Capítulo 2.

Estructura cristalina del dominio 2 de la carboxipeptidasa D de pato: un prototipo para la subfamilia de metalocarboxipeptidasas reguladoras

2.1. Summary

The crystal structure of domain II of duck carboxypeptidase D, a prohormone/propeptide processing enzyme integrated in a three repeat tandem in the natural system, has been solved, constituting a prototype for members of the regulatory metallocarboxypeptidase subfamily. It displays a 300 residue N-terminal α/β-hydrolase subdomain with overall topological similarity to and general coincidence of the key catalytic residues with the archetypal pancreatic carboxypeptidase A. However, numerous significant insertions/deletions in segments forming the funnel-like access to the active site explain differences in specificity towards larger protein substrates or inhibitors. This α/β-hydrolase 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. The structure described here establishes the fundamentals for a better understanding of the mechanism ruling events such as prohormone processing and will enable modelling of regulatory carboxypeptidases as well as a more rational design of inhibitors of carboxypeptidase D.

2.2. Introduction

The metal dependent carboxypeptidase (CP) family currently contains ~25 members, which can be subdivided into two subfamilies, the digestive enzymes and the regulatory enzymes (Fricker, 1988; Skidgel, 1988, 1996; Rawlings and Barrett, 1995). The latter are involved in more selective processing reactions than the mere digestion of intake proteins. 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. Among the digestive CPs, pancreatic CPA1, CPA2, CPB, mast cell CPA and plasma CPB (also known as CPU)

have been described (Avilés *et al.*, 1993). These proteins have a CP domain of ~300 amino acids and are secreted as inactive zymogens containing a 90-95 amino acid N-terminal prosegment (Avilés *et al.*, 1993). The regulatory CPs include CPD, CPE, CPM, CPN, CPZ, and proteins designated CPX-1, CPX-2 and AEBP-1 (He *et al.*, 1995; Song and Fricker, 1995, 1997; Skidgel, 1996; Xin *et al.*, 1998; Lei *et al.*, 1999). These enzymes perform a variety of important cellular functions, including prohormone processing, regulation of peptide hormone activity and alteration of protein-protein or protein-cell interactions (Skidgel, 1988, 1996; Fricker, 1991). Knowledge about the sequence and function of these CPs has increased dramatically over the past two decades as a consequence of the research on biologically active peptides and the discovery that some of these hydrolases are involved in their processing (Naggert *et al.*, 1995; Fricker *et al.*, 1996).

All members of the regulatory subgroup have a conserved region of ~400 residues, suggesting the presence of a second smaller subdomain of ~100 amino acids following the 300 amino acid catalytic subdomain. The active site, metal binding and substrate binding residues of pancreatic CPA/CPB are generally conserved in most members of the regulatory group, suggesting a common fold for the catalytic subdomains. However, CPX-1, CPX-2, AEBP-1 and the third repeat of CPD lack several of these residues critical for proteolytic activity. They have been proposed to function as binding proteins rather than active enzymes.

CPD is a 180 kDa single-chain protein containing three tandem repeats of ~390 amino acids linked by short bridge regions, which are followed by a transmembrane domain and a short 60 residue sequence that forms the cytosolic tail (Kuroki *et al.*, 1995; Tan *et al.*, 1997; Xin *et al.*, 1997). The first two repeats are catalytically active and have ~40-50% amino acid identity with other members of the regulatory subfamily (Tan *et al.*, 1997; Xin *et al.*, 1997). CPD was initially described and characterized from cattle (Song and Fricker, 1995). The duck homolog, gp180, was identified by its ability to bind the preS envelope protein of duck hepatitis B virus particles (Kuroki *et al.*, 1995). Recently, the rat and human homologs were cloned and sequenced (Tan *et al.*, 1997; Xin *et al.*, 1997). A comparison of the human and rat enzymes with duck gp180 reveals 66, 83 and 82% sequence identity among the three repeats. The second repeat is the most conserved of the two catalytically active repeats and was therefore chosen as the subject of this study. The third repeat is also highly conserved, implying that, albeit being probably catalytically unfunctional, this region may play an important role, either in the secretory pathway or on the cell surface, such as binding an

endogenous preS-like factor (Eng *et al.*, 1998). Furthermore, a two-repeat *Drosophila* homolog has been found as the *silver* gene product, associated with embryogenesis (Settle *et al.*, 1995), and a four-repeat form has been described for *Aplysia* (Fan *et al.*, 1999).

CPD removes C-terminal basic residues and prefers an alanine in the penultimate position. The majority of proteolytic activity has been reported to be membrane-bound in bovine and is only extracted in appreciable amounts using detergent. The optimal pH is 5.5-6.5 and the activity is stimulated by cobalt ions and inhibited by zinc chelating agents (Song and Fricker, 1995). The wide distribution of CPD in virtually all tissues examined (Song and Fricker, 1995, 1996; Tan *et al.*, 1997), and the enrichment in the *trans*-Golgi network and immature secretory vesicles, suggest a role for this enzyme in the processing of many prohormones/proproteins that transit either the regulated or constitutive secretory pathways by removing basic C-terminal residues following the action of furin and related endopeptidases. In addition, CPD could potentially function in endosomes, cleaving peptide hormones such as bradykinin and epidermal growth factor after receptor-mediated endocytosis (Tan *et al.*, 1997). The re-uptake of CPD from the cell surface may furthermore contribute to the pathogenesis of hepatitis B in the duck (Eng *et al.*, 1998). Given that CPD seems to be the primary cellular receptor for viral infection with hepatitis B viruses (Breiner *et al.*, 1998; Urban *et al.*, 1999), its structural knowledge could facilitate the design of antivirals.

Although crystal structures of metallocarboxypeptidases are available for both pancreatic CPAs and CPBs and their proforms (Rees *et al.*, 1983; Coll *et al.*, 1991; Guasch *et al.*, 1992; Gomis-Rüth *et al.*, 1995; García-Sáez *et al.*, 1997), for CPT from *Thermoactinomyces vulgaris* (Teplyakov *et al.*, 1992), CPG2 from *Pseudomonas* (Rowsell *et al.*, 1997) and *Streptomyces albus* muramoyl-peptapeptide CP (Dideberg *et al.*, 1982), no structure has been published for a member of the regulatory group. The low degree of amino acid sequence identity and the presence of a large number of insertions/deletions (see figure 2.1) observed when pairing the sequences of the latter to those of other proteins of known 3-dimensional structures does not permit extraction of accurate conclusions for the structural determinants of the cleavage mechanism and inhibition of members of the regulatory group or accurate modelling. Furthermore, the folding of the C-terminal subdomain within each repeat is unknown. In order to address these issues we have undertaken the three-dimensional structure analysis of gp180 carboxypeptidase D domain II (CPD-2) and describe here the main structural features.

2.3. Results and discussion

For the structural studies, CPD-2 was expressed in a *Pichia pastoris* system, which had previously been used to express large amounts of digestive CPs (Reverter *et al.*, 1998; Ventura *et al.*, 1999), using a similar construct to that previously employed for expression in the baculovirus system (Eng *et al.*, 1998). The *P. pastoris* system provided an efficient system for the production of tens of milligrams of CPD-2 protein per litre of culture, and rendered an enzyme with catalytic properties close to those of the full-length protein.

2.3.1. Overall structure of CPD-2

The CPD-2 polypeptide chain is defined from residue Gln4 to Thr383 (for numbering, see figure 2.1 and Materials and methods) and is folded into two distinct subdomains, the catalytic CP subdomain (Gln4-His303) and the C-terminal subdomain (CTSD; Arg304-Thr383). It is worth mentioning that the sequence of the C-terminal subdomain is highly conserved among all known regulatory carboxypeptidases and is thus present in the three duck CPD repeats.

The former, of approximate overall dimensions 50 × 50 × 40 Å, is formed by a doubly wound eight-stranded β-sheet of mixed topology (figure 2.2). The central sheet (strands I-VIII) is flanked on both sides by three and six helices, respectively (figure 2.2). This folding topology is in accordance with the α/β-hydrolase fold (Ollis *et al.*, 1992) or PLEES-protein family fold (Puente and López-Otín, 1997) already observed in other metallocarboxypeptidases (see below).

The CTSD consists of 80 amino acids and displays a rod-like shape with approximate overall dimensions 25 × 25 × 40 Å, with its N- and C-terminus located on opposite sides of the rod (figures 2.2A and 2.3A). It forms a β-barrel or β-sandwich of prealbumin-like folding topology made up by seven strands (strands IX-XV) connected by short loops, with the first four strands displaying a Greek-key topology. The β-barrel can be interpreted as composed of two parallel sheets, a three-stranded one comprising strands IX, XII and XV (see figure 2.2B),

and a four-stranded one encompassing strands X, XI, XIII and XIV. The structure of this subdomain is maintained by a buried hydrophobic cluster running across the whole CTSD made up by the side chains of Ile306, Val310, Ile319, Ala322, Ile324, Val326, Ala327, Ile329, Val333, Thr335, Tyr341, Leu345, Tyr350, Val352, Ala354, Tyr359, Val362, Val366, Val368, Val374, Val376, Phe370 and Leu380 (figure 2.3A). CTSD shares topological similarity and connectivity with transthyretin (prealbumin; PDB access code 1etb; Blake and Oatley, 1977) over 78 backbone C_α atoms (r.m.s.d. 2.4 Å) despite low sequence similarity (18% identity) and despite lacking an eighth strand sheet present in the latter structure. Transthyretin, one of the most strongly conserved plasma proteins, is associated with several forms of amyloidosis (Blake and Oatley, 1977). It binds to and distributes thyroid hormones and also forms a complex with retinol-binding proteins. Further structural relatedness of CTSD due to superimposable β-strands and connectivity is found with a 70 residue subdomain (r.m.s.d. 2.5 Å) of cyclodextrin glucanotransferase (Harata *et al.*, 1996; PDB access code 1pam), a 71 amino acid (r.m.s.d. 2.6 Å) glucoamylase fragment (Sorimachi *et al.*, 1997; PDB access code 1kum), and a subdomain of protocatechuate 3,4-dioxygenase (Orville *et al.*, 1997; PDB access code 3pch), despite some larger loop insertions in the latter (76 topologically equivalent C_α atoms, r.m.s.d. 2.4 Å).

Three N-linked glycosylation sites have been found in CPD-2 at Asn136, Asn321 and Asn377 (see figures 2.2 and 2.3A), in accordance with the sequence-based prediction for this domain in the duck (Kuroki *et al.*, 1995). For the human homologue the same sites have been proposed, as well as a further unique one (Tan *et al.*, 1997). All sites are surface located, freely exposed to solvent and not involved in inter- or intra-subdomain stabilization. One site has been localized in the connecting segment between helices D and E of the CP subdomain, not far away from the active site cleft. Two more sites are observed in the CTSD, on opposite sides of the longitudinal surface of the rod.

Three sulfate anions have been assigned in the CP subdomain based on the electron density. Their presence is chemically reasonable due to the excess of this anion in the crystallization solution. $SO_4$998 is located in the active site cleft (figure 2.2A) as in carboxypeptidase T (CPT; Teplyakov *et al.*, 1992). $SO_4$997 is placed on the surface in the vicinity of the N-terminus and is coordinated to both Arg292 and Arg293 side chains. Finally, $SO_4$996 is charge compensated by Lys223 Nζ and two solvent molecules (Wat642 and Wat644). The only disulfide bridge (out of a total of four cysteine residues present in the

structure) is observed in the catalytic moiety in the neighbourhood of this latter sulfate anion, $SO_4$996, and is established between Cys230 and Cys275. It connects loop αGβVII with loop βVIIIαI contributing to anchor the chain segment Tyr225-His241, a surface-located loop unique to regulatory CPs (see below), to the main subdomain body. Cys230 and Cys275 are conserved in all members of the regulatory CP subfamily.

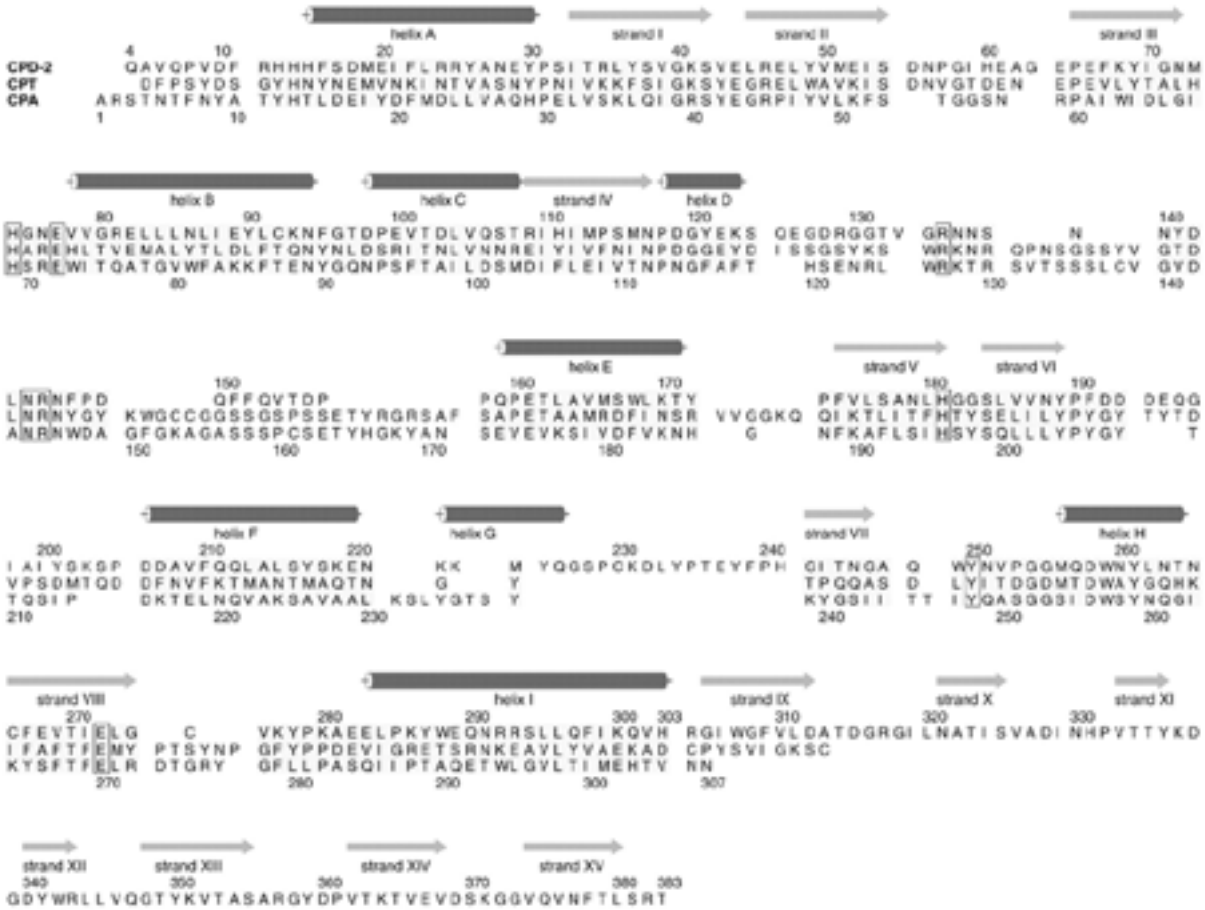

Figure 2.1. Sequence alignment based on topological equivalence of CPD-2, CPT (PDB access code 1obr) and CPA (PDB access code 6cpa). Equivalent regions in all three enzymes are displayed with a grey background. Conserved residues key for catalysis and general substrate binding are boxed. The regular secondary structure elements (helices and strands) shown correspond to the CPD-2 structure, as does the upper numbering. Numbering of CPA residues is presented in the lower line.

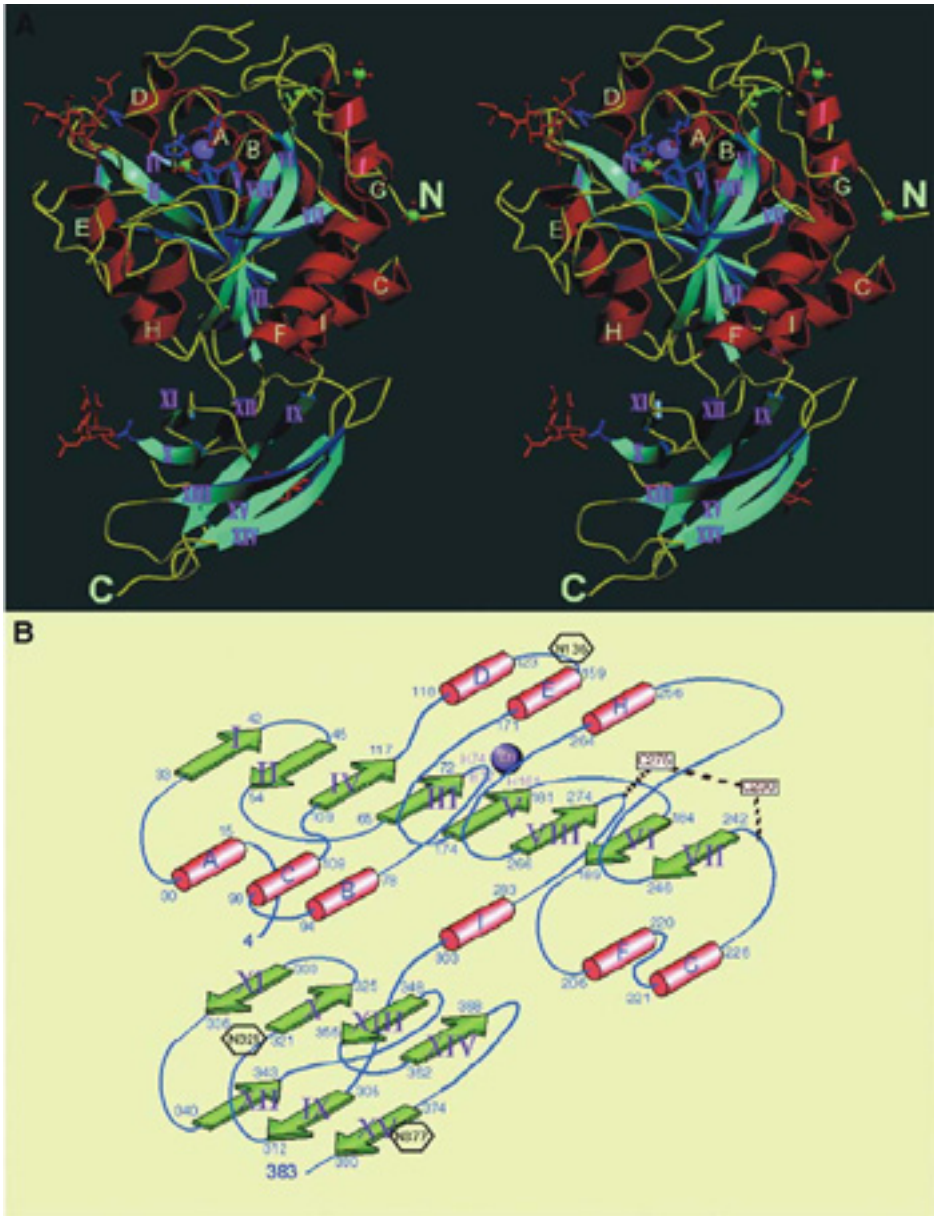

Figure 2.2. (A) Stereo ribbon cartoon of CPD-2. The termini are labelled, as are the helices (a-I) and strands (I-XV). The zinc ion, the zinc-coordinating protein site chains, the N-linked glycosylation sites, the disulfide bond and the sulfate anions are also displayed. (B) Topology scheme of the CPD-2 polipeptide fold, with α-helices indicated by cylinders and β-strands by arrows.

2.3.2. Subdomain interactions

CTSD is connected to the CP subdomain via helix I, which belongs to the latter and is bent for ~40° along its sequence, and is attached to the N-edge of the central β-sheet of the CP subdomain with the β-barrel forming strands almost perpendicular to the former (figure 2.2A). The interactions between both subdomains are mainly of a hydrophobic nature, burying a surface of 827 Å (calculated employing a probe radius of 1.6 Å). A total of 48 van der Waals interactions (<4 Å), nine hydrogen bonds and one (double) salt bridge (Asp206-Arg343) have been observed. The main segments implicated are the loops or connecting segments βIIβIII, αEβV, βVIαF (and the subsequent beginning of helix αF), αHβVIII (and the end of helix αH) of the CP subdomain, and strands βIX, βXI, βXII and βXV and loops βXβXI and βXIβXII of CTSD.

2.3.3. The active site

The active site cleft is placed at the C-edge of the central part of the β-sheet of the catalytic moiety. The catalytic zinc ion is pentahedrally coordinated by His74 Nδ1, His181 Nδ1, both Glu77 Oε1 and Oε2, and solvent molecule Wat601 (see figure 2.3B), which in turn is coordinated to the general base Glu272 (Wat601-Glu272 Oε1, 2.94 Å). In the absence of a substrate, a sulfate anion ($SO_4$998) occupies the position of the C-terminal carboxylate group of a substrate. Based on previous studies performed with CPs and their zymogens (Christianson and Lipscomb, 1989; Coll *et al.*, 1991; Kim and Lipscomb, 1991; Avilés *et al.*, 1993; García-Sáez *et al.*, 1997), the residues essential for catalysis have been identified.

The S_1' subsite (nomenclature according to Schechter and Berger, 1967), which forms the enzyme specificity pocket and the C-terminal anchoring of the substrate, is designed to accommodate positively charged bulky residues at the substrate C-terminus in a fashion similar to CPB (Coll *et al.*, 1991). This pocket is lined by residues Gly246-Gln257, Asn188-Asp192 and Phe267-Thr270. The electronegative character of this pocket is provided by Asp192, whose side chain points towards the interior of the subsite. This key residue is vicinal to the only *cis*-peptide bond found in the structure (Pro190-Phe191), which may adopt this unusual conformation to keep the acidic side chain in its appropriate position. This feature is

favoured by the establishment of an inter-main chain hydrogen bond between Phe191 N and Ser202 O and a further one between Pro190 O and Ser204 Oγ, both incompatible with a *trans*-peptide bond between Pro190 and Phe191. The specificity pocket is further closed by the aromatic ring of Tyr250 (equivalent to Tyr248 in CPA), fixed in the `down' conformation due to the presence of the sulfate anion as observed in other CPs with an `occupied' pocket (Teplyakov *et al.*, 1992; García-Sáez *et al.*, 1997), and by Asn144, whose Nδ2 atom would additionally coordinate a C-terminal carboxylate oxygen from a substrate. This latter group would be further anchored to the protein moiety through the side chains of Arg145 and Arg135, which in our structure coordinate the sulfate anion. The latter arginine (equivalent to Arg127 in CPA) would also be in charge of polarization of the scissile carbonyl and subsequent stabilization of the transition state following a general base mechanism (Christianson and Lipscomb, 1989).

The S_1 subsite is shaped by Gly182-Ser184, Glu272 (Glu270 in CPA), Tyr250 (which is probably involved in hydrogen-bonding of the P_1 amide nitrogen through its OH group), and notably by Trp249, which strongly restricts the size of this site and explains the enzyme preference for an alanine in P_1. The S_2 and S_3 subsites involved in substrate binding and torsion are constituted by Arg135, Arg145, Asp142, Lys277 (putatively involved in P_2 carbonyl oxygen binding) and Gly130, Gly131, Tyr234, respectively.

The formation of the `oxyanion hole' for the scissile carbonyl oxygen might be promoted by Asp142, hydrogen-bonded to the zinc ligand His74 side chain and salt-bridged to Arg135, as pointed out previously for CPT and digestive CPs (Phillips *et al.*, 1990; Teplyakov *et al.*, 1992).

2.3.4. Detailed comparison of CPD-2 catalytic subdomain with other proteases

The CPD-2 catalytic moiety shares the central twisted eight-stranded β-sheet flanked by α-helices with other proteases displaying the α/β-hydrolase fold (Ollis *et al.*, 1992; Puente and López-Otín, 1997) that belong to the cysteine-, serine- and metalloprotease families. In many of them, the topology and strand connectivity is the same for the first six strands, all of them parallel except strand II. The active sites, despite differences in the nature of their

constituting residues and hydrolytic mechanisms, reside in equivalent loci on the C-terminal site of the β-sheet and close to the end of strand V (see figure 2.2B). Interestingly, most of them are exopeptidases chopping off either N- or C-terminal residues from substrates. Key residues for catalysis are provided by loops connecting the regular secondary structure elements. The orientation of the last two β-strands enables a further subdivision of these α/β-hydrolases: (i) parallel to the previous four in prolyl iminopeptidase (Medrano *et al.*, 1998), prolyl oligopeptidase (Fülöp *et al.*, 1998), wheat and yeast serine carboxypeptidase (Liao *et al.*, 1992; Endrizzi *et al.* , 1994; these two display two additional strands at the outer end of the sheet, inserted between strands VII and VIII; see figure 2.2B); and (ii) antiparallel in CPD, CPA/CPB (Rees *et al.*, 1983; Coll *et al.*, 1991; Guasch *et al.*, 1992; García-Sáez *et al.*, 1997), CPT (Teplyakov *et al.*, 1992), pyrrolidone carboxypeptidase (Singleton *et al.*, 1999; this structure lacking the first two strands), *Aeromonas proteolytica* aminopeptidase (Chevrier *et al.*, 1994), *Streptomyces griseus* aminopeptidase (Greenblatt *et al.*, 1997), leucine aminopeptidase (Burley *et al.*, 1990) and carboxypeptidase G2 (Rowsell *et al.*, 1997).

CPD-2 catalytic subdomain displays closest topological similarity to both CPT and CPA/CPB. The main chain atoms of all three exopeptidases can be superimposed fairly well over most parts of the structures as denoted by r.m.s.ds of 1.6 Å over 264 topologically equivalent C atoms of CPD-2 and CPA deviating less than 3.5 Å, of 1.4 Å superimposing CPD-2 and CPT (261 common C_α atoms), and of 1.5 Å comparing CPD-2 and CPB (267 common C_α atoms). All regular secondary structure elements of CPD-2, except helix G, lie in regions topologically equivalent to CPA and CPT (figure 2.1). This similarity is nonetheless lower than that observed between CPA and CPT (1.2 Å; 290 common C_α atoms). The common fold and equivalence in the position of the residues described to be crucial for catalysis (figures 2.1 and 2.4A, C) is further underlined by the presence of a *cis*-peptide bond at an equivalent position, Pro190-Phe191 in CPD-2, Pro213-Tyr214 in CPT and Pro205-Tyr206 in CPA/CPB. However, CPT displays three additional *cis*-peptide bonds, while CPA/CPB have two more (Rees *et al.*, 1983; Coll *et al.*, 1991; Teplyakov *et al.*, 1992). The most important differences between the catalytic subdomains can be observed in the chain segments that form the funnel-like access to the active site cleft. Segment Ser124-Val133 is folded inwards over the molecular body in CPD, partially covering the access to the active site (top centre of figure 2.4A and C). On another side of the rim of the active site, an important deletion is observed between Asp149 and Pro158 in the regulatory enzyme

(figure 2.1), equivalent to Ala149 and Ser172 in CPA (top, facing side in figure 2.4A and C). Furthermore, the adjacent loop (Ser131-Val139) in CPA is cross-connected to the region Ala149-Ser172, whereas no equivalent crossconnection is present in CPD. Importantly, these two regional differences discussed previously clearly distinguish CPD from other known CPs since the affected loops are almost perfectly superimposable when comparing pancreatic and bacterial CPs. The most significant structural difference is found, however, in the unique insertion of segment Tyr225-His241 at the beginning of strand VII, a structural feature further shaping the funnel border.

All these differences suggest that CPD-2 may have a distinct selectivity towards larger substrates. For instance, the potato carboxypeptidase inhibitor, which forms a tight complex with CPA, could not bind CPD-2 due to severe steric hindrance (see figure 2.4C). Another interesting point is that CPD-2 does not display a surface complementarity to a prosegment, as observed in the zymogens of CPA and CPB, consistent with the absence of such a propart in CPD. In CPA and CPB (Rees *et al.*, 1983; Coll *et al.*, 1991; Guasch *et al.*, 1992), the prosegment has been proposed to perform a cotranscriptional chaperone function required for proper folding of the whole zymogen (Ventura *et al.*, 1999). In contrast to the digestive CPs, which must be stored before secretion in an environment where their activity is suppressed, regulatory CPs are produced directly in the appropriate tissues.

Although the residues responsible for catalysis are mainly conserved (figure 2.1), a closer inspection of the active sites reveals that the P_1 carbonyl oxygen stabilizing Arg71 (in CPA and CPT) is functionally replaced in CPD by Lys277, in approximately the same place as Phe279 in CPA/T (the corresponding C_α atoms are just 2 Å away), and whose $N\zeta$ atom is directed in our unliganded structure to meet a hypothetical P_1 carbonyl oxygen. Further differences arise when this detailed inspection is extended to the specificity pocket. In CPB, which displays a preference for basic C-termini similar to CPD, the electronegative character is provided by Asp255 (CPB numbering; Coll *et al.*, 1991; PDB access code 1nsa), at a position equivalent to Gln257 in CPD. In contrast, this role is carried out by Asp192 in the latter (Ser207 in CPB), despite the almost perfectly superimposable peptide chain folding in this region. The size of the pocket is also distinct: Thr194, Leu203, Gly243, Tyr248, Ala250, Gly253, Ser254, Asp255 and Thr268 render in CPB a pocket somewhat larger than the topologically equivalent Asn179, Asn188, Gly246, Tyr250, Val252, Gly255, Met256, Gln257 and Thr270 in CPD-2.

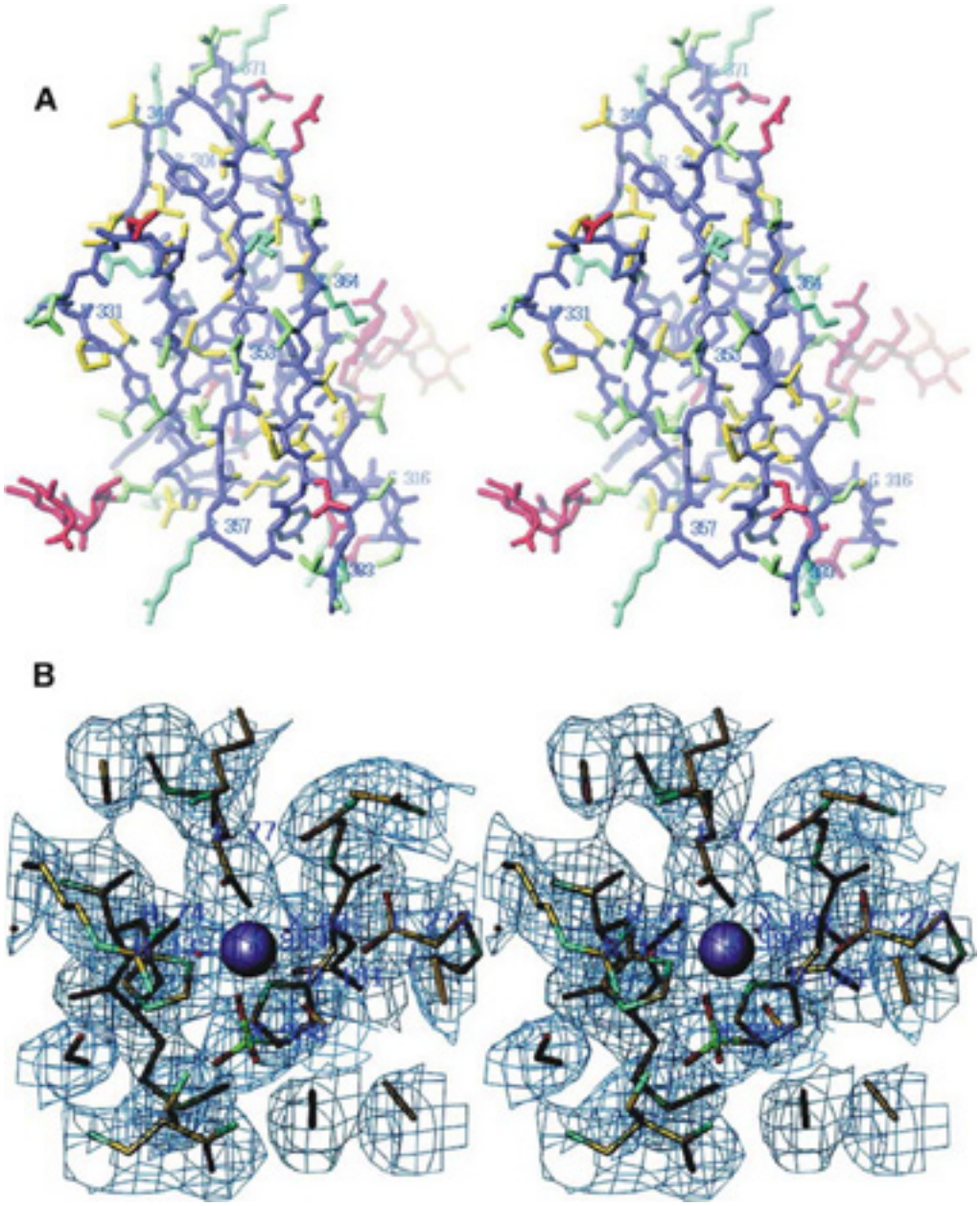

Figure 2.3. (A) Stereo drawing displaying the 80 residue C-terminal subdomain of CPD-2 as a stick model. Main chain atoms are coloured in violet, acidic residues in red, basic ones in cyan, hydrophobic in yellow and aromatic ones in violet. The two glycosylation sites present in this subdomain are also displayed in red. Some selected residues are labelled. (B) Stereo plot of the final ($2F_{obs}$-F_{calc}) electron density (contoured at 1σ) superimposed with the final coordinates drawn as atom-coloured sticks around the zinc binding site.

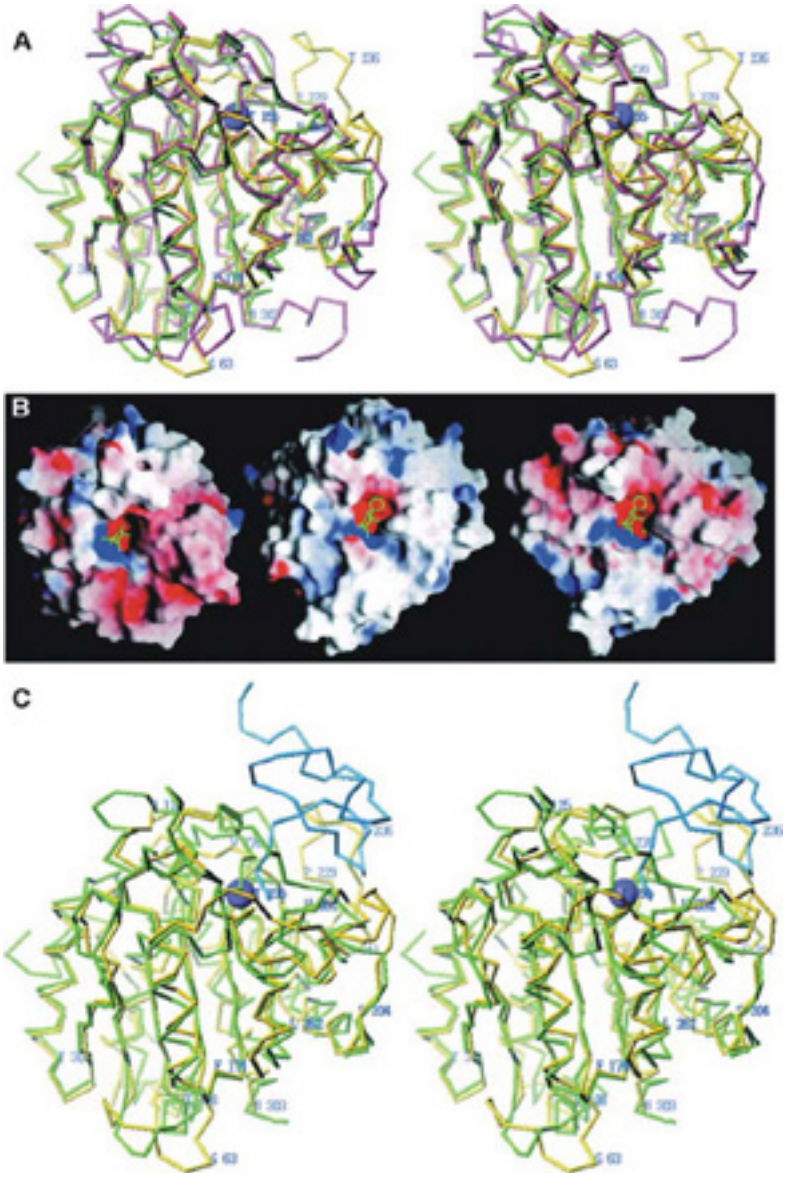

Figure 2.4. (A) Stereo view of the overlaid C_α-carbon structures of CPD-2 (CP subdomain only; yellow sticks), CPA (green sticks) and CPT (magenta sticks). The zinc ion (violet sphere) belongs to the CPD-2 structure. Selected CPD-2 residues are labelled. (B) View facing the funnel surrounding the active site cleft superimposed with its solid Connolly surface displaying the electrostatic potential [ranging from –15 k_BT/e (red) to +15 k_BT/e (blue)] of CPD-2 (left), CPA (centre) and CPT (right). The modelled substrate (CPD-2) and the coordinates (PDB access code 6cpa) of a phosphonate inhibitor (CPA and CPT) are also displayed respectively, to highlight the active site. (C) Stereo view of the C_α-carbon structures of CPD-2 (CP subdomain only; yellow sticks) and CPA (green sticks) in its complex with potato carboxypeptidase inhibitor (blue sticks) (PDB acces code 4cpa). Steric hindrance prevents the latter from binding the regulatory CP. The zinc ion (violet sphere) belongs to the CPD-2 structure. Some CPD-2 residues are labelled.

2.4. Conclusions

The structure of CPD-2, a prototype for the regulatory metallocarboxypeptidase subfamily, reveals that its CP subdomain, despite overall structural similarity to pancreatic CPs, possesses distinct features that affect the access to the active site cleft and might be responsible for a sharp selectivity on natural peptidic substrates. The P_1 carbonyl oxygen stabilizing residue Arg71 from the pancreatic forms is replaced by an asparagine in CPD-2, and the function of the former is probably performed by Lys277. The structure described here provides an important tool for the computational-based modelling of other regulatory carboxypeptidases (unfeasible until now) and for the development of CPD inhibitors for a variety of applications including antivirals and antibiotics. The 80 residue C-terminal subdomain, whose function has still to be fully confirmed, has been unveiled to display a seven-stranded β-sandwich topology.

2.5. Materials and methods

2.5.1. Cloning, expression and purification of CPD

To obtain large amounts of duck CPD domain II, the high level eukaryotic *P. pastoris* expression system (Invitrogen) was chosen as previously described for human proCPA2 (García-Sáez *et al.*, 1997; Reverter *et al.*, 1998). Briefly, a 1199 bp *Sna*BI-*Stu*I cDNA fragment encoding most of duck CPD domain II was inserted into the *Sna*BI site of pPIC9 *P. pastoris* expression system. Then a 144 bp PCR *Xho*I-*Sna*BI fragment, containing the 5' end of CPD domain II and vector sequences to encode an α-factor KEX2 cleavage site, was subcloned into the *Xho*I-*Sna*BI site. The CPD domain II sequence was confirmed by the DNA sequence facility of the Albert Einstein College of Medicine. The protein was expressed in the His^+Mut^s *P. pastoris* strain KM71. The protein produced was detected by an antiserum against full-length duck CPD using western blot analysis. Purification was achieved by adjusting the expression media to pH 5.5 with 0.1 M acetic acid and applying onto a *p*-aminobenzoyl-Arg Sepharose 6B substrate affinity column as previously described (Song and Fricker, 1995).

CPD-2 was eluted with 50 mM Tris-HCl pH 8.0, 100 mM NaCl and 0.01% CHAPS containing 25 mM arginine. The examination of carboxypeptidase enzymatic activity of the purified protein was performed as previously reported (Song and Fricker, 1995).

2.5.2. Crystallization and data collection

Cubic crystals ($P2_13$; a = 135.54 Å) were obtained at 20°C using the sitting-drop vapour diffusion method with Cryschem dishes (Charles Supper, MA, USA) from drops consisting of equal volumes of protein solution (10 mg/ml in 5 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.5 mM benzamidine) and a 2.1 M ammonium sulfate, 0.1 M sodium acetate pH 5.2 precipitating solution. Drops were allowed to equilibrate against 300 µl of the latter. Crystals contain a protein monomer per asymmetric unit (V_M = 4.6 Å^3/Da; Matthews, 1968; 73 % solvent content) and were harvested with 2.5 M ammonium sulfate, 0.1 M sodium acetate pH 5.2 to permit manipulation. An isomorphous mercury derivative (a = 136.14 Å) was prepared soaking native crystals for 23 h in harvesting buffer further containing 10 mM of *o*-phenanthroline in order to extract the catalytic zinc-ion and then for an additional 32 h in harvesting buffer containing 0.5 mM mercury acetate. A cryoprotecting protocol was worked out consisting of successive equilibration of the crystals for 10-15 min against harvesting solutions with increasing glycerol concentrations (5-20%). Diffraction data were collected from 100K-cryocooled crystals at EMBL beamlines X11 and BW7B, DESY (Hamburg), on MAR Research image plate area detectors. Data were processed with MOSFLM v. 6.0 (Leslie, 1991) and SCALA from the CCP4 suite (CCP4, 1994). Table 2.1 provides a summary of data collection and processing.

Table 2.1. Data collection, processing and phasing, and final model refinement

	Native	Mercury acetate derivative
Wavelength (Å)	0.8467	0.9114
Temperature	N_2 cryocooling	N_2 cryocooling
No. of measurements	270 100	432 567
No. of unique reflections	22 564	20 929
Statistics[a] whole resolution range (Å)	20.00-2.70	19.65-2.80
Completeness(%)/R_{merge}[b]	97.8/7.6	99.7/10.6
<I>/<σ(I)>/Average multiplicity	8.3/12.0	6.4/20.7
Anomalous R_{merge}/Anomalous fraction	-	4.7/97.1
Statistics[a] last resolution shell (Å)	2.85-2.70	2.95-2.80
Completeness(%)/R_{merge}[b]	86.6/60.8	99.7/68.7
<I>/<σ(I)>/Average multiplicity	1.2/6.0	1.1/16.2
Anomalous R_{merge}/Anomalous fraction	-	20.5/96.1
B-value from Wilson-plot (Å^2)	83.1	76.5
R_{Deriv}[c] (20-2.8 Å)	-	0.29
Refined Hg positions (20-2.8 Å)		
1 Fractional cell coordinates	-	0.115/0.057/0.438
Occupancy[d]/Anom. occupancy[d]/*B*-factor (Å^2)	-	0.463/0.438/54.1
2 Fractional cell coordinates	-	0.314/0.053/0.064
Occupancy[d]/Anom. occupancy[d]/*B*-factor (Å^2)	-	0.332/0.368/55.2
3 Fractional cell coordinates	-	0.037/0.259/0.166
Occupancy[d]/Anom. occupancy[d]/*B*-factor (Å^2)	-	0.206/0.322/37.1
Mean figure of merit (m)[e] (20-2.8 Å)	-	0.46
Phasing power[f] (centric/acentric)	-	1.16/1.41
R_{Cullis}[g] (centric/acentric/anomalous)	-	0.70/0.76/0.77
Resolution range used for refinement (Å)	20.0-2.70	
No. of reflections/completeness (%)	22 539/98.1	
Final crystallographic R_{factor}[h] (using all data)	0.198	
R_{factor}/free R_{factor}[h] (penultimate cycle)	0.203/0.236	

Non-hydrogen protein atoms	3057
Solvent molecules	130
Ions	1 Zn^{2+}, 3 SO_4
Glycosylation sites	Asn136, Asn321, Asn377
R.m.s. deviation from target values of	
bonds (Å)	0.011
angles (°)	1.59
bonded Bs main chain/side chain ($Å^2$)	1.46/2.19
Average (refined) occupancies/Average *B*-factors ($Å^2$)	
protein atoms	1.0/60.0
solvent molecules	1.0/55.4
sulfate anions	0.73/61.4
carbohydrate atoms	0.78/67.8
zinc cation	1.0/51.8

[a] For separated Friedel-mates in the mercury acetate derivative data.

[b] $R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$.

[c] $R_{Deriv} = \Sigma_{hkl} |F_P - F_{PH}| / \Sigma_{hkl} F_P$.

[d] On an arbitrary scale.

[e] $m = | F(hkl)_{best}| / |F(hkl)|$, with $\mathbf{F}(hkl)_{best} = \Sigma_\alpha P(\alpha)\, \mathbf{F}_{hkl}(\alpha) / \Sigma_\alpha P(\alpha)$.

[f] Phasing power = r.m.s. ($|F_H|/E$), with E the residual lack of closure.

[g] $R_{Cullis} = \Sigma_{hkl}||F_{PH} \pm F_P| - F_{H(calc)}| / \Sigma_{hkl}|F_{PH} - F_H|$.

[h] $R_{factor} = \Sigma_{hkl}||F_{obs}| - k\,|F_{calc}|| / \Sigma_{hkl}|F_{obs}|$; free R_{factor}, same for a test set of 1546 (7%) reflections not used during refinement (until last-but-one cycle).

2.5.3. Structure solution and refinement

The structure was solved by the SIRAS method using a single mercury derivative and anomalous diffraction. Density modification procedures were particularly effective due to the high solvent content of the crystals. Difference and anomalous Patterson maps and, posteriorly, difference-Fourier synthesis permitted the localization of three mercury sites (see Table I). These positions were refined with MLPHARE (CCP4, 1994) and phases were computed, rendering a mean figure of merit of 0.46 (20.0-2.8 Å). Posterior density modification (Cowtan and Main, 1996) considerably increased this value to 0.79 (20.0-3.0 Å). In parallel, molecular replacement calculations with AMoRe (Navaza, 1994) using the (appropriately truncated) coordinates of carboxypeptidase T (Teplyakov *et al.*, 1992, PDB access code 1obr) rendered a clear solution using data in the 15-4.5 Å range that additionally confirmed $P2_13$ as the correct space group ($\alpha = 68.6°$, $\beta = 28.8°$, $\gamma = 333.7°$, $x = 0.0461$, $y = 0.3214$, $z = 0.1170$, correlation coefficient = 29.4, $R_{factor} = 50.7$; 2nd highest solution: correlation coefficient = 21.0, $R_{factor} = 53.5$; α, β, γ are in Eulerian angles, x, y, z are in fractional cell coordinates). However, this correctly oriented and translated model was just used for assistance in chain tracing and not for phasing in order to completely omit any possible model bias. The sigma A-weighted map computed after density modification calculations was of excellent quality and permitted straightforward chain tracing of the whole molecule. Successive cycles of positional and temperature factor refinement performed with CNS v. 0.4 (Brünger *et al.*, 1998) applying bulk-solvent correction and anisotropic *B*-factor correction, and manual model building on a Silicon Graphics workstation using Turbo-Frodo (Roussel and Cambilleau, 1989), permitted gradual completion of the model and localization of solvent molecules and ions. Table I summarizes the refinement. The final model comprises residues Gln4 to Thr383 of the chemical sequence, 130 solvent molecules (labelled Wat601-Wat730), one zinc cation (residue Zn999) and three sulfate anions with (refined) partial occupancy ($SO_4$996-$SO_4$998). Three asparagine residues are glycosylated (Asn136, Asn321 and Asn377). For the first and last cases, two *N*-acetylglucosamide and one mannose residues (labelled Nag901-Nag902-Man903 and Nag921-Nag922-Man923, respectively), for the second, just two *N*-acetylglucosamide residues (Nag911-Nag912) have been traced. Standard β-1,4 links have been assumed and (grouped) occupancies have also been refined for each carbohydrate moiety. All protein residues are clearly defined by electron density and are placed in the most favourable and additionally allowed regions of the Ramachandran plot,

except Ala327, placed in a generously allowed region. One peptide bond (Pro190-Phe191) has been found in *cis* conformation. Two out of four cysteine amino acids join to establish a disulfide bond (Cys230-Cys275). Table I provides a summary of the final refined model parameters. As can be observed, the extremely high solvent content of the cell (73%) taken together with the bulk-solvent correction performed render above-average temperature factors for the whole structure, in accordance with the high temperature factor obtained from Wilson scaling (see Table I). Nonetheless, all structural features and parameters are correct and the final electron density is of excellent quality.

2.5.4. Miscellaneous

Numbering of CPD-2 residues is sequential with Gln4 corresponding to Gln503 of the published sequence of duck CPD (Kuroki *et al.*, 1995). Residue numbers of other proteins correspond to their respective PDB entries. Figures have been calculated with Turbo-Frodo (Roussel and Cambilleau, 1989), GRASP (Nicholls *et al.*, 1993), ALSCRIPT (Barton, 1993) and Bobscript (Esnouf, 1997). Superpositions were performed using Turbo-Frodo. Ascription of an amino acid to a regular secondary structure element (minimum four residues) was based on participation of at least one of its N or O main chain atoms in the hydrogen bond network characterizing the secondary structure element. Searches for protein structures with similar folding topology were made with the DALI (Holm and Sander, 1993) server at EBI in Hinxton (http://www.embl-heidelberg.de/dali/dali.html), the SCOP server (http://scop.mrc-lmb.cam.ac.uk/scop/), and the CATH server (http://www.biochem.ucl.ac.uk/bsm/cath/server/index.html). The final coordinates have been deposited with the Protein Data Bank (access code qmu).

2.6. References

Avilés 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.

Barton GJ (1993) ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng 6, 37-40.

Blake CCF, Oatley SJ (1977) Protein-DNA and protein-hormone interactions in prealbumin: a model of the thyroid hormone nuclear receptor? Nature 268, 115-120.

Breiner KM, Urban S, Schaller H (1998) Carboxypeptidase D (gp180), a Golgi-resident protein, functions in the attachment and entry of avian hepatitis B viruses. J Virol 72, 8098-8104.

Brünger AT *et al* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D 54, 905-921.

Burley SK, David PR, Taylor A, Lipscomb WN (1990) Molecular structure of leucine aminopeptidase and its complex with bestatin. Proc Natl Acad Sci USA 87, 6878-6882.

Chevrier B, Schalk C, D'Orchymont H, Rondeau JM, Moras D, Tarnus C (1994) Crystal structure of *Aeromonas proteolytica* aminopeptidase: a prototypical member of the co-catalytic zinc enzyme family. Structure 2, 283-291.

Christianson DW, Lipscomb WN (1989) Carboxypeptidase A. Acc Chem Res 22, 62-69.

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

Cowtan KD, Main P (1996) Phase combination and cross validation in iterated density-modification calculations. Acta Crystallogr D 52, 43-48.

Dideberg O, Charlier P, Dive G, Joris B, Frere JM, Ghuysen JM (1982) Structure of Zn^{2+}-containing D-alanyl-D-alanine-cleaving carboxypeptidase at 2.5 Å resolution. Nature 299, 469-470.

Endrizzi JA, Breddam K, Remington SJ (1994) 2.8 Å structure of yeast serine carboxypeptidase. Biochemistry 33, 11106-11120.

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

Esnouf RM (1997) An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. J Mol Graph 15, 133-138.

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

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.

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

Fülöp V, Böcskei Z, Polgár L (1998) Prolyl oligopeptidase: an unusual β-propeller domain regulates proteolysis. Cell 94, 161-170.

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.

Gomis-Rüth FX, Gómez M, Bode W, Huber R, Avilés FX (1995) The three-dimensional structure of the native ternary complex of bovine pancreatic procarboxypeptidase A with proproteinase E and chymotrypsinogen C. EMBO J 14, 4387-4394.

Greenblatt HM, Almog O, Maras B, Spungin-Bialik A, Barra D, Blumberg S, ShohamG (1997) *Streptomyces griseus* aminopeptidase: X-ray crystallographic structure at 1.75 Å resolution. J Mol Biol 265, 620-636.

Guasch A, Coll M, Avilés FX, Huber R (1992) Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. J Mol Biol 224, 141-157.

Harata K, Haga K, Nakamura A, Aoyagi M, Yamane K (1996) X-ray structure of cyclodextrin glucanotransferase from alkalophilic *Bacillus sp* 1011—comparison of two independent molecules at 1.8 Å resolution. Acta Crystallogr. D 52, 1136-1145.

He G-P, Muise A, Li AW, Ro H-S (1995) An eukaryotic transcriptional repressor with carboxypeptidase activity. Nature 378, 92-96.

Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. J Mol Biol 233, 123-138.

Kim H, Lipscomb WN (1991) Comparison of the structures of three carboxypeptidase A-phosphonate complexes determined by X-ray crystallography. Biochemistry 30, 8171-8180.

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.

Liao D-I, Breddam K, Sweet RM, Bullock T, Remington SJ (1992) Refined atomic model of wheat serine carboxypeptidase II at 2.2 Å resolution. Biochemistry 31, 9796-9812.

Matthews BW (1968) Solvent content of protein crystals. J Mol Biol 33, 491-497.

Medrano FJ, Alonso J, García JL, Romero A, Bode W,Gomis-Rüth FX (1998) Structure of proline iminopeptidase from *Xanthomonas campestris* pv. *citri*: a prototype for the prolyl oligopeptidase family. EMBO J 17, 1-9.

Naggert JK, Fricker LD, Varlamov O, Nishina PM, Rouille Y, Steiner DF, Paigen BJ, Leiter EH (1995) Hyperproinsulinaemia in obese fat/fat mice associated with a carboxypeptidase E mutation which reduces enzyme activity. Nature Genet 10, 135-142.

Navaza J (1994) AMoRe : an automated package for molecular replacement. Acta Crystallogr. A 50, 157-163.

Nicholls A, Bharadwaj R, Honig B (1993) GRASP: graphical representation and analysis of surface properties. Biophys J 64, A166-A166.

Ollis DL *et al* (1992) The α/β hydrolase fold. Protein Eng 5, 197-211.

Orville AM, Elango N, Lipscomb JD, Ohlendorf DH (1997) Structures of competitive inhibitor complexes of protocatechuate 3,4-dioxygenase-multiple exogenous ligand binding orientations within the active site. Biochemistry 36, 10039-10051.

Phillips MA, Fletterick R, Rutter WJ (1990) Arginine 127 stabilizes the transition state in carboxypeptidase. J Biol Chem 265, 20692-20698.

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

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

Rees DC, Lewis M, Lipscomb WN (1983) Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. J Mol Biol 168, 367-387.

Reverter D, Ventura S, Villegas V, Vendrell J, Avilés FX (1998) Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. J Biol Chem 273, 3535-3541.

Richardson JS (1981) The anatomy and taxonomy of protein structure. Adv. Protein Chem 34, 167-339.

Roussel A, Cambilleau C (1989) Turbo-Frodo, Silicon Graphics Geometry Partners Directory. Silicon Graphics, Mountain View, CA, pp. 77-79.

Rowsell S, Pauptit RA, Tucker AD, Melton RG, Blow DM, Brick P (1997) Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. Structure 5, 337-347.

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

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.

Singleton MR, Isupov MN, Littlechild JA (1999) X-ray structure of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon *Thermococcus litoralis.* Structure 7, 237-244.

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 (1996) Tissue distribution and characterization of soluble and membrane-bound forms of metallocarboxypeptidase D J Biol Chem 271, 20884-20889.

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

Sorimachi K, Legal-Coeffet MF, Williamson G, Archer DB, Williamson MP (1997) Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to β-cyclodextrin. Structure 5, 647-661.

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.

Teplyakov A *et. al.* (1992) Crystal structure of carboxypeptidase T from *Thermoactinomyces vulgaris.* Eur J Biochem 208, 281-288.

Urban S, Kruse C, Multhap G (1999) A soluble form of the avian hepatitis B virus receptor. J Biol Chem 274, 5707-5715.

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. J Biol Chem 274, 19925-19933.

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

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.

Anexo al capítulo 2

El trabajo presentado en este capítulo se ha realizado en colaboración con el grupo del Dr. Miquel Coll de la Unidad de Cristalografía del Instituto de Bioquímica y Biología Molecular del CSIC en Barcelona. En concreto, mi contribución en este trabajo ha sido la obtención de la proteína a cristalizar. La obtención de la proteína recombinante conllevó la obtención del DNA correspondiente al dominio CPD2 por PCR, su secuenciación y posterior clonación en *Pichia pastoris*, la puesta a punto del sistema de expresión y de la purificación. Cabe mencionar que la expresión y purificación de este dominio ha sido un trabajo arduo, debido principalmente a dos factores probablemente relacionados: primero, los bajos niveles de expresión obtenidos y, segundo, la tendencia a agregar de la proteína durante la purificación. Finalmente se consiguió la cantidad de proteína necesaria para describir por primera vez la estructura tridimensional de una carboxipeptidasa reguladora. Este trabajo posibilita el modelado de otras carboxipeptidasas reguladoras, proteínas cuya obtención en las cantidades requeridas para los estudios estructurales se presenta, en estos momentos, como una tarea de gran dificultad.