A Self-Compartmentalizing Hexamer Serine Protease from *Pyrococcus Horikoshii* – Substrate Selection Achieved through Multimerization*

Dóra K. Menyhárd[1], Anna Kiss-Szemán[2], Éva Tichy-Rács[2,3], Balázs Hornung[2,4], Krisztina Rádi[2,5], Zoltán Szeltner[6], Klarissza Domokos[6], Ilona Szamosi[6], Gábor Náray-Szabó[2], László Polgár[6], and Veronika Harmat[1,2]

[1]From Protein Modeling Research Group, Hungarian Academy of Sciences, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary

[2] Institute of Chemistry, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary

[3]Current address: Institute for Solid State Physics and Optics, Wigner Research Centre for Physics, Hungarian Academy of Sciences, Konkoly-Thege M. út 29-33, H-1121 Budapest, Hungary

[4]Current address: Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, OX1 3QZ, United Kingdom

[5]Current address: University of Warwick, School of Life Sciences, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom

[6]Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, 29. Karolina street, H-1113, Budapest, Hungary

*Running title: *Hexameric serine oligopeptidase*

To whom correspondence may be addressed: Veronika Harmat, Institute of Chemistry, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary, Tel.:(+36-1) 372-2500 ext. 6547, FAX: (+36-1) 372-2592 E-mail: veronika@chem.elte.hu

Keywords: Acylaminoacyl peptidase; X-ray crystallography; Self compartmentalization; Protein self-assembly, Substrate size selection, Amylogenic β-strand

Background: Oligopeptidases are serine proteases cleaving only short peptides.
Results: The complex channel system found within a hexameric oligopeptidase presents a rigid, double-gated model for size-based substrate selection.
Conclusion: Substrate selection mechanism applied by an oligopeptidase depends on its multimerization state.
Significance: Degradation of cytotoxic and misfolded proteins is aided by oligopeptidases, which are thus possible targets of cancer therapy.

SUMMARY

Oligopeptidases impose a size limitation on their substrates, the mechanism of which has long been in debate. Here we present the structure of a hexameric serine protease, an oligopeptidase from *Pyrococcus horikoshii* (PhAAP), revealing a complex, self-compartmentalized inner space, where substrates may access the monomer active sites passing through a double-gated "check-in" system: first passing through a pore on the hexamer surface, then turning to enter through an even smaller opening at the monomers' domain-interface. This substrate screening strategy is unique within the family. We found that among oligopeptidases a member of catalytic apparatus is positioned near an amylogenic β-edge, which needs to be protected to prevent aggregation and found different strategies applied to such end. We propose that self-assembly within the family results in characteristically different substrate selection mechanisms coupled to different multimerization states.

Members of the prolyl oligopeptidase family are serine proteases, such as prolyl oligopeptidase (POP) itself, dipeptidyl peptidase IV (DPP-IV), oligopeptidase B (OPB) and acylaminoacyl peptidase (AAP), that are able to distinguish between potential substrates based on their size, and only cleave those that do not exceed 30 amino acids in length (1,2). When the first crystal structure of a serine oligopeptidase, that of POP, was determined (3), size exclusion was explained by the domain structure of the enzyme, which later proved to be quite similar for all family members. Oligopeptidases consist of a half-sphere shaped hydrolase region containing the active site and a cylindrical propeller domain, which caps it, burying the Ser-His-Asp catalytic triad in a well-protected inner hole with dimension comparable to the size limitations imposed by the enzyme. The structure, however, raised the question of substrate access -and through it- how enzyme effectiveness and selectivity are simultaneously maintained in case of such a decidedly buried active site.

It was initially proposed that substrates must reach the active site through the narrow channel bisecting the propeller domain. Alternatively, substrate access routes must target the domain interface, either through a transient opening between the propeller and the hydrolase domains or through a permanent hole between the two.

POP, the first discovered and most extensively studied member of the family was revealed to have both a closed (4,5) (PDB codes: 1h2w, 3mun) and an open arrangement (6,7) (PDB codes: 3iul, 1yr2) in crystal structures of its unligated form, in the latter of which the cylindrical dome of the propeller region flips back exposing the active site entirely. This suggests, that an open-closed dynamic equilibrium, fine-tuned by the amino acid composition of the interacting domain surfaces, might describe the resting state of the enzyme, where both forms coexist. This notion recently gained support by the determination of the crystal structure of another oligopeptidase, AAP from *Aeropyrum pernix* (ApAAP) which contains both the open and closed forms of the unligated enzyme within the same crystal lattice, in form of 1:1 open-closed mixed dimers (8). In both cases a keen control mechanism is coupled to the opening – the active site is disassembled in the open form mainly by the destabilization and re-*formation* of the flexible loop containing the catalytic His (6,8). Substrates exceeding the size limit disrupt the closing of the enzyme, thus only encounter the impaired active site of the open form and leave the transient complex unchanged.

An alternate possible substrate selection strategy is formation of a permanent entrance hole and providing for its shielding. DPP-IV was shown to have such an architecture - substrates can access its active site either through its large propeller hole (9) or through a spacious side opening between its hydrolase and propeller domains stabilized and partly covered by dimerization (10,11,12).

In this work we present an even more complex example of the latter: that of acylaminoacyl peptidase from *Pyrococcus horikoshii* (PhAAP), which we found active in hexameric form (13). Three crystal structures were determined, which demonstrate that hexamerization results in self-compartmentalizing: a self-assembled double-gated channel system that may effectively screen the substrates of the six individual monomers all possessing a permanent entry at the domain-interface. While hexameric arrangement is quite unique among classical serine-proteases, it is reminiscent of other self-compartmentalizing intracellular proteases, such as Gal6/Bleomycin hydrolase (14), tricorn protease (15), lon protease (16) or the complex systems of proteasomes.

Thus the PhAAP structure provides a unique possibility drawing parallels among oligopeptidases sharing common domain structure but applying different substrate selection strategy, and among self-compartmentalizing proteases of different monomeric structure but of similar selection strategy. This comparison allowed us to propose that the flexibility of the loop holding the catalytic His residue, the accessibility of a sticky β-strand anchoring the His-loop and unobtrusive insertions or terminal extensions determine the substrate selection strategy and self-assembly in the oligopeptidase family. Structure determination of PhAAP and understanding the structural requirements and consequences of multimerization also takes us one step closer to the physiologically relevant mammalian enzyme, which is active as a tetramer.

AAP functions as an exopeptidase in mammals, removing N-terminally blocked amino acids from peptides, though other orthologs have endopeptidase activity too (13, 17). It has been shown to act in concert with, and exert regulation over the proteosome (18), while also being implicated in renal- and small-cell carcinoma (19, 20, 21), and it is also referred to as potential target of certain cognitive enhancers (22, 23).

EXPERIMENTAL PROCEDURES

Protein expression and purification for crystallization- Acylaminoacyl peptidase from *Pyrococcus horikoshii* (PhAAP) was expressed in E. coli and purified as previously described for PhAAP and *Aeropyrum pernix* acylaminoacyl peptidase (ApAAP) (13,17). The active site mutation was introduced in PhAAP (S466A) by the two-step PCR method as described for ApAAP in (17). The protein solutions were concentrated to 10 mg/ml in 20 mM Tris/HCl buffer (pH 8.0). Preparing the covalent complex of PhAAP with the inhibitor benzyloxycarbonyl-glycyl-glycyl-phenylalanyl-chloromethyl ketone (CMK) was carried out by adding the inhibitor in 2.4 fold excess to the protein solution.

Crystallization, data collection, phasing and refinement- The PhAAP structure was determined from crystals grown at 20°C by the hanging drop method. Crystal quality was improved by microseeding. The crystals were flash frozen and stored in liquid nitrogen prior data collection.

For the native crystal, drops were prepared by mixing 2µl of protein solution and 2µl of reservoir solution (3.0 M 1,6-hexanediol, 0.20 M $MgCl_2$, 0.1 M Tris/HCl buffer, pH=8.5). The protein solutions contained 10.5-13.0 mg/ml PhAAP in 20 mM Tris/HCl buffer (pH 8.0). Phase information was obtained using the following three derivatives: A crystal was soaked in 4.5 µl of reservoir mixed by 1.5 µl of 0.05 M $UO_2(NO_3)_2$ solution for 1 h prior to flash cooling. A crystal was soaked in 2 µl of reservoir mixed by 2 µl of 0.02 M K_2PtCl_4 solution for 2 h prior to flash cooling. After mixing 2 µl of 0.2 M KI solution in the crystallization drop a crystal was quick-soaked in it.

The orthorhombic crystal form was obtained by co-crystallizing trial of S466A mutant form of PhAAP and its substrate (13) N-benzyloxycarbonyl-Glu-Phe-Ser-Pro-(para-nitro-Phe)-Arg-Ala. The substrate was dissolved in the protein solutions (9.2 mg/ml protein in 20 mM Tris/HCl buffer, pH 8.0) at 10 fold excess. The drops contained 2µl of protein solution and 2µl of reservoir solution (3.0 M 1,6-hexanediol, 0.30 M $MgCl_2$, 0.1 M Tris/HCl buffer, pH=8.0). The structure, however does not contain any fragment of the substrate, possibly because 1,6 hexanediol competes for and occupies the S1 site. For studying binding of a substrate fragment we prepared crystals with a covalent inhibitor CMK.

The crystal of PhAAP/CMK complex was grown in a hanging drop of 4µl of the PhAAP/CMK solution (9.7 mg/ml of covalent protein/inhibitor complex in 20 mM Tris/HCl buffer, pH 8.0) and 4µl of reservoir solution (2.5 M 1,6-hexanediol, 0.20 M $MgCl_2$, 0.1 M Tris/HCl buffer, pH=8.5).

Data collection was carried out on a Rigaku diffractometer (Cu Kα radiation, wavelength 1.5418 Å) for the native, $UO_2(NO_3)_2$ and KI crystals (graphite monochromator, R-AXIS IIc detector) as well as the CMK complex (blue optics, R-AXIS IV++ detector). Data sets of the K_2PtCl_4 derivative crystal were collected at synchrotron sources at wavelengths 0.8162 Å and 1.0715 Å, at the EMBL beamline X11 of the DESY and at beamline BM14 of ESRF, respectively. The data set of a crystal of the orthorhombic crystal form was collected at beamline BM14 of ESRF (wavelength 1.0715 Å). All data were collected at 100 K. Data processing was carried out using the XDS and XSCALE programs (24) for all data sets. All but one crystals were isostructural with the hexamers of AAP formed by space group symmetry H 3 2 from the single monomer of the asymmetric unit. The asymmetric unit of the crystal in orthorhombic crystal form contains two hexamers of PhAAP (for this data set reflections for R_{free} calculation were selected in thin shells).

Data sets of the native, $UO_2(NO_3)_2$ and KI and K_2PtCl_4 derivatives were used for solving the phase problem using multiple isomorphous replacement with anomalous scattering (MIRAS). The initial set of sites was located by the SHELX package (25). Phases were refined with the program MLPHARE (26), and they were further improved with the program DM (27). An initial model was automatically generated by Buccaneer (28). The phase problem for the orthorhombic crystal form was solved by molecular replacement using the program MOLREP (29). Manual model building was performed for all the three structures using the program Coot (30). Refinement was carried out either by using REFMAC (31) (native and CMK complex) or Phenix (32) (orthorhombic crystal form); including TLS B-factor refinement for each individual protein domain. The model was validated using SFCHECK (33) and MolProbity (34). Percent of residues in the favored/disallowed Ramachandran regions are as follows. PhAAP native structure: 95.82% / 0.17%, PhAAP/CMK complex: 96.22% / 0.33%, PhAAP orthorhombic crystal form: 96.19% / 0.00%. The final statistics can be found in Table 1.

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under the accession codes 4hxe, 4hxf and 4hxg.

Molecular Dynamics (MD) simulations- 60 ns NPT MD simulations were carried out on monomers of POP (using PDB structure 1h2w (4) as a starting conformation) and PhAAP (using the presently determined native structure as starting conformation) at 300 K. The systems were pre-equilibrated at 300 K by decreasing the restraints on the protein atoms, and submitting the free system to a 200 ps NVT step to stabilize pressure. The GROMACS program suite (35) and CHARMM27 force field (36, 37) was utilized. Systems were solvated by approx. 28000 TIP3P waters, protein overall charge was neutralized and physiological salt concentration set using Na^+ and Cl^- ions. Average geometries and B-factors were calculated using the last 10 ns of the trajectories.
Figures were created using PyMOL (DeLano Scientific) and Maestro (Maestro, Version 9.2, Schrodinger, LLC, NY, 2011).

RESULTS

Crystal structures- Crystal structure of the PhAAP hexamer was determined in two crystal forms. Two sets of data originate from crystals bearing R32 space group symmetry where the six monomers of the hexamer are crystallographically equivalent. One of these shows the substrate-free form of the enzyme, while the other contains a chloromethyl ketone inhibitor (benzyloxycarbonyl-glycyl-glycyl-phenylalanyl-chloromethyl ketone; CMK) covalently bound to the active site serine and histidine residues, serving as an analogue of the intermediate of the substrate hydrolysis reaction. The second crystal form is orthorhombic containing two substrate-free hexamers in the asymmetric unit, with no crystallographic restrictions for the symmetry and topology of the monomers within the hexamer. (Table 1) The monomer units within each hexamer and in between monomers of the different structures are strikingly similar (after fitting to the native unligated structure, r.m.s. deviation of Cα atoms is 0.23 Å for the CMK complex structure, and it is between 0.25-0.41 Å for the 12 molecules of the orthorhombic structure, respectively). Furthermore, the overall hexameric form is also quite similar in all the three structures, indicating that the hexamer is a symmetric object with a well-defined structure.

Overall structure of the PhAAP monomer- The enzyme monomer consists of a dome-shaped, seven-bladed β-propeller domain and the C-terminal catalytic domain of α/β hydrolase fold. Each blade of the β-propeller domain contains four antiparallel β-strands. There is a large cavity with overall diameter of about 22 Å between the two domains, which contains the active site (Fig. 1A, B).

A *side opening* of this cavity (12×16 Å) is created under blade #2, gated by blades #1 and #3. (Fig. 1B). Blade #2 is both shortened and upwards shifted by 6-8 Å as compared with the corresponding blades of other family members such as POP, OPB or ApAAP, but similar to that of DPP-IV with an even greater interdomain opening of 15×22 Å formed by shortening and shift of its blades #1-#3 (9).

The single homologous structure to PhAAP available to date is that of the β-propeller domain of trilobed protease from *Pyrococcus furiosus* (TLP, sequence identity 79% and 85% for the propeller domain and the full protein chain, respectively (PDB code: 2gop (38)). Alignment of the propeller domains of PhAAP and TLP reveals high structural similarity (r.m.s. deviation of fitted Cα atoms is 0.94-1.06 Å) and the conformational differences, as well as disorder of certain loop regions of TLP can all be explained by the fact that in PhAAP these segments are involved in interactions with the hydrolase domain (that is missing from the TLP structure) or other monomers of the PhAAP multimer structure (Supplementary Fig.1).

The hydrolase domain (residues 1-14 and 347-618) that contains the catalytic triad, consists of a short N-terminal segment and the residues sequentially following the 7-bladed β-propeller domain inserted between the two (15-346). The N-terminal segment is unusually short within the family – it contains only 14 amino acid residues (as compared with the 22, 31, 72 and 94 amino acids of the same region of ApAAP, DPP-IV, POP and OPB, respectively) - while the structure of the core hydrolase domain is well conserved despite its relatively low sequence homology.

In PhAAP, the Ser466 of the catalytic triad is positioned on a short turn between a β-strand – member of an eight stranded β-sheet – and a helix, and is thus in a rather well defined environment. This feature is conserved among oligopeptidases, as is the fact that the other two members of the catalytic triad, His578 and Asp546, are placed on neighboring, longer loops (His-loop / Asp-loop), connecting β-stands of the very same β-sheet that the Ser is linked to, with helices that stand on either side of the sheet - at the point where it reaches the surface of the hydrolase domain. (Fig. 1C) Thus the

catalytic His and Asp residues provide direct coupling between the rather buried catalytic Ser and the domain surface. The Asp-loop (residues 543-552) is well structured with 4 intra-loop H-bonds, similarly to the corresponding loops of the other family members. However, the well-defined nature of the His-loop of PhAAP (residues 574-586) is unique, there are 6 intra-loop H-bonds contributing to its internal stability (Fig. 1C). The only other example of rigidified His loop among oligopeptidases is that of DPP-IV with 4 H-bonds in this region. In the rest, 0-2 intra-loop H-bonds can be found in the corresponding His-loops (0 in ApAAP (8), 1 in POP from porcine brain (4), muscle (3), and from *Aeromonas caviae* (5), 2 in POP from *Myxococcus xanthus* (39), OPB (40); PDB codes are 3o4j, 1h2w, 1qfm, 3mun, 2bkl, 2xe4, respectively); their conformations are mainly defined by their interaction with the neighboring Asp-loops and by inter-domain connections with the loops of the propeller domain. In PhAAP 5 H-bonds link the His-loop and the Asp-loop to one-another in addition to the catalytic His578—Asp546 interaction, but neither the His- nor the Asp-loop forms direct inter-domain H-bonds with the propeller domain. This is the case since, interestingly, these two loops line the entrance hole facing (through the cavity) the shortened blade #2 of the propeller – so besides participating in the catalytic reaction, these regions may partake in the screening and direction of the substrates too.

Accommodation of the covalently bound inhibitor and hexanediol molecules in the active site- In both crystal forms of the substrate-free enzyme the substrate specificity pocket was occupied by 1,6 hexanediol (Fig. 1D), which was present in the crystallization condition in high concentration (3 mol/dm^3). For studying substrate binding, a covalent adduct of PhAAP and a chloromethyl ketone inhibitor was crystallized and its structure solved. The covalently bound part, as well as the Phe and Gly moieties occupying the S1 and S2 subsites, are well defined by electron density and are bound in canonical position (Supplementary Fig. 2). The N-terminal benzyloxycarbonyl-glycyl moiety however cannot be located in the electron density map, suggesting that it does not specificly bind to the enzyme surface. Comparison of the structures containing the covalently bound inhibitor mimicking the reaction intermediate and that of the substrate-free form reveals high similarity of the two, indicating that the substrate binding site and the catalytic apparatus is pre-formed in the latter.

The active site structure is consistent with endopeptidase activity- The active site structure of PhAAP resembles that of ApAAP (Fig. 1D). The catalytic triad and the oxyanion site display active conformation in both enzymes. The conformation and interactions of the loop holding the conserved His stabilizing the oxyanion hole (His367) highly resemble that of ApAAP, as it was proposed that this structural element is characteristic to all AAPs within the prolyl oligopeptidase family (41). The substrate specificity pocket (S1) is deep, like in ApAAP (Supplementary Fig. 3) allowing the binding of large hydrophobic side chains, such as Phe or Leu. Most of the residues accommodating the P1 side chain are hydrophobic, and most of them are conserved or similar in PhAAP and ApAAP (Trp494/474, Val491/471, Tyr467/446, Val511/Leu492, Phe507/488, Ser501/481, Ile503/Phe485). However, an interesting difference can be observed. In the bottom of the S1 pocket of PhAAP a small polar (Ser497) and a negatively charged residue (Asp508) are located, while in ApAAP the corresponding side chains are Met477 and Ile489, respectively, suggesting that PhAAP may also cleave after positively charged residues (Supplementary Fig. 3). Indeed, we observed that in the absence of Phe at the P1 site, substrates with Arg or Lys can also be hydrolyzed (data not shown).

In addition to the acylaminoacyl peptidase activity, PhAAP also exhibits endopeptidase activity (13). The structure is in line with this finding. The substrate binding site is extended and does not have any steric or electrostatic blocking at the S2-S3 sites which true exopeptidases, like DPP-IV, do.

Stabilization of the PhAAP hexameric structure: trimer of dimers- In a previous study we have shown, using size exclusion chromatography, that PhAAP is a homohexameric enzyme in solution (13). Hexamers of PhAAP in the crystalline state are formed as trimers of dimers of the monomer enzyme (Fig. 2A,B). Analysis of molecular contacts revealed that the association within the dimers is much stronger than that securing trimerization. Both the area of the buried surfaces and the free enthalpy contribution of the dimers dominate over those of the trimers (Supplementary Table 1). Though the highly homologous enzyme, TLP from *Pyrococcus furiosus* was proposed to be trimeric (38), we suggest, it may form hexamers in solution because the sequence identity between PhAAP and TLP is even higher at the contact regions of the trimer interfaces formed by the propeller domains (81% compared to 79% for the propeller domain) as well as for the dimer interfaces formed by both domains (96%, compared to 85% for the whole protein) (Supplementary Table 2)

The *dimer* interface lies between the long loop insertion of blade #3 from the propeller domain of one monomer and the hydrolase domain of the other monomer (Fig. 2C). The long insertion of one molecule lines up to and extends the central 8-stranded β-sheet holding the catalytic residues of the other, by forming the 9th strand of it in an antiparallel manner (framed in Fig. 2C; Supplementary Fig. 4). In contrast, in ApAAP, which is active in the dimeric form, the dimer interface is formed between the two hydrolase domains so that the central β-sheets of the two hydrolase domains are connected by forming a 16-stranded continuous sheet, while the propeller domain of one momomer has no contacts with the either domains of the other (Supplementary Fig. 4).

The *trimer* interfaces are composed of the outer side of blade #1 of the propeller of one monomer and *the* outward facing part of blades #2 and #3 of its neighboring monomer (Fig. 2D, blades #1-#3 are labeled for the three monomers), and are dominated by side chain interactions. Blades #1-#3 are those lining the entrance of the monomer, thus trimerization contributes to fixing of the pore size, by restricting their upward movement.

Hexamerization creates further substrate-size screeners. Three large openings of about 20×30 Å can be found on the hexamer surface (Fig. 2B ,E), which continue in tunnels and are joined in a big cavity in the middle of the hexamer. The interdomain opening of each monomer is also connected to this central chamber of the hexamer, since all six cavities face inward. In principle, the substrate can reach one of the active sites in two ways: 1) first entering the opening of the hexamer and subsequently the side-chamber of a monomer, or 2) through the central pore of the propeller domain, provided that the pore can be widened (42). However, the trimerization contacts restrict flexibility of the propeller pore region, which makes this second option unlikely. (Fig. 2E) The two successive openings and the complex tunnel system between them seem sufficient for screening the substrates - those molecules that are small enough to enter the antechamber through the larger hole but are not flexible enough for entering the second, may simply pass through the central chamber and leave at the opposite end, however those that can proceed through the monomer-cavity to the active site, will be cleaved.

MD simulations for determining monomer stability- MD simulations were carried out to determine the monomer stability of two members of the prolyl oligopeptidase family, those of POP and PhAAP. Calculations were designed to test the significance of multimerization – POP is active and stable as a monomer, while PhAAP hexamerizes. Monomer structures were compared after 60 ns of simulation time. The two molecules behave quite similarly; the conformation seen in the crystal structure of each was preserved in the simulation. This has a special significance in case of the PhAAP structure since it also demonstrates that the 1,6-hexanediol molecule trapped in the active site in crystalline state and absent in the simulation does not have a structure determining role. The overall r.m.s. deviation of main chain atoms from their crystallographically determined position was found to be 1.34 Å and 1.44 Å in POP and PhAAP, respectively (Supplementary Fig. 5).

A notable difference between the crystal structures and the simulation results can be seen in the conformation of the catalytic triad – both in the closed structure of POP and in PhAAP with its wide side opening. The H-bonded triad of the crystal structures loosens in the fully solvated monomer. The Ser-Oγ—His-Nε2 distance increases from 2.9 Å and 2.6 Å to 4.7 Å and 4.6 Å, in case of POP and PhAAP, respectively – due mainly to the active Ser sampling a different rotamer, relaxing its strained, nearly eclipsed N_{+1}-C-Cα-Cβ torsion of 2-5° seen in the crystal structures to that of a more staggered 35°. Ser and His stay connected through a shared water molecule in 58 % and 19 % of the structures of the last 10 ns of the trajectory in POP and PhAAP, respectively, and through two or more coordinated water molecules in the rest. The His-Asp H-bond remains intact in all snapshots. Such a "latent" conformation of the Ser-His-Asp catalytic triad was seen in the crystal structure of ligand-free tricorn interacting factor F1 (43), where - in the absence of propeller domain - the hydrolase is only partially shielded from solvent. In crystal structures of other oligopeptidases, in addition to the strained backbone conformation of the catalytic Ser, the above mentioned N_{+1}-C-Cα-Cβ torsion is also in a high energy conformation of -6° – 7°. It seems that the partially "dried-out" condition of crystallization or the binding of a substrate is required to force the Ser into the conformation that guarantees its reactivity.

DISCUSSION

When comparing the structure of PhAAP presented here and those of oligopeptidases previously determined, three fundamentally different substrate admission routes can be detected: i) allowing access of substrates through the propeller hole of the closed form, ii) dynamic domain-flapping between closed and open conformation or iii) a channel system created by multimerization.

The first provides an almost too strict screener, the narrow pore of the propeller domain cannot be appropriate for all oligopeptidase substrates. However, DPP-IV, for example, with an eight-bladed propeller domain and a correspondingly wider pore than those of POP, ApAAP, OPB or PhAAP was proposed to function that way (9), although - since it can also cleave even 80 amino acids long peptides– it must apply other substrate admission routes too (10).

POP and ApAAP seem to utilize the second strategy. In the closed conformation of these enzymes the hydrolase and propeller domains sit on top of each other, burying their active site. In their open form the relative domain position is fixed only on one side by a hinge region, while at the opposite side of the molecules the domains flap away from each-other forming a 40-50° opening, causing an approximately 20 Å shift at the point of the greatest opening. This conformation provides an easy access to the active site. In all cases where such open form was observed, the active site was disassembled and inactivated by the opening, chiefly through the destabilization of the loop holding the catalytic His residue. In the closed form this loop is supported by inter-domain H-bonds that are lost when the opening takes place and the domains move away from each-other (6,8).

In the crystal structure of PhAAP described in this paper the third strategy can be witnessed. Hexamerization creates a compartmentalized inner space, with a complex, double-gated "check-in" system. Substrates first have to pass through the entrance at the hexamer surface. (Fig. 2E). Once inside, the active sites can be reached through the spacious side opening of the monomers created by the shortening and upward shift of #2 of the seven blades of the propeller domains.

The overall topology enables all oligopeptidases to widen their propeller pore in an induced fit step, multimerize, or to open – thus, to utilize any of the three outlined strategies. The interesting question is which structural features encode selecting one or the other.

A crucial feature to be considered is the flexibility of the loop containing the His residue of the catalytic triad (His-loop). Pliability of this region is essential for those oligopeptidases that open-up, since selectivity is assured by "switch-off" of the active site in the open form by destabilization of the His-loop. Accordingly, in case of POP, OPB and ApAAP the shape of the His loop and the position and orientation of the catalytic His is defined only by inter-loop contacts, both with the neighboring Asp-loop containing the third member of the triad, and with the facing loops of the propeller domain. Therefore the shape and stability of the His-loop is coupled to the proximity of the propeller domain. In contrast, in multimer enzymes with permanent entrance between the two domains (PhAAP and DPP-IV) the His-loop is rigidified in the active conformation by a number of intra-loop H-bonds.

In all oligopeptidase structures so far determined, the His-loop connects an α-helix and the terminal β-strand of the central β-sheet of the hydrolase domain - both at the surface of the monomer. Interestingly, this edge β-strand of 6-8 residues is of regular extended geometry in all cases, without twists, bulges or prolines to disrupt its H-bonding potential, thus it is an ideal aggregation primer (44). The corresponding sequences were recognized by the WALTZ predictor (45) as having high probability for amyloid formation – PhAAP scoring the highest value, 98.3, while the others between 79.9 and 89.0.

The propensity for amyloid-type aggregation is a generic property of proteins (46), which thus have evolved structural and sequential adaptations to protect their surface-close β-strands (44, 47, 48), because aggregation prompted by such segments, even in the native fold or in locally unfolded states, has been shown to be the first step of toxic aggregation processes (49, 50). Negative selection against conformers with significant tendency to aggregate is now regarded as one of the key determinants in the development of complex molecular machineries (51) such as seen in case of PhAAP in the present study. (Fig. 3).

In monomeric oligopeptidases POP and OPB, the sticky β-strand is covered by a 72 and 95 residue long N-terminal extension, respectively (Fig. 3A, left; Supplementary Fig. 4). Although the extensions join to the amino acids of the propellers in sequence, they run alongside the hydrolase domains and form a complex strap around it. Numerous H-bonds are established between them, the

majority of which fix the strap to the sticky β-strand and the two α-helices that stand on either side of it (12-14 and 17 in total, 8-11 and 9 to the α-helix/β-strand/α-helix triad in POPs from different sources and in OPB, respectively). Since the structure of POP was determined in the open conformation too, it could be ascertained that these interactions are unchanged by the opening. The N-terminal strap might also be a promoter of monomer stability, but our MD simulations showed that the PhAAP monomer, devoid of such a segment, preserved its structure just as well as POP, supporting the notion that the N-terminal extension might instead be a safeguard against aggregation.

In ApAAP, DPP-IV and in PhAAP the edge β-strand is unprotected in the monomer - these oligopeptidases multimerize. In ApAAP the N-terminal extension is only 22 amino acids long, it forms a helix that runs parallel to the α-helix/β-strand/α-helix triad and leaves the sticky-β-strand fully unprotected (Fig. 3A, right). However in the functional ApAAP dimer the β-strand of one monomer is covered by the corresponding strand of the other in an antiparallel manner joining their central 8 β-strands to form a 16-pleated continuous β-sheet (Supplementary Fig. 4). This arrangement results in the two active sites facing in the opposite direction, which allows them to function independently; they might even adopt strikingly different conformations as was seen in the open/closed mixed dimer structure of ApAAP, where one monomer was found in an open, the other in a closed conformation. Flexibility is retained since only the hydrolase domains participate in the dimerization – and this is essential, since the structure of the closed form of the ApAAP monomer has no side opening to allow access of substrates.

Small hindrances obstruct full access to the exposed β-strands of DPP-IV and PhAAP (Fig. 3B). In DPP-IV a 30 amino acids long insertion of blade #4 of the propeller is what reaches to the top of the edge-β-strand, while in PhAAP a 12 residue-long C-terminal extension comes in proximity of the bottom end of it. Both behave as a peg, making it impossible for the head-on dimerization seen in case of ApAAP to take place. DPP-IV dimerizes, in the dimer the insertions from each monomer form an X-shaped cross, while the 8-pleated central β-sheets meet in a perpendicular orientation. In PhAAP the loop insertion of blade #3 of one monomer lines up to the sticky β-strand of the other and forms a 9^{th}, antiparallel strand of the central β-sheet making a dimer, while the dimers are further organized into the hexamer (Supplementary Fig. 4). Since in both DPP-IV and PhAAP, both the hydrolase and the propeller domains participate in multimerization, domain movements necessarily become restricted, thus these enzymes must possess a fixed substrate entrance for effectiveness – which they do. The mode of multimerization also leads to self-compartmentalizing, especially in the hexamer structure of PhAAP, since the active site openings point inward toward a common central chamber formed by the association. This therefore provides for a sufficient substrate screening mechanism – for maintaining selectivity.

Experimental evidence also supports the suggested role of the N-terminal pegs of oligopeptidases. Removal of the N-terminal segment of POP (where no additional protection for the sticky-β-strand is available) results in aggregation and inactivation depending on the size of the removed segment (52) In contrast, removing the N-terminal extension of ApAAP causes only some destabilization of the enzyme with lowering the melting temperature by about 20 K (53).

The 8^{th} β-strand of the central β-sheet supporting the His-loop of α/β hydrolases is not amylogenic in all cases. Dehalogenases, haloperoxidases, epoxide hydrolases and hydroxynitrile lyase remove the aggregational risk by inserting prolines (PDB codes in parentheses for the following: 2had (54)) or charged amino acids (4fwb (55), 3a2m (56), 1a88 (57), 1a7u (57), 3koo (58), 2zjf (59), 3c70 (60)) into the strand holding the His-loop, lowering their amylogenic propensity considerably, or by introducing a significant twist to this segment (2bfn (61)). Sticky β-edge is buried by further, non-amylogenic β-strands in other α/β hydrolases like acetylcolin esterases, serine carboxypeptidases and lipases. Peroxisomal α/β hydrolase Lpx1 forms a dimer in a similar fashion to ApAAP, covering its unprotected amylogenic β-strand by the process (62). However, surprisingly, several monomeric α/β hydrolases (prolyl iminopeptidase (1qtr (63), 1azw (64)), thioesterase (3gro (65)), dienelactone hydrolase (1din (66)) and carboxylestherase (1auo (67)) provide no protection for their highly amylogenic β-edge (with WALTZ values (45) between 82.4 and 92.2). This might be because other interaction partners, cofactors, glycosylation or closeness of the membrane obstructs free access to the implicated β-edge in these enzymes.

Outside the α/β hydrolase clan, among self-compartmentalizing proteases, tricorn protease, for example, has a multidomain structure similar to that of PhAAP. This enzyme carries out the degradation of small (7-9 amino acids long) peptides and just as PhAAP, is also hexameric, although its multimer-architecture is of a bit simpler, toroidal shape. Its monomer, however, is more complex than that of PhAAP, consisting of 5 subdomains. Its α/β catalytic region is capped on both sides by propeller domains - substrates reach the active site by passing the great opening of the toroid then proceeding through the channel of a seven-bladed propeller – a permanently open route to its active site. Instead of a catalytic triad, a four residue H-bonded system of Ser-His-Ser-Glu performs the catalytic reaction. Glu1023 is situated on a long loop connecting an unprotected strand and a distant helix. The seven-bladed propeller is close enough to obstruct dimerization with the corresponding strand of another monomer, but too far for shielding, thus a β-turn-β type insert from the neighboring monomer covers it via hexamerization similarly to that seen in case of PhAAP.

The proteasome, a further self-compartmentalizing system, forms an immense multilayer, barrel shaped complex, where 4 rings, each made up of 7 monomers, stack on each-other to compose 2 antechambers and one central catalytic chamber. Proteasome monomers do not have a propeller domain (so they have a very large permanent hole above the hydrolase). The catalytically active Thr is located on the penultimate β-strand of a 5-membered sheet, where - in the heptamer - the β-edge of the sheet in one monomer is covered by an insert of the neighboring monomer, resulting in a self-assembled compartment system of inward turned active sites in this case too.

Based on the above we propose the mode of multimerization and self-assembly among oligopeptidases is fine-tuned by a shielding intent of a sticky β-edge, insertions and N- and C-terminal extensions, while to maintain effectiveness of catalysis and selectivity, pliability of the His-loop and the position of the propeller blades are adjusted (as summarized in Table 2.).

REFERENCES

1. Szeltner, Z., and Polgar L. (2008) Structure, function and biological relevance of prolyl oligopeptidase. *Curr Protein Pept Sci.* **9**, 96-107
2. Polgar, L. (2002) The prolyl oligopeptidase family. *Cell Mol. Life Sci.* **59**, 349-362
3. Fulop, V., Bocskei, Z., and Polgar, L. (1998) Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. *Cell* **94**, 161-170
4. Szeltner, Z., Rea, D., Renner, V., Fulop, V., and Polgar, L. (2002) Electrostatic effects and binding determinants in the catalysis of prolyl oligopeptidase: Site-specific mutagenesis at the oxyanion binding site. *J. Biol. Chem.* 277:42613–42622
5. Chiu, T. K., and Li, M. Route of substrate entry in prolyl endopeptidase. *(to be published);* PDB entries 3mun and 3muo
6. Li, M., Chen, C., Davies, D. R., and Chiu, T. K. (2010) Induced-fit mechanism for prolyl endopeptidase. *J. Biol. Chem.* **285**, 21487-21495
7. Shan, L., Mathews, I. I., and Khosla, C. (2005) Structural and mechanistic analysis of two prolyl endopeptidases: role of interdomain dynamics in catalysis and specificity. *Proc. Natl. Acad. Sci. USA* **102**, 3599-3604
8. Harmat, V., et al. (2011) Structure and catalysis of acylaminoacyl peptidase: closed and open subunits of a dimer oligopeptidase. *J. Biol. Chem.* **286**, 1987-1998
9. Engel, M., et al. (2003) The crystal structure of dipeptidyl peptidase IV (CD26) reveals its functional regulation and enzymatic mechanism. *Proc. Natl. Acad. Sci. USA* **100**, 5063-5068
10. Engel, M., et al. (2005) Rigidity and flexibility of dipeptidyl peptidase IV: crystal structures of and docking experiments with DPIV. *J Mol Biol* 355:768-783
11. Hiramatsu, H., et al. (2003) The structure and function of human dipeptidyl peptidase IV, possessing a unique eight-bladed beta-propeller fold. *Biochem. Biophys. Res. Commun.* **302**, 849–854
12. Aertgeerts, K., et al. (2004) Crystal structure of human dipeptidyl peptidase IV in complex with a decapeptide reveals details on substrate specificity and tetrahedral intermediate formation. *Protein Sci.* **13**, 412–421
13. Szeltner, Z., et al. (2009) Characterization of a novel acylaminoacyl peptidase with hexameric structure and endopeptidase activity. *Biochim. Biophys. Acta* **1794**, 1204–1210
14. Joshua-Tor, L., Xu, H. E., Johnston, S. A., and Rees, D. C. (1995) Crystal structure of a conserved protease that binds DNA: the bleomycin hydrolase, Gal6. *Science* **269**, 945-950
15. Kim, J. S., et al. (2002) Navigation inside a protease: substrate selection and product exit in the tricorn protease from Thermoplasma acidophilum. *J. Mol. Biol.* **324**, 1041-1050
16. Cha, S. S., et al. (2010) Crystal structure of Lon protease: molecular architecture of gated entry to a sequestered degradation chamber. *EMBO J.* **29**, 3520-3530
17. Kiss, A. L., et al. (2007) The Acylaminoacyl Peptidase from Aeropyrum pernix K1 Thought to Be an Exopeptidase Displays Endopeptidase Activity. *J. Mol. Biol.* **368**, 509-520
18. Palmieri, G., et al. (2011) Acylpeptide hydrolase inhibition as targeted strategy to induce proteasomal down-regulation. *PLoS One* **6(10)**, e25888.
19. Naylor, S. L, et al. (1989) The DNF15S2 locus at 3p21 is transcribed in normal lung and small cell lung cancer. *Genomics* **4**, 355-361
20. Scaloni, A., et al. (1992) Deficiency of acylpeptide hydrolase in small-cell lung carcinoma cell lines. *J. Lab. Clin. Med.* **120**, 546–552

21. Erlandsson, R., et al. (1991) The gene from the short arm of chromosome 3, at D3F15S2, frequently deleted in renal cell carcinoma, encodes acylpeptide hydrolase. *Oncogene* **6**, 1293–1295

22. Duysen EG, et al. (2001) Evidence for nonacetylcholinesterase targets of organophosphorus nerve agent: supersensitivity of acetylcholinesterase knockout mouse to VX lethality. *J. Pharmacol. Exp. Ther.* **299**, 528–535

23. Richards, P. G., Johnson, M. K., and Ray, D. E. (2000) Identification of acylpeptide hydrolase as a sensitive site for reaction with organophosphorus compounds and a potential target for cognitive enhancing drugs. *Mol Pharmacol* **58**:577–583

24. Kabsch, W. (2010) XDS. *Acta Crystallogr. D* **66**, 125-132

25. Sheldrick, G. M. (2008) A short history of SHELX. *Acta Crystallogr. A* **64**, 112-122

26. Otwinowski, Z. (ed) (1991) Isomorphous Replacement and Anomalous Scattering. Daresbury Laboratory, Warrington, UK

27. Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* 50:760–763

28. Cowtan, K. (2006) The Buccaneer software for automated model building. *Acta Crystallogr. D* **62**, 1002-1011

29. Vagin, A., and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30,** 1022–1025

30. Emsley, P., and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126-2132

31. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240-255

32. Afonine, P. V., et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352-356

33. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. D* **55**, 191-205

34. Chen, V. B., et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12-21

35. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435-447

36. MacKerell, A. D. Jr, Feig, M., and Brooks, C. L. 3rd (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400-1415.

37. Bjelkmar, P., Larsson, P., Cuendet, M. A., Hess, B., and Lindahl, E. (2010) Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *J. Chem. Theory Comput.* **6**, 459-466

38. Bosch, J., Tamura, T., Tamura, N., Baumeister, W., and Essen, L. O. (2007) The beta-propeller domain of the trilobed protease from Pyrococcus furiosus reveals an open Velcro topology. *Acta Crystallogr. D* **63**, 179-187

39. Shan, L., Mathews, I. I., and Khosla, C. (2005) Structural and mechanistic analysis of two endopeptidases: role of interdomain dynamics in catalysis and specifity. *Proc. Natl. Acad. Sci. USA* **102**, 3599-30604

40. McLuskey, K., Paterson, N. G., Bland, N. D., Isaacs, N. W., and Mottram, J. C. (2010) Crystal stucture of Leishmania major oligopeptidase B gives insight into the enzymatic properties of a trypanosomatid virulence factor. *J. Biol. Chem.* **285**, 39249-39259

41. Kiss, L., et al. (2008) Structural and kinetic contributions of the oxyanion binding site to the catalytic activity of acylaminoacyl peptidase. *J. Struct. Biol.* **162**, 312-323

42. Fuxreiter, M., et al. (2005) Flexibility of prolyl oligopeptidase: molecular dynamics and molecular framework analysis of the potential substrate pathways. *Proteins* **60**, 504-512

43. Goettig, P., Groll, M., Kim, J. S., Huber, R., and Brandstetter, H. (2002) Structures of the tricorn-interacting aminopeptidase F1 with different ligands explain its catalytic mechanism. *EMBO J.* **21**, 5343-5352

44. Richardson, J. S., and Richardson, D. C. (2002) Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. USA* **99**, 2754-2759

45. Maurer-Stroh, S., et al. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods.* **7**, 237-242

46. Dobson, C. M. (2003) Protein folding and misfolding. *Nature* **426**, 884-890

47. Clark, L. A. (2005) Protein aggregation determinants from a simplified model: Cooperative folders resist aggregation. *Protein Sci.* **14**, 653-662

48. Monsellier, E., and Chiti, F. (2007) Prevention of amyloid-like aggregation as a driving force of protein evolution. *EMBO reports*, **8**, 737-742

49. Soldi, G., Bemporad, F., Torrassa, S., Relini, A., Ramazzotti, M., Taddei, N., and Chiti, F. (2005) Amyloid formation of a protein in the absence of initial unfolding and destabilization of the native state. *Biophys. J.* **89**, 4234-4244

50. Chiti, F., and Dobson, C. M. (2009) Amyloid formation by globular proteins under native conditions. *Nature Chem. Biol.* **5**, 15-22

51. Stefani M., (2004) Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochim Biophys Acta* **1739**, 5-25

52. Juhasz, T., Szeltner, Z., and Polgar, L. (2007) Truncated prolyl oligopeptidase from Pyrococcus furiosus. *Proteins* **69**, 633-643

53. Zhang, Z., et al. (2008) The conserved N-terminal helix of acylpeptide hydrolase from archaeon Aeropyrum pernix K1 is important for its hyperthermophilic activity. *Biochim. Biophys. Acta* **1784**, 1176–1183

54. Franken, S. M., Rozeboom, H. J., Kalk, K. H., Dijkstra, B. W. (1991) Crystal structure of haloalkane dehalogenase: an enzyme to detoxify halogenated alkanes. *EMBO J.* **10**, 1297-1302

55. Lahoda, M., Stsiapanava, A., Mesters, J., Chaloupkova, R., Damborsky, J., and Kuta Smatanova, I. Crystallographic analysis of 1,2,3 trichloropropane biodegradation by haloalkane dehalogenase DhaA31. *(to be published);*PDB entry 4fwb.

56. Prokop, Z., et al. (2010) Enantioselectivity of haloalkane dehalogenases and its modulation by surface loop engineering. *Angew. Chem. Int. Ed. Engl.* **49**, 6111-6115

57. Hofmann, B., et al. (1998) Structural investigation of the cofactor-free chloroperoxidases. *J. Mol. Biol.* **279**, 889-900

58. Eldrup, A. B., Soleymanzadeh, F., Farrow, N. A., Kukulka, A., and De Lombaert, S. (2010) Optimization of piperidyl-ureas as inhibitors of soluble epoxide hydrolase. *Bioorg. Med. Chem. Lett*. **20**, 571-575

59. Biswal, B. K., et al. (2008) The molecular structure of epoxide hydrolase B from Mycobacterium tuberculosis and its complex with a urea-based inhibitor. *J. Mol. Biol.* **381**, 897-912

60. Schmidt, A., Gruber, K., Kratky, C., and Lamzin, V. S. (2008) Atomic resolution crystal structures and quantum chemistry meet to reveal subtleties of hydroxynitrile lyase catalysis. *J. Biol. Chem.* **283**, 21827-21836

61. Monincova, M., Prokop, Z., Vevodova, J., Nagata, Y., and Damborsky, J. (2007) Weak activity of haloalkane dehalogenase LinB with 1,2,3-trichloropropane revealed by X-Ray crystallography and microcalorimetry. *Appl. Environ. Microbiol.* **73**, 2005-2008

62. Thoms, S., Hofhuis, J., Thoing, C., Gartner, J., and Niemann, H. H. (2011) The unusual extended C-terminal helix of the peroxisomal α/β-hydrolase Lpx1 is involved in dimer contacts but dispensable for dimerization. *J. Struct. Biol.* **175**, 362-371

63. Rould M. A., Perona J.J., and Steitz T.A. (1991) Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. *Nature* **352**, 213-218

64. Medrano, F.J., et al. (1998) Structure of proline iminopeptidase from Xanthomonas campestris pv. citri: a prototype for the prolyl oligopeptidase family. *EMBO J.* **17**, 1-9

65. Dobrovetsky, E., Seitova, A., Tong, Y, Tempel, W., Dong, A., Arrowsmith, C. H., Edwards, A. M., Bountra, C., Weigelt, J., Bochkarev, A., Cossar, D., and Park, H. Human palmitoyl-protein thioesterase 1 *(to be published);* PDB entry 3gro.

66. Pathak, D., and Ollis, D. (1990) Refined structure of dienelactone hydrolase at 1.8 Å. *J. Mol. Biol.* **214**, 497-525

67. Kim, K. K., et al. (1997) Crystal structure of carboxylesterase from Pseudomonas fluorescens, an alpha/beta hydrolase with broad substrate specificity. *Structure* **5**, 1571-1584

Acknowledgments
This work was supported by the Hungarian Scientific Research Fund (OTKA) PD101095, NK101072, NK67800. We acknowledge the European Synchrotron Radiation Facility (Grenoble, France) and the EMBL outstation at DESY (Hamburg, Germany) for provision of synchrotron radiation facilities and we thank Dr. Hassan Belrhali and Dr. Santosh Panjikar for assistance in using beamlines ESRF BM14 and DESY X11, respectively.

Accession codes - Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under the accession codes 4HXE, 4HXF and 4HXG.

Abbreviations - AAP, acylaminoacyl peptidase; ApAAP, *Aeropyrum pernix* acylaminoacyl peptidase; CMK, benzyloxycarbonyl-glycyl-glycyl-phenylalanyl-chloromethyl ketone; DPP-IV, dipeptidyl peptidase IV; MD, molecular dynamics; OPB, oligopeptidase B; PhAAP, *Pyrococcus horikoshii* acylaminoacyl peptidase; PDB, protein data bank; POP, prolyl oligopeptidase; TLP, *Pyrococcus furiosus* trilobed protease;

FIGURE LEGENDS

FIGURE 1. Structure of PhAAP reveals two entrances to the interdomain cavity. The structure of the monomer is shown in cartoon representation (A) and its cross section is shown with molecular surface (wheat, B). The two entrances to the central cavity of the monomer are shown (the pore of the β-propeller domain and the side opening formed between blade #1 and insertion of blade #3 of the propeller domain). C. The loops holding the catalytic triad (Ser466, His578, Asp546; light blue, orange and green, respectively) are exposed to solvent and form only few contacts with the propeller domain (yellow), yet they are stabilized by intra- and inter-loop H-bonds. (H-bonds are color coded: red: intra-His-loop; green: intra-Asp-loop, light blue: between His- and Asp-loops, black: formed with the rest of the hydrolase domain; yellow: with the propeller domain). D: Comparison of the PhAAP structure with hexanediol (wheat) bound in the substrate specificity pocket, and that of PhAAP in complex with a covalently bound chloromethyl ketone inhibitor (magenta; with two hexanediol molecules outside the S1-S2 region shown in wheat) reveal the only significant difference is adjustment of the S1 pocket by rotation of Phe507 side chain in the covalent complex (in darker colors).

FIGURE 2. Structure of the symmetric hexamer of PhAAP allows for substrate selection by a rigid gate-system. Panels A and B show top and side views of the hexamer formed as trimer of dimers. There are three large side pores leading to a central chamber of the complex. C: The dimers are stabilized mainly by contacts of the insertion of the propeller domain and the hydrolase domain of the other monomer, forming extension of the central β-sheet. D: The trimers are stabilized by interactions of blades #1-#3 of the propellers. E: Cross section of the hexamer with molecular surfaces of PhAAP molecules (color code is similar to panel A, with the propeller domains in darker shades). Substrate oligopeptides entering the central chamber through the side pores of the hexamer should be flexible enough to pass through the narrower interdomain opening of the monomers to get to the active site.

FIGURE 3. Strategies of shielding the sticky β-edge of the hydrolase domain to prevent aggregation within the prolyl oligopeptidase family. Hydrolase domain cores are shown in light blue with a red fork at its β-edge. Regions covered by other monomers or segments are colored yellow. The propeller domains are shown in grey-blue. A: In monomeric enzymes, like prolyl oligopeptidase and oligopeptidase-B a long N-terminal segment (blue) wraps around the hydrolase core and shields the β-edge. In *Aeropyrum pernix* AAP dimer the other monomer covers the β-edge forming a continuous β-sheet. The propeller domains are not involved in these interactions and remain floppy, allowing dynamic open/close equilibrium for substrate screening. B: In DPP-IV and *Pyrococcus horikoshii* AAP bulky regions (grey mesh) hinder simple hydrolase/hydrolase β-sheet formation and both domains are involved in covering the sticky β-edge. As a consequence, domain positions are fixed. In these enzymes permanent pores or channel system ensure size selection of substrates.

Table 1 Data collection, phasing and refinement statistics (MIRAS)

	Native (unligated)	Uranyl derivative	Iodine derivative	Platinum derivative		CMK inhibitor complex	Orthorhombic unligated
Data collection[a]	In house	In house	In house	DESY X11	ESRF BM14	In house	ESRF BM14
Space group	H 3 2	H 3 2	H 3 2	H 3 2	H 3 2	H 3 2	P 2_1 2_1 2_1
Cell dimensions							
a, *b*, *c* (Å)	183.000, 183.000, 144.632	182.953, 182.953, 143.936	182.819, 182.819, 144.443	184.802, 184.802, 145.727	184.802, 184.802, 145.727	184.057, 184.057, 145.529	183.31, 183.80, 275.74
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 90
Wavelength	1.5418	1.5418	1.5418	0.8162	1.0715	1.5418	1.0715
Resolution (Å)[b]	20.0-1.90 (2.00-1.90)	20.0-2.5 (3.0-2.5)	20.0-2.25 2.5-2.25)	20.0-2.00 (2.15-2.00)	20.0-2.50 (2.80-2.50)	20.0-1.60 (1.64-1.60)	20.0-2.65 (2.72-2.65)
R_{meas} [b]	0.158 (0.686)	0.141 (0.319)	0.138 (0.357)	0.112 (0.414)	0.071 (0.212)	0.066 (0.738)	0.090 (0.690)
$I / \sigma I$ [b]	12.79(1.87)	10.94(5.18)	10.22(3.74)	15.35(5.02)	30.97(12.35)	22.25(2.57)	16.47(2.36)
Completeness (%)[b]	95.7(74.6)	99.2(100.0)	98.8(99.1)	99.4 (97.4)	99.5(99.0)	99.2(94.4)	94.9(96.8)
Redundancy [b]	8.32(2.84)	4.05(4.07)	2.80(2.66)	6.10(5.96)	11.53(11.43)	7.03(5.24)	4.39(4.33)
Refinement							
Resolution (Å)	19.78-1.91					19.74-1.60	19.98-2.70
No. reflections[c]	64801 (3425)					116219 (6218)	222127 (1751)
R_{work} / R_{free}	0.172 / 0.210					0.155 / 0.186	0.200 / 0.241
No. of monomers / asymmetric unit	1					1	12
No. atoms							
Protein	5082					5093	58518
Ligand/ion	51					101	231
Water	395					697	1161
B-factors							
Protein	21.23					21.33	67.98
Ligand/ion	39.16					40.05	57.74
Water	28.73					33.89	45.22
R.m.s deviations							
Bond lengths (Å)	0.018					0.025	0.005
Bond angles (°)	1.944					2.248	0.979

[a]Each data set was collected from one crystal.
[b]Values in parentheses are for highest-resolution shell.
[c]Reflection number used in R_{free} calculation is shown in parenthesis.

Table 2 Strategies of size selectivity and ways of shielding the sticky β-edge of the hydrolase domain in the prolyl oligopeptidase family

β-edge coverage in the monomer	**Functional unit**	**Permanent side entrance**	**Pliability of His-loop**	**Substrate screening strategy**	**Known structures**
Full	Monomer	No	Pliable	Dynamic: inactive open/active closed states	POP, OPB*
None	Dimer	No	Pliable	Dynamic: inactive open/active closed states	ApAAP
Partial	Self-assembly	Yes	Rigid	Permanent pores or self-compartmentalization	DPP-IV, PhAAP

*OPB is monomeric, however, no experimental evidence was, as of yet, obtained confirming its ability to open.

Figure 1.

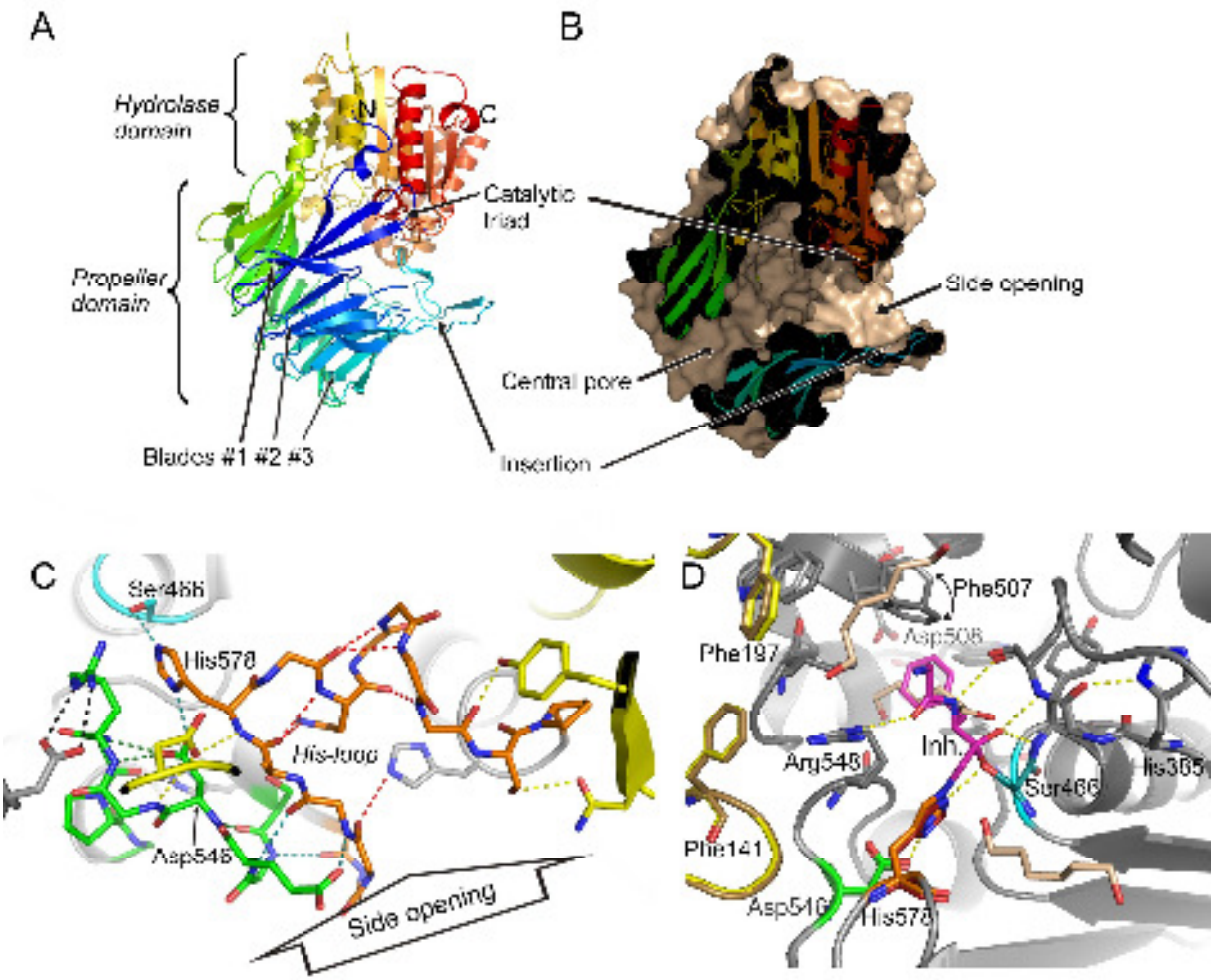

A
Hydrolase domain
N
C
Propeller domain
Blades #1 #2 #3
B
Catalytic triad
Side opening
Central pore
Insertion
C
Ser466
His578
His-loop
Asp545
Side opening
D
Phe507
Asp508
Phe197
Inh.
Arg543
Ser466
His365
Phe141
Asp510
His578

Figure 2.

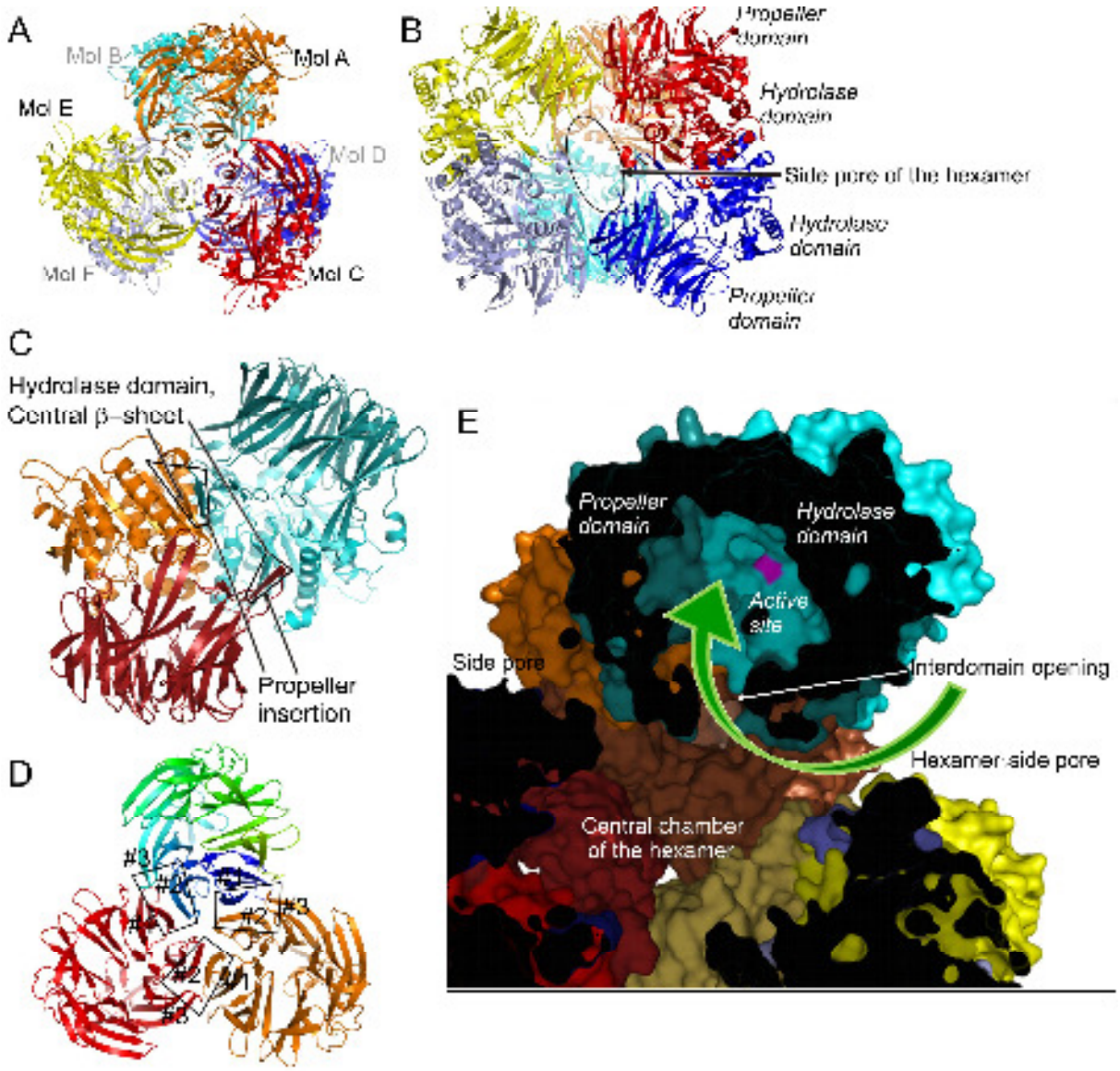
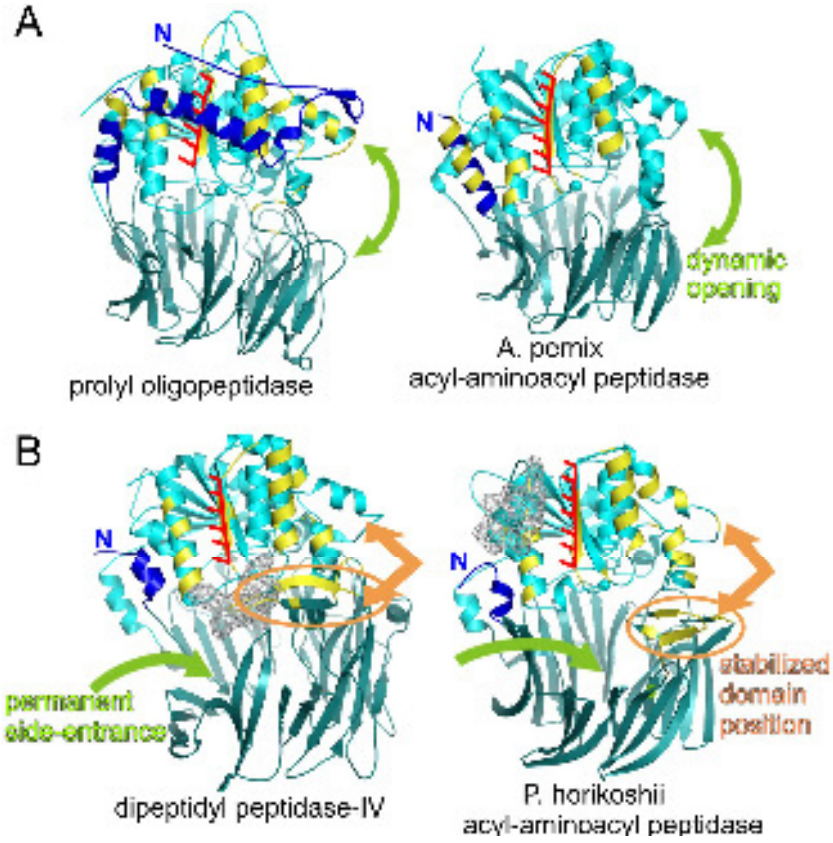

A
Mol B
Mol A
Mol E
Mol D
Mol F
Mol C
B
Propeller domain
Hydrolase domain
Side pore of the hexamer
Hydrolase domain
Propeller domain
C
Hydrolase domain, Central β–sheet
Propeller insertion
D
#3
#4
#1
#2
#3
#2
#1
#3
E
Propeller domain
Hydrolase domain
Active site
Side pore
Interdomain opening
Hexamer side pore
Central chamber of the hexamer

Figure 3.


A
N
prolyl oligopeptidase
N
dynamic opening
A. pernix
acyl-aminoacyl peptidase
B
N
permanent side-entrance
dipeptidyl peptidase-IV
N
stabilized domain position
P. horikoshii
acyl-aminoacyl peptidase