CMLS, Cell. Mol. Life Sci. 62 (2005) 1996–2014
1420–682X/05/171996-19
DOI 10.1007/s00018-005-5174-4

CMLS Cellular and Molecular Life Sciences

Research Article

Detailed molecular comparison between the inhibition mode of A/B-type carboxypeptidases in the zymogen state and by the endogenous inhibitor latexin

R. García-Castellanos [a], R. Bonet-Figueredo [b], I. Pallarés [b], S. Ventura [b], F. X. Avilés [b], J. Vendrell [b, *] and F. X. Gomis-Rüth [a, *]

[a] Institut de Biologia Molecular de Barcelona, C.I.D. – C.S.I.C., C/Jordi Girona 18–26, 08034 Barcelona (Spain)
[b] Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Facultat de Ciències, Universitat Autònoma de Barcelona, 08193 Bellaterra (Spain)

Received 28 April 2005; received after revision 1 June 2005; accepted 29 June 2005
Online First 9 August 2005

Abstract. Treatment of advanced stages of prostate carcinoma with histone-deacetylase inhibitors entails expression of human procarboxypeptidase-A4 (hPCPA4). The three-dimensional structure of hPCPA4 has been solved and shows the features of related metallocarboxypeptidase zymogens, with a preformed α/β-hydrolase active-enzyme moiety (hCPA4) and an inhibiting pro-domain (PD). The protease moiety recalls a sphere, out of which a spherical cone has been cut. This results in a funnel-like structure, at the bottom of which the active-site cleft resides. The border of this funnel is shaped by loops, which are responsible for the interaction with the PD, characterised by a large interface area and relatively few contacts. Such an inhibitory mode is evocative of the recently reported structure of the human inhibitor latexin in its complex with hCPA4. The main contacting structure of latexin is similar to the one employed for PD inhibition. In both cases, active-site blocking relies mainly on a loop provided by the central part of a β sheet.

Key words. Metallocarboxypeptidase; X-ray crystal structure; carboxypeptidase inhibitor; latexin.

Cancer, with more than seven million casualties in 2002, is the main cause of mortality before the age of 65 despite the overall advances in the understanding of the molecular basis underlying the pathology, and the development of new therapeutic approaches. Prostate cancer ranks sixth in the causes of death from malignant neoplasms in men [1]. The two lines of action are early detection to eliminate or prevent progression of the cancer through appropriate treatment and the search for the switches that turn cells malignant.

* Corresponding author e-mail: xgrcri@ibmb.csic.es; fax: +34 932 045 904 and e-mail: josep.vendrell@uab.es; fax: +34 935 811 264. R. Garcia-Castellanos and R. Bonet-Figueredo contributed equally to this study and share first authorship.

Within the prostate, androgens regulate both cell proliferation and the rate of epithelial cell death. Therefore, testosterone ablation is currently employed during therapy for advanced prostate carcinoma, as it may induce apoptosis in these tumours. However, cancer cells resistant to androgen ablation therapy may emerge within 2 years [2]. In these cases, an alternative is provided by differentiation or apoptosis therapy, which takes advantage of the capacity of resistant tumour cells to differentiate and to activate the apoptotic cascade upon injury caused by exogenous agents such as butyrates [3]. These compounds are inhibitors of histone deacetylases that affect histones as well as transcription factors, promoting selective transcriptional activation by DNA relaxation or repression relief [4]. Due to their growth arrest properties in cell prolifera-

tion models, inhibitors of histone deacetylases are currently under study as drugs against several cancer types, including prostate cancer [5]. Apoptosis therapy treatment of androgen-independent prostate cancer cell lines revealed that a new gene product, human procarboxypeptidase A4 (hPCPA4), the zymogen of a novel metallocarboxypeptidase (MCP), was up-regulated. Sodium butyrate or trichostatin A, a potent and specific inhibitor of histone deacetylase, was responsible for induction of hPCPA4 mRNA expression in the prostate cell lines PC-3, DU145 and BPH1. Induction further required transcriptional activity of $p21^{WAF1/CIP1}$. These findings suggested that *hPCPA4* gene induction was mediated by histone hyperacetylation as a downstream effect [3]. Northern hybridization and PCR analyses revealed low or no hPCPA4 expression in normal human tissues, including pancreas, prostate, ovary and testis. However, expression was found, in both mice and humans, in hormone-regulated pregnant uterus and fetal tissue. The gene was found to be located in the chromosome 7q32 imprinting domain and was thus suggested as a strong candidate for prostate cancer aggressiveness, possibly through its modulation of the function of peptide hormones required for growth or differentiation of prostate epithelial cells [3, 6].

MCPs (clan MC, family M14; see MEROPS database [7]) are zinc-dependent exopeptidases that excise C-terminal amino acids from protein or peptide substrates [8]. MCPs can be classified into two subfamilies, the A/B (M14A according to MEROPS) and the N/E forms (M14B), previously referred to as pancreatic and regulatory carboxypeptidases (CPs), respectively [9]. A third subfamily (M14C) is characterised by γ-D-glutamyl-(L)-meso-diaminopimelate peptidase I, distinct from glutamate carboxypeptidase or CPG (family M20). A/B-MCPs are digestive enzymes synthesised in the exocrine pancreas of mammals and include the first metalloprotease and the second zinc enzyme to be identified [10]. These enzymes break down peptides in the gut in a complementary manner to the serine endopeptidases, trypsin, elastase and chymotrypsin, in order to deliver single amino acids and dipeptides that can be absorbed [11–13]. Molecular prototypes are the carboxypeptidases A from bovine (bCPA), originally carboxy-polypeptidase [11], and B from porcine (pCPB), alias protaminase [13]. They show preference for C-terminal hydrophobic and basic side chains, respectively [14]. Mammal A-type MCPs were later split into CPA1 and CPA2, with the former preferring smaller aromatic and aliphatic residues and the latter the bulkier tryptophan side chain [15]. In recent years, A/B-MCPs have been found not just in mammals but in all kingdoms of life, from archaea and bacteria, protozoa, fungi, nematodes, insects and other invertebrates to plants, amphibians and birds [10]. As new spheres of action were found, functional and local ascription of A/B-MCPs has moved away from the mere proteolysis of intake proteins as a complement to the endopeptidases in the digestive tract [15]. A/B-MCPs have been found in plasma, brain, heart, stomach, colon, testis, liver and lung, as well as in the serum of patients with acute pancreatitis [15, 16]. They intervene in hormone-regulated tissue growth or differentiation (CPA4 [3]), in fibrinolysis inhibition and bradykinin activation in blood serum or in regulation of peptide hormone activity (CPU, alias thrombin-activatable fibrinolysis inhibitor (TAFI) [17]), and intracellularly in mast cells, dealing with the cellular response or complementing the serine endopeptidase chymase (mast cell CPA3 [18]).

Pancreatic A/B-MCPs are secreted as inactive zymogens [19] encompassing a 90/95-residue N terminal independently folding pro-domain (PD) and an already preformed active CP moiety. The PD acts as a chaperone to assist with the folding of the active CP domain [9] and also prevents access of larger substrates to the active-site cleft of the enzyme. PCPA bears residual activity against smaller substrates [20–22] but PCPB apparently does not [23, 24]. Activation occurs through limited proteolysis in a connecting segment at the end of the PD through, e.g. trypsin [19, 25]. This liberates the active CP from its pro-segment, which in the case of PCPA may act as an autologous inhibitor in trans [26]. Some heterologous MCP protein inhibitors have been reported, including the archetypal potato carboxypeptidase inhibitor (PCI), the orthologue from tomato (MCPI), inhibitors from the intestinal parasite *Ascaris suum* and the medical leech *Hirudo medicinalis* (LCI) [for a review, see ref. 9], and, most recently, the inhibitor from the tick *Rhipicephalus bursa* [27].

To date, however, the only specific endogenous inhibitor present in mammals is latexin, also known as tissue or endogenous carboxypeptidase inhibitor (ECI). This 222-residue protein (in humans) bears no sequence relationship with any structurally characterised protein. It was discovered in rodents, in which it has a widespread tissue expression pattern, by two independent research lines. First, it is expressed in neural tissues, such as in the lateral neocortex of rats (hence latexin [28, 29]), and is considered a marker of regionality and development in rodent central and peripheral nervous systems [28, 30]. It has also been suggested that it intervenes in the modulation of sensory perception and it is down-regulated in the brain of presenilin-1 knock-out mice, thus possibly contributing to Alzheimer's disease [31]. Second, other studies identified latexin as an inhibitor of CPA, responsible for the lack of proteolytic activity despite the presence of mRNA encoding CPA in a series of non-pancreatic tissues [16]. An animal model showed that latexin expression is also induced in rat acute pancreatitis, besides in inflammatory lung disease [32]. The tissue distribution of CPA and latexin correlate well in the rat [33]. Latexin has also been shown to play a role in inflammation, as it is expressed in rat mast cells and inhibits mast cell CP. It is

highly expressed upon induction in macrophages, together with other MCPs and cystatins, which are cysteine protease inhibitors [28, 29]. Latexin is also widely expressed in humans, though the tissue distribution is different: here, expression is high in heart, prostate, ovary, kidney, pancreas and colon and only moderate in brain [33–35].

Several three-dimensional (3D) structures are available for vertebrate and insect CPs of the N/E [36] and the A/B subfamilies, in their active or pro-enzyme form [37–43], as well as in a complex with protein inhibitors [44, 45]. We report here for the first time the structural analysis of hPCPA4 as well as additional experimental details and features of the complex of hCPA4 with latexin, on which a research report has been recently published [46]. As a protein-degrading enzyme potentially overexpressed in prostate cancer, the knowledge of the atomic structure of hCPA4, as well as the structural determinants of its inhibition within the zymogen and through an endogenous protein inhibitor, may provide another piece of the complex puzzle of the pathology.

Materials and methods

Recombinant overexpression and purification of hPCPA4

The cDNA for hPCPA4, containing the full pro-enzyme sequence, was kindly provided by D. I. Smith and H. Huang (Mayo Clinic, Rochester, Minn.). Heterologous overexpression and production of recombinant protein was performed in yeast as described previously for other PCPs [47, 48]. Accordingly, the hPCPA4 cDNA was cloned, expressed and secreted to the extracellular medium using vector pPIC9 of the methylotrophic yeast *Pichia pastoris*. pPIC9 provides the α-mating factor signal for secretion and the HIS4 gene for selection of the recombinant yeast clones. *P. pastoris* was transformed by the spheroplast method and histidine-independent transformants were selected and tested for production of hPCPA4. For analytical expression tests, single colonies were grown in 15 ml of BMGY medium (1% yeast extract/2% peptone/90 mM potassium phosphate pH 6.0/1.3% yeast nitrogen base/0.0004% biotin/1% glycerol) for 2 days at 30ºC during the growth phase, collected by centrifugation and gently resuspended in 3 ml of BMMY medium (same as BMGY medium but containing 1% methanol instead of glycerol) during the induction phase. The cultures were grown for another 2 days and the production of protein was monitored after 6, 24 and 48 h by SDS-PAGE. The identity and correct processing of the recombinant protein were confirmed by N terminal sequencing and its function was checked with the synthetic substrate N-(4-furylacryloyl)-Phe-Phe-OH (FAPP) after complete activation of the pro-enzyme with bovine trypsin at a ratio of 400:1 (w/w) for 2 h at 25ºC. After selection of the most productive clones, preparative expression was done in 1L shake-flasks. Cultures were grown for 2 days in BMGY (growth phase) and for a further 2 days in BMMY (induction phase), with an intermediate concentration step to 20% of the initial volume. At the end of the induction phase, the supernatant was separated from the cells by centrifugation and, after equilibration with ammonium sulphate at 30% of saturation, loaded onto a hydrophobic interaction column (Toyopearl butyl 650 M) and eluted with a decreasing gradient (from 30 to 0%) of ammonium sulphate. The elution of the protein was followed by absorbance measurement and the zymogen-containing fractions were selected after SDS-PAGE analysis. These fractions were further purified by FPLC using an anion exchange column (TSK-DEAE 5PW) with an increasing gradient (from 0 to 0.8 M) of ammonium acetate. The purified protein was dialysed against 50 mM Tris-HCl pH 7.5 and concentrated to ~5 mg/ml by Amicon centrifugal filter devices.

Recombinant overexpression and purification of human latexin

As previously published [46], the human latexin/ECI nucleotide sequence was amplified from human brain first-strand cDNAs by RT-PCR using sense (5'-GGAATGGAAATCCCGCCGACCAACTACC-3') and antisense (5'-GGCATCTAGCTTATGAGGCCAAATAATCCCA-3') oligonucleotide primers, designed on the basis of the genomic human latexin sequence (GenBank access code NM 020169). The PCR reaction was carried out in a 100-µl reaction mixture using Taq DNA polymerase. Amplification conditions comprised 30 cycles of 1 min at 94ºC, 1 min at 50ºC, and 1 min at 72ºC. The final reaction mixture was placed on a 1.2% agarose gel, separated by electrophoresis, and viewed by means of ethidium bromide staining. A discrete 762-base pair (bp) PCR product was excised from the gel, and the DNA fragment therein was isolated by spin filtration. A nested PCR using the same sense oligonucleotide and an internal antisense primer (5'-GCCGTCTGCCAAAGGAAGTACAACTGGAATAA-3') was performed, with the first PCR product as a template to ensure product specificity. Amplification conditions were as before, except for an annealing temperature of 60 instead of 50ºC. In this way, a 673-bp product encoding the complete latexin protein was amplified, isolated and subcloned into a T/A cloning vector for further manipulation. The cDNA sequence was verified in both senses. The cloned human brain latexin sequence was amplified by PCR and subcloned into the *Nco*I and *Eco*R1 restriction sites of the prokaryotic expression vector pGAT2 (EMBL, Heidelberg, Germany). In this vector, the target sequence is fused in frame downstream from the gene encoding glutathione-S-transferase (GST), which is preceded by a polyhistidine-tag (His-tag). Vectors containing the construct were transformed into com-

petent BL21(DE3) *Escherichia coli* cells. For protein expression, bacterial cultures were grown in 2YT medium (16 g tryptone/10 g yeast extract/5 g NaCl per litre of water) at 37 ºC to an OD of 0.8. The temperature was subsequently lowered to 18 ºC, and the culture induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) for 24 h. Cells were harvested by centrifugation, resuspended in PBS buffer (140 mM NaCl/2.7 mM KCl/8 mM Na_2HPO_4/2 mM KH_2PO_4/pH 7.4), treated with Triton X-100, DNAse and RNAse, and sonicated. The protein was finally recovered in the supernatant fraction after centrifugation at 17,000 rpm (35,000 g) and loaded onto a Chelating Sepharose Fast Flow (Amersham Biosciences) nickel-sepharose His-tag affinity column equilibrated with 5 mM imidazole/0.5 M NaCl/20 mM Tris-HCl/pH 7.9. After the binding step, the column was washed with 60 mM imidazole/0.5 M NaCl/20 mM Tris-HCl/pH 7.9 and the His-GST-latexin construct was eluted with a linear gradient (from 0.1 to 1 M imidazole) in 0.5 M NaCl/20 mM Tris-HCl/pH 7.9. The eluted protein was dialysed against PBS buffer before thrombin cleavage of the His-GST fusion through overnight incubation at room temperature (1 unit thrombin per milligram of His-GST-latexin). The resulting mixture was loaded again onto the His-tag affinity column to separate the cleaved latexin from the His-GST fusion construct. Latexin was finally purified by size-exclusion chromatography in 20 mM Tris-HCl pH 8.0 and its purity was assessed by SDS-PAGE and mass spectrometry.

Preparation of the hCPA4/latexin complex

The hCPA4/latexin complex was produced as reported using fresh preparations of both proteins: the imidazole-free His-GST-latexin product obtained after elution from the nickel affinity chromatography and subsequent dialysis, and the hCPA4 product obtained after hydrophobic interaction chromatography, dialysis, and subsequent trypsin activation (at a 200:1 w/w ratio) for 90 min at room temperature. Starting from equimolar amounts of both proteins, the formation of the complex was monitored spectrophotometrically by measuring the loss of hCPA4 activity against the FAPP substrate. The resulting complex was incubated overnight with thrombin (1 unit thrombin/mg His-GST-latexin) to cleave off the His-GST tag and subsequently purified by anion exchange chromatography (TSK-DEAE 5PW) using an FPLC-Äkta system with a linear salt gradient from 0 to 60% of 0.8 M ammonium acetate in 20 mM Tris-HCl pH 8.0 [44]. After its purity had been checked by SDS-PAGE, the complex was desalted and concentrated to ~6–7 mg/ml by Amicon centrifugal filter devices.

Latexin inhibitory assays

The K_i values for latexin against MCPs were calculated by pre-steady-state kinetics ($K_i = k_{off}/k_{on}$), using a very sensitive substrate and a continuous photometric assay. The following chromogenic substrates were used: N-(4-methoxyphenyl-azoformyl)-L-Phe-OH for hCPA1, hCPA2, bCPA, haCPA and hCPA4; N-(4-methoxyphenyl-azoformyl)-Arg-OH for hCPB and hTAFI; and N-(4-furylacryloyl)-Ala-Lys-OH for the hCPN assay. The assays were performed in 50 mM Tris-HCl/0.1 M NaCl/pH 7.5 with a substrate concentration of 100 µM and enzyme concentrations of 2 nM. In this assay, the inhibitor, at a concentration pre-established after different trials with varying amounts of enzyme and inhibitor, is added to a continuously monitored progress curve that follows the generation of product from an enzyme/substrate mixture. Values for k_{on}, k_{off} and subsequently K_i can be derived from mathematical processing of the data. To assess the type of inhibition and to confirm the K_i values obtained previously, steady-state inhibition tests were carried out by measuring the CPA4 activity at various enzyme/latexin ratios and at different substrate concentrations (fig. 1). The substrate used in this case was FAPP.

Crystallisation of hPCPA4, structure solution and refinement

Prism-shaped crystals were obtained by the vapour-diffusion crystallisation method from hanging drops consisting of 1 µl of hPCPA4 (5 mg/ml in 20 mM Tris-HCl pH 7.5) and 1 µl of reservoir solution (0.1 M bicine pH 9.0/ 0.2 M $MgCl_2$/14.5% PEG 4000) after 3 days at 4 °C. The cryoprotecting protocol consisted of soaking crystals in a mixture containing reservoir solution and increasing (5% steps) glycerol concentrations (up to 25% v/v) at 4 °C, allowing the crystals to equilibrate for several minutes after each step. A complete diffraction dataset was

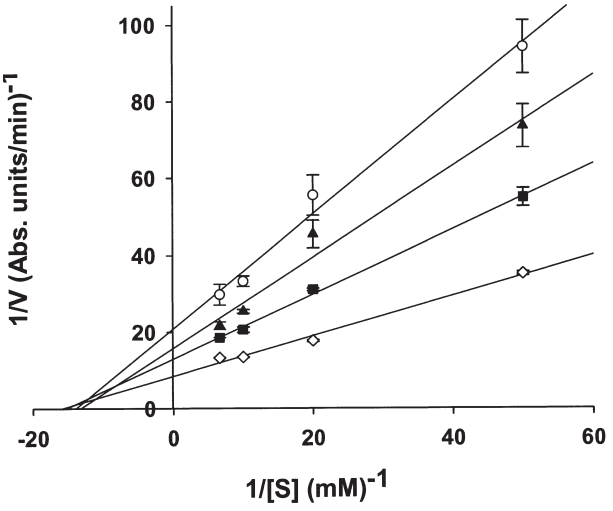

Figure 1. Latexin inhibits A/B-MCPs. Lineweaver-Burk plot for hCPA4 inhibition by latexin/ECI measured at four different hCPA4: ECI ratios. Enzyme concentration: 2nM; latexin concentrations: 0 (◇), 0.94 nM (■), 2.32 nM (▲), 4.70 nM (○). The experimental points shown are the mean of three repeated measurements.

collected at 100 K from a single N_2 flash-cryo-cooled (Oxford Cryosystems) crystal on an ADSC Q210 2D detector at beamline ID29 of the European Synchrotron Radiation Facility (ESRF; Grenoble, France) within the Block Allocation Group 'BAG Barcelona'. Crystals belong to the monoclinic space group C2, harbour two molecules per asymmetric unit, and diffract to 2.2 Å resolution. Crystals displayed such a high mosaicity (eventually refining to 2.5º) that automated indexation procedures employing programs HKL2000, XDS, and MOSFLM failed, although these programs did determine crystal cell parameters. Data could eventually be indexed manually with the latter program using two images on which the crystallographic two-fold axis became apparent. This enabled data processing, scaling, merging, and reduction with the programs MOSFLM v. 6.2.2. [49] and SCALA within the CCP4 suite [50] (see table 1).

The structure was solved by Patterson-search methods with the standalone version of program AMoRe [51], employing data in the 15 to 3.5-Å resolution range. Human PCPA2 (PDB access code 1aye [41]) was used as a searching model as it had the closest sequence similarity (see fig. 2 and Huang et al. [3]). A single solution was encountered for the two molecules present in the asymmetric unit at 91.9,63.1,107.5,0.1874,–0.0002,0.3302 and at 32.9,116.0,286.4,0.8088,0.4046,0.9146 (α,β,γ, in Eulerian angles; x,y,z, as fractional unit-cell coordinates), respectively, after rigid-body fitting employing the routine 'fiting' [51]. This double solution gave a cumulative correlation coefficient in structure factor amplitudes (CC_F) of 46.0% and a crystallographic R_{factor} of 42.1% (for definitions, see table 1 and Navaza [51]; second-highest unrelated peak, CC_F 23.3%, R_{factor} 48.5%). Subsequently, manual model building on a Silicon-Graphics graphic workstation using the TURBO-Frodo program alternated with crystallographic refinement with CNS v. 1.1 [52] and REFMAC5 within CCP4, until the final model was obtained. This model contains protein residues ArgA4 to GlnA99 of the PD and Glu1 to Tyr309 of the mature protease moiety (nomenclature according to pPCPB; see Coll et al. [38] and fig. 2) for each of the two hPCPA4 molecules (chains A and B) present in the asymmetric unit. All residues are placed in allowed regions of the Ramachandran plot except for Ser199 of each polypeptide ($\Phi = 157$, $\Psi = -12$ in Ser199A; 146/–11 in Ser199B), as seen in other carboxypeptidases of the A/B subfamily like hPCPA2 [41]. Each polypeptide chain bears one N-glycosylation attached to Asn148 Nδ2. In molecule A, the (omit) electron density suggested the presence of one N-acetylglucosamine module (Nag901A), while in molecule B two such units were modelled (Nag901B–Nag902B). Each polypeptide chain harbours one disulphide bond (Cys138–Cys161) and three cis-peptide bonds (Ser197–Tyr198, Pro205–Tyr206 and Arg272–Asp273). Three hundred and seventy-four solvent molecules (Hoh501W-Hoh874W) and four (tentatively assigned) glycerol molecules (Gol901W-Gol904W) are also present in the structure.

Crystallisation of the hCPA4/latexin complex, structure solution, and refinement.

As previously indicated, plate-like crystals were obtained for the complex from hanging drops containing 1 µl of protein solution (~7 mg/ml), 1 µl of reservoir solution [40% 2-methyl-2,4-pentanediol (MPD)/0.1 M Bis-Tris pH 6.5], and 0.2 µl 40% acetone after about 4 days at 20 ºC [46]. The presence of both proteins in the crystals was assessed by mass spectrometry. The crystals were flash-cryogenised in liquid N_2 directly out of the drop. The structure of the complex was solved by a combination of Patterson search and multiple-wavelength anomalous diffraction (MAD) at the zinc absorption K-edge. Therefore, a series of complete diffraction datasets was collected at 100 K from a single crystal on a marCCD 225 detector at ESRF beamline ID23–1. The anomalous absorption edge was precisely determined by a fluorescence scan, with which the wavelengths corresponding to the absorption maximum (f''_{max}), the inflection point (f'_{min}) and a high-energy remote were chosen. A further high-resolution dataset was collected from the same crystal (see table 1), which was later subjected to mass spectrometry analysis. The result of this assay was somewhat noisy but clearly revealed the presence of two peaks at 25,312 and 34,649 Da, compatible with the presence of both the inhibitor and the mature protease within the crystal (see figs. 2, 3). The measured crystal belongs to the monoclinic space group $P2_1$, harbours two complexes per asymmetric unit and diffracts beyond 1.6 Å resolution. Data were processed, scaled, merged, and reduced with MOSFLM v. 6.2.2. and SCALA (see table 1).

To calculate initial phases, a Patterson search was performed with AMoRe employing the coordinates corresponding to the protein part of the active enzyme (excluding the catalytic zinc cation) of the hPCPA4 structure and diffraction data in the 12–4.0 Å resolution range. A single solution was encountered for two molecules in the asymmetric unit at 10.8,90.3,325.8,0.0302,0.0001, 0.0652 and 149.6,89.2,146.1,0.9697,0.0451,0.5563, respectively. This solution gave an overall CC_F of 40.6% and a crystallographic R_{factor} of 45.9% (second-highest unrelated peak, CC_F 25.1%, R_{factor} 50.7%). The rotated and translated model was refined against the 1.6-Å resolution dataset and the position of the two catalytic zinc cations was determined by difference Fourier synthesis. These positions were then used to compute experimental phases to 2.0 Å using the three datasets of the MAD experiment and program MLPHARE within CCP4. These phases were combined (σ_A-weighted) with those from the positioned and refined hCPA4 moieties (also computed up to 2.0 Å resolution) and subjected to a density modification

Table 1. Crystallographic statistics on data collection and refinement

Dataset	hPCPA4	hCPA4/latexin (high resolution)[a]	hCPA4/latexin (f'_{min})	hCPA4/latexin (f''_{max})	hCPA4/latexin (remote)
Space group	C2	$P2_1$		$P2_1$	
Cell constants (a, b, c; in Å; β in °)	136.8, 87.2, 89.7, 117.2	78.9, 96.7, 92.4, 99.6		78.9, 96.7, 92.4, 99.6	
Wavelength (Å)	0.9791	1.0067	1.2828	1.2818	1.2770
Number of measurements	158,016	678,157	350,563	350,066	350,334
Number of unique reflections	46,526	179,401	92,181	92,184	92,185
Resolution range (Å) (outermost shell)[b]	62.0–2.20 (2.32–2.20)	48.3–1.60 (1.69–1.60)		60.6–2.00 (2.11–2.00)	
Completeness (%)/anom. completeness (%)	97.9 (95.9)	99.8 (99.3)		99.8 (99.2)/98.6 (96.8)	
R_{merge}[c]	0.094 (0.437)	0.068 (0.257)	0.062 (0.135)	0.057 (0.107)	0.057 (0.083)
$R_{anomalous}$[d]	–	–	0.044 (0.101)	0.044 (0.085)	0.042 (0.063)
Average intensity ($\langle[\langle I\rangle/\sigma(\langle I\rangle)]\rangle$)	10.5 (2.2)	11.8 (4.8)	10.6 (5.4)	13.2 (7.9)	13.8 (10.8)
B-factor (Wilson) (Å^2)	29.0	13.5	19.6	19.5	19.4
Average multiplicity	3.4 (3.0)	3.8 (3.7)		3.8 (3.8)	
Estimated values of f'/f'' (in electrons)	–	–	–9.2/2.3	–7.3/4.7	–4.6/4.9
Resolution range used for MAD phasing (Å)[e]	–	–		48.3–2.00	
Mean figure of merit (fom) MLPHARE/ phase comb./DM[f]	–	–		0.25 (2.0)/0.70 (2.0)/0.83 (1.6)	
Resolution range used for refinement (Å)	50.0–2.20	48.3–1.60			
Number of reflections used (test set)	45,815 (709)	178,706 (678)			
Crystallographic R_{factor} (free R_{factor})[g]	0.221 (0.293)	0.149 (0.178)			
Number of protein atoms (asymmetric unit)	6,440	8,463[h]			
Number of solvent molecules/ions/ other molecules	374/2 (Zn^{2+})/4 (glycerol)	1,153/2 (Zn^{2+})/ 12 (MPD), 1 (acetone)			
R.m.s. deviation from target values					
bonds (Å)	0.011	0.011			
angles (°)	1.33	1.35			
bonded B-factors (Å^2) main-chain/side-chain	0.55/1.55	0.88/2.10			
Average B-factors for protein (Å^2)	35.5	16.1[h]			

Friedel-pairs were treated separatedly in the processing of the MAD datasets.

[a] See also Pallarés et al. [46].

[b] Values in parentheses refer to the outermost resolution shell, unless otherwise indicated.

[c] $R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the i-th intensity measurement of reflection hkl, including symmetry-related reflections, and $\langle I(hkl)\rangle$ is its average.

[d] $R_{anomalous} = \Sigma_{hkl}|\langle I(hkl)\rangle - \langle I(-h-k-l)\rangle|/\Sigma_{hkl}(\langle I(hkl)\rangle + \langle I(-h-k-l)\rangle)$.

[e] The dataset at the inflection point was taken as a reference for phasing.

[f] fom = $|F(hkl)_{best}|/|F(hkl)|$, with $F(hkl)_{best} = \Sigma_\alpha P(\alpha)F_{hkl}(\alpha)/\Sigma_\alpha P(\alpha)$. After experimental phase calculation with MLPHARE, phases were combined (σ_A-weighted) with those from the refined models positioned according to the Patterson search solutions and subjected to density modification, including averaging and phase extension, with the DM program (in parentheses, high-resolution limits in Å of fom value).

[g] $R_{factor} = \Sigma_{hkl}||F_{obs}| - k|F_{calc}||/\Sigma_{hkl}|F_{obs}|$, with F_{obs} and F_{calc} as the observed and calculated structure factor amplitudes; free R_{factor}, the same for a test set of reflections (>500) not used during refinement.

[h] Including atoms present in double occupancy.

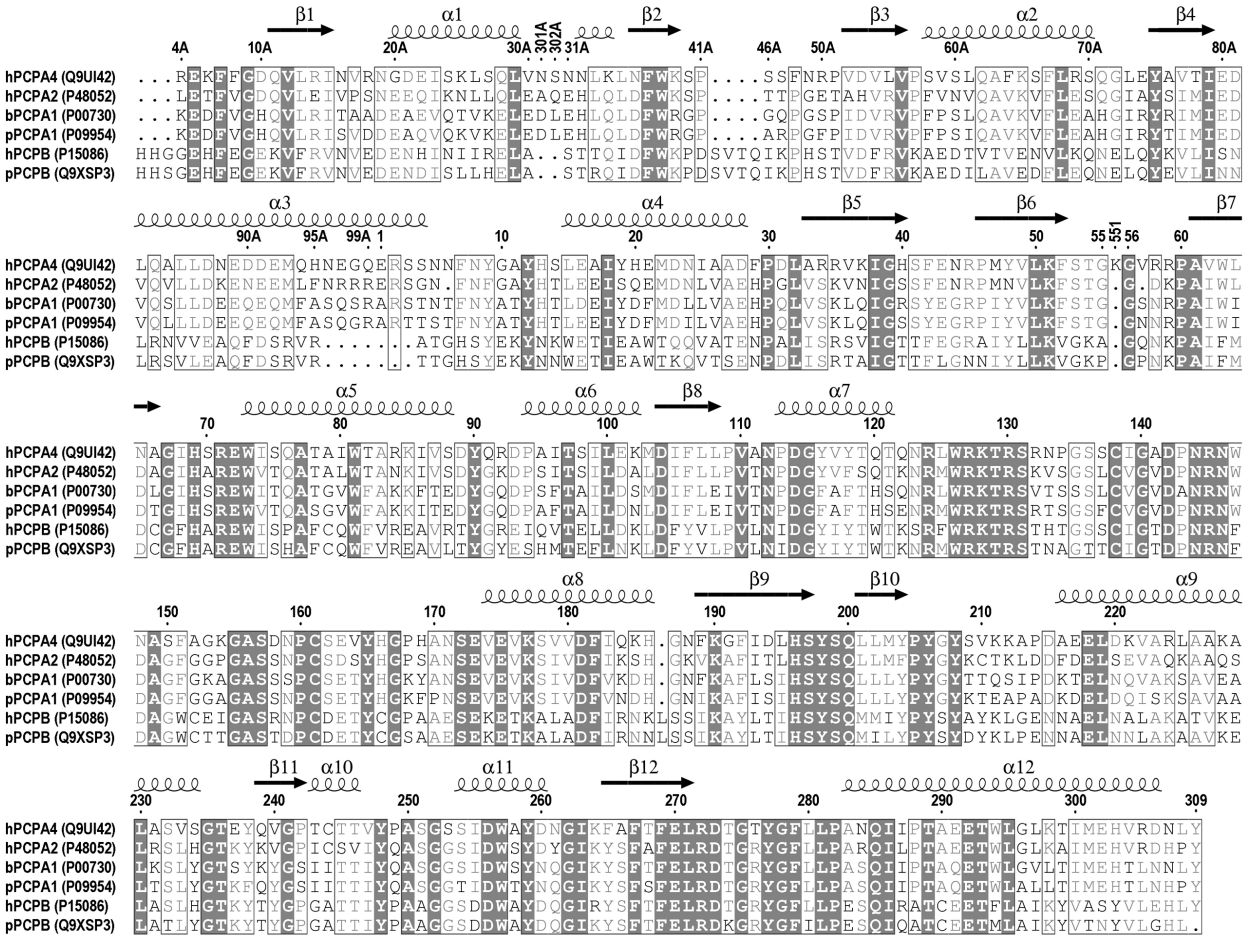

Figure 2. Procarboxypeptidases. Sequence alignment of zymogen members of the A/B-MCP subfamily, whose X-ray crystal structure has been worked out. h stands for human, b for bovine and p for porcine. hPCPA4 shares 63% sequence identity with hPCPA2, 55% with bPCPA1, 54% with pPCPA1, 41% with hPCPB, and 40% with pPCPB. The regular secondary structure elements of hPCPA4 (cringles indicating α helices and arrows for β strands) are shown and labelled (α1–α12 and β1–β12). The numbering corresponds to the same molecule and is based on the nomenclature established for porcine PCPB [38]; this results in insertions (between position 30A and 31A, at the end of the PD and at the end of the enzyme part) and deletions (at the N terminus of the PD, between 41A and 46A of the PD and at position 186 of the enzyme moiety). hPCPA4 has a single insertion between positions 55 and 56 of the active enzyme. The SwissProt/TrEMBL sequence access codes for each protein are given in parentheses. After similarity shading according to MULTALIN/ESPRIPT 2.2, strictly conserved positions are in white over dark grey, and those conserved over 50% (including residues of equivalent nature) are boxed and in light grey.

step (with DM within CCP4) under two-fold averaging and phase extension to the full resolution of the 1.6-Å high-resolution dataset. This produced an electron density map of excellent quality (see fig. 4A), which enabled straightforward tracing of the two complexes. Subsequently, manual model building alternated with crystallographic refinement until the final model, featuring protein residues Ser3 to Leu308 of each of the two hCPA4 molecules (chains A and C) and residues Met1 to Lys217 of the two latexin molecules (chains B and D), was obtained. The polypeptide chains were well defined by electron density, except for loops $\beta1\beta2$ (centred on Pro46B and Pro46D) and loops $\beta3\beta4$ (around Gly80B and Gly80D) of the latexin molecules, which are somewhat flexible at these points. All main-chain angles of the polypeptide chains were placed in permitted regions of the Ramachandran plot. Ser199 of both hCPA4 polypeptide chains A and C showed values close to those in the zymogen, though within a generously allowed region (Φ = 156, Ψ = –8 in Ser199A; 155/–6 in Ser199C). Some residues were modelled with double occupancy of their side chains. One N-acetylglucosamine module (Nag901A and Nag901C) was attached to Asn148 Nδ2 of each hCPA4 moiety. A free-standing valine residue was found at each of the hCPA4 active sites (Val998A and Val998C). Disulphide bonds and cis-peptide bonds of hCPA4 were as in the zymogen. Each latexin chain showed an additional cis-peptide bond (Ile122-Pro123). One thousand one hundred and fifty-three solvent molecules (Hoh1W–Hoh1153W), 12 MPD molecules

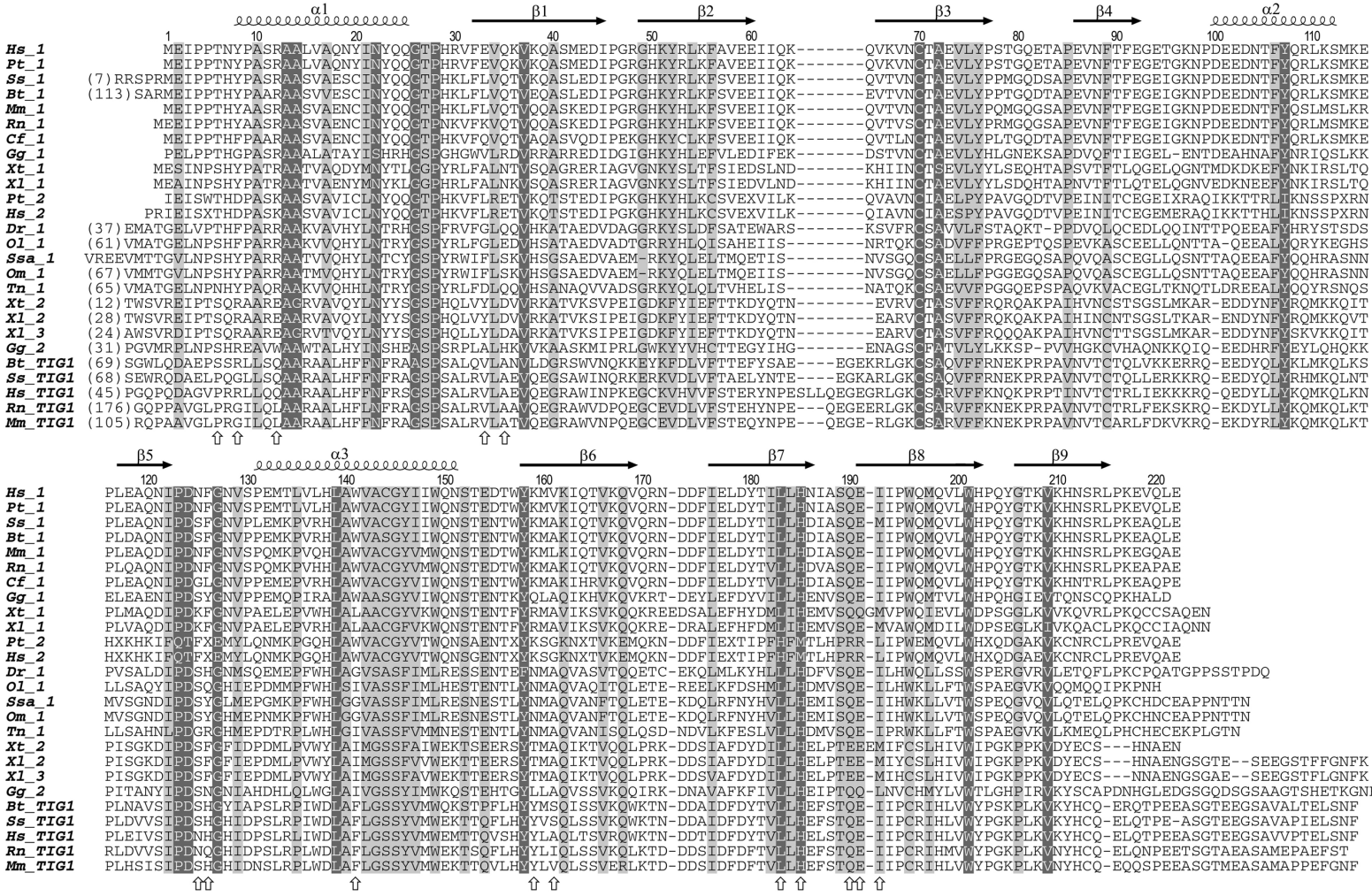

Figure 3. The latexin family. Sequence alignment of members of the latexin family of putative MCP inhibitors retrieved from SwissProt/TrEMBL (www.expasy.ch), GenBank (www.ncbi.nlm.nih.gov), TIGR (tigrblast.tigr.org/tgi) and ENSEMBL (www.ensembl.org). Hs, *Homo sapiens* (human); Pt, *Pan troglodytes* (chimpanzee); Ss, *Sus scrofa* (pig); Bt, *Bos taurus* (cattle); Mm, *Mus musculus* (mouse); Rn, *Rattus norvegicus* (rat); Cf, *Canis familiaris* (dog); Gg, *Gallus gallus* (chicken); Xt, *Xenopus tropicalis* (pipid frog); Xl, *Xenopus laevis* (African clawed frog); Dr, *Danio rerio* (zebrafish); Ol, *Oryzias latipes* (Japanese medaka fish); Ssa, *Salmo salar* (salmon); Om, *Oncorhynchus mykiss* (rainbow trout); Tn, *Tetraodon nigroviridis* (pufferfish). Some sequences are from unfinished genome projects. In these cases, they may still bear errors and include signal and/or pro-peptides; X stands for unidentified amino acids. Hs_2 and Pt_2 are probably processed pseudogens. Similar though incomplete sequences were also identified in opossum, killifish and Japanese pufferfish. Positions conserved in more than 90% of sequences are shown in white over dark grey; those with a maximum of three different amino acid types or with conservative substitutions (all apolar; i.e. aliphatic or aromatic) appear with a light-grey background. The number of additional N terminal protein residues is indicated in parentheses if they occur. Amino acid numbering and regular secondary structure elements correspond to the human latexin structure. Cringles stand for α helices (labelled α1–α3), and arrows for β strands (β1–β9). Small vertical arrows denote residues engaged in interactions with hCPA4 (see table 2).

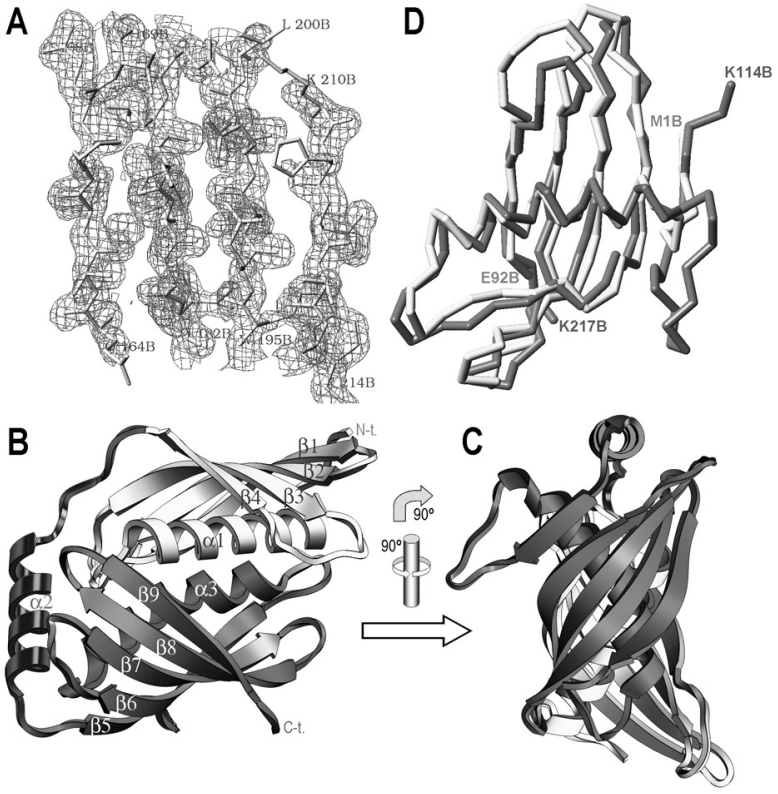

Figure 4. Structure of human latexin. (*A*) Detail of the original experimental electron density (resolution 1.6 Å; contour level 1σ), after phase combination with the Patterson-search solution and density modification, superimposed on the final refined model. The region around the centre of the C-terminal β sheet of latexin is shown, and some residues are labelled. (*B*) Ribbon plot of human latexin, view of the upper barrel surface. The N terminal subdomain (white), the connecting segment featuring helix α2 (black) and the C-terminal subdomain (dark grey) are shown. The constituent regular secondary structure elements are labelled. (*C*) Lateral view of the barrel. (*D*) Superimposition of the N- and C-terminal subdomains of latexin. Same grey tone coding as in (*B*).

(Mpd1L–Mpd12L) and an acetone molecule (Acn13L) were also identified.

Miscellaneous

The structures of both hPCPA4 and the latexin/hCPA4 complex have two copies in their respective asymmetric units. These copies are structurally equivalent, with an rmsd of 0.29 Å for the two hPCPA4 molecules (suffixes A and B) and of 0.27 Å for the two complexes (protease moieties are A and C and the inhibitor chains are B and D). Accordingly, discussion will consider only molecule A of hPCPA4 (residues ArgA4A to GlnA99A of the PD and Glu1A to Tyr309A of the enzyme moiety; nomenclature according to Coll et al. [38] and fig. 2) and the complex consisting of molecule A of hCPA4 (Ser3A to Leu308A) and molecule B of latexin (Met1B to Lys217B).

Figures were prepared with the TURBO-Frodo, GRASP, and SETOR programs; superimpositions and searches for structural similarity were performed with SIM, TURBO-Frodo and DALI; and multiple sequence alignments, with MULTALIN and ESPRIPT 2.2. Of particular help was program DBAli2.0 [53]. Close contacts and interaction surfaces were calculated with CNS. The final coordinates of hPCPA4 and the hCPA4/latexin complex have been deposited with the Protein Data Bank (PDB access codes 2boa and 2bo9) at the European Bioinformatics Institute (www.ebi.ac.uk/msd), Hinxton (UK).

Results and discussion

Inhibition profile of latexin

The results of inhibition studies of mature hCPA4, obtained after tryptic activation, with human latexin are shown in figure 1. The profile corresponds to non-competitive inhibition, since addition of latexin affects only V_{max} and not K_m. However, this is valid for synthetic, small substrates like those used here and in most CP activity determinations. The presence of a huge inhibitor molecule on top of the active site might preclude not only the binding of natural substrate molecules, but even their docking. In this case, the action of latexin in a cellular context is more difficult to define and its mechanism is better considered a mixed inhibition. Since obtaining kinetic data of CPs with natural peptides has always been

difficult, we suggest that, with the substrates and the data presented here, the definition of the action of latexin on CPs as tight-binding non-competitive inhibition is adequate.

As previously reported, all vertebrate A/B-type MCPs tested were inhibited. The kinetic inhibition constants (K_i), calculated by pre-steady-state analysis and confirmed by parallel steady-state measurements, are in the nanomolar range [46] and similar to those obtained with the LCI invertebrate inhibitor from leech [54]. Thus, K_i values calculated for bCPA, hCPA1, hCPA2, hCPB, hTAFI and hCPA4 ranged from 1.1 to 3.5 nM. However, human latexin fails to inhibit not only members of the N/E class like hCPN or duck (d)CPD domain 2, but also an invertebrate A/B-MCP from the cotton bollworm, *Helicoverpa armigera,* haCPA, despite 30% sequence identity of its zymogen with hPCPA4 [A. Bayés, personal communication]. The invertebrate enzyme was more efficiently inhibited by PCI than by LCI, unlike other vertebrate CPs [54]. This suggests further diversification in specificity between vertebrate and invertebrate A/B-MCPs. The inhibition results reported here are in accordance with those described for rat latexin against various MCPs. Those studies, however, established an IC_{50} value 100-fold larger for bCPB and 10-fold larger for mast cell CP than for bCPA [33]. It was also shown that rodent latexin does not inhibit N/E CPs like mouse (m)CPH and hCPM, nor other metalloproteases like gluzincins (aminopeptidases, neprilysin, dipeptidyl peptidase, leukotriene A4 hydrolase) and serine proteases (trypsin, chymotrypsin, elastase, and yeast CPY) [33].

The structure of hPCPA4

hPCPA4 has high sequential similarity with other mammal PCPs (37–63% identity) and somewhat lower similarity (27–43% identity) with non-vertebrate forms, such as those in *E. coli*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae* and *Drosophila melanogaster*. The closest relationship found was with hPCPA1 and hPCPA2, which confirms that it is a new member of the human CPA subfamily [3]. Similar to other PCPs (see fig. 2 and Vendrell at al. [9] for a review), the hPCPA4 zymogen structure consists of an N terminal 96-residue PD and a C-terminal 308-residue mature enzyme moiety. The PD blocks access of larger substrates to the preformed enzyme part, with a buried surface at the interface of 2240 Å^2. It consists of a globular portion (ArgA4A-AspA81A) and an adjacent connecting segment (LeuA82A-Ser4A), comprising a long α helix that includes the activation scissile bond (Arg2A-Ser3A) (fig. 5a). The globular portion adopts a two-layer open-face sandwich topology with a four-stranded antiparallel β sheet on the side facing the mature enzyme and two helices on the opposite side. A β ribbon, consisting of the second and the third strands and the loop in between (segment AsnA36A to LeuA55A; see figs. 2, and 5A), is the main structure responsible for latency. In total, PD establishes 47 interactions below 4 Å with the enzyme moiety, involving 12 hydrogen bonds, ten van-der-Waals interactions and one salt bridge. Among these contacts, a double hydrogen bond of the side chain of AsnA36A with Arg71A Nη1 and Glu163A Oε1 of the enzyme moiety is novel. In all other PCPs studied (see fig. 2), the position of AsnA36A is occupied by an aspartate which establishes a salt bridge, with arginine kept strictly at position 71 of the enzyme. The PD does not completely shield the mature enzyme, thus acting as a non-competitive inhibitor (fig. 1). This allows the access of smaller peptide substrates to the active site, as already noted for other CPA-like PCPs, and accounts for residual CP activity [39, 41].

The mature enzyme has a compact globular shape and recalls a sphere, out of which a spherical cone has been cut, resulting in a funnel-like structure (fig. 5A). The active-site cleft resides at the bottom of this funnel. The domain shows the classical α/β-hydrolase fold of A/B-MCPs [55], with a central mixed parallel-antiparallel eight-stranded β sheet flanked on both sides by several helices (fig. 5A). The active-site residues are close to the C terminus of the central parallel β strands (figs. 5A, 6, 7). The funnel border is shaped by several loops which connect regular secondary-structure elements and which are responsible for interactions with PD and protein inhibitors such as PCI and LCI [44, 45]. The catalytic zinc ion is tetrahedrally coordinated by His69A, Glu72A, His196A, and a catalytic solvent molecule (Hoh501W), further polarised by the side chain of the general base, Glu270A. Further residues traditionally identified as responsible for substrate binding and catalysis [9, 10] are Arg145A, Tyr248A, Arg127A and Glu270A shaping active-site subsite S_1; Arg71A, Ser197A, Ty198A and Ser199A for S_2 and Phe279A for S_3. The terminal carboxylate group of a substrate is fixed by Asn144A, Arg145A and Tyr248A, while the scissile carbonyl group is near Glu270A, Arg127A and the catalytic zinc. Typical CPA-like specificity towards hydrophobic side chains in substrates is accomplished by a hydrophobic S_1 pocket, shaped by the side chains of Met203A, Thr243A, Val247A, Ala250A, Ile255A, Thr268A and Tyr248A.

The human latexin structure and comparison with the mouse orthologue

Human latexin, as seen in its complex with hCPA4 [46], has an elongated shape, maximum dimensions of about 65×40×40 Å (fig. 4) and an α/β topology within a single 222-residue polypeptide chain. The latter five residues are not visible in the electron density maps. The protein can be split into two subdomains, an N terminal subdomain (NTS; from Met1B to Glu92B; see fig. 4B, D) and a C-terminal one (CTS; Lys114B–Lys217B). The sub-

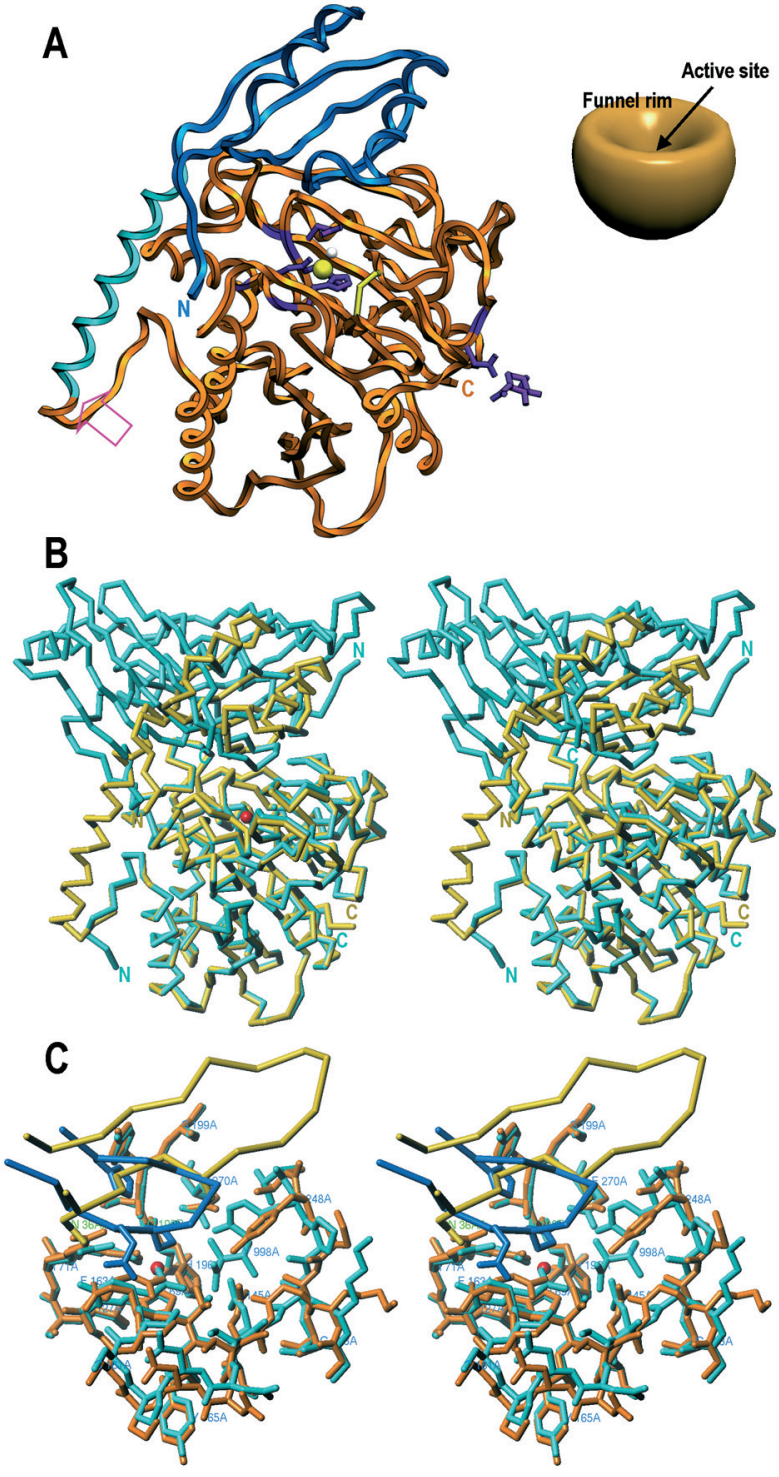

Figure 5. hPCPA4 and comparison with the hCPA4/latexin complex. (*A*) (left) Ribbon plot of hPCPA4, with the PD (globular part in blue, connecting segment in cyan) and the mature enzyme moiety (in orange). The protein residues coordinating the catalytic zinc ion (yellow sphere) are shown in violet, as is the glycosylation site (Asn148A). The catalytic solvent molecule is shown as a white sphere. The activation site is denoted by a magenta arrow. (right) Cartoon illustrating the overall shape of the protease moiety. (*B*) Stereo cartoon of the Cα plots of hPCPA4 (yellow sticks) and the latexin/hCPA4 complex (cyan) after superimposition of the protease moieties. The termini of each molecule are labelled. (*C*) Close-up of the active-site environment of the protease showing segments slightly rearranged on going from the zymogen to the complex. The protease in the complex is shown as cyan sticks (the latexin inhibitory loop as a blue Cα trace) and in the zymogen as orange sticks (inhibition loop of PD as yellow Cα trace). Some residues in the complex are labelled for orientation.

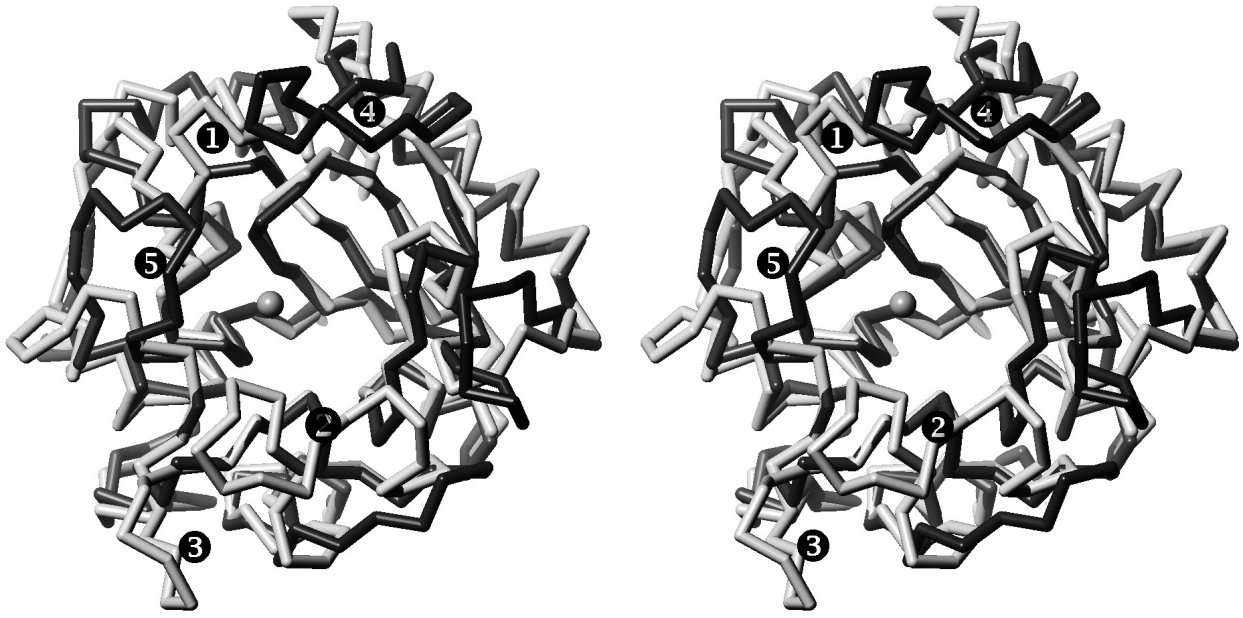

Figure 6. The funnel rims in MCPs. Stereo superimposition of the Cα backbones of hCPA4 (white), as in its complex with latexin, and dCPD domain 2 (black; PDB 1h81), as representatives of A/B- and N/E-type MCPs, respectively. The orientation corresponds to that in figure 5B, rotated clockwise vertically 45º and then horizontally 90º. The regions shown are those shaping the border of the funnel that leads to the active site of the enzymes, pinpointed by the catalytic zinc ion of hCPA4 (grey sphere in the middle). The five main insertions/deletions accounting for lack of interactions (❶ to ❸) or potential steric clashes (❹ to ❺) with latexin, on going from hCPA4 to dCPD domain 2 are: ❶, Thr274A–Tyr277A; ❷, Ser150A–Asn171A; ❸, Ser131A–Ile139A (all of hCPA4); ❹, Gln226–His241; ❺, Ser124–Val133 (both of dCPD domain 2).

domains share an equivalent fold despite minute sequence similarity (14% identity), such that 87 of their Cα atoms can be superimposed with an *rmsd* of 2.1 Å (fig. 4D). The subdomains comprise a long α helix (α1 in NTS, α3 in CTS) and a strongly twisted four-stranded antiparallel β sheet of simple up-and-down connectivity (β1–β4 in NTS, β6–β9 in CTS), which accommodates the helix on its concave side (see fig. 4B). The topology of this ensemble of regular secondary-structure elements recalls a left hand, with the helix mimicking the thumb and the β strands the four fingers (fig. 4D). The β sheet of CTS contains an additional strand, β5, which runs antiparallel to β6 and is sequentially N terminal to the cognate α helix. The side chains of both helices, α1 and α3, are aliphatic, aromatic or uncharged, except for one residue (Arg12B and Glu132B, respectively). This causes these structural elements to interact with their cognate β sheets mainly through hydrophobic contacts. The subdomains are bridged by a connecting segment running along the molecule surface (Gly93B–Met113B). It is mainly constituted by an α helix, α2, which is predominantly charged, with 6 out of 13 residues with formal charges (see fig. 3). The quaternary structure of latexin results from packing the two subdomains against each other through the α helices. These run antiparallel to each other, with their axes ~7.5 Å away and rotated ~50º relative to each other around the point at which Ile21B of α1 and Gly145B of α3 meet. This architecture entails that both curved β sheets are placed on the molecular surface, together forming a flat, incomplete β barrel with two main surfaces, an 'upper' and a 'lower' one. Each surface is shaped by the central parts of helices α1 and α3 on opposite faces and the beginning and ends of the β strands (fig. 4B).

The structure of unbound mouse latexin has been reported recently [29]. The rodent orthologue displays the same number of residues as the human form and shares 85% sequence identity (fig. 3). The Cα backbones of the two proteins fit well and may be superimposed with an *rmsd* of 0.63 Å over 213 residues out of the 217 amino acid positions defined by electron density in the two structures. The mouse structure reported has three additional residues preceding the N terminal methionine due to cloning strategies. The high similarity of the unbound and the bound forms suggests that there is no induced fit of the inhibitor upon complex formation. This is akin to the structure of the target protease moieties, already preformed in the zymogens (see above). The inhibitory loop of latexin and, in particular, Gln190B, is in the same conformation in both structures, as are all the major interaction points with the protease (see below). Minor details worth mentioning that could be attributable to complex formation include position 125B, whose side chain rotates about 260º around its χ_1 angle in the transition from the unbound mouse to the complexed human structure to prevent collision with the side chains of Arg124A and

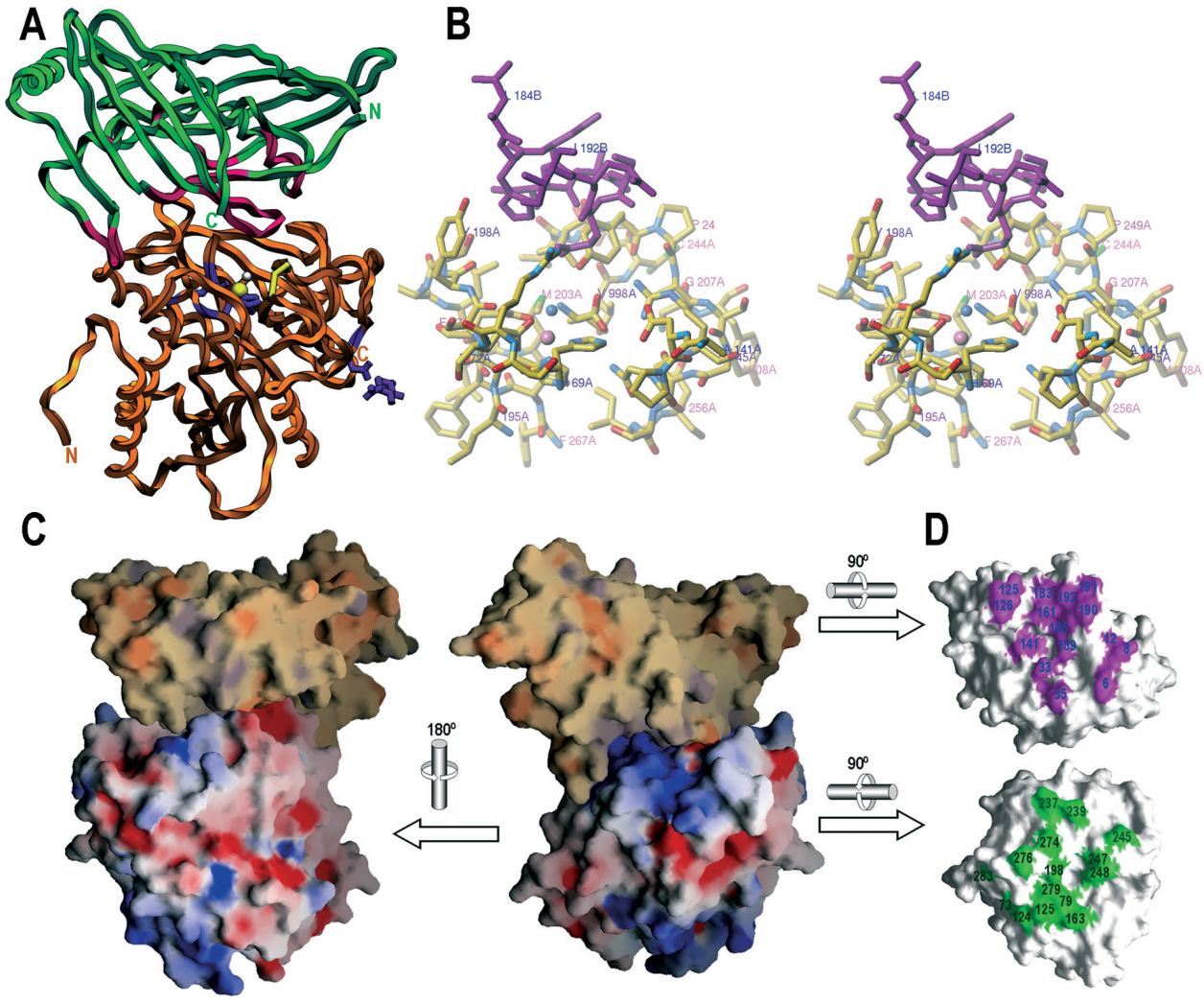

Figure 7. The hCPA4/latexin complex. (*A*) Ribbon plot of the inhibitor/enzyme complex showing hCPA4 in orange and latexin in green, with segments containing residues intervening in intermolecular contacts in magenta. The protease residues coordinating the catalytic zinc ion (yellow sphere) are shown in violet, as is the glycosylation site (Asn148A). The catalytic solvent molecule is shown as a white sphere. The protease moiety has similar orientation to that in figure 5A. (*B*) Close-up in stereo of the section around the catalytic active-site cleft and the region of main peptidase/inhibitor interactions. Latexin is in magenta and hCPA4 is coloured according to atom type. The catalytic zinc ion (violet) and solvent molecule (cyan) are shown as little spheres. (*C*) Lateral view of the latexin/hCPA4 complex, as in *A*, on the right; and a view of the back surface on the left. In both cases, the structure is superimposed with its Connolly surface showing the electrostatic potential [ranging from –10 k_BT/e (red) over 0 k_BT/e (white in hCPA4, beige in latexin) to +10 k_BT/e (blue)]. (*D*) Interacting surfaces in the complex. Residues of the inhibitor (above; magenta) and the protease (below; green) engaged in complex formation are matched on the respective Connolly surface (see also table 2). This view comes from *C* after 90º horizontal rotation of each of the components.

Trp73A of the protease. Glu191B rotates 90º around χ_1, which brings it in contact with the protease main chain at Glu163A N. The remainder is very similar in the two structures and significant deviations can only be found at the tip of loop $\beta3\beta4$ (2.4 Å at Gly80B Cα) and at the beginning of CTS (1.5 Å at Lys114B Cα), possibly due to crystal contacts.

Structural details of the latexin/hCPA4 complex

In the human complex structure, the inhibitor is placed on top of the funnel bordering the active site and traps loop $\beta12\alpha12$ (encompassing residues Asp273A–Pro282A) of the protease through its lower barrel surface at the subdomain interface (fig. 7). This lower surface of the inhibitor has a greater indentation than the upper, shallower surface. This results in a cavity of 235 Å^3, once the complex with hCPA4 is formed, which is partially occupied by an MPD molecule. The complex buries a total surface of 2,340 Å^2 at the protein interface with a shape complementarity of 71%, a value which indicates a good fit between interacting surfaces [56]. The surface is rather large if compared with typical protease-inhibitor complexes, which span between 1250 and 1750 Å^2 [57], and results in several solvent and MPD molecules being en-

Table 2. Interaction scheme between hCPA4 and latexin.

Latexin	hCPA4															
	Arg71-Nη2	Trp73-Cζ3	Arg124-Nε	Leu125-Cδ2	Glu163-N	Tyr198-Oη	Glu237-Oε2	Gln239-Cβ	Thr245-O	Val247-O	Tyr248-Oη	Thr274-Oγ1	Thr274-O	Thr276-Oγ1	Thr276-Cγ2	Phe279-Cζ	Ala283-N
Thr6-Cγ2								3.79									
Tyr8-N									2.90								
Arg12-Nη1										2.94							
Glu33-Oε2												3.17		2.79			
Gln35-Nε2							2.81										
Asn125-O			2.94														
Phe126-O																	3.67
Phe126-Cε2		3.77															
Trp141-Cη2															3.92		
Lys159-Nζ													3.04				
Val161-Cγ1				3.90													
Leu183-Cδ1				3.68													
His185-Nδ1						2.68											
His185-Cε1																3.37	
Gln190-O	2.83																
Gln190-Nε2											2.91						
Gln190-Oε1	2.89																
Glu191-Oε1					2.91												
Ile192-Cγ2				3.95													

The distance (in Å) is shown at the intersection between interacting atoms. There are a total of 47 contacts below 4 Å, with 13 hydrogen bonds and seven hydrophobic interactions.

closed between the protein molecules. However, complex formation entails rather few direct interactions, as only 48 intermolecular contacts below 4 Å are observed, among them 13 hydrogen bonds and 7 hydrophobic interactions (see table 2). The number of hydrogen bonds is within the range of the numbers of previously described protease-inhibitor complexes, which, however, are joined through roughly half the surface area. The interaction involves 15 residues of both hCPA4 and latexin, which again is unlike other protease-inhibitor complexes, in which, characteristically, a similar number of inhibitor residues (10–15) interact with twice as many protease residues [57]. These peculiarities can be attributed to the rather shallow interacting surface of latexin and to the protease mainly interacting though the border of the funnel that gives access to the active site (see above).

The most interesting feature of the interaction is that none of the latexin termini is involved in protease inhibition. This is in contrast to what has been described for other complexes of MCPs with exogenous inhibitors like PCI and LCI, in which a conserved seven-residue C-terminal segment contributes to inhibition in a substrate-like binding manner. Substrate-like binding occurs in other complexes through the N terminal tails of inhibitors, as observed in metalloprotease/inhibitor complexes of matrix metalloproteinases (MMPs) with tissue inhibitors of metalloproteinases (TIMPs) and of serralysins with their cognate inhibitors [44, 45, 58, 59]. Rather, latexin inhibition is mainly caused by an inhibitory loop provided by the central part of the β sheet of the CTS, reminiscent of the inhibition through PD in the zymogen. This loop is shaped by the end of strand β7, the beginning of β8 and the connecting loop β7β8, which penetrates the protease moiety. This behaviour is also found in cystatins, inhibitors of cysteine proteases which also employ a β ribbon structure for inhibition, as seen in the steffin/papain complex [60] (see below). Gln190B Nε2, at the tip of loop β7β8, most closely approaches the active site (5.8 Å away from the catalytic zinc ion). In the latter, a free valine (Val998A), not found in the zymogen and probably left behind after a proteolytic event during purification, is encountered in the specificity pocket (fig. 7B). This finding was described previously for pPCPA1 [39]. Strikingly important for complex stability is the interaction of Gln190B with Arg71A (see table 2). The latter is a basic residue present throughout A/B-MCPs and is required for maintenance of the crucial salt bridge with AspA36 (AsnA36A in PCPA4) of the PDs of PCPs (see above). This basic residue is absent in N/E forms, which are not secreted as pro-enzymes. Gln190B, furthermore, interacts with Tyr248A in 'down' conformation (see below). This tyrosine is a residue involved in hydrogen bonding of the P_1 amide nitrogen of substrates through its Oη group. The position of Gln190B within the inhibitory loop is maintained by an intramolecular interaction with the Nε2 atom of the preceding His185B (2.8 Å). This histidine further interacts via an edge-to-face van-der-Waals contact with the side chains of Phe279A and Tyr198A of the protease moiety. Among the structurally important interactions of the inhibitory loop are also the one established by Glu191B, approaching Glu163A, and by both Ile192B and Leu183B, contacting Leu125A. The position of the inhibitory loop is further fixed through intramolecular main-chain interactions with the neighbouring strands β6 and β9, as well as with loop α3β6. Here, Val161B makes a hydrophobic interaction with Leu125A,

and Lys159B a hydrogen bond with Thr274A. Further complex stabilisation is provided by latexin loop $\beta5\alpha3$ of the CTS, which comes close to the funnel border of the protease. In particular, the tip of the loop (around Phe126B) contacts Arg124A, Trp73A and, weakly, Ala283A (table 2). Two other regions, belonging to the NTS of latexin, participate in contacts, the beginning of helix $\alpha1$ and the preceding residue, and the adjacent central part of strand $\beta1$. Here we observe six interactions (Tyr8B N-Thr245A O; Arg12B Nη1-Val247A O; Thr6B Cγ2-Gln239A Cβ; Glu33B Oε2-Thr274A Oγ1 and -Thr276A Oγ1; Gln35B Nε2-Glu237A Oε2). Last but not least, Trp141B makes an additional isolated hydrophobic contact with the methyl group of Thr276A.

The mode of latexin inhibition described may explain why those CPs with the characteristic PD, the A/B-MCPs, are inhibited by latexin, whereas CPs without PD, e.g. N/E-MCPs like hCPM and dCPD domain 2, are not inhibited. In general, five loops in the regions furnishing the funnel border can be distinguished for either the A/B- or the N/E-MCPs, and potential steric clashes and lack of interactions may justify the selective inhibitory profile of latexin. The absence of a long insertion in hCPA4, Ser150A–Asn171A, reduced to Ala142–Ser149 in hCPM and Gln150–Pro157 in dCPD domain 2 [see fig. 6 and refs. 36, 61], and the adjacent Ser131A–Ile139A, assisting in hCPA4 to strut the previous loop and short-circuited to a single residue in hCPM (Asn131) and dCPD domain 2 (Asn139), dramatically diminish the interaction possibilities. The same occurs in the loop region Thr274A–Tyr277A in hCPA4, absent in hCPM and dCPD domain 2. However, the most important features are two characteristic loop insertions of N/E-MCPs, Lys221–Asn233 (hCPM) or Gln226–His241 (dCPD domain 2), as well as Val116–Ser124 (hCPM) or Ser124–Val133 (dCPD domain 2), which would sterically clash with the inhibitor (figs. 6, 7). In the case of haCPA, shown not to be inhibited by latexin despite being an A/B-MCP (see above), the only structural difference observed with hCPA4 is an insertion in the invertebrate protein between the positions equivalent to Leu271A and Gly278A of the human counterpart. This results in a loop that is four residues longer and that would collide with latexin helix $\alpha3$.

Differences of hCPA4 between hPCPA4 and the latexin complex

The overall structure of hCPA4 in the zymogen and in the inhibitor complex is, in general, equivalent (303 topologically equivalent Cα atoms from Phe7A to Leu308A; rmsd 0.40 Å; see fig. 5B, C). However, closer inspection revealed some noteworthy structural differences. First, the N terminal four-residue segment of the mature enzyme is rotated (mainly around Asn6A C-Cα) away from its position in the zymogen such that the new N terminus, Ser3A N, is anchored to the molecular surface by a triple interaction, with Arg28A Oδ1 (3.2 Å) and Oδ2 (2.8 Å) and with Asn24A Oδ1 (2.9 Å). Other differences arise, on the one hand, from the different interactions made with the PD and latexin and, on the other, from the left-behind free valine (Val998A) mentioned above. This valine, present in the complex but not in the zymogen, causes a rearrangement of the side chain of Tyr248A from an 'up' or 'open' conformation (as in Coll et al. [38]) to a 'closed' or 'down' conformation, as observed in (P)CPs with an 'occupied' pocket [36, 39, 41]. This puts Tyr248A Oη in a position to interact with both a carboxylate oxygen group (2.5 Å) and the free amino group of valine (3.0 Å), besides the mentioned interaction with Gln190B of latexin (see above). This rearrangement of Tyr248A entails also a displacement of the whole polypeptide chain from Val247A to Ser251A (maximal at Pro249A Cα with 1.0 Å; see fig. 5C). In a concerted movement, the adjacent segment Gly153A–Val164A is also shifted (maximal at Gly155A Cα with 1.6 Å). However, this second rearrangement may also be influenced by a contact with the N terminus of a crystallographically related protease moiety. Other minor changes include the side chain of Arg71A, slightly reallocated from the position adopted to exchange a hydrogen bond with AsnA36A in the zymogen to interact with Gln190B of the inhibitor (see above). Also, Arg127A replaces its hydrogen bond with the main-chain carbonyl of Glu163A in the zymogen by a salt bridge with the carboxylate group of the free valine. Finally, the side chain of Arg145A is also rearranged to contact this carboxylate group (see fig. 5C).

Structural similarities of latexin

Interestingly, the overall topology of latexin resembles the structure of serralysin inhibitors [58, 62], which target the serralysin family of bacterial metalloendopeptidases (MEPs), part of the metzincin clan of MEPs [63]. In these cases, the endogenous inhibitors are also β barrels of simple antiparallel up-and-down connectivity, but they are shaped by eight strands and closed, with a helix inserted between the third and fourth strands. All in all, these bacterial inhibitors are about half the size of latexin and their inhibition mode is different. Latexin also has some similarities with cystatins in the way it functions and bioinformatic searches have de facto identified chicken egg-white cystatin (PDB 1cew [64]) and monellin (PDB 1 mol [65]) as the closest structural homologues to latexin (see fig. 8). Monellin is the main agent responsible for the sweetness of plant berries and is among the sweetest substances that have been identified so far. It bears close structural similarity to cystatins. The structural similarity of latexin with the latter is, however, limited to less than half of the complete molecule, as it encompasses only one of the two subdomains, and entails low sequential similarity. In particular, 87 Cα atoms of

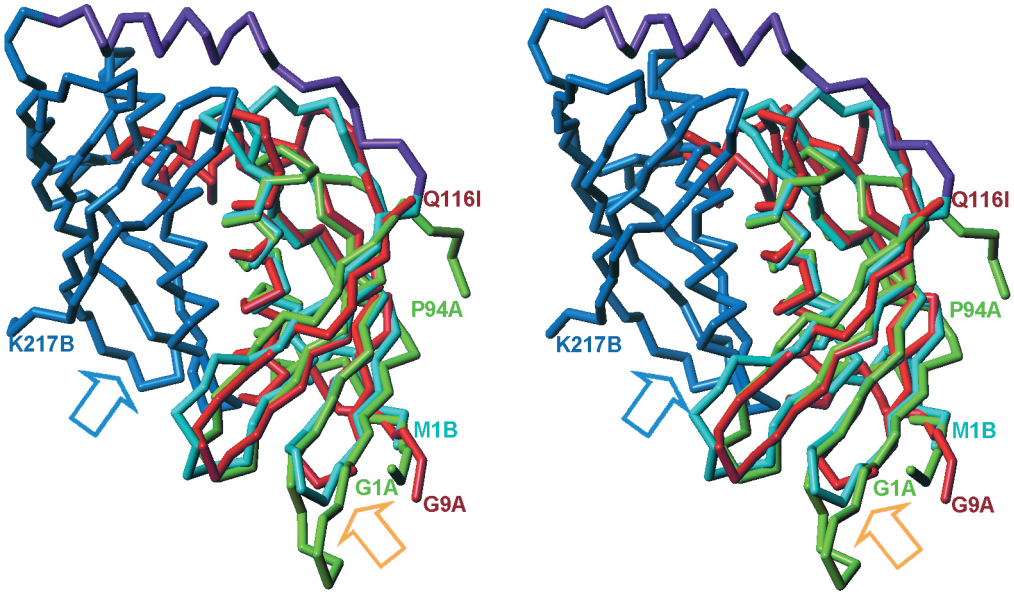

Figure 8. Structural similarities to latexin. Stereo-plot showing the superimposition of human latexin (NTS in cyan, connecting segment in violet, CTS in blue) on chicken egg-white cystatin (red; PDB 1cew [64]) and sweet protein monellin from the serendipity berry (*Dioscoreophyllum cumminsii)* (green; PDB 1 mol [65]). The termini of each chain are labelled. The approximate position of a cysteine proteinase when inhibited by a cystatin (as inferred from the steffin/papain complex; PDB 1stf, [60]) is indicated by a yellow arrow; and the location of hCPA4 in the complex with latexin, by a blue arrow.

cystatin and latexin can be aligned with an rmsd of 2.2 A (16% sequence identity), as can 81 Cα atoms of monellin and latexin (rmsd 2.4 Å; sequence identity 14%). Closer inspection reveals that the structural unit consisting of the α helix and the surrounding β sheet can be superimposed. One main difference is that cystatin displays an additional α helix between the strands equivalent to β2 and β3 of latexin, while monellin shows shorter strands equivalent to β1, β2 and β3. Inhibition of cysteine proteinases through cystatins, as observed in the steffin/papain complex [60], occurs through the tips of the fingers following the left-hand similitude in the fold, through the loop connecting the strands equivalent to β1β2 and β3β4 of latexin NTS (or β6β7 and β8β9 considering CTS), and through the N terminus. In contrast, latexin inhibits by elements participating in the interface between the two subdomains located on the opposite face of the inhibitor (fig. 8). Furthermore, slight conformational differences, mainly due to deletions/insertions, of both previously mentioned loop pairs of NTS and CTS suggest that these should not inhibit cysteine proteinases, a finding experimentally proved in the case of papain [our unpublished results].

Interestingly, two tandem copies of cystatin-like repeats have been predicted for inhibitors of metzincins belonging to the astacin and adamalysin/ADAM families. In particular, the endogenous protein inhibitor of carp nephrosin, an astacin, and two proteins isolated from snake venoms, *Trimeresurus flavoviridis* Habu serum factor and *Bothrops jararaca* BJ46a, inhibitors of snake venom adamalysins/ADAMs, have been described [66]. They are all members of the fetuin family of proteins and, as in the case of latexins, are unable to inhibit cysteine proteases despite structural similarity with cystatins. However, no structural information on their mode of inhibition is currently available.

A novel family of potential A/B-MCP inhibitors

Bioinformatic retrieval within complete and unfinished genome sequence projects has revealed a number of putative proteins which have significant sequence similarity to human latexin (fig. 3). These sequences go beyond mammals to include other vertebrates like frogs, birds and fish. The aligned sequences were compiled from EST libraries from diverse tissues and organs, including kidney, spleen, liver, brain, lung, ovary, testis, aorta, cartilage, placenta, thymus, gut and small intestine. Accordingly, the broad tissue distribution strongly suggests temporal and spatial ubiquity in the transcription of these genes. Inspection of the residues mainly engaged in inhibition suggests that they are mostly highly conserved or conservatively substituted. Therefore, it has recently been proposed that these sequences be grouped in the latexin family, which comprises biochemically characterised inhibitors (from human, rat and mouse, grouped into MEROPS inhibitor family I47 [67]) as well as potential MCP inhibitors [46], and may also harbour structurally related proteins with other functions.

Apart from the closer rodent orthologues, a sequence relative that has been studied at the biochemical level is ovocalyxin-32, a chicken eggshell protein highly ex-

pressed and secreted by surface epithelial cells of the isthmus and the uterus during eggshell formation. Its molecular interacting partners have not been discovered, but it has been ascribed to the matrix [68]. Further similarity can be found with a protein whose expression is induced by retinoids in skin raft cultures, retinoic acid receptor responder protein 1 (RARRES1 alias tazarotene-induced gene 1 protein, TIG1), a putative tumour suppressor gene with various splice variants [69]. Retinoids are known to down-regulate the expression of metalloproteases, cytokines and other genes involved in cell proliferation and inflammation through ligand-dependent transcription factors that are retinoic acid receptors [70]. Thus, RARRES1/TIG1 could be a response to hyper-proliferative and inflammatory processes like skin diseases, putatively acting against MCPs engaged in inflammation. RARRES1/TIG1 orthologues were found in cattle, pig, rat and mouse, with sequence identities ranging from 28 to 31% of human latexin (fig. 3) in a common stretch of ~220 residues. They show additional N terminal residues that may encode membrane anchors [29, 70].

Interestingly, multiple sequences very similar to latexin have been found in several organisms (see fig. 3). In the African clawed frog there are even three potential paralogues, with 49, 34 and 33% sequence identity with human latexin. These sequences could encode putative latexin isoforms. This is reminiscent of TIMPs, for which four different variants have been reported (TIMP-1 to TIMP-4), with multiple presence in several organisms [71].

Conclusions

There are some similarities in the way latexin and the PD block the hCPA4 enzyme moiety, basically a comparably large interface area and relatively few contacts. The main interaction area includes the lower barrel surface of latexin around the CTS β sheet. This arrangement is somewhat reminiscent of the outer face of the four-stranded β sheet of the globular part employed by PD for inhibition. In both cases, active-site blocking occurs with participation of an inhibitory loop provided by the central part of the mentioned β sheets. However, while latexin completely blocks access to hCPA4, thus inhibiting in a non-competitive manner, the activation segment in hPCPA4 does not shield wholly the active-site cleft, which smaller peptide substrates can reach, as occurs in other type-A PCPs, though not type-B ones [41].

The striking structural similarity of latexin subdomains and cystatins suggests that the latter can be seen as evolutionary ancestors. Genetic duplication of such a cystatin module, including the incorporation of a 41-residue linker in between and subsequent sequential divergence may have given rise to the current structure of latexin. The latter, however, employs different regions to perform the common task of inhibiting proteases.

Latexins seem to be restricted to vertebrates and have been found in organisms ranging from fish and frogs to mammals and other higher vertebrates. No sequences could be retrieved from invertebrates or other less complex organisms. Invertebrate A/B-MCPs and N/E subfamily MCPs are not inhibited, as happens with several endopeptidases studied. Our molecular studies show that structural determinants simultaneously confer vertebrate-linked specificity to latexins and sufficient variability to strongly inhibit all vertebrate forms tested, even across the species barrier.

There is a parallel evolution in the studies of A/B-MCPs and MMPs. These MEPs were found because of their non-specific degradation of extracellular matrix components in vertebrates, as occurs during tissue turnover and developmental and metamorphic processes, but also during tissue destruction in pathologies such as cancer and arthritis. They were located in the extracellular space where these events take place. However, decades of work led to a re-examination of the roles of MMPs as they were observed in a specific context to be engaged in limited proteolysis events of single or few peptide bonds. Furthermore, they were expressed in several organs and tissues and also in non-vertebrates, even possibly in bacteria and viruses [63]. Examples of these new selective functions are the ectodomain shedding of non-matrix substrates like growth factors and growth-factor-binding proteins to liberate membrane-anchored substrates and thus facilitate autocrine and paracrine interactions. Other substrates include pro-proteinases and clotting factors that must be activated, proteins regulating apoptosis, and others engaged in cell migration and in intercellular communication [71]. Likewise, the perceived importance of A/B-MCPs has evolved from indiscriminate protein crunchers to sophisticated regulators of hormone homeostasis or fibrinolyisis, as well as of other important functions associated with selective limited proteolysis. MMPs are inhibited by their endogenous tissue inhibitors, the TIMPs, which show complementary spatial and temporal expression patterns. The putative latexin member ovocalixin-32 was identified as a matrix protein, like TIMPs and MMPs. All these findings suggest that latexins are the endogenous inhibitors of A/B-MCPs in vertebrates, as TIMPs are of MMPs.

Acknowledgements. This study was supported by the following grants: BIO2003-00132, GEN2003–20642 and BIO2001–2046, from the Spanish Ministry for Science and Education (formerly Science and Technology); ON03–7-0 from the 'Fundació La Caixa'; FP6 Integrated Project LSHC-CT-2003–503297 'CANCERDEGRADOME' and 2001SGR-00346 from the National Government of Catalunya. I. P. is recipient of a postgraduate fellowship from DGR (Generalitat de Catalunya); both R. B. and R. G.-C. acknowledge FPI Ph.D. fellowships from the Spanish Ministry for

Science and Technology. We are grateful for the help provided by EMBL and ESRF synchrotron local contacts during data collection (ESRF, Grenoble). For these data collections, funding was provided by ESRF and by a grant from the European Union (Access to Research Infrastructures – contract No. RII3-CT-2004–506008). R. Rycroft and X. Puente are thanked for helpful contributions to the manuscript, and F. Canals for carrying out the mass spectrometry analysis.

1 Jemal A., Tiwari R. C., Murray T., Ghafoor A., Samuels A., Ward E. et al. (2004) Cancer statistics, 2004. CA Cancer J. Clin. **54:** 8–29
2 Kyprianou N., English H. F. and Isaacs J. T. (1990) Programmed cell death during regression of PC-82 human prostate cancer following androgen ablation. Cancer Res. **50:** 3748–3753
3 Huang H., Reed C. P., Zhang J. S., Shridhar V., Wang L. and 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
4 Wolffe A. P. (1996) Histone deacetylase: a regulator of transcription. Science **272:** 371–372
5 Vigushin D. M. and Coombes R. C. (2002) Histone deacetylase inhibitors in cancer treatment. Anticancer Drugs **13:** 1–13
6 Bentley L., Nakabayashi K., Monk D., Beechey C., Peters J., Birjandi Z. et al. (2003) The imprinted region on human chromosome 7q32 extends to the carboxypeptidase A gene cluster: an imprinted candidate for Silver-Russell syndrome. J. Med. Genet. **40:** 249–256
7 Rawlings N. D., Tolle D. P. and Barrett A. J. (2004) MEROPS: the peptidase database. Nucleic Acids Res. **32:** D160-D164
8 Willstätter R., Waldschmidt-Leitz E., Harden A., Keilin D., Haldane J. B. S., Quastel J. H. et al. (1932) Discussion on recent advances in the study of enzymes and their action. Proc. R. Soc. Lond. B **111:** 280–297
9 Vendrell J., Querol E. and Avilés F. X. (2000) Metallocarboxypeptidases and their protein inhibitors: structure, function and biomedical properties. Biochim. Biophys. Acta **1477:** 284–298
10 Auld D. S. (2004) Carboxypeptidase A. In: Handbook of Proteolytic Enzymes, vol. 1, pp. 812–821, Barrett A. J., Rawlings N. D. and Woessner J. F. Jr (eds), Elsevier, London
11 Waldschmidt-Leitz E. and Purr A. (1929) Über Proteinase und Carboxy-Polypeptidase aus Pankreas. (XVII. Mitteilung zur Spezifizität tierischer Proteasen.). Ber. Dt. Chem. Ges. **62:** 2217–2226
12 Anson M. L. (1935) Crystalline carboxypolypeptidase. Science **81:** 467–468
13 Waldschmidt-Leitz E., Ziegler F., Schäffner A. and Weil L. (1931) Über die Struktur der Protamine. I. Protaminase und die Produkte ihrer Einwirkung auf Clupein und Salmin. Hoppe-Seyler Z. Physiol. Chem. **197:** 219–236
14 Neurath H. (1960) Carboxypeptidases A and B. In: Enzymes, vol. 4, pp. 11–36, Boyer P.D., Lardy H.A. and Myrback K. (eds), Academic Press, New York
15 Reznik S. E. and Fricker L. D. (2001) Carboxypeptidases from A to Z: implications in embryonic development and Wnt binding. Cell. Mol. Life Sci. **58:** 1790–1804
16 Normant E., Gros C. and Schwartz J. C. (1995) Carboxypeptidase A isoforms produced by distinct genes or alternative splicing in brain and other extrapancreatic tissues. J. Biol. Chem. **270:** 20543–20549
17 Nesheim M. E. (1999) TAFI. Fibrinol. Proteol. **13:** 72–77
18 Miller L. A., Cochrane D. E., Feldberg R. S. and Carraway R. E. (1998) Inhibition of neurotensin-stimulated mast cell secretion and carboxypeptidase A activity by the peptide inhibitor of carboxypeptidase A and neurotensin-receptor antagonist SR 48692. Int. Arch. Allergy Immunol. **116:** 147–153
19 Anson M. L. (1937) Carboxypeptidase. II. Partial purification of pro-carboxypeptidase. J. Gen. Physiol. **20:** 777–780
20 Aberhalden E. and Hanson H. (1938) Carboxypeptidase and acylase. Fermentforschung **16:** 37–47
21 Yamasaki M., Brown J. R., Cox D. J., Greenshields R. N., Wade R. D. and Neurath H. (1963) Procarboxypeptidase A-S6: further studies of its isolation and properties. Biochemistry **128:** 859–866
22 Lacko A. G. and Neurath H. (1970) Studies on procarboxypeptidase A and carboxypeptidase A of the spiny pacific dogfish (*Squalus acanthias*). Biochemistry **9:** 4680–4690
23 Burgos F. J., Salva M., Villegas V., Soriano F., Mendez E. and Aviles 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
24 Wintersberger E., Cox D. J. and Neurath H. (1962) Bovine pancreatic procarboxypeptidase B. I. Isolation, properties, and activation. Biochemistry **1:** 1069–1078
25 Uren J. R. and Neurath H. (1972) Mechanism of activation of bovine procarboxypeptidase A S 5: alterations in primary and quaternary structure. Biochemistry **11:** 4483–4492
26 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: isolation and characterisation of the activation peptide. Biochim. Biophys. Acta **707:** 74–80
27 Arolas J. L., Lorenzo J., Rovira A., Castella J., Avilés F. X. and Sommerhoff C. P. (2004) A carboxypeptidase inhibitor from the tick *Rhipicephalus bursa*: isolation, cDNA cloning, recombinant expression, and characterization. J. Biol. Chem. **280:** 3441–3448
28 Arimatsu Y. (1994) Latexin: a molecular marker for regional specification in the neocortex. Neurosci. Res. **20:** 131–135
29 Aagaard A., Listwan P., Cowieson N., Huber T., Ravasi T., Wells C. et al. (2005) An inflammatory role for the mammalian carboxypeptidase inhibitor latexin: relationship to cystatins and the tumor suppressor TIG1. Structure **13:** 309–317
30 Miyasaka N., Hatanaka Y., Jin M. and Arimatsu Y. (1999) Genomic organization and regulatory elements of the rat latexin gene, which is expressed in a cell type-specific manner in both central and peripheral nervous systems. Brain Res. Mol. Brain Res. **69:** 62–72
31 Mirnics Z. K., Mirnics K., Terrano D., Lewis D. A., Sisodia S. S. and Schor N. F. (2003) DNA microarray profiling of developing PS1-deficient mouse brain reveals complex and coregulated expression changes. Mol. Psychiatry **8:** 863–878
32 Ji B., Chen X. Q., Misek D. E., Kuick R., Hanash S., Ernst S. et al. (2003) Pancreatic gene expression during the initiation of acute pancreatitis: identification of EGR-1 as a key regulator. Physiol. Genomics **14:** 59–72
33 Normant E., Martres M. P., Schwartz J. C. and Gros C. (1995) Purification, cDNA cloning, functional expression, and characterization of a 26-kDa endogenous mammalian carboxypeptidase inhibitor. Proc. Natl. Acad. Sci. USA **92:** 12225–12229
34 Liu Q., Yu L., Gao J., Fu Q., Zhang J., Zhang P. et al. (2000) Cloning, tissue expression pattern and genomic organization of latexin, a human homologue of rat carboxypeptidase A inhibitor. Mol. Biol. Rep. **27:** 241–246
35 Uratani Y., Takiguchi-Hayashi K., Miyasaka N., Sato M., Jin M. and Arimatsu Y. (2000) Latexin, a carboxypeptidase A inhibitor, is expressed in rat peritoneal mast cells and is associated with granular structures distinct from secretory granules and lysosomes. Biochem. J. **346:** 817–826
36 Gomis-Rüth F. X., Companys V., Qian Y., Fricker L. D., Vendrell J., Avilés F. X. et al. (1999) Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallocarboxypeptidase subfamily. EMBO J. **18:** 5817–5826
37 Quiocho F. A. and Lipscomb W. N. (1971) Carboxypeptidase A: a protein and an enzyme. Adv. Protein Chem. **25:** 1–78

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

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

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

41 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

42 Estébanez-Perpiñá E., Bayes A., Vendrell J., Jongsma M. A., Bown D. P., Gatehouse J. A. et al. (2001) Crystal structure of a novel mid-gut procarboxypeptidase from the cotton pest *Helicoverpa armigera.* J. Mol. Biol. **313:** 629–638

43 Pereira P. J. B., Segura-Martín S., Oliva B., Ferrer-Orta C., Avilés F. X., Coll M. et al. (2002) Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). J. Mol. Biol. **321:** 537–547

44 Reverter D., Fernandez-Catalan C., Baumgartner R., Pfander R., Huber R., Bode W. et al. (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

45 Rees D. C. and Lipscomb W. N. (1980) Structure of the potato inhibitor complex of carboxypeptidase A at 2.5-A resolution. Proc. Nat. Acad. Sci. USA **77:** 4633–4637

46 Pallarés I., Bonet R., García-Castellanos R., Ventura S., Avilés F. X., Vendrell J. et al. (2005) Structure of human carboxypeptidase A4 with its endogenous protein inhibitor, latexin. Proc. Natl. Acad. Sci. USA **102:** 3978–3983

47 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. **28:** 19925–19933

48 Reverter D., Ventura S., Villegas V., Vendrell J. and Aviles F. X. (1998) Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. J. Biol. Chem. **273:** 3535–3541

49 Leslie A. G. (1999) Integration of macromolecular diffraction data. Acta Crystallogr. D **55:** 1696–1702

50 CCP4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D **50:** 760–763

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

52 Brünger A. T., Adams P. D., Clore G. M., DeLano W. L., Gros P., Grosse-Kunstleve R. W. et al. (1998) Crystallography & NMR System: a new software suite for macromolecular structure determination. Acta Crystallogr. D **54:** 905–921

53 Martí-Renom M. A., Ilyin V. A. and Sali A. (2001) DBAli: a database of protein structure alignments. Bioinformatics **17:** 746–747

54 Bayes A., Sonnenschein A., Daura X., Vendrell J. and Avilés F. X. (2003) Procarboxypeptidase A from the insect pest *Helicoverpa armigera* and its derived enzyme: two forms with new functional properties. Eur. J. Biochem, **270:** 3026–3035

55 Ollis D. L., Cheah E., Cygler M., Dijkstra B., Frolow F., Franken S. M. et al. (1992) The α/β hydrolase fold. Prot. Eng. **5:** 197–211

56 Lawrence M. C. and Colman P. M. (1993) Shape complementarity at protein/protein interfaces. J. Mol. Biol. **234:** 946–950

57 Janin J. and Chothia C. (1990) The structure of protein-protein recognition sites. J. Biol. Chem. **265:** 16027–16030

58 Baumann U., Bauer M., Letoffe S., Delepelaire P. and Wandersman C. (1995) Crystal structure of a complex between *Serratia marcescens* metallo-protease and an inhibitor from *Erwinia chrysanthemi.* J. Mol. Biol. **248:** 653–661

59 Gomis-Rüth F. X., Maskos K., Betz M., Bergner A., Huber R., Suzuki K. et al. (1997) Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. Nature **389:** 77–81

60 Stubbs M. T., Laber B., Bode W., Huber R., Jerala R., Lenarcic B. et al. (1990) The refined 2.4 Å X-ray crystal structure of recombinant human stefin B in complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. EMBO J. **9:** 1939–1947

61 Reverter D., Maskos K., Tan F., Skidgel R. A. and Bode W. (2004) Crystal structure of human carboxypeptidase M, a membrane-bound enzyme that regulates peptide hormone activity. J. Mol. Biol. **338:** 257–269

62 Hege T., Feltzer R. E., Gray R. D. and Baumann U. (2001) Crystal structure of a complex between *Pseudomonas aeruginosa* alkaline protease and its cognate inhibitor: inhibition by a zinc-NH_2 coordinative bond. J. Biol. Chem. **276:** 35087–35092

63 Gomis-Rüth F. X. (2003) Structural aspects of the metzincin clan of metalloendopeptidases. Mol. Biotech. **24:** 157–202

64 Bode W., Engh R., Musil D., Thiele U., Huber R., Karshikov A. et al. (1988) The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. EMBO J. **7:** 2593–2599

65 Somoza J. R., Jiang F., Tong L., Kang C. H., Cho J. M. and Kim S. H. (1993) Two crystal structures of a potently sweet protein: natural monellin at 2.75 Å resolution and single-chain monellin at 1.7 Å resolution. J. Mol. Biol. **234:** 390–404

66 Tsai P. L., Chen C. H., Huang C. J., Choo C. M. and Chang G. D. (2004) Purification and cloning of an endogeneous inhibitor of carp nephrosin, an astacin metallo proteinase. J. Biol. Chem. **279:** 11146–11155

67 Rawlings N. D., Tolle D. P. and Barrett A. J. (2004) Evolutionary families of peptidase inhibitors. Biochem. J. **378:** 705–716

68 Gautron J., Hincke M. T., Mann K., Panheleux M., Bain M., McKee M. D. et al. (2001) Ovocalyxin-32, a novel chicken eggshell matrix protein: isolation, amino acid sequencing, cloning, and immunocytochemical localization. J. Biol. Chem. **276:** 39243–39252

69 Youssef E. M., Chen X. Q., Higuchi E., Kondo Y., García-Manero G., Lotan R. et al. (2004) Hypermethylation and silencing of the putative tumor suppressor tazarotene-induced gene 1 in human cancers. Cancer Res. **64:** 2411–2417

70 Nagpal S., Patel S., Asano A. T., Johnson A. T., Duvic M. and Chandraratna R. A. (1996) Tazarotene-induced gene 1 (TIG1), a novel retinoic acid receptor-responsive gene in skin. J. Invest. Dermatol. **106:** 269–274

71 Woessner J. F. Jr and Nagase H. (2000). Matrix Metalloproteinases and TIMPs. Protein Profile series, Sheterline P. (ed.), Oxford University Press, New York