In this project we set out to gain a better understanding of the substrate binding and selection mechanisms of acyl-aminoacyl peptidase (AAP), an enzyme that functions as an exopeptidase (removing N-terminally blocked amino acids from peptides), while possessing endopeptidase activity too. It has been shown to act in concert with and exert regulation over the proteasome while also being implicated in renal and small cell carcinoma – thus is an important target of cancer therapy. AAP can also cleave amyloid-beta peptides, thus it might present a new tool for therapy aimed at reducing neurodegeneration in the Alzheimer's brain.

AAP is an oligopeptidase, which means that it is able to distinguish between potential substrates based on their size and only cleave those that do not exceed 30 amino acids in length. Structurally, similarly to all other oligopeptidases, it consist of a half-sphere-shaped hydrolase domain 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 dimensions comparable with the size limitations imposed by the enzyme.

The crystal structure of mammalian AAP has long been sought after but, as of yet, without success. Diffracting crystals were reported to have been obtained in 1993 and in 2005, but the structure could not be determined in either case due to their instability in the X-ray beam. The only structural model available at the beginning of this project, was of an archeal source - AAP of *Aeromonas punctata* (ApAAP).

During the project we determined and analyzed the structure of a hexameric AAP from *Pyrococcus horikoshii* (PhAAP), determined further forms of the dimeric ApAAP which helped elucidate its substrate binding process and obtained crystals of the tetrameric mammalian enzyme (of porcine liver – pAAP) that remained stable during the X-ray diffraction measurements and collected data of sufficient resolution for determining its structure – data processing is currently in progress.

PhAAP is a hexamer enzyme. Its structure was determined by us in three different forms: unligated and inhibitor-bound states in crystals of hexameric symmetry and in an orthorhombic form. Binding of the peptide inhibitor or crystallization in a less symmetrical environment resulted only in minor structure modifications – supporting our notion that the hexameric structure restricts internal movements of the monomers constituting it, and that pre-formed substrate binding site must exists within each. 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. Once inside, the active sites can be reached through the spacious side opening of the monomers created by the shortening and upward shift of the second of the seven blades of the propeller domains. The Z-Gly-Gly-Phe-CMK (chloromethyl-ketone) inhibitor binds to the active site by forming two covalent bonds with it, ligating to both the active Ser and His sidechains. The obtained structures were deposited in the Protein Data Bank under the accession codes 4HXE, 4HXF and 4HXG.

We found that both the crystalline and functional state PhAAP is a novel, hexameric structure, previously unseen among serine proteases. In case of PhAAP, it had to be determined if the reason for hexamerization was the monomer unit of the enzyme being structurally unstable. Therefore molecular dynamics (MD) simulations were carried out on

two systems: the prolyl oligopeptidase (POP) monomer (sharing considerable structural homology with acylaminoacyl peptidases) which functions as a monomer, and a monomer of PhAAP that was shown to hexamerize. During the run-time no difference in the stability of the two molecules could be detected. This finding lent support to our hypothesis that the driving force of hexamerization concerns protecting an amylogenic β-strand that is fully exposed in PhAAP, but is covered by an unusually long N-terminal extension in POP.

Some interesting results concerning the conformation of the active sites were also derived from the simulations. Both catalytic triads (those of POP and PhAAP) were found in a "latent" conformation, where the active site residues are, though close to each-other, but not directly H-bonded. Such a "latent" conformation of the Ser-His-Asp catalytic triad was seen in the crystal structure of ligand-free tricorn interacting factor F1 too, where - in the absence of propeller domain - the hydrolase is only partially shielded from solvent. 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.

ApAAP, as opposed PhAAP described above, functions as a dimer. Previously determined crystal structures of ApAAP revealed a true novelty – the enzyme was proven to possess both an open and a closed conformation, and the crystal structures demonstrated that these two forms might exist simultaneously. The active site of the open form is readily accessible, but is disassembled (in-activated), while the closed form is in a catalytically active conformation, but its active site is buried within the enzyme interior. A crystal form where an "open" monomer and a "closed" monomer formed the dimers of the crystal lattice was determined in our laboratory. The results suggest that the resting state of the enzyme should be thought of as the mixture of the two forms, however it remained to be clarified whether closing and activation is induced by the binding of a ligand or closing is independent of ligand binding but is a means of substrate selection: those ligands that are not small enough for the enzyme to re-close upon, do not encounter a potentiated active site, and thus will not be cleaved. Therefore we attempted to use the crystallization conditions resulting in the mixed open/closed dimer structure of the unligated state to crystallize the enzyme-inhibitor complex to see how the open form interacts with the ligand molecule. These trials however all lead to crystals of closed/closed dimers (ascertained as such based on crystal morphology). On the other hand, soaking pre-formed crystals in the solution containing the inhibitor was successful, and thus, open/closed mixed dimer form the substrate complexed state was determined. Data was collected at the synchrotron radiation source ESRF at Grenoble, the obtained structure was deposited in the Protein Data Bank under the accession code 4RE6.

We found that the substrate was unable to bind to the open form, but was found ligated to the active site of each closed monomer – thus demonstrating that specific binding only takes place within the closed form – the open form is indeed catalytically disabled. Since crystal packing would not allow an open→closed transformation in the crystalline state, we concluded that an entrance spacious enough for a 3-4 aminoacid peptide must permanently exist on the surface on the closed form and identified the so-called "propeller channel" as the site of entrance. A further crystal structure of the inhibitor-bound form of ApAAP has also been determined. Co-crystallization of ApAAP with the inhibitor lead to the formation of, yet

unseen, closed/closed ligand-bound dimers, even if 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 (deposited in the Protein Data Bank under the accession code 4RE5). The overall binding mode of the ligand and the accommodation of the P1 Phe residue is similar to that seen in case of the open/closed dimer, but a 1.22Å rms of inhibitor atoms within the overlaid protein matrices indicates subtle differences. The protein matrix of the closed monomer itself is remarkably similar in both cases with a backbone rms of 0.34 Å, therefore it is not the alteration of the binding site that causes the deviation – it is an apparent conformational freedom in the binding process. Such promiscuity of ligand-binding has not yet been reported concerning oligopeptidases.

In case of ApAAP, the question we addressed using molecular modeling tools was whether all substrates of the enzyme would be able to use the propeller-channel as an entrance and exit site. The propeller domain of acylaminoacyl peptidases covers the hydrolase domain containing the active site residues, limiting access to it. However, the propeller domain is pierced by a narrow passage that is perpendicular to the plane of the active site and is positioned right above it, such that a molecule small enough to pass through it would arrive directly at the site of catalysis. The 3-residue long covalently bound inhibitor (Z-Gly-Gly-Phe-CMK) in the crystal of the mixed open/closed dimer determined by us, must have passed through this channel. To test the general applicability of such an entrance, a 7 and a 9 amino-acid long peptide substrate was docked into a monomer unit of the enzyme using the MCMM (monte Carlo Multiple Minimum) methodology. The models thus derived were subjected to MD simulations, relaxing the structures in presence of explicit solvent molecules and ions mimicking the physiological conditions. Principal component analysis and clustering of the trajectories resulted in the identification of dominant protein movements and rearrangements attainable during the equilibrium. For the first time in this enzyme class, we were able to identify the domain-shift movement of the opening, where the two rigid domains undergo parting in a clamshell-like fashion. Promiscuity of binding was also further supported by the results, demonstrating, for example, how both a hydrophobic (Pro) and a polar (Arg) residue can be accommodated in the P4’ position.

As one of the major goals of the project, we planned to determine the structure of mammalian AAP too, an enzyme in the focal point of interest both as a pharmaceutical target and as structural prototype of substrate differentiation though small-scale self-assembly. During the first year of the project, we have been able to obtain, for the first time, crystals that withstand radiation damage – as tested on our in-house diffractometer. In cooperation with the Biostruct Laboratory of BUTE we have tested 3 x 96 crystallization conditions (the PACT-Premier, JSCG-Plus and MIDAS screens were used) on pAAP with and without inhibitor using the Mosquito crystallization robot of Biostuct Lab. After optimization, thin plate crystals were grown at our laboratory from a solution of the enzyme in complex with its specific inhibitor (N-Ac-Ala-CMK), which was kindly synthesized by Zoltán Bánóczi of the Organic Chemistry Department of ELTE. We were able to grow thin needle crystals from the uncomplexed state too. Both were heavy-atom derivatized, and the unligated crystals were also soaked in a solution of a widely-used carbaprem-type antibiotic, which was reported to

have inhibitory effect on the enzyme. Due to both states producing relatively small crystals (0.2 x 0.2 x 0.005mm – thin plates, 0.2 x 0.1 x 0.1mm – thin needles), we decided that their measurement should be carried out at a microfocused synchrotron beam and were accordingly measured at ESRF, Grenoble. However, only the inhibitor-bound crystals diffracted appreciably - albeit anisotropically - to a resolution of 3.5 Å in two directions, but only to ~8 Å in the third, very similarly to our test measurements at our in-house diffractometer. Since we were unable to determine the diffraction pattern of the heavy-atom derivatives of the crystal, we had to build a homology model of the enzyme to attempt solving the structure by a molecular replacement method.

Homology modeling lead to unexpected results. The hydrolase domain could be assigned to a sequence segment with certainty (based on a 29% identity and 46% homology with that of ApAAP), and was verified by the unconstrained – voluntary- formation of the catalytic triad, made up of three sequentially distant but through-space-ligated amino acids, when the model of this domain was built. Anchoring the sequence segment of the hydrolase domain showed that in pAAP it is only preceded by a single α-helix as in ApAAP– which probably leaves its amylogenic, sticky β-edge uncovered – in accordance with pAAP functioning not in a monomeric but in a tetrameric form. The propeller domain – on the other hand – showed no appreciable homology with any other propeller, but based on its secondary structure prediction, could be best built as an eight-bladed propeller. This was a surprising finding, since both ApAAP and PhAAP contain seven-bladed propeller domains.

Unfortunately, probably due to the low resolution of our data set, no conclusive result was reached in our molecular replacement attempts with any of the homology models, however using the eight-bladed form did produce better results than the others (seven-bladed propellers with long insertions) and possible modes of tetramerization could be derived - but not with enough certainty to accept any solution. Thus, molecular replacement was unsuccessful, so obtaining new crystals, possibly in a more stable crystal form became necessary.

However, collaboration between our research partner – the Polgar group of the Institute of Enzymology of the HAS – has meanwhile ceased due to the reorganization of that Institute, so we had to learn and adept to our laboratory the purification process of the mammalian AAP enzyme. Numerous attempts have been carried out in the past to use a recombinant technique for the expression of AAP both by the Polgar group and others, but without success. Thus, the single methodology currently available is purification of the enzyme from porcine liver, which was carried out successfully by us. Crystallization of the freshly purified samples resulted in crystals of two distinct crystal forms – we were able to reproduce the thin plates obtained in earlier attempts, but were able to derive crystals belonging to a new morphology too. Several different forms of this new variant could be derived: inhibitor and antibiotic-bound, and heavy-atom (Pt^{2+}, $[Ta_6Br_{12}]^{2+}$) derivatized forms too. We carried out their measurements at the DESY synchrotron of Hamburg. We have obtained datasets with resolution in the 2.6 - 3.0 Å range, which - we hope- will allow for direct structure determination (instead of molecular replacement methods). Data processing and structure solution is currently under way.

Considering the results of this project along with the previously determined structures, we conclude that the modular build-up of oligopeptidases can lead to three fundamentally different substrate admission routes in this enzyme-family: 1.) allowing access of substrates through the propeller hole of the closed form, 2.) dynamic domain flapping between the closed and open conformation, or 3.) a channel system created by multimerization. The present results provided the first verification that the open form of flexible monomers belonging to type 2.) is catalytically disabled (the dimer structure of ApAAP), and showed that substrate selection in case of rigid monomers of type 3.), which contain a permanent entrance to their active site, is carried out by their intricate, multi-gated inner structure created by oligomerization (as in the hexameric structure of PhAAP).

This set of structures provides a unique framework for the study of the modes and significance of oligomerization in biological systems – and how these processes are able to significantly alter the physiological function of enzymes. In case of AAP, we found that the presence of a sticky β-edge, loop-insertions or terminal elongations and their flexibility attenuate the final oligomerization state, whereby also determine the substrate selection mechanism of the enzyme.

We also participated in collaborations that studied the same phenomenon in case of two different proteins, those of podocin and a Trp-cage miniprotein. In case of the former, based on the results of MD simulations, we were able to propose a dimerization model of the WT protein which could be disrupted by certain single amino acid mutations, while withstanding others – thus providing an atomic level explanation for the pathogenicity of some genetically inherited mutations, intricately linked to the dimerization capabilities of the resultant proteins. In case of the miniprotein, oligomerization was found to be the on-path intermediate of aggregation, thus phosphorylation of a Ser in the WT sequence which promoted oligomer formation (as seen in our MD simulations too) resulted in the loss of the physiological form and function. Taken together, our results indicate that oligomerization can effectively function as a control mechanism, easily disrupted or enhanced, that can switch between the active and inactive form of proteins and enzymes.

In case of AAP, we could verify that in case of flexible monomers (like those of ApAAP) substrate binding and substrate cleaving steps are carried out by different forms of the enzyme - while substrate scavenging may take place in the easily-accessed cavity of the open form, the catalytic bond-breaking reaction is the exclusive task of the closed form. Rigid monomers, possessing a permanent side entrance (like PhAAP), multimerize to simultaneously maintain selectivity and effectiveness. In spite of our continued effort, we have not been able to determine the structure of mammalian AAP yet, but we have been successful in obtaining data-sets of sufficient resolution to now attempt structure solution. With a tetrameric structure non-homologous to any previously determined oligopeptidases, this is still a challenging task – one that we are determined to see through.