

Supporting Information

# Light-Emitting Dehalogenases: Reconstruction of Multifunctional Biocatalysts

*Radka Chaloupkova<sup>1</sup>, Veronika Liskova<sup>1</sup>, Martin Toul<sup>1</sup>, Klara Markova<sup>1</sup>, Eva Sebestova<sup>1</sup>, Lenka Hernychova<sup>2</sup>, Martin Marek<sup>1</sup>, Gaspar P. Pinto<sup>1,4</sup>, Daniel Pluskal<sup>1</sup>, Jitka Waterman<sup>3</sup>, Zbynek Prokop<sup>1,4</sup>, Jiri Damborsky<sup>\*1,4</sup>*

<sup>1</sup>Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic

<sup>2</sup>Regional Centre for Applied Molecular Oncology, Masaryk Memorial Cancer Institute, 656 53 Brno, Czech Republic

<sup>3</sup>Diamond Light Source, Harwell Science and Innovation Campus, Didcot, OX11 0DE, United Kingdom

<sup>4</sup>International Clinical Research Center, St. Anne's University Hospital Brno, Pekarska 53, 656 91 Brno, Czech Republic

\*Correspondence should be addressed to J.D. ([jiri@chemi.muni.cz](mailto:jiri@chemi.muni.cz))

**Table S1.** Diffraction data collection and refinement statistics.

	<b>AncHLD-RLuc</b>
<b>X-ray diffraction data collection statistics</b>	
X-ray source	Diamond light source - I02
Space group	<i>P</i> 12 <sub>1</sub> 1
Cell parameters <i>a</i> , <i>b</i> , <i>c</i> (Å) <i>α</i> , <i>β</i> , <i>γ</i> (°)	44.55, 60.51, 123.63 90, 90.9, 90
Number of molecules in AU	2
Wavelength (Å)	0.9795
Resolution range (Å)	21.48 - 1.39 (1.44 - 1.39)
Observed reflections	1502788 (148042)
Number of unique reflections	124697 (11940)
Completeness (%)	92.62 (91.12)
Multiplicity	12.1
<i>I</i> / <i>σ</i> ( <i>I</i> )	20.6 (1.65)
<i>R</i> <sub>merge</sub> <sup>a</sup>	0.6 (1.8)
CC <sub>1/2</sub> <sup>b</sup>	0.8 (0.324)
<b>Refinement statistics</b>	
Resolution (Å)	1.39
No. of reflections in working set	122118 (11933)
<i>R</i> <sub>work</sub> <sup>c</sup>	0.15
<i>R</i> <sub>free</sub> <sup>d</sup>	0.19
RMSD bond length (Å)	0.005
RMSD angle (°)	0.821
<b>Contents of asymmetric unit</b>	
No. of atoms in AU	5338
No. of protein atoms in AU	4867
No. of water molecules in AU	465
No. of ligand atoms	36
<b>Mean B values</b>	
Protein (Å <sup>2</sup> )	16.99
Ligands O <sub>2</sub> /MPD/Tris (Å <sup>2</sup> )	46.68
Waters (Å <sup>2</sup> )	32.87
<b>PDB code</b>	6G75

Values in parentheses are for the highest resolution shell.

<sup>a</sup> $R_{\text{merge}} = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$ , where the  $I_i(\text{hkl})$  is the  $i$ th observation of reflection  $\text{hkl}$  and  $\langle I(\text{hkl}) \rangle$  is the weighted average intensity for all observations of reflection  $\text{hkl}$ .

<sup>b</sup>As described by Karplus and Diederichs (2012)<sup>1</sup>.

<sup>c</sup> $R_{\text{work}}$  value =  $\|F_o\| - \|F_c\| / \|F_o\|$ , where  $F_o$  and  $F_c$  are the observed and calculated structure factors, respectively.

<sup>d</sup> $R_{\text{free}}$  is equivalent to  $R_{\text{work}}$  value, but is calculated for 5% of the reflections chosen at random and omitted from the refinement process.

**Table S2.** Specific activities<sup>a</sup> of AncHLD-RLuc, LinB, RLuc8 and RLuc8+Ala54Pro toward a set of 29 halogenated substrates and coelenterazine.

Substrate	AncHLD-RLuc [nmol s <sup>-1</sup> mg <sup>-1</sup> ]	LinB <sup>b</sup> [nmol s <sup>-1</sup> mg <sup>-1</sup> ]	RLuc8 [nmol s <sup>-1</sup> mg <sup>-1</sup> ]	RLuc8+Ala54Pro [nmol s <sup>-1</sup> mg <sup>-1</sup> ]
1-chlorobutane	0.782	23.1	nd	0.0442
1-chlorohexane	1.78	27.0	nd	0.0245
1-bromobutane	13.7	48.9	nd	0.0501
1-bromohexane	3.81	29.3	nd	0.0239
1-iodopropane	22.3	66.5	nd	0.103
1-iodobutane	11.8	56.5	nd	0.0339
1-iodohexane	2.18	46.0	nd	0.0241
1,2-dichloroethane	0.954	nd	nd	nd
1,3-dichloropropane	3.44	20.4	nd	0.103
1,5-dichloropentane	7.38	28.8	nd	0.0118
1,2-dibromoethane	7.79	133	nd	0.418
1,3-dibromopropane	31.5	92.5	nd	0.257
1-bromo-3-chloropropane	28.5	86.0	nd	0.301
1,3-diiodopropane	17.9	47.9	nd	0.0741
2-iodobutane	2.82	10.1	nd	nd
1,2-dichloropropane	0.568	nd	nd	nd
1,2-dibromopropane	14.6	62.5	nd	0.270
1,2,3-trichloropropane	nd	nd	nd	0.00691
<i>bis</i> (2-chloroethyl)ether	1.48	17.7	nd	0.0441
chlorocyclohexane	0.956	7.40	nd	nd
bromocyclohexane	2.59	24.9	nd	nd
(1-bromomethyl)cyclohexane	5.44	8.50	nd	0.00674
1-bromo-2-chloroethane	10.7	93.9	nd	0.375
chlorocyclopentane	2.08	5.87	nd	nd
4-bromobutyronitrile	21.9	57.8	nd	0.412
1,2,3-tribromopropane	9.43	93.6	nd	0.139
1,2-dibromo-3-chloropropane	7.33	nd	nd	0.0698
3-chloro-2-methylpropene	4.25	35.1	nd	nd
2,3-dichloropropene	3.24	15.5	nd	nd
	AncHLD-RLuc [RLUs <sup>-1</sup> mg <sup>-1</sup> ]	LinB [RLUs <sup>-1</sup> mg <sup>-1</sup> ]	RLuc8 [RLUs <sup>-1</sup> mg <sup>-1</sup> ]	RLuc8+Ala54Pro [RLUs <sup>-1</sup> mg <sup>-1</sup> ]
coelenterazine	58.9	nd	1.52 x 10 <sup>7</sup>	1.13 x 10 <sup>6</sup>

<sup>a</sup>Each activity was measured in three to five independent replicates with standard deviations of less than 10%. <sup>b</sup>Data from Koudelakova et al. 2011<sup>2</sup>, nd – not detected under tested experimental conditions.

**Table S3.** Steady-state kinetic parameters of AncHLD-RLuc and LinB with six selected halogenated substrates.

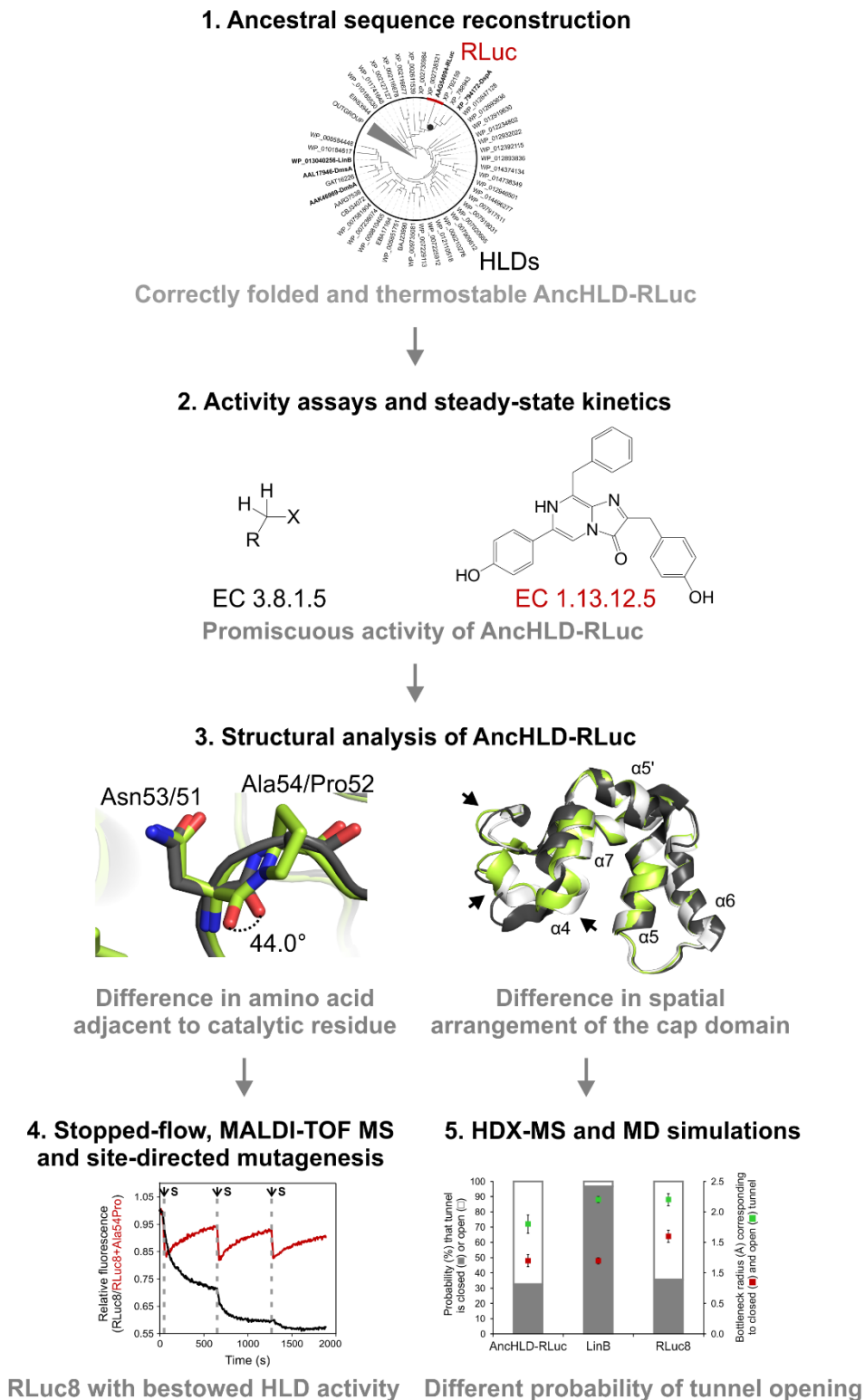
Substrate	AncHLD-RLuc				LinB			
	$k_{\text{cat}}$ [s <sup>-1</sup> ]	$K_{\text{m}}$ [mM]	$K_{\text{si}}$ [mM]	$n$	$k_{\text{cat}}$ [s <sup>-1</sup> ]	$K_{\text{m}}$ [mM]	$K_{\text{si}}$ [mM]	$n$
1-bromobutane	0.55 ± 0.03	0.043 ± 0.002	11 ± 1	- <sup>a</sup>	2.26 ± 0.07 <sup>d</sup>	0.121 ± 0.002 <sup>d</sup>	- <sup>a</sup>	0.440 ± 0.002 <sup>d</sup>
1-iodopropane	0.82 ± 0.07	0.15 ± 0.02	2.8 ± 0.9	- <sup>a</sup>	2.05 ± 0.05	0.041 ± 0.002	1.7 ± 0.1	- <sup>a</sup>
1,2-dibromoethane	0.290 ± 0.004	1.72 ± 0.04	- <sup>a</sup>	- <sup>a</sup>	12 ± 4 <sup>e</sup>	1.7 ± 0.2 <sup>e</sup>	13.0 ± 0.6 <sup>e</sup>	- <sup>a</sup>
1,3-dibromopropane	1.6 ± 0.3	0.07 ± 0.01	0.2 ± 0.1 <sup>b</sup>	1.7 ± 0.2	41 ± 5 <sup>d</sup>	24 ± 3 <sup>d</sup>	0.49 ± 0.06 <sup>d</sup>	- <sup>a</sup>
4-butyronitrile	1.0 ± 0.2	0.18 ± 0.06	2 ± 1 <sup>c</sup>	1.8 ± 0.5	5.03 ± 0.06	0.31 ± 0.02	39 ± 4	1.55 ± 0.01
1,2,3-tribromopropane	0.034 ± 0.009	0.039 ± 0.007	- <sup>a</sup>	1.45 ± 0.07	4.0 ± 0.1	0.59 ± 0.05	- <sup>a</sup>	0.82 ± 0.04

$k_{\text{cat}}$  turnover number,  $K_{\text{m}}$  Michaelis constant (concentration of substrate at half maximal velocity),  $K_{\text{si}}$  substrate inhibition constant,  $n$  Hill coefficient (measure of cooperativity; in monomeric enzymes with single-ligand binding sites, cooperativity is associated with slow conformational transitions between different enzyme conformers<sup>3</sup>). <sup>a</sup>Not applicable, <sup>b</sup>hyperbolic substrate inhibition with  $b$  value = 0.57 ± 0.06, <sup>c</sup>hyperbolic substrate inhibition with  $b$  value = 0.59 ± 0.02, <sup>d</sup>data from Chaloupková et al. 2014<sup>4</sup>, <sup>e</sup>data from Brezovsky et al. 2016<sup>5</sup>. The standard deviation values are presented whenever applicable.

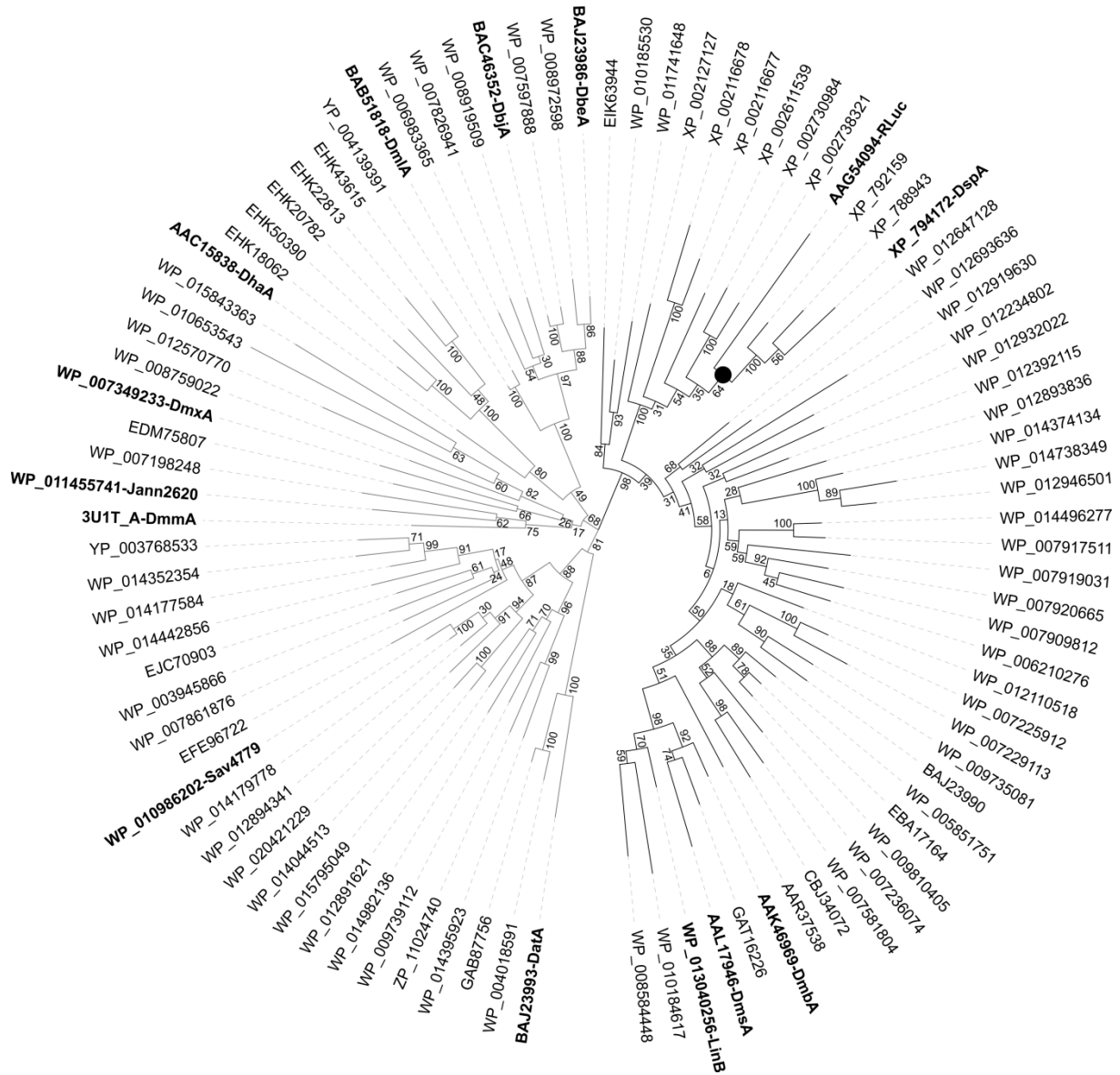
**Table S4.** Comparison of steady-state kinetic parameters of RLuc8+Ala54Pro toward 1,2-dibromoethane and coelenterazine with corresponding parameters of AncHLD-RLuc, LinB and RLuc8.

<b>Dehalogenation</b>	<b>1,2-dibromoethane</b>			
	$k_{cat}$ [s <sup>-1</sup> ]	$K_m$ [mM]	$K_{si}$ [mM]	$k_{cat}/K_m$ [mM <sup>-1</sup> s <sup>-1</sup> ]
RLuc8+Ala54Pro	0.052 ± 0.002	1.2 ± 0.2	2.7 ± 0.3	0.043 ± 0.007
AncHLD-RLuc	0.290 ± 0.004	1.72 ± 0.04	- <sup>a</sup>	0.169 ± 0.005
LinB	12 ± 4 <sup>b</sup>	1.7 ± 0.2 <sup>b</sup>	13.0 ± 0.6 <sup>b</sup>	7 ± 2 <sup>b</sup>
<b>Monoxygenation</b>	<b>coelenterazine</b>			
	$V_{lim}$ [RLUs <sup>-1</sup> mg <sup>-1</sup> ]	$K_m$ [μM]	$K_{si}$ [μM]	$V_{lim}/K_m$ [μM <sup>-1</sup> RLUs <sup>-1</sup> mg <sup>-1</sup> ]
RLuc8+Ala54Pro	(4.9 ± 0.6) x 10 <sup>6</sup>	5.4 ± 0.9	6.6 ± 0.4 <sup>c</sup>	(9 ± 2) x 10 <sup>5</sup>
AncHLD-RLuc	73 ± 2	1.3 ± 0.1	- <sup>a</sup>	53.8 ± 0.09
RLuc8	(1.63 ± 0.06) x 10 <sup>7</sup>	1.1 ± 0.1	- <sup>a</sup>	(1.4 ± 0.2) x 10 <sup>7</sup>

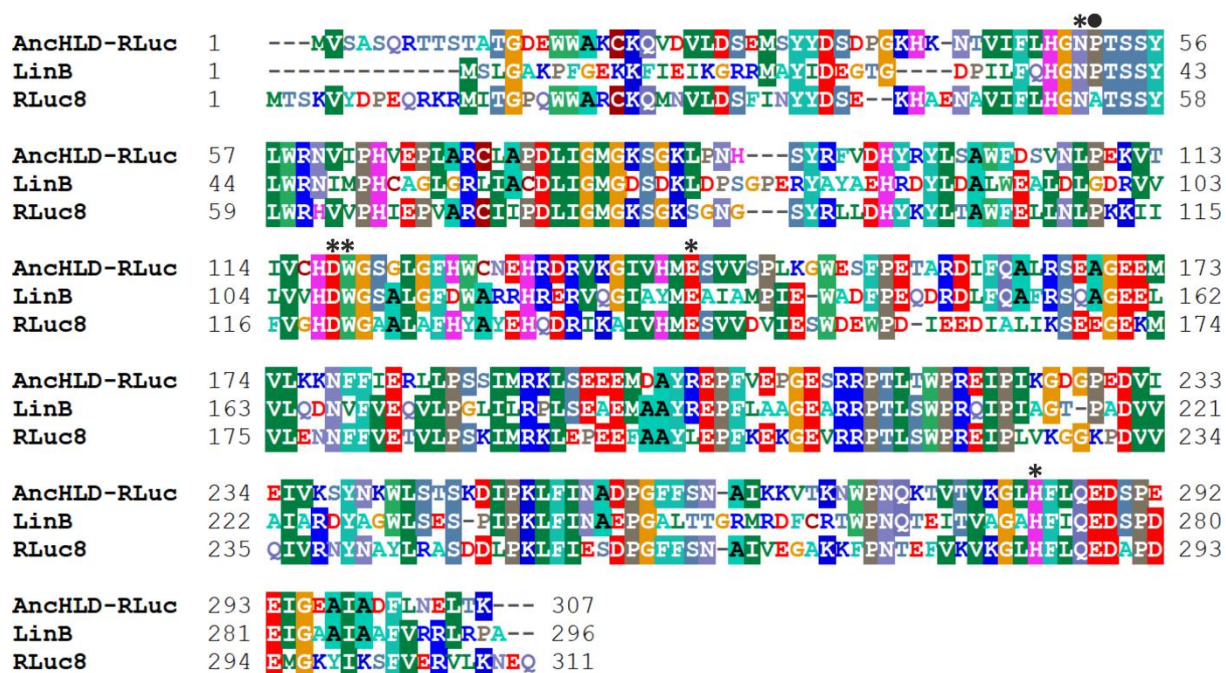
$k_{cat}$  turnover number,  $V_{lim}$  limiting value of maximal velocity,  $K_m$  Michaelis constant (concentration of substrate at half maximal velocity),  $K_{si}$  substrate inhibition constant,  $k_{cat}/K_m$  catalytic efficiency. <sup>a</sup>Not applicable, <sup>b</sup>data from Brezovsky et al. 2016<sup>5</sup>, <sup>c</sup>cooperative substrate inhibition with Hill coefficient  $m = 3.44 ± 0.34$ . The standard deviation values are presented whenever applicable.



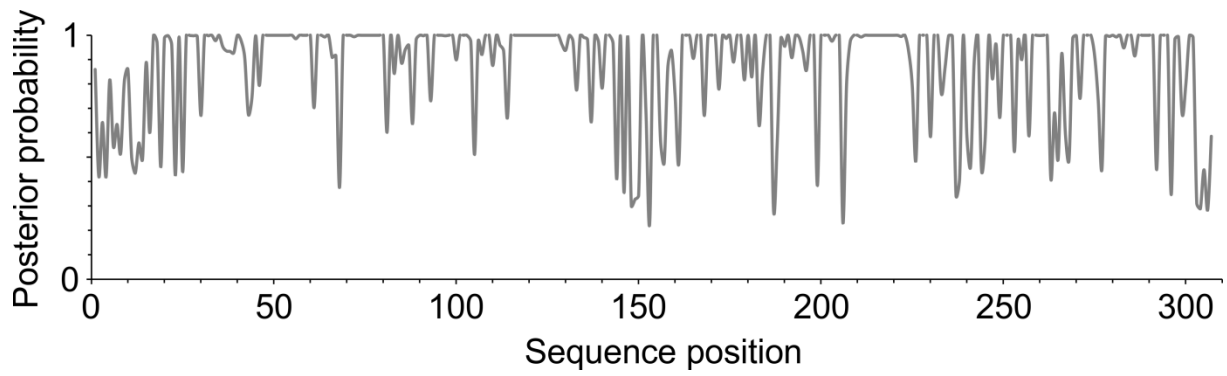
**Figure S1.** Schematic workflow of the study. The study employed ancestral sequence reconstruction, X-ray crystallography, steady-state and pre-steady-state enzyme kinetics, matrix-assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF MS), site-directed mutagenesis, hydrogen-deuterium exchange mass spectrometry (HDX-MS) and molecular dynamics (MD). The data collected in five steps provided detailed insights into the structural determinants of hydrolase and monooxygenase activities.



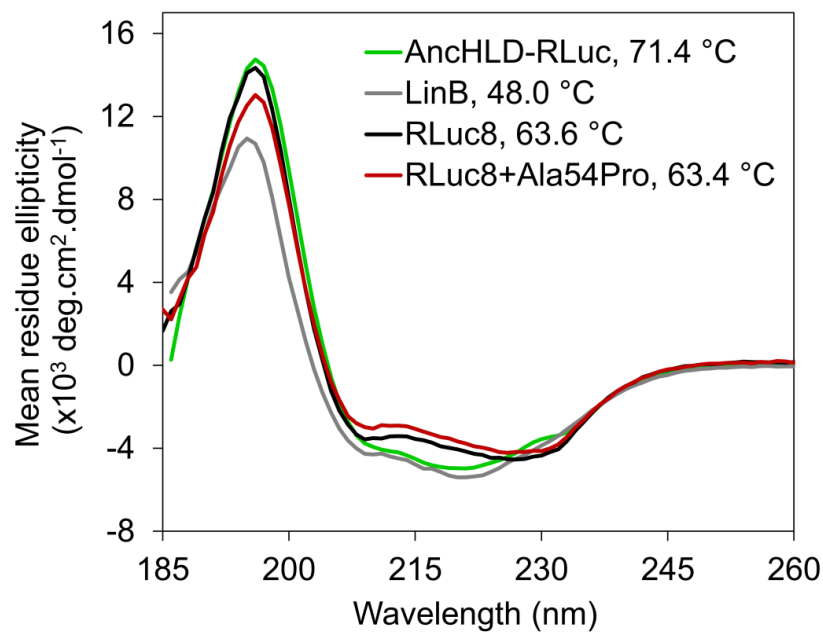
**Figure S2.** Maximum-likelihood phylogenetic tree of the HLD-II subfamily. The position of reconstructed ancestor AncHLD-RLuc is indicated by the black circle. Experimentally characterized extant enzymes are labeled in bold. Putative HLD-II subfamily members are labeled by their accession numbers. Black branches mark the LinB/RLuc HLD-II subgroup; grey branches the second HLD-II subgroup used for rooting the LinB/RLuc subgroup tree. Bootstrap support values are depicted above branches. The figure was generated with iTOL<sup>6</sup>. Based on the tree topology, RLuc probably evolved from HLD enzymes.



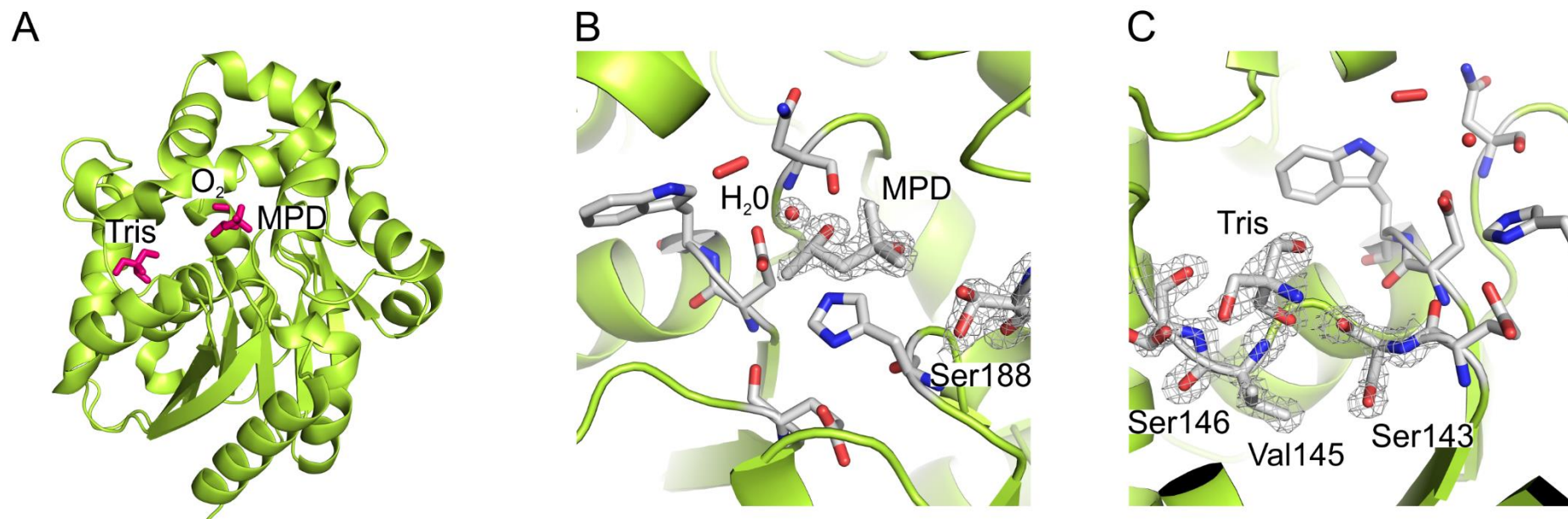
**Figure S3.** Multiple sequence alignment of haloalkane dehalogenase LinB, *Renilla luciferase* RLuc8 and predicted ancestor AncHLD-RLuc. Amino acids are coloured according to their physicochemical properties. Identical residues are highlighted. Conserved catalytic residues are marked by a star, while the position of proline/alanine residue adjacent to the halide/oxygen-stabilizing residue is marked by a circle. The figure was generated with BioEdit<sup>7</sup>.



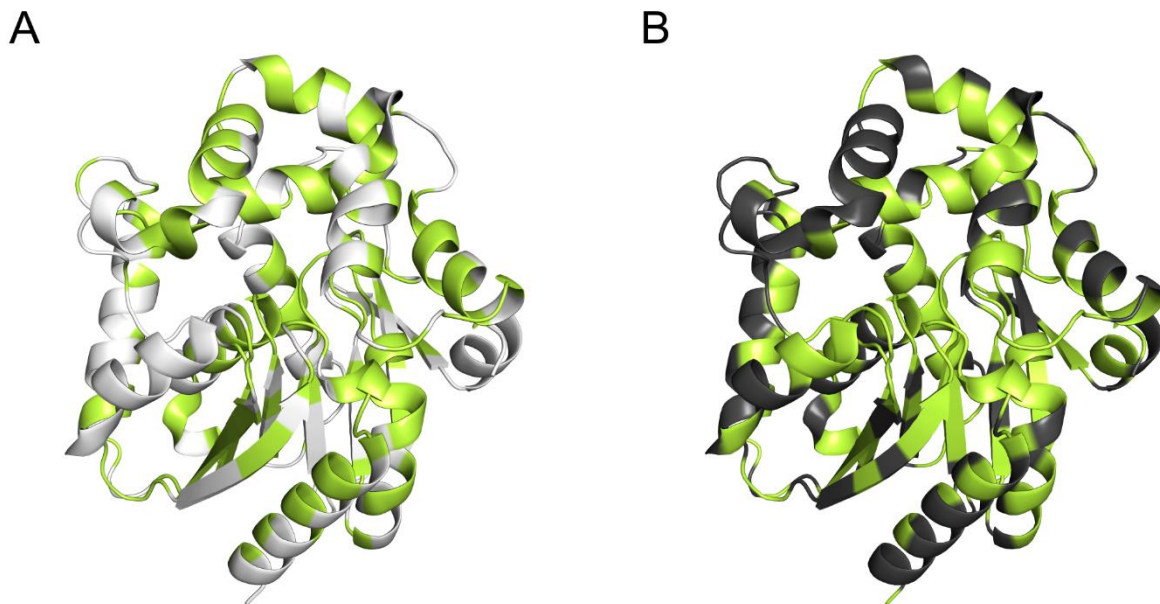
**Figure S4.** Posterior probability distribution of inferred amino acids across all sites of AncHLD-RLuc. The inferred amino acid at each position of ancestral protein is the amino acid with the highest-weighted posterior probability at a given site.



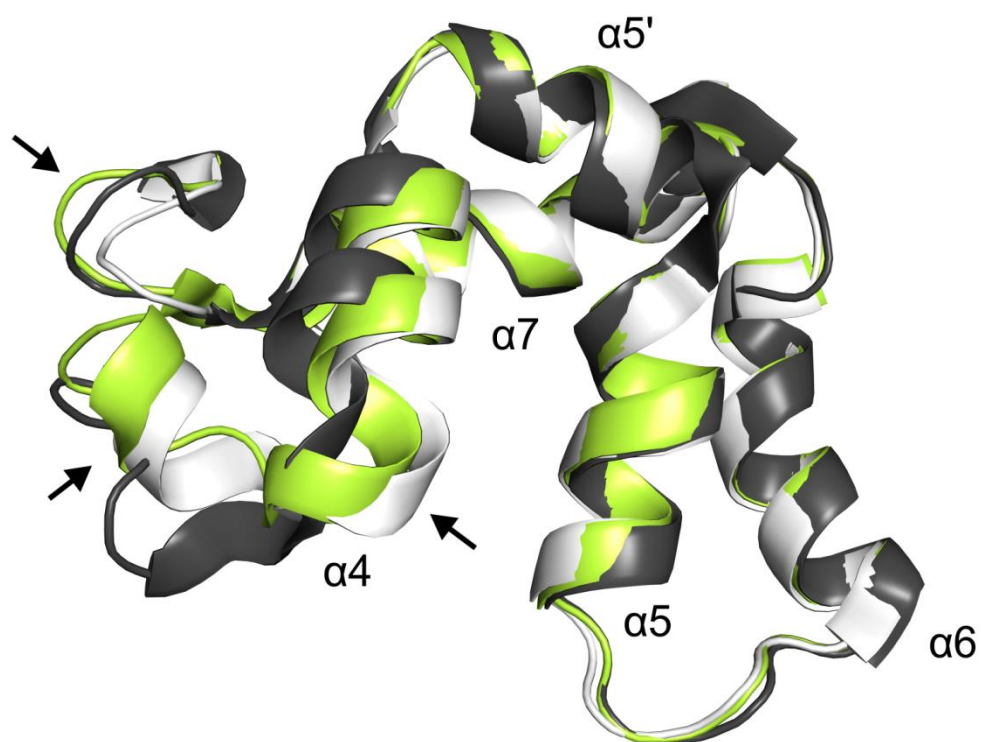
**Figure S5.** Far-UV CD spectra and thermostability of predicted ancestor AncHLD-RLuc, haloalkane dehalogenase LinB, *Renilla* luciferase RLuc8 and its variant RLuc8+Ala54Pro. Apparent melting temperatures of particular enzymes are included in the figure legend. The values were determined from three independent measurements with standard deviations of less than 3%.



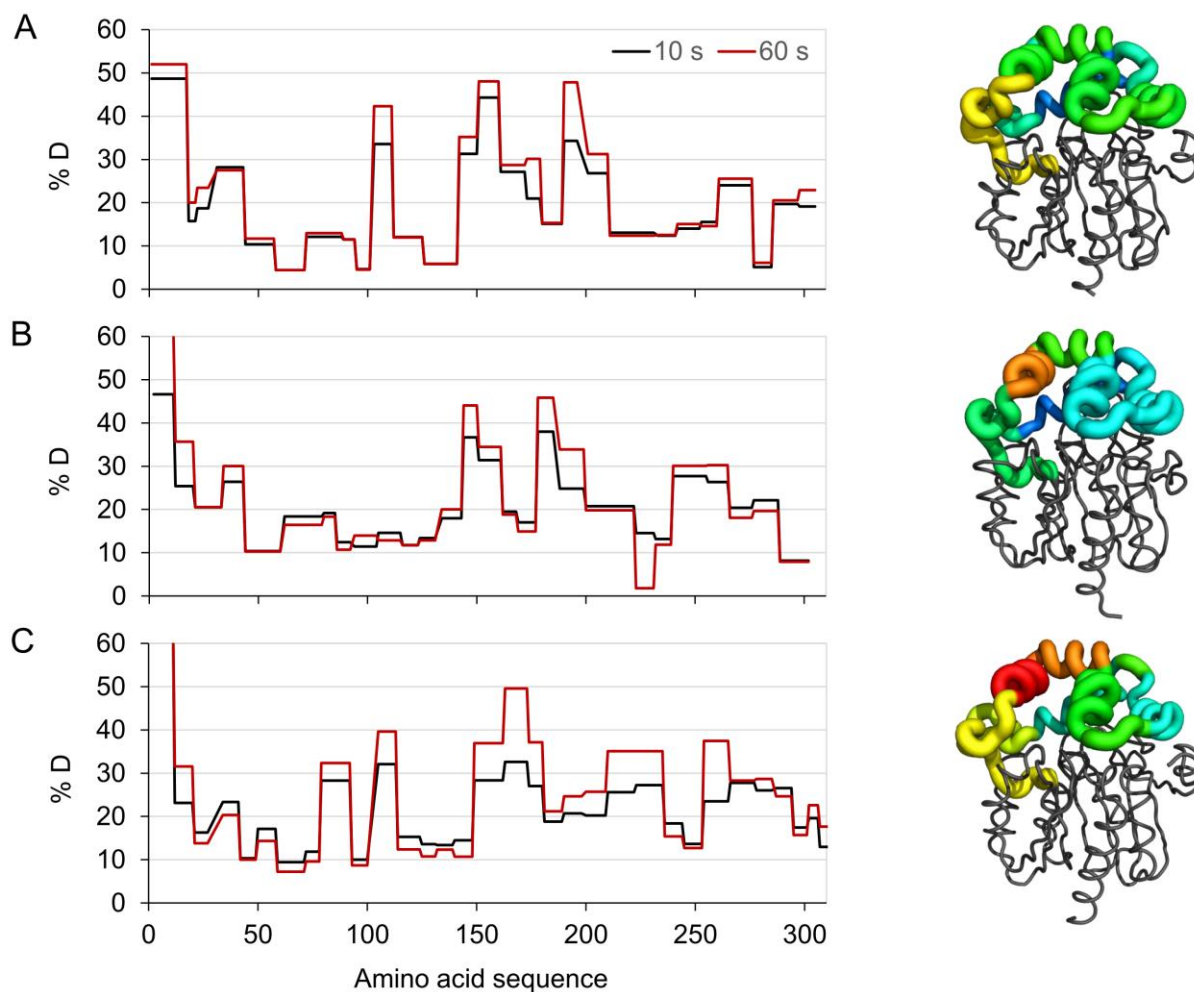
**Figure S6.** Ligands identified in the crystal structure of AncHLD-RLuc. (A) Cartoon representation of AncHLD-RLuc structure with highlighted O<sub>2</sub>, (4*S*)-2-methyl-2,4-pentanediol (MPD) and tris(hydroxymethyl)aminomethane (Tris) ligands. Two of the ligands (MPD and Tris) that are not coupled to the enzyme reactivity because they originate from the storage and cryoprotectant buffers. (B) The first ligand, interpreted as MPD, is located approximately 5 Å far from the molecular oxygen and is coordinated by the same water molecule mediating the contact with nucleophile and the oxygen. The distance between the water molecule O atom and the MPD O<sup>4</sup> is 2.36 Å. Further coordination is provided by Ser188 O<sup>ε</sup> with the distance between MPD O<sup>2</sup> and Ser188 O<sup>ε</sup> 3.00 Å. (C) The second ligand, interpreted as Tris, is located in the slot tunnel, approximately 9 Å far from the molecular oxygen. The ligand interacts with Ser143 main chain carbonyl O, Val145 main chain N and Ser146 O<sup>γ</sup>. The distances between the Tris O<sup>8</sup> and Ser143 O, Tris O<sup>8</sup> and Val145 N, and Tris O<sup>4</sup> and Ser146 O<sup>γ</sup> are 2.40, 2.67 and 2.99 Å, respectively. The ligands, catalytic and interacting residues are displayed in stick representation; water molecule is displayed in sphere representation. 2F<sub>o</sub>-F<sub>c</sub> electron density map contoured at 2.0, 1.3 and 1.0 σ is shown for the interacting residues with bound MPD and Tris ligands, respectively.



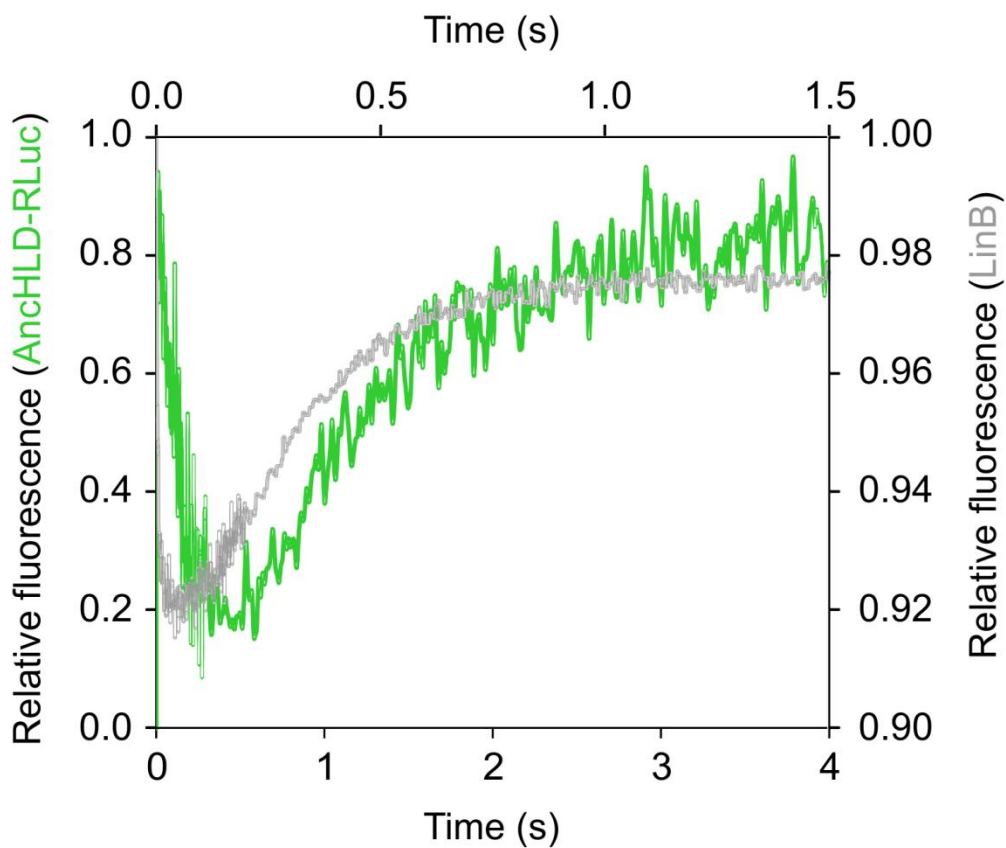
**Figure S7.** Structural comparison of the ancestral and modern-day enzymes. AncHLD-RLuc structure with highlighted amino acid positions occupied by different residues (A) in comparison of AncHLD-RLuc and LinB (white) and (B) in comparison of AncHLD-RLuc and RLuc8 (black). The amino acid differences between the predicted ancestor and both present-day enzymes mapped on AncHLD-RLuc structure depict that the substitutions are spread throughout the protein structure including substitutions in the protein core, on the protein surface and in the active site.



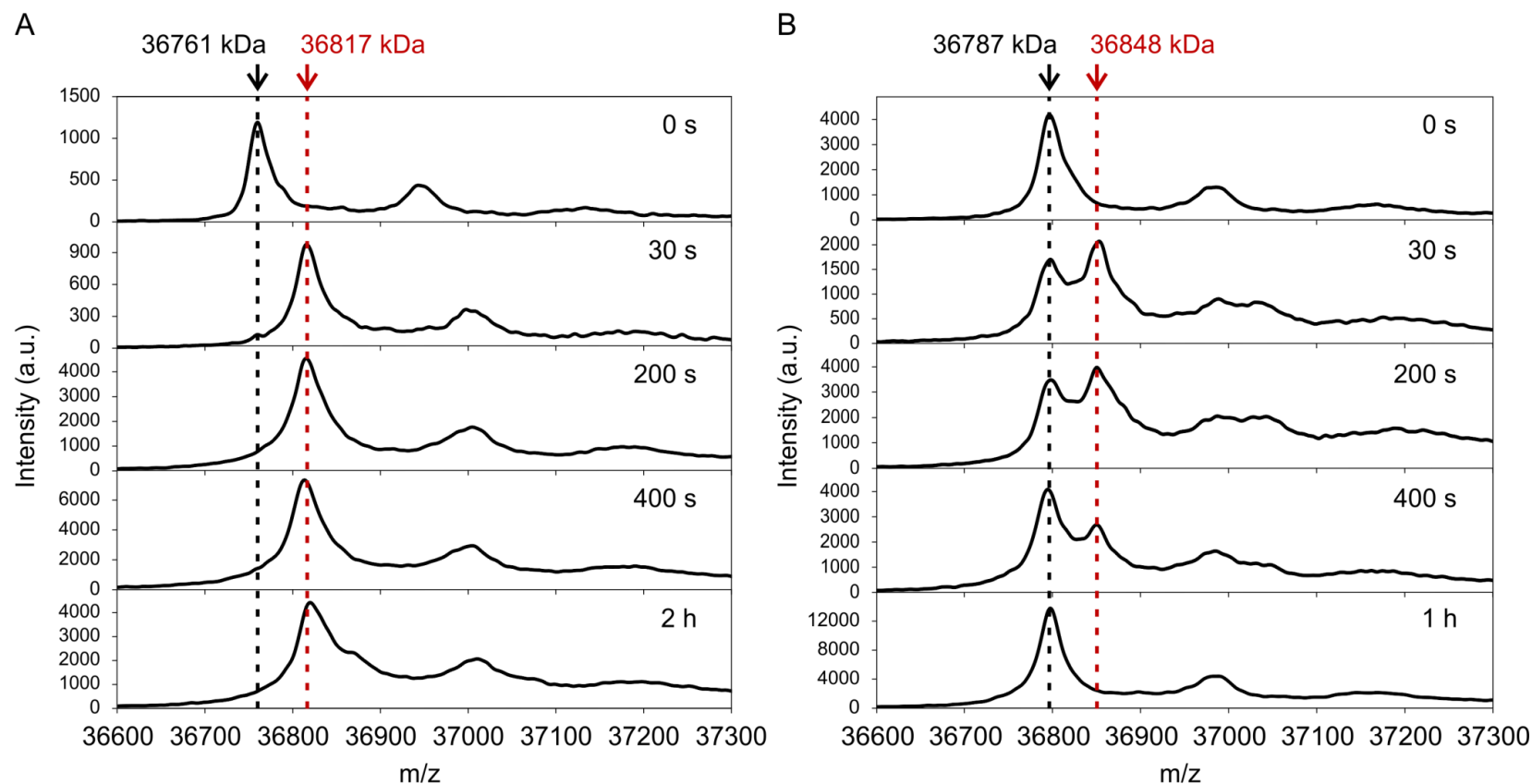
**Figure S8.** Expanded view of superposition of cap domains of AnchLD-RLuc, LinB and RLuc8, depicting the most significant structural deviations among the three enzyme structures. Crystal structures of AnchLD-RLuc, LinB and RLuc8 are shown in green, white and black, respectively. Arrows indicate identified differences the spatial arrangements of  $\alpha 4$  helices of the cap domain, an adjacent loop connecting the  $\alpha 4$  helix of the cap domain with the main domain, and a loop connecting the  $\alpha 7$  helix of the cap domain with the main domain.



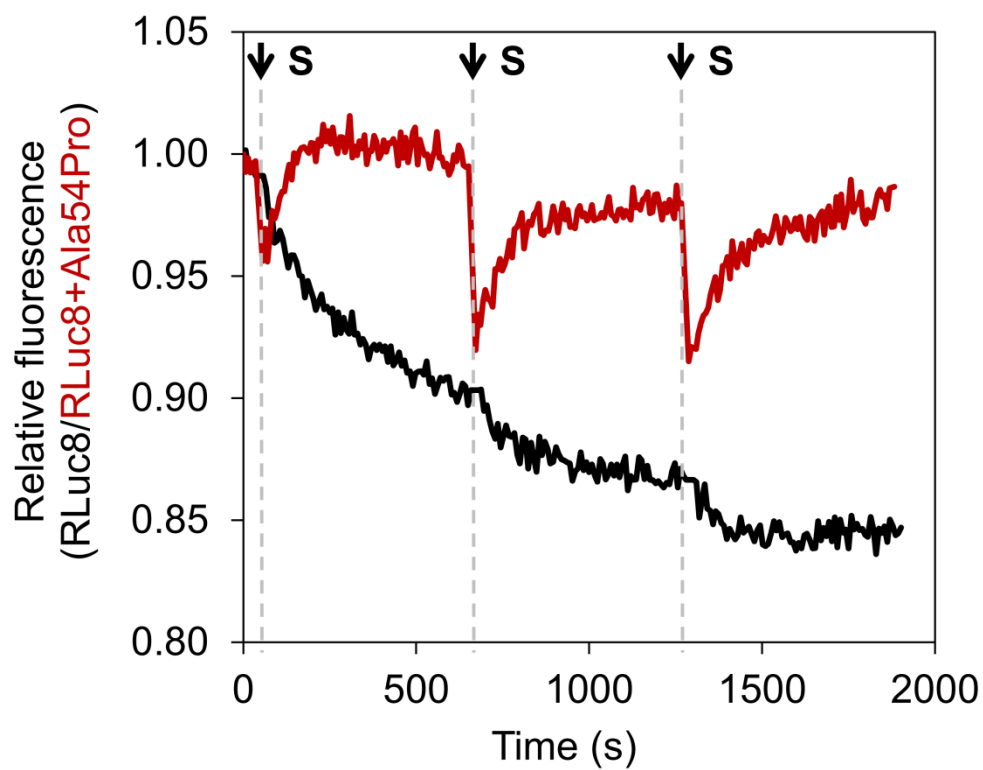
**Figure S9.** Hydrogen-deuterium protection plots and conformational dynamics of (A) AncHLD-RLuc, (B) LinB, and (C) RLuc8. Left: Deuteration profile of peptides over the amino acid sequence of tested enzymes at the time intervals 10 s (black) and 60 s (red). Right: Visualization of dynamics for the cap domains based on molecular dynamics simulations. The B-factors were calculated for individual amino acids and then averaged for the secondary structure elements. Color ranges from low (blue) to high (red) B-factor with a common maximum. A thickness of the ribbon ranges from low (thin) to high (thick) B-factor for each protein separately. The main domains are in grey.



**Figure S10.** Single-turnover stopped-flow analysis of AncHLD-RLuc and LinB reactions with 1-bromobutane. Fluorescence traces recorded following rapid mixing of 75  $\mu\text{M}$  AncHLD-RLuc and 124  $\mu\text{M}$  LinB with 12.7  $\mu\text{M}$  and 40.9  $\mu\text{M}$  1-bromobutane, respectively.



**Figure S11.** MALDI-TOF MS spectra of RLuc8 and its variant RLuc8+Ala54Pro collected at different time intervals of their reaction with 1-bromobutane. (A) Formation of the covalent complex upon interaction of RLuc8 with 1-bromobutane, where the complex is not further hydrolysed and thus the enzyme stays in its complexed-form in all reaction times. (B) Formation of the covalent complex upon interaction of RLuc8+Ala54Pro with 1-bromobutane, where the complex is further hydrolysed and thus the enzyme is recovered back to its free form during time course of the reaction. The peaks at 36761 and 36787 kDa, highlighted by black dashed lines, correspond to the free forms of RLuc8 and RLuc8+Ala54Pro, respectively. The peaks at 36817 and 36848 kDa, highlighted by red dashed lines, correspond to the complexed forms of RLuc8 and RLuc8+Ala54Pro, respectively. Observed differences in the molecular masses between the free and the complexed forms of RLuc8 (56 Da) and RLuc8+Ala54Pro (61 Da) corresponds to the mass of butyl chain (57 Da) covalently bound to the enzymes.



**Figure S12.** Repetitive single-turnover stopped-flow analysis of RLuc8 and RLuc8+Ala54Pro reactions with 1,2-dibromoethane. Fluorescence traces recorded following mixing 89  $\mu\text{M}$  RLuc8 and 767  $\mu\text{M}$  RLuc8+Ala54Pro with 116.8  $\mu\text{M}$  and 747.1  $\mu\text{M}$  1,2-dibromoethane (S), respectively, repeatedly injected into the same enzyme solution in 630-s time intervals.

## REFERENCES

- (1) Karplus, P. A.; Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* **2012**, *336*, 1030–1033.
- (2) Koudelakova, T.; Chovancova, E.; Brezovsky, J.; Monincova, M.; Fortova, A.; Jarkovsky, J.; Damborsky, J. Substrate Specificity of Haloalkane Dehalogenases. *Biochem. J.* **2011**, *435*, 345–354.
- (3) Porter, C. M.; Miller, B. G. Cooperativity in Monomeric Enzymes with Single Ligand-Binding Sites. *Bioorganic Chem.* **2012**, *43*, 44–50.
- (4) Chaloupkova, R.; Prudnikova, T.; Rezacova, P.; Prokop, Z.; Koudelakova, T.; Daniel, L.; Brezovsky, J.; Ikeda-Ohtsubo, W.; Sato, Y.; Kutý, M.; Nagata, Y.; Smatanova, I. K.; Damborsky, J. Structural and Functional Analysis of a Novel Haloalkane Dehalogenase with Two Halide-Binding Sites. *Acta Crystallogr. D Biol. Crystallogr.* **2014**, *70*, 1884–1897.
- (5) Brezovsky, J.; Babkova, P.; Degtjarik, O.; Fortova, A.; Gora, A.; Iermak, I.; Rezacova, P.; Dvorak, P.; Smatanova, I. K.; Prokop, Z.; Chaloupkova, R.; Damborsky, J. Engineering a de Novo Transport Tunnel. *ACS Catal.* **2016**, *6*, 7597–7610.
- (6) Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL): An Online Tool for Phylogenetic Tree Display and Annotation. *Bioinformatics* **2007**, *23*, 127–128.
- (7) Hall, T. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.