Supporting Information

McKee et al. 10.1073/pnas.1117686109

Materials and Methods.

Cloning, Expression, Mutagenesis, and Purification of *Hi*AXHd3. The gene encoding mature *Hi*AXHd3 (residues 28 to 558) was synthesized by Blue Heron Biotechnology Inc., with codons optimized for *Escherichia coli*; the translational stop codon was deleted and the gene was flanked with NcoI and XhoI restriction sites. The gene was inserted into the *E. coli* expression vector pET21(a) at the NcoI and XhoI restriction sites to generate pLM1. *Hi*AXHd3 was expressed in *E. coli* TUNER cells harboring pLM1; the recombinant bacterium was grown to midexponential phase ($A_{600} = 0.6$) and expression of *Hi*AXHd3 was induced with 0.2 mM isopropyl-β-D-thiogalactopyranoside and incubated for 16 h at 16 °C. The *E. coli* derived *Hi*AXHd3 construct contained a C-terminal His_6-tag and was thus purified from the *E. coli* cell-free extracts to electrophoretic homogeneity by immobilized metal ion affinity chromatography (IMAC) purification essentially as described previously (1), except that His select HF Nickel Affinity gel matrix (Sigma Aldrich) was used instead of Talon Fast Flow. The enzyme was generally eluted with 100 mM imidazole. For structural studies the protein was further purified by size exclusion chromatography using a HiLoad 16/60 Superdex 200 gel filtration column (GE Healthcare) as described previously (1). Site-directed mutagenesis of *E. coli*-derived *Hi*AXHd3 was carried out employing a PCR-based QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's instructions, using appropriate primers. Randomization of residues Gly183, Asn184, Tyr166, and Phe493 was achieved by carrying out four individual site-directed mutagenesis reactions in which the codon for the respective target amino acid was randomized. The four libraries (approximately 200 colonies per library) were screened for increased xylanase activity (compared to the Y166A variant) using a Blue-xylan overlay as described previously (2), except that the colony-containing agar plates were not preheated but overlayed with Blue-xylan agar and incubated at 37 °C for 8 h.

Data Collection and Structure Solution of *Hi*AXHd3. A seleno-methionine derivative of *Hi*AXHd3 (Se-Met *Hi*AXHd3), prepared as described in ref. 1, was crystallized at 20 mg/mL in 0.1 M Bistris (pH 6.5), 22% (wt/vol) PEG 3350, and 2 M ammonium sulfate. The D43A variant of the protein crystallized at 10 mg/mL in 0.1 M Na-Hepes (pH 7.5), 22.5% (wt/vol) PEG 4000, and 0.1 M NaCl and subsequently soaked with 20 mM ligand (Araf-Xylp3) for 2 h, while the Y166A/E216A mutant, also at 10 mg/mL, was crystallized in 0.1 M Bistris propane (pH 8.5), 20% (wt/vol) PEG 3500, and 0.2 M sodium fluoride. Crystals of *Hi*AXHd3 were cryoprotected using the reservoir solution supplemented with 20% (wt/vol) glycerol. Data for the Se-Met protein were collected on beam line I04 at Diamond Light Source, whereas data for the other proteins were collected at the Advanced Photon Source on beam line 22-ID or 22-BM. The diffraction images were integrated using iMosfilm, and the results were scaled and merged using SCALA (3). The space group was determined to be $P2_1$. Selenium positions and phases were determined using the SHELXC/D/E pipeline [recently reviewed (4)]. The resulting phases were used in model building using the Collaborative Computational Project No. 4 implementation of REFMAC/ARP-wARP (5). Minor manual corrections were performed using the program COOT (6) interspersed with cycles of maximum likelihood refinement using REFMAC (5) and PHENIX Refine (7). The structure of the *Hi*AXHd3 mutants D43A in complex with ligands, and apo Y166A/E216A, were determined by molecular replacement using Phaser (8) with the Se-Met form of wild-type *Hi*AXHd3 as the search model. Final data and structure quality statistics for these proteins are given in Table S2, which also provides the Protein Data Bank codes.

Generation of Ligands for Structural and Biochemical Studies. To generate the arabinooligosaccharides used in the experiment described in Fig. S2 sugar beet arabinan (Megazyme) was added to 0.5 M HCl to a final concentration of 20 mg/mL. The mixture was heated to 100 °C for 10 min, followed by 20 min at room temperature. To neutralize the solution, sodium hydroxide (NaOH) was added to a final molarity matching that of the acid. Samples were then analyzed by high performance anion exchange chromatography (HPAEC) and fractions where 25% of the glycosidic bonds were hydrolyzed were retained for further use. Unhydrolyzed arabinan was removed from these samples by ethanol precipitation using two to three volumes of the alcohol and incubation at 4 °C for 2–16 h. The soluble arabinooligosaccharide fraction was then lyophilized and resuspended in a small volume of deuterium oxide.

To generate oligosaccharide substrates for structural studies arabinoxylan (1 g) was dissolved in 40 mL of 50 mM sodium phosphate buffer (pH 7.0) to which was added 2 mL BSA (1 mg mL^{-1} final) and the GH10 xylanase *Cj*Xyn10A to a final concentration of 50 nM final. The reaction was incubated for 1 h. The digested products were separated by size exclusion chromatography using P2 Biogel (BioRad) matrix packed in a Glass Econo-Column (2.5 cm × 80 cm, total volume 500 mL, BioRad). The column was equilibrated with three column volumes of water and the mixture of xyloarabinooligosaccharides generated by enzyme treatment were lyophilized, resuspended in 2 mL of water, and loaded onto the column. The column was run by gravity flow and 1 mL fractions were collected after the void volume. Aliquots of the fractions were subjected to HPAEC before and after treated with *Hi*AXHd3 to identify those that contain substrates for the enzyme. Fractions with the simplest oligosaccharide profile were freeze dried and used in crystal soaking experiments.

1. Charnock SJ, et al. (2000) The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain. *Biochemistry* 39:5013–5021.
2. Andrews SR, et al. (2004) The use of forced protein evolution to investigate and improve stability of family 10 xylanases. The production of Ca2þ-independent stable xylanases. *J Biol Chem* 279:54369–54379.
3. Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D* 62:72–82.
4. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64:112–122.
5. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* 53:240–255.
6. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D* 60:2126–2132.
7. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* 66:213–221.
8. McCoy AJ (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D* 63:32–41.

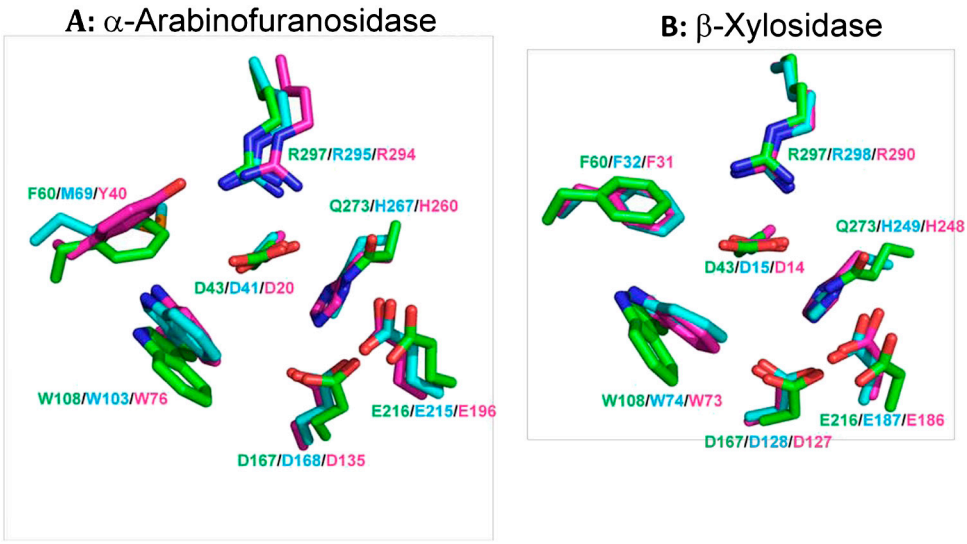

Fig. S1. Structural conservation within the active site of GH43. The key residues, shown in stick format at the −1 subsite of *Hi*AXHd3 (carbon atoms shown in green) are overlayed with the following enzymes. (*A*) α-L-arabinofuranosidases *Cj*Abf43A [Protein Data Bank (PDB), 3QED; carbons in cyan] and *Sa*Araf43A (PDB, 3AKG; carbons in pink). (*B*) β-D-xylosidases XynB3 (PDB, 2EXK; carbons in cyan) and *Sr*Xyn43A (PDB, 3C2U; carbons in pink).

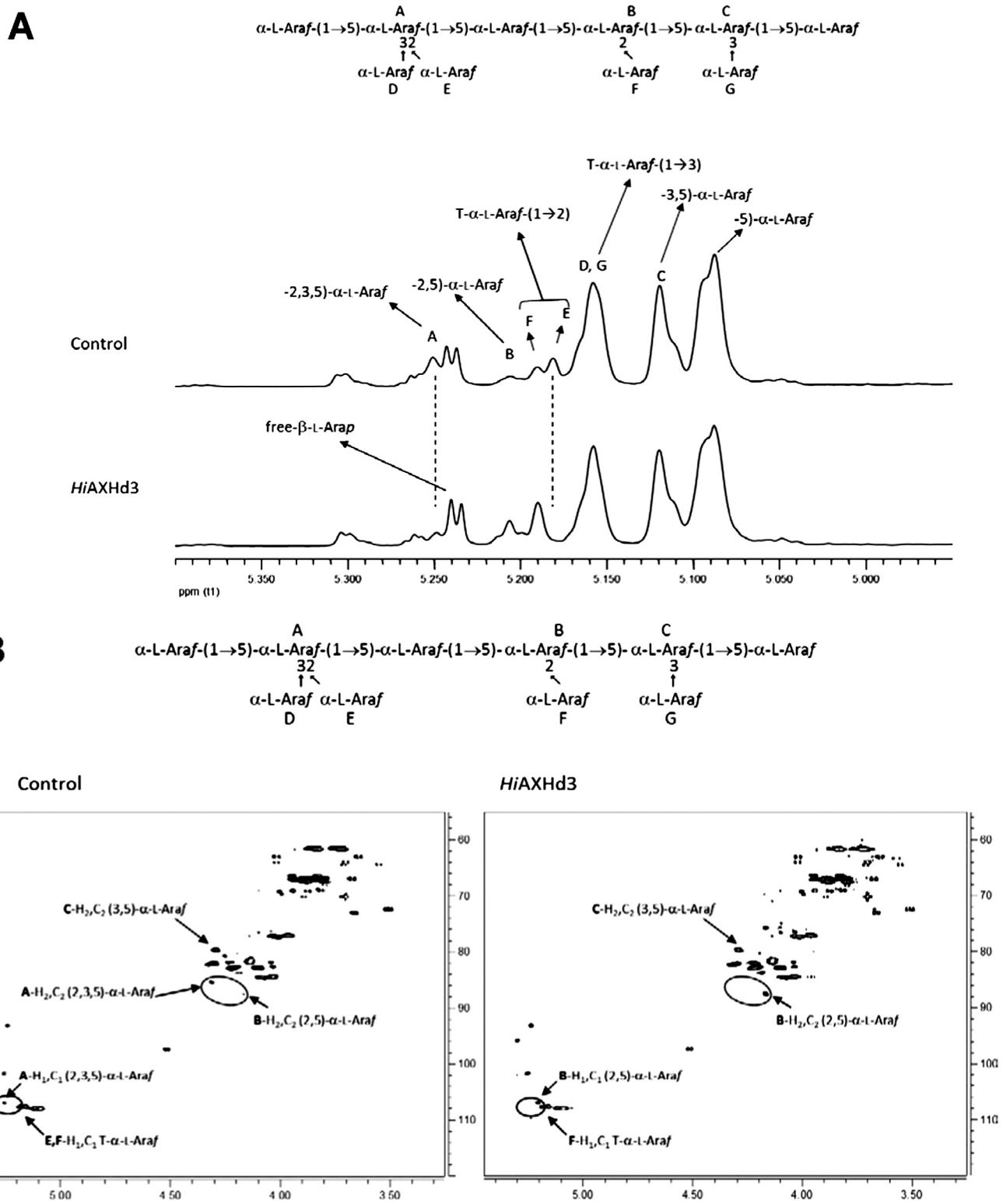

Fig. S2. NMR analysis of enzyme treated arabinan. (*A*) displays partial ^{1}H-NMR spectra of sugar beet arabinan-derived oligosaccharides incubated without (control) and with *Hi*AXHd3 (100 nM) for 90 min. The letters in the spectra correspond to residues in the structures in the top of both *A* and *B*. The structures show the types of substitutions present in the substrates. (*B*) Shows 2D NMR (gHSQC) of the two samples. The disappearance of *A* and the increase in *B* shows that *Hi*AXHd3 attacks the O3-linked Ara*f* in double substitutions (at O2 and O3) in arabinan.

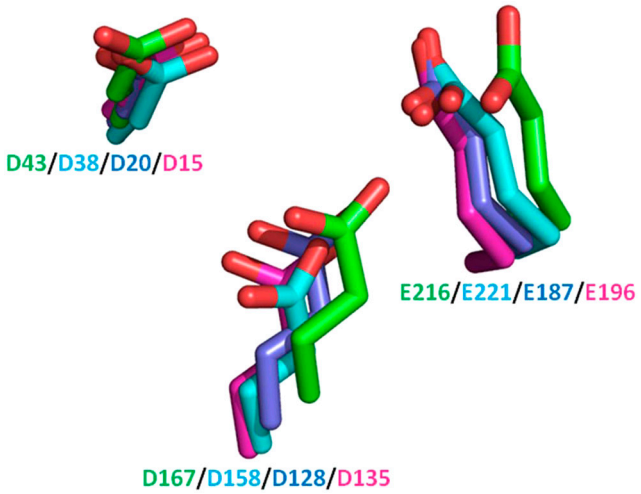

Fig. S3. Conservation of the catalytic residues with GH43. The catalytic residues, shown in stick format, of *Hi*AXHd3 (carbon atoms shown in green) are overlayed with the β-D-xylosidase XynB3 [Protein Data Bank (PDB), 2EXK; carbons in pink], the α-L-arabinofuranosidase *Sa*Araf43A (PDB, 3AKG; carbons in slate blue), and the arabinanase *Cj*Arb43A (PDB, 1GYD; carbons in cyan). The catalytic base, catalytic acid, and pKa modulator of the catalytic acid in *Hi*AXHd3 are D43, E216, and D167, respectively.

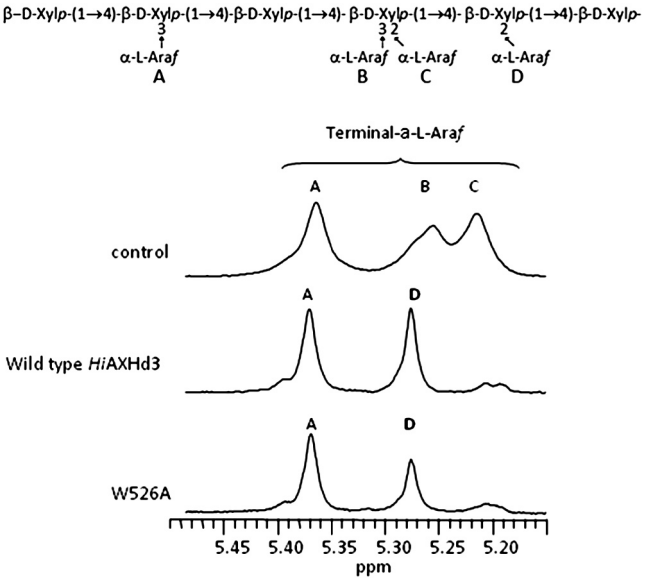

Fig. S4. 1D NMR of *Hi*AXHd3 wild type and W526A variant wheat arabinoxylan reaction products. In the schematic, linkages within the polysaccharide (A–D) are labeled to correspond with peaks in the NMR spectra. The control spectrum is arabinoxylan that has not been subjected to enzymatic digestion. In the spectrum of wheat arabinoxylan incubated with wild-type *Hi*AXHd3 (or the Y166A mutant) peaks A and D are roughly equal in size (ratio 1:0.84), because cleavage of the double substitution has generated single O2 substitutions that are almost as numerous as the preexisting single O3 substitution. In the spectra of wheat arabinoxylan incubated with the W526A mutant of *Hi*AXHd3 the same profile is observed but the ratio of peak size for A and D is 1:0.47. This indicates that cleavage of the double substitution by the mutant has similar quantities of additional single O3 and O2 substitutions in the arabinoxylan polysaccharide.

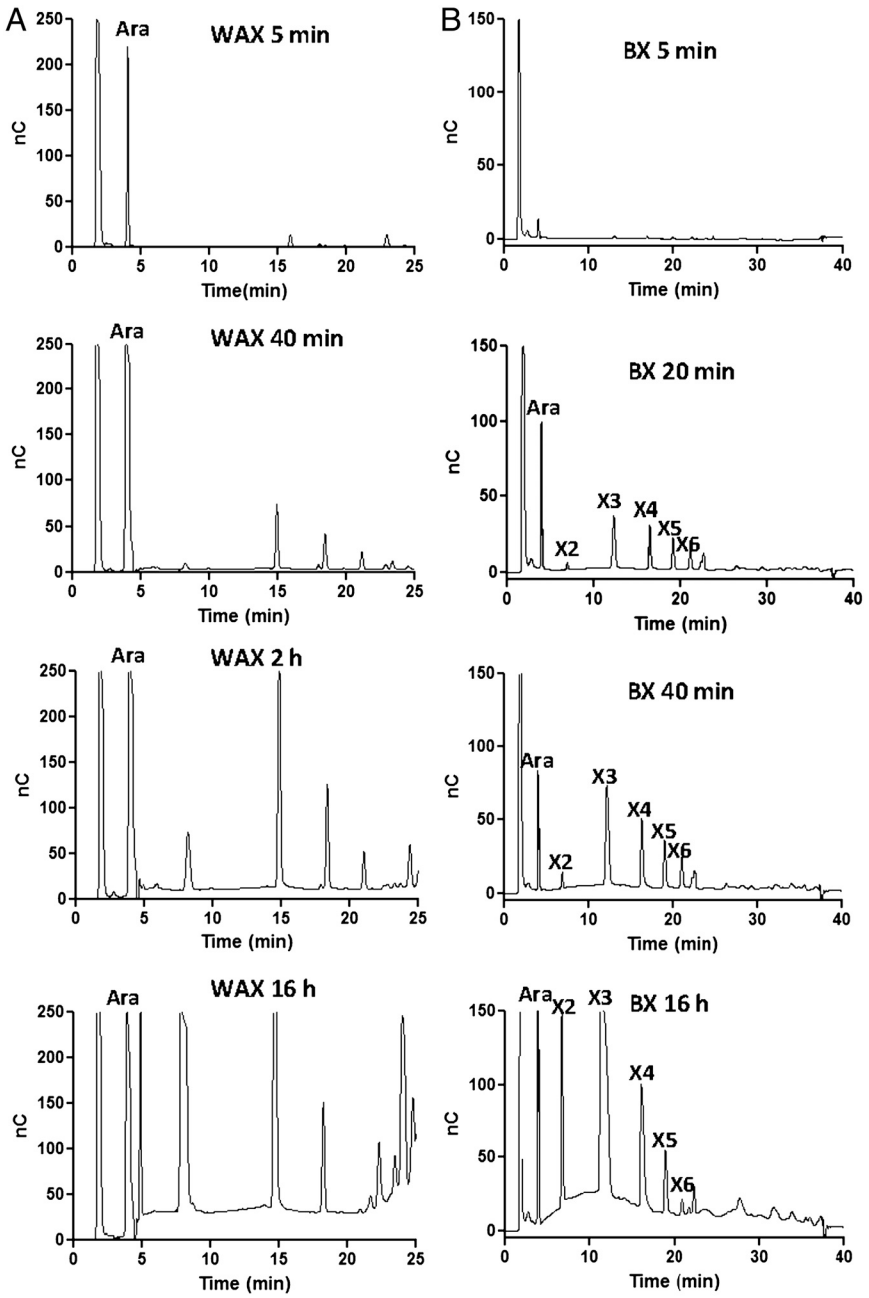

Fig. S5 HPAEC analysis of xylan hydrolysis by the Y166A mutant. Birchwood xylan (BX) and wheat arabinoxylan (WAX) at 4 mg/mL were incubated with 10 nM of the Y166A mutant in 50 mM sodium phosphate buffer (pH 7.0). At regular time intervals aliquots of the WAX-Y166A (WAX) and BX-Y166A (BX) were analyzed by HPAEC. Peaks, identified through comigration with standards, were labeled appropriately; Ara, arabinose; X2, xylobiose; X3, xylotetraose, X5, xylopentaose, X6, xylohexaose.

Table S1. The catalytic activity of wild type and variants of *Hi*AXHd3

	Wheat arabinoxylan			Sugar beet arabinan		
	k_{cat} (min^{-1})	K_M(M)	k_{cat}/K_M (min^{-1} M^{-1})	k_{cat} (min^{-1})	K_M (M)	k_{cat}/K_M (min^{-1} M^{-1})
Wild type	$1.0(\pm0.13)\times10^{4}$	$3.5(\pm1.4)\times10^{-4}$	3.0×10^{7}	$1.7(\pm0.1)\times10^{3}$	$1.3(\pm0.3)\times10^{-4}$	1.4×10^{7}
D43A		Inactive			Inactive	
D167A		Inactive			Inactive	
E216A		Inactive			Inactive	
H272A	11.0 ± 2.35	$3.9(\pm2.7)\times10^{-4}$	2.9×10^{4}	3.2 ± 1.0	$2.8(\pm0.3)\times10^{-4}$	1.2×10^{4}
F289A	25.7 ± 2.7	$2.0(\pm0.7)\times10^{-4}$	1.3×10^{5}	15.5 ± 2.4	$1.1(\pm0.8)\times10^{-4}$	1.4×10^{5}
T232A	17.7 ± 2.8	$8.1(\pm5.3)\times10^{-5}$	2.2×10^{5}	4.7 ± 1.1	$2.3(\pm0.4)\times10^{-5}$	2.0×10^{5}
T232A/H272A/F289A		Inactive			Inactive	
S58A	$7.1(\pm1.4)\times10^{2}$	$2.6(\pm2.0)\times10^{-4}$	2.7×10^{6}	75.8 ± 6.4	$3.1\pm1.2\times10^{-5}$	2.5×10^{6}
R297A		Inactive			Inactive	
Q273A		Inactive			Inactive	
D291A		Inactive			Inactive	
W108A		Inactive			Inactive	
W526A	25.4 ± 6.6	$1.9(\pm1.7)\times10^{-4}$	1.4×10^{5}	40.4 ± 4.8	$1.2(\pm0.4)\times10^{-3}$	3.4×10^{4}
Y166F	n.d.	n.d.	$2.9(\pm0.9)\times10^{3}$		Inactive	
Y166W	16.6 ± 2.5	$1.5(\pm0.4)\times10^{-5}$	1.1×10^{6}	n.d.	n.d.	$2.1(\pm0.4)\times10^{3}$

n.d., not determined.

Table S2. Crystallographic data and refinement statistics

HiAraF		Wildtype Se-Met	D43A/Ara*f*-Xylp_3	Y166A/E216A
Data collection				
	Space group	$P2_1$	$P2_1$	$P2_1$
	Wavelength (Å)	0.9795	1.0000	1.0000
	Unit cell (Å)	95.32 78.26 95.77 $\beta = 103.13$	65.38 52.97 146.87 101.69	65.31 83.93 97.89 102.81
	Resolution (Å)	47.75-1.85 (1.90–1.85)	33.68-1.44(1.47–1.44)	1.87-48.27(1.89–1.87)
	No. observations	820,508 (100,003)	426,983 (61,153)	443,472 (75,433)
	unique	80,563 (11,718)	177,611 (25,617)	89,446 (13,300)
	Multiplicity	10.2 (8.5)	2.4 (2.4)	4.8 (4.2)
	Anomalous multiplicity	5.1 (4.3)	N/A	N/A
	completeness	100 (100)	99.2 (98.3)	99.7 (99.3)
	anomalous completeness	99.4 (100)	N/A	N/A
	Average I/sigma I	13.8 (5.1)	16.2 (9.8)	14.7 (8.6)
	R_{sym}	0.121 (0.463)	0.084 (0.172)	0.097 (0.186)
Refinement				
	Number of reflections	76,187	167,521	82,744
	R_{cryst} (%)	15.13	16.67	14.16
	R_{free} (%)	18.47	20.81	18.70
	No. of protein atoms	8,406	8,382	8,414
	No. of solvent atoms	847	1,081	1,108
	No. of ligand atoms	N/A	74	N/A
rmsd				
	Bond lengths (Å)	0.011	0.012	0.017
	Bond angles (°)	1.290	1.408	1.555
Ramachandran				
	Favored (%)	96.21	96.11	95.7
	Allowed (%)	4.79	3.89	4.00
	Disallowed (%)	0.00	0.00	0.30
B-factors				
	Wilson B ($Å^2$)	16.09	8.08	15.9
	Main chain ($Å^2$)	16.52	9.36	16.26
	Side chain ($Å^2$)	19.65	10.59	22.10
	Water ($Å^2$)	17.77	25.35	30.38
	Ligand ($Å^2$)	N/A	25.30	N/A
PDB codes		3ZXJ	3ZXK	3ZXL

PDB, Protein Data Bank.