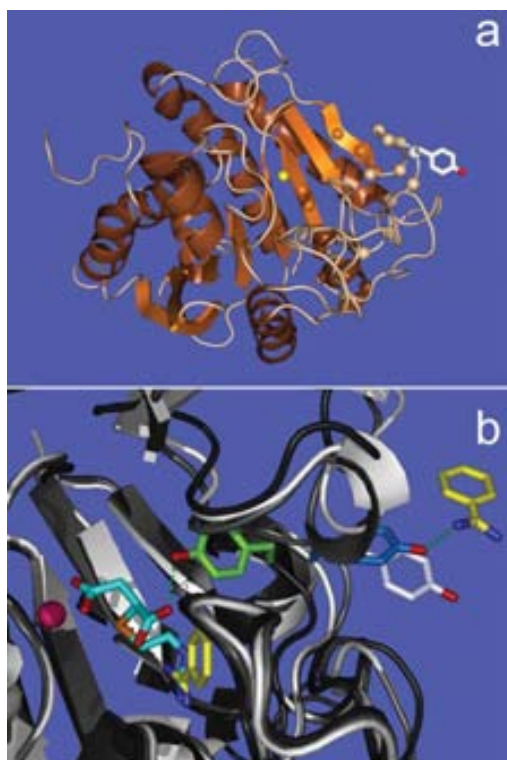
*Biopolymers*

Next, the physicochemical nature of the wild type sequence and its mutants was computationally investigated using the aggregation propensity scale implemented in the AGGRESCAN software, which assigns high aggregation propensities to hydrophobic amino acids such as Ile, Phe, Val, Leu, Trp, Tyr, and Val and low aggregation propensities to Asp, Glu, Asn, Arg, Gln, His, and Lys.<sup>28</sup> Considering that peptidases bind to peptide substrates during their normal mechanism of action, the aggregation propensity scale used here, regarded in the broader sense of protein–protein interactions, may be a valid measure of the ligand–protease association process. Figure 4 shows the predicted hydrophobicity value for the 239–268 sequences depending on the nature of the side chain at position 248. This single change suffices to introduce a dramatic shift in the overall properties of the

F4



**FIGURE 4** Sequence normalized hydrophobicity for the 239–268 sequence of CPB as computed by AGGRESCAN. The overall hydrophobicity value for each one of 20 mutants at position 248 was estimated from the experimental aggregation propensity calculated using a 5-residue sliding window. The data shown are absolute, normalized values that result from subtraction of the overall average value from the normalized hydrophobicity of each mutant.



**FIGURE 5** (a) Cartoon representation of PCPB structure highlighting the region limited by hinge residues 239 and 268 that contains Tyr248. The position of the C $\alpha$  atoms is indicated by a sphere and the side chain of Tyr248 is shown in stick model. Shown here is the structure corresponding to the white Tyr248 in part (b) of the Figure. (b) A close-up view of Tyr248 aromatic ring displacements in three superimposed CPB structures: unbound-CPB, this work (gray cartoon, Tyr248 in white); benzamidine-bound CPB (1nsa; white cartoon, Tyr248 light blue); and a complex with a bi-product analog inhibitor (3D67; black cartoon, Tyr248 light green). The catalytic zinc ion is shown as a magenta sphere (far left). Benzamidine molecules (yellow carbons) and the bi-product inhibitor GEMSA (cyan carbons) are in sticks model. Other atoms are red (oxygen), blue (nitrogen), and orange (sulfur). Hydrogen bonds of Tyr248 to small-molecule ligands are indicated as green dashed lines. The displacement of Tyr248 aromatic ring may be viewed from right to left (white, blue, green Tyr side chain and gray, white, black main chain).

sequence. Phe, Ile, Leu, and Val confer the highest hydrophobicity to the whole sequence. The wild type Tyr248 sequence scores in a second subset, together with Trp and Met, which differs widely from the more hydrophilic sequences created by the polar side chains as well as by Ala, Gly, and Thr. Kinetic data using peptide substrates showed that the *k<sub>cat</sub>*/

*K<sub>M</sub>* ratios are reduced about 50- and 300-fold for the Tyr248-Phe and Tyr248Ala mutations, respectively,<sup>21</sup> indicating the preference for a tyrosyl residue at the entrance of the active site cleft. Aside from considerations about the involvement of the hydroxyl group in catalysis (see further on), we might propose that part of the effects of the mutations over the enzymic activity may arise from ligand recognition and binding. The presence of Tyr at position 248 inside the sequence 239–268 suggests a choice over the more hydrophobic Phe-containing sequence, to which a peptide ligand may be more prone to stick and remain bound, and the less hydrophobic Ala-containing sequence that would be less suited to desolvate the peptide substrate. As a consequence, for Ala or Phe substitutions, the overall physicochemical nature of the sequence would impair the catalytic efficiency due to a poorer or stronger desolvation effect as compared to Tyr. These aspects of ligand binding may be complemented by the action of the phenolic hydroxyl group as discussed next.

A superimposition of three PCPB crystal structures, two in the open conformation and one in the closed conformation, shows that, in spite of overall structural similarity, minor local differences are visible. In the unbound PCPB (this work) corresponding to the triclinic unit cell (Figure 5b; gray trace), the phenolic hydroxyl oxygen engages two solvent water molecules and lies at the largest distance from the zinc ion, while in the tetragonal cell (PDB code 1nsa) the benzamidine molecule would displace the corresponding water molecules (Figure 5b; white trace). Here, the Tyr248 aromatic ring is within a different local environment due to the presence of a second symmetry-related benzamidine molecule (see also Figure 3) and approaches the active site cavity entrance. In the human plasma PCPB structure bound to the bi-product analog *DL*-guanidinoethylmercaptosuccinic acid (GEMSA) (PDB code 3d67),<sup>29</sup> the Tyr248 aromatic ring adopts the closed conformation with the phenolic hydroxyl pointing inward to the catalytic zinc ion (Figure 5b; black trace). From these observations, we may envisage that the phenolic hydroxyl initially serves as a point of anchorage and later as a pivot that moves inward to the catalytic zinc ion after a swing of  $\sim 120^\circ$  from the solvent-exposed open conformation. At the same time, the protein main chain tracks the movement of Tyr248 as can be observed in Figure 5 by the displacement of the backbone from right to left.

## CONCLUSION

The new triclinic structure of the porcine CPB zymogen displays a packing different from the previously known tetragonal form and has been determined at a higher crystallographic resolution. The new structure has allowed for the

accurate assignment of the complete polypeptide chain in accordance to current genomic sequence information. Sequence 239–268, guarding the access to the catalytic zinc ion, is only present in M14A peptidases and encompasses a rigid region in the protein flanked by hinge residues. This sequence holds the catalytically important position 248, whose side chain is hydrogen bonded to solvent molecules in the new structure. Comparison of this region with the corresponding ones in other ligand-bound structures by a combination of structural and bioinformatic approaches confirms that a phenolic ring is important at this position for controlling the binding of ligands and the access to the active site cleft in the following terms: (i) the phenolic hydroxyl group establishes a hydrogen bond to solvent water molecules in the absence of ligand; when the latter is present, water is substituted in a process that brings about a shift in the side-chain conformation from a fully solvated open state to the dry closed one; and (ii) the Tyr248 phenolic ring modulates the hydrophobicity of its neighboring sequence, which guards the entrance to the active site cleft. Mutations at this position affect the overall physicochemical properties of the sequence and probably its ligand binding abilities.

Summarizing, the use of a combination of techniques, beyond the redetermination of the crystal structure of a model zinc-dependent peptidase at high resolution, has brought further insight into its reactivity and mode of action in an approach that may be generally applicable to a broad range of different enzymes.

## MATERIALS AND METHODS

### Protein Preparation and Crystallization

The porcine pancreatic CPB was recombinantly expressed in the methylotropic *P. pastoris* system as published.<sup>18</sup> This yielded large amounts of protein sample that was purified through a hydrophobic interaction butyl column and eluted with an increasing gradient of ammonium sulfate. The zymogen-containing fractions were selected and then desalted and concentrated in Centricon centrifugal filter devices (Millipore, Billerica, MA) against 5 mM Tris, pH 7.5. The protein was crystallized in the sitting drop setting at room temperature. Small prism-shaped crystals appeared in drops containing a 1:1  $\mu$ l mixture of protein (in 5 mM Tris, pH 7.5, at 14 mg/ml) and reservoir solution (comprising 1.6M ammonium sulfate and 0.1M sodium citrate, pH 5.6). Some of these crystals were harvested and used as seeds to grow larger ones in the same reservoir solution.

### Data Collection and Processing

X-ray diffraction experiments were performed at the X11 and X13 beamlines of the EMBL (European Molecular Biology Laboratory)-Outstation at DESY (Deutsches Elektronen Synchrotron), Hamburg, Germany. Crystals of adequate size for X-ray diffraction were

harvested and then cryoprotected (in the reservoir solution containing glycerol 30%) and flash-frozen in the N<sub>2</sub> stream. The structure was solved by molecular replacement using 1nsa as the model. X-ray diffraction data were processed with XDS and XSCALE.<sup>30</sup> Programs included in the CCP4 suite<sup>31</sup> were used at different stages of refinement, structure determination, and data analysis. Model building and map fitting were made with COOT.<sup>32</sup> All the figures were prepared with Pymol, DeLano Scientific© (www.pymol.org). Full crystallographic details are listed in Table I.

### Bioinformatic Analysis

HingeProt<sup>24</sup> is found at <http://www.prc.boun.edu.tr/appserv/prc/hingeprot/>. Sequence searches were performed using BLAST at <http://www.expasy.ch/tools/blast/>. From Hingeprot results, the 29 amino acid sequence query included residues 240–268: NH<sub>2</sub>-YGP GATTIYPAAGGSDDWAYDQGIKYSFT-COOH. AGGRESKAN,<sup>28,33</sup> based on a scale of relative aggregation propensities of the 20 natural amino acids experimentally derived from the analysis of mutants in the central position of the CHC in amyloid- $\beta$ -protein, was accessed at the URL: <http://bioinf.uab.es/aggrescan/>. The program assigns the experimentally derived aggregation propensity value to each amino acid in the sequence in the first place to calculate the final aggregation value from the average of the sum of the values in a 5-residue sliding window, and normalizing the average by the length of the polypeptide chain multiplied by 100. The normalized total average value was used in this work for the hydrophobicity prediction of the 239–268 AA sequence.

Natàlia Sánchez de Groot is thanked for useful comments on the use of AGGRESKAN. The authors wish to express their gratitude to the local staff at the EMBL (European Molecular Biology Laboratory)-Outstation at DESY (Deutsches Elektronen Synchrotron), Hamburg, Germany, for excellent assistance during X-ray data collection.

## REFERENCES

- Vendrell, J.; Aviles, F. X.; Fricker, L. D. In Handbook of Metalloproteins; Messerschmidt, A.; Bode, W.; Cygler, M., Eds.; Wiley: Chichester, 2004, pp 176–189.
- Muller, C. A.; Appelros, S.; Uhl, W.; Buchler, M. W.; Borgstrom, A. Gut 2002, 51, 229–235.
- Pezzilli, R.; Morselli-Labate, A. M.; Barbieri, A. R.; Plate, L. JOP 2000, 1, 58–68.
- Regner, S.; Appelros, S.; Hjalmarsson, C.; Manjer, J.; Sadic, J.; Borgstrom, A. Pancreatolgy 2008, 8, 42–49.
- Borgstrom, A.; Regner, S. Pancreatolgy 2005, 5, 530–536.
- Saez, J.; Martinez, J.; Trigo, C.; Sanchez-Paya, J.; Company, L.; Laveda, R.; Grino, P.; Garcia, C.; Perez-Mateo, M. World J Gastroenterol 2005, 11, 7261–7265.
- Tsakiris, I.; Soos, G.; Nemes, Z.; Kiss, S. S.; Andras, C.; Szanto, J.; Dezzo, B. J Cancer Res Clin Oncol 2008, 134, 439–451.
- Vairaktaris, E.; Yapijakis, C.; Nkenke, E.; Vassiliou, S.; Vylliotis, A.; Nixon, A. M.; Derka, S.; Ragos, V.; Spyridonidou, S.; Tsigris, C.; Neukam, F. W.; Patsouris, E. Am J Hematol 2007, 82, 1010–1012.
- Willemse, J. L.; Chen, D.; Hendriks, D. F. Clin Chim Acta 2008, 389, 181–182.

8 *Fernández et al.*

10. Kitagawa, N.; Yano, Y.; Gabazza, E. C.; Bruno, N. E.; Araki, R.; Matsumoto, K.; Katsuki, A.; Hori, Y.; Nakatani, K.; Taguchi, O.; Sumida, Y.; Suzuki, K.; Adachi, Y. *Diabetes Res Clin Pract* 2006, 73, 150–157.
11. Rigla, M.; Wagner, A. M.; Borrell, M.; Mateo, J.; Foncuberta, J.; de Leiva, A.; Ordonez-Llanos, J.; Perez, A. *Metabolism* 2006, 55, 1437–1442.
12. Mari, D.; Coppola, R.; Provenzano, R. *Exp Gerontol* 2008, 43, 66–73.
13. Leebeek, F. W.; Goor, M. P.; Guimaraes, A. H.; Brouwers, G. J.; Maat, M. P.; Dippel, D. W.; Rijken, D. C. *J Thromb Haemost* 2005, 3, 2211–2218.
14. Leung, L. L.; Myles, T.; Nishimura, T.; Song, J. J.; Robinson, W. H. *Mol Immunol* 2008, 45, 4080–4083.
15. Rooth, E.; Wallen, H.; Antovic, A.; von Arbin, M.; Kaponides, G.; Wahlgren, N.; Blomback, M.; Antovic, J. *Blood Coagul Fibrinolysis* 2007, 18, 365–370.
16. Coll, M.; Guasch, A.; Aviles, F. X.; Huber, R. *EMBO J* 1991, 10, 1–9.
17. Zen, A.; Carnevale, V.; Lesk, A. M.; Micheletti, C. *Protein Sci* 2008, 17, 918–929.
18. Ventura, S.; Villegas, V.; Sterner, J.; Larson, J.; Vendrell, J.; Hersberger, C. L.; Aviles, F. X. *J Biol Chem* 1999, 274, 19925–19933.
19. García-Sáez, I.; Reverter, D.; Vendrell, J.; Avilés, F. X.; Coll, M. *EMBO J* 1997, 16, 6906–6913.
20. Gardell, S. J.; Hilvert, D.; Barnett, J.; Kaiser, E. T.; Rutter, W. J. *J Biol Chem* 1987, 262, 576–582.
21. Cho, J. H.; Kim, D. H.; Lee, K. J.; Choi, K. Y. *Biochemistry* 2001, 40, 10197–10203.
22. Kalinina, E.; Biswas, R.; Berezniuk, I.; Hermoso, A.; Aviles, F. X.; Fricker, L. D. *FASEB J* 2007, 21, 836–850.
23. Rodriguez de la Vega, M.; Sevilla, R. G.; Hermoso, A.; Lorenzo, J.; Tanco, S.; Diez, A.; Fricker, L. D.; Bautista, J. M.; Aviles, F. X. *FASEB J* 2007, 21, 851–865.
24. Emekli, U.; Schneidman-Duhovny, D.; Wolfson, H. J.; Nussinov, R.; Haliloglu, T. *Proteins* 2008, 70, 1219–1227.
25. Liebman, M. N.; Venanzi, C. A.; Weinstein, H. *Biopolymers* 1985, 24, 1721–1758.
26. Pritchard, L.; Cardle, L.; Quinn, S.; Dufton, M. *Protein Eng* 2003, 16, 87–101.
27. Firth-Clark, S.; Kirton, S. B.; Willems, H. M.; Williams, A. *J Chem Inf Model* 2008, 48, 296–305.
28. Sanchez de Groot, N.; Pallares, I.; Aviles, F. X.; Vendrell, J.; Ventura, S. *BMC Struct Biol* 2005, 5, 18.
29. Marx, P. F.; Brondijk, T. H.; Plug, T.; Romijn, R. A.; Hemrika, W.; Meijers, J. C.; Huizinga, E. G. *Blood* 2008, 112, 2803–2809.
30. Kabsch, W. *J Appl Crystallogr* 1988, 21, 916–924.
31. Collaborative Computational Project, Number 4. *Acta Crystallogr D Biol Crystallogr* 1994, 50, 760–776.
32. Emsley, P.; Cowtan, K. *Acta Crystallogr D Biol Crystallogr* 2004, 60, 2126–2132.
33. Conchillo-Sole, O.; de Groot, N. S.; Aviles, F. X.; Vendrell, J.; Daura, X.; Ventura, S. *BMC Bioinformatics* 2007, 8, 65.

*Reviewing Editor: Eric Toone*

AQ4

Author Proof



## Aromatic Organic Compounds as Scaffolds for Metalloproteinase Inhibitor Design

Daniel Fernández, Francesc X. Avilés and Josep Vendrell\*

Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, and Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain

\*Corresponding author: Josep Vendrell, josep.vendrell@uab.cat

**We have identified and characterized a set of quinoline, naphthalene and quinazoline derivatives as inhibitors of metalloproteinases, a class of metal-dependent proteolytic enzymes. The aromatic organic compounds were selected from a high-throughput screening survey and, with some exceptions, showed a good correlation between inhibitory potency and docking energy value. The *in vitro* inhibition tests gave  $K_i$  values in the lower micromolar range for metalloproteinases with different specificities, and a tendency to behave as more powerful inhibitors of CPB was observed for most of the compounds tested. The kinetic results were further analyzed by structural analysis via molecular docking. The most potent aromatic organic inhibitor docks to human CPB mostly through burial of its hydrophobic moiety deep into the enzyme's active site cleft and by interacting with the catalytic zinc ion. The significance of our results in designing inhibitors against disease-related CPs from the identified ligands is examined herein.**

**Key words:** aromatic organic derivative, enzyme inhibition, human carboxypeptidase B, *in silico* screening, metalloproteinases, molecular docking

Received 24 July 2008, revised and accepted for publication 13 November 2008

Aromatic organic compounds have attracted special interest in the design of metalloproteinase (CP) ligands. Naphthalene derivatives of alanine (1), cysteine (2) and dansylated amino acids and peptides (3), were shown to exert potent CPA inhibition. A deep structural analysis of this class of compounds has not yet been performed, but docking of a naphthaldehyde hydrazone inhibitor ( $IC_{50} = 29 \mu M$ ) showed that the compound binds CPA by burying the naphthalene ring into a hydrophobic pocket along the S1–S2 subsites of the enzyme (4). A series of internally quenched dansylated tryptophan peptides were developed to monitor CPA proteolytic activity

(5) and dansylated arginine/lysine peptides were obtained for the chromatographic determination of CPB-like protease activity (6). Other bicyclic aromatic systems seem to act in a different manner, as exemplified by folate analogs. Some of them, like the antineoplastic methotrexate inhibit CPB through a non-competitive and rather complicated kinetics involving several units per enzyme molecule (7). Although more potent than the monocyclic counterparts, the fact that folates bind unspecifically to different classes of enzymes and their intricate mode of action may have prevented further developments within this class of compounds (8).

Metalloproteinases are metal-dependent exopeptidases with a wide range of physiological functions (9). More than 25 genes coding CPs in the human genome have been identified so far and their number tends to grow as novel CPs are being discovered (10). Among metalloproteinases of biomedical interest, human CPA1 is a pancreatic enzyme that has been shown to be a marker and a valid diagnostic tool of early stage pancreatic carcinoma (11,12), while human plasma CPB, alias carboxypeptidase U (CPU) or thrombin-activatable fibrinolysis inhibitor (TAFI), is a plasma circulating enzyme which functions as an antifibrinolytic agent and also participates in inflammatory processes (13,14). TAFI has become a promising target for the treatment of thrombosis (15,16) and the interest to develop pharmaceuticals targeted to it is reflected in the wealth of information available from scientific and patent literature (17–20). One recently disclosed molecule, (*S*)-2-[3-(aminomethyl)phenyl]-3-(hydroxyl [(*R*)-2-methyl-1-[[[3-phenylpropyl)sulfonyl] amino propyl] phosphoryl propanoic acid, (BX 528), showed promising properties when tested in animal models of thrombolytic disease (21,22). In spite of all these advances, no compound of this class has reached the clinical practice yet. Regarding other biotechnological interests, invertebrate CPs are putative targets for biological control of crop pest insects. We recently described a corn earworm (*Helioverpa zea*) CPB (HzCPB) as a digestive protease specifically expressed by the organism in response to plant proteinaceous inhibitors (23). The previous structural characterization of inhibitor-insensitive HzCPB (24) provides a template for the development of novel chemical agents to protect agronomically important crops from pest infestation. Similarly, the availability of the x-ray crystal structures of canonical CPs, like human CPA1 in complex with a drug-protecting polymer (25) and the human CPB zymogen (26), might assist in the discovery of novel agents targeted to metal-dependent proteases.

We aimed our research to the discovery of novel scaffolds for the development of drugs targeted to CPs, prompted by the lack of available drugs and the need to improve the properties of the currently described inhibitors. Bioavailability is a key problem in

protease inhibitor optimization (27). In this regard, the potent CPU inhibitor, BX 528, displayed a poor bioavailability when tested *in vivo* (21). When addressing this question by a high-throughput screening survey of scaffolds in a major screening collection database, we focused on the search of new hits that might be used to improve the oral absorption of CP inhibitors. A number of aromatic organic compounds, containing quinoline, naphthalene and quinazoline functionalities were retrieved using this procedure and further analyzed, together with two aminoisoquinoline compounds as simpler, smaller molecular weight versions of the selected molecules. Compounds bearing those functional groups have found widespread application in some therapeutic areas. For example, quinapril is an isoquinoline antihypertensive inhibitor of angiotensin-converting enzyme (ACE) (28). Doxazosin, a quinazoline selective  $\alpha_1$ -adrenergic blocker, is used for the treatment of hypertension and benign prostatic hypertrophy (29). The aromatic organic compounds identified from a database screening procedure were subsequently characterized by molecular docking as well by *in vitro* inhibition assays against CPs of different specificities and natural origin. Kinetic determinations included the benchmark protease bovine CPA (bCPA), the inhibitor-insensitive insect HzCPB, and human CPB, as a model for human disease-linked proteases.

## Methods and Materials

### General procedures and materials for the biological experiments

Recombinant human CPB, and insect CPB from *H. zea*, were expressed in the zymogen form in the methylotrophic *Pichia pastoris* system. They were then purified and proteolytically activated as published (24,30). Bovine pancreatic CPA was from Sigma (St Louis, MO, USA). Enzyme concentrations for the kinetic studies were kept fixed at typically 5–50 nM. The substrates, *N*-(4-methoxyphenylazoformyl)-Phe-OH (Aaf-Phe-OH;  $K_m$  for CPA: 100  $\mu$ M) and *N*-(4-methoxyphenylazoformyl)-Arg-OH (Aaf-Arg-OH;  $K_m$  for CPB: 60  $\mu$ M) were from Bachem (Bubendorf, Switzerland). The final Aaf-Phe-OH and Aaf-Arg-OH concentrations used were 100 and 60 or 200  $\mu$ M, respectively. The experimental assays were performed in 50 mM Tris, 0.5 M NaCl, pH 7.5 buffer and 20 mM Tris, 0.1 M NaCl, pH 7.5 buffer for CPA and CPB, respectively, with variable concentrations of the analyzed compounds between 50 nM and 1  $\mu$ M. The initial velocity measurements were performed at least per triplicate at room temperature in a Victor3 (PerkinElmer, Waltham, MA, USA) microtiter plate reader. The changes in absorbance were followed continuously at 340 nm. The GraphPad version 5.0 package (GraphPad Software, San Diego, CA, USA) was used to process the data.

### Docking procedures, structure analysis and visualization

The X-ray crystal structures of human CPB (PDB code: 1ZLI), corn earworm (*H. zea*) CPB (2C1C) and bovine CPA (2CTC) were used for docking. The approach was validated by redocking the small-molecule ligands from known bound CPA (2CTC) and CPB (2JEW) complex structures. The high-throughput screening was performed using human CPB as the target (PDB accession 1ZLI). The August 2006 version of ASINEX Platinum screening collection<sup>a</sup> (APC; size

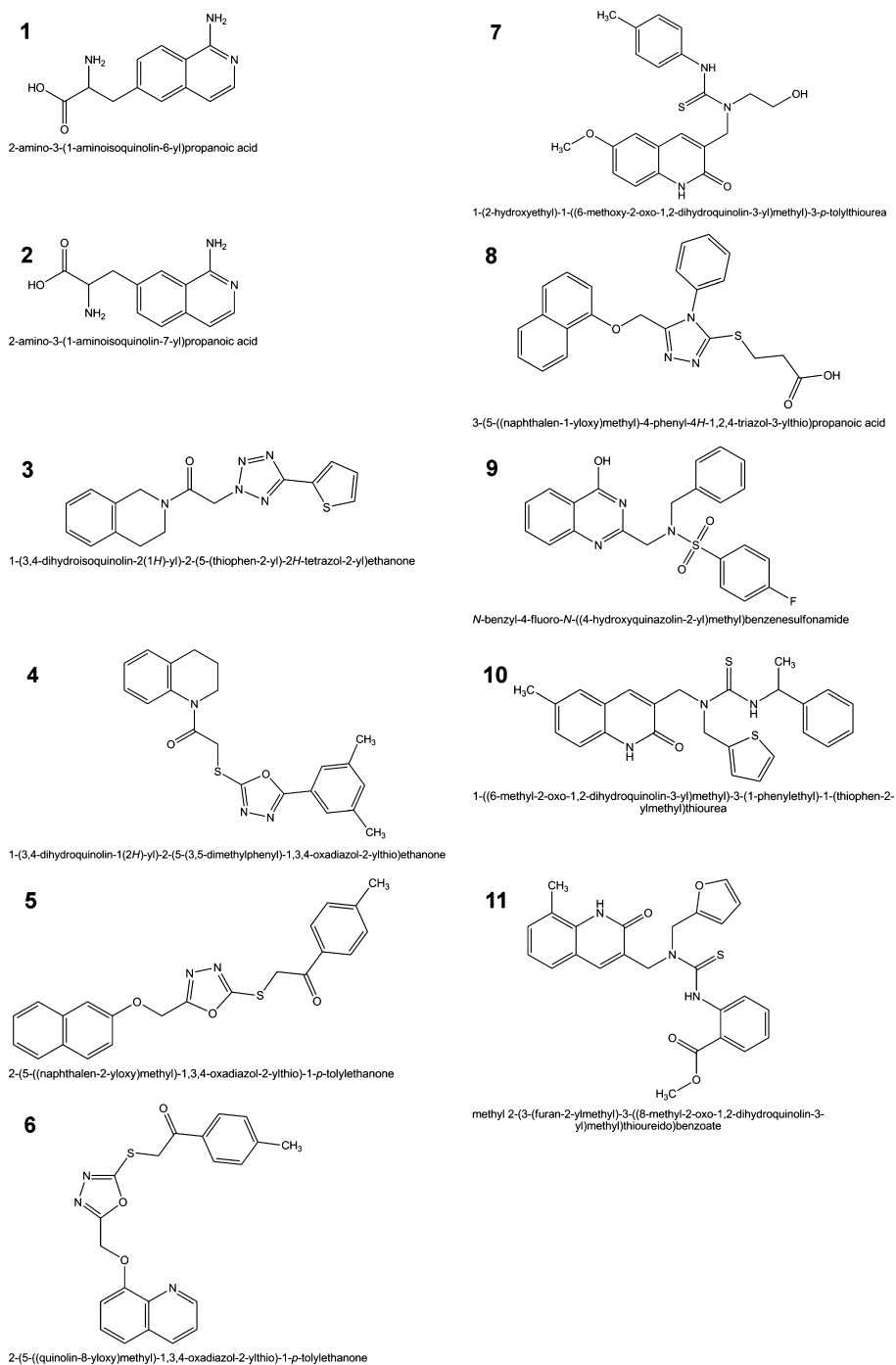
>125 000) was queried, dividing it into small size sublibraries (2500 compounds each) to ease the screening procedure. Polishing of the APC, creation of the sublibraries and 3D model building were performed with CORINA (31). The 10 lowest energy poses were retained from each sublibrary and were stored for further analysis. Flexible ligand docking was performed and the final set of poses was obtained after coarse minimization, reclustered and ranked. The 500 lowest energy poses were filtered with the FAF-Drugs using the default settings<sup>b</sup> (32). The properties monitored include molecular weight, number of hydrogen bond acceptors and donors, number of flexible and rigid bonds, number of carbon and hetero atoms, estimation of the partition coefficient ( $\log P$ ) and calculation of polar surface area. This was followed by a careful manual filtering to yield a selected subset, guided by the functional and structural knowledge that the narrow cavity in the CP active site cleft is able to accept short, linear polypeptide chains as substrates, whereas large structures bind only to the surface of the protein and do not pass into the active site cleft, as is the case of the globular part of the proteinaceous CP inhibitors. Therefore, the manual filtering included the visual inspection of the structures to remove polycyclic systems and compounds bearing large aromatic moieties. Structures containing a chemical linkage susceptible to be hydrolyzed by the target enzyme were also eliminated and a final filtering removed structurally closest structures. Docking of selected compounds was carried out using AUTODOCK 4 implementing AUTOGRID 4 using the default settings (33). Programs of the Collaborative Computational Project Number 4 suite (34) were used for protein–ligand interaction analysis. Visualization and molecular graphics were done with ARGUSLAB 4.0.1<sup>c</sup> graphical user interface or with PYMOL<sup>d</sup>. All the final figures were rendered with PYMOL.

## Results and Discussion

### Screening of small-molecule compounds by molecular docking

Among commercial screening collections, the APC queried in search for small-molecule ligands potentially able to inhibit CP activity contains a high percentage of drug-like molecules (35). The ArgusDock molecular modelling engine, implemented within the package ARGUSLAB and which approximates an exhaustive search method with similarities to DOCK and Glide (36) was applied using human CPB (PDB code 1ZLI) as the target structure. All ligands, except the zinc ion, were removed. A binding site box of size 20  $\times$  15  $\times$  20 Å centered at the zinc ion and encompassing all the known subsites for substrate binding, was defined for the screening. Careful filtering from the 500 lowest energy poses yielded a discrete number of hits, from which nine were available for purchase. The molecular weight of the selected compounds falls in the range 325–460 Da, and they contain between four and seven flexible bonds. The hydrogen bond donors and acceptors range from 0 to 3 and from 4 to 7, respectively. With only one exception (compound **8**), none of them displays ionizable groups, and their partition coefficient falls well below 5 (except compound **5**, whose  $\log P$  is 5.39). None of the selected compounds has a stereocenter, in contrast with known inhibitors like BX 528, which has two chiral atoms. From the observation that the selected compounds resemble some aminoisoquinoline derivatives which were incorporated into thrombin inhibitors

## Inhibition of MCPs by Aromatic Compounds

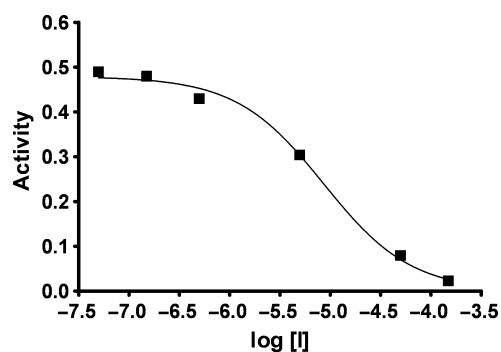


**Figure 1:** Chemical structures of the compounds tested, ordered according to increasing molecular weight.

with the results of improved membrane transport and oral bioavailability (37,38), two additional aminoisoquinoline amino acids (molecular weight = 230 Da) were added to this set to compare them with those selected in the survey. These two additional compounds contain a chiral C and were assayed as a racemic mixture. Overall, 11 compounds (Figure 1) were assayed as CP inhibitors.

#### Inhibition of metalloproteases

The inhibitory activity experiments were performed as published (39). The  $K_i$  values were derived from  $IC_{50}$  applying the equation of Cheng and Prusoff (40), using a fixed concentration of substrate and variable concentrations of inhibitor ranging from 50 nM to 1 mM. A non-linear global fit to a competitive one-site model gave the best fitting of the experimental observations. Figure 2 shows the fit of the experimental data for compound **5** as an example and the complete results for the  $K_i$  values of the aromatic derivatives against different CPs are collected in Table 1. Some of the aromatic derivatives behave as potent CP inhibitors, with  $K_i$  in the low micromolar range. Among them, the simple, low molecular weight



**Figure 2:** Inhibition of human CPB by compound **5**. The experimental measurements are shown as full squares and the fitting to a competitive inhibition model is drawn as a continuous line.

**Table 1:** Inhibitory constants ( $K_i$  in  $\mu\text{M}$ ) of aromatic derivatives

Compound	bCPA	hCPB	H <sub>z</sub> CPB
<b>1</b>	280 (50)	220 (35)	100 (13)
<b>2</b>	290 (75)	145 (20)	129 (11)
<b>3</b>	520 (240)	32 (15)	62 (30)
<b>4</b>	6.5 (2.2)	7.4 (2.5)	4.1 (1.5)
<b>5</b>	58 (15)	1.9 (0.4)	1.6 (0.5)
<b>6</b>	32 (13)	50 (22)	14 (5)
<b>7</b>	420 (180)	22.5 (5.8)	20.5 (5.0)
<b>8</b>	10.4 (4.0)	11.8 (3.0)	35 (12)
<b>9</b>	37 (15)	14.3 (5.0)	82 (30)
<b>10</b>	3.8 (0.6)	1.6 (0.2)	0.58 (0.12)
<b>11</b>	15 (5)	13.3 (4.5)	1.2 (0.4)

Values inside parentheses indicate the standard error of the mean. bCPA, bovine carboxypeptidase A; hCPB, human carboxypeptidase B; H<sub>z</sub>CPB, carboxypeptidase B from corn earworm (*Helicoverpa zea*).

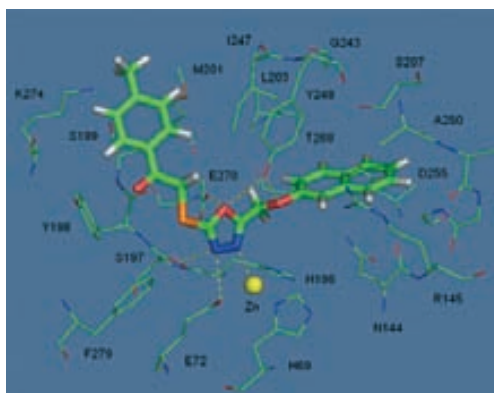
structures, aminoisoquinoline amino acids displayed the weakest inhibitory activity. In most cases, the best  $K_i$  values were those obtained against the B forms of CP. Quinolinone **10**, together with naphthalene **5** were the most active compounds against human CPB. In contrast to compound **10**, inhibitor **5** showed a 35-fold improvement in  $K_i$  value for CPB, suggesting some specific mode of action. Compounds **4** and **10** are equipotent inhibitors of both CPA and CPB, and the latter is a submicromolar inhibitor of insect CPB; both may be regarded as the most potent CPA inhibitors of the set, with  $K_i$  values below 7  $\mu\text{M}$ . Although the virtual screening was performed using CPB as the template, the inhibition of CPA is expectable given the close homology between CPA and CPB active sites. Considering the inhibition data against CPB, we observe that, starting from the aminoisoquinoline amino acid structure, additional fragments attached to the aromatic rings contribute significantly to binding. Interestingly enough, the two most potent CPB inhibitors represent two diverse chemical structures, distinct from known inhibitors.

For some compounds, the measured  $K_i$  values reflect the energy ranking in the virtual screening set. This is the case of top ranked **5**, which is only second in inhibitory potency against CPB (a value almost equal to that of compound **10**; see Table 1). Compound **4** is third in both rankings, while compound **10** is fourth in docking energy score and first in  $K_i$  value. Compound **9** deviates from this trend: it is second in docking energy, but only sixth in inhibitory potency. This results in that three of the four top-ranked compounds display the best  $K_i$  values (ranging from 1.6 to 11.8  $\mu\text{M}$ ). On the contrary, among compounds ranked 5–9 in docking energy, only one is among the top four in  $K_i$  values (inhibitor **8**, ranked 4 in  $K_i$  value but only 7 in docking energy). Aminoisoquinoline amino acids, not derived from the survey, showed poor docking energy values that would have prevented them to be within the 500 lowest energy poses, suggesting an efficient performance of the computational screening. The behavior of compound **9** might be rationalized as follows. The structures of the top four ranked compounds in docking energy, namely **4**, **5**, **9** and **10**, display almost similar properties (number of donor/acceptors, flexible bonds and polar surface area), but differ in the zinc-binding group. The zinc-binding groups in compounds **4** and **5** (oxadiazole), and **10** (thiourea) settle in a less-crowded environment as compared with **9**. In the latter, the *N*-benzyl benzenesulfonamide moiety might be less suited for binding to the zinc ion, possibly because of steric hindrance that might decrease its inhibitory potency. We might conceive that this particular kind of bonding, or the binding to the co-ordination sphere of the metal cation, cannot be managed adequately by the docking algorithm.

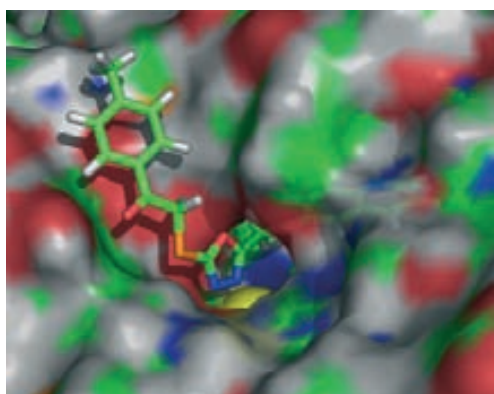
#### Analysis of the binding of aromatic derivatives to CPs

The binding of the selected set of compounds was analyzed through molecular docking, and the case of inhibitor **5**, which displays a preference to inhibit B-type CPs, is described below. The binding of inhibitor **5** into human CPB active site (Figure 3), spans subsites S1' (the environment around Asp255), S1 (defined by Arg145 and Glu270) and S2 (Arg71, Ser197, Tyr198 and Ser199). The representation in Figure 4 shows that the naphthyl group is buried, the

### Inhibition of MCPs by Aromatic Compounds



**Figure 3:** Putative binding of compound **5** (sticks model) in the active site of human CPB (only residues in a shell at a distance  $<4.5$  Å from the inhibitor are shown). Hydrogen bonds are depicted as dashed lines (see the text for details). Some residues relevant for catalysis or binding are labelled. The figure is coloured in atom type representation: green (carbon), red (oxygen), orange (sulfur), white (hydrogen) and blue (nitrogen). The zinc is depicted as a yellow sphere.



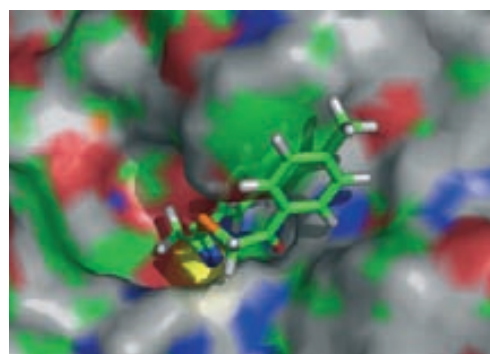
**Figure 4:** Putative binding of compound **5** (sticks model) in the active site of human CPB (shown as a translucent surface coloured as the atom types beneath). The active site cleft is pointing towards the reader. The inhibitor is twisted at the position of the oxadiazole ring, which is in contact to the zinc (yellow sphere). The figure is colored in atom type representation as in Figure 3.

oxadiazole ring locates near the zinc ion and the phenyl ring is at the protein surface. The bicyclic aromatic system locates in a hydrophobic environment defined by Ile203, Gly243, Ile247, Ala250, Gly253 and Thr268, at the bottom of the active site cleft. The naphthyl ring is positioned near the specificity-determinant residue Asp255 (the distance between the closest atoms is  $3.75$  Å). Polar interactions warrant a firm attachment of the oxadiazole ring to the

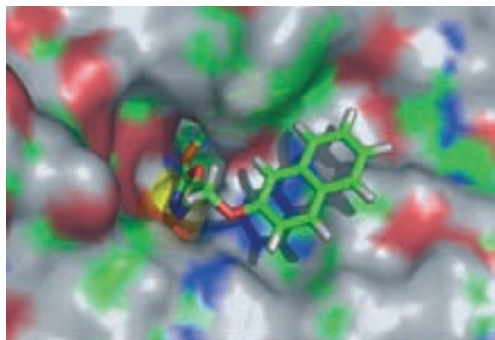
active site of the protein. The oxadiazole ring nitrogen atoms are in the  $Zn^{2+}$  ion environment (at  $2.5$  Å), and interact with one Glu72 carboxyl O atom (at  $2.8$  Å) and Ser197 carbonyl (at  $2.9$  Å), while the oxadiazolyl oxygen atom forms a hydrogen bond with the catalytically essential residues Tyr248 (the distance with the phenolic hydroxyl is  $3.2$  Å) and Glu270 (the distance with the carboxyl O atom is  $2.4$  Å). The O atom linker of the naphthyl and oxadiazolyl moieties is hydrogen bonded to Tyr248 OH (at  $2.8$  Å). There is an on-face interaction between the phenyl ring of the ligand and Ser199 OH group. Asn 144 and Arg145, a point of anchorage of ligands, are in the proximity of the naphthyl group. The aromatic ring of the phenyl moiety is surrounded by Tyr198, Ser199 and Ile247 in an open depression walled by Tyr198 and Tyr248 over the protein's surface.

Docking of compound **5** into insect HzCPB (Figure 5) shows an overall orientation of the ligand similar to that in human CPB, i.e. with the naphthyl moiety pointing in the direction of the bottom of the active site cleft. However, a slight offset in the orientation results in that the naphthyl moiety locates closer to the  $Zn^{2+}$  ion while the oxadiazole ring is far away from it. There is also a change in the position of the phenyl ring of the inhibitor that now locates in a surface patch close to, and making a face-to-face interaction with the Tyr248 phenolic ring. In the second case, the inhibitor must have adopted a more bent conformation than that observed for human CPB.

A superimposition of hCPB and HzCPB structures reveals that some residue substitutions at the bottom of the active site cleft might explain the different behaviour of inhibitor **5**. The changes that occur among residues surrounding the ligand (Gly243/Ser243, Ala250/Ile250, Gly253/Ser253 and Asp255/Glu255 in hCPB and HzCPB, respectively) turn the binding pocket narrower in the insect CPB than in the human counterpart. As a possible consequence, the naphthyl group docks differently to avoid clashing with several residues in HzCPB. Notwithstanding this, compound **5** showed an equipotent hCPB/HzCPB inhibitory profile. The binding of compound **5**



**Figure 5:** Putative binding of compound **5** (sticks model) in the active site cleft of insect CPB (HzCPB). The ligand is twisted to accommodate the phenyl ring in the vicinity of Tyr248. The view is approximately in the same orientation as in Figure 3. For details, see the legend of Figure 3.



**Figure 6:** Putative binding of compound **5** (sticks model) in the active site cleft of bovine CPA. The ligand is in opposite orientation by comparison with CPB binding. The view is approximately in the same orientation as in Figure 3. For details, see the legend of Figure 3.

to CPA was predicted to differ from that to CPB (Figure 6). In CPA, the naphthyl moiety of the inhibitor points towards the entrance of the active site cleft, i.e. in the opposite direction as compared with CPB, while the phenyl ring locates at the bottom of the active site cleft contacting Ile255, the specificity determinant residue. The oxadiazole ring interacts with the  $Zn^{2+}$  ion as in CPB. The fact that the naphthyl moiety cannot be buried by CPA, possibly to avoid colliding with Ile243 side chain (a glycine is found at the equivalent position in CPB), might cause a suboptimal binding and thus a noticeable drop in inhibitory potency.

## Conclusions

Aromatic organic derivatives, identified from a virtual screening approach have been found to display *in vitro* biological activity against metalloproteases of biomedical and biotechnological interest. Besides, two compounds added independently from the screening behaved poorly as CP inhibitors, indicating an efficient performance of the virtual screening over a huge database. The use of a massive screening collection, rather than of focused or pre-filtered databases, yields a large number of poses that have to be selected by manual, knowledge-based considerations. We consider that with this limited set of 11 compounds, nine obtained from virtual screening and two included independently, the performance of the virtual screening approach is efficient, as reflected by the overall correspondence between the measured  $K_i$  values and the docking energy. Improvements to the virtual screening might include a better handling of the binding of zinc chelating groups to the metal co-ordination sphere.

Some compounds display interesting features such as low molecular weight, powerful inhibition of the target enzymes and specificity. In addition, they obey Lipinski's rule-of-five and harbor functional groups that have been employed elsewhere to improve membrane transport and oral bioavailability. Among the tested compounds,

inhibitor **5** behaves as a potent and selective inhibitor towards a B-type CPs. Human CPB and CPB-like enzymes are among the proteases which are under active investigation in the search for novel therapeutic agents to treat thrombosis (15,16). In this context, compound **5** represents a new kind of scaffold for a CP inhibitor that can be considered an interesting starting point for improvements. For instance, the bulky naphthyl makes a suboptimal interaction with the specificity determinant aspartate at the bottom of the active site cleft. Besides, the phenyl ring at the opposite end of the inhibitor which docks into an open surface patch around subsite S2 leaving some space available. Thus, variants of this ligand might be generated by modification and extension at both ends to screen possible new interactions to the target CPB protein and to enhance the inhibitory potency. These appealing possibilities are currently under active exploration.

## Acknowledgments

The authors are indebted to Jos Rewinkel (Schering-Plough, Oss, the Netherlands) for providing the aminoisoquinoline aminoacids (compounds **1** and **2**) as well as for fruitful comments and suggestions. This study was supported by grants BIO2007-68046 from the Spanish Ministerio de Educación y Ciencia, 2005SGR-1037 from the Generalitat de Catalunya and CAMP project, VI EU Framework Programme, ref. 018830. DF wishes to thank the European Union *Alfa* Programme for a PhD fellowship (contract no. E04D035150AR).

## References

- Dunn F.W., Dittmer K. (1950) Action of carboxypeptidase toward peptides containing unnatural aromatic amino acids. *Science*;111:173–175.
- Park J.D., Kim D.H. (2002) Cysteine derivatives as inhibitors for carboxypeptidase A: synthesis and structure-activity relationships. *J Med Chem*;45:911–918.
- Martin M.T., Holmquist B., Riordan J.F. (1989) An angiotensin converting enzyme inhibitor is a tight-binding slow substrate of carboxypeptidase A. *J Inorg Biochem*;36:39–50.
- Lanthier C.M., Parniak M.A., Dmitrenko G.I. (1997) Inhibition of carboxypeptidase A by n-(4-t-butylbenzoyl)-2-hydroxy-1-naphthaldehyde hydrazone. *Bioorg Med Chem Lett*;7:1557–1562.
- Latt S.A., Auld D.S., Vallee B.L. (1972) Fluorescence determination of carboxypeptidase A activity based on electronic energy transfer. *Anal Biochem*;50:56–62.
- Yajima R., Chikuma T., Kato T. (1995) Highly sensitive high-performance liquid chromatography-fluorimetric assay method for carboxypeptidase H activity. *J Chromatogr B Biomed Appl*;667:333–338.
- Vogel W.H., Snyder R., Schulman M.P. (1964) Inhibition of several enzymes by folic acid, aminopterin and amethopterin. *Biochim Biophys Acta*;85:164–166.
- Vogel W.H., Snyder R., Schulman M.P. (1964) Effects of aromatic and nonaromatic model compounds and drugs on enzymic activities. *J Pharmacol Exp Ther*;146:66–73.

## Inhibition of MCPs by Aromatic Compounds

9. Vendrell J., Aviles F.X., Fricker L.D. (2004) Metallo-carboxypeptidases. In: Messerschmidt A., Bode W., Cygler M., editors. Metallo-carboxypeptidases. Chichester: John Wiley & Sons, Ltd; p. 176–189.
10. Rodriguez de la Vega M., Sevilla R.G., Hermoso A., Lorenzo J., Tanco S., Diez A., Fricker L.D., Bautista J.M., Aviles F.X. (2007) Nna1-like proteins are active metallo-carboxypeptidases of a new and diverse M14 subfamily. *FASEB J*;21:851–865.
11. Matsugi S., Hamada T., Shioi N., Tanaka T., Kumada T., Satomura S. (2007) Serum carboxypeptidase A activity as a biomarker for early-stage pancreatic carcinoma. *Clin Chim Acta*;378:147–153.
12. Shamamian P., Goldberg J.D., Ye X.Y., Stewart J.D., White P.J., Gilvarg C. (2006) Evaluation of pro-carboxypeptidase A and carboxypeptidase A as serologic markers for adenocarcinoma of the pancreas. *HPB (Oxford)*;8:451–457.
13. Rooth E., Wallen H., Antovic A., von Arbin M., Kaponides G., Wahlgren N., Blomback M., Antovic J. (2007) Thrombin activatable fibrinolysis inhibitor and its relationship to fibrinolysis and inflammation during the acute and convalescent phase of ischemic stroke. *Blood Coagul Fibrinolysis*;18:365–370.
14. Willemse J.L., Hendriks D.F. (2007) A role for procarboxypeptidase U (TAFI) in thrombolysis. *Front Biosci*;12:1973–1987.
15. Arolas J.L., Vendrell J., Aviles F.X., Fricker L.D. (2007) Metallo-carboxypeptidases: emerging drug targets in biomedicine. *Curr Pharm Des*;13:349–366.
16. Mittl P.R., Grutter M.G. (2006) Opportunities for structure-based design of protease-directed drugs. *Curr Opin Struct Biol*;16:769–775.
17. Barrow J.C., Nantermet P.G., Stauffer S.R., Ngo P.L., Steinbeiser M.A., Mao S.S., Carroll S.S. *et al.* (2003) Synthesis and evaluation of imidazole acetic acid inhibitors of activated thrombin-activatable fibrinolysis inhibitor as novel antithrombotics. *J Med Chem*;46:5294–5297.
18. Buchanan M.S., Carroll A.R., Edser A., Sykes M., Fechner G.A., Forster P.I., Guymer G.P., Quinn R.J. (2008) Lysianadioic acid, a carboxypeptidase B inhibitor from *Lysiana subfalcatata*. *Bioorg Med Chem Lett*;18:1495–1497.
19. Suzuki K., Muto Y., Fushihara K., Kanemoto K., Iida H., Sato E., Kikuchi C., Matsushima T., Kato E., Nomoto M., Yoshioka S., Ishii H. (2004) Enhancement of fibrinolysis by EF6265 [(S)-7-amino-2-[[[(R)-2-methyl-1-(3-phenylpropanoylamino)propyl]hydroxyphosphino]methyl]heptanoic acid], a specific inhibitor of plasma carboxypeptidase B. *J Pharmacol Exp Ther*;309:607–615.
20. Kallus C., Broenstrup M., Czechtizky W., Evers A., Follmann M., Halland N., Schreuder H. (2008) Urea and sulfamide derivatives as TAFIA inhibitors and their preparation, pharmaceutical compositions and use in the treatment of diseases, WO Patent No. 2008067909.
21. Wang Y.X., Zhao L., Nagashima M., Vincelette J., Sukovich D., Li W., Subramanyam B. *et al.* (2007) A novel inhibitor of activated thrombin-activatable fibrinolysis inhibitor (TAFIa) – part I: pharmacological characterization. *Thromb Haemost*;97:45–53.
22. Wang Y.X., da Cunha V., Vincelette J., Zhao L., Nagashima M., Kawai K., Yuan S., Emayan K., Islam I., Hosoya J., Sullivan M.E., Dole W.P., Morser J., Buckman B.O., Vergona R. (2007) A novel inhibitor of activated thrombin activatable fibrinolysis inhibitor (TAFIa) – part II: enhancement of both exogenous and endogenous fibrinolysis in animal models of thrombolysis. *Thromb Haemost*;97:54–61.
23. Bayes A., de la Vega M.R., Vendrell J., Aviles F.X., Jongsma M.A., Beekwilder J. (2006) Response of the digestive system of *Helicoverpa zea* to ingestion of potato carboxypeptidase inhibitor and characterization of an uninhibited carboxypeptidase B. *Insect Biochem Mol Biol*;36:654–664.
24. Bayes A., Comellas-Bigler M., Rodriguez de la Vega M., Maskos K., Bode W., Aviles F.X., Jongsma M.A., Beekwilder J., Vendrell J. (2005) Structural basis of the resistance of an insect carboxypeptidase to plant protease inhibitors. *Proc Natl Acad Sci USA*;102:16602–16607.
25. Pallares I., Fernandez D., Comellas-Bigler M., Fernandez-Rocio J., Ventura S., Aviles F.X., Bode W., Vendrell J. (2008) Direct interaction between a human digestive protease and the mucoadhesive poly(acrylic acid). *Acta Crystallogr D Biol Crystallogr*;64:784–791.
26. Barbosa Pereira P.J., Segura-Martin S., Oliva B., Ferrer-Orta C., Aviles F.X., Coll M., Gomis-Ruth F.X., Vendrell J. (2002) Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). *J Mol Biol*;321:537–547.
27. Turk B. (2006) Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov*;5:785–799.
28. Kaplan H.R., Cohen D.M., Essenburg A.D., Major T.C., Mertz T.E., Ryan M.J. (1984) CI-906 and CI-907: new orally active nonsulfhydryl angiotensin-converting enzyme inhibitors. *Fed Proc*;43:1326–1329.
29. Roehrborn C.G., Siegel R.L. (1996) Safety and efficacy of doxazosin in benign prostatic hyperplasia: a pooled analysis of three double-blind, placebo-controlled studies. *Urology*;48:406–415.
30. Ventura S., Villegas V., Sterner J., Larson J., Vendrell J., Hershberger C.L., Aviles F.X. (1999) Mapping the pro-region of carboxypeptidase B by protein engineering. Cloning, overexpression, and mutagenesis of the porcine proenzyme. *J Biol Chem*;274:19925–19933.
31. Gasteiger J., Rudolph C., Sadowski J. (1990) Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput Methodol*;3:537–547.
32. Miteva M.A., Violas S., Montes M., Gomez D., Tuffery P., Villou-treix B.O. (2006) FAF-drugs: free ADME/tox filtering of compound collections. *Nucleic Acids Res*;34:W738–W744.
33. Morris G.M., Goodsell D.S., Halliday R.S., Huey R., Hart W.E., Belew R.K., Olson A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem*;19:1639–1662.
34. Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*;50:760–776.
35. Krier M., Bret G., Rognan D. (2006) Assessing the scaffold diversity of screening libraries. *J Chem Inf Model*;46:512–524.
36. Joy S., Nair P.S., Hariharan R., Pillai M.R. (2006) Detailed comparison of the protein-ligand docking efficiencies of GOLD, a commercial package and ArgusLab, a licensable freeware. *In Silico Biol*;6:601–605.
37. Rewinkel J.B., Lucas H., Smit M.J., Noach A.B., van Dinther T.G., Rood A.M., Jenneboer A.J., van Boeckel C.A. (1999) Design, synthesis and testing of amino-bicycloalyl based orally

**Fernández et al.**

- bioavailable thrombin inhibitors. *Bioorg Med Chem Lett*;9:2837–2842.
38. Rewinkel J.B., Lucas H., van Galen P.J., Noach A.B., van Dinther T.G., Rood A.M., Jenneboer A.J., van Boeckel C.A. (1999) 1-Aminoisoquinoline as benzamidine isoster in the design and synthesis of orally active thrombin inhibitors. *Bioorg Med Chem Lett*;9:685–690.
39. Fernandez D., Illa O., Aviles F.X., Branchadell V., Vendrell J., Ortuno R.M. (2008) Thioxophosphoranyl aryl- and heteroaryloxiranes as the representants of a new class of metalcarboxypeptidase inhibitors. *Bioorg Med Chem*;16:4823–4828.
40. Cheng Y., Prusoff W.H. (1973) Relationship between the inhibition constant (K<sub>1</sub>) and the concentration of inhibitor which causes 50 per cent inhibition (I<sub>50</sub>) of an enzymatic reaction. *Biochem Pharmacol*;22:3099–3108.

**Notes**

- <sup>a</sup><http://www.asinex.com>  
<sup>b</sup><http://bioserv.rpbs.jussieu.fr/FAFDrugs.html>  
<sup>c</sup><http://www.arguslab.com>  
<sup>d</sup><http://www.pymol.org>





Contents lists available at ScienceDirect

## Bioorganic &amp; Medicinal Chemistry

journal homepage: [www.elsevier.com/locate/bmc](http://www.elsevier.com/locate/bmc)

## Cyclobutane-containing peptides: Evaluation as novel metalloprotease inhibitors and modelling of their mode of action

Daniel Fernández<sup>a</sup>, Elisabeth Torres<sup>b</sup>, Francesc X. Avilés<sup>a</sup>, Rosa M. Ortuño<sup>b,\*</sup>, Josep Vendrell<sup>a,\*</sup><sup>a</sup> *Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, and Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain*<sup>b</sup> *Departament de Química, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain*

## ARTICLE INFO

## Article history:

Received 13 February 2009

Revised 14 April 2009

Accepted 17 April 2009

Available online 23 April 2009

## Keywords:

Metalloprotease  
Inhibitor screening  
Ligand docking  
Human carboxypeptidase B  
Cyclobutane  $\beta$ -peptide  
Non-natural amino acid

## ABSTRACT

Different types of cyclobutane-containing peptides (CBPs) were screened for the first time as ligands of metalloproteases (MCPs). CBPs are conformationally constrained, low molecular-weight compounds which showed moderate yet selective inhibitory activity against mammalian MCPs. The most potent compound was a carboxypeptidase B inhibitor. Docked protein–ligand complexes indicated that CBPs may bind to the target proteases via electrostatic interactions and aromatic stacking to catalytically crucial residues and that the placement of functional groups seems to be assisted by the rigid CBP backbone. The easily obtainable CBPs may offer a valuable alternative in the design of novel inhibitors to disease-linked metalloproteases like human plasma carboxypeptidase B.

© 2009 Elsevier Ltd. All rights reserved.

### 1. Introduction

Since their discovery, carboxypeptidase A (CPA) and carboxypeptidase B (CPB), members of the M14 family of metalloproteases, have been intensively investigated as a mechanistic model of metal-dependent proteases.<sup>1–4</sup> The assessment of the potential roles that MCPs play in different pathologies, like acute pancreatitis<sup>5–7</sup> inflammation,<sup>8,9</sup> diabetes,<sup>10</sup> and cancer<sup>11–14</sup> is a highly explored field today. Investigations on gastrointestinal proteolytic enzymes—such as pancreatic CPA and CPB—are currently underway to discern potential ways to overcome the problem of poor oral absorption of peptide and protein drugs.<sup>15–17</sup> In another context, CPA and CPB are considered to be good biomarkers for the early detection of acute pancreatitis and cancer.<sup>18,19</sup> These findings, together with the recent discovery of a complete new subfamily,<sup>20</sup> may warrant further developments of chemotherapeutic or bioimaging agents targeted to M14 proteases.<sup>21</sup>

Peptidomimetics containing non hydrolyzable bonds have been widely exploited in the design of extremely potent protease inhibitors.<sup>22,23</sup> Compound BX528 (a phosphinate-containing mimic of the tripeptide Phe-Val-Lys), an inhibitor to human plasma CPB, appears as a recent example of application of this strategy.<sup>24</sup> An alter-

native approach may involve the use of  $\beta$ - and  $\gamma$ -peptides. Such compounds proved to be resistant to hydrolytic cleavage by benchmark proteases, including CPA.<sup>25</sup>  $\beta$ - and  $\gamma$ -peptides share some common properties such as natural origin, stability, and propensity to form folding structures.<sup>26</sup>

Among the broad structural diversity known for natural and designed oligomers, CBPs are compounds scarcely explored to account for their biological properties. The CBP backbone contains the conformationally rigid cyclobutyl group instead of the freely rotatable ( $-\text{CH}_2-$ ) units found in  $\beta$ - and  $\gamma$ -peptides. Moreover, CBPs constitute an attractive starting point for screening because of they bear some resemblance to the  $\alpha$ -peptide substrates of MCPs: a low molecular weight, and an easy synthetic preparation.<sup>27,28</sup> We, therefore, decided to test the ability of CBPs to block the enzymic reaction carried out by M14 proteases and to investigate whether their rigid backbone might impose a demanding constraint on binding as shown by our recent determination of an  $\alpha$ -peptide-bound MCP three-dimensional structure.<sup>29</sup> We obtained a series of CBPs and screened them against two prototypical MCPs with different specificities. Some CBPs were found to perform an inhibitory activity in vitro against bovine CPA and human CPB. Selected CBPs were computationally docked to the enzymes and their binding properties analysed in depth. As a result of this work, the identification and characterisation of a novel class of MCP inhibitors, the cyclobutane-containing peptides is described herein.

\* Corresponding authors. Tel.: +34 93 581 1602; fax: +34 93 581 1265 (R.M.O.), tel.: +34 93 581 2375; fax: +34 93 581 1264 (J.V.).

E-mail addresses: [rosa.ortuno@uab.cat](mailto:rosa.ortuno@uab.cat) (R.M. Ortuño), [josep.vendrell@uab.cat](mailto:josep.vendrell@uab.cat) (J. Vendrell).

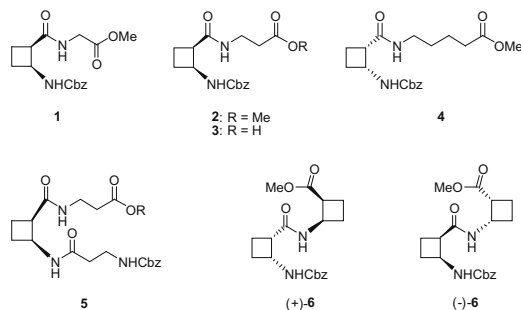


Chart 1. Structures of the CBPs analysed in this study.

## 2. Results and discussion

The structures of the cyclobutane containing peptides studied in this work are shown in Chart 1. Compounds **1–4** differ on the length of the linear residue. Thus, compound **1** is an  $\alpha,\beta$ -dipeptide, **2** and **3** are  $\beta$ -dipeptides differing in the protected or unprotected terminal carboxyl group, **4** is a mixed  $\beta,\delta$ -dipeptide, **5** is a  $\beta$ -tripeptide containing two  $\beta$ -alanine segments joined by a cyclobutane residue, and (+)- and (–)-**6** are enantiomeric bis(cyclobutane)  $\beta$ -dipeptides.

These compounds range in MW from 320 to 405 Da and their structures display a central cyclobutane ring flanked by a Cbz moiety and a carboxylate group at both ends. According to previous work, the backbone of the CBPs is bent around the cyclobutane ring and their functionalities may engage in a number of intermolecular interactions with neighbour molecules.<sup>27,28</sup>

### 2.1. Inhibition of carboxypeptidases

The results from seven CBPs acting against the two prototypical metallo-carboxypeptidases, CPA and CPB, are collected in Table 1.

CBPs were active against CPA and CPB in the micromolar range. Dipeptide **1**, containing a glycine residue and synthesized to bear resemblance to the natural MCP substrates, showed the weakest inhibitory effect of the set. On the contrary, the  $\beta$ -dipeptide **2**, which is a  $-\text{CH}_2-$  unit longer than **1**, behaved as a more effective inhibitor, especially against CPA. The short alkyl branch and the blocked carboxylate in compounds **1** and **2**, may weaken binding strength. Analogue **3**, with a free carboxylate, showed the lowest  $K_i$  value for both enzyme forms, with a preference for CPB. Compound **4**, with an alkyl branch which only differs from **2** in being two  $-\text{CH}_2-$  units longer, performed as a twofold more potent inhibitor towards CPA, while its effect on CPB remained similar to that of compound **2**. These observations highlight the preference of CPA to bind and/or process aliphatic compounds. However, the fact that

compound **4** displays a methyl ester instead of a free carboxylate may also explain its poor  $K_i$  against CPB and further investigations are necessary to clarify this issue. Compound **5**, the largest structure of the set with two alkyl branches, is similar to the latter in binding CPA, suggesting that the extra branch might have a minimal contribution to binding. Enantiomeric compounds (+)- and (–)-**6**, which contain a second cyclobutyl unit, showed also moderate inhibitory potency against both target proteins. Nevertheless, while (+)-**6** did not show any selectivity towards CPA and CPB, the enantiomer (–)-**6** was slightly CPA selective. A close inspection of CPA and CPB **6**-docked structures suggested a common mode of binding, with the Cbz moiety buried into the active site cleft while the blocked carboxylate is located at the protein's surface (not shown), in clear contrast with the mode of binding of compounds displaying a single cyclobutyl unit (see below). The settlement of the Cbz moiety is similar in the chiral active site cleft which imposes a symmetrical binding fashion of the solvent exposed groups in both enantiomers. Although the fitting in the active site is analogous in both structures, some minor offset was observed to occur, which may possibly be related to the inhibitory activity.

In summary, cyclobutane containing peptides were identified as a new kind of ligands for metallo-carboxypeptidases from the M14 family. The most potent compound was **3**, which showed some preference to interact to CPB as compared to the other members of the series. Other slightly less potent CBPs acted with a high selectivity ratio against CPA, as is the case of compound **4**. Thus, both compounds **3** and **4** were chosen for further analysis of their binding mode.

### 2.2. Analysis of $\beta$ -dipeptide **3** binding

$\beta$ -Dipeptide **3** occupies three subsites within CPB's active site cleft: S1' (around Asp255), S1 (defined by Arg145 and Glu270, a subsite for substrate anchoring and processing) and S2 (shaped by Arg71, Ser197, Tyr198 and Ser199) (Fig. 1).

The inhibitor backbone is bent at the cyclobutyl ring. This might result from the rigid conformation dictated by the four-membered ring itself and from the protein–ligand intermolecular interactions. The free carboxylate binds to the substrate-anchoring residues Asn144 and Arg145 via one of its oxygen atoms (at distances 3.0 and 2.9 Å, respectively). This functionality locates more than 5 Å apart from Asp255, the specificity-determinant residue in CPB.

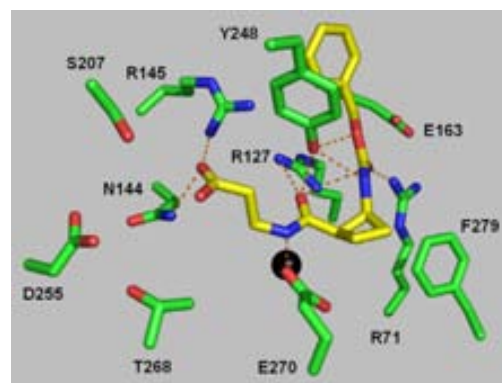
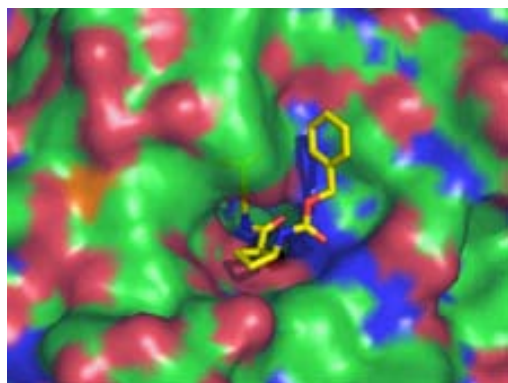


Figure 1. Predicted binding of compound **3** to human CPB structure (1zli). The ligand (yellow carbons) and selected CPB side-chains (green carbons) are shown in sticks model. Other atoms are coloured blue (nitrogen) and red (oxygen). Intermolecular interactions are as dashed orange lines. The catalytic zinc ion is shown as a black sphere.

Table 1  
Biological activity of CBPs

Compound	$K_i^a$ ( $\mu\text{M}$ )	
	CPA	CPB
<b>1</b>	550 (100)	500 (95)
<b>2</b>	180 (30)	400 (70)
<b>3</b>	70 (12)	43.0 (7.5)
<b>4</b>	72.5 (9.5)	410 (80)
<b>5</b>	70 (15)	165 (55)
(+)- <b>6</b>	67 (10)	70 (10)
(–)- <b>6</b>	95 (18)	275 (45)

<sup>a</sup> Values inside parentheses indicate the standard error of the mean (SEM).



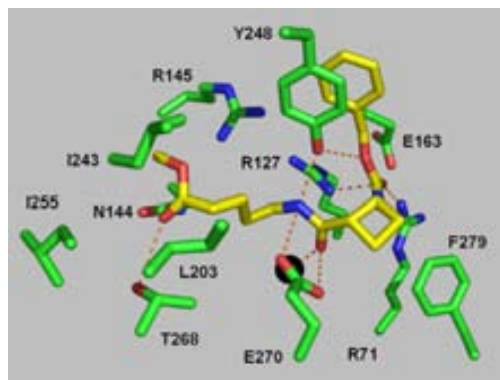
**Figure 2.** Placement of **3** over the CPB surface. The view shows the active site cleft pointing towards the reader, in approximately the same orientation as in Figure 1 (Tyr248 at about 12 o'clock). The ligand (yellow carbons) is shown in sticks model. Notice the bend of the ligand at the cyclobutyl unit. The figure is rendered as a translucent surface coloured as the atoms types beneath (see Fig. 1 legend for atom codes). The catalytic zinc ion is shown as a black sphere.

The amide bond of the carboxylate branch of the inhibitor makes a number of interactions to several residues at the active site: the N-atom is located in the second  $Zn^{2+}$  coordination sphere (at 3.1 Å) and hydrogen bonded to Glu270, the presumed proton donor during catalysis; the carbonyl oxygen interacts with the Arg127 side-chain. The cyclobutyl ring is close to two aromatic residues, Tyr198 and Phe279, and to the Ser199 side-chain. A further short interaction takes place between the carbonyl oxygen of the amide bond at the Cbz branch and Arg71 (separated by 2.7 Å). Most of the intermolecular interactions in the Cbz branch involve Tyr248 phenolic ring. This catalytically essential residue appears to be wrapped by **3**, with the Cbz aromatic ring facing the protein surface where it is accommodated in a hydrophobic depression (see Fig. 2). The Cbz ring sustains an edge-to-face interaction with Tyr248 phenolic ring on one face, while the side-chain carbon atoms from Thr164 and Glu163 are located near it at the solvent side. The binding of **3** to CPA proceeds much like as in CPB (not shown). The ligand is easily superimposable in both structures, and most interactions occur far from the specificity pocket of the enzyme. This can explain the almost similar  $K_i$  value for inhibition of CPA and CPB.

### 2.3. Analysis of $\beta,\delta$ -dipeptide **4** binding

Dipeptides **3** and **4** differ in the extension of the alkyl branch ending with either a free carboxylate in **3** or a methyl ester in **4**. Dockings predicted an opposite mode of binding for **4** into CPA and CPB. In CPA, the phenolic Tyr248 aromatic ring plays an important role in binding due to the number of interactions this residue makes to the ligand. These involve hydrogen bonds to the functionalities adjacent to the cyclobutane moiety as well as an face-to-face stacking of aromatic rings (Fig. 3).

One of the oxygen atoms of the methyl ester is hydrogen bonded to Thr268 hydroxyl (3.4 Å) and to one side chain amide nitrogen from Asn144 (3.2 Å). The long alkyl branch is within a hydrophobic environment lined by Leu203, Ile243, Ile247, and Ala250. The blocked carboxylate moiety is at the S1' subsite, in the proximity of Ile255, the specificity determinant residue of CPA. The inner environment has a different nature in CPB (Table 2) and might provide an unsatisfactory fit of the alkyl branch. This probably caused the reversed orientation predicted for the CPB–**4** complex (not shown), where the Cbz branch binds into the S1' pocket.



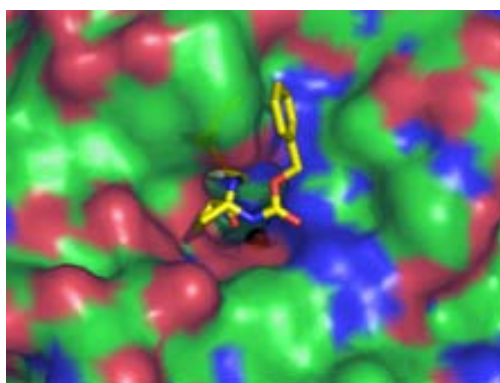
**Figure 3.** Predicted **4** binding to the CPA structure (2ctc). The view is in approximately the same orientation as in Figure 1. Other details as in the legend to Figure 1.

**Table 2**  
MCP ligand-binding residues

Position	203	207	243	251	255
CPA	Leu	Gly	Ile	Ser	Ile
CPB	Ile	Ser	Gly	Ala	Asp

Residues at distance <5 Å from the ligand are tabulated.

Beyond the cyclobutyl ring and heading to the active site, the amide bond engages the catalytically crucial Glu270 (via interactions with the carbonyl O and N atoms) and the zinc ion with the carbonyl O-atom (at a distance of 2.6 Å), suggesting that the ligand may be coordinating the metal ion. The cyclobutyl ring faces two aromatic residues, Tyr198 and Phe279, while the carbonyl O atom from the Cbz moiety interacts with Arg71 and Arg127 (at distances 2.7 and 2.6 Å, respectively). Finally, the Cbz aromatic ring is in a face-to-face arrangement with the Tyr248 phenolic ring, and is surrounded by other carbon atoms from Glu163 and Thr164 side-chains. As shown in Figure 4, the  $\beta,\delta$ -dipeptide **4** backbone is bent halfway between the Cbz and the blocked carboxylate group, i.e., at the position of the cyclobutyl ring. This suggests that conformationally constrained ligands are able to bind to into a narrow active site cavity such as that found in M14 proteases.



**Figure 4.** A view of the interactions between **4** and CPA. Same orientation as in Figure 2. Other details as in the legend to Figure 1.

### 3. Conclusion

We have recently identified a novel type of small-molecule ligands of MCPs from a collection of oxiranes.<sup>30</sup> As a continuation of the search for novel molecules, we have show herein that cyclobutane-containing peptides (CBPs) exert an inhibitory action against two prototypical M14 family proteases, CPA and CPB, in the micromolar range. Compound **3** was the most potent compound and behaved better as a CPB inhibitor, while compound **4**, provided with a longer alkyl chain with a blocked carboxylate, showed a preference to inhibit CPA. Based on the analysis of enzyme–inhibitor docked structures, further design of CBP-based inhibitors must consider, among other issues, the length of the alkyl branch and the chemical functionality displayed at its end, since both appear to interplay to define selectivity and binding strength towards one of the two types of MCPs. Docked structures also showed, as an unprecedented feature, that CBPs wrap the catalytically essential Tyr248 side chain, thus blocking the enzymic reaction. This would be a result from the constrained CBP conformation and the intermolecular interactions to the partner molecule. The most attractive inhibitors are ~350 Da compounds which may provide a convenient scaffold for chemical modification. CBPs might thus represent an alternative design to currently available MCP inhibitors, and might be considered interesting leads for further development in terms of potency and selectivity.

### 4. Experimental

#### 4.1. Synthesis of cyclobutane containing peptides

Compounds **2–6** were prepared according to the methods previously described.<sup>27,28</sup> Compound **1** is a new dipeptide that was synthesized as follows:

To a solution of (1*R*,2*S*)-2-*N*-Bezyloxycarbonylaminocyclobutane-1-carboxylic acid (85 mg, 0.34 mmol) in anhydrous DMF (25 mL) dry Et<sub>3</sub>N (0.2 mL), Gly-OMe (50.5 mg, 0.4 mmol), EDAC (200 mg, 1.04 mmol) and HOBt (70 mg, 0.52 mmol) were successively added. The mixture was stirred at room temperature under nitrogen atmosphere and the reaction was monitored by TLC. After 20 h, EtOAc (20 mL) was added and the combined organic layers were washed with saturated aqueous NaHCO<sub>3</sub> (3 × 15 mL). The organic phase was dried over MgSO<sub>4</sub> and solvents were removed under reduced pressure. The residue was chromatographed through Baker Silica gel using EtOAc as eluent to afford dipeptide **1** (40 mg, 44% yield) as a white solid crystals, mp 109–111 °C (AcOEt/pentane).  $[\alpha]_D^{25} -55.7$  (*c* 1.43, CH<sub>2</sub>Cl<sub>2</sub>). IR (solid): 3318 (NH), 2952 (CH), 1732(C=O), 1690 (C=O), 1542. <sup>1</sup>H NMR (250 MHz, CDCl<sub>3</sub>)  $\delta$  1.93 (m, 1H), 2.10 (m, 1H), 2.32 (complex absorption, 2H), 3.32 (m, 1H, H<sub>14</sub>), 3.74 (s, 3H, OMe), 3.89–4.02 (complex absorption, 2H, H<sub>17a,b</sub>), 4.53 (t, *J* = 6.82 Hz, 1H, H<sub>11</sub>), 5.10 (s, 2H, H<sub>7a,b</sub>), 5.88 (d, *J* = 5.67 Hz, 1H, NH<sub>10</sub>), 6.18 (broad s, 1H, NH<sub>16</sub>), 7.37 (m, 5H, H<sub>arom</sub>). <sup>13</sup>C NMR (62.5 MHz, CDCl<sub>3</sub>):  $\delta$  = 18.2, 29.6, 41.0, 46.0, 46.5, 52.2, 66.4, 128.0, 128.4, 136.7, 155.7, 170.5, 173.3. Anal Calcd for C<sub>16</sub>H<sub>20</sub>N<sub>2</sub>O<sub>5</sub>: C, 59.99; H, 6.29; N, 8.74. Found: C, 59.93; H, 6.56; N, 8.77.

#### 4.2. Biology

Recombinantly expressed human carboxypeptidase B was obtained as a zymogen, converted to the active form and purified as reported.<sup>31</sup> Bovine pancreatic carboxypeptidase A (CPA) was from Sigma. Enzyme concentrations for the kinetic studies were kept fixed at typically 5–50 nM. The chromogenic substrates, *N*-(4-methoxyphenylazofornyl)-*L*-phenylalanine and *N*-(4-methoxyphenylazofornyl)-*L*-arginine were from Bachem (Bubendorf,

Switzerland). The experimental assays were performed at room temperature in 50 mM Tris (Tris = 2-amino-2-(hydroxymethyl)propane-1,3-diol), 0.5 M NaCl, pH 7.5, and 20 mM Tris, 0.1 M NaCl, pH 7.5, for CPA and CPB, respectively. Typically, the enzymic activity was measured using a fixed concentration of substrate and variable concentrations of inhibitor (five concentrations in the range from 0.15  $\mu$ M to 750  $\mu$ M). The initial velocity measurements were performed per triplicate in a 96-well microtiter plate or in 1 ml cuvettes. The changes in absorbance due to substrate breakdown were followed continuously at 340 nm. The GRAPHPAD Version 5.0 ([www.graphpad.com](http://www.graphpad.com)) program was used to process the kinetic data. A non-linear global fit to a competitive one-site inhibition model gave the best fitting of the experimental observations. The enzyme inhibition *K*<sub>i</sub> values were obtained from IC<sub>50</sub> following the Cheng and Prusoff's equation.<sup>32</sup>

#### 4.3. Bioinformatics: docking procedures

The binding properties of the two selected CBPs, **3** and **4**, were analysed by docking procedures with the program AUTODOCK Version 4.0 (The Scripps Research Institute, La Jolla, CA, USA). The binding of both **6** enantiomers was also investigated. The compounds were constructed, converted to 3D structures, and energy minimised with CHEM3D Version 9.0 (<http://www.cambridgesoft.com>). Then, the ligands and the target macromolecules were prepared with AUTODOCK. To validate the approach, two separate experiments were performed with the known 3D structures of bovine CPA bound to the ester substrate *L*-phenyllactate (2ctc) and porcine CPB bound to a small-molecule inhibitor (2jew). The results from this dockings were in excellent agreement with the crystallographic determinations. The PDB coordinates of CPA and human CPB (2ctc and 1zli, respectively) were used for our ligands. The protein–ligand interactions of the predicted complexes were inspected with PyMol ([www.pymol.org](http://www.pymol.org)) graphical user interface. All the final figures for publication were prepared with PyMol.

#### Acknowledgements

Financial support from the Ministerio de Ciencia e Innovación, Spain (Grants CTQ2007-61704/BQU and BIO2007-68046), Generalitat de Catalunya (Grants 2005SGR-103 and 2005SGR-1037) and CAMP project 108830 (VI EU Framework Programme) is gratefully acknowledged.

#### References and notes

- Cho, J. H.; Kim, D. H.; Lee, K. J.; Choi, K. Y. *Biochemistry* **2001**, *40*, 10197.
- Cross, J. B.; Vreven, T.; Meroueh, S. O.; Mobashery, S.; Schlegel, H. B. *J. Phys. Chem. B* **2005**, *109*, 4761.
- Liu, Y. H.; Konermann, L. *FEBS Lett.* **2006**, *580*, 5137.
- Phoon, L.; Burton, N. A. *J. Mol. Graphics Modell.* **2005**, *24*, 94.
- Borgstrom, A.; Regner, S. *Pancreatology* **2005**, *5*, 530.
- Muller, C. A.; Appellos, S.; Uhl, W.; Buchler, M. W.; Borgstrom, A. *Gut* **2002**, *51*, 229.
- Regner, S.; Manjer, J.; Appellos, S.; Hjalmarsson, C.; Sadic, J.; Borgstrom, A. *Pancreatology* **2008**, *8*, 600.
- Asai, S.; Sato, T.; Tada, T.; Miyamoto, T.; Kimbara, N.; Motoyama, N.; Okada, H.; Okada, N. *J. Immunol.* **2004**, *173*, 4669.
- Rooth, E.; Wallen, H.; Antovic, A.; von Arbin, M.; Kaponides, G.; Wahlgren, N.; Blomback, M.; Antovic, J. *Blood Coagul. Fibrinol.* **2007**, *18*, 365.
- Jeffrey, K. D.; Alejandro, E. U.; Luciani, D. S.; Kalynyak, T. B.; Hu, X.; Li, H.; Lin, Y.; Townsend, R. R.; Polonsky, K. S.; Johnson, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8452.
- O'Malley, P. G.; Sangster, S. M.; Abdelmagid, S. A.; Bearne, S. L.; Too, C. K. *Biochem. J.* **2005**, *390*, 665.
- Vovchuk, I. L.; Petrov, S. A. *Biomed. Khim.* **2008**, *54*, 167.
- Fialka, F.; Gruber, R. M.; Hitt, R.; Opitz, L.; Brunner, E.; Schliephake, H.; Kramer, F. *J. Oral. Oncol.* **2008**, *44*, 941.
- Vovchuk, I. L.; Chernadchuk, S. S.; Petrov, S. A. *Biomed. Khim.* **2007**, *53*, 205.
- Bernkop-Schnurch, A.; Schmitz, T. *Curr. Drug Metab.* **2007**, *8*, 509.
- Fredholt, K.; Adrian, C.; Just, L.; Hoj Larsen, D.; Weng, S.; Moss, B.; Juel Friis, G. *J. Controlled Release* **2000**, *63*, 261.

17. Pallares, I.; Fernandez, D.; Comellas-Bigler, M.; Fernandez-Recio, J.; Ventura, S.; Aviles, F. X.; Bode, W.; Vendrell, J. *Acta Crystallogr. D Biol. Crystallogr. Sect. D* **2008**, *64*, 784.
18. Matsugi, S.; Hamada, T.; Shioi, N.; Tanaka, T.; Kumada, T.; Satomura, S. *Clin. Chim. Acta* **2007**, *378*, 147.
19. Stewart, J. D.; Gilvarg, C. *Clin. Chim. Acta* **1999**, *281*, 19.
20. Rodriguez de la Vega, M.; Sevilla, R. G.; Hermoso, A.; Lorenzo, J.; Tanco, S.; Diez, A.; Fricker, L. D.; Bautista, J. M.; Aviles, F. X. *FASEB J.* **2007**, *21*, 851.
21. Golz, S.; Brueggemeier, U.; Geerts, A. WO Patent No. 2006010499, 2006, p 95.
22. Bartlett, P. A.; Marlowe, C. K.; Giannousis, P. P.; Hanson, J. E. *Cold Spring Harb. Symp. Quant. Biol.* **1987**, *52*, 83.
23. Fan, H.; Zhao, Y.; Byers, L.; Hammer, R. P.. In *Peptides for the New Millennium*; Fields, G. B., Tam, J. P., Barany, G., Eds.; Springer: Netherlands, 2000; Vol. 6, p 91.
24. Wang, Y. X.; Zhao, L.; Nagashima, M.; Vincelette, J.; Sukovich, D.; Li, W.; Subramanyam, B.; Yuan, S.; Emayan, K.; Islam, I.; Hrvatin, P.; Bryant, J.; Light, D. R.; Vergona, R.; Morser, J.; Buckman, B. O. *Thromb. Haemost.* **2007**, *97*, 45.
25. Frackenhohl, J.; Arvidsson, P. I.; Schreiber, J. V.; Seebach, D. *ChemBiochem.* **2001**, *2*, 445.
26. Seebach, D.; Beck, A. K.; Bierbaum, D. J. *Chem. Biodivers.* **2004**, *1*, 1111.
27. Izquierdo, S.; Kogan, M. J.; Parella, T.; Moglioni, A. G.; Branchadell, V.; Giralt, E.; Ortuno, R. M. *J. Org. Chem.* **2004**, *69*, 5093.
28. Izquierdo, S.; Rua, F.; Sbai, A.; Parella, T.; Alvarez-Larena, A.; Branchadell, V.; Ortuno, R. M. *J. Org. Chem.* **2005**, *70*, 7963.
29. Bayes, A.; Fernandez, D.; Sola, M.; Marrero, A.; Garcia-Pique, S.; Aviles, F. X.; Vendrell, J.; Gomis-Ruth, A. F. *Biochemistry* **2007**, *46*, 6921.
30. Fernandez, D.; Illa, O.; Aviles, F. X.; Branchadell, V.; Vendrell, J.; Ortuno, R. M. *Bioorg. Med. Chem.* **2008**, *16*, 4823.
31. Ventura, S.; Villegas, V.; Sterner, J.; Larson, J.; Vendrell, J.; Hershberger, C. L.; Aviles, F. X. *J. Biol. Chem.* **1999**, *274*, 19925.
32. Cheng, Y.; Prusoff, W. H. *Biochem. Pharmacol.* **1973**, *22*, 3099.





Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: <http://www.elsevier.com/locate/ejmech>

Original article

## A new type of five-membered heterocyclic inhibitors of basic metalloproteinases

Daniel Fernández, Francesc X. Avilés, Josep Vendrell\*

Departament de Bioquímica i Biologia Molecular, Facultat de Biociències and Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain

## ARTICLE INFO

## Article history:

Received 29 December 2008

Received in revised form

20 February 2009

Accepted 26 March 2009

Available online 5 April 2009

## Keywords:

Metalloproteinase

Protease M14 family

Human metalloproteinase B

High-throughput screening

Molecular docking

1,3,4-Oxadiazole

## ABSTRACT

A structure-based virtual screening survey was used to identify potential inhibitors of the human M14 family of metalloproteinases. A good correlation between docking energy scores and measured  $K_i$  values was observed, indicating an efficient performance of the screening procedure. Among various compounds displaying  $K_i$  values in the low micromolar range, *N*-(3-chlorophenyl)-4-((5-(3-methoxybenzylthio)-1,3,4-oxadiazol-2-yl)methyl)thiazol-2-amine emerged as the most powerful inhibitor for human metalloproteinase B (CPB). According to molecular docking, this compound fits into CPB active site cleft through coordination of the catalytic zinc ion with the 1,3,4-oxadiazole moiety. This represents a novel five-membered heterocyclic type of inhibitor for disease-linked metalloproteinases and an interesting lead for further development.

© 2009 Elsevier Masson SAS. All rights reserved.

## 1. Introduction

The M14 family of proteases (metalloproteinases, CPs) constitutes a widely distributed group of enzymes that perform different physiological functions [1]. This large family of metalloproteinases contains two subfamilies represented by the prototypical CPA and CPB forms and has been expanded by the recent discovery of a completely new subfamily [2]. Different CPs may participate in normal tissue organogenesis [3,4] as well as in pathological processes like pancreatic diseases [5], inflammation [6], fibrinolysis [7] and cancer [8,9]. Pharmaceutically relevant M14 members include human CPA2, mast cell metalloproteinase (CPA3), and plasma CPB (also known as CPB2, CPR, CPU or TAFI) [10–12]. The development of novel diagnostic or chemotherapeutic tools may warrant a better understanding of M14 protease activities [13,14]. Several groups reported on the discovery of small-molecule inhibitors targeted to CPs [15,16]. One of these molecules, a peptidomimetic TAFI inhibitor, (*S*)-2-[3-(aminomethyl)phenyl]-3-[[hydroxyl(*R*)-2-methyl-1-[[3-(phenylpropyl)sulfonyl] amino propyl] phosphoryl propanoic acid showed promising properties when tested in animal models of thrombotic disease [17,18]. However, despite much effort that has been cast in the development of chemotherapeutic agents targeted to M14 proteases, a clinically useful molecule awaits to be discovered.

The work discussed here was aimed to the discovery of novel small-molecule organic ligands that might be useful leads for M14 protease drug design [19]. We employed a structure-based virtual screening methodology to query a diverse, non-focused, non-pre-filtered, small-molecule database using the 3D crystal structure of human CPB as the template. This high-throughput screening survey afforded a number of hits out of which 25 were purchased and their  $K_i$  values measured *in vitro*. Although the correlation between the measured  $K_i$  values and the docking energy score is good for most compounds, a few exceptions suggest some drawbacks in the selection process. A five-membered heterocyclic compound with low micromolar CPB inhibitory potency emerged from the kinetic activity measurements. Molecular docking of such inhibitor indicated that binding to the catalytic zinc ion is mediated by its 1,3,4-oxadiazole moiety, a feature that distinguishes it from other known inhibitors. Thus, we report here on the discovery of a new type of metalloproteinase inhibitor scaffold and discuss on its potential for further developments.

## 2. Results and discussion

## 2.1. Bioinformatics: screening of small-molecule compounds

The molecules indicated in Fig. 1 are those which were available for purchase from the final set of compounds retrieved after careful automatic and manual filtering. The manual filtering took into

\* Corresponding author. Tel.: +34 93 581 2375; fax: +34 93 581 1264.  
E-mail address: [josep.vendrell@uab.cat](mailto:josep.vendrell@uab.cat) (J. Vendrell).

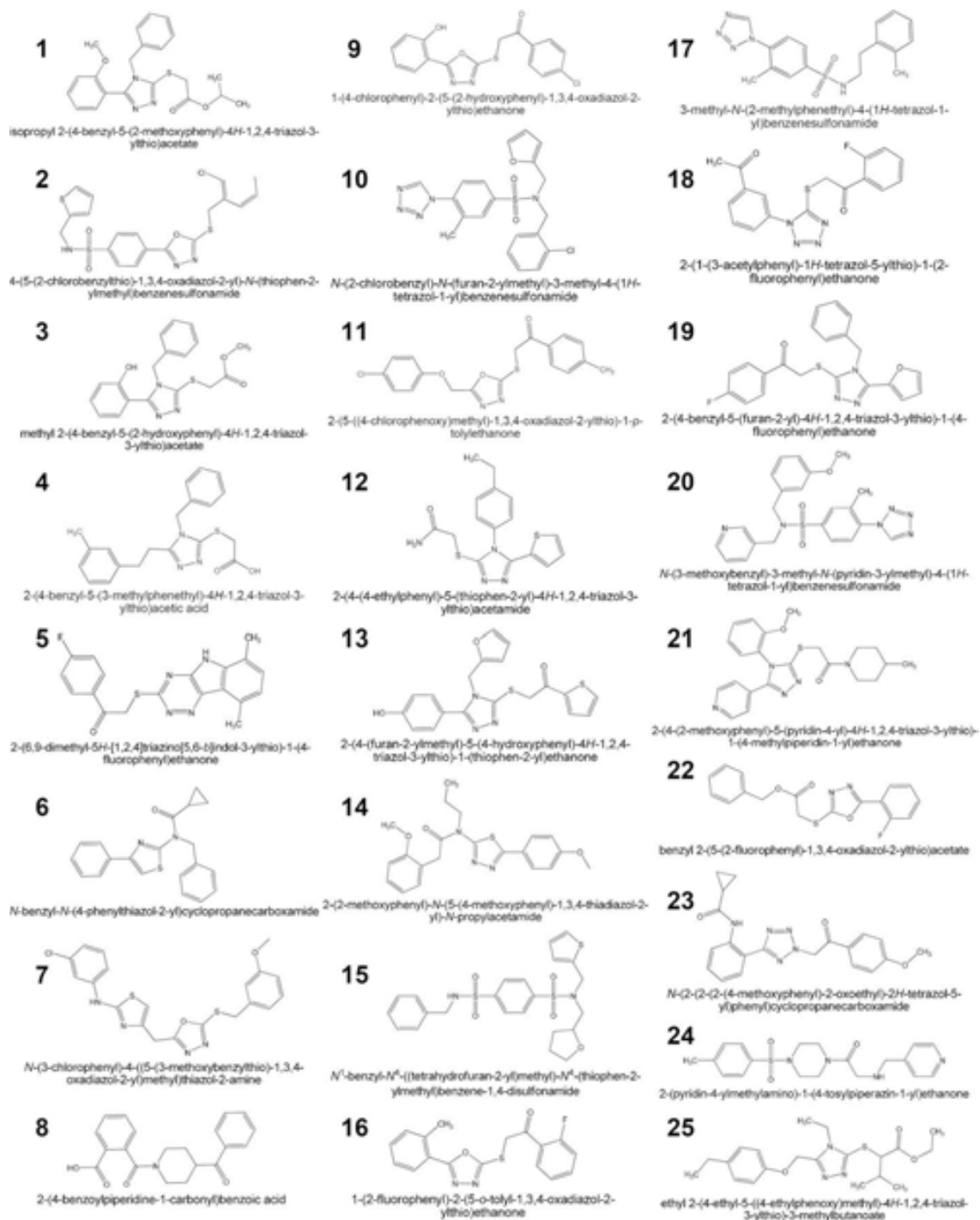


Fig. 1. Chemical structures of the tested compounds.



consideration a number of factors such as the number of functionalities of the molecules, their bulkiness, stereochemistry and susceptibility to protease attack, as outlined under **Experimental protocols**. The molecular weight of the selected compounds falls in the range 328–506 Da, with an average value of 388 Da. They contain between 4 and 11 flexible bonds, with the number of hydrogen bond donors and acceptors ranging from 0 to 1 and from 3 to 9, respectively. With only two exceptions (**4** and **8**), no compounds display ionisable groups. The log *P* values range from 0.34 (compound **23**) to 4.60 (compound **16**). Two compounds (**15** and **25**) possess a stereocentre. By comparison, a known inhibitor, BX 528 (MW = 496.3; log *P* = 1.51), possesses 4 and 8 hydrogen bond donors and acceptors, 13 flexible bonds, a carboxylate, and two chiral atoms. Interestingly, this indicates that the retrieved molecules are structurally dissimilar, and have a lower molecular weight than a recognized CP inhibitor.

The presence of five-membered hetero-ring systems is remarkable since 20 out of 25 (representing 80% of the total) contain an oxadiazole, tetrazole, thiadiazole or triazole ring system in their structure. Such ring systems are not uncommon in the literature. 1,3,4-oxadiazoles have been used as amide bond surrogates in benzodiazepine receptor agonists [20], muscarinic receptor agonists [21], neurokinin 1 receptor antagonists [22], antiinflammatory agents [23], and Phe-Gly peptidomimetics [24]. Several non-peptide angiotensin II receptor antagonists like (1-((2'-(2*H*-tetrazol-5-yl)biphenyl-4-yl)methyl)-2-butyl-4-chloro-1*H*-imidazol-5-yl)methanol (Losartan, known by the trade name Cozaar) bear a tetrazole ring [25] while 1,3,4-thiadiazole compounds showed potent matrix metalloproteinase inhibition [26]. Posaconazole, a broad-spectrum orally active antifungal [27], and the hypnotic/sedative benzodiazepine triazolam (marketed under brand names Halcion/Novodorm/Songar) [28], are examples of therapeutic agents containing a triazolyl nucleus. This fact might be related to the high percent of drug-like structures within the screening collection database. Further, it suggests that the high-throughput screening of large databases may provide valuable known fragments that can be assayed against unspotted biological targets.

## 2.2. Inhibition of metalloproteinases

The inhibitory ability of the selected test set was evaluated in vitro against two B-type metalloproteinases, namely corn earworm (*Helicoverpa zea*) CPB (HzCPB), an inhibitor-resistant digestive protease from a lepidopteran pest insect [29], and human CPB (hCPB), a disease-linked metalloproteinase. Bovine CPA (bCPA1), a model zinc-dependent protease, was included in the assay given the structural proximity among M14 A- and B-type proteins. As can be seen in Table 1, the inhibitory potency falls in the lower micromolar range, with five compounds (20% of the total) showing *K<sub>i</sub>* values below 10 μM. Inhibition of bCPA1 was also detected, but it was weaker as compared to that of CPB in most cases. The values for the best compounds are comparable to those for two well-known reference inhibitors [30,31] also included in Table 1.

Interestingly, the three most potent CPB inhibitors, **3**, **6** and **7**, harbor three different chemical classes of five-membered hetero-rings: 1,2,4-triazole, thiazole, and 1,3,4-oxadiazole, respectively. Compound **3**, methyl 2-(4-benzyl-5-(2-hydroxyphenyl)-4*H*-1,2,4-triazol-3-ylthio)acetate, shows a 30-fold CPB/CPA inhibitory ratio, suggesting some specific mode of action. This ratio increases to 160-fold in the case of **6**, *N*-benzyl-*N*-(4-phenylthiazol-2-yl)cyclopropanecarboxamide, while **7**, *N*-(3-chlorophenyl)-4-((5-(3-methoxybenzylthio)-1,3,4-oxadiazol-2-yl)methyl)thiazol-2-amine, exerted the most potent human CPB inhibitory effect, its *K<sub>i</sub>* being 1.2 μM, although with a moderate 20-fold CPB/CPA ratio. The latter

is an extended molecule (four planar rings) whereas compounds, **3** and **6** display three planar systems, suggesting that **7**, which is also the best among 1,3,4-oxadiazoles (namely, **2**, **9**, **11**, **16**, **22**), is ideally suited to bind the CPB active site cleft. Compound **2** is comparable in size to **7**, but it behaves as a poor human CPB inhibitor (*K<sub>i</sub>* = 40.0 μM), probably indicating a less well tailored structure.

Compounds retrieved from the high-throughput virtual screening survey are listed in descending docking energy value in Table 1. It is clear that docking energy ranking scores do not strictly correlate with the measured *K<sub>i</sub>* values. For example, for the 25% top ranked in docking energy (compounds **1–7**), only four (57%) are within the 25% best *K<sub>i</sub>* values, and only one of them, **3**, matches the ranking in both categories, while **7**, which ranks first in *K<sub>i</sub>* value and is the most potent human CPB inhibitor, is only 7th in docking energy. Three of the top 25% best scores in docking energy (compounds **2**, **4**, and **5**) fall well below the tenth position in the *K<sub>i</sub>* value ranking. Compounds **2** and **7**, belonging to the same 1,3,4-oxadiazole class showed opposite behaviours as, although **2** is a poor human CPB inhibitor, it scores better than **7**. An inspection of their structures revealed that the 1,3,4-oxadiazole ring is directly joined to the bulky benzenesulfonamide moiety in **2**, while the substituent group in **7** is bonded through a methylene spacer. This could reduce the flexibility of **2** and impair the placing of its 1,3,4-oxadiazole in the zinc coordination sphere. Otherwise, we might conceive that the contribution of the sulfonamide group to binding was overestimated or even misplaced during the screening process.

Compounds **1**, **3**, and **4** are 1,2,4-triazoles. While **1** and **3** show a good agreement between docking energy score and *K<sub>i</sub>* value, **4** deviates from this trend, and is indeed one of the poorest inhibitors. The structures of these compounds are similar, but **4** displays a free carboxylate while the other compounds are the methyl and isopropyl esters. A possible explanation for these data is that the free carboxylate might have been overestimated as a binding group by the docking algorithm. Overall, the fact that the predictions were poor for some compounds might be ascribed to some weaknesses of the docking algorithm to handle certain types of chemical bonding (i.e., the coordination sphere of the zinc ion and its ligands) or to our manual knowledge-based filtering procedure, which was based on structural inspection rather than on docking energy scores.

Although others had reported on five-membered hetero-rings like imidazole as CPA and CPB inhibitors [32–36], none of them has yet reached the clinical practice for the treatment of diseases related to the M14 family of proteases. Thus, the 1,3,4-oxadiazole containing compounds can be regarded as a new type of small-molecule inhibitors of various proteases belonging to the M14 family. To better understand the behaviour of this new type of compounds we performed molecular docking of selected 1,3,4-oxadiazole inhibitors to different M14 target proteins.

## 2.3. Analysis of 1,3,4-oxadiazole binding to metalloproteinases

Compound **7** was chosen for molecular docking based on its behaviour as the most potent inhibitor of the tested set. The docking procedure was validated by redocking of known crystal complex structures. Although the overall quality of the redocking was satisfactory in terms of rmsd values (see **Experimental protocols** for details), a failure to predict a salt bridge in one of the cases implies that the reliability of the atomic distances derived from the docking procedures is not absolute.

Since compound **7** inhibits both the human and insect CPB counterparts with fairly similar potency, docking was initially carried out in parallel. As expected, the binding to insect CPB proceeds much like that to the human counterpart, so we focused

**Table 1**  
Inhibitory constants<sup>a</sup> and docking energy values.<sup>b</sup>

Compound	bCPA1 <sup>c</sup>	hCPB	HzCPB	Docking energy	Compound	bCPA1 <sup>c</sup>	hCPB	HzCPB	Docking energy
<b>1</b>	160.0	7.4	36.1	-12.63	<b>15</b>	19.4	23.0	-	-11.03
<b>2</b>	67.3	40.0	1.4	-12.06	<b>16</b>	39.3	14.7	40.0	-11.01
<b>3</b>	159.0	4.9	41.6	-11.75	<b>17</b>	138.0	25.9	38.0	-10.82
<b>4</b>	217.0	196.0	57.0	-11.67	<b>18</b>	550.0	33.6	53.8	-10.82
<b>5</b>	86.3	15.0	9.0	-11.64	<b>19</b>	236.0	74.5	-	-10.65
<b>6</b>	800.0	4.5	1.2	-11.53	<b>20</b>	148.0	23.9	28.4	-10.58
<b>7</b>	22.5	1.2	3.6	-11.27	<b>21</b>	381.0	20.8	30.2	-10.47
<b>8</b>	320.0	20.7	19.6	-11.25	<b>22</b>	32.0	67.2	33.4	-10.46
<b>9</b>	2.1	7.5	1.0	-11.23	<b>23</b>	41.3	292.0	31.0	-10.44
<b>10</b>	159.0	16.6	12.6	-11.22	<b>24</b>	250	364.0	18.0	-10.42
<b>11</b>	36.7	11.7	12.0	-11.13	<b>25</b>	22.0	10.8	11.9	-10.05
<b>12</b>	155.0	11.6	27.2	-11.11	<b>BzISA<sup>d</sup></b>	0.5	10.0	11.9	
<b>13</b>	74.5	9.1	-	-11.08	<b>GEMSA<sup>e</sup></b>	nd	1.1	nd	
<b>14</b>	128.0	25.1	19.7	-11.07					

<sup>a</sup> Measured  $K_i$  value ( $\mu\text{M}$ ); (-), no inhibition observed at the highest concentration tested; nd: no data available.

<sup>b</sup> Compounds are ordered in descending docking energy value, in kcal/mol.

<sup>c</sup> bCPA1, bovine carboxypeptidase A1; hCPB, human carboxypeptidase B; HzCPB, carboxypeptidase B from corn earworm (*Helicoverpa zea*).

<sup>d</sup> BzISA, *DL*-benzylsuccinic acid.

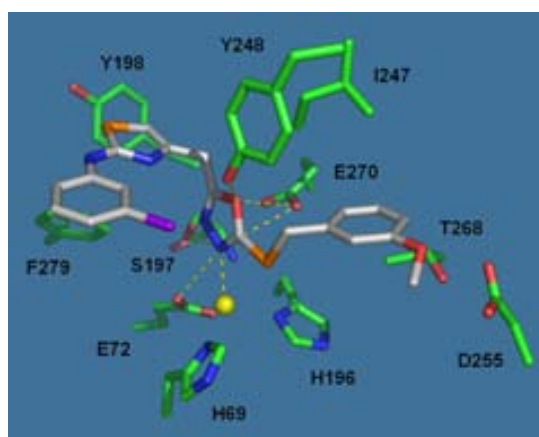
<sup>e</sup> GEMSA, *DL*-guanidinoethylmercaptosuccinic acid.

our analysis on the vertebrate enzyme, comparing the putative CPB/7 complex structure with a known CPB/peptide inhibitor structure. Besides, to better characterize the fairly selective CPB binding mode, we also analyzed the 7-docked CPA structure. From that comparison, striking differences in the binding of 7 to CPA and CPB arised that deserved further analysis, as will be discussed below.

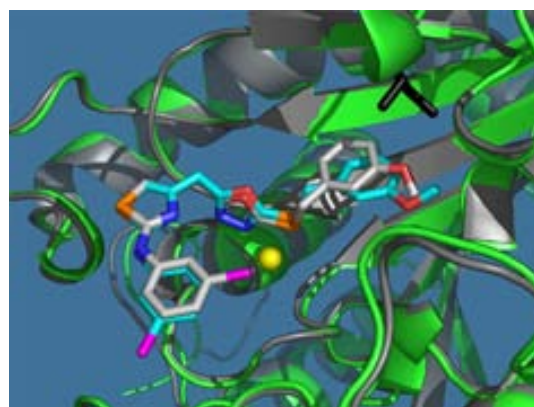
The predicted binding of 7 to human CPB suggests that it involves the zinc ion and key residues lining the active site cleft (Fig. 2). One of the nitrogen atoms of the 1,3,4-oxadiazole ring coordinates the zinc ion (at 2.3 Å), while the other forms a strong hydrogen bond to the carbonyl O atom of Ser197 (a residue at subsite S2). The oxadiazolyl O atom forms hydrogen bonds (at 2.9 Å) to the Tyr248 phenolic hydroxyl and to the Glu270 O $\epsilon$ 2 atom (3.1 Å). In this way, the 1,3,4-oxadiazole hetero-ring is engaged in interactions with key catalytic residues at the S1' and S1 subsites. The 1,3,4-oxadiazole hetero-ring seems to bisect the molecule in two parts: the methoxyphenyl moiety becomes buried into the CPB

specificity pocket while the chlorobenzyl group is left just at the entrance of the active site cleft. In the former, the methoxy oxygen atom forms a strong hydrogen bond to Ser207 hydroxyl (at 2.7 Å), while the aromatic ring is placed in a wide hydrophobic environment sustained by Ile203, Gly243, Ile247, Ala250, and Thr268. The methoxy oxygen atom appears too far (at a distance > 3.5 Å) and probably is not hydrogen bonded to Asp255, the specificity determinant residue in CPB. Some other atoms of the chlorobenzyl and the adjacent thiazole ring seem to be less favourably engaged in interactions to residues lining the active site cleft. The sulfur atom of the thiazole ring linking the chlorobenzyl and the 1,3,4-oxadiazole hetero-ring contributes marginally to binding (no interaction of this atom with any protein residue was observed). The sulfur atom is at 3.2 Å from Tyr198 phenolic hydroxyl and at almost 3.9 Å from a carbon atom of Phe279 aromatic ring (Tyr198 and Phe279 are in the S2 and S3 subsites, respectively).

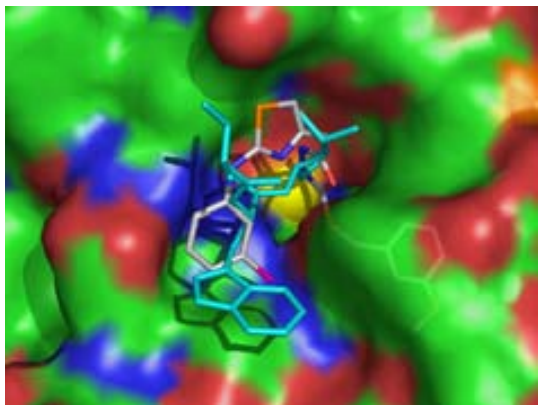
A clear difference was evident from the comparison of the complexes of CPA and CPB with compound 7 (Fig. 3). In bCPA1, the methoxyphenyl aromatic ring is placed in an environment similar to that in CPB. However, the bulky Ile243 side chain in bCPA1



**Fig. 2.** A view of the putative binding of inhibitor 7 (stick model, white carbons) in the active site of human CPB. The 1,3,4-oxadiazole ring engages two catalytically important residues, Tyr248 and Glu270 as well as the  $\text{Zn}^{2+}$  ion. CPB side chains at a distance <5 Å from compound 7 are depicted in stick model. The  $\text{Zn}^{2+}$  ion is shown as a yellow sphere. Other residues important for binding are labelled. Carbon: green, sulfur: orange, nitrogen: blue, oxygen: red and chlorine: magenta.



**Fig. 3.** Superimposition of the complexes of inhibitor 7 with CPA and CPB. Compound 7 (white carbons in CPB, light blue in CPA) is shown in stick model. CPB (green) and CPA (grey) are in cartoon representation. The  $\text{Zn}^{2+}$  ion is shown as a yellow sphere. Ile243 in CPA is highlighted black (a glycine is in the equivalent position in CPB) to mark the offset of 7. Approximately the same orientation as in Fig. 2.



**Fig. 4.** CPB in complex with inhibitor **7** (white carbons) compared to the complex with tick proteinaceous inhibitor, TCI. Only the three C-terminal residues from TCI, Val72–Trp73–Leu74 (cyan carbons), are shown. The Zn<sup>2+</sup> ion is coloured yellow and CPB as a translucent surface in atom-type colour. The bulge located south-east from the Zn ion corresponds to the catalytically crucial Tyr248. The view corresponds approximately to the orientation displayed in Fig. 2 after a 70° rotation downwards.

(the equivalent CPB residue is Gly243) is in close proximity to the phenyl ring of the methoxyphenyl moiety, leading to a steric clash between them. As a result, this group is displaced, probably resulting in an impaired binding of the inhibitor to bCPA1 and causing a drop in inhibitory potency. This emphasizes that an aromatic fragment placed at this position in the 1,3,4-oxadiazole scaffold might be of help in the design of CPB specific variants.

As compared to the structure of CPB bound to the proteinaceous inhibitor from ticks, TCI, the binding of the chlorobenzyl moiety recalls that of the penultimate tryptophan residue from the peptidic inhibitor (Fig. 4). From this comparison another point of optimization of the 1,3,4-oxadiazole scaffold arises, as the chlorobenzyl group makes a suboptimal interaction that demands a larger aromatic moiety. The outcome from the molecular docking experiment illustrates that the 1,3,4-oxadiazole scaffold is ideally adapted to attach to key residues at the active site of B-type carboxypeptidases, while there are points beyond this scaffold that perform a less advantageous exploitation of anchorage points and may therefore be taken into consideration as a subject for further improvement.

### 3. Conclusion

Following a high-throughput virtual screening survey, we obtained new five-membered hetero-ring inhibitors of metal-carboxypeptidases. The poses retrieved from the screening display chemical fragments found in several marketed drugs, indicating the drug-like content of the screening database. Once obtained, these known chemical fragments were then challenged against untried biological targets. Some of the selected poses performed an inhibitory activity against the target protein. The screening procedure performed efficiently regarding the top ranked compounds, as the docking energy scores reflect the measured  $K_i$  values. For other compounds the predictions were poor, either due to some weaknesses of the docking algorithm or to our particular knowledge-based filtering methodology.

The most active compounds show different five-membered hetero-ring structures, and exert potent human CPB inhibition. In particular, **7**, *N*-(3-chlorophenyl)-4-((5-(3-methoxybenzylthio)-1,3,4-oxadiazol-2-yl)methyl)thiazol-2-amine, is a low micromolar

inhibitor of this enzyme. This compound features an unprecedented 1,3,4-oxadiazole five-membered hetero-ring scaffold, distinct from known M14 small-molecule inhibitors. From molecular docking simulations of **7** and comparison to a known peptide inhibitor, some points of improvement can be suggested that might lead to an optimized version. We envisage that these modifications could yield second generation compounds with improved features in terms of inhibition profiles and selectivity. Therefore, the 1,3,4-oxadiazole scaffold is proposed as an interesting lead for further development.

### 4. Experimental protocols

#### 4.1. General procedures and materials for the biological experiments

Human CPB, and corn earworm (*H. zea*) CPB, were recombinantly expressed as the zymogen form in the methylotrophic *Pichia pastoris* system. They were then proteolytically converted to the mature, enzymatically active, form as published [29,37]. Bovine pancreatic CPA was from Sigma. Enzyme concentrations for the kinetic studies were kept fixed at typically 5–50 nM. Two chromogenic synthetic substrates (Bachem, Bubendorf, Switzerland) were used: *N*-(4-methoxyphenylazoformyl)-Phe-OH (Aaf-Phe-OH;  $K_m$  for CPA: 100  $\mu$ M) and *N*-(4-methoxyphenylazoformyl)-Arg-OH (Aaf-Arg-OH;  $K_m$  for CPB: 60  $\mu$ M). The assay concentrations for Aaf-Phe-OH and Aaf-Arg-OH were 100 and 60 or 200  $\mu$ M, respectively. The experimental assays were performed at room temperature in 50 mM Tris, 0.5 M NaCl, pH 7.5 buffer for CPA and in 20 mM Tris, 0.1 M NaCl, pH 7.5 buffer for CPB, with variable concentrations of the analyzed compounds between 50 nM and 250  $\mu$ M. The initial velocity measurements were performed at least per triplicate at room temperature in a Victor3 microtiter plate reader (PerkinElmer, Waltham, MA, USA), following the changes in absorbance continuously at 340 nm. Data were fed into the GraphPad Version 5.0 package (GraphPad Software, San Diego, CA, USA) that was used to derive the kinetic parameters. The inhibition was fitted to a competitive, one-site model, with a good agreement to the experimental observations. The  $K_i$  value was then calculated from the IC<sub>50</sub> following the procedure by Cheng and Prusoff [38].

#### 4.2. Bioinformatics: screening procedures

The ArgusDock molecular modelling engine [39] implemented within the package ArgusLab Version 4.0.1 ([www.arguslab.com](http://www.arguslab.com)) was used to query the ASINEX Platinum (APC) screening collection (version of August 2006; size > 125,000). The APC is a diverse, non-focused, non-prefiltered, small-molecule screening collection and contains a high percentage of drug-like molecules [40]. The screening procedure was run on bench top PCs under Windows. To ease the screening procedure, the APC was divided into small size sub-libraries (2500 compounds each) with CORINA [41]. The final set of poses was obtained after coarse minimization, re-clustering and ranking. The 10 lowest-energy poses were retained from each sub-library, and were stored for further analysis.

The 500 lowest-energy poses were then analyzed with the FAF-Drugs server (<http://bioserv.rpbs.jussieu.fr/FAFDrugs.html>) [42]. The properties monitored included molecular weight, number of hydrogen bond acceptors and donors, number of flexible and rigid bonds, number of carbon and hetero-atoms, estimation of the partition coefficient ( $\log P$ ), and calculation of polar surface area. This automatic procedure was followed by a careful, knowledge-based, manual filtering to yield a selected final set. The knowledge-based manual filtering was performed by visual inspection of the compounds, which were selected from the preference of CPs to

bind to short peptides, applying the following general rules: i) compounds containing a linear and extended backbone having 3–5 functionalities (i.e. ring systems and/or trigonal planar  $sp^2$ -carbons, joined by single bonds) were preferred as the first choice; ii) compounds having bulky systems, like those containing three or more fused rings, and tribenzene-containing molecules, were discarded; iii) compounds having stereogenic centres were also disregarded; and iv) structures containing a chemical linkage susceptible to be hydrolyzed by the target enzyme were also eliminated.

The 3D crystal structure of human CPB (coordinates file code: 1zli) was used for the screening. After removal of ligands, a box centred at the catalytic zinc was defined as the binding site. The size of the box,  $20 \times 15 \times 20 \text{ \AA}$ , encompasses all the known subsites for substrate binding, thus allowing a complete sampling of the CPB active site cleft. A flexible ligand docking was employed.

#### 4.3. Bioinformatics: molecular docking

The structural analysis of the most potent inhibitors was carried out by molecular docking with the AutoDock 4 package [43]. A binding box was built, centred at the catalytic zinc ion with a size defined to encompass all known binding sites in the active site. The lowest-energy conformer was used for the discussion and a careful analysis of hydrogen bonds and intermolecular interactions was performed. The approach was validated by redocking of small-molecules appearing in known crystal complex structures: *L*-phenyllactate for CPA (PDB code 2ctc) and an imidazole inhibitor for CPB (PDB code 2jew). Redocking was correct in terms of global conformation, orientation of the ligand and rmsd values (0.22 Å and 0.62 Å for *L*-phenyllactate and the imidazole, respectively). In the latter case, the docking correctly predicted the coordination of the zinc and the hydrogen bond between the terminal  $\text{NH}_2$  and Asp255. However, it failed to predict a salt bridge between the ligand and Arg145. A superimposition of compound 7 and the known imidazole inhibitor showed a similar orientation and location of the five-membered rings. The human CPB structure in complex to a proteinaceous inhibitor from ticks (PDB code 1zli) was used for comparison and superimposition. Programs of the Collaborative Computational Project Number 4 suite [44] were used to analyze protein–ligand interactions. Visualization and molecular graphics were done with ArgusLab 4.0.1 ([www.arguslab.com](http://www.arguslab.com)) or with PyMol ([www.pymol.org](http://www.pymol.org)). All the final figures were prepared with PyMol.

#### Acknowledgements

Support by grants BIO2007-68046 (Ministerio de Educación y Ciencia, Spain), 2005SGR-1037 (Generalitat de Catalunya) and CAMP project 108830 (VI EU Framework Programme) is gratefully acknowledged. DF wishes to thank the European Union A1þan Programme for a PhD fellowship (contract no. E04D035150AR).

#### References

- J. Vendrell, F.X. Aviles, L.D. Fricker, Metalloproteinases, in: A. Messerschmidt, W. Bode, M. Cygler (Eds.), *Handbook of Metalloproteins*, John Wiley and Sons Ltd., Chichester, 2004, pp. 176–189.
- M. Rodríguez de la Vega, R.G. Sevilla, A. Hermoso, J. Lorenzo, S. Tanco, A. Diez, L.D. Fricker, J.M. Bautista, F.X. Aviles, *FASEB J.* 21 (2007) 851–865.
- H. Wang, Q. Zhou, J.W. Kesinger, C. Norris, C. Valdez, *Exp. Biol. Med.* (Maywood) 232 (2007) 1170–1180.
- Q. Zhou, A.C. Law, J. Rajagopal, W.J. Anderson, P.A. Gray, D.A. Melton, *Dev. Cell* 13 (2007) 103–114.
- C.A. Muller, S. Appelos, W. Uhl, M.W. Buchler, A. Borgstrom, *Gut* 51 (2002) 229–235.
- S. Asai, T. Sato, T. Tada, T. Miyamoto, N. Kimbara, N. Motoyama, H. Okada, N. Okada, *J. Immunol.* 173 (2004) 4669–4674.
- J.L. Willemse, D.F. Hendriks, *Front. Biosci.* 12 (2007) 1973–1987.
- F. Fialka, R.M. Gruber, R. Hitt, L. Opitz, E. Brunner, H. Schliephake, F.J. Kramer, *Oral Oncol.* 44 (2008) 941–948.
- O. Hataji, O. Taguchi, E.C. Gabazza, H. Yuda, C.N. D'Alessandro-Gabazza, H. Fujimoto, Y. Nishii, T. Hayashi, K. Suzuki, Y. Adachi, *Am. J. Hematol.* 76 (2004) 214–219.
- P. Cronet, W. Knecht, C.A.-C. Malmberg Hager, M. Andersson, C. Furebring, Sequences of modified human carboxypeptidase U (CPU) with increased thermal stability for treating hemorrhage disorders, Astrazeneca AB, Swed., Astrazeneca UK Limited, Application: WO. Patent No. 2005052149, 2005.
- S. Golz, U. Brueggemeier, A. Geerts, Diagnostics and therapeutics for diseases associated with human carboxypeptidase A3 based on tissue expression profiling, Bayer Healthcare AG, Germany, Application: WO. Patent No. 2006010495, 2006.
- S. Golz, U. Brueggemeier, A. Geerts, Diagnostics and therapeutics for diseases associated with human carboxypeptidase A2 based on tissue expression profiling, Bayer Healthcare AG, Germany, Application: WO. Patent No. 2006010499, 2006.
- S. Matsugi, T. Hamada, N. Shioi, T. Tanaka, T. Kumada, S. Satomura, *Clin. Chim. Acta* 378 (2007) 147–153.
- O. Yanes, J. Villanueva, E. Querol, F.X. Aviles, *Nat. Protoc.* 2 (2007) 119–130.
- M.S. Buchanan, A.R. Carroll, A. Edser, M. Sykes, G.A. Fechner, P.I. Forster, G.P. Guymer, R.J. Quinn, *Bioorg. Med. Chem. Lett.* 18 (2008) 1495–1497.
- S.H. Wang, S.F. Wang, W. Xuan, Z.H. Zeng, J.Y. Jin, J. Ma, G.R. Tian, *Bioorg. Med. Chem.* 16 (2008) 3596–3601.
- Y.X. Wang, V. da Cunha, J. Vincelette, L. Zhao, M. Nagashima, K. Kawai, S. Yuan, K. Emayan, I. Islam, J. Hosoya, M.E. Sullivan, W.P. Dole, J. Morser, B.O. Buckman, R. Vergona, *Thromb. Haemostasis* 97 (2007) 54–61.
- Y.X. Wang, L. Zhao, M. Nagashima, J. Vincelette, D. Sukovich, W. Li, B. Subramanyam, S. Yuan, K. Emayan, I. Islam, P. Hrvatin, J. Bryant, D.R. Light, R. Vergona, J. Morser, B.O. Buckman, *Thromb. Haemostasis* 97 (2007) 45–53.
- D. Fernandez, O. Illa, F.X. Aviles, V. Branchadell, J. Vendrell, R.M. Ortuno, *Bioorg. Med. Chem.* 16 (2008) 4823–4828.
- W.R. Tully, C.R. Gardner, R.J. Gillespie, R. Westwood, *J. Med. Chem.* 34 (1991) 2060–2067.
- B.S. Orlek, F.E. Blaney, F. Brown, M.S. Clark, M.S. Hadley, J. Hatcher, G.J. Riley, H.E. Rosenberg, H.J. Wadsworth, P. Wyman, *J. Med. Chem.* 34 (1991) 2726–2735.
- T. Ladduwahetty, R. Baker, M.A. Cascieri, M.S. Chambers, K. Haworth, L.E. Keown, D.E. MacIntyre, J.M. Metzger, S. Owen, W. Rycroft, S. Sadowski, E.M. Seward, S.L. Shephard, C.J. Swain, F.D. Tattersall, A.P. Watt, D.W. Williamson, R.J. Hargreaves, *J. Med. Chem.* 39 (1996) 2907–2914.
- M.D. Mullican, M.W. Wilson, D.T. Connor, C.R. Kostlan, D.J. Schrier, R.D. Dyer, *J. Med. Chem.* 36 (1993) 1090–1099.
- S. Borg, R.C. Vollaing, M. Labarre, K. Payza, L. Terenius, K. Luthman, *J. Med. Chem.* 42 (1999) 4331–4342.
- D.H. Smith, *Drugs* 68 (2008) 1207–1225.
- J.Y. Winum, A. Scozzafava, J.L. Montero, C.T. Supuran, *Med. Res. Rev.* 26 (2006) 767–792.
- J.E. Frampton, L.J. Scott, *Drugs* 68 (2008) 993–1016.
- G.E. Pakes, R.N. Brogden, R.C. Heel, T.M. Speight, G.S. Avery, *Drugs* 22 (1981) 81–110.
- A. Bayes, M. Comellas-Bigler, M. Rodriguez de la Vega, K. Maskos, W. Bode, F.X. Aviles, M.A. Jongsma, J. Beekwilder, J. Vendrell, *Proc. Natl. Acad. Sci. U.S.A.* 102 (2005) 16602–16607.
- L.D. Byers, R. Wolfenden, *Biochemistry* 12 (1973) 2070–2078.
- T.J. McKay, A.W. Phelan, T.H. Plummer Jr., *Arch. Biochem. Biophys.* 197 (1979) 487–492.
- J.C. Barrow, P.G. Nantermet, S.R. Stauffer, P.L. Ngo, M.A. Steinbeiser, S.S. Mao, S.S. Carroll, C. Bailey, D. Colussi, M. Bosserman, C. Burlein, J.J. Cook, G. Sitko, P.R. Tiller, C.M. Miller-Stein, M. Rose, D.R. McMasters, J.P. Vacca, H.G. Selnick, *J. Med. Chem.* 46 (2003) 5294–5297.
- M.E. Bunnage, J. Blagg, J. Steele, D.R. Owen, C. Allerton, A.B. McElroy, D. Miller, T. Ringer, K. Butcher, K. Beaumont, K. Evans, A.J. Gray, S.J. Holland, N. Feeder, R.S. Moore, D.G. Brown, *J. Med. Chem.* 50 (2007) 6095–6103.
- M.S. Han, D.H. Kim, *Bioorg. Med. Chem. Lett.* 11 (2001) 1425–1427.
- K.J. Lee, K.C. Joo, E.J. Kim, M. Lee, D.H. Kim, *Bioorg. Med. Chem.* 5 (1997) 1989–1998.
- P.G. Nantermet, J.C. Barrow, S.R. Lindsley, M. Young, S.S. Mao, S. Carroll, C. Bailey, M. Bosserman, D. Colussi, D.R. McMasters, J.P. Vacca, H.G. Selnick, *Bioorg. Med. Chem. Lett.* 14 (2004) 2141–2145.
- S. Ventura, V. Villegas, J. Sterner, J. Larson, J. Vendrell, C.L. Hershberger, F.X. Aviles, *J. Biol. Chem.* 274 (1999) 19925–19933.
- Y. Cheng, W.H. Prusoff, *Biochem. Pharmacol.* 22 (1973) 3099–3108.
- S. Joy, P.S. Nair, R. Hariharan, M.R. Pillai, *In Silico Biol.* 6 (2006) 601–605.
- M. Krier, G. Bret, D. Rognan, *J. Chem. Inf. Model.* 46 (2006) 512–524.
- J. Gasteiger, C. Rudolph, J. Sadowski, *Tetrahedron Comput. Methodol.* 3 (1990) 537–547.
- M.A. Miteva, S. Violas, M. Montes, D. Gomez, P. Tuffery, B.O. Villoutreix, *Nucleic Acids Res.* 34 (2006) W738–W744.
- G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, *J. Comput. Chem.* 19 (1998) 1639–1662.
- Collaborative Computational Project, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 50 (1994) 760–776.



## Thioxophosphoranyl aryl- and heteroaryloxiranes as the representants of a new class of metallo-carboxypeptidase inhibitors

Daniel Fernández,<sup>a</sup> Ona Illa,<sup>b</sup> Francesc X. Avilés,<sup>a</sup> Vicenç Branchadell,<sup>b</sup>  
Josep Vendrell<sup>a,\*</sup> and Rosa M. Ortuño<sup>b,\*</sup>

<sup>a</sup>Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, and Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain

<sup>b</sup>Departament de Química, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain

Received 4 February 2008; revised 12 March 2008; accepted 20 March 2008

Available online 23 March 2008

**Abstract**—A novel and potent family of metallo-carboxypeptidase inhibitors based on thioxophosphoranyl oxiranes is presented. These compounds bear aryl or heteroaryl substituents with *trans*-stereochemistry with respect to the phosphorylated group and they have been synthesized by the addition of [bis(diisopropylamino)phosphino](trimethylsilyl)carbene to the corresponding aldehydes and the subsequent thiolation of the phosphine. These oxiranes contain a tetrahedral P atom harboring shielded *N,N*-groups. The screening of their biological activity as metallo-carboxypeptidase inhibitors and some structural studies, as well as full experimental details for the new compounds, is disclosed. Thus, from the analysis of their activity against the prototypical metallo-carboxypeptidases A and B (CPA and CPB), we have observed that hydrophobic phosphorylated oxiranes perform better as CPB inhibitors, reaching *K<sub>i</sub>* values comparable to classical synthetic carboxypeptidase inhibitors. X-ray diffraction analysis revealed that the packing in the structure of one phosphorylated oxirane is mediated mainly by hydrophobic contacts and that the *N,N*-groups are highly flexible. Consequently, phosphorylated oxiranes might constitute an attractive material for subsequent improvements in the design of novel inhibitors against human proteolytic enzymes with enhanced oral availability.

© 2008 Elsevier Ltd. All rights reserved.

### 1. Introduction

The search for drugs targeted to proteases is an extremely fertile field as suggested by the current developments,<sup>1</sup> and much more progress would be attained if the bioavailability problem associated to most new designs could be conveniently tackled.<sup>2</sup>

Phosphorylated compounds are widely used in the design of enzyme inhibitors. This class of products bear charged groups that endow them with high affinity toward the biological target. In particular, phosphorus (V) containing organic compounds have attracted much attention as inhibitors of metalloproteases because this functionality can yield a variety of metal-coordinating possibilities. The phosphorus-containing moiety provides strong binding to the metal ion at the active site

in metalloenzymes, while the rest of the molecule accounts for selectivity issues. In the case of the model metalloprotease carboxypeptidase A (CPA), the phosphonic and thiophosphonic class of inhibitors have been considered as multisubstrate analogues, or as transition-state mimetics, due to the presence of a tetrahedrally hybridized phosphorus atom at the position occupied by either a trigonal planar carbon atom in the ground state of the substrate or a tetrahedrally hybridized carbon atom in the transition state. Analysis of the mode of binding of these inhibitors has provided a more profound knowledge on the catalytic mechanism of CPA in particular and of mechanistically related enzymes in general.<sup>3</sup>

Probably, the first examples of phosphorylated organic compounds acting as CP inhibitors are the phosphoramidates **I–III** (Fig. 1). Their design can be exemplified by the *N*-phosphorylated amino acid phosphoryl-L-phenylalanine, P-Phe-OH, **I** (R = OH), which fulfills the requirements of strong chelation of the active site zinc atom by the phosphoryl group and a hydrophobic interaction between the side chain of the P1' amino acid res-

**Keywords:** Thioxophosphoranyl oxiranes; Metallo-carboxypeptidases; Human carboxypeptidase B; Inhibitor screening.

\* Corresponding authors. Tel.: +34 93 581 16 02; fax: +34 93 581 12 65 (R.M.O.); e-mail addresses: [josep.vendrell@uab.es](mailto:josep.vendrell@uab.es); [rosa.ortuno@uab.es](mailto:rosa.ortuno@uab.es)

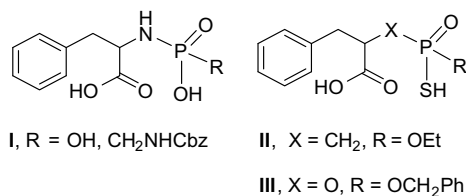


Figure 1. Structure of some representative CP inhibitors.

idue and the S1' subsite (the specificity pocket) of the enzyme. Thiophosphorylated compounds such as **II**<sup>4</sup> and **III**<sup>5</sup> have also been shown to inhibit CPA. All these compounds resemble the structure of benzylsuccinic acid, a standard inhibitor to CPA and, not surprisingly, display similar inhibitory potency. Remarkably, an earlier study showed that *O,O*-disubstituted *N*-phosphoryl derivatives of amino acids did not inhibit CPA even at high concentrations.<sup>6</sup> This strongly suggests that a phosphoryl group accommodating free oxygen atoms is required for binding and that blocking the phosphoryl oxygens eliminated the inhibition.

From the mechanistic standpoint, much of our knowledge on the catalytic mechanism of metalloenzymes is based on that of CPA. This knowledge was exploited in the design of potent CPA inhibitors in the form of transition-state analogues containing a phosphorus center.<sup>7</sup>

Thus, *N*-[[[(benzyloxycarbonyl)amino]methyl]hydroxyphosphinyl]-*L*-phenylalanine (ZG<sup>P</sup>P), **I** (R = CH<sub>2</sub>NHCbz), the phosphoramidate analogue of the CPA substrate carbobenzoxyglycyl-*L*-phenylalanine, represents the first of a series of extremely potent CPA inhibitors.<sup>3</sup> Further achievements within this series led to phosphonate tripeptides, in which the last amide bond is replaced by a phosphorus-containing moiety. These compounds can reach *K<sub>i</sub>* values in the femtomolar range<sup>8–10</sup> and their crystal structures in complex with CPA support the notion of their behavior as transition-state analogues.<sup>11,12</sup> The design of tripeptide phosphoramidates has inspired a series of tight-binding inhibitors to human plasma CPB (alias CPU or TAFI),<sup>13</sup> a protein that belongs to the CPA family and is regarded as a target for therapeutic intervention because of its involvement in blood hemostasis.

A highly desirable goal of our approach is to search for a suitable solution to the design of compounds that could overcome the problem of poor oral availability associated to phosphorylated drugs. Since phosphorus-containing compounds have been shown to interact primarily with the zinc at the enzymes' active site, we wished to investigate the effects of a complete shielding of the charges in the phosphonate functionality on the binding affinity. Therefore, a discovery program was launched in search of novel phosphorus-containing scaffolds, such as the phosphorylated oxiranes. These compounds have displayed relevant activities as antibacterial agents<sup>14–16</sup> and as antibiotics. Among the

latter, phosphomycin, (1*R*,2*S*)-epoxypropylphosphonic acid (Fig. 2), is the most prominent instance.<sup>17,18</sup>

Starting from an easy and stereoselective synthetic strategy to prepare differently substituted phosphoranyl oxiranes bearing alkyl<sup>19,20</sup> and aryl<sup>21</sup> substituents, we decided to extend this methodology to the synthesis of new heteroaryl substituted epoxides and to test the inhibitory activity toward CPA and CPB of some representatives of these families. From the initial screening, we found a series of phosphorus-containing oxiranes which displayed biological activity against CPA and CPB, the prototypical metalcarboxypeptidases belonging to the MC clan of proteases.<sup>22</sup> Thus, we report herein on the synthesis, the identification, and the biological characterization of a new class of metalcarboxypeptidase inhibitors, the thioxophosphoranyl aryl- and heteroaryloxiranes.

## 2. Results and discussion

### 2.1. Chemistry: synthesis of the thioxophosphoranyl oxiranes

Oxiranes that contain aromatic rings bearing electron-withdrawing (NO<sub>2</sub>, CF<sub>3</sub>) or electron-releasing (OMe, Me) substituents as well as several heterocycle moieties were envisioned as representatives of the thioxophosphoranyl aryl- and heteroaryloxiranes. Compounds **3a–d** were prepared according to the protocol previously developed in our laboratory<sup>19–21</sup> that involves the stereoselective addition of nucleophilic phosphinocarbene **2**,<sup>23</sup> resulting from photochemical decomposition of the precursor diazoalkane **1**, to aldehydes followed by the thiolation of the produced phosphinoepoxide with elemental sulfur (Scheme 1). This methodology has been successfully applied to the synthesis of the new heteroaryloxiranes **3e–k**, which were isolated in 56–92% yield. The reaction progress was monitored by <sup>31</sup>P NMR spectroscopy following the disappearance of the carbene signal of **2** (δ = –46). Thiolation was achieved in about 30 min, and the corresponding epoxy thioxophosphonamides **3e–k** were characterized by chemical shifts at 83–86 ppm. No significant differences were observed in the behavior of heteroaromatic aldehydes. *trans*-Stereochemistry was assigned by comparison with other previously synthesized epoxides whose relative configuration was assigned by X-ray diffraction structural analysis.<sup>19–21</sup> The <sup>3</sup>J<sub>P–H</sub> coupling constants for compounds **3e–k** were between 7.5 and 9.0 Hz, in good agreement with known derivatives.

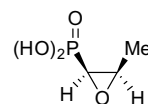
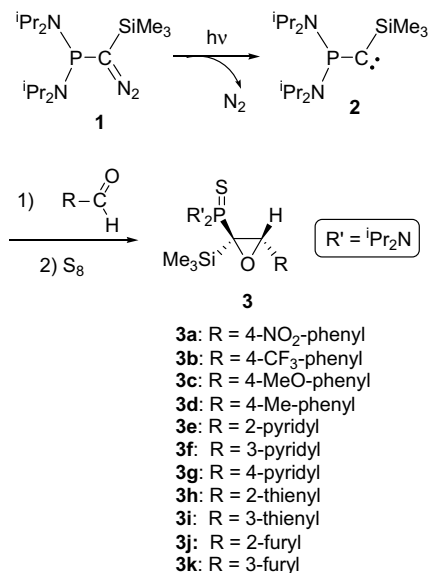


Figure 2. The antibiotic phosphomycin.



Scheme 1.

## 2.2. Biology: inhibition of carboxypeptidases

The enzyme inhibition assays were performed applying the equation of Cheng and Prusoff,<sup>24</sup> using a fixed concentration of substrate and variable concentrations of inhibitor (usually five concentrations in the range 100 nM–500 μM). A nonlinear global fit to a competitive one-site model gave the best fitting of the experimental observations. Figure 3 shows the results for the two most potent inhibitors against CPB. The model and the experimental data were in excellent agreement ( $R^2 > 0.95$ ) for all the assayed compounds.

The  $K_i$  values obtained for the 11 thioxophosphoranyl aryl- or heteroaryloxiranes acting against two prototyp-

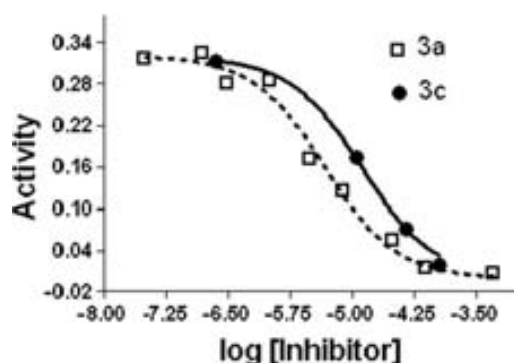


Figure 3. Plot of CPB activity, measured as the slope of the absorbance change vs time, as a function of inhibitor concentration for compounds **3a** and **3c**.

ical metallocarboxypeptidases (bovine CPA and human CPB) are shown in Table 1. The capacity of thioxophosphoranyl aryl- and heteroaryloxiranes to act as inhibitors of both CPA and CPB is unprecedented, since none of the previously known *O,O*-disubstituted phosphoramidates were able to inhibit CPA.<sup>6</sup> A further unsuspected feature is that the thioxophosphoranyl aryl- and heteroaryloxiranes perform better as CPB inhibitors. CPB and CPB-like enzymes like CPU/TAFI have a preference to excise C-terminal basic lysine or arginine residues, whereas CPA cleaves substrates ending with aromatic and aliphatic residues.<sup>25</sup> In fact, all the thioxophosphoranyl oxiranes tested in this work are potent CPB inhibitors, the best examples belonging to the aryl series (**3a–d**), with slight differences probably caused by the different substituents of the phenyl ring. Thus, there seems to be a preference for groups containing an oxygen atom, as evidenced by **3a** and **3c**. The pyridine series (**3e–g**) ranks next, with best inhibitory potency with the nitrogen in the 4-position. The heteroaryl oxiranes-containing a thienyl moiety (**3h–i**) seem to be more potent than those containing a furyl ring (**3j–k**).

Compounds that selectively interact with one type of protease are interesting from the point of view of drug design, and specially in the case of CPB-type metallocarboxypeptidases as most enzymes of this subfamily are involved in physiologically relevant processes like the maturation of hormones or neuropeptides or, as commented above, in maintaining blood homeostasis.<sup>25</sup> The inhibitory potency of oxiranes **3a–k** against CPB is comparable to the phosphoramidates containing a free phosphoryl group. However, the structure of the former is clearly different from the known CPB peptide substrates, and this is probably one of the reasons why there are several orders of magnitude poorer CP inhibitors as compared to the tripeptide phosphonates. This fact might suggest that, although the phosphorus-containing moiety is disubstituted, and its capability to coordinate to the zinc ion is thus diminished in comparison with a free group, the substituents have a non-deleterious effect on binding.

It should be added that, although no systematic screening of our series of compounds against other families of proteases was undertaken, a control performed using the model serine protease trypsin as a target showed no inhibitory capacity for any of the compounds at any of the assayed concentrations.

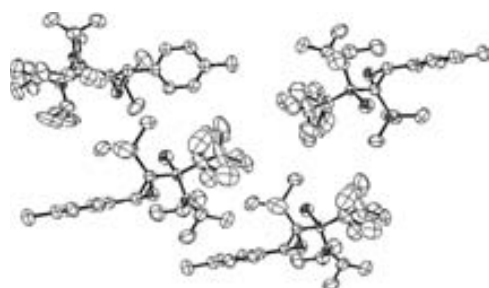
## 2.3. X-ray crystallography: structure of oxirane **3d**

Although attempts to crystallize CPA and CPB in complex with our small-molecule inhibitors were not successful, a large single crystal grew in a drop containing CPB and inhibitor **3d**. After physically treating the crystal to rule out that it was not of protein content, it was subjected to X-ray diffraction. The subsequent structural analysis revealed that it corresponded to none of the possible inorganic components in the crystallization solution (acetate, cacodylate, PEG, or Tris) but instead to compound **3d**, the packing structure of which is shown in Figure 4.

**Table 1.** Biological activity of thioxophosphoranyl aryl- or heteroaryloxiranes **3a–k**

Compound	$K_i$ ( $\mu\text{M}$ )	
	CPA	CPB
<b>3a</b>	46.0 (9)	1.06 (0.15)
<b>3b</b>	300.0 (20)	3.9 (0.6)
<b>3c</b>	70.0 (14)	3.0 (0.4)
<b>3d</b>	43.5 (8)	9.1 (2.3)
<b>3e</b>	53.3 (8.5)	15.1 (2.2)
<b>3f</b>	26.2 (4.5)	16.7 (2.4)
<b>3g</b>	29.6 (4.5)	9.7 (1.5)
<b>3h</b>	150.0 (18)	7.2 (2.1)
<b>3i</b>	62.0 (14)	9.2 (2.6)
<b>3j</b>	74.0 (16)	23.5 (3.9)
<b>3k</b>	124.0 (22)	16.7 (3.9)

Values inside parentheses indicate the standard error of the mean (SEM).

**Figure 4.** Crystal packing for oxirane **3d** as determined by X-ray structural analysis.

The crystal structure of **3d** is different from that of the published phenyl compound.<sup>21</sup> One of the diisopropylamino groups is disordered and was refined using a two sites model with a 0.60:0.40 occupancy ratio. The disorder is reflected in the large atomic displacement parameters of all the carbon atoms in this group. The possible H-bonds that **3d** can form are all intramolecular, and no electrostatic intermolecular interactions can be detected. Furthermore, the packing diagram revealed that **3d** associates in the unit cell mainly by the interactions mediated by the hydrophobic groups. The disordered diisopropyl group is located in a hydrophobic environment close to the phenyl ring. The high flexibility of the diisopropyl moiety might allow for the adoption of different conformations in accordance with the environment with which it is in contact. This can have important consequences, for example, upon binding the active site of an enzyme one might speculate that steric strains introduced by the bulky phenyl substituents in the *O,O*-disubstituted phosphoramidates would impair the binding of the compound. By contrast, our *N,N*-disubstituted compounds harbor flexible groups that can be better adapted to fit in a hydrophobic pocket at the active site of the enzyme.

### 3. Conclusion

We have found a novel and potent family of metallo-carboxypeptidase inhibitors based on oxiranes bearing

a bis(diisopropylamino) thioxophosphoranyl group and aryl- or heteroaryl substituents with *trans*-stereochemistry. Our results show, for the first time, that a fully shielded P moiety might act as an inhibitor of a metal-dependent enzyme. Regarding CPs, it was also unanticipated that the hydrophobic phosphorylated oxiranes perform better as inhibitors toward CPB, a protease with preference to cleave residues with basic character. Phosphorylated oxiranes can be regarded as a starting point in the design of novel inhibitors against human CPB or a more relevant pharmaceutical target like the CPB-like variant plasma CPU/TAFI. The benefits of our approach include the versatility of the scaffold, its low molecular weight, and the possibility to control the hydrophilic nature without compromising their activity. This can be devised by choosing the appropriate pattern of substitutions, for example, by screening different groups at the oxirane ring or different *N*-substituents in the amino phosphine functions. Structural factors controlling the efficiency in the binding to the enzyme are under active investigation in our laboratories.

## 4. Experimental

### 4.1. General procedures and materials for the biological experiments

Bovine pancreatic CPA was from Sigma. Human CPB was prepared as published.<sup>26,27</sup> Enzyme concentrations for the kinetic studies were kept fixed at typically 5–50 nM. Magnesium acetate (Merck), sodium cacodylate, Tris (Tris = 2-amino-2-(hydroxymethyl)propane-1,3-diol), and PEG 8000 (Polyethyleneglycol) were from Sigma. The substrates, *N*-(4-methoxyphenylazofonyl)-Phe-OH (Aaf-Phe-OH) and *N*-(4-methoxyphenylazofonyl)-Arg-OH (Aaf-Arg-OH) were from Bachem (Switzerland). The  $K_m$  values for Aaf-Phe-OH and Aaf-Arg-OH substrates for CPA and CPB are 100 and 60  $\mu\text{M}$ , respectively. The final Aaf-Phe-OH and Aaf-Arg-OH concentrations used were 100 and 200  $\mu\text{M}$ , respectively. The experimental assays were performed at 50 mM Tris, 0.5 M NaCl, pH 7.5 buffer and 20 mM Tris, 0.1 M NaCl, pH 7.5 buffer for CPA and CPB, respectively. The initial velocity measurements were performed at least per triplicate at room temperature in a Victor3 (PerkinElmer, USA) microtiter plate reader. The changes in absorbance were followed continuously at 340 nm. The GraphPad Version 5.0 package ([www.graphpad.com](http://www.graphpad.com)) was used to process the data.

### 4.2. Synthesis of thiophosphoranyloxiranes (3a–k)

**4.2.1. General procedures and materials.** All the manipulations were performed under an inert atmosphere of nitrogen by using standard Schlenk techniques. Dry, oxygen-free solvents were employed. Carbene **2** was prepared according to Ref. 1. All employed aldehydes were distilled before each reaction. <sup>31</sup>P NMR downfield chemical shifts are expressed with a positive sign, in ppm, relative to external 85% H<sub>3</sub>PO<sub>4</sub>. Microanalysis of the synthesized compounds used to afford erratic results



since the combustion of carbon was systematically incomplete. Purity criterion was assessed from cut-range melting points and homogeneity of  $^{31}\text{P}$ ,  $^1\text{H}$ , and  $^{13}\text{C}$  NMR spectra of these products.

As a general procedure for the synthesis of the thiophosphoranyloxiranes, 0.7 mmol of the dried and freshly distilled aldehyde was added to a solution of carbene **2**<sup>23</sup> (0.7 mmol) in pentane (3 mL) and the resultant solution was stirred at room temperature for 10 min under nitrogen atmosphere. Then an excess of elemental sulfur was added and the mixture was stirred for 30 min. The solvent was removed at reduced pressure and the residue was chromatographed on neutral silica gel (9:1 hexane/ether) to afford the pure product.

The new heteroaryl oxiranes **3e–k** were fully characterized as follows.

**4.2.2. (2*R,S*,3*R,S*)-3-(2-Pyridyl)-2-trimethylsilyl-2-oxiran-2-yl-*N,N,N',N'*-tetraisopropylthiophosphonodiamide (3e).** 77% Yield. Crystals, mp 137–139 °C (from methanol).  $^{31}\text{P}$  NMR (101.2 MHz, acetone-*d*<sub>6</sub>):  $\delta$  82.1.  $^1\text{H}$  NMR (250 MHz, acetone-*d*<sub>6</sub>):  $\delta$  0.06 (s, 9H, Si(CH<sub>3</sub>)<sub>3</sub>), 1.40–1.51 (m, 24H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 3.97–4.19 (m, 4H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 4.53 (d,  $^3J_{\text{P-H}} = 9.0$  Hz, 1H, H<sub>3</sub>), 7.37 (m, 2H, H<sub>3'</sub> and H<sub>5'</sub>), 8.60 (m, 2H, H<sub>4'</sub> and H<sub>6'</sub>).  $^{13}\text{C}$  NMR (62.5 MHz, acetone-*d*<sub>6</sub>):  $\delta$  1.6 (Si(CH<sub>3</sub>)<sub>3</sub>), 24.1–25.1 ((CH<sub>3</sub>)<sub>2</sub>CHN–), 48.3 (d,  $^2J_{\text{C-P}} = 5.7$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 49.5 (d,  $^2J_{\text{C-P}} = 5.2$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 61.7 (d,  $^1J_{\text{C-P}} = 70.1$  Hz, C<sub>2</sub>), 63.5 (C<sub>3</sub>), 122.0 and 123.7 (C<sub>3'</sub> and C<sub>5'</sub>), 136.7 (C<sub>4'</sub>), 149.9 (C<sub>6'</sub>), 156.8 (C<sub>2'</sub>). MS (*m/e*): 456.2 (M+H<sup>+</sup>).

**4.2.3. (2*R,S*,3*R,S*)-3-(3-Pyridyl)-2-trimethylsilyl-2-oxiran-2-yl-*N,N,N',N'*-tetraisopropylthiophosphonodiamide (3f).** 56% Yield. Crystals, mp 121–123 °C (from methanol).  $^{31}\text{P}$  NMR (101.2 MHz, acetone-*d*<sub>6</sub>):  $\delta$  82.6.  $^1\text{H}$  NMR (250 MHz, acetone-*d*<sub>6</sub>):  $\delta$  0.03 (s, 9H, Si(CH<sub>3</sub>)<sub>3</sub>), 1.41–1.51 (m, 24H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 3.98–4.22 (m, 4H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 4.58 (d,  $^3J_{\text{P-H}} = 8.6$  Hz, 1H, H<sub>3</sub>), 7.40 (ddd,  $^3J_{\text{H-H}} = 7.9$  Hz,  $^3J_{\text{H-H}} = 5.0$  Hz,  $J_{\text{H-H}} = 0.5$  Hz, 1H, H<sub>5'</sub>), 7.74 (m, 1H, H<sub>4'</sub>), 8.57 (m, 2H, H<sub>2'</sub> and H<sub>6'</sub>).  $^{13}\text{C}$  NMR (62.5 MHz, acetone-*d*<sub>6</sub>):  $\delta$  2.1 (Si(CH<sub>3</sub>)<sub>3</sub>), 24.4–25.5 ((CH<sub>3</sub>)<sub>2</sub>CHN–), 48.6 (d,  $^2J_{\text{C-P}} = 5.7$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 49.9 (d,  $^2J_{\text{C-P}} = 4.8$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 61.4 (C<sub>3</sub>), 62.1 (d,  $^1J_{\text{C-P}} = 70.6$  Hz, C<sub>2</sub>), 123.8 (C<sub>5'</sub>), 132.5 (C<sub>3'</sub>), 135.4 (C<sub>4'</sub>), 149.3 and 150.2 (C<sub>2'</sub> and C<sub>6'</sub>). MS (*m/e*): 456.2 (M+H<sup>+</sup>).

**4.2.4. (2*R,S*,3*R,S*)-3-(4-Pyridyl)-2-trimethylsilyl-2-oxiran-2-yl-*N,N,N',N'*-tetraisopropylthiophosphonodiamide (3g).** 66% Yield. Crystals, mp 164–165 °C (from methanol).  $^{31}\text{P}$  NMR (101.2 MHz, acetone-*d*<sub>6</sub>):  $\delta$  27.18.  $^1\text{H}$  NMR (250 MHz, acetone-*d*<sub>6</sub>):  $\delta$  0.05 (s, 9H, Si(CH<sub>3</sub>)<sub>3</sub>), 1.42–1.50 (m, 24H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 3.99–4.19 (m, 4H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 4.52 (d,  $^3J_{\text{P-H}} = 8.9$  Hz, 1H, H<sub>3</sub>), 7.37 (m, 2H, H<sub>3'</sub> and H<sub>5'</sub>), 8.60 (m, 2H, H<sub>2'</sub> and H<sub>6'</sub>).  $^{13}\text{C}$  NMR (62.5 MHz, acetone-*d*<sub>6</sub>):  $\delta$  2.0 (Si(CH<sub>3</sub>)<sub>3</sub>), 24.3–25.4 ((CH<sub>3</sub>)<sub>2</sub>CHN–), 48.6 (d,  $^2J_{\text{C-P}} = 4.8$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 50.0 (d,  $^2J_{\text{C-P}} = 4.8$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 62.1 (C<sub>3</sub>), 62.5 (d,  $^2J_{\text{C-P}} = 69.6$  Hz, C<sub>2</sub>), 123.0 (C<sub>3'</sub> and

C<sub>5'</sub>), 145.6 (C<sub>4'</sub>), 150.4 (C<sub>2'</sub> and C<sub>6'</sub>). MS (*m/e*): 456.2 (M+H<sup>+</sup>).

**4.2.5. (2*R,S*,3*R,S*)-3-(2-Thienyl)-2-trimethylsilyl-2-oxiran-2-yl-*N,N,N',N'*-tetraisopropylthiophosphonodiamide (3h).** 92% Yield. Crystals, mp 134–136 °C (from methanol).  $^{31}\text{P}$  NMR (101.2 MHz, acetone-*d*<sub>6</sub>):  $\delta$  86.5.  $^1\text{H}$  NMR (250 MHz, acetone-*d*<sub>6</sub>):  $\delta$  0.13 (s, 9H, Si(CH<sub>3</sub>)<sub>3</sub>), 1.38–1.49 (m, 24H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 3.94–4.15 (m, 4H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 4.66 (d,  $^3J_{\text{P-H}} = 7.7$  Hz, 1H, H<sub>3</sub>), 7.04–7.46 (m, 2H, H<sub>heterocycle</sub>), 7.46 (m, 1H, H<sub>heterocycle</sub>).  $^{13}\text{C}$  NMR (62.5 MHz, acetone-*d*<sub>6</sub>):  $\delta$  1.8 (Si(CH<sub>3</sub>)<sub>3</sub>), 24.3–25.5 ((CH<sub>3</sub>)<sub>2</sub>CHN–), 48.5 (d,  $^2J_{\text{C-P}} = 5.7$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 49.8 (d,  $^2J_{\text{C-P}} = 4.7$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 60.6 (C<sub>3</sub>), 63.2 (d,  $^1J_{\text{C-P}} = 68.7$  Hz, C<sub>2</sub>), 126.4, 126.9, and 127.8 (C<sub>3'</sub>, C<sub>4'</sub> and C<sub>5'</sub>), 139.8 (C<sub>2'</sub>). MS (*m/z*): 461.1 (M+H<sup>+</sup>).

**4.2.6. (2*R,S*,3*R,S*)-3-(3-Thienyl)-2-trimethylsilyl-2-oxiran-2-yl-*N,N,N',N'*-tetraisopropylthiophosphonodiamide (3i).** 77% Yield. Crystals, mp 115–117 °C (from methanol).  $^{31}\text{P}$  NMR (101.2 MHz, acetone-*d*<sub>6</sub>):  $\delta$  82.2.  $^1\text{H}$  NMR (250 MHz, acetone-*d*<sub>6</sub>):  $\delta$  0.07 (s, 9H, Si(CH<sub>3</sub>)<sub>3</sub>), 1.39–1.50 (m, 24H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 3.97–4.49 (m, 4H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 4.50 (d,  $^3J_{\text{P-H}} = 8.1$  Hz, 1H, H<sub>3</sub>), 7.11 (d,  $^3J_{4',5'} = 5.0$  Hz, 1H, H<sub>4'</sub>), 7.31 (m, 1H, H<sub>2'</sub>), 7.51 (dd,  $^3J_{4',5'} = 5.0$  Hz,  $^4J_{2',5'} = 3.0$  Hz, 1H, H<sub>5'</sub>).  $^{13}\text{C}$  NMR (62.5 MHz, acetone-*d*<sub>6</sub>):  $\delta$  1.8 (Si(CH<sub>3</sub>)<sub>3</sub>), 24.4–25.5 ((CH<sub>3</sub>)<sub>2</sub>CHN–), 48.5 (d,  $^2J_{\text{C-P}} = 4.8$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 49.8 (d,  $^2J_{\text{C-P}} = 4.8$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 60.9 (C<sub>3</sub>), 61.5 (d,  $^1J_{\text{C-P}} = 70.6$  Hz, C<sub>2</sub>), 123.4, 127.1, and 128.0 (C<sub>2'</sub>, C<sub>4'</sub> and C<sub>5'</sub>), 138.0 (C<sub>3'</sub>). MS (*m/e*): 461.1 (M+H<sup>+</sup>).

**4.2.7. (2*R,S*,3*R,S*)-3-(2-Furyl)-2-trimethylsilyl-2-oxiran-2-yl-*N,N,N',N'*-tetraisopropylthiophosphonodiamide (3j).** 67% Yield. Crystals, mp 107–108 °C (from methanol).  $^{31}\text{P}$  NMR (101.2 MHz, acetone-*d*<sub>6</sub>):  $\delta$  82.7.  $^1\text{H}$  NMR (250 MHz, acetone-*d*<sub>6</sub>):  $\delta$  0.12 (s, 9H, Si(CH<sub>3</sub>)<sub>3</sub>), 1.37–1.49 (m, 24H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 3.93–4.16 (m, 4H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 4.42 (d,  $^3J_{\text{P-H}} = 7.5$  Hz, 1H, H<sub>3</sub>), 6.38 (m, 2H, H<sub>heterocycle</sub>), 6.47 (m, 1H, H<sub>heterocycle</sub>), 7.61 (m, 1H, H<sub>5'</sub>).  $^{13}\text{C}$  NMR (62.5 MHz, acetone-*d*<sub>6</sub>):  $\delta$  0.8 (Si(CH<sub>3</sub>)<sub>3</sub>), 23.8–25.0 ((CH<sub>3</sub>)<sub>2</sub>CHN–), 48.0 (d,  $^2J_{\text{C-P}} = 5.5$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 49.3 (d,  $^2J_{\text{C-P}} = 5.0$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 57.8 (C<sub>3</sub>), 62.1 (d,  $^1J_{\text{C-P}} = 69.6$  Hz, C<sub>2</sub>), 109.9 and 111.6 (C<sub>3'</sub> and C<sub>4'</sub>), 143.7 (C<sub>5'</sub>), 150.4 (C<sub>2'</sub>). MS (*m/e*): 445.2 (M+H<sup>+</sup>).

**4.2.8. (2*R,S*,3*R,S*)-3-(3-Furyl)-2-trimethylsilyl-2-oxiran-2-yl-*N,N,N',N'*-tetraisopropylthiophosphonodiamide (3k).** 91% Yield. Crystals, mp 145–147 °C (from methanol).  $^{31}\text{P}$  NMR (101.2 MHz, acetone-*d*<sub>6</sub>):  $\delta$  82.0.  $^1\text{H}$  NMR (250 MHz, acetone-*d*<sub>6</sub>):  $\delta$  0.13 (s, 9H, Si(CH<sub>3</sub>)<sub>3</sub>), 1.38–1.49 (m, 24H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 3.95–4.14 (m, 4H, (CH<sub>3</sub>)<sub>2</sub>CHN–), 4.35 (dd,  $^3J_{\text{P-H}} = 8.2$  Hz,  $^4J_{\text{H-H}} = 1.1$  Hz, 1H, H<sub>3</sub>), 6.49 (dd,  $J_{\text{H-H}} = 1.8$  Hz,  $J'_{\text{H-H}} = 0.8$  Hz, 1H, H<sub>heterocycle</sub>), 7.48 (m, 1H, H<sub>heterocycle</sub>), 7.58 (dd,  $J_{\text{H-H}} = J'_{\text{H-H}} = 1.8$  Hz, 1H, H<sub>heterocycle</sub>).  $^{13}\text{C}$  NMR (62.5 MHz, acetone-*d*<sub>6</sub>):  $\delta$  1.9 (Si(CH<sub>3</sub>)<sub>3</sub>), 24.4–25.5 ((CH<sub>3</sub>)<sub>2</sub>CHN–), 48.4 (d,  $^2J_{\text{C-P}} = 5.2$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 49.8 (d,  $^2J_{\text{C-P}} = 5.2$  Hz, (CH<sub>3</sub>)<sub>2</sub>CHN), 57.8 (C<sub>3</sub>), 61.5 (d,

$^1J_{C-P} = 71.0$  Hz,  $C_2$ ), 111.2 ( $C_4$ ), 122.0 ( $C_3$ ), 141.5 and 144.3 ( $C_7$  and  $C_5$ ). MS (*m/e*): 445.1 (M+H<sup>+</sup>).

#### 4.3. Crystallization and structure determination of **3d**

Co-crystallization procedures were taken from the literature.<sup>12,28</sup> CPA and CPB solutions were added to millimolar amounts of each inhibitor, and 1  $\mu$ l aliquots from the solutions were put in contact with an equal amount of reservoir solution (0.2 M magnesium acetate, 0.1 M sodium cacodylate, 0.02 M Tris-HCl (Tris: 2-amino-2-(hydroxymethyl)propane-1,3-diol), and 20% PEG 8000 (Polyethyleneglycol)). The vapor diffusion method was used, with hanging drops containing 1  $\mu$ l:1  $\mu$ l mixture of protein and reservoir solution. Twenty-four well Greiner plates, containing 500  $\mu$ l reservoir volume and sealed with siliconized glass cover slides were left in a crystal farm at 16 °C.

Data collection of the small-molecule inhibitor **3d** was carried out on a Bruker diffractometer at the MoK $\alpha$  wavelength. The Wingx suite of programs<sup>29</sup> was used to solve, refine, and analyze the structure. This compound has been deposited at the Cambridge Crystallographic Data Centre and allocated the deposition number CCDC 671796.

#### Acknowledgments

The authors are indebted to S. Bartolomé (Servei de Fotodocumentació, UAB) for technical assistance in the use of Victor3, and to Dr. Á. Álvarez-Larena for the data collection at the Servei de Raigs X (UAB). D.F. thank the Europea Union Al $\beta$ n Programme for a PhD fellowship (contract no. E04D035150AR). Financial support from Ministerio de Educación y Ciencia (CTQ2007-61704/BQU, BIO2004-05879) and Generalitat de Catalunya (2005SGR-103, 2005SGR-1037, Laboratori Europeu Associat LTPMM), is gratefully acknowledged.

#### References and notes

- Mittl, P. R.; Grutter, M. G. *Curr. Opin. Struct. Biol.* **2006**, *16*, 769.
- Turk, B. *Nat. Rev. Drug Discov.* **2006**, *5*, 785.
- Bartlett, P. A.; Marlowe, C. K.; Giannousis, P. P.; Hanson, J. E. *Cold Spring. Harb. Symp. Quant. Biol.* **1987**, *52*, 83.
- Grobelyny, D.; Goli, U. B.; Galardy, R. E. *Biochem. J.* **1985**, *232*, 15.
- Hill, J. M.; Lowe, G. *J. Chem. Soc. Perkin Trans 1: Org. Bioorg. Chem.* **1995**, 2001.
- Kam, C. M.; Nishino, N.; Powers, J. C. *Biochemistry* **1979**, *18*, 3032.
- Jacobsen, N. E.; Bartlett, P. A. *J. Am. Chem. Soc.* **1981**, *103*, 654.
- Hanson, J. E.; Kaplan, A. P.; Bartlett, P. A. *Biochemistry* **1989**, *28*, 6294.
- Kaplan, A. P.; Bartlett, P. A. *Biochemistry* **1991**, *30*, 8165.
- Phillips, M. A.; Kaplan, A. P.; Rutter, W. J.; Bartlett, P. A. *Biochemistry* **1992**, *31*, 959.
- Kim, H.; Lipscomb, W. N. *Biochemistry* **1990**, *29*, 5546.
- Kim, H.; Lipscomb, W. N. *Biochemistry* **1991**, *30*, 8171.
- Wang, Y. X.; Zhao, L.; Nagashima, M.; Vincelette, J.; Sukovich, D.; Li, W.; Subramanyam, B.; Yuan, S.; Emayan, K.; Islam, I.; Hrvatin, P.; Bryant, J.; Light, D. R.; Vergona, R.; Morser, J.; Buckman, B. O. *Thromb. Haemost.* **2007**, *97*, 45.
- Chemerdá, J. M.; Sletzing, M. Application: DE Patent 69-1924172, 1970, 24 pp.
- Christensen, B. G.; Cama, L. D. Application: DE Patent 70-2002415, 1970, 89 pp.
- Firestone, R. A. Application: DE Patent 69-1924093, 1970, 28 pp.
- Hendlin, D.; Stapley, E. O.; Jackson, M.; Wallick, H.; Miller, A. K.; Wolf, F. J.; Miller, T. W.; Chaiet, L.; Kahan, F. M.; Foltz, E. L.; Woodruff, H. B.; Mata, J. M.; Hernandez, S.; Mochales, S. *Science* **1969**, *166*, 122.
- Hori, T.; Horiguchi, M.; Hayashi, A., Eds.; *Biochemistry of Natural Carbon-Phosphorus Compounds*, 1984.
- Illa, O.; Alvarez-Larena, A.; Baceiredo, A.; Branchadell, V.; Ortuno, R. M. *Tetrahedron: Asymmetry* **2007**, *18*, 2617.
- Illa, O.; Gornitzka, H.; Branchadell, V.; Baceiredo, A.; Bertrand, G.; Ortuno, R. M. *Eur. J. Org. Chem.* **2003**, 3147.
- Illa, O.; Gornitzka, H.; Baceiredo, A.; Bertrand, G.; Branchadell, V.; Ortuno, R. M. *J. Org. Chem.* **2003**, *68*, 7707.
- Rawlings, N. D.; Morton, F. R.; Barrett, A. J. *Nucleic Acids Res.* **2006**, *34*, D270.
- Igau, A.; Baceiredo, A.; Trinquier, G.; Bertrand, G. *Angew. Chem.* **1989**, *101*, 617.
- Cheng, Y.; Prusoff, W. H. *Biochem. Pharmacol.* **1973**, *22*, 3099.
- Vendrell, J.; Aviles, F. X.; Fricker, L. D. In *Handbook of Metalloproteins*; Messerschmidt, A., Bode, W., Cygler, M., Eds.; John Wiley & Sons, Ltd.: Chichester, 2004; Vol. 3, p 176.
- Reverter, D.; Ventura, S.; Villegas, V.; Vendrell, J.; Aviles, F. X. *J. Biol. Chem.* **1998**, *273*, 3535.
- Ventura, S.; Villegas, V.; Sterner, J.; Larson, J.; Vendrell, J.; Hershberger, C. L.; Aviles, F. X. *J. Biol. Chem.* **1999**, *274*, 19925.
- Adler, M.; Bryant, J.; Buckman, B.; Islam, I.; Larsen, B.; Finster, S.; Kent, L.; May, K.; Mohan, R.; Yuan, S.; Whitlow, M. *Biochemistry* **2005**, *44*, 9339.
- Farrugia, L. D. *J. Appl. Crystallogr.* **1999**, *32*, 837.

## Structural and Functional Characterization of Binding Sites in Metalloproteases Based on Optimal Docking Area Analysis

Daniel Fernández,<sup>1,2</sup> Josep Vendrell,<sup>1,2</sup> Francesc X. Avilés,<sup>1,2</sup> and Juan Fernández-Recio<sup>3,4\*</sup>

<sup>1</sup>Institut de Biotecnologia i Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

<sup>2</sup>Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

<sup>3</sup>Institute for Research in Biomedicine, Parc Científic de Barcelona, 08028 Barcelona, Spain

<sup>4</sup>Life Sciences Department, Barcelona Supercomputing Center, 08034 Barcelona, Spain

**ABSTRACT** The metalloproteases (MCPs) belonging to the clan MC were studied by the Optimal Docking Area (ODA) method to evaluate protein–protein binding sites and to provide a basis for the identification of binding partners for this class of enzymes. The ODA method identifies surface patches with optimal desolvation energy based on the selection of low-energy docking regions, generated from a set of surface points around the protein. With few exceptions, the ODA method identified surface patches with a significant low-energy docking surface for all the MCPs with known three-dimensional structure. Overall, in 14 out of 24 cases, the detected ODA patches were correctly located (i.e. more than 50% of the predicted residues were in known protein–protein binding sites), yielding a global success rate of 58%. More specifically, the success rate increased up to 80% on the ODA patches detected for the catalytic domains of the M14A subfamily, independently on the partner. Interestingly, the ODA residues on the catalytic domain were correctly located in the interface with the N-terminal pro domain in all MCPs. The spatial distribution of the ODA patches for the different members of the family is in relation to the origin and function of the particular MCP, which allowed distinguishing between them. In good agreement with the experimentally characterized protein interfaces, the total average surface area of the theoretically derived ODA patches for the catalytic domain of MCPs is around 1700 Å<sup>2</sup> and their content in hydrophobic residues is about 40%. As a particular case, the average surface area of the ODA patches in MCPs of crop insect pests is about twice that of the MCPs of vertebrates, which might be related to their particular function. We recognized two binding regions for the catalytic domain of the MCPs, one of them accounting for nearly all the known intermolecular interactions made up by the enzymes. Protein inhibitors seem to have evolved to dock on this subset of ODA patches, evoking the binding mode of the N-terminal pro domains. The second binding region detected, for which no ligands have been identified

so far, seems to be related to the acquisition/maintenance of the native structure of the peptidase. Overall, the ODA method has been successful in identifying low-energy docking areas in a set of structurally and functionally related proteins, suggesting that it can be easily extended to other families in the search for protein–protein binding sites and for their functional significance. *Proteins* 2007;68:131–144. © 2007 Wiley-Liss, Inc.

**Key words:** ODA method; metalloprotease; protein–protein association; peptidase MC clan; peptidase M14 family; low-energy docking area; protein–ligand recognition

### INTRODUCTION

Metalloproteases (MCPs) are enzymes that hydrolyze peptide bonds for a broad range of substrates and may act in a variety of roles, from the general digestion of peptides released from food intake to the regulation of cellular metabolism or the processing of prohormones and neuroproteins. The MCPs are widely distributed throughout almost all kind of organisms with the exception of viruses. Within mammals, there are more than twenty members of the clan MC (according to definition in MEROPS database; <http://merops.sanger.ac.uk>) in the human genome, several of which have been deeply characterized structurally and functionally.<sup>1</sup> The known three-dimensional structures in the clan belong to members of the M14 family, of which human carboxypeptidase A1 (CPA1) is the prototypical form. Although at least one MCP is

Grant sponsor: European Union Programme of High Level Scholarships for Latin America (Alfian Programme); Grant number: E04D035150AR; Grant sponsor: Marie Curie European Reintegration Grant; Grant sponsor: Ministerio de Educación y Ciencia, Spain; Grant numbers: BIO2004-05879, GEN2003-20642-C09-05.

\*Correspondence to: Juan Fernández-Recio, Institute for Research in Biomedicine, Parc Científic de Barcelona, C/ Josep Samitier 1-5, 08028 Barcelona, Spain. E-mail: [juan@mmb.pcb.uab.es](mailto:juan@mmb.pcb.uab.es)

Received 6 September 2006; Revised 17 November 2006; Accepted 12 December 2006

Published online 3 April 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21390

present in all tissues or fluids that have been examined, most MCPs have fairly restricted patterns of tissue distribution, reflecting a specific physiological function.<sup>1</sup> Their involvement in many deregulations or pathological processes also prompted their investigation and use in biomedical strategies.

Most of the reported three-dimensional (3D) structures of MCPs correspond to the immature pro forms (zymogens) or to complexes of the mature enzyme with protein inhibitors because this is a situation that both provides a detailed functional information and gives rise to more stable or readily foldable forms, and also avoids the instability of enzyme–substrate complexes as a consequence of the hydrolytic activity of the enzymes. The MCP catalytic moiety (the catalytic domain) is around 300 residues long and is folded in a central mixed parallel/antiparallel eight-strand  $\beta$ -sheet over which eight  $\alpha$ -helices are packed. Depending of the subfamily, the catalytic domain displays extensions either at the N-terminus (a 90–95 residues pro domain) or at the C-terminus (the C-terminal domain), which frequently consists of an 80 residues long transthyretin-like subdomain and additional extensions. The N-terminal pro domain (also known as activation segment) is an independent folding unit with no disulfide bridges that blocks the entrance to the active site cleft.<sup>2</sup> In some instances, it has been found that the N-terminal pro domain behaves as an inhibitor of the catalytic domain once severed from the enzyme.<sup>1</sup> The C-terminal domain, which folds in a manner reminiscent of transthyretin, is involved in the anchoring of the enzyme moiety to the plasma membrane and in intracellular routing.<sup>3</sup>

A number of proteinaceous inhibitors of MCPs have been identified in different species, and several of them characterized in depth.<sup>1</sup> In plant species, they constitute a defense system against herbivores, as exemplified by the potato carboxypeptidase inhibitor (PCI), a small polypeptide for which the expression levels in the plant have been found to depend on external signals, such as leaf wounds caused by phytophagous insects.<sup>4</sup> The pioneering work on the 3D structure of the complex of bovine CPA1 with PCI derived by Rees and Lipscomb revealed in detail the substrate-like binding mode of this protein inhibitor.<sup>5</sup> Other protein inhibitors of MCPs characterized at the structural level are found in hematophagous animals, where they presumably play a role in keeping the sucked blood in the appropriate state for its digestion.<sup>6,7</sup> Human latexin, also called Endogenous Carboxypeptidase Inhibitor (ECI), is the first characterized member of a putative family of endogenous mammalian MCP inhibitors, and consists of a single polypeptide chain of 25 kDa that is expressed at high levels in heart, prostate, ovary, kidney, pancreas, and colon. Also in this case, the 3D structure of the complex of human CPA4 and human latexin (ECI) has unveiled the mode of action of latexin, closer to the binding of the catalytic domain and the N-terminal pro domain, and which has distinctive features with respect to other known protein inhibitors.<sup>8</sup>

Given that an important number of detailed crystal structures of MCPs, either unbound or mostly bound to

N-terminal pro domains or proteinaceous inhibitors, is available, we chose them to investigate the potential protein binding sites on the surface of the MCPs by means of the Optimal Docking Area (ODA) method. The method identifies continuous surface patches with optimal docking desolvation energy based on atomic solvation parameters adjusted for protein–protein docking. It has recently been benchmarked on a set of nonhomologous proteins involved in nonobligate protein–protein interactions.<sup>9</sup> The application of this method to the M14 family (MCPs) of structurally and functionally related enzymes should give us a more profound knowledge of the energetics and mechanistic features by which these proteins are able to recognize and interact specifically with different partner molecules. The MCPs are considered biological targets for the development of therapeutic alternatives in the treatment of several human diseases. Examples include carboxypeptidase U, also called plasma carboxypeptidase B (CPB) or thrombin-activatable fibrinolysis inhibitor (TAFI)<sup>10–13</sup> and human carboxypeptidase A4 (CPA4).<sup>14</sup> Therefore, the knowledge of the identity of potential MCP ligands has further biomedical applications.<sup>15</sup> In the same context, this information should encounter application in the discovery of newer inhibitory agents to combat insect pests of agronomically important crop plants.<sup>16,17</sup>

## METHODS

### Metallo-carboxypeptidase Dataset

Currently, there are more than 50 crystal structures of MCPs deposited in the Brookhaven Protein Data Bank.<sup>18</sup> We retrieved the PDB entries corresponding to known MCP protein–protein complexes solved by X-ray diffraction, disregarding the structures of those in complex with small-molecule ligands. The metallo-carboxypeptidase dataset was thus created from 16 structures of protein–protein complexes and also included two unbound structures for validation purposes. Table I lists the PDB codes, as well as the structural quality assurance and completeness of the dataset.

The structural quality assessment was performed carefully in all members of the dataset. Two structures were disregarded because either the unbound or bound structure was incomplete (PDB codes: 1AYE and 1PCA). The side-chain atoms missing in the structures 1DTD, 1ZLH, 1ZLI, and 2C1C were added using the program SCWRL Version 3.0.<sup>19</sup> Only two structures that are lacking main-chain atoms were used because the missing bonds are in a solvent-exposed loop (residues 134–135 in 1ZLH), or in a partially solved C-terminal segment (residues 382–402 in 1UWY), far from any known binding site. All the heteroatoms, water molecules, ions, degraded peptide fragments, and sugar moieties were removed from the original PDB file to create the final MCP dataset. This comprises one separate coordinate file that contains the computational analysis below described for each of the following: (a) the catalytic domain of the MCP, (b) the pro domain located N-terminal, (c) the transthyretin-like domain located C-terminal and, (d) the protein inhibitor

TABLE I. The 3D Structures of MCPs Studied in This Work

Catalytic domain <sup>a</sup>	PDB code	Resolution (Å) <sup>b</sup>	B-factor (Å <sup>2</sup> ) <sup>c</sup>	Binding partner <sup>d</sup>	Source organism	Production method <sup>e</sup>
hCPA2	1DTD	1.65	23.6	lci	<i>Homo sapiens</i>	<i>Pichia pastoris</i>
hCPA4	2BOA	2.2	35.0	pro	<i>Homo sapiens</i>	<i>Pichia pastoris</i>
	2BO9	1.6	16.2	eci	<i>Homo sapiens</i>	<i>Pichia pastoris</i>
bCPA1	1PYT	2.57	40.2	pro	<i>Bos taurus</i>	Bovine pancreas
	4CPA	2.5	9.6 <sup>f</sup>	pci	<i>Bos taurus</i>	Bovine pancreas*
	1ZLH	1.7	14.1	tci	<i>Bos taurus</i>	Bovine pancreas*
	1M4L	1.25	17.1	—	<i>Bos taurus</i>	Bovine pancreas*
iCPA	1JQG	2.5	29.2	pro	<i>Helicoverpa armigera</i>	<i>Pichia pastoris</i>
hCPB	1KWM	1.6	20.5	pro	<i>Homo sapiens</i>	<i>Pichia pastoris</i>
	1ZLI	2.09	22.1	tci	<i>Homo sapiens</i>	<i>Pichia pastoris</i>
pCPB	1NSA	2.3	23.0	pro	<i>Sus scrofa</i>	Porcine pancreas
	1Z5R	1.4	14.9	—	<i>Sus scrofa</i>	Porcine pancreas*
iCPB	2C1C	2.3	19.2	—	<i>Helicoverpa zea</i>	<i>Pichia pastoris</i>
dCPD	1QMU	2.7	55.4	ctd	<i>Anas platyrhynchos</i>	<i>Pichia pastoris</i>
hCPM	1UWY	3.0	53.8	ctd	<i>Homo sapiens</i>	<i>Spodoptera frugiperda</i>
mCPT	1OBR	2.3	12.6 <sup>f</sup>	—	<i>Thermoactinomyces vulgaris</i>	<i>Thermoactinomyces vulgaris</i>

<sup>a</sup>h, human; b, bovine; p, porcine; i, insect; d, duck; m, microbial.

<sup>b</sup>Crystallographic resolution.

<sup>c</sup>Average B-factor of the protein atoms.

<sup>d</sup>pro, pro domain; lci, leech carboxypeptidase inhibitor; eci, human endogenous carboxypeptidase inhibitor; pci, potato carboxypeptidase inhibitor; tci, tick carboxypeptidase inhibitor; ctd, C-terminal domain.

<sup>e</sup>Source of MCP used for crystallization. An asterisk (\*) marks if the protein was obtained from a commercial supplier.

<sup>f</sup>In these cases, the average B-factor value is not comparable with others in the Table because of a particular recording method (e.g. film).

chain. In the case of molecules with more than one X-ray structure independently determined, the computational results resulted to be very similar irrespective of the specific molecule, and only one of them was chosen for further analysis. Overall, this represents 28 unique polypeptide chains, 24 of which have their protein–protein complex structure available.

### Calculation of ODA Patches

We computed the ODA method in search for areas with favorable energy change when buried in a protein–protein association. The ODA method identifies continuous surface patches with optimal desolvation energy based on atomic solvation parameters adjusted for protein–protein docking. We followed the published procedure<sup>9</sup> with some modifications. First, the center of coordinates of each residue side-chain was used to define the starting points around the protein, instead of using the systematically distributed surface points generated in the original method. Then the docking surface energy of the different-sized patches was calculated at each point. Thus, the docking surface energy of the patch is dependent on the atomic solvation energy contribution of the residues that lie within a sphere of radius  $d$  ( $d = 1, 2, \dots, 20$  Å) from the starting point. Among all surface patches generated from each starting point, the patch with lowest energy value is called an ODA, and its optimal energy value is assigned to that point, and therefore to the residue from which that point was generated. ODAs were considered only for residues at the surface of the protein, that is, when the relative solvent-accessible surface area (SASA) of the residue is greater than 0.25 (value obtained in previous

internal tests in order to disregard poorly exposed surface residues, which do not significantly contribute to the desolvation effects). Zn atoms were not considered in the calculations since they were not solvent-exposed (relative SASA less than 0.25). Finally, the ODA energy values per residue were used to define a region over the protein surface most likely to be involved in protein–protein binding. An ODA was considered significant when its optimal docking energy value was  $< -10.0$  kcal/mol. Significant ODAs can be graphically displayed on the surface of the proteins and can be used to analyze if they are conserved in the family.

The catalytic domain is central to all MCPs, and thus the ligand molecule in the complex is referred to as the binding or interacting partner. The unbound structures, namely 1M4L, 1Z5R, 2C1C, and 1OBR, were used to build models of complexes with different binding partners. This procedure involved the structural superimposition of the main-chain atoms of the unbound to the bound molecule using the command “superimpose align” of the ICM-Browser (www.molsoft.com). In the cases of the structures of the bovine CPA1 (1M4L) and the porcine CPB (1Z5R), it served the purpose of checking the accuracy of our predictions against the available structure of the complexes 1PYT and 1NSA, respectively. On the contrary, there is no structural information available on complexes of mCPT (1OBR) and iCPB (2C1C), and we shall thus evaluate these predictions with caution.

The positive predictive value and the global success rate were the two parameters used to assess the performance of the ODA method. For a given unbound molecule, the positive predictive value was defined as the percentage of ODA hot spots (i.e. formed by up to a maximum of

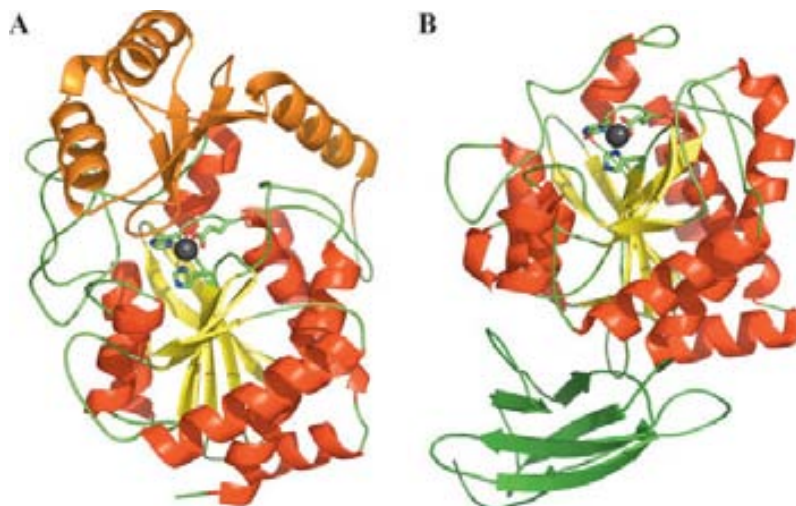


Fig. 1. View of the 3D structures of representative MCPs. **A:** A member of the M14A subfamily, bovine CPA1 (PDB code 1PYT) with the N-terminal pro domain shown in orange. **B:** A member of the M14B subfamily, human CPM (PDB code 1UWY) with the C-terminal domain depicted in green. The catalytic domain is in the same orientation in both structures. The HEH triad of residues coordinating the catalytic zinc ion (gray sphere) is shown as a stick model. The figure was made with the program PyMol ([www.pymol.org](http://www.pymol.org)).

ten residues with ODA energy  $< -10$  kcal/mol) that were correctly located in the real protein–protein interface (i.e. that at least one atom of the given residue is located at a distance of  $< 5$  Å from any heavy atom of the interacting partner after superimposing the structure of the unbound protein onto the equivalent molecule in the structure of the protein–protein complex). The success rate is the percentage of cases where the positive predictive value is  $\geq 50\%$ , meaning that at least half of the ODA residues are located at the real interface.

## RESULTS AND DISCUSSION

### Structures of the MCPs and Their Complexes

We shall briefly review the structural characteristics of the MCPs, with special focus on the known protein–protein interactions. According to the MEROPS classification scheme,<sup>20</sup> MCPs belong to clan MC of peptidases where they are grouped within the subfamilies M14A, M14B, and M14C, based on overall domain structure and sequence similarities. Structural knowledge is available for members of the M14A and M14B subfamilies, which appear to be the best characterized ones. The function of the MCPs in the M14A subfamily, also known as A/B-type carboxypeptidases, is typically concerned with the processing of the peptides from the diet (and hence their common names of “pancreatic” or “digestive” carboxypeptidases), but some members carry out important physiological activities such as that of plasma proCPB (or TAFI), which is involved in the fibrinolytic processes. The MCPs in the M14B subfamily, known as N/E-type carboxypeptidases, are typically involved in the maturation process of

prohormones and neuropeptides and were thus originally termed “regulatory” carboxypeptidases.

The catalytic domain common to all MCPs displays the  $\alpha/\beta$  hydrolase fold shown in Figure 1, where the active center is located in a cavity that contains the catalytic  $Zn^{2+}$  ion. Residues involved in the catalytic machinery and substrate binding include His69, Glu72, and His196 (for coordination of the zinc ion); Arg127 and Glu270 (directly involved in the catalysis); as well as Asn144, Arg145, and Tyr248, which anchor and neutralize the C-terminal residue of the substrate. The specificity among MCPs is significantly dependent on the residue at position 255 (bovine CPA1 numbering).

The MCPs of the M14A subfamily are synthesized as proenzymes with an N-terminal noncatalytic, 90–95 residues, pro domain that behaves as an independent folding globular unit that serves the purposes of (i) assisting the correct folding of the catalytic domain and (ii) maintaining the inactive state of the protease by blocking the entrance to the active site. The N-terminal pro domain in its isolated state is a potent inhibitor of some MCPs, and particularly in the case of CPA1, it is able to rejoin to the catalytic domain once it is trimmed off from it. On the contrary, no inhibitory effect of the N-terminal pro domain against the cognate human CPA2 or the porcine CPB could be found.<sup>21</sup> In the M14B subfamily, there is an extension of the 80-residue long C-terminal domain, which in certain cases anchors the catalytic domain to the plasma membrane. The positioning of the catalytic domain onto the membrane is mediated by positively charged side chains of the C-terminal domain and the GPI anchor post-translationally attached to it.<sup>22</sup>

TABLE II. Analysis of Protein-Protein Interactions in the Catalytic Domain of MCPs

	Catalytic domain <sup>a</sup>	Chain length <sup>b</sup>	ODA residues <sup>c</sup>	Energy (kcal/mol) <sup>d</sup>	Binding partner <sup>e</sup>	Interface residues <sup>f</sup>	ODA positive predictive value (%) <sup>g</sup>	Rank of ODA at interface <sup>h</sup>
1	hCPA2	303	10	-19.31	pro*	22	80	1
					lci	14	30	1
2	hCPA4	310	8	-17.12	pro	25	88	1
					eci	19	60	2
3	bCPA1	309	21	-18.69	pro	27	50	2
		307	17	-18.21	pci	10	20	1
		301	22	-23.58	tci	15	60	1
		307	16	-17.98	pro*	29	80	1
					pci*	12	20	4
tci*	17	70	2					
4	iCPA	317	54	-51.32	pro	25	90	1
5	hCPB	307	12	-16.75	pro	25	80	1
		306	23	-28.48	tci	19	90	1
6	pCPB	306	17	-16.66	pro	23	80	1
		304	12	-19.46	pro*	23	50	1
7	iCPB	312	38	-30.33	pro*	23	90	1
					tci*	20	70	1
					ctd	16	10	7
8	dCPD	300	21	-30.42	ctd	16	10	7
9	hCPM	295	28	-30.96	ctd	20	10	1
10	mCPT	323	18	-18.57	pro*	27	80	1
					pci*	9	40	1

<sup>a</sup>h, human; b, bovine; p, porcine; i, insect; d, duck; m, microbial.

<sup>b</sup>Number of residues contributing to ODA calculation.

<sup>c</sup>Number of significant ODAs.

<sup>d</sup>Energy value of the lowest-energy ODA residue.

<sup>e</sup>pro, pro domain; lci, leech carboxypeptidase inhibitor; eci, human endogenous carboxypeptidase inhibitor; pci, potato carboxypeptidase inhibitor; tci, tick carboxypeptidase inhibitor; ctd, C-terminal domain. An asterisk (\*) denotes that the analysis was performed on a modeled complex formed by structural superimposition of the unbound onto the bound molecule.

<sup>f</sup>Number of residues of the catalytic domain that are at the interface.

<sup>g</sup>Percentage of the ODA residues (among the significant top 10) of the catalytic domain that are at the interface.

<sup>h</sup>Rank (according to ODA energy) of the lowest-energy ODA residue located at interface.

The protein inhibitors of the MCPs are small to medium range polypeptides, from 39 to 222 residues, found in a number of tissues and organisms that bind to the enzymes with nano- and subnanomolar affinities.<sup>1</sup> The protein inhibitors of the MCPs have a compact globular fold, generally sustained by disulphide bridges. PCI is a small polypeptide (39 residues) with a five-residue long C-terminal tail that makes interactions with the primary binding site of the catalytic domain and also coordinates to the zinc ion in a substrate-like manner.<sup>5</sup> Similar characteristics are found in other small polypeptides, such as the leech carboxypeptidase inhibitor (LCI, 66 residues)<sup>6</sup> and the tick carboxypeptidase inhibitor (TCI, 76 residues).<sup>7</sup> Human latexin, or ECI, is a 222-residue protein sequentially and structurally unrelated to the latter forms and is devoid of disulphide bridges.<sup>8</sup> It behaves differently because none of its residues is perturbing the Zn coordination of the catalytic domain nor having primary binding site interactions. All known protein inhibitors establish additional interactions at the interface between the inhibitor and the catalytic domain surfaces. The inhibitors identified until now are specific and potent towards the MCPs of the M14A subfamily but are ineffective against MCPs in the M14B subfamily.

### Experimental Interfaces in Bound and Unbound MCPs

Protein interactions were analyzed for the available structures of bound and unbound MCPs. The analysis of the protein interactions was performed using the program ICM-Browser, defining the real interface as all residues in one molecule that have at least one atom within a distance of 5.0 Å from any nonhydrogen atom of the interacting partner. We have considered here domain-domain interactions, that is, between the catalytic domain and its N- or C-terminal domain, as well as protein-protein interactions, i.e. in a complex with a heterologous protein.

As can be seen from Table II, the experimental interactions of the catalytic domain involve 22 to 29 residues with the N-terminal pro domain (entries 1-7, and 10; column 7) and slightly fewer residues with the C-terminal domain (18 on average for entries 8, 9). In contrast, the number of residues found at the interface with the protein inhibitors is quite variable and ranges from 9 to 20. The results obtained for the modeled protein-protein complexes formed upon structural superimposition of the unbound to the bound catalytic domain are in good agreement with respect to the real complexes. This can be seen,

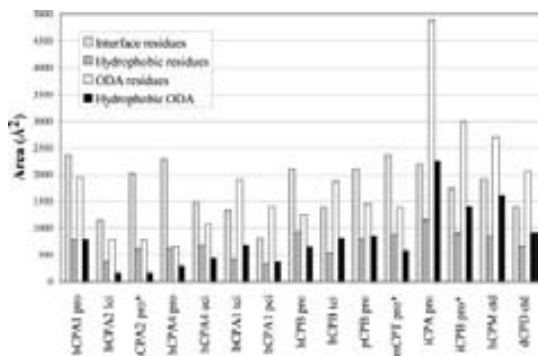


Fig. 2. Surface area of the residues at the predicted ODA patches and at the experimental protein–protein interfaces. The values shown correspond to the total set of predicted ODA residues (open bars) and to the subset of hydrophobic ODA residues (filled bars). Also shown are the values corresponding to the residues at the experimental interfaces (light gray bars) and to the subset of hydrophobic residues at the experimental interfaces (dark gray bars). The complexes of human CPA2, microbial CPT, and insect CPB modeled upon structural superimposition are marked with an asterisk (\*).

for instance, by comparing the data of the real complexes of bCPA1 (entry 3 in Table II) to the modeled ones obtained by structural superimposition of the unbound to the bound catalytic domain (marked by an asterisk). Besides, the values obtained for the modeled complexes of iCPB (entry 7) and mCPT (entry 10) are comparable with those obtained for the real complexes.

The surface area formed by the residues at the known experimental interfaces ranges from 800 Å<sup>2</sup> (in the complexes of the catalytic domain with the protein inhibitors) to almost 2500 Å<sup>2</sup> (in the interaction between the catalytic domain and the N-terminal pro domain) (Fig. 2). The average percentage of hydrophobic residues at the known experimental interfaces is close to 40% (ranging between 33% and 53%), which correlates well with the average of 42% of the total surface area that is hydrophobic in nature. Except for the insect MCP iCPA, there are fewer hydrophobic residues at the interface with the binding partner for CPA1, CPA2, and CPA4 (close to 35%) than for hCPB, pCPB, dCPD, and mCPT (between 45–50%). Notably, the interface in the insect CPA has an above average hydrophobic content (53%). A similar trend is observed for the insect iCPB (52%) in the complex modeled by structural superimposition.

### Analysis of ODA Residues in MCPs

The previously described set of structures of MCPs was analyzed through the ODA method in search of putative protein-binding areas. We evaluated whether the predicted protein interaction sites defined by the significant low-energy ODA residues (see Methods) were located in the experimentally determined protein–protein interfa-

ces. Since in some cases more than 10 ODA residues with  $< -10$  kcal/mol were identified, and they defined surface areas much larger than the average protein–protein interfaces, it was decided to use the 10 lowest-energy ODAs for the predictions.

The ODA positive predictive value for the ODA patches in the central catalytic domain (Table II), although being generally good among MCPs, is quite variable and seems to be dependent on the kind of binding partner that is present in the complex. In 10 out of 10 cases (100% success rate), including the unbound structures for which the complex structure had to be modeled, the ODA residues were correctly located (positive predictive value  $\geq 50\%$ ) at the interface with the N-terminal pro domain. Similarly, in four out of four cases (100% success rate), the ODA residues correctly predicted the interface with tick carboxypeptidase (TCI) inhibitor. However, in none of the cases did the ODA residues fully predict the location of the interface with the LCI, ECI, or PCIs. Interestingly, in 16 of 21 cases, the rank 1 ODA residue (according to ODA energies) was at the interface with either the N-terminal pro domain or any inhibitor (including those inhibitors for which the overall ODA positive prediction values were below the 50% cutoff). Moreover, in 19 of 21 cases (90% success rate), an ODA residue with either rank 1 or 2 (according to energies) is located at the experimental interface. This indicates that the lowest-energy ODAs were indeed located at a surface region that might be important not only for binding to the N-terminal pro domain, but also to any of the inhibitors, in spite of the different atomic interactions that every type of inhibitor may be involved in. In the majority of the M14A cases (excluding insect and bacteria MCPs), the residue with lowest-energy ODA is either aspartate or glutamate at position 163 (bovine CPA1 numbering). As it is discussed later, this position is quite conserved in the M14A subfamily, and is located in a loop that is one of the easiest points of cleavage by endoproteases.

When all the predictions (in catalytic domains and in their partners) are compared with only the experimentally determined interfaces (i.e. no complex models included), in a total of 14 over 24 cases, the predicted ODA patches (see the “Overall catalytic domain” and “Overall partner” entries in the Table III) were successfully located (i.e., more than 50% of the predicted residues were at the interface), which yields an overall success rate of 58% for the analyzed set of structures. The success rates of the ODA predictions for the catalytic domain were in general high, except in the case of its interactions with the C-terminal domain. Indeed, the ODA predictions in the catalytic domain of the members of the M14A subfamily (entries M14A<sup>h</sup> and M14A<sup>i</sup> in Table III) were correctly located in 8 out of 10 cases (i.e. success rate of 80%). In spite of being sequentially and structurally unrelated, the success rate is similarly high for the MCP protein inhibitors (there are no significant ODA residues in the TCI structure, which is present in two complexes). On the contrary, the predictions for the N-terminal pro domains were poor; only in one case (the pro domain of



TABLE III. Summary of ODA Analysis for the MCPs<sup>a</sup>

MCP subfamily	ODA range <sup>b</sup>	Energy (kcal/mol) <sup>c</sup>	Success rate (%) <sup>d</sup>	Type of binding partner	ODA range <sup>e</sup>	Energy (kcal/mol) <sup>f</sup>	Success rate <sup>g</sup>
M14A <sup>h</sup>	8–54	–51.32	100 (5/5)	Pro domain	1–24	–24.79	20 (1/5)
M14A <sup>i</sup>	10–23	–28.48	60 (3/5)	Inhibitor	0–25	–38.43	60 (3/5)
M14B	21/28	–30.96	0 (0/2)	C-terminal domain	8/15	–25.41	100 (2/2)
Overall catalytic domain	8–54	–51.32	67 (8/12)	Overall binding partner	0–25	–38.43	50 (6/12)

<sup>a</sup>Complexes formed by structural superimposition are not summed up here.

<sup>b</sup>Maximum and minimum number of ODA residues of the catalytic domain.

<sup>c</sup>Energy value of the first ranked ODA residue of the catalytic domain.

<sup>d</sup>Percentage of cases where the ODA positive predictive value for predicting the interface is above the cutoff of 50%. Values in brackets indicate the number of cases above the minimum and the total number of cases.

<sup>e</sup>Maximum and minimum number of ODA residues in the binding partner molecule.

<sup>f</sup>Energy value of the first ranked ODA residue of the binding partner molecule.

<sup>g</sup>Ratio of cases where the ODA positive predictive value is above the 50% level over the total number of binding partner molecules. Values in brackets indicate the number of cases above the minimum and the total number of cases.

<sup>h</sup>MCPs of the M14A subfamily where the pro domain structure is present.

<sup>i</sup>MCPs of the M14A subfamily that are in complex with inhibitor.

the porcine CPB) the ODA positive predictive value is above the 50% minimum level.

The overall average number of ODA residues in a catalytic domain is 16. The maximum ODA patch area is found in the insect MCP iCPA, covering almost 5000 Å<sup>2</sup> (Fig. 2) and involving 54 ODA residues, where 44% of them are hydrophobic. This is followed by the other insect MCP iCPB (with an area of 3000 Å<sup>2</sup> and 38 ODA residues, where 50% of them are hydrophobic). For the catalytic domain of the complete M14A subfamily (which comprises MCPs where the N-terminal pro domain structure is present and also those that are in complex with an inhibitor), there are between 8 and 23 ODA residues (50% and 43% of them hydrophobic, respectively) at the surface of the protein, forming an ODA patch area that extends from 800 to 2000 Å<sup>2</sup> (Fig. 2). The MCPs of the M14B subfamily show intermediate values, with 21 and 28 ODA residues (52% and 54% of them hydrophobic, respectively) that display ODA patch areas of 2100 and 2700 Å<sup>2</sup> for the duck CPD and the human CPM, respectively. Thus, there seems to be an increasing trend in the number of ODA residues/covered surface area from CPA/Bs < CPN/Es < insect MCPs.

Previous tests indicate that the ODA method is not sensitive to small conformational changes in the binding interfaces. For instance, the ODA results do not significantly vary between the unbound and the complexed states, irrespective of the conformational changes in the interface side-chains.<sup>9</sup> In addition, as described in Methods, when several structures of the same protein were available, the ODA results were practically the same, in spite of the existence of small side-chain conformational changes. Finally, we have built alternative models modifying the side-chain conformations with SCWRL, and the ODA results do not vary when only a few side-chains are modified. The only exception is that, in some cases, a protein structure solved by NMR can have significant conformational changes as compared with the same protein structure solved by X-ray crystallography, which may

artificially modify the ODA results,<sup>9</sup> but it is not the case in this study.

### Conserved ODA Segments in the MCPs

The alignment of the sequences of the MCPs used in this study found very few residues fully conserved along all MCPs, and most of them were buried from solvent and therefore not relevant for ODA analysis. Actually, only Tyr248 (bovine CPA1 numbering) is fully conserved at an interface in all MCPs, and happens to be predicted by ODA. Within the M14A subfamily, in addition to Tyr248, there are other residues quite conserved: position 163 can be either Glu or Asp, and is at the interface in all members (except in the bCPA/pro structure), and position 198 can be either Tyr or Phe and is at the interface in all members. All of these interface conserved residues in the M14A subfamily are predicted by ODA. However, although the conservation at the individual residue level is not high, our results show that the ODA residues in the catalytic domain are grouped in sequence segments that are quite conserved along the MCPs (Fig. 3). The ODA residues identified are involved in nearly all the protein–protein interactions established by the catalytic domain (>70% in the cases of the N-terminal pro domain, the C-terminal domain, and the protein inhibitors). We analyzed the intermolecular interactions established by each of the sequence segments of ODA residues. The first low-energy ODA segment (labeled S1 in Fig. 3) is at the N-terminus of the catalytic domain, and all interactions found here are with the N-terminal pro domain. Probably because of the smaller size of the protein inhibitors, none of their residues are interacting with this region. On the contrary, and in the rest of the ODA-based residue segments, the protein inhibitors seem to interact with similar residues as the N-terminal pro domains. The sequence segment extending for more than 10 residues after Pro205 contains a number of significant ODA residues that do not seem to be involved in known protein–protein interactions, with the exception of a short segment of

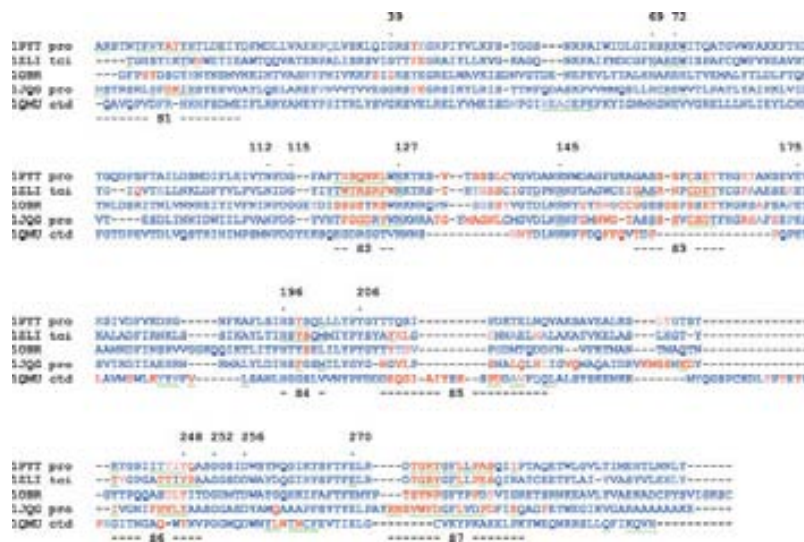


Fig. 3. Sequence alignment based on topological equivalence of the catalytic domain of selected representative MCPs of the dataset. The numbers label important residues in the structure. The different sequence segments derived from the significant ODA residues are labeled from S1 to S7. The residues underlined in green are located at the known experimental interface (i.e. they have at least one atom within 5 Å distance from any atom from the partner molecule). Residue segment number 5 (S5) is a putative binding region. The alignment was made with CLUSTALW ([www.ebi.ac.uk/clustalw](http://www.ebi.ac.uk/clustalw)) and was manually refined to account for experimental information (location of catalytic relevant residues, metal binding residues, and secondary structure elements). Sequence similarity values within A/B MCPs (first four entries) ranged between 24% and 46%, whereas the average sequence similarity between A/B MCPs and the only N/E MCP in the figure (1QM0) was 18%. The alignment has been colored with an in-house script according to ODA energy values from red (energy value  $< -10.0$  kcal/mol) to blue (energy value = 0.0).

residues that are at the interface with the C-terminal domain in the members of the M14B subfamily. Thus, the 209–221 residue segment (labeled S5 in Fig. 3) was defined as a “putative binding region”.

#### Regionalization and Functional Significance of the ODA Residues

The analysis of the conservation of the ODA regions at the sequence level indicates that they are formed by segments of residues interspersed throughout the whole catalytic domain. However, they generally appear surrounding the active site cleft in the tertiary structure (Fig. 4). What follows is a brief description of the main structural characteristics of the ODA-based predicted binding region

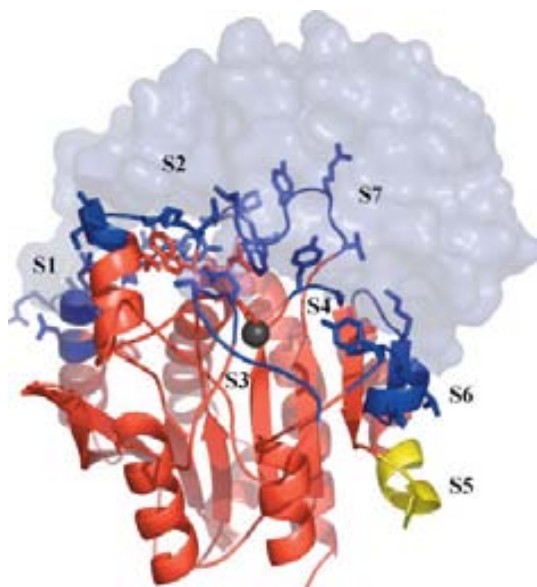


Fig. 4. View of the binding regions in a representative MCP, bovine CPA1 (1PYT in PDB). The structure of the CP domain is represented as a cartoon where the different residue segments, numbered according to their location in the sequence (see the text for a complete discussion) are shown in blue. The side-chains of the experimentally known interface residues (those with at least one atom within 5 Å distance from any atom of the partner molecule) are shown in blue stick. The ODA-based putative binding region (segment S5) is in yellow. The zinc ion (gray sphere) and the N-terminal pro domain (depicted as a surface) are also shown. The figure was prepared with the program PyMol ([www.pymol.org](http://www.pymol.org)).

(the segments of residues numbered S1–S4,S6,S7 in Fig. 4) that accounts for nearly all the known protein–protein interactions established by the catalytic domain, and of the features of the ODA-based putative binding region (segment S5 in Fig. 4). Residue segment S1 consists of the first 20 N-terminal residues of the catalytic domain, some of which establish noncovalent interactions to the N-terminal pro domain. Segment S2 contains residues 120–125, all of which are involved in protein–protein interactions either with the N-terminal pro domain or with the protein inhibitors. In the “reference” structure of bovine CPA1, this region spans the end of helix 4 and the first part of the following loop.

Segment S3, containing residues 154–164, is entirely a loop. It has been described that this loop is one of the easiest points of cleavage of the MCPs by endoproteases, a fact that might be related to both easy accessibility and regulatory processes. Thus, it has been reported that a specific cleavage of bovine CPA1 by subtilisin at the Ala154–Gly155 bond in this loop promoted catalytic and conformational changes to the catalytic domain.<sup>23</sup> Members of the M14A subfamily display different interactions with either the N-terminal pro domain or the protein inhibitor in this region. Remarkably, the residue at position 163 is an acidic one (aspartate/glutamate) in all MCPs, except for those in the M14B subfamily. In the M14A subfamily, this residue seems to change its conformation depending on the kind of binding partner molecule since it establishes a hydrogen bond to the LCI and TCI protein inhibitors (but not to PCI), although such interaction does not take place with the N-terminal pro domain.

Two residues in segment S4, Tyr198 and Ser199, are involved in four interactions with either the N-terminal pro domain or the protein inhibitors. Binding region 5 runs from residues 234 to 249 across the last part of helix 6, the loop connecting it with  $\beta$ -strand 7,  $\beta$ -strand 7, and a short segment in the next loop. The only strictly conserved residue is Tyr248, which makes interactions with residues of the N-terminal pro domain and the protein inhibitors. It is worth mentioning that the loop containing segment S6 is longer for the members of the M14B subfamily and might be a determinant for substrate or ligand selectivity. Residue segment S7, which includes residues 274–284, is found at the C-terminal part of the catalytic domain. It is essentially a loop containing residues that are to a greater extent involved in hydrophobic interactions with the binding partner molecules.

The residue segment S5 covers the loop between  $\beta$ -strand 6 and helix 6 of the catalytic domain. Neither the N-terminal pro domain nor the protein inhibitors seem to dock to this region. A superimposition with the MCPs of the M14B subfamily reveals that this region overlaps with a limited number of residues that are interacting with the C-terminal domain. However, since no C-terminal domain has been described for any MCP of the M14A subfamily, this ODA-based putative binding region may be involved in interactions with partners of different nature, yet to be discovered. Again, the loop is longer and contains more ODA residues in the members of the M14B

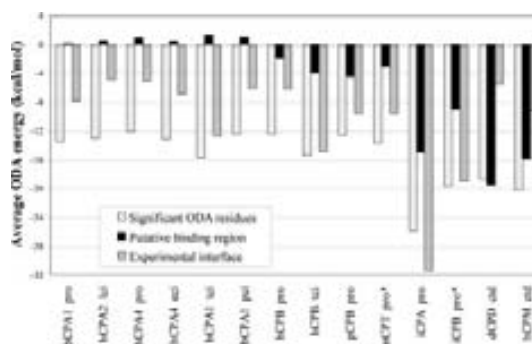


Fig. 5. Average ODA energy values for different groups of residues of the catalytic domain: significant ODA residues (i.e. those with ODA energy  $< -10$  kcal/mol), residues at the ODA-based putative binding regions, and residues at the experimental interfaces. The complexes of insect CPB and microbial CPT with a pro domain modeled upon structural superimposition are marked with an asterisk (\*).

subfamily. One might hypothesize that this binding region performs an as yet unknown general function. A potential role is related to its susceptibility to aggregation. It is known that the MCPs of the M14B subfamily might suffer an aggregation/solubilization transition along the sorting process at the secretory pathway.<sup>24</sup> Besides, binding of a monoclonal antibody recognizing the bovine CPA1 epitope located in the sequence 209–218 inside the loop was shown to protect the enzyme from aggregation.<sup>25,26</sup> Taken together, these observations could suggest that the putative binding region might play a functional and structural role in the folding process of the MCPs.

Previous molecular dynamics (MD) simulations of bCPA1<sup>27</sup> showed pronounced fluctuations in two regions: residues 120–126 (which is a region making interactions with extended oligopeptides) and residues 270–281 (which are lining the active site). Interestingly, these regions showing high mobility in MD calculations are two of our ODA-based predicted binding regions (segments S2 and S6). It would be interesting to include protein fluctuations in the ODA analysis, but this is ongoing work and will be the object of a separate large systematic study.

### Energetics and Binding Properties of the ODA Patches

The regions defined by the significant ODA residues, located either at the experimentally known interfaces or in putative binding regions, seem to contribute differently to the interaction with the corresponding partners, according to the particular type of MCP (Fig. 5). The average ODA energy of the significant ODA residues (i.e. those with ODA energy  $< -10$  kcal/mol) is around  $-12.0$  kcal/mol for all catalytic domains of the M14A subfamily (all the A and B MCPs) with the exception of the insect variants, whereas it shows slightly lower energy values (around  $-20$  kcal/mol) for the MCPs of the M14B subfamily (dCPD and hCPM species, here), and reaches a mini-

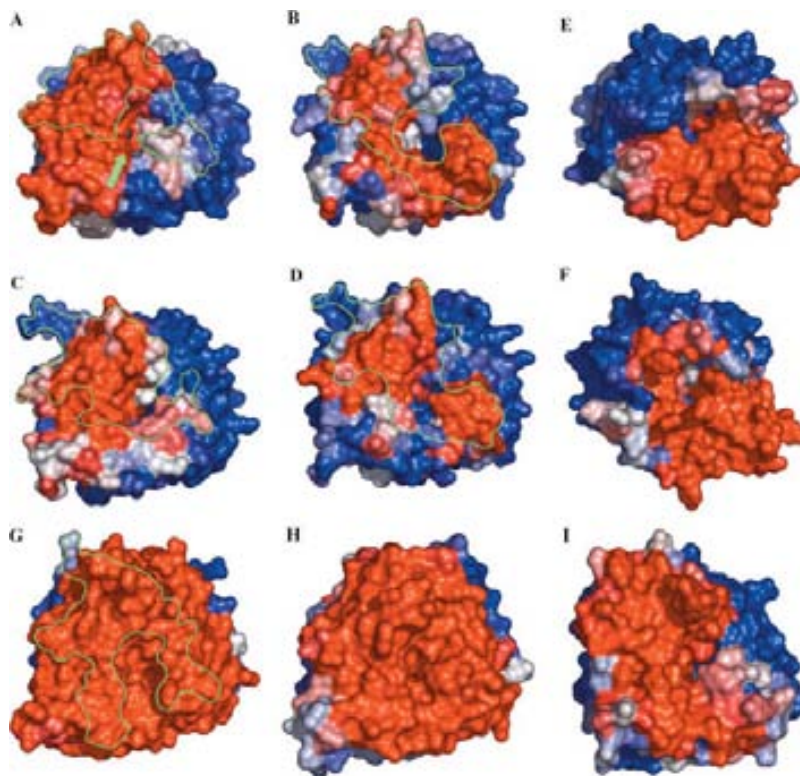


Fig. 6. Representation of the surfaces of the ODA patches for the MCPs. The view is with the active site cleft pointing towards the reader. In panel A, the zinc ion is shown in green at the bottom of the active site cleft, which is indicated by a green arrow. The structures have similar orientation and are grouped into four classes. Upper left side and center, mammalian M14A MCPs: (A) bCPA1, bovine CPA1, (B) pCPB, porcine CPB, (C) hCPA4, human CPA4, (D) hCPB, human CPB; upper right side, M14B MCPs: (E) dCPD, domain II of duck CPD, (F) hCPM, human CPM; lower left side, insect M14A MCPs: (G) iCPA, CPA from *H. armigera*, (H) iCPB, CPB from *H. zea*; lower right side, a bacterial M14A MCP: (I) mCPT, CPT from *T. vulgaris*. The color ramp is from red (ODA energy value  $< -10.0$  kcal/mol) to blue (ODA energy value = 0.0). The experimental interface is shown in green contour line for those cases where the structure of the complex is known (A-D,G). The figure was prepared with the program PyMol ([www.pymol.org](http://www.pymol.org)), and the color ramp was generated with an in-house script.

mum value of  $-26.0$  kcal/mol for the insect iCPA. These results emphasize that the average energy profile of the ODA patches is similar among the MCPs, in spite of the changes at the primary sequence level. In contrast, the average ODA energy of the residues forming the putative binding region (as previously defined) is above zero for CPA1, CPA2, and CPA4, and reaches values below zero for the rest of MCPs, again being much lower (larger in absolute value) for the two insect MCPs and for the members of the M14B subfamily.

The energy-based colored ODA residues for selected MCPs are shown in Figure 6. The most prominent feature among all the MCPs is the large ODA area displayed by the insect peptidases. Other differences are evident when comparing the pattern of the ODA patches between the MCPs of the M14A [Fig. 6(A-D,G-I)] and M14B [Fig. 6(E,F)] subfamilies. Firstly, the surroundings of the en-

trance to the active site cleft for all MCPs of vertebrate origin [Fig. 6(A-F)] appear to be covered by regions of ODAs of different energy levels. On the contrary, for the insect MCPs [Fig. 6(G,H)], this face of the enzyme is a homogeneous low-energy ODA patch. Secondly, when comparing the ODA patches between the MCPs of the M14A [Fig. 6(A-D)] and M14B [Fig. 6(E,F)] subfamilies (with the exception of the insect and bacterial variants), we found that for the MCPs of the M14A subfamily there are a number of regions of ODAs of different energy levels, whereas those on the M14B MCPs tend to be clustered. As the experimental evidence is against the inhibition of any of the MCPs of the M14B subfamily by the currently known natural protein inhibitors, and in favor that the MCPs in this subfamily contain no N-terminal pro domain comparable to the M14A ones, we might hypothesize that the regionalization of ODA patches is an important

determinant for the binding of a particular partner molecule. Together with the quantitative analysis in the first part of this section, it seems that the pattern of ODA patches distinguishes between the different subfamilies (e.g. in M14A the average value of ODA residues is  $-12$  kcal/mol, whereas in M14B the average value is  $-20$  kcal/mol; Fig. 5) and, within them, between the type and origin of the different forms (e.g. within M14A, differences in average ODA values are found between vertebrate, insect, and bacterial variants, in which the average ODA values are around  $-12$ ,  $-23$ , and  $-14$  kcal/mol, respectively). This might be useful for predicting, besides the area of the studied molecule that is likely to be involved in putative interactions with ligands, the type and surface of those yet unknown partner molecules.

#### Analysis of ODA Residues for the Catalytic Domain and its Binding Partners

We then analyzed the contribution of the ODA residues to the protein–protein interactions in the MCPs. The percentage of significant ODA residues that are located at the real interface (i.e. the positive predictive value) is a good indication of the involvement of the low-desolvation surface areas in the interaction. When we calculated this value for both the catalytic domain and the binding partners, we found that at least one, or in some instances both, of the interacting partners contribute to the binding energy with favorable desolvation. Besides, this effect may be related to the kind of binding partner (Fig. 7). In the subfamily M14A there are a large number of ODA residues belonging to the catalytic domain, but only a few of the top ten ODA residues of the N-terminal pro domain at the interface (with the sole exception of the pro domain of the porcine CPB). The reverse situation was observed for the MCPs of the M14B subfamily: only a few ODA residues belonging to the catalytic domain are at the interface with the C-terminal domain; however, the latter domain contributes with at least seven out of ten ODA residues to the known interface (see entries dCPD and hCPM in Fig. 7).

#### ODA Hot Spots: Correlation With Real Protein Interaction Sites

The method used for the identification of low-energy docking regions for protein binding is based on the selection of ODAs among a series of different-sized surface patches generated from a set of surface points around the protein. In practice, the method calculates the optimal size of the surface patch that would yield the best possible docking surface energy on different regions of the protein surface. We can thus identify which regions (if any) would have a favorable energy change when buried by a hypothetical binding partner. Previous results indicate that, for more than half of the proteins tested, a region can be found in which the binding of another protein would be highly favored.<sup>9</sup> For all the MCPs analyzed here, it was possible to find at least one ODA hot-spot (i.e. residues with ODA value  $< -10$  kcal/mol) on the protein surface.

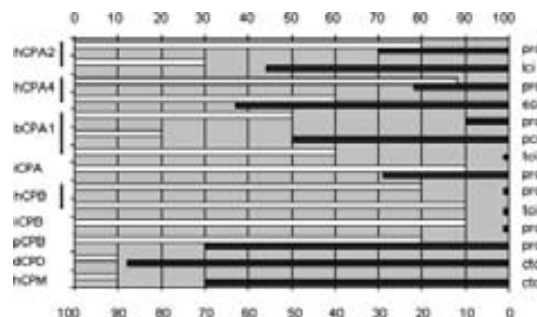


Fig. 7. Accuracy (positive predictive value) of ODAs in MCPs on prediction of protein–protein interactions. The white boxes, from left to right, show the ODA positive predictive values for the catalytic domain. The dark grey boxes, from right to left, indicate the ODA positive predictive value levels of the corresponding interacting partners. The complexes of human CPA2 and insect CPB with a pro domain modeled by structural superimposition are marked with an asterisk (\*).

These ODA hot spots defined a region on the protein surface that, remarkably, was located in or in the vicinity of the known binding site in most of the complexes. In some cases, the catalytic domain is known to be involved in interactions with different proteins using the same binding site, as for human CPB [Fig. 8(A)], bovine CPA1 [Fig. 8(B)], and human CPA4, where the predicted ODA points are correctly located in the interfaces with the N-terminal pro domain and a protein inhibitor. In other cases, the ODAs were correctly located (as compared with other MCPs), but only a limited number of ODAs are at the real interface [e.g. with the small-size protein inhibitor PCI; Fig. 8(C)] or they are found far from the known site of interaction as in the case of the MCPs of the M14B subfamily, where only a few ODAs are located at the interface with the C-terminal domain [Fig. 8(D)].

In some cases, an accurate prediction was not possible even though a conspicuous number of ODA residues were found on the protein surface. The complexes between the catalytic domain and the small-size protein inhibitors PCI and LCI are examples in which positive predictive values lower than 50% were obtained. This might be related to the shape or size of the ligand (for the same number of significant ODA residues, fewer of them could be statistically involved in an interaction with a smaller ligand as compared with that of a larger one). In these cases, the contribution to the binding energy provided by the coordination to the zinc ion may outweigh the other available electrostatic contributions. On the other hand, for the MCPs of the M14B subfamily, the predicted low-energy docking surface is close to the active site, with only a few ODAs located at the interface with the C-terminal domain, suggesting that the catalytic activity is energetically favored with respect to the mechanical function, that is, the attachment to the plasma membrane for certain members. In a broader generalization, a low ODA positive predictive value could just be an artifact of the

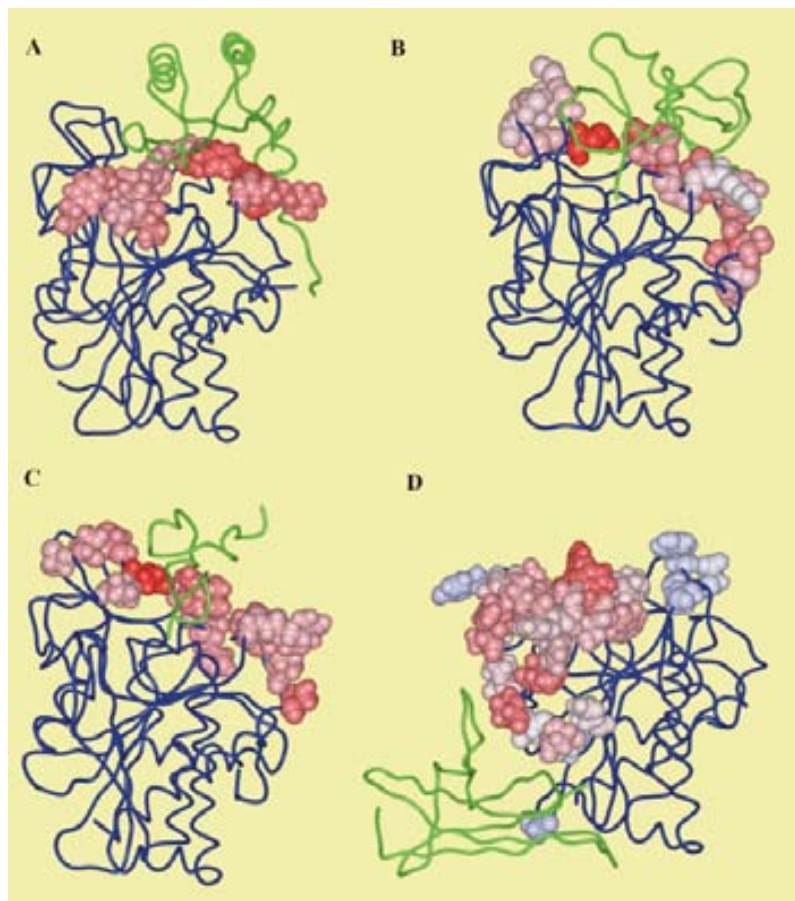


Fig. 8. View of ODA residues in the 3D structures of selected MCPs. The ODA residues are located in the vicinity of the active site cleft of the catalytic domain (shown in blue) in the 3D structure of the MCPs and are involved in protein–protein interactions with different partner molecules (green). The significant ODA residues are depicted as CPK models and colored from red (highest ODA energy value) to white (lowest ODA energy value). The structures are in a similar orientation and correspond to: (A) human CPB with the N-terminal pro domain (PDB code 1KWM), (B) bovine CPA1 in complex with TCI (PDB code 1ZLH), (C) bovine CPA1 in complex with PCI (PDB code 4CPA), and (D) human CPM with the C-terminal domain (PDB code 1UWY). The figure was made with the program ICM-Browser ([www.molsoft.com](http://www.molsoft.com)).

method, but we cannot disregard the existence of putative docking sites for potential ligands yet to be discovered.

#### The Dimensions of the ODA Hot Spots Among the MCPs

The surface area of the residues of the catalytic domain at the experimental interfaces is below  $2500 \text{ \AA}^2$  (Fig. 2). The average ODA patch area ( $1700 \text{ \AA}^2$ ) agrees well with the peak value of area distribution found in a survey of protein–protein complexes of known 3D structure.<sup>28</sup> In some instances, the area of the ODA patches parallels the surface area at the experimental interfaces, but there are

exceptions, as is the case of the ODA patches in human CPA2 and CPA4 where the surface areas are under  $1000 \text{ \AA}^2$ , near the lower limit found in the above mentioned survey. On the other hand, the areas of the ODA patches in the MCPs of insect origin are above the maximum value of the survey. One possible explanation for this feature is that such an enlarged area of the ODA patches may help the insect to “concentrate” the peptide and protein substrates from the diet in the midgut (depending on the availability of food or other factors). One form of control mechanism, different from the interaction with the N-terminal pro domain or a protein inhibitor, might also be envisaged. Since the ODA low-energy surface area

surrounds the active site cleft, a complex that may regulate the accessibility of the substrates to the protease could be built through the interaction of the catalytic domain with the membrane of the insect midgut (or with another protein receptor).

### CONCLUSIONS

We analyzed the protein–protein interactions among MCPs, integrating for the first time structural data and novel energy-based computational methods in a systematic way. The ODA method was found to be an accurate and fast algorithm for predicting the protein–protein binding sites, previously benchmarked on a set of three-dimensional structures corresponding to known nonobligate heterocomplexes, which can help to propose a new structural and functional classification of the members of the MC clan. For the central catalytic domain among the MCPs, the predictions were found to be correct in the 80% of cases. Poorer results were obtained for other kinds of partner molecules, notably the N-terminal pro domains. However, when we analyzed both interacting partner molecules, the ODA positive predictive values were high at least for one, and even in some cases, for both interacting members in the complex. This emphasizes the importance of the complementarity of the protein–protein interactions and is consistent with the view that the receptor and ligand molecules have attractive regions that can help them to approach each other, and perhaps drive the formation of an encounter complex.

When the ODA patches are mapped over the surface of some representative MCPs, we observed some important differences in energy and localization of ODAs that allow distinguishing between MCPs of the M14A and M14B subfamilies. This may be especially important for the known protein inhibitors, which are able to interact with the subset of ODA patches that belong to the interface between the catalytic domain and the N-terminal pro domain in the zymogen. The pattern of ODA patches of the MCPs of the M14A subfamily is different from the one in the MCPs of the M14B subfamily, and could explain why the known protein inhibitors are unable to block the activity of the latter. Exceptionally, the ODA patches in the insect MCPs display a surface area more than twice larger than the vertebrate counterparts. The functional and mechanistic relevance of this enlarged low-energy docking area is not yet known but, in addition to raising the possibility that the insect MCPs might possess additional features compared with those of vertebrate origin, this information could be exploited in the computer-assisted design of inhibitors directed against proteases and in the bioinformatics screening of novel peptide and protein partners able to specifically interfere with the digestive process of the crop pest insects. This is of particular interest for the digestive MCPs of two crop insect pests,<sup>17,29</sup> cotton bollworm and corn earworm, the latter being apparently unaffected by the known protein inhibitors from plants.<sup>29,30</sup>

### ACKNOWLEDGMENTS

We are grateful to Molsoft for the use of ICM-Browser.

### REFERENCES

- Vendrell J, Aviles FX, Fricker LD. Metallo-carboxypeptidases. In: Messerschmidt A, Bode W, Cygler M, editors. Handbook of metalloproteins, Vol. 3. Chichester: Wiley; 2004. pp 176–189.
- Coll M, Guasch A, Aviles FX, Huber R. Three-dimensional structure of porcine procarboxypeptidase B: a structural basis of its inactivity. *EMBO J* 1991;10:1–9.
- Varlamov O, Fricker LD. The C-terminal region of carboxypeptidase E involved in membrane binding is distinct from the region involved with intracellular routing. *J Biol Chem* 1996;271:6077–6083.
- Graham JS, Ryan CA. Accumulation of a metallo-carboxypeptidase inhibitor in leaves of wounded potato plants. *Biochem Biophys Res Commun* 1981;101:1164–1170.
- Rees DC, Lipscomb WN. Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J Mol Biol* 1982;160:475–498.
- Reverter D, Fernandez-Catalan C, Baumgartner R, Pfander R, Huber R, Bode W, Vendrell J, Holak TA, Aviles FX. Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nat Struct Biol* 2000;7:322–328.
- Arolas JL, Popowicz GM, Lorenzo J, Sommerhoff CP, Huber R, Aviles FX, Holak TA. The three-dimensional structures of tick carboxypeptidase inhibitor in complex with A/B carboxypeptidases reveal a novel double-headed binding mode. *J Mol Biol* 2005;350:489–498.
- Pallares I, Bonet R, Garcia-Castellanos R, Ventura S, Aviles FX, Vendrell J, Gomis-Ruth FX. Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin. *Proc Natl Acad Sci USA* 2005;102:3978–3983.
- Fernandez-Recio J, Totrov M, Skorodumov C, Abagyan R. Optimal docking area: a new method for predicting protein–protein interaction sites. *Proteins* 2005;58:134–143.
- Brouwers GJ, Leebeek FW, Tanck MW, Wouter Jukema J, Kluff C, de Maat MP. Association between thrombin-activatable fibrinolysis inhibitor (TAFI) and clinical outcome in patients with unstable angina pectoris. *Thromb Haemost* 2003;90:92–100.
- Montaner J, Ribo M, Monasterio J, Molina CA, Alvarez-Sabin J. Thrombin-activatable fibrinolysis inhibitor levels in the acute phase of ischemic stroke. *Stroke* 2003;34:1038–1040.
- Muto Y, Suzuki K, Sato E, Ishii H. Carboxypeptidase B inhibitors reduce tissue factor-induced renal microthrombi in rats. *Eur J Pharmacol* 2003;461:181–189.
- van Tilburg NH, Rosendaal FR, Bertina RM. Thrombin activatable fibrinolysis inhibitor and the risk for deep vein thrombosis. *Blood* 2000;95:2855–2859.
- Bentley L, Nakabayashi K, Monk D, Beechey C, Peters J, Birjandi Z, Khayat FE, Patel M, Preece MA, Stanier P, Scherer SW, Moore GE. The imprinted region on human chromosome 7q32 extends to the carboxypeptidase A gene cluster: an imprinted candidate for Silver-Russell syndrome. *J Med Genet* 2003;40:249–256.
- Arolas JL, Vendrell J, Aviles FX, Fricker LD. Metallo-carboxypeptidases: emerging drug targets in biomedicine. *Curr Pharm Des* 2007;15:347–364.
- Chilcutt CF, Wilson LT, Lascano RJ. Field evaluation of a *Helicoverpa zea* (Lepidoptera: Noctuidae) damage simulation model: effects of irrigation, *H. zea* density, and time of damage on cotton yield. *J Econ Entomol* 2003;96:1174–1183.
- Estebanez-Perpina E, Bayes A, Vendrell J, Jongsma MA, Bown DP, Gatehouse JA, Huber R, Bode W, Aviles FX, Reverter D. Crystal structure of a novel mid-gut procarboxypeptidase from the cotton pest *Helicoverpa armigera*. *J Mol Biol* 2001;313:629–638.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. *Nucleic Acids Res* 2000;28:235–242.

19. Canutescu AA, Shelenkov AA, Dunbrack RL, Jr. A graph-theory algorithm for rapid protein side-chain prediction. *Prot Sci* 2003;12:2001–2014.
20. Rawlings ND, Morton FR, Barrett AJ. MEROPS: the peptidase database. *Nucleic Acids Res* 2006;34:D270–D272 (database issue).
21. Segundo BS, Martínez MC, Vilanova M, Cuchillo CM, Aviles FX. The severed activation segment of porcine pancreatic pro-carboxypeptidase A is a powerful inhibitor of the active enzyme. Isolation and characterisation of the activation peptide. *Biochim Biophys Acta* 1982;707:74–80.
22. Reverter D, Maskos K, Tan F, Skidgel RA, Bode W. Crystal structure of human carboxypeptidase M, a membrane-bound enzyme that regulates peptide hormone activity. *J Mol Biol* 2004;338:257–269.
23. Solomon BM, Larsen KS, Riordan JF. Catalytic and conformational changes induced by limited subtilisin cleavage of bovine carboxypeptidase A. *Biochemistry* 1990;29:7303–7309.
24. Song L, Fricker LD. Calcium- and pH-dependent aggregation of carboxypeptidase E. *J Biol Chem* 1995;270:7963–7967.
25. Solomon B, Koppel R, Kenett D, Fleminger G. Localization of a highly immunogenic region of carboxypeptidase A recognized by three different monoclonal antibodies and their use in the detection of subtle conformational alterations in this enzyme region. *Biochemistry* 1989;28:1235–1241.
26. Katzav-Gozansky T, Hanan E, Solomon B. Effect of monoclonal antibodies in preventing carboxypeptidase A aggregation. *Biotechnol Appl Biochem* 1996;23(Part 3):227–230.
27. Makinen MW, Troyer JM, van der Werff H, Berendsen HJ, van Gunsteren WF. Dynamical structure of carboxypeptidase A. *J Mol Biol* 1989;207:201–216.
28. Janin J. Kinetics and thermodynamics of protein–protein interactions. In: Kleanthous C, editor. *Protein–protein recognition*, Vol. 31: *Frontiers in molecular biology*. New York: Oxford University Press; 2000. pp 1–32.
29. Bayes A, Comellas-Bigler M, Rodriguez de la Vega M, Maskos K, Bode W, Aviles FX, Jongsma MA, Beekwilder J, Vendrell J. Structural basis of the resistance of an insect carboxypeptidase to plant protease inhibitors. *Proc Natl Acad Sci USA* 2005;102:16602–16607.
30. Bayes A, de la Vega MR, Vendrell J, Aviles FX, Jongsma MA, Beekwilder J. Response of the digestive system of *Helicoverpa zea* to ingestion of potato carboxypeptidase inhibitor and characterization of an uninhibited carboxypeptidase B. *Insect Biochem Mol Biol* 2006;36:654–664.