A new L-proline amide hydrolase with potential application within the Amidase Process

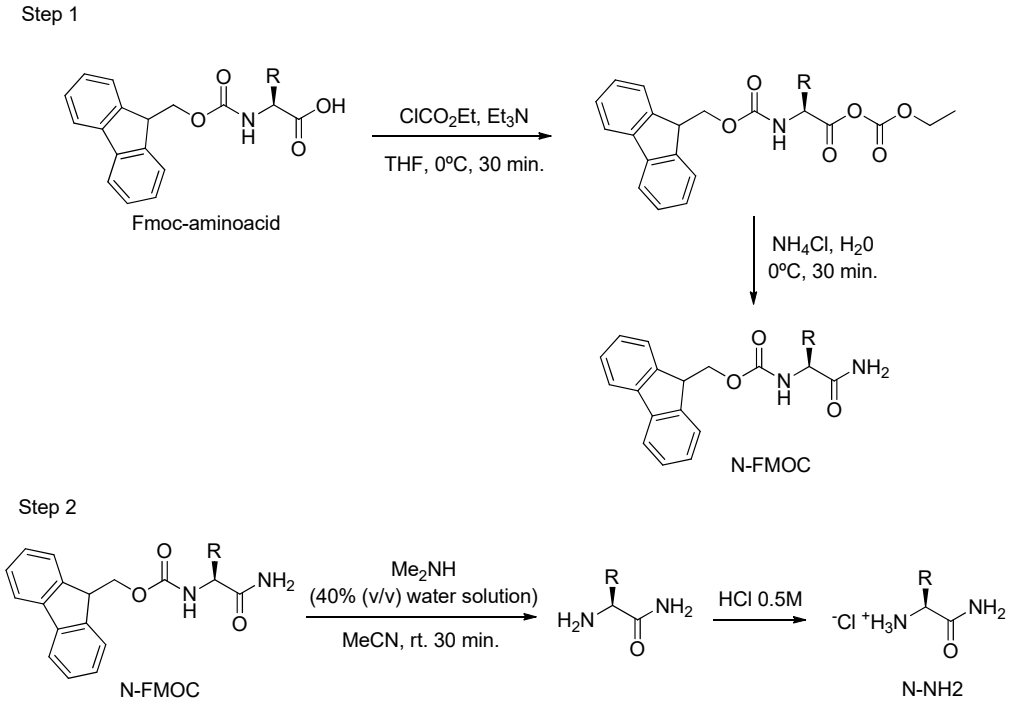
Sergio Martinez-Rodríguez[a,b], Rafael Contreras-Montoya[c], Jesús M. Torres de Pinedo[a], Luis Álvarez de Cienfuegos[c] and Jose Antonio Gavira[b]

[a]Departamento de Bioquímica y Biología Molecular III e Inmunología, Universidad de Granada, Granada, 18071, Spain

[b]Laboratorio de Estudios Cristalográficos, CSIC, Armilla, Granada, 18100, Spain

[c]Departamento de Química Orgánica, Universidad de Granada, Granada, 18071, Spain

Synthesis of different amino-acid amides



General procedure:

Step 1. To a colorless solution of 0.50 mmol of Fmoc-amino acid in 10 mL of THF were added 67 μL (0.70 mmol, 1.4 equiv) of $ClCO_2Et$ and 209 μL (1.5 mmol, 3.0 equiv) of Et_3N at 0 ºC. After stirring for 30 min at 0 ºC, 0.75 ml of a 1.0 M aqueous solution of NH_4Cl (0.75 mmol, 1.5 equiv) were added at 0 ºC to the colorless suspension. The mixture was stirred for 30 min at 0 ºC and 5 mL of H_2O were added. The colorless clear solution was extracted with 30 mL of EtOAc and the

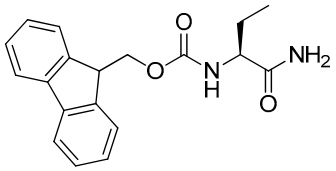
aqueous layer was extracted with 20 mL of EtOAc. The organic layers were combined, washed with 5 mL of brine, and dried over anhydrous Na_2SO_4. The product purified via flash column chromatography on silica gel with mixtures of hexane and EtOAc to afford the Fmoc-amino acid-NH_2 (N-FMOC).

Step 2. To a colourless solution of 0.50 mmol of Fmoc-amino acid-NH_2 in 100 mL of MeCN,10 mL of Me_2NH 40% (v/v) water solution were added and the mixture was stirred for 30 min at room temperature. The resulting solution was then vaccum evaporated to obtain a white solid. The crude product was extracted with 100 mL of H_2O and washed with three portions of 30 of diethyl ether. HCl 0.5 M was then added dropwise to the aqueous solution to reach a pH 2. The resulting acid solution was then freeze dried to obtain the final product.

Product characterization:

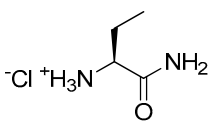
NMR spectra were recorded on 500 MHz spectrometer (Bruker); chemical shift (δ), referenced to the residual protonated solvent as an internal standard are quoted in ppm; coupling constants (*J*) are reported in Hz. High Resolution Mass Spectrometry was recorded using an Xevo G2-XS QTof (Waters, USA) mass spectrometer.

Fmoc-L-2-ABA amide:



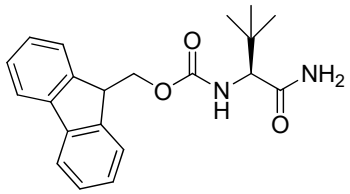
^{1}H NMR (500 MHz, Methanol-d_4) δ 7.81 (d, *J* = 7.5 Hz, 2H), 7.69 (t, *J* = 7.5 Hz, 2H), 7.41 (t, *J* = 7.5 Hz, 2H), 7.33 (t, *J* = 7.5 Hz, 2H), 4.46 – 4.37 (m, 2H), 4.24 (t, *J* = 6.9 Hz, 1H), 4.03 (dd, *J* = 8.6, 5.3 Hz, 1H), 1.89-1.79 (m, 1H), 1.72-1.61 (m, 1H), 0.99 (t, *J* = 7.4 Hz, 3H). ^{13}C NMR (126 MHz, Methanol-d_4) δ 177.56, 158.51, 145.34, 145.21, 142.61, 128.77, 128.15, 128.13, 126.19, 126.17, 120.91, 67.84, 57.64, 48.46, 26.55, 10.68. Yield: 87%. HRMS (ESI): m/z calculated for $C_{19}H_{21}N_2O_3$ $[M+H]^+$: 325.1552; found: 325.1549.

L-2-ABA amide · HCl:



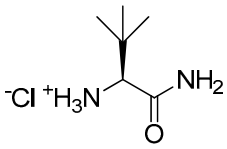
^{1}H NMR (400 MHz, DMSO-d_6) δ 8.28 (s, 3H), 8.00 (s, 1H), 7.48 (s, 1H), 3.72-3.62 (m, 1H), 1.91 – 1.68 (m, 2H), 0.89 (t, *J* = 7.5 Hz, 3H). 13C NMR (101 MHz, DMSO-d_6) δ 170.24, 53.11, 24.03, 8.97. Yield: 90%. HRMS (ESI): m/z calculated for $C_4H_{11}N_2O$ $[M+H]^+$: 103.0871; found: 103.0873.

Fmoc-L-*tert*-Leu amide:



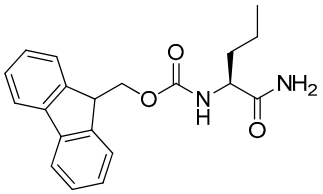
^{1}H NMR (500 MHz, DMSO-d_6) δ 7.89 (d, J = 7.5 Hz, 2H), 7.76 (d, J = 7.5 Hz, 2H), 7.41 (t, J = 7.5, 2H), 7.38 (s, 1H), 7.34-7.28 (m, 2H), 7.16 (d, J = 9.8 Hz, 1H), 7.05 (s, 1H), 4.33 – 4.26 (m, 1H), 4.25 – 4.19 (m, 2H), 3.89 (d, J = 9.8 Hz, 1H), 0.93 (s, 9H). 13C NMR (126 MHz, DMSO-d_6) δ 172.15, 156.06, 143.88, 143.82, 140.70, 140.68, 127.65, 127.06, 125.41, 120.08, 65.78, 62.35, 33.72, 26.71. Yield: 86%. HRMS (ESI): m/z calculated for $C_{21}H_{25}N_2O_3$ $[M+H]^+$: 353.1865; found: 353.1865.

L-*tert*-Leu amide · HCl:



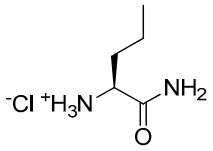
^{1}H NMR (400 MHz, DMSO-d_6) δ 8.18 (s, 3H), 7.94 (s, 1H), 7.54 (s, 1H), 3.49 (d, J = 4.9 Hz, 1H), 1.00 (s, 9H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 169.51, 60.80, 32.92, 26.91. Yield: 95%. HRMS (ESI): m/z calculated for $C_6H_{15}N_2O$ $[M+H]^+$: 131.1184; found: 131.1180.

Fmoc-L-*nor*Val amide:



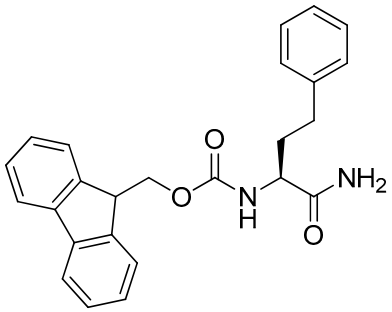
^{1}H NMR (400 MHz, DMSO-d_6) δ 7.89 (d, J = 7.5 Hz, 2H), 7.73 (dd, J = 7.5, 4.2 Hz, 2H), 7.42 (t, J = 7.5, 2H), 7.36-7.29 (m, 3H), 7.27 (s, 1H), 6.95 (s, 1H), 4.32-4.18 (m, 3H), 3.92 (td, J = 8.8, 5.1 Hz, 1H), 1.67 – 1.43 (m, 2H), 1.40-1.20 (m, 2H), 0.86 (t, J = 7.3 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 174.53, 156.41, 144.41, 144.28, 141.19, 128.10, 127.52, 125.79, 120.56, 66.01, 54.67, 47.20, 34.47, 19.17, 14.04. Yield: 84%. HRMS (ESI): m/z calculated for $C_{20}H_{22}N_2O_3Na$ $[M+Na]^+$: 361.1528; found: 361.1521

L-*nor*Val amide · HCl:



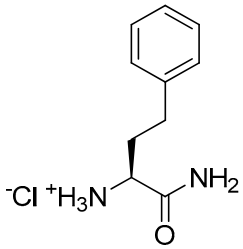
^{1}H NMR (400 MHz, DMSO-d_6) δ 8.27 (s, 3H), 7.99 (s, 1H), 7.49 (s, 1H), 3.70 (s, 1H), 1.80 – 1.64 (m, 2H), 1.42 – 1.24 (m, 2H), 0.88 (t, J = 7.3 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 170.44, 52.01, 32.94, 17.56, 13.60. Yield: 91%. HRMS (ESI): m/z calculated for $C_5H_{13}N_2O$ [M+H]$^+$: 117.1028; found: 117.1017.

Fmoc-L-*homo*Phe amide:



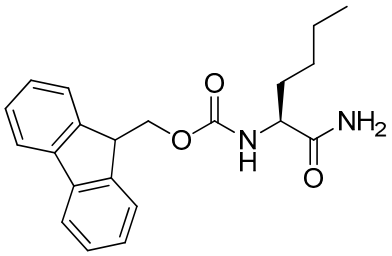
^{1}H NMR (400 MHz, DMSO-d_6) δ 7.90 (d, J = 7.5 Hz, 2H), 7.76 (dd, J = 7.5, 4.9 Hz, 2H), 7.50 (d, J = 8.3 Hz, 1H), 7.42 (d, J = 7.5, 2H), 7.38 – 7.24 (m, 5H), 7.19 (d, J = 7.2 Hz, 2H), 7.17 (s, 1H), 7.02 (s, 1H), 4.38 – 4.20 (m, 3H), 3.95 (td, J = 9.0, 4.6 Hz, 1H), 2.71 – 2.53 (m, 2H), 2.01 – 1.77 (m, 2H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 173.76, 155.99, 143.91, 143.78, 141.38, 140.71, 128.28, 128.25, 127.61, 127.03, 125.79, 125.27, 120.07, 65.53, 54.28, 46.71, 33.70, 31.67. Yield: 87%. HRMS (ESI): m/z calculated for $C_{25}H_{24}N_2O_3$ [M+H]$^+$: 401.1865; found: 401.1862.

L-*homo*Phe amide · HCl:



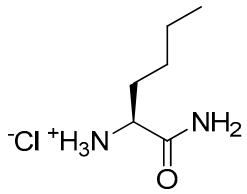
^{1}H NMR (400 MHz, DMSO-d_6) δ 8.37 (s, 3H), 8.03 (s, 1H), 7.55 (s, 1H), 7.31 – 7.24 (m, 2H), 7.22 – 7.15 (m, 3H), 3.80 (s, 1H), 2.69-2.55 (m, 2H), 2.09-1.95 (m, 2H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 170.15, 140.68, 128.46, 128.11, 126.09, 52.13, 32.81, 30.37. Yield: 70%. HRMS (ESI): m/z calculated for $C_{10}H_{15}N_2O$ [M+H]$^+$: 179.1184; found: 179.1201.

Fmoc-L-*nor*Leu amide:



^{1}H NMR (400 MHz, DMSO-d_6) δ 7.89 (d, J = 7.5 Hz, 2H), 7.74 (dd, J = 7.6, 3.9 Hz, 2H), 7.42 (t, J = 7.5, 2H), 7.37-7.30 (m, 3H), 7.27 (s, 1H), 6.96 (s, 1H), 4.35 – 4.19 (m, 3H), 3.91 (td, J = 8.8, 5.0 Hz, 1H), 1.70 – 1.46 (m, 2H), 1.35-1.21 (m, 4H), 0.88 (t, J = 7.3 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 174.00, 155.90, 143.92, 143.78, 140.69, 127.60, 127.02, 125.29, 120.06, 65.51, 54.43, 46.69, 31.60, 27.64, 21.78, 13.85. Yield: 87%.

Fmoc-L-*nor*Leu amide · HCl:



^{1}H NMR (400 MHz, DMSO-d_6) δ 8.25 (s, 3H), 7.97 (s, 1H), 7.48 (s, 1H), 3.69 (t, J = 6.4 Hz, 1H), 1.79-1.66 (m, 2H), 1.33-1.22 (m, 4H), 0.86 (t, J = 7.3 Hz, 3H). ^{13}C NMR (101 MHz, DMSO-d_6) δ 170.44, 52.15, 30.53, 26.22, 21.82, 13.71. Yield: 90%. HRMS (ESI): m/z calculated for $C_6H_{15}N_2O$ $[M+H]^+$: 131.1184; found: 131.1180.

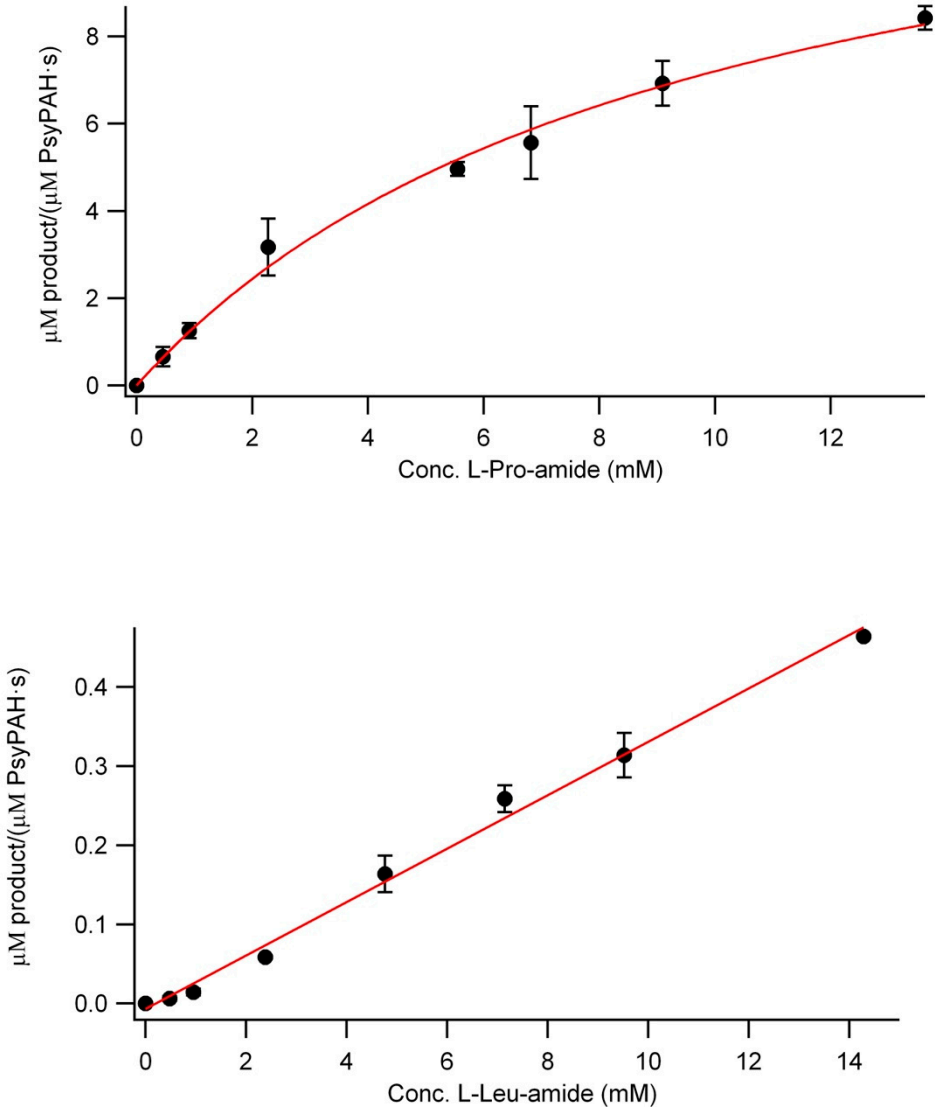


Figure S1. Kinetic determinations for L-Pro- and L-Leu-amide (35ºC, pH 7.0).

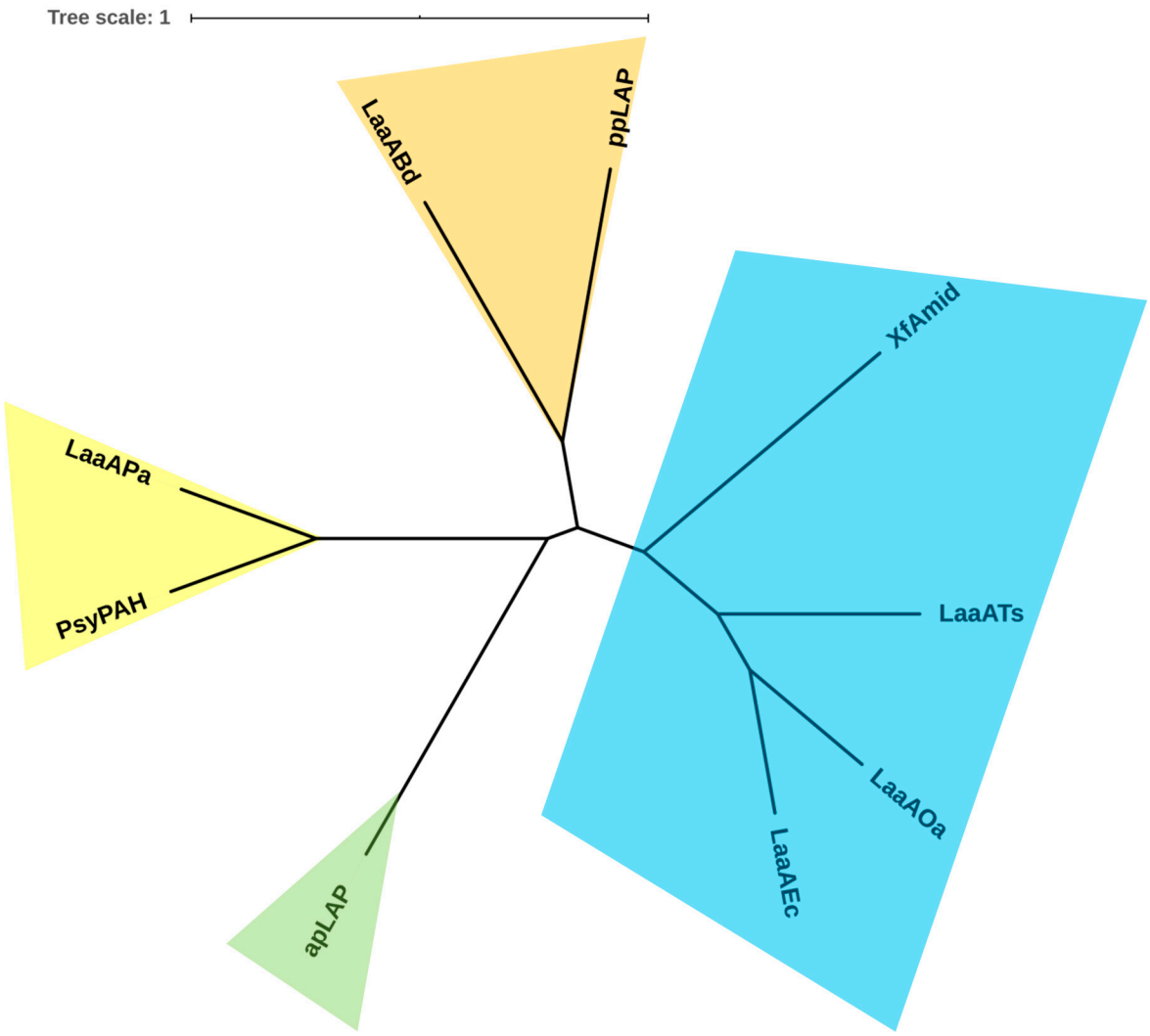


Figure S2. Phylogenetic analysis of different enzymes with proven L-amidase activity. The representative structures of each group are as follows: acetamidase/formamidase clan (blue, PDB ID: 3TKK), aminopeptidase M17 clan (orange, PDB ID: 3H8E), aminopeptidase M28 clan (green, PDB 1AMP) and peptidase S33 clan (yellow, PsyPAH, PDB 7A6G). Image was generated with i-tol (Letunic and Bork, 2019).

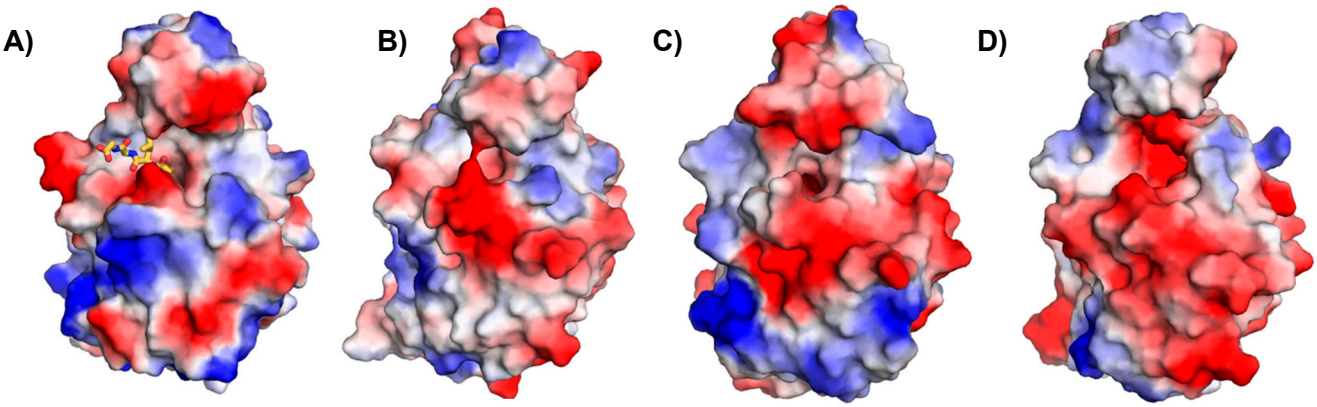


Figure S3. Surface representation of different peptidase S33 family members. A) APF1 from *Thermoplasma acidophilum* (E213Q mutant) bound to PLGG peptide (PDB 1XRP, peptide in stick mode, yellow tones), B) PsyPAH (PDB 7A6G), C) amidohydrolase VinJ (PDB 3WMR), and D) putative uncharacterized PAP from *Mycobacterium smegmatis* (PDB 3NWO). The orientation is exactly the same for the four structures. Qualitative charge-smoothed potential was generated with Pymol.

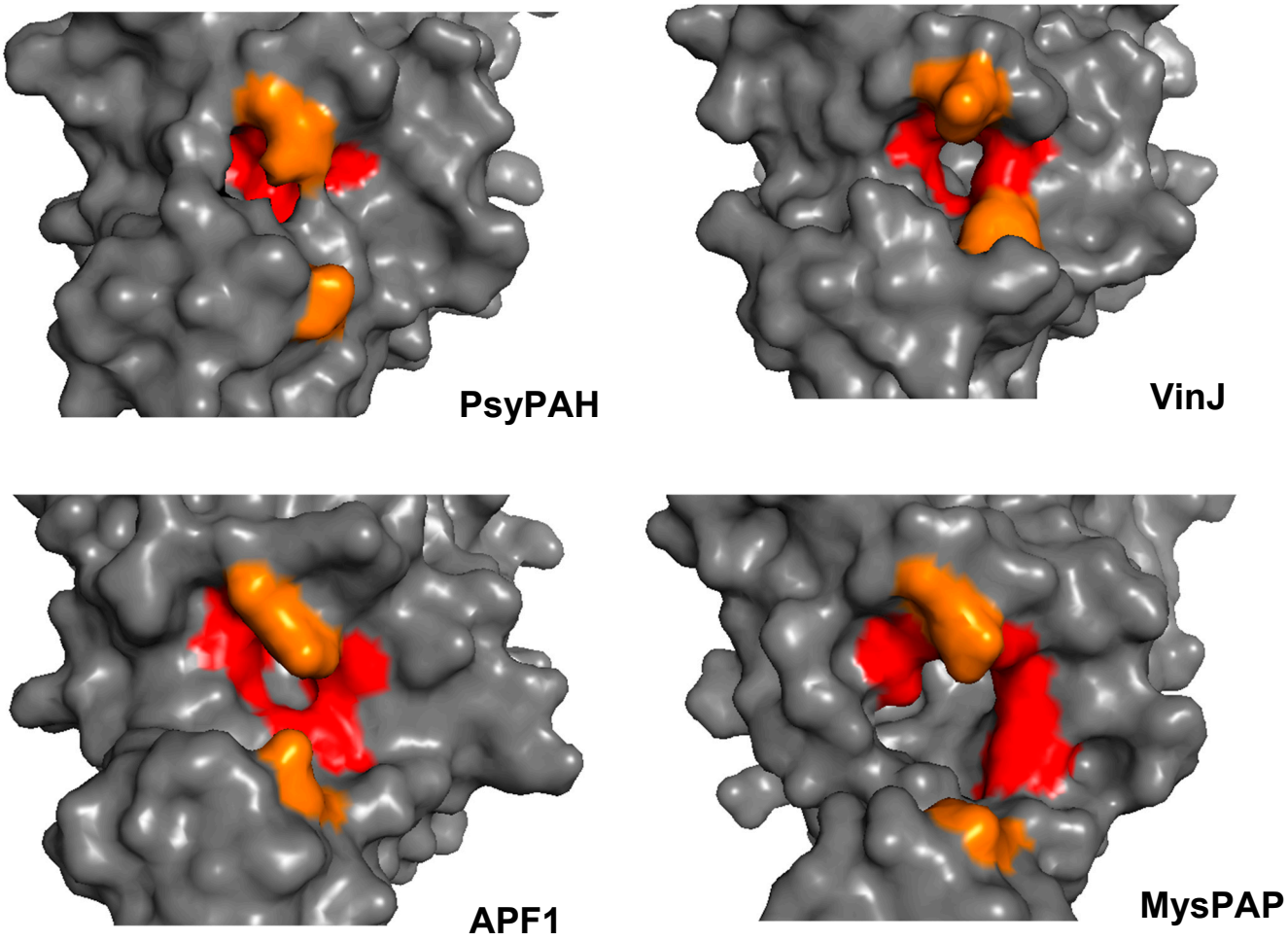


Figure S4. Comparison of the polyketide substrate binding site of VinJ (PDB 3WMR) with that of APF1 (PDB 1XRP), MysPAP (PDB 3NWO) and PsyPAH (PDB 7A6G). The position of the hydrophobic tunnel (Shinohara et al., 2014) is shown in red, whereas additional residues suggested to bind the polyketide moiety in VinJ are shown in orange.

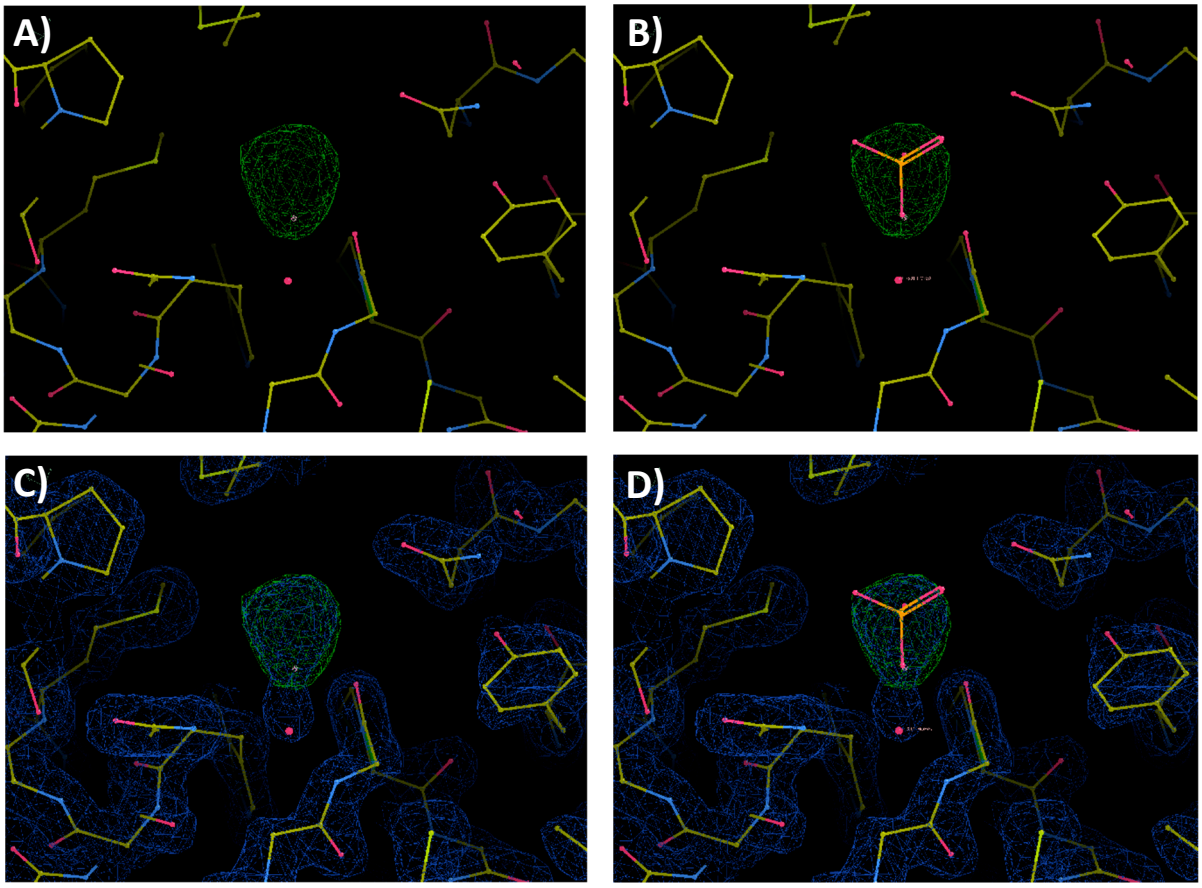


Figure S5. Omit maps calculated for PsyPAH structure. A) |mFo-DFc| omit maps (green) calculated using the deposited model PDB ID. 7A6G omitting the phosphate molecule and contoured at a 4.1 σ level. The current phosphate moiety is shown in B. C and D also shows the |2mFo-DFc| omit maps of the whole model (blue) contoured at a 1.5 σ level with (C) and without (D) the phosphate molecule.

CHAIN	Z	rmsd	lali	nres	%id	Nomenclature
3WMR-A	48.6	1	296	296	55	PROLINE IMINOPEPTIDASE;
3NWO-A	43.1	2.2	291	299	50	PROLINE IMINOPEPTIDASE;
1MU0-A	41.9	1.7	292	294	34	PROLINE IMINOPEPTIDASE;
5H3H-B	28.3	2.4	245	269	20	ABHYDROLASE DOMAIN-CONTAINING PROTEIN;
2YYS-A	28.2	2.4	257	283	19	PROLINE IMINOPEPTIDASE-RELATED PROTEIN;
3FOB-A	27.4	2.3	241	277	20	BROMOPEROXIDASE;
1A8Q-A	27.3	2.3	241	274	19	BROMOPEROXIDASE A1;
4UHC-A	27.2	2.3	237	278	21	ESTERASE;
2XUA-H	27.1	2.7	244	261	23	3-OXOADIPATE ENOL-LACTONASE;
1A88-A	27	2.5	244	275	16	CHLOROPEROXIDASE L;
1J1I-A	26.9	2.5	241	258	20	META CLEAVAGE COMPOUND HYDROLASE;
1A8S-A	26.9	2.4	242	273	17	CHLOROPEROXIDASE F;
3OM8-B	26.8	2.5	238	265	23	PROBABLE HYDROLASE;
6EB3-B	26.6	3.1	245	268	22	EST1;
1ZOI-A	26.6	2.5	244	275	14	ESTERASE;
1VA4-A	26.5	2.4	243	271	18	ARYLESTERASE;
4RNC-A	26.5	2.6	251	281	20	ESTERASE;
4I3F-A	26.5	2.6	246	283	17	SERINE HYDROLASE CCSP0084;
4DGQ-A	26.5	2.5	245	277	15	NON-HEME CHLOROPEROXIDASE;
4LXH-A	26.4	2.7	251	276	20	MCP HYDROLASE;
4QES-A	26.3	2.5	243	441	19	NON-HAEM BROMOPEROXIDASE BPO-A2, MATRIX PROTEIN 1
4X00-A	26.2	2.6	253	273	18	PUTATIVE HYDROLASE;
5EGN-A	26.1	2.7	244	260	21	ESTERASE;
1IUN-B	26	2.7	240	276	17	META-CLEAVAGE PRODUCT HYDROLASE;
1C4X-A	26	2.6	250	281	18	PROTEIN (2-HYDROXY-6-OXO-6-PHENYLHEXA-2,4-DIENOAT
1U2E-A	25.9	2.8	251	286	18	2-HYDROXY-6-KETONONA-2,4-DIENEDIOIC ACID
2OG1-A	25.8	2.6	247	285	20	2-HYDROXY-6-OXO-6-PHENYLHEXA-2,4-DIENOATE HYDROLA
4OPM-A	25.8	2.6	244	299	17	LIPASE;
6AZD-A	25.7	2.5	242	266	17	PPKAI2-LIKE H;
6I8W-A	25.6	2.7	245	310	18	ALPHA/BETA FOLD HYDROLASE;
2OCI-A	25.5	2.6	236	255	16	VALACYCLOVIR HYDROLASE;
4NS4-A	25.5	2.4	238	271	20	ALPHA/BETA HYDROLASE FOLD PROTEIN;
3E0X-A	25.4	2.3	228	245	16	LIPASE-ESTERASE RELATED PROTEIN;
5FRD-A	25.3	2.4	239	256	17	CARBOXYLESTERASE (EST-2);
4G8B-A	25.3	2.8	244	277	17	ALPHA/BETA HYDROLASE FOLD PROTEIN;
1Q0R-A	25.1	3.1	252	297	18	ACLACINOMYCIN METHYLESTERASE;
4PW0-A	25.1	3	244	275	18	ALPHA/BETA HYDROLASE FOLD PROTEIN;
4F0J-A	25	2.5	246	312	17	PROBABLE HYDROLYTIC ENZYME;
3V48-A	25	2.8	240	268	15	PUTATIVE AMINOACRYLATE HYDROLASE RUTD;
6BRT-A	25	2.5	242	285	15	D3-CTH-D14-D-RING;
1HKH-A	24.9	2.6	240	279	18	GAMMA LACTAMASE;
6EIC-B	24.9	2.5	236	279	18	MYCOBACTERIUM TUBERCULOSIS MONOGLYCERIDE LIPASE;
3E3A-B	24.8	2.6	240	275	19	POSSIBLE PEROXIDASE BPOC;
3OOS-A	24.7	3.6	256	278	18	ALPHA/BETA HYDROLASE FAMILY PROTEIN;
1AZW-A	24.6	2.9	263	313	18	PROLINE IMINOPEPTIDASE;
4Q3L-B	24.6	2.8	239	280	21	MGS-M2;
4PSU-A	24.6	2.8	245	286	13	ALPHA/BETA HYDROLASE;
1WOM-A	24.6	2.7	241	271	18	SIGMA FACTOR SIGB REGULATION PROTEIN RSBQ;
4DNQ-I	24.6	2.6	241	264	17	DAD2;
3QVM-B	24.4	2.7	243	280	16	OLEI00960;
6F9O-A	24.4	2.7	247	309	13	HALOALKANE DEHALOGENASE;
3QYJ-A	24.4	3	251	291	14	ALR0039 PROTEIN;
4MJ3-B	24.3	2.5	242	303	14	HALOALKANE DEHALOGENASE;
5NG7-A	24.3	3.1	251	292	19	EPOXIDE HYDROLASE;
4IH4-A	24.2	2.5	239	261	18	AT3G03990/T11I18_10;
4NVR-A	24.2	3	253	302	15	PUTATIVE ACYLTRANSFERASE;
5LKA-A	24.2	2.9	257	301	15	HALOALKANE DEHALOGENASE;
5A62-A	24.2	2.7	242	272	22	PUTATIVE ALPHA/BETA HYDROLASE FOLD PROTEIN;
3R3U-A	24.2	3	255	304	18	FLUOROACETATE DEHALOGENASE;
5YHP-B	24.2	2.9	266	319	17	COLD ACTIVE PROLINE IMINOPEPTIDASE;

CHAIN	Z	rmsd	lali	nres	%id	Nomenclature
2VF2-A	24.1	2.6	242	284	21	2-HYDROXY-6-OXO-6-PHENYLHEXA-2,4-DIENOATE
5Z7W-B	24.1	2.7	246	271	15	HYPOSENSITIVE TO LIGHT 1;
1QTR-A	24.1	3.1	264	314	16	PROLYL AMINOPEPTIDASE;
5O7G-A	24	3	249	310	16	ALPHA/BETA HYDROLASE FAMILY PROTEIN;
5Z7Z-B	24	2.6	238	266	16	DWARF 14;
1Y37-A	23.9	3	252	294	16	FLUOROACETATE DEHALOGENASE;
4HAI-A	23.9	2.7	239	548	19	BIFUNCTIONAL EPOXIDE HYDROLASE 2;
4KAJ-A	23.9	3.1	254	300	16	PENTAETHYLENE GLYCOL;
5CBK-A	23.9	2.8	242	271	14	SHHTL5;
4C6H-A	23.9	2.9	250	291	18	HALOALKANE DEHALOGENASE;
6ATX-A	23.8	2.7	242	265	16	PPKAI2-LIKE C;
6QE2-A	23.8	2.5	224	257	16	MONOACYLGLYCEROL LIPASE;
4HTA-A	23.8	2.7	243	270	14	HYDROLASE, ALPHA/BETA FOLD FAMILY PROTEIN;
3BWX-A	23.8	3.1	248	285	14	ALPHA/BETA HYDROLASE;
5HDF-B	23.7	3.1	249	320	18	HYDROLASE;
2QVB-A	23.7	3	256	296	13	HALOALKANE DEHALOGENASE 3;
6J2R-A	23.7	2.6	238	267	16	HYPOSENSITIVE TO LIGHT 8;
1B6G-A	23.7	2.8	248	310	17	HALOALKANE DEHALOGENASE;
2XMZ-A	23.6	2.8	232	266	21	HYDROLASE, ALPHA/BETA HYDROLASE FOLD FAMILY;
3WI7-A	23.6	3.1	254	291	14	HALOALKANE DEHALOGENASE;
3U1T-B	23.6	3	253	299	14	DMMA HALOALKANE DEHALOGENASE;
6G75-A	23.6	3	250	296	14	COMMON ANCESTOR OF HALOALKANE DEHALOGENASE AND RE
5Z7X-A	23.6	2.8	242	270	16	HYPOSENSITIVE TO LIGHT 4;
1M33-A	23.5	2.8	232	255	19	BIOH PROTEIN;
4CCY-A	23.5	3	243	285	15	CARBOXYLESTERASE YBFK;
6A9D-A	23.4	2.8	243	275	16	HYPOSENSITIVE TO LIGHT 7;
3KXP-A	23.4	3.2	241	268	13	ALPHA-(N-ACETYLAMINOMETHYLENE)SUCCINIC ACID
5NFQ-A	23.4	2.9	249	298	18	EPOXIDE HYDROLASE BELONGING TO ALPHA/BETA HYDROLA
3AFI-B	23.4	3.1	254	302	16	HALOALKANE DEHALOGENASE;
5XM6-A	23.4	2.6	238	318	16	EPOXIDE HYDROLASE;
3HJU-A	23.4	2.6	232	291	13	MONOGLYCERIDE LIPASE;
4UFO-A	23.4	2.8	241	321	15	EPOXIDE HYDROLASE;
3FSG-A	23.4	3.3	239	267	18	ALPHA/BETA SUPERFAMILY HYDROLASE;
2R11-D	23.3	3	243	288	17	CARBOXYLESTERASE NP;
3SK0-A	23.2	3	252	308	16	HALOALKANE DEHALOGENASE;
3KDA-A	23.2	2.9	256	298	14	CFTR INHIBITORY FACTOR (CIF);
2XT0-A	23.2	2.9	245	297	16	HALOALKANE DEHALOGENASE;
4B9A-A	23.2	3.1	251	298	14	PROBABLE EPOXIDE HYDROLASE;
2PSF-A	23.2	3	248	307	14	RENILLA-LUCIFERIN 2-MONOOXYGENASE;
4IO0-A	23.1	3.2	252	288	17	SOLUBLE EPOXIDE HYDROLASE;
6K5E-B	23.1	2.7	231	253	20	PIMELOYL-[ACYL-CARRIER PROTEIN] METHYL ESTER ESTE
4NMW-A	23.1	2.9	232	254	19	PIMELYL-[ACYL-CARRIER PROTEIN] METHYL ESTER ESTER
4OSE-B	23.1	3.1	242	284	15	PUTATIVE HYDROLASE;
4ZWN-B	23	2.8	241	311	12	MONOGLYCERIDE LIPASE;
6XY9-A	23	3	250	303	16	HALOALKANE DEHALOGENASE;
6KHM-K	23	3	249	312	18	HYDROLASE, ALPHA/BETA DOMAIN PROTEIN;
5ESR-A	22.9	2.8	243	308	12	HALOALKANE DEHALOGENASE;
5MXP-A	22.8	3.2	254	294	14	ALPHA/BETA HYDROLASE;
5W15-A	22.8	2.9	242	270	17	ALPHA/BETA HYDROLASE FOLD PROTEIN;
3BDI-A	22.8	2.2	197	207	18	UNCHARACTERIZED PROTEIN TA0194;
5CW2-A	22.7	2.8	251	319	17	PUTATIVE EPOXIDE HYDROLASE EPHA;
5DNW-A	22.7	2.8	236	274	17	SHKAI2IB;
5JD6-A	22.7	2.7	211	219	20	MGS-MCHE2;
2E3J-A	22.6	2.7	249	346	15	EPOXIDE HYDROLASE EPHB;
1EHY-A	22.4	3.3	243	282	14	PROTEIN (SOLUBLE EPOXIDE HYDROLASE);
5ALJ-A	22.3	2.6	235	523	19	BIFUNCTIONAL EPOXIDE HYDROLASE 2;
5XWZ-B	22.3	2.9	244	264	12	UNPLACED GENOMIC SCAFFOLD SUPERCONT1.36, WHOLE GE
5XO7-F	22.2	2.9	245	264	13	LACTONASE FOR PROTEIN;
5XMW-A	22.2	3	248	270	14	ZEARALENONE LACTONASE;
3PF8-A	22.1	2.8	217	251	16	CINNAMOYL ESTERASE;

CHAIN	Z	rmsd	lali	nres	%id	Nomenclature
2WTM-A	22.1	2.8	216	250	15	EST1E;
6LZH-A	22	2.8	232	289	13	GRGF;
4L0C-A	22	2.7	225	255	16	DEFORMYLASE;
3QIT-B	21.9	3.9	234	283	20	POLYKETIDE SYNTHASE;
4KE6-E	21.9	2.9	208	226	16	THERMOSTABLE MONOACYLGLYCEROL LIPASE;
5URO-A	21.7	3	243	336	14	PREDICTED PROTEIN;
3C5V-A	21.6	3	236	294	15	PROTEIN PHOSPHATASE METHYLESTERASE 1;
5XMD-B	21.6	3.1	240	322	15	EPOXIDE HYDROLASE A;
4MEA-A	21.6	2.9	249	324	12	PREDICTED PROTEIN;
6RA3-A	21.5	3	240	267	14	PUTATIVE DIOXYGENASE (1H-3-HYDROXY-4-OXOQUINALDIN
4Y7D-A	21.4	3.1	236	289	17	ALPHA/BETA HYDROLASE FOLD PROTEIN;
4LHE-A	21.4	3.4	223	250	17	THERMOSTABLE MONOACYLGLYCEROL LIPASE;
2WM2-C	21.3	3.1	242	277	12	1-H-3-HYDROXY-4-OXOQUINALDINE 2,4-DIOXYGENASE;
5G59-A	21.1	2.7	212	274	17	ESTERASE;
6NY9-B	21.1	4	228	250	18	MYCOPHENOLIC ACID ACYL-GLUCURONIDE ESTERASE, MITO
6T6H-A	21	2.8	228	253	14	BOTH;
3R0V-A	21	2.8	217	257	19	ALPHA/BETA HYDROLASE FOLD PROTEIN;
3LLC-A	21	3.5	224	261	17	PUTATIVE HYDROLASE;
3IBT-A	21	3.1	234	263	14	1H-3-HYDROXY-4-OXOQUINOLINE 2,4-DIOXYGENASE;
5BOV-B	20.7	3.2	248	302	18	PUTATIVE EPOXIDE HYDROLASE PROTEIN;
4GDM-C	20.7	3.1	227	259	19	2-SUCCINYL-6-HYDROXY-2,4-CYCLOHEXADIENE-1-CARBOXY
5UGQ-A	20.6	3	259	463	14	CARBOXYLESTERASE A;
4FBM-A	20.6	3	214	252	15	LIPS LIPOLYTIC ENZYME;
2JBW-B	20.4	2.8	215	360	19	2,6-DIHYDROXY-PSEUDO-OXYNICOTINE HYDROLASE;
3BF7-A	20.2	3.3	225	255	16	ESTERASE YBFF;
1IMJ-A	20.1	2.2	191	208	16	CCG1-INTERACTING FACTOR B;

Table S1. Homolog structures of PsyPAH obtained with the DALI server (Holm, 2020). PAP and PAP-like enzymes are highlighted.

PDB ID	Structural model
1MU0	APF1 Complex with PCK
1XQW	APF1 -mutant S105A complex with Phe-Leu
1XQX	APF1-mutant S105A complex with PCK
1XQY	APF1-mutant S105A complex with Pro-Leu-Gly-Gly
1XRL	APF1-mutant Y205F complex with inhibitor PCK
1XRM	APF1-mutant E213Q soaked with peptide Ala-Phe
1XRN	APF1-mutant E213Q soaked with peptide Phe-Ala
1XRO	APF1-mutant E213Q soaked with peptide Phe-Leu
1XRP	APF1-mutant E213Q soaked with peptide Pro-Leu-Gly-Gly
1XRQ	APF1-mutant E245Q soaked with peptide Phe-Leu
1XRR	APF1-mutant E245Q soaked with peptide Pro-Pro

Table S2. Different ligand-bound structures of Tricorn Interacting Factor F1 (APF1)

Structural motifs	APF1 (1XRP)	PsyPAH (7A6G)	VinJ (3WMR)	MysPAP
E1 site	Y178	Y186	Y183	Y186
	W188	W196	W193	T196
	L196	F204	F201	V204
	Y205	Y214	Y211	Y214
	N212	T221	N218	-[#]
S1 site (cap domain)	V134	M142	M139	M142
	T137	W145*	W142	W145
	M141	A149	M146	A149*
	N209	N218	N215	N218
	I216	V225	V222	V225
	I220	M229	L226	L229
	W223	W232	W229	W232
S1 site (catal. dom.)	P38	P43	P40	P43
	I79	V87	V84	P88
	Y106[b]	W114[b]	W111[b]	W114[b]
	A109	M117	M114	M117P1
	L131	P139	P136	P139
	V246	A255	A252	A255
	V250	V259	T256	T259
S1′ site	L182	V190	V187	V190
	L183	C191	C188	C191
	W188	W196	W193	T196
	V192	V200	F197	F200
	S195	T203	T200	S203
	M40	C45	S42	M45
	Y44	Y49	Y46	Y49
	L272	M281	V278	C281
P site	S104	Q112	Q109	Q112
	S128	A136	A132	C136
	T239	I248	L245	I248
	D244[a]	**D253**[a]	**D250**[a]	**D253**[a]
	H271[a]	**H280**[a]	**H277**[a]	**H280**[a]
	T273	P282	P279	T282
Other catalytic residues	G37[b]	G42[b]	G39[b]	G42[b]
	S105[a]	**S113**[a]	**S110**[a]	**S113**[a]
	E213[c]	E222	E219	-[#]
	E245[c]	E254	E251	E254

Table S3. Residues involved in substrate binding and catalysis in the different pockets of APF1 (Goettig et al., 2002 and 2005). The counterpart residues for PsyPAH and other S33 family members were obtained by structural superposition. Numbering is that from the different PDBs used for superposition: APF1 (PDB 1XRP), PsyPAH (PDB 7A6G), amidohydrolase VinJ (PDB 3WMR) and MysPAP (PDB 3NWO). [#]Displaced loop. [a]Catalytic triad. [b]Oxyanion hole. [c]Interaction with peptide N termini.

APF1 (1XRP)	PsyPAH (7A6G)	VinJ (3WMR)	MysPAP
Y178	Y186	Y183	Y186
V192	V200	F197	F200
L196	F204	F201	V204
A199	I207	I204	M207
Y205	Y214	Y211	Y214
Additional residues			
Q171*	T180*	F176	L179*
E200	D208	Y205	E208

Table S4. Comparison of residues proposed in the polyketide binding tunnel in VinJ (Shinohara et al., 2014). The counterpart residues for PsyPAH, APF1 and MysPAP were obtained by structural superposition. Numbering is that from the different PDBs used for superposition: APF1 (PDB 1XRP), PsyPAH (PDB 7A6G), amidohydrolase VinJ (PDB 3WMR) and MysPAP (PDB 3NWO). *Lateral chains displaced with respect to $F176^{VinJ}$.

References

- Goettig, P.; Groll, M.; Kim, J.S.; Huber, R.; Brandstetter, H. Structures of the tricorn-interacting aminopeptidase F1 with different ligands explain its catalytic mechanism. *EMBO J* **2002**, *21*, 5343-5352, doi:10.1093/emboj/cdf552.

- Goettig, P.; Brandstetter, H.; Groll, M.; Göhring, W.; Konarev, P.V.; Svergun, D.I.; Huber, R.; Kim, J.S. X-ray snapshots of peptide processing in mutants of tricorn-interacting factor F1 from Thermoplasma acidophilum. *J Biol Chem* **2005**, *280*, 33387-33396, doi:10.1074/jbc.M505030200

- Holm, L. DALI and the persistence of protein shape. *Protein Sci* **2020**, *29*, 128-140, doi:10.1002/pro.3749

- Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **2019**, *47*, W256-W259, doi:10.1093/nar/gkz239.

- Shinohara, Y.; Miyanaga, A.; Kudo, F.; Eguchi, T. The crystal structure of the amidohydrolase VinJ shows a unique hydrophobic tunnel for its interaction with polyketide substrates. *FEBS Lett* **2014**, *588*, 995-1000, doi:10.1016/j.febslet.2014.01.060.