Caught after the Act: A Human A-Type Metallocarboxypeptidase in a Product Complex with a Cleaved Hexapeptide†

Álex Bayés,§ Daniel Fernández,§ Maria Solà,‡ Aniebrys Marrero,‡ Sonia García-Piqué,‡ Francesc X. Avilés,§ Josep Vendrell§ and F. Xavier Gomis-Ruth*‡

Institut de Biotechnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Facultat de Biociències, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain and Department de Biologia Estructural, Institut de Biologia Molecular de Barcelona CSIC, Jordi Girona 18-26, E-08034 Barcelona and Parc Científic de Barcelona, Josep Sanitier 1-5, E-08028 Barcelona, Spain

Received March 9, 2007; Revised Manuscript Received April 2, 2007

ABSTRACT: A/B-type metallocarboxypeptidases (MCPs) are among the most thoroughly studied proteolytic enzymes, and their catalytic mechanisms have been considered as prototypes even for several unrelated metalloprotease (in)ase families. It has long been postulated that the nature of the side chains of at least five substrate residues, i.e., P₄–P₁, influence Kᵢₐ and kₐₑₐ and that once the peptide or protein substrate is cleaved, both products remain in the first instance bound to the active-site cleft of the enzyme in a double-product complex. Structural details of binding of substrate to the nonprimed side of the cleft have largely relied on complexes with protein inhibitors and peptidomimetic small-molecule inhibitors that do not span the entire groove. In the former, the presence of N-terminal globular protein domains participating in large-scale interactions with the surface of the cognate catalytic domain outside the active-site cleft mostly conditions the way their C-terminal tails bind to the cleft. Accordingly, they may not be accurate models for a product complex. We hereby provide the structural details of a true cleaved double-product complex with a hexapeptide of an MCP engaged in prostate cancer, human carboxypeptidase A4, employing diffraction data to 1.6 Å resolution (R₊ₑₑₑ = 0.159 and 0.176, respectively). These studies provide detailed information about subsites S₅–S₇ and contribute to our knowledge of the cleavage mechanism, which is revisited in light of these new structural insights.

Zinc-dependent metallocarboxypeptidases (MCPs)† remove C-terminal amino acids from protein and peptide substrates and can be classified into A/B and N/E forms, previously known as pancreatic and regulatory MCPs, respectively (1, 2). At the onset of biochemistry, A/B-type MCPs (A/B-MCPs) were identified as digestive enzymes produced by the exocrine pancreas of mammals (3–5). More recently, homologues with functions distinct from the degradation of intake proteins have been found in several other mammalian tissues and fluids and within cells, as well as in every realm of life (5). They participate in delicately regulated processes such as insect attack—vegetal defense strategies, parasite invasion, blood coagulation and fibrinolysis, inflammation, local anaphylaxis, fertilization, hormone and neuropeptide processing, resistance to bee and snake poison, and carcinogenesis and tumor invasion, among others (2, 6–9).

A/B-MCPs exhibit distinct substrate specificity patterns. While B-type CPs (CPBs) favor substrates with C-terminal basic side chains, A-type CPs (CPAs) are subdivided into A₁ and A₂ forms and prefer smaller aliphatic residues and bulkier aromatic side chains, respectively (6, 10). The archetypal and most thoroughly studied MCP is bovine pancreatic CPA (bCPA), actually an A₁ form in its specificity, discovered in the 1920s (4). A more recently identified family member is human carboxypeptidase A4 (hCPA4), detected in human hormone-regulated tissues and thought to play a role in prostate cancer (11, 12). The coding gene may be responsible for prostate-cancer aggressiveness, as it is imprinted and upregulated via the histone hyperacetylation pathway as suggested by the analysis of downstream effects of sodium butyrate treatment of prostate cancer cell lines (12). A/B-MCPs are synthesized and secreted as zymogens [procarboxypeptidases (PCPs)] to prevent activity in the
Scheme 1: Catalytic Mechanism of MCP-Mediated Peptide Cleavage

absence of the appropriate environment and stimuli. The structural effector for latency is an N-terminal 90–95-residue independently folding prodomain that prevents substrate access to the catalytic CP domain, which is already in an active conformation in the zymogen. Activation occurs through limited proteolysis in a connecting segment at the end of the prodomain through endopeptidases such as trypsin (13, 14). A number of endogenous and exogenous protein inhibitors of A/B-MCPs have been reported: the archetypal potato carboxypeptidase inhibitor (PCI) and its orthologue from tomato (MCPI), inhibitors from the intestinal parasite Ascaris suum and the medical leech Hirudo medicinalis (LCI), the inhibitor from the tick Rhipicephalus bursa (TCI), and rodent and human latexin alias ECI (2, 15–17). These protein inhibitors display equilibrium dissociation constants in the nanomolar range against vertebrate A/B-MCPs but are inactive against N/E-type MCPs (15–19).

The enzymatic activity of A/B-MCPs has been the subject of a number of studies, chiefly on bCPA, that have led to the following consensus mechanism of catalysis (5, 20–26) (Scheme 1). In the Michaelis complex, the substrate side chains (P₁–P₃, etc.; the nonprimed side) preceding the scissile bond interact with substrate-binding subsites S₁–S₃, etc., of the enzyme (nomenclature from refs 27 and 28). The C-terminal residue to be removed (side chain P₁) interacts with cognate enzyme substrate-binding subsite S₁, the specificity pocket, on the primed side of the groove (stage I) (27). This is followed by the nucleophilic attack of a catalytic solvent molecule, further bound to and thus polarized by Glu270, on the scissile carbon atom backbone. This generates a negatively charged tetrahedral reaction intermediate (stage II). The scissile bond is cleaved, and the cleavage fragments are subsequently removed so that the enzyme is poised for a new catalytic round. This mechanism entails two product complexes, a double-product complex (stage III), in which both the upstream and downstream cleavage products are still bound to the enzyme, and a single-product complex (stage IV), in which only the excised C-terminal residue remains anchored to the specificity pocket (see Scheme 1 and Figure 68.3 in ref 24).

This mechanistic view is supported by studies with peptidomimetic compounds with a phosphinato group [-P(O)- (O⁻) - (alias -PO₂⁻)] replacing the sp³-hybridized diolate group [-C(OH)₂(O⁻); Scheme 1, stage II] of the reaction intermediate, as observed in phosphonamidates (-PO₂-NH₂), phosphonates (-PO₂-O⁻), and phosphinates (-PO₂-CH₂). To date, phosphate complexes with zinc-dependent proteases have only been reported for astacin and matrix metallopeptinase 11, but not for MCPs (29). In contrast, bCPA has been the subject of studies with tripeptide-mimicking phosphonates with the generic formula ϕ-CH₂-O-CO-NH-CH(R₁)-CO-NH-CH(R₂)-PO₂-O-CH(CH₂-ϕ)-COO⁻, where R₁ = CH₂-ϕ and R₂ = CH(CH₃)₂ for O-{{[(1R)-(N-(phenylmethoxycarbonyl)-L-phenylalanyl)amino]isobutyli}-hydroxyphosphinyl}-L-3-phenyllactate [alias ZVF(O)F]; R₁ = R₂ = CH₃ for O-{{[(1R)-(N-(phenylmethoxycarbonyl)-L-alanyl)amino]ethyl}hydroxyphosphinyl}-L-3-phenyllactate [alias ZAG(P-O)F]; and R₁ = CH₃ and R₂ = H for O-{{[(1R)-(N-(phenylmethoxycarbonyl)-L-alanyl)amino]methyl}hydroxyphosphinyl}-L-3-phenyllactate [alias ZAGP(O)F] (26, 30). These and similar phosphorus-containing compounds were initially identified as high-affinity inhibitors of CPA with dissociation constants in the pico- or femtomolar range (26, 31). Other transition-state analogues based on sulfur-containing compounds have also been described and their structures determined (32, 33). In addition, structures representing the single-product complex of stage IV are available for porcine PCPA [Protein Data Bank (PDB) entry 1PCA (34)] and hCPA4 in complex with latexin [PDB entry 2BO9 (17)]. Here, leftover valine residues are found occupying the S₁ pocket. In contrast, structural information about stage III of the mechanism is restricted to complexes with the inhibitors PCI, LCI, and TCI (35–37). These protein inhibitors recognize the surface of the target molecule via large contact areas with prefixed structural cores, and their C-terminal tails are mostly disordered in the unbound structures (36, 38). These tails become rigid upon complex formation, and reminiscent of a stopperlike mechanism, they penetrate the active-site cleft in a substratelike manner that allows for an occasional hydrolysis of the C-terminal residues (35–37). However, superimposition of the complex structures of PCI, LCI, and TCI reveals that the conformation of these tails is conditioned by globular protein scaffolds, which accounts for the variability in the polypeptide chain traces.
Human Carboxypeptidase A4 in Complex with a Hexapeptide in the positions preceding S4. Early studies of hCPA had revealed that the nature of the side chains of at least five residues, P2–P1, influences K_{in} and K_{cat} (21, 28). Accordingly, a more accurate model for a bound and cleaved substrate, i.e., of a double-product complex, would be better provided by a structurally unrestrained oligopeptide.

We have determined the high-resolution three-dimensional structure of the first complex of an A/B-MCP with a free-standing oligopeptide in the form of hCPA4 bound to a hexapeptide emanating from the prodomain processed during the activation process. The peptide is tightly bound to the enzyme, and its C-terminal residue is severed and occupies the specificity pocket. Given the lack of restraints due to preceding peptide domains, this complex can be envisaged as a true double-product complex. The structure enables the structural determinants of the subsites shaping the nonprimed region of the active-site cleft to be assessed in detail and contributes to the unraveling of novel details of the catalytic mechanism of A/B-MCPs that had not been addressed previously.

**EXPERIMENTAL PROCEDURES**

**Protein Production and Purification.** hCPA4 was produced as a zymogen (hPCPA4) through recombinant heterologous overexpression using vector pPIC9 and the methylotrophic yeast *Pichia pastoris* as an expression host and purified as described elsewhere (17, 39). The active enzyme was obtained through tryptic activation (at a 200:1 w/w ratio) for 90 min at room temperature, and the resulting protein—product complex was subsequently purified by anion-exchange chromatography (TSK-DEAE 5PW) using a FPLC-Akt system with a linear salt gradient from 0 to 60% of 0.8 M ammonium acetate in 20 mM Tris-HCl (pH 9.5). Eluted fractions were analyzed by SDS–PAGE, and the purest samples containing the complex were pooled, desalted, and concentrated to ~15 mg/mL by Amicon centrifugal filter devices.

**Activity and Mass Spectrometry Measurements.** hCPA4 proteolytic activity was analyzed using peptides Phe-Asn-Arg-Pro-Val-Asp, Phe-Asn-Arg-Pro-Val-Val, and Phe-Asn-Arg-Pro-Val (purchased from EZBiolabs). The reaction took place in a 20 μL solution of 150 mM NaCl and 10 mM Tris-HCl (pH 8.0) with peptide and enzyme concentrations described elsewhere (39) with the chromogenic substrate Arg-Pro-Val (purchased from EZBiolabs). The reaction took 600

**Table 1: Crystallographic Statistics on Data Collection and Refinement**

<table>
<thead>
<tr>
<th>Space group</th>
<th>Crystallographic Statistics on Data Collection and Refinement</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2_12_12</td>
<td>no. of reflections</td>
</tr>
<tr>
<td>37869 (609)</td>
<td>50.0–1.60</td>
</tr>
<tr>
<td>1 thiocyanate, 1 glycerol</td>
<td>average intensity</td>
</tr>
<tr>
<td>(Zn^{2+})</td>
<td>13.0 (3.3)</td>
</tr>
<tr>
<td></td>
<td>estimated overall coordinate error</td>
</tr>
<tr>
<td></td>
<td>(Å)</td>
</tr>
<tr>
<td></td>
<td>no. of protein atoms, peptide atoms</td>
</tr>
<tr>
<td></td>
<td>rmsd from target values</td>
</tr>
<tr>
<td></td>
<td>2439/59/2671</td>
</tr>
<tr>
<td></td>
<td>solvent molecules, ions, other molecules</td>
</tr>
<tr>
<td></td>
<td>1 thiocyanate, 1 glycerol</td>
</tr>
</tbody>
</table>

* Values in parentheses refer to the outermost resolution shell if not otherwise indicated. b R_{free} = Σ_{i=1}^{n=1} [I_{i}(hkkl) – 〈I_{i}(hkkl)〉] / Σ_{i=1}^{n=1} [I_{i}(hkkl) + 〈I_{i}(hkkl)〉], and R_{free} = Σ_{i=1}^{n=1} [I_{i}(hkkl) – 〈I_{i}(hkkl)〉] / Σ_{i=1}^{n=1} [I_{i}(hkkl) + 〈I_{i}(hkkl)〉], where 〈I_{i}(hkkl)〉 is the rth intensity measurement, n_{hkkl} is the number of observations of reflection hkl, including symmetry-related reflections, and 〈I_{i}(hkkl)〉 is its average intensity. R_{free} and R_{free} are improved multiplicity-weighted indicators of the quality of the data, the redundancy-independent merging R-factor and the precision-indicating merging R-factor, the latter computed after averaging the multiple measurements (69, 70). c Crystallographic R_{factor} = Σ_{i=1}^{n=1} |F_{obs} – k|F_{calc}|/Σ_{i=1}^{n=1} F_{obs}|, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively. R_{free} same for a test set of reflections (>500) not used during refinement. d Including atoms in an alternate conformation.

Crystallography crystals were obtained by sitting drops subjected to vapor diffusion and containing 100 nL of hCPA4–hexapeptide complex [15 mg/mL in 5 mM Tris-HCl (pH 7.5)] and 100 nL of reservoir solution (0.2 M potassium thiocyanate and 20% PEG 3350). Crystallization drops were dispensed on 96 × 3-well Greiner plates by a Tecan robot and a Cartesian nanodrop robot (Genomic Solutions) at the joint IMBB-CSIC/Barcelona Science Park Automated Crystallization Platform (PAC). Crystals appeared after incubation for 10–15 days in a Bruker steady-temperature crystal farm at 4 °C. The cryoprotection protocol included soaking crystals in a mixture containing reservoir solution and 20% glycerol. A complete diffraction data set was collected at 100 K from a single N_{2}- flash-cryocooled (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 were orthorhombic, harbored one complex per asymmetric unit, and diffracted beyond 1.6 Å resolution. Diffraction data were integrated, scaled, merged, and reduced with MOSFLM (40) and SCALAM (41) within the CCP4 suite (42) (see Table 1).

**Structure Solution and Refinement.** The structure was determined by Patterson search methods with AMoRe (43) within ccp4i (44), employing all diffraction data up to 4 Å resolution. The coordinates of hCPA4, as in its complex with latexin (pdb entry 2BO9), were used as a searching model. A single solution was encountered at 50.5, 83.8, and 302.0
(α, β, and γ, respectively, in Eulerian angles) and 0.030, 0.437, and 0.136 (x, y, and z, respectively, as fractional unit-cell coordinates) after rigid-body fitting with the routine “fitting” (43). This solution gave a correlation coefficient in structure factor amplitudes (CCy) of 36.3% and a crystallographic R-factor of 40.8% (for definitions, see Table 1 and ref 43), with a second-highest unrelated peak, CCy = 10.8%, R-factor = 57.4%. Subsequently, manual model building on a SiliconGraphics graphic workstation using TURBO-Frodo (45) alternated with crystallographic refinement with REFMAC5 (46) within CCP4, until the final model was obtained (see Table 1). This model contained protein residues Asn5A-Leu308A of the mature protease moiety (for the numbering and nomenclature of hCPA4, see refs 17 and 39) and hexapeptide residues Phe1B-Asn2B-Arg3B-Pro4B-Val5B-Asp6B. All residues were in most favored and additionally favored regions of the Ramachandran plot except for Ser199A, which was in a generously allowed region, as seen in other h(PCPA4 structures and other (P)CPAs of the A/B subfamily like hPCPA2 (17, 39, 47). An N-glycosylation site consisting of a single N-acetylgalactosamine module (Nag901A) was modeled, attached to Asn148A N̈O2 based on weak electron density, in particular for the distal part of the sugar ring. Three peptide bonds were in the cis conformation (Ser197A—Tyr198A, Pro205A—Tyr206A, and Arg272—Asp273A), and a disulfide bond linked Cys138A and Cys161A. Residues Arg35A, Met47A, Val62A, Ser75A, Cys244A, His303A, and Asp6B were modeled with two conformations. The 50 loop (Gly55A—Arg58A) was flexible, and the electron density map did not allow the side chains to be placed unambiguously. Two hundred sixty-seven solvent molecules (Hoh503—Hoh769), one (tentatively assigned) glycerol molecule (Gol501), and one thiocyanate molecule (Scn502) were also identified. As there is only one polypeptide chain and one oligopeptide in the asymmetric unit, the chain identifiers (A and B, respectively) were omitted in the following.

Miscellaneous. Figures were prepared with TURBO-Frodo and SETOR (48); superimpositions were performed with TURBO-Frodo. The final coordinates of the hCPA4—hexapeptide complex were deposited with the Protein Data Bank at the RCSB (PDB entry 2PCU).

RESULTS AND DISCUSSION

Complex Preparation. Human CPA4 was produced as azymogen by heterologous recombinant overexpression in a P. pastoris system, purified, and activated with trypsin to produce crystals suitable for inhibitory drug design, given its potential role in human pancreatic cancer. To our surprise, the initial unbiased electron density map calculated after structure solution by Patterson search unambiguously revealed the presence of a hexapeptide bound to the active-site cleft, with its C-terminal residue severed and occupying the S1 pocket, i.e., featuring a double-product complex (stage III in Scheme 1; see Figure 1A). This finding is reminiscent of studies performed on γ-chymotrypsin, for which a residual tetrapeptide was encountered bound to the catalytic serine, i.e., featuring the acylenzyme intermediate (49, 50). This would be a precedent for the observation of a substrate in a protease active site following activation.

The excellent quality and high resolution of the experimental diffraction data of our complex with an unexpected ligand bound to the active site enabled us to identify the chemical sequence of the first five residues of the peptide as [Phe,His]-[Asp,Asn]-Arg-Pro-[Val,Thr]. The electron density map corresponding to the C-terminal residue clearly evinced four atoms in a plane adopting a “Y” structure adjacent to a sp3-hydrized atom (see Figure 1A). The shape of the map clearly indicated the presence of a planar group; i.e., it excluded the presence of a molecule only composed of tetrahedral carbon atoms as found in, for example, glycerol, a molecule that was considered during the initial stages of structure solution and refinement. The density map, taken together with the surrounding protein residues (in particular Arg145; see below and Figure 1E), suggested a carboxymethylene moiety attached (1.51 Å distance) to a carbon atom in sp3 hybridization. After several attempts to interpret this density, it became clear that no further atom was attached to the carbon next to the carboxyamide group. This ruled out the possibility that it was an α-carbon atom and restricted the identity of the residue to an aspartate that interacted with Arg145 through its side chain carboxylate and displayed alternate occupancy for its α-amino and α-carboxylate groups (see Figure 1A). This train of thoughts was confirmed by the refined B-factor values of the intervening atoms (overall B-factor for the whole residue, 23.2 Å2; whole peptide, 20.3 Å2; lowest for the carboxylate oxygens (13.3 and 14.6 Å2), as they constitute the main anchor points to the protein; average for the adjacent carboxylate and methylene carbon atoms (18.2 and 23.3 Å2, respectively); and highest for the remaining distal atoms in alternate occupancy (ranging from 23.8 to 28.5 Å2). Although acidic residues are not among the favorite residues of A-type CPs in P1, it has been reported that human CPA2 cleaves the C-terminal glutamate residue of LCI (36). Furthermore, the oxygen atoms of the α-carboxylate group evince rotational disorder and are undefined by electron density (Figure 1A). This finding correlates well with the different orientations found for isolated residues in S1 in the complexes of MCPs with PCI, LCI, and ECI and in porcine PCPA1 (17, 34, 36, 51) and accounts for a certain degree of liberty of the severed C-terminal residues to reallocate within the specificity pocket after cleavage, in particular if the preference of the enzyme for a hydrophobic side chain in the specificity pocket is not fulfilled. Sequence-pattern searches based on the previous sequence information performed on proteins of the expression host and the full-length hCPA4 protein found only one hit, Phe48-Asn49-Arg50-Pro51-Val52-Asp53 from the prodomain of hPCPA4 [for the complete sequence and numbering of h(PCPA4, see ref 39)]. Accordingly, this hexapeptide, whose sequence and numbering are Phe1-Asn2-Arg3-Pro4-Val5-Asp6, is likely to be a collateral product generated during the enzymatic activation of the metalloenzyme that was trapped and cleaved in the active-site cleft of hCPA4, where it withheld the final anion-exchange chromatography purification step and the crystallization conditions to give a true double-product complex, as depicted in Scheme 1. All atoms of the hexapeptide except the α-amino and α-carboxylate group of Asp6 and the α-carbon of Phe1 are unambiguously defined by excellent electron density. Their average temperature factor is lower than the average of the protease atoms (Table 1), thus accounting for the marked stability and rigidity of the complex. Indeed, the hexapeptide participates in building up
the crystal through a total of six crystal contacts with a neighboring hCPA4 molecule.

Cleavage and Inhibitory Assays in Vitro. The sequence of the active-site-bound hexapeptide corresponds to a region
where the activation domain folds in a loop and the beginning of strand $\beta_3$ (39). Mapping hCPA4 for tryptic digestion indicates that there are two putative cleavage sites for trypsin, 4 and 12 residues upstream of the position of the hexapeptide sequence, respectively. Thus, the hexapeptide detected is likely to be a mature product of a series of cleavage events of undetermined sequence within the prodomain, or, alternatively, a byproduct due to a minor chymotryptic activity present in the trypsin used for activation. To test the hypothesis of peptide uptake and cleavage, the ability of hCPA4 to cleave the C-terminal residue of the molecule bound at the enzyme active site was examined with a synthetic hexapeptide identical to the one found in the crystal (Phe-Asn-Arg-Pro-Val-Asp). In addition, a second hexapeptide with a C-terminal valine instead of an aspartate was used to fit the preference of A-type MCPs for aliphatic residues.

Figure 2 depicts some points of the time course of peptide degradation as followed by MS analysis (52). A detailed analysis of the time course degradation of the peptide Phe-Asn-Arg-Pro-Val-Asp shows that, although the MALDI-TOF analysis lacks quantitative precision, a decrease in the amount of the resulting pentapeptide occurred with time, suggesting some kind of enzyme inhibition.

Activity measurements with chromogenic substrates of the enzyme incubated with increasing quantities of both hexapeptides gave the IC$_{50}$ values that are shown in Figure 2B. Both peptides behave as inefficient inhibitors of hCPA4 in the high micromolar range, with their common pentapeptide product, Phe-Asn-Arg-Pro-Val, being the least inhibitory species. The observable, albeit low, inhibitory activity of the Phe-Asn-Arg-Pro-Val-Asp peptide may explain both the persistence of the double-product complex and its stability during purification and crystallization procedures. These studies were complemented with the inhibitory analysis of further pentapeptides and hexapeptides (see Figure 2B) that confirmed (i) the significantly higher inhibitory activity of hexapeptides over pentapeptides, (ii) the requirement of a bulky aromatic residue in P$_5$, and (iii) the fact that proline is not an absolute requirement in P$_2$.

**Overall Structure of hCPA4.** The mature hCPA4 enzyme has a compact globular shape and shows the classic $\alpha/\beta$-hydrolase fold of A/B- and N/E-type MCPs. It has a central eight-stranded $\beta$-sheet made up of strands $\beta_5-\beta_{12}$ (numbering of the regular secondary structure elements according to the zymogen; see ref 39) of strand connectivity $+1,+2,-1x,-2x,-2,+1x,-2$ according to ref 53. Metalloprote(in)ases (MPs) are best shown in a canonical standard orientation.
giving rise, together with the protein ligands, to the cation being coordinated in a distorted trigonal-bipyramidal fashion by four oxygen and two nitrogen atoms. Val5 O is also within binding distance of atoms Oe1 (2.96 Å) and Oe2 (2.90 Å) from the general base/acid Glu270, suggesting that either the latter residue or Val5 O is protonated; i.e., the second proton transfer to the α-amino group of Asp6 in S1 has not yet occurred (see Scheme 1). Val5 OT, in turn, establishes a salt bridge with Arg127 Nη2 (2.85 Å), a residue that stabilizes the negatively charged reaction intermediate as seen in the ZFV\textsuperscript{3}(O)F complex (see below). The P1 side chain of Val5 is placed in the S\textsubscript{1} subsite shaped in hCPA4 and in bCPA by the side chains of Phe279 and Tyr198. Val5 N is bound by Ty248 Oη (2.77 Å). The preceding residue in P2 (Pro4) shows its main chain torsion angles in the β-region of a Ramachandran plot (Φ = −66°; Ψ = 157°), and its carbonyl oxygen is bound by Arg71 Nη2 (3.06 Å) and, more weakly, by Arg127 Nη2 (3.28 Å). The P2 side chain is accommodated in shallow subsite S\textsubscript{2}, shaped by the apolar side chain parts of Arg127, Glu163, and Val164 (Thr164 in bCPA). Arg3, the peptide residue in P3 (Φ = −130°; Ψ = 154°, also in the β-region), is bound via its carbonyl oxygen and its α-amino group with Tyr198 Oη through a solvent-mediated (Hoh697) hydrogen bond. The P3 side chain protrudes away, running down along the enzyme surface, and is held in place by Glu163 (Arg3 Nη1−Glu163 Oε1, 3.09 Å), which adopts a novel conformation (see below). Interestingly, the position equivalent to 163 is always acidic in the reported CP sequences, except for human CPO, where a glutamine is found. The side chain of Arg71, Leu125, and, to a lesser extent, Phe279 further contribute to the shallow S\textsubscript{3} site. P4 residue Asn2 points to the bulk solvent, while Phe1, featuring P3, nestles snugly into a mainly hydrophobic pocket formed by the aromatic part of Tyr198, Tyr248, Val247 (Ile247 in bCPA), and Ser199, as well as the side chain of the product P1 residue Val5. The participation of Ser199 and Tyr198 in subsite shaping may provide an explanation for (i) the main chain angles of the former residue systematically corresponding to disallowed or generously allowed regions of a Ramachandran plot and (ii) Ser197 and Tyr198 being linked by a conserved cis peptide bond.

On the opposite side of the groove, Asp6 fits into the S\textsubscript{1} or specificity pocket, which is conceived to accommodate hydrophobic side chains. This residue binds the primary substrate-anchoring residue in MCPs, Arg145, through a bidentate salt bridge with its β-carboxylate group instead of its α-carboxylate group (Asp6 Oδ2−Arg145 Nη2, 2.77 Å; Asp6 Oδ1−Arg145 Nη1, 2.89 Å). Asp6 Oδ2 is further anchored to Tyr248 Oη (2.58 Å) and Arg127 Nη1 (3.11 Å), while Asp6 Oδ1 additionally contacts Asn144 Nδ2 (2.92 Å). These interactions are usually performed by the α-carboxylate group in non-aspartate C-terminal residues of substrates (see refs 17, 26, and 34 for examples). As Asp6 is present with multiple conformations for its α-amino- and -carboxylate groups, it can be stated in merely tentative terms that interactions of these groups may occur with Glu270 Oε2, Asn144 Nδ2, and Thr268 Oγ1. In accordance with previous descriptions for bCPA, the S1 pocket is delimited further in hCPA4 by the side chains of Asp194 (Ser194 in bCPA), Met203 (Leu203), Thr243 (Ile243), Val247 (Ile247), Ala250, Ser253 (Gly253), Ser254, Ile255, and Thr268. These residues
render a dead-end pocket characteristic of exopeptidases.

Among these residues, Asp194 may be hallmarkpered as being unique for hCPA4, since its position is always occupied by a serine or a threonine in mammalian A/B-MCPs (see ref 39). Here, it is simultaneously strongly bound to one of the zinc-binding protein ligands (Asp194 Oδ2—His196 Ne2, 2.60 Å) and to Ser254, at the bottom of the cleft (Asp194 Oδ2—Ser254 Oγ2, 2.73 Å). This arrangement is reminiscent of the catalytic triad of serine proteases, with the relative positions of the aspartate and the histidine residues swapped. The strong interaction between the latter two residues may contribute to the enhancement of the nucleophilicity of the zinc ligand, which, in turn, affects the catalytic efficiency of hCPA4, when compared with hCPA. One of the few instances in which position 194 is occupied by an aspartate is in insect CPA from *Helicoverpa armigera*. Interestingly, this enzyme exhibits broad substrate specificity and excises C-terminal acidic residues (57) with an efficiency comparable to that of a human CPB specifically engineered to display such specificity (58).

**Comparison with a Stage III Model.** As reported, structural information about the reaction intermediate of MCP catalysis relies on complexes with phosphinyl-derived peptidomimetic compounds. A search through the Cambridge Structural Database for small-molecule structures (59) with the motif C-PO2-[O,NH,CH2]-C found that phosphinyl P—O bonds span 1.49 ± 0.04 Å (n = 50), while the bonds mimicking the scissile peptide bond in the compounds found, P—O, P—N, and P—C, span 1.60 ± 0.03 (n = 25), 1.69 (n = 1), and 1.81 ± 0.01 Å (n = 2), respectively. In contrast, the only available structure mimicking a tetrahedral reaction intermediate, 1,1,1,3,3,3-hexafluoropropanediolate [(CF3)2-C(O)-OH], evinces C—C bonds that are 1.54 ± 0.01 Å long (n = 2) and C—O bonds that are 1.38 ± 0.02 Å long (n = 2) (60). Despite these differences in bond lengths and the pentavalence of phosphorus, phosphinyl-based compounds are widely accepted as valid models for a reaction intermediate and provide an unrivaled aid to the understanding of enzyme catalysis and specificity (61) due to the similarities in hybridization and geometry (29, 62). Furthermore, they constitute a validated scaffold for the development of potent inhibitors of zinc proteases (63).

Among the three compound classes, phosphonamidates are less well suited due to the intrinsic instability of the P—N bond (61, 63). Moreover, phosphinates should be preferred over phosphonates due to the reported capacity of the phosphinyl-adjacent methylene to participate in substratelike nitrogen atom. A second proton is transferred to the amide nitrogen atom of the peptide bond between P1 and P2 and the catalytic zinc ion coordinates the scissile carbonyl oxygen atom and to provide an update (see Scheme 1). As reported, the zinc ion coordinates the scissile carbonyl oxygen atom and the catalytic solvent molecule polarized by Glu270. Substrate binding further entails that Tyr248 makes a large movement from a superficial “up” position in the unbound state to a “down” position (34, 47, 64, 65), although there is some controversy about this point (66). In the down position, the tyrosine hydroxyl group hydrogen bonds the amide nitrogen atom of the peptide bond between P1 and P2 and the C-terminal carboxylate group (26). This group is further anchored to the active-site cleft by a bidentate salt bridge with Arg145 and by a hydrogen bond with Asn144 Nδ2. This process can be described in four steps. (i) In the first step, Glu270 acts as a general base and abstracts a proton from the catalytic solvent molecule, which conducts a nucleophilic attack on the scissile carbonyl atom. (ii) This leads to a negatively charged tetrahedral reaction intermediate that is stabilized by the side chain of Arg127 (67) and possibly interacts in a bidentate manner with the zinc ion that becomes pentacoordinate (26). This basic residue replaces structural elements forming what is known as the “oxyanion hole”, which stabilizes similar types of intermediates during cleavage reactions performed by other proteases (68). Glu270 subsequently acts as a general acid catalyst delivering the captured proton to the scissile amide nitrogen atom. A second proton is transferred to the amide nitrogen, and (iii) this results in the metal-bound tetrahedral intermediate collapsing to products that initially remain bound to the enzyme in the form of a double-product complex, described here for the first time. At this stage, the enzyme can perform the reverse reaction, i.e., the synthesis of a peptide bond (24). The new C-terminus of the upstream part of the substrate interacts with the catalytic zinc in a bidentate manner, and the amide nitrogen of the preceding peptide bond remains hydrogen-bonded to Tyr248 Oη. In turn, the novel N-terminus probably interacts with Glu270 Oε2 and the catalytic zinc ion, but the isolated amino acid in S5 could be free to reallocate within the cleft if it does not fit the preference of the enzyme for a hydrophobic side chain. (iv) The N-terminal product leaves, and the catalytic solvent molecule is replenished; the C-terminal product is still bound to the zinc ion, Glu270, and Arg145 (and for the hydrophobic nature of S5, comparison of the complexes strongly suggests that the absence of residues and/or atoms mimicking the upstream part of a substrate leads the Z moiety to rotate ~180° around the bond equivalent to the Arg3—Pro4 peptide bond (Figure 1D). This rotation entails the aromatic Z group occupying S5.

A further noteworthy point is the novel conformation of the Glu163 side chain due to, we believe, a hitherto unseen interaction with the side chain of the residue nestling in S1. This interaction is particularly important, as it involves the acidic side chain rotating anticlockwise ~80° around Y2, in comparison with the reaction-intermediate-mimicking ZFV^P(O)F complex of bCPA. This conformation of Glu163 is buttressed by a bidentate salt bridge with Arg71 (Glu163 Oe1—Arg71 Nη1, 3.06 Å; Glu163 Oe2—Arg71 Nη2, 2.89 Å) and can be envisaged as being pivotal for P1 recognition.

**An Updated Reaction Mechanism for MCPs.** The structure described here, along with the number of structural studies performed on MCPs, enables us to add some details to the consensus mechanism, in particular for secondary players, and to provide an update (see Scheme 1). As reported, the zinc ion coordinates the scissile carbonyl oxygen atom and the catalytic solvent molecule polarized by Glu270. Substrate binding further entails that Tyr248 makes a large movement from a superficial “up” position in the unbound state to a “down” position (34, 47, 64, 65), although there is some controversy about this point (66). In the down position, the tyrosine hydroxyl group hydrogen bonds the amide nitrogen atom of the peptide bond between P1 and P2 and the C-terminal carboxylate group (26). This group is further anchored to the active-site cleft by a bidentate salt bridge with Arg145 and by a hydrogen bond with Asn144 Nδ2. This process can be described in four steps. (i) In the first step, Glu270 acts as a general base and abstracts a proton from the catalytic solvent molecule, which conducts a nucleophilic attack on the scissile carbonyl atom. (ii) This leads to a negatively charged tetrahedral reaction intermediate that is stabilized by the side chain of Arg127 (67) and possibly interacts in a bidentate manner with the zinc ion that becomes pentacoordinate (26). This basic residue replaces structural elements forming what is known as the “oxyanion hole”, which stabilizes similar types of intermediates during cleavage reactions performed by other proteases (68). Glu270 subsequently acts as a general acid catalyst delivering the captured proton to the scissile amide nitrogen atom. A second proton is transferred to the amide nitrogen, and (iii) this results in the metal-bound tetrahedral intermediate collapsing to products that initially remain bound to the enzyme in the form of a double-product complex, described here for the first time. At this stage, the enzyme can perform the reverse reaction, i.e., the synthesis of a peptide bond (24). The new C-terminus of the upstream part of the substrate interacts with the catalytic zinc in a bidentate manner, and the amide nitrogen of the preceding peptide bond remains hydrogen-bonded to Tyr248 Oη. In turn, the novel N-terminus probably interacts with Glu270 Oε2 and the catalytic zinc ion, but the isolated amino acid in S5 could be free to reallocate within the cleft if it does not fit the preference of the enzyme for a hydrophobic side chain. (iv) The N-terminal product leaves, and the catalytic solvent molecule is replenished; the C-terminal product is still bound to the zinc ion, Glu270, and Arg145 (and
putatively to Asn144 and Tyr248) (34, 39) featuring the single-product complex. Eventually, the amino acid in $S_1$ leaves and the enzyme is poised for a new catalytic round.

ACKNOWLEDGMENT

We are grateful to Robin Rycroft for helpful discussions on the manuscript and acknowledge the help provided by EMBL and ESRF synchrotron local contacts during data collection.

REFERENCES


distance geometry, and restrained molecular dynamics, *Biochemistry* 26, 8012–8023.


