Supporting Information

Functional and Mechanistic Characterization of an Enzyme Family by an Integrated Strategy Combining Bioinformatics and Microfluidics

Michal Vasina[1,2,#], Pavel Vanacek[1,2,#], Jiri Hon[2,3], David Kovar[1,2], Hana Faldynova[1], Antonin Kunka[1,2], Tomas Buryska[1], Christoffel P. S. Badenhorst[4], Stanislav Mazurenko[1,2], David Bednar[1,2], Stavros Stavrakis[5], Uwe T. Bornscheuer[4], Andrew deMello[5], Jiri Damborsky[1,2,*], Zbynek Prokop[1,2,*]

[1] Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, 625 00 Brno, Czech Republic

[2] International Clinical Research Centre, St. Ann's Hospital, 656 91 Brno, Czech Republic

[3] IT4Innovations Centre of Excellence, Faculty of Information Technology, Brno University of Technology, 612 66 Brno, Czech Republic

[4] Department of Biotechnology & Enzyme Catalysis, Institute of Biochemistry, Greifswald University, Greifswald 17487, Germany

[5] Institute for Chemical and Bioengineering, ETH Zürich, 8093 Zürich, Switzerland

# P.V. and M.V. contributed equally

* Authors for correspondence: jiri@chemi.muni.cz, ORCID, 0000-0002-7848-8216; zbynek@chemi.muni.cz, ORCID 0000-0001-9358-4081

Table of Contents

Supporting Tables

Table S1. Sequence annotations according to individual HLD subfamilies.

Table S2. Summary of the selection procedure.

Table S3. The number of candidates assigned for the particular selection criteria**.**

Table S4. Annotations of selected HLDs.

Table S5. Analysis of protein solubility.

Table S6: Screening of the enzymatic activity with whole cells.

Table S7. Microscale analysis of thermostability.

Table S8. Microfluidic analysis of temperature profiles of catalytic activity.

Table S9: Representative set of halogenated substrates.

Table S10. Classification and labelling of halogenated substrates.

Table S11. Microfluidic analysis of substrate specificity.

Table S12. The assessment of substrate specificity range.

Table S13: The steady-state kinetic and thermodynamic parameters.

Table S14. Determination of oligomeric states.

Supporting Figures

Figure S1. Analysis of volumes of catalytic pockets in different HLD subfamilies.

Figure S2. Restriction analysis of 45 recombinant pET-24a::HLD plasmids.

Figure S3. Small scale expression analysis.

Figure S4. Analysis of protein purity.

Figure S5. Comparison of different methods for thermostability analysis.

Figure S6. Microfluidic analysis of temperature profiles of HLDs activity.

Figure S7. Analysis of thermostability and temperature profiles.

Figure S8. Analysis of substrate preferences.

Figure S9. Validation of the specific activities.

Figure S10. Principal component analysis.

Figure S11. Global fitting of microfluidic kinetic and thermodynamic data.

Figure S12. Calibration of the microfluidic platform for reaction thermodynamics characterization.

Supporting Tables

Figure S13. Kinetic resolution of 2-bromopentane.

Figure S14. Kinetic resolution of ethyl-2-bromopropionate.

Figure S15. Analysis of temperature-induced unfolding.

Figure S16. Analysis of the quaternary structure.

References

Supporting Tables

Table S1. **Sequence annotations according to individual HLD subfamilies.** Sequence annotations were extracted from the sequence header line in the NCBI database.

	Number of sequences			
Annotation	HLD-I	HLD-II	HLD-III	HLD-IV
haloalkane dehalogenase	596	764	263	9
dehalogenase	0	19	0	0
renilla-luciferin 2-monooxygenase-like	0	6	0	0
alpha/beta hydrolase	54	22	568	0
putative hydrolase/hydrolase/hydrolase protein	21	5	11	0
hypothetical protein/unnamed protein	32	104	107	7
CMP deaminase/tRNA-specific adenosine deaminase	47	0	0	0
epoxide hydrolase	0	2	0	0

Supporting Tables

Table S2. Summary of the selection criteria. Proposed sequence and structural characteristics used for probing HLD diversity.

Category	Description of category
Selected genes	All selected sequences
Full dataset	All identified sequences
Archaea	Sequences from Archaea
Temperature	Sequences from organisms with an extreme optimum temperature (psychrophilic, cryophilic or thermophilic)
Salinity	Sequences from organisms with extreme salinity (moderate or extreme halophilic)
Biotic relationship	Sequences from organisms with biotic relationship annotation
Disease	Sequences from disease-related organisms
Top 5% by active site volume (smallest)	Top 5% of sequences by the active site volume (smallest)
Top 5% by active site volume (largest)	Top 5% of sequences by the active site volume (largest)
Top 5% by main tunnel throughput	Top 5% of sequences by the throughput of the main tunnel
Top 5% by main tunnel bottleneck radius	Top 5% of sequences by the bottleneck radius of the main tunnel
Occluded active site	Sequences with no tunnels detected in the homology model
Unique sequences	Outlier sequences selected using the sequence similarity network visualization
Multi-domains	Sequences with additional Pfam domains or predicted transmembrane regions

Supporting Tables

Table S3. The number of candidates assigned based on a given selection criterion. If more candidates were available in the selected categories, the sequences with the highest predicted solubility and those with known tertiary structures were prioritized.

Criterion	n
Archaea	4
Biotic relationships	4
Closed tunnels	5
Disease-related	4
Large active site	5
Large tunnel throughput	5
Psychrophile	4
Salinity	6
Small active site	6
Thermophile	2
Sum	45

Table S4. Annotations of selected HLDs. Accession numbers are from the NCBI database.

Accession number	Protein code
ELP68488.1	DstA
WP_008318309.1	DhmuA
SFR53052.1	DhliA
WP_053043363.1	DtacA
WP_071046332.1	DfxA
WP_057462346.1	DpxA
WP_030471311.1	DlaA
WP_020421229.1	DaxA
WP_025046666.1	DsmA
ACC42520.1	DmmarA
WP_027934580.1	DathA
ABL06043.1	DmuA
KIE53139.1	DmaA
WP_037967458.1	DspoA
WP_007164271.1	DexA
WP_007637529.1	DppsA
WP_009367015.1	DhsA
WP_007323654.1	DgarA
WP_044544029.1	DmgoA
SDD02355.1	DnlA
WP_053551405.1	DdsA
WP_065375349.1	DeaA
WP_007235390.1	DmgaA
WP_060978844.1	DprxA
WP_018448660.1	DrgA
AAR37538.1	DmbaA
ACC38583.1	DmmarB
KKP01761.1	DthA
WP_058332862.1	DphxA
KKP07098.1	DthB
WP_014982136.1	DnbA
KJS36764.1	DhxA
ODT89409.1	DphexA
WP_066483168.1	DspxA
WP_011189873.1	DdpA
WP_068372589.1	DrxA
OLS22338.1	DchA
WP_006031557.1	DmoxA
WP_006231583.1	DpproA
WP_012277899.1	DshA
ACL23491.1	DcagA
ELU04432.1	DctA
XP_002730984.1	DskA
XP_792159.3	DspB
XP_019625274.1	DbbA

Table S5. **Analysis of protein solubility.** Comparison of *in silico* predictions with *in vitro* protein solubility. Pearson's correlation coefficient of the predicted and experimental solubility is 0.263. The accuracy of the prediction is 66.7%. To calculate the accuracy, we used a threshold of 50% for the predicted solubility and a threshold of 20% for the experimental solubility. There were 22 true positives, 8 true negatives, 4 false-negative and 11 false-positive predictions.

Name	*In silico*	*In vitro*
	Predicted solubility (%)	Experimental solubility (%)
DstA	82.4	76.0
DhmuA	81.9	0
DhliA	77.7	0
DtacA	75.6	0
DfxA	73.4	24.2
DpxA	72.2	0
DlaA	70.9	61.0
DaxA	70.7	71.1
DsmA	68.8	66.2
DmmarA	68.4	35.5
DathA	68.3	69.7
DmuA	67.3	0
DmaA	67.3	91.2
DspoA	66.3	95.3
DexA	65.5	81.2
DppsA	65.5	85.6
DhsA	64.5	13.4
DgarA	63.6	0
DmgoA	63.5	0
DnlA	63.0	0
DdsA	62.5	67.4
DeaA	61.5	75.7
DmgaA	61.1	39.6
DprxA	59.7	87.5
DrgA	59.6	94.2
DmbaA	56.5	61.0
DmmarB	54.9	0
DthA	54.1	83.7
DphxA	53.6	66.7
DthB	53.4	49.5
DnbA	53.1	63.8
DhxA	51.9	99.0
DphexA	51.6	15.5
DspxA	49.5	53.4
DdpA	49.3	0
DrxA	49.0	0
DchA	47.7	45.2
DmoxA	40.8	0
DpproA	38.3	27.6
DshA	36.8	10.4
DcagA	35.6	0
Dcta	42.3	33.6
DskA	35.7	0
DspB	24.7	0
DbbA	20.2	0

Table S6. **Screening of enzymatic activity with whole cells.** The maximum fluorescence signal reached (relative fluorescence units) for each enzyme/substrate combination is shown. The reaction time varied for individual substrates: 135 min (#47), 175 min (#48), 165 min (#20), 110 min (#119), and 17 h (#154). The signal values are coloured on a scale from blue to red. The measurements were carried out in triplicates.

	#47	#48	#20	#119	#154
DstA	n.d.	2428	8833	n.d.	838
DhmuA	4097	5123	1655	1538	6334
DhliA	4852	7080	5978	3915	5833
DtacA	n.d.	2735	8048	n.d.	n.d.
DfxA	n.d.	n.d.	2163	n.d.	n.d.
DpxA	n.d.	8176	9428	1641	2842
DlaA	n.d.	n.d.	n.d.	n.d.	n.d.
DaxA	n.d.	2504	6897	n.d.	n.d.
DsmA	7663	7757	7702	8966	11951
DmmarA	1580	6899	5848	5843	3023
DathA	n.d.	2676	5009	n.d.	n.d.
DmuA	n.d.	760	n.d.	n.d.	n.d.
DmaA	7664	7212	8422	11146	12491
DspoA	7860	7815	9066	10784	11486
DexA	7834	7903	8384	11301	11768
DppsA	7643	7158	8077	11076	10050
DhsA	3933	5350	3023	2214	7447
DgarA	744	4364	3668	1283	n.d.
DmgoA	n.d.	962	n.d.	n.d.	n.d.
DnlA	n.d.	n.d.	n.d.	n.d.	n.d.
DdsA	1163	2423	n.d.	n.d.	n.d.
DeaA	7252	7883	7642	10564	9112
DmgaA	1846	7275	8304	7375	3144
DprxA	7656	7735	9507	10936	10934
DrgA	n.d.	1275	2464	n.d.	n.d.
DmbaA	7322	6751	6943	5643	11366
DmmarB	4891	5258	4185	2076	8021
DthA	6777	6641	4598	4878	8508
DphxA	6619	7671	7291	10479	10144
DthB	3728	7133	6215	7914	8961
DnbA	n.d.	5186	6913	1271	1022
DhxA	6771	7927	9506	11824	11372
DphexA	5103	6555	7913	3734	8084
DspxA	7683	7757	7145	8469	11472
DdpA	810	2179	n.d.	n.d.	n.d.
DrxA	n.d.	n.d.	n.d.	n.d.	n.d.
DchA	5879	6891	4942	6560	9499
DmoxA	973	4028	4139	n.d.	n.d.
DpproA	n.d.	n.d.	n.d.	n.d.	n.d.
DshA	n.d.	n.d.	n.d.	n.d.	n.d.
DcagA	n.d.	n.d.	n.d.	n.d.	n.d.
Dcta	n.d.	n.d.	n.d.	n.d.	n.d.
DskA	n.d.	3848	1972	808	1639
DspB	6238	7170	7166	5474	8371
DbbA	2712	5738	4821	1617	2283

n.d. –not detected under the tested conditions (the activity values were below the detection limit)

Supporting Tables

Table S7. Microscale analysis of thermostability. The apparent melting temperatures (T_m^{app}) were obtained by using high-throughput screening methods. Thermal shift assays (TSA) utilizing extrinsic fluorescence (SYPRO orange dye), differential scanning fluorimetry (DSF) employing the intrinsic fluorescence of tryptophan in capillaries (Nano DSF Prometheus NT.48 scanning fluorometer, NanoTemper Technologies), and microcuvette (Nano DSF UNcle, Unchained labs) assays are compared to conventional circular dichroism (CD) spectroscopy.

	Capillary DSF		Microcuvette DSF		TSA		CD spectroscopy	
	T_m^{app}	Std.	T_m^{app}	Std.	T_m^{app}	Std.	T_m^{app}	Std.
DstA	43.4	0.1	43.5	0.5	42.5	0.3	41.4	0.7
DfxA	40.6	0.5	41.0	0.5	n.a	n.d.	41.4	0.7
DlaA	48.1	0.6	47.8	0.3	47.8	0.3	50.6	0.4
DaxA	48.7	0.1	48.8	0.5	45.1	0.2	53.6	1.6
DsmA	35.7	0.1	35.0	0.3	34.5	0.2	35.6	0.7
DmmarA	42.1	0.2	40.7	0.6	42.6	0.3	42.9	0.6
DathA	46.4	0.1	45.4	0.7	46.7	0.2	50.5	0.5
DmaA	40.2	0.3	39.1	0.1	41.4	0.1	42.8	1.2
DspoA	58.7	0.6	57.4	0.3	n.a	n.d.	60.0	0.3
DexA	47.5	0.4	46.5	0.1	47.8	0.1	50.4	0.7
DppsA	38.1	0.2	38.0	0.1	43.8	1.5	38.1	1.0
DeaA	52.2	0.2	50.7	0.5	52.4	0.1	54.8	0.6
DmgaA	44.7	0.9	44.0	0.5	43.7	0.8	47.0	0.8
DprxA	51.8	0.3	51.8	0.5	50.5	0.4	54.3	0.8
DrgA	44.2	0.4	43.1	1.0	44.2	0.3	46.1	0.5
DmbaA	46.6	0.2	45.4	0.6	45.6	0.8	47.5	0.5
DthA	49.9	0.9	49.2	0.4	49.6	0.4	52.8	0.6
DphxA	55.4	0.2	54.0	0.7	n.a	n.d.	56.4	0.8
DthB	53.4	0.4	52.4	0.3	n.a	n.d.	55.4	1.2
DnbA	47.8	0.4	46.5	0.1	46.8	0.5	49.6	0.4
DhxA	53.1	0.3	52.4	1.5	53.1	0.7	56.7	0.3
DspxA	53.3	0.2	52.0	0.1	54.1	0.3	54.6	0.4
DchA	55.2	0.8	53.2	0.6	55.4	0.3	55.0	0.8
DctA	39.8	0.6	40.0	0.6	40.5	0.8	45.4	0.2

Std. – standard deviation, n.d. – not determined, n.a. – not applicable

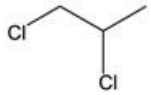
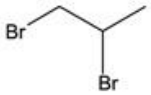
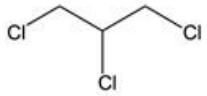
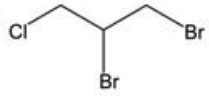
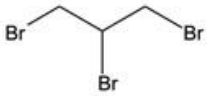
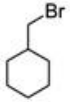
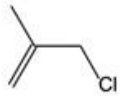
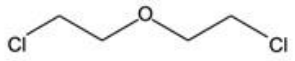
Supporting Tables

Table S8. **Microfluidic analysis of temperature profiles of catalytic activity.** Specific activities were recorded at different temperatures ranging from 5 to 55 °C. The level of activity is depicted by a colour scale (from blue to red) in nmol.s^{-1}.mg^{-1}. All measurements were carried out in 5-10 replicates. The reactions were conducted in 1 mM HEPES buffer at pH 8.2.

Enzyme	Substrate	5	10	15	20	25	30	35	40	45	50	55	T_{max} (°C)	T_{onset} (°C)
DstA	#20		13.8	25.2	36.2	44.1	60.3	53.0	35.2	18.9			30	30.9
DfxA	#20		1.7	3.0	4.0	5.0	5.8	6.7	3.8	2.5			35	30.5
DlaA	#20		2.3	2.6	3.0	4.0	4.3	3.7	3.5	1.9			30	35.6
DaxA	#20		0.5	1.0	1.4	2.3	3.1	3.6	3.1	1.9			35	42.9
DsmA	#47		21.3	30.2	35.2	39.3	36.6	16.1	15.7				25	27.6
DmmarA	#20		3.0	5.1	5.8	8.4	9.2	9.0	3.2	2.2			30	32.3
DathA	#20		0.7	1.3	1.8	2.9	3.4	3.6	3.4	1.6			35	38.1
DmaA	#47	137.2	153.1	167.0	179.0	184.2	193.8	211.8	136.9	24.2			35	32.5
DspoA	#47				233.7	246.7	257.4	318.5	341.6	378.5	427.5	234.5	50	50.8
DexA	#47				144.7	313.6	349.4	405.3	469.3	475.7	53.2		45	43.4
DppsA	#47		10.9	27.1	28.5	30.3	32.7	38.8	31.6	3.6			35	24.7
DeaA	#47				45.9	53.5	59.6	79.7	85.1	106.5	84.1	0.0	45	45.3
DmgaA	#20		2.9	4.8	5.3	6.5	7.1	7.2	7.3	2.0			40	38.2
DprxA	#47				105.6	174.3	202.3	215.9	256.8	269.2	53.3		45	44.3
DrgA	#47			0.0	2.1	5.9	16.8	17.2	15.9	4.7	0.0	0.0	35	36.8
DmbaA	#47			90.5	91.0	148.8	146.9	163.9	176.0	176.5	58.5	46.6	45	36.8
DthA	#47	52.6	65.3	90.3	107.3	152.7	181.0	208.4	199.3	135.1	30.8		35	40.4
DphxA	#47				433.4	422.7	480.9	679.7	658.5	228.9	214.1		35	47.0
DthB	#47				15.5	16.7	20.7	34.3	38.0	49.3	47.5	0.0	45	44.8
DnbA	#47			0.0	1.2	3.3	7.8	7.8	7.5	3.0	0.0	1.0	40	37.3
DhxA	#47					339.7	452.2	498.1	390.0	257.2	96.1	2.0	35	44.1
DspxA	#47				199.7	202.8	243.6	328.9	283.5	126.7	74.2		35	44.2
DchA	#47					135.2	163.0	208.6	227.9	95.0	90.5		40	47.0
Dcta	#47	1.6	1.9	2.0	2.7	2.8	3.4	5.3	3.0	1.6			35	31.6

Supporting Tables

Table S9. Representative set of substrates. Altogether 27 halogenated substrates were used to characterize the substrate specificity of the novel HLDs. The substrates were selected using multivariate statistics based on their diverse physico-chemical properties. The names of the selected substrates, classification and safety labelling are provided in **Table S10**.

	Chlorinated	Chlorinated & brominated	Brominated	Iodinated
Ethanes	**37** 	**137** 	**47** 	
Propanes	**67** 		**72** 	**28** 
	38 	**52** 	**48** 	**54** 
	80 	**155** 	**154** 	
Butanes	**4** 		**18** 	**29** 
Pentanes	**40** 			
Hexanes	**6** 		**20** 	**31** 
Cyclic alkanes	**138**  **115** 		**117**  **119** 	
Propenes	**209** 			
Miscellaneous alkanes	**111** 		**141** 	

Supporting Tables

Table S10. **Classification and labelling of selected halogenated substrates.** Globally Harmonized System of Classification and Labelling of Chemicals (GHS) with one of the following dangerous properties: Corrosive (GHS 05), Acute toxic (GHS 06), Health hazard (GHS 08), or Environmental hazard (GHS 09). The classification of carcinogenicity is based on the International Agency for Research on Cancer (IARC).

Substrate name (number)	GHS symbols	IARC carcinogenicity
1-bromobutane (#18)	09	–
1-bromohexane (#20)	09	–
1-iodobutane (#29)	06	–
1-iodohexane (#31)	05	–
1,2-dichloroethane (#37)	08	Group 2B
1,5-dichloropentane (#40)	06	–
1,2-dibromoethane (#47)	06, 08, 09	Group 2A
1,3-dibromopropane (#48)	09	–
1-bromo-3-chloropropane (#52)	06, 08	Group 2B
1,2-dichloropropane (#67)	08	Group 1
1,2,3-trichloropropane (#80)	08	Group 2A
bis(2-chloroethyl)ether (#111)	06, 08	Group 3
1-bromo-2-chloroethane (#137)	06, 08	–
4-bromobutyronitrile (#141)	06	–
1,2,3-tribromopropane (#154)	08	–
1,2-dibromo-3-chloropropane (#155)	06, 08	Group 2B
3-chloro-2-methylpropene (#209)	05, 09	Group 3

Supporting Tables

Table S11. Microfluidic analysis of substrate specificity. Specific activities of newly identified enzymes are depicted by a colour scale (from blue to red) in nmol.s^{-1}.mg^{-1}. All measurements were carried out in 5-10 replicates. The reactions were conducted in 1 mM HEPES buffer (pH 8.2) at a temperature close to optimal temperature (**Table S8**).

Enzyme	#4	#6	#18	#20	#28	#29	#31	#37	#38	#40	#47	#48	#52	#54	#67	#72	#80	#111	#115	#117	#119	#137	#138	#141	#154	#155	#209
DstA	n.d.	1.4	6.5	26.1	2.2	3.0	5.2	0.6	n.d.	0.5	0.5	2.5	0.6	n.d.	n.d.	0.3	n.d.	n.d.	0.5	0.4	n.d.	0.4	n.d.	n.d.	n.d.	n.d.	1.5
DfxA	0.5	n.d.	0.7	5.8	2.0	0.8	2.8	1.9	2.1	1.0	2.7	2.9	3.1	n.d.	n.d.	n.d.	0.5	n.d.	1.0	n.d.	n.d.	n.d.	n.d.	n.d.	0.4	0.5	2.0
DlaA	4.9	2.6	5.7	4.0	n.d.	1.1	1.0	1.0	n.d.	n.d.	2.0	3.9	n.d.	n.d.	n.d.	1.1	n.d.	0.6	1.6	1.2	n.d.	0.9	n.d.	n.d.	n.d.	n.d.	0.3
DaxA	n.d.	n.d.	0.2	0.6	1.1	1.3	1.4	0.9	0.5	n.d.	1.4	1.1	2.0	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	0.2	1.0
DsmA	3.3	4.6	29.8	114.5	24.4	29.0	30.4	1.0	n.d.	11.2	30.3	86.1	27.9	6.1	n.d.	5.9	n.d.	0.3	0.7	1.4	1.2	5.5	0.9	1.7	4.9	1.2	73.1
DmmarA	n.d.	0.9	6.0	8.4	2.6	5.5	2.9	0.7	n.d.	n.d.	1.9	5.9	5.2	5.8	n.d.	1.3	n.d.	n.d.	0.8	n.d.	n.d.	0.6	n.d.	n.d.	n.d.	n.d.	0.6
DathA	n.d.	0.4	1.8	3.6	2.2	n.d.	n.d.	0.3	n.d.	0.4	2.6	5.7	n.d.	2.8	0.3	0.2	n.d.	n.d.	n.d.	n.d.	0.4	0.6	n.d.	0.6	n.d.	n.d.	3.3
DmaA	20.0	8.8	53.3	9.9	79.1	105.4	20.0	30.3	4.8	12.1	184.2	211.1	202.9	59.4	1.9	62.2	n.d.	n.d.	2.7	23.4	15.7	177.1	11.2	35.4	58.4	30.5	147.2
DspoA	n.d.	79.0	568.3	176.8	749.7	668.8	177.4	n.d.	17.5	279.7	341.6	860.7	942.1	861.0	0.5	184.3	0.6	7.7	n.d.	17.3	55.2	248.5	22.9	164.0	50.6	14.1	633.4
DexA	39.9	297.5	705.0	553.5	661.8	574.3	579.0	n.d.	9.4	443.4	469.3	572.7	700.0.	313.7	n.d.	151.7	n.d.	n.d.	n.d.	27.5	168.5	82.7	155.7	78.5	54.9	165.6	779.2
DppsA	3.7	9.1	76.4	77.8	21.6	30.5	70.7	1.7	1.5	17.3	32.7	29.0	24.7	20.9	0.6	2.9	n.d.	n.d.	0.4	n.d.	1.8	10.4	n.d.	1.9	1.6	n.d.	42.5
DeaA	7.2	81.1	227.3	286.8	263.4	242.2	323.9	1.3	9.3	233.8	85.1	405.0	423.4	401.8	n.d.	22.3	n.d.	7.7	n.d.	14.6	9.5	66.0	20.9	290.2	1.9	n.d.	257.6
DmgaA	3.5	3.5	5.7	6.5	1.7	3.5	2.9	2.4	1.5	n.d.	5.5	6.1	1.6	1.7	1.6	3.1	2.6	1.6	3.4	1.7	0.5	3.3	0.5	2.4	2.9	n.d.	1.9
DprxA	48.6	176.4	636.3	431.9	515.6	584.1	508.2	1.4	32.0	385.5	256.8	630.1	695.5	308.2	n.d.	127.1	n.d.	5.8	n.d.	9.0	28.4	94.1	42.7	144.5	55.5	8.6	569.1
DrgA	6.4	4.8	3.9	5.6	n.d.	1.1	n.d.	4.9	2.6	10.9	17.2	1.8	n.d.	2.4	7.2	5.4	4.7	5.8	6.5	3.9	4.3	6.1	4.3	6.6	6.2	4.3	n.d.
DmbaA	45.3	8.0	105.7	19.1	151.6	101.9	89.5	13.9	55.2	80.7	176.0	132.5	219.0	118.2	6.1	83.9	8.4	44.7	0.6	43.3	0.9	88.4	47.2	121.8	192.1	110.5	187.8
DthA	4.9	3.1	2.8	2.4	8.1	7.4	n.d.	7.7	12.0	2.8	217.0	31.3	14.5	12.8	n.d.	46.9	n.d.	10.9	1.1	1.5	1.4	184.7	1.2	11.4	7.6	n.d.	11.4
DphxA	83.5	32.8	304.4	32.7	346.8	338.7	30.9	14.5	42.5	189.5	679.7	595.7	412.5	498.1	7.5	415.5	5.4	45.4	5.9	77.6	17.2	168.6	97.1	523.2	123.0	55.1	468.5
DthB	n.d.	5.3	51.2	43.0	93.1	76.1	39.6	n.d.	n.d.	81.1	38.0	121.8	155.9	145.7	n.d.	11.4	n.d.	2.2	n.d.	1.9	3.8	31.9	4.6	40.9	12.6	5.7	99.0
DnbA	4.3	2.9	4.1	5.6	6.6	4.2	1.5	1.7	n.d.	3.2	7.8	6.5	4.8	2.5	0.4	4.4	0.4	n.d.	1.6	0.7	1.1	4.4	0.8	4.7	1.4	0.2	1.4
DhxA	16.6	77.2	326.3	148.9	168.4	223.1	183.6	20.0	63.7	149.3	452.2	610.8	560.0	143.2	5.5	81.5	n.d.	n.d.	n.d.	n.d.	52.5	97.0	23.3	30.8	80.6	44.0	177.2
DspxA	34.9	23.7	63.9	24.3	82.3	64.4	36.1	32.3	48.3	42.8	243.6	81.9	81.0	69.7	7.5	45.2	n.d.	20.5	n.d.	33.9	10.5	83.8	27.3	54.3	73.5	64.4	285.0
DchA	9.2	5.6	43.4	4.6	29.6	43.8	n.d.	2.2	7.7	9.2	208.6	143.3	99.7	34.6	n.d.	38.4	1.0	11.6	2.1	2.9	5.2	178.2	9.2	59.7	55.2	5.3	22.6
Dcta	1.7	0.5	n.d.	2.0	2.6	2.0	3.9	3.0	2.4	4.7	3.4	5.0	3.4	n.d.	n.d.	n.d.	1.3	n.d.	2.0	n.d.	0.6	n.d.	n.d.	n.d.	1.0	1.4	2.8

n.d. –not detected under the tested conditions (the activity values were below the detection limit)

Supporting Tables

Table S12. The assessment of substrate specificity range. The substrate specificity range quantifies the relative number of substrates converted by the particular enzyme based on testing with 27 halogenated substrates (**Table S9** and **S11**). Red bars encode the broader specificity enzymes and blue bars narrower specificity enzymes. The names of benchmark HLDs are in grey.

Enzyme	Substrate specificity range (%)
DmbaA	100
DphxA	100
DhaA	96
DbjA	93
DchA	93
DmaA	93
DmgaA	93
DspxA	93
DnbA	89
DprxA	89
DsmA	89
DspoA	89
DeaA	85
DhlA	85
DhxA	85
DrgA	85
DthA	85
LinB	85
DatA	81
DbeA	81
DexA	81
DppsA	81
DthB	78
DmbA	74
DctA	67
DfxA	63
DstA	59
DathA	56
DmmarA	56
DlaA	56
DrbA	56
DaxA	44

Supporting Tables

Table S13. The steady-state kinetics and thermodynamics parameters. The values of the parameters ± standard errors obtained from the global fit (reference temperature 310.15 K). The lower and upper limits for each parameter were derived from the confidence contours for the χ^2 threshold at boundary 0.95 (Figure S11).

	k_1 (k_{cat}/K_m) ($mM^{-1}.s^{-1}$)	k_2 (k_{cat}) (s^{-1})	k_{-3} (x 1000 s^{-1})	K_m (k_2/k_1) (mM)	K_P ($1000/k_{-3}$) (mM)
DspoA	254 ± 8	13.2 ± 0.5	11.3 ± 0.7	0.052 ± 0.003	0.09 ± 0.01
lower; upper limit	228; 290	11.8; 15.3	9.1; 14.4		
E_a ($kJ.mol^{-1}$)	29 ± 1	88 ± 2	n.d.		
lower; upper limit	24; 35	80; 99	n.d.		
DexA	13.3 ± 0.1	64 ± 5	1.95 ± 0.05	4.8 ± 0.4	0.51 ± 0.1
lower; upper limit	13.0; 13.7	51; 86	1.78; 2.14		
E_a ($kJ.mol^{-1}$)	17.8 ± 0.8	82 ± 5	n.d.		
lower; upper limit	15.5; 20.1	66; 99	n.d.		
DeaA	90 ± 10	23 ± 2	6.3 ± 0.8	0.26 ± 0.04	0.16 ± 0.02
lower; upper limit	67; 141	17.3; 35.6	4.1; 10.5		
E_a ($kJ.mol^{-1}$)	80 ± 7	61 ± 7	101 ± 7		
lower; upper limit	54; 109	37; 89	72; 126		
DprxA	130 ± 20	80 ± 10	10.7 ± 1.3	0.6 ± 0.1	0.09 ± 0.01
lower; upper limit	93; 218	48; 184	7.0; 18.3		
E_a ($kJ.mol^{-1}$)	64 ± 10	40 ± 20	60 ± 10		
lower; upper limit	25; 96	0; 90	12; 91		
DphxA	290 ± 30	54 ± 3	9.5 ± 0.9	0.19 ± 0.02	0.11 ± 0.1
lower; upper limit	241; 538	44; 64	7.7; 18.1		
E_a ($kJ.mol^{-1}$)	55 ± 7	95 ± 4	55 ± 8		
lower; upper limit	39; 97	79; 106	37; 96		
DhxA	147 ± 8	74 ± 4	10.6 ± 0.6	0.50 ± 0.04	0.09 ± 0.01
lower; upper limit	122; 177	63; 92	8.6; 13.1		
E_a ($kJ.mol^{-1}$)	32 ± 4	33 ± 6	65 ± 5		
lower; upper limit	17; 46	11; 52	48; 80		

n.d. not determined

Supporting Tables

Table S14. Determination of oligomeric states. The proportions of oligomeric states (in %) of novel HLDs were analyzed by gel filtration.

Enzyme	Oligomeric state			
	Monomer (%)	Dimer (%)	Tetramer (%)	Hexamer (%)
DstA	25	75	-	-
DfxA	100	-	-	-
DlaA	100	-	-	-
DaxA	100	-	-	-
DsmA	100	-	-	-
DmmarA	4	96	-	-
DathA	100	-	-	-
DmaA	100	-	-	-
DspoA	96	4	-	-
DexA	96	4	-	-
DppsA	100	-	-	-
DeaA	100	-	-	-
DmgaA	100	-	-	-
DprxA	53	16	13	17
DrgA	100	-	-	-
DmbaA	100	-	-	-
DthA	100	-	-	-
DphxA	100	-	-	-
DthB	100	-	-	-
DnbA	100	-	-	-
DhxA	94	6	-	-
DspxA	100	-	-	-
DchA	100	-	-	-
DctA	100	-	-	-
DstA	100	-	-	-

Supporting Figures

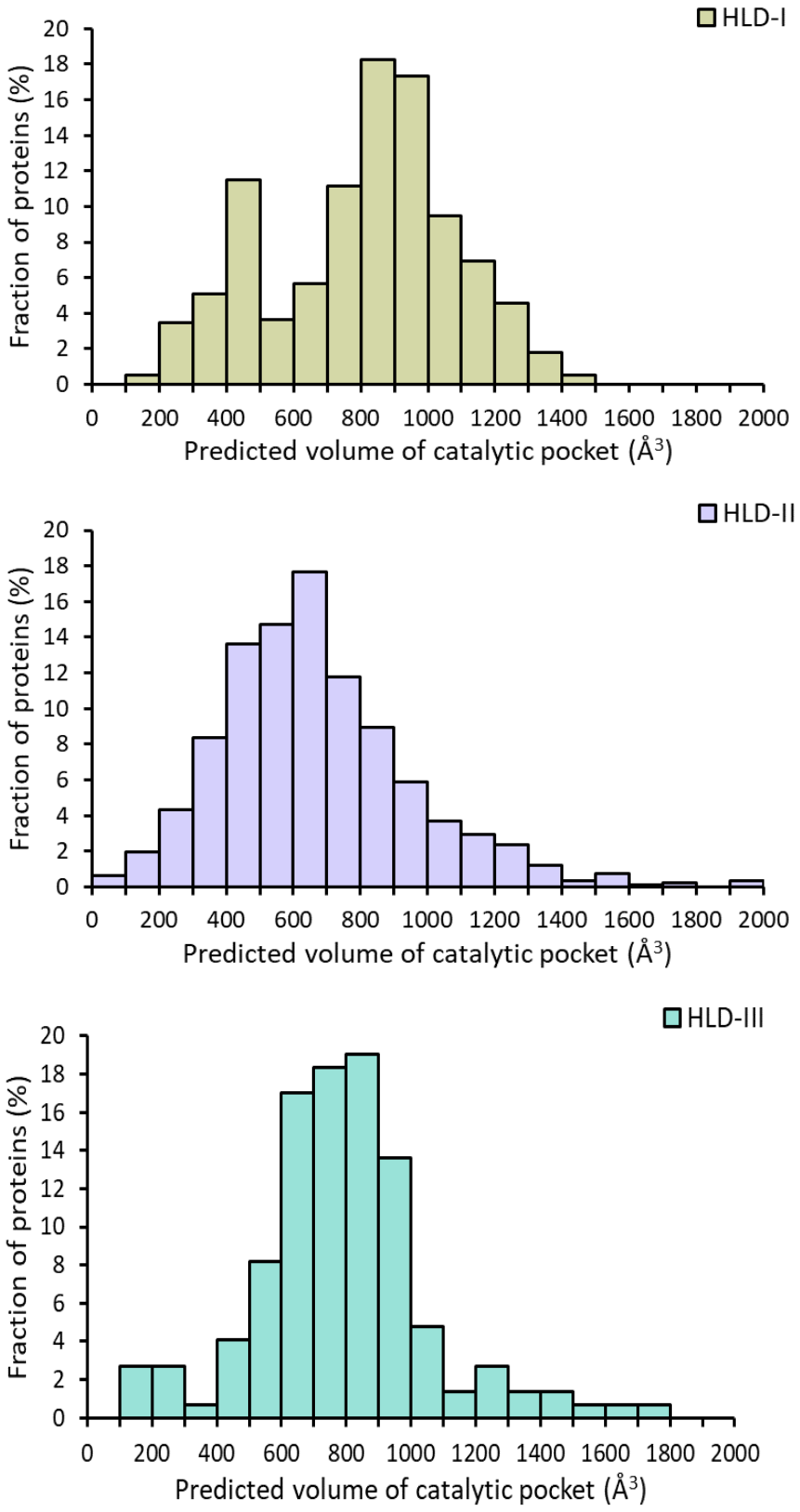


Figure S1. Estimated volumes of catalytic pockets in different HLD subfamilies. The analysis was performed for homology models with template scores ≤ 100. The volumes of the active sites strongly depend on the correct prediction of the protein backbone as well as side chains of the first shell residues and must therefore be interpreted carefully.

Supporting Figures

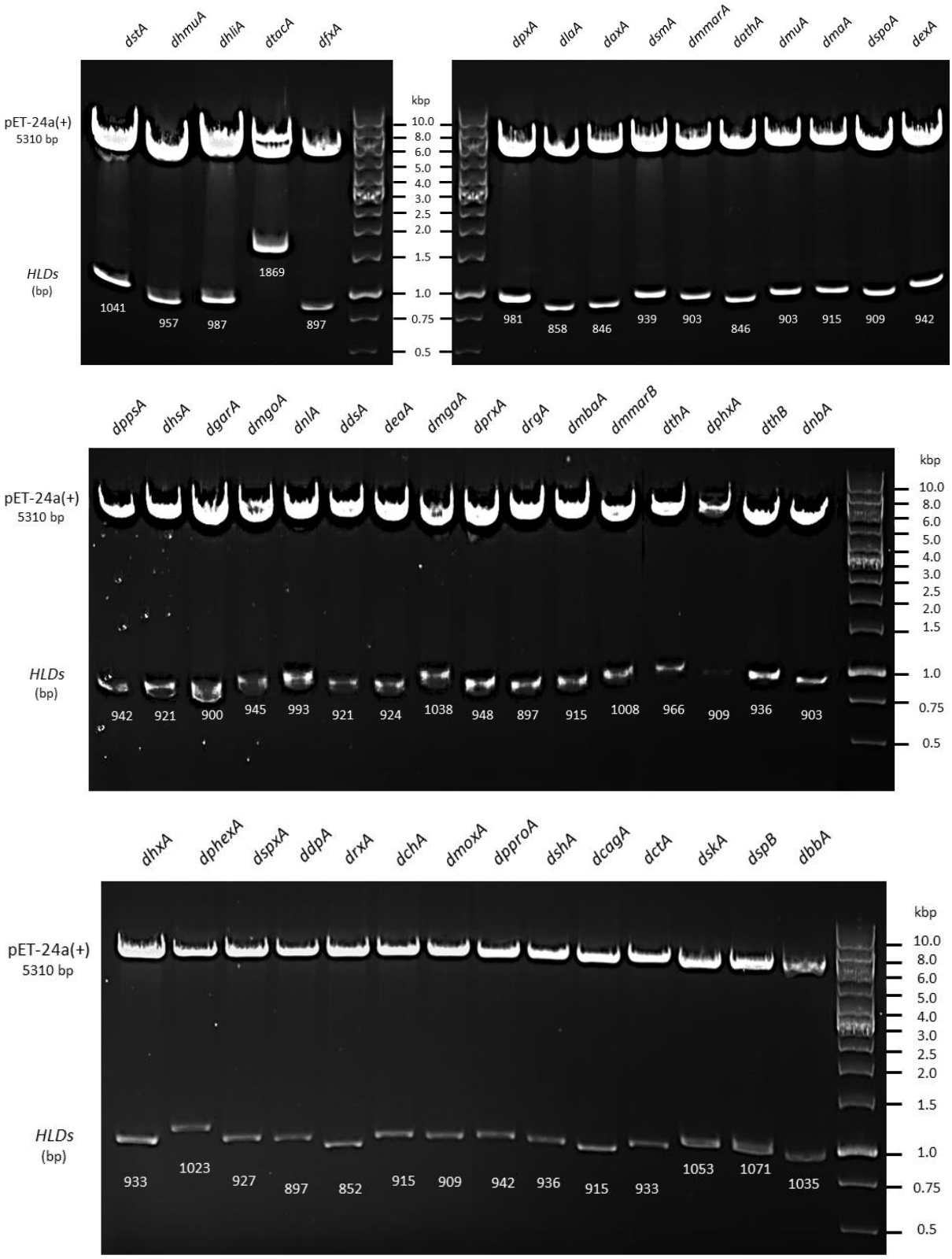


Figure S2. Restriction analysis of 45 recombinant pET-24a::*HLD* plasmids. NdeI/XhoI restriction enzymes were used for verification. pET-24(+) – 5310 bp, *HLD* genes (bp).

Supporting Figures

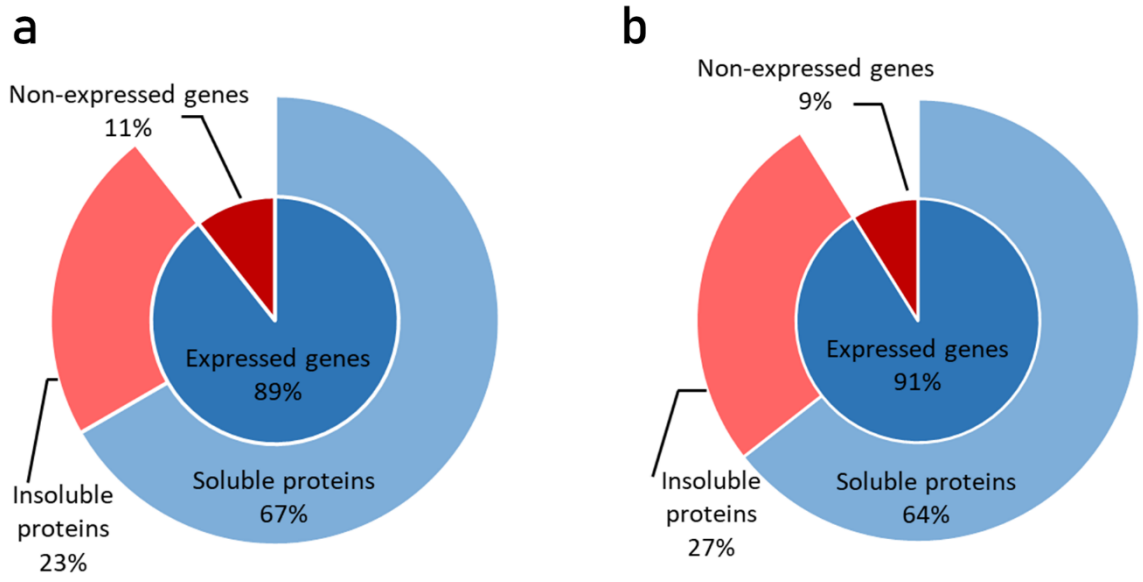


Figure S3. Small scale expression analysis. Quantitative comparison of *in vivo* (**a**) and *in vitro* (**b**) expression analysis.

Supporting Figures

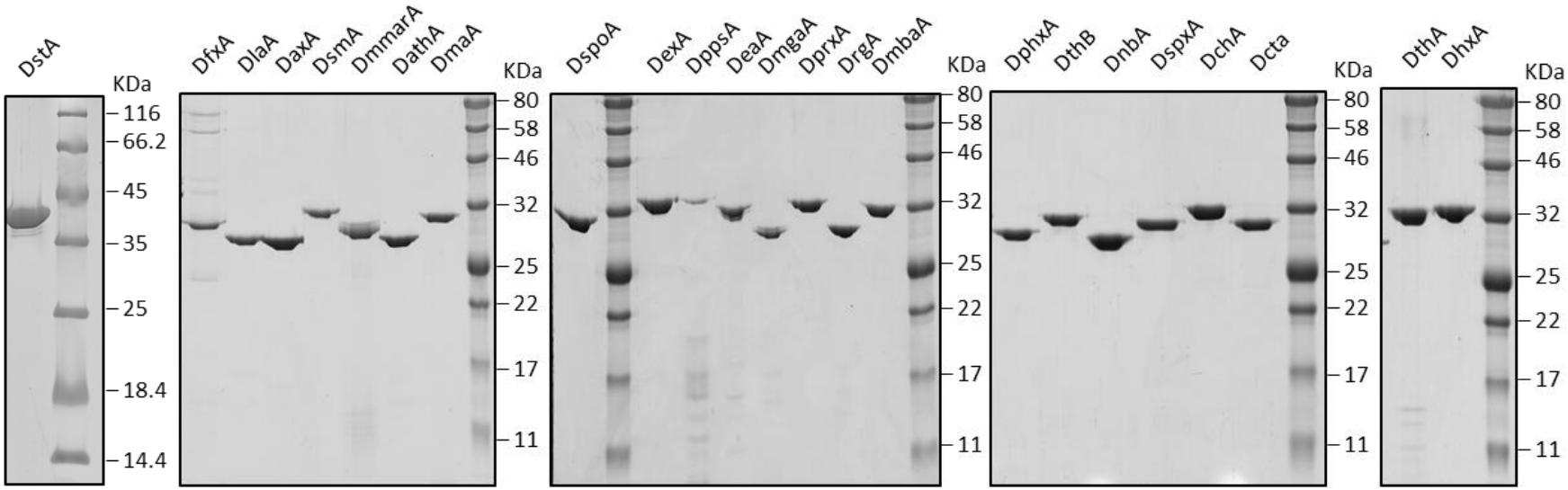


Figure S4. Analysis of protein purity. SDS-PAGE analysis of purified HLDs with a characteristic monomeric size between 28 to 37 kDa. M: Color Pre-stained Protein Standard, Broad Range (11–245 kDa) (New England Biolabs).

Supporting Figures

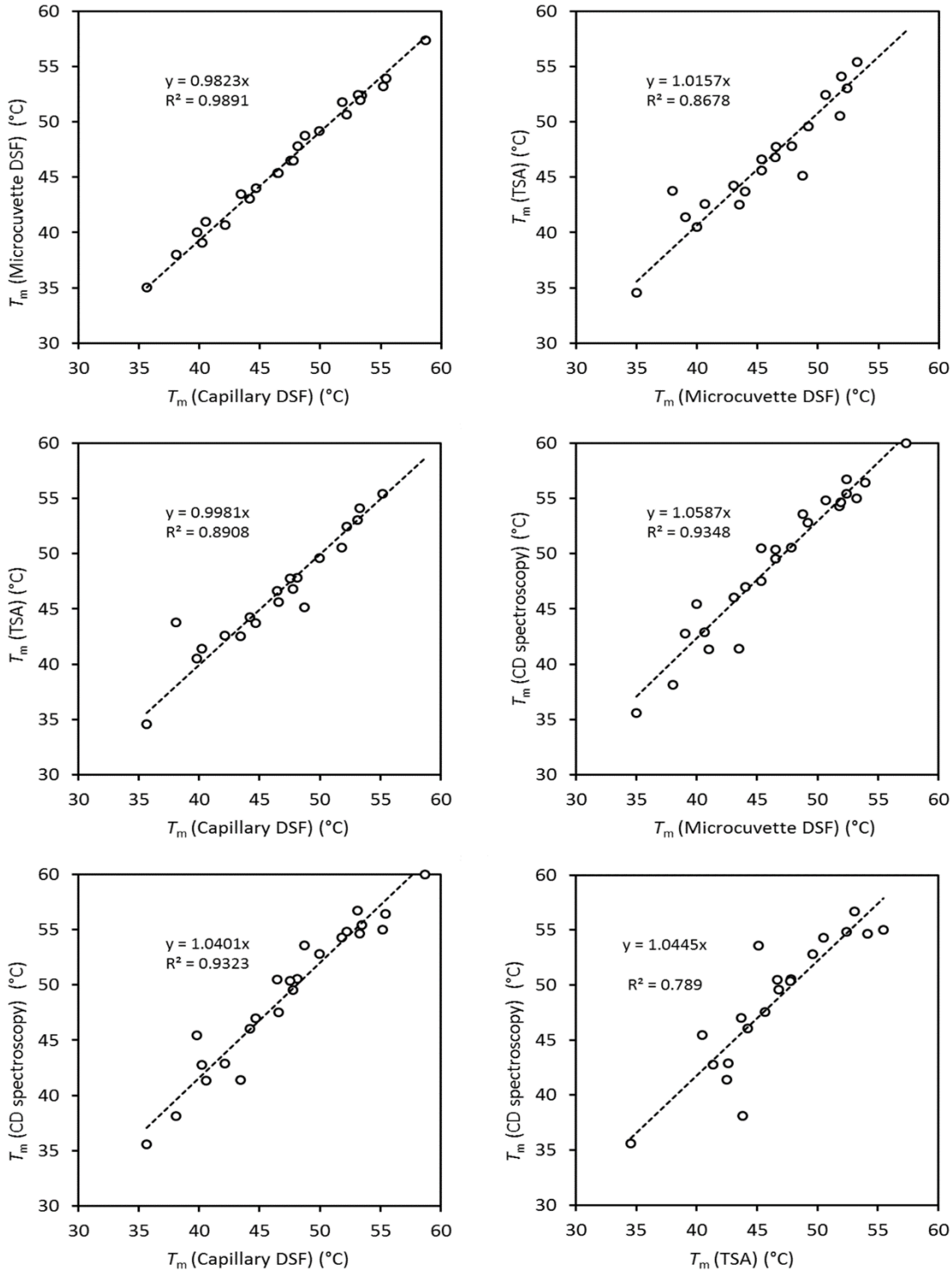


Figure S5. Comparison of different methods for thermostability analysis. The correlation between the thermostability datasets obtained by (i) Thermal shift assay (TSA), (ii) capillary differential scanning fluorimetry (DSF), (iii) microcuvette DSF, and (iv) circular dichroism (CD) spectroscopy. The Pearson correlation coefficient is reported.

Supporting Figures

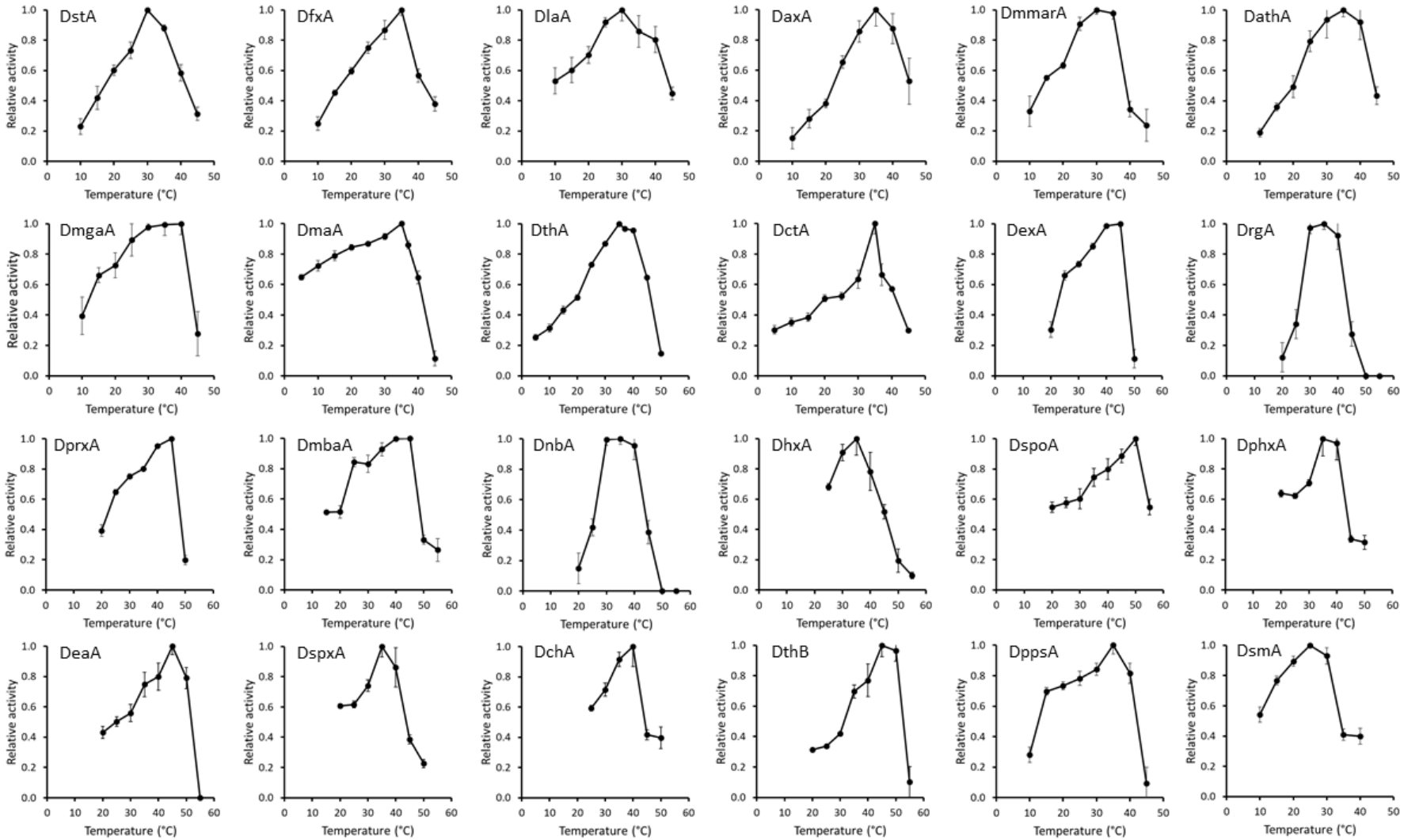


Figure S6. Microfluidic analysis of temperature profiles of HLD activities. Temperature profiles expressed as relative activities of HLDs in the temperature range from 5 to 55 °C. The relative activities are related to the maximum activity of each dataset (set to 1). Error bars represent the standard deviation from at least 5 experimental replicates. The reactions were conducted in 1mM HEPES buffer at pH 8.2.

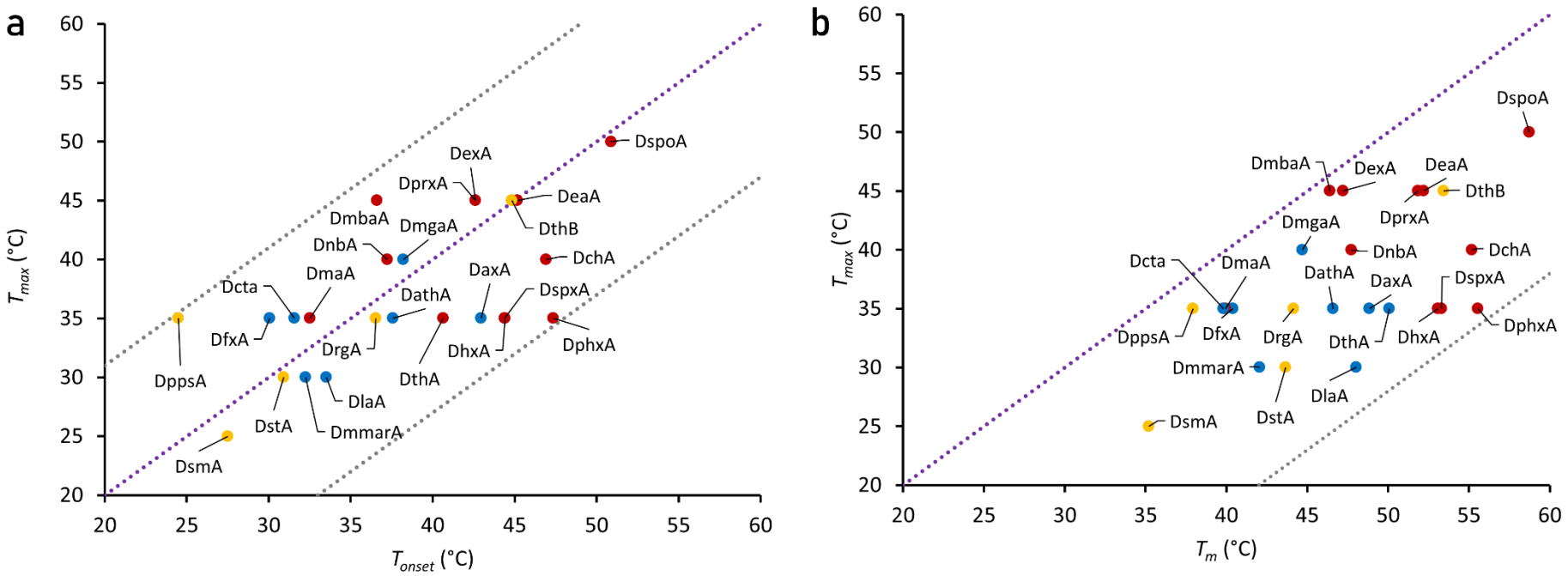
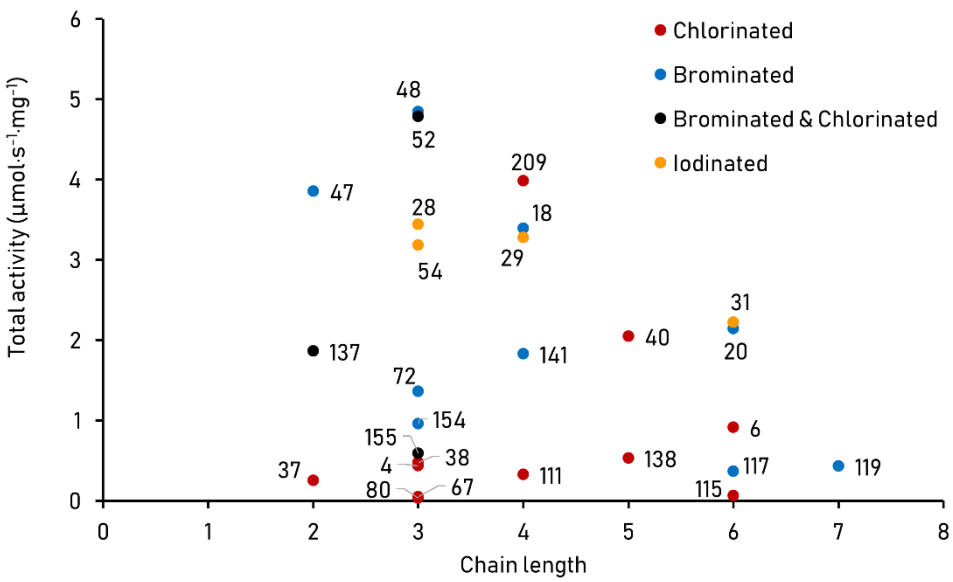


Figure S7. Analysis of thermostability and temperature profiles. The temperature of maximum activity (T_{max}) was compared (**a**) with the temperature at which protein denaturation starts (T_{onset}) and (**b**) with the melting temperature (T_m). The individual enzymes are coloured according to the level of activity at T_{max}: hundreds (blue), tens (yellow) and units (red) of nmol·s^{-1}·mg^{-1}. The purple dotted line stands for y=x correlation, while the grey dotted lines display the maximum deviation from the y=x correlation, specifically +12 °C and -14°C for **a** and -22°C for **b**.

Supporting Figures

a



b

No.	Substrate name	Convertibility	No.	Substrate name	Convertibility	No.	Substrate name	Convertibility
47	1,2-dibromoethane	100	137	1-bromo-2-chloroethane	91	119	(1-bromomethyl)cyclohexane	75
20	1-bromohexane	100	52	1-bromo-3-chloropropane	88	37	1,2-dichloroethane	75
48	1,3-dibromopropane	97	31	1-iodohexane	88	38	1,3-dichloropropane	72
18	1-bromobutane	97	40	1,5-dichloropentane	84	138	chlorocyclopentane	72
29	1-iodobutane	97	154	1,2,3-tribromopropane	81	155	1,2-dibromo-3-chloropropane	63
209	3-chloro-2-methylpropene	94	141	4-bromobutyronitrile	81	111	bis(2-chloroethyl)ether	56
6	1-chlorohexane	94	4	1-chlorobutane	81	115	chlorocyclohexane	56
28	1-iodopropane	94	54	1,3-diidopropane	81	67	1,2-dichloropropane	41
72	1,2-dibromopropane	91	117	bromocyclohexane	78	80	1,2,3-trichloropropane	34

Figure S8. Analysis of substrate preferences. **a,** Dependence of total activity (sum of specific activities of 24 newly characterized and 8 benchmark HLDs) on the chain length of the halogenated substrate. The individual colours of the data points stand for the type of substrate: chlorinated (red), brominated (blue), brominated & chlorinated (black), and iodinated (orange). **b,** Convertibility of the substrate quantifies the fraction of HLDs which convert a particular substrate.

Supporting Figures

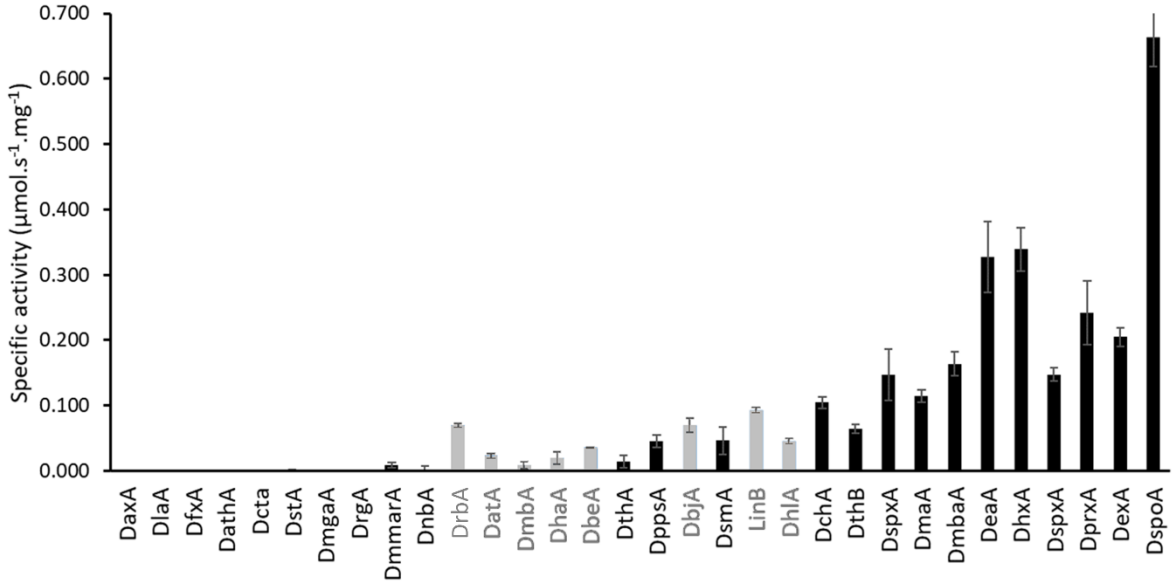


Figure S9. **Validation of the specific activities**. The specific activities of HLDs towards 1,3-dibromopropane were analyzed by the conventional analytical method based on the detection of halide product using the Iwasaki assay.[1] The reactions were conducted in 100 mM glycine buffer (pH 8.6) at temperatures close to the optimal temperature of each enzyme (**Table S8**). The values are averages of triplicate measurements.

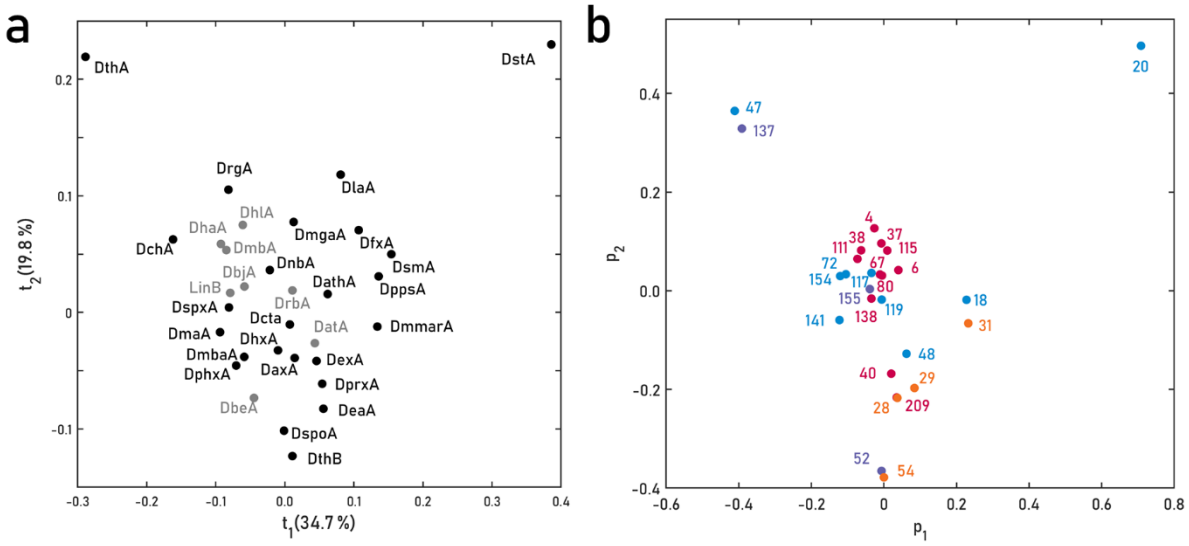


Figure S10. **Principal component analysis. a,** Comparison of the score plot t_1/t_2 and the corresponding loading plot p_1/p_2 of the log-transformed specific activities. The score plot is a two-dimensional projection into the multidimensional space, where HLDs with similar or unique substrate specificity profiles can be visualized. The t_1/t_2 score plot describes 54.6% of the variance in the dataset. Previously known HLDs and newly identified enzymes are depicted by black and grey colour, respectively. **b,** The corresponding loading plot p_1/p_2 from the principal component analysis shows the substrates that govern the distribution of HLDs on the two-dimensional window. Variables distinct from the origin contribute to the principal component more than the variables localized close to the origin of the plot. Brominated substrates are shown in blue, chlorinated in red, iodinated in orange, and brominated+chlorinated in violet.

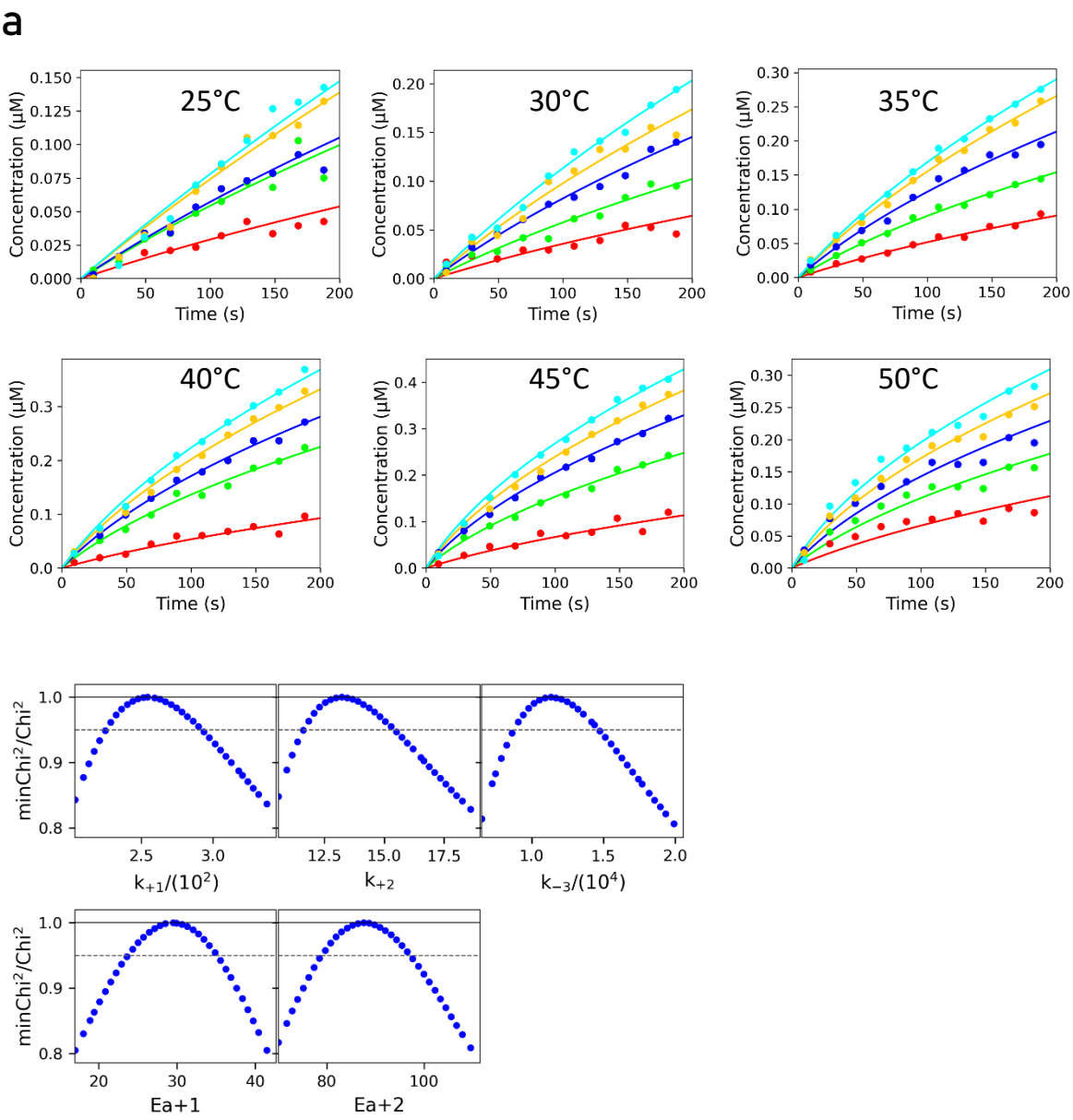


Figure S11. Global fitting of microfluidic kinetic and thermodynamic data. The time traces were obtained by monitoring the enzymatic reaction in the temperature range from 25 to 50 °C in 5 °C increments, each at six different substrate concentrations (0 - 1 mM 1,3-dibromopropane) using droplet-based microfluidics. The complex concentration and temperature dependence datasets were recorded for selected enzymes DspoA (**a**), DexA (**b**), DeaA (**c**), DprxA (**d**), DphxA (**e**), DhxA (**f**). All reactions were performed in 1 mM HEPES buffer, pH 8.2. Each data point represents an average of 20 repetitions, the solid lines represent the best global fit. Confidence contour analysis was carried out for the fitted parameters of the kinetic constants and energy barriers obtained from the global fit. The distribution graphs (blue dots) represent the dependence of the error on the individual fitted parameter (either *k* or E_a) while varying all other parameters to achieve the best fit. The dashed line shows the $\chi 2$ threshold 0.95 used to establish confidence intervals as reported in the **Supporting Table S13**.

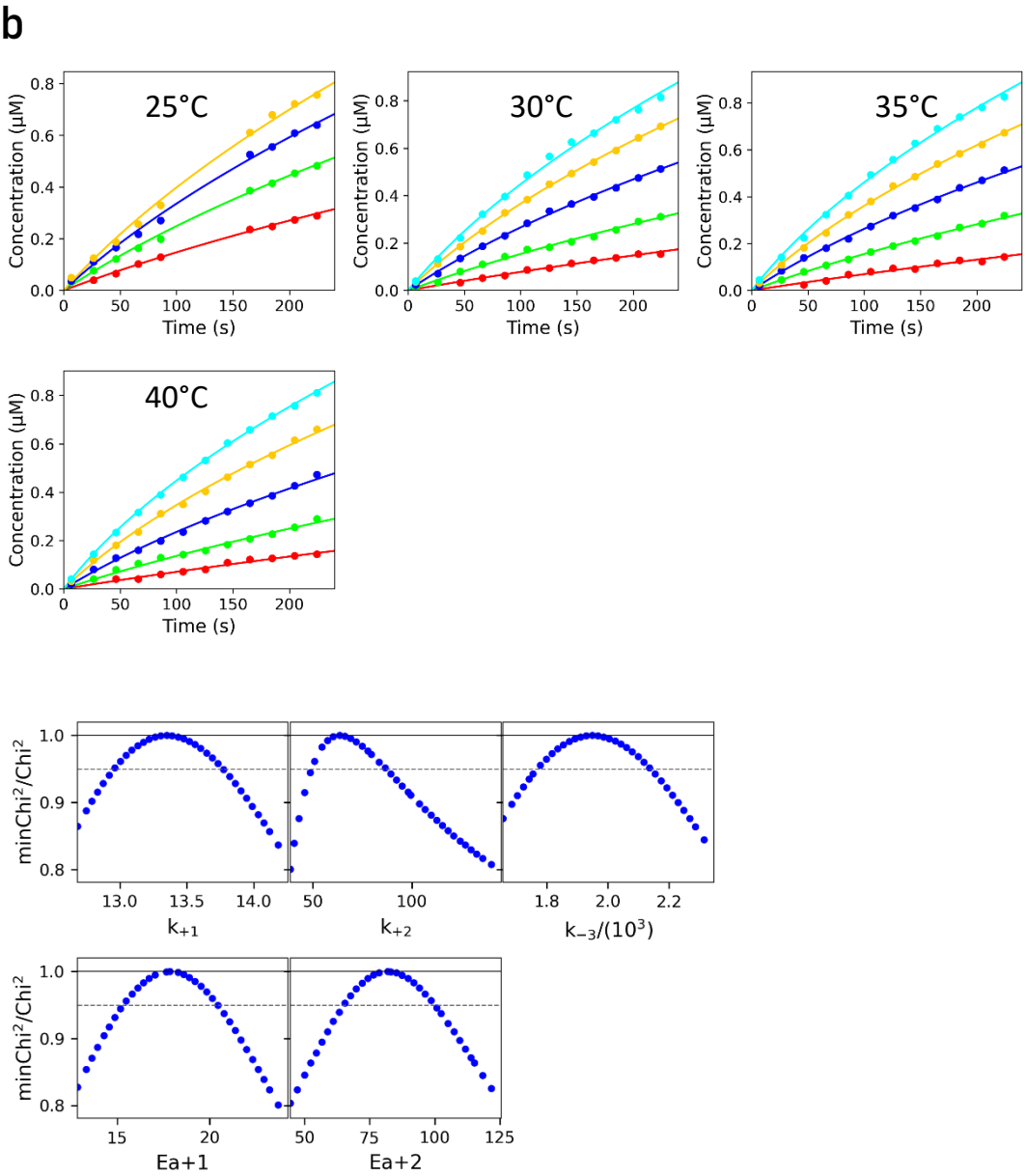


Figure S11 Global fitting of microfluidic kinetic and thermodynamic data. DexA (Continued)

C

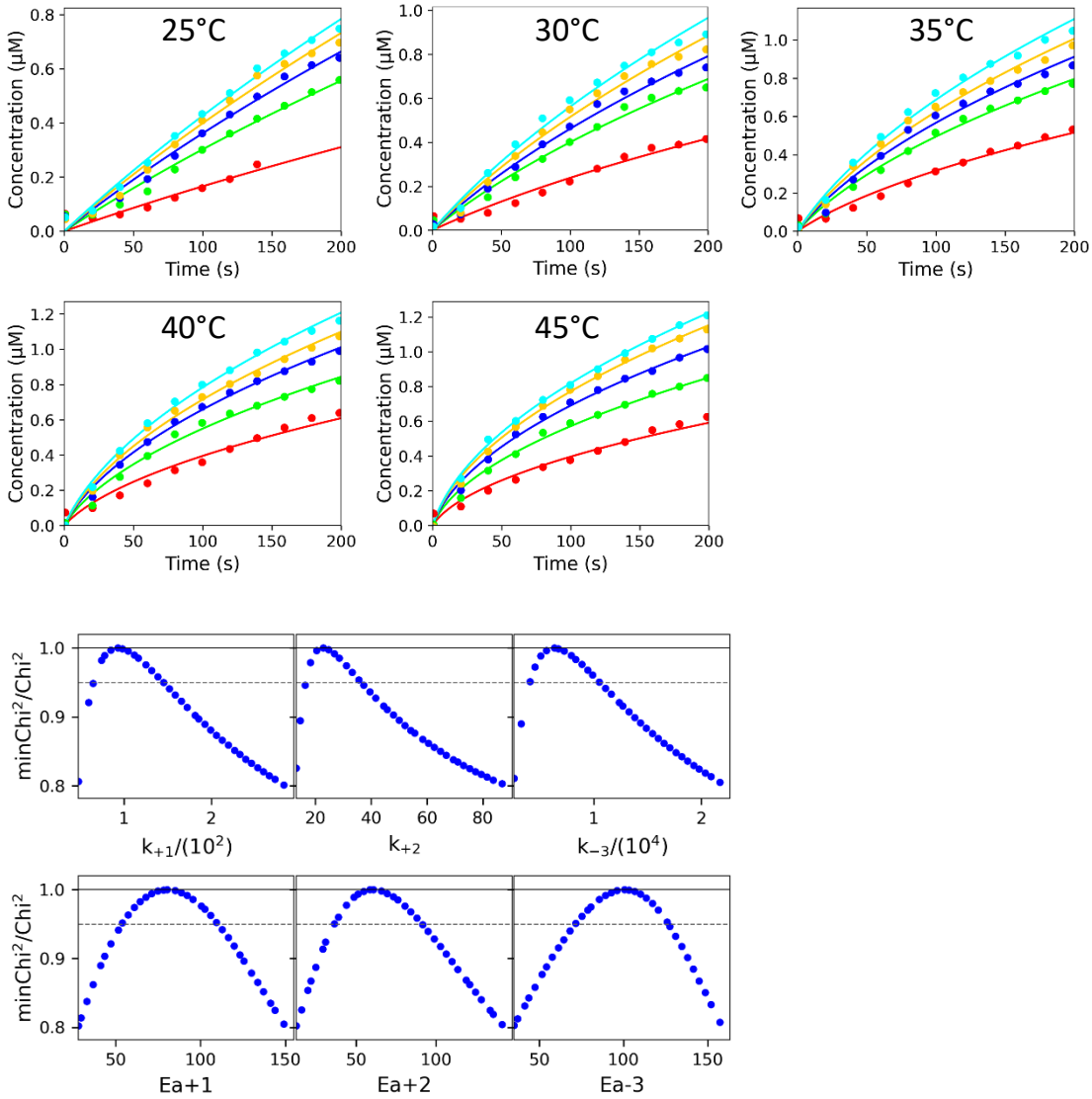


Figure S11 Global fitting of microfluidic kinetic and thermodynamic data. DeaA. (Continued)

d

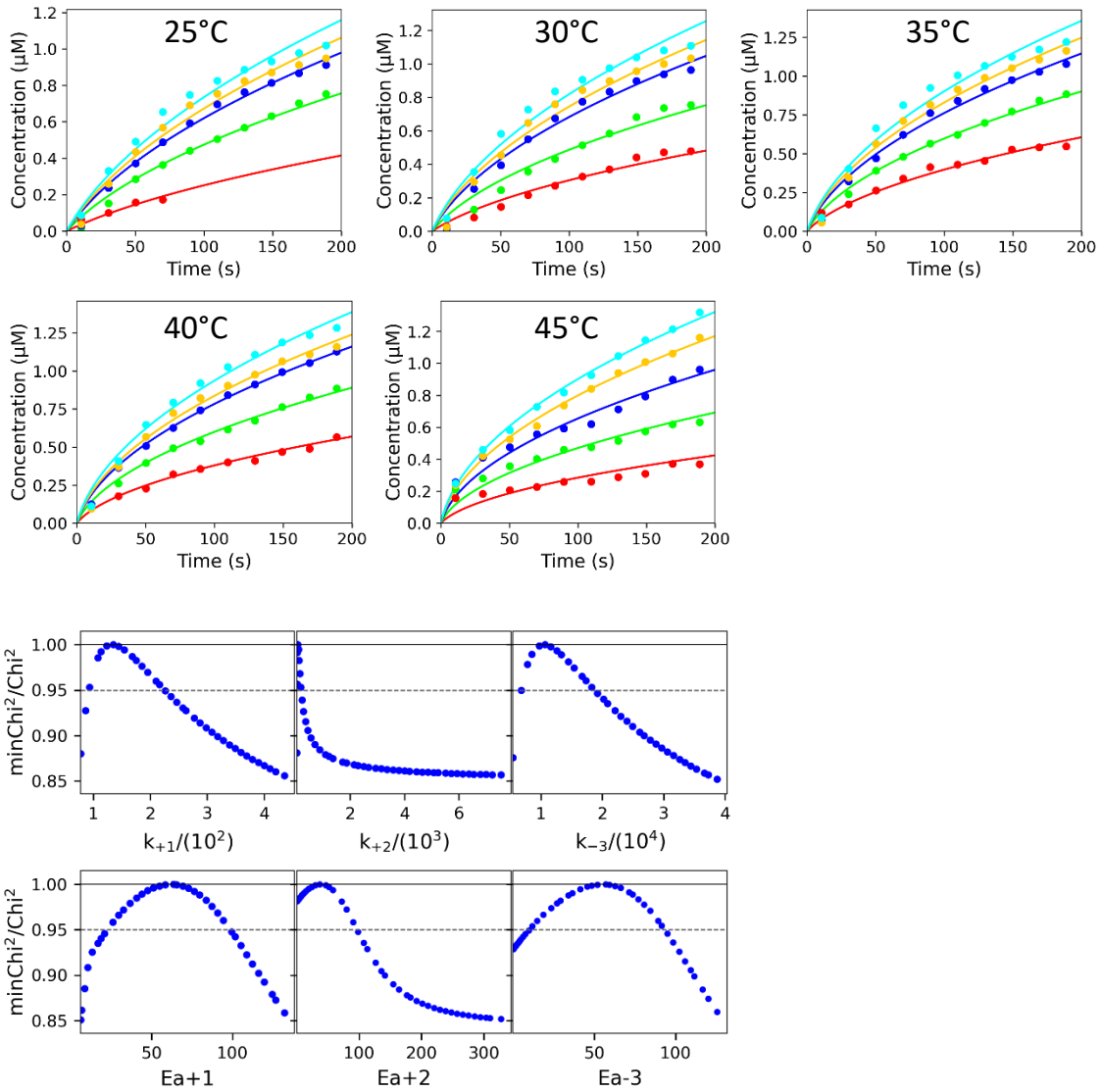


Figure S11 Global fitting of microfluidic kinetic and thermodynamic data. DprxA. (Continued)

e

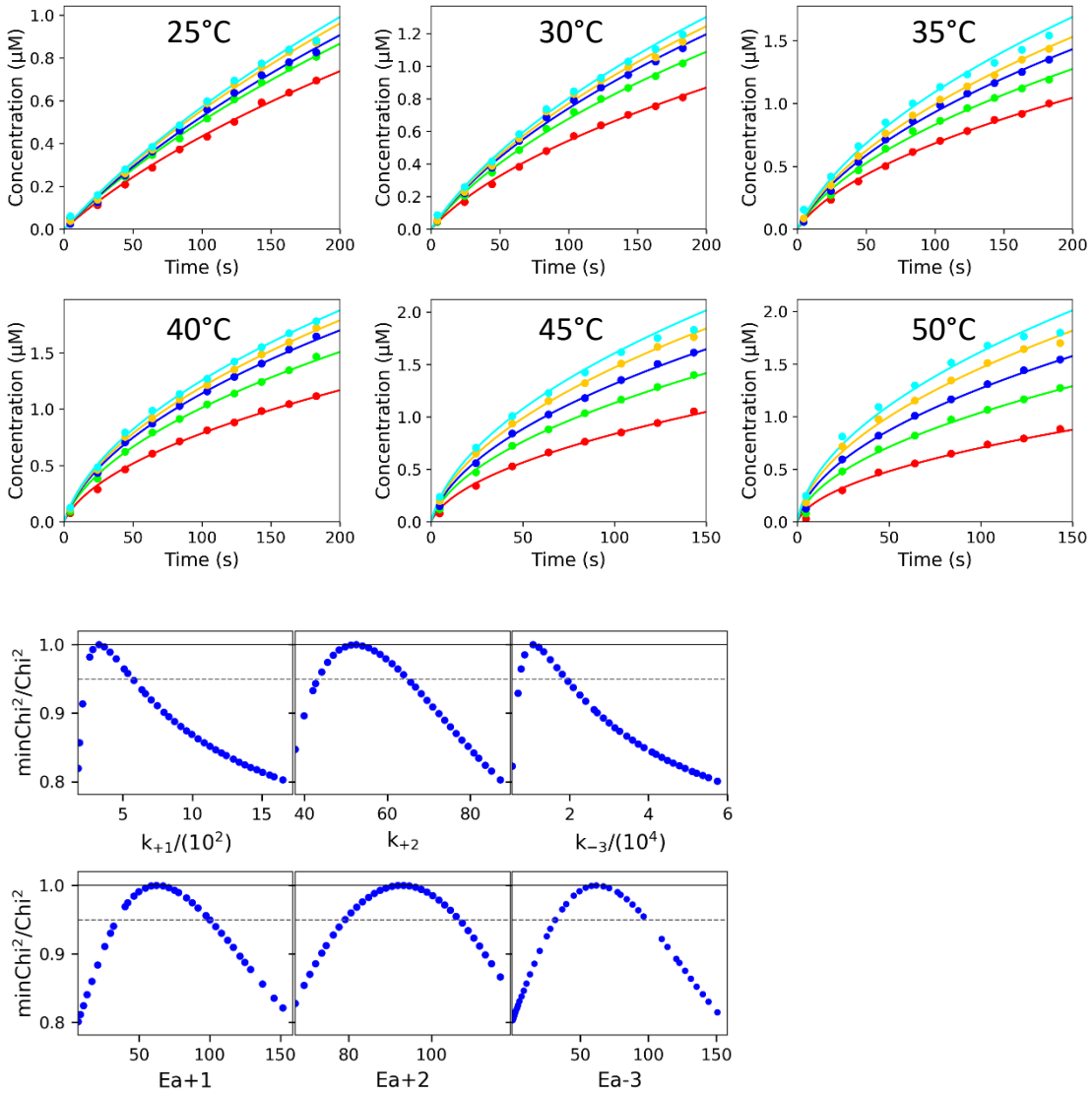


Figure S11 Global fitting of microfluidic kinetic and thermodynamic data. DphxA. (Continued)

f

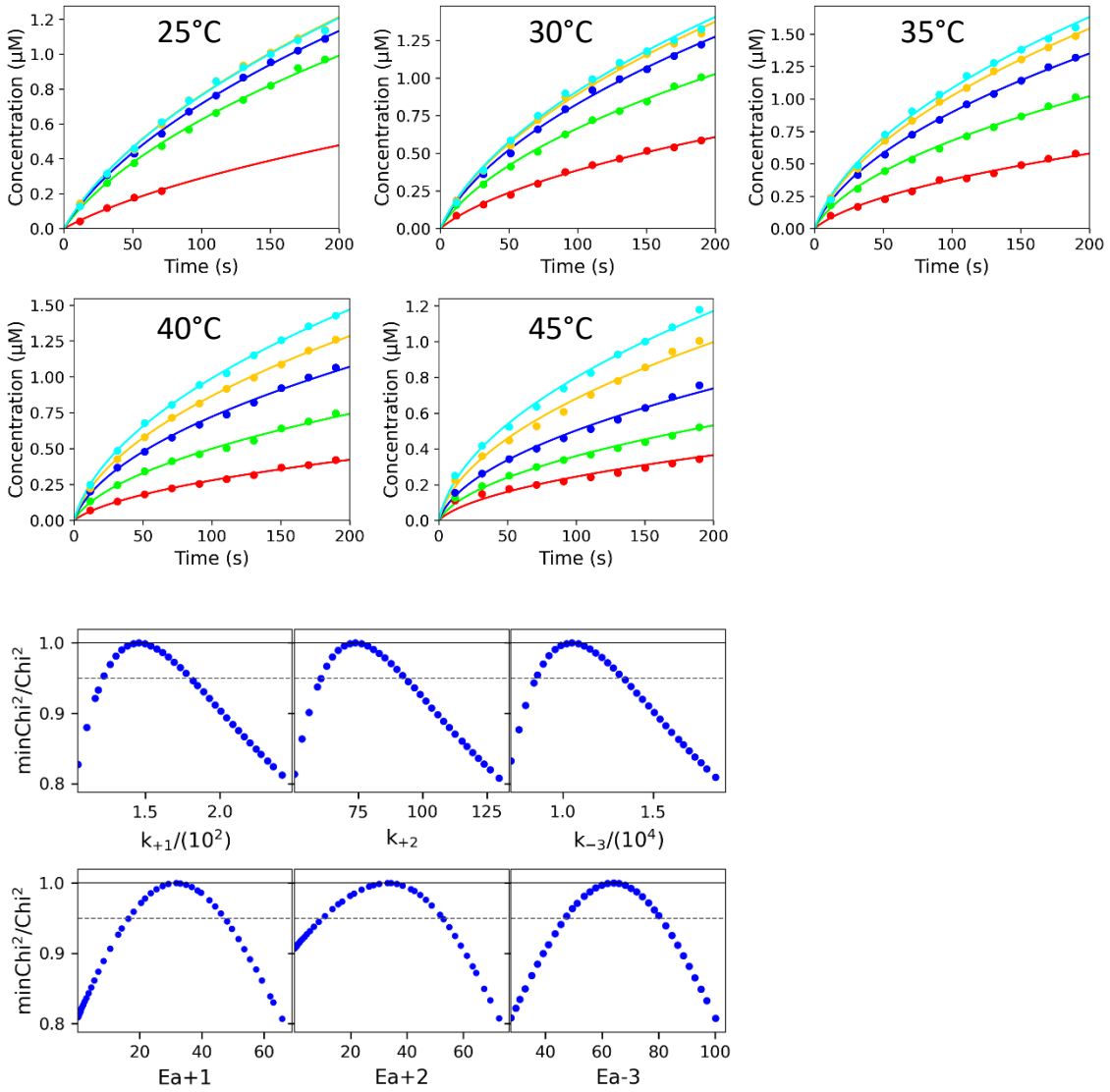


Figure S11 Global fitting of microfluidic kinetic and thermodynamic data. DhxA. (Continued)

Supporting Figures

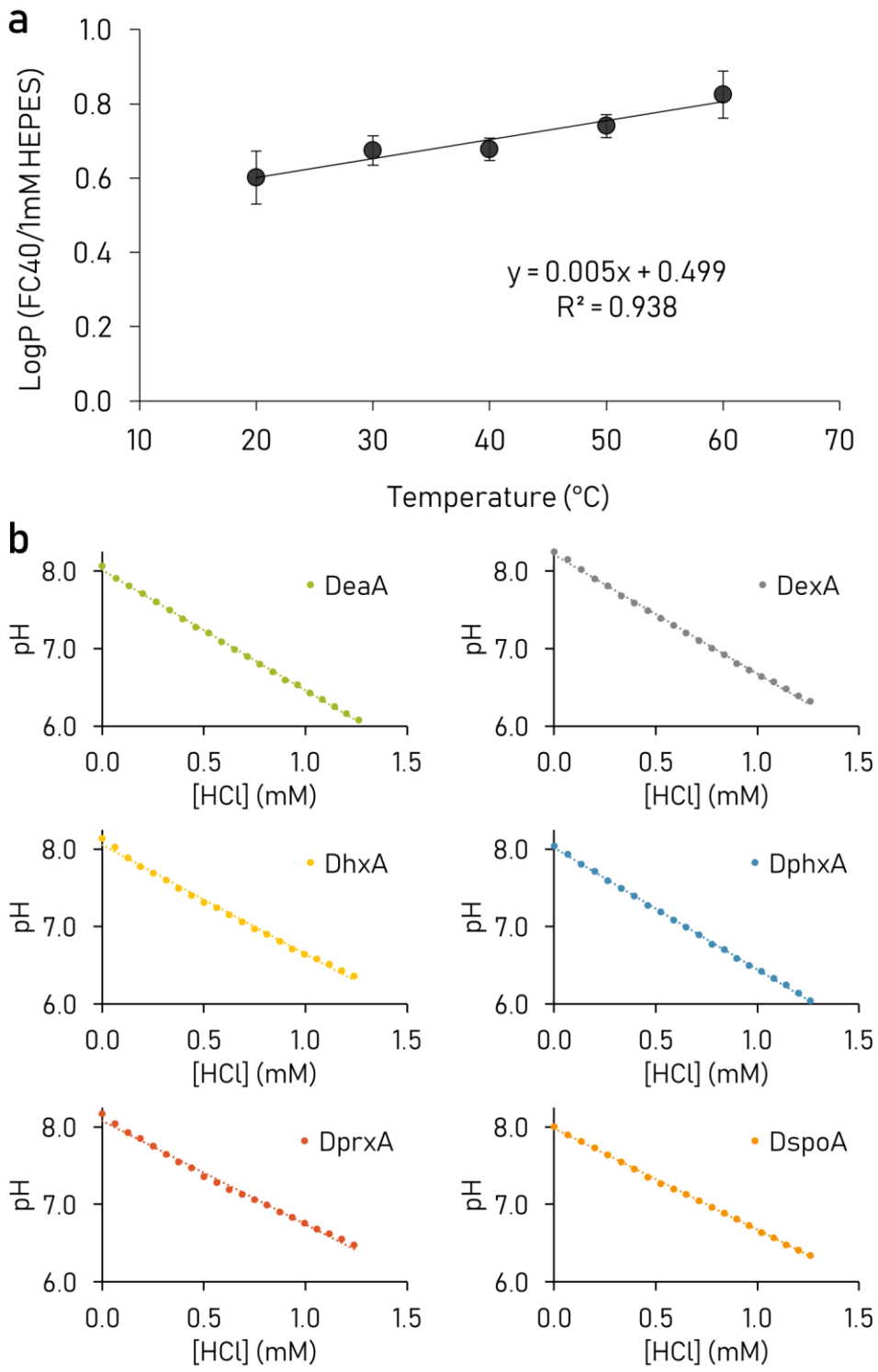


Figure S12. Calibration of the microfluidic platform. a, The substrate concentration was back-calculated using partition coefficients for an FC-40/1mM HEPES buffer system exhibiting a linear temperature dependence. Each data point represents an average of 6-9 independent measurements. **b,** The calibration of pH with the hydrochloric acid (reaction product) concentration shown for individual enzymes.

Supporting Figures

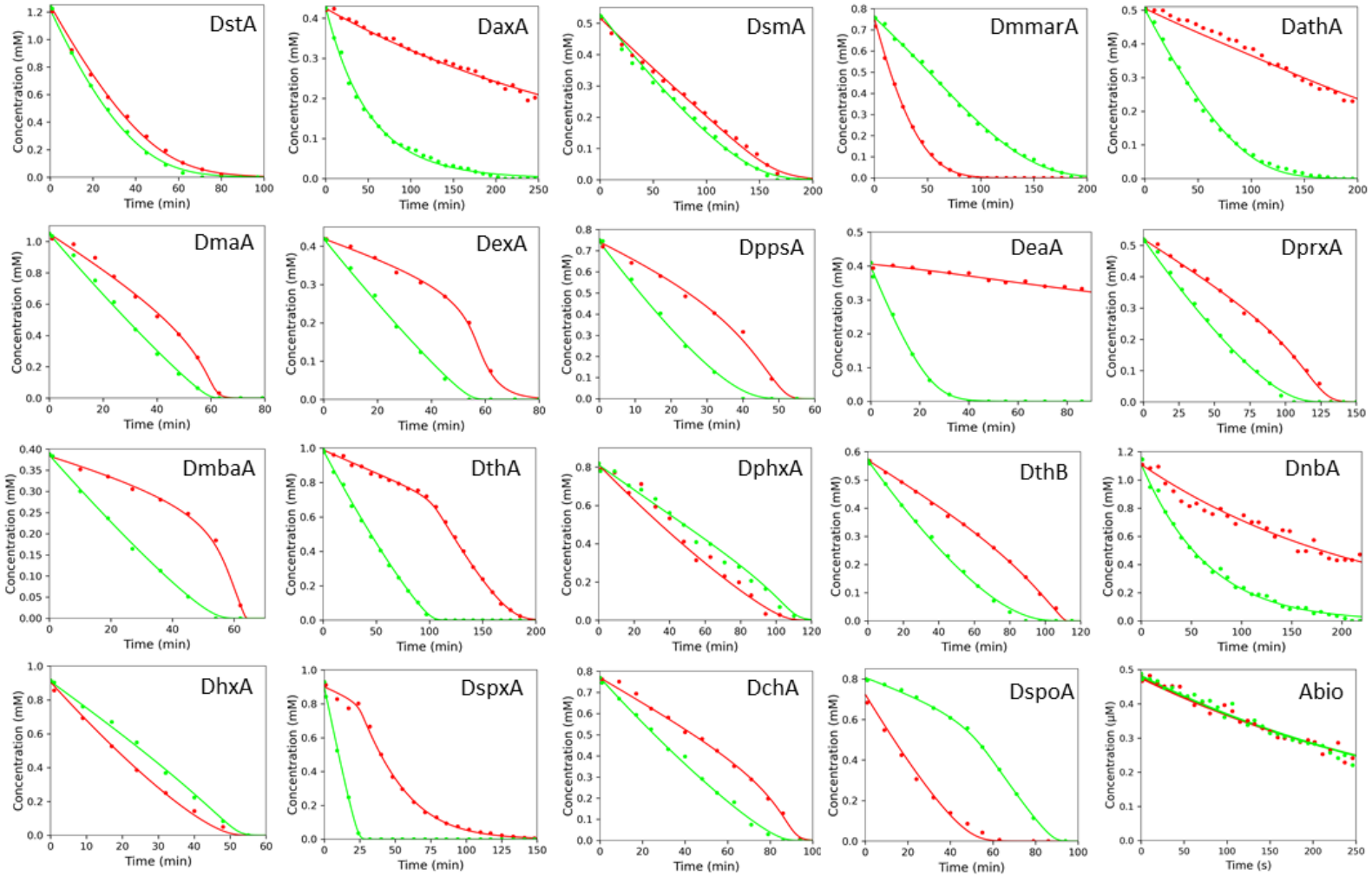


Figure S13. Kinetic resolution of 2-bromopentane. The (*R*)-enantiomer is shown in green and the (*S*)-enantiomer in red. The solid lines represent the best fit to the data. Spontaneous hydrolysis of the racemic substrate was tested (Abio) and used in the global data analysis. The reactions were conducted in 100 mM glycine buffer at 20 ˚C and pH 8.6.

Supporting Figures

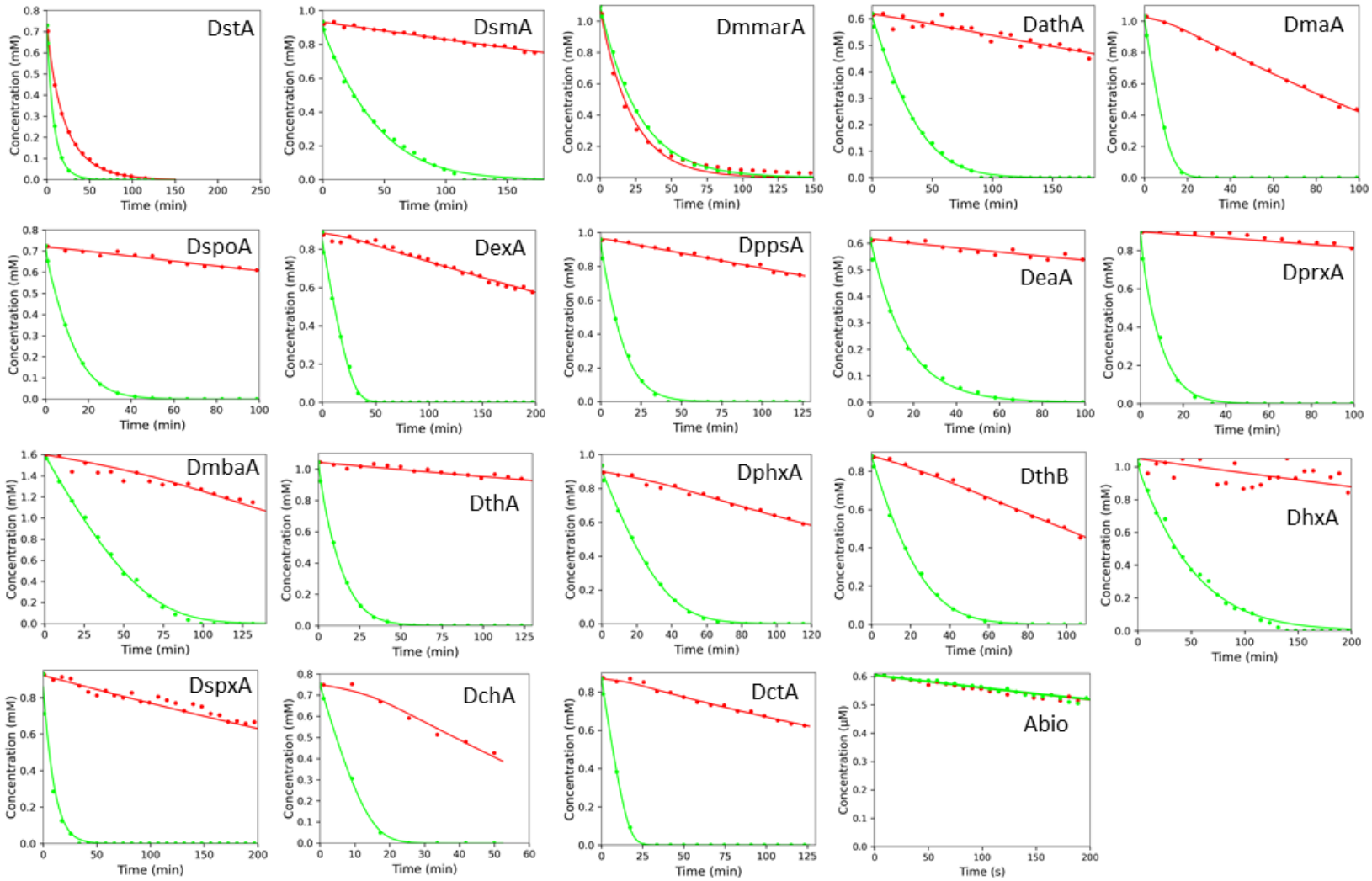


Figure S14. Kinetic resolution of ethyl 2-bromopropionate. The (*R*)-enantiomer is shown in green and the (*S*)-enantiomer in red. The solid lines represent the best fit to the data. Spontaneous hydrolysis of the racemic substrate was tested (Abio) and used in the global data analysis. The reactions were conducted in 100 mM glycine buffer at 20 ˚C and pH 8.6.

Supporting Figures

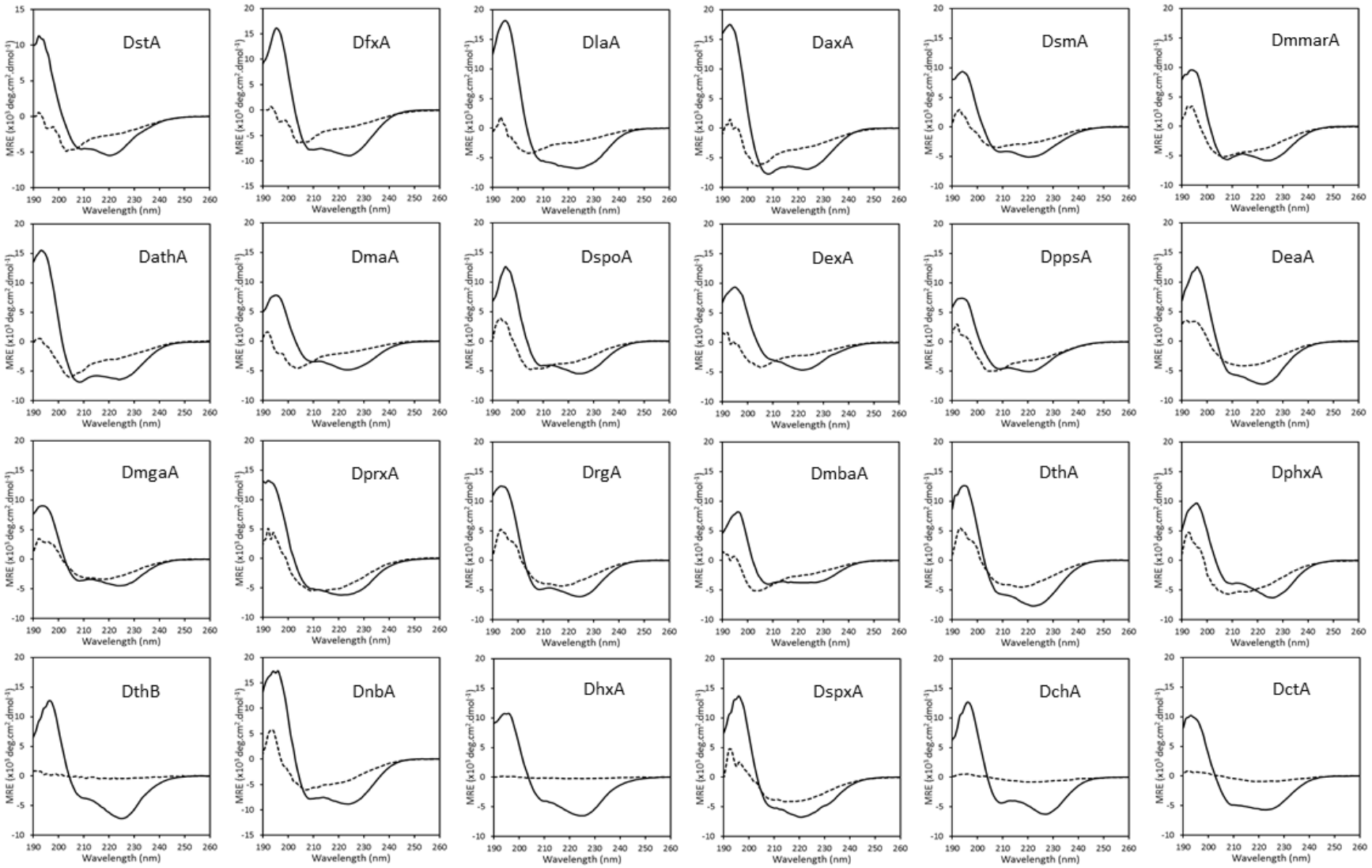


Figure S15. **Analysis of temperature-induced unfolding.** Far-UV CD spectra of HLDs with the depicted native state (black line) and denatured state (dashed line) during temperature scan. Each spectrum shown is the average of five individual scans and has been corrected for the absorbance of the buffer.

Supporting Figures

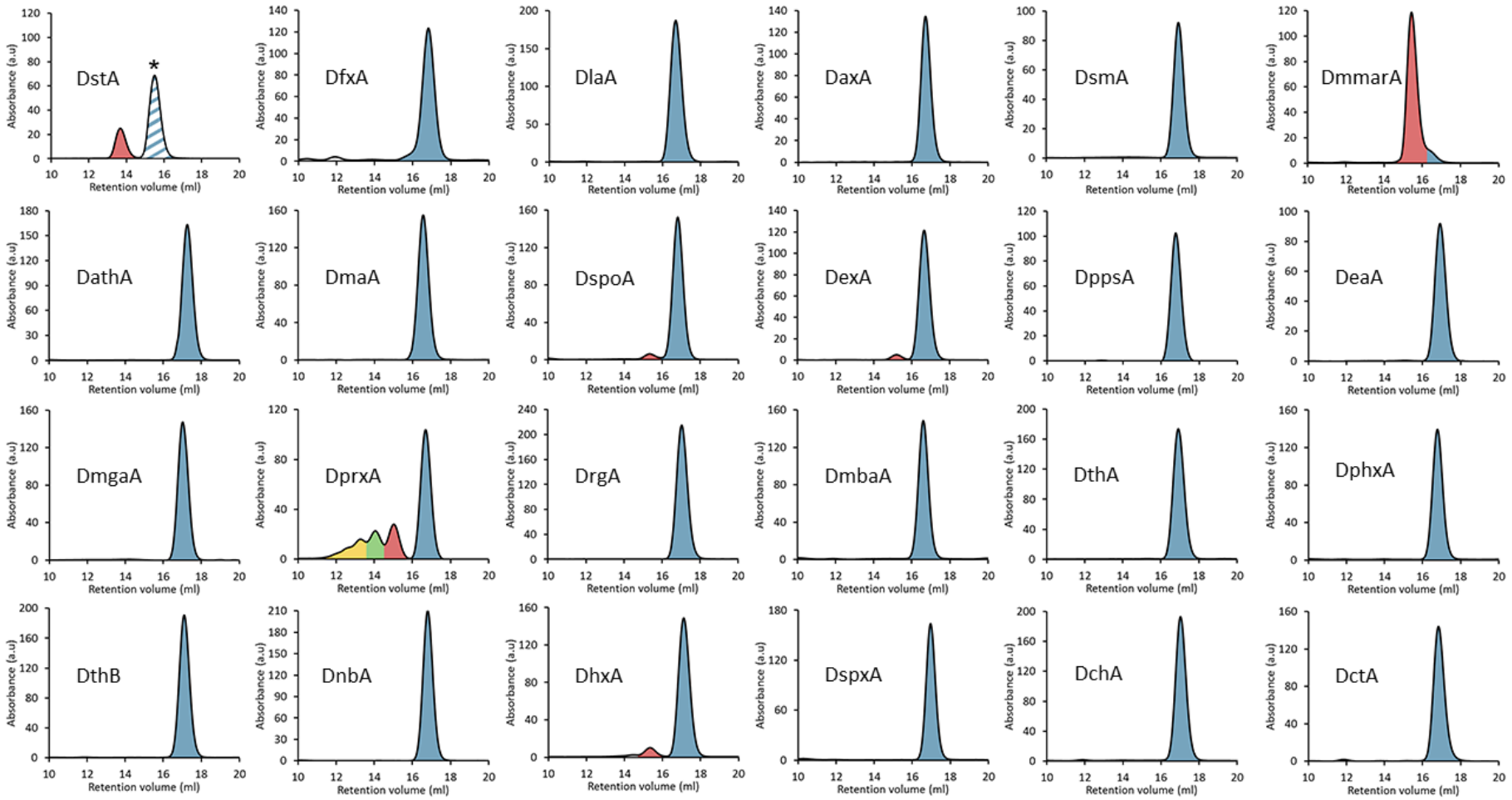


Figure S16. **Analysis of the quaternary structure**. The quaternary structure was determined by size-exclusion chromatography: monomeric (blue), dimeric (red), higher oligomeric states (green and yellow).*DstA – the fraction contains a mixture of monomer and oxidized dimer under oxidative conditions. UV_{280} response *versus* time is presented.

References

1. Iwasaki, I., Utsumi, S. & Ozawa, T. New Colorimetric Determination of Chloride using Mercuric Thiocyanate and Ferric Ion. *Bulletin of the Chemical Society of Japan* **25**, 226–226 (1952).