V- Treball 2. The prosequence of potato carboxypeptidase inhibitor does not influence the mature peptide folding

V.A Introduction

All proteins contain within their amino acid sequence the required information for their folding. However, other factors may be required for a fast and efficient folding *in vivo*. Molecular chaperones facilitate folding by decreasing the tendency of partially folded proteins to go into non productive pathways (Thomas, 1997). The protein disulfide isomerase and the peptidyl-prolyl-isomerase can function as folding catalysts as well (Ferrari, 1999 and Gothel, 1999). Apart from these components that have a broad substrate specificity, the folding process may be also facilitated specifically by the precursor protein. Many proteins are synthesized *in vivo* as precursors in the form of pre-pro-proteins. Pre or signal peptides are often involved in export while propeptides can regulate many processes (Martoglio, 1998). Depending on their function they can be classified in two groups (Shinde, 2000): The class I propeptides, which are required for the correct folding of the proteins to which they are attached (Van den Hazel et al, 1993; Ikemura et al, 1987; Rattenholl et al, 2000) and the class II propeptides, which influence other cellular processes, such as secretion, protein activity or molecular assembly (Baardsnes, 1998; Van Wart, 1990; Alvares, 1999).

Class I propeptides have also been termed as intramolecular chaperones, and its role in folding has been demonstrated *in vitro* and *in vivo* (Ikemura, 1987; Winther, 1991; Lee, 1991 and Zhu, 1989). There are not many examples of the role of the pro-region in small disulfide-rich proteins. In these proteins the folding process differs from that of larger proteins in that is mainly driven by the formation of the disulfide bridges. Among the most studied proteins of this group we find the bovine pancreatic trypsin inhibitor (BPTI), that is generated as a pre-pro-protein, which contains a central 58 residue mature protein with 3 disulfide bonds, a 15-residue N-terminal pro-region and a 7-residue C-terminal pro-region. The N-terminal pro-region contains a cysteine residue which appears to increase both the yield of properly folded mature BPTI and the rate of the folding process *in vitro* (Weissmann, 1992), providing an intramolecular thiol-disulfide reagent. Nevertheless, it did not appear to have any positive effect under more physiological conditions (Creighton et al, 1993). Other experimental evidences in favour or against the prominent role of pro-regions in small disulfide-rich proteins have been reported. The pro-region of the guanylyl cyclase activating peptide (GCAPII) contributes significantly to the correct disulfide-coupled folding of the mature protein and the dimerization of the molecule (Hidaka, 2000). In contrast, the studies performed with ω-conotoxin-MVIIA suggest that this protein does not rely on the propeptide region of its precursor to facilitate folding (Price-Carter, 1996). This molecule is synthesized as a pre-pro-protein containing a 24-residue N terminal pro-region, a 1-residue C terminal pro-region and a 26-residue mature protein with 3 disulfide bonds. Further studies performed with other conotoxins demonstrated that the mature forms of these molecules contain sufficient information to direct their folding and correct disulfide pairing *in vitro* (Price-Carter,1996).

In all cases, the *in vitro* folding of the mature disulfide-rich proteins is characterized by a low efficiency and a slow kinetics. This fact suggest that these proteins need other factors, apart from their mature amino acidic sequence, in order to fold efficiently and rapidly into the native form i*n vivo*. The nature of these factors, whether those are found in its pro-region or in other cellular components (protein disulfide isomerases, molecular chaperones...), seems to depend on each protein. This class of proteins present many biotechnological applications as hormones, growth factors, venomes... therefore, it seems clear that further research is required to fully understand their folding process.

Potato carboxypeptidase inhibitor (PCI) is a 39 residue globular protein that inhibits several metallocarboxypeptidases (Hass, 1982). It has a 27-residue central core with 3 disulfide bridges that forms a T-Knot scaffold, also found in other proteins such as many growth factors (Lin and Nussinov). This molecule is synthesized as a pre-pro-protein, that contains a 27-residue N terminal pro-region of unknown function, a 7-residue C terminal pro-region, probably involved in transport to the vacuole and the 39-residue mature protein (Villanueva, 1998). The folding and unfolding pathways of mature PCI have been previously studied and are well characterized (Chang, 1994; Chang, 2000). The extremely unefficient folding of PCI *in vitro* (Chang, 1994), together with the presence of a relatively long pro-sequence, as deduced from its c-DNA sequence (Villanueva, 1998), suggested a possible involvement of the pro-region in the *in vivo* folding of PCI.

In order to determine whether the prosequence of PCI has an effect on its rate of disulfide formation or its folding efficiency, refolding studies *in vitro* of the mature protein with the N terminal extension (ProNtPCI) and with the N and C terminal extensions (ProPCI) were performed by RP-HPLC analysis of the acid-trapped disulfide intermediates collected during the refolding processes. Since some proteins can achieve its prosequence-dependent folding intermolecularly, the experiments were also performed with the mature PCI in the presence of the N terminal prosequence (ProNt) in *trans* (Marie-Claire, 1999 and Zhu, 1989). In addition, the folding *in vivo* in *Escherichia coli* of wt PCI and some of their site-directed mutants containing the N-terminal pro-peptide were studied and compared to their forms without the pro-peptide to evaluate any influence of the pro-region. From both, *in vitro* and *in vivo* studies, we can conclude that the propeptide region does not influence the folding of mature PCI.

V.B Experimental procedures

V.B.1 Plasmid constructions and mutagenesis

A plasmid containing a synthetic gene encoding wtPCI (Molina et al 1992) cloned into pINIII-OmpA3 vector (Ghrayeb et al 1984), was used as a template to generate the plasmidic constructions used for the expression of the PCI forms studied. ProNtPCI was obtained by means of one-step PCR. The PCR fragment was subcloned into pGEM-T Vector System (Promega),

restricted with XbaI and EcoRI and ligated into pIN-III-OmpA3 vector. Similarly, this construction was used as a template to generate the Δ3, Y37G and G35P/P36G ProNtPCI mutant genes, by means of one-step PCR. Constructs for Δ3, Y37G and G35P/P36G PCI mutant genes were achieved by PCR of wt-mature PCI (Venhudová et al, 2000).

ProNtPCI was also cloned into pBAT4 expression vector, derived from the pET plasmids (Studier et al 1990), with and without the leader sequence OmpA. ProPCI was generated from ProNtPCI by means of one-step PCR and cloned into pBAT4 vector without the leader sequence OmpA. All constructions cloned into pINIII vector were transformed into *Escherichia coli* strain MC1061 and those cloned into pBAT4 vector were transformed into *Escherichia coli* strain BL21(DE3).

V.B.2 Protein expression and purification

For production of PCI and ProNtPCI mutant genes, wtPCI and wtProNtPCI cloned into pINIIIOmpA3 vector, *Escherichia coli* MC1061 cells were grown at 37ºC and expression was induced at 0.1 UDO_{550nm} by the addition of 1 mM isopropyl-β-D-thiogalactoside (IPTG), and they were harvested by centrifugation 20 hours after protein expression induction. ProNtPCI was purified from the culture medium by a Sep Pak C_{18} (Waters) cartridge and eluted with 70% isopropanol containing 0.1% trifluoroacetic acid. Isopropanol was evaporated and the protein was finally purified by HPLC, on a Protein C4 0.46x25 cm, 5 μm column (Vydac). The conditions used were: solvent A was water containing 0.1% trifluoroacetic acid, solvent B was acetonitrile containing 0.1% trifluoroacetic acid and the gradient was 25%-55%B in 60 minutes. Details regarding the PCI purification protocol have been published elsewhere (Molina et al, 1992). ProNtPCI mutant proteins were only used in the *in vivo* refolding experiments and the supernatant was directly analyzed by HPLC on a Nova-Pak C8 3.9x150 mm column (Waters) after sample acidification with trifluoroacetic acid and filtration through 4 mm, 0.2 μm PVDF filters (National Scientific).

For protein production in *Escherichia coli* BL21(DE3) cells, the cultures were grown until they reached a value of 1 UDO_{550nm}, induced by addition of 0.2 mM IPTG, and cells were harvested by centrifugation 2.5 hours after protein expression induction. The cell pellet from a 1 L culture was resuspended in 50 mL 20mM Tris.HCl, 0.5 mM EDTA (pH 8.5) and was maintained on ice for 10 minutes. The solution was sonicated for 10 minutes on ice at 50Hz at half power, on a Labsonic-Braun sonicator and centrifuged at 22000 g for 25 minutes. The pellet was resuspended in 50 mL 20 mM Tris.HCl, 0.5 mM EDTA and 2% Triton x-100 (pH 8.5) and centrifuged at 22000g for 25 minutes. The pellet was resuspended in 10 mL 6 M GdnCl and 30 mM dithiothreitol (pH 8.5). After 6 hours the sample was centrifuged at 3000 g for 10 minutes and the supernatant was dialyzed against 0.1 M Tris HCl (pH 8.5) for 12 hours and then renaturation was performed by dialysis in the presence of a redox system containing 4 mM Cys and 2 mM Cys-Cys at pH 8.5 for 48 hours at 4ºC with a 3500 Da cut-off membrane (Spectrum medical industries Inc). After

dialysis, the sample was centrifuged at 3000 g for 10 minutes and the supernatant was purified by HPLC, on a Protein C4 0.46x25 cm, 5 µm column (Vydac), using a linear gradient 20%-30% in 5 minutes and 30%-50% solvent B in 20 minutes for ProPCI and a linear gradient 25-55% B solvent in 60 minutes for ProNtPCI. The peptide corresponding to the N-terminal pro-sequence (ProNt) was obtained by solid-phase chemical synthesis. The released peptide was purified by HPLC on a Protein C4 1x25 cm, 5µm column (Vydac), in a linear gradient 20-27% in 7 minutes and 27-40% solvent B in 26 minutes.

V.B.3 *In vitro* folding experiments

100 µg lyophilized aliquots of PCI, 185 µg lyophilized aliquots of ProPCI and 171 µg lyophilized aliquots of ProNtPCI were used in each folding experiment. The proteins were dissolved in 0.5 mL Tris.HCl (0.5 M, pH 8.5) containing 5M GdnCl and 30 mM dithiothreitol, to a final protein concentration of 46.5 µM. After 2 hours at 25ºC, the reduced and denatured proteins were passed through a PD-10 (Pharmacia) column equilibrated with 0.1 M Tris.HCl buffer (pH 8.5). The proteins were eluted in 1.2 mL and split in 3 parts which were diluted to a final protein concentration of 14.5 µM, with the 0.1 M Tris.HCl buffer (pH 8.5), the same buffer containing 1 mM Cys and the same buffer containing 4 mM Cys and 2 mM Cys-Cys, respectively. In the experiments where the N terminal pro-sequence was tested in *trans*, the peptide was added to the denatured and reduced PCI in the dilution buffer, to a final concentration of 14.5 µM. Samples of all reaction mixtures were collected in a time-course manner for up to 24 hours and trapped by mixing with an equal volum of (a) 1% trifluoroacetic acid in water (reversible trapping) and were analyzed by HPLC on a Protein C4 0.46x25 cm, 5µm column (Vydac). The gradient was linear, 20%-40% solvent B in 30 minutes for PCI, 25%-35% solvent B in 10 minutes and 35-45% solvent B in 40 minutes for ProNtPCI and 20%-30% solvent B in 5 minutes and 30%-50% solvent B in 30 minutes for ProPCI. (b) 0.1 M iodoacetic acid in Tris.HCl buffer (0.5 M, pH 6.5) containing 40% (by volume) of dimethylformamide (irreversible trapping) (Chatrenet et al, 1993). Carboxymethylation was allowed to proceed for 30 minutes at 25ºC.

V.B.4 Inhibitory activity

The substrate used to perform the carboxypeptidase activity was furyl-acryloyl-L-phenylalanyl-L-phenylalanine (FAPP) 0.2mM and the buffer was 50 mM Tris.HCl, 0.5 M NaCl, pH7.5. To 985µl of substrate, 5 µl of carboxypeptidase A (CPA) (Sigma) at 0.02 mg/ml were added and the absorbance change at 330 nm was measured during 2 minutes, then 10 µl of increasing concentrations of PCI or ProNtPCI were added and the measures were continued for 2 minutes. The slope of the first part of the assay corresponded to v_o and the slope of the second part to v_i. The residual enzymatic activity was calculated (v_o to v_i) and plotted as function of the inhibitor concentration.

V.B.5 Mass spectrometry

Molecular mass was determined by matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry on a BRUKER-BIFLEX spectrometer. Ionization was accomplished with a 337 nm pulsed nitrogen laser and spectra were acquired in the linear positive ion mode, using a 19 Kv acceleration voltage. Samples were prepared mixing equal volumes of the protein solution and a saturated solution of sinapic acid, used as a matrix , in aqueous 30% acetonitrile with 0.1% trifluoroacetic acid.

V.B.6 Circular Dichroism Spectroscopy

Circular dichroism spectra were collected on a Jasco-J715 spectropolarimeter at 25ºC, using a 2 mm path length cell, a band width of 2 nm, a step size of 0.5 nm and an averaging time of 1 second. Samples were prepared by dissolving lyophilized proteins to a final concentration of 100µg/mL in 0.1% trifluoroacetic acid (pH 2) or 50 mM Na_2HPO_4 (pH 7).

V.B.7 Deuterium to proton (D/H) exchange

15µg of lyophilized samples of PCI or ProNtPCI were ressuspended in 5 µL of D_2O and incubated for 3 hours at 50ºC in order to exchange completely all labile protons and afterwards were maintained for 30 minutes at room temperature to refold properly. The native deuterated proteins were diluted with 4 volumes of 15 mM glycine pH 3.0 in H_2O to start the hydrogen exchange and samples were taken in a time-course manner and analyzed by MALDI-TOF MS.

V.B.8 Exoproteolysis

15 µg lyophilized aliquots of ProNtPCI were dissolved in 10 µL 10 mM Tris.HCl buffer (pH 8.5) containing 5 µg of leucine aminopeptidase (Sigma). Samples were collected in a time-course manner, diluted with water containing 0.1% trifluoroacetic acid (1:2) and the proteolysized products present in the mixture were identified by MALDI-TOF mass spectrometry.

V.B.9 Nuclear Magnetic Resonance

NMR spectra were recorded on a -Bruker AMX- spectrometer operating at 400 MHz. 2 mg of PCI, 2 mg of the N terminal prosequence peptide and 100 µg of ProNtPCI were resuspended in 500 µL of NaH_2PO_4 pH 4.00 20 mM containing 10% D_2O and the spectra were acquired at 35ºC.

V.C Results

V.C.1 Expression in *Escherichia coli*

Potato carboxypeptidase inhibitor (PCI) is synthesized as a precursor form, which contains a 27-residue N terminal extension, a 7-residue C terminal extension and the mature 39-residue protein[2] (Fig. 1).

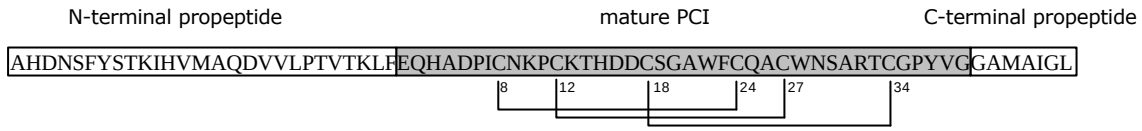

Figure 1. Primary structure of PCI precursor form. The aminoacids are indicated in one-letter code and the disulfide pairing of the native PCI is displayed.

In order to study the *in vitro* refolding behaviour of the precursor form of PCI containing the mature peptide and the N terminal pro-sequence (ProNtPCI), two strategies for the production of this molecule were followed. Expression of ProNtPCI in *Escherichia coli* MC1061 using a pINIII-OmpA3-derived secretion vector led to a low yield of purified protein due to the proteolysis of the prosequence during the protein expression period. The initial ProNtPCI protein was largely degraded to PCI that accumulated in the culture medium. The final yield of ProNtPCI was very low (50 μg/L) and the major protein found in the cell culture was the mature protein (PCI) (Fig. 2a). This form of ProNtPCI was used exclusively for the experiments requiring small amounts of protein. Another expression system was developed in order to produce ProNtPCI in amounts large enough to perform refolding experiments. ProNtPCI was cloned into pBAT4 vector with and without the signal peptide OmpA. The extracellular expression of the molecule resembled that of ProNtPCI in MC1061 cells using the pINIII-OmpA3 vector. The intracellular expression led to the formation of inclusion bodies, as expected, since PCI contains three disulfide bonds that can not be efficiently formed inside the host cells, leading to the accumulation of protein aggregates. The inclusion bodies were reduced with dithiothreitol and denatured with GdnCl. Later, renaturation was performed by dialysis in the presence of a redox system containing Cys and Cys-Cys.

---

[2] PCI pro-forms used in this study are:

PCI: mature PCI

ProNtPCI: PCI with the N terminal pro-sequence

ProCtPCI: PCI with the C terminal pro-sequence

ProPCI: PCI with the N and C terminal pro-sequences

ProNt + PCI: PCI plus the N terminal pro-sequence *in trans*

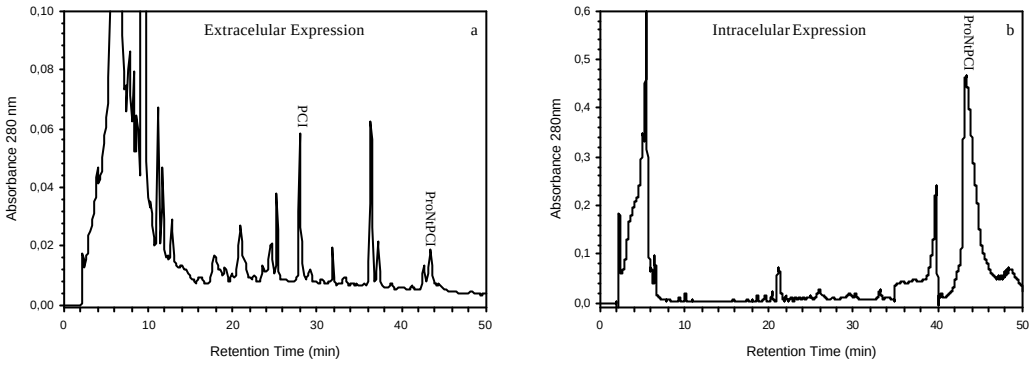

Figure 2. RP-HPLC chromatographic expression pattern of ProNtPCI on a Vydac C$_4$ column. Sample amounts equivalent to 2 mL of cell culture were analyzed in each chromatogram. a Sample corresponding to an extracelular expression achieved with the pINIII-OmpA3 expression vector. Proteins present in the cell culture medium were purified by a Sep pak C$_{18}$ cartridge and then subjected to RP-HPLC. b Sample corresponding to an intracelular expression achieved with the pBAT4 expression vector. The inclusion bodies were purified, denatured and dialized against 0.1M Tris.HCl pH 8.5 and then RP-HPLC was performed (See Experimental Procedures).

The final yield of ProNtPCI produced with this latter system was 3.4 mg/L (Fig.2b). The same expression protocol was used for the intracellular expression of ProCtPCI using a pBAT derived vector, giving a final yield of 3.0 mg/L of protein.

V.D.2 Refolding *in vitro*

PCI refolding undergoes a two-stage process, a first stage of fast unspecific disulfide formation is followed by a second stage (rate limiting step) of disulfide reshuffling which leads to the native form (Chang et al, 1994). This refolding pattern was apparently reproduced when different variants of recombinant PCI pro-forms were assayed.

The yield of fully refolded (native) forms achieved after 7 hours of refolding, in absence of an external thiol was small (<5%) and similar in all the molecules tested (Figure 3a, left). The RP-HPLC chromatogram profiles from the 7 hours refolding mixture of PCI and PCI plus the ProNt extension in *trans* were indistinguishable. Thus, we can assume that the molar ratio among the scrambled and native species is not affected by the addition of the ProNt in *trans*. The RP-HPLC chromatographic profiles of ProNtPCI and ProPCI were different from those of PCI, since they have a different RP-HPLC mobility. It should be mentioned that the scrambled species of these molecules were not further characterized, so we can not assure that the amount of each scrambled specie present in the refolding mixture of ProNtPCI and ProPCI are equivalent to those observed for PCI refolding. Nevertheless, the ratio between the native form and the ensemble of scrambled species remains constant for all the molecules. Neither the N terminal nor the C terminal PCI pro-sequences have effects on the final yield of fully refolded (native) PCI, indicating that the overall folding process is similar among all the molecules tested under these conditions.

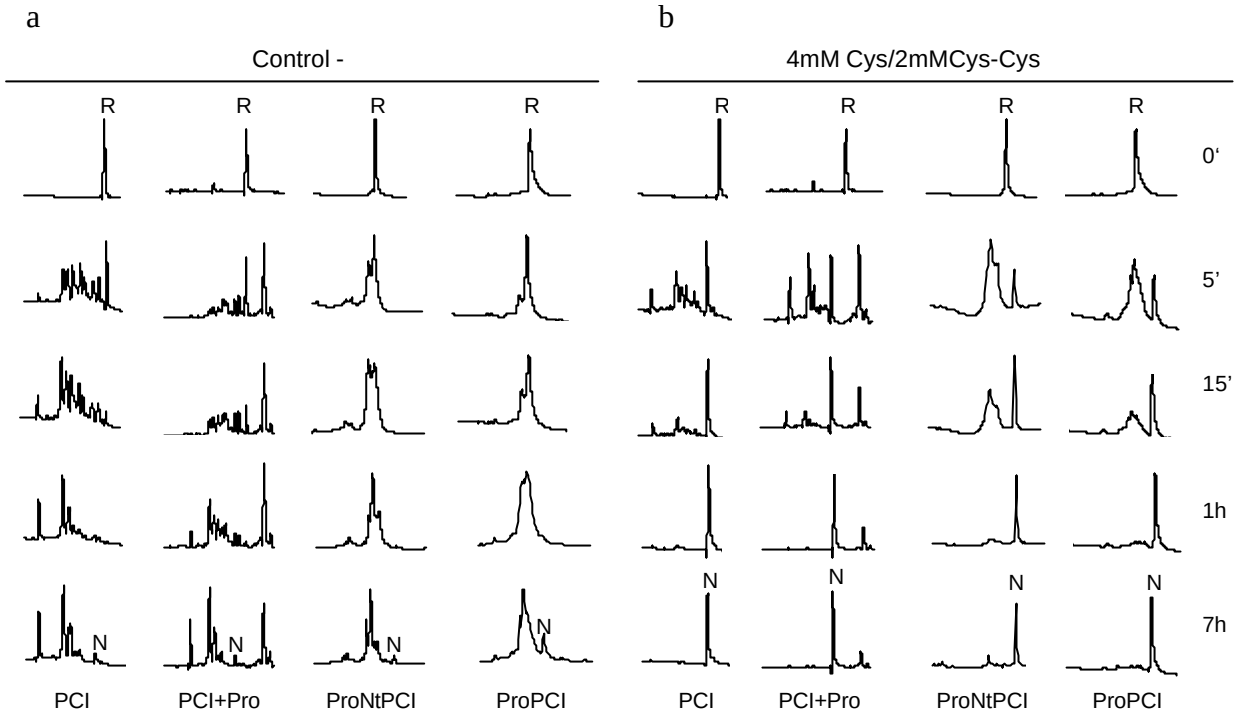

Figure 3. *In vitro* folding of PCI, PCI plus the ProNt *in trans*, ProNtPCI and ProPCI in the presence of selected redox potentials. Reduced/Denatured proteins were refolded in the absence (no added thiols) or presence (4mMCys/2mMCys-Cys) of exthernal thiols. Folding intermediates were acid-trapped and analyzed by RP-HPLC. Elution positions of native (N) and reduced (R) forms are indicated

The folding of PCI is accelerated by the presence of external thiols in the folding mixture. While the addition of Cys accelerates the second stage of disulfide reshuffling of scrambled forms to native PCI, the addition of Cys-Cys enhances the first stage of disulfide formation, which leads to the formation of scrambled species. An optimized combination of 4 mM Cys and 2 mM Cys-Cys enables mature PCI to refold quantitatively within 3 hours (Chang, 1994). A similar overall view was found when these conditions were applied to the different pro-forms, and the refolding processes were followed by RP-HPLC. The yield of native forms achieved in the presence of an optimized combination of Cys and Cys-Cys is superior to 70% after 1 hour of refolding (Fig. 3b). The folding kinetics and efficiency of PCI and PCI plus ProNt peptide in *trans* are similar, but for ProNtPCI and ProPCI the folding kinetics are a little slower (Fig. 3b and 4). Nevertheless, they have the same folding efficiency, since the final yield of native form at 24 hours of refolding are nearly identical (Fig. 4).

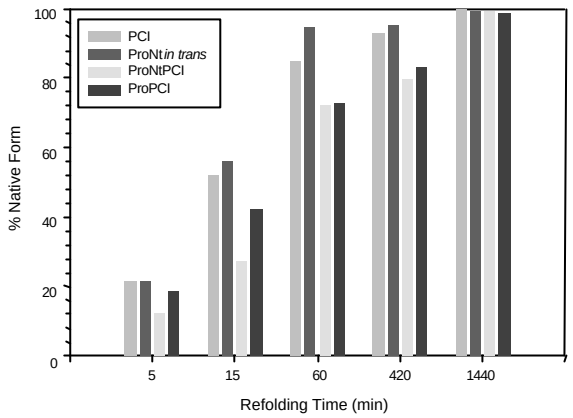

Figure 4. Folding efficiency of PCI, PCI plus the ProNt peptide *in trans*, ProNtPCI and ProPCI. Reduced/denatured proteins were refolded in the presence of 4 mM Cys/2mM Cys-Cys and acid-trapped folding intermediates were analyzed by RP-HPLC. The yield of native form was calculated in each time point from the peak areas in the corresponding RP-HPLC chromatograms.

The flow of intermediates of one, two and three disulfides was followed by MALDI-TOF mass spectrometry, through the analysis of the iodoacetate-trapped folding intermediates of all four sets of molecules tested. The flow of refolding intermediates is characterized by a progression from the reduced state through more thermodynamically stable species: 1-disulfide, 2-disulfide and 3-disulfide species (scrambled species). The rate of disulfide formation is similar in all the molecules tested under the same conditions.

V.D.3 Influence of prosequences *in vivo*

Some recombinant mature PCI variants with mutations at the C-tail have been reported to give a low expression yield or a low folding efficiency compared to wt mature PCI (C terminal Δ3, Y37G and G35P/P36G PCI) (Venhudova, 2000). DNA constructs encoding these proteins plus an added N terminal prosequence (Δ3, Y37G and G35P/P36G ProNtPCI) were designed to determine whether this region influenced their *in vivo* expression or folding in *E.coli*. Since the C terminal pro-region seems to be involved in transport to the vacuoles (Villanueva, 1998), the mutants were designed only with the N terminal prosequence. The expression of each ProNtPCI mutant proteins and wt ProNtPCI was performed in parallel to the corresponding PCI mutant proteins and wt PCI. 24 hours after induction of protein expression, the supernatant was collected and the species present in the supernatant were analyzed by RP-HPLC (Fig. 5) and identified by MALDI-TOF mass spectrometry. Since the N terminal pro-regions were largely degraded in all the species, the proteins found in the cultures were the mature PCI forms.

The amount of each expressed protein was calculated by comparison of the corresponding RP-HPLC peak areas (Data not shown). The final yield of the refolded forms for each ProNtPCI mutant, the ratio between the refolded form and the ensemble of scrambled species and the ratio

among the scrambled species present in the cell culture supernatants, were compared with those of PCI mutants lacking the N-terminal prosequence to evaluate any influence of the pro-region. Under the conditions of the experiment, the values obtained were similar between PCI and ProNtPCI mutant proteins, indicating that the pro-region of PCI does not affect neither the expression level nor the folding efficiency *in vivo* in *Escherichia coli*.

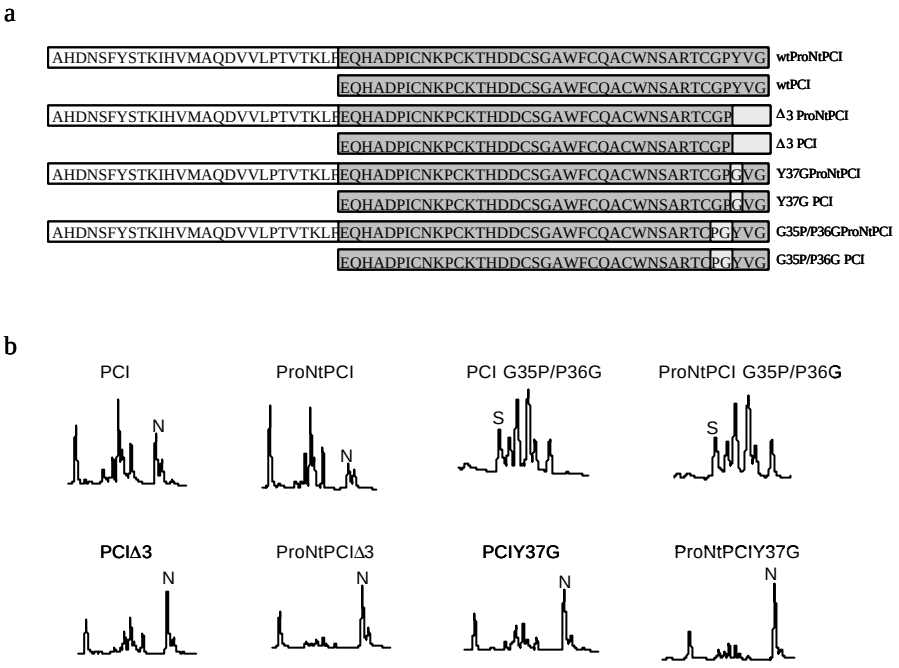

Figure 5. ***In vivo*** **expression of G35P/P36GPCI, D3PCI, Y37GPCI, wtPCI and G35P/P36GProNtPCI, D3ProNtPCI, Y37GProNtPCI and wtProNtPCI molecules. a. Schematic representation of the recombinant proteins used in this study. Aminoacids are in one-letter code. The N terminal prosequence is indicated with a white box, the mature protein with a dark shaded box and the mutated aminoacids with a light shaded box. b. The recombinant proteins were expressed in** ***Escherichia coli*** **MC1061cells in the expression vector pINIII-OmpA3. The supernatants after 24 hour induction were analized by RP-HPLC. The quantity of each native and scrambled form was calculated from the peak area of its corresponding RP-HPLC chromatogram. The elution position of each native form is indicated (N). In case of G35P/P36G mutants the disulfide pairing is not the same as wtPCI (Venhudovà et al,2000) and S stands for the most stable form.**

V.D.4 Inhibitory activity

To test whether ProNtPCI presents or not the same biological capabilities than mature PCI, inhibition of bovine carboxypeptidase1 (CPA1) by ProNtPCI and PCI was performed. Both proteins show very similar inhibitory activity towards CPA1 using the substrate FAPP and they have a similar IC_{50} value (about 100nM) (Fig.6). According to these results, the mature PCI region within ProNtPCI should keep its native three-dimensional structure and the proper disulfide pairing.

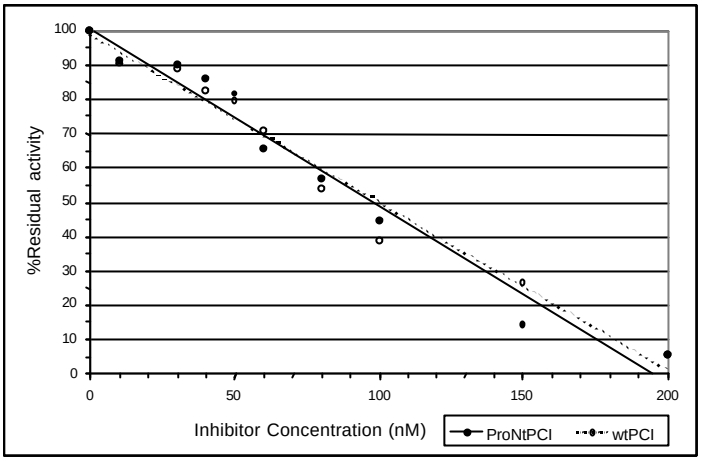

Figure 6. Inhibition of CPA by ProNtPCI and PCI. The inhibitory activity of each molecule was determined using the CPA1 substrate FAPP and testing different concentrations of the molecules (See *Experimental Procedures*).

V.D.5 CD spectroscopy

Despite the lack of regular secondary structures in wtPCI native form, far-UV CD spectroscopy may be helpful to indicate its folding state, since it shows a characteristic positive ellipticity band at 228 nm when it is properly folded. This band is not found in the mutant PCI proteins that lack the Tyr 37 residue, neither in wtPCI scrambled forms (Venhudova et al, 2000). These facts suggest that the characteristic maximum band at 228 nm is probably shown by a PCI molecule that contains the Y37 residue in a specific correctly folded environment.

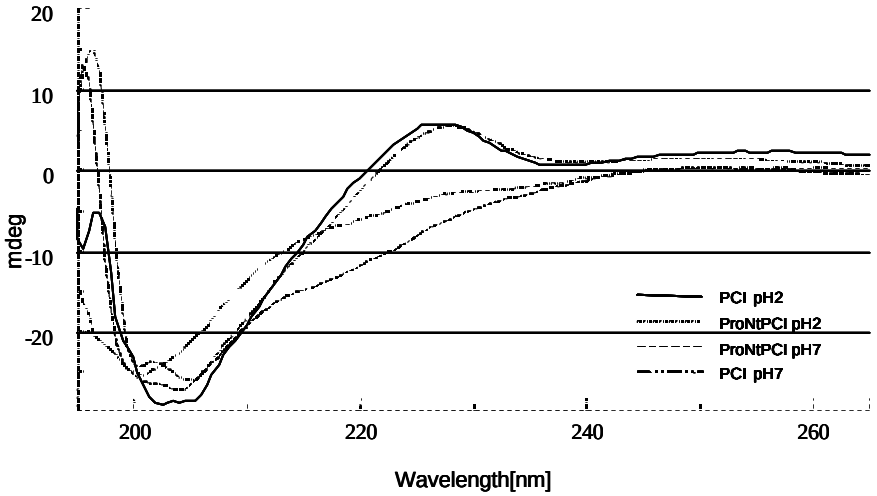

Figure 7. Circular dichroism analysis. CD analysis of native PCI and ProNtPCI was carried out in 20 mM phosphate buffer, pH 7.0 and 0.1% TFA, pH 2.0 and 100mg of protein were used in each spectrum.

When CD spectra of ProNtPCI were recorded, the maximum at 228 nm was observed neither at pH 7.0 nor at pH 2.0. Instead, the spectra presented two turnings at 228 nm and 214 nm, which could be observed more clearly at pH 7.0 (Fig. 7).

V.D.6 D/H Exchange

Recently, we demonstrated that hydrogen exchange monitored by MALDI-TOF mass spectrometry is a useful approach for the rapid characterization of protein conformation and requires minimal amounts of sample (Villanueva, 2000). The determination of the number of slow-exchanging hydrogens provides a parameter to discriminate between the stabilities of different protein variants. PCI contains a total of 65 labile hydrogens and it has been demonstrated by NMR that 6 of them form the slow exchange core (Clore, 1987). To test whether ProNtPCI has a higher number of slow-exchanging hydrogens than PCI or not, D/H exchange experiments were performed with both proteins. All labile protons were first exchanged to deuterons and then, the protein was diluted in a proton containing buffer. The kinetics of the D/H exchange was followed in a time-course manner by MALDI-TOF MS.

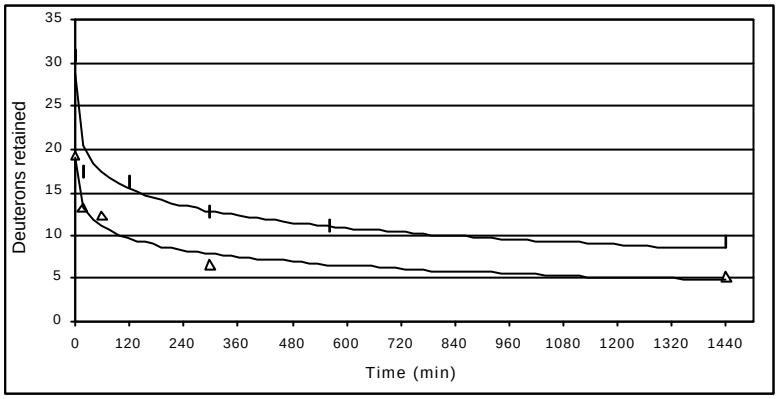

Figure 8. Kinetic plots of the D/H exchange monitored by MALDI-TOF MS. Decrease in deuteration of PCI (triangles) and ProNtPCI (diamonds) after dilution (1/5) of the deuterated samples in a proton buffer (15mM Glycine, pH 3.0).

The results indicate that the hydrogen exchange kinetics of both proteins follow a similar pattern (Fig. 8). For each protein a major subpopulation of protons exchange rapidly (within 2 hours) and the overall equilibrium is reached after 24 hours. However, the number of slow exchanging deuterons is significantly different. While PCI retains 5 deuterons (protected from exchange), ProNtPCI retains 9. This difference shows that ProNtPCI presents a different conformational state than PCI and/or presents a larger number of residues involved in compact structures.

V.D.7 Exoproteolysis

Leucine aminopeptidase is a metalloexoprotease widely used in peptide sequence analysis (Caprioli and Fan, 1986), but its use in whole mature proteins has not been so largely studied. When this enzyme is used in limited proteolysis experiments and the production of protein fragments is followed in a time-course manner by MALDI-TOF MS, an accumulation of certain stable protein fragments can be observed, which correspond to the beginning of the secondary structures present

in the protein. The identification of these peptides leads to an easy mapping of the secondary structures present in the protein (Villanueva et al, 2002) becoming a particularly useful tool in case of proteins which can not be obtained in high enough amounts. Due to the small yield of native ProNtPCI achieved in recombinant expression, this approach was chosen in order to identify possible secondary structure elements in the N terminal pro-sequence of PCI.

A lyophilized aliquot of ProNtPCI was dissolved in a Tris.HCl buffer containing leucine aminopeptidase (See Experimetnal Procedures) and samples collected in a time-course manner were analyzed by MALDI-TOF mass spectrometry in order to identify which protein fragments accumulated during the proteolysis (Fig 9a). Along the 27-residue N terminal pro-sequence of PCI, only two accumulated fragments (-10 and -15 in Fig 9b) could be identified. This is an indication of the presence of secondary structures in this region of the propeptide.

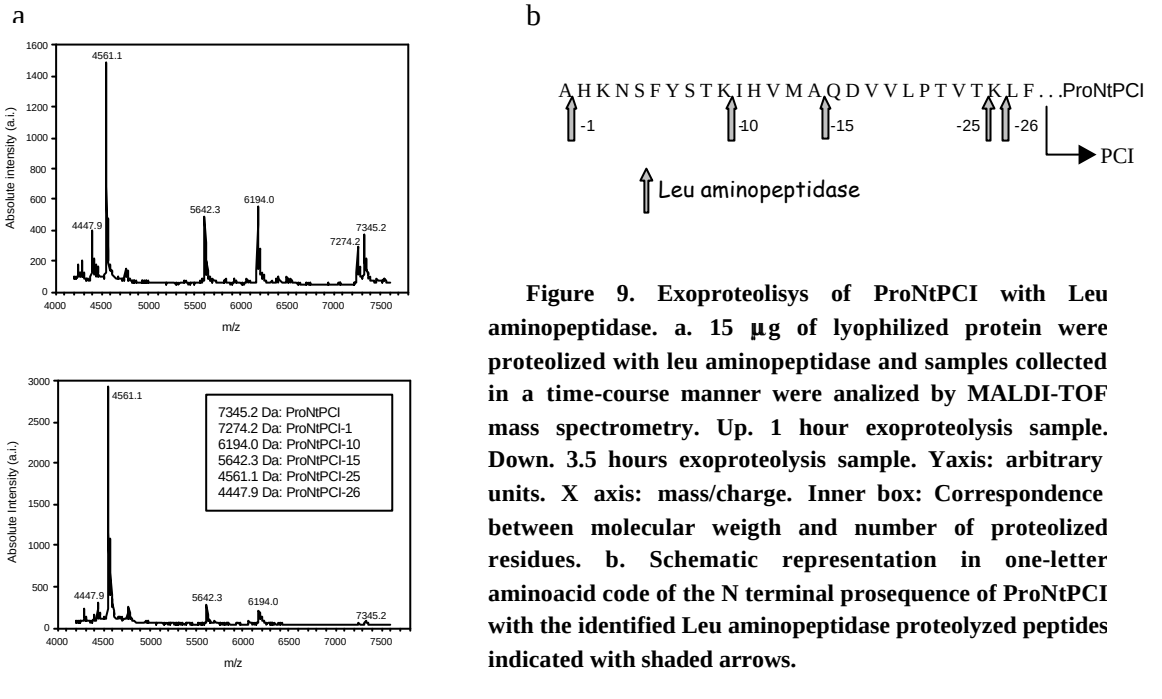

Figure 9. Exoproteolisys of ProNtPCI with Leu aminopeptidase. a. 15 µg of lyophilized protein were proteolized with leu aminopeptidase and samples collected in a time-course manner were analized by MALDI-TOF mass spectrometry. Up. 1 hour exoproteolysis sample. Down. 3.5 hours exoproteolysis sample. Yaxis: arbitrary units. X axis: mass/charge. Inner box: Correspondence between molecular weigth and number of proteolized residues. b. Schematic representation in one-letter aminoacid code of the N terminal prosequence of ProNtPCI with the identified Leu aminopeptidase proteolyzed peptides indicated with shaded arrows.

V.D.8 NMR analysis

The monodimensional proton-NMR spectra of PCI, the N terminal propeptide and ProNtPCI were recorded under the same conditions (pH 4.00 and 35ºC). In these conditions the ^{1}H-NMR spectra of the isolated N terminal pro-peptide (Fig 10b) does not show a large dispersion of resonances neither at low or at high fields. In the 0-1 ppm region there are not a significant number of potential shifted methyl protons, neither dispersion is observed in the NH region (9-12 ppm).

a

b

c

d

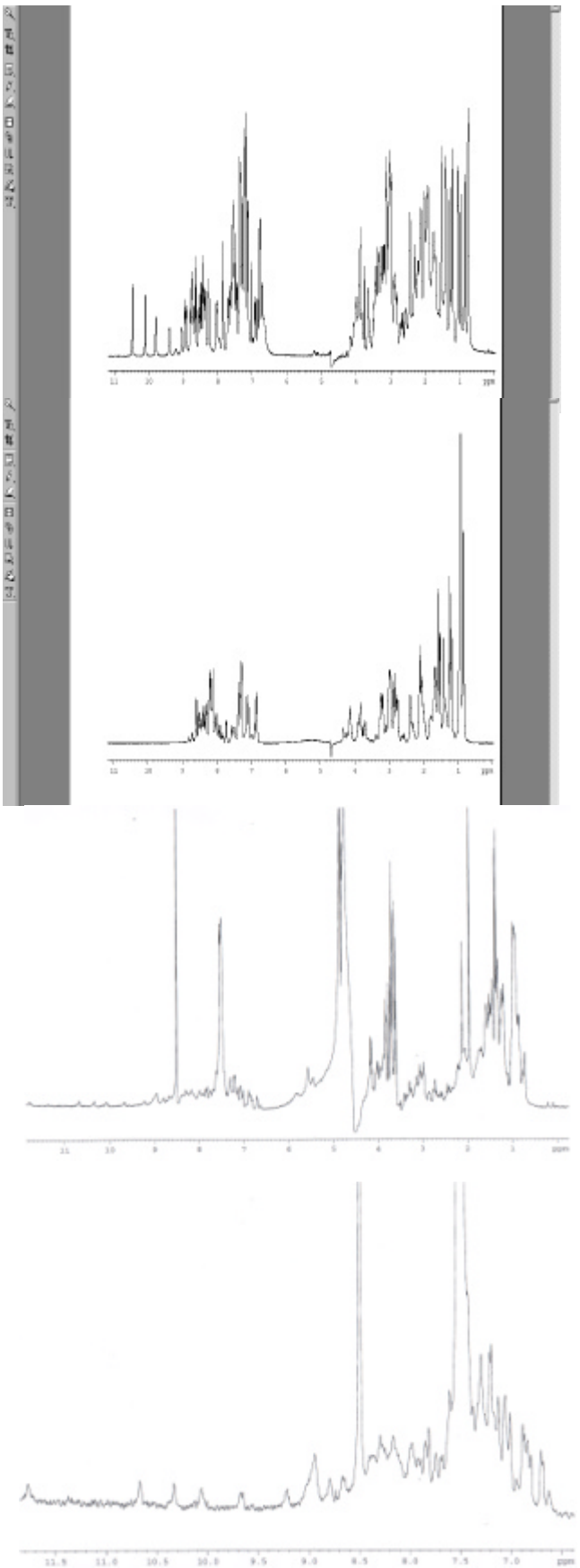

Figure 10. NMR spectres of PCI (a), the N terminal pro-sequence (Pro) (b), the ProNtPCI (c) and an ampliation of the amide region of ProNtPCI spectra (d). 2mg of PCI and N terminal pro-sequence, and 100 µg of ProNtPCI were ressuspended in 500 µl of Na_2HPO_4 20 mM containing 10% D20 and the spectra were recorded at 30ºC.

The TOCSY and NOESY two-dimensional spectra of this molecule were also recorded (data not shown). Comparison of both spectra indicates the lack of non-sequential interactions.

In contrast, the ^{1}H-NMR analysis of ProNtPCI (Fig 10c) shows that this molecule is richer in exchangeable resonances in the amide region (9-12 ppm) and presents a higher dispersion in the methyl region as well, although the signal intensity is lower due to the small quantity of protein available to perform the experiment. The dispersion of resonances in both regions differs from that of PCI (Fig 10a), indicating that the protein stablishes internal interaction additional or different than those present in mature PCI, but keeping a well-defined three-dimensional structure.

V.E Discussion

Previous work in our laboratory has demonstrated that PCI can correctly fold *in vitro* with kinetics and efficiencies depending on the redox conditions used (Chang, 1994). Since its rate of folding *in vitro* is very slow, other folding helpers, as molecular chaperones, isomerases or pro-regions are expected to catalyze its folding *in vivo*. Both, its N and C terminal pro-peptides are removed *in vivo* (in the plant) before the protein is secreted, so it was a plausible possibility that these extensions might be involved in the folding process. This seemed more expectable for the N terminal pro-sequence, given that the C terminal pro-region is probably involved in sorting to vacuoles (Villanueva,1998). In this work we have studied the influence of PCI pro-regions on mature PCI folding rates and efficiency, using RP-HPLC to analyze the acid-trapped disulfide intermediates collected during the refolding process.

The *in vitro* folding of mature PCI was compared with that of ProNtPCI, ProPCI and with the ProNt in *trans*. The final yield of fully refolded (native) form achieved in all cases was similar under the conditions tested. In fact, the kinetic rates for ProNtPCI and ProPCI were slightly slower than those for PCI. Since the kinetics for the PCI plus the ProNt *in trans* are similar to that of PCI, it seems that the presence of the pro-sequence extension *in cis* might cause a slight decrease on the overall folding kinetics. Nevertheless, neither the progression of disulfide intermediates nor the final percentage of native form achieved (folding efficiency) are altered with respect to mature PCI. Thus, neither the N terminal nor the C terminal pro-sequence appear to have a positive effect on the rate of disulfide formation, the flow of disulfide species or the folding efficiency of PCI under the conditions tested.

Related results have been reported in the study of the folding of ω-conotoxins (Price-Carter, 1996a and b), where the N terminal pro-sequence has no effect on the mature protein folding and only the addition of a glycine residue at the C terminus (C terminal pro-sequence) enhances the yield of properly folded ω-conotoxin MVIIA. A glycine residue, placed after the last cysteine of the T-knot core, is conserved among all the known members of the squash inhibitor family and in PCI as well (both belong to the T-knot superfamily), where it plays an important role on its folding

(Venhudova, 2000). Unlike the case of conotoxins, in the latter molecules this glycine does not belong to the C terminal pro-sequence, since it is already present in the mature protein.

The *in vivo* studies in *E.coli* presented here show that for PCI and for the mutant proteins tested the N terminal pro-sequence has no effect on the expression level or the final yield of native form produced. In addition, the ratio between the scrambled species and the native form that were secreted into the cell culture supernatants were similar. Thus, we can assume that the N terminal pro-region does not affect the *in vivo* folding in *Escherichia coli* of the mature PCI. In the case of BPTI, another well studied small disulfide-rich protein, its N-terminal pro-region has been reported not to play a substantial role in disulfide bond formation or rearrangement within microsomes (Creighton et al, 1993), even though it seems to have a larger effect on its folding rate *in vitro* (Weissmann and Kim, 1992). The pro-region of PCI could play a role in the *in vivo* folding in the potato, interacting with other cellular factors. Such a role has been suggested for the propeptide of barnase that interacts with the molecular chaperone GroEl (Gray, 1993). Although it does not affect the folding pathway of its mature enzyme and lacks any indices of ordered structure, the pro-region of barnase provides an extra site or sites for binding to GroEL.

The presence of ordered structural elements in the N terminal extension of PCI has been investigated by several techniques. The CD spectra were recorded in order to determine the secondary structure present in the N terminal pro-region. Wt native PCI presents a characteristic maximum at 228 nm, which is not found in the mutant PCI proteins that lack the Tyr 37 residue, neither in wtPCI scrambled forms (Venhudova et al 2000), indicating that this band is present in the PCI molecules which present the Y37 residue in a correctly folded environment. The information obtained indicated that both ProNtPCI and PCI did not present the same secondary structure, since their CD spectra were different. The characteristic maximum band of PCI at 228 nm disappeared in the case of ProNtPCI, in spite of displaying CPA inhibitory activity and containing the Y37 residue. This could be explained by the presence in this region of another signal which would mask the characteristic band at 228 nm. The D/H exchange experiments also confirmed the presence of an extra structural element in the N terminal pro-region. It has been hypothesized that the last few amide hydrogens to exchange constitute the slow exchange core, that is formed by the secondary structure elements that are more tightly packed in a protein (Kim, 1993 and Woodward, 1993). Given that the number of slow exchanging protons of ProNtPCI was significantly higher than those of PCI, this indicates that the pro-region does not adopt a random coil conformation, but presents additional structural elements. Actually,

The leucine aminopeptidase exoproteolysis experiments also provided evidences of the ocurrance of an extra structural element in the N terminal pro-region of PCI. Interestingly, the presence of two accumulated peptides in the center of ProNtPCI pro-region (-10, -15) gives us information about the possible position of this element, which coincided with the secondary structure predictions of several algorithms.

Finally, the NMR analysis also prove that the ProNtPCI has a well-defined tridimensional structure and that presents some extra interactions in addition to those belonging to the mature PCI.

PCI N and C terminal extensions do not appear to have any substantial role in its *in vitro* folding neither in its *in vivo* folding in *Escherichia coli*, eventhough the ProNtPCI seems to posses a well-defined structure. Other roles not related with the folding of the mature protein have also been described for some pro-regions. For example, the pro-region of acetyclholine esterase in *Pichia pastoris* modulates the protein secretion (Morel, 1997) and the pro-region of caspase-8 interacts with the tumor necrosis factor receptor (Chaudhary, 1999). The propeptides might play their role during PCI biosynthesis or shortly after this process and subsequently ProPCI would undergo a proteolytic processing to remove these extensions. Whether these extensions might be involved in targeting PCI within the cell or in facilitating its interaction with other proteins still remains to be established.

V.F References

Alvares K, Siddiqui F, Malone J and Veis A (1999) Assembly of the type 1 procollagen molecule: selectivity of the interactions between the alpha 1(I) and alpha 2(I)-carboxylpropeptides. Biochemistry 38:5401-5411

Baardsnes J, Sidhu S, MacLeod A, Elliott J, Morden D, Watson J and Borgford T (1998) *Streptomyces griseus* protease B: secretion correlates with the length of the propeptide. Bacteriol 180:3241-3244

Caprioli RM, Fan T (1986) Peptide sequence analysis using exopeptidases with molecular analysis of the truncated polypeptides by mass spectrometry. Anal Biochem 154:596-603

Chang J-Y, LI L, Canals F and Aviles FX (2000) The unfolding pathway and conformational stability of potato carboxypeptidase inhibitor. J Biol Chem 275:14205-14211

Chang J-Y, Canals F, Schindler P, Querol E and Aviles FX (1994) The disulfide folding pathway of potato carboxypeptidase inhibitor. J Biol Chem 269: 22087-22090

Chatrenet B and Chang J-Y (1993) The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. 268:20988-20996

Chaudhary PM, Eby MT, Jasmin A and Hood L (1999) Activation of the c-Jun N-terminal Kinase/stress-activated protein Kinase pathway by overexpresssion of caspase-8 and its homologs. J Biol Chem 274:19211-19219

Clore GM, Gronenborn AM, Nilges M and Ryan CA (1987) Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. Biochemistry 26:8012-8023

Creighton TE, Bagley CJ, Cooper L, Darby NJ, Freedman RB, Kemmink J and Sheikh A (1993) On the biosynthesis of bovine pancreatic trypsin inhibitor (BPTI). Structure, processing, folding and disulphide bond formation of the precursor *in vitro* and in microsomes. J Mol Biol 232:1176-1196

Ferrari DM and Soling HD (1999) The protein disulphide-isomerase family: unravelling a string of folds. Biochem J 339:1-10

Ghrayeb J, Kimura H, Takahara M, Hsiung H, Masui Y and Inouye M (1984) Secretion cloning vectors in Escherichia coli. EMBO J 3:2437-2442

Gothel SF and Marahiel MA (1999) Peptidyl-prolyl *cis-trans* isomerases, a superfamily of ubiquitous folding catalysts. Cell Mol Life Sci 55:423-436

Gray TE, Eder J, Bycroft M, Day AG and Fersht AR (1993) Refolding of barnase mutants and pro-barnase in the presence and absence of GroEl

Hass GM and Ryan CA (1981) Carboxypeptidase inhibitor from potatoes. Methods in enzimology 80:778-791

Hidaka Y, Shimono C, Ohno M, Okumura N, Adermann K, Forssmann W-G and Shimonishi Y (2000) Dual function of the propeptide of prouroguanylin in the folding of the mature peptide. J Biol Chem 275:25155-25162

Ikemura H, Takagi H and Inouye M (1987) Requirement of pro-sequence for the production of active subtyilisin E in *Escherichia coli*. J Biol Chem 262:7859-7864

Kim KS, Fuchs JA and Woodward CK (1993) Hydrogen exchange identities native-state motional domains important in protein folding. Biochemistry 32:9600-9608

Lee YC, Miyata Y, Terada I, Ohta T and Matsuzawa H (1991) Involvement of NH2-terminal pro-sequence in the production of active aqualysin I (a thermoplilic serine protease) in *Escherichia coli*. Agric Biol Chem 55:3027-3032

Lin SL and Nussimov R (1995) A disulphide-reinforced structural scaffold shared by small proteins with diverse functions. Nat Struct Biol 2:835-837

Marie-Claire C, Ruffet E, Beaumont A and Roques BP (1999) The posequence of thermolysin acts as an intramolecular chaperone when expressed in trans with the mature sequence in Escherichia coli. J Mol Biol 285:1911-1915

Martoglio B and Dobberstein B (1998) Signal sequences: more than just greasy peptides. Trends Cell Biol 8:410-415

Molina MA, Aviles FX and Querol E (1992) Expression of a synthetic gene encoding potato carboxypeptidase inhibitor using a bacterial secretion vector. Gene 116:129-138

Morel N and Massoulie J (1997) Expression and processing of vertebrate acetylcholinesterase in the yeast Pichia pastoris. Biochem J 328:121-129

Price-Carter M, Gray WR and Goldenberg DP (1996a) Folding of ω-conotoxins.1.Efficient disulfide-coupled folding of mature sequence in vitro. Biochemistry 35:15537-15546

Price-Carter M, Gray WR and Goldenberg DP (1996b) Folding of ω-conotoxins.1.Influence of precursor sequences and protein disulfide isomerase. Biochemistry 35:15547-15557.

Rattenholl A, Ruoppolo M, Flagiello A, Monti M, Vinci F, Marino G, Lilie H, Schwaerz E and Rudolph R (2001) Pro-sequence assisted folding and disulfide bond formation of human neerve growth factor. J Mol Biol 305:523-533.

Shinde U and Inouye M (2000) Intramolecular chaperone: polypeptide extensions that modulate protein folding. Cell Developmental Biology 11:35-44

Studier FW, Rosenberg AH, Dunn JJ and Dubendorf JW (1990) . Use of T7 RNA polimerase to direct expression of cloned genes. Methods in enzimology 185:60-89

Thomas JG, Ayling A and Baneyx F (1997) Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli.* To fold or to refold. Appl Biochem Biotechnol 66:197-238

Van den Hazel HB, Kielland-Brandt MC and Winther JR (1993) The propeptide is required for *in vivo* formation of stable active yeast proteinase A and can function even when not covalently linked to the mature region. J Biol Chem 268:18002-18007

Van Wart HE and Birkedal-Hansen H (1990) The cysteine switch: a principle of regulation of metalloproteinase activity with gene family. Proc Natl Acad Sci USA 87:5578-5582

Venhudová G, Canals F, Querol E and Aviles FX (2001) Mutations in the N- and C-terminal tails of potato carboxypeptidase inhibitor influence its oxidative refolding process at the reshuffling stage. Journal Biol Chem 276:11683-11690

Villanueva J, Canals F, Villegas V, Querol E and Aviles FX (2000) Hydrogen exchange monitored by MALDI-TOF mass spectrometry for rapid characterization of the stability and conformation of proteins. FEBS Lett 472:27-33

Villanueva J, Canals F, Prat S, Ludevid D, Querol E and Aviles FX (1998) Characterization of the wound-induced metallocarboxypeptidase inhibitor from potatoe. cDNA sequence, induction of gene expression, subcellular immunolocalization and potential roles of the C terminal popeptide. FEBS Lett 440:175-182

Weissman JS and Kim PS (1992) The pro region of BPTI facilitates folding. Cell 71:841-851

Winther JR and Sorensen P (1991) Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of the enzymatic activity. Proc Natl Acad Sci USA 88:9330-9334

Woodward C (1993) Is the slow exchange core the protein folding core? TIBS 18:359-360

Zhu X, Ohta Y, Jordan F and Inouye M (1989) Pro-sequence of subtilisin can guide the folding of denatured subtilisin in an intermolecular process. Nature 339-:483-484