# Catalytically distinct states captured in a crystal lattice: the substrate-bound and scavenger states of acylaminoacyl peptidase and their implications for functionality

Dóra Karancsiné Menyhárd,[a] Zoltán Orgován,[b] Zoltán Szeltner,[c] Ilona Szamosi[c] and Veronika Harmat[a,b]*

[a]MTA-ELTE Protein Modelling Research Group, Pázmány Péter sétány 1/A, 1117 Budapest, Hungary, [b]Laboratory of Structural Chemistry and Biology, Institute of Chemistry, Eötvös Loránd University, Pázmány Péter sétány 1/A, 1117 Budapest, Hungary, and [c]Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Magyar tudósok körútja 2, 1117 Budapest, Hungary. *Correspondence e-mail: veronika@chem.elte.hu

Acylaminoacyl peptidase (AAP) is an oligopeptidase that only cleaves short peptides or protein segments. In the case of AAP from *Aeropyrum pernix* (ApAAP), previous studies have led to a model in which the clamshell-like opening and closing of the enzyme provides the means of substrate-size selection. The closed form of the enzyme is catalytically active, while opening deactivates the catalytic triad. The crystallographic results presented here show that the open form of ApAAP is indeed functionally disabled. The obtained crystal structures also reveal that the closed form is penetrable to small ligands: inhibitor added to the pre-formed crystal was able to reach the active site of the rigidified protein, which is only possible through the narrow channel of the propeller domain. Molecular-dynamics simulations investigating the structure of the complexes formed with longer peptide substrates showed that their binding within the large crevice of the closed form of ApAAP leaves the enzyme structure unperturbed; however, their accessing the binding site seems more probable when assisted by opening of the enzyme. Thus, the open form of ApAAP corresponds to a scavenger of possible substrates, the actual cleavage of which only takes place if the enzyme is able to re-close.

## 1. Introduction

Acylaminoacyl peptidase [AAP; also called acylpeptide hydrolase (APEH)] is a member of the prolyl oligopeptidase protein family, which consists of serine protease enzymes that impose size selection on their substrates, cleaving only relatively short, unstructured peptides or protein segments (Polgár, 2002). Mammalian AAP is specifically able to remove N-terminal acetylated amino acids from peptides, but is also able to hydrolyze longer peptide chains. This lends the enzyme its physiological significance: AAP has been shown to exert regulation of protein maturation and degradation processes (Perrier *et al.*, 2005; Arnesen, 2011; Forte *et al.*, 2011; Sandomenico *et al.*, 2012) and cell survival (Fujino *et al.*, 2000; Shimizu *et al.*, 2003; Adibekian *et al.*, 2011), most probably through being an upstream modulator of the proteosome (Shimizu *et al.*, 2004; Palmieri *et al.*, 2011). It has also even been shown to be able to degrade multimeric forms of the amyloid β-peptide present in the Alzheimers brain (Yamin *et al.*, 2009).

Since the crystal structure of mammalian AAP has not yet been determined, structural considerations must rely on the structures of two archaeal analogues: AAP from *Pyrococcus*
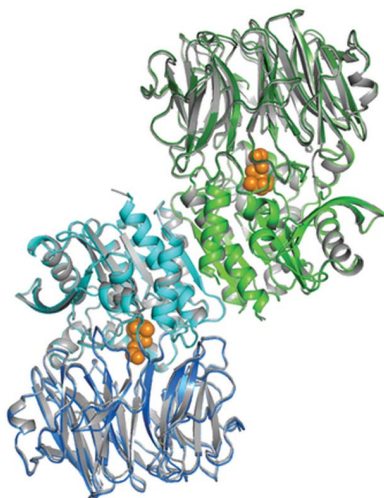
*horikoshii* (PhAAP) and *Aeropyrum pernix* (ApAAP) (Bartlam *et al.*, 2004; Menyhárd *et al.*, 2013). All oligopeptidases share a common modular structure, being composed of a hydrolase and a propeller domain. The hydrolase domain contains the Ser–His–Asp amino acids of the catalytic triad, while the propeller domain provides the means of size selection, shielding the active site from the exterior. Between the two, a spacious substrate-binding cavity can be found, which can be accessed by one of the following pathways: (i) through a narrow channel in the propeller domain, (ii) *via* a side gate (Rasmussen *et al.*, 2003; Engel *et al.*, 2003; Menyhárd *et al.*, 2013) or (iii) by the opening of the enzyme, as indicated by different X-ray structures, molecular-dynamics and enzyme-kinetic studies (Szeltner *et al.*, 2004; Fuxreiter *et al.*, 2005; Shan *et al.*, 2005; Li *et al.*, 2010; Kaszuba *et al.*, 2012; Canning *et al.*, 2013; Fig. 1). It is evident that if reaching the active site is an elaborate process (such as passage through a long, narrow passage, as in the first case; Fig. 1*a*) then enzyme efficiency is in danger; on the other hand, if ready access is provided to the active site (such as in the latter two cases, Figs. 1*b* and 1*c*) then the exclusion of larger substrates, and thus selectivity, is hard to maintain. Different oligopeptidases adopt different strategies for simultaneous management of both size selection and access to the active site. They also assume different multimerization states: prolyl oligopeptidase (PREP) and certain forms of oligopeptidase B are monomeric (Fülöp *et al.*, 1998; McLuskey *et al.*, 2010), ApAAP, DPP-IV and oligopeptidase B from other sources form dimers (Bartlam *et al.*, 2004; Rasmussen *et al.*, 2003; Canning *et al.*, 2013) and the active form of PhAAP is a hexamer (Szeltner *et al.*, 2009), while mammalian AAP is tetrameric (Mitta *et al.*, 1989).

In the case of ApAAP, previous studies revealed that the monomers of the ApAAP dimer can adopt an open or a closed conformation independently of one another. The active-site conformation is distorted in the open state, which led us to propose a model in which a clamshell-like opening of the enzyme provides the means of approach (especially for larger ligands), while this same movement simultaneously deactivates the catalytic triad through the loss of interdomain contacts (Harmat *et al.*, 2011). The crystallographic and simulation results presented here show that the open form of ApAAP is indeed functionally disabled and indicate that ApAAP provides different entrance mechanisms for substrates of differing lengths and structural complexities: substrates that are small enough to pass through the propeller channel may directly approach the fully functional active site of the closed form, while bulkier candidates may ligate to the catalytically impaired but approachable active site of the open form, where their compliance with the closure of the enzyme will provide the means of size selection.

## 2. Materials and methods

### 2.1. Crystallization of ApAAP–CMK complexes

ApAAP was prepared and purified for crystallization as described in Kiss *et al.* (2007). Previously published crystal-

lization conditions of ApAAP open/closed crystal forms were optimized to obtain crystals of ApAAP with its covalent chloromethyl ketone inhibitor (benzyloxycarbonyl-glycyl-glycyl-phenylalanyl-chloromethyl ketone; CMK). The crystals were grown at 293 K using the hanging-drop vapour-diffusion method. Crystal 1 was grown in a drop obtained by mixing 3 µl protein solution (221 µ*M* ApAAP and 560 µ*M* CMK in 20 m*M* Tris pH 7.5 buffer) and 3 µl reservoir solution [78 m*M* sodium acetate buffer pH 4.5, 2.4%(*w/v*) PEG 4000, 6.7 m*M* dithiothreitol, 0.44 m*M* EDTA]. The crystal was cryoprotected in a solution with a similar composition to that of the reservoir condition but also containing 23%(*v/v*) 2-methyl-2,4-pentanediol. Crystal 2 was grown in a drop obtained by mixing 3 µl protein solution (158 µ*M* ApAAP in 20 m*M* Tris pH 7.5 buffer) and 3 µl reservoir solution [78 m*M* sodium acetate buffer pH 5.0, 2.2%(*w/v*) PEG 4000, 5.2 m*M* dithiothreitol, 0.34 m*M* EDTA]. Crystal 2 was soaked in reservoir condition containing 1.58 m*M* CMK overnight, cryoprotected in perfluoropolyether oil (MiTeGen LV CryoOil) and flash-cooled in liquid nitrogen.

### 2.2. Data collection, processing and refinement of crystal structures

All diffraction data were collected at 100 K. A data set was collected from crystal 1 on a home diffractometer (Rigaku RU200 generator, blue optics, R-AXIS IV++ detector; wavelength 1.5418 Å, φ range of 100.5° with an increment of 0.5°; crystal-to-detector distance 100.03 mm). The data set from
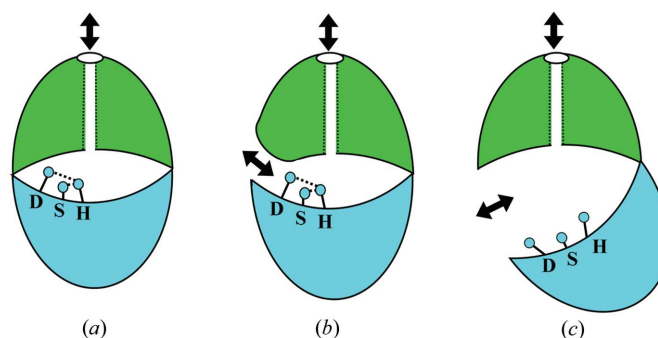


**Figure 1**
Schematic models of oligopeptidase monomers. The hydrolase domain with the catalytic triad is shown in cyan, while the propeller domain is coloured green. (*a*) The closed form found in structures of both the ligand-free and the ligand-bound states of prolyl oligopeptidases from various organisms (for examples, see Fülöp *et al.*, 1998, 2001; Szeltner, Rea, Juhász *et al.*, 2002; Szeltner, Rea, Renner *et al.*, 2002; Shan *et al.*, 2005; Li *et al.*, 2010), oligopeptidase B (McLuskey *et al.*, 2010; Canning *et al.*, 2013), puromycin hydrolase (Matoba *et al.*, 2011) and AAP from *A. pernix* (Bartlam *et al.*, 2004; Kiss *et al.*, 2007, 2008; Harmat *et al.*, 2011). The catalytic triad is only accessible through the propeller channel. (*b*) A monomer with a permanent side entrance as found in DPP-IV (for examples, see Rasmussen *et al.*, 2003; Engel *et al.*, 2003; Thoma *et al.*, 2003; Nordhoff *et al.*, 2006; Nakajima *et al.*, 2008), tripeptidyl aminopeptidase (Ito *et al.*, 2006) and AAP from *P. horikoshii* (Menyhárd *et al.*, 2013). These enzymes are active in multimer forms, and the permanent entry is shielded by the other interacting monomers. (*c*) The open form seen in the case of certain bacterial prolyl oligopeptidases (Shan *et al.*, 2005; Li *et al.*, 2010), oligopeptidase B (Canning *et al.*, 2013) and AAP from *A. pernix* (Harmat *et al.*, 2011).

**Table 1**
Crystallographic data-collection and refinement statistics.

Values in parentheses are for the outer shell.

| Crystal | 1 | 2 |
|---|---|---|
| Global conformation of the AAP monomers within the dimer | Closed/closed | Open/closed |
| PDB entry | 4re5 | 4re6 |
| Data-collection statistics | | |
|   Diffraction source | Rigaku RU200 rotating anode | ESRF beamline ID14-1 |
|   Wavelength (Å) | 1.5418 | 0.9334 |
|   Detector | Rigaku R-AXIS IV$^{++}$ image plate | ADSC Quantum 210 CCD |
|   Space group | $P2_12_12_1$ | $P1$ |
|   Unit-cell parameters | | |
|     $a$ (Å) | 63.81 | 71.58 |
|     $b$ (Å) | 104.43 | 97.30 |
|     $c$ (Å) | 170.08 | 99.16 |
|     $\alpha$ (°) | 90 | 105.15 |
|     $\beta$ (°) | 90 | 103.96 |
|     $\gamma$ (°) | 90 | 100.26 |
|   Resolution range (Å) | 20.0–1.90 (1.95–1.90) | 20.0–2.55 (2.62–2.55) |
|   Total No. of reflections | 354000 (25117) | 149085 (6410) |
|   No. of unique reflections | 90022 (6591) | 68281 (2964) |
|   Completeness (%) | 99.7 (99.9) | 86.4 (51.1) |
|   Multiplicity | 3.93 (3.81) | 2.18 (2.16) |
|   $\langle I/\sigma(I)\rangle$ | 17.58 (2.27) | 13.38 (2.16) |
|   $R_{\text{r.i.m.}}$ | 0.062 (0.682) | 0.071 (0.505) |
|   $CC_{1/2}$ | 0.999 (0.750) | 0.997 (0.775) |
|   Overall $B$ factor from Wilson plot (Å$^2$) | 32.79 | 41.30 |
| Refinement statistics | | |
|   No. of reflections, working set | 85520 (6271) | 64833 (2782) |
|   No. of reflections, test set | 4499 (315) | 3441 (126) |
|   Final $R_{\text{work}}$ | 0.177 (0.294) | 0.212 (0.285) |
|   Final $R_{\text{free}}$ | 0.218 (0.352) | 0.260 (0.350) |
|   No. of non-H atoms | | |
|     Protein | 8591 | 17351 |
|     Ligand | 64 | 43 |
|     Ion | 5 | 6 |
|     Water | 662 | 674 |
|   R.m.s.d. from ideal geometry | | |
|     Bonds (Å) | 0.016 | 0.011 |
|     Angles (°) | 1.720 | 1.470 |
|   Average $B$ factors (Å$^2$) | | |
|     Protein | 36.4 | 34.6 |
|     Ligand | 44.4 | 34.7 |
|     Ion | 57.4 | 41.7 |
|     Water | 38.0 | 30.8 |
|   Ramachandran plot | | |
|     Most favoured (%) | 97.1 | 96.9 |
|     Allowed (%) | 2.9 | 3.1 |

The structures were solved using the difference Fourier method with the *CCP*4 package (Winn *et al.*, 2011) using the protein atoms of the isostructural structures with PDB entries 2hu5 (Kiss *et al.*, 2007; resolution 2.0 Å) and 3o4h (Harmat *et al.*, 2011; resolution 1.8 Å), respectively. 5.0% of the reflections for a test set for monitoring the refinement process were imported from the data sets of the isostructural model structures. The models were refined by likelihood-based refinement using *REFMAC* v.5.7.032 (Murshudov *et al.*, 2011), in which TLS refinement (Winn *et al.*, 2001) was carried out for quaternary-structure regions (one TLS group per domain for crystal 1 and one TSL group per protein monomer for crystal 2) and noncrystallographic restraints were introduced for regions with similar conformations in each monomer. Model building was carried out with *Coot* (Emsley *et al.*, 2010) using graphical comparison of noncrystallographically related monomers (NCS ghost control option). Water molecules were added to the model using the water-picking mode of *Coot* (within hydrogen-bonding distances of hydrogen-bond donor or acceptor atoms, with $mF_o - DF_c$ density of greater than $3.0\sigma$ and checked for refined $B$ factors of less than 80 Å$^2$).

For crystal 1 the asymmetric unit contains one ApAAP dimer, with both molecules in the closed conformation with convalently bound CMK inhibitors in their active sites, 669 water molecules, three 2-methyl-2,4-pentanediol molecules, one acetate ion and one chloride ion. 19 residues and seven water molecules are modelled in alternate conformations/positions. For crystal 2 the asymmetric unit contains two ApAAP dimers with one of the monomers in the open conformation and the other in the closed conformation, two CMK inhibitors covalently bound to the two closed monomers, 679 water molecules and six chloride ions. 15 residues were modelled in alternate conformations: serines and a valine with electron density between the two rotamers and longer side chains with clear electron density for both rotamers (at the $1.2\sigma$ contour level of $2mF_o - DF_c$ density). Refinement statistics and model quality are compiled in Table 1. The coordinates and structure factors have been deposited in the Protein Data Bank as entries 4re5 and 4re6, respectively. The quality of the final structures was validated using *MolProbity* (Chen *et al.*, 2010). The r.m.s. fit of backbone atoms of crystal

crystal 2 was collected on beamline ID14-1 at the European Synchrotron Radiation Facility (ESRF; wavelength 0.9334 Å, $\varphi$ range of 200.0° with an increment of 0.5°; crystal-to-detector distance 298.55 mm). The data sets were processed to 1.90 and 2.55 Å resolution for crystal 1 and crystal 2, respectively, using the *XDS* and *XSCALE* programs (Kabsch, 2010.). The data-collection statistics are summarized in Table 1. Note that the data set from crystal 2 has low completeness in the 2.8–2.55 Å resolution range (Supplementary Table S1), but these data were also used in refinement as a slight improvement in map quality (Supplementary Fig. S1) and $R$ factors could be achieved (at the early stage of refinement paired refinement to 2.8 and 2.55 Å resolution resulted in a decrease in $R$ and $R_{\text{free}}$ of 0.002 and 0.005, respectively).

structures was calculated within the *CCP*4 package (Winn *et al.*, 2011). Figures were prepared using *PyMOL* v.0.99

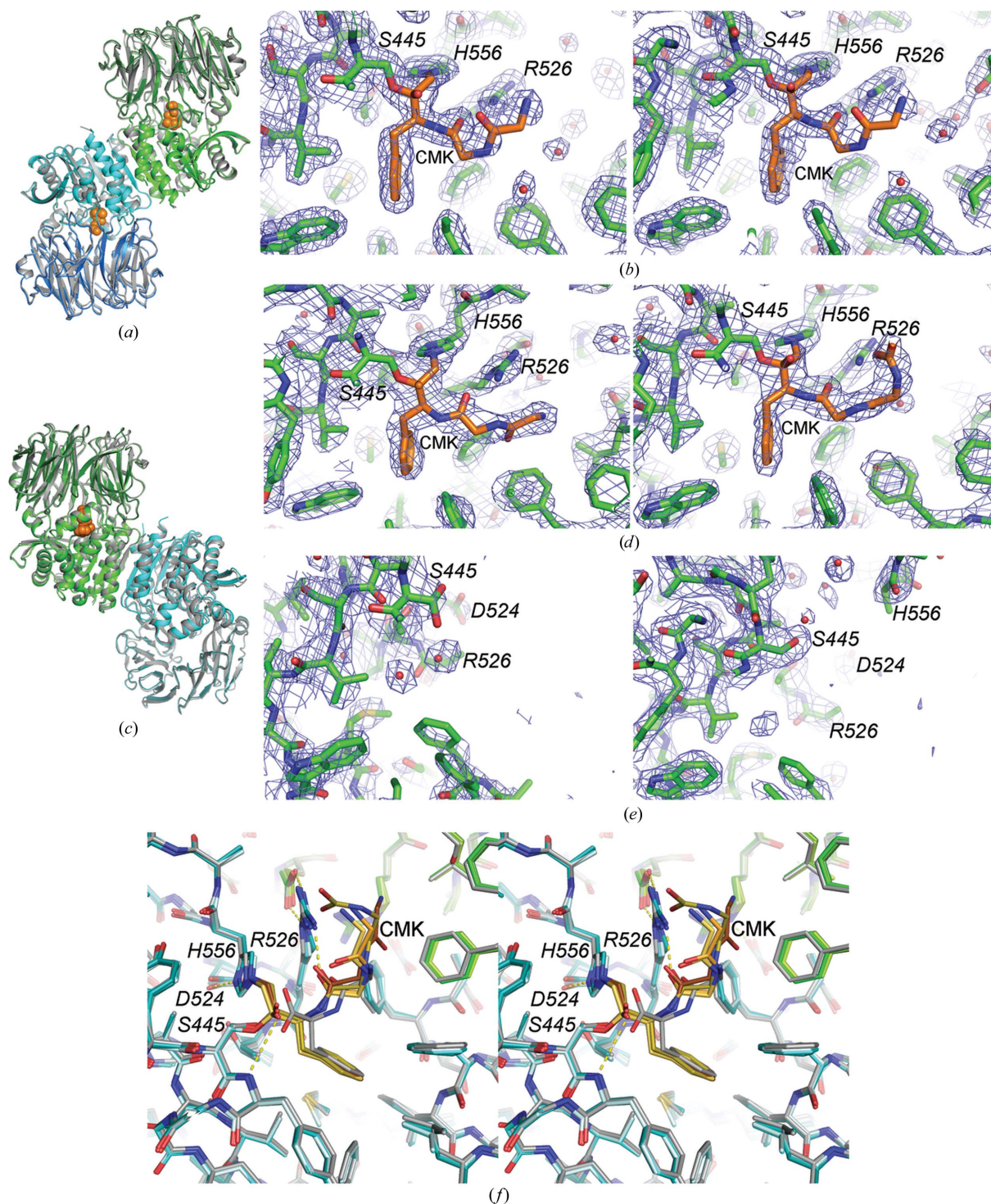(http://www.pymol.org) and *Maestro* from the Schrödinger package.



**Figure 2**
Crystal structures of the ApAAP–chloromethyl ketone complex. (*a*) Overall conformation of the closed/closed ApAAP dimer (cyan and green, with the propeller domains in darker colours) with the bound inhibitor (orange). (*b*) The bound inhibitors in the active sites shown with $2mF_o - DF_c$ electron density contoured at 1.2σ. (*c*) One of the two ApAAP dimers of the open/closed structure shown with the inhibitor bound only at the active site of the closed monomer. (*d*, *e*) The bound inhibitors in the active sites of the closed monomers (*d*) and the empty active sites of the open monomers (*e*) of the open/closed structure shown with $2mF_o - DF_c$ electron density contoured at 1.2σ. (*f*) Comparison of the conformations of the bound CMK inhibitors (orange tones) with the hydrolase domains shown in cyan and the propeller domain shown in green (stereoview). The structure of the ApAAP–gylcyl-phenylalanine enzyme–product-type complex (PDB entry 2hu5) is shown in grey for reference. Residues of the catalytic triad (Ser445, His556 and Asp524) as well as Arg526 are labelled.

## 2.3. Monte Carlo multiple minimum docking of substrates

10 000-step Monte Carlo multiple minimum (MCMM) searches (Chang *et al.*, 1989), as implemented in *MacroModel* (v.10.2; Schrödinger), were carried out involving the random variation (within the range 0–180°) of a randomly selected subset of all internal torsional angles of the substrates and a random rigid-body translation (0–5 Å) and rotation (0–180°) step of the substrate with respect to the enzyme within the fixed interior of the enzyme in the Monte Carlo step. The perturbed structures were energy-minimized allowing atoms of the substrate to move freely, while applying a 200 kJ mol$^{-1}$ Å$^{-2}$ restraint to side chains of the propeller domain and freezing the rest of ApAAP. Unique structures were stored within a 21 kJ mol$^{-1}$ energy window above the global minimum. Calculations were carried out using the AMBER force field (Cornell *et al.*, 1995). The solvent effect was modelled by the *GB/SA* algorithm (using water as the solvent).

## 2.4. Molecular-dynamics simulations

The MCMM-derived structures were subjected to molecular-dynamics (MD) simulations as implemented in *GROMACS* (Hess *et al.*, 2008) using the CHARMM27 force field (MacKerell *et al.*, 2004; Bjelkmar *et al.*, 2010). Systems were solvated by approximately 17 000 TIP3P water molecules, the total charge of the system was neutralized and a physiological salt concentration (0.15 *M*) was set using Na$^+$ and Cl$^-$ ions. Conjugate-gradient energy minimization of starting structures was followed by sequential relaxation of constraints on protein atoms in three steps and an additional NVT step (all of 200 ps) to stabilize the pressure. Trajectories of 200 ns NPT simulations [where heavy-atom hydrogen bonds were constrained using the *LINCS* algorithm (Hess, 2008) but no other geometrical restraints were applied] at 300 K and 1 bar [applying the velocity-rescale algorithm (Bussi *et al.*, 2007) and a Berendsen barostat] were recorded for further analysis (collecting snapshots every 4 ps). Average structures for the last 100 ns were calculated and used for comparison of models and *B*-factor analysis. For figures, the snapshot with the lowest backbone root-mean-square deviation (r.m.s.d.) compared with the average structure was selected and energy-minimized. Clustering of conformations (Daura *et al.*, 1999) was carried out using a cutoff of 0.8 Å. Principal component analysis of backbone motions was also carried out as implemented in *GROMACS*.

## 2.5. Channel systems

Channels between the active site and the solvent were calculated using *MolAxis* (Yaffe *et al.*, 2008). The bottleneck radius for PREP was measured and averaged in the following characteristic crystal structures: PDB entries 1qfm (Fülöp *et al.*, 1998) and 1h2w (Szeltner *et al.*, 2002a). For ApAAP, values obtained for PDB entries 2hu5, 2hu7, 2hu8 (Kiss *et al.*, 2007), 1ve6, 1ve7 (Bartlam *et al.*, 2004), 3o4g (Harmat *et al.*, 2011) and the four copies of the closed form determined in this study were averaged. In the case of the MD-derived structures,

conformers of the 100–200 ns time range were clustered based on the conformation of the residues defining the three channels considered (propeller channel, residues 23–43, 64–72, 111–121, 153–161, 194–204, 246–257 and 283–293; channel by loopA, 82–89, 111–116, 129–134, 468–490, 520–529 and 551–560; channel by the His-loop, 42–46, 58–64, 83–90 and 551–560); the bottleneck radius was measured in the mid-structure of each cluster and weighted by the occupancy of that cluster.

## 3. Results

### 3.1. Crystal structures of ApAAP–CMK complexes

**3.1.1. Crystals with the open and closed enzyme form, and the overall conformation of ApAAP.** Previously determined ApAAP structures contain dimers composed of closed/closed, open/closed or open/open monomer pairs. Of these, the open/closed dimer of ApAAP provides a unique structural model system for the study of ApAAP, since this arrangement allows the simultaneous testing of various effects on two different conformations of the enzyme under identical experimental conditions. Thus, it was our goal to crystallize the open/closed dimer structure of ApAAP in complex with a substrate-like inhibitor in order to determine the structural roles of domain closure and substrate binding on reconstructing the active site. The covalently binding inhibitor benzyloxycarbonyl-glycyl-glycyl-phenylalanyl-chloromethyl ketone (CMK) was chosen as our probe because (i) the initial chemical step of its binding is similar to that of the enzymatic hydrolysis reaction catalyzed by the enzyme and requires the residues of the catalytic apparatus in active conformations (reviewed in Powers *et al.*, 2002) and (ii) the structure of the covalent complex highly resembles that of an enzyme–substrate complex with minimal shift of the serine and histidine side chains; thus, it is suitable for studying enzyme–substrate contacts.

Two different crystal structures were determined. Our first attempt to produce crystals of open/closed dimers was as follows: the ApAAP–CMK complex was formed in solution and then crystallized using the crystallization conditions of the open/closed crystal form. Surprisingly, crystals of a different, previously encountered crystal form (Bartlam *et al.*, 2004; Kiss *et al.*, 2007) were obtained containing enzyme dimers in only the closed conformation (closed/closed form; Fig. 2*a*), albeit in this case of the CMK inhibitor-bound form (Fig. 2*b*). This supports the hypothesis that ligand binding stabilizes the closed form of ApAAP (Li *et al.*, 2010; Harmat *et al.*, 2011). The enzyme backbone could be fitted to previously determined crystal structures with r.m.s.d. values of 0.29–0.56 Å for all backbone atoms, indicating that no significant rearrangement is required for ligand coordination.

We were successful in attaining the inhibitor-bound form of the open/closed dimer structure by soaking pre-formed crystals of the open/closed crystal form of the unligated enzyme in a solution of CMK (Fig. 2*c*). The enzyme backbone could be fitted to previously determined crystal structures with r.m.s.d. values of 0.27–0.53 and 0.39–0.61 Å for all backbone atoms of the closed and open monomers, respectively. Addition of the

inhibitor to the crystal of the open/closed form thus did not result in the closure of the open monomer. Such large-scale rearrangement has been observed previously in the case of *Aeromonas caviae* PREP. In that case, addition of benzyloxy-
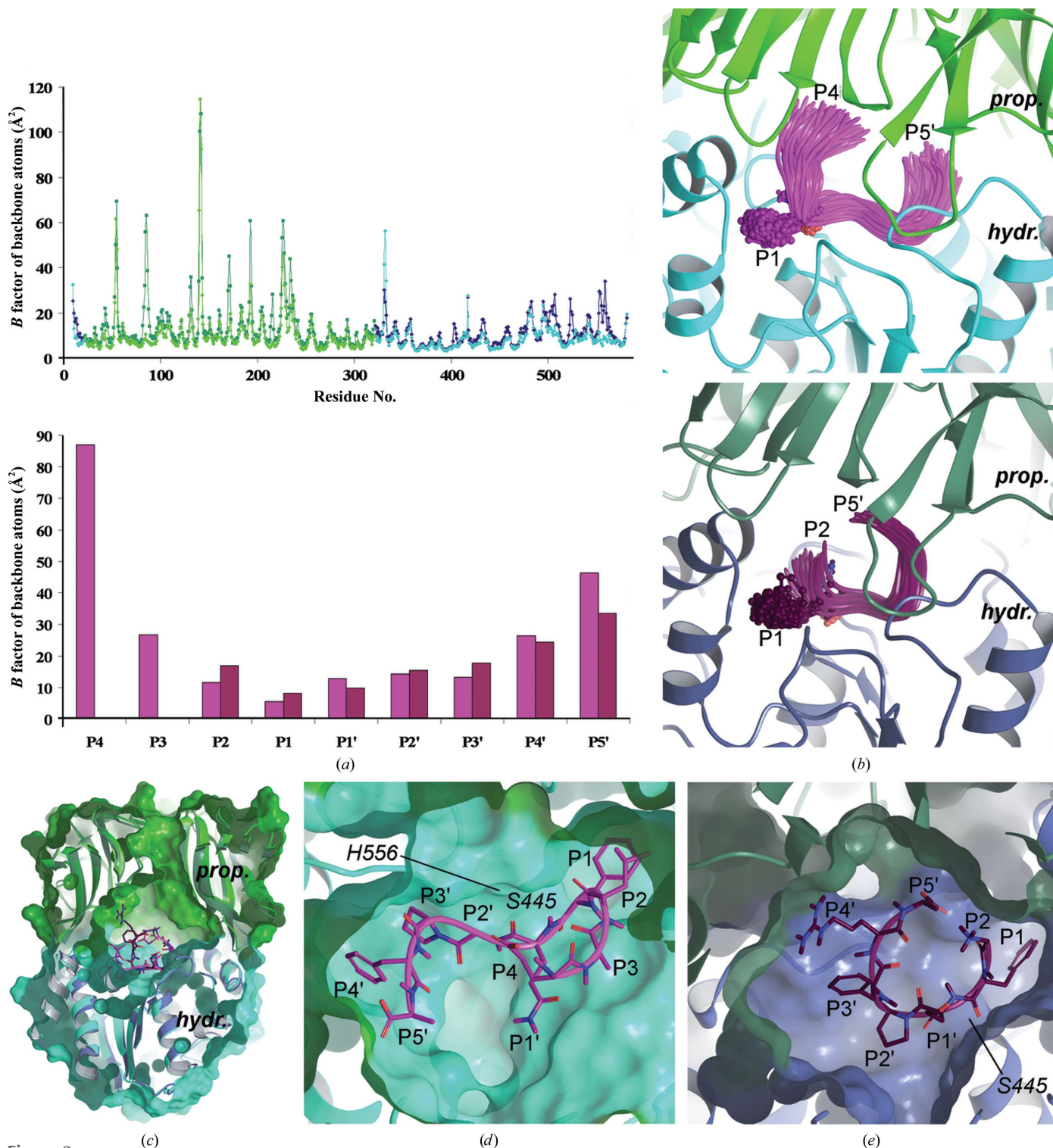


Figure 3
Molecular dynamics-derived structures of the ApAAP–L1 and ApAAP–L2 complexes. (*a*) *B* factors of backbone atoms derived from the 100–200 ns time span of the MD simulations for the enzyme (propeller domain, green; hydrolase domain, cyan and blue; ligand, magenta and purple; residues labelled according to Schechter & Berger, 1967). (*b*) Clustering of the substrate conformations in the cases of ApAAP–L1 (top) and ApAAP–L2 (bottom). (*c*) The superimposed complexes shown as the molecular surface of ApAAP, with substrates bound in the spacious inner cavity below the propeller channel (cross-section; the propeller domain is shown in green, the hydrolase domain in cyan and the L1 and L2 substrates in magenta and purple, respectively). (*d*, *e*) Close-up of the substrate-binding cavity. P1 Phe of L1 (*d*) is immersed in the specificity pocket, while P2 Leu and P4′ Phe are docked into hydrophobic pockets formed by both domains. The P4′ Arg residue of L2 (*e*) pulls away from the hydrophobic surface towards the propeller channel. Catalytic residues of AAP are labelled in italics.

**Table 2**
Accommodation of CMK (PDB entry 4re5, chain *A*) and substrate peptides L1 and L2 (molecular-dynamics simulations) by ApAAP.

Residue positions of the CMK inhibitor and substrates are also labelled according to Schechter & Berger (1967). Contacting residues (atoms within 4.5 Å distance) for each residue position of CMK and the substrates are listed, with hydrogen-bonding residues shown in italics.

| Residue position | Ligand name | Residue type | Contacting residues |
|---|---|---|---|
| | *Substrate/inhibitor ligand* | | |
| P4 | CMK | — | — |
| | L1 | Glu | Ala557 |
| | L2 | — | — |
| P3 | CMK | Gly | — |
| | L1 | Ala | Pro370, Phe371, Ala372, Phe488 |
| | L2 | — | — |
| P2 | CMK | Gly | *Arg526*, Phe485, Phe488 |
| | L1 | Leu | *Arg526*, Phe153, Phe155, Phe169, Phe485, Phe488 |
| | L2 | Gly | *Arg526*, **P1′** or **P5′**, **P1**, Phe153, Phe155, Phe485 |
| P1 | CMK | Phe | *Gly369*, Tyr446, Val471, Trp474, Met477, Leu482, Phe485, Phe488, Ile489 |
| | L1 | Phe | *Gly369*, Tyr446, Val471, Trp474, Met477, Leu482, Phe485, Phe488, Ile489 |
| | L2 | Phe | *Gly369*, **P2**, Tyr446, Val471, Trp474, Met477, Leu482, Phe485, Phe488, Ile489, His556 |
| P1′ | CMK | — | — |
| | L1 | Gln | Ala372 |
| | L2 | Glu | — |
| P2′ | CMK | — | — |
| | L1 | Gly | Tyr444 |
| | L2 | Pro | *Ile558*, Phe381, Tyr444, Ile558 |
| P3′ | CMK | — | — |
| | L1 | Pro | **P4′**, Phe381, Tyr444, Ile567, Ile558 |
| | L2 | Phe | — |
| P4′ | CMK | — | — |
| | L1 | Phe | Ile20, Phe381, Ala564, Leu568 |
| | L2 | Arg | *Ser66* |
| P5′ | CMK | — | — |
| | L1 | Ala | — |
| | L2 | Ala | **P2**, *Arg526*, Leu115, Phe153, Phe155 |

carbonyl-prolyl-prolinal (zPP) to a crystal consisting of open monomers of PREP resulted in the closure of the enzyme in the crystalline state, coupled to the shortening of one of the unit-cell axes. The authors argued that the loose packing and the spacious channel systems of that particular crystal form allowed such grand rearrangements (Li *et al.*, 2010). In the case of the PREP crystals, the direction of domain motion is nearly parallel to one of the unit-cell axes, allowing concerted movement of layers of domains within the crystal. In contrast, in ApAAP there are two different directions for potential domain closure of the open monomers, neither of which is parallel to a cell axis. Moreover, the closed monomers between the open monomers obstruct concerted movement of the latter, creating a more formidable restriction to such movement (Supplementary Fig. S2).

Since the overall structures of the ApAAP molecules in the newly determined ApAAP–CMK complexes highly resemble those of previously determined unligated and noncovalent inhibitor-bound states for both of these two crystal forms, our findings support the two-state model of ApAAP with only two possible relative domain orientations. In the case of *A. caviae* PREP, various extents of opening were found in different ligation states (Li *et al.*, 2010), which is not the case here. In ApAAP crystals (obtained in four different space groups

across a pH range of 4.6–8.0 in unligated and ligated states with wild-type and mutant sequences), monomers with closed and open conformations can be fitted with a backbone r.m.s.d. not exceeding 0.68 and 1.18 Å, respectively. In the closed state the catalytic apparatus is in the active conformation in all cases, while in all of the open states the catalytic apparatus is disassembled and the loop holding the catalytic histidine is highly flexible.

**3.1.2. Binding of CMK inhibitor in the crystal structures.** The CMK molecules are clearly shown in the electron-density maps bound at the substrate-binding sites of the AAP molecules (Fig. 2). Interestingly, they are bound only in the closed enzyme structures (in all monomers in the closed/closed crystal form and in only half of the monomers in the open/closed crystal form; Figs. 2*b*, 2*d* and 2*e*). It is clear from the electron density that a chemical reaction has occurred between the closed AAP monomers and CMK: the inhibitor is covalently bound to the catalytic serine and histidine (Figs. 2*b* and 2*d*). The P1 phenylalanine side chain of the inhibitor is bound in the S1 pocket (the nomenclature of Schechter & Berger, 1967 is used throughout the text). However, the P2–P3 glycyl-glycine moiety is less well defined by electron density, suggesting looser binding of this part of the inhibitor, while the benzyloxycarbonyl moiety is disordered, with the conformation and position of the P3 glycine being somewhat different in the four ApAAP–CMK monomer structures (Fig. 2*f*). The reason for this flexibility/disorder may be that neither S3–P3 contacts that orient its backbone nor side-chain-accommodating contacts exist for the small P3 glycine residue. Comparing the four forms of CMK ligation with that of enzyme–product complexes reveals that the accommodation and the position of the P1–P2 moiety is similar, as expected, showing conserved hydrogen bonds to the oxyanion site and Arg526 (Fig. 2*f*).

Observing the inhibitor molecules in the closed monomers of the open/closed crystal form proves that the inhibitor can reach the active site of closed monomers of ApAAP by diffusion within the crystal and that the enzyme is active in the crystalline state. Interestingly, the inhibitor molecule did not bind to the open monomers. As the active sites of the open monomers are solvent-exposed in the crystal, the reason for this cannot be the limited access to the active site but must be the conformational differences affecting the substrate affinity and catalytic efficiency: the disassembled state of the catalytic triad, the large shift and destabilization of the loop containing

His556, the shift of the Asp524–Arg526 region and the small changes in the S1 pocket.

## 3.2. Molecular-dynamics studies

**3.2.1. Binding of substrates.** MD simulations were carried out involving the binding of two known substrates of ApAAP to the enzyme: one of nine amino acids in length (sequence EALFQGPFA; from here on referred to as L1) and another of seven amino acids in length (sequence GFEPFRA; from here on referred to as L2). ApAAP cleaves the peptide bond of its oligopeptide substrates following phenylalanine residues (and, with significantly less activity, leucine residues). In the case of L1 the primary cleavage site is at Phe4, while L2 is cleaved after Phe2 (Kiss *et al.*, 2007).

L1 and L2 were placed into the ligand-binding pocket of a closed ApAAP monomer with the P1 residues of their primary cleavage site overlaying that of the CMK inhibitor in the crystal structure. An MCMM search was then applied to dock substrates into the enzyme interior, followed by a 200 ns MD simulation of the derived complexes. The final results showed substrates bound in the spacious cavity between the domains without disturbing the structure of the closed form of the enzyme (Fig. 3): the backbone of the closed ApAAP monomer in the crystal structures (four copies; two in PDB entry 4re5 and two in PDB entry 4re6) and those of the simulation-derived complexes could be fitted with r.m.s.d. values ranging from 1.1 to 1.2 Å. Two conserved hydrogen bonds to the oxyanion pocket and a tailored hydrophobic pocket for the P1 side chain were able to direct and stabilize substrates in an ideal position for cleavage (with distances between the Ser445 $O^\gamma$ atom and the carbonyl C atom of the substrate P1 residue of $3.4 \pm 0.2$ and $3.3 \pm 0.2$ Å in the case of ApAAP–L1

and ApAAP–L2, respectively) without applying constraints during the MD simulations. This is also well reflected in the backbone *B* factors calculated for the substrates (Fig. 3*a*). The P2–P1′ segment shows characteristically low *B*-factor values, with the P1 backbone being most restrained, indicating its specific and strong coordination by the enzyme; while the chain ends were quite mobile in the simulations, the P1 Phe residue of both substrates was found to be most ordered. The backbones of the P2–P1′ segments of L1 and L2 run quite similarly, but the remaining segments were found to be accommodated in diverse ways (Figs. 3*b*, 3*d* and 3*e*). The carbonyl O atoms of the P2 residue of both substrates form a conserved hydrogen bond to Arg526. Additionally, the Phe side chain of P2 of L1 docks into a spacious hydrophobic crevice created by Phe153, Phe155 and Phe169 of the propeller, Phe485 and Phe488 of the hydrolase domain and the P1 residue. The Gly residue of L2 in the same position, the N-terminus of this substrate, forms $NH_3^+\cdots O$-type intramolecular hydrogen bonds mainly to the P1′ residue (in 84% of the snapshots of the last 100 ns of the MD simulation) or the C-terminus (in 3%). The carbonyl O atom of the P1 phenylalanine of both substrates is bound by two hydrogen bonds to the oxyanion hole (backbone amides of Gly369 and Tyr446) while immersing into the hydrophobic S1 pocket lined by the residues of the hydrolase domain (Pro370, Tyr446, Trp474, Met477, Phe485, Phe488 and Ile489). P2′ Pro of L2 is hosted by Phe381, Ile558 and Tyr444. The carbonyl O atom of P3′ of L1 is stabilized by a main-chain hydrogen bond to Ile558; its Pro side chain, while surrounded by hydrophobic amino acids of the hydrolase domain (Phe381, Tyr444, Ile558 and Ile567), forms a stacking interaction with Tyr444 (contacting AAP residues for L1 and L2 binding in comparison with those of CMK are listed in Table 2.).
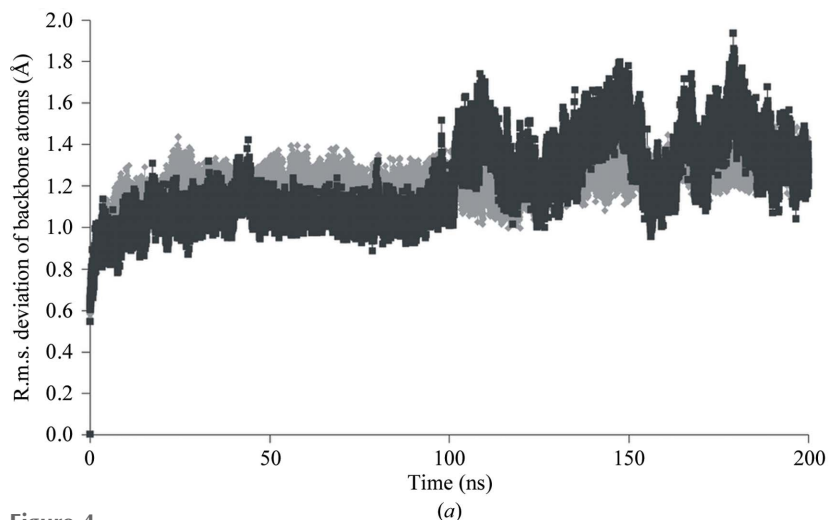


**Figure 4**
Domain motion of ApAAP complexes during MD simulation. (*a*) Backbone r.m.s. values (compared with the starting conformation) along the trajectory of the MD simulation of the ApAAP–L1 (grey) and ApAAP–L2 (black) complexes. (*b*) The most significant motion of ApAAP–L2 in the last 100 ns of its MD simulation as derived by principal component analysis. The movement is shown in ten stages coloured from green to yellow. The black dot indicates the position of the hinge region and the black arrow indicates the site of the greatest opening.

**3.2.2. Small-scale opening of the enzyme structure.** As is apparent in the backbone r.m.s.d. graph of the simulations (Fig. 3a), after approximately 100 ns the complex formed with L2 started to oscillate between different states. While clustering of the conformers of the complex formed with L1 belonging to the 100–200 ns time span resulted in two quite similar clusters, 47 clusters were found in case of ApAAP–L2, which could be separated into three characteristically different groups based on their backbone values: two minor populations with high and low r.m.s.d. (15 and 8% of all conformers, respectively) and a major group with intermediate values (77% of all conformers).

Average structures of the groups were calculated and compared. In case of the minor groups we found that while the domains individually are very similar (the backbone of the propeller domains could be fitted with an r.m.s.d. of 0.4 Å and that of the hydrolase domains with an r.m.s.d. of 0.3 Å), the fit of the backbone atoms of the hydrolase domain increased by nearly a factor of three (to 1.1 Å with the propeller domains overlaid), demonstrating that it is not the inner structure of the domains that is different in the two groups but their respective orientation. Within the hydrolase domain, the greatest change is observed at residues 503–562 (r.m.s.d. of 1.7 Å), while the hinge region connecting the two domains stays less disturbed. Comparing the calculated backbone $B$ factors of ApAAP–L1 and ApAAP–L2 (Fig. 4a), the greatest difference was found in the flexibility of residues 82–89, a loop of the propeller pointing toward the hydrolase domain, and two loops of the hydrolase domain facing it: residues 522–525 (members of the loop carrying the catalytic Asp524) and residues 552–562 (the loop with the catalytic His556 residue). These regions are those that are most destabilized by the opening of the structure (Harmat *et al.*, 2011), supporting the notion that it is a small-scale opening and closing that takes place in the second half of the simulation of ApAAP-L2. Principal component analysis also showed that the largest amplitude motion of the 100–200 ns segment of the simulation is a clamshell-like opening of the structure (Fig. 4b).

The fact that we only observed such motion in the simulation of the complex formed with L2 could be rationalized by the different binding modes of the two substrates. Both the P2 and P4′ residues of L1 bind to hydrophobic pockets formed by the domain interfaces, lined on one side by residues of the propeller and on the other by residues of the hydrolase domain. Thus, by filling the void between the two domains the P2 Leu and P4′ Phe of L1 reinforce the interlocking of the clamshell-like structure. These interactions also place the carboxy-terminus of L1 within hydrogen-bonding distance of Lys24 of the propeller, and since the other part of the substrate is rooted in the hydrolase domain, this also has a domain movement-restricting effect. L2 has a Gly at the P2 position and an Arg at P4′, neither of which can perform the hydrophobic linker function: the end of the chain dislocates into the space below the propeller channel. Being associated chiefly with the hydrolase domain, a slight opening does not influence the accommodation of L2; thus, its presence does not hinder domain movements as much.

## 4. Discussion

ApAAP, just like its mammalian orthologue, functions both as an exopeptidase and an endopeptidase. Its substrate-selection mechanism involves the opening of the enzyme, in which the propeller lid moves in a clamshell-like fashion: it is fixed on the hinge side by interconnecting segments while opening by a considerable extent (10–15 Å) on the other side. The parts travelling furthest from their closed conformation are loop-(43–46) and loop-(83–88) of the propeller domain and loop-(521–529) and loop-(551–560) of the hydrolase domain, which are connected by numerous interdomain hydrogen bonds and thus form the buckle region of the closed form. Incidentally, these two surface-close loops of the hydrolase domain are those carrying the His and Asp residues of the catalytic triad. These loops lose their interdomain interactions as the opening takes place, and as a consequence the His-loop becomes flexible (with the side chain of His itself disappearing from the electron-density map) and the Asp-loop restructures, thus disassembling the catalytic triad (as indicated by the spreading pegs representing the triad residues in Fig. 5). However, previous studies were not able to distinguish whether it is the binding of the substrate or the closure of the enzyme that assembles the catalytic triad.

In this study, two crystal structures of an inhibitor-bound form of ApAAP have been determined. Co-crystallization of ApAAP with the CMK inhibitor led to the formation of closed/closed dimers, even if the crystallization conditions of the previously obtained open/closed dimers were adhered to. In this structure the covalently bound inhibitor is present in both monomers of the dimers (Fig. 5b).

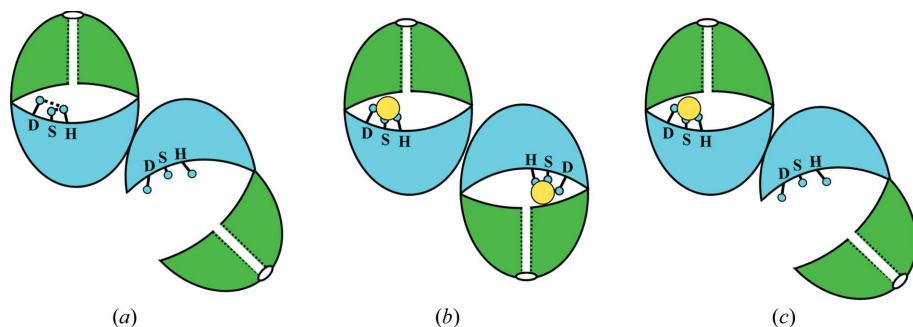Addition of the CMK inhibitor to the pre-formed crystal lattice of open/closed dimers resulted in the formation



**Figure 5**
Schematic view of the dimers of ApAAP discussed in this paper. (a) The asymmetric unligated open/closed dimer of ApAAP. (b) The inhibitor-bound closed/closed dimer formed by co-crystallization. (c) The open/closed dimer formed when the inhibitor was added to crystals of asymmetric dimers.

of asymmetric dimers composed of one closed monomer with the inhibitor covalently linked to the catalytic Ser and His residues and an open, unligated monomer (Fig. 5c). Here, remarkably, the inhibitor migrated to the active site and formed two covalent bonds to two amino acids of the catalytic triad within the rigid frame of the crystal lattice. The monomer units making up this crystal lattice thus represent opposite ends of the catalytic cycle: an open, scavenger conformer with a deactivated active site and a substrate-bound, closed conformer. The appearance of both in the same crystal lattice demonstrates that these forms co-exist under identical experimental conditions, and supports the notion that the native state of the enzyme is best described as a dynamic equilibrium between open and closed states. Since the open or closed nature of the adjoining monomer did not have an effect on the conformation of the substrate-bound closed form, the results also show that the monomer units function independently (without signs of allostery), so that ligand binding and catalysis at one site does not require the same at the other.

The overall binding mode of the CMK ligand and the accommodation of its P1–P2 Gly-Phe moiety in the open/closed structure is similar to that seen in case of the closed/closed dimer, but a 0.92–1.79 Å r.m.s.d. of inhibitor-atom positions within the overlaid protein matrices indicates subtle differences (Fig. 2). The protein matrix of the closed monomer itself is remarkably similar in both cases, with a backbone r.m.s.d. of 0.27–0.56 Å; therefore, it is not the alteration of the binding site that causes the variation of CMK binding, rather it is the conformational freedom of the binding process. This short ligand has no side chains in its P2 and P3 positions (both glycines) that would be able to form specific interactions with the host protein matrix.

The fact that the backbone structure of the closed monomer remained the same in the presence of the inhibitor also indicates that the rigid structure of the closed form must be penetrable to small ligands. Several different access routes have been proposed that could be used by ligands to reach the active site of the closed form of various oligopeptidases. The channel piercing through the propeller domain was initially suggested as the entrance, being the only opening on the surface of PREP, the first enzyme in the family for which a structure was determined. The propeller domain is the structurally more variable of the two, resulting in an array of different sized channels that even differ within orthologues and might help to explain their differing substrate profiles (Kaushik & Sowdhamini, 2011). Also, the modest widening of this channel does not require backbone rearrangement; it can be achieved by reorientation of the side chains clogging its interior (Kaushik & Sowdhamini, 2011). Dynamic restructuring of the backbone segments is needed in the case of the other proposed admittance mechanisms. It has recently been shown that another gate might open by the flip of a 20-amino-acid, loosely structured loop in PREP (loopA formed by residues 189–209) at the domain interface (St-Pierre et al., 2011; Kaszuba et al., 2012). Another short helical segment that belongs sequentially to the propeller domain, but runs along the hydrolase domain and effects interdomain communication

(Papaleo et al., 2012), might also shift and expose yet another tunnel towards the active site (Papaleo & Renzetti, 2012).

On the surface of the closed form of the ApAAP structure, three narrow channels can be observed that connect to the buried active site (Supplementary Fig. S3). In place of the long loopA of PREP, a much shorter loop can be found (residues 130–133) that would leave this accession route ungated. However, in the crystal structures of ApAAP an interdomain hydrogen-bonded network of Asp482 (hydrolase domain)–Arg113 (propeller domain)–Glu88 (conserved residue of the propeller)–Arg526 (conserved residue of the hydrolase domain) can be found which blocks the entrance under the shortened loopA at the protein surface (Supplementary Fig. S4a). In contrast, in PREP the corresponding residues are Thr590, Lys172, Asp149 and Arg643, of which only Asp149 and Arg643 form hydrogen-bond associations and the other two are not stabilized by a hydrogen-bond network; therefore, if loopA fluctuates to an open position a wide entrance may open up towards the interior (Supplementary Fig. S4b). The channel that passes through this region of ApAAP is directly accessible from the solvent owing to the shortness of loopA, but has a bottleneck radius of 0.7 Å owing to the hydrogen-bonded network of the aforementioned quartet. Another, wider side entrance by the loop carrying the catalytic His556 (His-loop) has a bottleneck radius of 1.1 Å. In contrast, the bottleneck radius of the propeller channel, a third possible accession route, is 1.7 Å, which thus provides a quite narrow, but still the most unhindered, passage to the active site. Thus, we propose that the most likely route of the inhibitor to the active site of the closed monomer within the crystal is the propeller channel.

It is also noteworthy that the inhibitor did not bind to the open form, despite the considerable driving force of the possibility of forming two covalent bonds. It has been shown previously that the binding of covalent inhibitors is coupled to the closure of oligopeptidases (Li et al., 2010). In that case, crystals of the open form of PREP monomers were soaked overnight in solution of an inhibitor, which led to the emergence of a crystal with a ligand bound to the active site and the closure of all monomers. This result, however, still has not clarified whether the active site can be ordered to form a catalytically active conformation in the absence of the propeller domain just by the binding of a substrate or whether the closure of the enzyme is required for re-activation of the active site. (To state it another way: whether the closure of the enzyme is a consequence or a condition of the chemical reaction with the ligand.) Our densely packed crystals of the open/closed dimers of ApAAP did not allow such a global rearrangement as seen in the case of PREP (Li et al., 2010). Since the molecules of the inhibitor were able to reach the active site of the closed monomer within the crystal lattice, there is no reason why they should not have reached the clearly more accessible active site of the open form. The fact that the chemical reaction did not take place shows, for the first time, that the open form is indeed catalytically disabled and that closure of the enzyme is required for catalytic activity.

In order to study the binding of longer substrates to ApAAP, MD simulations were carried out for the binary complexes of ApAAP formed with two of its known peptide substrates (Kiss *et al.*, 2007; Fig. 3, Table 2). The inner cavity of the closed monomer proved to be amply spacious for accommodating either peptide without disturbing the overall structure. Even though in the fully solvated state the channels leading to the enzyme interior did widen (with bottleneck radii of 2.0 Å for the propeller channel, 1.2 Å for the side entrance near the His-loop and 0.9 Å under loopA), the propeller channel remained the most spacious entrance route to the active site. A force-biased MD study of PREP has demonstrated that small substrates (Z-Pro-prolinal in this case) can be pulled through the propeller channel, but only with the application of an extremely high force (St-Pierre *et al.*, 2011). The propeller channel of ApAAP is wider than that of PREP and thus might be more suitable for allowing small ligands to reach the active site, but it is highly unlikely that longer oligopeptide substrates would be able to utilize this pathway; therefore, we postulate that longer and more structured ligands might reach the active site assisted by the opening of the structure. Simulations also showed that opening and closing is a trait of ApAAP that is not lost even when peptide ligands are bound within the ligand-binding cavity, indicating that after catalytic cleavage, the dissociated products, if too large for the propeller channel, might leave the enzyme interior by the unhindered route provided by the opening.

Similar opening and closing domain movements of an oligopeptidase have recently been observed in MD simulations of PREP monomers, prompting the authors to also propose this motion as the major substrate-admittance/release mechanism of that enzyme. In their model, both the opening and closing of the enzyme is prompted by the presence of the substrate (Kaushik *et al.*, 2014). This mechanism might be modulated by the flexible loops (loopA, for example) near the domain interface that also appear to have limited size-regulation function (Szeltner *et al.*, 2013). Our results indicate that in the case of ApAAP the presence of a ligand is not necessary to ignite domain movements; rather, it is the dynamic equilibrium mixture of the open and closed states that the ligands encounter and may attempt to interact with. Whether this mixture is dominated by one form or another might be fine-tuned by the amino-acid composition of the domain interfaces.

Besides that of ApAAP, the crystal structures of the open forms of further oligopeptidases have been determined: those of PREP from lower taxa (Shan *et al.*, 2005; Li *et al.*, 2010) and oligopeptidase B (McLuskey *et al.*, 2010; Canning *et al.*, 2013). Thus, clarifying the molecular details involved in the opening and closing of AAP has a universal appeal, while also allowing a better understanding of such a physiologically relevant enzyme. There are two structural models presently available for the study of AAP: PhAAP with a permanent side entrance (Menyhárd *et al.*, 2013) and the flexible ApAAP that utilizes domain movements for substrate selection (Harmat *et al.*, 2011). PhAAP is a hexamer in its active form, the structure of which is of a complex self-compartmentalized inner channel-and-chamber system where only suitably short and unstructured ligands are able to reach the side entrance and, through it, the active site of its monomers. As the tetrameric active form of mammalian AAP might not provide enough building blocks for such an intricate system as observed in the case of PhAAP, we suggest that it consists of ApAAP-like flexible monomers and also resorts to an opening–closing mechanism for substrate screening.

## 5. Conclusion

In this paper, crystallography was used as a tool for the sequential sampling of the favourable conformations of a substrate-binding process. Firstly, we established that adding a substrate-like inhibitor to a solution of ApAAP results in the formation of closed/closed ligand-bound dimers. The fact that an open/closed asymmetric dimer of the unligated form of ApAAP could also be crystallized under very similar experimental conditions in the absence of the ligand adds additional support to a model in which an equilibrium mixture of the open and closed forms must be present in the solution of the resting state of the enzyme: this equilibrium is shifted towards the closed form in the presence of the ligand. The final sampling step was carried out by the inhibitor itself, which, when added to the rigid mixture of these two states in the crystal, chose only one of them, the closed conformer, for binding. This way, we were able to capture both the unligated and the substrate-bound form of the enzyme within the same crystal lattice, *i.e.* two catalytically distinct states. Catalysis itself can only take place if the enzyme assumes its active conformation by transferring from an open, scavenger state to its activated, closed form.

### References

Adibekian, A., Martin, B. R., Wang, C., Hsu, K.-L., Bachovchin, D. A., Niessen, S., Hoover, H. & Cravatt, B. F. (2011). *Nature Chem. Biol.* **7**, 469–478.

Arnesen, T. (2011). *PLoS Biol.* **9**, e1001074.

Bartlam, M., Wang, G., Yang, H., Gao, R., Zhao, X., Xie, G., Cao, S., Feng, Y. & Rao, Z. (2004). *Structure*, **12**, 1481–1488.

Bjelkmar, P., Larsson, P., Cuendet, M. A., Hess, B. & Lindahl, E. (2010). *J. Chem. Theory Comput.* **6**, 459–466.

Bussi, G., Donadio, D. & Parrinello, M. (2007). *J. Chem. Phys.* **126**, 014101.

Canning, P., Rea, D., Morty, R. E. & Fülöp, V. (2013). *PLoS One*, **8**, e79349.

Chang, G., Guida, W. C. & Still, W. C. (1989). *J. Am. Chem. Soc.* **111**, 4379–4386.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). *J. Am. Chem. Soc.* **117**, 5179–5197.

Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F. & Mark, A. E. (1999). *Angew. Chem. Int. Ed.* **38**, 236–240.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Engel, M., Hoffmann, T., Wagner, L., Wermann, M., Heiser, U., Kiefersauer, R., Huber, R., Bode, W., Demuth, H. U. & Brandstetter, H. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 5063–5068.

Forte, G. M., Pool, M. R. & Stirling, C. J. (2011). *PLoS Biol.* **9**, e1001073.

Fujino, T., Watanabe, K., Beppu, M., Kikugawa, K. & Yasuda, H. (2000). *Biochim. Biophys. Acta*, **1478**, 102–112.

Fülöp, V., Böcskei, Z. & Polgár, L. (1998). *Cell*, **94**, 161–170.

Fülöp, V., Szeltner, Z., Renner, V. & Polgár, L. (2001). *J. Biol. Chem.* **276**, 1262–1266.

Fuxreiter, M., Magyar, C., Juhász, T., Szeltner, Z., Polgár, L. & Simon, I. (2005). *Proteins*, **60**, 504–512.

Harmat, V., Domokos, K., Menyhárd, D. K., Palló, A., Szeltner, Z., Szamosi, I., Beke-Somfai, T., Náray-Szabó, G. & Polgár, L. (2011). *J. Biol. Chem.* **286**, 1987–1998.

Hess, B. (2008). *J. Chem. Theory Comput.* **4**, 116–122.

Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. (2008). *J. Chem. Theory Comput.* **4**, 435–447.

Ito, K., Nakajima, Y., Xu, Y., Yamada, N., Onohara, Y., Ito, T., Matsubara, F., Kabashima, T., Nakayama, K. & Yoshimoto, T. (2006). *J. Mol. Biol.* **362**, 228–240.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kaszuba, K., Róg, T., Danne, R., Canning, P., Fülöp, V., Juhász, T., Szeltner, Z., St Pierre, J.-F., García-Horsman, A., Männistö, P. T., Karttunen, M., Hokkanen, J. & Bunker, A. (2012). *Biochimie*, **94**, 1398–1411.

Kaushik, S., Etchebest, C. & Sowdhamini, R. (2014). *Proteins*, **82**, 1428–1443.

Kaushik, S. & Sowdhamini, R. (2011). *PLoS One*, **6**, e26251.

Kiss, A. L., Hornung, B., Rádi, K., Gengeliczki, Z., Sztáray, B., Juhász, T., Szeltner, Z., Harmat, V. & Polgár, L. (2007). *J. Mol. Biol.* **368**, 509–520.

Kiss, A. L., Palló, A., Náray-Szabó, G., Harmat, V. & Polgár, L. (2008). *J. Struct. Biol.* **162**, 312–323.

Li, M., Chen, C., Davies, D. R. & Chiu, T. K. (2010). *J. Biol. Chem.* **285**, 21487–21495.

MacKerell, A. D. Jr, Feig, M. & Brooks, C. L. III (2004). *J. Comput. Chem.* **25**, 1400–1415.

Matoba, Y., Nakayama, A., Oda, K., Noda, M., Kumagai, T., Nishimura, M. & Sugiyama, M. (2011). *Proteins*, **79**, 2999–3005.

McLuskey, K., Paterson, N. G., Bland, N. D., Isaacs, N. W. & Mottram, J. C. (2010). *J. Biol. Chem.* **285**, 39249–39259.

Menyhárd, D. K., Kiss-Szemán, A., Tichy-Rács, É., Hornung, B., Rádi, K., Szeltner, Z., Domokos, K., Szamosi, I., Náray-Szabó, G., Polgár, L. & Harmat, V. (2013). *J. Biol. Chem.* **288**, 17884–17894.

Mitta, M., Asada, K., Uchimura, Y., Kimizuka, F., Kato, I., Sakiyama, F. & Tsunasawa, S. (1989). *J. Biochem.* **106**, 548–551.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nakajima, Y., Ito, K., Toshima, T., Egawa, T., Zheng, H., Oyama, H., Wu, Y.-F., Takahashi, E., Kyono, K. & Yoshimoto, T. (2008). *J. Bacteriol.* **190**, 7819–7829.

Nordhoff, S., Cerezo-Gálvez, S., Feurer, A., Hill, O., Matassa, V. G., Metz, G., Rummey, C., Thiemann, M. & Edwards, P. J. (2006). *Bioorg. Med. Chem. Lett.* **16**, 1744–1748.

Palmieri, G., Bergamo, P., Luini, A., Ruvo, M., Gogliettino, M., Langella, E., Saviano, M., Hegde, R. N., Sandomenico, A. & Rossi, M. (2011). *PLoS One*, **6**, e25888.

Papaleo, E. & Renzetti, G. (2012). *J. Mol. Graph. Model.* **38**, 226–234.

Papaleo, E., Renzetti, G. & Tiberti, M. (2012). *PLoS One*, **7**, e35686.

Perrier, J., Durand, A., Giardina, T. & Puigserver, A. (2005). *Biochimie*, **87**, 673–685.

Polgár, L. (2002). *Cell. Mol. Life Sci.* **59**, 349–362.

Powers, J. C., Asgian, J. L., Ekici, Ö. D., Özlem, D. & James, K. E. (2002). *Chem. Rev.* **102**, 4639–4750.

Rasmussen, H. B., Branner, S., Wiberg, F. C. & Wagtmann, N. (2003). *Nature Struct. Biol.* **10**, 19–25.

Sandomenico, A., Russo, A., Palmieri, G., Bergamo, P., Gogliettino, M., Falcigno, L. & Ruvo, M. (2012). *J. Med. Chem.* **55**, 2102–2111.

Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.

Shan, L., Mathews, I. I. & Khosla, C. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 3599–3604.

Shimizu, K., Fujino, T., Ando, K., Hayakawa, M., Yasuda, H. & Kikugawa, K. (2003). *Biochem. Biophys. Res. Commun.* **304**, 766–771.

Shimizu, K., Kiuchi, Y., Ando, K., Hayakawa, M. & Kikugawa, K. (2004). *Biochem. Biophys. Res. Commun.* **324**, 140–146.

St-Pierre, J.-F., Karttunen, M., Mousseau, N., Róg, T. & Bunker, A. (2011). *J. Chem. Theory Comput.* **7**, 1583–1594.

Szeltner, Z., Juhász, T., Szamosi, I., Rea, D., Fülöp, V., Módos, K., Juliano, L. & Polgár, L. (2013). *Biochim. Biophys. Acta*, **1834**, 98–111.

Szeltner, Z., Kiss, A. L., Domokos, K., Harmat, V., Náray-Szabó, G. & Polgár, L. (2009). *Biochim. Biophys. Acta*, **1794**, 1204–1210.

Szeltner, Z., Rea, D., Juhász, T., Renner, V., Fülöp, V. & Polgár, L. (2004). *J. Mol. Biol.* **340**, 627–637.

Szeltner, Z., Rea, D., Juhász, T., Renner, V., Mucsi, Z., Orosz, G., Fülöp, V. & Polgár, L. (2002). *J. Biol. Chem.* **277**, 44597–44605.

Szeltner, Z., Rea, D., Renner, V., Fülöp, V. & Polgár, L. (2002). *J. Biol. Chem.* **277**, 42613–42622.

Thoma, R., Löffler, B., Stihle, M., Huber, W., Ruf, A. & Hennig, M. (2003). *Structure*, **11**, 947–959.

Winn, M. D. et al. (2011). *Acta Cryst.* D**67**, 235–242.

Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* D**57**, 122–133.

Yaffe, E., Fishelovitch, D., Wolfson, H. J., Halperin, D. & Nussinov, R. (2008). *Proteins*, **73**, 72–86.

Yamin, R., Zhao, C., O'Connor, P. B., McKee, A. C. & Abraham, C. R. (2009). *Mol. Neurodegeneration*, **4**, 33.