The EMBO Journal Vol.16 No.23 pp.6906–6913, 1997

The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen

Isabel García-Sáez, David Reverter[1], Josep Vendrell[1], Francesc X.Avilés[1,2] and Miquel Coll[2]

Departament de Biologia Molecular i Cel.lular, Centre d'Investigació i Desenvolupament-CSIC, Jordi Girona 18-26, 08034 Barcelona, Spain and [1]Departament de Bioquímica i Institut de Biologia Fonamental, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

[2]Corresponding authors

The three-dimensional structure of human procarboxypeptidase A2 has been determined using X-ray crystallography at 1.8 Å resolution. This is the first detailed structural report of a human pancreatic carboxypeptidase and of its zymogen. Human procarboxypeptidase A2 is formed by a pro-segment of 96 residues, which inhibits the enzyme, and a carboxypeptidase moiety of 305 residues. The pro-enzyme maintains the general fold when compared with other non-human counterparts. The globular part of the pro-segment docks into the enzyme moiety and shields the S2–S4 substrate binding sites, promoting inhibition. Interestingly, important differences are found in the pro-segment which allow the identification of the structural determinants of the diverse activation behaviours of procarboxypeptidases A1, B and A2, particularly of the latter. The benzylsuccinic inhibitor is able to diffuse into the active site of procarboxypeptidase A2 in the crystals. The structure of the zymogen–inhibitor complex has been solved at 2.2 Å resolution. The inhibitor enters the active site through a channel formed at the interface between the pro-segment and the enzyme regions and interacts with important elements of the active site. The derived structural features explain the intrinsic activity of A1/A2 pro-enzymes for small substrates.

Keywords: activation/carboxypeptidase A2/inhibition/pro-enzyme/three-dimensional structure

Introduction

Pancreatic carboxypeptidases are a family of zinc-containing exopeptidases, involved in the digestion process, which catalyse the hydrolysis of alimentary proteins and esters from their C-terminus (see Avilés *et al.*, 1993 and references therein). They are secreted as zymogens which are proteolytically activated in the duodenum, producing the active form. In the process of activation of these pro-enzymes, a segment of ~94–95 residues is released (Rinderknecht, 1993). This segment, known as the 'pro-segment' or 'activation segment', acts as a clamp of the active site of the carboxypeptidase, impeding most or all enzymatic activity. Traditionally, carboxypeptidases were classified into A and B forms with regard to their different substrate specificity: CPA has a preference for aliphatic and aromatic C-terminal residues, and CPB for basic C-terminal residues. More recently, two different isoforms, A1 and A2, were distinguished in the rat (Gardell *et al.*, 1988; Oppezzo *et al.*, 1994) and in humans (Pascual *et al.*, 1989; Catasús *et al.*, 1995). The A1 isoform, which corresponds to the previously named 'A', shows a preference for aliphatic C-terminal residues of peptide substrates, while the A2 isoform has a clear specificity for bulkier aromatic C-terminal residues, being the only isoform which shows specificity towards tryptophan. Human procarboxypeptidase A2 (PCPA2) has been cloned by one of our laboratories (Catasús *et al.*, 1995) and shows an 89% sequence identity to rat PCPA2, 64% identity to human PCPA1 and 42% to human PCPB.

It is presently accepted that the peptide hydrolysis mechanism of pancreatic metallo-carboxypeptidases follows the water-promoted pathway (Christianson and Lipscomb, 1989) in which Zn^{2+}, active site water, Glu270 and Arg127 play an essential role. These elements, together with other important active centre residues, such as Asn144, Arg145 and Tyr248, are found in all active forms where they seem to play the same, well-characterized role. The knowledge of the precise and differential role that other residues play in the several active centre subsites (S1′, S1, S2, S3 and S4) and main specificity subsite (formally enclosed in S1′) proposed for metallo-carboxypeptidases is moderate because their variability is greater and because most of the crystal structures of enzyme and enzyme–inhibitor complexes have been obtained only for bovine CPA1 or homologous enzymes and for small inhibitors (Christianson and Lipscomb, 1989; Kim and Lipscomb, 1991; Mangani *et al.*, 1992). An exception to this is the crystal structure of rat CPA2 (Faming *et al.*, 1991). From the crystal structures of zymogens, such as porcine PCPA1 (Guasch *et al.*, 1992) and porcine PCPB (Coll *et al.*, 1991), and of the CPA1–potato carboxypeptidase inhibitor complex (Rees and Lipscomb, 1982), it has been possible to further delineate such extended active centre subsites. From the former structures it is also possible to understand some features of the inhibition and activation mechanisms of metallo-carboxypeptidase zymogens. However, significant questions remain to be answered. Among others, it has been reported that the A2 zymogens show inhibition properties and activation mechanisms much more similar to the B than to the A1 forms (Pascual *et al.*, 1989; Oppezzo *et al.*, 1994), in spite of the fact that the sequences and modelled structures are more similar in the A1/A2 forms (Catasús *et al.*, 1995).

In this work, the three-dimensional structure of a human carboxypeptidase, CPA2, in its pro-form (PCPA2) is presented for the first time. This is also the first crystallographic report on a procarboxypeptidase of the A2 type.

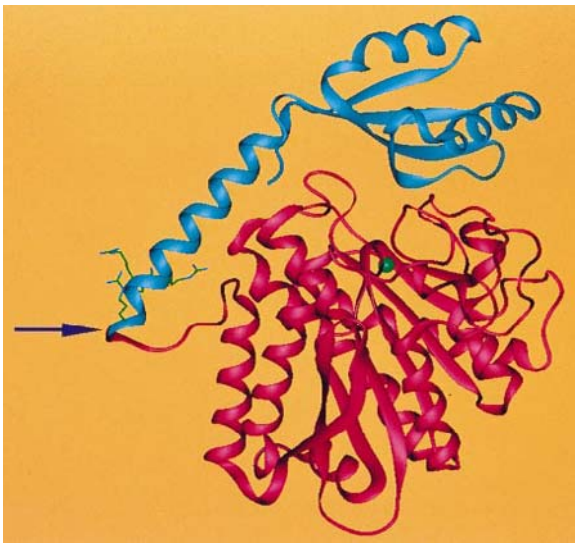

Fig. 1. Ribbon diagram of the three-dimensional crystal structure of recombinant human PCPA2. The globular part of the pro- (activation) segment covers the active centre and is linked to the enzyme by a helical connecting segment followed by a loop. Activation with trypsin occurs between the latter two structures (the primary cleavage site is indicated with an arrow). Arginine residues at this boundary are depicted.

Additionally, the crystal structure of this zymogen complexed with the inhibitor L-benzylsuccinic acid (PCPA2–BZS) has been derived. This inhibitor has been used as a probe to test the accessibility and functionality of the preformed active site of the zymogen. BZS acts as a competitive inhibitor of both peptide and ester hydrolysis, with $K_i = 4.5 \times 10^{-7}$ M (Byers and Wolfenden, 1973), and mimics the product of peptide hydrolysis, since it contains a moiety analogous to L-Phe and a group which emulates a peptide (Mangani *et al.*, 1992). A diffusion experiment was performed over the capillary-mounted crystal after the collection of the unliganded structure data set. To our surprise, the analogue is able to enter the active centre pocket despite the presence of the activation (or inhibition) domain capping the pocket, promoting a concomitant movement of Tyr248 as in the fully activated enzyme. The derived structures clarify significant properties of metallo-carboxypeptidase zymogens. In addition, they are of biotechnological interest in providing the structural basis for the redesign of human carboxypeptidases to act as optimum pro-drug activators for tumour therapy, following the recently proposed antibody-directed pro-drug chemotherapy (Huennekens *et al.*, 1994).

Results and discussion

Human PCPA2 has a globular shape with two clearly separated moieties: the pro- or activation segment of 96 residues, and the carboxypeptidase A2 moiety with 305 residues (Figure 1). The distance between both moieties is shifted two residues towards the C-terminus with respect to previous sequence alignments (Catasús *et al.*, 1995), and to the structure of the A1 and B homologous forms (Avilés *et al.*, 1993), as deduced from the analysis of the conformation and the activation sites carried out in this work (see below). The numbering of both moieties is based on 3-D structure alignments, and an explanation of the conventions used can be found in the last paragraph of Materials and methods.

Activation segment of human PCPA2

The activation segment is divided into two different regions, as described for porcine PCPB (Coll *et al.*, 1991). These two regions are the large N-terminal globular domain, which shields the preforming active site of the enzyme, and a long α-helix, or connecting segment, which links the globular domain with the enzyme. The connecting segment is where proteases (mainly trypsin) perform the activation by cleavage. The globular N-terminal domain (residues 4A–80A) has an open-sandwich antiparallel-α/antiparallel-β fold, with two α-helices and four β-strands which follow the β1α1β2β3α2β4 topology (Figure 1). When comparing this domain with its equivalent in porcine PCPB, an insertion of two residues (Leu34B and Gln34C) occurs and four residues (43A–46A) are absent. This deletion, which corresponds to a 3_{10} helix in PCPB, was also observed in porcine PCPA1 (Guasch *et al.*, 1992) and probably has important consequences for facilitating the presence of a remnant (intrinsic) activity for small substrates, as discussed below. The C-terminal connecting segment comprises residues Val82A to Arg2 (CPA2 numbering) forming the longest α-helix described for pro-segments of carboxypeptidases at this location (α-helix 3). This helix has five turns and is followed by a short loop at the border with the CPA2 moiety. At the C-terminal end of the α3 helix, there are three arginines in a row—Arg97A, Arg98A and Arg99A—followed by Glu1 and a fourth arginine residue, Arg2. All are well defined in the electron density and are solvent-accessible. This unusual concentration of arginines is the target of the sequential activation by trypsin (D.Reverter, unpublished results), which initially cuts at Arg2, the most exposed of the four (Figure 1). In contrast, the connecting segment of porcine PCPA1 has only a four-turn helix, followed by a short loop where only one arginine (Arg99A) is present and exposed. In porcine PCPB, the equivalent α3 helix is even shorter, with only two helical turns, and the region linking to the enzyme is in a loose conformation (see comparison in Figure 2).

CPA2 moiety of human PCPA2

Human CPA2, the core of the pro-enzyme, maintains the same fold as its other non-human homologues (Rees *et al.*, 1983): eight α-helices (α4–α11) and a mixed eight-stranded β-sheet (β5–β12) forming a globular α/β protein (Figure 1). In comparison with porcine PCPA1, used to solve the crystal structure, there are two deletions at positions 6 and 57 (threonine and serine, in CPA1). The first deletion is located in the loop which follows the α3 helix-connecting segment, the zone at which the enzyme moiety starts. It is at the beginning of this loop, which is longer in porcine PCPA1, where the first proteolytic cut takes place. Deletion at position 57 is located in the coil between strands β6 and β7.

As in porcine PCPA1 (Guasch *et al.*, 1992) and bovine PCPA1 (Gomis-Rüth *et al.*, 1995), in human PCPA2 there is a disulfide bond between Cys138 and Cys161. However, a second disulfide bond exists in the present structure linking Cys210 and Cys244. This feature is unique for

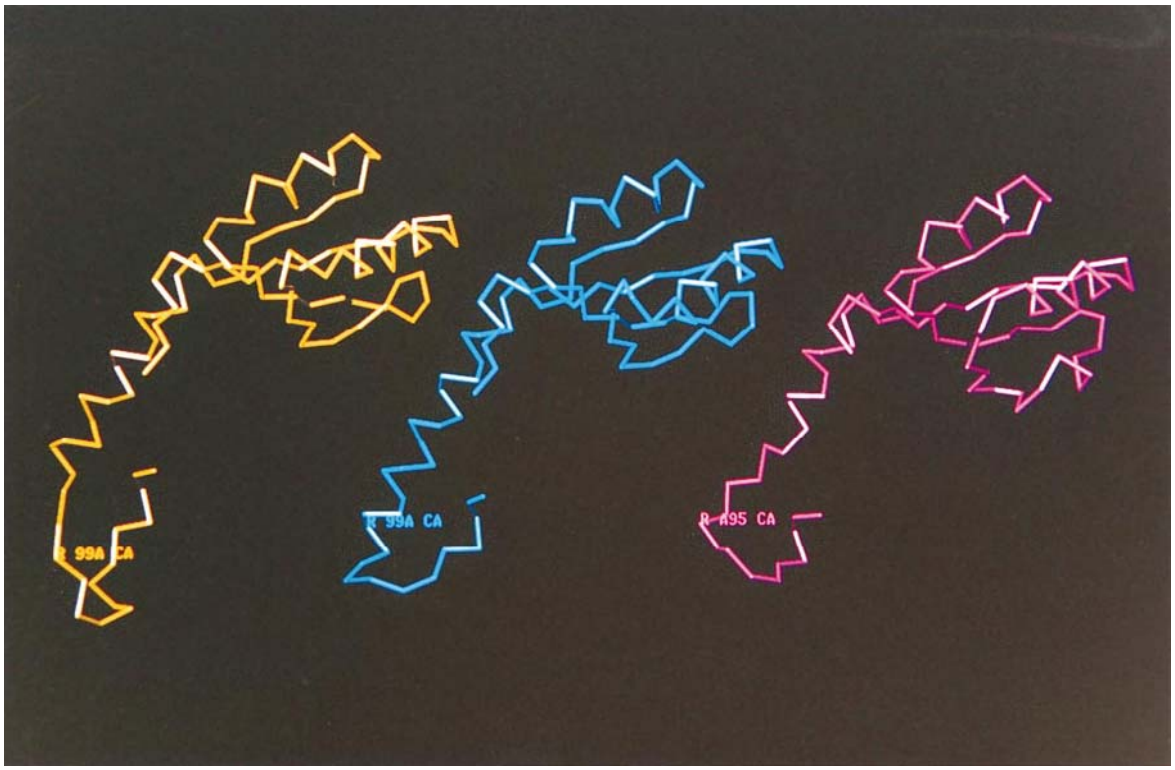

Fig. 2. Comparative diagram showing the C tracings of the pro-segments of porcine PCPA1 (left), human PCPA2 (centre) and porcine PCPB, followed by the loops linking them with the corresponding enzyme moieties. The pro-segments have a similar topology in the globular part—except for a 3_{10} helix present only in the B isoform—but have different lengths in the α-helix in the connecting segment. The linking loop (spanning the boundary between the pro- and enzyme regions in the A1 and B forms, and fully at the enzyme region in A2) also shows a significant difference in conformation, being longer in PCPB.

PCPA2. There are three disulfide bridges in porcine PCPB (Coll *et al.*, 1991), but none at residues 210–244. This link in the present structure affects the position of a loop region, residues 210–212 being displaced 1.8 Å when compared with PCPA1 or PCPB. Furthermore, the disulfide bridge is spatially close to residues involved in the active site, Ile247 and Tyr248, which are also slightly displaced and could influence the specificity of the enzyme. Moreover, as previously reported for other pancreatic carboxypeptidases, three *cis*-peptide bonds are located between residues Ser197–Tyr198, Pro205–Tyr206 and Arg272–Asp273.

In the preformed active site of PCPA2 the typical coordination of the Zn^{2+} atom described for carboxypeptidases was found. This atom is liganded to a catalytic water molecule and to the enzyme by His69, His196 and Glu72, the latter in a bidentate form. The well-structured arrangement of His196 coordinated with Zn^{2+} forces the formation of the Ser197–Tyr198 *cis*-peptide bond between the two following residues; we should also note that the other two *cis*-peptide bonds are very close to residues involved in the active site which need to be kept in their appropriate positions.

Residues which form the different active-centre subsites are substantially conserved if compared with other pancreatic metallo-carboxypeptidases. The S1′ subsite, which forms a pocket involved in the anchoring and neutralization of the C-terminal carboxylate of the substrate, is still formed by Arg145, Asn144 and Tyr248. The latter caps this pocket once the benzylsuccinic acid inhibitor is bound (see below). Subsites S1 (Arg127 and Glu270), which contribute to the polarization of the carbonyl of the scissile peptide bond of the substrate and to the proton exchange, S2 (Arg71, Ser197, Tyr198, Ser199) and S3 (Phe279), involved in binding and torsion of peptide substrates, also remain unchanged. The only difference is at subsite S4, where Gln122, Arg124 and Lys128 are present in bovine CPA1 (Catasús *et al.*, 1995), while Lys122, Arg 124 and Lys 128 are found in human PCPA2.

The main putative specificity pocket of the enzyme moiety was found to be empty in the crystal structure of PCPA2. This cavity is formed in this enzyme by Asn144, Ile194, Met203, Ile243, Ile247, Ala250, Gly253, Ser254, Ile255, Asp256 and Ala268, which are better delineated in the complex with BZS (see later). Substitutions at positions 194, 203 and 268 for smaller or more hydrophobic residues with respect to the A1 forms are probably responsible for the higher specificity of CPAs for bulky aromatic residues.

Interaction of the pro-region with the enzyme: the inhibition mechanism

The pro-region establishes an extensive interaction with the enzyme moiety (Table I), burying a surface of 1375 $Å^2$, of which 760 $Å^2$ correspond to the interaction of the N-terminal globular domain of the former with the enzyme. The binding between both regions occludes the S2, S3 and S4 active centre subsites of the enzyme, a fact which should make the recognition of large peptide substrates by the enzyme difficult. However, the S1′ and S1 subsites (Asn144, Arg145, Tyr248, Glu270 and Arg127) are not directly shielded in the pro-enzyme. This could explain the

Table I. Interactions[a] between the pro-segment and the CPA2 moiety in human pro-CPA2

Globular domain/CPA2	Distance (Å)	Connecting segment/CPA2	Distance (Å)
N(Glu5A)--O(Lys122)	3.0	$C^{\gamma 1}$(Val82A)...$C^{\delta 1}$(Leu280)	4.0
O(Glu5A)--W2--N(Lys124)	3.0/2.9	C^{β}(Leu85A)...$C^{\delta 1}$(Leu280)	3.9
O(Glu5A)--W2--O(Lys124)	3.0/3.5	O(Leu85A)--W64--O (Leu280)	3.4/2.9
$C^{\varepsilon 1}$(Phe7A)...C^{γ}(Met125)	3.7	$C^{\delta 1}$(Leu86A)...C^{γ}(Arg124)	3.8
C^{ζ}(Phe7A)...C^{γ}(Met125)	3.7	$O^{\varepsilon 1}$(Glu89A)--$N^{\eta 2}$(Arg124)	2.9
$O^{\varepsilon 2}$(Glu33A)--$N^{\delta 2}$(Asn159)	3.3	$O^{\varepsilon 2}$(Glu89A)--N^{ε}(Arg124)	2.8
$O^{\delta 1}$(Asp36A)--$N^{\eta 2}$(Arg71)	2.8	C^{δ}(Glu89A)--$C^{\zeta 3}$(Trp73)	3.7
$O^{\delta 1}$(Asp36A)--$N^{\eta 1}$(Arg71)	3.4	$O^{\varepsilon 1}$(Glu89A)--W65--O(Leu281)	2.8/2.9
$O^{\delta 2}$(Arg36A)--$N^{\eta 2}$(Arg71)	3.0	$O^{\varepsilon 2}$(Glu92A)--$N^{\eta 2}$(Arg284)	3.3
$O^{\delta 1}$(Asp36A)--W55--$O^{\delta 2}$(Asp163)	2.7/2.8	$O^{\varepsilon 1}$(Glu92A)--W132--N^{ε}(Ala283)	2.9/3.0
O(Trp38A)--W120--$N^{\eta 1}$(Arg71)	3.1/3.1	$O^{\varepsilon 1}$(Glu92A)--W100--N(Arg284)	2.8/2.9
C^{γ}(Trp38A)...$C^{\varepsilon 2}$(Phe279)[b]	3.5	$O^{\varepsilon 1}$(Glu92A)--W00--N^{ε}(Arg284)	2.8/2.9
$C^{\zeta 3}$(Trp38A)...C^{α}(Gly278)	4.0	O(Met93A)--W87--O(Tyr12)	2.9/2.9
N^{ζ}(Lys39A)--W256--O(Val246)	2.8/2.7	O(Asn96A)--N(Glu1)[e]	2.8
N(Ser40A)--W13--O(Tyr248)[c]	3.3/3.2	$N^{\delta 2}$(Asn96A)--$O^{\varepsilon 2}$(Glu1)	2.9
O(Ser40A)--W13--O(Tyr248)[c]	3.2/3.2	$O^{\delta 1}$(Asn96A)--$N^{\delta 2}$(Asn8)	3.0
$C^{\gamma 2}$(Thr42A)...C(Ile247)	3.9	C^{γ}(Asn96A)--C^{β}(Ala11)	3.6
$C^{\gamma 2}$(Thr42A)...$C^{\varepsilon 2}$(Tyr248)[d]	3.5	O(Arg97A)--N(Arg2)[e]	3.5
$C^{\gamma 2}$(Thr42A)...C^{β}(Tyr248)[c]	3.8	O(Arg97A)--N(Glu1)[e]	3.3
$C^{\gamma 2}$(Thr47A)...$C^{\varepsilon 2}$(Tyr248)[d]	4.0	$N^{\eta 2}$(Arg97A)--W36--O(Tyr12)	2.7/3.5
$N^{\delta 1}$(His53A)--O^{η}(Tyr198)	2.7	Nε(Arg97A)--W36--O(Tyr12)	2.9
S^{δ}(Met78A)...O(Gly275)	3.7	O(Arg98A)--N(Glu1)[e]	3.3
S^{δ}(Met78A)...C(Gly275)	4.0	O(Arg98A)--N(Arg2)[e]	3.9
		O(Arg99A)--N(Gly4)[e]	3.1
		O(Arg99A)--N(Gly3)[e]	3.1
		O(Met93A)--W87--O(Tyr12)	2.9/2.9
		$N^{\delta 2}$(Asn96A)--W87--O(Tyr12)	2.7/2.9
		N^{ε}(Arg97A)--W36--O(Tyr12)	2.9/3.5
		$N^{\eta 2}$(Arg97A)--W36--O(Tyr12)	2.7/3.5

[a](...):van der Waal's contacts ⩽4.0 Å; (--): H-bonds ⩽3.5 Å.
[b]Aromatic–aromatic interaction; only the shortest distance between the two rings is listed.
[c]In the benzylsuccinate complexed structure.
[d]In the uncomplexed structure.
[e]Intrahelical H-bond.

small, but clearly detectable, intrinsic hydrolytic activity of human PCPA2 towards small peptide substrates, which attains a 0.5–10% level with respect to the active enzyme according to the nature and size of the substrate (D.Reverter, unpublished results). These results show that the active centre is preformed in PCPA2, and that the pro-region is unable to fully inhibit the enzyme moiety. This situation is in contrast to porcine PCPB in which no intrinsic activity towards small peptide substrates is measurable; the strong inhibition of PCPB is probably due to the partial shielding of the S1′ and S1 subsites (Coll *et al.*, 1991).

The important consequences of the absence of residues 43A–46A in human PCPA2, in comparison with porcine PCPB, should be noted. This deletion (or insertion in PCPB) is located in the loop which connects β2 and β3 strands of the activation segment, just over the active site. In the PCPA2 case, the loop adopts a straight conformation and does not physically cap the active site, as occurs in porcine PCPB where it forms a 3_{10} helix which covers the active site pocket. PCPB also shows a salt bridge between Asp41A and Arg145, a bridge which probably prevents the anchorage of substrates by its C-terminal carboxyl to the enzyme (to Arg145), and which does not exist in human PCPA2. The above deletions in human PCPA2 allow the formation of a clear channel at the interface between the pro-segment and the enzyme which make the active site accessible (see Figure 3). This channel is flanked by Tyr248 in its open position in the unliganded PCPA2, pointing out towards the solvent. We shall see later that in the PCPA2–BZS complex the side chain of Tyr248 swings into the channel, closing it.

Comparison between the unliganded human PCPA2 and its complex with benzylsuccinic acid

The three-dimensional structure of the PCPA2–BZS complex, obtained by diffusion *in situ* of the inhibitor in the crystal of the unliganded pro-enzyme after data collection, has been solved at 2.19 Å resolution. The r.m.s. deviation between unliganded PCPA2 and the PCPA2–BZS complex has been checked, giving an overall r.m.s. difference of 0.254 Å for the backbone of the zymogen. The main differences in the main chains are found in the loop which connects strands β2 and β3 at the pro-segment, from residue Pro40A to Pro48A, where Thr42A moves 1.3 Å from its position in the unliganded structure. The differences are caused by the presence of the inhibitor in the active site pocket and the concomitant rotation of the Tyr248 side chain. In its closed or inner position (see above), this side chain would clash with the Thr42A side chain in its unliganded position. Also other rearrangements at the active site and its environment occur to accommodate the inhibitor, but they mainly affect side chains.

Presumably, BZS diffuses into the active site through the above-mentioned channel at the surface of the zymogen. This channel is closed in the PCPA2–BZS complex by Tyr248 due to a rotation of 165° of its χ_1 torsion angle, by which its hydroxyl group moves 12.7 Å and points

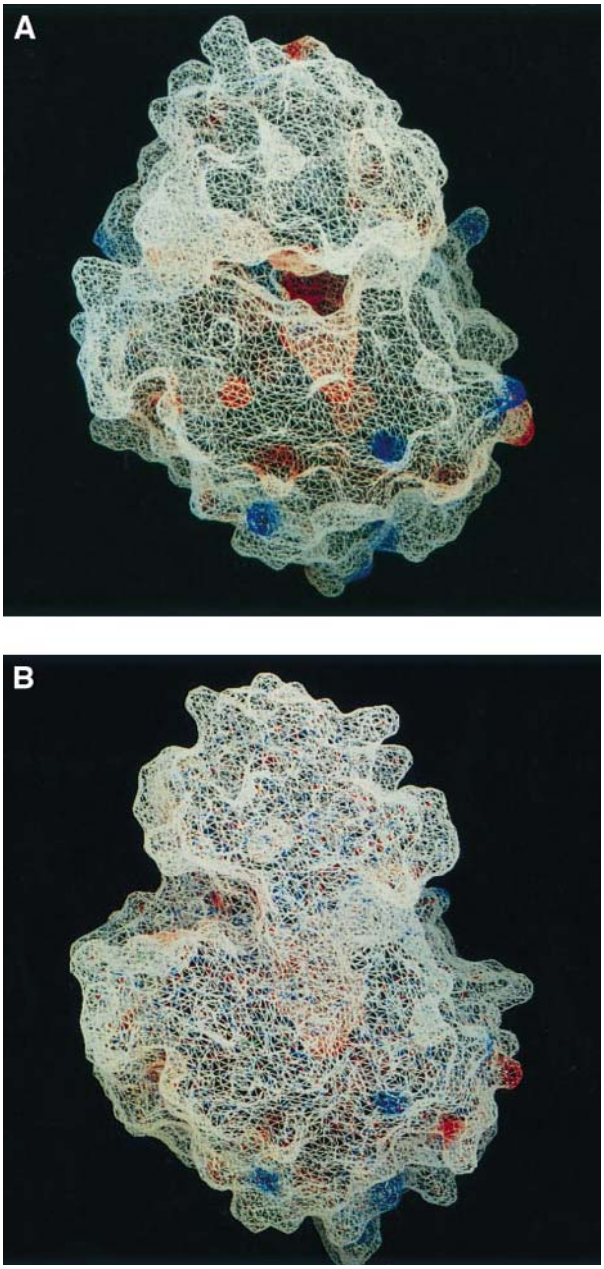

Fig. 3. Surface representation and electrostatic potential, calculated with GRASP (Nicholls *et al.*, 1993), of human PCPA2 in the unliganded form (this work) (**A**), and porcine PCPB (Coll *et al.*, 1990) (**B**). The pro-segment is on top, covering the active centre of the enzyme moiety. A channel between both regions, making the active site accessible for small substrates, is present in PCPA2 but not in PCPB. This channel becomes closed in the inhibitor-complexed structure of PCPA2 due to a swing movement of Tyr248.

towards one carboxyl group of the inhibitor (Figure 4). The structure of the complex demonstrates that this movement does not have steric restrictions even though it takes place in the zymogen form of the protein. This is another proof of the intrinsic functionality of the preformed active centre of PCPA2, in agreement with its residual activity towards small peptide substrates. It is worth mentioning that the long-range movement of Tyr248 in carboxypeptidases (from the up to the down position), and of its associated loop, has been used as a model example of the induced fit of enzymes during substrate binding (Christianson and Lipscomb, 1989). Other important active site residues, such as Arg127, Arg145 and Glu270, have also suffered deviations in the PCPA2–BZS complex in comparison with those in unliganded proenzymes.

It has been demonstrated that bovine CPA1 shows preference for the L-form in a mixture of L- and D-benzylsuccinic acid (Byers and Wolfenden, 1973; Mangani *et al.*, 1992). In the present case, a molecule of L-benzylsuccinic acid diffused selectively into the preformed active site of PCPA2, displacing four molecules of water, including the catalytic one which was coordinated with the zinc atom in the unliganded structure. Carboxylic oxygens O1 and O2, of L-BZS, are coordinated in a symmetric bidentate form with the Zn^{2+}, at 2.5 Å distance. The inhibitor is also bound to the enzyme by its other carboxylic oxygens, O3 and O4, which establish hydrogen bonds with the guanidinium groups of Arg145, Arg127 and Tyr248. Arg127 in PCPA2–BZS also interacts with O1. The interaction which exists between the $N^{\delta 2}$ atom of Asn144 and the BZS benzyl ring, at 3.4 Å distance, should also be noted. This interaction helps to position the benzyl ring in the hydrophobic pocket.

It has been described for bovine CPA–inhibitor complexes, including transition state analogues, that the coordination of Glu72 with the Zn atom changes from bidentate in the unliganded structure to monodentate, as a consequence of the slight movement of the zinc atom (Christianson and Lipscomb, 1986; Shoham *et al.*, 1988; Mangani *et al.*, 1992). We do not find such a change in our structures.

The presence of BZS in the preformed active site of PCPA2 helps to delineate the specificity pocket. In this pocket the benzyl group of the inhibitor, which mimics the side chain of substrates bound at S1′, is surrounded by residues Asn144, Ile194, Met203, Ile243, Ala250, Gly253, Ser254 and Ala268. Residues Ile247 and Ile255 are also close to BZS. Residue 268 is a threonine or a serine in bovine or porcine PCPA1. This fact determines, for human PCPA2, a wider pocket which can accommodate bulkier residues in comparison with PCPA1. It also has a stronger hydrophobic character because of this substitution and the presence of an isoleucine residue at position 194, which is a serine in PCPA1.

Activation mechanism in human PCPA2

A faster activation by trypsin cleavage for PCPBs than for PCPA1s (Avilés *et al.*, 1993; Villegas *et al.*, 1995) has been described. On the other hand, human and rat PCPA2 show much closer activation rates to PCPB than to PCPA1 (Pascual *et al.*, 1989; Oppezzo *et al.*, 1994). These differences could be primarily due to a slower degradation of the pro-segment in PCPA1s as well as to a remaining inhibitory capability of large fragments of the segment after cleavage. In contrast to PCPA1, the whole severed segment and fragments would not be inhibitory in PCPBs and PCPA2s.

In human PCPA2, the target of activation by trypsin cleavage is initially the arginyl peptide bond between Arg2 and Ser3 (enzyme numbering) located at the end of the well-structured connecting segment, in the C-terminus of helix α3 (D.Reverter, unpublished results). Subsequent cleavages occur at Arg99A and Arg97A, at the preceding

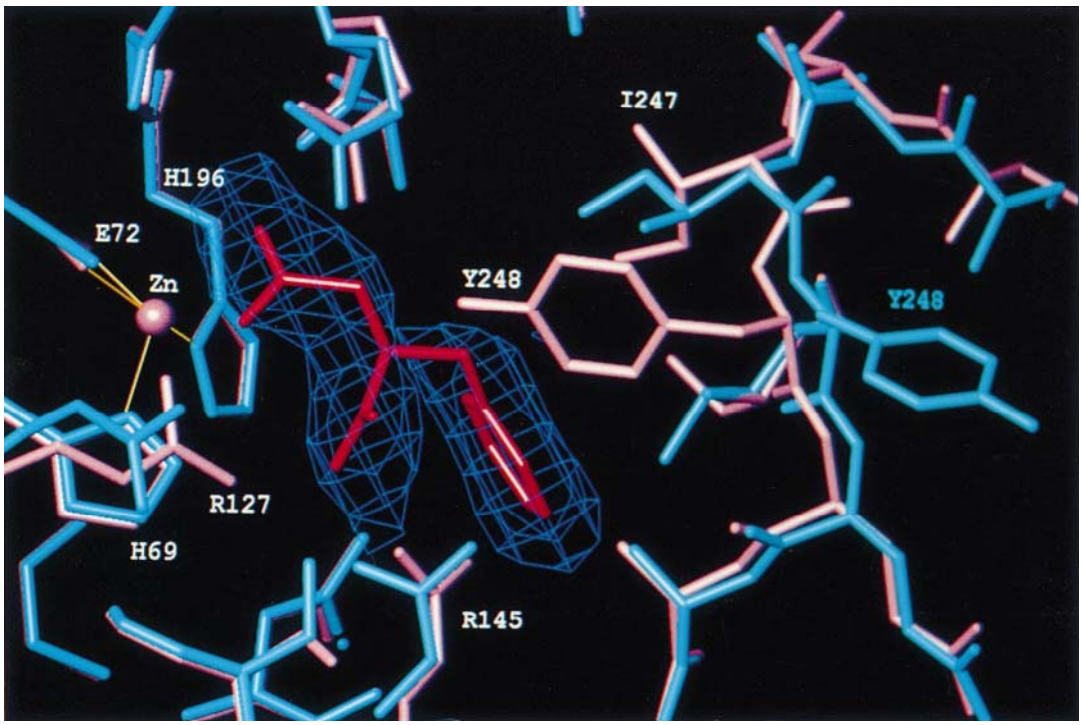

Fig. 4. Differences in three-dimensional structure of human PCPA2 at its active centre when unliganded (blue) or liganded (pink) with benzylsuccinic acid (red). Important residues for binding and catalysis are labelled. The major change is detected for the Tyr248 side chain, which rotates 165° around its χ_1 angle, moving its hydroxyl group 12.7 Å, and enters the active site in the complexed structure. The electron density corresponds to a 2.5 Å resolution F_o–F_c Fourier map, contoured at 2σ, and calculated before adding the inhibitor molecule.

helical turn of the well-structured α3 helix (Figure 1). No other cleavages take place in the pro-segment. It has also been shown that the first cleavage is sufficient for the release of the pro-segment and the full generation of activity (D.Reverter, unpublished results). Why does this release take place in spite of the extensive interaction (Table I) between pro-segment and enzyme? What are the precise structural determinants of the activation?

In both porcine PCPA1 and PCPB, the first tryptic cleavage occurs at an arginyl peptide bond located in a flexible region at the end of the connecting segment, at the boundary with the enzyme. The first cleavage in PCPA1 is followed by a trimming of the C-terminus of the pro-segment by the generated CPA1 and by a second tryptic cleavage at the peptide bond ArgA74–TyrA75, far upstream within strand β4 at the globular domain of the pro-segment. Both types of cleavage are required for full activation (Vendrell *et al.*, 1990). The first cleavage in PCPB is followed by a trimming of the C-terminal residue of the pro-segment by the generated CPB, and by a second tryptic cleavage at the ArgA83–SerA84 peptide bond, in the N-terminal part of helix α3. However, in this case, the first tryptic cleavage is sufficient to generate full activity (Burgos *et al.*, 1991; Villegas *et al.*, 1995). It has been proposed (Avilés *et al.*, 1993) that this differential behaviour is mainly due to the different degrees of structuration of the connecting segment of both pro-enzymes and of the extension and nature of their contacts with the enzyme moiety. Nevertheless, this hypothesis does not seem to have a general validity for PCPs in light of the crystal structure of PCPA2 reported here: the connecting segment and its long α3 helix in the human PCPA2 three-dimensional structure show an interactive surface with the enzyme moiety (692 Å^2) more extensive than the equivalent pieces in porcine PCPA1 and PCPB (566 Å^2 and 539 Å^2, respectively). This interaction is strengthened by two salt bridges in PCPA2, between the $O^{\varepsilon 2}$ of Glu89A and the $N^{\eta 2}$ atom of Arg 124, and the $O^{\varepsilon 2}$ of Glu92A and the $N^{\eta 2}$ atom of Arg 284 (Table I); in porcine PCPA1 only the salt bridge between GluA89 and Arg124 is present, and in porcine PCPB both salt bridges are lacking.

Given that the connecting segment of PCPA2 is the better structured one within PCPs and the one with stronger contacts with the enzyme moiety, we should conclude that the different velocities of the activation process in PCPs are mainly determined by its complementary piece, that is by the inhibitory capability of the pro-segment globular domain when cut from the pro-enzyme. The larger buried surface between the pro-segment globular domain and the enzyme moiety in PCPA1 (861 Å^2) with respect to PCPA2 (760 Å^2) and PCPB (755 Å^2) is in agreement with this interpretation. In particular there are four amino acids, spatially located very close together, in the β-sheet of the globular domain of the pro-segment of PCPA1 which strongly interact with the enzyme moiety (Arg14A–Thr274, Arg39A--Ser199, Arg47A--Ile244 and Phe50A... Val246). None of these interactions are present in human PCPA2. Also, the entire electrostatic energy of interaction between both pieces, as measured by the GROMOS program (van Gunsteren *et al.*, 1996) is clearly negative (attractive) in PCPA1, clearly positive (repelling) in PCPA2, and nearly null in PCPB. Overall, this could lead to the easy release of the pro-segment globular domain of PCPA2 when its connecting segment is cleaved and its interaction with the enzyme is thus weakened. The above-

described channel between the two moieties of the pro-enzyme should also facilitate the penetration of water and small molecules, and dissociation.

We should also consider the possibility that dynamic effects and long range conformational effects take place after the first proteolytic cleavage in PCPs, leading to the repositioning of the two regions of the pro-segment over the enzyme. This could impair fitting and eventually lead to the release of the pro-segment (in PCPA2 and PCPB) or to its further trimming and final release (in PCPA1). Further work is required to confirm this hypothesis.

It is worth mentioning that native and redesigned active forms of pancreatic carboxypeptidases, such as bovine CPA1 and human CPA1 and CPA2, have been proposed to be used in cancer therapy following the ADEPT strategy (Huennekens, 1994; Vitols *et al.*, 1995; Laethem *et al.*, 1996). According to this approach, an antibody specific for tumoral tissues is covalently linked to a carboxypeptidase-like enzyme able to remove a C-terminal amino acid (or analogue) from a pro-drug, such as those based on the methotrexate molecule, transforming it into the active form near the tumoral tissue. The derived three-dimensional structures of recombinant human PCPA2 and the PCPA2–BZS complex could be very useful in the redesign of human CPA2 to increase its specificity for these or other pro-drugs, improving its biomedical applicability.

Materials and methods

Protein purification

Human recombinant proCPA2 was secreted extracellularly in a *Pichia pastoris* heterologous system and purified by chromatographic approaches (D.Reverter *et al.*, to be published elsewhere in detail). In summary, the supernatant from a *P.pastoris* expression medium was fractionated by hydrophobic interaction chromatography at atmospheric pressure, followed by FPLC anion exchange chromatography. The peaks containing the pro-enzyme were located and characterized by enzymatic analysis on furyl-acryloyl-L-phenylalanyl-L-phenylalanine (FAPP) and benzoyl-glycyl-L-phenylalanine (BGP), with or without a previous activation with trypsin, as previously described (Pascual *et al.*, 1989).

Crystallization

Crystals of unliganded PCPA2 were obtained after mixing 4 µl of fresh concentrated protein solution with 2 µl of 20% PEG, 0.01 M nickel chloride and 0.1 M Tris–HCl, pH 8.0 as buffer solution, at 4°C. The crystals were monoclinic and belonged to space group $P2_1$, with unit cell parameters a = 42.20 Å, b = 87.07 Å, c = 59.02 Å and β = 99.35°. The calculated V_m value is 2.23 Å^3/Dalton with one molecule per asymmetric unit. The nature of the crystals was checked by PAGE, confirming the presence of the pro-enzyme without any activation cleavage. To obtain the complex with benzylsuccinic acid (PCPA2–BZS), the same crystal used for data collection of the unliganded PCPA2 was diffused with the inhibitor. After the initial native data collection, the X-ray glass capillary was filled with a solution of 10 mM D,L-benzylsuccinic acid (Sigma) in a stabilization buffer. The capillary was closed with wax and the crystal was left to soak in this solution for one week. After this, the solution was removed from the capillary and the crystal was directly used for X-ray data collection again. The unit cell dimensions were minimally affected, being a = 42.61 Å, b = 87.41 Å, c = 60.27 Å and β = 99.27°.

Data collection

For the unliganded structure, a 1.8 Å X-ray data set was collected, at room temperature, using a 300 mm MarResearch Image Plate Detector and CuK_α radiation from an RU200 Rigaku rotating anode source. For the PCPA2–BZS complex, a 2.2 Å data set was collected, also at room temperature, using the same X-ray source and the scanner set in the 180 mm mode. Both data sets were processed with the program DENZO-SCALEPACK (Otwinowski, 1993). The completeness of the data was checked with PROTEIN (Steigemann, 1991) (Table II).

Table II. Data collection statistics

	PCPA2[a]	PCPA2–BZS[b]
Number of measurements	327073	136319
Number of unique reflections ($F_o > 2\sigma F_o$)	37420	20232
Completeness (%) ($F_o > 2\sigma F_o$)		
overall	94.0	90.9
last shell (resolution range)	87.9 (1.86–1.8 Å)	80.5 (2.27–2.2 Å)
$I/\sigma(I)$		
overall	32.6	16.7
last shell (resolution range)	9.1 (1.86–1.8 Å)	8.0 (2.27–2.2 Å)
R merge (%)	4.7	6.1

[a]Procarboxypeptidase A2.
[b]Procarboxypeptidase A2–benzylsuccinic acid complex.

Table III. Refinement statistics

	PCPA2[a]	PCPA2–BZS[b]
Resolution range (Å)	8.0–1.8	8.0–2.2
Number of reflections ($F_o > 2\sigma F_o$)	37092	19802
Total No. of atoms (excluding hydrogens)	3393	3339
Number of solvent molecules	218	150
Metal ions	1 Zn	1 Zn
R-cryst (%)[c]	19.8	19.4
R-free (%)[d]	25.3	27.3
R.m.s. deviation from target values		
bond lengths (Å)	0.008	0.008
bond angles (°)	1.6	1.5

[a]Procarboxypeptidase A2.
[b]Procarboxypeptidase A2–benzylsuccinic acid complex.
[c]R-cryst = $\Sigma_{hkl}||F_{obs}|-|F_{calc}||/\Sigma_{hkl}|F_{obs}|$, for 90% of the data included in the refinement.
[d]R-free, for 10% of the data excluded from the refinement.

Structure determination and refinement

The unliganded human PCPA2 structure was determined by molecular replacement using the structure of porcine pancreatic PCPA1 (PDB code 1PCA; Guasch *et al.*, 1992) as a starting model, with the AMoRe program (Navaza, 1994). Diffraction data from 8 Å to 4 Å were used to calculate the rotation and translation functions. The correlation factor of the best solution was 55.9% and the R-factor 38.6% after the rigid body 'fitting' routine. Combined rounds of positional refinement using the X-PLOR program (Brünger, 1992) and model building using the graphic TURBO program (Rousell and Cambillau, 1991) followed. Ten-percent randomly selected reflections were set aside for the R-free calculation. The area which required more model building was the connecting segment of the activation domain. Omit maps were calculated excluding several residues in the area, and the α-helix was built according to the residual electron density. Further refinement cycles followed, increasing the resolution, adding solvent molecules and including restrained B-factor refinement. The final R-factor is 19.8%, and R-free is 25.3% (Table III).

The PCPA2–BZS structure was solved using the coordinates of the previously determined native PCPA2 as a starting model. Rigid body refinement was performed with the X-PLOR program using data between 10 Å and 4 Å resolution, giving an R-factor of 27.7%. During the entire refinement process, 10% of the reflections were set aside for the R-free calculation. The 3.0 Å Fourier maps, with coefficients $2F_o–F_c$ and $F_o–F_c$, were calculated after a few steps of positional refinement in order to check whether the benzylsuccinic acid inhibitor had diffused into the active site of the pro-enzyme, in the crystal. Residual electron density, close to the atom of Zn, fitted very well with the inhibitor coordinates in its L-form (which were obtained from the complex structure bovine CPA1/L-benzylsuccinate; Mangani *et al.*, 1992), and revealed that the small molecule had reached the active site. Several cycles of positional refinement with XPLOR and model building with TURBO followed, increasing the resolution. After the final restrained B-factor refinement cycle, the R-factor was 19.4% and the R-free of 27.3% (Table III).

Both uncomplexed and complexed models had good stereochemistry, as assessed with PROCHECK (Laskowski *et al.*, 1993). Atomic coordinates and structure factors for both structures have been deposited with the Brookhaven protein data bank.

The numbering of the pro-segments of procarboxypeptidases is based on that used in Guasch *et al.* (1992) where (i) numbering is referenced to that of PCPB, the first 3-D structure obtained for a procarboxypeptidase, and the number of the amino acid positions are followed by 'A' to distinguish them from the positions in the enzyme moiety; (ii) the first amino acid in PCPA1 and A2 is in position number 4; (iii) a two-amino acid insertion after position number 34 in the A forms is numbered 34B and 34C; (iv) amino acids 43A to 46A of PCPB are absent in PCPA1 and A2. The numbering of bovine CPA1 is kept for the enzyme moiety.

Acknowledgements

We thank A.Párraga and J.M.Nas for their help in the elaboration of the figures. This work has been supported by grants BIO95-0848 and PB95-0224 (Ministerio de Educación y Ciencia, Spain) and by the Centre de Referència en Biotecnologia (Generalitat de Catalunya). D.Reverter is a recipient of a fellowship from the CIRIT (Generalitat de Catalunya). Support from Fundación Francisca de Roviralta is gratefully acknowledged.

References

Avilés,F.X., Vendrell,J., Guasch,A., Coll,M. and Huber,R (1993) Advances in metallo-procarboxypeptidases. Emerging details on the inhibition mechanism and activation processes. *Eur. J. Biochem.*, **211**, 381–389.

Burgos,F.J., Salvà,M., Villegas,V., Soriano,F., Méndez,E. and Avilés,F.X. (1991) Analysis of the activation process of porcine procarboxypeptidase B and determination of the sequence of its activation segment. *Biochemistry*, **30**, 4082–4089.

Brünger,A.T. (1992) XPLOR, version 3.1, a system for X-ray crystallography and NMR. Yale University Press, New Haven and London.

Byers,L.D. and Wolfenden,R (1973) Binding of the by-product analogue benzylsuccinic acid by carboxypeptidase A. *J. Biol. Chem.*, **247**, 606–608.

Catasús,L., Vendrell,J., Avilés,F.X., Carreira,S., Puigserver,A. and Billeter,M. (1995) The sequence and conformation of human pancreatic procarboxypeptidase A2. *J. Biol. Chem.*, **270**, 6651–6657.

Christianson,D.W. and Lipscomb,W.N. (1986) Structure of the complex between an unexpectedly hydrolyzed phosphonamidate inhibitor and carboxypeptidase A. *J. Am. Chem. Soc.*, **108**, 545–546.

Christianson,D.W. and Lipscomb,W.N. (1989) Carboxypeptidase A. *Acc. Chem. Res.*, **22**, 62–69.

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

Faming,Z., Kobe,B., Stewart,C.B, Rutter,W.J. and Goldsmith,E.J. (1991) Structural evolution of an enzyme specificity. The structure of rat carboxypeptidase A2 at 1.9-Å Resolution. *J. Biol. Chem.*, **266**, 24606–24612.

Gardell,S.J., Craik,C.S., Clauser,E., Goldsmith,E.J., Stewart,C.-B., Graf,M. and Rutter,W.J. (1988) A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family. *J. Biol. Chem.*, **263**, 17828–17836.

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

Guasch,A., Coll,M., Avilés,F.X. and 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.

Huennekens,F.M. (1994) Tumor targeting: activation of prodrugs by enzyme-monoclonal antibody conjugates. *Trends Biotechnol.*, **12**, 234–239.

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

Laethem,R.M., Blumenkopf,T.A., Cory,M., Elwell,L., Moxham,C.P., Ray,P.H., Walton,L.M. and Smith,G.K. (1996) Expression and characterization of human pancreatic preprocarboxypeptidase A1 and preprocarboxypeptidase A2. *Arch. Biochem. Biophys*., **332**, 8–18.

Laskowski,R.A, MacArthur,M.W., Moss,D.S. and Thorton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structure coordinates. *J. Appl. Crystallogr.*, **A42**, 140–149.

Mangani,S., Carloni,P. and Orioli,P. (1992) Crystal structure of the complex between carboxypeptidase-A and the biproduct analog inhibitor L-benzylsuccinate at 2.0-Å resolution. *J. Mol. Biol.*, **223**, 573–578.

Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.

Nicholls,A. (1993) *GRASP: Graphical Representation and Analysis of Surface Properties*. Columbia University Press, New York.

Oppezzo,O., Ventura,S., Bergman,T., Vendrell,J., Jornvall,H. and Avilés,F.X. (1994) Procarboxypeptidase in rat pancreas. Overall characterization and comparison of the activation processes. *Eur. J. Biochem.*, **222**, 55–63.

Otwinowski,Z. (1993) Oscillation data reduction program. In Sawyer,L., Isaacs,N. and Bailey,S. (eds), *Proceedings of the CCP4 Study Weekend: 'Data Collection and Processing,' 29–30 January 1993*. SERC Daresbury Laboratory, UK, pp. 56–62.

Pascual,R., Burgos,F.J., Salvà,M., Soriano,F., Méndez,E. and Avilés,F.X. (1989) Purification and properties of five different forms of human procarboxypeptidases. *Eur. J. Biochem.*, **179**, 609–616.

Rees,D.C. and Lipscomb,W.N. (1982) Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J. Mol. Biol.*, **160**, 475–498.

Rees,D.C., Lewis,M. and Lipscomb,W.N. (1983) Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. *J. Mol. Biol.*, **168**, 367–387.

Rinderknecht,H. (1993) Pancreatic secretory enzymes. In Go,V.L.W. *et al.* (eds), *The Pancreas*: *Biology*, *Pathobiology and Diseases.* Raven Press, New York, pp. 219–252.

Rousell,A. and Cambillan,C. (1991) Silicon Graphics Directory, Silicon Graphics, Mountain View, CA.

Shoham,G., Christianson,D.W. and Oren,D.A. (1988) Complex between carboxypeptidases A and hydrated ketomethylene substrate analogue. *Proc. Natl Acad. Sci. USA*, **85**, 684–688.

Steigemann,W. (1991) *PROTEIN. A Program System for the Crystal Structure Analysis of Proteins. Version 3.1.* Max-Planck-Institut für Biochemie, Martinsried bei München.

van Gunsteren,W.F., Billeter,S.R., Eising,A.A., Hünenberger,P.H., Krüger,P., Mark,A.E., Scott,W.R.P. and Tironi,I.G. (1996) *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. vdf Hochschulverlag AG an der ETH Zürich and BIOMOS b.v., Zürich, Groningen.

Vendrell,J., Cuchillo,C.M. and Avilés,F.X. (1990) The tryptic activation pathway of monomeric procarboxypeptidase A. *J. Biol. Chem.*, **256**, 6949–6953.

Villegas,V., Vendrell,J. and Avilés,F.X. (1995) The activation pathway of procarboxypeptidase B from porcine pancreas: Participation of the active enzyme in the proteolytic processing. *Protein Sci.*, **4**, 1792–1800.

Vitols,K.S., Haag-Zeino,B., Baer,T., Montejano,Y.D. and Huennekens, F.M. (1995) Methotrexate-alpha-phenylanine-optimization of methotrexate prodrug for a activation by carboxypeptidase A monoclonal antibody conjugate. *Cancer Res.*, **55**, 478–481.

Received on July 1, 1997; revised on August 28, 1997