doi:10.1006/jmbi.2001.5076 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2001) 313, 629–638

JMB

Crystal Structure of a Novel Mid-gut Procarboxypeptidase from the Cotton Pest *Helicoverpa armigera*

Eva Estébanez-Perpiñá[1], Alex Bayés[2], Josep Vendrell[2] Maarten A. Jongsma[3], David P. Bown[4], John A. Gatehouse[4] Robert Huber[1], Wolfram Bode[1]*, Francesc X. Avilés[5]* and David Reverter[1]*

[1]*Abteilung für Strukturforschung, Max-Planck-Institut für Biochemie Am Klopferspitz 18a D-82152 Planegg-Martinsried Germany*

[2]*Departament de Bioquímica i Biologia Molecular, Facultat de Ciències Universitat Autònoma de Barcelona, 08193 Bellaterra Barcelona, Spain*

[3]*Business Unit Cell Cybernetics, Plant Research International B.V., PO Box 16 NL-6700 AA, Wageningen, The Netherlands*

[4]*Department of Biological Sciences, University of Durham, South Road, Durham DH1 3LE, UK*

[5]*Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain*

**Corresponding authors*

The cotton bollworm *Helicoverpa armigera* (Hubner) (Lepidoptera: *Noctuidae*) is one of the most serious insect pests in Australia, India and China. The larva causes substantial economical losses to legume, fibre, cereal oilseed and vegetable crops. This pest has proven to be difficult to control by conventional means, mainly due to the development of pesticide resistance. We present here the 2.5 Å crystal structure from the novel procarboxypeptidase (PCPAHa) found in the gut extracts from *H. armigera* larvae, the first one reported for an insect. This metalloprotease is synthesized as a zymogen of 46.6 kDa which, upon *in vitro* activation with Lys-C endoproteinase, yields a pro-segment of 91 residues and an active carboxypeptidase moiety of 318 residues. Both regions show a three-dimensional structure quite similar to the corresponding structures in mammalian digestive carboxypeptidases, the most relevant structural differences being located in the loops between conserved secondary structure elements, including the primary activation site. This activation site contains the motif $(Ala)_5Lys$ at the C terminus of the helix connecting the pro- and the carboxypeptidase domains. A remarkable feature of PCPAHa is the occurrence of the same $(Ala)_6Lys$ near the C terminus of the active enzyme. The presence of Ser255 in PCPAHa instead of Ile and Asp found in the pancreatic A and B forms, respectively, enlarges the S1′ specificity pocket and influences the substrate preferences of the enzyme. The C-terminal tail of the leech carboxypeptidase inhibitor has been modelled into the PCPAHa active site to explore the substrate preferences and the enzymatic mechanism of this enzyme.

Keywords: carboxypeptidases; crystal structure; *Helicoverpa armigera*; pests

Introduction

Helicoverpa armigera (Lepidoptera: *Noctuidae*), also known as bollworm or corn earworm, has a widespread distribution in tropical, subtropical and warm temperature regions. *H. armigera* is a polyphagous pest of 181 plant species, including cotton (*Gossypium* species) and maize (*Zea mays*), and is predicted to become an important pest in other crops such as tobacco (*Nicotiana tabacum*).[1] It is one of the most serious insect pests in important cotton-producing countries like Australia, India and China. The economic problems caused by this voracious insect are enormous, as exemplified by the crisis in the cotton production in China after the occurrence of mass outbreaks of *H. armigera* in 1992.[2]

Abbreviations used: PCPAHa, *H. armigera* procarboxypeptidase; PCP, procarboxypeptidase; CPA and CPB, carboxypeptidase A and B; LCI, leech carboxypeptidase inhibitor; PCI, potato carboxypeptidase inhibitor.

E-mail addresses of the corresponding authors: fx.aviles@blues.uab.es; reverter@biochem.mpg.de

0022-2836/01/030629–10 $35.00/0

Until recently, studies of the protein digestion in the insect gut have concentrated mainly on the initial phases of protein and peptide degradation through endopeptidase (trypsin and chymotrypsin-like proteases) activities.[3] However, the products of such endopeptidase cleavage are large peptides, which must be further degraded to be taken up by the insect gut cells.[4] Consequently, exopeptidases such as carboxypeptidases are also significant factors in the overall process of protein digestion.

Recently, a novel procarboxypeptidase from *H. armigera* (PCPAHa), the first enzyme of this class from a lepidopteran insect, has been characterised by expressing its encoding cDNA in insect cells.[5]. The pre-proprotease sequence contains 426 amino acid residues and exhibits sequence homology with the metallo-carboxypeptidases from mammalian species, and with carboxypeptidases from other invertebrates. Removal of a predicted signal peptide and of the activation peptide results in a product of approximately 35.5 kDa, which is comparable in size to the pancreatic mammalian carboxypeptidases.[6] PCPAHa belongs to the A form of carboxypeptidases and preferentially cleaves aliphatic and aromatic residues. This carboxypeptidase is active over a broad pH range (7.5-10), displaying maximal activity at pH 8.0. This is in agreement with the report of a maximal peptidase activity in the alkaline region determined in other phytophagous lepidopteran species.[7–9]

We have determined the three-dimensional structure of this novel procarboxypeptidase from *H. armigera*. This is the second structure reported for an insect proteinase[10] and the first for an insect carboxypeptidase. The structure of PCPAHa resembles the mammalian pancreatic procarboxypeptidases, containing a long pro-segment that inhibits the active enzyme without having any direct interaction with the zinc and the active-site region.

Results and Discussion

Overall structure

Like other mammalian procarboxypeptidases (PCPs),[6] the *H. armigera* procarboxypeptidase has a globular shape with two independent domains: the large pro-segment or activation segment, consisting of 91-97 residues, and the active enzyme moiety with 312-318 residues (Figure 1). Upon proteolytic cleavage of the zymogen, the interaction of the pro-segment with the enzyme moiety is weakened, allowing the release of the pro-segment. The possible presence of more than one activation site would be in agreement with the connecting loop (Figure 1) being considerably long, lacking any regular structure and containing three basic residues (see also Figure 2(b)). One particular feature of PCPAHa when compared to the related pancreatic PCPs (PCPA2, PCPB and PCPA1)[11–13] is the slight rotation shift of 2-3° of the whole pro-segment relative to the carboxypeptidase moiety (see Figure 2(a)). This shift does not seem to compromise the inhibitory capacity of the pro-segment, as the proenzyme has no detectable intrinsic activity (A.B. *et al.*, unpublished results).

Activation segment of PCPAHa

The activation segment of PCPAHa can be described as being composed of two different regions (Figure 1): the large N-terminal globular domain, which shields the pre-formed wide active-site depression of the enzyme, and the connecting segment consisting of a long α-helix (α3) that links

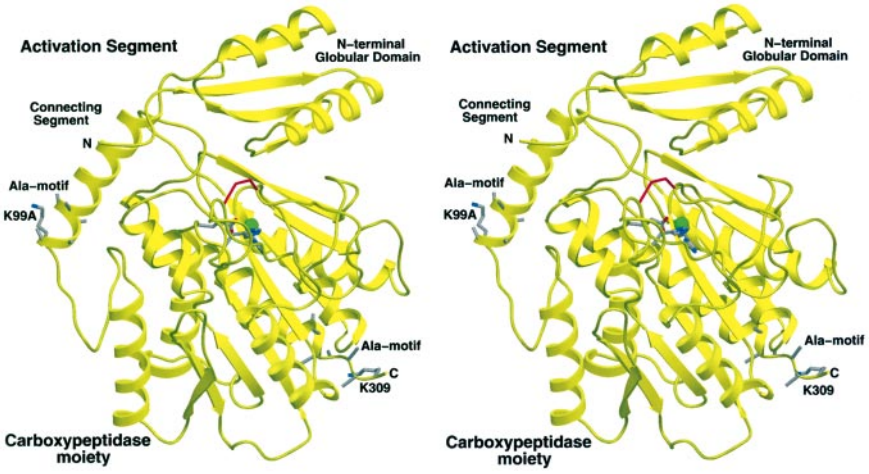

Figure 1. Three-dimensional structure of PCPAHa. Stereo ribbon plot representation of PCPAHa. The pro-segment is folded into a globular N-terminal domain consisting of an antiparallel-α/antiparallel-β open-sandwich, and an α-helical connecting segment. The catalytic moiety is formed by a central twisted eight-stranded β-sheet flanked by eight α-helices, together forming a globular α/β protein. The zinc ion (green sphere) is coordinated in the active site by two histidine residues and one glutamic acid residue. The single disulfide bridge contained is shown in red. Lys99A, Lys309 and the Ala-motifs are shown as grey stick models.

this globular domain with the carboxypeptidase moiety.[12] The globular N-terminal domain (residues His4A to Val82A) exhibits an open-sandwich antiparallel-α/antiparallel-β-fold, made up by two α-helices and four β-strands arranged with a β1α1β2β3α2β4 topology (Figure 1). Despite the relatively low sequence identity of the pro-segment of PCPAHa to those of mammalian carboxypeptidases (around 20 %), the overall folding topology is conserved (Figure 2). This structural similarity is higher to the A forms, which lack the 3_{10}-helix insertion (43A-46A) that is typical in the B forms. The C-terminal α-helical connecting segment comprising residues Lys83A to Lys99A is followed by a loop that links the pro-peptide with the carboxypeptidase (CP) moiety. Noteworthy is the presence of five consecutive Ala residues followed by an exposed Lys at the end of the connecting segment (Figure 1). This endoprotease cleavage motif that forms the last turn of the α3-helix has never been described before for procarboxypeptidases, and its role is still unclear. Due to its four turns, this α3-helix resembles the connecting helices of PCPA forms, where four and five-turn helices are found (PCPA1 and PCPA2) compared to the two-turn α3-helix observed in PCPB.

Due to the sequence homology and from previous biochemical studies (A.B. *et al.*, unpublished results) it is known that the *in vitro* activation by Lys-C endoproteinase occurs at Lys99A, located at the end of the activation segment (Figures 1 and 2). Although the activation with Lys-C endoproteinase is more efficient, the proenzyme can also be activated *in vitro* by bovine trypsin at Arg4, located just five residues behind Lys99A. On the other hand, the sequence of the carboxypeptidase purified from larval gut extracts is LSFDKIHSYEEVDAYLQELAKEFPNVVTVV, which coincides with residues numbered 7-36 in Figure 2(b).

Carboxypeptidase moiety of PCPAHa

The PCPAHa catalytic moiety exhibits the same fold as the other known carboxypeptidase structures:[11–14] a central twisted eight-stranded β-sheet (β5-β12) flanked by eight α-helices (α4-α11), which together form a globular α/β protein (Figure 1). Compared to the structure of the search model (human PCPA2), several insertions and deletions located in the loops that connect the secondary structure elements are required to align both enzymes (Figure 2(a)), namely: a three residue insertion in the loop between β6 and β7, a four residue deletion between α5 and α6, a three residue insertion between α9 and β11 and a final three residue insertion after β12. Another $(Ala)_6Lys$ motif, similar to that found in the connecting segment (see above), is located at the end of the last α-helix (α11) which contains six consecutive Ala residues compared to the five described for the activation segment (Figures 1 and 2(b)).

PCPAHa possesses only one disulfide bond (Cys138-Cys161), being the only one conserved in all known procarboxypeptidases. However, this protease has an additional buried free cysteine residue (Cys70), following the His69 from the zinc-binding motif, replacing the Ala70 currently found in the other enzymes (Figures 2(b) and 3). No catalytic role has been ascribed to this cysteine residue, despite its proximity to the active-site residues.

The active site cleft of the PCPAHa is placed at the C-edge of the central part of the β-sheet of the catalytic moiety. The active site is composed of a zinc atom and the different active subsites (S1′, S1, S2, S3 and S4, nomenclature according to Schechter & Berger[15]). In the PCPAHa, the typical Zn^{2+} coordination also described for the other CPs is found (Figure 3): the zinc ion is pentahedrally coordinated by His69 $N^{\delta1}$, His196 $N^{\delta1}$, both Glu72 $O^{\varepsilon1}$ and $O^{\varepsilon2}$ atoms, and the ''catalytic'' water molecule, which in addition is linked to the general base Glu270 (distance 3.7 Å). As in the other CP structures, the correct arrangement of His196, one of the catalytic Zn^{2+} ligands, requires the formation of a characteristic *cis*-peptide bond between Ser197 and Phe198. In the absence of a substrate, the substrate-binding site is apparently empty in the crystal structure of PCPAHa, but contains several disordered solvent molecules. The essential residues for catalysis have been identified on the basis of studies made with CPs and their zymogens.[6,11,12,16,17]

The S1′ subsite, which has a dual positive-charged/hydrophobic nature to anchor and neutralize the scissile C-terminal residue of the substrate, is lined by residues Leu203-Gly207, Ile243-Asp256 and Tyr267-Thr268 (Figures 3 and 4). The hydrophobic character of the pocket is mainly provided by Ile243, Ala250 and Leu247 (i.e. residues also conserved in the other CPAs). However, the Ser255 at the bottom of the pocket replacing the aspartic acid or isoleucine residue in CPB and CPA, renders this S1′ pocket larger as well as more polar. The residues Arg145 and Asn144, which extend into this pocket, anchor the negative C-terminal carboxylate of the substrate. The nearby Tyr248 is found in an ''up'' conformation in the PCPAHa structure, as currently seen in the empty active sites of the PCP structures,[19,20,22] interacting with the side-chains of the residues Val23A and His27A of the pro-segment (see below). The adjacent S1 subsite formed by Arg127 and Glu270 extends, *via* the S2 and S3 subsites, into the bulk water.

Probable interaction with peptidic substrates

To explore the probable binding geometry of a peptide substrate and the potential enzymatic mechanism of CPAHa, the C-terminal tail Ile62-Pro63-Tyr64-Val65↓Glu66 of leech carboxypeptidase inhibitor (LCI) has been specifically modelled into the active site of PCPAHa based on the structure of LCI in complex with human CPA2.[26] In this

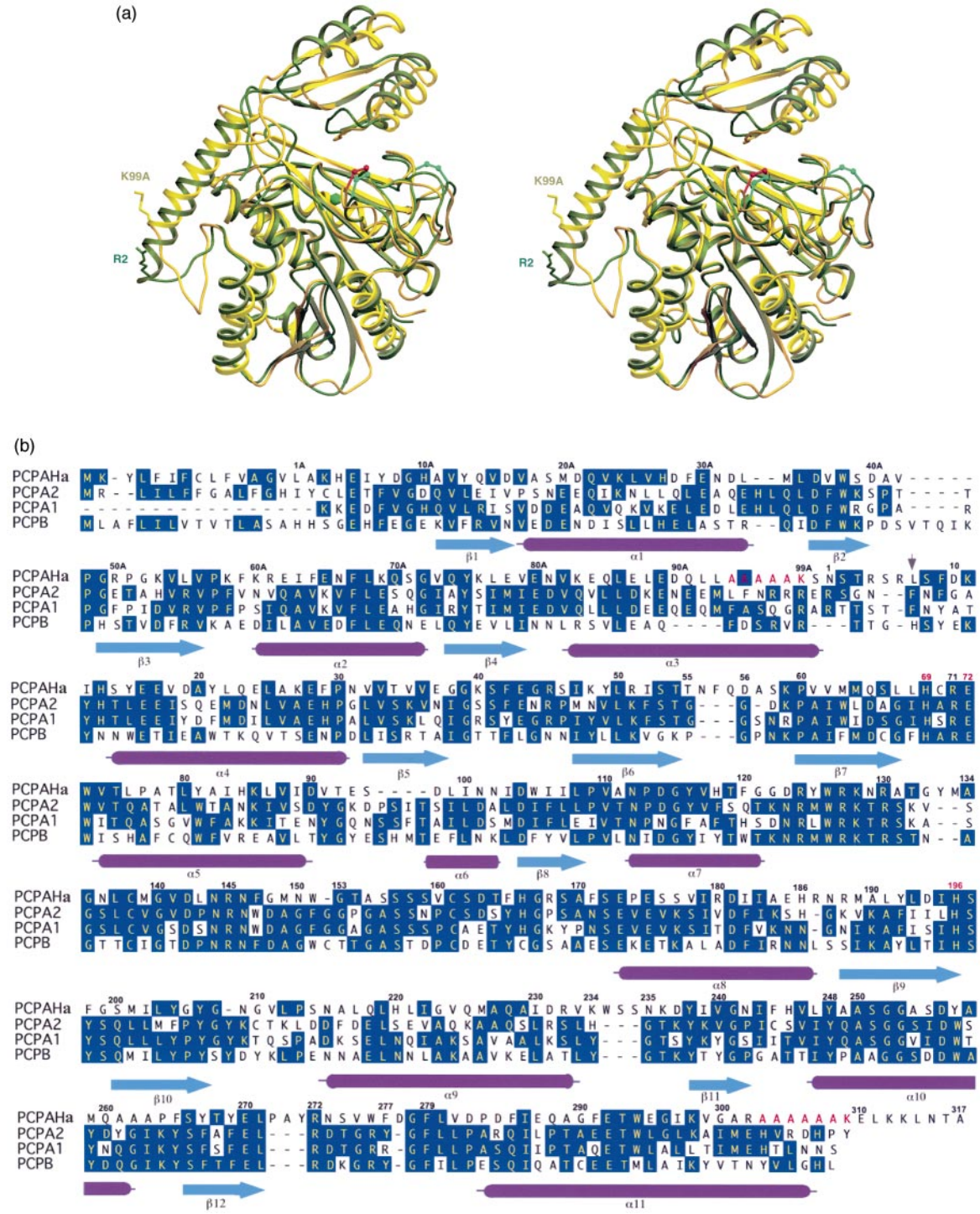

Figure 2. Topological and sequence comparison of PCPAHa and human PCPA2. (a) Stereo ribbon plot superimposition of PCPAHa (yellow) and human PCPA2 (green). The main differences between these enzymes are located in the loops between the secondary structure elements. A slight shift is observed in the activation segment of PCPAHa compared to human PCPA2. The Figure was prepared with SETOR.[34] (b) Structure-based amino acid sequence alignment between *H. armigera* PCPAHa, human PCPA2, porcine PCPA1 and porcine PCPB. The numbering of both moieties of the proenzyme is based on three-dimensional topological alignments with the other known procarboxypeptidases structures.[11–13] The numbers of the amino acid positions in the activation segment are followed by an A to distinguish them from positions in the enzyme moiety. An arrow on Leu7 indicates the N terminus of the mature enzyme isolated from larval extracts. The β-sheets (β1 to β12) and the α-helices (α1-α11) are shown by arrows and cylinders, respectively. The alignment was prepared with the DNASTAR software package (Madison, USA).

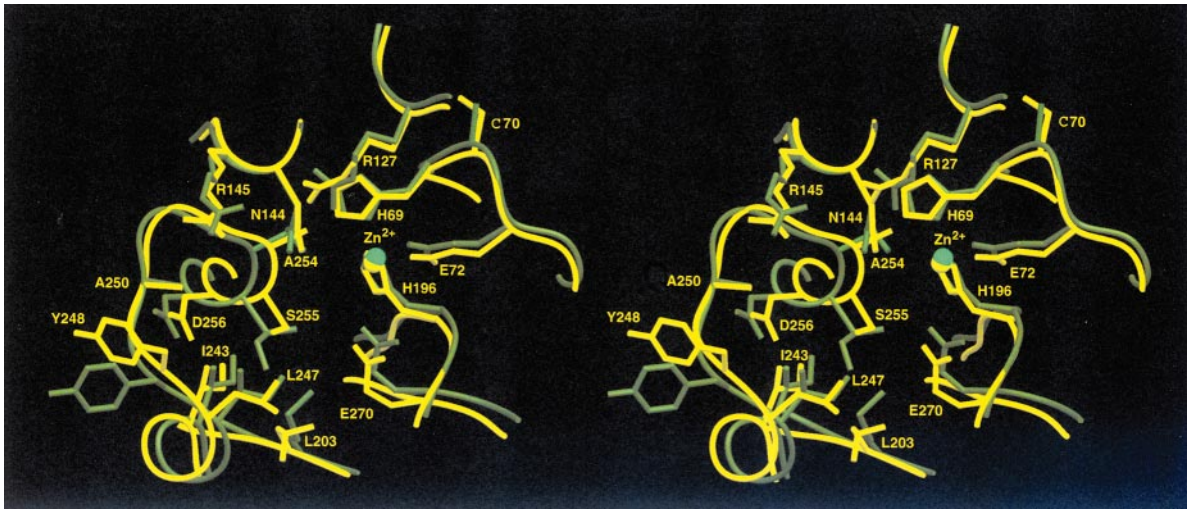

Figure 3. Section around the S1′ pocket of PCPAHa (yellow) compared to PCPA2 (green). The broader substrate specificity of PCPAHa can be explained by the substitution of Ile255 by Ser at the base of the substrate-binding pocket. This residue change renders the pocket less hydrophobic and larger. The zinc ion is shown as a green sphere. The Figure was prepared with SETOR.[34]

structure, the last five C-terminal residues of LCI represent the ''primary'' interaction region believed to interact with the zinc and the subsites S1′, S1, S2 and S3 in a substrate-like manner. The last residue of the inhibitor (Glu66i) is cleaved off by the carboxypeptidase but remains trapped in the S1′ cavity. In the PCPAHa model the Glu66i has been substituted by Gln in order to remove its charge but allowing polar interaction with the lining residues of the pocket. The model was submitted to several rounds of energy minimisation with XPLOR[18] in order to avoid clashes and find the most favourable conformation.

The current model of the enzyme complex formed with the LCI C-terminal peptide mimics the optimal geometrical conformation and interactions of the substrate in the active site in the pretransition state (Figure 4). The most accepted model of peptide hydrolysis is the so-called water-promoted mechanism.[17] In such a model, the catalytic zinc ion together with the other electrophile, the guanidyl group of the Arg127, polarise the carbonyl group of the scissile peptide bond, and simultaneously, the catalytic water molecule is rendered to a potent nucleophile by donation of one proton to the catalytic Glu270. The resulting

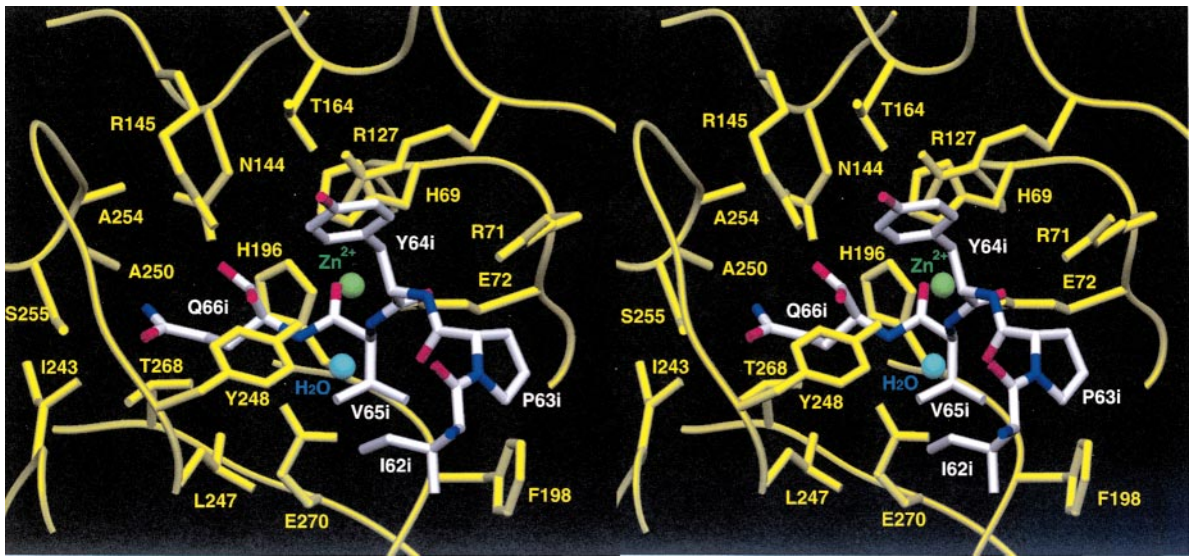

Figure 4. Probable interaction of the Ile-Pro-Tyr-Val↓Gln C-terminal pentapeptide of LCI with the active site of PCPAHa. The residues of the active site of PCPAHa are shown in yellow. The modelled C-terminal segment of LCI is shown in a charged-coloured stick representation. The ''catalytic'' water molecule and zinc atom are shown as blue and a green spheres, respectively. The Figure was prepared with SETOR.[34]

hydroxyl can attack the carbonyl carbon atom of the scissile peptide bond of the substrate.

In the S1′ subsite the accommodation of the C-terminal Gln66i produces not only a slight shift of the Arg145 and Asn144 side-chains, but also a dramatic (and well-known) conformational change of Tyr248 from the ''up'' (Figure 3) to the ''down'' position (Figure 4). Arg145 and Asn144 side-chains are at interaction distances to the carboxylate oxygen atom of Gln66i, and the hydroxyl side-chain oxygen atom of Tyr248 interacts with the amino group of Gln66i. Such a ''down'' conformation of the Tyr248 has been frequently reported in CP and PCP structures containing a substrate/inhibitor inside of the active site.[11,16,19] The side-chain of Gln66i fits perfectly to the polar environment created at the bottom of the S1′ pocket by Ser255 and Thr268 (Figure 5).

It is worth mentioning that whilst some discrepancies appeared recently in the literature for the structures of active carboxypeptidases with respect to the ''up'' and ''down'' positions of Tyr248,[20,21] the view is much more homogeneous in the case of the zymogen structures. Thus, it has been reported the ''up'' conformation (interacting with residues of the pro-segment) in porcine PCPB and human PCPA2.[11,12] Only in the case of porcine PCPA1 (with a free valine bound to the S1′ subsite[13]) and in the benzylsuccinic-complexed PCPA2,[11] the Tyr248 is found in the ''down'' conformation.

In the S1 subsite the carbonyl group of the P1-Val65i is directed just to the side of the zinc ion, with its oxygen atom also located close to the guanidinium group of Arg127. The ''catalytic'' water molecule is sandwiched between the carbonyl carbon atom of Val65i and the Glu270 carboxylate (Figure 4). The geometry of the modelled substrate presents a good scenario for the nucleophilic attack of the ''catalytic'' water on the scissile peptide bond of the substrate (Val65i-Gln66i). The Arg127 and the zinc atom are optimally placed to stabilize the negative charges formed during the transition state. The amino group of Val65i is also in contact with the hydroxyl oxygen atom of the Tyr248 (as in the S1′ subsite), and its side-chain is located inside a small hydrophobic pocket created by Phe198, Phe279 and Leu247.

The major interaction in the S2 subsite is the hydrogen bond between the carbonyl oxygen atom of the P2-Tyr64i and the guanidinium group of Arg71. This carbonyl group also interacts with the guanidinium group of Arg127, and the phenyl ring

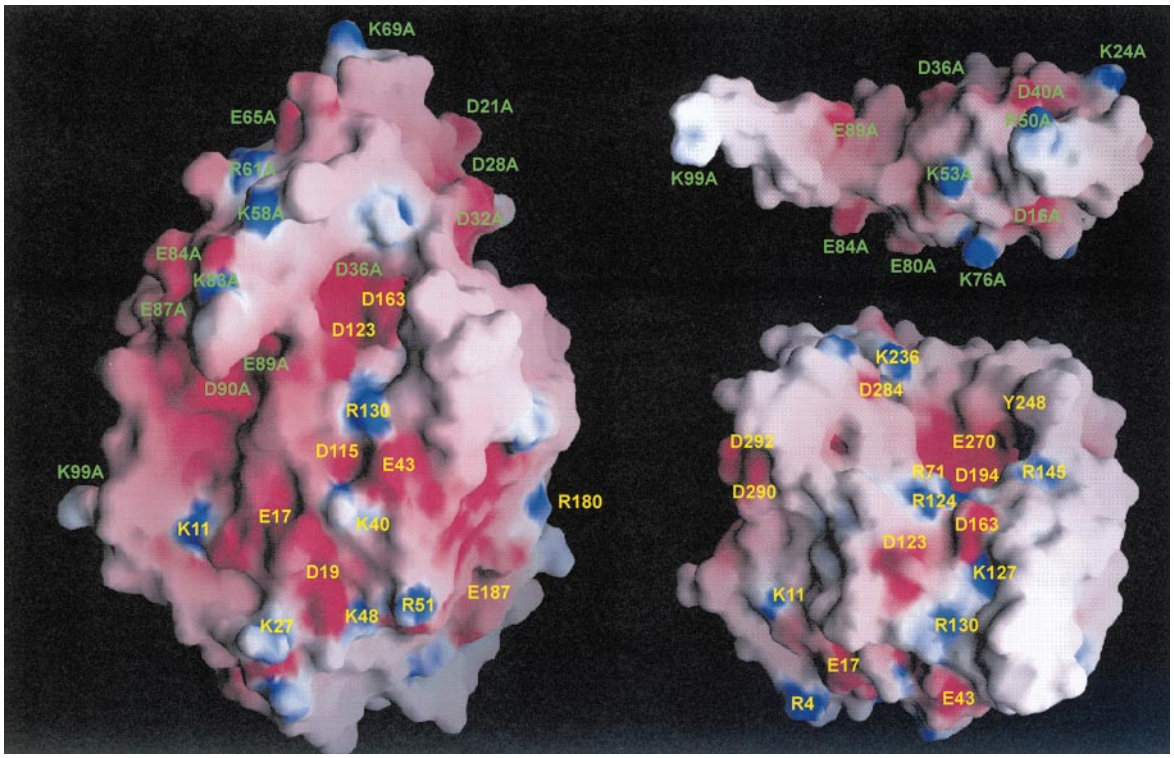

Figure 5. Solid surface representation of PCPAHa. (a) The PCPAHa molecule is shown in an identical orientation as in Figures 1 and 2, with the exposed Lys99A at the end of the connecting segment being clearly visible. PCPAHa has many acidic residues as indicated by the mainly red colouring of the surface, with colours indicating positive (blue) and negative (red) electrostatic surface potential contoured from +15 kT/e to 15 kT/e. (b) Surface of the pro-segment docking the carboxypeptidase moiety. (c) Surface of the carboxypeptidase moiety interacting with the pro-segment. Both molecules described in (b) and (c) have moved apart and rotated by 90° in opposite directions from each other around the horizontal axis to show the respective interacting surfaces. The salt-bridges between residues Asp36A, Lys53A and Glu89A from the activation segment and residues Arg71, Asp284 and Arg124 from the carboxypeptidase moiety, respectively, are observable. The Figure was prepared with GRASP.[35]

of Phe279, which belong to the S1 and S3 subsites, respectively. The side-chain of Tyr64i is in contact with Arg127, Arg145 and Thr164. In the S3 subsite, the only interactions observed are the contact of the methylene C^{γ} and C^{δ} of the P3-Pro63i with the phenyl rings of Phe198 and Phe279.

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

The inhibition of PCPAHa is accomplished through blocking the wide active-site depression by the globular part of the activation segment (Figures 1 and 5). The activation segment occludes the S2, S3 and S4 subsites of the enzyme, preventing substrates from entering into the active-site groove of the enzyme. The main interactions in this region are the salt-bridge between the side-chain groups of Asp36A with Arg71 (around the S2 subsite), the hydrophobic interactions between the indole group of Trp37A with the phenyl groups of Phe198 and Phe279 (around the S2 and S3 subsites), and the interaction of the His24A and Val20A side-chains with the side-chain of Tyr248 (belonging to the S1′ subsite, but in the ''up'' conformation). There is, however, no direct contact of the pro-segment with the zinc atom or with the residues involved in the formation of the S1′ and the S1 subsites. This inhibition mechanism differs from the substrate-like inhibition mechanism of the CP protein inhibitors LCI and potato carboxypeptidase inhibitor (PCI), where a direct contact with the catalytic residues (S1′ and S1 subsites) and the Zn^{2+} atom of the enzyme occurs.[19,22] In Figure 5, the electrostatic interactions between the pro-segment with the carboxypeptidase moiety are shown. In contrast to the structure of porcine PCPB[12] but in agreement with the PCPAs,[11,13] PCPAHa does not possess any salt-bridge between the active-site residue Arg145 and any acidic residue from the pro-segment. But in the case of PCPAHa, no intrinsic activity is detected in contrast to PCPA1 and PCPA2.

The globular activation domain and the connecting segment (helix α3) contribute differently to the inhibition of the carboxypeptidase moiety. The interaction surfaces of the globular domain (residues His4A-Val82A) and the connecting segment (Lys83A-Lys99A) with the carboxypeptidase moiety are approximately 730 Å^2 and 515 Å^2, respectively, which lie within the range of other PCPs structures.[6] Combination of structural information and knowledge about the activation processes of mammalian pancreatic A1, A2 and B forms,[23–25] have helped us to understand the contribution of both parts of the pro-segment in the inhibition of the enzyme.

Conclusions

Here, we present the three-dimensional structure of a procarboxypeptidase from *H. armigera*, a devastating pest of important crop plants. The structure shows the preservation of the overall fold of exopeptidases throughout the animal kingdom. Thus, the characteristic zinc procarboxypeptidase fold, the catalytic mechanism and the inhibition exerted by the pro-segment seem to be universal. However, despite the general scaffold similarity, structural differences can be found in the loops between the conserved secondary structures, including the loop where the activation processing occurs.

The presence of the amino acid region with the sequence $(Ala)_nLys$ located where the proteolytic activation cleavage of the enzyme is likely to occur, is a remarkable feature of this procarboxypeptidase. The fact that this sequence not only appears at the activation-site of the protease but also near the C terminus, suggests that post-translational protein processing, different from the activation process might occur. We have observed that one of them, from Ala94 to Lys99, is the *in vitro* activation point for a Lys-specific endoproteinase (A.B. *et al.*, unpublished results). Most probably the removal of an octapeptide by the cleavage after the lysine in the second $(Ala)_nLys$ sequence (from Ala303 to Lys309) is a process catalysed *in vivo* by the same enzyme responsible for activation of PCPAHa. Although we still do not have *in vivo* evidence for this cleavage, structural data support this possibility, since such a sequence is exposed and thus accessible to this second proteolytic attack.

Trypsin can also activate the pro-enzyme *in vitro* at Arg4, located just five residues behind Lys99A, but this activation is not as efficient as in the case of LysC endopeptidase. The stability and proteolytic activities of the LysC and trypsin-cleaved CPAHa forms in front of the synthetic substrate *N*-(3-(2-furyl)acryloyl)-Phe-Phe (FAPP) is similar, at 25 °C. The purification of a carboxypeptidase enzyme in *H. armigera* gut extracts with an N-terminal sequence starting at Leu7, as a result of cleavage at Arg6, confirms that trypsin-like enzymes can sever the unstructured carboxypeptidase N terminus. Our hypothesis about the *in vivo* activation process would involve a controlled cleavage at the specific $(Ala)_5$-Lys site by the natural activator, followed by subsequent trimming of the unstructured N terminus by a trypsin-like action up to the sequence found in the carboxypeptidase purified from the larva extracts.

The modelled structure with the reconstructed C-terminal tail of LCI has revealed the optimal geometrical conformation of the residues involved in the substrate-binding subsites of the enzyme. The model provides a good scenario for the water-promoted catalytic mechanism attained before the transition state occurs, including the interaction between the catalytic residues Glu270 and Arg127 with the catalytic water, the zinc atom and the scissile peptide bond of the C-terminal residue of the substrate. The presence of Ser255 in the S1′ pocket enlarges the classical aliphatic and aromatic substrate preference found in the A forms. Only the

carboxypeptidase sequence from the earthworm *Lumbricus rubellus* possesses a similar residue in its hydrophobic pocket (Protein Data Bank accession number Y09625), but no functional information is available in this case. Functional studies performed with recombinant PCPAHa confirm this broad substrate preference for this enzyme, including polar residues (A.B. *et al.*, unpublished results).

The quests for an effective, safe and lasting pest management program should be the primary target in the development of new products with a minimal undesirable impact on the environment, which could replace conventional toxic pesticides. The development of pesticide resistance in insects causes the need to inhibit more specific enzymatic targets. All attempts to eliminate the *H. armigera* pest by conventional means have been unsuccessful to date due the build up of resistance to pesticides.[26] Knowledge of the regulation of digestive proteinases at a biochemical and structural level will help to develop new tools to combat this pest.

Materials and Methods

Purification of the carboxypeptidase from *H. armigera* larval gut

The mature enzyme was purified from *H. armigera* larval gut extracts (prepared as described by Bown[27]) by affinity chromatography using PCI as ligand. After equilibrating the column with 50 mM Tris-HCl (pH 7.5), 0.1 M NaCl, 10 mM benzamidine, 2 mM phenylmethylsulphonyl fluoride, the larval gut extract was applied to the column, which was subsequently washed with equilibration buffer, and with 0.1 M glycine-NaOH (pH 12.0) (>five column volumes each). Specifically bound proteins were eluted with 6 M guanidine-HCl, acetone-precipitated and centrifuged. The pellet was redissolved in SDS sample buffer and, after SDS-PAGE, the protein was blotted onto a PVDF membrane, identified by staining with Coomassie blue, excised from the membrane and subjected to N-terminal sequencing on an Applied Biosystems 477A protein sequencer.

Heterologous expression and purification of *H. armigera* procarboxypeptidase

The cDNA corresponding to the PCPAHa[5] was amplified by PCR with the primers PCP-Ha5′: GATTCTCTCGAGAAAAGAAAACATGAAATTTATGATGG and PCP-Ha3′: CATCCTAGGTTAAGCAGTATTGAGTTTCTTC and cloned into the extracellular *Pichia pastoris* expression vector pPIC9. Purification to homogeneity of the proenzyme was made by hydrophobic interaction chromatography (Butyl-Sepharose, TosoHaas) at atmospheric pressure, followed by anion-exchange chromatography (TSK-DEAE 5PW) using an FPLC system with a linear gradient from 0-15 % of 0.8 M ammonium acetate in 10 mM Tris-HCl (pH 8.0), 1 μM $ZnCl_2$. Prior to crystallization, the sample was dialysed against 5 mM Tris-HCl (pH 8.0), 1 μM $ZnCl_2$, and concentrated to approximately 10 mg/ml (Ultrafree® filter device from Millipore).

Crystallization and structure determination

A single small crystal grew after two months in 30 % (w/v) PEG 8000, 0.2 M sodium acetate, 0.1 M sodium cacodylate (pH 5.5) by mixing 2 μl of the protein sample with 2 μl of the crystallization buffer. Using micro and macro-seeding techniques, large crystals could be produced after a few days, which reached a maximum size (0.4 mm × 0.3 mm × 0.5 mm) in approximately one week. The largest crystal was harvested using the reservoir solution and mounted in a thin-walled glass capillary tube. A complete data set up to 2.5 Å was collected at room temperature in-house using a 300 mm Mar Research image plate detector. The crystal belonged to space group $P2_1$ with cell constants $a = 48.03$ Å, $b = 86.39$ Å, $c = 50.55$ Å and $\beta = 100.70°$, and contained one molecule per asymmetric unit. Diffraction data (see Table 1) were evaluated, scaled and reduced with DENZO-SCALEPACK.[29] The PCPAHa structure was solved by molecular replacement with AMoRe,[13] using the structure of human procarboxypeptidase A2 (PDB code 1AYE) as a search model. Diffraction data from 12.0 Å to 3.5 Å were used to calculate the rotation and translation functions. The correlation factor and the *R*-factor of the best solution after rigid body refinement was 25.5 and 48.5 %, respectively. Model building was done on an SGI graphics workstation using TurboFRODO.[30] Calculation of the electron density maps and crystallographic refinement were performed with X-PLOR[18] and REFMAC/CCP4 program.[31] After several cycles of model building, conjugate gradient minimisation and simulated annealing, a structure with good stereochemistry was obtained. The target parameters of Engh & Huber[32] were used. A Ramachandran plot calculated using the program PROCHECK[33] shows that all the residues fall into the most favoured (86.6 %) or additionally favoured regions (13.4 %).

Table 1. X-ray data and refinement statistics

Crystal cell constants	$a = 48.032$ Å, $b = 86.395$ Å $c = 50.550$ Å, $\beta = 110.7°$
Space group	$P2_1$
Limiting resolution (Å)	2.5
Total number of reflections measured	230,224
Unique reflections (12.0-2.5 Å)	21,449
R_{merge}[a] (12.0-2.5 Å) (%)	9.2
R_{merge}[a] (2.6-2.5 Å) (%)	34.4
Completeness (%) (12.0-2.5 Å)	93.2
Completeness (%) (2.6-2.5 Å)	93.0
Solvent molecules	157
Zinc ions	1
Reflections used for refinement	18,816
Sigma cut-off for refinement	2.0
Resolution range used (Å)	12.0-2.5
R-value[b] overall (%)	18.01
R_{free}[c] (%)	23.81
r.m.s standard deviations	
Bond length (Å)	0.011
Bond angles (deg.)	1.5

[a] $R_{merge} = [\Sigma_h \Sigma_i \, |I(h,i) - \langle I(h)\rangle| / \Sigma_h \Sigma_i I(h,i)] \times 100$, where $I(h,i)$ is the intensity value of the ith measurement of h and $\langle I(h)\rangle$ is the corresponding mean value of h for all i measurements of h. The summation is over all measurements.

[b] R-value $= (\Sigma |F_o - F_c| / \Sigma F_o) \times 100$.

[c] R_{free} was calculated randomly omitting 7 % of the observed reflections from refinement and *R*-factor calculation.

The numbering system for the activation segments of procarboxypeptidases is derived from the alignment of all pancreatic-like procarboxypeptidases as described by García-Sáez.[11] The numbers of the amino acid positions at the pro-segment are followed by an A to distinguish them from positions in the enzyme moiety. Following this system, the activation segment of PCPAHa spans from His4A to Lys99A, even though the real length of the activation segment is 91 residues (see Figure 1 for deletions). The numbering of bovine carboxypeptidase A1 is kept for the enzyme moiety.

Protein Data Bank accession number

The coordinates of the *H. armigera* procarboxypeptidase A (PCPAHa) have been deposited with the PDB ID code 1JQG.

Acknowledgments

Thanks to Carlos Fernández-Catalán, Pablo Fuentes-Prior and John Richardson for helpful discussions. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 413 and SFB 469) and by the E.C. Biotechnology Programme (DGXII-E/BH/009294). D. R. has been a recipient of a fellowship from the European Community Biotechnology Marie Curie program. E. E. acknowledges the EU ''Training and Mobility'' programme financial support (COSSAC, FMRX-CT-98-0193).

References

1. Harsulkar, A. M., Giri, A. P., Patankar, A. G., Gupta, V. S., Sainani, M. N., Ranjekar, P. K. *et al*. (1999). Successive use of non-host plant proteinase inhibitors required for effective inhibition of *Helicoverpa armigera* gut proteinases and larval growth. *Plant Physiol.* **121**, 497-506.
2. Shen, Y., Jinliang, Y., Wu, Y., Tan, J., Zhou, B., Chen, J. *et al*. (1993). Comparison of two monotoring methods for pyrethroid resistance in cotton bollworm (Lepidoptera: *Noctuidiae*). *Resistant Pest Manage.* **5**, 5-7.
3. Terra, W. R. & Ferreira, C. (1994). Insective digestive enzymes: properties, comparmentalization and function. *Comp. Biochem. Physiol. B,* **109**, 1-9.
4. Billingley, P. F. (1990). Blood digestion in the mosquito, *Anopheles stephensi* Liston (Diptera: *Culicidae*): partial characterisation and post feeding activity of midgut aminopeptidase. *Arch. Insect Biochem. Physiol.* **15**, 149-163.
5. Bown, P. B., Wilkinson, H. S. & Gatehouse, J. A. (1998). Midgut carboxypeptidase from *Helicoverpa armigera* (Lepidoptera: *Noctuidae*) larvae: enzyme characterisation, cDNA cloning and expression. *Insect Biochem. Mol. Biol.* **28**, 739-749.
6. Vendrell, J., Querol, E. & Aviles, F. X. (2000). Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. *Biochim. Biophys. Acta,* **1477**, 284-298.
7. Lenz, C. J., Kang, J., Rice, W. C., McIntosh, A. H., Chippendale, G. M. & Schubert, K. R. (1991). Digestive proteinases of larvae of the corn earthworm, *Heliothis zea:* characterisation, distribution, and dietary relationships. *Arch. Insect Biochem. Physiol.* **16**, 210-212.
8. Ortego, F., Novillo, C. & Castanera, P. (1996). Characterisation and distribution of digestives proteases of the stalk corn borer, *Sesamia nonagrioides Lef*. (Lepidoptera: *Noctuiiae*). *Arch. Insect Biochem. Physiol.* **32**, 163-180.
9. Ferriera, C., Capella, A. N., Sitnik, R. & Terra, W. R. (1994). Properties of the digestive enzymes and the permeability of the peritrophic membrane of *Spodoptera frugiperda* (Lepidoptera) larvae. *Comp. Biochem. Physiol.* **107**, 631-640.
10. Botos, I., Meyer, E., Nguyen, M., Swanson, S. M., Koomen, J. M., Russell, D. H. *et al*. (2000). The structure of an insect chymotrypsin. *J. Mol. Biol.* **298**, 895-901.
11. García-Sáez, I., Reverter, D., Vendrell, J., Avilés, F. X. & Coll, M. (1997). The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of its inhibition, activation and intrinsic activity of the zymogen. *EMBO J.* **23**, 6906-6913.
12. Coll, M., Guasch, A., Aviles, F. X. & Huber, R. (1991). Three-dimensional structure of porcine procarboxypeptidase B: a structural basis of its inactivity. *EMBO J.* **10**, 1-9.
13. Guasch, A., Coll, M., Huber, R. & Aviles, F. X. (1992). Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. *J. Mol. Biol.* **224**, 141-157.
14. Gomis-Rüth, F. X., Gomez, M., Bode, W., Huber, R. & Aviles, F. X. (1995). The three-dimensional structure of the native ternary complex of bovine pancreatic procarboxypeptidase A with proproteinase E and chymotrypsinogen C. *EMBO J.* **14**, 4387-4394.
15. Schechter, I. & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162.
16. Christianson, D. W. & Lipscomp, W. N. (1989). Carboxypeptidase A. *Accts Chem. Res.* **22**, 62-69.
17. Kim, H. & Lipscomb, W. N. (1990). Crystal structure of the complex of carboxypeptidase A with a strongly bound phosphonate in a new crystalline form: comparison with structures of other complexes. *Biochemistry,* **29**, 5546-5555.
18. Brünger, G. J. (1993). *XPLOR, Version 3.1. A System for X-ray Crystallography and NMR*, Yale University Press, New Haven, CT.
19. Reverter, D., Fernandez-Catalan, C., Baumgartner, R., Pfänder, R., Huber, R. *et al*. (2000). Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nature Struct. Biol.* **7**, 322-328.
20. Bukrinsky, J. T., Bjerrum, M. J. & Kadziola, A. (1998). Native carboxypeptidase A in a new crystal environment reveals a different conformation of the important tyrosine 248. *Biochemistry,* **37**, 16555-16564.
21. van Aalten, D. M., Chong, C. R. & Joshua-Tor, L. (2000). Crystal structure of carboxypeptidase A complexed with D-cysteine at 1.75 Å - inhibitor-induced conformational changes. *Biochemistry,* **39**, 10082-10089.
22. Rees, D. C. & Lipscomb, W. N. (1980). Structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *Proc. Natl Acad. Sci. USA,* **77**, 4633-4637.
23. Vendrell, J., Cuchillo, C. M. & Aviles, F. X. (1990). The tryptic activation pathway of monomeric procarboxypeptidase A. *J. Biol. Chem.* **265**, 6949-6953.

24. Reverter, D., Ventura, S., Villegas, V., Vendrell, J. & Avilés, F. X. (1998). Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. *J. Biol. Chem.* **273**, 3535-3541.
25. Ventura, S., Villegas, V., Sterner, J., Larson, J., Vendrell, J., Hershberger, C. L. *et al*. (1999). Mapping the pro-region of carboxypeptidase B by protein engineering. Cloning, overexpression, and mutagenesis of the porcine proenzyme. *J. Biol. Chem.* **274**, 19925-19933.
26. Fitt, G. P. (1989). The ecology of the *Heliothis* species in relation to agroecosystems. *Annu. Rev. Entomol.* **34**, 17-52.
27. Bown, D. P., Wilkinson, H. S. & Gatehouse, J. A. (1997). Differentially regulated inhibitor-sensitive and insensitive protease genes from the phytophagous insect pest, *Helicoverpa armigera*, are members of complex multigene families. *Insect Biochem. Mol. Biol.* **27**, 625-638.
28. Otwinowski, Z. (1993). SERC. Daresbury Laboratory, pp. 52-62, , England.
29. Navaza, J. (1994). An automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157-163.
30. Roussel, A. & Cambillau, C. (1989). TurboFRODO . In *Silicon Graphics Geometry Partner Directory*, Silicon Graphics, Mountain View, CA.
31. Collaborative Computational Project No. 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.
32. Engh, R. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A,* **47**, 392-400.
33. Laskowski, R., MacArthur, M., Hutchinson, E. & Thorton, J. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
34. Evans, S. V. (1990). SETOR: hardware lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134-138.
35. Nicholls, A., Bharadwaj, R. & Honig, B. (1993). GRASP - graphical representation and analysis of surface properties. *Biophys. J.* **64**, A166.

Edited by D. Rees

(Received 11 June 2001; received in revised form 27 August 2001; accepted 6 September 2001)