Supporting Information

Structure-guided reprogramming of a hydroxylase to halogenate its small molecule substrate

Andrew J. Mitchell[1], Noah P. Dunham[1], Jonathan A. Bergman[1], Bo Wang[2], Qin Zhu[3], Wei-chen Chang[2†], Xinyu Liu[3*], Amie K. Boal[1,2*]

[1]Department of Biochemistry and Molecular Biology and [2]Department of Chemistry, The Pennsylvania State University, University Park PA 16802; [3]Department of Chemistry, University of Pittsburgh, Pittsburgh PA 15260

Corresponding Author

Correspondence should be addressed to or X.L. (xinyuliu@pitt.edu) or A.K.B. (akb20@psu.edu).

Present Address

[†]Department of Chemistry, North Carolina State University, Raleigh, NC 27695

Structural-based identification of SadA as a halogenation scaffold

A search of the Protein Data Bank (PDB) for structural homologs of WelO5 (PBD accession code 5IQS) was accomplished by submission of coordinates to the DaliLite structural alignment online server.[1] The SadA structure (PDB accession code 3W21) emerged as the top hit with a *Z*-score of 17.6 (**fig. S2**). Functionally-related carrier-dependent Fe/2OG halogenases appeared as well but with significantly lower *Z*-scores (9-11 range). Pairwise sequence alignments were performed with Clustal Omega[2] and secondary structure matching (SSM) superpositions were carried out with the PDBeFold server.[3] The PDB files were analyzed with Coot[4] and figures were generated with the PyMOL Molecular Graphics Software Package (Schrödinger LLC).

Expression and purification of SadA

The *sadA* gene was codon-optimized for heterologous overexpression in *Escherichia coli* and purchased as a synthetic construct (GeneArt/Life Technologies). The sequence of SadA is shown below with the start codon underlined and highlighted in boldface.

GGATCC**ATG**CAGCATACCTATCCGGCACAGCTGATGCGTTTTGGCACCGCAGCAC
GTGCAGAACATATGACCATTGCAGCAGCAATTCATGCACTGGATGCAGATGAAGCA
GATGCAATTGTTATGGATATTGTTCCGGATGGTGAACGTGATGCATGGTGGGATGA
TGAAGGTTTTAGCAGCAGCCCGTTTACCAAAAATGCACATCATGCAGGTATTGTTG
CAACCAGCGTTACCCTGGGTCAGCTGCAGCGTGAACAGGGTGATAAACTGGTTAG
CAAAGCAGCAGAATATTTTGGTATTGCCTGCCGTGTTAATGATGGTCTGCGTACCA
CCCGTTTTGTTCGTCTGTTTAGTGATGCCCTGGATGCCAAACCGCTGACCATTGGT

CATGATTATGAAGTTGAATTTCTGCTGGCAACCCGTCGTGTTTATGAACCGTTTGAA
GCACCGTTTAACTTTGCACCGCATTGTGATGATGTTAGCTATGGTCGTGATACCGT
TAATTGGCCTCTGAAACGTAGCTTTCCGCGTCAGCTGGGTGGTTTTCTGACCATTC
AGGGTGCAGATAATGATGCCGGTATGGTTATGTGGGATAATCGTCCGGAAAGCCG
TGCAGCGCTGGATGAAATGCATGCAGAATATCGTGAAACCGGTGCAATTGCCGCA
CTGGAACGTGCAGCCAAAATCATGCTGAAACCGCAGCCTGGCCAGCTGACACTGT
TTCAGAGCAAAAATCTGCATGCCATTGAACGTTGTACCAGCACCCGTCGTACCATG
GGTCTGTTTCTGATTCATACCGAAGATGGTTGGCGTATGTTTGATTGAGCGGCCGC

The coding region was amplified and cloned into the pQTEV expression vector using standard molecular biology procedures. The primers for PCR amplification are shown below with the BamHI/NotI restriction sites underlined.

Forward: 5’ TAGGATCCATGCAGCATACCTATCCGGCACA 3’

Reverse: 5’ GCGGCCGCAATCAAACATACGCCAATCTT 3’

PCR procedures were performed on a Bio-Rad S1000 thermal cycler. Sanger sequencing reactions were carried out at the Genomics Core Facility in the Huck Institutes of the Life Sciences at the Pennsylvania State University. DH5α *E. coli* cells were transformed with pQTEV/SadA and the sequences of plasmids that were isolated from successful transformants were confirmed with Sanger sequencing.

The D157G and G179S mutations were introduced consecutively via inverse PCR[5] using the following primers. The sequences of the resulting plasmids were verified by Sanger sequencing.

Forward (D157G): 5' GGCGATGTTAGCTATGGTCGTGATACCG 3'

Reverse (D157G): 5' ACAATCCGGTGCAAAGTTAAACG 3'

Forward (G179S): 5' AGCGGTTTTCTGACCATTCAGGGTG 3'

Reverse(G179S): 5' CAGCTGACGCGGAAAGCTAC 3'

BL21(DE3) *E. coli* cells were transformed with pQTEV containing wt and variant *sadA* sequences for protein overexpression. A 1 L culture of LB medium containing 100 µg/mL ampicillin was inoculated with 30 mL of an overnight starter culture and allowed to grow with shaking at 180 rpm at 37 ^{o}C to an OD_{600} of 0.6-0.8. Protein expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) followed by cold shock of the medium and subsequent growth with shaking at 180 rpm (18 ^{o}C for 18 hours). Cultures were harvested by centrifugation and the resulting cell paste (~10 g/L) was stored frozen at -80 ^{o}C.

Frozen cell paste was resuspended in lysis buffer (100 mM HEPES pH 7.5, 20 mM imidazole) containing 1 mg/mL lysozyme and 1 mM phenylmethylsulfonyl fluoride. Cells were lysed via sonication (QSonica instrument, 60% power, 30 sec pulse time with 45 sec rest interval) and the cell debris was pelleted via centrifugation at 22,000*g*. The resulting lysate was applied to Ni-nitrilotriacetic acid (NTA) agarose resin (MC Labs)

that was pre-equilibrated in lysis buffer. Bound proteins were eluted in 100 mM HEPES pH 7.5, 250 mM imidazole. The fractions containing SadA, as assessed by SDS-PAGE, were concentrated to 3-5 mM using an Amicon 10K MWCO centrifugal filter device and dialyzed against 100 mM HEPES pH 7.5 containing 5 mM ethylenediamine tetraacetic acid (EDTA) to chelate any adventitiously-bound metal ions. Inductively-coupled plasma atomic emission spectroscopic (ICP-AES) analysis shows 0.018 equivalents of iron and 0.015 equivalents of zinc bound to the protein after EDTA treatment. Two subsequent rounds of dialysis were performed against 100 mM HEPES pH 7.5 to remove EDTA. The purified protein was rendered anoxic via repetitive cycles of vacuum and purge with $Ar_{(g)}$. All subsequent manipulations were performed in an oxygen-free glovebox (Coy Laboratory Products). The final protein concentration was determined by UV-visible spectrophotometry using a calculated ε_{280} of 39,545 $M^{-1}cm^{-1}$. SadA variant proteins were overexpressed and purified via the same procedure.

Synthesis of *N*-succinyl-L-leucine

Synthetic methods were adapted from previously published procedures in Sakai et al.[6] L-leucine (1.31 g, 0.1 mol) and succinic anhydride (1.20 g, 0.12 mol) were added to glacial acetic acid and allowed to stir for 16 hours at 55°C under an N_2 atmosphere. The reaction mixture was concentrated by rotary evaporation and subjected to silica gel chromatography using CH_2Cl_2/CH_3OH (9:1) as the mobile phase. The product was further purified by HPLC on an Agilent Pursuit 10 C18 (10 x 250 mm, 10 µm particle size) column with a mobile phase composed of 85% solvent A (0.1% TFA in H_2O) and 15% solvent B (CH_3CN). Fractions containing the product were pooled and dried over

N_2. ^{1}H NMR (500 MHz, D_2O): δ 4.34 (t, 1H), 2.61 (t, 2H), 2.56 (t, 2H), 1.63 (m, 3H), 0.89 (d, 3H), 0.84 (d, 3H).

SadA enzymatic activity assays

All solutions were degassed to remove oxygen, cycled into an anoxic glovebox and used under those conditions unless stated otherwise. SadA-catalyzed reactions were carried out in 100 mM HEPES buffer, pH 7.5, with 100 μM enzyme in a reaction mixture that included 100 μM $Fe(NH_4)_2(SO_4)_2$, 2 mM 2OG, 1 mM **1**, and 50 mM NaCl. Reactions were prepared in a total volume of 200 μL under anoxic conditions prior to removal from the glovebox. The solution was mixed with an equal volume of oxygen-saturated buffer to initiate turnover. After a 3-hr incubation at room temperature, the enzyme was removed using a 10K MWCO spin filter (Pall). In assays for bromination activity, NaBr was substituted for NaCl at the same concentration. The reactions were analyzed using an Agilent 1200 series HPLC attached to an Agilent 6410 LC-MS and equipped with a Zorbax Extend-C18 1.8 μm particle size 4.6 x 50 mm column. Isocratic elution of the products was carried out by using an 85:15 ratio of 0.1% formic acid in water (solvent A) to acetonitrile (solvent B) as the mobile phase, pumped at a rate of 0.8 mL/min. All mass spectra were analyzed with the MassHunter software package (Agilent Technologies). A control reaction lacking enzyme revealed a strong peak associated with **1** and no evidence of production of **2** or **3**, consistent with our peak assignments. Assays to probe the origin of uncoupling in the D157G SadA variant were modified to include 1 μM enzyme, 5 μM $Fe(NH_4)_2(SO_4)_2$, 1 mM **1**, 2 mM 2OG, and 0-5 mM ascorbate. Samples were analyzed both prior to oxygen exposure and at 1, 2, 5, and 24

hr time points. All assays were performed in triplicate and a representative experiment is shown in **fig. S5**.

To evaluate the relationship between substrate consumption and succinate production, wt SadA and its D157G variant were assayed under single turnover conditions. Various concentrations of 2OG (0-160 μM) were tested in reactions containing 250 μM SadA, 200 μM $Fe(NH_4)_2(SO_4)_2$, 500 uM **1**, 50 mM NaCl, 250 μM d_4-succinate, 1.25 mM glycine, 50 mM HEPES pH 7.5. Glycine and d_4-succinate were included as an external and internal standard, respectively. All solutions were prepared anaerobically prior to mixing with an equal volume of O_2-saturated buffer to initiate turnover and quenched with formic acid after incubation at room temperature for 15 minutes. The resulting samples were filtered to remove the enzyme and analyzed via LC-MS. The elution method was modified to include a gradient from 95% to 60% solvent A over the first 10 minutes followed by a wash step at 40% solvent A for 5 min and a 10 min equilibration at 95% solvent A prior to the next injection. Mass chromatograms were analyzed with MassHunter and the plot and equations shown in **fig. S4** were generated with the KaleidaGraph software package.

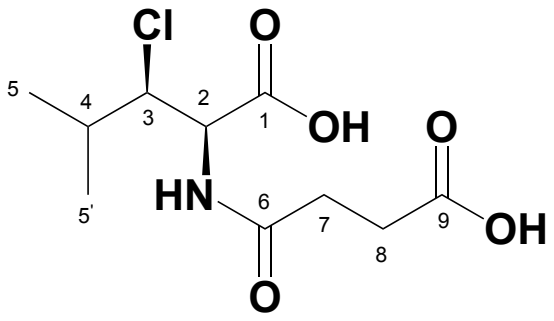
Purification of reaction products

A solution of 100 μM SadA D157G/G179S, 100 μM $Fe(NH_4)_2(SO_4)_2$, 10 mM **1**, 20 mM 2OG, and 100 mM NaCl was prepared under anoxic conditions and the reaction volume was brought to 12 mL with air-saturated buffer to initiate the reaction and incubated under a steady stream of compressed air. An additional aliquot of 100 μM Fe^{II}-SadA

was added every two hours until the total concentration of SadA reached 400 μM. The reaction was incubated overnight at room temperature and the enzyme removed using a 10K MWCO centrifugal filter device. The hydroxylated product was prepared similarly in a 3 mL reaction with wt SadA but with no extra addition of enzyme at later time points. The filtered reaction products were dried and dissolved in water for HPLC purification. The products were isolated using an Agilent 1260 Infinity series HPLC equipped with a Zorbax SB-C18 5 μm particle size 4.6 x 150 mm column. The chlorination product was purified by isocratic elution with a 90:10 ratio of solvent A to solvent B as the mobile phase, pumped at 2 mL/min, and monitored by UV absorbance at 210 nm with the OpenLAB software package (Agilent Technologies). The hydroxylated product was purified similarly but with a 93:7 mixture of the mobile phase components. Under these conditions, **2** eluted at 3 minutes and **3** eluted at 6 minutes. The purified products were collected and dried under N_2 for NMR analysis.

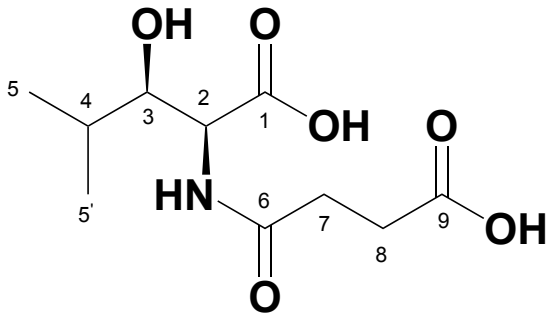
NMR characterization

NMR spectra were recorded on a Bruker AV-3-HD-500 instrument and calibrated using residual solvent peaks as an internal reference. Multiplicities are recorded as: d = doublet, dd = doublet of doublets, m = multiplet. ^{1}H, ^{13}C, COSY, and HSQC NMR spectra are provided for **2** and **3** appended to the end of the Supporting Information document. The full mass spectrum of purified **2** and **3** (obtained over a range of 50-500 *m/z*) is shown in **figs. S6** and **S7**.



^{1}H NMR (500 MHz, D_2O) δ 4.88 (d, *J* = 2.7 Hz, 1H, H-2), 4.28 (dd, *J* = 9.3, 2.7 Hz, 1H, H-3), 2.78 – 2.59 (m, 4H, H-7 & H-8), 1.98 – 1.86 (m, 1H, H-4), 1.09 (d, *J* = 6.6 Hz, 3H, H-5), 0.96 (d, *J* = 6.7 Hz, 3H, H-5'). (1H_compound_3)

^{13}C NMR (126 MHz, D_2O) δ = 176.99 (C-1), 174.89 (C-9), 166.84 (C-6), 69.57 (C-3), 56.25 (C-2), 32.35 (C-4), 30.05, 29.25 (C-7 & C-8), 19.64 (C-5), 18.87 (C-5'). (13C_compound_3)



^{1}H NMR (500 MHz, D_2O) δ 4.62 (d, *J* = 2.4 Hz, 1H, H-2), 3.77 (dd, *J* = 9.1, 2.5 Hz, 1H, H-3), 2.75 – 2.54 (m, 4H, H-7 & H-8), 1.73 – 1.60 (m, 1H, H-4), 0.99 (d, *J* = 6.6 Hz, 3H, H-5), 0.86 (d, *J* = 6.7 Hz, 3H, H-5'). (1H_compound_2)

^{13}C NMR (126 MHz, D_2O) δ = 177.02 (C-1), 174.94 (C-9), 167.15 (C-6), 77.01 (C-3), 55.62 (C-2), 30.48 (C-4), 30.17, 29.31 (C-7 & C-8), 18.28 (C-5), 18.18(C-5'). (13C_compound_2)

Table S1. Single ion monitoring (SIM) parameters for the chromatograms depicted in Fig. 2.

Compound	**Ion**	***m/z***	**elution time (min)**
Succinate	$[succ.-H]^-$	117.1	0.91
1	$[\mathbf{1}-H]^-$	230.1	3.43
2	$[\mathbf{2}-H]^-$	246.1	1.38
3	$[\mathbf{3}-H-CO_2-Cl]^-$	184	5.52
4	$[\mathbf{4}-H-CO_2]^-$	264	6.88

Table S2. SIM parameters for the chromatogram depicted in Figure 3a.

compound	Ion	*m/z* value (ratio)[a]
3	[**3**-H]$^-$	264/266 (2.81)
	[**3**-H-CO_2]$^-$	220/222 (2.80)
	[**3**-H-CO_2-Cl]$^-$	184
4	[**4**-H]$^-$	308/310 (1.10)
	[2*(**4**-H)-CO_2]$^{2-}$	286/288 (1.01)
	[**4**-H-CO_2]$^-$	264/266 (1.01)

[a]The *m/z* + 2 values for the stable and naturally-occurring heavier isotope of the halogen are also reported, when applicable, along with the ratio of *m/z* : *m/z* + 2 values. This ratio corresponds to the relative abundance of ^{35}Cl/^{37}Cl or ^{79}Br/^{81}Br pairs in compounds **3** and **4**. For comparison, calculated ratios are 3.0 for Cl and 1.0 for Br.

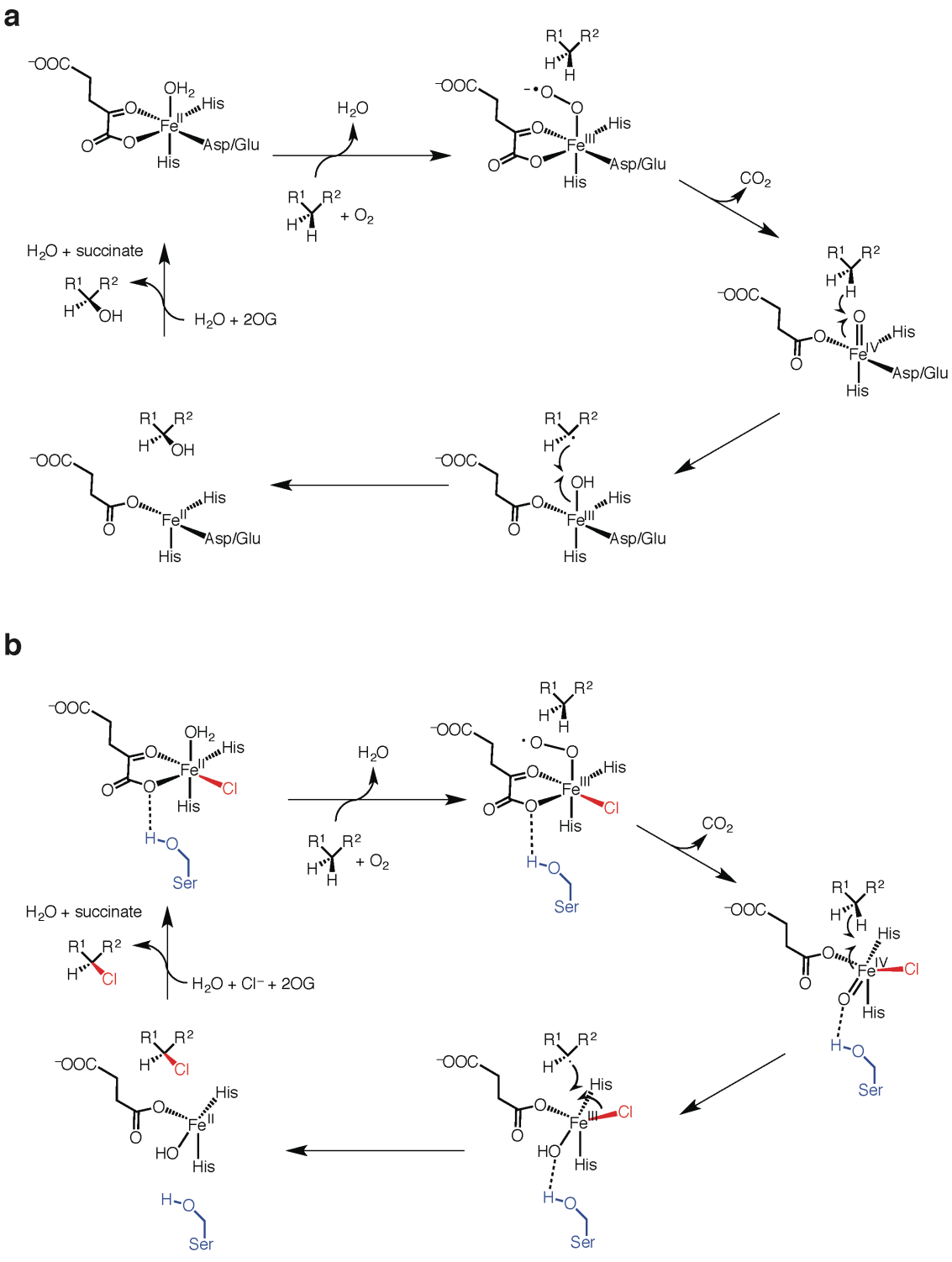


Figure S1. General mechanistic scheme for Fe/2OG hydroxylases (**a**) and carrier-protein-independent halogenase WelO5 (**b**).

Z-score	protein name (PDB accession code)	% sequence ID
17.8	**SadA (3W20)**	**19**
14.4	PHKD-like (3DKG)	17
14.1	P4H (4IW3)	15
14.1	P4H (4J25)	15
13.8	HIF (4BQW)	14
13.7	SnoK (5EPA)	17
…	…	…
10.7	**SyrB2 (2FCT)**	**10**
11.4	CurA-HAL (3NNF)	15
9.9	CytC3 (3GJA)	10

Figure S2. Dali alignment results after query with WelO5 (PDB accession code WIQS). The top six hits are listed at the top of the table and the SadA protein investigated in this work is shown in boldface. Other structurally characterized halogenases (example SyrB2 is shown in boldface) are listed after the break denoted by an ellipsis.

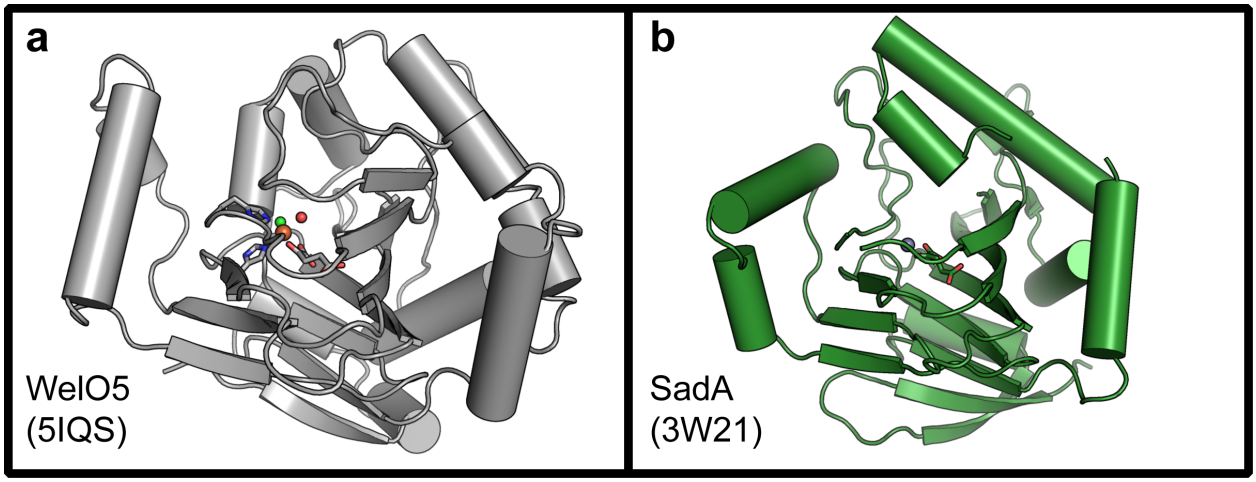


Figure S3. Comparison of overall topology in WelO5 (**a**) and SadA (**b**).

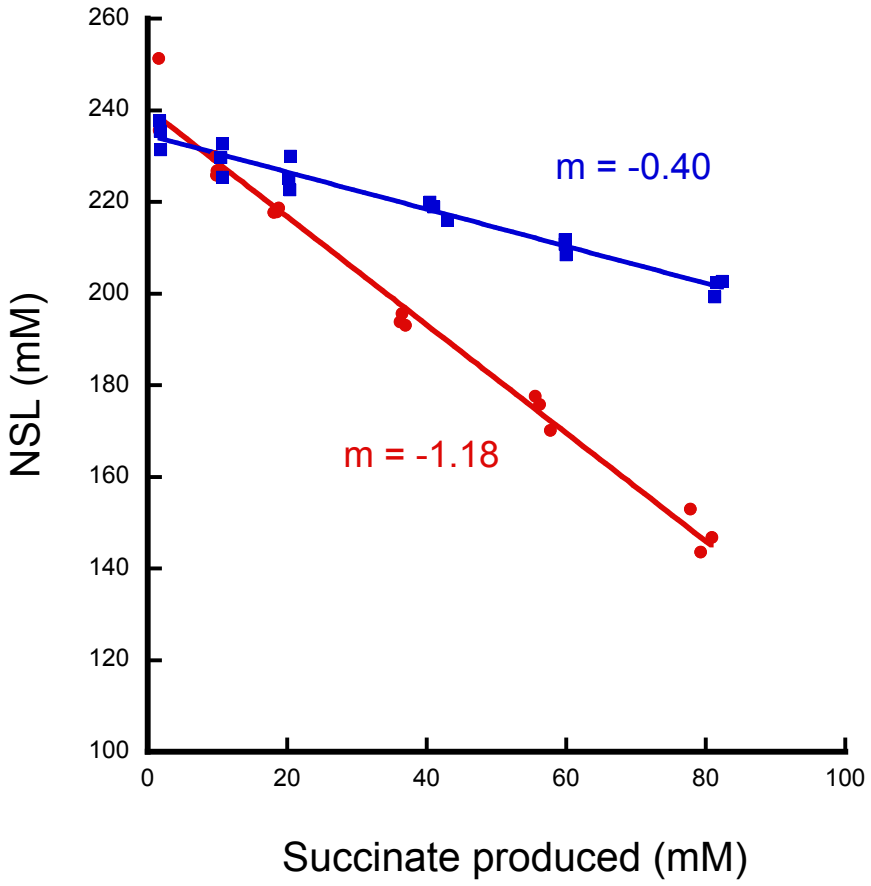


Figure S4. A comparison of succinate production and *N*-succinyl-Leu (NSL) consumption for wt SadA (red) and D157G SadA (blue). Coupling ratios were obtained by varying the amount of the 2OG reactant present. All experiments were performed in triplicate and the resulting datapoints were fit to a linear equation with slope (m) stated in the plot. The wt enzyme fully couples 2OG processing to substrate consumption (slope ~ -1) while the D157G variant exhibits a coupling ratio of only 0.4. The variant protein inactivates ~60% of the catalyst pool on every turnover.

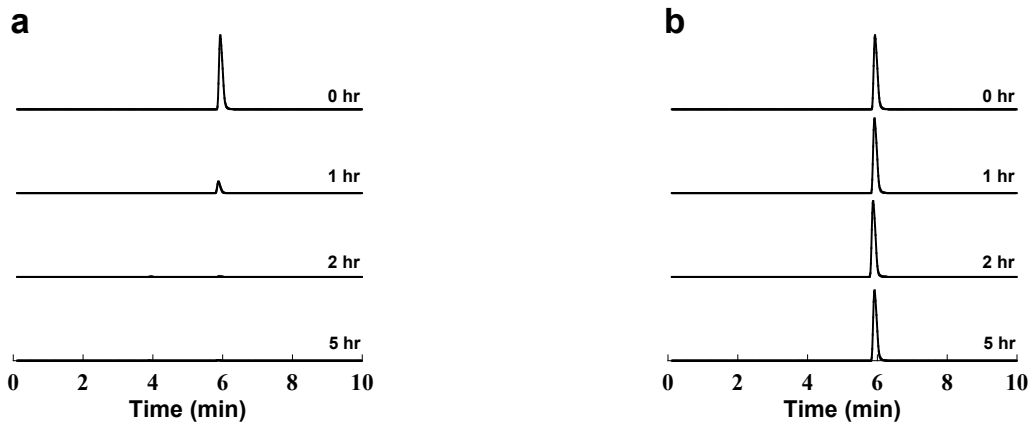


Figure S5. Single ion monitoring trace for consumption of **1** (230.1 *m/z*, peak at 6 min) by wt SadA (**a**) and D157G SadA (**b**) with 1 μM enzyme and 1 mM substrate in the presence of ascorbate. The wt enzyme accomplishes ~1000 turnovers within 2 hours and D157G SadA undergoes only 2-3 turnovers in the same period of time. The variant enzyme exhibits little change in substrate peak height at later time points, consistent with rapid uncoupling that cannot be rescued by addition of an Fe(III) reductant. Turnover numbers were obtained by averaging over at least three individual experiments. Similar metrics are difficult to quantify in other Fe/2OG halogenases because of the requirement for a carrier protein and the challenge in obtaining substrates, but initial measurements for WelO5 indicate that the system does not rapidly self-inactivate and is capable catalyzing ~75 turnovers (see ref. 12 in the main text).

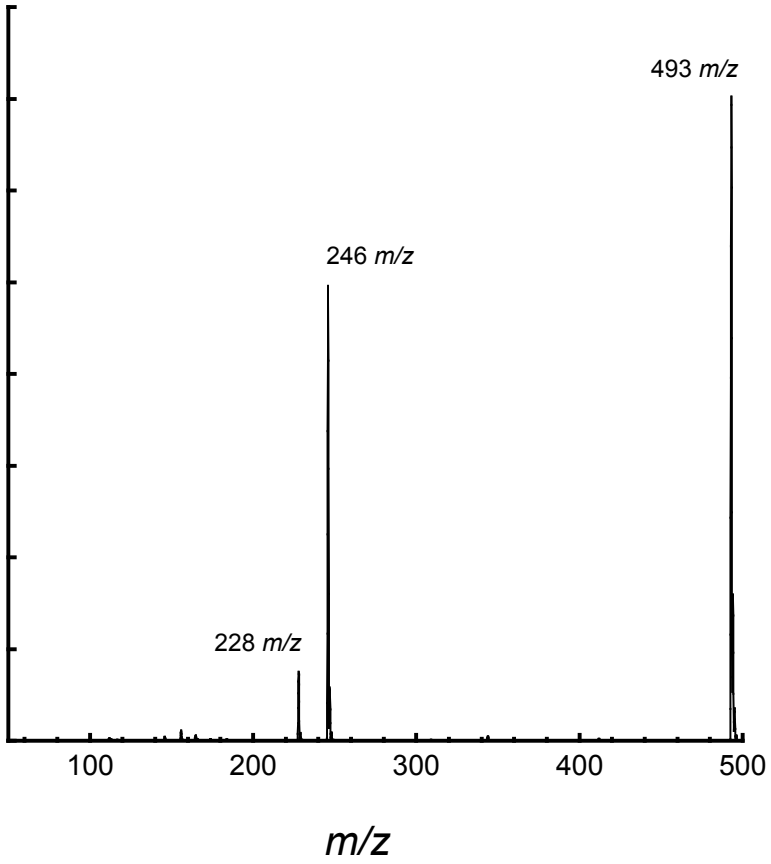


Figure S6. Full mass spectrum of HPLC purified **2** obtained in negative ion mode over 50-500 *m/z*.

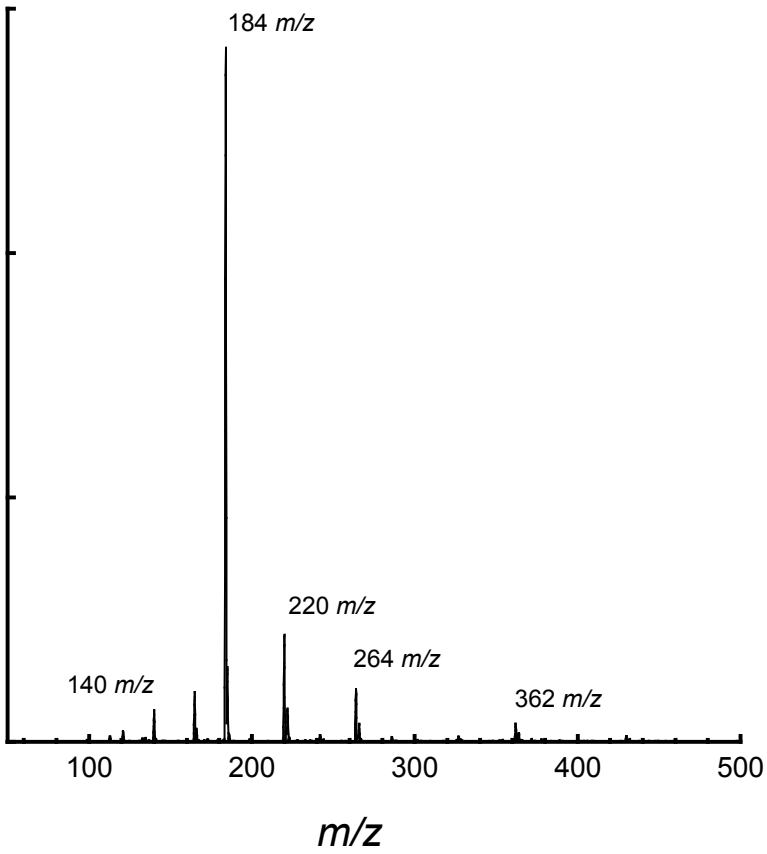


Figure S7. Full mass spectrum of HPLC-purified **3** obtained in negative ion mode over 50-500 *m/z*. The unlabeled peak at 165 *m/z* is a background signal found throughout the elution profile.

SI references

(1) Holm, L.; Rosenström, P. *Nucleic Acids Res* **2010**, *38*, W545.
(2) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G. *Mol Syst Biol* **2011**, *7*, 539.
(3) Krissinel, E.; Henrick, K. *Acta Crystallogr* **2004**, *D60*, 2256.
(4) Emsley, P.; Cowtan, K. *Acta Crystallogr.* **2004**, *D60*, 2126.
(5) Hemsley, A.; Arnheim, N.; Toney, M. D.; Cortopassi, G.; Galas, D. J. *Nucleic Acids Res* **1989**, *17*, 6545.
(6) Sakai, A.; Xiang, D. F.; Xu, C.; Song, L.; Yew, W. S.; Raushel, F. M.; Gerlt, J. A. *Biochemistry* **2006**, *45*, 4455.