research papers

Direct interaction between a human digestive protease and the mucoadhesive poly(acrylic acid)

Irantzu Pallarès,[a] Daniel Fernández,[a] Mireia Comellas-Bigler,[b] Juan Fernández-Recio,[c] Salvador Ventura,[a] Francesc X. Avilés,[a] Wolfram Bode[b]* and Josep Vendrell[a]*

[a]Departament de Bioquímica i Biologia Molecular, Facultat de Biociències and Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain, [b]Proteinase Research Group, Max-Planck Institute of Biochemistry, D-82152 Martinsried, Germany, and [c]Life Sciences Department, Barcelona Supercomputing Center, E-08034 Barcelona, Spain

Correspondence e-mail:
bode@biochem.mpg.de, josep.vendrell@uab.es

Carboxypeptidase A1 has been the subject of extensive research in the last 30 y and is one of the most widely studied zinc metalloenzymes. However, the three-dimensional structure of the human form of the enzyme is not yet available. This report describes the three-dimensional structure of human carboxypeptidase A1 (hCPA1) derived from crystals that belong to the tetragonal space group $P4_32_12$ and diffract to 1.6 Å resolution. A description of the ternary complex hCPA1–Zn^{2+}–poly(acrylic acid) is included as a model of the interaction of mucoadhesive polymers with proteases in the gastrointestinal tract. The direct mode of interaction between poly(acrylic acid) and the active site of the target protease was confirmed by *in vitro* inhibition assays. The structure was further analyzed *in silico* through the optimal docking-area method. The characterization of binding sites on the surface of hCPA1 and a comparison with other available carboxypeptidase structures provided further insights into the formation of multiprotein complexes and the activation mechanisms of carboxypeptidase zymogens. The high-resolution structure of hCPA1 provides an excellent template for the modelling of physiologically relevant carboxypeptidases and could also contribute to the design of specific agents for biomedical purposes.

1. Introduction

Metallocarboxypeptidases (MCPs) are Zn^{2+}-dependent exoproteases of widespread distribution which are present in almost every tissue or fluid examined in mammals and perform a wide variety of physiological functions (Vendrell *et al.*, 2004). The protein-degrading activity of CPs requires tight levels of control, involving their synthesis as inactive precursor proteins (proenzymes or zymogens), their release into separate compartments and the expression of endogenous CP inhibitors, amongst other strategies.

Based on their primary and tertiary structure, the most thoroughly studied CPs of the M14 family can be divided into subfamilies A/B and N/E, which differ in their domain distribution and the architecture of the loops surrounding the active site (Wei *et al.*, 2002). Further, the mature enzymes that catalyze the removal of C-terminal amino acids from proteins and/or peptides can be roughly classified into two groups: the forms that preferentially hydrolyze the peptide bonds before C-terminal aliphatic and aromatic residues (various CPA forms) and those that are specific for basic residues (such as CPB, CPN, CPM, CPU/TAFI *etc.*).

A battery of pancreatic enzymes catalyze the degradation of the protein content of the diet into free amino acids in the gastrointestinal tract and their activity may impair the bio-

Table 1
X-ray data collection and processing.

Values in parentheses are for the last resolution shell.

Space group	$P4_32_12$
Unit-cell parameters (Å, °)	$a = 129.6$, $b = 129.6$, $c = 90.9$, $\alpha = 90.0$, $\beta = 90.0$, $\gamma = 90.0$
Molecules per ASU	2
Resolution (Å)	20.0–1.60 (1.68–1.60)
Reflections measured/unique	521057/101271
$I/\sigma(I)$	33.5 (5.0)
R_{merge}†	0.054 (0.439)
Completeness (%)	98.8 (83.1)
Reflections used for refinement (total/test set)	100653/2012
R factor‡ (%)/R_{free}§ (%)	17.8/19.2
No. of non-H atoms	5018
No. of residues	616
No. of solvent molecules	779
No of ligands (Zn^{2+}/PAA/HEPES)	2/2/6
R.m.s.d. bond lengths (Å)	0.007
R.m.s.d. bond angles (°)	1.4
R.m.s.d. bonded B factors (Å^2)	0.94
Average B factor for protein atoms (Å^2)	20.0
Average B factor for PAA (Å^2)	23.0
Average B factor for metal ions (Å^2)	15.4
Ramachandran core conformations (%)	87.3
Additionally allowed conformations (%)	12.5
Disallowed conformations (%)	0.2

† $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. ‡ R factor $= \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$. § R_{free} was calculated omitting 2% of the observed reflections from refinement and R-factor calculation.

availability of therapeutic peptides and proteins (Lee *et al.*, 2000). An increasingly explored area in the pharmaceutical field is the development of strategies to enhance the absorption of protein and peptide medications. In this context, the co-administration of bioadhesive polymers is used to prolong the contact time of many drugs with body tissue, improving their performance (Lee *et al.*, 2000). In addition to their mucoadhesive properties and their permeation-enhancing effects, certain polymers are able to inhibit the protease activities in the intestine. Among them, the mucoadhesive polymer poly(acrylic acid) (PAA), used as a co-administered product, has been shown to inhibit CPA activity in the gastrointestinal tract (Luessen *et al.*, 1996).

The enzyme mechanism of CPs has essentially been elucidated (Kim & Lipscomb, 1990; Christianson & Lipscomb, 1986) and three-dimensional structures are available for many different CPs (Vendrell *et al.*, 2004). However, although full structural details of human pancreatic CPA2 and CPB are known (PDB codes 1aye and 1kwm, respectively), the three-dimensional structure of human CPA1 is not yet available. In addition to being considered the reference or 'canonical' form for functional and mechanistic studies of proteases, mature hCPA1 and its proregion have been linked to human diseases and are considered to be serum markers for the diagnosis of pancreatic cancer (Bar-Or *et al.*, 2006; Shamamian *et al.*, 2006; Matsugi *et al.*, 2007). The availability of three-dimensional structures of peptidases related to disease may be beneficial to the drug-discovery process (Mittl & Grutter, 2006). For instance, full details of the active site of hCPA1 and of its differences compared with that of hCPA2 or hCPB might aid in the design of new prodrugs for cancer therapy (Smith *et al.*, 1997).

The binding of ligands to CPs has been the subject of extensive research, with special emphasis on inhibitors of either natural or synthetic origin (Kim, 2004; Christianson & Lipscomb, 1989). In the case of pancreatic CPs, the concept of ligand binding can be extended to the interaction with their own activation domain and with accompanying proteases in quaternary complexes (Pascual *et al.*, 1989). Modern technologies such as the optimal docking-area (ODA) method are valuable tools for the identification of exosites or potential ligand-binding sites on the surface of the protein (Fernandez-Recio *et al.*, 2005). The continuous surface patches with optimal docking desolvation energy defined by ODA could be regarded as docking sites that might be exploited in the identification of atypical CP inhibitors, *i.e.* compounds that do not coordinate to the metal ion. The recent successful application of the ODA method to the CP M14 family based on complexes of known three-dimensional structure (Fernandez *et al.*, 2007) permits its extrapolation to newly released structures.

A description of the newly determined three-dimensional structure of human CPA1 is presented here, together with a study of potential binding sites for multiprotein complex formation and a description of the mode of binding of PAA to the active site of the enzyme in what, to our knowledge, is the first example of a complex between a mucoadhesive polymer and a target digestive protease of the human gastrointestinal tract.

2. Materials and methods

2.1. Protein preparation and crystallization

Human pancreatic pro-CPA1 (hPCPA1) was overexpressed in the supernatant medium of the yeast *Pichia pastoris* heterologous system following established procedures (Bayes *et al.*, 2003). Briefly, after hydrophobic interaction chromatography on a butyl-Toyopearl column followed by preparative chromatography on an anion-exchange TSK-DEAE column (60 min gradient from 0 to 0.15 *M* ammonium acetate in 20 m*M* Tris pH 7.0 at a flow rate of 4 ml min^{-1}), recombinant hPCPA1 was activated with trypsin at a 40:1(*w*:*w*) ratio and the prodomain was removed by gel filtration. After concentration, the purified protein (1 μl) dissolved at 6 mg ml^{-1} in 5 m*M* Tris–HCl, 50 m*M* NaCl pH 7.5 was mixed in a 1:1 ratio with the crystallization solutions. Crystals grew in 22%(*w/v*) PAA 5100, 0.1 *M* HEPES, 0.02 *M* magnesium chloride pH 7.5 crystallization buffer using a sitting-drop vapour-diffusion setup at 293 K.

2.2. Data collection and processing

For data collection, crystals were soaked in a cryobuffer containing 30%(*v/v*) glycerol before freezing in a nitrogen stream at 100 K (Oxford Cryosystems Cryostream). High-resolution data to 1.6 Å were collected on the DESY BW6 beamline (Deutsches Elektronen Synchrotron, Hamburg, Germany) at a wavelength of 1.050 Å. Indexing and integration of the diffraction data was carried out with *DENZO*

(Otwinowski & Minor, 1997). The data were merged, scaled and truncated using programs supported by CCP4 (Collaborative Computational Project, Number 4, 1994). Table 1 provides a summary of the data collection and processing.

2.3. Structure determination and processing

The structure of hCPA1 was solved by molecular replacement using the *AMoRe* program suite (Navaza, 1994) and a bovine CPA1 search model (PDB code 5cpa; Rees *et al.*, 1983) which consisted of the residues that were conserved in both enzymes, with all other residues mutated to alanine. After rigid-body fitting, the best solution reached a correlation coefficient and *R* factor of 36.1% and 46.1%, respectively, for data in the resolution range 15.0–3.5 Å. Model building was performed on a Silicon Graphics Indigo 2 workstation using the program *MAIN* (Turk, 1992). Positional and individual temperature-factor refinement was carried out with the program *CNS* (Brünger *et al.*, 1998) until convergence without imposing noncrystallographic symmetry constraints and the final model yielded final *R* and R_{free} values of 17.8% and 19.2%, respectively. The two hCPA1 molecules present in the crystal asymmetric unit are structurally equivalent, with an r.m.s.d. of 0.23 Å for C^{α} atoms. Therefore, only molecule *A* is considered in the description of the structure. The hCPA1 chain encompasses protein residues Ala1–Pro308 and a Zn^{2+} ion which was identified and refined with full occupancy. Additional electron density near the active site was interpreted as a PAA molecule and three HEPES molecules were found on the surface. During the final steps of refinement, alternative conformations were identified and built for a few side chains.

The Ramachandran plot calculated for the final hCPA1 model with the program *PROCHECK* (Laskowski *et al.*, 1993) shows that 87.3% of the residues are in the most favoured regions of the φ, ψ plane, 12.5% of residues are in additionally allowed regions and only one amino-acid residue appears in a disallowed region. Only residual electron density was detected beyond the last observable methylene of PAA and it was not possible to construct a reliable model accounting for a putatively more extended polymer chain.

2.4. Inhibition assays

Poly(acrylic acid, sodium salt) 5100, the substrate *N*-[3-(2-furyl)acryloyl]-L-phenylalanyl-L-phenylalanine (FAPP) and bovine pancreatic CPA1 were from Sigma; the substrate *N*-(4-methoxyphenylazoformyl)-L-phenylalanine (AAFP) was obtained from Bachem. To calculate the K_i values for PAA against hCPA1, measurements of enzymatic activity were carried out at an enzyme concentration of 5 n*M* and variable concentrations of ligand (100 n*M* to 100 µ*M*) and FAPP substrate (25–250 µ*M*) in 50 m*M* Tris, 0.5 *M* NaCl pH 7.5 at room temperature in 1 ml cuvettes. Changes in absorbance were followed at 330 nm. Assays using the bovine form of the enzyme (bCPA1) were also carried out under the same buffer and temperature conditions but using the substrate AAFP and following the absorbance changes at 350 nm. In this case, a continuous photometric assay was used at an enzyme concentration of 50 n*M*, a substrate concentration fixed at its K_{m} value of 0.11 m*M* (Mock *et al.*, 1996) and ligand concentrations that spanned from 10 n*M* to 1 m*M*. The measurements were carried out in a CARY 100 Bio Spectrophotometer (Varian). Data were analyzed with the *GraphPad Prism* 5.0 package. A nonlinear regression fitting to a single-site ligand competitive model yielded kinetic parameters with excellent agreement with the experimental data.
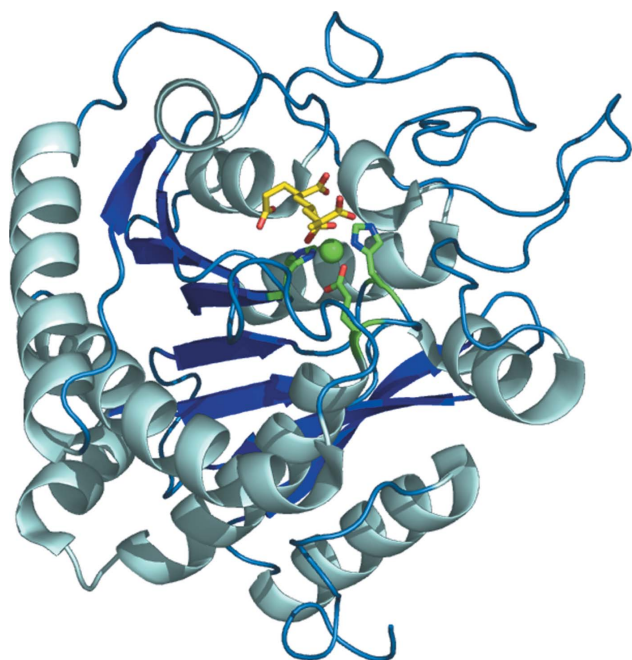
2.5. Calculation of ODA patches

To identify continuous surface patches with optimal desolvation energy, we followed the procedure established in a previous publication (Fernandez *et al.*, 2007). Briefly, a number of starting points were defined around the protein using the centre of coordinates of each residue side chain. The docking surface energy of the different-sized patches was then calculated at each point as the sum of the solvation energy contribution of the atoms that lie within a sphere of radius *d* ($d = 1, 2, \ldots, 20$ Å) from the starting point. Among all surface patches generated in this way, that with the lowest energy value is called an ODA and its optimal energy value is assigned to that point and subsequently to the residue from which that point was generated. ODAs were considered only for residues at the surface of the protein, *i.e.* when the relative solvent-accessible surface area (SASA) of the residue is greater than 0.25. An ODA was considered to be significant when its optimal docking energy value was $< -42 \mathrm{~kJ~mol}^{-1}$. Finally, the ODA energy values per residue were used to define regions of the protein surface (ODA patches) that are likely to be involved in protein–protein binding.

3. Results and discussion

3.1. Overall structure of human CPA1 and comparison with other CP structures

The crystals belonged to the tetragonal space group $P4_32_12$, with unit-cell parameters $a = b = 129.6$, $c = 90.9$ Å, and contained two molecules per asymmetric unit. Two other entries in the PDB belong to the same space group and correspond to MCPs bound to protein inhibitors (PDB codes 1zli and 2abz). This paper reports the determination of a new structure of a CP bound to a synthetic compound that behaves as an inhibitor of the protease. The statistics for the structure at the final 1.6 Å resolution are listed in Table 1.

Like its other nonhuman homologues (Vendrell *et al.*, 2004), the tertiary fold of hCPA1 corresponds to the $\alpha\beta$-hydrolase fold and is formed by a central mixed parallel/antiparallel eight-stranded β-sheet with a 120° twist between the first and the last strand, over which eight α-helices pack on both sides to form a globular molecule (Fig. 1). Superimposition of the hCPA1 structure onto that of its bovine and porcine orthologues (PDB codes 5cpa and 1pca, respectively) gave r.m.s.d. values of 0.4 Å for 308 C^{α} atoms in both cases. Of the human CP forms, hCPA2 is the enzyme that bears the higher degree of similarity to hCPA1 (Fig. 2) and pairwise comparison of the
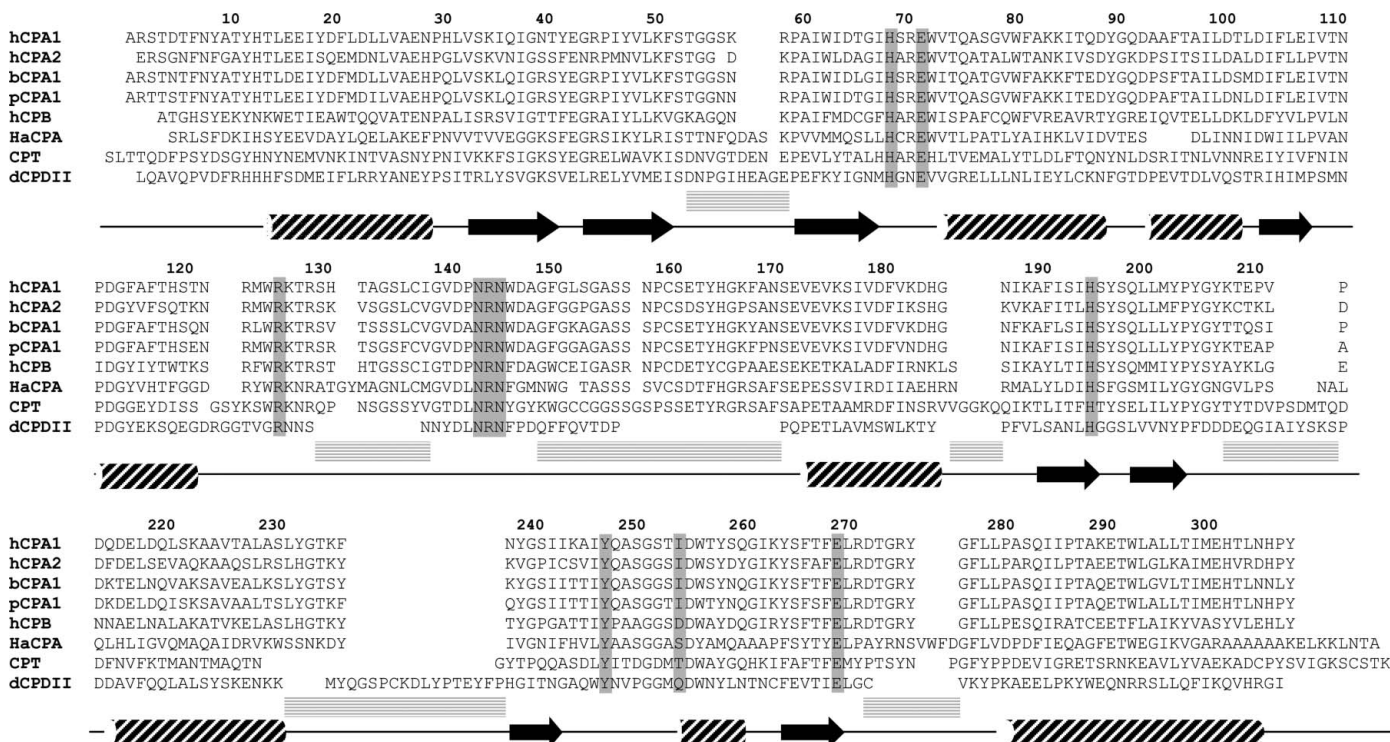
Figure 1
Cartoon representation of hCPA1. The PAA molecule is shown as a yellow stick model, the Zn atom is represented as a green sphere and the three Zn^{2+}-coordinating residues at the active site (His196, Glu72 and His69, from left to right in this view) are shown as green sticks. This figure was generated with *PyMOL* (http://www.pymol.org).

three-dimensional structures gave an r.m.s.d. value of 0.6 Å for 301 equivalent C^α atoms. The specificity pocket of both enzymes is lined by residues that are at equivalent positions in the three-dimensional structure. Gly253, Ser254 and Ala268 in hCPA2 substitute for larger or more polar residues in the A1 form (García-Sáez *et al.*, 1997), defining a wider pocket for hCPA2 that is able to accommodate bulkier residues and that also has a stronger hydrophobic character.

The amino-acid residue at position 255 appears to be the main determinant of specificity for the A/B subfamily of MCPs. As in all mammalian CPAs, residue 255 is an isoleucine in hCPA1, while hCPB (PDB code 1kwm) has an aspartic acid at this position, accounting for the specificity of B-type enzymes for C-terminal basic residues. The surface loops of hCPB also show some differences compared with the equivalent loops in hCPA1, which, together with the different orientation of the N-termini, gives an r.m.s.d. value of 1.4 Å for 301 C^α atoms.

Most members of the N/E subfamily of CPs function in the processing of biologically active peptides, whereas many of those from the A/B subfamily act in the digestion of dietary protein. These functional differences are also reflected in the three-dimensional structures. The structures of hCPA1 and the second catalytic domain of duck CPD (dCPD-II; PDB code 1qmu; a representative member of the N/E subfamily) can be superimposed with an r.m.s.d. of 1.5 Å for 239 topologically

Figure 2
Structure-based sequence alignment of hCPA1 and other CPs: h, human; b, bovine; p, porcine; Ha, *Helicoverpa armigera*; d, duck; CPT, carboxypeptidase from *Thermoactinomyces vulgaris*. The alignment was made with *ClustalW* (http://www.ebi.ac.uk/clustalw) and manually refined to account for experimental information. Conserved residues for catalysis are boxed in grey. The regular secondary-structure elements (helices and strands) indicated below the sequences correspond to the hCPA1 structure. Between the sequences and the symbols for secondary structure, grey lined boxes indicate regions of topological divergence between hCPA1 and dCPDII, the only representative of the N/E subfamily of CPs. The sequence of dCPDII is truncated at its C-terminal end, showing only the residues corresponding to its CP moiety.

Table 2
Intermolecular contacts between the PAA molecule and the enzyme moiety.

Enzyme atom	PAA atom	Distance (Å)
Zn^{2+}	$O^{\beta 1}$	2.5
Zn^{2+}	$O^{\beta 2}$	2.1
His69 $N^{\delta 1}$	$O^{\beta 2}$	3.0
Arg71 $N^{\eta 1}$	$O^{\alpha 1}$	3.3
Glu72 $O^{\varepsilon 1}$	$O^{\beta 1}$	3.2
Glu72 $O^{\varepsilon 1}$	$O^{\beta 2}$	3.1
Arg127 $N^{\eta 1}$	$O^{\gamma 1}$	3.2
Arg127 $N^{\eta 1}$	$O^{\gamma 2}$	3.2
Arg127 $N^{\eta 2}$	$O^{\beta 2}$	2.8
Arg127 $N^{\eta 2}$	$O^{\alpha 1}$	3.3
Asn144 $N^{\delta 2}$	$O^{\delta 1}$	3.0
Asn144 $N^{\delta 2}$	$O^{\gamma 2}$	2.9
Arg145 $N^{\eta 2}$	$O^{\gamma 1}$	2.8
Arg145 $N^{\eta 1}$	$O^{\gamma 2}$	2.8
Tyr248 O^{η}	$O^{\gamma 1}$	2.6
Thr268 $O^{\gamma 1}$	$O^{\delta 2}$	2.6
Glu270 $O^{\varepsilon 1}$	$O^{\delta 2}$	3.1
Glu270 $O^{\varepsilon 1}$	$O^{\beta 1}$	2.7
Glu270 $O^{\varepsilon 2}$	$O^{\beta 1}$	3.1

equivalent C^{α} atoms, whereas the r.m.s.d. value rises to 6.7 Å when all C^{α} atoms are considered. The most important differences between the catalytic subdomains can be observed in the chain segments that form the funnel-like access to the active-site cleft. As can be seen in Fig. 2, most disagreements in the alignment come from the entry corresponding to dCPD-II. The insertions and deletions shape an entrance to the active site of dCPD-II that is clearly distinguishable from that in hCPA1 and other A/B enzymes (Gomis-Ruth *et al.*, 1999). All these differences, which are also observed in comparisons with the recently determined structures of hCPM (Reverter *et al.*, 2004) and hCPN (Keil *et al.*, 2007), suggest that members of the N/E subfamily have a distinct selectivity towards peptide substrates.

3.2. The active site

The active site is located in a cavity formed by parallel strands of the internal β-sheet at the switch point between strands 3 and 5, two helices and a non-regular extension that partially covers it. It has been reported that the active site of CPAs is found in two different conformations depending on the binding of substrates, substrate analogues or inhibitors (Rees & Lipscomb, 1981; a sketch of the two conformations is shown in Supplementary Fig. 1[1]). In the presence of substrates, inhibitors or anions that occupy the active site, Tyr248 rotates about its χ_1 torsion angle in a movement that spans about 12 Å and fixes it in the 'down' position, closing the access to the active site. Tyr248 has been found in the 'down' position in the PCI–bCPA1 complex (Rees & Lipscomb, 1982; PDB code 4cpa), porcine PCPA1 (Guasch *et al.*, 1992; PDB code 1pca), the LCI–hCPA2 complex (Reverter *et al.*, 2000; PDB code 1dtd) and now in hCPA1.
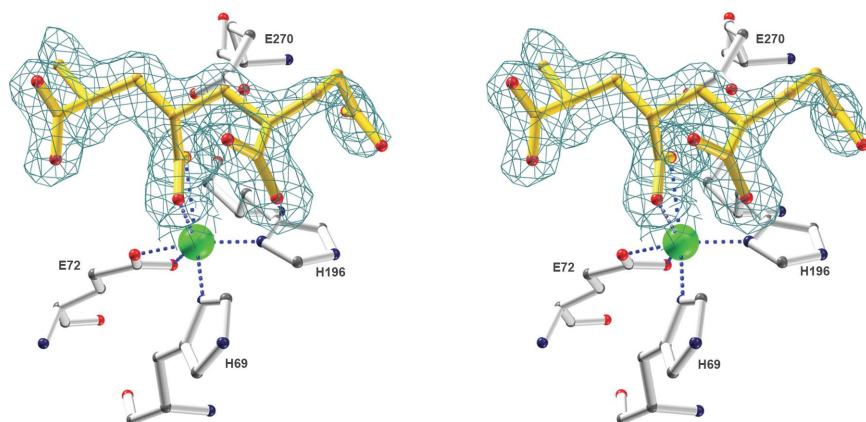
A strong density at the active site was interpreted as a PAA molecule, a ligand present in the crystallization solution. As discussed in §3.3, the binding of a PAA molecule to the active site delineates the positions of the S1′ and S1 subsites. Similar interactions involving the two main residues of the S1 site, Arg127 and Glu270, were also observed in the LCI–hCPA2 complex. PAA was found to interact with most of the active-site residues and correspondingly some residues are found in the conformation induced by substrate or inhibitor binding.

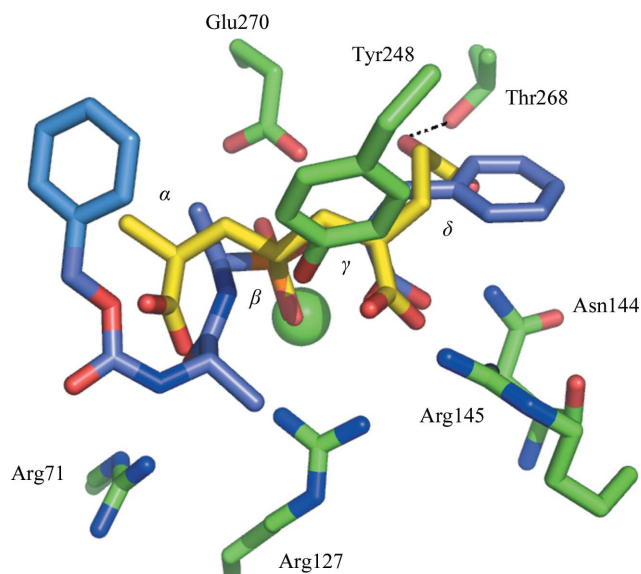## 3.3. The interaction of the PAA molecule in the active site of human CPA1

The coordination of Zn^{2+} in the active site is conserved compared with other MCPs. The first coordination shell of the Zn atom usually consists of four protein atoms, Glu72 $O^{\varepsilon 1}$, Glu72 $O^{\varepsilon 2}$, His69 $N^{\delta 1}$ and His196 $N^{\delta 1}$, and one water molecule. In the present structure (Fig. 3) the catalytic water molecule has been displaced by the PAA molecule and substituted by the $O^{\beta 1}$ and $O^{\beta 2}$ atoms which interact with Zn^{2+}. A water molecule 3.8 Å away from the zinc interacts with Arg71 $N^{\delta 1}$ (2.9 Å), Arg127 $N^{\eta 2}$ (3.0 Å) and with the $O^{\alpha 1}$ group of the PAA molecule (2.6 Å).

PAA exploits the flexibility of its carbon chain to accommodate its functional groups so as to satisfy all possible interactions within the active site, resulting in its binding to hCPA1 being mediated by the four carboxylate groups (Table 2). As shown in the complementary view of PAA binding presented in Fig. 4, starting from the far-right carboxylate group (the δ-carboxylate), the dihedral angles are *gauche/gauche/trans/trans/trans/trans*, suggesting that the carbon-chain structure is twisted so as to allow the β- and γ-carboxylates to interact with Zn^{2+} and Arg145. The mode of action of the two central β- and γ-carboxylates, the first of which coordinates bidentately with the Zn^{2+} ion and the other establishing a salt bridge with Arg145, resembles that of typical CP inhibitors such as L-benzylsuccinic acid (PDB code 1cbx) and the pseudo-substrate ZAA^P(O)F (PDB code 6cpa; Kim & Lipscomb, 1990). A superimposition of the latter structure with that presented here shows that the β-carboxylate is located exactly over the POO— moiety that coordinates the Zn^{2+}. The α-carboxylate is oriented towards Arg71 in a fashion similar to the carbonyl O atom of the first Ala residue in ZAA^P(O)F.

The most striking difference between PAA and the typical phenylalanine-based inhibitors is the presence of the δ-carboxylate at the hydrophobic binding pocket of hCPA1. Fig. 4 shows that it occupies a space where the benzyl ring of a phenylalanine-based ligand is commonly found. Probably to satisfy a hydrogen-bonding interaction with the polar ligand, Thr268 orients its side chain so as to contact PAA *via* its OH group. This feature results in the Thr268 χ_1 dihedral angle being *trans* ($\chi_1 = -177.1°$), in clear contrast to other CP structures, in which the most common rotamer conformation is *gauche* ($\pm 60°$). A further stabilization of the Thr268 conformation may be provided by the hydrogen bond between atom $O^{\gamma 1}$ and the main-chain carbonyl O atom of Ser194. Thr268 is

Figure 3
Stereo plot of active-site residues of hCPA1 in complex with PAA (yellow carbon chain). The figure shows an OMIT electron-density map of bound PAA contoured at the 1σ level. The Zn^{2+} atom is represented by a green sphere. Residues in the zinc-coordination sphere, as well as the catalytically relevant Glu270, are shown. This figure was rendered with *BobScript* (Esnouf, 1997).

Figure 4
Close-up view of the active site of hCPA1 in complex with PAA (yellow carbon chain) superimposed on bCPA1 in complex with the phosphonate inhibitor ZAA^P(O)F (PDB code 6cpa). For the latter structure only the inhibitor is shown, with a light blue carbon chain and the P atom in orange. Some residues involved in catalysis and substrate binding in hCPA1 are shown as green stick models. Try248 adopts the 'down' conformation (see Supplementary Fig. 1). Other residues commented upon in the text describing this figure are not shown for clarity. The Zn^{2+} atom is represented as a green sphere. The carboxylate groups in the PAA moiety are labelled as they are named in the text. The hydrogen bond between the δ-carboxylate and Thr268 is marked with a dashed line. This figure was prepared with the program *PyMOL* (http://www.pymol.org).

considered to be an important determinant of specificity in CPs; its uncommon rotamer conformation might create a different binding pocket for ligands, as has been reported in a previous drug-design investigation (Smith *et al.*, 1997).

Other investigations have dealt with the inhibitory and thus drug-protective effects of some polymers, such as poly-acrylates, against a number of digestive proteases including

trypsin, α-chymotrypsin, CPA, CPB and cytosolic leucine aminopeptidase (Luessen *et al.*, 1995, 1996). It has been argued that protease inhibition mainly originates from the sequestration of active-site Zn^{2+} or Ca^{2+} ions, thus rendering the enzymes inactive. However, some metalloenzymes require higher concentrations of polymer to achieve an inhibitory effect (this is the case for cytosolic leucine aminopeptidase, a Zn^{2+}-dependent metalloenzyme) or appear not to be inhibited at all by polymers (as in the case of microsomal leucine aminopeptidase) and consequently alternative modes of interaction cannot be ruled out (Luessen *et al.*, 1995, 1996).

The structure reported here reveals a direct interaction between a mucoadhesive polymer and a digestive protease and suggests that the mode of action of mucoadhesive polymers, apart from a putative metal-depriving effect, could also involve occupation of the protease substrate-binding site. Those polymers that combine delivery capabilities with protease-inhibitory effects might be exploited in a structure-based drug-design strategy aimed at the development of novel and specific excipients to improve the bioavailability of non-invasively administered peptide drugs (Bernkop-Schnurch & Kast, 2001).

In light of the structural evidence, hCPA1–PAA crystals were picked up from the drops, carefully washed and dissolved in a substrate-containing solution. No carboxypeptidase activity was detected by enzymatic measurements. Subsequently, *in vitro* inhibition assays were performed with a fresh sample of hCPA1 using various enzyme:PAA ratios at different substrate concentrations. The profile of the double-reciprocal representation shows that addition of PAA affects K_m and not V_{max} and thus corresponds to competitive inhibition with a K_i value of 8.7 μM (see Supplementary Fig. 2). To investigate whether PAA is also able to inhibit hCPA1 orthologues, we also calculated the K_i value for bCPA1 using a continuous assay method (Bieth, 1995), which rendered a similar K_i value of 8 μM after fitting to a competitive-inhibition model. For comparison, the benchmark inhibitor DL-benzylsuccinic acid has a K_i value of about 1.0 μM against the same target. These observations are consistent with the structural data and indicate that PAA behaves as an inhibitor and interacts directly with the protease.

3.4. Binding sites and complex formation of hCPA1 as studied by the optimal docking-area method

The ODA (optimal docking-area) method is an efficient algorithm that identifies areas on the surface of a protein that have a higher probability of being involved in protein–protein interactions (Fernandez-Recio *et al.*, 2005). We have applied ODA analysis to the structure of hCPA1 and compared the results with those obtained from a set of known three-

dimensional structures of the M14 family of CPs that were the subject of a recent structural and functional characterization of binding sites using this method (Fernandez *et al.*, 2007). In the mentioned study it was found that although the set of CP-enzyme domains were structurally very similar, the protein–protein docking surfaces of different subclasses differed substantially and could be distinguished based on analysis of ODA residues over the surface of the protein.

Alignment of the sequences of hCPA1 and bCPA1 indicates that regardless of the conservation of individual residues, the grouping of ODA residues in the catalytic domains is very similar in both proteins (see Supplementary Fig. 3). All segments of clustered ODA residues in the catalytic domains correspond to areas in which interactions with either the activation domain or protein inhibitors are known to take place (Fernandez *et al.*, 2007). A surface representation indicates that the distribution of ODA patches is very similar in the two proteins (Fig. 5).

Pancreatic CPs are synthesized as zymogens (procarboxypeptidases, PCPs) that occur as binary or ternary complexes with other proproteinase subunits. These associations differ between organisms; for instance, ternary complexes have only been observed in ruminants, while binary complexes have a wider occurrence. Human PCPA1 is known to occur as a binary complex with proproteinase E (PPE), the zymogen of a digestive serine proteinase, whereas bPCPA1 occurs both as a binary complex similar to the human form and a ternary complex in which the additional subunit is chymotrypsinogen C (CTGC; see Aviles *et al.*, 1993). Fig. 6 compares the structures of a virtual hPCPA1 zymogen and the bovine ternary zymogen complex (PDB code 1pyt). For the construction of the virtual hPCPA1 zymogen, the structures of the enzyme regions of hCPA1 and pPCPA1, the sequentially closer orthologue, were superimposed and the coordinates for the porcine enzyme moiety were subsequently eliminated from the figure, rendering a virtual chimera built up of the enzyme moiety of hCPA1 and the pro-region of pPCPA1. Most protein–protein interactions in the ternary complex are mediated *via* the prodomain of the CP, which sustains almost all the interactions to dock both proproteinases; the contribution of the CP moiety is only marginal. A visual observation of the ODA-coloured structures indicates that the constructed hPCPA1 displays a PPE-binding surface
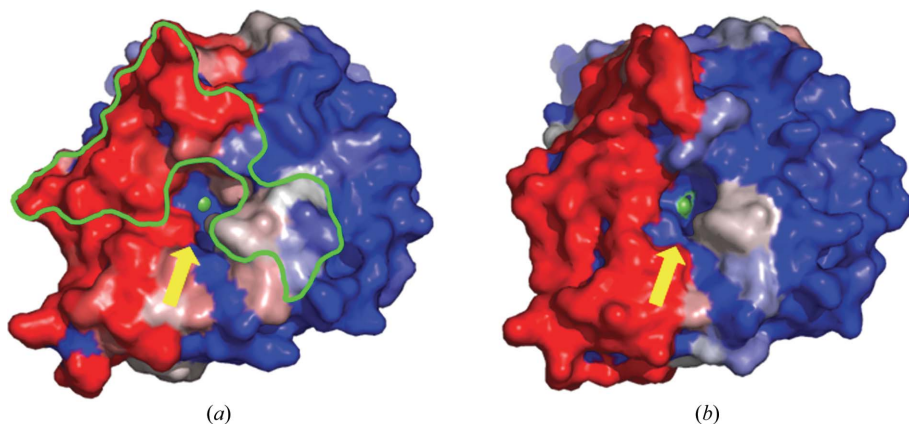

Figure 5
Representation of the surfaces of ODA patches for (*a*) bCPA1 and (*b*) hCPA1. The view points the active-site cleft towards the reader. A yellow arrow marks the entrance to the active site, where a green sphere depicts the Zn^{2+} ion. The colour ramp ranges from red (ODA energy value <-42 kJ mol^{-1}) to blue (ODA energy value = 0.0). The experimental ligand interface is indicated with a green line for bCPA1. This figure was prepared with the program *PyMOL* (http://www.pymol.org) and the colour ramp was generated using an in-house script.
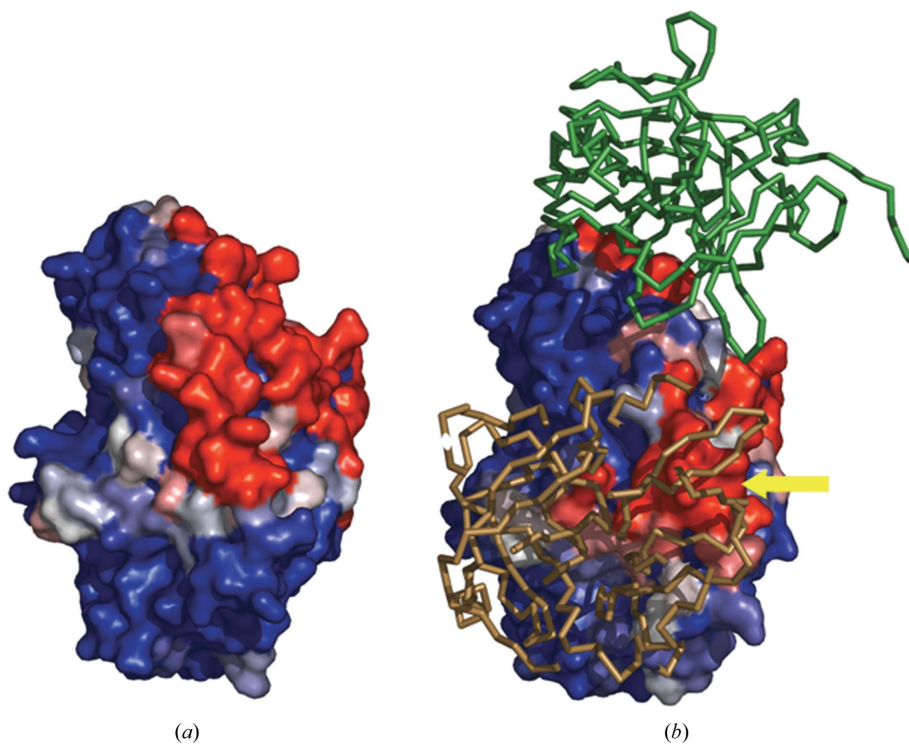

Figure 6
Representation of the surfaces of ODA patches for (*a*) a virtual hPCPA1 (constructed from hCPA1 and the pro-region of pPCPA1) and (*b*) the natural bovine ternary complex built up of bPCPA1, PPE (green backbone representation) and CTGC (gold backbone). The yellow arrow indicates the location of the centre of the ODA-rich region between residues 90 and 99 in bCPA1. The view must be approximately rotated 180° vertically and 90° downwards horizontally to correspond with the orientation in Fig. 5. Other details are as in Fig. 5. This figure was prepared with the program *PyMOL* (http://www.pymol.org).

with ODA energy patches extensive enough to accommodate a PPE subunit, as could be expected from the natural occurrence of the porcine zymogen as a binary complex with PPE (Ventura *et al.*, 1997). On the other side, bPCPA1 is richer in ODA residues at the surface at which CTGC is accommodated, where the residues contained in the long α-helix (Val82–Arg99) connecting the activation and the catalytic domains display the highest ODA energy values (residues 90–99) and are of particular importance in bPCPA1–CTGC binding. It is conceivable that as the long α-helix contains the cleavable arginine residue, species-dependent binding of proproteinases to this region and the subsequent formation of oligomeric zymogen complexes modulates the kinetics of active-enzyme generation (Gomis-Ruth *et al.*, 1998). The computation of ODA energy patches for the known structures of CPs and PCPs is in agreement with the observed occurrence of complexes which are responsible for the differences in activation behaviour between ruminant and mammal PCPs.

References

Aviles, F. X., Vendrell, J., Guasch, A., Coll, M. & Huber, R. (1993). *Eur. J. Biochem.* **211**, 381–389.

Bar-Or, D., Rael, L. T., Bar-Or, R., Slone, D. S. & Craun, M. L. (2006). *Clin. Chim. Acta*, **365**, 346–349.

Bayes, A., Sonnenschein, A., Daura, X., Vendrell, J. & Aviles, F. X. (2003). *Eur. J. Biochem.* **270**, 3026–3035.

Bernkop-Schnurch, A. & Kast, C. E. (2001). *Adv. Drug Deliv. Rev.* **52**, 127–137.

Bieth, J. G. (1995). *Methods Enzymol.* **248**, 59–84.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.

Christianson, D. W. & Lipscomb, W. N. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 7568–7572.

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

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Esnouf, R. M. (1997). *J. Mol. Graph. Model.* **15**, 132–134.

Fernandez-Recio, J., Totrov, M., Skorodumov, C. & Abagyan, R. (2005). *Proteins*, **58**, 134–143.

Fernandez, D., Vendrell, J., Aviles, F. X. & Fernandez-Recio, J. (2007). *Proteins*, **68**, 131–144.

García-Sáez, I., Reverter, D., Vendrell, J., Avilés, F. X. & Coll, M. (1997). *EMBO J.* **16**, 6906–6913.

Gomis-Ruth, F. X., Companys, V., Qian, Y., Fricker, L. D., Vendrell, J., Aviles, F. X. & Coll, M. (1999). *EMBO J.* **18**, 5817–5826.

Gomis-Ruth, F. X., Gomez-Ortiz, M., Vendrell, J., Ventura, S., Bode, W., Huber, R. & Aviles, F. X. (1998). *Eur. J. Biochem.* **251**, 839–844.

Guasch, A., Coll, M., Aviles, F. X. & Huber, R. (1992). *J. Mol. Biol.* **224**, 141–157.

Keil, C., Maskos, K., Than, M., Hoopes, J. T., Huber, R., Tan, F., Deddish, P. A., Erdos, E. G., Skidgel, R. A. & Bode, W. (2007). *J. Mol. Biol.* **366**, 504–516.

Kim, D. H. (2004). *Curr. Top. Med. Chem.* **4**, 1217–1226.

Kim, H. & Lipscomb, W. N. (1990). *Biochemistry*, **29**, 5546–5555.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lee, J. W., Park, J. H. & Robinson, J. R. (2000). *J. Pharm. Sci.* **89**, 850–866.

Luessen, H. L., de Leeuw, B. J., Pérard, D., Lehr, C. M., de Boer, A. G., Verhoef, J. C. & Junginger, H. E. (1996). *Eur. J. Pharm. Sci.* **4**, 117–128.

Luessen, H. L., Verhoef, J. C., Borchard, G., Lehr, C. M., de Boer, A. G. & Junginger, H. E. (1995). *Pharm. Res.* **12**, 1293–1298.

Matsugi, S., Hamada, T., Shioi, N., Tanaka, T., Kumada, T. & Satomura, S. (2007). *Clin. Chim. Acta*, **378**, 147–153.

Mittl, P. R. & Grütter, M. G. (2006). *Curr. Opin. Struct. Biol.* **16**, 769–775.

Mock, W. L., Liu, Y. & Stanford, D. J. (1996). *Anal. Biochem.* **239**, 218–222.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pascual, R., Burgos, F. J., Salva, M., Soriano, F., Mendez, E. & Aviles, F. X. (1989). *Eur. J. Biochem.* **179**, 609–616.

Rees, D. C., Lewis, M. & Lipscomb, W. N. (1983). *J. Mol. Biol.* **168**, 367–387.

Rees, D. C. & Lipscomb, W. N. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 5455–5459.

Rees, D. C. & Lipscomb, W. N. (1982). *J. Mol. Biol.* **160**, 475–498.

Reverter, D., Fernandez-Catalan, C., Baumgartner, R., Pfander, R., Huber, R., Bode, W., Vendrell, J., Holak, T. A. & Aviles, F. X. (2000). *Nature Struct. Biol.* **7**, 322–328.

Reverter, D., Maskos, K., Tan, F., Skidgel, R. A. & Bode, W. (2004). *J. Mol. Biol.* **338**, 257–269.

Shamamian, P., Goldberg, J. D., Ye, X. Y., Stewart, J. D., White, P. J. & Gilvarg, C. (2006). *HPB*, **8**, 451–457.

Smith, G. K., Banks, S., Blumenkopf, T. A., Cory, M., Humphreys, J., Laethem, R. M., Miller, J., Moxham, C. P., Mullin, R., Ray, P. H., Walton, L. M. & Wolfe, L. A. III (1997). *J. Biol. Chem.* **272**, 15804–15816.

Turk, D. (1992). PhD Thesis. Technische Universität, München, Germany.

Vendrell, J., Aviles, F. X. & Fricker, L. D. (2004). *Handbook of Metalloproteins*, Vol. 3, edited by A. Messerschmidt, W. Bode & M. Cygler, pp. 176–189. Chichester: John Wiley & Sons.

Ventura, S., Gomis-Ruth, F. X., Puigserver, A., Aviles, F. X. & Vendrell, J. (1997). *Biol. Chem.* **378**, 161–165.

Wei, S., Segura, S., Vendrell, J., Aviles, F. X., Lanoue, E., Day, R., Feng, Y. & Fricker, L. D. (2002). *J. Biol. Chem.* **277**, 14954–14964.