Supplementary Information for

Cryo-EM structure of acylpeptide hydrolase reveals substrate selection by multimerization and a multi-state serine-protease triad.

Anna J. Kiss-Szemán, Pál Stráner, Imre Jákli, Naoki Hosogi, Veronika Harmat, Dóra K. Menyhárd* and András Perczel*

Corresponding authors: Dora K. Menyhard, Andras Perczel
*e-mail: dora.k.menyhard@ttk.elte.hu ; perczel.andras@ttk.elte.hu

This file includes:

Detailed methods
Tables S1 – S5
Figures S1 – S15
References for Supplementary

Detailed Methods

Cryo-EM data processing

Movies were subjected to beam-induced motion correction using RELION-3.0 [1] and contrast transfer function parameters were estimated by CTFFIND4 [2]. All of the following processes were performed using RELION-3.0. Particles were auto-picked and two rounds of reference-free two-dimensional (2D) classification was performed. In total, 345,001 particles from 1157 micrographs were auto-picked and subjected to 3D initial model building and 3D classification. 3D refinement was performed applying C2 and D2 symmetry. 50,604 particles were used for final reconstruction. The 3D map and particle set yielded from the D2 symmetry refinement were subjected to per-particle defocus refinement, beam-tilt refinement, Bayesian polishing and 3D refinement. Final 3D refinement and postprocessing yielded a map with an overall resolution of 3.27 Å, estimated by the gold-standard FSC = 0.143 criterion.

Reprocessing out data with cryoSPARC [3] yielded a map with no symmetry applied (C1) with overall resolution of 4.12 Å. 3D variability analysis [4] of this map was carried out to obtain a more thorough picture of the fluctuation of the flexible segments of our final model.

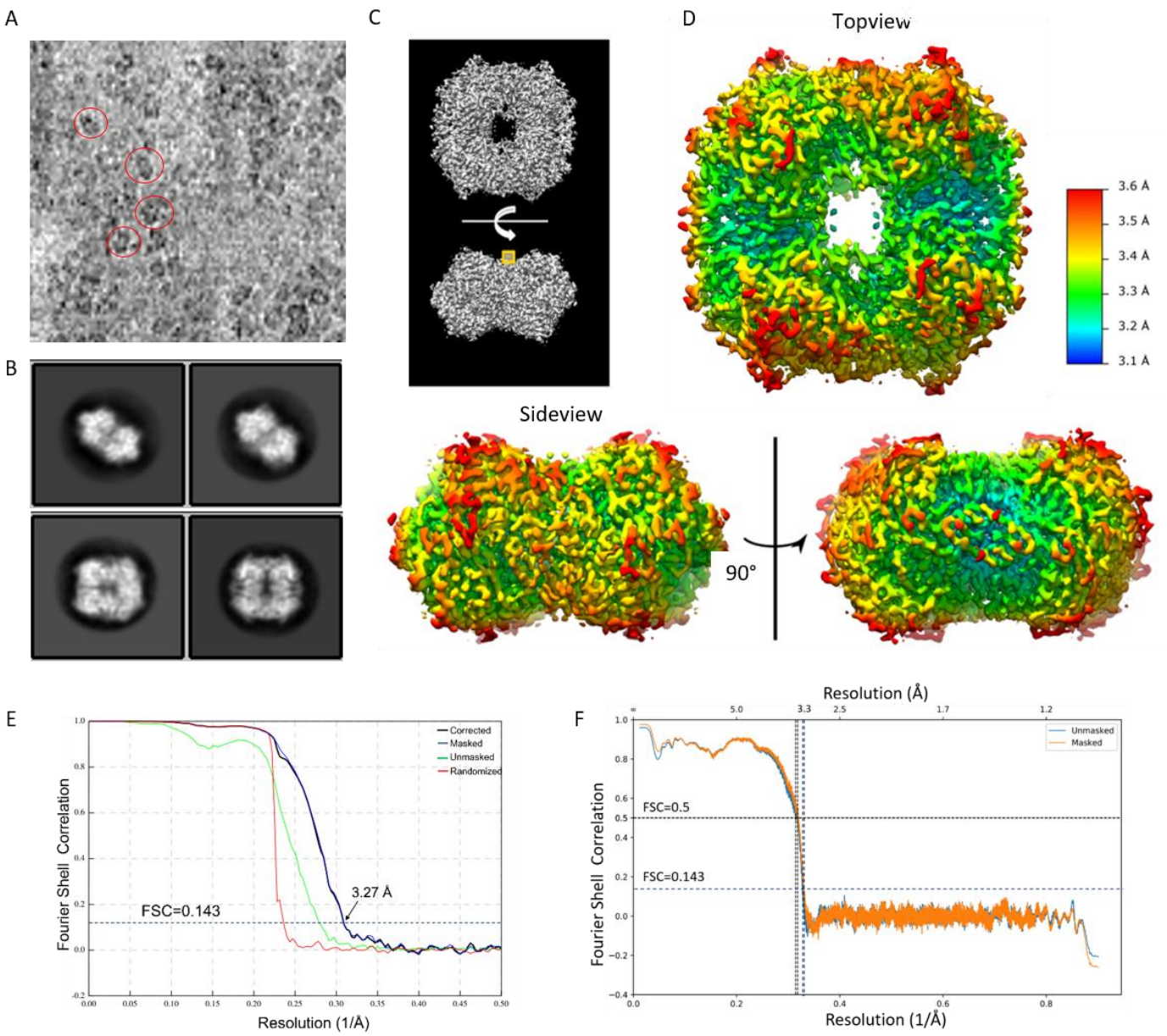


Figure S1. Overview of single-particle cryo-EM for AAP. A) Example cryo-EM micrograph for AAP with example particles circled in red. B) Reference-free 2D class averages of different particle orientations. C) C1 reconstruction of cryoEM map: 3D reconstruction of AAP specimen with no symmetry applied (resolution: 4.12Å). D) Local resolution of the D2 symmetry-averaged map (of overall 3.27Å resolution) contoured at a 0.010 threshold level (figure was made using UCSF Chimera [5]) calculated using ResMap [6]. The highest resolution region is observed in the hydrolase core. This map was used for structure determination. E) Map to map and F) model to map FSC curves calculated, the resolution of the map corresponding to FSC=0.143 is indicated.

Model building and refinement

The processed map was inverted by EMAN2 [7]. Automatic model building was carried out by ARP/wARP [8] and REFMAC5 [9] and manual finishing the monomer model in Coot [10] using the sequence and D2 point group symmetry of the tetramer was used. The tetramer was created by using Phenix Dock in Map [11], manually finished in Coot [10] and refined with real space refinement in Phenix [12].

Structure characterization

Analysis of interaction surfaces wax carried out with the PISA server [13]. Amyloidogenic properties were characterized using predictor servers: Aggrescan3D [14], Waltz [15], MetAmyl [16], RFAmyloids [17], BAP [18] and Zipper DB [19]. Sequence alignments were carried out in Uniprot [20].

Molecular dynamic simulations

Starting model for the molecular dynamics (MD) simulation was built from the cryo-EM determined structure. Missing residues were built manually and the longest of these new loops (residues 183-198) was subjected to 500 step MCMM-LMOD conformational search [21] as implemented in the Schrödinger Modeling Suite [22], keeping the rest of the model frozen. MD simulation was started from low energy conformers (for the entire tetrameric system, including the missing loop in identical conformation to all four chains) derived from the search, selecting the one that resulted in a structure most similar to the experimental result. The simulation was carried out suing GROMACS [23], applying the AMBER-ff99SBildnp* forcefield [24] with the OPC water model [25]. The total charge of the system was neutralized and physiological salt concentration was set using Na^+ and Cl^- ions. Energy minimization of the starting structures was followed by relaxation of constraints on protein atoms in three steps, with an additional NVT step (all of 200 ps) to stabilize pressure. Trajectories of 500 ns NPT simulations at 310 K and 1 bar were recorded for analysis (collecting snapshots at every 4 ps). Snapshots from the last 300 ns of the most optimal simulation were clustered based on the conformation of the backbone of the protein components using a 1 Å cutoff both for the tetrameric system and the monomers alone (using all four chains of every snapshot independently). For analyzing protein-protein contacts within the tetramer with PISA [9] a hybrid model was used: the cryo-EM structure attenuated with the missing residues in their most frequently sampled conformations (carrying out separate clustering for the missing residues of the 1st helix (residues 1-18), the inner gate (105-123) and the outer gating region (residues 171-201)). For determining the heterogeneity of the active site, clustering was carried out based on the conformation of residues 582-294, 673-677, 705-709. Figures showing conformational ensembles were composed using the mid-structures of the most populated clusters that account for over 95% of the snapshots. Principle component analysis was carried out based on main chain conformations along the full sequence using GROMACS. Hydrogen bonds were considered to be present when donor-acceptor distance did not exceed 3.2 Å.

Table S1. Sequence identity and similarity (in parentheses) among S9 family members as compared to human AAP. AAPs with available crystal structures were included and selected representatives of the S9 protease family. Comparison was calculated along the full sequence of human AAP (residues 1-732), the propeller domain (residues 22-461) and the hydrolase domain (residues 462-732, excluding the highly varying *N*-terminal segment (residues1-21)).

Protein name	full sequence	propeller	hydrolase
Sus scrofa (liver) AAP (pAAP)	92% (98%)	87% (92%)	91% (94%)
Pyrococcus horikoshii AAP (*Ph*AAP)	15% (31%)	14% (28%)	21% (40%)
Aeropyrum pernix AAP (*Ap*AAP)	18% (33%)	13% (27%)	26% (43%)
Sporosarcina psychrophila AAP (*Sp*AAP)	13% (30%)	10% (24%)	20% (40%)
Streptomyces morookaense AAP (*Sm*AAP)	16% (26%)	13% (21%)	23% (34%)
Deinococcus radiodurans R1 S9-carboxypeptidase (*Dr*CP)	20% (33%)	16% (26%)	26% (45%)
human POP	12% (29%)	11% (27%)	18% (35%)
human DPP4	11% (27%)	10% (24%)	14% (34%)
Leishmania major OPB	13% (30%)	13% (28%)	14% (31%)

Table S2. Length and location of the blades of the propeller domains of the S9 serine proteases.

Protein		Propeller domain	Blade1	Blade2	Blade3	Blade4	Blade5	Blade6	Blade7	Blade8
AAP *(porcine)* PDB id: 7px8	Residue numbers	23-457	23-88	89-134	135-238	239-298	299-345	346-405	406-457	-
	No. of amino acids	435	66	46	104	60	47	60	52	-
POP *(human)* PDB id: 3ddu	Residue numbers	72-426	72-114	115-168	169-227	228-284	285-330	331-377	378-426	-
	No. of amino acids	355	43	54	59	57	46	47	49	-
OPB *(L. major)* PBD id: 2xe4	Residue numbers	95-448	95-152	153-201	202-250	251-299	300-347	348-399	400-448	-
	No. of amino acids	354	58	49	49	49	48	52	49	-
DPP4 *(human)* PDB id: 1j2e	Residue numbers	56-496	56-90	91-150	151-191	192-287	288-350	351-401	402-453	454-496
	No. of amino acids	441	35	60	41	96	63	51	52	43
PTP *(P. gingivalis)* PDB id: 2eep	Residue numbers	50-471	50-107	108-145	146-192	193-282	283-331	332-379	380-427	428-471
	No. of amino acids	422	58	38	47	90	49	48	48	44
AAP *(A. pernix)* PDB id: 3o4g	Residue numbers	23-320	23-62	63-109	110-149	150-191	192-232	233-281	282-320	-
	No. of amino acids	298	40	47	40	42	41	49	39	-
AAP *(P. horikoshii)* PDB id: 4hxe	Residue numbers	15-338	15-59	60-103	104-171	172-216	217-258	259-299	300-338	-
	No. of amino acids	324	45	44	68	45	42	41	39	-
AAP *(S. psychrophyla)* PDB id: 5l8s	Residue numbers	16-331	16-59	60-106	107-152	153-199	200-244	245-287	288-331	-
	No. of amino acids	316	44	47	46	47	45	43	44	-
IsoP *(S. alaskensis)* PDB id: 5jrk	Residue numbers	42-413	42-89	90-131	132-224	225-277	278-320	321-367	368-413	
	No. of amino acids	372	48	42	93	53	43	47	46	-
AAP (*S. morookaensis*, PDB id: 3azo	Residue numbers	25-380	25-66	67-120	121-185	186-239	240-283	284-333	334-380	
	No. of amino acids	356	42	54	65	54	44	50	47	-
DrCP *(D. radiodurans)* PDB id: 5yzn	Residue numbers	17-389	17-75	76-122	123-195	196-249	250-292	293-351	352-389	
	No. of amino acids	373	59	47	73	54	43	59	38	-

Table S3. **Residues participating in the formation of the substrate specificity pocket S1 of pAAP in comparison with other S9 oligopeptidases**. (Based on the presently determined structure and PDB entries: 4hxe, 1ve6, 5yzn, 5l8s, 3azo, 5yzn, 3ddu, 1j2e, 2xe4).

hAAP*	Pro510	His588	Val613	Trp628	Glu632	Val678
pAAP (PDB id: 7px8)	*cis*Pro510	His588	Val613	Trp628	Glu632	Val678
*Ph*AAP (PDB id: 4hxe)	*cis*Pro388	Tyr467	Ile491	Phe507	Val511	Cys549
*Ap*AAP (PDB id: 1ve6)	*cis*Pro370	Tyr446	Val471	Phe488	Leu492	Thr527
*Sp*AAP (PDB id: 5l8s)	*cis*Pro381	Tyr459	Pro484	Trp505	-	Val543
*Sm*AAP (PDB id: 3azo)	*cis*Pro434	Ala512	Val536	Tyr554	Leu588	Cys596
*Dr*CP (PDB id: 5yzn)	*cis*Pro436	Tyr515	Ile539	Phe555	Glu559	Cys600
hPOP PDB id: 3ddu)	Phe476	Asn555	Val580	Trp595	Tyr599	Val644
hDPP4 (PDB id: 1j2e)	Pro550	Tyr631	Val656	Tyr666	Tyr670	Val711
*Lm*OPB (PDB id: 2xe4)	Ser498	Ala578	Phe603	Glu621	Glu624	Val665

* estimated based on sequence alignment

Table S4. Predicted amyloidogenicity of the outermost β-strand of the central β-sheet of the hydrolase domain in proteins of the S9 serine protease family: amyloid propensity was evaluated from high to low probability and was visualized by coloring from red through yellow to green.

Enzyme name	**PDB id**	**residue number** *	**sequence**	**Aggrescan3D** ** (aa)	**WALTZ** ** (75< score)	**MetAmyl** ** (6aa)	**RFAmyloid** ** (Y/N)	**BAP** ** (6aa)	**ZipperDB** ** (6aa)	**oligomeric state***	**PISA CSS** ***
pAAP (*Sus scrofa*)	7px8	695-704	VPVRLLLYPK	++++	++	-	+	-+-+-	+	tetramer (edge-shielding)	1.000
POP (*human*)	3ddu	668-678	NPLLIHVDTK	+++++	+	+++++	+	++++-	+-	monomer (intramolecular edge-shielding)	0.000
OPB (*L. major*)	2xe4	685-695	NEILLNIDMES	+++++	+++	++++-	+	+++---	+++	monomer (intramolecular edge-shielding)	0.000
OPB (*T. brucei*)	4bp8	671-681	NEVLLKMDLES	+++++	++	-	+	-----	+++	monomer (intramolecular edge-shielding)/ dimer (non edge-shielding)	0.000
DPP4 (*human*)	1j2e	729-738	DFQAMWYTDE	++-	+++	+	+	-+---	+++	dimer (edge-shielding)	1.000
PTP	2eep	698-708	YPDYYVYPSH	++++-	++	+	+	---+-	----	dimer (edge-shielding)	0.561
AAP (*A. pernix*)	3o4g	545-553	TFEAHIIPD	+++++	++	+	+	---+	+-	dimer (edge-shielding)	0.925
AAP (*P.horikoshii*)	4hxe	566-575	KEVYIAIFKK	-++++-	+++	+++	+	+-+++	++++-	hexamer (edge-shielding)	0.892
AAP (*S.psychrophyla*)	5l8s	560-569	RDVEYLVLED	+	+	+	+	--+++-	+++++	dimer (non edge-shielding) /tetramer(could be edge-shielding-model)?	0.528
IsoP (*S. alaskensis*)	5jrk	673-682	VATQISYYPG	++++-+++-	+++	+++	+	--+--	+++	dimer (edge-shielding)	1.000
PMH/AAP (*S.morookaensis*)	3azo	613-622	VPHAYLSFEG	+-++	++	-	+	-----	++-	monomer (intramolecular edge-shielding)	0.011
DrCP (*D.radiodurans*)	5yzn	618-625	VPVRFVRFPE	++	+	*not available*[+]	*not available*[+]	+-+--	---	tetramer (edge-shielding)	1.000

* based on PDB structure

** Aggrescan3D [14] predicts from whole protein sequence for individual residues (A3D score,color coded in the sequence in column 4, column 5. + for residues with A3D score > 0.0000 (aggregation-prone residue), + for residues with A3D score =0 and – for residues with A3D score <0.

**Waltz [15] predicts scores for aggregation propensities, number being higher than 75 indicates that submitted sequence is prone to amyloid-type aggregation. number being 75-80: +, 80-90: ++, 90<+++ and – for number being <75.

**MetAmyl [16] predicts amyloid hotspots from whole protein sequence. The score for the meta-prediction, p(x), is obtained using a logistic regression model, so that it can be interpreted as the probability for a fragment to form amyloid fiber. The score p(x) have been calculated for all possible hexapeptides. MetAmyl uses a sliding window to screen the whole submitted sequence. 0.6< + for each hexapeptide in the selected sequence (column4). + for probability 0.5-0.6 and – for probability <0.5.

**RFAmyloid Server [17] predicts yes: +/no: - for submitted entries.

**The Budapest Amyloid Predictor (BAP) Server [18] predicts amyloidogenicity for hexapeptides yes: +/no: - (for every hexapeptide in sliding sequence).

**ZipperDB [19] predicts fibrinogen propensity for hexapeptides from a protein sequence. + for high fibrillation propensity > -23kcal/mol, + from -23kcal/mol – (-6 kcal/mol) and – for no fibrillation propensity profile predicted for hexapeptides contained in the sequence (column 4).

***PISA [13] calculates interactions of molecular surfaces between submitted structures. The Complex Formation Significance Score for the interaction (CSS ranges from 0 to 1 as interface relevance to complex formation increases. Low CSS implies that the interface is not significant for complex formation and may be solely a result of crystal packing.

[+] Servers not available in December 2021.

Table S5. **Cryo-EM data collection, refinement and validation statistics**

Data collection	
Magnification	X80,000
Voltage (kV)	300
Electron exposure (e/Å^2)	40
Defocus range (µm)	1-3
Pixel size (Å)	0.95
Data processing	**EMD-13691, PDB 7px8**
Symmetry imposed	D2
Initial particle images (no.)	345,001
Final particle images (no.)	50,604
Map resolution (Å)	3.27
FSC threshold	0.143
Refinement	
Model resolution (Å)	3.27
Model composition	
Non-hydrogen atoms	21,409
Protein residues	2798
B-factors (min/max/mean)	
Protein	24.27/122.04/46.20
r.m.s. deviations	
Bond length (Å)	0.002
Bond angles (°)	0.529
Validation	
MolProbility score	1.18
Clashscore	3.45
Rotamer outliers (%)	0
CaBLAM outliers (%)	1.87
Ramachandran plot (%)	
Outliers	0
Allowed	69 (2%)
Favored	2731 (98%)

Figure S2. Topology of blades and inserts of the propeller domains of the S9 serine proteases compared to the presently determined structure. Numbers indicate the length of the segments creating turns, loops and β-sheets. Distorted or partially unstructured β-strands are labeled with an asterisk (*). Insertions are colored orange. Insertions of pAAP referred to as insert1 and insert2 are the 12 residues long insert is in blade 1; and the 54 residues long insert is in blade 3, respectively.

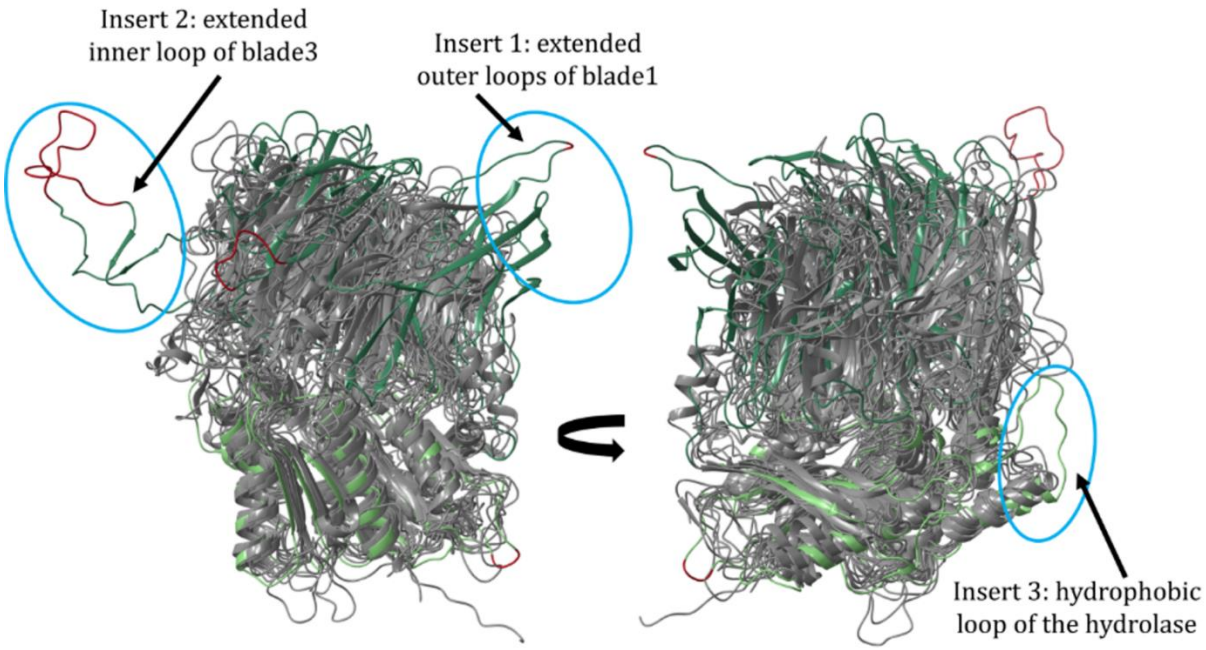


Figure S3. Comparing the structures of S9 oligopeptidases with that of pAAP. Structural alignment of members of the S9 family reveals that insertions to the pAAP sequence create significant protrusions from the surface allowing them to form extra contact surfaces. Monomeric units of pAAP (shown in green: hydrolase lighter, propeller darker), *Pyrococcus horikoshii* AAP (PDB id: 4hxe), *Aeropyrum pernix* AAP (PDB id: 3o4g), *Sporosarcina psychrophila* AAP (PDB id: 5l8s), *Streptomyces morookaensis* AAP (PDB id: 3azo), *Deinococcus radiodurans R1* S9-peptidase (PDB id: 5yzn), human POP (PDB id: 3ddu), human DPP4 (PDB id: 1j2e) and *Leishmania major* OPB (PDB id: 2xe4) are overlaid, fitting their hydrolase domains. For representing pAAP the cryo-EM structure was used, completing missing segments (residues 1-9, 39, 110-115, 183-198, 496-497) (shown in red) based on the results of the MD simulations. The N-terminal segment – a highly variable region that belongs to the hydrolase but sequentially precedes the propeller domain is not shown for any of the structures, for clarity.

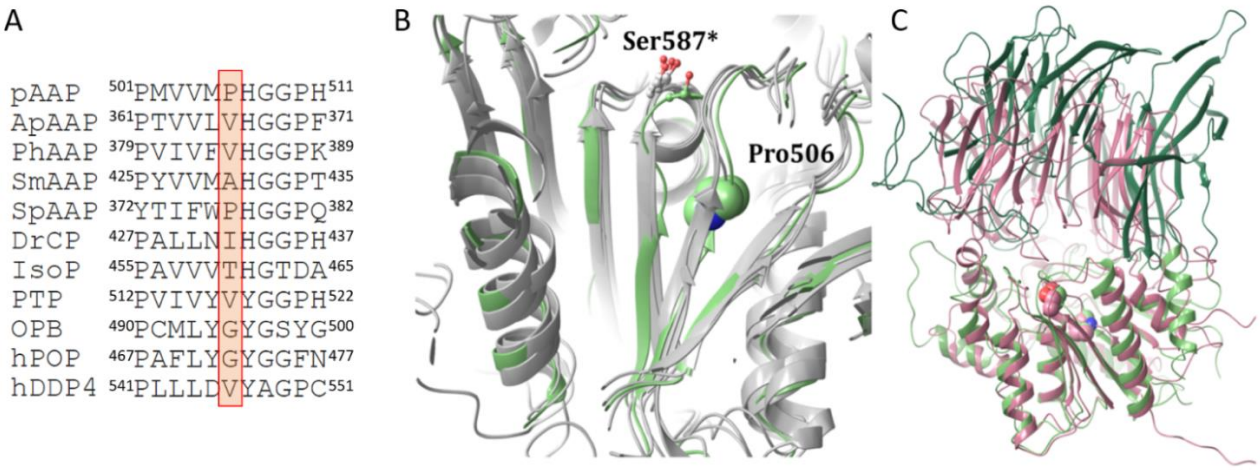
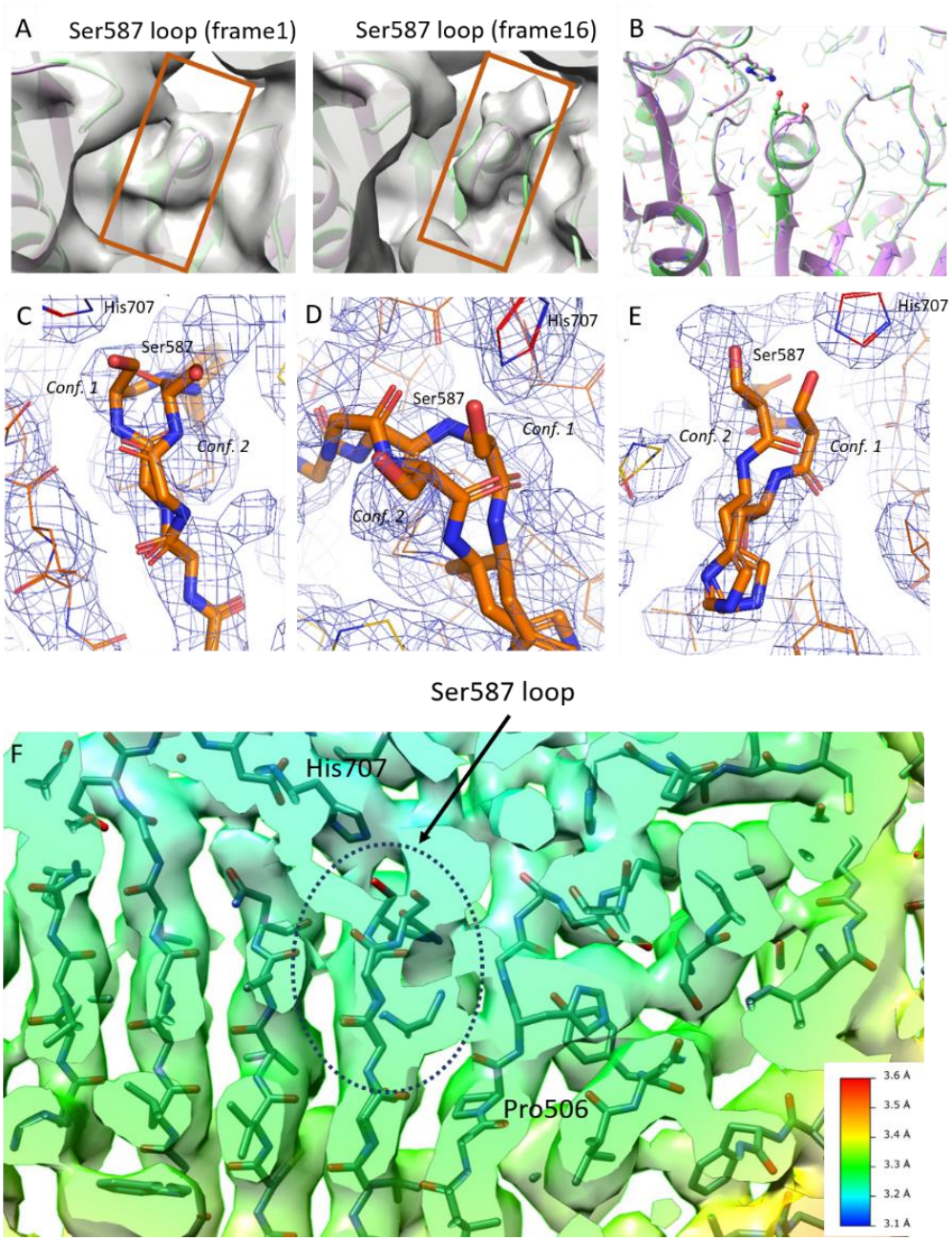


Figure S4. Destabilization of the catalytic Ser587. (*A*) Sequence alignment of the 4th strand of the hydrolase domain core β-sheet and the following loop (forming the oxyanion site) among the S9 protease family members (those with known structures) reveals that Pro506 of pAAP is rather unusual in this position (with the only one other such protein: *Sp*AAP). (*B*) Presence of Pro506 considerably shortens the 4th strand of the core of pAAP and also the adjacent 5th strand that leads up to the catalytic Ser587 (green: pAAP, grey: all other AAPs: *Ap*AAP, *Ph*AAP, *Sm*AAP, *Sp*AAP; PDB id: 3o4g, 4hxe, 3azo and 5l8s, respectively). (*C*) Comparison of pAAP (green) and *Sp*AAP (pink) shows that the closed and more compact fold of the *Sp*AAP monomer provides stabilization to the catalytic Ser, while in pAAP the loss of backbone H-bonds between the 4th and 5th strands allow destabilization of the loop holding the catalytic Ser because of the considerably more open structure, with reduced number of stabilizing propeller-hydrolase interactions.



Figures S5. Model building for the disordered Ser587-loop. A) 3D variability analysis (carried out using cryoSPARC [4] shows fluctuation of the C1 symmetry map near the active site (green and magenta backbone traces show the final models (active and latent state, respectively), the map is shown at a threshold of 0.015 (image created with Chimera)). B) Final model built in the D2 symmetry map with two conformers of the 584-591 segment is shown in the same orientation and coloring as of panel A. C -E) The "active" conformation (Conf.1) and the "latent" conformation (Conf.2.) of the Ser-loop in different orientations in the D2 symmetry map (contoured at treshold of 0.016. F) Local resolution (contoured at a 0.010 threshold level in Chimera, calculated by ResMap [6]) of the hydrolase domain core. Chain A was used to represent the averaged structures because of the D2 symmetry. (Panels C-E were created with PyMol [Schrödinger, L., & DeLano, W. (2020). *PyMOL*. Retrieved from http://www.pymol.org/pymol])

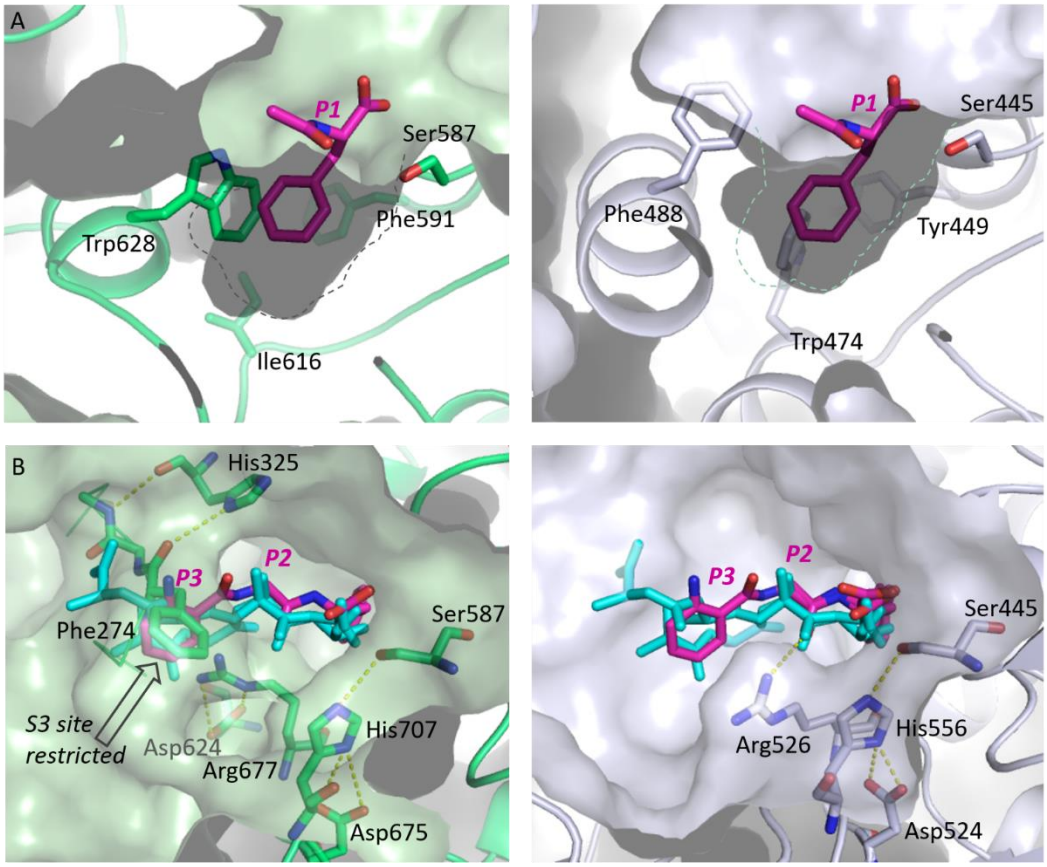


Figure S6. Substrate binding subsites of AAPs tailoring specificity. (*A*) Molecular surfaces of the S1 substrate specificity pocket of pAAP (left, green) is narrowed by Trp628 as compared to *Ap*AAP (right, grey; PDB id: 2hu7). Note: In oligopeptidase structures, beside pAAP (and the human variant) only POP and *Sp*AAP carry a Trp in this position. However, in POP, this Trp residue is in a different conformation forming a stacking interaction with P1 Pro residue by which it becomes the major determinant of specificity; in *Sp*AAP it is considerably shifted, widening the S1 pocket. (*B*) S2 and S3 subsites of pAAP (left, green) and *Ap*AAP (right, grey) are shown. The S3 subsite of pAAP is restricted by Phe274-Cys275 barrier, while this region of archaeal AAPs is open. Bound peptides overlaid are shown for different oligopeptidases: for *Ap*AAP in magenta (PDB id: 2hu7, 2hu8); backbone of the peptide fragment in cyan for OPB and POP (PDB ids: 4bp9; 5n4c, respectively).

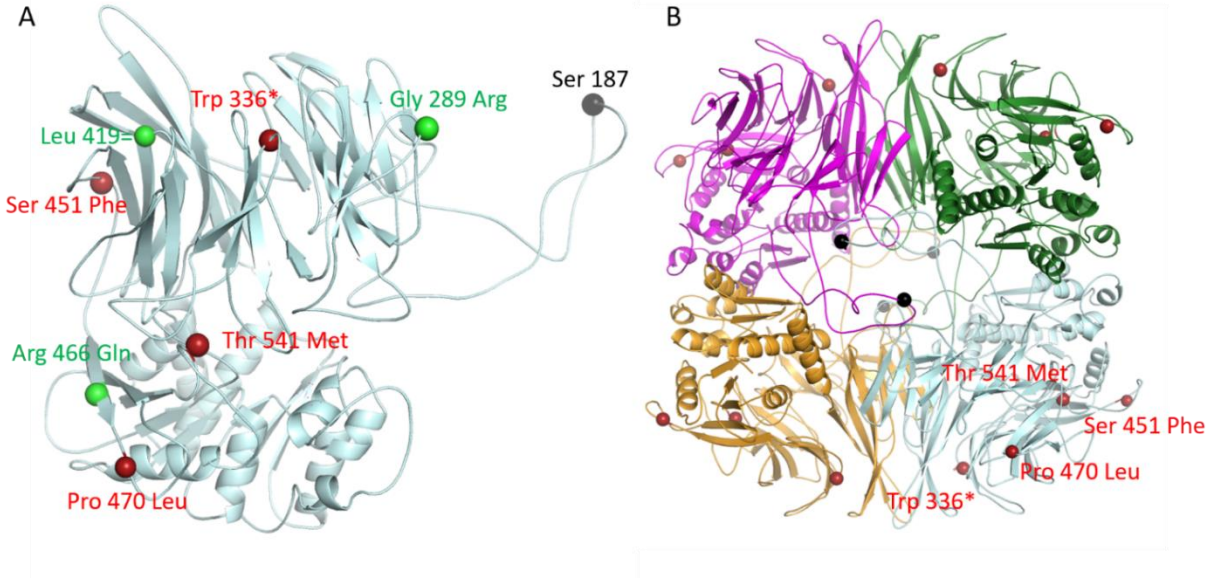


Figure S7. Reported natural mutations of human AAP. (*A*) The location of the mutations is shown on the monomer: mutations not altering function in green, mutations leading to loss of enzyme activity in red. (*B*) The latter are shown on the tetramer (red), together with the position of the predicted phosphorylation site Ser187 (black). Trp336* terminates the sequence in the middle of blade 5 so the loss of enzyme activity is due to the absence of the entire hydrolase domain. Thr541Met mutation could result in catalytic Ser-loop disorder (Fig. S6). Pro470Leu and Ser451Phe mutations are on the outer surface of the tetramer not blocking the entrance or causing any disorder near the catalytic triad - although they could play an essential role in effecting interactions with other proteins or could lead to aggregation or misfolding and thus cause disfunction of the enzyme.

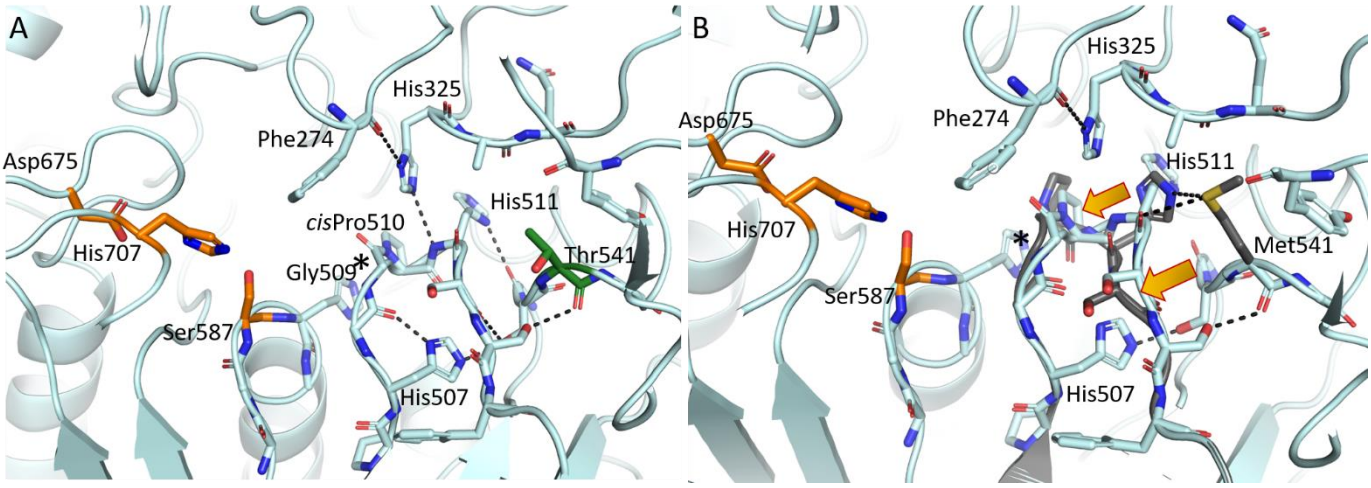


Figure S8. Possible structural effects of the Thr541Met mutation. (*A*) Hydrogen bonding network including Thr541 (green) and the catalytic triad (orange) is shown. There is a more hydrophilic and quite loose loop (residues 507-513) between the mutation site of Thr541 and the loop holding the catalytic Ser587. Methionine sulfur is known to form S…O interaction where the divalent sulfur serves as an electrophile, while it is also known to form H-bonds, especially with heterocyclic N-H moieties like the His sidechain [26]. (*B*) Possible changes caused by Thr541Met mutation (dark grey). The Met541...His511 side chain interaction could modify the 507-513 loop conformation. This may cause effectivity loss through repositioning the Gly509 residue (oxyanion binding site labeled with an asterisk), which will thus not be able to take part in stabilizing the transition state of the catalytic reaction.

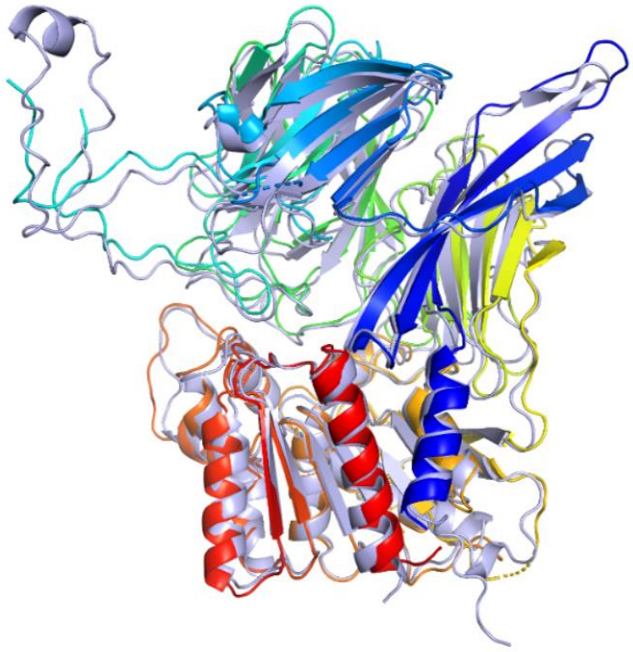


Figure S9. **The comparison of the cryo-EM structure of pAAP and the model generated by AlphaFold** [27]. The domain opening was slightly underestimated, the *cis* conformation of Pro510 was not predicted, neither was the liberation of the Ser-loop - or the mode of tetramerization. (Cryo-EM structure: rainbow colors; the structure found in AlphaFold database: light blue).

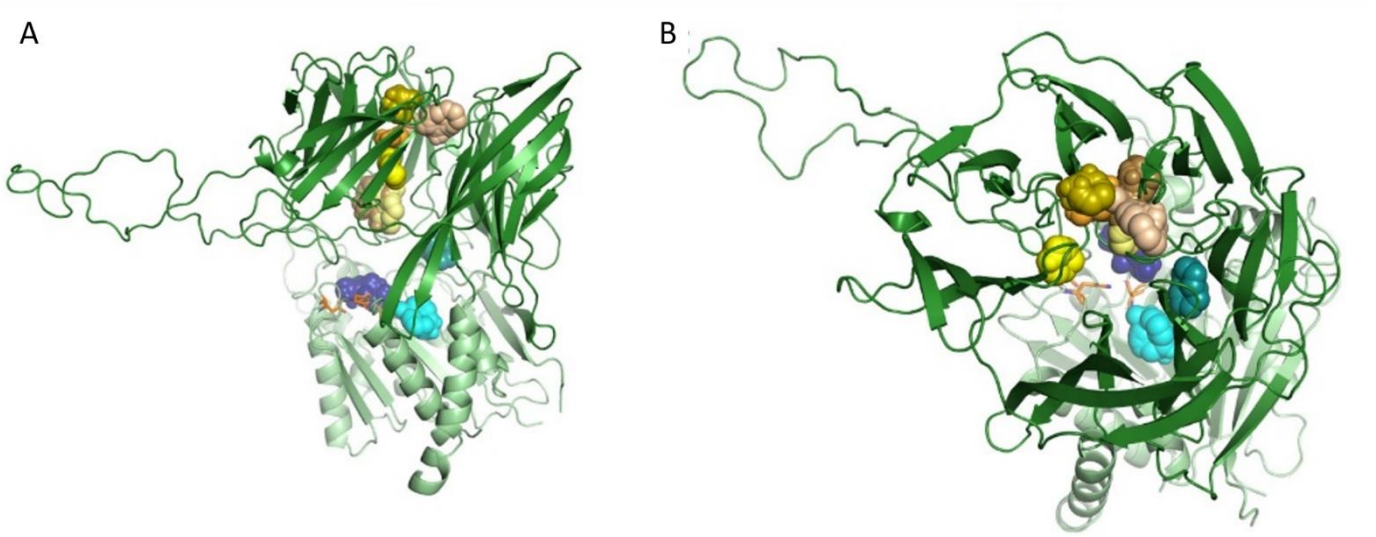


Figure S10. The propeller channel of the mammalian AAP may serve as a pathway towards the catalytic cavity. Ligand binding sites along the channel calculated by the FTMap server [28] (yellow and orange) and within the substrate specificity pockets (S1 blue, S2 cyan) are shown ((*A*) side view and (*B*) top view).

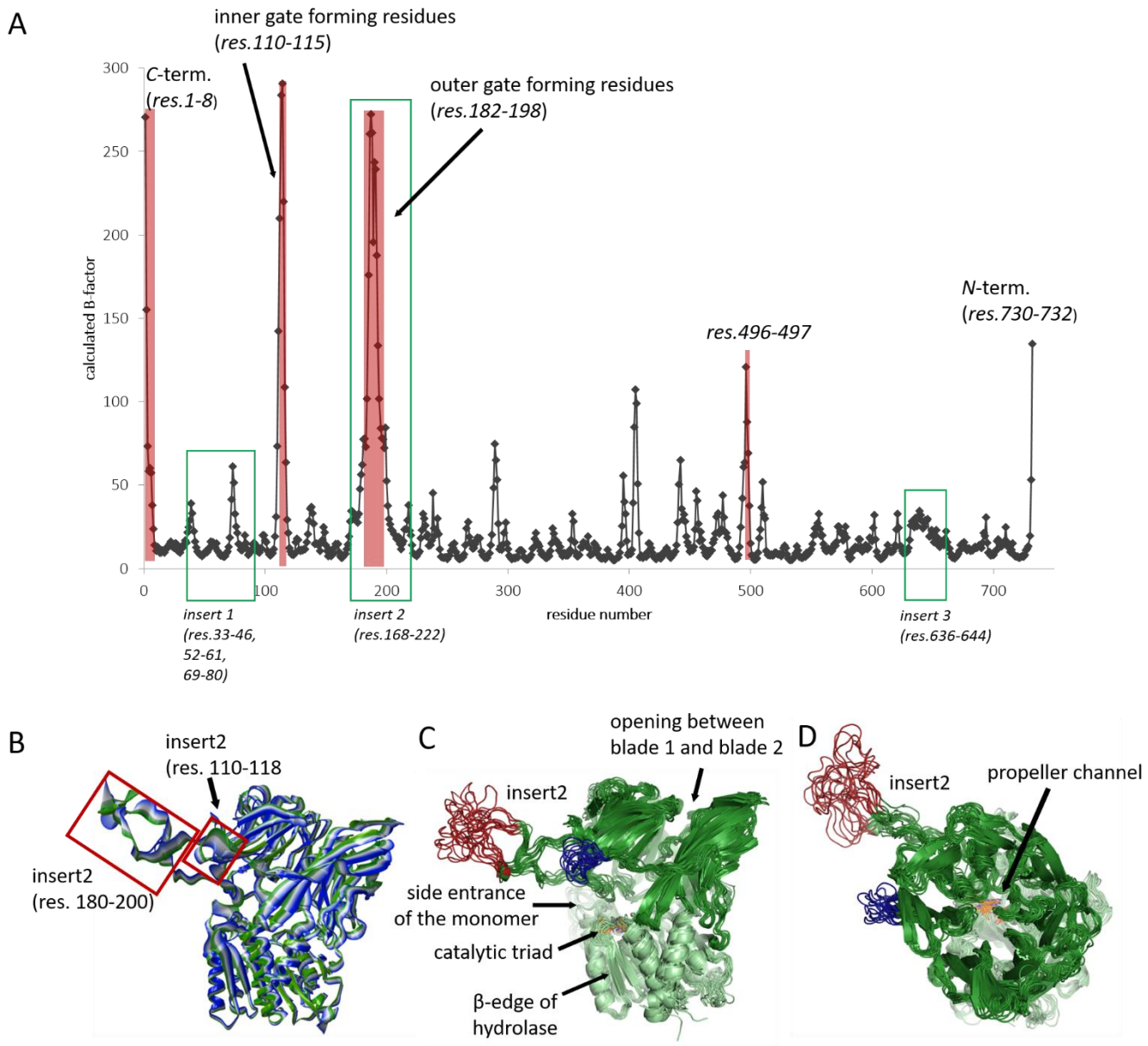


Figure S11. MD-derived B-factors of pAAP. (*A*) B factors were calculated based on the average deviation of main-chain atoms (averaged over all chains of the tetramer, for the last 300ns of the molecular dynamics simulation). Red rectangles show the regions that could not be resolved in the cryo-EM map. Green rectangles show the three unique inserts characteristic to mammalian AAPs. (*B*) Conformational changes associated with the first component of the PCA or the MD trajectory (carried out for all monomers of the tetrameric assembly). The figure shows transition between the two endpoints of the fluctuations (green corresponding to the closed structure, blue to the open form), with the most significant variations in the insert 2 (residues 180-200) and the 110-118 regions. (*C*, *D*) Cluster mid-structures are shown for the monomeric unit (side and top view, respectively; hydrolase domain: light green, propeller domain: dark green, catalytic triad: orange). Two segments were found especially highly mobile during MD simulations (and also missing in the cryo-EM map): insert2 (red) and residues 110-117 (blue), respectively.

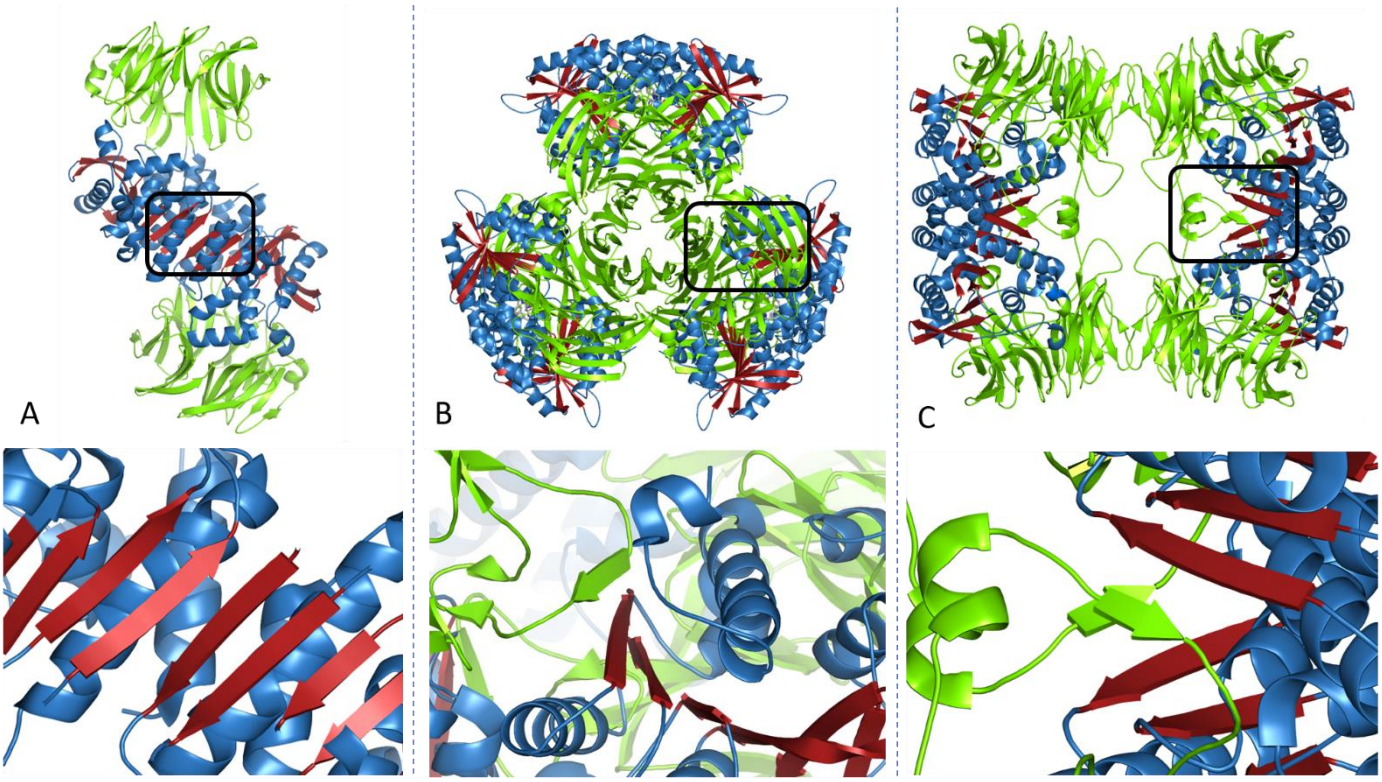


Figure S12. Structural elements shielding the sticky edge of the central β-sheet of the hydrolase domain in multimeric oligopeptidases. (*A*) In the dimeric structure of *Ap*AAP (PDB id: 3o4g) the two monomers (hydrolase domain in skyblue, propeller domain in green) create a 16-stranded large β-sheet (red). (*B*) In the *Ph*AAP (PDB id: 4hxg) hexamer and (*C*) in the *Dr*CP tetramer (PDB id: 5yzn) the β-edge is covered by a loop extension of the propeller of a neighboring monomer.

```
SP|P19205|ACPH_PIG   MERQVLLSEPEEAAALYRGLSRQPALSAACLGPEVTTQYGGRYRTVHTEWTQRDLERMEN 60
SP|P13798|ACPH_HUMAN MERQVLLSEPEEAAALYRGLSRQPALSAACLGPEVTTQYGGQYRTVHTEWTQRDLERMEN 60
                     ************************************************:*****************

SP|P19205|ACPH_PIG   IRFCRQYLVFHDGDSVVFAGPAGNSVETRGELLSRESPSGTMKAVLRKAGGTGTAEEKQF 120
SP|P13798|ACPH_HUMAN IRFCRQYLVFHDGDSVVFAGPAGNSVETRGELLSRESPSGTMKAVLRKAGGTGPGEEKQF 120
                     ***************************************************** .*****

SP|P19205|ACPH_PIG   LEVWEKNRKLKSFNLSALEKHGPVYEDDCFGCLSWSHSETHLLYVADKKRPKAESFFQTK 180
SP|P13798|ACPH_HUMAN LEVWEKNRKLKSFNLSALEKHGPVYEDDCFGCLSWSHSETHLLYVAEKKRPKAESFFQTK 180
                     ********************************************:*:*************

SP|P19205|ACPH_PIG   ALDVTGSDDEMARTKKPDQAIKGDQFLFYEDWGENMVSKSTPVLCVLDIESGNISVLEGV 240
SP|P13798|ACPH_HUMAN ALDVSASDDEIARLKKPDQAIKGDQFVFYEDWGENMVSKSIPVLCVLDVESGNISVLEGV 240
                     ****:.****:** ************:************* *******:***********

SP|P19205|ACPH_PIG   PESVSPGQAFWAPGDTGVVFVGWWHEPFRLGIRFCTNRRSALYYVDLTGGKCELLSDESV 300
SP|P13798|ACPH_HUMAN PENVSPGQAFWAPGDAGVVFVGWWHEPFRLGIRFCTNRRSALYYVDLIGGKCELLSDDSL 300
                     **.************:****************************** *********:*:

SP|P19205|ACPH_PIG   AVTSPRLSPDQCRIVYLRFPSLVPHQQCGQLCLYDWYTRVTSVVVDIVPRQLGEDFSGIY 360
SP|P13798|ACPH_HUMAN AVSSPRLSPDQCRIVYLQYPSLIPHHQCSQLCLYDWYTKVTSVVVDVVPRQLGENFSGIY 360
                     **:**************::***:**:**.*********:*******:*******:*****

SP|P19205|ACPH_PIG   CSLLPLGCWSADSQRVVFDSPQRSRQDLFAVDTQMGSVTSLTAGGSGGSWKLLTIDRDLM 420
SP|P13798|ACPH_HUMAN CSLLPLGCWSADSQRVVFDSAQRSRQDLFAVDTQVGTVTSLTAGGSGGSWKLLTIDQDLM 420
                     ******************* *************:*:*******************:***

SP|P19205|ACPH_PIG   VVQFSTPSVPPSLKVGFLPPAGKEQAVSWVSLEEAEPFPDISWSIRVLQPPPQQEHVQYA 480
SP|P13798|ACPH_HUMAN VAQFSTPSLPPTLKVGFLPSAGKEQSVLWVSLEEAEPIPDIHWGIRVLQPPPEQENVQYA 480
                     *.******:**:******* *****:* *********:*** *.********:**:****

SP|P19205|ACPH_PIG   GLDFEAILLQPSNSPEKTQVPMVVMPHGGPHSSFVTAWMLFPAMLCKMGFAVLLVNYRGS 540
SP|P13798|ACPH_HUMAN GLDFEAILLQPGSPPDKTQVPMVVMPHGGPHSSFVTAWMLFPAMLCKMGFAVLLVNYRGS 540
                     ***********.. *:*********************************************

                                                                  *
SP|P19205|ACPH_PIG   TGFGQDSILSLPGNVGHQDVKDVQFAVEQVLQEEHFDAGRVALMGGSHGGFLSCHLIGQY 600
SP|P13798|ACPH_HUMAN TGFGQDSILSLPGNVGHQDVKDVQFAVEQVLQEEHFDASHVALMGGSHGGFISCHLIGQY 600
                     ************************************.:***********:********

SP|P19205|ACPH_PIG   PETYSACVVRNPVINIASMMGSTDIPDWCMVEAGFSYSSDCLPDLSVWAAMLDKSPIKYA 660
SP|P13798|ACPH_HUMAN PETYRACVARNPVINIASMLGSTDIPDWCVVEAGFPFSSDCLPDLSVWAEMLDKSPIRYI 660
                     **** ***.**********:*********:***** :************ *******:*

                                 *                                *
SP|P19205|ACPH_PIG   PQVKTPLLLMLGQEDRRVPFKQGMEYYRVLKARNVPVRLLLYPKSTHALSEVEVESDSFM 720
SP|P13798|ACPH_HUMAN PQVKTPLLLMLGQEDRRVPFKQGMEYYRALKTRNVPVRLLLYPKSTHALSEVEVESDSFM 720
                     ****************************.**:****************************

SP|P19205|ACPH_PIG   NAVLWLCTHLGS 732
SP|P13798|ACPH_HUMAN NAVLWLRTHLGS 732
                     ****** *****
```

91.9% sequence identity
673 identical positions
45 similar positions
14 different positions

– Propeller domain
– Hydrolase domain
– Blade3 insertion
* Catalytic triad

Figure S13. Sequence alignment of porcine ((Uniprot [20] code P19205) **and human AAP** (P13798**)** with the positions of the domain border, as well as large insertion of propeller blade3 and the catalytic triad indicated**.**

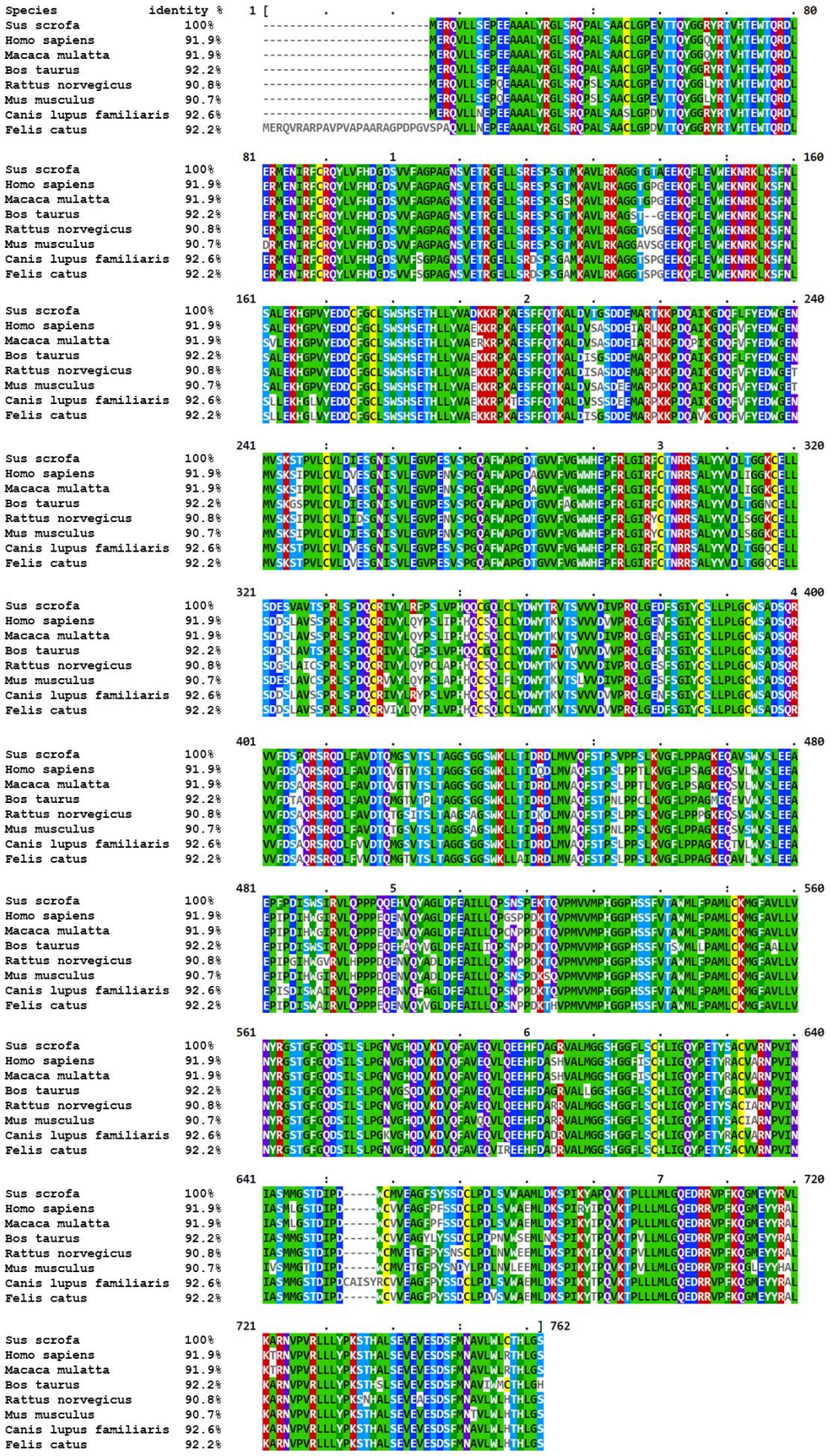
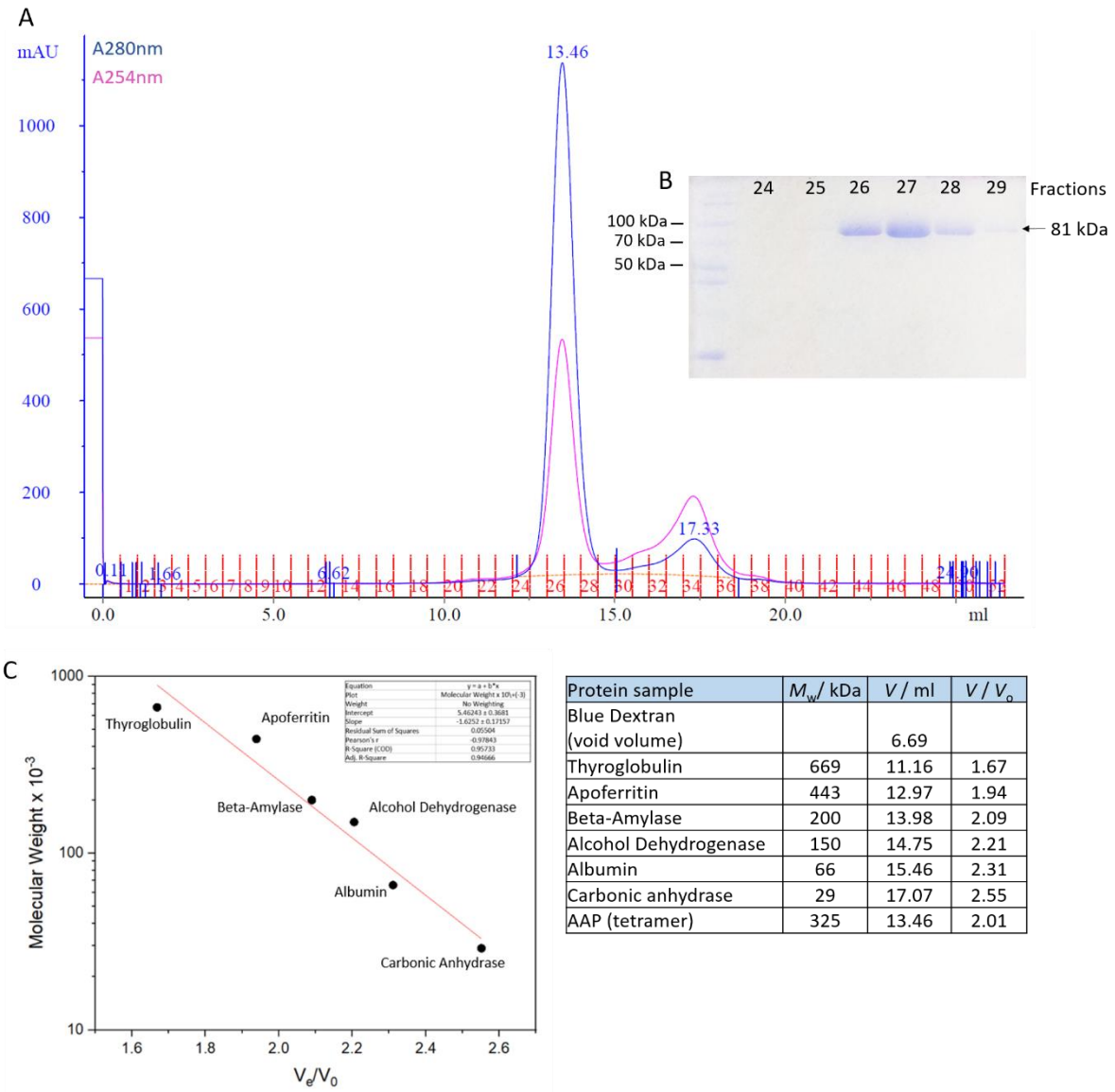


Figure S14. Sequence alignment of AAPs of mammals. Results of sequence alignment carried out with Clustal Omega [29] are presented with MView [30]. Sequence identity % with porcine AAP is shown (backgrounds of residues identical with that of porcine AAP are colored). Uniprot IDs are as follows: Sus scrofa: P19205, Homo sapiens: P13798, Macaca mulatta: F7G744, Bos taurus: P80227, Rattus norvegicus: P13676, Mus musculus: Q8R146, Canis lupus familiaris: E2R7E8, Felis catus: A0A5F5XJB2.



Protein sample	M_w / kDa	V / ml	V / V_o
Blue Dextran (void volume)		6.69	
Thyroglobulin	669	11.16	1.67
Apoferritin	443	12.97	1.94
Beta-Amylase	200	13.98	2.09
Alcohol Dehydrogenase	150	14.75	2.21
Albumin	66	15.46	2.31
Carbonic anhydrase	29	17.07	2.55
AAP (tetramer)	325	13.46	2.01

Figure S15. Characterization of the AAP sample. A) Size exclusion chromatography (Superose6 30/100 column, 20 mM TRIS, pH=8, 0.15M NaCl, 1mM EDTA, 1mM DTT) shows the homogeneity and the main peak retention is corresponding to the approximate mass of the AAP tetramer (325 kDa), B) SDS-PAGE of the main peak fractions is shown. SDS disassembles the tetramer and the bands are corresponding to the mass of the monomer (81 kDa). C) Calibration of Superose6 30/100 column (20 mM TRIS, pH=8, 0.15M NaCl, 1mM EDTA, 1mM DTT) with Gel Filtration Molecular Weight Markers Kit for Molecular Weights 29,000-700,000 Da (Merck, recommended concentrations were used). From elution volume (V_e) of our AAP sample (same sample volume and flow rate was used). The calibration peaks obtained in Superose6 30/100 column calibration: column3 shows the peak centroids (integrated with UNICORN™ software) and the estimated M_w of our protein sample based on the calibration curves. Using the standard semilog calibration curve the molecular mass of AAP specimen was calculated to be 341 kDa.

References cited in Supplementary

1. J. Zivanov et al., New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife.* **7**, e42166 (2018).
2. A. Rohou & N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol.* **192**(2), 216-21 (2015).
3. A. Punjani, *et al.,* cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods.* **14,** 290–296 (2017).
4. A. Punjani & D. Fleet, 3D Variability Analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol*. **213**(2), 107702 (2021).
5. E.F. Pettersen et al., UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* **25**(13), 1605-12 (2004).
6. A. Kucukelbir, F. Sigworth & H. Tagare, Quantifying the local resolution of cryo-EM density maps. *Nat Methods*. **11,** 63–65 (2014).
7. G. Tang et al., EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol.* **157**, 38-46 (2007).
8. G. Chojnowski, J. Pereira & V.S. Lamzin, Sequence assignment for low-resolution modelling of protein crystal structures. *Acta Crystallogr. D Struct. Biol*. **75**, 753-763 (2019).
9. G.N. Murshudov et al., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol Crystallogr*. **67**, 355-367 (2011).
10. P. Emsley & K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol Crystallogr*. **60**, 2126-2132 (2004).
11. D. Liebschner et al., Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**(10), 861-877 (2019).
12. P.V. Afonine et al., Real-space refinement in *PHENIX* for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol*. **74**, 531-544 (2018).
13. E. Krissinel & K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol*. **372**, 774—797 (2007). Protein interfaces, surfaces and assemblies' service PISA at the European Bioinformatics Institute. (http://www.ebi.ac.uk/pdbe/prot_int/pistart.html)
14. N.S. de Groot, V Castillo, R. Graña-Montes & S. Ventura, AGGRESCAN: method, application, and perspectives for drug design. *Methods Mol Biol*. **819**, 199-220 (2012).
15. S. Maurer-Stroh et al., Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods.* **7**, 237–242 (2010).
16. M. Emily, A. Talvas & C. Delamarche, MetAmyl: a METa-predictor for AMYLoid proteins. *PLoS One*. **8**(11), e79722 (2013).
17. Niu, M., Li, Y., Wang, C. & Han, K. RFAmyloid: A Web Server for Predicting Amyloid Proteins. *Int. J. Mol. Sci.* **19**(7), 2071 (2018).
18. L. Keresztes et al., The Budapest Amyloid Predictor and Its Applications. *Biomolecules*. **11**(4), 500 (2021).
19. L. Goldschmidt, P.K. Teng, Riek, R. & Eisenberg, D. The amylome, all proteins capable of forming amyloid-like fibrils. *Proc. Natl. Acad. Sci. U S A.* **107**(8), 3487-3492 (2010).
20. The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021**.** *Nucleic Acids Res*. 49(D1), 480–489 (2021).
21. I. Kolossváry & W.C. Guida, Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides *J. Am. Chem. Soc.* **118**, 5011–5019 (1996).
22. Schrödinger Release 2021-2: MacroModel, Schrödinger, LLC, New York, NY, 2021.
23. S. Pronk et al., GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **29**(7), 845–854 (2013).
24. A.E. Aliev et al., Motional Timescale Predictions by Molecular Dynamics Simulations: Case Study Using Proline and Hydroxyproline Sidechain Dynamics. *Proteins* **82**, 195–215 (2014).
25. S. Izadi, R. Anandakrishnan & A.V. Onufriev, Building Water Models: A Different Approach. *J. Phys. Chem. Lett.*, **5** (21), 3863–3871 (2014).
26. D. Pal & P. Chakrabarti, Non-hydrogen Bond Interactions Involving the Methionine Sulfur Atom. *J. Biomol. Struct. Dyn*. **19**(1), 115–128, (2001).
27. J. Jumper et al., Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
28. D. Kozakov et al., The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* **10**(5), 733-755 (2015).
29. F. Sievers et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol*. **7**, 539 (2011).
30. N.P. Brown, C. Leroy and Sander C. MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics* **14**(4), 380-1 (1998).