Capítulo 3.

Estructura cristalina del dominio 2 de CPD acomplejado con un inhibidor y modelado de otras carboxipeptidasas reguladoras

3.1. Abstract

The three-dimensional crystal structure of duck carboxypeptidase D (CPD) domain II has been solved in a complex with the peptidomimetic inhibitor guanidinoethylmercaptosuccinic acid occupying the specificity pocket. This structure allows a clear definition of the substrate binding sites and the substrate funnel-like access. As it represents the first and only structure available from a member of the regulatory carboxypeptidase family, it has been used to model the structures of domains I and III from the same protein and of human carboxypeptidase E (CPE). The models obtained show that the overall topology is similar in all cases, the main differences being local and due to insertions in non-regular loops. Both in CPD domain I and CPE slightly different shapes of the access to the active site are predicted, implying some kind of structural selection of protein or peptide substrates. Furthermore, emplacement of the inhibitor structure in the active site of the constructed models showed that the inhibitor fits very well in all of them and that the relevant interactions observed with domain II are conserved in domain I and CPE but not in the non-active domain III, due to the absence of catalytically indispensable residues in the latter protein. However, in domain III some of the residues potentially involved in substrate binding are well preserved, together with others of unknown role which also are highly conserved among all carboxypeptidases. These observations, taken together, suggest that domain III might play a role in the binding and presentation of proteins or peptide substrates, as proposed for the pre-S domain of the large envelope protein of duck hepatitis B virus.

3.2. Introduction

Carboxypeptidases (CP) are enzymes that catalyse the cleavage of carboxy-terminal peptide bonds in proteins and peptides. From a mechanistic point of view, carboxypeptidases can be classified in two groups: metallocarboxypeptidases and serine carboxypeptidases. The metallocarboxypeptidases possess a Zn^{2+} cofactor in the active site. In mammals, this family currently contains 13 members subdivided into two subfamilies, the digestive enzymes and the regulatory enzymes (Fricker, 1988; Skidgel, 1988, 1996; Rawlings & Barret, 1995; Vendrell *et al*., 2000). While the biological function of the digestive CPs is to contribute to protein degradation, the regulatory ones are generally involved in physiological processes that require a higher specificity. Within each group, the members have 25-63% amino acid sequence identity, but it decreases to only 15-25% when comparison is performed between subfamilies. This low overall homology between subfamilies implies that they diverged early in time.

The digestive CPs are soluble, non-glycosylated proteins that are synthesized as inactive precursors containing a 90-95 amino acid N-terminal activation pro-segment (Aviles *et al*., 1993; Vendrell *et al*., 2000). The regulatory CPs have been purified and characterised from biological fluids and tissues, where they are found in soluble or in membrane-attached forms, in minor quantities. This subfamily includes CPD, CPE, CPM, CPN, CPZ, AEBP-1 and novel proteins with a unknown function designated as CPX-1 and CPX-2 (He *et al*., 1995; Song & Fricker, 1995, 1997; Skidgel, 1996; Xin *et al*., 1998; Lei *et al*., 1999). These enzymes perform a variety of important physiological functions, including neuropeptide and prohormone processing, regulation of peptide hormone activity and alteration of protein-protein or protein-cell interactions (Skidgel, 1988, 1996; Fricker, 1991).

Carboxypeptidase E (CPE), also known as enkephalin convertase or CPH (EC.3.4.17.10), is a CPB-like enzyme associated with the biosynthesis of many peptide neurotransmitters and hormones. It was purified for the first time from bovine brain (Fricker and Snyder, 1983). Later, cDNAs corresponding to CPEs from cattle (Fricker *et al*., 1986), rat (Fricker *et al*. 1989; Rodríguez *et al*., 1989), human (Manser *et al*., 1990), *Aplysia californica*

(Fan & Nagle, 1996), and the fish *Lophius americanus* (Roth *et al*., 1991) have been cloned and sequenced. The amino acid sequence homology among vertebrate species is greater than 80%. The molecular weight of CPE is 55 kDa and it is formed by 476 residues, of which 25 correspond to the signal peptide and 17 to the pro-segment. However, and in contrast with digestive CPs, scission of the pro-segment is not necessary for expression of the activity (Parkinson, 1990). Also, in contrast with the great majority of metallocarboxypeptidases, whose optimum pH is around neutrality, CPE has its maximum activity at acidic pH, between 5 and 5,5 (Greene *et al*., 1992), coincident with the internal pH of the secretory granules. It has also been observed that its activity is regulated by the presence of Co^{2+} (Fricker *et al*., 1988). Several analogs of arginine and lysine, that were originally designed as active site directed inhibitors of CPB and CPN, were found to be potent inhibitors of CPE (Fricker *et al*., 1983). Two of these compounds, GEMSA (guanidinoethylmercaptosuccinic acid) and APMSA (aminopropylmercaptosuccinic acid), are several hundred fold more potent as inhibitors of CPE than of either CPB or CPN (Fricker *et al*., 1983).

It has been shown that mice with the mutation Cpe^{fat}/Cpe^{fat} have deficient proinsulin processing due to the absence of CPE activity in the pancreatic islets and the pituitary, caused by a point mutation Ser202Pro (Naggert *et al*., 1995). Mice containing such mutations in the CPE gene also show a reduced ability to process other hormones (Fricker *et al*., 1996). However, the observation that Cpe^{fat}/Cpe^{fat} mice are still able to process a small quantity of insulin suggested that another carboxypeptidase was also involved in peptide processing .

A search for additional CPE-like enzymes led to the discovery of metallocarboxypeptidase D (CPD; EC. 3.4.17.22) (Song & Fricker, 1995). Carboxypeptidase D (CPD) is a 180 kDa protein containing a signal peptide, three CPE-like domains of ~390 residues separated by short bridge regions, a transmembrane domain, and a 60 residue C-terminal cytosolic tail (Kuroki *et al*., 1995; Tan *et al*., 1997; Xin *et al*., 1997). The cDNAs corresponding to CPD of human (Tan *et al*., 1997), rat (Xin *et al*., 1997), mouse (Ishikawa *et al*, 1998), duck (Kuroki *et al*, 1995), *Drosophila melanogaster* (Settle *et al*, 1995), and *Aplysia californica* (Fan *et al*, 1999) have been cloned. All species contain 3 CPE-like domains (here named CPD-I, CPD-II and CPD-III), suggesting that their distinct physiological functions are important. The characterization of the first and second domains of CPD has shown that both possess catalytic activity and have somewhat complementary activities. Specifically, the first domain is optimally active at pH 6.3-7.5 and prefers substrates with C-terminal Arg, whereas the second domain is optimally active at pH 5.0-6.5 and prefers substrates with C-terminal Lys (Novikova *et al*., 1999). In contrast, the third

domain is inactive towards a variety of standard carboxypeptidase substrates (Eng *et al*., 1998; Novikova *et al*., 1999). Duck CPD, also named gp180, was identified by its ability to bind the preS envelope protein of duck hepatitis B virus particles (Kuroki *et al*., 1995). A comparison of human and duck CPD reveals 66, 83 and 82% amino acid sequence identity among the first, second, and third carboxypeptidase repeats, respectively. Recent studies with mutants lacking the first, second, or third carboxypeptidase-like domains have shown that the third domain of duck CPD is responsible for binding of the duck preS protein from hepatitis B virus, and that this binding does not require carboxypeptidase activity (Eng *et al*., 1998). Despite the absence of activity in the third domain, the fact that it is highly conserved among duck and mammals suggests the existence of a biological function for it.

Crystallization of both carboxypeptidase E and the complete three-domain carboxypeptidase D has been attempted. However, low yields in the protein recovery and the occurrence of glycosylations, together with the fact that the inter-domain linker peptides in CPD are probably highly flexible, have precluded direct 3D determination. The only crystal structure from the regulatory metallocarboxypeptidase subfamily that has been solved is that of the second domain of duck gp180 (Gomis-Rüth *et al*., 1999). It displays a 300 residue N-terminal α/β-hydrolase with overall topological similarity to the pancreatic carboxypeptidase A. This subdomain is followed by a C-terminal 80 residue β-sandwich subdomain, unique for these regulatory metalloenzymes and topologically related to transthyretin and sugar-binding proteins (Gomis-Rüth *et al*., 1999). In order to further investigate and better define the enzyme substrate pocket and to provide a basis for the rational design of specific inhibitors of regulatory carboxypeptidases, we have solved the crystal structure of carboxypeptidase D domain II in complex with the peptidomimetic inhibitor GEMSA. Based on this structure, overall models and detailed ones of the respective active sites have been built for human CPE and domains I and III of duck carboxypeptidase D. These models permit hypotheses about the structural basis of enzyme specificity and biological activity.

3.3. Materials and methods

3.3.1. Crystallization

Crystals of native CPD-II were produced as previously mentioned (Gomis-Rüth *et al*, 1999). The CPD-II/GEMSA complex was obtained by soaking native crystals in a solution 2.5 M in ammonium sulphate, buffered with 0.15 M sodium acetate to pH 5.2, and containing 10 mM 2-guanidinoethyl-mercaptosuccinic acid (GEMSA; purchased from Calbiochem), for 3 days. Diffraction data to 2.6 Å resolution were collected from a single N_2-cryocooled complex crystal that belongs to the same spacegroup ($P2_13$) as the native ones at the EMBL synchrotron beamline BW7B (DESY, Hamburg). Data were processed with MOSFLM v. 6.0.1 (Leslie, 1991) and SCALA from the CCP4 suite (CCP4, 1994). After a rigid body refinement step and positional/temperature refinement employing CNS v. 0.9 (Bruenger *et al*., 1998) and using maximum-likelihood as minimisation criterium and employing the coordinates of native CPD-II (after removal of solvent molecules and sulphate anion 998 located in the active site cleft), initial omit maps (sigma A-weighted $2F_{obs}$-F_{calc} and F_{obs}-F_{calc}) were computed. The difference density map clearly revealed the location of the bound inhibitor (see Fig. 1) permitting its model building using the Turbo-Frodo program (Roussel & Cambilleau, 1989). The complex was submitted to further positional/temperature refinement after setup of appropriate inhibitor parameter and topology files. The refinement of the occupancy of the latter revealed 100% presence, in accordance with the very high affinity of the inhibitor (in the nmol range). The final model comprises residues Gln4-Thr383 of the chemical sequence, 195 solvent molecules (labelled 601-795), one zinc cation (residue 999), one sulfate anion (998) with partial occupancy, and the 15-atom inhibitor GEMSA (residue Gem801). Three asparagine residues were found to be glycosylated (Asn136, Asn321 and Asn377; partial occupancies refined). One peptide bond (Pro190-Phe191) has been found in the *cis* conformation. Table I provides a summary of the data processing and final model refinement.

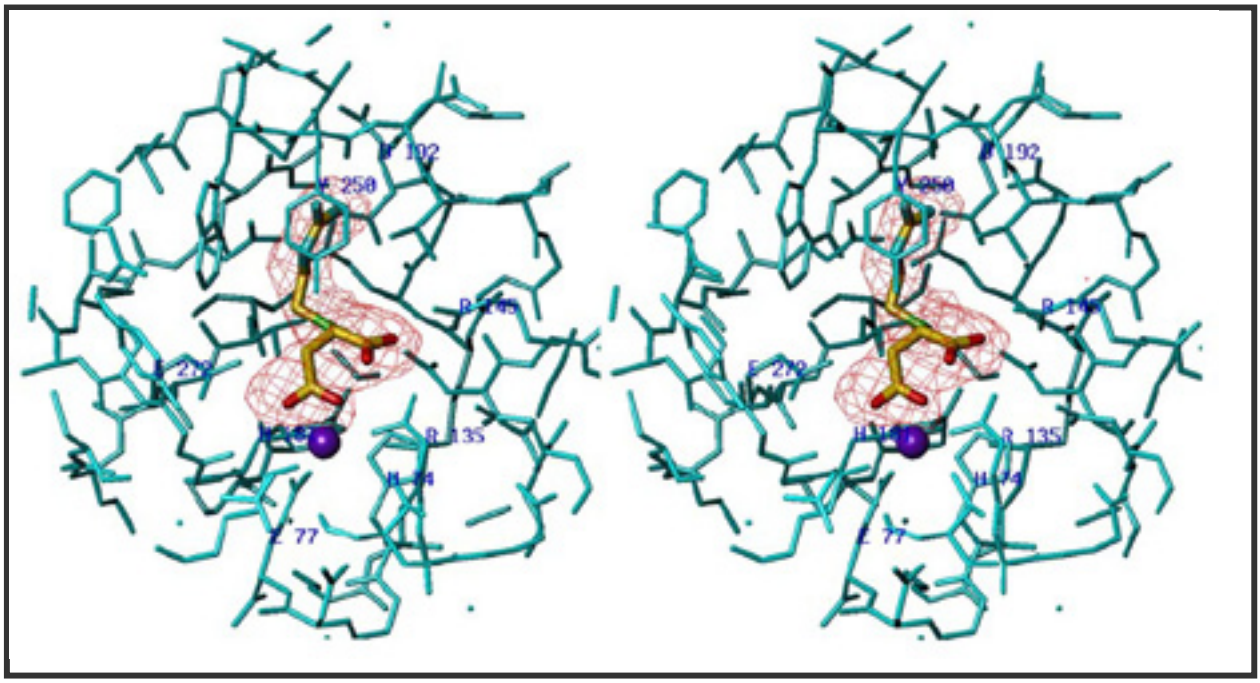

Figure 3.1. Stereo plot of CPD-II in complex with GEMSA displaying the final structure around the active site cleft superimposed with the initial sigma-weighted omit map density (F_{obs}-F_{calc}) contoured at 2.5 σ. The two inhibitor carboxylate groups coordinate the catalytic zinc ion (magenta sphere) in a bidentate manner and Arg 145/Arg 135 from the protein, respectively. The 2-guanidinoethyl moiety of the inhibitor is placed in the specificity pocket.

Table 1 - Data collection, processing, and final model refinement

Wavelength (Å)	0.8437
Temperature	N_2 cryocooling
No. of measurements	362,074
No. of unique reflections	26,360
Statistics - whole resol. range (Å)	43.03-2.60
Completeness(%) / R_{merge}[1]	100.0 / 8.3
$<I>/<\sigma(I)>$ / Average multiplicity	7.8 / 13.7
Statistics - last resol. shell (Å)	2.74-2.60
Completeness(%) / R_{merge}[1]	100.0 / 47.5
$<I>/<\sigma(I)>$ / Average multiplicity	1.6 / 13.9
B-value from Wilson-plot ($Å^2$)	59.4
Resolution range used for refinement (Å)	43.03-2.60
No. of reflections / completeness (%)	26,328 / 100.0
Final crystallographic R_{factor} [2] (using all data)	0.195
R_{factor} / free R_{factor} [2] (penultimate cycle)	0.199 / 0.239
Non-hydrogen protein atoms (incl. glucides and inhibitor)	3,138
Solvent molecules / Ions	195 / 1 Zn^{2+}, 1 SO_4
Glycosylation sites	Asn136, Asn321, Asn377
R.m.s. deviation for the protein part (incl. glucides) from target values of	
bonds (Å)	0.007
angles (°)	1.34
bonded Bs main chain / side chain ($Å^2$)	1.92 / 2.95
Average (refined) occupancies / Average B-factors ($Å^2$)	
protein atoms	1.0 / 43.5
solvent molecules	1.0 / 48.6
sulphate anion	0.94 / 51.3
carbohydrate atoms (attached to Asn136, Asn321, and Asn377)	
	0.93 / 49.3; 0.69 / 62.3; 0.73 / 61.3
zinc cation	1.0 / 33.3
inhibitor atoms	1.0/ 39.6

[1] $R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - <I(hkl)>| / \Sigma_{hkl}\Sigma_i I_i(hkl)$.

[2] $R_{factor} = \Sigma_{hkl} ||F_{obs}| - k |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$; free R_{factor}, same for a test set of 1950 (7 %) reflections not used during refinement (till cycle last-but-one). This test set includes the same reflections as the native dataset (np), further extended to cover the non-common resolution shell (2.70 - 2.60 Å).

3.3.2. Model building

A preliminary multiple alignment was performed by means of the program PILEUP (Fenger & Doolitte, 1987; Higgins & Sharp,1989) for the three duck CPD structural repeats (domains). This alignment was used as a "seed" to build a hidden Markov model profile with the program HMMER (Eddy, 1998) that was used to align eight additional homologous sequences. Expert knowledge and experimental information were also used in order to improve the quality of the alignment in several segments. The primary and three dimensional structures of duck CPD-II were used as a template to build the models. A segment of 30 residues from carboxypeptidase T (PDB access code 1obr) was aligned to CPE in order to model a 23-residues insertion (residues 202-224; see footnote to Table II for the conventions used on numbering the different sequences). Finally, a 25-residues stretch from the sequence of Adenovirus coat protein (1dhx) was also aligned to the insertion observed in CPD-I (residues 96-124). Using the multiple alignment for the three CPD domains (CPD-I, CPD-II and CPD-III) and CPE as a starting point, a method of comparative modelling by satisfaction of spatial constraints was used to build the 3D structure of CPD-I, CPD-III and CPE. This method is implemented in the program MODELLER (Sali & Blundell, 1993). The spatial constraints are derived by transferring the spatial features from the structures of known proteins to the sequence of the unknown ones. The program PROSA-II (Sippl, 1993) was used to check the quality of the models as described in a previous work (Aloy *et al*., 2000). The regions with non near-native fold were identified by the high positive values of pseudo-potential energy, independently of the crystallographic structure. Once the three models were built automatically, manual intervention was required for re-modelling those regions identified by PROSA-II with non near-native fold. The program FRAZER, developed in our laboratory (unpublished), was used to reconstruct the problematic regions. The overall RMSD calculations and superimposition of the three modelled structures with respect to the crystallographic one (CPD-II) were obtained according to the structural alignment given by the program SSAP (Orengo *et al*., 1992). The active sites superposition and GEMSA inhibitor replacement in the three models were also performed with FRAZER.

3.4. Results

3.4.1. Structure of the complex CPD-II-GEMSA

The CPD-II polypeptide chain in the complex is folded into two distinct subdomains, a 300-residue catalytic carboxypeptidase subdomain displaying the α/β-hydrolase fold reminiscent of the carboxypeptidase A structure, and a 80-residue C-terminal subdomain of all-β prealbumin-like β-sandwich folding topology, the so-called transthyretin subdomain (Gomis-Rüth *et al*, 1999). The complex structure displays no significant deviations from the native protein (Gomis-Rüth *et al*., 1999), as denoted by a RMSD of 0.19 Å for all Cα atoms. Only the catalytic zinc ion is somewhat moved away (0.7 Å) from its position in the isolated domain structure forced by the presence of the inhibitor. This movement is accompanied by a similar displacement (0.7 Å) on the coordinating residue, His74. Interestingly, the catalytic solvent molecule attached to the zinc ion (Wat601) in the unliganded structure is moved 2.3 Å away upon inhibitor binding (Wat684 in the present structure, see Fig. 2). The peptidomimetic GEMSA molecule occupies the primed side of the catalytic active site cleft, emulating a bound C-terminal amino acid after proteolytic cleavage of a substrate. The guanidinoethylmercapto group is reminiscent of a substrate arginine side chain (CPD-II displays a CPB-like preference for basic residues in P_1') and occupies the same position in the specificity pocket. It is anchored through its atoms Nη1 and Nη2 to the side chain of Asp192 and the main chain carbonyls of Gly246 and Tyr250, the latter one present in the "down" conformation as in the native structure. This planar guanidinoethylmercapto group establishes an additional van-der-Waals interaction (3.8 Å) with Val252. The inhibitor carboxylate group mimicking a peptide substrate C-terminus is anchored to both Arg145 and Arg135. The second carboxylate group is similar to a scissile carbonium ion in the transition state and coordinates the catalytic zinc ion in a bidentate manner. One of its oxygens is further bonded by Arg135 and His74 (see Fig. 2).

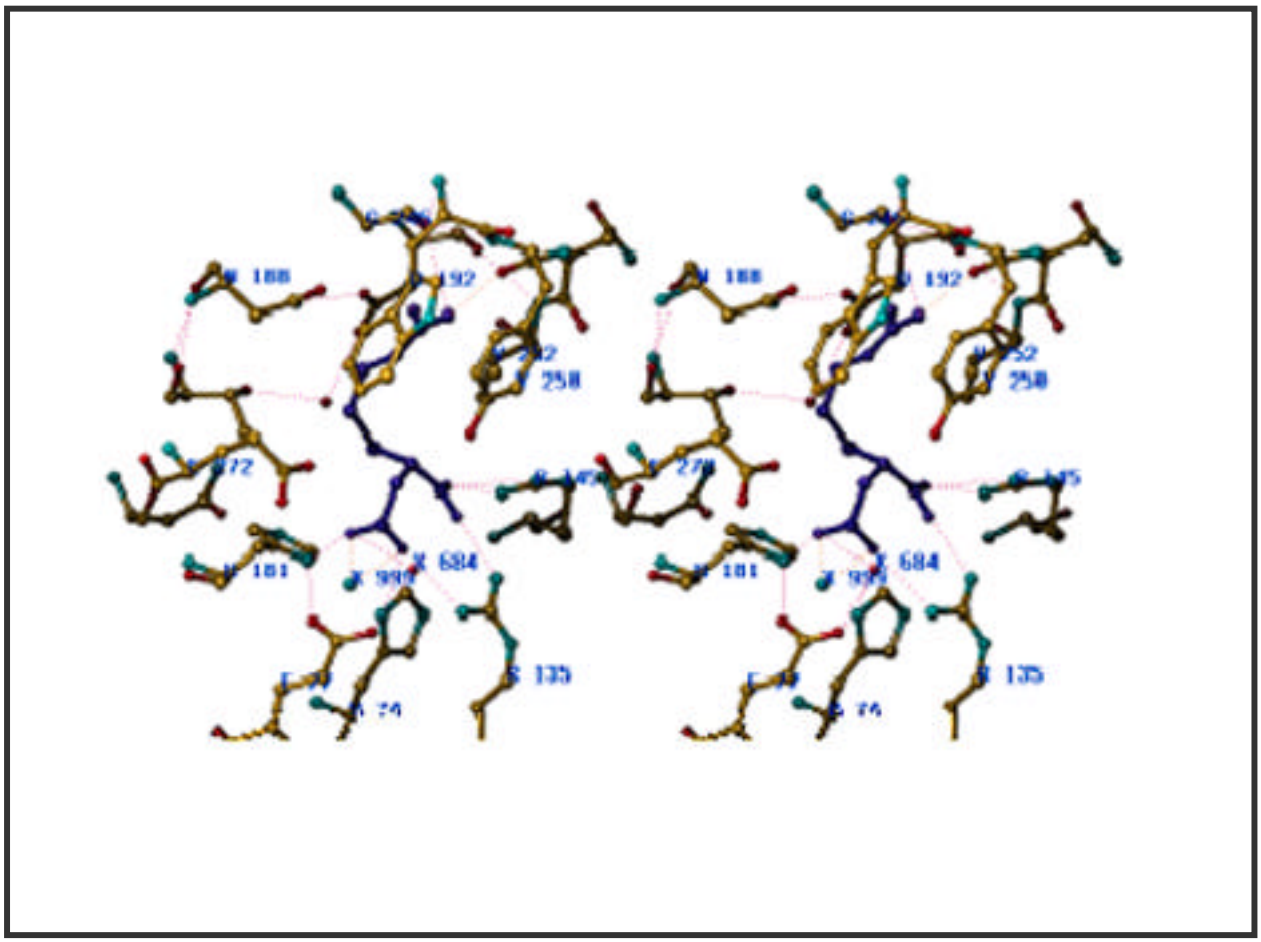

Figure 3.2. Stereo plot of CPD-II in complex with GEMSA as a ball-and-stick model displaying the hydrogen bond network around the active site. The catalytic zinc ion (999) is displayed as a cyan sphere, the inhibitor moiety in violet, and solvent molecules as red spheres. Intervening residues are labelled.

3.4.2. Sequence alignments, model building and refinement

Figure 3.3 shows the multiple sequence alignment of the three domains of duck CPD and human CPE. This alignment, performed as indicated in the Methods section, allows us to derive accurate models for these proteins. The alignment reveals a 42% and 32% sequence identity between duck CPD domain II and domains I and III respectively. The percentage identity for human CPE with respect to duck CPD-II is even higher, 50%. These identity levels allow homology modelling of the three dimensional structures of these proteins.

The alignment in Figure 3.3 shows that all the residues experimentally described as important (those at the active site, at the metal binding site and at the substrate binding subsites) are essentially conserved among all the sequences, except for CPD-III. Several insertions and deletions, however, can be observed in the alignment. First of all, a large insertion (29 residues) in domain I of duck CPD can be detected. This inserted stretch is extremely charged, with 5 basic and 15 acidic residues, and the only sequences in the data banks that show a certain level of homology with it belong to proteins that interact with nucleic acid binding proteins. This suggests that this segment could adopt a helical topology and, somehow, mimic nucleic acid molecules in order to attach the binding proteins. In any case, only one of these related sequences has its 3D structure determined (adenovirus type 2 hexon, pdb code 1dhx), and the low percentage of identity observed is not sufficient to model this stretch. Another large insertion (23 residues) can be found in human CPE. This insertion is present in all species of CPE as well as most other members of the CP family, although the length varies from 14-15 residues for CPA, CPB and the bacterial CP to 27 residues for CPX-1, CPX-2, and AEPB-1. Because the three-dimensional structure of this loop in CPA, CPB, and CPT is known, this region of CPD can be modelled using the crystal structure and alignment of the other CPs. The rest of indels are much shorter and can be modelled with reasonable confidence by energy optimisation.

The pseudo-energies of the original models were calculated with PROSA-II (Sippl, 1993) in order to identify the incorrect chain tracings. As expected, the regions that presented higher energy were those of the indels (data not shown). In the case of the large highly charged insertion in duck CPD-I, the energy tended to infinite and, consequently, the loop was removed from the model. For the insertion in human CPE, the pseudo-energy was corrected to acceptable values by manual modification of the human CPE model and energy minimisation.

The overall RMSDs calculations between the 3D structure of duck CPD-II and the

three different models gave the following values: 0.5Å for domain I (once the non-modelled loop was removed), 1.3 Å for domain III and 0.3 Å for human CPE. Taking into account that only one crystal structure was used to model the three sequences, the RMSDs correlate well with the percentage of sequence identity obtained in the multiple alignment. The RMSD was also calculated for the three models using only the active site residues. In this case, the results were: 0.1 Å for duck CPD-I and CPE and 0.5 Å for the duck CPD-III.

Figure 3.4 shows the modelled 3D-structures of domains I and III of duck CPD and human CPE compared to the crystal structure of CPD-II. The RMDS values indicate that, although a number of local differences are obviously present (discussed below) the models share a common topology and the relative positions of the two subdomains is maintained in all of them. A close inspection of the models also shows that the major structural features in CPD II that suggest a different selectivity of regulatory carboxypeptidases towards large protein substrates as compared to the pancreatic enzymes are also present in the other members of the regulatory family studied here. Thus, those loops in the funnel-like access to the active site, that are probably responsible for the different selectivity of the two families of carboxypeptidases, conform an opening of the solvent-exposed surface which, beyond individual characteristics that will be discussed below, is very similar in all cases.

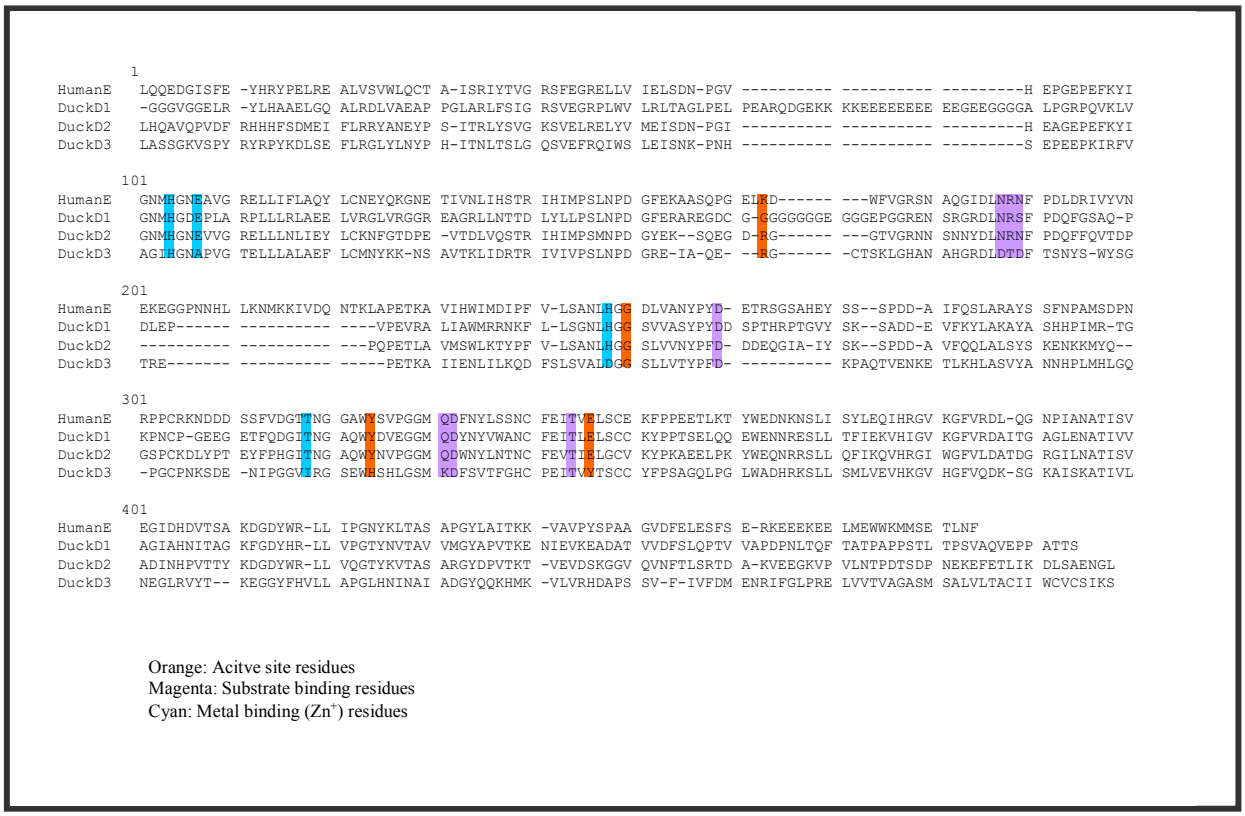

Figure 3.3. Multiple alignment of duck carboxypeptidase D subunits one, two and three (DuckD1, DuckD2 and DuckD3) and human carboxypeptidase E (HumanE). The alignment was performed using the programs PILEUP, HMMER and manually refined to account for experimental information. Key residues for catalysis (orange), substrate binding (magenta) and zinc binding (cyan) are boxed.

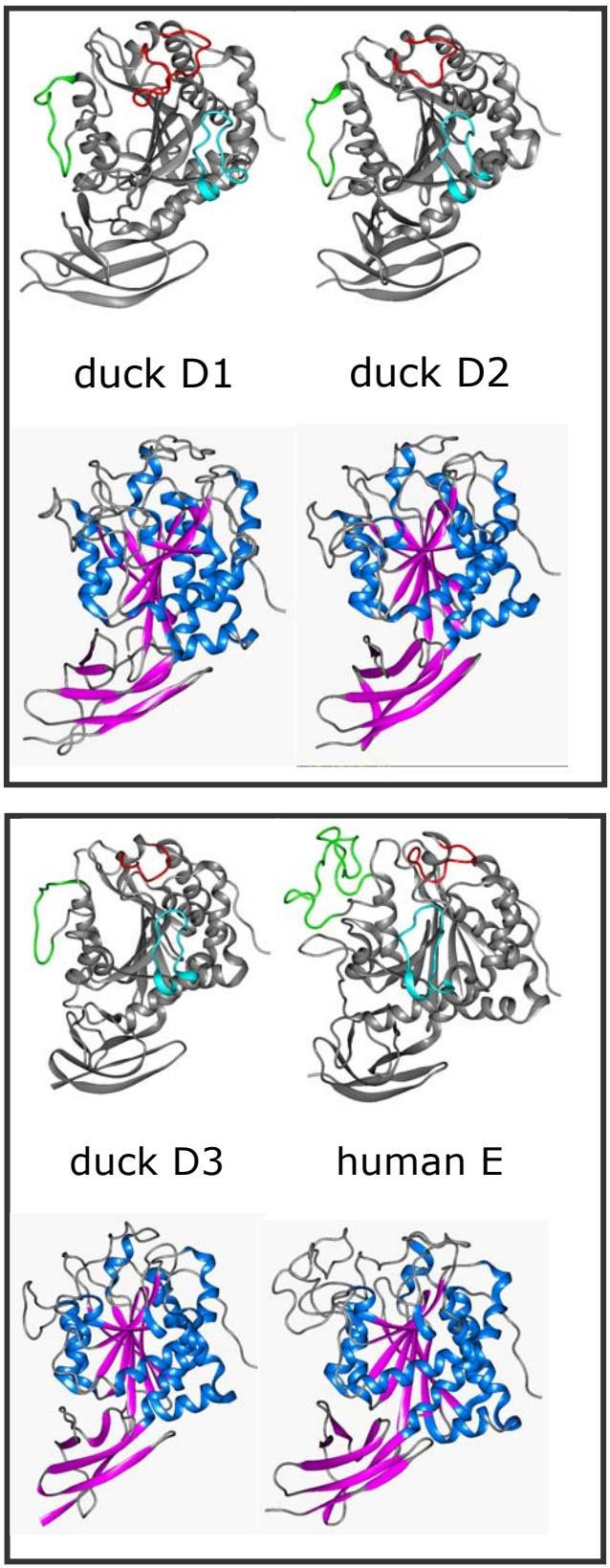

Figure 3.4. Ribbon representation of the modelled structures of duck carboxypeptidase D (subunits I and III) and human carboxypeptidase E compared to the crystal structure of duck CPD II. Top: the three loops that shape the entrance to the active site are colored in red (124-134), green (149-159) and cyan (225-241). Bottom: modelled structures showing the regular secondary structures:α-helix (blue) and β-strands (magenta). The residue numbering corresponds to the CPD-II structure.

3.4.3. Conserved interactions between GEMSA and the different models

After superimposing the four active sites, the GEMSA inhibitor was emplaced in the three models in order to find and rationalize its possible interactions with the enzymes. The fit was excellent in all three cases. The residues in the x-ray structure of duck CPD-II interacting with the GEMSA inhibitor and their equivalents in the three modelled structures are shown in Table II. As can be seen, in duck CPD-I and human CPE all the hydrogen bonds found in the co-crystal are conserved in the complex between the inhibitor and the protein residues. In contrast, in duck CPD-III several critical interactions are lost due to the different residues found at the active site.

3.5. Discussion

3.5.1. Structural basis of the inhibitor action

The carboxypeptidase inhibitor GEMSA has been frequently used as a potent inhibitor of regulatory carboxypeptidases (CPN, CPE, CPD) but this is the first time that a crystal structure of its complex with one of these enzymes has been reported. Such structure clearly explains the powerful action of this inhibitor, previously inferred from a crystal structure derived from a complex with a pancreatic carboxypeptidase: the catalytic water and the essential Zn^{2+} are both displaced, the latter one being bound in a bidentate manner by one of the carboxylate groups of the inhibitor. In addition, the inhibitor is bound to residues of duck CPD-II that are essential for substrate binding and polarization: Tyr250, Arg145 and Arg135. Therefore, several structural elements indispensable for the catalytic action of the enzyme are perturbed or shielded by the inhibitor.

Taking into account the similarity of CPD-II to the modelled structures, and the easy way in which GEMSA has been fitted on them, it is expected that the inhibitor binds in a very similar way to CPD-I and CPE. However, it is unlikely that GEMSA binds CPD-III due to the absence of critical residues (discussed below).

3.5.2. Overall comparison of the models

The derived models of domains I and III of avian CPD and human CPE show an overall similarity with the recently described crystal structure of domain II of avian CPD (Gomis-Rüth *et al*.,1999). In all models two subdomains are clearly visible: the carboxypeptidase (CP) subdomain and the C-terminal subdomain (CTSD) which shares topological similarity and connectivity with transthyretin (Gomis-Rüth *et al*., 1999). The CP subdomain shows the α/β-hydrolase fold common to many proteases from the cysteine- serine- and metalloprotease families. It is formed by a doubly wound eight-stranded β-sheet flanked on both sides by three and six helices, respectively. Meanwhile, the CTSD subdomain displays a rod-like shape forming a β-barrel or β-sandwich of prealbumin-like folding topology made up by seven strands connected by short loops. This is valid for all the models studied here.

Like in CPD-II, the interactions between both subdomains are mainly of a hydrophobic nature in all models. Most of the van der Waals interactions described for CPD-II are also found in CPD-I and human CPE. Albeit containing a smaller number of such interactions, CPD-III still conserves the most significant ones. A number of hydrogen bonds also contribute to subdomain interactions in CPD-I, CPD-III and human CPE, most of them being exactly conserved in CPD-I and CPE versus CPD-II, and greater differences being found for CPD-III. It is worth mentioning that the only salt bridge between subdomains described for CPD-II, Asp206-Arg343, is also conserved in the modelled structures between pairs of Asp/Arg at equivalent positions in CPD-I and CPE, and between Glu1123 and His1258 at equivalent positions in CPD-III. On the other hand, the unique disulfide bond found in domain II between Cys230 and Cys275 is also the only one found in the three models, whereas the three glycosylation sites observed in domain II probably are absent in domain III, and only one (the equivalent to Asn321 in domain II) is expected to be found in domain I and in CPE if the residue substitutions are considered.

CPE is the protein with the highest homology to CPD domain II and also the one whose model has the lowest RMSD value with the experimental 3D structure. However, it should be taken into account that the RMSD value was calculated only on the structurally equivalent residues given by the alignment performed with the program SSAP. This means that, for instance, the 23-residues insertion in CPE, spanning from Glu158 to Lys189, was not considered. As compared to CPD-II, pancreatic and bacterial carboxypeptidases have also a

14 amino acid insertion in this region, forming a loop that shapes one side of the entrance to the active site and establishes cross-connections to an adjacent loop. This feature is considered to be one of the distinctive determinants of specificity between regulatory and pancreatic carboxypeptidases. The 23 extra residues in CPE form a turn-rich region rather exposed to the solvent in the model built, according to the well-defined structure of this loop in *T. vulgaris* carboxypeptidase (1obr).

The main difference between domains I and II of CPD is the above mentioned long insertion of 29 residues in domain I that contains 20 net charges. This sequence was eliminated from the calculations as no homologous sequences and 3D-structures were found to model it with a sufficient degree of confidence. A further significant difference is a glycine-rich insertion of 9 residues in one of the loops that shape the active site entrance (residues 165 to 173 in Fig. 1). This insertion does not generate a substantial change in the surface of the active site cleft in our model and is folded inwards over the molecular body of the enzyme.

The study of the model of the domain III of avian CPD is particularly important because of its lack of enzymatic activity (Eng *et al*., 1998), probably due to the absence of key residues for carboxypeptidase catalysis. However, alignment of the sequences and superimposition of the three-dimensional structures shows that other residues with yet unknown function are highly conserved. When comparing the models and the structure, three categories of residues can be defined. The first one is formed by the residues essential for catalysis. In domain II, these essential residues are His 74, Glu77, His181 (coordinators of Zn^{2+}), Glu272 and Arg135. Only the first His is conserved in domain III, while the other residues are replaced by Ala, Asp, Tyr and His, respectively. The enzymatic machinery of domain III is, therefore, disabled because neither proper coordinators of Zn^{2+}, nor a general base or a polarizing residue are present, respectively (Aviles *et al*., 1993).

Those residues that are necessary for substrate binding are included in the second category. The triad Asn144, Arg145 and Asn146, that is responsible for the anchoring of the C-terminal carboxylate (COO^-) of the substrate, is generally conserved in all carboxypeptidases that are enzymatically active towards peptides, including the pancreatic and bacterial CPs (Table II). In CPD-III, this triad is replaced by Asp, Thr and Asp, rendering a domain that has lost the ability to anchor the carboxyl group. Interestingly, a peptidase in the bacteria *Bacillus sphaericus* is a distant member of the metalloCP family which also lacks this Asn-Arg-Asn triad (Guinand, 1998). Instead of cleaving substrates with a C-terminal carboxylate group, as in other CP, the *B. sphaericus* peptidase hydrolyzes C-terminal *meso*-diaminopimelic acid. This substrate has an amino group in place of the carboxylate of a

typical peptide, consistent with the replacement of the Asn-Arg-Asn with an Asn-Asp-Gln. Thus, the differences in this sequence between CPD-II and CPD-III is predicted to be critical for defining the binding specificity of each domain.

Some other residues necessary for substrate binding in CPD-II are also different in domain III. For example, Tyr250, Val252, and Gly255 of domain II are replaced by His, His and Ser, respectively, in domain III. However, in spite of these replacements, when the model of domain III and the crystal structure of domain II are superimposed, it can be observed that the different residues in domain III occupy exactly the same position of their homologues in domain II. Taken together, it is unlikely that CPD-III binds with high affinity to peptides that are substrates of the other two domains.

The rest of the residues that are highly conserved in almost all carboxypeptidases, either regulatory or pancreatic, like Gly40, Asn117, Gly120 and Pro190 (numbering system of CPD-II) would belong to the third category. None of them has been related to any specific function and their role is more likely purely structural.

Thus, to summarize, the catalytic machinery of domain III has been suppressed by replacement of the key residues for carboxypeptidase activity, and there are also substantial differences in the residues responsible for substrate binding. The high conservation of sequence and structure in the enzymatically incompetent domain III suggests another biological function, possibly related to the binding of proteins or other molecules.

3.5.3. Active site and substrate binding subsites

All residues involved in metal binding and catalysis are conserved in domains I and II of CPD and in CPE. Domain III is the already commented exception because it lacks most of the residues involved in Zn^{2+} binding and the Arg that bind the terminal carboxylate (here a threonine) and polarize the scissile peptide bond (a His in domain III). Also, the general base (Glu272 in domain II) has its position occupied by a Tyr in domain III.

The loops that form the specificity pocket in domain II (S1' subsite) (Asn188 – Asp192, Gly246 – Gln257 and Phe267 – Thr270) have the same length in all the models; amino acid residue identity in these loops is high for domain I and CPE and low for domain III. There is also low identity between these loops of domain II and those of the pancreatic carboxypeptidases. On the other hand, it is worth to note that Tyr250 (equivalent to Tyr248 in pancreatic carboxypeptidases, the one that caps the active site, facilitating the proper location

of the substrate over it, and that fluctuates between two conformations depending on substrate binding), is replaced by histidine in domain III, supporting the idea that this domain is unable to catalyze peptide bond hydrolysis.

A key residue that is essential for the specificity of digestive CPB for C-terminal basic residues is Asp 255 (Coll *et al*., 1991, Aviles *et al*., 1993), which is replaced by an Ile or Leu in the digestive enzymes that prefer C-terminal aliphatic and aromatic residues (Table 2). In CPD-I, CPD-II and CPE, which all are highly specific for basic C-terminal amino acids, the residue in a position sequentially equivalent to this Asp255 of CPB is a Gln (Table 2), which is functionally unable to perform a similar role as the Asp. Instead, the electronegative character required for the selectivity for C-terminal Lys and Arg residues is provided by Asp192, located in a spatially comparable position. This Asp192 is conserved in all regulatory CP, including CPD-III.

The relevant residues at the S2 subsite in domain II are also found in equivalent positions in the three models of domains I, III and CPE. The residues that line this subsite are considerably different in pancreatic carboxypeptidases, suggesting that a general specificity for either sequence or volume of the substrates is shared in all regulatory enzymes, including the inactive domain III. As an example, Gly182 and Gly183 (CPD domain II numbering) are present in all models of the regulatory forms at the same positions found in the crystal structure, whereas the equivalent residues in pancreatic carboxypeptidases are Ser197 and Tyr198, also highly conserved in such pancreatic enzymes.

Variation is also observed in all proteins for those residues putatively involved in subsite S3. However, one remarkable difference involves Lys277, conserved in CPD-I and CPE and that was putatively involved in P_2 carbonyl oxygen binding in CPD-II (Gomis-Rüth *et al*, 1999), which is replaced by a Tyr in CPD-III.

3.5.4. Accessibility of the active site

One of the most significant structural differences between the crystal structure of CPD domain II and pancreatic carboxypeptidases is the long insertion Tyr225-His241 that shapes the border of the funnel that leads to the active site and that hinders the binding of potato CP inhibitor (PCI) to CPD. PCI is a 39 residue peptide that potently inhibits several of the digestive carboxypeptidases including CPA, CPB, and CPU (see Gomis-Rüth *et al*, 1999). Although particular residues are not conserved, the loop is present in all models suggesting that restrictions in specificity may be common to all regulatory enzymes. However, two further loops are also critical in shaping the funnel border and significant differences are observed in these cases (Fig. 4). The insertion of a 9-residues glycine rich sequence at loop Ser124-Val133 (domain II numbering) does not seem to affect the accessible surface in domain I of CPD. In all cases the loop is longer than that observed in pancreatic carboxypeptidases and is folded inwards, partially covering the access to the active site. CPE has a much longer insertion between residues 157 and 158 of CPD-II (Figure 3.3) that coincides with an equivalent, albeit shorter, insertion in pancreatic CPA and B. Taken together, all these observations suggest that, within a general frame of specificity, regulatory carboxypeptidases have developed variations in the structural determinants that lead to selection of substrates that are far more sophisticated than the mere selectivity of C-terminal residues observed in the pancreatic enzymes. Work is in progress to test this hypothesis.

The information collected or derived in the present study might facilitate the understanding of the differential biological roles of regulatory carboxypeptidases, and the design of specific inhibitors for them. These would be interesting tools to experimentally analyze the properties and roles of these enzymes, and to produce lead compounds for drug design, given the potential biotechnological and biomedical interest in the modulation of their activities.

Table 2. Selected residues in duck CPD-II and their equivalents in bovine CPA, bovine CPB, human CPE, and full-length duck CPD

Duck	Full-length duck CPD[2]			Human	Bovine	Bovine
CPD-II[1]	I	II	III	CPE[3]	CPA[4]	CPB[4]
H74*	H139	H574	H997	H114	H69	H69
E77*	E142	E577	A1000	E117	E72	E72
R135*	R212	R634	H1058	R179	R127	R127
N144	N221	N643	D1067	N188	N144	N144
R145	R222	R644	T1068	R189	R145	R145
N146	S223	N645	D1069	N190	N146	N146
H181*	H260	H680	D1105	H248	H196	H196
G183	G262	G682	G1107	G250	Y198	Y198
N188*	S267	S687	T1112	D255	L203	M203
D192*	D271	D691	D1116	D259	G207	S207
Y250*	Y331	Y749	H1168	Y320	Y248	Y248
V252*	V333	V751	H1170	V322	A250	A250
Q257	Q338	Q756	K1175	Q327	I255	D255
E272*	E353	E771	Y1190	E342	E270	E270

*) Residues interacting with the inhibitor GEMSA.

1) The numbering system is based on the construct of CPD-II alone that was expressed in *Pichia* and used for the determination of the crystal structure (Gomis-Rüth *et al*, 1999).

2) The numbering system is based on the full-length amino acid sequence deduced from the cDNA sequence (Kuroki *et al*., 1995). In this system, amino acid number 1 is assigned to the initiation Met of the signal peptide.

3) The numbering system is based on the full-length amino acid sequence deduced from the cDNA sequence (Manser *et al*, 1990). In this system, amino acid number 1 is assigned to the initiation Met of the signal peptide.

4) This numbering system is the standard for metallocarboxypeptidases, and assigns amino acid number 1 to the first residue of the mature form, after removal of the signal peptide and pro-segment.

3.6. References

Aloy P, Mas J M, Martí-Renom MA, Querol E, Aviles FX, Oliva B (2000) Refinement of modelled structures by knowledge-based energy profiles and secondary structure prediction: Application to the human procarboxypeptidase A2. JCAMD 14, 83-92.

Aviles FX, Vendrell J, Guasch A, Coll M, Huber R (1993) Advances in metallocarboxypeptidases. Emerging details on the inhibition mechanism and on the activation process. Eur J Biochem 211, 381-389.

Brünger AT *et al*. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 52, 43-48.

Coll M, Guasch A, Aviles FX & Huber R (1991) Three-dimensional structure of porcine procarboxypeptidase B : a structural basis of its inactivity. EMBO J. 10, 1-9.

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14, 755-763.

Eng FJ, Novikova EG, Kuroko K, Ganem D, Fricker LD (1998) gp180, a protein that binds hepatitis B virus particles, has metallocarboxypeptidase D-like enzymatic activity. J Biol Chem 273, 8382-8388.

Fan X, Nagle GT (1996) Molecular cloning of *Aplysia* neuronal cDNAs that encode carboxypeptidases related to mammalian processing enzymes. DNA Cell Biol. 15, 937-945.

Fan X, Qian Y, Fricker LD, Akalal DBG, Nagle GT (1999) Cloning and expression of *Aplysia* carboxypeptidase D, a candidate prohormone-processing enzyme. DNA Cell Biol., 18, 121-132.

Fenger DF, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J Mol Evol 25, 351-360.

Fricker LD and Snyder SH (1983) Purification and characterization of enkephalin convertase, an enkephalin-synthesizing carboxypeptidase. J Biol. Chem. 258, 10950-10955.

Fricker LD, Adelman JP, Douglass J, Thompson RC, von Strandmann RP, Hutton J (1989) Isolation and sequence analysis of cDNA for rat carboxypeptidase E [EC 3.4.17.10], a neuropeptide processing enzyme. Molecular Endocrinology, 3, 666-673.

Fricker LD, Berman YL, Leiter EH, Devi LA (1996) Carboxypeptidase E activity is deficient in mice with the *fat* mutation: effect on peptide processing. J Biol Chem 271, 30619-30624.

Fricker LD, Evans CJ, Esch FS, Herbert E (1986) Cloning and sequence analysis of cDNA for bovine carboxypeptidase E. Nature 323, 461-464.

Fricker LD, Snyder SH (1983) Purification and characterization of enkephalin convertase, an enkephalin synthesizing carboxypeptidase. J Biol Chem 258, 10950-10955.

Fricker LD (1988) Carboxypeptidase E. Annu Rev Physiol 50, 309-321.

Fricker LD (1991) Peptide processing exopeptidases: amino- and carboxypeptidases involved in peptide biosynthesis. In Fricker LD (ed.) Peptide Biosynthesis and Processing. CRC Press Boca Raton FL, pp.199-230.

Gomis-Rüth FX, Companys V, Quian Y, Fricker LD, Vendrell J, Aviles FX, Coll M (1999) Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallocarboxypeptidase subfamily. EMBO J 18, 5817-5826.

Greene D, Das B, Fricker LD (1992) Regulation of carboxypeptidase E: effect of pH, temperature and Co^{2+} on kinetic parameters of substrate hydrolysis. Biochem J 285, 613-618.

He GP, Muise A, Li AW, Ro HS (1995) An eukaryotic transcriptional repressor with carboxypeptidase activity. Nature 378, 92-96.

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. Comput Appl Biosci 5, 151-153.

Ishikawa T, Murakami K, Kido Y, Ohnishi S, Yazaki Y, Harada F, Kuroki K (1998) Cloning, functional expression, and chromosomal localization of the human and mouse gp180-carboxypeptidase D-like enzyme. Gene. 215, 361-370.

Kuroki K, Eng F, Ishikawa T, Turck C, Harada F, Ganem D (1995) gp180, a host cell glycoprotein that binds duck hepatitis B virus particles, is encoded by a member of the carboxypeptidase gene family. J Biol Chem 270, 15022-15028.

Lei Y, Xin X, Morgan D, Pintar JE, Fricker LD (1999) Identification of mouse CPX-1, a novel member of the metallocarboxypeptidase family with highest similarity to CPX-2. DNA Cell Biol 18, 175-185.

Leslie AGW (1991) Macromolecular data processing. In Moras D, Podjarny AD, Thierry JC (eds). Crystallographic computing V. Oxford University Press, Oxford, UK, pp. 27-38.

Manser E, Fernandez D, Loo L, Goh PY, Monfries C, Hall C, Lim L (1990) Human carboxypeptidase E: Isolation and characterization of the cDNA, sequence conservation, expression and processing *in vitro*. Biochem J 267, 517-525.

Naggert JK, Fricker LD, Varlamov O, Nishina PM, Rouille Y, Steiner DF, Carroll RJ, Paigen BJ, Leiter EH (1995) Hyperproinsulinemia in obese fat/fat mice associated with a point mutation in the carboxypeptidase E gene and reduced carboxypeptidase E activity in the pancreatic islets. Nature Genet 10, 135-142.

Novikova EG, Eng FJ, Yan L, Quian Y, Fricker LD (1999) Characterization of the enzymatic properties of the first and second domains of metallocarboxypeptidase D. J Biol Chem 274, 28887-28892.

Orengo CA, Brown NP, Taylor WR (1992) Fast structure alignment for databank searching. Proteins 14, 139-167.

Parkinson D (1990) Two soluble forms of bovine carboxypeptidase H have different NH2-terminal sequences. J. Biol. Chem., 265, 17101-17105.

Rawlings ND, Barret AJ (1995) Evolutionary families of metallopeptidases. Methods enzymol 248, 183-228.

Rodríguez C, Brayton KA, Brownstein M, Dixon JE (1989) Rat preprocarboxypeptidase H: cloning, characterization and sequence of the cDNA and regulation of the mRNA by corticotropin releasing factor. J Biol Chem 264, 5988-5995.

Roth WW, Mackin RB, Spiess J, Goodman RE, Noe BD (1991) Primary structure and tissue distribution of anglerfish carboxypeptidase E. Molec Cell Endocrinol 78, 171-178.

Roussel A, Cambilleau C (1989) Turbo-Frodo, Silicon Graphics Geometry Partners Directory. Silicon

Graphics, Mountain View, CA, pp. 77-79.

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restrains. J Mol Biol 234. 779-815.

Settle SH Jr, Green MM, Burtis KC (1995) The silver gene of *Drosophila melanogaster* encodes multiple carboxypeptidases similar to mammalian prohormone-processing enzymes. Proc. Natl. Acad. Sci. USA, 92, 9470-9474.

Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. Proteins 17, 355-362.

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

Skidgel RA (1996) Structure and function of mammalian zinc carboxypeptidases. In Hooper, NM (ed.) Zinc Metalloproteases in Health and Disease. Taylor and Francis, London, UK, pp. 241-283.

Song L, Fricker LD (1995) Purification and characterization of carboxypeptidase D, a novel carboxypeptidase E-like enzyme from bovine pituitary. J Biol Chem 270, 25007-25013.

Song L, Fricker LD (1997) Cloning and expression of human carboxypeptidase Z, a novel metallocarboxypeptidase. J Biol Chem 272, 10543-10550.

Strittmatter SM, Lynch DR, Snyder SH (1984) [^{3}H]Guanidinoethylmercaptosuccinic acid binding to tissue homogenates. J Biol Chem 259, 11812-11817.

Tan F, Rehli M, Krause SW, Skidgel RA (1997) Sequence of human carboxypeptidase D reveals it to be a member of the regulatory carboxypeptidase family with three tandem active site domains. Biochem J 327, 81-87.

Vendrell J, Querol E, Aviles FX (2000) Metallocarboxypeptidases and their protein inhibitors. Biochim. Biophys. Acta 1477, 284-298.

Xin X, Day R, Dong W, Lei Y, Fricker LD (1998) Identification of mouse CPX-2, a novel member of the metallocarboxypeptidase gene family: cDNA cloning, mRNA distribution and protein expression and characterization. DNA Cell Biol 17, 897-909.

Xin X, Varlamov O, Day R, Dong W, Bridget MM, Leiter EH, Fricker LD (1997) Cloning and sequence analysis of cDNA encoding rat carboxypeptidase D. DNA Cell Biol 16, 897-905.

Anexo al capítulo 3

El trabajo presentado en este capítulo, al igual que el presentado en el capítulo anterior, se ha realizado en colaboración con el grupo del Dr. Miquel Coll y del Dr. Xavier Gomis-Rüth de la Unidad de Cristalografía del Instituto de Bioquímica y Biología Molecular del CSIC en Barcelona. Mi contribución en este trabajo ha tenido dos vertientes: por un lado, la experimental y, por otro, la teórica. Las proteínas objeto de estudio (CPD dominio 2, CPE silvestre, Δ23 CPE y E270Q CPE) se clonaron y expresaron en *Pichia pastoris*. Desafortunadamente, la única proteína que rindió niveles de expresión que permitieran la purificación en las cantidades necesarias para realizar la cristalización y posterior resolución de su estructura tridimensional fue la CPD2 (detallada en el capítulo 2 de la presente Tesis Doctoral). Es por ello que me animé a confeccionar los modelos de los dominios 1 y 3 de CPD y CPE con la supervisión técnica de algunos compañeros de grupo. Finalmente, realicé el análisis comparativo de la estructura de la CPD2 con los modelos obtenidos.

Discusión general

El trabajo sobre metalocarboxipeptidasas presentado en esta Tesis Doctoral comprende tres partes diferenciadas. La primera parte se centra en un estudio sobre la activación de la PCPB pancreática basado en algunos mutantes de la misma. La segunda y tercera partes contienen estudios estructurales sobre las carboxipeptidasas reguladoras CPD y CPE.

Las procarboxipeptidasas pancreáticas expresan su actividad una vez se ha eliminado su segmento pro, o segmento de activación, por un ataque proteolítico de la tripsina. Además, tanto en el caso de la PCPA1 como en el de la PCPB se ha demostrado la intervención de la propia enzima en la degradación de su segmento pro. La considerable longitud de este segmento pro (94-95 residuos) distingue a las procarboxipeptidasas de otros zimógenos digestivos como el tripsinógeno o la proelastasa, con propéptidos mucho más cortos. Esta característica diferencial sugiere que el segmento de activación juega un papel clave en la modulación de la activación de estos proenzimas. Por otro lado, es bastante probable que los prosegmentos grandes y estructurados que presentan algunas proteasas funcionen como chaperonas internas que guíen el plegamiento hacia la forma nativa de la parte activa de la enzima (Fabre et al., 1992; Eder et al., 1993).

En la primera parte de este trabajo, el uso conjunto de técnicas de expresión heteróloga de proteínas y mutagénesis dirigida ha permitido ahondar un poco más en las relaciones estructura-función que rigen los mecanismos de activación e inhibición de la PCPB de páncreas porcino. Es precisamente el segmento de activación de este proenzima, el que determina la falta de actividad del correspondiente zimógeno y la evolución de su proceso de activación y, por tanto, el principal objeto de este estudio. Dicho segmento de activación consta de dos partes estructuralemnte diferenciadas, el dominio de activación y el segmento de conexión. El dominio de activación se corresponde con la región globular, resistente a la proteolisis, que tapa el centro activo en la forma de proenzima. El segmento de conexión, estructurado en hélice α, conecta el dominio de activación con el enzima.

Para llevar a cabo el trabajo, se utilizó un sistema recombinante de expresión que se había desarrollado con anterioridad y que permitía la obtención de PCPB heteróloga en una conformación nativa y activable, tanto la forma silvestre como las mutantes. Este sistema es el de la levadura *Pichia pastoris*, que, utilizando el vector

adecuado, permite que la proteína heteróloga se secrete al medio de cultivo, facilitando así su purificación. Mediante el sistema de expresión de *P. pastoris* se ha obtenido cantidad suficiente de proteína silvestre y mutantes para el análisis de sus procesos de activación por diferentes técnicas. Este análisis ha intentado abordar tres cuestiones diferentes.

En primer lugar, se han mutado dianas trípticas del segmento de conexión para observar diferencias en la activación con respecto a la PCPB silvestre. El segmento de conexión contiene tres Arg en las posiciones 83, 93 y 95. Aunque todas ellas se sitúan en la superficie, su grado de exposición al solvente difirie, por lo que el primer corte tríptico del procesamiento proteolítico de PCPB ocurre sólo en Arg 95. Este corte inicial va seguido de la escisión del residuo C-terminal del prosegmento por la CPB generada. Un segundo corte tríptico tiene lugar en el enlace Arg 83-Ser 84. Estudios anteriores han demostrado que Arg 93 puede también actuar como diana tríptica primaria cuando se añaden inhibidores de carboxipeptidasa a la mezcla de activación (Villegas et al., 1995). En un trabajo más reciente (Ventura et al., 1999) se generaron diferentes mutantes para individualizar el efecto de las diferentes dianas trípticas. Se demostró que, aunque la Arg 95 es normalmente la diana inicial para el procesamiento proteolítico, Arg 93 puede también actuar como diana inicial en ausencia de Arg 95 bajo condiciones de activación fuertes. Cuando no están presentes ni Arg 93 ni Arg 95, no es posible la conversión de proenzima a enzima activo en ningún caso.

En este trabajo se han estudiado dos nuevos mutantes de las dianas trípticas: un mutante doble R83Q/R93Q y el mutante triple R93Q/R95Q/T97R. El doble mutante sólo deja intacta la diana tríptica primaria Arg 95 y su proceso de activación no varía con respecto a la proteína silvestre, lo cual nos indica que un solo corte tríptico es capaz de generar totalmente la actividad carboxipeptidasa. El comportamiento de este mutante es muy similar al del mutante sencillo R83Q analizado previamente. En ese caso previo, sin embargo, se observó una deceleración en la curva de activación visualizada gráficamente como una ligera bifasicidad (Ventura *et al*, 1999), la cual se interpretó como una capacidad del fragmento liberado de inhibir la actividad carboxipeptidasa. En el mutante doble R83Q/R93Q no se ha observado esta capacidad inhibidora del segmento liberado (1-94). Esta ligera diferencia de comportamiento debe ser atribuida a la mutación R93Q, que podría haber provocado una desestabilización del extremo C-terminal de la hélice 3 por la eliminación de cargas positivas que compensaban el dipolo eléctrico característico de las hélices α. La desestabilización de esta región dificultaría

su interacción con el dominio enzimático. La diferencia entre el mutante doble R83Q/R93Q se hace patente en el seguimiento por HPLC de los fragmentos generados por su activación. Debido a la ausencia de Arg 93, el fragmento 84-94, uno de los principales productos de degradación de la activación de la PCPB silvestre, no se genera en este mutante.

En el mutante triple se ha eliminado la diana primaria Arg 95 y la alternativa Arg 93 y se ha introducido una Arg en la posición 97, imitando la posición en que se encuentra la diana tríptica en PCPA1 y PCPA2 (Arg 99). Se pretende observar si sería posible la activación de la PCPB mediante un único corte tríptico en el lugar donde éste se produciría en la PCPA1 y 2. La respuesta es negativa. Podemos hacer alineamientos basados en homología con un alto nivel de identidad aminoacídica, pero ello no significa que todos los residuos que ocupan una determinada posición en el alineamiento realicen la misma función o se encuentren en un mismo entorno estructural. Por ejemplo, en este caso, tal y como muestran las estructuras tridimensionales de PCPB y PCPA2 (Coll *et al,* 1991; García-Sáez *et al,* 1997), la situación de estos residuos y su exposición al solvente es muy diferente: mientras Arg 99 de PCPA2 se encuentra al final de la hélice de conexión, expuesta a la acción de proteasas activadoras, la Arg 97 artificialmente introducida en PCPB queda en medio de un "loop" enterrada en el dominio enzimático, lugar difícilmente accesibles para la tripsina. De todo ello se infiere que la situación de las dianas de activación en las regiones de conexión de las PCPBs siguen un patrón individual para cada una de las formas de las proenzimas.

La segunda cuestión que se ha tratado en esta parte del trabajo es la relevancia del segmento de conexión en la activación tríptica de la PCPB. El segmento de conexión de PCPB presenta diferencias con respecto a su región correspondiente en PCPA1. Mientras en PCPB hay un puente de hidrógeno entre Gln 89 y Arg 124 de la enzima activa, los residuos equivalentes en PCPA1 forman un puente salino (Avilés et al., 1993). Por otro lado, la hélice más larga del segmento de conexión de PCPA1 parece ser responsable de mantener la interacción entre la región pro y la carboxipeptidasa después de la acción tríptica (Vendrell et al., 1993). En cambio, la región pro de CPB una vez escindida es incapaz de inhibir a la proteína. Esta falta de capacidad inhibidora puede deberse a que existen pocos contactos entre el segmento de conexión y la enzima activa, a la debilidad de un puente de hidrógeno frente al puente

salino que presenta PCPA1 y al hecho de que una hélice corta tienda a desestructurarse en ausencia de continuidad en la cadena polipeptídica.

Para probar la relevancia del segmento de conexión en las diferencias entre los procesos de activación de PCPB y PCPA1, se diseñó un mutante de PCPB en el que su segmento de conexión fue sustituido por el de la PCPA1, que es una hélice más larga y estructurada. La velocidad de activación de este mutante es más lenta que la de la PCPB silvestre. Debido a la ausencia de la diana tríptica secundaria R83, no hay conversión del fragmento F1(1-95, 94) a F2(1-83, 82) y F3(1-81), como ha podido observarse tanto electroforéticamente como por HPLC. La velocidad de activación del mutante triple es también más lenta que la de los mutantes anteriormente comentados R83Q y R83Q/R93Q, que carecían igualmente de la diana tríptica secundaria. Sin embargo, el comportamiento de este mutante no reproduce la capacidad inhibidora mostrada por el segmento de activación de PCPA1 ni la consecuente curva bifásica de la activación de este zimógeno. Por tanto, PCPB no es capaz de reproducir en proceso de activación de PCPA1 con la simple sustitución de su segmento de conexión. La acción inhibidora del segmento de activación liberado no depende sólo de su integridad covalente, sino también de su estructura, estabilidad y capacidad para unirse a superficies complementarias de la enzima activa. En el diseño de este mutante se intentó respetar la estructura helicoidal local del segmento de conexión, pero no se realizaron estudios previos de modelado molecular que tuvieran en cuenta las interacciones entre las superficies del segmento liberado y el centro activo de la enzima.

La tercera cuestión que se ha abordado en esta parte del trabajo referente a PCPB es averiguar cuáles son los determinantes de la ausencia de actividad intrínseca de PCPB en contraste con la PCPA1, que presenta algo de actividad sin necesidad de sufrir ningún corte tríptico. En la PCPB porcina, la parte de la región pro que queda sobre el centro activo está formada por la cadena β2, un giro cis-pro y una hélice 3_{10} con un aspartato en posición 41. Este aspartato forma un puente salino con la Arg 145 de la enzima, lo cual podría explicar en parte la ausencia de actividad intrínseca en PCPB frente a sustratos peptídicos pequeños (Avilés *et al*., 1993). En PCPA1, la actividad intrínseca es del 10% de la V_{max} para sustratos pequeños y es debida a un centro activo preformado en el zimógeno y a la capacidad de difusión de los sustratos para alcanzar el centro activo pasando a través de cavidades que quedan entre las superficies de los dominios (García-Sáez *et al*., 1997). Intuimos que la hélice $3_{10,}$ ausente en PCPA1, debía tener algún papel en esta diferencia entre PCPA1 y PCPB y por eso diseñamos un

mutante en el que esta estructura que presenta PCPB fue sustituida por la secuencia correspondiente por alineamiento de PCPA1. Simultáneamente, el puente salino existente en PCPB entre Asp 41 y la enzima fue eliminado.

Al medir la actividad intrínseca de este mutante, se ha observado que es sólo ligeramente mayor a la de PCPB pero queda muy lejos de la que presenta PCPA1. Por tanto, la hélice 3_{10} de PCPB tiene que ver con su ausencia de actividad intrínseca, pero deben intervenir además otros factores.

Sintetizando, en la primera parte de este trabajo se ha estudiado la activabilidad de la PCPB, la importancia de los contactos de su segmento de conexión con la enzima y la inhibición en la forma de proenzima. La conclusión general es que, aunque es posible mimetizar el comportamiento de una proteína de la misma familia mediante rediseño por mutagénesis dirigida, el contexto en el que se acomodan las mutaciones determina diferencias funcionales.

La segunda y tercera partes de esta Tesis Doctoral consisten en estudios estructurales de las carboxipeptidasas reguladoras CPD y CPE. Estos estudios son muy novedosos y tienen un gran interés por la importancia fisiológica de estas proteínas; requieren un trabajo previo de expresión heteróloga bastante arduo como consecuencia de la dificultad de obtener estas proteínas en cantidad y grado de pureza suficientes. El dominio 2 de CPD, CPE silvestre, Δ23 CPE y E270Q CPE se clonaron y expresaron en *Pichia pastoris*. Desafortunadamente, la única proteína que proporcionó niveles de expresión que permitieran la purificación en las cantidades necesarias para realizar estudios estructurales fue la CPD-2.

En una primera fase del proyecto, se ha resuelto la estructura cristalográfica del dominio 2 de la CPD de pato (al cual nos referiremos indistintamente como CPD-2), que ha supuesto un prototipo para la subfamilia de las metalocarboxipeptidasas reguladoras. Basándose en las estructuras cristalinas similares de las carboxipeptidasas pancreáticas CPA y CPB y con su pariente lejano de bacteria CPT (Teplyakov el al., 1992) era predecible que CPE, CPD y otras metalocarboxipeptidasas tendrían una estructura similar. Sin embargo, el alineamiento de las secuencias de varias carboxipeptidasas mostró algunas inserciones en CPD y CPE no presentes en CPA o CPB. Además, la CPD y todos los demás miembros de su subfamilia contienen una región extra de unos 75 residuos localizada en el extremo C-terminal. La estructura de esta región, así como de las inserciones no se podía predecir, lo cual hacía necesaria la

estructura cristalina de alguno de los miembros de la subfamilia de las carboxipeptidasas reguladoras.

El dominio 2 de CPD de pato pudo expresarse en grandes cantidades usando el sistema de expresión de *P. pastoris*, y la proteína purificada pudo concentrarse sin sufrir agregación.

La estructura de CPD-2 se divide claramente en dos subdominios: un subdominio carboxipeptidasa y un subdominio C-terminal que comentaremos más adelante. Como se había predicho, la estructura general del subdominio carboxipeptidasa de CPD-2 es similar a las estructuras de las carboxipeptidasas pancreáticas y bacterianas. Este subdominio contiene una lámina β de ocho cadenas rodeada por hélices α, común a las proteasas del tipo de plegamiento de α/β-hidrolasa. Todos los residuos catalíticamente esenciales en las carboxipeptidasas pancreáticas están presentes en CPD-2 en una posición espacialmente similar. Una excepción es la Arg 71 de CPA, que interacciona con el sustrato y en CPD-2 es funcionalmente sustituido por Lys 277. Otra diferencia radica en los residuos que unen al sustrato y que proporcionan especificidad para residuos aromáticos (CPA) o básicos (CPD, CPB). En las carboxipeptidasas pancreáticas, esta especificidad está proporcionada por el residuo 255, que es un Asp en CPB y una Leu o Ile en CPA (Folk, 1971). Sin embargo, el aminoácido en la posición similar secuencialmente en CPD-2 es Gln 257. El residuo que proporciona la especificidad de sustrato en CPD-2 es Asp 192, que se encuentra en la posición correspondiente a Ser 207 en CPB. De este modo, CPD-2 y CPB contienen un residuo polar (Gln 257 en CPD-2, Ser 207 en CPB) y un residuo ácido (Asp 192 en CPD-2, Asp 255 en CPB) en el sitio de unión a sustrato, lo cual confiere especificidad para aminoácidos C-terminales básicos.

Una diferencia importante entre la estructura de CPD-2 y la de las carboxipeptidasas pancreáticas está en los lazos que delimitan el túnel de acceso al centro activo. CPD-2 contiene un lazo de 16 residuos (Tyr 255-His 241) que no está presente en las carboxipeptidasas pancreáticas. Otro lazo de CPD-2 (Ser 124-Val 133) se encuentra doblado hacia adentro y cubre parcialmente la entrada al centro activo. CPD-2 ha sufrido una deleción de 14 residuos también en la zona de acceso al centro activo, lo cual acaba de alterar la estructura de esta zona. Las diferencias en los lazos que rodean este túnel pueden explicar parcialmente por qué CPD-2 no hidroliza rápidamente sustratos dipeptídicos como el benzoil-Gly-Arg, sino que prefiere sustratos de al menos tres aminoácidos (Song y Fricker, 1995).

La estructura cristalina de CPD-2 ha revelado que el subdominio C-terminal ausente en las carboxipeptidasas pancreáticas forma un barril β con 7 cadenas conectadas por lazos. Esta estructura es similar a la de la transtiretina, una proteína sérica que tranporta hormonas tiroideas a través del plasma. La función de este subdominio no se conoce aún, aunque podría estar relacionado con el transporte de moléculas pequeñas. El hecho de que este subdominio se encuentre conservado en las carboxipeptidasas reguladoras de especies muy distantes sugiere que es funcionalmente esencial para la proteína. Una posibilidad es que intervenga en el plegamiento. Se ha observado (Varlamov y Fricker, 1996) que la deleción de una porción pequeña del extremo C-terminal del subdominio de transtiretina elimina la actividad de la CPE y causa la degradación intracelular de la proteína, que es típico de proteínas mal plegadas.

En una segunda fase del proyecto relacionado con las carboxipeptidasas reguladoras, la estructura de CPD-2 sumada a otra estructura de la misma molécula acomplejada con un inhibidor sirven de base para el modelado de los dominios 1 y 3 de CPD de pato (CPD-1 y CPD-3) y de la CPE humana.

El inhibidor de carboxipeptidasa GEMSA es muy potente para las carboxipeptidasas reguladoras. La nueva estructura del complejo CPD-2/GEMSA muestra que el inhibidor está unido a residuos de CPD-2 esenciales para la unión a sustrato: Tyr 250, Arg 145 y Arg 135.

Los modelos de CPD-1, CPD-3 y CPE son en general bastante similares a CPD-2; todos cuentan con el subdominio de transtiretina. Las interacciones entre subdominios son principalmente hidrofóbicas en todos los modelos.

Las diferencias más importantes entre modelos y estructuras son locales y se deben a inserciones y deleciones. Por ejemplo, la CPE humana contiene una inserción de 23 residuos. Las carboxipeptidasas pancreáticas contienen también una inserción de 14 residuos en la misma región. La principal diferencia entre CPD-1 y CPD-2 es una inserción de 29 residuos en CPD-1 que contiene 20 cargas, cuyo significado se desconoce.

El estudio del modelo de CPD-3 es especialmente importante por su falta de actividad, probablemente debida a la ausencia de los residuos clave para la catálisis y para la unión a los sustratos de carboxipeptidasas. Sin embargo, el alineamiento de secuencias y la superposición de las estructuras tridimensionales muestra que otros residuos de función desconocida se encuentran altamente conservados. Se supone que esa función sería básicamente estructural.

Pese a que la maquinaria catalítica de este dominio no es funcional, su secuencia y estructura conservadas sugieren alguna otra función biológica, posiblemente relacionada con la unión a proteínas o a otras moléculas y su transporte intracelular.

Los lazos que forman el bolsillo de especificidad (subsitio S1') en CPD-2 tienen la misma longitud en todos los modelos y su identidad aminoacídica es alta para CPD-1 y baja para CPD-3. La Tyr 250, que facilita la ubicación del sustrato en el centro activo, se encuentra sustituida por una His en CPD-3, lo cual concuerda con la inactividad de este dominio.

Un residuo esencial para la especificidad de CPB por residuos C-terminales básicos es Asp 255 (Coll et al., 1991). En CPD-1, CPD-2 y CPE, el residuo en la posición secuencialmente equivalente a Asp 255 es una Gln, que es funcionalmente incapaz de realizar la misma función que el Asp. En lugar de ello, el carácter electronegativo requerido para la selección de Arg y Lys C-terminales viene proporcionado por Asp 192, que se encuentra cercano espacialmente. Este Asp 192 está conservado en todas las reguladoras, incluida CPD-3.

Una de las diferencias estructurales más significativas entre CPD-2 y las carboxipeptidasas pancreáticas es la inserción Tyr 225-His224 que forma el túnel que lleva hasta el centro activo y que impide la unión del PCI (inhibidor de carboxipeptidasa de patata) a la CPD. El PCI es un péptido de 39 residuos que inhibe fuertemente las carboxipeptidasas digestivas CPA y CPB.

Todas estas observaciones que se han realizado sugieren que las carboxipeptidasas reguladoras han desarrollado variaciones estructurales que les permiten seleccionar los sustratos de forma más sofisticada que la simple selección de residuos C-terminales de las carboxipeptidasas pancreáticas.

La información que ha proporcionado esta parte del trabajo podría facilitar la comprensión de los diferentes papeles biológicos de las carboxipeptidasas reguladoras y el diseño de inhibidores específicos para ellas. Como ejemplo, encontrar un inhibidor específico de la CPD de pato podría tener un gran interés económico e industrial en la lucha contra la hepatitis B en esta especie. Si se consiguiera inmovilizar a la CPD en la superficie celular, se obstaculizaría su reciclaje impidiendo así la internalización del virus.

En definitiva, los estudios estructurales proporcionan una herramienta básica para acercarnos al conocimiento de las propiedades y funciones de las carboxipeptidasas

reguladoras, que por su importante papel fisiológico y su especificidad pueden crear un interés creciente a medida que se vayan conociendo mejor.

Referencias

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. Eur J Biochem 211, 381-389.

Coll M, Guasch A, Avilés FX, Huber R (1991) Three-dimensional structure of porcine procarboxipeptidase B: a structural basis of its inactivity. EMBO J 10, 1-9.

Eder J, Rheinhecker M, Ferhst AR (1993) Folding of subtilisin BPN': characterization of a folding intermediate. Biochemistry 33, 18-26.

Fabre E, Tharaud C, Gaillardin C (1992) Intracellular transit of a yeast protease is rescued by trans-complementation with its prodomain. J Biol Chem 267, 15049-15055.

García-Sáez I, Reverter D, Vendrell J, Avilés FX, Coll M (1997) The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. EMBO J 16, 6906-6913.

Song L, Fricker LD (1995) Calcium- and pH-dependent aggregation of carboxypeptidase E. J Biol Chem 270, 25007.

Varlamov O, Fricker LD (1996) The C-terminal region of carboxypeptidase E involved in membrane-binding is distinct from the region involved with intracellular routing. J Biol Chem 271, 6077-6083.

Vendrell J, Catasús Ll, Opezzo O, Ventura S, Villegas V, Avilés FX (1993) Ed. Walter de Gruyter. Berlin, NY. Capítulo 16, 279-297.

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. J Biol Chem 274, 19925-19933.

Villegas V, Vendrell J, Avilés FX (1995) The activation pathway of procarboxypeptidase B from porcine pancreas: participation of the active enzyme in the proteolytic processing. Protein Science 4, 1792-1800.

Conclusiones

Estudio de los mecanismos de activación e inhibición de la PCPB mediante mutagénesis dirigida.

1. La disponibilidad de un sistema de expresión recombinante eficaz, el de la levadura *Pichia pastoris*, ha permitido obtener formas mutantes de la PCPB porcina en diferentes regiones del segmento de activación.

2. Un único corte tríptico en la Arg 95 de PCPB es capaz de provocar la activación completa de la proteína. La rotura del enlace Arg 95-Thr 1 es suficiente para generar actividad CPB sin necesidad de la existencia de otras dianas. Este hecho se ha comprobado con una proteína mutante a la que se han eliminado las dianas trípticas alternativas dejando sólo intacta la diana primaria.

3. La activación por proteólisis limitada de la PCPB pancreática es un proceso altamente específico. La ausencia de la diana tríptica primaria genera un zimógeno no activable, aún cuando se haya introducido en él una diana artificial emulando el lugar de corte en la PCPA2.

4. Se puede modificar la velocidad de activación de la PCPB cambiando su segmento de conexión por el de la PCPA1. Sin embargo, la proteína resultante no muestra la velocidad de activación ni la naturaleza bifásica de la PCPA1. De aquí se deduce que el segmento de conexión, aunque es muy importante en el proceso de activación, no es el único factor determinante en dicho proceso.

5. La ausencia de actividad intrínseca en la forma silvestre de PCPB es parcialmente originada por el impedimento estérico producido por la hélice 3_{10} presente en el segmento de conexión y al puente salino establecido entre un giro cis-pro que precede a dicha hélice y el centro activo de la enzima (Asp 41 y Arg 145). Sin embargo, el mutante en que se han sustituido estas dos estructuras por la secuencia existente en PCPA1 no muestra un nivel de actividad intrínseca tan elevado como el de esta proteína, lo cual nos indica que otros factores e interacciones deben de intervenir en este fenómeno.

Estructura tridimensional de CPD-2.

6. El dominio 2 de CPD de pato pudo expresarse en grandes cantidades usando el sistema de expresión de *Pichia pastoris* y la proteína purificada pudo concentrarse sin sufrir agregación. Ello contrasta con la imposibilidad de expresar la CPE, tanto en la forma silvestre como en forma de dos mutantes, E270Q CPE y Δ23 CPE. El primero de ellos debía facilitar la co-cristalización con el sustrato, mientras que el mutante con la delección C-terminal hipotéticamente debía mostrar una mayor solubilidad.

7. La estructura de CPD-2, un prototipo de la subfamilia de las metalocarboxipeptidasas reguladoras, está dividida claramente en dos subdominios: un subdominio carboxipeptidasa y un subdominio C-terminal que presenta homología con la proteína plasmática transtirretina.

8. El subdominio carboxipeptidasa, aunque presenta una estructura general similar a las carboxipeptidasas pancreáticas, posee características diferenciales que afectan a la accesibilidad del centro activo y podrían ser las responsables de la alta selectividad de sustratos peptídicos naturales.

9. La Arg 71 de CPA, que interacciona con el sustrato, se encuentra sustituida funcionalmente en CPD-2 por la distante Lys 277.

10. CPD-2 y CPB contienen un residuo polar (Gln 257 en CPD-2, Ser 207 en CPB) y un residuo ácido (Asp 192 en CPD-2, Asp 255 en CPB) en el sitio de unión a sustrato, lo cual confiere especificidad frente aminoácidos C-terminales básicos.

Estructura tridimensional de CPD-2/GEMSA y modelado de CPD-1, CPD-3 y CPE.

11. El inhibidor GEMSA se une a CPD-2 por los residuos esenciales para unión a sustrato y polarización: Tyr 250, Arg 145 y Arg 135.

12. Los modelos de CPD-1, CPD-3 y CPE son, a grandes rasgos, bastante similares a CPD-2. Todos cuentan con el subdominio de transtirretina.

13. La carencia de actividad que presenta CPD-3 es debida a la ausencia de los residuos clave para la catálisis y para la unión a los sustratos de carboxipeptidasas. Sin embargo, CPD-3 contiene otros residuos altamente conservados en la familia de las carboxipeptidasas, lo cual sugiere que una función de este dominio podría ser básicamente estructural. Adicionalmente, la sustitución de la tríada responsable de la unión a sustrato en carboxipeptidasas Asn-Arg-Asn por Asp-Thr-Asp podría indicar una posible función de unión a otras moléculas, por el momento desconocida.

14. Las carboxipeptidasas reguladoras no pueden ser inhibidas por el PCI debido a la presencia de la inserción Tyr 255-His 224, que bloquea en parte el túnel de acceso al centro activo en estas proteínas.

15. Las diferentes inserciones y delecciones observadas en el túnel de acceso al centro activo de las reguladoras sugieren que estas han desarrollado variaciones estructurales que les permiten seleccionar los sustratos de forma más sofisticada que las carboxipeptidasas pancreáticas. El procesamiento de hormonas y neurotransmisores en que intervienen algunas de las carboxipeptidasas reguladoras requiere más especificidad que la función digestiva que desempeñan las pancreáticas.

16. La obtención de la primera estructura tridimensional de una carboxipeptidasa reguladora ha permitido el modelado molecular de otras proteínas de misma familia. Con ello se abren las puertas al estudio de la relación estructura/función en la familia de las carboxipeptidasas reguladoras, proteínas que desempeñan un papel fisiológico relevante. Ello ha de permitir el diseño de inhibidores que modulen dichas actividades.