Supporting Information

Gorelik et al. 10.1073/pnas.1719834115

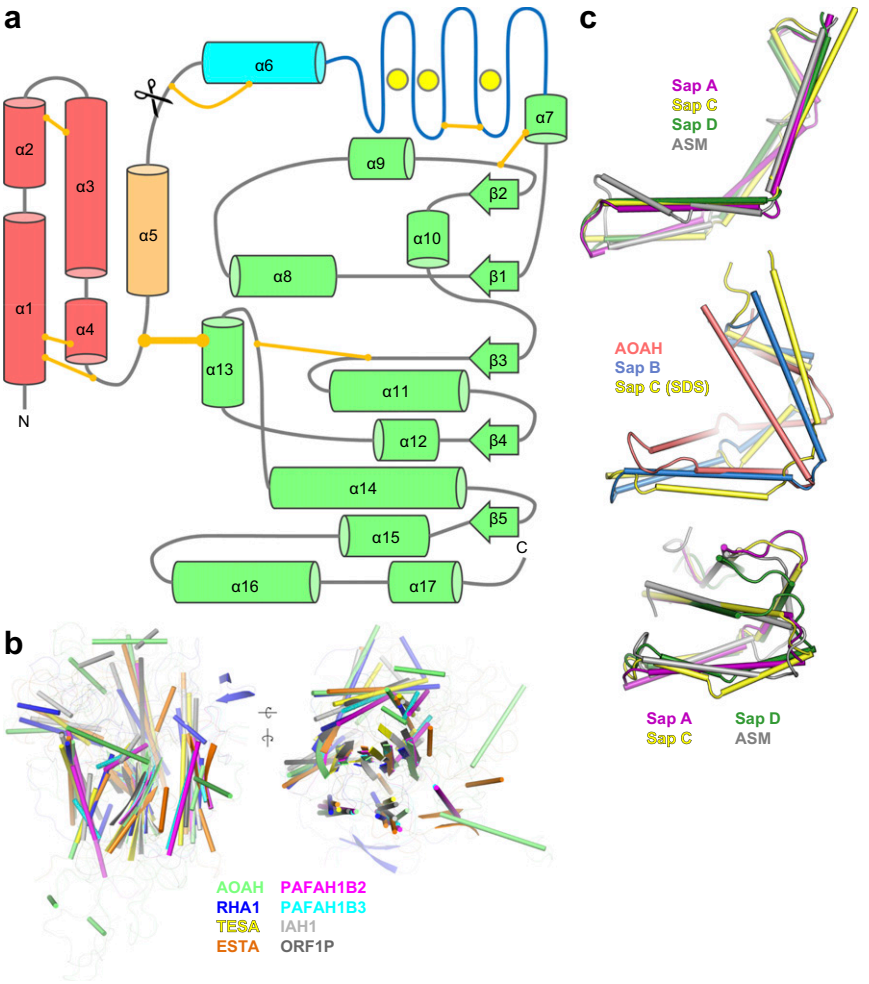

Fig. S1. AOAH structure and comparison with structurally related proteins. (*A*) Topology of AOAH, with α-helices (cylinders), β-strands (arrows), disulfide bonds (orange connectors), and calcium ions (yellow spheres). The size of secondary structure elements is proportional to sequence length. The native proteolytic processing site is marked by scissors. The saposin domain is shown in red, the saposin extension in orange, the esterase domain in green, the calcium-binding region in dark blue, and the amphipathic helix in cyan. (*B*) Superimposition of the AOAH esterase domain with other GDSL esterases: rhamnogalacturonan acetylesterase (RHA1) from *Aspergillus aculeatus* (PDB ID code 1DEO, 16% identity, rmsd of 4.8 Å over 306 main-chain atoms); thioesterase 1/protease 1/lysophospholipase L1 (TESA) from *E. coli* (PDB ID code: 1IVN, 26% identity, rmsd of 4.8 Å over 440 main-chain atoms); human platelet-activating factor acetylhydrolase IB subunit β (PAFAH1B2) (PDB ID code: 1VYH, 17% identity, rmsd of 2.4 Å over 75 main-chain atoms from helices and strands); bovine platelet-activating factor acetylhydrolase IB subunit γ (PAFAH1B3) (PDB ID code: 1WAB, 15% identity, rmsd of 1.9 Å over 73 main-chain atoms from helices and strands); autotransporter esterase EstA (ESTA) from *Pseudomonas aeruginosa* (PDB ID code: 3KVN, 12% identity, rmsd of 4.2 Å over 304 main-chain atoms); isoamyl acetate-hydrolyzing esterase (IAH1) from *Saccharomyces cerevisiae* (PDB ID code: 3MIL, 20% identity, rmsd of 5.2 Å over 486 main-chain atoms); zebrafish esterase from retrotransposon ZfL2-1 (ORF1P) (PDB ID code: 4C1B, 20% identity, rmsd of 3.7 Å over 403 main-chain atoms). Additional domains of the homologs and the saposin domain of AOAH are hidden for clarity. (*C*) The saposin domain of AOAH is compared with the open conformation structures of saposin A (PDB ID code: 4DDJ), B (PDB ID code: 1N69), C (PDB ID code: 2QYP), C in a detergent environment (PDB ID code: 1SN6), D (PDB ID code: 5U85), and the saposin domain of acid sphingomyelinase (ASM) (PDB ID code: 5FIB). The closed forms of saposin A (PDB ID code: 2DOB), C (PDB ID code: 1M12), and D (PDB ID code: 2RB3) and the saposin domain of acid sphingomyelinase (PDB ID code: 5FI9) are also displayed.

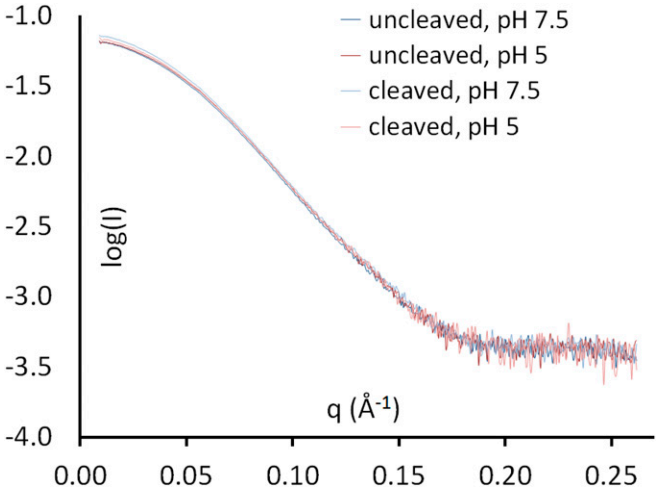

Fig. S2. X-ray scattering by AOAH in solution. The X-ray scattering profile from SEC-SAXS experiments for uncleaved or cleaved human AOAH under different conditions.

a

metal type	#	occupancy	ligands B factor	metal B factor	ligands	valence	geometry	bidentate	alternative metal
	1	1	15.9	14.6	O6	2.1	octahedral	0	
Ca	2	1	14.1	16.2	O7	2.0	octahedral	1	
	3	1	16.6	14.9	O6	2.1	octahedral	0	
	1	1	15.0	20.7	O6	1.4	octahedral	0	Ca
Mn	2	1	13.4	22.6	O7	1.3	octahedral	1	Ca
	3	1	15.5	21.8	O6	1.4	octahedral	0	Ca
	1	1	14.8	25.7	O6	1.1	octahedral	0	Ca
Zn	2	1	13.1	27.6	O7	1.0	octahedral	1	Ca
	3	1	15.3	26.6	O6	1.1	octahedral	0	Ca

b

```
P28039   Homo sapiens                 184  DVDSDKY-SVFPTLRGYHWRGRDCNDSDESVYPGRRPNN-WDVHQDSNCNGIWGVDPKDGVPYE
Q95P05   Trypanosoma brucei brucei    171  DKDGDRF-STYPAKRGTDWRGRDCDDDDARVYPGRNS---TDSQRDENCNGIYGVEESSKKTYE
Q4CUX3   Trypanosoma cruzi           174  DEDGDLF-STYPTKRGSDWRGRDCDDKDPKVYPGRNT---TDAFRDENCNGIFGVDNVTGKTYE
Q4D5N4   Trypanosoma cruzi           174  DEDGDLF-STYPTKRGSDWRGRDCDDRDPKVYPGRNT---TDAFRDENCNGIFGVDNVTGKTYE
I7MLC4   Tetrahymena thermophila     290  DLDNDGF-STIPTLRGYNWRGYDCNDFHKDIRPGVKTYA-GDQTIDYNCNGIWGVDTVSGKPYK
I7MA17   Tetrahymena thermophila     295  DFDNDGF-STIGPLRGYNWRGKDCDDFQNTIYPGRKIWEGSNPSIDYDCNGIWGVNTVTGKNYE
L8HD62   Acanthamoeba castellanii    204  DLDSDKFGALTQTLRGTFWRGKDCNDFASDIYPGRSKTN-RGPEFDHNCNGISGTDPQ-GRSYE
L8H4D7   Acanthamoeba castellanii    160  DDDDDRF-STFATLRGTSWRGKDCDDHDPNVYPGRKAAPSGKESADYNCNGINGVDPTSGRPWK
Q86B07   Dictyostelium discoideum    181  DFDGDHH-SLEATFRGYNWRGKDCNDFDKSSYPGTIEGT-GDASVDKNCNGIFGVNNQS-QSYE
D3BK82   Polysphondylium pallidum    166  DADADRF-SDAPTFRGYDWRGRDCDDANNNVYPGASPVA-GQPETDWNCNGIVGMNTTSGISWE
D3BK83   Polysphondylium pallidum    201  DLDGDRY-SLDSTFRGYSWRGKDCDDTAKHIYPGAENDP-NFPDIDWNCNGIMGSNALD-KSWE
D2VJ96   Naegleria gruberi           239  DDDHDLFPGSHMLLRNADWRGMDCDDKNPGIKPGIYDN--NQNNRDMNCNGVYGTEPKTGESYE
D2VIY6   Naegleria gruberi           150  DIDGDTY-STEFAFRGSAWRGKDCNDNNAAIYPGRKTSS-LSPLYDHNCNGVFGYD-KSGKSYE
                                          * * * . .     *.  *** **:*      **          * :**: * :      ::
```

Fig. S3. Calcium binding by AOAH. (*A*) To verify the identity of metals bound by AOAH, the 1.75-Å resolution structure of murine AOAH was refined with the three metal ions replaced by Ca, Mn, or Zn. The models were validated by the CheckMyMetal server (1). For each metal type, coordination properties are colored as acceptable (green), borderline (yellow), or outlier (red). Ligands B factor is the valence-weighted average B factor of coordinating atoms. Valence is the summation of bond valence values, accounting for metal–ligand distances. Bidentate is the number of residues that form a bidentate interaction instead of being considered as multiple ligands. An alternative metal type is sometimes proposed by the CheckMyMetal server based on the ion's environment. (*B*) The calcium-binding region of human AOAH is aligned with homologs from unicellular organisms. Conserved residues that coordinate calcium ions via their side chain are highlighted in red. UniProt identifiers are indicated. The asterisks indicate fully conserved residues, per standard notation for sequence alignments.

1. Zheng H, et al. (2014) Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nat Protoc* 9:156–170.

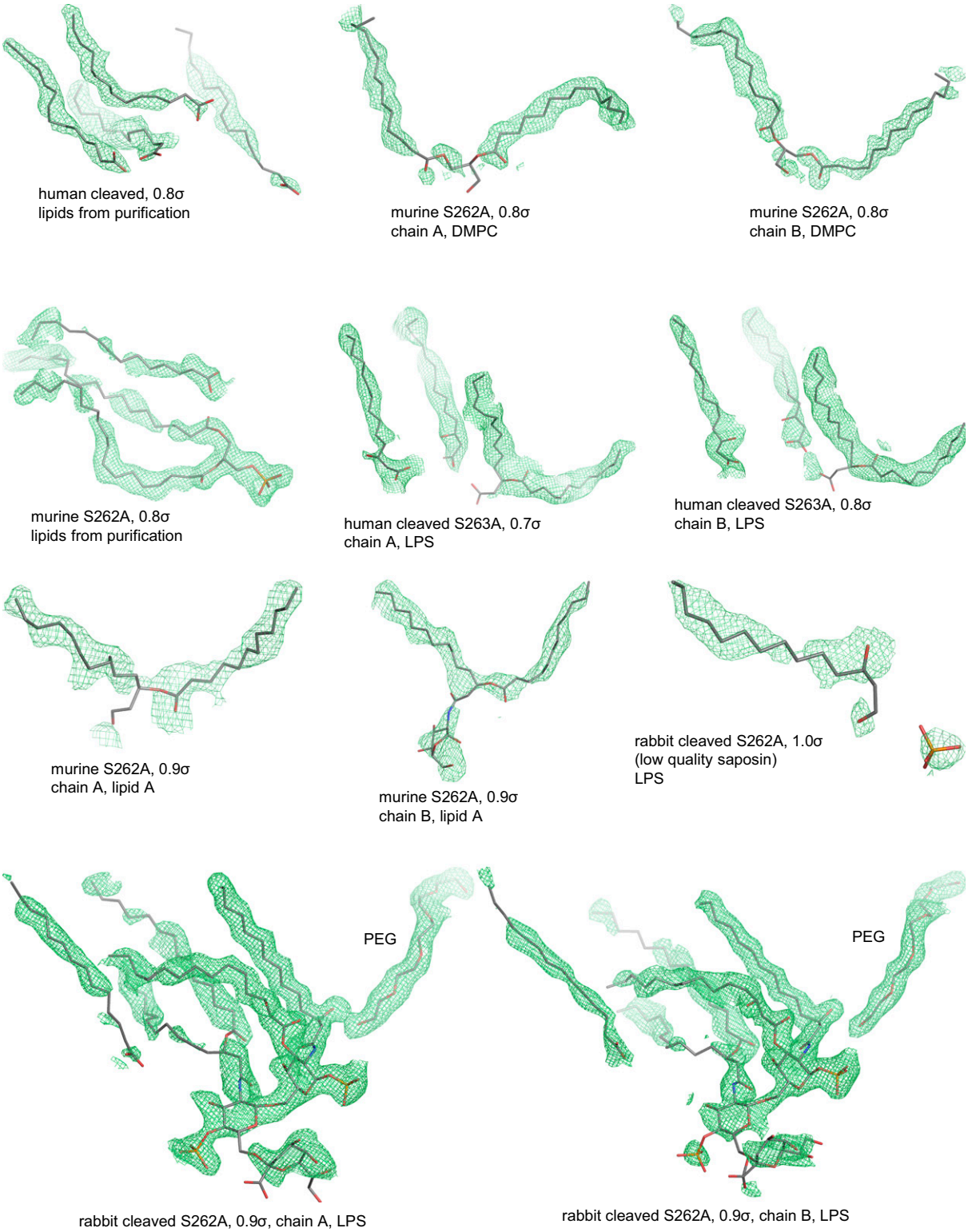

Fig. S4. Electron density maps of lipids and related ligands. $2F_o - F_c$ electron density maps are contoured at the indicated σ level. All maps were generated before the addition of ligands to the models. Most ligands are only partially visible in the electron density. Lipids carried over from the purification are modeled as free fatty acids or phosphatidic acid.

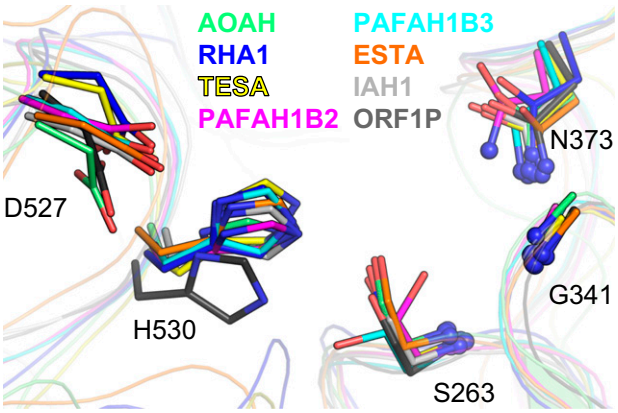

Fig. S5. Comparison of the active site with other GDSL esterases. The active-site residues of AOAH are superimposed with other GDSL esterases. The amide groups forming the oxyanion hole are shown as blue spheres. The nitrogen and oxygen atoms of the asparagine side chain were inverted for TESA and ORF1P from their initial orientations in the Protein Data Bank models. Human AOAH residues are numbered.

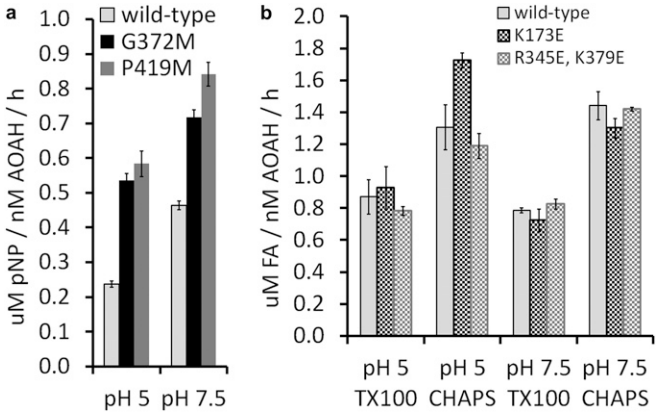

Fig. S6. Effect of mutants that block the hydrophobic tunnel on enzymatic activity against a small molecule substrate and impact of mutations of the phosphate-interacting residues. (*A*) Activity of human AOAH against p-nitrophenyl acetate was assayed at different pHs. (*B*) Activity of human AOAH against LPS in Triton X-100 or CHAPS micelles was assayed at different pHs. Data are the mean ± SD of three replicates representative of one of two experiments. FA, free fatty acids; pNP, p-nitrophenol.

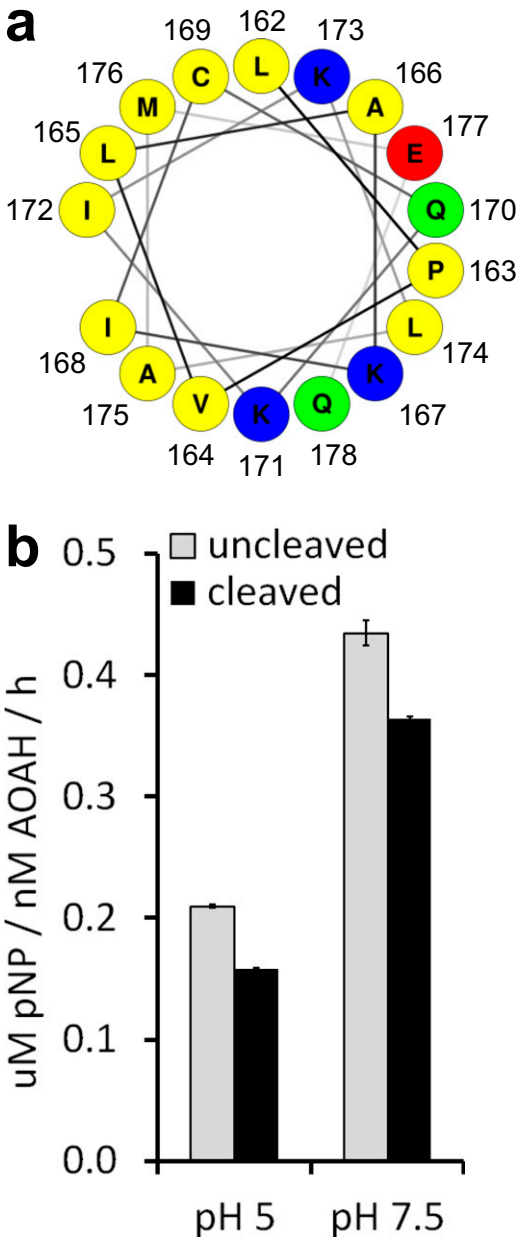

Fig. S7. Amphipathic helix of AOAH. (*A*) A helical wheel representation of the human AOAH amphipathic helix with residues colored by type: hydrophobic, yellow; polar, green; cationic, blue; and anionic, red. The diagram was created by the NetWheels server. (*B*) The effect of proteolytic processing on enzymatic activity against the small molecule substrate p-nitrophenyl acetate was assayed at different pHs. Data are the mean ± SD of three replicates representative of one of two experiments. pNP, p-nitrophenol.

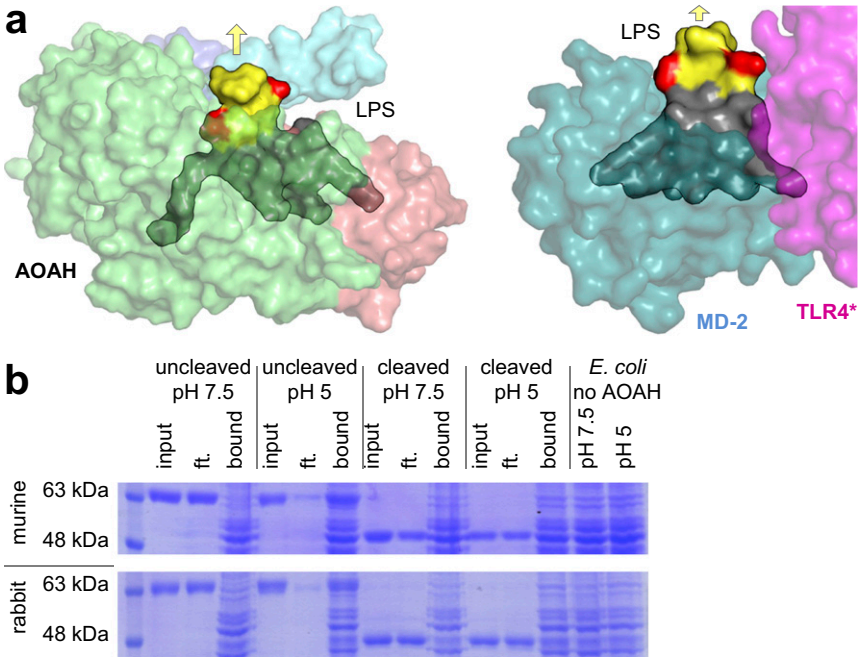

Fig. S8. AOAH binding to LPS and to bacteria. (*A*) The LPS-binding mode of AOAH is compared with that of the TLR4–MD-2 complex (PDB ID code: 3FXI). LPS is shown with acyl chains (gray), carbohydrate moiety (yellow), and phosphate groups (red). Yellow arrows indicate the expected direction of the core oligosaccharide extension. The TLR4* molecule displayed is the second (nonprimary) copy from the heterotetramer. (*B*) To test binding of murine or rabbit AOAH to intact *E. coli*, the enzyme was incubated with bacteria at different pHs, centrifuged at low speed, and visualized on reducing gels. Ft, flowthrough.

Table S1. Crystallographic data collection and structure refinement statistics

Parameters	Human cleaved	Human cleaved S263A, LPS	Rabbit cleaved S262A, LPS, low-quality saposin	Rabbit cleaved S262A, LPS	Murine S262A	Murine S262A, DMPC	Murine S262A, lipid A
PDB ID code	5W78	5W7C	5W7A	5W7B	5W7D	5W7E	5W7F
Data collection							
Space group	$P\ 2_1\ 2_1\ 2_1$	$P\ 2_1\ 2_1\ 2_1$	$C\ 2\ 2\ 2_1$	$P\ 2_1\ 2_1\ 2$	$P\ 2_1$	$P\ 2_1$	$P\ 2_1$
Unit cell a, b, c, Å	50.40, 87.36, 146.51	88.17, 104.08, 145.28	89.58, 123.20, 125.92	109.52, 138.81, 89.53	46.53, 98.59, 73.49	81.43, 85.37, 93.40	81.80, 88.35, 93.45
Unit cell α, β, γ, °	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 99.68, 90	90, 103.64, 90	90, 103.01, 90
X-ray wavelength, Å	1.91	0.97	0.97	0.98	0.98	0.98	0.98
Resolution range, Å	50–2.27 (2.35–2.27)	50–2.23 (2.31–2.23)	50–2.30 (2.38–2.30)	50–1.90 (1.97–1.90)	50–1.75 (1.81–1.75)	50–1.83 (1.90–1.83)	50–2.80 (2.90–2.80)
Total reflections	327,986 (9,054)	836,593 (90,618)	471,301 (46,667)	1,558,834 (130,838)	491,555 (44,737)	403,003 (40,041)	119,141 (10,243)
Unique reflections	30,265 (2,663)	61,346 (6,518)	31,484 (3,132)	106,913 (10,467)	66,110 (6,579)	106,148 (10,537)	31,850 (3,104)
Multiplicity	10.8 (3.4)	13.6 (14.5)	15.0 (14.9)	14.6 (12.5)	7.4 (6.8)	3.8 (3.8)	3.7 (3.3)
Completeness, %	99 (88)	100 (96)	100 (100)	100 (99)	100 (100)	100 (100)	100 (99)
R_{meas}, %	12 (>100)	29 (>100)	6 (>100)	12 (>100)	8 (>100)	7 (>100)	12 (95)
R_{pim}, %	5 (70)	12 (>100)	2 (39)	6 (91)	3 (82)	4 (58)	6 (51)
Mean I/σ(I)	15.6 (1.1)	9.4 (1.7)	45.0 (2.5)	19.5 (1.1)	24.4 (1.1)	19.9 (1.3)	14.1 (1.5)
$CC_{1/2}$	(0.34)	(0.59)	(0.85)	(0.59)	(0.53)	(0.75)	(0.72)
Wilson B factor, $Å^2$	31.1	38.8	25.8	20.5	15.2	15.1	38.7
Refinement							
Protein copies per ASU	1	2	1	2	1	2	2
Resolution range, Å	47.66–2.27 (2.35–2.27)	44.08–2.23 (2.31–2.23)	31.40–2.30 (2.38–2.30)	44.76–1.90 (1.97–1.90)	36.10–1.75 (1.81–1.75)	40.07–1.83 (1.90–1.83)	40.47–2.80 (2.90–2.80)
Reflections used	28,017 (1,688)	65,771 (6,518)	30,588 (2,467)	94,674 (4,114)	58,484 (3,073)	94,792 (3,508)	28,172 (1,553)
Reflections for R_{free}	1,402 (85)	3,290 (326)	1,530 (124)	1,912 (83)	2,924 (153)	1,948 (72)	1,409 (78)
R_{work}, %	16.6 (28.3)	21.2 (34.0)	24.0 (31.2)	18.5 (27.9)	15.2 (23.8)	15.7 (23.7)	21.6 (34.0)
R_{free}, %	20.1 (31.0)	24.5 (34.3)	26.1 (34.1)	21.4 (33.6)	18.2 (27.8)	19.8 (25.1)	26.5 (44.9)
Nonhydrogen atoms	4,519	9,076	4,017	9,307	5,093	9,608	8,520
Protein	4,193	8,478	3,821	8,131	4,308	8,545	8,397
Glycans, ligands, ions	150	302	85	728	143	134	104
Water	176	295	111	448	642	929	19
Average B factor, $Å^2$	36.3	47.5	48.4	34.3	21.9	26.2	47.5
Protein	35.5	47.0	48.5	32.6	19.6	24.8	47.7
Glycans, ligands, ions	60.2	66.3	59.2	53.3	44.4	55.2	42.7
Water	35.0	43.0	37.3	34.4	32.1	34.4	18.3
Rmsd bond lengths, Å	0.006	0.002	0.004	0.006	0.011	0.008	0.003
Rmsd bond angles, °	0.75	0.60	0.75	0.90	1.03	0.89	0.52
Ramachandran favored, %	96.7	96.1	94.5	96.2	97.0	97.6	96.7
Ramachandran allowed, %	3.1	3.9	5.1	3.6	2.4	2.1	3.1
Ramachandran outliers, %	0.2	0	0.4	0.2	0.2	0.3	0.2
Rotamer outliers, %	0	0.8	0.7	0.3	0.6	0.6	0.4
Clash score	1.0	1.1	3.5	3.4	2.0	2.6	2.3

Values in parentheses are for the highest-resolution shell. ASU, asymmetric unit.

Table S2. Predicted impact of missense SNPsin AOAH

Probably benign									
S36L	V48L	V48M	V50A	I51V	L54F	N59D (glycan loss)		S60L	S60W
T61M (glycan loss)		Q63R	A64T	M66T	Y72N	P74L	P74T	F78L	L79M
I87N	K96Q	S99R	A100T	A104T	H109Y	L111Q	E112V	F113S	K115E
T118A	T118P	Q120R	P121Q	P121T	H124Q	L128P	E131K	Q138R	K139E
Q142K	S147F	S147Y	P148L	L150V	S153Y	P163L	A166T	M176V	Q178E
Q178H	S179A	K183E	V192I	F193S	W201R	D210N (glycan gain)		E211K	R218G
H225R	P238T	V242I	P243L	F248L	G251D	G251V	P254R	I257N	A264V
A266G	H267Y	I270V	S271C	T276A	S278L	M280V	N283K	L292F	L301V
G303Y	A304T	T311S	I314F	K317R	R322H	L323F	W324R	R345Q	I351L
S353G	L354S	Y362H	A364T	V366I	I367T	M370L	I371T	V375G	V375L
S377N	G378R	S380I	P382T	M386I	L392R	N395S	V396I	L400M	L407M
N409D (glycan loss)		L415F	Y416H	P419A	T422N	T422S	D426N	H429R	R431K
Q437L	K440E	M442V	A445V	S449F	F450V	G461C	G461S	S464F	T473A
K485R	L498V	Y500C	Y500H	M501V	A504S	H506R	E507G	G516D	P519L
L522F	E524Q	N532S	A535S	A540V	W544R	K546N	K546Q	Q548R	W551C
E558Q	P560L	D571H							
Uncertain impact									
Q56R	L69P	Y126C	R197Q	P219L	G235D	D237H	D240N	H269D	W274R
N287S	N287T	P289S (glycan gain)		E295D	E295K	D297V	R322C	R322S	D333H
N340H	N357Y	D361N	M397T (glycan gain)		R431S	G436S	Y444C	A445P	W462R
T468M (glycan loss)		Q479P	N497S (glycan gain)		E533K	G574R			
Probably harmful									
C41S (disulfide)		C108R (disulfide)		C123W (disulfide)		D186N (calcium binding)			R197W
R202Q	R202L	R202W	R204G	S263L (catalysis)		L296P	T305I	S318Y	R322P
R326I	N327S	C329R (disulfide)		R356G	D361Y	T387P	G436V	R470W	S474L
G517R	D527Y (catalysis)		D527V (catalysis)		H530Q (catalysis)		H530Y (catalysis)		

Human AOAH missense variants found in the dbSNP database are categorized according to their predicted impact on the enzyme's fold, stability, or function, with specific effects in parentheses. Predictions are based on the crystal structure as well as sequence conservation in vertebrates.