# CAPÍTOL 1:

# EXPRESSIÓ DE LES PROCARBOXIPEPTIDASES A1 I B HUMANES EN *P. pastoris* I ESTUDI DETALLAT DELS SEUS PROCESSOS D´ACTIVACIÓ

# 1.1.- SUMMARY

The cDNAs of human PCPA1 and PCPB have been cloned, expressed and secreted using the pPIC9 vector of the methylotrophic yeast *Pichia pastoris*. In the case of PCPB a major protein of identical size and N-terminal sequence as the form isolated from natural sources is obtained, while several side forms of the proenzyme, containing C-terminal fragments of the $\alpha$-MF signal for secretion, are also expressed along with the mature PCPA1. Both recombinant procarboxypeptidases were purified to homogeneity using a two-step chromatographic method.

*Pichia pastoris* provided an efficient system for the production of tens of milligrams of both proteins per liter of cell culture and rendered enzymes with identical properties to those isolated from natural sources. The amount of recombinant human PCPA1 and PCPB obtained allowed crystallization studies and a detailed characterization of their tryptic activation processes.

Previous detailed studies on the tryptic activation mechanism of the porcine proenzymes showed that the proteolytic processing of PCPB to a mature enzyme is a much faster process than that of PCPA1, the former following a monotonic curve and the latter a biphasic curve, as a result of the activation segment of porcine PCPA1 being a powerful competitive inhibitor of the enzyme moiety. The study reported here shows that, while human PCPB follows the general behaviour described above, this is not the case for human PCPA1 whose proteolytic processing, in terms of activation conditions, velocity of the process and shape of the activation curve, ressembles that of PCPB, as happens in the activation process of PCPA2.

A reversed-specificity mutant of human PCPB, named D255K, was expressed and purified in the same manner as the wild-type form in order to study the effect of this point mutation on the tryptic activation process of PCPB. These studies demonstrate the involvement of the generated CPB in the trimming of the severed pro-segment.

# 1.2.- EXPERIMENTAL PROCEDURES

## 1.2.1.- CLONING, EXPRESSION IN *P. pastoris* AND PURIFICATION OF HUMAN PCPA1, PCPB AND THE D255K MUTANT

### 1.2.1.1.- Plasmid constructs

DNA manipulations were performed essentially as described by Sambrook *et al.* (1989), using the *E. coli* strand MC1061 as host. The cDNAs of both human PCPA1 and PCPB were amplified by PCR from a pUC9 vector in order to introduce a *Xho I* site at the 5´ end and an *EcoRI* site at the 3´ end of the cDNAs. For PCPA1 the primers used were: sense primer, 5´-GTATCTCTCGAGAAAAGAAAGGAGGACTTTGTGGGG-3´, antisense primer, 5´-CATGAATTCTTTGGTTGCCTGGATGGG-3´; and for PCPB: sense primer, 5´-GTATCTCTCGAGAAAAGACATCATGGTGGTGAGCAC-3´, antisense primer, 5´-CATGAATTCTGAAACAAGGCCATCAGC-3´. After restriction enzyme digestion of the PCR products, the cDNAs were subcloned in the pBluescript SK+ vector in order to confirm the entire sequence and the changes made in the PCR reaction. Sequencing was performed on an automated DNA sequencer (Amersham Pharmacia Biotech). Both constructs in pBluescript were digested by *XhoI* and *EcoRI* and the cDNAs were ligated to the *P. pastoris* shuttle expression vector pPIC9, which provides the $\alpha$-mating factor (MF) signal for secretion and the *HIS4* gene for selection of the recombinant yeast clones. A scheme of the cloning procedure is presented in Figure 1.1.

### 1.2.1.2.- PCR mutagenesis

A reversed-specificity mutant form of PCPB was obtained by means of a two-step PCR mutagenesis procedure using pBluescript-PCPB as template. The sense and antisense external primers were the same as the ones used to clone the wild-type form. The two following internal primers were designed to contain the single mutation D255K: sense, 5´-GGAATTATAATCCAAGTGACAGAGCTTCTC-3´; antisense, 5´-CACTTGGATTATAATTCCATAGGTACG-3´. The PCR product was first cloned into pBluescript to confirm the construct by automated DNA sequencing and subsequently cloned into pPIC9, in the same way as for the wild-type form, to obtain pPIC9-D255K.

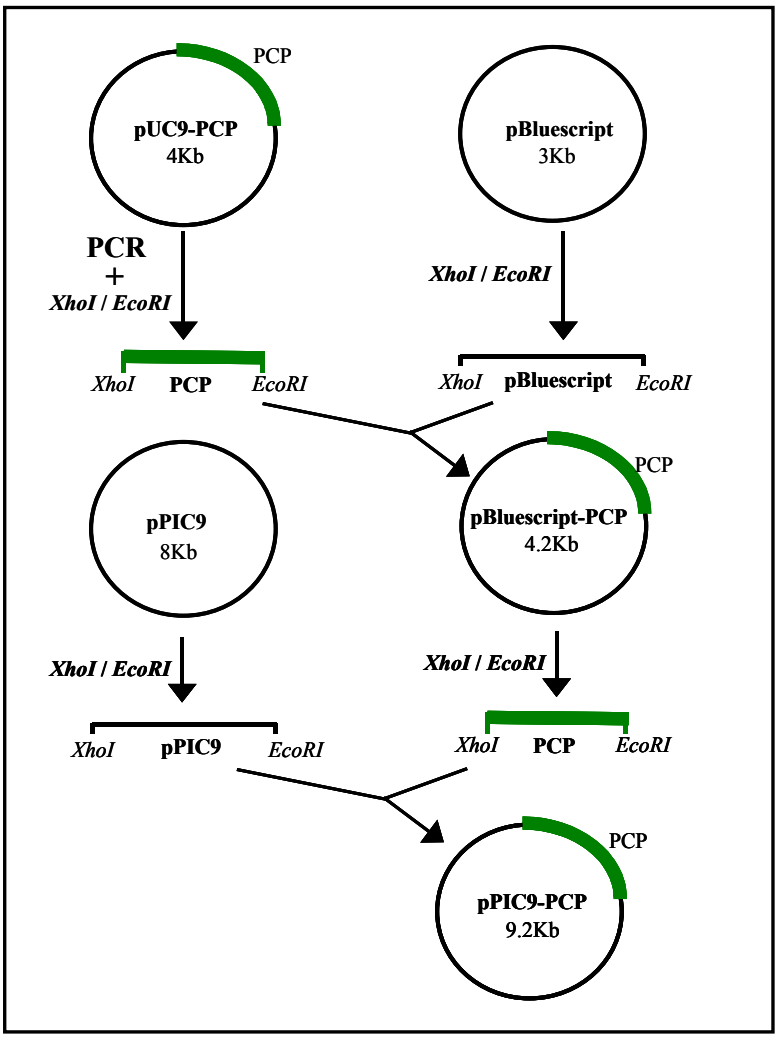


**Figure 1.1.- Cloning of the cDNAs of human PCPA1 and PCPB in the *P. pastoris* shuttle expression vector pPIC9. The cDNA of either PCPA1 or PCPB is denoted by PCP.**

### 1.2.1.3.- Transformation and selection of productive clones

pPIC9-PCPA1, pPIC9-PCPB and pPIC9-D255K were linearized by *SacI* digestion and transformed into *P. pastoris* KM71 (*arg4*, *his4*, *aox1::ARG4*) strain using the spheroplast method. Histidine-independent transformants were selected and tested for production of protein. Colonies were grown in 15 mL of BMGY buffered liquid medium (1% yeast extract, 2% peptone, 1% glycerol, 1.34% Yeast Nitrogen Base, $4x10^{-5}$ % biotin and 100 mM potassium phosphate, pH 6.0) at 30 ºC for two days. Cells were collected by centrifugation and gently resuspended in 3 mL of BMMY buffered liquid medium (same as BMGY medium but containing

1% methanol instead of 1% glycerol) and cultured for another two days to induce the expression of recombinant protein. The production of the clones was followed by SDS-PAGE analysis of the supernatants on 12% polyacrylamide gels. Western blots were carried out as previuosly described using 1:500 anti-human pancreatic procarboxypeptidases antiserum (Catasús *et al.*, 1995). The functionality of the expressed proteins was analyzed with synthetic substrates after activation of the proenzymes with trypsin. FAPP was used for PCPA1, BGA for PCPB and benzoyl-alanyl-L-glutamic acid for the D255K mutant.

### 1.2.1.4.- Expression and purification of recombinant human PCPA1, PCPB and the D255K mutant

One-liter shake-flask cultures were grown for three days in buffered glycerol medium BMGY. Cells were collected by centrifugation at 3000xg, gently resuspended in 200 mL of methanol containing medium, BMMY, and cultured for another two days to induce production of recombinant protein. For the three studied proteins, the same two-step chromatographic purification approach was used: a hydrophobic interaction chromatography on a Toyopearl butyl 650M column followed by an FPLC chromatography on a preparative anion exchange column (TSK-DEAE 5PW). The supernatant, where the recombinat protein is secreted, was separated from the cells by centrifugation and was brought to 30% saturation of ammonium sulphate before loading onto the butyl column. Elution was performed with a decreasing gradient between buffer A (Tris-HCl 20 mM, 30% saturation of ammonium sulphate pH 7.0) and buffer B (Tris-HCl 20 mM pH 7.0). Fractions containing the proenzymes were selected by detection of carboxypeptidase activity after trypsin activation of a 50 μL aliquot of each fraction. These fractions were dialysed overnight at 4 ºC against the buffer A used in the following FPLC chromatography. In this second chromatography, the proenzymes were eluted using an increasing gradient of ammonium acetate between buffer A (Tris-acetate 20 mM pH 8.0 for human PCPA1 and Tris-acetate 20 mM pH 7.5 for human PCPB and the D255K mutant) and buffer B (Tris-acetate 20 mM, ammonium acetate 0.8 M pH 8.0 for human PCPA1 and Tris-acetate 20 mM, ammonium acetate 0.4 M pH 7.5. for human PCPB and the D255K mutant). The gradient used for human PCPA1 was: 0% B from 0 to 10 min, 10% B from 10 to 20 min and 30% B from 20 to 100 min, and for human PCPB and the D255K mutant: 0% B from 0 to 20 min and 10% B from 20 to 120 min. The identity and correct precessing of the different recombinant proteins was confirmed by automated Edman degradation analysis of their N-terminal sequence after blotting the samples on polyvinylidene difluoride membranes. A Beckman LF3000 Protein Sequencer was used. In order to avoid spontaneous activation, the proenzymes were kept precipitated in 43% ammonium sulphate at 4 ºC.

## 1.2.2.- ACTIVATION STUDIES OF RECOMBINANT HUMAN PCPA1, PCPB AND THE D255K MUTANT

### 1.2.2.1.- Activation experiments

The precipitated proenzymes were recovered by centrifugation at 7000xg during 30 min at 4ºC, resuspended in 50 mM Tris-HCl,1 μM $ZnCl_2$ pH 8.0 buffer and dialysed against this buffer overnight. Activation was performed using bovine TPCK treated trypsin from Worthington. Trypsin was dissolved in 1 mM HCl, 1 mM CaCl2 pH 3.0 buffer to avoid autolysis. The different proenzymes at 1 mg/mL were treated with trypsin at 40/1 or 400/1 ratios (w/w) at 25 or 0ºC. Aliquots were removed at given times after trypsin addition for activity measurements and for electrophoretic, reversed-phase HPLC and mass spectrometry analysis. For human PCPA1, samples were also removed for quantification of the released amino acids. For activity measurements 10 μL of the activation mixture were added to 190 μL aprotinin (BPTI) at 0.1 mg/mL in 20 mM Tris/0.1 M NaCl (pH 7.5) and 10 μL of this new mixture were used to perform spectrophotometric activity measurements with the synthetic substrates mentioned above. In the case of the D255K mutant, which shows poor enzymatic activity, 100 μL of the activation mixture were added to 2 μL of BPTI at 6 mg/mL and 20 μL of this mixture were used for activity measurements with benzoyl-alanyl-L-glutamic acid. For electrophoretic analysis, 20 μL of the activation mixture were mixed with 2 μL of 22 mM TLCK in water to reach a final trypsin inhibitor concentration of 2 mM. For reversed-phase HPLC analysis, 120 μL samples were removed from the activation mixture, made 0.5% in TFA to inhibit proteolysis and kept at -20ºC until analysis. For quantification of the amino acids released into the activation mixture, 100-μL samples were taken (2 nmol of initial proenzyme).

### 1.2.2.2.- Electrophoretic and N-terminal sequence analyses

Each sample removed from the activation mixture was mixed with electrophoretic loading buffer (containing 1% SDS and 3% β-mercaptoethanol), heated at 100 ºC for 1 min and stored at –20 ºC until analysis. Electrophoresis was carried out in polyacrylamide tricine gels. N-terminal sequence analysis was performed by blotting the samples on polyvinylidene difluoride membranes after SDS-PAGE analysis. A Beckman LF3000 Protein Sequencer was used.

### 1.2.2.3.- Chromatographic analysis by reversed-phase HPLC and mass spectrometry

Samples removed from the activation mixture were analyzed by reversed-phase HPLC on a Vydac C4 column (250 x 54 mm, 5 μm particle size, 300 Å pore). Elution was followed at 214

nm. Chromatographies were performed in 0,1% TFA with an eluting linear gradient between water (solvent A) and 90% acetonitrile (solvent B), according to the following steps: 10% B from 0 to 10 min and 57% B at 130 min. HPLC purified fragments were analyzed by mass spectrometry using a MALDI-TOF spectrometer (Biflex with reflectron, from Bruker) using 50% synapinic acid as matrix.

#### 1.2.2.4.- Amino acid analysis

To each sample removed from the activation mixture, 1 nmol of methionine was added as a quantitative reference before the addition of six volumes of ethanol to precipitate proteins and large peptides. The supernatant was lyophilized and analyzed for amino acid composition using the dabsyl derivatization method (Vendrell and Avilés, 1986). A reversed-phase NovaPak $C_{18}$ column was used to separate the amino acids produced during the activation.

## 1.3.- RESULTS AND DISCUSSION

### 1.3.1.- CONSTRUCTION OF THE EXPRESSION VECTORS, EXPRESSION IN *Pichia pastoris* AND PURIFICATION OF RECOMBINANT HUMAN PCPA1, PCPB AND THE D255K MUTANT

The cDNAs encoding human PCPA1 and PCPB have been previously cloned from a pancreatic library by immunological and radioactive approaches (Catasús *et al.*, 1992; Aloy *et al.*, 1998). The PCPB sequence obtained corresponds to that of the allelomorphic B1 form (Pascual *et al.*, 1989). Since the differences between the B1 and B2 forms are limited to a slightly different isoelectric point and to a few amino acid substitutions and the complete sequence of the B2 form is unknown, the name PCPB will be kept for clarity. Both cDNAs of PCPA1 and PCPB were modified by PCR to add a *XhoI* site at the 5´ end and an *EcoRI* site at the 3´ end in order to be cloned into the *P. pastoris* expression vector pPIC9. On the other hand, a reversed-specificity mutant of PCPB, D255K, was obtained using a PCR approach and cloned into pPIC9. The three recombinant plasmids were linearized by *SacI* digestion and transformed into KM71 (*arg4, his4, aox1::ARG4*) *P. pastoris* strain by the spheroplast method.

Transformed colonies were screened for protein secretion. In the three cases a dominant 45 kDa protein was secreted upon induction by methanol, as detected by means of SDS-PAGE analysis of the extracellular medium. These proteins were identified as human

procarboxipeptidases by Western Blot analysis. N-terminal sequence analysis of the secreted recombinant proteins showed that for PCPB and the D255K mutant the α-MF signal for secretion was correctly processed, but in the case of PCPA1 several protein products were identified which contained in their N-termini fragments of this signal peptide. This heterogeneity in the processing of the α-MF signal for secretion has also been reported for recombinant human PCPA2 expressed in *P. pastoris* (Reverter *et al.*, 1998). High productivity clones were selected for human PCPA1, PCPB and the D255K mutant and used for large scale productions of recombinant protein. The follow-up of the expression of recombinant human PCPA1 and PCPB after methanol induction by SDS-PAGE is presented in Figure 1.2.

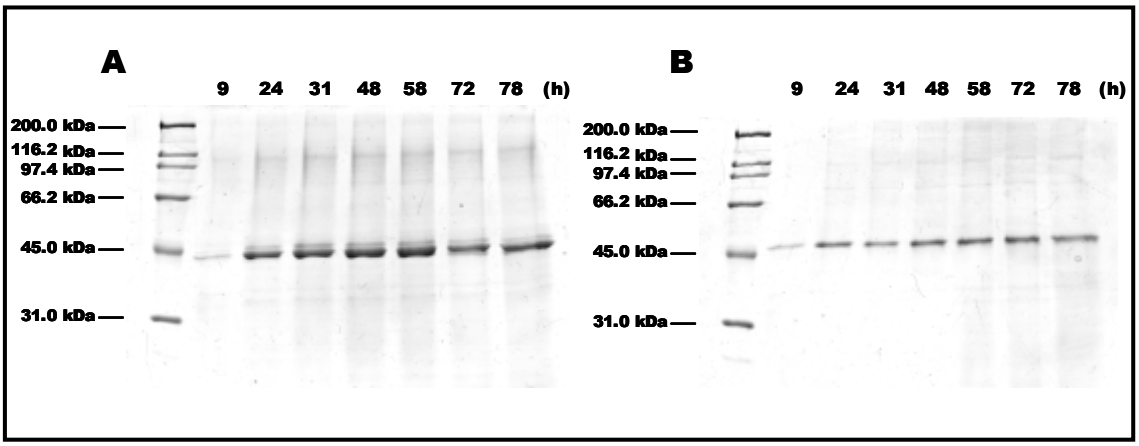


**Figure 1.2.- Analysis of the production of recombinant human PCPA1 (A) and PCPB (B) in *P. pastoris* by SDS-PAGE. Supernatant samples of the cultures were taken at given times after methanol induction. The samples were directly treated with electrophoesis loading buffer for 2 minutes at 100 ºC and a volume equivalent to 15 μL was loaded in the electrophoresis wells. Time is indicated in hours.**

Purification of the proenzymes was performed with a two-step chromatographic approach, consisting of an atmospheric hydrophobic interaction chromatography on a butyl column, followed by FPLC anion-exchange chromatography on a TSK DEAE-5PW preparative column, as indicated in the "Experimental procedures" section. For PCPB, addition of protease inhibitors such as PMSF and Pefabloc to the supernatant prior to purification was necessary to avoid unespecific proteolytic activation of the proenzyme. Figure 1.3 shows the hydrophobic interaction and FPLC chromatograms obtained during purification of PCPB. On the other hand, in order to purify the native, well-processed PCPA1 from the incorrectly processed forms, optimization of the first chromatographic step was necessary, which was achieved by eluting the protein with a two-step slow decreasing gradient of ammonium sulphate. In both cases, the

chromatographic procedure used was sufficient to obtain high purity protein, suitable for activation and crystallization studies of the proenzymes, which were fully functional after tryptic activation. Starting from one liter cultures, a final yield of 15-20 mg of purified PCPA1 and 10-15 mg of purified PCPB was obtained. The D255K mutant was purified in the same manner as the wild-type form, and the final yield obtained was similar in both cases.

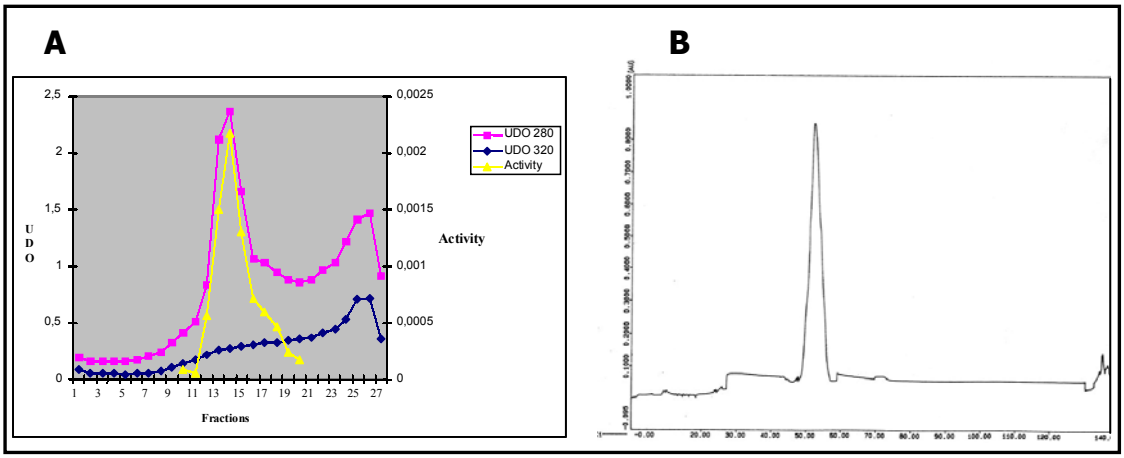


**Figure 1.3.- Hydrophobic interaction (A) and FPLC (B) chromatograms obtained for the purification of recombinant human PCPB. For details see "Experimental procedures".**

Human carboxypeptidases are enzymes of biomedical and biotechnological interest and their application as prodrug activators for the antibody-directed enzyme prodrug therapy (ADEPT) of cancer has been studied for the wild-type forms as well as for mutants designed to cleave specific prodrugs (Smith *et al.*, 1997; Edge *et al.*, 1998; Wolfe *et al.*, 1999; Wright and Rosowsky, 2002). Hence, the development of an efficient recombinant expression system for these enzymes is highly desirable. Since the pro-regions of carboxypeptidases, through their high folding capability (Villegas *et al.*, 1995b), act as intramolecular chaperones that drive the correct folding of their own protease domains, the best approach to obtain active carboxypetidases is the recombinant expression of the proenzymatic forms followed by proteolytic activation. The first overexpression system optimized for a human procarboxypeptidase was that of PCPA2, developed by our group using *P. pastoris* (Reverter *et al.*, 1998). Now, efficient expression of human PCPA1 and PCPB has also been achieved using *P. pastoris* and, therefore, a high expression system for the obtention of recombinant protein is available for the whole pancreatic human system. In addition, a reversed-specificity variant of human PCPB, named D255K, which has the potential to be used in ADEPT (Edge *et al.*, 1998), was also efficiently expressed with the *P. pastoris* system.

### 1.3.2.- TRYPSIN ACTIVATION OF RECOMBINANT HUMAN PCPA1

In contrast to previous studies on porcine PCPA1 (Vendrell *et al.*, 1990c) the action of trypsin on human PCPA1 at 25 ºC and at a 40/1 (w/w) PCPA1/trypsin ratio is too quick to allow the follow-up of the activation process. Screening of the temperature and PCPA1/trypsin ratios led to the conclusion that activation must be performed at 0 ºC and at a 400/1 ratio to obtain a detailed analysis of the process. The appearance of carboxypeptidase activity followed a monotonic activation curve which correlated with the proteolytic severing of the pro-segment from the active enzyme, as observed with Tricine electrophoresis analysis (Figure 1.4). The time needed to reach 50% of potential carboxypeptidase activity is about 20 minutes. The generated pro-segment appears as a single band in the Tricine/SDS-PAGE, but, as shown by reversed-phase HPLC and mass spectrometry analysis, two species with similar molecular masses are formed. N-terminal sequence analysis of the generated CPA1 indicates that the primary tryptic cleavage of the proenzyme occurs at the Arg99A-Ala1 peptide bond (porcine PCPA1 numbering, Guasch *et al.*, 1992, Reverter *et al.*, 1998).

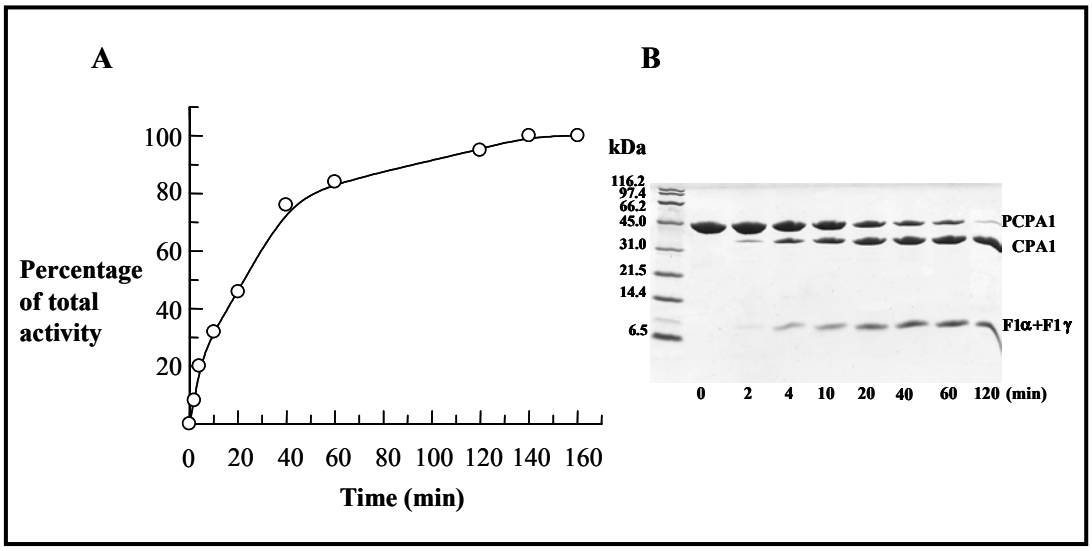


**Figure 1.4.- Generation of mature CPA1 during tryptic activation of human PCPA1. (A) Measure of activity generation using 0,2 mM FAPP as substrate. (B) Follow-up of the activation process using Tricine/SDS-PAGE. The species generated during the proteolytic processing of PCPA1 are identified on the right side of the gel. On the left side, the relative masses of the molecular weight markers are indicated.**

The processing of the pro-segment was followed by reversed-phase HPLC (Figure 1.5). These chromatographic analyses show that only two fragments are generated throughout the

activation process. These fragments were identified by mass spectrometry as the 94-residues primary fragment (F1$\alpha$, Lys4A-Arg99A), generated as a result of the first tryptic cleavage, and a 92-residues fragment (F1$\gamma$, Lys4A-Arg97A), which appears after trimming of the complete fragment, F1$\alpha$. Analysis of the amino acids released during the activation course was performed and generation of an arginine and a serine was detected. The identity, relative quantities and kinetics of appearance of these amino acids indicate that they are released as a result of the C-terminal processing of fragment F1$\alpha$ by the action of the CPA1 generated, giving rise to fragment F1$\gamma$.

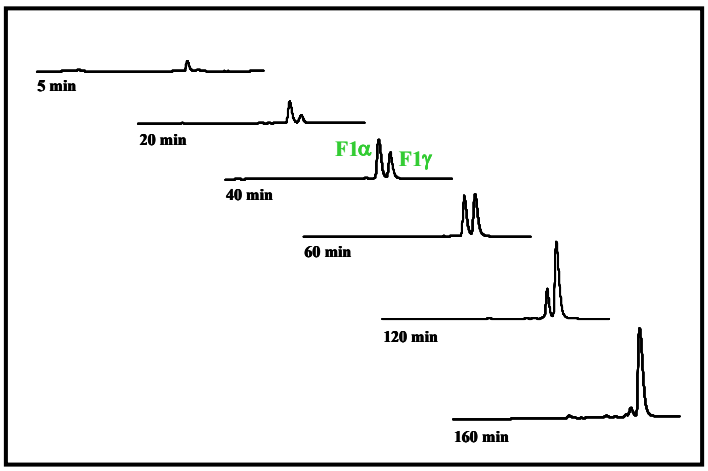


**Figure 1.5.- Analysis by RP-HPLC of the fragments generated during tryptic activation of human PCPA1. F1$\alpha$ corresponds to the 94-residues primary fragment and F1$\gamma$ to the 92-residues fragment.**

Generation of CPA1 activity is very quick for human PCPA1 at 25 ºC and at a 40/1 (w/w) proenzyme/trypsin ratio, while, in the same conditions, it is rather slow for the porcine proenzyme (Vendrell *et al.*, 1990c). In fact, in order to allow the follow-up of the activation process of the human proenzyme, activation must be performed at 0 ºC and at a 400/1 (w/w) proenzyme/trypsin ratio. A schematic representation of the tryptic activation mechanism of human PCPA1 is presented in Figure 1.6. The first target for trypsin action is Arg99A, releasing the intact pro-segment of 94 residues, named F1$\alpha$. The mature human CPA1 generated during the activation course sequentially cleaves the C-terminal residues Arg99A and Ser98A of the complete fragment F1$\alpha$, as shown by the release of these free amino acids into the activation medium. As a result of the carboxypeptidase action, fragment F1$\gamma$ is formed, which is resistant to further proteolysis under these conditions. The unexpected ability of human CPA1, an enzyme specific for C-terminal hydrophobic residues, to remove the arginine residue of the C-

terminal end of the pro-segment has also been observed for the porcine enzyme (Vendrell *et al.*, 1990c).

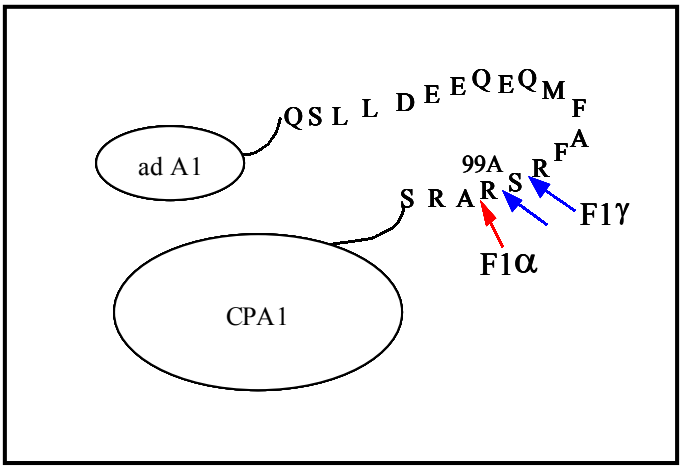


**Figure 1.6.- Schematic representation of the cleavage points in the tryptic activation process of human PCPA1. The sequence of the mainly α-helical connecting region between the globular activation domain and the active enzyme is shown. Cleavages performed by trypsin and carboxypeptidase action are indicated by red and blue arrows, respectively. The fragments detected as a result of these cleavages are indicated. The globular activation domain of PCPA1 is denoted by *ad A1* (activation domain A1).**

A detailed study on the tryptic maturation course of porcine PCPA1 has been previously reported (Vendrell *et al.*, 1990c). It was observed for this proenzyme a slow, biphasic activation behaviour, which was attributed to the inhibitory action of the released pro-segment and other large fragments derived from it on the generated enzyme (San Segundo *et al.*, 1982; Vilanova *et al.*, 1985). Thus, a second internal tryptic cleavage at Arg74A was necessary to attain full activity. In contrast to these observations in the porcine proenzyme, the appearance of mature human CPA1 during the activation process, as detected by Tricine/SDS-PAGE, coincides with the increase of CPA1 activity, indicating that the first trypsin cleavage is sufficient to achieve maximum CP activity. Therefore, it can be concluded that the severed pro-segment is not an inhibitor of the active human enzyme. In fact, a secondary tryptic target is not present in the human pro-segment neither around position 74A nor in the α-helix of the connecting region that links the globular activation domain to the enzyme moiety (Figure 1.7). An arginine residue is present at position 97A, but this is not a tryptic cleavage point, as shown by our studies. The

absence of an internal secondary target in the human proenzyme indicates, in accordance to our studies, that such a cleavage is not necessary for the enzyme to attain full activity.

Human and porcine PCPA1 share an 82% of sequential identity and, therefore, they must have very similar overall structures. However, some particular interactions must be different in order to explain the observation that the pro-segment of the human proenzyme does not bind to the enzyme once severed from it, in contrast to the corresponding segment of porcine PCPA1, which strongly binds and inhibits the enzyme, determining a slow, biphasic activation process. Thus, human PCPA1 shows an activation behaviour similar to that of B and A2 forms, as in all these cases a fast monotonic activation process is observed, partly because of the pro-segment being unable to inhibit the generated enzyme.

The resolution of the crystal structures of porcine PCPB (Coll *et al.*, 1991), porcine PCPA1 (Guasch *et al.*, 1992) and human PCPA2 (Garcia-Saez *et al.*, 1997) has been of key importance to unveil the structural determinants of the differential activation behaviour of these three proenzymes. From these three-dimensional structures, two distinct regions of the pro-segments could be identified as responsible for the interaction with the enzyme: the globular activation domain and the connecting region that covalently links both globular moieties. Although different studies pointed out to one or the other region as the most relevant in determining the activation behaviour, recent studies led to the conclusion that both regions play an important role (Ventura *et al.*, 1999). The connecting region is structured in an $\alpha$-helix followed by a loop where the first tryptic action takes place. It has been described that the length of this $\alpha$-helix, which comprises 14 residues in PCPA1 and 8 residues in PCPB, is of key importance in determining the differential activation behaviour observed between porcine PCPA1 and PCPB, since, in the former case, it is better structured and this may help to maintain the strong contacts between the connecting region and CPA1. A prediction of the secondary structure of this $\alpha$-helix for human PCPA1 has been performed (Rost *et al.*, 1994) and its predicted length is the same as that of the porcine proenzyme, indicating that an explanation for their differential activation behaviour must be found somewhere else. It is even clear that the calculated helical propensities (Muñoz and Serrano, 1997) do not point to a lack of helical structure for human PCPA1. An analysis of the sequence of the human proenzyme reveals changes in some residues that were found in porcine PCPA1 to be involved in important interactions between the pro-segment and the carboxypeptidase moiety (Figure 1.7). The altered interactions are Lys4A--Asp122, Glu5A--Asp122, Arg47A--Ile244 and Phe50A--Val246 between the globular activation domain and the CP moiety and Ser96A--Asn8, Ser96A--Ala10 and Ser96A--Thr11 involving the connecting region (Guasch *et al.*, 1992), as in the human proenzyme Thr122, His47A, Ser50A and Phe96A are found. However, until the three-dimensional structure of human PCPA1 is solved, it is not possible to assess the relevance of these altered interactions. We think that,

since the human and porcine structures must be very similar, the peculiarities detected when comparing both structures are likely to be of essential importance in determining their differential activation behaviour. Currently, crystallization studies on human PCPA1 are being performed.

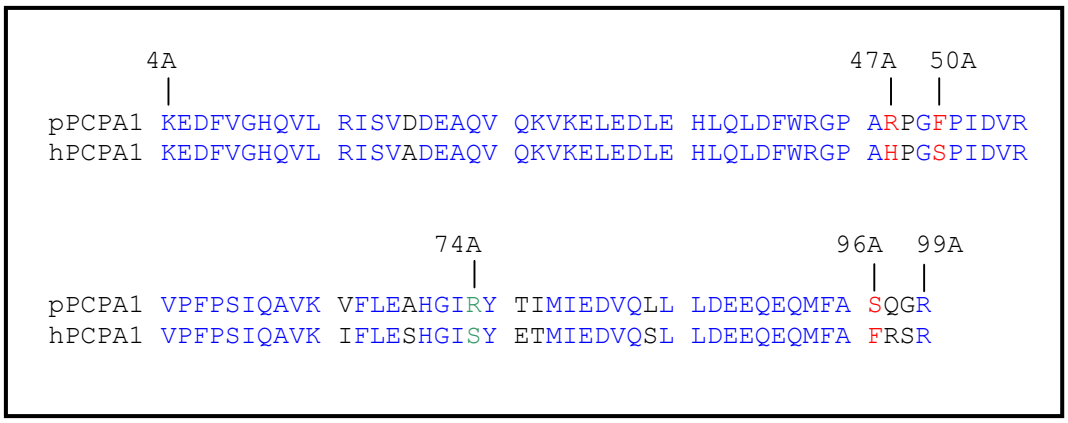


**Figure 1.7.- Comparison of the amino acid sequences of the pro-segments of porcine and human PCPA1. Identical residues are coloured in blue, residues at positions 47A, 50A and 96A, involved in important interactions between the pro-segment and the CP moiety in porcine PCPA1 and altered in the human proenzyme, are highlighted in red and residue at position 74A, where the second tryptic action takes place in porcine PCPA1, is coloured in green. Porcine PCPA1 numbering (Guasch *et al.*, 1992).**

### 1.3.3.- COMPARATIVE STUDY OF THE ACTIVATON PROCESSES OF RECOMBINANT HUMAN PCPB AND ITS REVERSED-SPECIFICITY VARIANT, D255K

It has been previously reported that human PCPB requires the addition of relatively large quantities of trypsin to generate its maximum activity (Pascual *et al.*, 1989), in contrast to other homologous PCPB from other species, which are fully activated with small amounts of trypsin. Thus, activation of recombinant human PCPB was first studied at 25 ºC at 4/1 (w/w) PCPB/trypsin ratio, but under these conditions the activation process is excessively quick for detailed studies. Several experiments at 0 ºC using different PCPB/trypsin ratios showed that the optimum conditions for tryptic activation of human PCPB are 0 ºC at a 400/1 ratio, as in the case of other B forms previously studied (Burgos *et al.*, 1991; Villegas *et al.*, 1995a; Opezzo *et al.*, 1994). This discrepancy may be due to the difficulty in obtaining sufficient and sufficiently pure quantities of PCPB from natural sources.

The maturation course of human PCPB follows a quick and monotonic activation curve and the generation of CPB, as observed with Tricine/SDS-PAGE, coincides with the generation of activity

(Figure 1.8 and Figure 1.9A). The first tryptic cleavage occurs at the Arg95A-Ala4 peptide bond (porcine PCPB numbering, Coll *et al.*, 1991), as shown by N-terminal sequence analysis of the generated CP. The electrophoretic follow-up of the process shows the generation of two species derived from the pro-segment. Reversed-phase HPLC analysis (Figure 1.10A) reveals that the high molecular weight band contains two different species that were identified by mass spectrometry analysis as the complete 95-residues fragment generated by the first tryptic cleavage (F1$\alpha$, His1A-Arg95A) and a 94-residues fragment (F1$\beta$, His1A-Val94A) resulting from

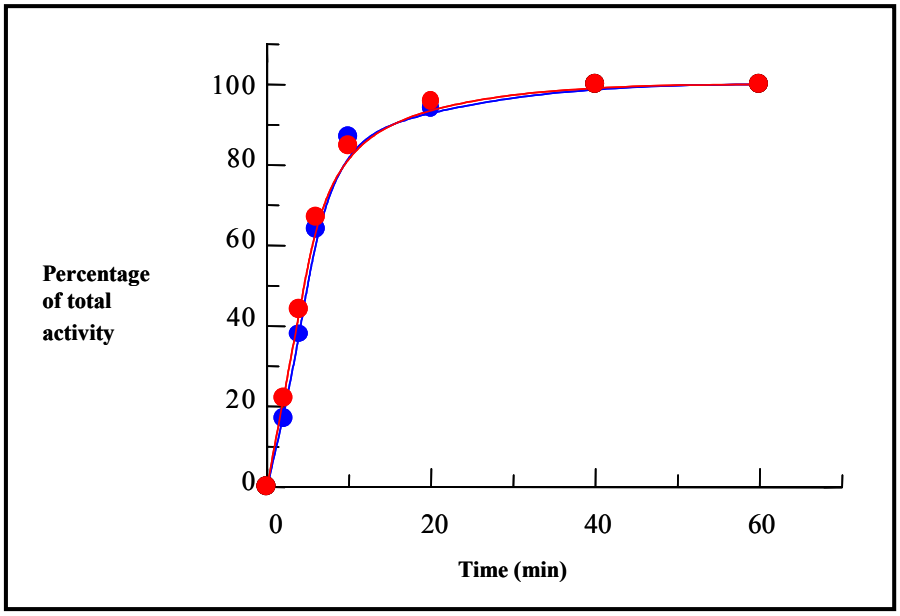


**Figure 1.8.- Comparison of the time course of CPB activity generation after tryptic activation of wild-type human PCPB (in blue) and its reversed-specificity variant, D255K (in red). Activity measures were performed using 1 mM BGA and 0.75 mM benzoyl-alanyl-glutamic acid as substrate for the wild-type and mutant form, respectively.**

the C-terminal processing of Arg95A of the complete pro-segment. The other fragment detected by Tricine/SDS-PAGE and reversed-phase HPLC was identified as a 82-residues fragment (F2$\beta$, His1A-Leu82A). This fragment is formed by a second tryptic cleavage of the high molecular species at Arg83A and subsequent rapid removal of this C-terminal arginine.

The participation of the generated carboxypeptidase in its own activation process through the trimmimg of the C-terminal arginines that are left after trypsin action was proved for porcine PCPB by comparing the activation process in the absence and the presence of CP inhibitors (Villegas *et al.*, 1995a). In order to study the differences in the activation course of a PCPB variant that generates a mature enzyme unable to cleave C-terminal arginines, trypsin activation of the reversed-specificity human PCPB mutant, D255K, was performed using the

same conditions as for the wild-type form (0ºC, 400/1 (w/w) proenzyme/trypsin ratio). The time course of CP activity generation was very similar to that of the wild-type form (Figure 1.8) and for both cases the time required to attain 50% of potential CP activity was about 10 min. The N-terminal sequence of the electrophoretic D255K-CPB bands, obtained from samples at different activation times, was always Ala4-Thr5-Gly6 (porcine PCPB numbering, Coll *et al.*, 1991; residue at position 4 corresponds in B forms to the first residue of the enzyme moiety). This result pointed to the Arg95A-Ala4 peptide bond as the first point of cleavage of the mutant proenzyme by trypsin, as in the case of the wild-type form. Two activation fragments were detected by means of Tricine/SDS-PAGE (Figure 1.9B) and reversed-phase HPLC analysis (Figure 1.10B, which were identified by mass spectrometry analysis as a 93-residues fragment (F1, His1A-Arg93A) and a 83-residues fragment (F2, His1A-Arg83A). While F2 fragment is formed as a consequence of the same secondary tryptic cleavage at Arg83A that takes place in the wild-type form activation process, F1 fragment arises from an unexpected trypsin action at Arg93A.

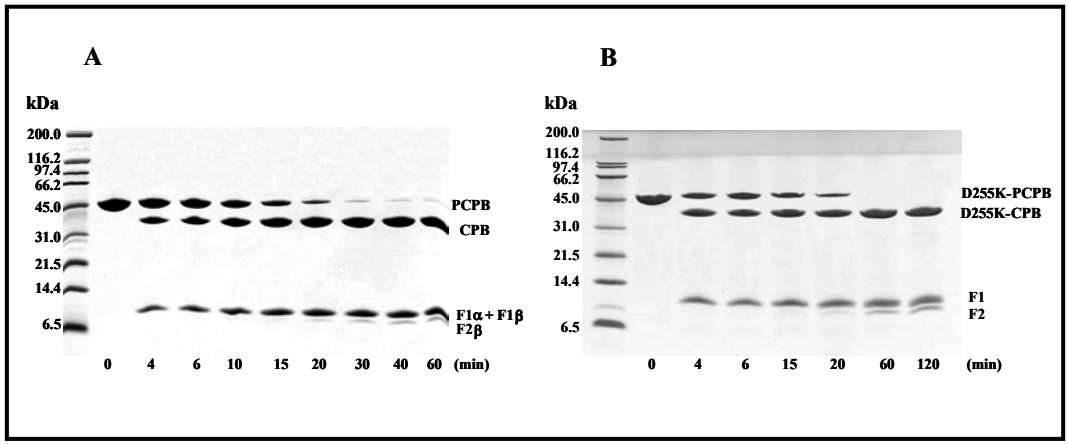


**Figure 1.9.- Electrophoretic follow-up of the activation process of human PCPB (A) and the D255K mutant (B) on Tricine/SDS-polyacrylamide gel. The species generated during the proteolytic processing of the proenzymes are identified on the right side of the gels. On the left side, the relative masses of the molecular weight markers are indicated.**

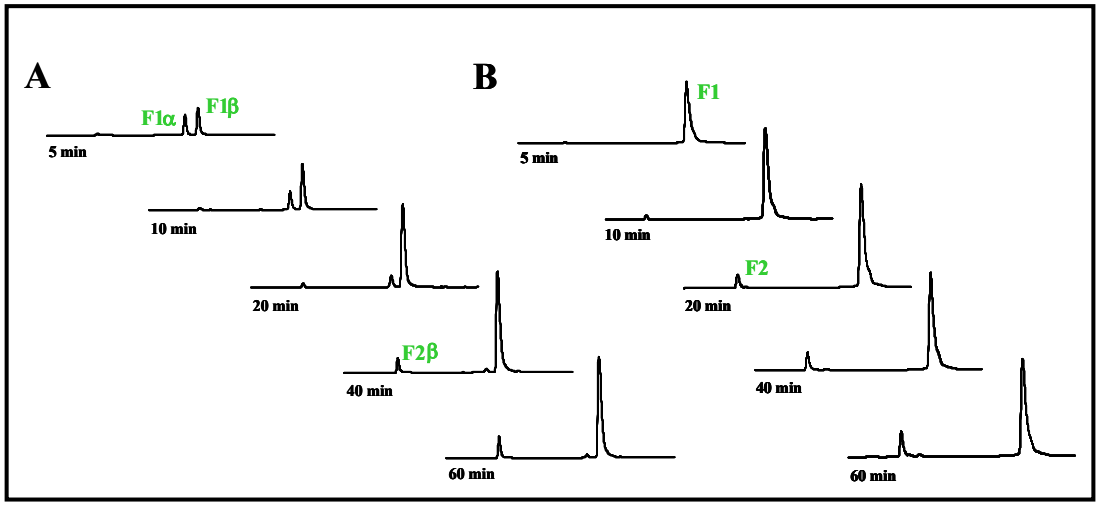


**Figure 1.10.- Analysis by RP-HPLC of the fragments generated during tryptic activation of human PCPB (A) and the D255K mutant (B). In (A), F1α corresponds to the 95-residues primary fragment, F1β to the 94-residues fragment and F2β to the 82-residues fragment and in (B), F1 and F2 correspond to the 93- and 83-residues fragments, respectively.**

Previous studies on the activation process of B forms showed that mild conditions are required (0 ºC, 400/1 (w/w) proenzyme/trypsin ratio) to allow the follow-up of the process (Burgos *et al.*, 1991; Villegas *et al.*, 1995a; Opezzo *et al.*, 1994). Human PCPB seemed an exception to this behaviour, as derived from the preliminary study on the activation process of the human procarboxypeptidase complement (Pascual *et al.*, 1989), which showed that for PCPB a 4/1 (w/w) proenzyme/trypsin ratio was necessary to generate its maximum activity. It must be pointed out that these studies were carried out with a proenzyme concentration of 45 μg/mL, instead of 1 mg/mL, which is the concentration that is generally used for procarboxypeptidase activation experiments. The work reported here shows that, using the standard proenzyme concentration of 1 mg/mL, human PCPB follows a quick activation course and attains its maximum potential activity at 0 ºC and at a 400/1 (w/w) proenzyme/trypsin ratio. These results indicate that the kinetics of the activation process of human PCPB is sensitive to changes in zymogen concentration.

A scheme of the cleavages that take place during human PCPB tryptic activation process is presented in Figure 1.11. Human PCPB presents a general activation behaviour very similar to that of porcine PCPB, whose activation mechanism has been thoroughly studied (Burgos *et al.*, 1991; Villegas *et al.*, 1995a; Ventura *et al.*, 1999; Companys, 2002). Both proenzymes follow a quick monotonic activation process, where appearance of the CPB species coincides with

generation of activity, indicating that the pro-segment does not inhibit the enzyme moiety, and the CP generated participates in the proteolytic processing of the activation fragments released during the activation process. Figure 1.12 shows a comparison of the sequences of the pro-segments of porcine and human PCPB.

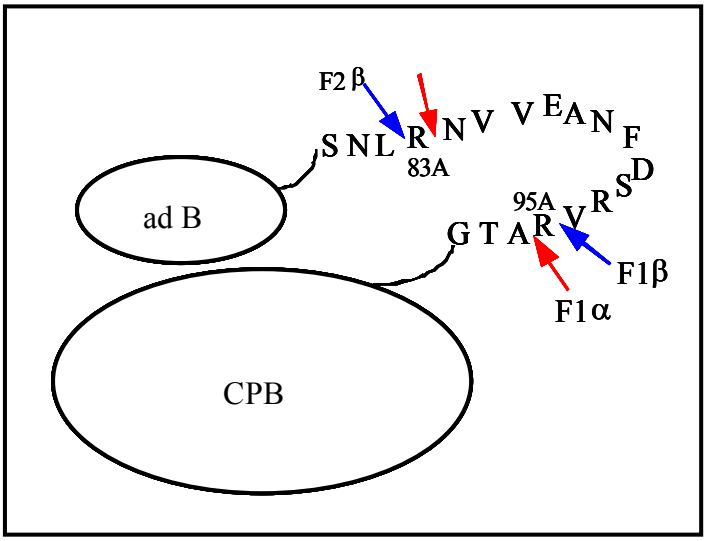


**Figure 1.11.- Schematic representation of the cleavage points in the tryptic activation process of human PCPB. The sequence of the connecting region between the globular activation domain and the active enzyme is shown. Cleavages performed by trypsin and carboxypeptidase action are indicated by red and blue arrows, respectively. The fragments detected as a result of these cleavages are indicated. The globular activation domain of PCPB is denoted by *ad B* (activation segment B).**

Some peculiarities must be pointed out in the activation process of human PCPB. First, a relatively slow transformation of the primary fragment F1α into F1β by CP action on the C-terminal Arg95A residue is observed. This allows detection and isolation of this primary fragment, which cannot be detected during the activation course of porcine PCPB. This observation lead us to the conclusion that human CPB is less efficient in the digestion of its activation segment. On the other hand, for human PCPB it is not observed a total conversion of the high molecular species of the activation segment (F1β) to the final globular domain resistant to proteolysis (F2β), even in the long run of the activation process. In contrast, for porcine PCPB this conversion is almost complete after 60 minutes of trypsin treatment. This has, however, no effect on the overall activity generated since the maximum activity is attained when the primary

activation cleavage has been completed, independently of the degradation state of the pro-segment.

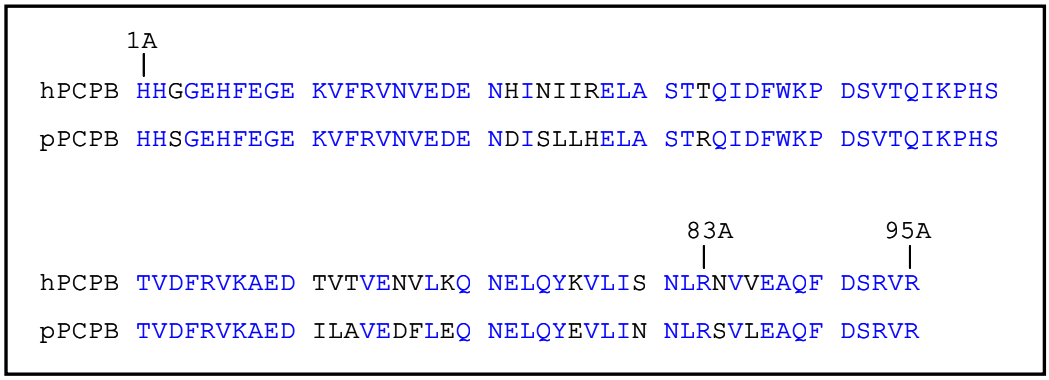


**Figure 1.12.- Comparison of the amino acid sequences of the pro-segments of porcine and human PCPB. Identical residues are coloured in blue. Porcine PCPB numbering (Coll *et al.*, 1991).**

The comparative study on the activation processes of human PCPB and its reversed-specificity variant, D255K, shows that, while the kinetics of activity generation and the chromatographic profiles of appearance of the primary and secondary activation fragments are very similar, a differential processing of these fragments occurs. Fragments F1β and F2β detected during the activation process of the wild-type form are formed as a result of the C-terminal trimming of Arg95A and Arg83A of fragments F1α and F2α, respectively. In the case of the D255K mutant, which is unable to cleave C-terminal arginines, the activation fragments detected keep the C-terminal arginine residue, confirming that wild-type human CPB participates in the processing of its own activation segment.

Interestingly, one of the fragments generated during the activation of the D255K mutant, F1 (His1A-Arg93A), arises from an unexpected tryptic action at Arg93A, which never occurs in the activation process of any wild-type B form. F1 formation implies either a first tryptic action at Arg93A (Figure 1.13A) or a first tryptic action at Arg95A and subsequent rapid action at Arg93A (Figure 1.13B). If the first mechanism is the one that occurs, we should expect detection of a CPB species with the following N-terminus: Val94A-Arg95A-Ala4-Thr5-Gly6 (porcine PCPB numbering, Coll *et al.*, 1991), but, as reported above, the N-terminal sequence detected was always Ala4-Thr5-Gly6. However, it could be possible that a second rapid action of trypsin took place at the Arg95A-Ala4 peptide bond, not allowing the detection of the expected N-terminus (Figure 1.13A). Since no species containing Val94A or Arg95A colud be detected, it is not possible to tell where the first tryptic cleavage takes place, either at the Arg95A-Ala4 or

Arg93AVal94A peptide bond. Though the first case is the one observed for the wild-type form, the second possibility also deserves consideration, since such a primary tryptic action was observed in the activation process of porcine PCPB in presence of CP inhibitors (Villegas *et al.*, 1995a). It was shown in this study that presence of organic or peptidic CP inhibitors in the activation mixture generated a new tryptic target at Arg93A by diffusion of the inhibitors to the active site of the enzyme, affecting the conformation of the activation loop (C-terminal part of the pro-segment) by a long-range effect. For our mutant, it could be possible that the introduction of a mutation in the active site of the enzyme caused a distortion of the interaction between the pro-segment and the carboxypeptidase moiety, altering the conformation of the activation loop and generating a new tryptic target.

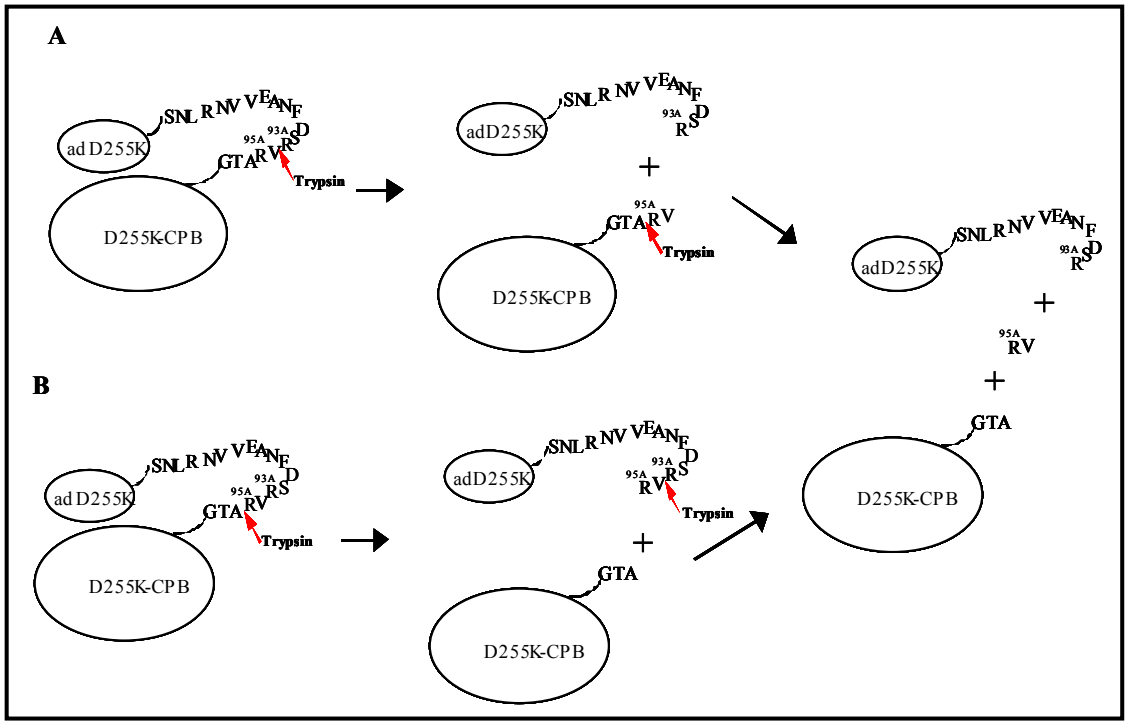


**Figure 1.13.- Schematic representation of the two possible mechanisms of tryptic activation of the D255K mutant. In (A) the first tryptic action takes place at the Arg93A-Val94A peptide bond and in (B) at the Arg95A-Ala4 peptide bond. The globular activation domain is denoted by *ad* (activation domain).**

### 1.3.4.- COMPARISON OF THE TRYPTIC ACTIVATION PROCESSES OF THE MEMBERS OF THE HUMAN PROCARBOXYPEPTIDASE SYSTEM

The tryptic activation course of human PCPA1, PCPA2 and PCPB has been studied at 0 ºC and at a 400/1 (w/w) proenzyme/trypsin ratio. All three forms follow a monotonic activation curve, but activity generation is quicker for PCPA2 and PCPB (Figure 1.14). None of the pro-segments of the members of the procarboxypeptidase system inhibit the activity of the active enzyme once the activating cut takes place, as shown in all cases by the correlation of mature enzyme generation with the increase of CP activity. Finally, due to the high specificity shown by CPA2, no release of free amino acids is observed during the activation course of its proenzyme (Reverter *et al.*, 1998), in contrast to the corresponding activation processes of PCPA1 and PCPB, where the generated mature enzymes participate in the processing of their activation segments.

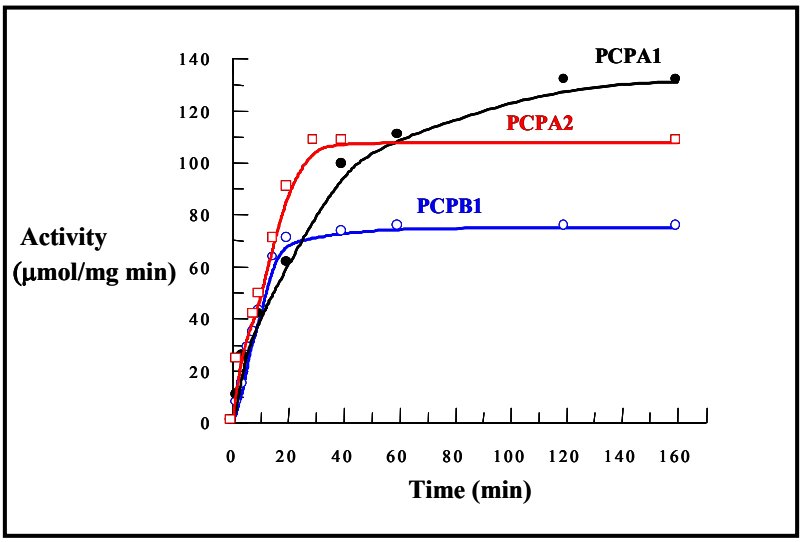


**Figure 1.14.- Time course of activity generation during tryptic activation of human procarboxypeptidases. CP activity was measured using 0,2 mM FAPP as substrate for CPA1 and CPA2 and 1 mM BGA for CPB; units are expressed as μmol of substrate hydrolyzed per minute/mg of protein.**

# CAPÍTOL 2:

# ESTRUCTURA TRIDIMENSIONAL DE LA PROCARBOXIPEPTIDASA B HUMANA I MODELAT DE LA TAFI I DE TRES NOUS MEMBRES DE LA FAMÍLIA DE METAL·LOCARBOXIPEPTIDASES HUMANES

# 2.1.- SUMMARY

The three-dimensional structure of human pancreatic procarboxypeptidase B (PCPB) has been solved and refined employing X-ray diffraction data to 1.6 Å resolution. Human pancreatic PCPB is the prototype for those human exopeptidases that cleave off basic C-terminal residues and are secreted as inactive zymogens. One such protein is thrombin-activatable fibrinolysis inhibitor (TAFI), a recently described fibrinolysis inhibitor that circulates in plasma as a zymogen bound to plasminogen and is highly regulated, therefore involved in the fibrinolytic pathway.

The structure of human pancreatic PCPB displays a 95-residue pro-segment consisting of a globular region with an open-sandwich antiparallel-$\alpha$ antiparallel-$\beta$ topology and a C-terminal $\alpha$-helix, which connects to the enzyme moiety. The latter is a 309-amino acid catalytic domain of $\alpha/\beta$ hydrolase topology and a preformed active site, which is shielded by the globular domain of the pro-segment. The fold of the proenzyme is similar to previously reported procarboxypeptidase structures, also in that the most variable region is the connecting segment that links both globular moieties. However, the empty active site of human procarboxypeptidase B has two alternate conformations in one of the zinc-binding residues, which account for subtle differences in some of the key residues for substrate binding. On the basis of this structure and others previously reported, a three-dimensional model of human TAFI has been built in order to generate a structure that may help us to understand its behaviour and also as a basis for the design of drugs to modulate its biological activity.

On the other hand, amino acid homology searches of the human genome were performed and revealed three members of the metallocarboxypeptidase family that had not been described in the literature, in addition to the 14 known genes. One of these three, named *CPA5*, is present in a gene cluster with *CPA1*, *CPA2*, and *CPA4* on chromosome 7. The deduced amino acid sequence of human CPA5 has highest amino acid identity (60%) to CPA1. Modelling analysis shows the overall structure to be very similar to that of other members of the A/B subfamily of metallocarboxypeptidases. The active site of CPA5 is predicted to cleave substrates with C-terminal hydrophobic residues, as do CPA1, 2 and 3. Two additional members of the human CP gene family were also studied. Modelling analysis indicates that both contain the necessary amino acids required for enzymatic activity. The CP on chromosome 8 is predicted to have a CPA-like specificity for C-terminal hydrophobic residues, and was named CPA6. The CP on chromosome 2, named CPO, is predicted to cleave substrates with C-terminal acidic residues, a unique activity among human CPs.

# 2.2.- EXPERIMENTAL PROCEDURES

## 2.2.1.- RESOLUTION OF THE THREE-DIMENSIONAL STRUCTURE OF HUMAN PCPB

### 2.2.1.1.- Crystallization and data collection

Recombinant human PCPB was expressed and purified as described in the "Materials and Methods" section of the present work. Orthorhombic crystals were obtained at 20 ºC by the vapour diffusion method from 4 μL hanging drops containing equal volumes of PCPB solution (11 mg/mL in 5 mM MOPS pH 7.0, 150 mM sodium chloride, 0.1 mM benzamidine) and precipitant (0.1 M MES pH 6.0, 1 M sodium citrate). The crystals belong to space-group $P2_1$ (a=43.0 Å, b=83.6 Å, c=127.4 Å, β=99.1°) and have two molecules per asymmetric unit (solvent contents 49.5 %). Diffraction data were collected on a 345-mm MAR Research Imaging plate and on 165-mm marCCD and ADSC Quantum4 CCD detectors at DESY (Hamburg) and ESRF (Grenoble) synchrotrons. Data were processed with DENZO (Otwinowski and Minor, 1993) and SCALA (Collaborative Computational Project No. 4., 1994). Table 2.1 provides a summary of data collection and processing.

### 2.2.1.2.- Structure solution and refinement

The structure was solved by molecular replacement with AmoRe (Navaza, 1994), using the coordinates of porcine procarboxypeptidase B (PDB 1NSA; (Coll *et al.*, 1991)) as a search model, with all non-identical non-glycine residues truncated to alanine. Calculations were performed with data in the 15 to 3.5 Å resolution range. The rotation function (RF) showed two clear peaks at Eulerian angles α1= 35.1º, β1= 71.8º, γ1= 66.1º and α2= 165.4º, β2= 107.1º, γ2= 245.2º . The correlation coefficient in amplitudes (CC; (Navaza, 1994)) for these solutions equals 19.4% and 19.2%, respectively; the crystallographic $R_{factor}$ (see Table 2.1 for a definition; computed after data expansion to $P_1$) equals 69.9% and 70.2% (RF=10.8 and 20.7, respectively). The translation functions rendered equally clear solutions (35.1, 71.8, 66.1, 0.3976, 0.2177, 0.5370 and 165.4, 107.1, 245.2, 0.1527, 0.000, 0.0176; α,β,γ in degrees and x, y, z in fractional cell co-ordinates; CC=28.7% and 29.2%; $R_{factor}$=67.4% and 67.3%; highest background peak CC=21.7%, $R_{factor}$=69.7%). The cumulative values were CC=49.4% and $R_{factor}$=59.7%. A final rigid-body refinement with FITING (Navaza, 1994) led to the figures (cumulative for both solutions) CC=56.1% and $R_{factor}$=56.6% (35.5, 70.9, 67.3, 0.4016, 0.2167, 0.5354 and 165.0, 107.0, 246.8, 0.1540, -0.0003, 0.0182). The input co-ordinates were

rotated and translated according to these solutions and subjected to a simulated-annealing refinement step and to positional/temperature-factor refinement using maximum-likelihood as a criterion, after performing bulk-solvent correction and anisotropic temperature-factor refinement ($R_{factor}$=30.0%; free $R_{factor}$= 33.2%). Program CNS v. 1.0 (Brunger *et al.*, 1998) was

| **Data collection and processing** | |
|---|---|
| Space group | $P2_1$ |
| Unit cell dimensions | a=43.0 Å, b=83.6 Å, c=127.4 Å, β=99.1° |
| Resolution limits (Å) (overall / last shell) | 27.4 – 1.60 / 1.68 – 1.60 |
| Number of observations (total / unique) | 303,465 / 104,805 |
| Multiplicity (overall / last shell) | 2.9 / 2.4 |
| $R_{merge}$[1] (overall / last shell) | 5.0 / 24.1 |
| Completeness (%) (overall / last shell) | 89.7 / 81.9 |
| I / σ(I) (overall / last shell) | 8.4 / 2.7 |
| | |
| **Refinement** | |
| Resolution range (Å) | 27.4 – 1.60 |
| Reflections used for refinement (total / test set) | 104,785 / 5,241 |
| $R_{factor}$[2] | 13.6 |
| Free $R_{factor}$[3] | 18.5 |
| R. m. s. d. bond lengths (Å) | 0.006 |
| R. m. s. d. bond angles (º) | 1.670 |
| R. m. s. d. bonded B-factors (Å$^2$) | 2.5 |
| Average B factor for protein atoms (Å$^2$) | 20.8 |
| Average B factor for solvent atoms (Å$^2$) | 35.9 |
| Average B factor for metallic ions (Å$^2$) | 49.4 |
| Non-hydrogen protein atoms | 6,578 |
| Solvent molecules | 814 |
| Inorganic ions | 2 ($Zn^{2+}$) |

[1] $R_{merge}$ = { $\Sigma_{hkl}\Sigma_i$ $|I_i$ (hkl) - <I (hkl)>| / $\Sigma_{hkl}\Sigma_i$ $I_i$ (hkl) }×100, where $I_i$ (hkl) is the observed intensity of the i$^{th}$ measurement of reflection (hkl), including symmetry-equivalent ones, and <I (hkl)> its mean intensity over all measurements of $I_i$ (hkl).

[2] $R_{factor}$ = { $\Sigma_{hkl}$ $||F_{obs}|$ - k $|F_{calc}||$ / $\Sigma_{hkl}$ $|F_{obs}|$ }×100

[3] free $R_{factor}$= $R_{factor}$ for 5% of reflections not used during refinement.

**Table 2.1.- Data processing and refinement statistics.**

used for this purpose. The subsequently computed $\sigma_A$-weighted ($2F_{obs}$-$F_{calc}$)- and ($F_{obs}$-$F_{calc}$)-type electron-density maps, inspected together with the initially refined model on a SGI Graphic Workstation with program TURBO-FRODO (Roussell and Cambilleau, 1989), showed the correctness of the solution and positive difference density for the regions diverging from the phasing model. Crystallographic refinement alternated with cycles of manual model building until completion of the model. The final isotropically refined model was further refined with SHELX97-2 (Sheldrick and Schneider, 1997). An initial round of 10 conjugate gradient least-square (CGLS) cycles using isotropic thermal parameters was followed by refinement with anisotropic displacement parameters (ADPs) for all atoms.

The final model, obtained after a refinement run with all the data without free $R_{factor}$ monitoring, comprises residues 1A-95A and 4-309 of each of the two PCPB molecules present in the asymmetric unit (the numbering of human PCPB, used throughout this work, is based on that used in Coll *et al.*, 1991), one zinc ion per propeptidase molecule and 814 solvent molecules. For both molecules, all residues fall in the allowed regions of the Ramachandran plot except Ser199, which is nevertheless unambiguously defined in the electron density maps. The sidechains of residues 5A, 6A, 34A, 47A, 57A, 59A, 70A, 57, 98, and 214 of molecule A and of residues 5A, 28A, 57A, 59A, 69A, 70A, 72A, 83A, 11, 57, 102 and 122 of molecule B are not defined in the electron-density maps and their occupancy was set to zero. All other residues, except segments 93A-95A (molecule A) and 92A-94A (molecule B) of the connecting segments, are clearly defined in the final electron-density map. Disulfide bonds are observed between residues 66 and 79, residues 138 and 161 and residues 152 and 166, and *cis*-peptide bonds are observed between residues 39A-40A, 197-198, 205-206 and 272-273. Table 2.1 presents a summary of the refinement parameters.

### 2.2.2.- DATABASE SEARCHES FOR NEW MEMBERS OF THE HUMAN METALLOCARBOXYPEPTIDASE GENE FAMILY

The NCBI Web site was used to search the public human genome data base with various human CP sequences using the tblast-n program and the default parameters. To identify additional exons and to determine the nucleotide sequence of the exon/intron junctions, the genomic sequences that corresponded to the novel CP-like genes on chromosomes 7, 8 and 2 were downloaded from the NCBI site. These sequences were translated in all three reading frames and the deduced amino acid sequences were searched for homology to various CPs using the GenePro program (Hoeffer Scientific). Because the amino acid sequence similarity within the N-terminal precursor regions of the CPs is low, these regions could not be identified

from this homology-based approach. Searches of the expressed sequence tag (EST) data base on the NCBI Web site using the blast-n program revealed matches to the 5' region of the novel CP-like genes on chromosomes 7 and 8. Several of these EST sequences extended in the 5' region into the putative signal peptide and pro region, and so they were used to search the human genome data base and obtain the 5' exons for this N-terminal region. Predictions on the presence of N-terminal signal peptides and putative cleavage sites of these peptides were performed using the Signal-P website (Nielsen *et al.*, 1997).

## 2.2.3.- MODEL BUILDING OF TAFI AND THE THREE NEW MEMBERS OF THE HUMAN METALLOCARBOXYPEPTIDASE FAMILY

### 2.2.3.1.- Template selection and alignments

The set of template structures used for modelling the target sequences was chosen in every case based on closest sequence similarity. The following templates were used:

- for human TAFI: human PCPB, porcine PCPB (Coll *et al.*, 1991), porcine PCPA1 (Guasch *et al.*, 1992), bovine PCPA1 (Gomis-Ruth *et al.*, 1995) and human PCPA2 (Garcia-Saez *et al.*, 1997)
- for human PCPA5: bovine PCPA1, porcine PCPA1 and human PCPA2
- for human PCPA6: human PCPB, porcine PCPB, human PCPA2 and bovine PCPA1
- for human CPO: human CPB, porcine CPB, human CPA2 and bovine CPA1

A preliminary multiple alignment was performed for the four sets of template proteins plus the target sequence by means of the program CLUSTALW (Thompson *et al.*, 1994) using a BLOSUM 62 matrix for weighting. In order to correctly align the target and the template sequences, data on the secondary and tertiary structures for this set of proteins was added. The secondary structure of the target proteins was predicted with the program PHD (Rost *et al.*, 1994) and the real secondary structure of the templates was calculated with the program DSSP (Kabsch and Sander, 1983). The 3D structures of the chosen templates were structurally superimposed by means of the program SSAP (Orengo *et al.*, 1992) and the multiple alignments of the templates were modified according to the 3D superimposition of the templates. The resulting alignments were used as "seeds" to build hidden Markov model profiles with the program HMMER (Eddy, 1998). Subsequently, the sequence of each target was aligned to the profile and a final multiple alignment accounting for the structural information obtained from the templates superimposed structures, plus the predicted and real secondary structure, was obtained.

### 2.2.3.2.- Building and evaluation of models

Using the alignment of each target with the selected proteins of known structure as a starting point, a method of comparative modelling by satisfaction of spatial constraints was used to build the three-dimensional structure of TAFI, PCPA5, PCPA6 and CPO. This method is implemented in the program MODELLER (Sali and Blundell, 1993). The spatial constraints were derived by transferring the spatial features from the known protein structures to the sequence of the unknown ones. Idealisation of bond geometry and removal of unfavourable non-bonded contacts was performed by energy minimization with the GROMOS force field for unsolvated systems (Oliva *et al.*, 1991; van Gunsteren and Berendsen, 1987). Models were refined using 1000 steps of steepest descent. Overall r.m.s.d. calculations and superimposition of the modelled structures with respect to the crystallographic ones were obtained by means of program SSAP. Secondary structure calculations of the models were performed with DSSP (Kabsch and Sander, 1983) and compared with the predicted secondary structure. After energy minimisation to avoid clashes between atoms, the models were evaluated by means of mean force potential with the program PROSA-II (Sippl, 1993) to identify the incorrect chain tracings and to assess model quality. This program allows identification of regions with non-near-native fold by the high positive values of pseudo-potential energy. For each target, the model showing the smallest pseudo-energy was taken for the final modelled structure.

## 2.2.4.- MISCELLANEOUS

The numbering of the pro-segments of procarboxypeptidases is based on that used in Coll *et al.*, 1991, the first three-dimensional structure obtained for a procarboxypeptidase, where the number of the amino acid positions are followed by "A" to distinguish them from the positions in the enzyme moiety. The numbering of bovine CPA1 is kept for the enzyme moiety. Figures were prepared with ALSCRIPT (Barton, 1993), SETOR (Evans, 1993), TURBO-FRODO (Roussell and Cambilleau, 1989) and PREPI (http://www.bmm.icnet.uk/prepi/index.html). Structural homology searches have been performed with the DALI (Holm and Sander, 1993) server at EBI (http://www.ebi.ac.uk/dali/).

# 2.3.- RESULTS AND DISCUSSION

## 2.3.A.- THREE-DIMENSIONAL STRUCTURE OF HUMAN PCPB AND MODELLING OF TAFI

### 2.3.A.1.- STRUCTURE OF HUMAN PCPB

For the structural studies, human PCPB was recombinantly expressed in a *Pichia pastoris* system designed to secrete the protein to the extracellular medium. This provided an efficient system for the production of tens of milligrams of protein per liter of cell culture, and rendered an enzyme with identical properties to that isolated.

The two molecules in the asymmetric unit are very similar, with a r.m.s.d. of only 0.43 Å, for 398 aligned Cα atoms. Most of the differences observed are due to crystal contacts and occur in the rather flexible and partly disordered connecting segment (see below). For this reason, only the arbitrarily termed molecule A will be discussed.

Human PCPB is a globular protein, which comprises a 95-residue N-terminal pro-segment and the 309-residue catalytic domain of the enzyme (Figure 2.1a). The pro-segment consists of an N-terminal globular pro-domain and a connecting region that links it to the carboxypeptidase moiety. The pro-domain covers the active site of the enzyme, shielding it from substrates. The two protein moieties interact through an intricate network of direct and water-mediated hydrogen bonds, together with extensive van-der-Waals interactions (see Figure 2.2). A summary of the polar interactions observed is provided in Table 2.2.

The globular pro-domain of human PCPB encompasses the first 80 residues of the pro-segment (1A-80A), while residues 81A to 95A form the partly α-helical connecting segment. The globular region displays a β1α1β2β3α2β4 topology, in which the N-terminus is exceptionally well ordered, with a $3_{10}$-helix preceding the first β-strand unique among PCPs (Figures 2.1a and 2.3). The connecting segment starts with a short two-turn α-helix, which is followed by a loosely arranged section containing Arg95A, the target for trypsin cleavage during proenzyme activation. In our human proenzyme there is an additional arginine residue at position 93A, which may also be a target for trypsin (Villegas *et al.*, 1995a). The whole region is very exposed and probably rather flexible, as indicated by above-average B-values. The sidechain of Arg93A is disordered, as are all atoms of Arg95A. The connecting segment of molecule B displays a different arrangement, which is brought about by crystal packing. The

proximity of a neighbouring protein molecule changes the position of residues 92A-95A, again as a consequence of the high flexibility of this region, which in turn favours crystal packing.

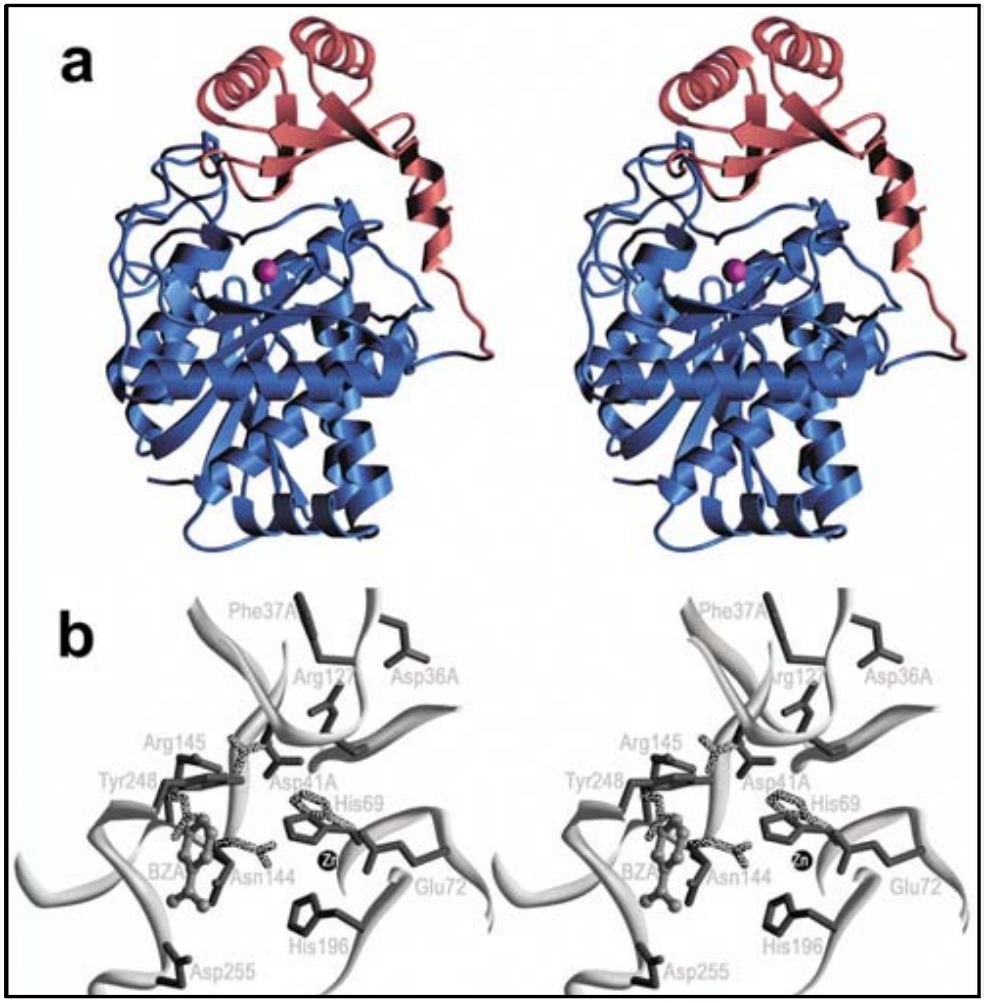


**Figure 2.1**.- **Structure of human PCPB. (a) The proenzyme is composed of two distinct domains. Overall structure of human PCPB as a ribbon plot with the secondary structure elements shown ribbons (α-helices) and arrows (β-strands). The zinc ion is shown as a magenta sphere. The ribbon corresponding to the 95-residue activation segment is displayed in red, the active enzyme moiety in blue. (b) Close-up stereo view of the active-site of human PCPB. The sidechains of important residues are shown as stick models. A modelled benzamidine molecule (labelled BZA and shown as a ball-and-stick model) has been placed based on its structure with porcine PCPB (PDB access code 1NSA; (Coll *et al.,* 1991)) to line out the active-site cleft, in particular the $S_1'$ specificity pocket. Side-chains present with a second occupancy have been displayed with a dotted pattern.**

Attached to the pro-domain is the enzyme moiety, already preformed in the complex and shielded by the pro-segment. The catalytic domain has the same fold as its porcine counterpart (Coll *et al.*, 1991) and shows the typical central mixed eight-stranded twisted β–sheet surrounded by eight α–helices. The arrangement and identity of the residues at the specificity pocket and at the different substrate-binding subsites is also reminiscent of the homologous porcine proenzyme and it also contains the three expected disulfide bridges. Over the preformed active site lie strand β2, the *cis*-proline turn and the $3_{10}$-helix elements of the activation domain (residues Asp36A-Lys47A), covering the active site in a fashion that avoids a substrate-like interaction and impedes peptide bond hydrolysis of the inhibiting pro-domain. The overall structure of the human protein closely resembles that of its porcine homologue (Coll *et al.*, 1991), with which the human protein shares 84% of sequence identity (see Figure 2.3). The backbones and most of the common side chains coincide upon least-squares fitting (r.m.s.d. of 0.53 Å for 393 common Cα atoms deviating less than 3.5 Å).

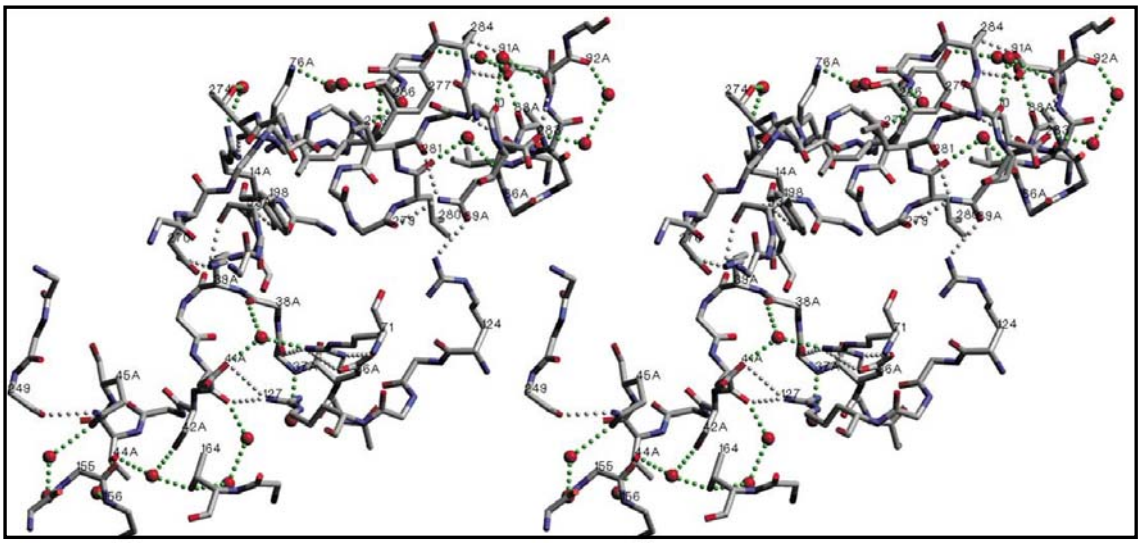

**Figure 2.2.- Hydrogen-bonding network between domains. Close-up view of the PCPB structure displaying the interface between the enzyme moiety and the pro-segment. Direct hydrogen bonds are displayed as grey spherical dashes, solvent (red spheres)-mediated ones in green. The intervening residues corresponding to the mature enzyme moiety and the pro-domain (suffix A) are labelled.**

### 2.3.A.1.1.- Active-site architecture

At the active-site of human PCPB the catalytically active zinc ion is penta-coordinated by the side chains of His69, His196, a water molecule, and Glu72, the latter in a bidentate manner.

Although this is a classical arrangement, similar to that observed in porcine PCPB (Coll *et al.*, 1991), the active-site environment of human PCPB displays subtle differences (see Figure 2.4). First, the side chain of the zinc-coordinating His69 displays two alternate conformations, with approximately equal occupancies (Figure 2.1b). Probably due to this unusual movement of His69, the side chain of Arg127, which in the active enzyme is involved in stabilising the carbonyl oxygen of the scissile peptide bond, moves further apart from the zinc ion than in porcine PCPB. In our structure it performs new interactions with three pro-domain residues, namely the side chain of Asp36A, and the carbonyl oxygen atoms of Phe37A and Asp41A. The latter has also double occupancy, and while one of its rotamers retains the interaction observed

| | **Activation segment** | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *CPB* | R14 | D36 | F37 | W38 | K39 | D41 | T44 | Q45 | D53 | K76 | L78 | V85 | A88 | Q89 | D91 | S92 |
| E10 | | | | | | | | | | | | | | | | 2(1) |
| R71 | | 2 | 1 | 1(1) | | 1(1) | | | | | | | | | | |
| R124 | | | | | | | | | | | | | | 1 | | |
| R127 | | 1 | 1 | | | 2 | | | | | | | | | | |
| R145 | | | | | | 3 | | | | | | | | | | |
| G155 | | | | | | | 1(1) | 2(1) | | | | | | | | |
| T164 | | | | | | 1(1) | | | | | | | | | | |
| Y198 | | | | | 1 | | | | 1 | | | | | | | |
| Y248 | | | | | | 3 | | | | | | | | | | |
| P249 | | | | | | | | 1 | | | | | | | | |
| E270 | | | | | 2 | | | | | | | | | | | |
| T274 | 1 | | | | | | | | | | | | | | | |
| R276 | | | | | | | | | | 1(1) | 1(1) | | | | | |
| Y277 | | | | | | | | | | | | | 1(1) | | | |
| F279 | | | | | | | | | | | | | | 1 | | |
| L280 | | | | | | | | | | | | 1(1) | | | | |
| L281 | | | | | | | | | | | | | | 1 | | |
| E283 | | | | | | | | | | | | | | 1 | | 1(1) |
| S284 | | | | | | | | | | | | | | | 2 | |

**Table 2.2.- Hydrogen bond network between the two domains of human PCPB. The figures in the table indicate the total number of hydrogen bonds, in parenthesis are those mediated by solvent molecules.**

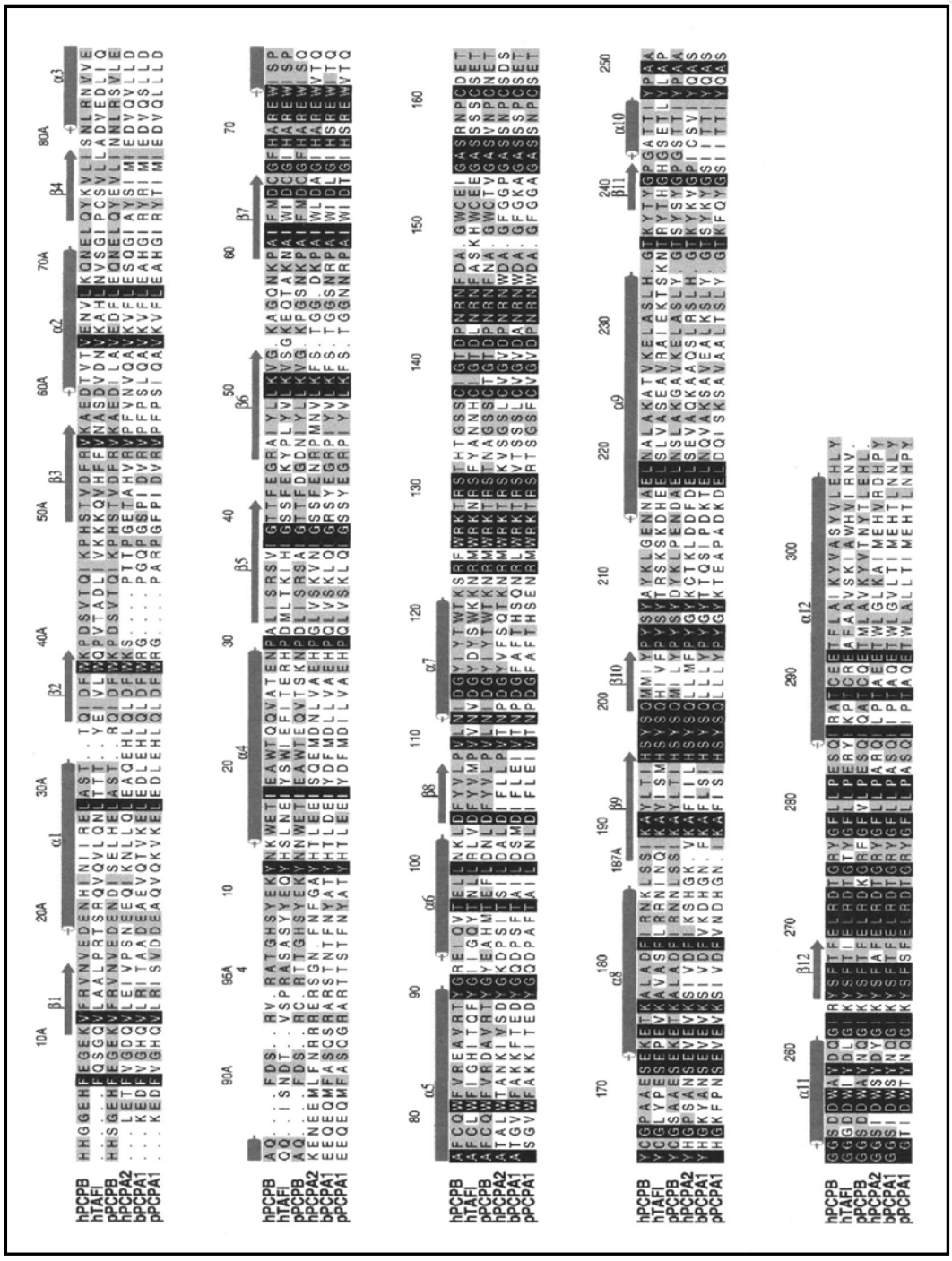

**Figure 2.3.- Amino acid sequence alignment between human TAFI and other PCPs. Presented are, from top to bottom, human PCPB, human TAFI, porcine PCPB, human PCPA2, bovine PCPA1 and porcine PCPA1. The numbering, as defined by Coll *et al.*, 1991, corresponds to human PCPB (activation segment from 1A-95A, peptidase domain from 4 to 309), as do the depicted secondary structure elements (dark arrows represent β-strands, labelled β1- β12, rods represent α-helices, termed α1- α12). Identical residues in all sequences are displayed as white letters over black background. Residues identical in at least human PCPB and TAFI and/or porcine PCPB are displayed with a grey background.**

in porcine PCPB (Coll *et al.*, 1991) with Arg145, it is probably less favourably oriented to form a strong salt bridge. Arg145, on the other hand, also displays an alternating side-chain conformation. Besides the "standard" orientation, by which it interacts with Asp41A, and presumably accounts to a great extent for the inactivity of PCPBs, it partially occupies the $S_1$ selectivity pocket (nomenclature according to Rees and Lipscomb, 1982), where a benzamidine molecule, absent in our structure, is found in the three-dimensional structure of the porcine enzyme (Coll *et al.*, 1991). In this orientation, Arg145 makes a polar interaction, mediated by two water molecules, with Asp255, located at the bottom of the $S_1$ pocket and thus determining the specificity for basic C-terminal side chains. The movement of the His69 side chain allows that of Asn144 to adopt a second conformation, moving closer to the metal ion. Also, Tyr248 presents a side-chain flip and is in the "down" conformation, when compared to the porcine enzyme, displaying a rotamer similar to that found in the crystallographic structures of duck carboxypeptidase D (PDB entry 1QMU; (Gomis-Ruth *et al.*, 1999)) and bovine carboxypeptidase A in complex with potato carboxypeptidase inhibitor (PDB entry 4CPA; (Rees and Lipscomb, 1982)). In the bovine enzyme, the side chain of Tyr248 stabilises the carboxylate group of the inhibitor molecule, while in duck carboxypeptidase D it is fixed by a sulfate anion found in the active site. However, in human PCPB the Tyr248 side chain does not interact with any "left-behind" molecule after hydrolysis (Garcia-Saez *et al.*, 1997), as the pocket is empty. Instead, it is within hydrogen-bonding distance to the side-chain carboxylate and the main-chain nitrogen of Asp41A. This may explain the slight change in orientation of Asp41A towards Arg145 and it may be a further consequence of the alternate conformation of His69.

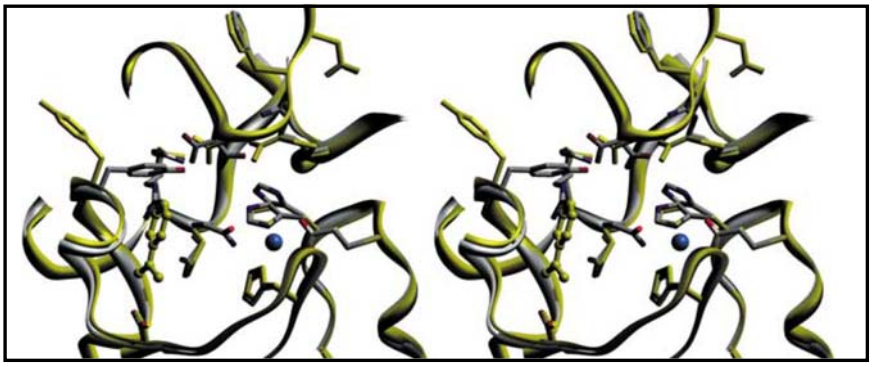

**Figure 2.4.- Comparison between human and porcine pancreatic PCPBs. Displayed are the active-site centres of both enzymes, that of the porcine proenzyme in yellow, with the benzamidine molecule identified in the active-site cleft, and that of the human enzyme in grey.**

### 2.3.A.2.- MODELLING OF TAFI

Our group has solved the three-dimensional structure of several pancreatic PCPs, namely porcine PCPB (Coll *et al.*, 1991), porcine PCPA1 (Guasch *et al.*, 1992), bovine PCPA1 (Gomis-Ruth *et al.*, 1995), human PCPA2 (Garcia-Saez *et al.*, 1997) and now human PCPB. In terms of primary structure, TAFI is 42% identical to both human and porcine B forms and 38% identical to A forms. Since the identity levels between TAFI and the pancreatic A and B forms are all very similar and sufficiently high to allow modelling, all the crystallographic structures available were used as templates for modelling.

The sequence alignment (Figure 2.3) shows that the region corresponding to the pro-segments is less conserved than the enzyme moieties. Thus, while the identity levels between TAFI and the pancreatic procarboxypeptidases in their pro-segments is only about 20%, it increases to 44% and 48% for A and B forms, respectively, when considering the enzyme moieties alone. The alignment also shows that some key residues that are always conserved among digestive carboxypeptidases are also conserved in TAFI: the residues involved in the coordination of the active-site zinc cation (His69, Glu72 and His196; bovine carboxypeptidase A1 numbering; (Rees *et al.*, 1983)), and a series of residues important for substrate binding and catalysis (Arg71, Arg124, Arg127, Lys128, Asn144, Arg145, Ser197, Tyr198, Ser199, Tyr248, Glu270, Phe279; bovine carboxypeptidase A1 numbering), as previously determined from kinetic studies on rat CPB (Clauser *et al.*, 1988) and crystallographic studies on porcine PCPB (Coll *et al.*, 1991). The connecting-segment region between the pro-domain and the enzyme domain comprises an $\alpha$–helix and an exposed loop, where the first target for trypsin activation is located. Structural superimposition of the templates showed that the connecting regions of the A and B forms have completely different spatial conformations, due to the presence of a longer $\alpha$-helix in the A forms. The program MODELLER was furnished with the spatial constraints determined by all the structures present in the alignment of Figure 2.3 to build up to 50 models, with the exception of the connecting segment region, which was only constrained by the structure of PCPB forms. Therefore, special attention was focused on this region when evaluating the models. Loops containing insertions are also likely to present a variety of conformations due to the lack of restraints. Therefore, these structurally variable regions were further checked against a general loops database (Oliva *et al.*, 1997). However, in the absence of clear topological coincidences, no additional structures were considered to restraint the model of TAFI.

### 2.3.A.2.1.- Evaluation and refinement of the TAFI model

Based on the differences in primary sequence in the distinct segments to be modelled, most of the detected errors should be expected to belong to the variable regions. However, it was observed that the energy profile for most of the enzyme moiety of TAFI was very similar to those observed for the known structures of PCPs. The main differences in the energy profile were found in the pro-segment of TAFI, but this variation also occurred when comparing the energy profiles of the templates. Therefore, as the accuracy of the model for the connecting region could not be determined unequivocally, the model with the lowest energy was chosen as the final one (Figure 2.5A). In order to improve model accuracy, TAFI secondary structure was predicted. A few differences were found between the predicted secondary structure and the DSSP-calculated secondary structure of the proposed model. These differences arise from inaccuracy of the secondary structure prediction, as they were also apparent when comparing prediction and DSSP-calculation for the experimental PCP structures. The prediction for the N-capping (Aurora and Rose, 1998) of the connecting $\alpha$-helix enlarges it, but this does not significantly improve the energy profile of the model, as tested with PROSA-II. Finally, some models were built to study the effect of constricting the connecting region using the structures of both A and B forms and, as expected, this introduced local instabilities in the TAFI models.

### 2.3.A.2.2.- Analysis of the TAFI model and comparison with the experimental PCP structures

The overall r.m.s.d. values between the final refined model and the various templates are: 0.52 Å for porcine PCPB, 0.57 Å for human PCPB, 1.11 Å for human PCPA2, 1.15 Å for porcine PCPA1 and 1.0 Å for bovine PCPA1. These values indicate that TAFI, a CPB-like enzyme, is more closely related to the B than to the A forms of carboxypeptidases. TAFI, like all other B forms, has a characteristic two-residue deletion between $\alpha$1 and $\beta$2 (Figure 2.3) and a four-residue insertion between β2 and β3. This insertion contains the $3_{10}$-helix that covers the active site in the B forms and is one of the elements responsible for the null intrinsic activity of their zymogens. When compared to the pancreatic PCPBs, TAFI has six point insertions. Three of them are located in the loop of the connecting region (Ile89B, Ser89C, Pro94B), making it unusually longer. The rest of the insertions are found in the enzyme moiety, the first (Gly53A) being located between two $\beta$-strands (β6 and β7) , the second (Lys149A) in a large loop and the last (Lys234A) between an $\alpha$-helix and a $\beta$-strand ($\alpha$9 and $\beta$11).

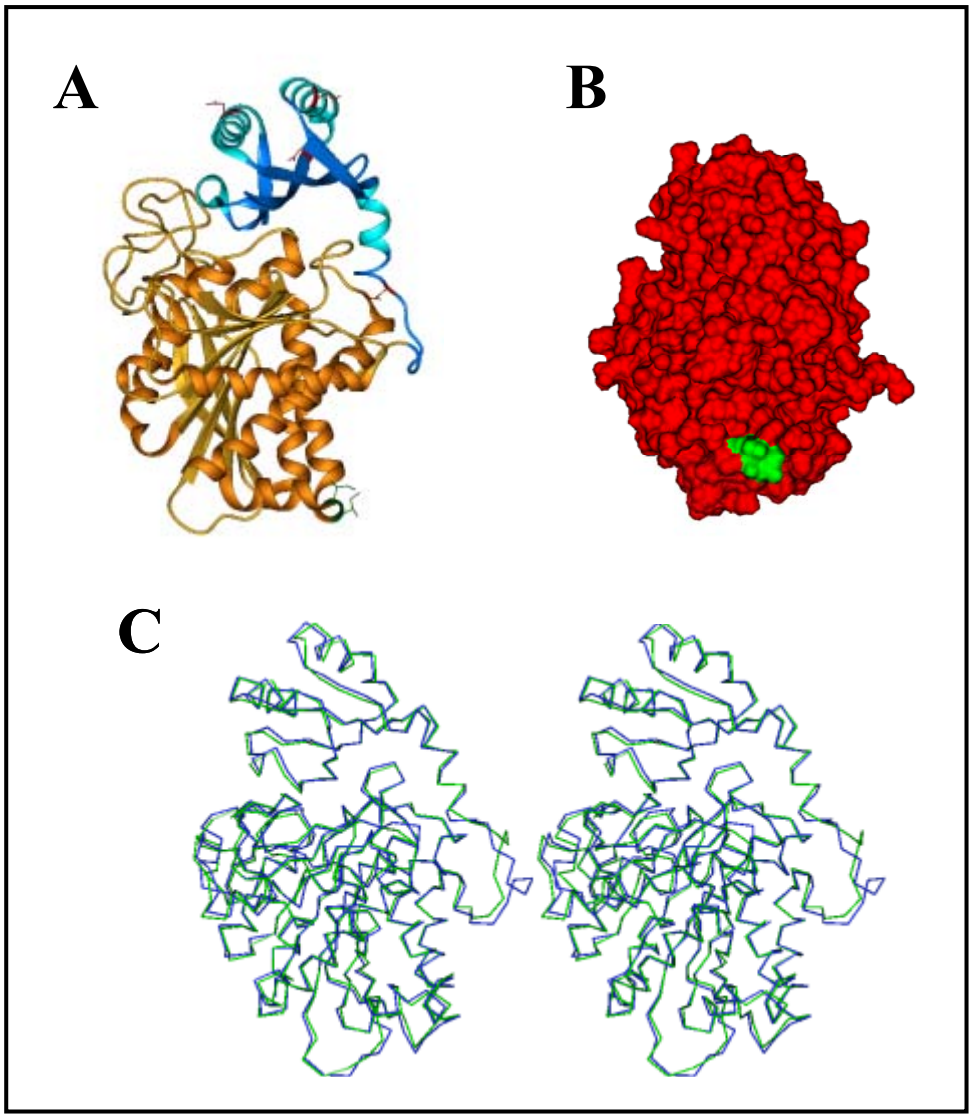


**Figure 2.5.- Representation of the modelled TAFI structure. (A) Ribbon drawing of the modelled folding of TAFI; the pro- and enzyme domains are shown in blue and gold, respectively, and the different colour intensities indicate α or β secondary structures. The positions and side-chains of the four glycosylation points in the pro-segment are marked in red (Asn28A, Asn57A, Asn69A and Asn90A) and the side-chains of the two sequentially consecutive isoleucine residues at proenzyme positions 182 and 183 are highlighted in green (positions 92 and 93 according to the active enzyme numbering). (B) Surface representation of TAFI, in which Ile182 and Ile183 are highlighted in green as in (A). (C) Stereo representation of the superposition of the Cα traces of human PCPB (green) and the TAFI model (blue).**

The three intra-chain disulphide bonds present in both human and porcine CPBs (Cys66-Cys79, Cys138-Cys161 and Cys152-Cys166) are conserved in human TAFI. The specificity-determining aspartic residue at position 255 is also present, together with two other residues typically present in B forms: a glycine at 243 and a serine at 207. In contrast, a conserved residue present in the pro-segment of all B forms, Asp41A, is replaced by a valine in TAFI. Asp41A forms a salt bridge with a key residue in $S_1$'-pocket shaping, Arg145 (see above), and this interaction is thought to account for the null intrinsic activity of PCPBs. Thus, TAFI may show intrinsic carboxypeptidase activity, even containing the $3_{10}$-helix motif that follows Asp41A in other B-type PCPs. On the other hand, the series of key residues that are conserved in the sequence alignment of PCPs are also conserved in the structural superimposition of TAFI and its templates.

The activated form of TAFI, TAFIa, is sensitive to proteolytic cleavage by thrombin and plasmin and, recently, residue Arg210 (numbering of the enzyme moiety, corresponding to 302 of the complete sequence; see Figure 2.3) has recently been identified as the cleavage site (Boffa *et al.*, 2000). This arginine, which is located in a loop, is the most exposed residue after Arg95A, the target for thrombin activation, and it is unique to TAFI. In the other B forms it is replaced by a tyrosine and in the A isoforms by cysteine or threonine. This feature indicates that a proteolytically sensitive point is constitutively present in the structure of the protein. However, thermal-dependent conformational changes render the protein less stable favouring degradation. In fact, inactivation of TAFI does not require a previous proteolytic cleavage (Boffa *et al.*, 2000; Marx *et al.*, 2000). Thermal instability thus appears to be the key factor in the regulation of the antifibrinolytic activity of TAFI (Boffa *et al.*, 2000) , but no structural explanation for this observation is yet available. In the absence of crystallographic data, preliminary analysis of the TAFI model suggests several TAFI-specific hydrophobic residues on its surface, which could cause its conformational instability (Figure 2.5B). Residues Ile92-Ile93 of the enzyme moiety (corresponding to positions 182-183 in the complete sequence and absolutely unique for TAFI) show high positive values of pseudo-potential energy when the model is analyzed with PROSA II (Sippl, 1993). This is the most remarkable difference observed in the pseudo-potential energy profiles calculated for all the enzyme regions analyzed in this work. The two hydrophobic residues are located in a surface-located loop preceeding an exposed $\alpha$–helix with little intrinsic tendency to form such a secondary structure, as measured by the program AGADIR (Muñoz and Serrano, 1997). Destabilisation of the region containing and surrounding the exposed isoleucine residues could lead to TAFI inactivation.

# 2.3.B.- IDENTIFICATION AND MODELLING OF THREE NEW MEMBERS OF THE HUMAN METALLOCARBOXYPEPTIDASE FAMILY

## 2.3.B.1.- DATABASE SEARCHES

When CPE was used to search the public human genome data base using the tblast-n program, each known member of the N/E subfamily was detected, but no additional genes were found (Table 2.3). None of the N/E subfamily genes appeared to be present as a cluster; although chromosome 4 contains both *CPE* and *CPZ*, and chromosome 10 contains both *CPN* and *CPX2*, these genes are located at a considerable distance from each other within the chromosome. Similarly, none of the members of the N/E family were located in close proximity to members of the A/B family. Searches with CPE produced a limited number of hits to members of the A/B subfamily, as expected due to the generally low amino acid identity between subfamilies.

Searches with CPA1 and CPB yielded generally similar results to each other, with hits to genes on chromosomes 7, 3, 8, 13, and 2 (Table 2.3). The four genes found on chromosome 7 are found within a cluster, with the coding region of each gene located from 3 to 23 kb apart from the coding region of the adjacent gene. The genes for the pancreatic proteins CPA1 and CPA2 are located at the ends of this cluster. A cDNA sequence reported in the literature as CPA3 (Huang *et al.*, 1999), and renamed CPA4 by the human genome nomenclature committee, is also present in this cluster. This sequence presumably encodes an active CPA-like enzyme based on consideration of the active site region; the enzymatic properties have not yet been reported. In addition, a novel human CP-like gene was found within this cluster and named *CPA5* based on predictions of the active site region. The genomic sequence of the appropriate region of chromosome 7 was searched in all three reading frames for homology to various CPs using the GenePro program. This analysis identified 11 exons for the entire predicted coding region of *CPA5*, consistent with the 11 exons found in most other members of the A/B subfamily of CPs (Clauser *et al.*, 1988; Gardell *et al.*, 1988). All of the introns in the coding region of *CPA5* begin with GT and end with AG, and their locations exactly match those in other members of the A/B subfamily.

| Gene | Chromosome | Comments |
|---|---|---|
| A/B Subfamily | | |
| *CPA1* | 7 | |
| *CPA2* | 7 | |
| *CPA3* | 3 | Previously "mast cell CPA," now CPA3 in GenBank. |
| *CPA4* | 7 | Previously CPA3, changed to CPA4 in GenBank. |
| *CPA5* | 7 | Similar to mouse cDNA sequence AK015256. |
| *CPA6* | 8 | Partial similarity to human cDNA sequence AF221594. |
| *CPB* | 3 | |
| *CPB2* | 13 | Protein named TAFI, plasma CPB, CPU and CPR. |
| *CPO* | 2 | Not previously identified in public cDNA data bases. |
| N/E Subfamily | | |
| *CPE* | 4 | Not near CPZ gene on chromosome 4. |
| *CPN* | 10 | CPN protein associates with 83-kDa noncatalytic subunit. |
| *CPM* | 12 | |
| *CPD* | 17 | |
| *CPZ* | 4 | Not near CPE gene on chromosome 4. |
| *CPX1* | 20 | Protein may not be an active CP. |
| *CPX2* | 10 | Protein may not an active CP. |
| *AEBP1* | 7 | Also named ACLP. May not be an active CP. |

**Table 2.3. - Genomic location of human metallocarboxypeptidase-related genes.**

*CPA5* has >98% nucleotide identity with a total of thirteen sequences in the GenBank human EST data base (as of November 30, 2001). Of these, nine were from libraries that included RNA from testis, three were from libraries of adult brain, and one was from fetal heart. One of the testis sequences, and all three of the brain sequences were 560 nucleotides longer in the 5' direction than the coding sequence predicted from homology to other CPs. All four of these clones contained two additional 5' non-coding exons and the testis clone contained a third non-coding exon. The intron/exon junctions of the two common exons obeyed the GT/AG rule while the junction of the exon found only in the testis clone had a GG in place of the GT. The 560 nucleotides upstream region contains 8 ATGs, however, only one of these upstream ATG motifs

is predicted to be a good consensus sequence for transcription initiation (Kozak, 1984), and this is followed by an in-frame stop codon. The second ATG that is predicted to be a good consensus sequence for transcription initiation encodes a protein of 436 amino acids.

In addition to *CPA5*, two other sequences that have not previously been reported in the literature were detected on chromosomes 8 and 2. The first eight exons of the CP-like gene on chromosome 8 exactly match the sequence of a cDNA clone reported in GenBank (Accession number AF221594). However, this GenBank entry lacks the sequence corresponding to exons 9-11 and has a 3'-end that does not have any homology with known carboxypeptidases. Instead, this 3'-end perfectly matches the putative intronic region found in the human genome sequence. Searches of the downstream genomic sequence for CP-like domains revealed the missing exons ~50 kb away. Although large for an intron, this is considerably smaller than the first and second introns which are 121.8 and 98.4 kb. All of the predicted introns begin with GT and end with AG, and their locations exactly match those found in other A/B subfamily CPs. Based on consideration of the active site region, this chromosome 8 CP-like gene has been named *CPA6*.

Similar analysis of the novel gene on chromosome 2 revealed eight exons corresponding to the active CP domain of the protein. The amino acid homology within the N-terminal "prepro" regions of the carboxypeptidases was not sufficient to identify this portion of the gene based on homology searches. Because no cDNA sequence was found in public data bases that corresponds to this gene, it was not possible to determine the N-terminal exons. The downstream exons were numbered based on the related CPs, with the assumption that there would be three additional exons for the signal peptide and pro domains. The introns within this chromosome 2 CP-like gene all begin with GT and end with AG, and their positions exactly match one or more of the other family members. Based on consideration of the active site region, which suggests that this enzyme has a markedly distinct substrate specificity from CPA and CPB, this gene was named *CPO*.

Of the three CP genes identified in the present study, only two of them were expected based on previous database searches of mRNA-based cDNA libraries. However, none of these mRNA-based sequences predicted an active CP protein due to either frameshift errors or incomplete sequence information. It is likely that these are the last human metalloCP-like genes that will be identified because no other sequences were found in numerous searches of genome and cDNA libraries using a variety of query CP sequences and search parameters. The total number of CP family members is 17, with 9 of these in the A/B subgroup and 8 in the N/E subgroup (although CPD consists of three CP-like domains, so the total number of CP domain sequences is 19). However, it is possible that additional human proteins have structural homology to the CP family that is not detected by the amino acid homology searches used in the present study.

The location of introns in the coding regions of the various members of the A/B subfamily of CP is generally conserved in all of the previously identified CPs as well as the three novel ones found in the present study (although the exons encoding the putative pro region of CPO could not be identified and so no conclusions about the intron locations within this part of CPO can be made). It is likely that the predicted intron splicings are correct because the sites exactly match one of the previously identified CP genes, the splice sites all obey the GT-AG rule, and for CPA5 and 6, cDNA sequences that cover much or all of the predicted coding region are present in the data base. Of the three additional CPs described in this study, CPA5 had the most matches to cDNA clones in the database.

Although full length sequences of CPA6 have not been reported in GenBank, it is likely that this gene also is expressed because of the matches to cDNA sequences in the data base. The longest of these cDNA sequences corresponds to the first 8 exons of CPA6 but then continues into the putative intron and would not therefore encode an active CP as it is missing the catalytically important exons 9-11. Although no matches were found to this C-terminal portion of CPA6 in human or rodent cDNA databases, several matches were found to this region in chicken (*Gallus gallus*) and frog (*Xenopus laevis*) cDNA databases. The amino acid sequence identity is extremely high, with 95% identity over 128 amino acids for the longer of the two chicken clones, and 89% identity over 113 residues for a similar region of the frog sequence; this implies that they represent homologs of human CPA6 and that this protein has been highly conserved during evolution.

### 2.3.B.2.- ALIGNMENTS OF THE NOVEL CPs

Human CPA5 is predicted to be produced as a precursor containing a 33-residues N-terminal signal peptide. Following the signal peptide is a predicted 93-96-residue pro domain that has some homology to the corresponding pro region of other A/B subfamily CPs (Figure 2.6). Based on sequence alignments with other PCPs and on consideration of the specificity of various endopeptidases, cleavage of the pro domain is predicted to occur either between Arg98A and Leu99A (porcine PCPB numbering, (Coll *et al.*, 1991)) or between Arg101A and Ser3. Both sites are related to the consensus site for furin and other prohormone convertases (PCs) in which cleavage occurs C-terminal to the second Arg of the sequence Arg-Arg or Arg-Xaa-Xaa-Arg (Nakayama, 1997; Zhou *et al.*, 1999; Seidah and Chretien, 1998). Although the first site has an Arg-Arg motif, the presence of the Leu in the $P_1$ position is not ideal for many of these PCs. The second site, with the Arg-Xaa-Xaa-Arg motif, may be more readily cleaved due to the $P_1$ Ser. Also, this second site is favoured from the modelling analysis.


**Figure 2.6.- Alignment of the deduced amino acid sequence of human (h) PCPA5 and human PCPA2, bovine (b) PCPA1 and porcine (p) PCPA1. Pro-segments residues are coloured in green (for hPCPA5 the pro peptide cleavage site may occur at either Arg98A or Arg101A, but only the Arg101A site is considered). The zinc-binding residues and the active site residues coloured in red, Cys residues forming disulphide bridges in blue and specificity pocket residues in positions 194, 253, 255 and 268 in pink. Porcine PCPB numbering (Coll *et al.*, 1991).**

The deduced amino acid sequence of CPA6 encodes a protein of 437 residues that is predicted to have a 30-residues signal peptide. The putative pro domain is nearly identical in length to that of PCPA5 and other family members (Figure 2.7). The predicted cleavage site of the pro domain of PCPA6 is a perfect furin/PC consensus site, with cleavage expected between Arg101A and Ser3 (Figure 2.7). Because the N-terminal exons of *CPO* could not be identified due to the relatively low homology among the prepro regions of the various CPs, no information is available regarding the presence of a signal peptide and pro region in CPO. The identified exons encode for a protein of 352 amino acids, whose 25 last residues show no homology to CPs. There is no furin/PC consensus site within the region of CPO that aligns with the pro domain cleavage site of the other family peptides (Figure 2.8). Interestingly, Asp312, within the C-terminal region with no homology to CPs, is predicted to be a glycosylphosphatidylinositol

(GPI) anchoring site, according to the big-PI Predictor website (Eisenhaber *et al.*, 1999). GPI anchoring is a common post-translational modification of extracellular eukaryotic proteins. Attachment of the GPI moiety to the carboxyl terminus of the polypeptide occurs after proteolytic cleavage of a C-terminal propeptide. Thus, the last 22 residues of CPO would be processed and absent in the mature form of the enzyme. By means of GPI anchors proteins are


**Figure 2.7.- Alignment of the deduced amino acid sequence of human PCPA6 and porcine PCPB, human PCPB, human PCPA2 and bovine PCPA1. Pro-segments residues are coloured in green. The zinc-binding residues and the active site residues are marked in red, Cys residues forming disulphide bridges in blue, specificity pocket residues in positions 243, 255 and 268 in pink and the pro segment residue 41A in orange. Porcine PCPB numbering (Coll *et al.*, 1991).**

attached to membrane surface. A membrane-bound metallocarboxypeptidase has already been described, CPM, which belongs to the N/E subfamily and is involved in the local control of peptide hormone activity.

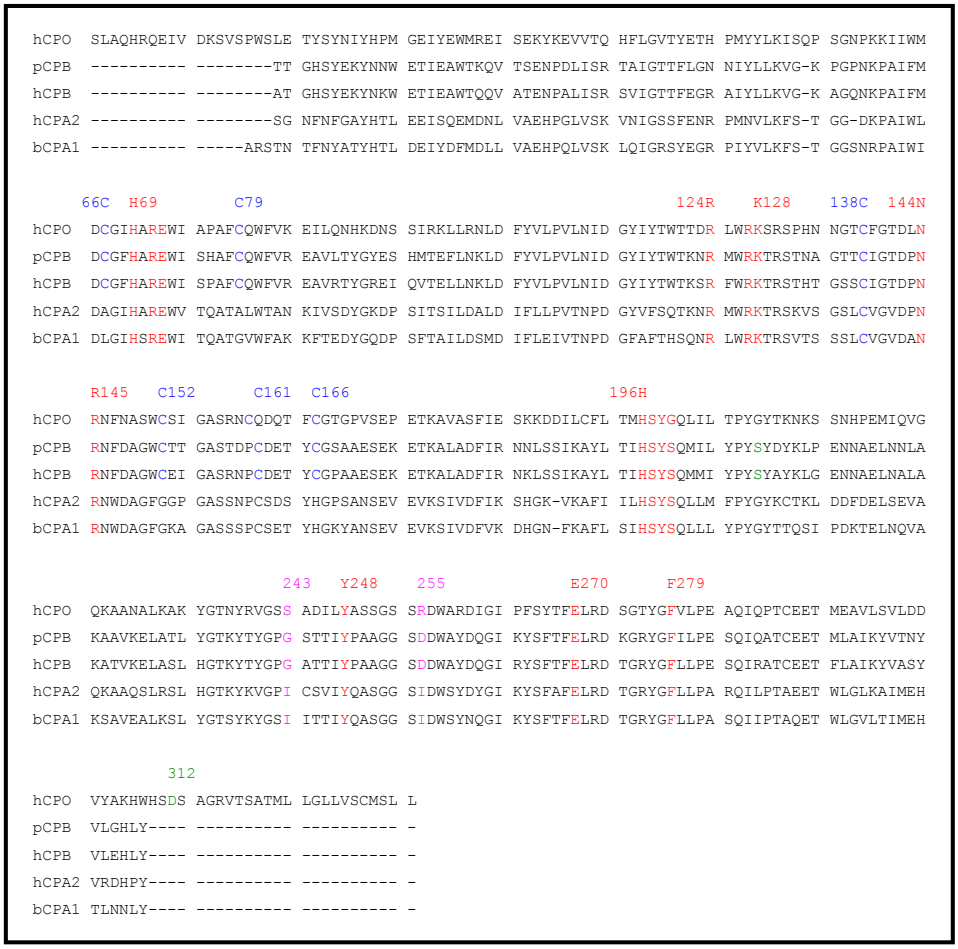


**Figure 2.8.- Alignment of the deduced amino acid sequence of human CPO and porcine CPB, human CPB, human CPA2 and bovine CPA1. The zinc-binding residues and the active site residues are coloured in red, Cys residues forming disulphide bridges in blue, specificity pocket residues in positions 243 and 255 in pink and residue in position 312, where GPI anchoring is predicted to occur, in green. Numbering according to bovine CPA1 (Rees *et al.*, 1983).**

Alignment of the amino acid sequences shows that the key residues for catalytic activity and substrate binding of other CPs are generally present in comparable positions in CPA5, CPA6 and CPO. Based on kinetic studies with rat CPB (Clauser *et al.*, 1988) and crystallographic studies with porcine CPB and CPA1 (Coll *et al.*, 1991; Guasch *et al.*, 1992), the important residues

include those involved in the coordination of the active site zinc atom (His69, Glu72 and His196) and a series of residues important for substrate binding and catalysis (Arg71, Gln122, Arg124, Arg127, Lys128, Asn144, Arg145, Ser197, Tyr198, Ser199, Tyr248, Glu270 and Phe279; the numbering and residue names used are those of bovine CPA1). All of the residues involved in zinc coordination and substrate catalysis are conserved in CPA5, CPA6 and CPO. Most of the residues involved in substrate binding are also conserved, and the differences observed in some of them are likely to be related to the individual specificities of the three forms. The length of the active CP domain is nearly identical in all members of the A/B subfamily, with the major differences in overall length due to small differences in the number of residues of the signal peptide. The overall amino acid identity of pre-PCPA5 is highest to pre-PCPA1 (60%), slightly lower to pre-PCPA2 and pre-PCPA4 (53 and 51%) and lower (34-36%) to the remaining members of the gene family (Table 2.4). Similar results are obtained when comparing just the active CP domains, although the amino acid identities are 4-7% higher without the signal peptides and the pro regions (Table 2.4). Pre-PCPA6 shows slightly higher amino acid sequence identity to pre-PCPA3 (41%) and pre-PCPB (40%) than the other members of the subfamily (Table 2.4). The amino acid identities are also higher when only the active CP domains are compared. Because the prepro region of CPO could not be determined, amino acid comparisons of this protein were done only with the active CP domains. CPO shows highest amino acid sequence identity to CPB (46%), CPA6 (45%), CPA3 (45%) and TAFI (44%).

### 2.3.B.3.- MODELLING OF THE NOVEL CPs AND PREDICTION OF THEIR SPECIFICITIES

To gain a better understanding of whether the novel CPs encode active enzymes as well as to predict the optimal substrates, a method of comparative modelling by satisfaction of the spatial restraints determined by sequence alignments was used to build the three-dimensional structure of the PCPs using the program MODELLER. For each particular protein, a subset of sequences with known PDB coordinates, chosen based on closest sequence similarity, was used as the starting alignment which, in every case, contained the restraints shown in the alignments in Figure 2.6, Figure 2.7 and Figure 2.8. In order to improve the model accuracy, a prediction of secondary structure was performed in every case. For PCPA5 and PCPA6, the prediction showed that the $\alpha$-helix of the connecting segment (the region that links the pro and enzyme domains) is similar to human PCPA2 and slightly longer than the $\alpha$-helix of PCPA1. Thus, several models were built in which the connecting region was only restrained by the structure of human PCPA2. These models showed an improved pseudo-energy profile with respect to those obtained using all restraints in this region. Due to the lack of information about the N-terminal pro region, CPO

was modelled on CP templates only. In all cases, the structure with the lowest pseudo-energy was chosen as the final model among the proposed ones.

| | CPA1 | CPA2 | CPA3 | CPA4 | CPA5 | CPA6 | CPB | TAFI |
|---|---|---|---|---|---|---|---|---|
| **CPA1** | 100 | 62 | 36 | 54 | 60 | 37 | 42 | 36 |
| **CPA2** | 66 | 100 | 37 | 63 | 53 | 38 | 40 | 38 |
| **CPA3** | 40 | 42 | 100 | 35 | 34 | 41 | 48 | 39 |
| **CPA4** | 58 | 68 | 42 | 100 | 51 | 38 | 40 | 35 |
| **CPA5** | 64 | 57 | 38 | 56 | 100 | 35 | 36 | 36 |
| **CPA6** | 41 | 42 | 47 | 42 | 42 | 100 | 40 | 38 |
| **CPB** | 47 | 46 | 53 | 47 | 42 | 48 | 100 | 41 |
| **TAFI** | 41 | 43 | 47 | 42 | 43 | 46 | 48 | 100 |
| **CPO** | 37 | 38 | 45 | 40 | 33 | 45 | 46 | 44 |

**Table 2.4.- Percentage of amino acid identity among members of the A/B subfamily. The upper right half of the table indicates identity among the full length (i.e. preprocarboxypeptidase) forms; the lower left half (blue) indicates identity only of the putative active forms (i.e. without the prepro sequence).**

The overall r.m.s.d. values calculated between the enzyme moieties of the final refined models and the different CP templates are as follows. CPA5: 0.28 Å, 0.34 Å and 0.37 Å for bovine CPA1, porcine CPA1 and human CPA2, respectively; CPA6: 0.49 Å, 0.39 Å, 0.66 Å and 0.69 Å for human CPB, porcine CPB, human CPA2, and bovine CPA1, respectively; CPO: 0.46 Å, 0.41 Å, 0.62 Å and 0.68 Å for human CPB, porcine CPB, human CPA2, and bovine CPA1, respectively. In the case of the two modelled proenzymes, the r.m.s.d. values are higher in all cases when the complete models are compared with their templates. For instance, the r.m.s.d. values for the PCPA5 model are 1.10 Å, 1.06 Å and 0.50 Å for bovine PCPA1, porcine PCPA1 and human PCPA2, respectively. This indicates that, while the enzyme domains of the PCPA5 or PCPA6 models and the templates are very similar, the pro segments are significantly different, especially when compared with PCPA1 structures. In the case of PCPA5, the main differences observed are allocated within the connecting segment, with its $\alpha$-helix being five turns long

(Figure 2.9), as in the three-dimensional structure of PCPA2, and spanning from Ile82A to Leu99A. This helix is followed by a short loop at the border with the CPA5 moiety, where a highly exposed arginine (Arg101A) is located. This arginine may be the first target for endopeptidase-mediated activation of the proenzyme. The connecting segment of PCPA1 has only a four-turn helix and the loop next to it is longer than that of PCPA5. As in PCPB, TAFI and PCPA3, the pro domain of PCPA6 contains a two residue deletion after position 34A and a four residues insertion after position 42A (porcine PCPB numbering) relative to all other CPAs (Figure 2.7). Thus, the model shows the one-turn $3_{10}$ helix located between beta strands 2 and 3 that is characteristic of the CPB pro region. However, unlike the pro region of CPB, in which Asp41A forms a salt bridge with the active site residue Arg145 (bovine CPA1 numbering) to completely inactivate the proenzyme toward all substrates, in PCPA6 an acidic amino acid is absent in this position (instead, a Ser residue is present, see in Figure 2.7).

In the predicted active sites of the three CP models, the residues involved in the coordination of the active site zinc ion and the series of conserved key residues that form the different active-center subsites have essentially the same conformation described for other CPs. Several residues at positions 194, 243, 253, 255 and 268 have been identified to influence enzyme specificity (bovine CPA1 numbering). The predictions of individual specificities are therefore based largely on the conformational and space-filling effects of these amino acid positions. CPAs hydrolyze hydrophobic C-terminal residues. CPA2 exhibits a preference for C-terminal bulky aromatic residues, whereas CPA1 prefers smaller aliphatic residues. This difference is thought to be primarily due to two substitutions in the specificity pocket of human CPA2 at positions 194 and 268 (Garcia-Saez *et al.*, 1997). Human CPA5 has the same residues as CPA1 in these positions (Figure 2.6). Two other residues in the specificity pocket of CPA5 are Ile253 and Val255, which correspond to residues Gly253 and Ile255 in CPA1 and CPA2. These changes would render the specificity pocket of CPA5 smaller than in the other CPs. CPA5 has only one predicted disulfide bond (between Cys138 and Cys161) which is in the same position as the disulfide bond in CPA1 and the other CPs. Human CPA2 contains a second disulfide bond between Cys210 and Cys244, which is unique for CPA2 and whose presence is thought to influence the specificity of the enzyme. This, together with the fact that the residues identified as responsible for the higher specificity of CPA2 for bulky aromatic residues are not present in CPA5 and that the substitution at positions 253 and 255 make the specificity pocket smaller, indicates that CPA5 is likely to exhibit a specificity for small aliphatic C-terminal residues.

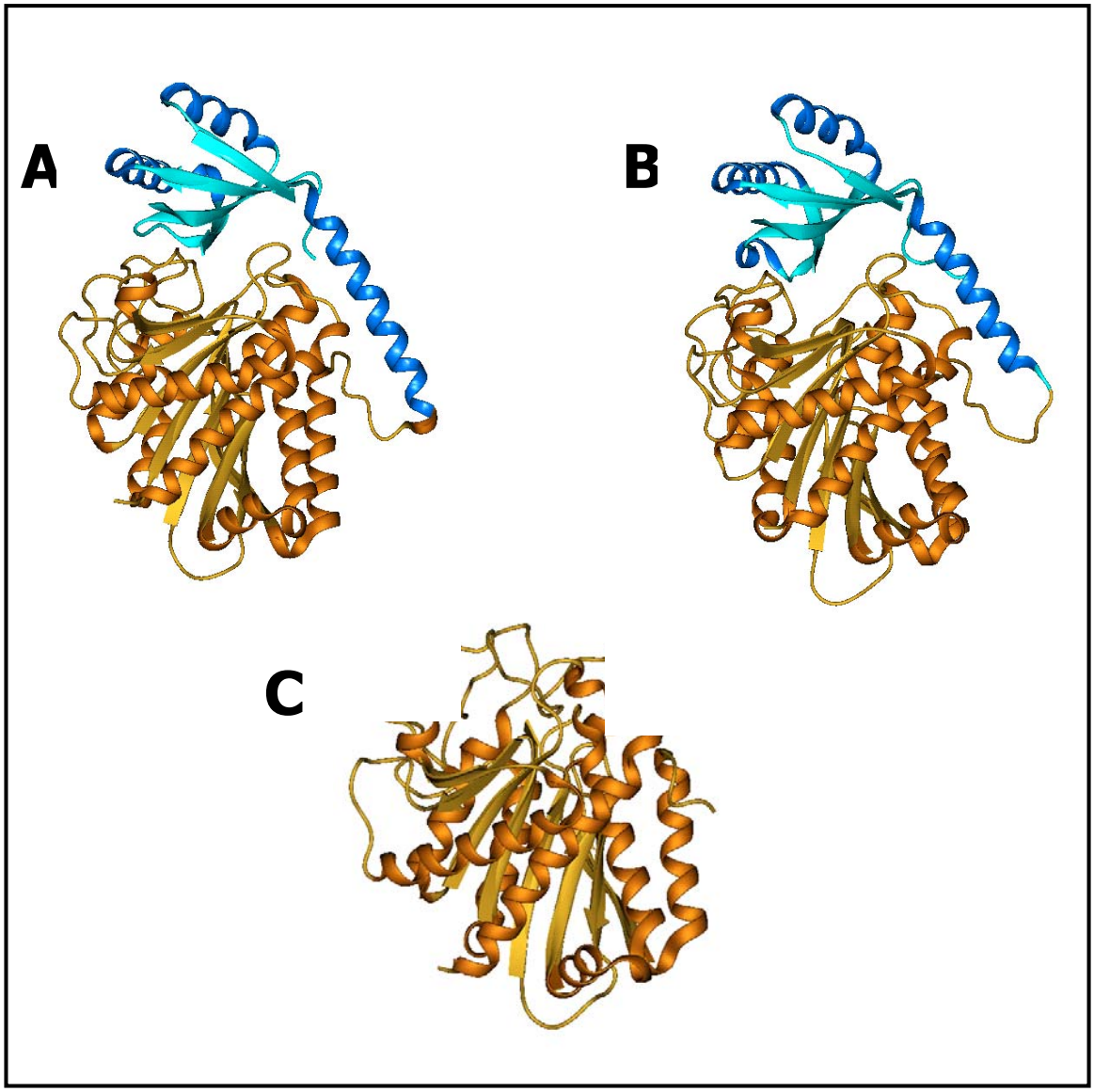


**Figure 2.9.- Ribbon representation of the modelled structures of PCPA5 (A), PCPA6 (B) and CPO (C). The pro and enzyme domains are shown in blue and gold, respectively, and the different color intensities indicate α or β secondary structures.**

There is a large overall coincidence in the hydrophobic nature of most of the residues that line the substrate binding pocket of CPA6 and CPA1. The CPA1-like activity of CPA6 may, however, be modulated by three substitutions: Ala243, Met255 and Ala268 in CPA6 correspond, respectively, to Ile243, Ile255 and Thr268 in CPA1 (bovine CPA1 numbering). These residues line the cavity where the side chain of the S1' residue sits. The smaller size of the residues lining the specificity pocket may indicate that CPA6 is more efficient in the hydrolysis of large,

probably branched hydrophobic C-terminal residues. The critical substrate-binding residue Met255 is conserved in the chicken and frog CPA6 sequences and, as it has been conserved during evolution, it may give the protein a unique CPA-like specificity relevant for its function. Interestingly, two of the eleven drosophila CPA/B-like genes have a Met in this position, but previously no known mammalian CPs had this residue.

While for CPA5 and CPA6 the predicted active sites are generally similar to other CPAs, in the case of CPO it is unique. CPO resembles CPB in the nature of the amino acids that form the subsites for substrate anchoring but with significant substitutions at four residues present in the lining of the S1' pocket: Gly243, Ala250, Gly253 and Asp255 (bovine CPA1 numbering). In CPO, these residues are substituted by Ser243, Ser250, Ser253 and Arg255, indicating that the substrate binding funnel is of similar nature but the specificity is reversed. Residue at position 255 is thought to be a key one in determining the substrate specificity of carboxypeptidases (Skidgel, 1988; Reynolds *et al.*, 1989b). This consideration was recently proved to be true with the obtention of reversed-specificity variants of human PCPB, where the single mutation of Asp255 to either Lys or Arg was sufficient to generate enzymes that were able to cleave substrates prior to a C-terminal acidic residue rather than a C-terminal basic residue (Edge *et al.*, 1998). Thus, it is very likely that CPO cleaves C-terminal acidic amino acids, a specificity not yet described for any known CP. Interestingly, searches of the *Drosophila* and the *C. elegans* genomes revealed metallo-CP-like genes with Lys and His in the position corresponding the active site residue 255; it is therefore possible that these proteins have specificities similar to that of human CPO.

# Annex al capítol 2

El treball presentat en aquest capítol s´ha realitzat en col.laboració amb el grup del Dr. Miquel Coll de la Unitat de Cristal·lografia de l´Institut de Biologia Molecular del C.S.I.C. de Barcelona i el grup del Dr. Lloyd D. Fricker del Department of Molecular Pharmacology, Albert Einstein College of Medicine, New York. Concretament, la resolució de l´estructura tridimensional de la PCPB humana ha estat realitzada pel Dr. Pedro J. Barbosa en el laboratori del Dr. Miquel Coll, mentre que la cerca de nous gens de metal·locarboxipeptidases humanes s´ha dut a terme en el laboratori del Dr. Lloyd D. Fricker. La meva contribució en aquest treball presenta una vessant experimental, pel que fa a l´obtenció i cristal·lització de la PCPB humana, i una de teòrica, quant a l´avaluació del significat dels alineaments i de les seves implicacions funcionals, i a la construcció dels diferents models presentats, aquest últim treball realitzat amb la supervisió del Dr. Baldomero Oliva.

# RESUM I DISCUSSIÓ GENERAL

L´estudi portat a terme en la present tesi doctoral tracta sobre la relació estructura/funció de metal·locarboxipeptidases humanes de la subfamília A/B. El treball realitzat s´ha dividit en dues parts diferenciades: una primera part que es centra en l´estudi detallat dels processos d´activació de la PCPA1 i PCPB humanes i una segona part que conté estudis estructurals de la PCPB humana pancreàtica i de diferents carboxipeptidases reguladores de la subfamília A/B, com són la TAFI i tres nous enzims identificats en aquest treball: la CPA5, la CPA6 i la CPO.

Les metal·locarboxipeptidases es poden dividir en dues subfamílies d´acord amb criteris estructurals i evolutius: la subfamília A/B i la subfamília N/E. La primera es caracteritza per estar composta d´enzims sintetitzats en forma de zimògens, amb segments *pro* d´una longitud considerable (75-76 residus per a les procarboxipeptidases de bacteris i fongs, 91-96 residus per a la resta), els quals han d´ésser separats del domini enzimàtic mitjançant un atac proteolític per part d´una proteasa activant. En contrast, dins la subfamília N/E aquests propèptids són absents o bé tenen una petita mida (5-17 residus) i una funció desconeguda. A més, mentre que la subfamília N/E només conté enzims de funció reguladora i amb especificitat per a residus C-terminals bàsics, la subfamília A/B conté enzims digestius, també anomenats pancreàtics, i reguladors i les especificitats que trobem són variades.

En la primera part d´aquest treball, l´optimització de sistemes d´expressió heteròloga i purificació per a la PCPA1 i PCPB humanes ha permès l´obtenció de quantitats suficients de proteïna per a realitzar un estudi detallat dels seus mecanismes d´activació. Així mateix, s´ha aplicat el mateix sistema d´expressió i purificació a un mutant de la PCPB humana obtingut mitjançant mutagènesi dirigida. Aquest mutant, que presenta una especificitat reversa respecte el tipus salvatge de la PCPB humana, ha estat utilitzat per realitzar un estudi comparatiu entre el seu procés d´activació i el del tipus salvatge.

En treballs previs s´havien caracteritzat de forma detallada els processos d´activació de la PCPA1 i PCPB porcines (Vendrell *et al.*, 1990c; Burgos *et al.*, 1991; Villegas *et al.*, 1995a; Ventura *et al.*, 1999; Companys, 2002), així com el de la PCPA2 humana (Reverter *et al.*, 1998). Aquests estudis havien mostrat un comportament diferencial entre els processos d´activació de la PCPB i la PCPA2, els quals són ràpids i monotònics, respecte del de la PCPA1, que presenta una cinètica d´activació lenta i bifàsica, tot i utilitzar-se condicions d´activació més dràstiques. Aquest enlentiment en l´aparició d´activitat carboxipeptidasa s´atribueix a la capacitat inhibidora que té el segment *pro*, i altres fragments llargs derivats d´ell, sobre l´enzim un cop ha tingut lloc l´atac proteolític. Per tal que el segment *pro* deixi d´interaccionar amb la carboxipeptidasa cal que tingui lloc un segon tall tríptic més intern. En canvi, per a la PCPB i

PCPA2, el tall tríptic primari és suficient per alliberar el segment *pro* de la carboxipeptidasa. Aquest treball mostra que, en contrast amb el que s´havia determinat per a la PCPA1 porcina, la PCPA1 humana presenta un mecanisme d´activació més semblant al de les formes B i A2. Així, per fer un seguiment adequat del seu procés d´activació calen unes condicions d´activació suaus (0 ºC, relació proenzim/tripsina 400/1 (p/p)), tal i com succeeix amb la PCPB i la PCPA2. En aquestes condicions, la cinètica d´activació de la PCPA1 humana no arriba a ser tan ràpida com la de la PCPB i PCPA2, però en tots els casos s´observa que el procés és monotònic. Aquesta observació indica que el segment *pro* del proenzim humà no presenta activitat inhibidora sobre la carboxipeptidasa generada, essent el tall tríptic primari suficient per alliberar-lo de l´enzim, fet que concorda amb l´absència d´una diana tríptica secundària en posició 74A. Així, el mecanisme d´activació de la PCPA1 humana es presenta com un procés senzill al llarg del qual tan sols es detecta l´alliberament del segment d´activació intacte de 96 residus i un fragment de 94 residus derivat d´ell, com a conseqüència de l´acció carboxipeptidàsica sobre els residus C-terminals Arg99A i Ser98A.

Els resultats obtinguts pel mecanisme d´activació de la PCPA1 humana divergeixen molt dels obtinguts pel proenzim porcí, cosa que resulta sorprenent, ja que són proenzims del mateix tipus i, donat l´elevat percentatge d´identitat seqüencial que comparteixen (82%), han de presentar estructures molt semblants. Tanmateix, han d´existir particularitats en la naturalesa i nombre d´interaccions entre el segment d´activació i el domini enzimàtic, les quals han de constituir els determinants estructurals responsables de què els dos proenzims presentin cinètiques d´activació tan diferents. Aquesta anàlisi estructural comparativa no s´ha pogut realitzar, perquè, de moment, no ha estat possible resoldre l´estructura tridimensional de la PCPA1 humana.

El segment *pro* consta de dues parts estructuralment diferenciades que estableixen interaccions característiques amb el domini carboxipeptidasa: el domini globular d´activació i el segment de connexió que uneix aquest domini globular amb la carboxipeptidasa. El segment de connexió és format per una hèlix $\alpha$ seguida d´un llaç exposat on té lloc l´atac tríptic primari. La llargada d´aquesta hèlix $\alpha$, que és major en la PCPA1 que en la PCPB, s´ha relacionat amb la capacitat inhibidora del segment *pro* de la PCPA1 porcina, degut a què la seva millor estructuració pot ajudar a mantenir els forts contactes de la regió de connexió amb la carboxipeptidasa (Avilés *et al.*, 1993). La predicció de l´estructura secundària de la seqüència de la PCPA1 humana indica que l´hèlix $\alpha$ del segment de connexió presenta una llargada equivalent a la de la forma porcina, de manera que no trobem una explicació de la seva activació diferencial en l´estructuració del segment de connexió. D´altra banda, l´anàlisi de la

seqüència del proenzim humà mostra la presència d´algunes substitucions en determinats residus que en el proenzim porcí estan implicats en interaccions importants entre el segment pro i el domini enzimàtic. Les interaccions alterades són Lys4A--Asp122, Glu5A--Asp122, Arg47A--Ile244 i Phe50A--Val246 entre el domini globular del segment d´activació i la carboxipeptidasa i les interaccions Ser96A--Asn8, Ser96A--Ala10 i Ser96A--Thr11 entre la regió de connexió i la carboxipeptidasa (Guasch *et al.*, 1992), ja que en el proenzim humà trobem Thr122, His47A, Ser50A i Phe96A. Tanmateix, no es pot avaluar la importància que poden tenir aquestes variacions fins que no es disposi de l´estructura tridimensional de la PCPA1 humana.

Estudis preliminars del procés d´activació de la PCPB humana realitzats amb una concentració de zimogen de 45$\mu$g/mL indicaven la necessitat d´una quantitat elevada de tripsina per tal d´assolir l´expressió de l´activitat màxima de l´enzim. El present estudi demostra que, emprant la concentració de proenzim estàndard d´1 mg/mL, la PCPB humana assoleix activitat màxima amb les condicions suaus d´activació típiques de les formes B (0 ºC, relació proenzim/tripsina 400/1 (p/p)). Aquests resultats indiquen que el procés d´activació de la PCPB humana presenta una cinètica que és sensible a la concentració de zimogen emprada. Sota les condicions abans esmentades, el proenzim humà segueix el característic procés d´activació ràpid i monotònic de la PCPB porcina, amb la particularitat de què el proenzim humà és menys eficient en el processament del seu segment *pro*. Aquesta observació es deriva del fet que el segment *pro* del proenzim humà és detectable en la seva forma completa, a diferència del corresponent segment del proenzim porcí, que pateix un ràpid processament carboxipeptidàsic de l´arginina C-terminal que queda després de l´acció tríptica primària, de manera que no es detecta el segment *pro* intacte. D´altra banda, per al proenzim humà no s´observa una evolució completa del fragment d´activació d´alt pes molecular (F1$\beta$) cap al fragment de menor pes molecular (F2$\beta$), tot i deixar llargs temps d´activació. Això contrasta amb el que s´observa pel proenzim porcí, en què al cap de 60 minuts aquesta transformació és quasi completa.

La diferent velocitat d´aparició de les activitats A i B conseqüència de diferències en els seus processos d´activació, ha estat abundosament discutida en els sistemes de porc i bou i sol ser considerada un reflex de la necessitat d´intervenció seqüencial a la digestió de les diferents activitats carboxipeptidasa. Així, l´aparició primerenca d´activitat CPB permetria la digestió de les cues C-terminals de pèptids alliberats per acció de la tripsina i sobre els productes d´aquesta digestió actuarien altres carboxipeptidases seqüencialment. És a dir, la diferent activabilitat de les procarboxipeptidases reflexaria una adaptació a la dieta de l´individu. Que aquestes hipòtesis tinguin una explicació en el marc del context de la nutrició humana és un tema que queda fora de l´abast d´aquest treball.

L´estudi comparatiu dels processos d´activació de la PCPB humana tipus salvatge i del seu mutant amb especificitat reversa D255K mostra que, tot i que segueixen una cinètica d´activació molt similar, el processament dels fragments generats durant l'activació és diferent. En concret, per al mutant D255K, que és incapaç de tallar arginines C-terminals, s´observa que els fragments generats com a conseqüència de l´acció tríptica mantenen el residu d´arginina en posició C-terminal, mentre que, en el cas del tipus salvatge, aquests residus són tallats. Aquests resultats demostren la participació de la CPB en el processament del seu segment *pro*. Anteriorment s´havien realitzat en el nostre grup experiments d´activació de la PCPB porcina en presència d´inhibidors de l´activitat carboxipeptidasa, els quals, així mateix, demostraven la participació de la CPB generada en la digestió del seu propi segment *pro* (Villegas *et al.*, 1995a). D´altra banda, s´ha observat que durant el procés d´activació del mutant D255K té lloc una acció tríptica sobre l´Arg93A que en el tipus salvatge no té lloc. Degut a què no s´ha detectat cap espècie que contingui els residus Val94A i Arg95A al llarg de l´esmentat procés, no s´ha pogut determinar si l´atac tríptic primari té lloc sobre l'Arg95A o l'Arg93A. De tota manera, la generació d´una nova diana tríptica pot indicar que la introducció de la mutació D255K en la butxaca d´especificitat provoca una distorsió de les interaccions entre el segment *pro* i el domini enzimàtic, afectant la conformació del llaç de la regió de connexió entre el domini d´activació globular i l´enzim. Un efecte d´aquest tipus es va descriure en l´estudi comentat anteriorment, en què l´activació de la PCPB porcina es va realitzar en presència d´inhibidors de la carboxipeptidasa (Villegas *et al.*, 1995a). En aquest cas, es va descriure el residu Arg93A com a lloc d´atac proteolític primari i es va atribuir la generació d´aquesta nova diana tríptica a l´alteració de la conformació del llaç de la regió de connexió com a causa de la difusió dels inhibidors cap al centre actiu.

La segona part del treball es centra en estudis estructurals de membres de la subfamília A/B de les metal·locarboxipeptidases. D´una banda, el desenvolupament d´un sistema d´expressió recombinant eficient ha permès la purificació i concentració d´elevades quantitats de PCPA1 i PCPB humanes per tal de dur a terme experiments de cristal·lització per a ambdós proenzims. Tot i això, només la forma B ha produït cristalls amb capacitat de difractar raigs X, de manera que només ha estat possible resoldre l´estructura de la PCPB humana.

L´estructura de la PCPB humana és molt semblant a la d´altres procarboxipeptidases prèviament estudiades i s´observa que la zona més variable és la regió de connexió entre el domini d´activació globular i el domini enzimàtic, la qual és una regió molt exposada i amb elevada flexibilitat. Un tret característic de l´estructura de la PCPB humana és que, a diferència

del proenzim porcí, presenta una regió N-terminal molt ben ordenada, formant una hèlix $3_{10}$ característica que precedeix la primera cadena $\beta$. Al igual que el que s´ha observat per a la PCPB porcina, el domini d´activació globular interacciona amb el domini carboxipeptidasa a través d´una complexa xarxa de ponts d´hidrogen directes i mediats per molècules d´aigua, així com de nombroses interaccions de van-der-Waals. Pel que fa a l´arquitectura del centre actiu, s´observa que el residu His69, implicat en la coordinació de l´ió de zinc catalític, presenta dues conformacions alternatives que causen petites diferències en les conformacions de residus dels subsetis S1´ i S1, involucrats en la catàlisi i en la fixació del residu C-terminal del substrat, com són els residus Arg127, Asn144, Arg145 i Tyr248. Aquests canvis de conformació provoquen petites alteracions en les interaccions característiques que aquests residus estableixen en altres estructures de carboxipeptidases estudiades. L´efecte que puguin tenir aquestes variacions sobre l´activitat de l´enzim humà és , de moment, desconegut.

La PCPB humana pancreàtica constitueix un prototip per a aquelles carboxipeptidases humanes que processen residus C-terminals bàsics i es sintetitzen en forma proenzimàtica, com és el cas de la TAFI. Així, la resolució de l´estructura tridimensional de la PCPB humana, junt amb la d´altres estructures de procarboxipeptidases prèviament obtingudes, ha permès l´obtenció d´un model acurat de la TAFI. El model obtingut mostra que la TAFI comparteix certs trets estructurals característics de les formes B, com són una deleció de dos residus i una inserció de quatre residus en el segment *pro* i la formació de tres ponts disulfur en el domini enzimàtic. D´altra banda, la TAFI presenta una deleció dels sis residus N-terminals presents a les formes B i sis insercions puntuals, tres de les quals es troben repartides en el domini enzimàtic i les altres tres es troben localitzades en el llaç de la regió de connexió on té lloc l´activació per acció de la trombina.

L´estudi del model de la TAFI permet l´elaboració d´una hipòtesi per explicar-ne una propietat funcional interessant, com és la inestabilitat conformacional intrínseca de la seva forma activa. El perfil d´energia pseudo-potencial del model calculat amb el programa PROSA (Sippl, 1993) mostra que dos residus del domini enzimàtic característics de la TAFI , Ile92 i Ile93 (corresponents a Ile182 i Ile183 segons la numeració correlativa del proenzim), presenten elevats valors d´energia pseudo-potencial, indicant que ocupen una posició dins de l´estructura que no és favorable. Efectivament, es va observar que aquests residus hidrofòbics es troben situats en un llaç exposat a superfície precedint una hèlix $\alpha$ també exposada, la qual presenta una baixa propensió per formar aquesta estructura. Una desestabilització de la regió on es troben localitzades aquestes isoleucines podria conduir a la inactivació de la forma activa de la TAFI. Un proper pas per comprovar la validesa d´aquesta hipòtesi seria l´obtenció d´una forma

mutant de la TAFI on es substituïssin els residus Ile92 i Ile93 pels que hi ha presents en aquesta posició, per exemple, en la PCPB humana, la qual presenta residus carregats en ambdues posicions.

Dins del capítol dedicat a estudis estructurals, s´han modelat les estructures de tres nous membres de la família M14 de les metal·locarboxipeptidases, els quals han estat identificats mitjançant cerques en la base de dades del genoma humà. El percentatge d´identitat entre les noves seqüències identificades i diferents membres de la família de les metal·locarboxipeptidases indica que tots tres enzims formen part de la subfamília A/B. D´acord amb la predicció de les seves especificitats els enzims s´han anomenat CPA5, CPA6 i CPO.

La CPA5 es sintetitza amb un pèptid senyal en posició N-terminal, seguit d´un segment d´activació de 96 residus. La comparació del model de la PCPA5 amb les estructures d´altres procarboxipeptidases mostra que aquest proenzim presenta major simil·litud amb la PCPA1 i PCPA2 pancreàtiques. L´anàlisi de la butxaca d´especificitat del domini enzimàtic indica que la substitució dels residus Gly253 i Ile255, presents en la CPA1 i la CPA2, pels residus més voluminosos Ile253 i Val255 de la CPA5, determina que la butxaca sigui més petita, cosa que fa pensar que la CPA5 pugui presentar especificitat per residus C-terminals alifàtics petits.

La seqüència aminoacídica deduïda per a la CPA6 codifica una proteïna de 437 residus, que, com en el cas de la CPA5, presenta un pèptid senyal i un segment d´activació de 99 residus. El model de la PCPA6 presenta una estructura més semblant a la PCPB, però l´especificitat predita per a l´enzim actiu és de tipus A, és a dir, presentaria preferència per al processament de residus C-terminals hidrofòbics. Una situació d´aquest tipus també es dóna en el cas de la PCPA3 de mastòcits, amb qui la PCPA6 comparteix el percentatge d´identitat més elevat. La butxaca d´especificitat de la CPA6 és força semblant a la de la CPA1, però les substitucions dels residus Ile243, Ile255 i Thr268 de la CPA1 pels residus Ala243, Met255 i Ala268 en la CPA6, de grandària menor, indiquen que la CPA6 podria presentar major eficiència en el processament de residus C-terminals hidrofòbics grans. Cal assenyalar que el residu Met255 pot tenir una funció clau en l´activitat de l´enzim, ja que es troba conservat en les seqüències de la CPA6 de pollastre i granota.

Els exons codificants de la regió N-terminal de la CPO no van poder ser identificats, de manera que no s´ha pogut determinar si aquest enzim és sintetitzat en forma proenzimàtica. Els exons identificats codifiquen una proteïna de 352 aminoàcids, que presenta una cua C-terminal de 25 residus sense homologia amb altres carboxipeptidases, per a la qual s´ha predit (Eisenhaber *et*

*al*., 1999) que pot actuar com a pèptid senyal d´unió a membrana mitjançant una àncora de glicosilfosfatidilinositol (GPI). Dins la família de les metal·locarboxipeptidases s´ha descrit que la CPM es troba unida a membrana mitjançant ancoratge amb GPI, des d´on actua exercint un control local de l´activitat d´hormones peptídiques.

La CPO comparteix el percentatge d´identitat més elevat amb la CPB pancreàtica i, tal com mostra el seu model, presenten estructures semblants. Com a tret característic d´aquest enzim cal assenyalar que en la seva butxaca d´especificitat trobem una sèrie de residus claus alterats respecte la CPB. Així, els residus Gly243, Ala250, Gly253 i Glu255 presents en la CPB són substituïts en la CPO per Ser243, Ser250, Ser253 i Arg255, canvis que indiquen, d´acord amb estudis de mutagènesi realitzats sobre la PCPB humana per obtenir variants amb especificitat reversa (Edge *et al*., 1998), que aquest enzim molt probablement mostra preferència per residus C-terminals àcids en comptes de bàsics. En base a aquesta observació aquest enzim ha estat anomenat CPO (*Other*), ja que presentaria una especificitat nova no descrita fins ara entre els membres de la família M14 de les metal·locarboxipeptidases.

# CONCLUSIONS

**Expressió de les procarboxipeptidases A1 i B humanes en *P. pastoris* i estudi detallat dels seus processos d´activació.**

1) L´ús del sistema d´expressió heteròloga que combina el llevat *P. pastoris* i el vector de secreció extracel·lular pPIC9 ha permès la producció eficient de les procarboxipeptidases A1 i B humanes recombinants, les quals són activables per tripsina, donant lloc a enzims amb les mateixes propietats que les formes aïllades de pàncrees. D´altra banda, fent ús del mateix sistema d´expressió, s´ha aconseguit produir un mutant de la PCPB humana, D255K, amb la mateixa eficiència que el tipus salvatge.

2) La PCPA1 humana recombinant presenta heterogeneïtat a l´extrem N-terminal degut al processament incorrecte de la senyal de secreció del factor d´aparellament $\alpha$ per part de les proteases de *P. pastoris*. Per tal de separar la forma nadiva de la PCPA1 de les formes mal processades, cal modificar el procés de purificació general mitjançant l´elució de la proteïna amb un gradient lent en dues etapes durant el primer pas de cromatografia.

3) L´estudi del procés d´activació tríptica de la PCPA1 humana mostra que calen unes condicions suaus per tal de poder fer un seguiment del procés, el qual segueix una corba monofàsica, en contrast amb la PCPA1 porcina, que necessita unes condicions d´activació més dràstiques i, tot i això, segueix una cinètica d´activació lenta i bifàsica. Així, la PCPA1 humana presenta un comportament en l´activació més semblant al de les formes B i A2.

4) El segment *pro* de la PCPA1 humana no és un inhibidor de l´enzim actiu un cop s´ha donat el tall tríptic activant, a diferència del corresponent segment del proenzim porcí, el qual interacciona fortament amb la carboxipeptidasa inhibint-la. Donat l´elevat nivell d´identitat seqüencial entre els dos proenzims (82%), les particularitats estructurals que es puguin extreure de la comparació de les dues estructures poden ser de gran rellevància per establir els trets estructurals determinants de la cinètica d´activació tríptica.

5) El procés d´activació de la PCPB humana és sensible a canvis en la concentració de proenzim, com indica el diferent comportament d´activació observat en emprar concentracions de proenzim de 45 $\mu$g/mL i d´1 mg/mL.

6) El procés d´activació de la PCPB humana és semblant al de la PCPB porcina, excepte en què la forma humana es mostra menys eficient en el processament del seu segment *pro*, essent possible detectar el seu segment *pro* complet.

7) L´estudi del procés d´activació del mutant D255K, incapaç de tallar arginines C-terminals, demostra la participació de la CPB tipus salvatge en el processament del seu segment *pro*, ja que, pel mutant, els fragments generats com a conseqüència de l´acció tríptica mantenen el residu d´arginina en posició C-terminal, mentre que, en el cas del tipus salvatge, aquests residus són tallats.

8) La generació d´una nova diana tríptica sobre el residu Arg93A durant el procés d´activació del mutant D255K pot ser deguda a què la introducció de la mutació en la butxaca d´especificitat provoca una distorsió de les interaccions entre el segment *pro* i el domini enzimàtic, afectant la conformació del llaç de la regió de connexió.

**Estructura tridimensional de la procarboxipeptidasa B humana i modelat de la TAFI i de tres nous membres de la família de metal·locarboxipeptidases humanes.**

9) Gràcies al desenvolupament d´un sistema d´expressió recombinant eficient s´han pogut obtenir cristalls de la PCPB humana amb capacitat de difractar raigs X, els quals han permès recollir dades fins a una resolució d´1,6 Å.

10) La PCPB humana presenta una estructura molt similar a la d´altres procarboxipeptidases prèviament estudiades i les diferències més remarcables es troben a la regió de connexió entre el domini d´activació globular i el domini enzimàtic. Aquests dos dominis interaccionen a través d´una xarxa complexa de ponts d´hidrogen directes i mediats per molècules d´aigua, així com d´interaccions de van-der-Waals. D´altra banda, les dues conformacions alternatives presentades pel residu His69, causen petites diferències en les conformacions de residus catalítics i involucrats en la fixació de substrat.

11) La disponibilitat de l´estructura de la PCPB humana, juntament amb la d´altres procarboxipeptidases prèviament obtingudes, ha permès l´obtenció d´un model acurat de la TAFI, que, en absència d´una estructura experimental, ha de permetre aprofundir en el coneixement de la seva relació estructura/funció, així com constituir la base per al disseny de fàrmacs que modulin la seva activitat biològica.

12) L´anàlisi del model de la TAFI obtingut permet l´elaboració d´una hipòtesi que expliqui la inestabilitat conformacional intrínseca de la seva forma activa. Concretament, s´ha detectat la presència en el domini enzimàtic de dos residus hidrofòbics característics de la TAFI, Ile92 i Ile93, en un llaç exposat a superfície, els quals presenten elevats valors d´energia pseudo-potencial. Una desestabilització de la regió on es troben localitzades aquestes isoleucines podria conduir a la inactivació de la forma activa de la TAFI.

13) Tres noves metal·locarboxipeptidases pertanyents a la subfamília A/B han estat identificades mitjançant cerques en la base de dades del genoma humà, les quals han estat anomenades, d´acord amb la predicció de les seves especificitats, CPA5, CPA6 i CPO. Prenent com a referència les estructures cristal·logràfiques de diferents membres de la subfamília A/B s´ha obtingut un model estructural per a cada un dels nous enzims.

14) La PCPA5 presenta major simil·litud amb la PCPA1 i PCPA2 pancreàtiques i la presència de dues substitucions en les posicions 253 i 255 de la butxaca d´especificitat per residus més voluminosos determinen que aquesta sigui més petita, cosa que indica que la CPA5 pot presentar especificitat per residus C-terminals alifàtics petits.

15) Tot i que per a la PCPA6 es prediu una preferència de l´enzim actiu per residus C-terminals hidrofòbics, estructuralment és més semblant a la PCPB. Respecte la butxaca d´especificitat, aquesta és força semblant a la de la CPA1, però existeixen tres substitucions en posicions 243, 255 i 268 per residus de grandària menor en la CPA6, que indiquen que aquest enzim podria presentar major eficiència en el processament de residus C-terminals hidrofòbics grans, probablement ramificats.

16) El model de la CPO presenta una estructura semblant a la de la CPB pancreàtica. Tanmateix, una sèrie de substitucions en la butxaca d´especificitat de la CPO en les posicions 243, 250, 253 i 255 indiquen que aquest enzim pot presentar una especifitat reversa respecte la CPB, mostrant preferència per residus C-terminals àcids en comptes de bàsics. Aquest tipus d´especificitat constituiria un cas únic entre tots els membres de la família de les metal·locarboxipeptidases.

# BIBLIOGRAFIA

Aloy, P., Catasús, Ll., Villegas, V., Vendrell, J. and Avilés, F.X. (1998) "Comparative analysis of the sequence and three-dimensional models of human procarboxypeptidases A1, A2 and B" *Biol. Chem.* 379, 145-149.

Álvarez-Santos, S., González-Lafont, A., Lluch, J.M., Oliva, B. and Avilés, F.X. (1994) "On water-promoted mechanism of peptide cleavage by carboxypeptidase A. A theoretical study" *Canadian J. Chemistry* 72, 2077-2083.

Anson, M.L. (1935) "Crystalline carboxypeptidase" *Science* 21, 467-468.

Aurora, R. and Rose, G.D. (1998) "Helix capping" *Prot. Sci.* 7, 21-38.

Avilés, F.X., Vendrell, J., Burgos, J.B., Turner, C., Cary, P.D. and Crane-Robinson, C. (1985) "Nuclear magnetic resonance studies on the isolated activation segment from porcine pancreas procarboxypeptidase A" *Biochem. Soc. Trans.* 13, 344-345.

Avilés, F.X., Vendrell, J., Guasch, A., Coll, M. and Huber, R. (1993) "Advances in metalloprocarboxypeptidases" *Eur. J. Chem.* 211, 381-389.

Bajzar, L., Manuel, R. and Nesheim, M.E. (1995) "Purification and characterization of TAFI, a thrombin-activable fibrinolysis inhibitor" *J. Biol. Chem.* 270, 14477-14484.

Bajzar, L., Morser, J. and Nesheim, M. (1996) "TAFI, or plasma procarboxypeptidase B, couples the coagulation and fibrinolytic cascades through the thrombin-thrombomodulin complex" *J. Biol. Chem.* 271, 16603-16608.

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

Birnboim, H.C. and Doly, J. (1979) "A rapid alkaline extraction procedure for screening recombinant plasmid DNA" *Nucleic Acids Res*. 6, 1513-1523.

Bodary, P.F, Wickenheiser, K.J. and Eitzman, D.T. (2002) "Recent advances in understanding endogenous fibrinolysis: implications for molecular-based treatment of vascular disorders" *Exp. Rev. Mol. Med*. http://www.expertreviews.org/02004362h.htm

Boffa, M.B., Wang, W., Bajzar, L. and Nesheim, M.E. (1998) "Plasma and recombinant thrombin-activable fibrinolysis inhibitor (TAFI) and activated TAFI compared with respect to glycosylation, thrombin/thrombomodulin-dependent activation, thermal stablility and enzymatic properties" *J. Biol. Chem.* 273, 2127-2135.

Boffa, M.B., Bell, R., Stevens, W.K. and Nesheim, M.E. (2000) "Roles of thermal instability and proteolytic cleavage in regulation of activated thrombin-activable fibrinolysis inhibitor" *J. Biol. Chem.* 275, 12868-12878.

Bown, D.P., Wilkinson, H.S. and Gatehouse, J.A. (1998) "Midgut carboxypeptidase from *Helicoverpa armigera* (Lepidoptera: *Noctuidae*): enzyme characterisation, cDNA cloning and expression" *Insect. Biochem. Mol. Biol.* 28, 739-749.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) "Crystallography & NMR system: A new software suite for macromolecular structure determination" *Acta Crystallogr. sect. D* 54, 905-921.

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.

Campbell, W. and Okada, H. (1989) "An arginine specific carboxypeptidase generated in blood during coagulation or inflammation which is unrelated to carboxypeptidase N or its subunits" *Biochem. Biophys. Res. Commun.* 162, 933-939.

Campbell, W., Yonezu, K., Shinohara, T. and Okada, H. (1990) "An arginine carboxypeptidase generated during coagulation is diminished or absent in patients with rheumatoid arthritis" *J. Lab. Clin. Med.* 115, 610-612.

Campbell, W., Okada, N. and Okada, H. (2001) "Carboxypeptidase R is an inactivator of complement-derived inflammatory peptides and an inhibitor of fibrinolysis" *Immunol. Rev.* 180, 162-167.

Campbell, W.D., Lazoura, E., Okada, N., Okada, H. (2002) "Inactivation of C3a and C5a octapeptides by carboxypeptidase R and carboxypeptidase N" *Microbiol. Immunol.* 46, 131-134.

Catasús, L., Villegas, V., Pascual, R., Avilés, F.X., Wicker-Planquart, C., and Puigserver, A. (1992) "cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1" *Biochem. J.* 287, 299-303.

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.* 263, 17837-17845.

Catasús, Ll. (1995) Tesi doctoral UAB "Procarboxipeptidases de pàncrees humà".

Casadaban, M.J. and Cohen, S.N. (1980) "Analysis of gene signals by DNA fusion and cloning in *Escherichia coli* " *J. Mol. Biol.* 138, 179-207.

Chen, C.C., Wang, S.S., Chen, T.W., Jap, T.S., Chen, S.J., Jeng, F.S. and Lee, S.D. (1996) "Serum procarboxypeptidase B, amylase and lipase in chronic renal failure" *J. Gastroenterol. Hepatol.* 11, 496-499.

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

Clauser, E., Gardell, S.J., Craick, C.S., McDonald, R.J. and Rutter, W.J. (1988) "Structural characterization of the rat carboxypeptidase A1 and B genes" *J. Biol. Chem.* 263, 17837-17845.

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.

Collaborative Computational Project No. 4. (1994) *Acta Cryst.* D50, 760-763.

Companys, V., Aloy, P., Vendrell, J., Aviles, F.X., Fricker, L.D., Coll, M. and Gomis-Ruth, F.X. (2001) "The crystal structure of the inhibitor-complexed carboxypeptidase D domain II and the modeling of regulatory carboxypeptidases" *J. Biol. Chem.* 276, 6177-6184.

Companys, V. (2002) Tesi doctoral UAB "Aproximación experimental y teórica a la carboxipeptidasas pancreáticas y reguladoras"

Cregg, J.M., Vedvick, T.S. and Raschke, W.C. (1993) "Recent advances in the expression of foreign genes in *Pichia pastoris*" *Biotechnology* 11, 905-910.

Csorba, T.R. (1991) "Proinsulin: biosynthesis, conversion, assay methods and clinical studies" *Clin. Biochem*. 24, 447-454.

Dominguez, D.I., De Strooper, B. and Annaert, W. (2001) "Secretases as therapeutic targets for the treatment of Alzheimer's disease" *Amyloid* 8, 124-142.

Eaton, D., Malloy, B., Tsai, S., Henzel, W. and Drayna, D. (1991) "Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase B from human plasma" *J. Biol. Chem.* 266, 21833-21838.

Eddy, S.R. (1998) "Profile hidden Markov models" *Bioinformatic*s 14, 755-763.

Edge, M., Forder, C., Hennam, J., Lee, I., Tonge, D., Hardern, I., Fitton, J., Eckersley, K., East, S., Shufflebotham, A., Blakey, D. and Slater, A. (1998) "Engineered human carboxypeptidase B enzymes that hydrolyse hippuryl-L-glutamic acid: reversed-polarity mutants" *Protein Eng.* 12, 1229-1234.

Edwards, M.J., Lemos, F.J., Donnelly-Doman, M. and Jacobs-Lorena, M. (1997) "Rapid induction by a blood meal of a carboxypeptidase gene in the gut of the mosquito *Anopheles gambiae*" *Insect .Biochem. Mol. Biol.* 27, 1063-1072.

Edwards, M.J., Moskalyk, L.A., Donelly-Doman, M., Vlaskova, M., Noriega, F.G., Walker, V.K. and Jacobs-Lorena, M. (2000) "Characterization of a carboxypeptidase A gene from the mosquito, *Aedes aegypti*" *Insect. Mol. Biol.* 9, 33-38.

Eisenhaber, B., Bork, P. and Eisenhaber, F. (1999) "Prediction of potential GPI-modification sites in proprotein sequences" *J. Mol. Biol*. 292, 741-758.

Eng, F.J., Novikova, E.G., Kuroki, K., Ganem, D. and Fricker, L.D. (1998) "gp180, a protein that binds duck hepatitis B virus particles, has metallocarboxypeptidase D-like enzymatic activity" *J. Biol. Chem.* 273, 8382-8388.

Erdös, E.G. (1979) A: "Handbook of experimental pharmacology" Erdös EG (Ed) Springer Verlag, Heidelberg, pp 427-448.

Estébanez-Perpiñá, E., Bayes, A., Vendrell, J., Jongsma, M.A., Bown, D.P., Gatehouse, J.A., Huber, R., Bode, W., Aviles, F.X. and Reverter, D. (2001) "Crystal structure of a novel mid-gut procarboxypeptidase from the cotton pest *Helicoverpa armigera*" *J. Mol. Biol.* 313, 629-638.

Evans, S. V. (1993) "SETOR: hardware-lighted three-dimensional solid model representations of macromolecules" *J. Mol. Graphics* 11, 134-138.

Fan, X., Qian, Y., Fricker, L.D., Akalal, D.B and Nagle, G.T. (1999) "Cloning and expression of Aplysia carboxypeptidase D, a candidate prohormone-processing enzyme" *DNA Cell. Biol.* 18, 121-32.

Folk, J.E., Piez, K.A., Carrol, W.R. and Gladner, J.A. (1960) "Carboxypeptidase B. IV. Purification and characterization of the porcine enzyme" *J. Biol. Chem.* 235, 2272-2277.

Franco, R.F., Fagundes, M.G., Meijers, J.C., Reitsma, P.H., Lourenco, D., Morelli, V., Maffei, F.H., Ferrari, I.C., Piccinato, C.E., Silva, W.A. Jr. and Zago, M.A. (2001) "Identification of polymorphisms in the 5'-untranslated region of the TAFI gene: relationship with plasma TAFI levels and risk of venous thrombosis" *Haematologica* 86, 510-517

Fricker, L.D., Adelman, J.P., Douglass, J., Thompson, R.C., von Strandmann, R.P. and Hutton, J. (1989) "Isolation and sequence analysis of cDNA for rat carboxypeptidase E (EC 3.4.17.10), a neuropeptide processing enzyme" *Mol. Endocrinol.* 3, 666-673.

Fricker, L.D., Berman, Y.L., Leiter, E.H. and Devi, L.A. (1996) "Carboxypeptidase E activity is deficient in mice with the fat mutation. Effect on peptide processing" *J. Biol. Chem.* 271, 30619-30624.

García-Sáez, I., Reverter, D., Vendrell, J., Avilés, F.X. and 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.

Gardell, S.J., Craick, C.S., Clauser, E., Goldsmith, E.J., Stewart, C.B., Graf, M. and Rutter, W.J. (1985) "Site-directed mutagenesis shows that tyrosine 248 of carboxypeptidase A does not play a crucial role in catalysis" *Nature* 317, 551-555.

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

Gomis-Rüth, F.X., Gómez-Ortiz, M., Vendrell, J., Ventura, S., Bode, W., Huber, R. and Avilés, F.X. (1997) "Crystal structure of an oligomer of proteolytic zymogens: detailed conformational analysis of the bovine ternary complex and implications for their activation" *J. Mol. Biol.* 269, 861-880.

Gomis-Rüth, F.X., Companys, V., Qian, Y., Fricker, L.D., Vendrell, J., Aviles, F.X. and Coll, M. (1999) "Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallocarboxypeptidase subfamily" *EMBO J.* 18, 5817-5826.

Görisch, H. (1988) "Drop dialysis: time course of salt and protein exchange" *Anal. Biochem*. 173, 393-398.

Greene, D., Das, B. and Fricker, L.D. (1992) "Regulation of carboxypeptidase E. Effect of pH, temperature and Co2+ on kinetic parameters of substrate hydrolysis" *Biochem. J.* 285, 613-618.

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 zimogens and their determinants for inhibition and activation" *J. Mol. Biol.* 224, 141-157.

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

Hendriks, D., Wang, W., Scharpe, S., Lommaert, M.P. and van Sande, M. (1990) "Purification and characterization of a new arginine carboxypeptidase in human serum" *Biochim. Biophys. Acta* 1034, 86-92.

Henry, M., Aubert, H., Morange, P.E., Nanni, I., Alessi, M.C., Tiret, L. and Juhan-Vague, I. (2001) "Identification of polymorphisms in the promoter and the 3' region of the TAFI gene: evidence that plasma TAFI antigen levels are strongly genetically controlled" *Blood* 97, 2053-2058.

Hillenkamp, F., Karas, M., Beavis, R.C. and Chait, B.T. (1991) "Matrix-assisted laser desorption/ionisation mass spectrometry" *Anal. Chem.* 63, 1193-1203.

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

Huang, H., Reed, C.P., Zhang, J.S., Shridhar, V., Wang, L., Smith, D.I. (1999) "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone deacetylase inhibitors during differentiation of prostate epithelial cancer cells" *Cancer Res.* 59, 2981-2988.

Joshi, L. and St. Leger, R.J. (1999) " Cloning, expression, and substrate specificity of MeCPA, a zinc carboxypeptidase that is secreted into infected tissues by the fungal entomopathogen *Metarhizium anisopliae*" *J. Biol. Chem.* 274, 9803-9811.

Juvvadi, S., Fan, X., Nagle, G.T. and Fricker, L.D. (1997) "Characterization of Aplysia carboxypeptidase E" *FEBS Lett.* 408, 195-200.

Kabsch, W. and Sander, C. (1983) "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features" *Biopolymers* 22, 2577-2637.

Kassell, B. and Kay, J. (1973) "Zymogens of proteolytic enzymes. These enzyme precursors, formerly thought to be inert substances, have inherent proteolytic activity" *Science* 180, 1022-1027.

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.

Kokame, K., Zheng, X. and Sadler, E. (1998) "Activation of thrombin-activable fibrinolysis inhibitor requires epidermal growth factor-like domain 3 of trombomodulin and is inhibited competitively by protein C" *J. Biol. Chem.* 273, 12135-12139.

Koschinsky, M.L., Boffa, M.B., Nesheim, M.E., Zinman, B., Hanley, A.J., Harris, S.B., Cao, H. and Hegele, R.A. (2001) "Association of a single nucleotide polymorphism in CPB2 encoding the thrombin-activable fibrinolysis inhibitor (TAFI) with blood pressure" *Clin. Genet.* 60, 345-349.

Kozak, M. (1984) "Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs" *Nuc. Acids Res.,* 12, 857-568.

Kuroki, K., Cheung, R., Marion, P.L. and Ganem, D. (1994) "A cell surface protein that binds avian hepatitis B virus particles" *J. Virol.* 68, 2091-2096.

Laemmli, U.K. (1970) "Cleavage of structural proteins during the assembly of the head of bacteriophage T4" *Nature* 227, 680-685.

Layne, M.D., Endege, W.O., Jain, M.K., Yet, S.F., Hsieh, C.M., Chin, M.T., Perrella, M.A., Blanar, M.A., Haber, E. and Lee, M.E. (1998) "Aortic carboxypeptidase-like protein, a novel protein with discoidin and carboxypeptidase-like domains, is up-regulated during vascular smooth muscle cell differentiation" *J. Biol. Chem.* 273, 15654-15660.

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

Leiter EH, Kintner J, Flurkey K, Beamer WG, Naggert JK (1999) "Physiologic and endocrinologic characterization of male sex-biased diabetes in C57BLKS/J mice congenic for the fat mutation at the carboxypeptidease E locus" *Endocrine* 10, 57-66.

Lisman, T., Leebeek, F.W., Mosnier, L.O., Bouma, B.N., Meijers, J.C., Janssen, H.L., Nieuwenhuis, H.K. and De Groot, P.G. (2001) "Thrombin-activatable fibrinolysis inhibitor deficiency in cirrhosis is not associated with increased plasma fibrinolysis" *Gastroenterology* 121, 131-139.

Mandel, M. and Higa, A. (1970) "Calcium-dependent bacteriophage DNA infection" *J. Mol. Biol.* 53, 159-162.

Marx, P.F., Hackeng, T.M., Dawson, P.E., Griffin, J.H., Meijers, J.C. and Bouma, B.N. (2000) "Inactivation of active thrombin-activable fibrinolysis inhibitor takes place by a process that involves conformational instability rather than proteolytic cleavage" *J. Biol. Chem.* 275, 12410-12415.

Masuda, K, Yoshioka, M., Hinode, D. and Nakamura, R. (2002) "Purification and characterization of arginine carboxypeptidase produced by *Porphyromonas gingivalis*" *Infect. Immun.* 70, 1807-1815.

McDonald, J.K. (1985) "An overview of protease specificity and catalytic mechanisms: aspects related to nomenclature and classification" *Histochem. J.* 17, 773-785.

McGuire, G.B. and Skidgel, R.A. (1995) "Extracellular conversion of epidermal growth factor (EGF) to des-Arg53-EGF by carboxypeptidase M" *J. Biol. Chem.* 270, 17154-17158.

Mosnier, L.O., Lisman, T., van den Berg, H.M., Nieuwenhuis, H.K., Meijers, J.C. and Bouma, B.N. (2001) "The defective down regulation of fibrinolysis in haemophilia A can be restored by increasing the TAFI plasma concentration" *Thromb. Haemost.* 86, 1035-1039.

Muise, A.M. and Ro, H.S. (1999) "Enzymic characterization of a novel member of the regulatory B-like carboxypeptidase with transcriptional repression function: stimulation of enzymic activity by its target DNA" *Biochem. J.* 343, 341-345.

Muñoz, V., and Serrano, L. (1997) "Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms" *Biopol.* 41, 495-509.

Nakayama, K. (1997) "Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins" *Biochem. J.* 327, 625-635.

Narahashi, Y. (1990) "The amino acid sequence of zinc-carboxypeptidase from *Streptomyces griseus*" *J. Biochem.* 107, 879-886.

Navaza, J. (1994) "AMoRe: an automated package for molecular replacement" *Acta Cryst.* sect. A50, 157-163.

Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" *Protein Eng.* 10, 1-6.

Oliva, B., Wastlund, M., Nilsson, O., Cardenas, R., Querol, E., Aviles, F.X. and Tapia, O. (1991) "Stability and fluctuations of the potato carboxypeptidase A protein inhibitor fold: a molecular dynamics study" *Biochem. Biophys. Res. Commun.* 176, 616-621.

Oliva, B., Bates, P. A., Querol, E., Aviles, F. X., and Sternberg, M. J. (1997) "An automated classification of the structure of protein loops" *J. Mol. Biol.* 266, 814-830.

Opezzo, O., Ventura, S., Bergman, T., Vendrell, J., Jörnvall, H. and Avilés, F.X. (1994) "Overall characterization and comparison of the activation processes of procarboxypeptidases from rat pancreas" *Eur. J. Biochem.* 222, 55-63.

Orengo, C.A., Brown, N.P. and Taylor, W.R. (1992) "Fast structure alignment for protein databank searching" *Proteins* 14, 139-167.

Otwinowski, Z. and Minor, W. (1993) "*DENZO: a film processing for macromolecular crystallography"*, Yale University, New Haven.

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.

Pascual, R., Vendrell, J., Avilés, F.X., Bonicell, J., Wicker, C. and Puigserver, A. (1990) "Autolysis of proproteinase E in bovine procarboxypeptidase A ternary complex gives rise to subunit III" *FEBS Letters* 277, 37-41.

Peterson, L.M., Holmquist, B. and Bethune, J.L. (1982) "A unique activity assay for carboxypeptidase A" *Biochemistry* 15, 2501-2508.

Philips, M.A., Fletterich, R. and Rutter, W.J. (1990) "Arginine-127 stabilizes the transition state in carboxypeptidase" *J. Biol. Chem.* 265, 20692-20698.

Phillips, M.A. and Rutter, W.J. (1996) "Role of the prodomain in folding and secretion of rat pancreatic carboxypeptidase A1" *Biochemistry* 35, 6771-6776.

Puigserver, A., Chapus, C. and Kerfelec, B. (1986) "Pancreatic exopeptidases". En: "Molecular and cellular basis of digestion" Eds. Desnuelle P. *et al*. Elsevier Publ. Capítulo 12, 235-247.

Quiocho, F.A. and Lipscomb, W.N. (1971) "Carboxypeptidase A: a protein and an enzyme" *Adv. Protein Chem.* 25, 1-78.

Ramos, A., Mahowald, A. and Jacobs-Lorena, M. (1993) "Gut-specific genes from the black fly *Simulium vittatum* encoding trypsin-like and carboxypeptidase-like proteins" *Insect. Mol. Biol.* 1, 149-163.

Rath-Wolfson, L. (2001) "An immunocytochemical approach to the demonstration of intracellular processing of mast cell carboxypeptidase" *Appl. Immunohistochem. Mol. Morphol.* 9, 81-85.

Rawlings, N.D. and Barret, A.J. (1995) "Evolutionary families of metallopeptidases" *Methods Enzymol.* 248, 183-228.

Rawlings, N.D. and Barret, A.J. (1999) "MEROPS: the peptidase database" *Nucleic Acids Res.* 27, 325-331.

Rawlings, N.D., O´Brien, E. and Barret, A.J. (2002) "MEROPS: the protease database" *Nucleic Acids Res.* 30, 343-346.

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

Rees, D.C., Lewis, M. and Lipscomb, W.N. (1983) "Refined Crystal Structure of carboxypeptidase A at 1.5 Å resolution" *J. Mol. Biol.* 168, 367-387.

Renart, J., Reiser, J. and Stark, G.R. (1979) "Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure" *Proc. Natl. Acad. Sci.* 76, 3116-3120.

Reverter, D., García-Sáez, I., Catasús, Ll., Vendrell, J., Coll, M. and Avilés, F.X. (1997) "Characterization and preliminary X-ray diffraction analysis of human procarboxypeptidase A2" *FEBS Letters* 420, 7-10.

Reverter, D., Ventura, S., Villegas, V., Vendrell, J. and Avilés, F.X. (1998) "Human procarboxypeptidase A2: overexpression in *Pichia pastoris* and detailed characterization of its activation process" *J. Biol. Chem.* 273, 3535-3541.

Reverter, D., Fernandez-Catalan, C., Baumgartner, R., Pfander, R., Huber, R., Bode, W., Vendrell, J., Holak, T.A. and Aviles, F.X. (2000) "Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2" *Nat. Struct. Biol.* 7, 322-328.

Reynolds, D.S., Stevens, R.L., Gurley, D.S., Lane, W.S., Austen, K.F. and Serafin, W.E. (1989a) "Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member of the carboxypeptidase gene family" *J. Biol. Chem.* 264, 20094-20099.

Reynolds, D.S., Gurley, D.S., Stevens, R.L., Sugarbaker, D.J., Austen, K.F. and Serafin, W.E. (1989b) "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison of the protein with mouse mast cell carboxypeptidase A and rat pancreatic carboxypeptidases" *Proc. Natl. Acad. Sci. USA* 86, 9480-9484.

Rost, B., Sander, C. and Schneider, R. (1994) "PHD: an automatic mail server for protein secondary structure prediction" *Comput. Appl. Biosci*. 10, 53-60.

Roussel, A. and Cambilleau, C. (1989) in Silicon Graphics geometry partners directory pp. 77-79, Silicon Graphics, Mountain View, CA.

Sali, A. and Blundell, T.L. (1993) "Comparative protein modelling by satisfaction of spatial restraints" *J. Mol. Biol*. 234, 779-815.

Sambrook, J. Fritsch, E.F. and Maniatis, T. (1989) "Molecular cloning: a laboratory manual" 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

San Segundo, B., Martínez, M.C., Vilanova, M., Cuchillo, C.M. and Avilés, F.X. (1982) "The severed activation segment of porcine pancreatic procarboxypeptidase A is a powerful inhibitor of the active enzyme" *BBA* 707, 74-80.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.

Schägger, H. and von Jagow, G. (1987) "Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa" *Anal. Biochem.* 166, 368-379.

Schatteman, K.A., Goossens, F.J., Scharpe, S.S., Neels, H.M. and Hendriks, D.F. (1999) "Assay of procarboxypeptidase U, a novel determinant of the fibrinolytic cascade, in human plasma" *Clin. Chem*. 45, 807-813.

Schatteman, K.A., Goossens, F.J., Leurs, J., Kasahara, Y., Scharpe, S.S. and Hendriks, D.F. (2001) "Fast homogeneous assay for plasma procarboxypeptidase U" *Clin. Chem. Lab. Med.* 39, 806-810.

Schmid, M.F. and Herriot, J.R. (1976) "Structure of carboxypeptidase B at 2.8 Å resolution" *J. Mol. Biol.* 103, 175-190.

Schneider, M., Boffa, M., Stewart, R., Rahman, M., Koschinsky, M. and Nesheim, M. (2002) "Two naturally occurring variants of TAFI (Thr-325 and Ile-325) differ substantially with respect to thermal stability and antifibrinolytic activity of the enzyme" *J. Biol. Chem.* 277,1021-1030.

Seemuller, E., Lupas, A., Stock, D., Lowe, J., Huber, R. and Baumeister, W. (1995) "Proteasome from *Thermoplasma acidophilum*: a threonine protease" *Science* 268, 579-582.

Seidah, N.G. and Chretien, M. (1998). *In: Handbook of Proteolytic Enzymes,* Barrett, A.J., Rawlings, N.D., and Woessner, J.F.(ed.), pp.345-357, Academic Press, San Diego.

Serra, M.A. (1992) Tesi Doctoral UAB "Estudi de la cinètica d´hidròlisi de pèptids per la carboxipeptidasa A pancreàtica".

Settle, S.H.J., Green, M.M. and Burtis, K.C. (1995) "The silver gene of *Drosophila melanogaster* encodes multiple carboxypeptidases similar to mammalian prohormone-processing enzymes" *Proc. Natl. Acad. Sci. USA* 92, 9470-9474.

Shapiro, A.L., Viñuda, E. and Maizel, J.V. (1967) "Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels" *Biochem. Biophys. Res. Commun.* 28, 815-820.

Sheldrick, G. M. and Schneider, T. R. (1997) in *Methods in Enzymology* (Carter, C. W., and Sweet, R. M., eds) Vol. 277, pp. 319-343, Academic Press, San Diego.

Silveira, A., Schatteman, K., Goossens, F., Moor, E., Scharpe, S., Stromqvist, M., Hendriks, D. and Hamsten, A. (2000) "Plasma procarboxypeptidase U in men with symptomatic coronary artery disease" *Thromb. Haemost*. 84, 364-368.

Sippl, M.J. (1993) "Recognition of errors in three-dimensional structures of proteins" *Proteins* 17, 355-362.

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

Skidgel, R.A., Davis, R.M. and Tau, F. (1989) "Human carboxypeptidase M. Purification and characterization of a membrane-bound carboxypeptidase that cleaves peptide hormones" *J. Biol. Chem.* 264, 2236-2241.

Skidgel, R.A., Tan, F., Deddish, P.A. and Li, X.Y. (1991) "Structure, function and membrane anchoring of carboxypeptidase M" *Biomed. Biochem. Acta* 50, 815-820.

Smith, G.K., Banks, S., Blumenkopf, T.A., Cory, M., Humphreys, J., Laethem, R.M., Miller, J., Moxham, C.P., Mullin, R., Ray, P.H., Walton, L.M. and Wolfe, L.A. (1997) "Toward antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1 and novel *in vivo* stable prodrugs of methotrexate" *J. Biol. Chem.* 272,15804-15816.

Smulevitch, S.V., Osterman, A.L., Galperina, O.V., Matz, M.V., Zagnitko, O.P., Kadyrov, R.M., Tsaplina, I.A., Grishin, N.V., Chestukhina, G.G. and Stepanov, V.M. (1991) "Molecular cloning and primary structure of *Thermoactinomyces vulgaris* carboxypeptidase T. A metalloenzyme endowed with dual substrate specificity" *FEBS Lett.* 291, 75-78.

Song, L. and Fricker, L.D. (1995a) "Calcium- and pH-dependent aggregation of carboxypeptidase E" *J. Biol. Chem.* 270, 25007-25013.

Song, L. and Fricker, L.D. (1995b) "Processing of procarboxypeptidase E into carboxypeptidase E occurs in secretory vesicles" *J. Neurochem.* 65, 444-453.

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

Sorensen, P., Winther, J.R., Kaarsholm, N.C. and Poulsen, F.M. (1993) "The pro region required for folding of carboxypeptidase Y is a partially folded domain with little regular structural core" *Biochemistry* 32, 12160-12166.

Stone, T.E., Li, J.P. and Bernasconi, P. (1994) "Purification and characterization of the *Manduca sexta* neuropeptide processing enzyme carboxypeptidase E" *Arch. Insect. Biochem. Physiol.* 27, 193-203.

Tan, F., Rehli, M., Krause, S.W. and Skidgel, R.A. (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., Polyakov, K., Obmolova, G., Strokopytov, B., Kuranova, I., Osterman, A., Grishin, N., Smilevitch, S., Zagnitko, O. and Galperina, O. (1992) "Crystal structure of carboxypeptidase T from *Thermoactinomyces vulgaris*" *Eur. J. Biochem*. 208, 281-288.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice" *Nucleic Acids Res.* 22, 4673-4680.

Towbin, H., Staehelin, T. and Gordon, J. (1979) "Electrophoretic transfer of proteins from acrylamide gels to nitrocellulose sheets: procedure and some applications" *Proc. Natl. Acad. Sci*. 76, 4350-4354.

Tregouet, D.A., Aubert, H., Henry, M., Morange, P., Visvikis, S., Juhan-Vague, I. and Tiret, L. (2001) "Combined segregation-linkage analysis of plasma thrombin activatable fibrinolysis inhibitor (TAFI) antigen levels with TAFI gene polymorphisms" *Hum. Genet.* 109, 191-7.

van Gunsteren, W. F. and Berendsen, H. J. C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual.*, BIOMOS B. V., Groningen, The Netherlands.

van Tilburg, N.H., Rosendaal, F.R. and Bertina, R.M. (2000) "Thrombin activatable fibrinolysis inhibitor and the risk for deep vein thrombosis" *Blood* 95, 2855-2859.

van Thiel, D.H., George, M. and Fareed, J. (2001) "Low levels of thrombin activatable fibrinolysis inhibitor (TAFI) in patients with chronic liver disease" *Thromb. Haemost.* 85, 667-670.

Vendrell, J., Avilés, F.X., San Segundo, B. and Cuchillo, C.M. (1982) "Isolation and reassociation of the procarboxypeptidase A/proproteinase E binary complex from pig pancreas" *Biochem. J.* 205, 449-452.

Vendrell, L. and Avilés, F.X. (1986) "Complete amino acid analysis of proteins by dabsyl derivatization and reversed-phase liquid chromatography" *J. Chrom.* 358, 401-413.

Vendrell, J., Wider, G., Avilés, F.X. and Wütrich, K. (1990a) "Sequence specific $^{1}$H NMR assignements and determination of the secondary structure for the activation domain isolated from pancreatic procarboxypeptidase B" *Biochemistry* 29, 7515-7522.

Vendrell, J., Avilés, F.X., Vilanova, M., Turner, C.H. and Crane-Robinson, C. (1990b) "$^{1}$H NMR studies of the isolated activation segment from pig carboxypeptidase A" *Biochem. J.* 267, 213-220.

Vendrell, J., Cuchillo, C.M. and Avilés, F.X. (1990c) "The tryptic activation pathway of monomeric procarboxypeptidase A" *J. Biol. Chem.* 265, 6949-6953.

Vendrell, J., Billeter, M., Wider, G., Avilés, F.X. and Wütrich, K. (1991) "The NMR Structure of the activation domain isolated from porcine procarboxypeptidase B" *EMBO J.* 10, 11-15.

Ventura, S., Gomis-Rüth, F.X., Puigserver, A., Avilés, F.X. and Vendrell, J. (1997) "Pancreatic procarboxypeptidases: oligomeric structures and activation processes revisited" *Biol. Chem.* 378, 161-165.

Ventura, S., Villegas, V., Sterner, J., Larson, J., Vendrell, J., Hershberger, C.L. and Avilés, F.X. (1999) "Mapping the pro-region of carboxypeptidase B by protein engineering" *J. Biol. Chem.* 274, 19925-19933.

Vilanova, M., Vendrell, J., López, M.T., Cuchillo, C.M. and Avilés, F.X. (1985) "Preparative isolation of the two forms of pig pancreatic procarboxypeptidase A and their monomeric carboxypeptidase A" *Biochem. J.* 229, 605-609.

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

Villegas, V., Azuaga, A., Catasus, L., Reverter, D., Mateo, P.L., Aviles, F.X. and Serrano, L. (1995b) "Evidence for a two-state transition in the folding process of the activation domain of human procarboxypeptidase A2" *Biochemistry* 34, 15105-15110.

Wang, W., Hendriks, D.F. and Scharpe, S.S. (1994) "Carboxypeptidase U, a plasma carboxypeptidase with high affinity for plasminogen" *J. Biol. Chem.* 269, 15937-15944.

Wang, W., Boffa, M.B., Bajzar, L., Walker, J.B. and Nesheim, M.E. (1998) "A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor" *J. Biol. Chem.* 273, 27176-27181.

Wei, S., Segura, S., Vendrell, J., Aviles, F.X., Lanoue, E., Day, R., Feng, Y. and Fricker, L.D. (2002) "Identification and characterization of three members of the human metallocarboxypeptidase gene family" *J. Biol. Chem.* 277, 14954-14964.

Wolfe, L.A., Mullin, R.J., Laethem, R., Blumenkopf, T.A., Cory, M., Miller, J.F., Keith, B.R., Humphreys, J. and Smith, G.K. (1999) "Antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1: in vitro and in vivo studies with prodrugs of methotrexate and the thymidylate synthase inhibitors GW1031 and GW1843" *Bioconjug. Chem.* 10, 38-48.

Wright, J.E. and Rosowsky, A. (2002) "Synthesis and enzymatic activation of N-[N(alpha)-(4-amino-4-deoxypteroyl)-N(delta)-hemiphthaloyl-L-ornithiny]-L-phenylalanine, a candidate for antibody-directed enzyme prodrug therapy (ADEPT)" *Bioorg. Med. Chem.* 10, 493-500.

Wun, T.C., Ossowski, L. and Reich, E. (1982) "A proenzyme form of human urokinase" *J. Biol. Chem.* 257, 7262-7268.

Xin, X., Day, R., Dong, W., Lei, Y., and Fricker, L.D. (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.

Yamamoto, K.K., Pousette, A., Chow, P., Wilson, H., el Shami, S. and French, C.K. (1992) "Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Identification of pancreas-specific protein as pancreatic procarboxypeptidase B" *J. Biol. Chem.* 267, 2575-2581.

Zhao, L., Morser, J., Bajzar, L., Nesheim, M. and Nagashima, M. (1998) "Identification and characterization of two thrombin-activatable fibrinolysis inhibitor isoforms" *Thromb. Haemost.* 80, 949-955.

Zhou, A., Webb, G., Zhu, X. and Steiner, D.F. (1999) "Proteolytic processing in the secretory pathway" *J. Biol. Chem.* 274, 20745-20748.

Zsebo, K., Lu, H., Fieschko, J.C., Goldstein, L., Davis, J., Duker, K., Suggs, S.V., Lai, P. and Bitter, G.A. (1986) "Protein secretion from *Saccharomyces cerevisiae* directed by the prepro-alpha-factor leader region" *J. Biol. Chem.* 261, 5858-5865.