Crystallization and structure determination

Bpro0530 - The protein (13 mg/ml, in 50 mM Tris-SO_4 pH 8.5) was crystallized by hanging drop vapor diffusion against 1 M Li_2SO_4 in 0.1 M Tris-HCl, pH 8.5. The crystals were coated in paratone N as cryoprotectant during data collection at 100 K. Diffraction data were collected in-house on a MicroMax-007 X-ray generator (Rigaku, The Woodlands, TX, USA) with Varimax optics (Rigaku) and a mar345 image plate detector (Marresearch, Norderstedt, Germany). Data were reduced in XDS (1), and initially processed in $P6_n22$ but various indicators suggested the likelihood of twinning. In the case of hemihedral twinning (*i.e.* merohedral twinning of two domains), the true space group may be $P6_n$, $P3_n12$ or $P3_n21$. The phases were solved by molecular replacement with Phaser (2) using a homology model built from the structure of PDB code 1AQ6 by SWISSMODEL (3). This procedure resulted in a unique solution in space group $P6_5$. The correct space group was also confirmed by the program Zanuda on the York Structural Biology Laboratory web server (http://www.ysbl.york.ac.uk/YSBLPrograms/index.jsp). Refinement was performed in REFMAC5 (4) using amplitude-based twin refinement (the twin fraction refined to 44.1%) and in Coot (5). Because of twinning and non-crystallographic symmetry, the random FreeR flags were selected in thin resolution shells using XPREP (Bruker AXS Inc, Madison, USA) to avoid bias. Additionally, the FreeR flags were expanded from $P6_522$ to $P6_5$ using the program CAD in the CCP4 software package (6) to ensure that twin-related reflections share identical FreeR flags.

Rha0230 - Both the native protein and the Se-Met derivative (24 mg/ml, in 50 mM TrisSO_4 pH 8.5) were crystallized by hanging drop vapor diffusion against 0.6 M trisodium citrate and 0.1 M Na-HEPES, pH 7.5. The reservoir buffer supplemented with 16% glycerol was used as cryoprotectant. Diffraction data were collected at beamline X8C at the National Synchrotron Light Source (Brookhaven, USA). The experiment was performed at the peak wavelength of the selenium absorption edge (λ = 0.9795 Å), and the dataset was reduced using HKL2000 (7). The heavy atom sites and phases were determined by single wavelength anomalous dispersion from the Se-Met protein crystals using SOLVE/RESOLVE (8) and the

initial model was built using Arp/wArp (9). The model was then further improved manually, and finally refined in REFMAC5 (4) and Coot (5) against the native dataset collected at beamline 14BM-C at the Advanced Photon Source (Argonne National Laboratory, Argonne, USA).

PA0810 - The protein (20 mg/ml, in 0.3 M NaCl, 1 mM TCEP and 10 mM Na-HEPES pH7.5) was crystallized by hanging drop vapor diffusion against 20% PEG 3350 and 0.2 M diammonium citrate. Paratone N was used as cryoprotectant. Diffraction data were collected at beamline 19ID at the Advanced Photon Source at the peak wavelength of the selenium absorption edge (λ = 0.979 Å) and reduced with XDS (1). Using SOLVE/RESOLVE (8), the 22 heavy atoms from the four chains were located by single wavelength anomalous dispersion, and the initial phases were optimized by histogram matching and solvent flattening to yield an electron density map of high quality. The initial model was built using Autobuild in Phenix (10) and manually completed in Coot (5). The final refinement was performed in REFMAC5 (4).

RSc1362 - The protein (30 mg/ml, in 0.3 M NaCl, 1 mM TCEP and 10 mM Na-HEPES pH7.5) was crystallized by sitting drop vapor diffusion using 15% isopropanol, 0.1 M KCl, 25 mM $MgCl_2$, 2% 1,4-dioxane and 50 mM sodium cacodylate, pH 6.0. Paratone N was used as cryoprotectant. Diffraction data were collected in-house on an FR-E+ Superbright X-ray generator equipped with Varimax optics and a Saturn CCD area detector (Rigaku), and reduced with XDS (1). The structure was solved by applying the molecular replacement program MOLREP (11) using chain A of PDB code 1QH9 as search model. The structure was then refined using Phenix (10) and Coot (5). Water molecules were added by ARP/wARP (9) after several refinement cycles. The final refinement was performed in REFMAC5 (4).

Molecular Dynamics Simulations

All-atom molecular dynamics (MD) simulations were performed for five dehalogenases, Bpro0530 (PDB code 3UM9), Rha0230 (PDB code 3UMG), PA0810 (PDB code 3UMC),

RSc1362 (PDB code 3UMB), and ST2570 (PDB code 2W43), in explicit water. The N- and C-termini of the enzymes were modeled as neutral NH_2 and -CO_2H groups, respectively, to reflect the fact that some terminal residues are missing in several of the PDB structures. Hydrogen atoms were added using the default settings of the GROMACS "pdb2gmx" tool, with all titratable groups modeled in their standard protonation state at pH 7.0. Each individual enzyme was solvated in a box of water with periodic boundary conditions to obtain an initial water density close to 1000 g/l for all systems. Between zero and eleven Na^+ counterions were added to neutralize the simulation box (detailed simulation cell composition described in Table S4). Protein and ions were modeled with the optimized potentials for liquid simulations (OPLS) all-atom force field (13,14), and the transferable intermolecular potential 3-point (TIP3P) model (15) was used for water molecules. The calculations were performed using the macromolecular simulation program GROMACS, version 4.0.7 (12). Covalent bonds of water and other molecules were constrained with SETTLE (20) and P-LINCS (21), respectively. Lennard–Jones interactions were evaluated using a group-based cut-off for separation distances less than 10 Å and otherwise ignored. Coulombic interactions were calculated using the smooth particle-mesh Ewald (PME) method (16,17) with a real-space cut-off of 1.2 Å and a Fourier grid spacing of 1.4 Å. During 500-step steepest-descent energy minimizations performed initially, all heavy atoms of the enzymes were kept restrained to their crystallographic structure using a force constant of 1000 $kJ \cdot mol^{-1} \cdot nm^{-2}$. Unrestrained simulations were then performed on all five enzymes. Each simulation system was replicated nine times (yielding ten in total) and simulated for 300 ns, for a total simulation time of 3 μs per enzyme. Each replica simulation was assigned initial velocities generated from a different random seed. Simulation in the isothermal-isobaric (NpT) ensemble was achieved by isotropic coupling to a Berendsen barostat (18) at 1 bar with coupling constants of 4 ps and temperature coupling using velocity Langevin dynamics (19) at 300 K with a collision frequency of 1 ps^{-1}. The integration time step was 2 fs and the non-bonded pair-list was updated every 20 fs. The time trajectories were saved every 25 ps. The first 50 ns of each trajectory were removed from analysis as equilibration based on the analysis of backbone RMSD vs. time (Figure S4) and all structural properties

were computed from the remaining data from t=50 ns to t=300 ns timeframes for all fifty trajectories.

Structural fluctuations in the respective halide-binding pockets were analyzed. Pairwise C_α - C_α distances for conserved residues Asp10, Arg41, and Asn119 (Bpro0530 numbering) were computed using the g_dist tool implemented in GROMACS. The Euclidian distance (square root of the sum of the three squared distances) between these three residues was used as a single metric to quantify the size of the binding pocket in each enzyme at each time step. The individual distributions of these Euclidian distances for each simulation repeat are shown in Figure S3.

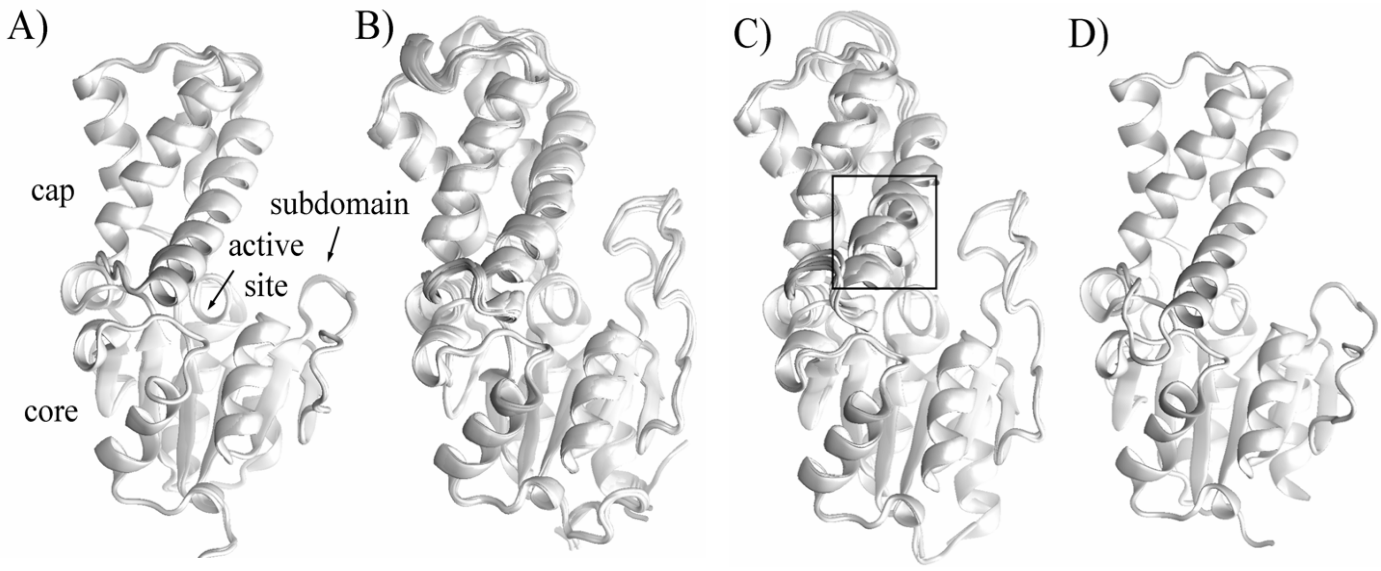

Figure S1. Ribbon representations of HAD monomers. Although all four enzymes are dimeric proteins all crystallographically independent monomers in an asymmetric unit were superimposed. Only the Cα atoms of the core domain were used in the superposition to reveal any potential interdomain displacement. Different subunits of the same proteins were also superposed when the crystal's asymmetric unit contains multiple protein subunits. **A**) Bpro0530 – 2 copies. Only a short loop assumes the position of the subdomain insert. Superposition of the 2 subunits reveals no obvious interdomain displacement. **B**) Rha0230 – 8 copies. This enzyme contains a 'flap' subdomain outside the active site entrance. The 8 cap domains display the largest positional difference at the tip of the protein, suggesting the possibility of hinge-like interdomain motion. **C**) PA0810 – 4 copies. Its overall fold, including the 'flap' subdomain, closely resembles that of Rha0230. The positional differences of the 4 cap domains also hint at similar hinge-like interdomain motions. Additionally, the variations in α2 conformations (Arg44-Ala50) reveal inherent flexibility in this region of the cap domain (boxed). **D**) RSc1362 – 1 copy. This protein again lacks a subdomain insert.

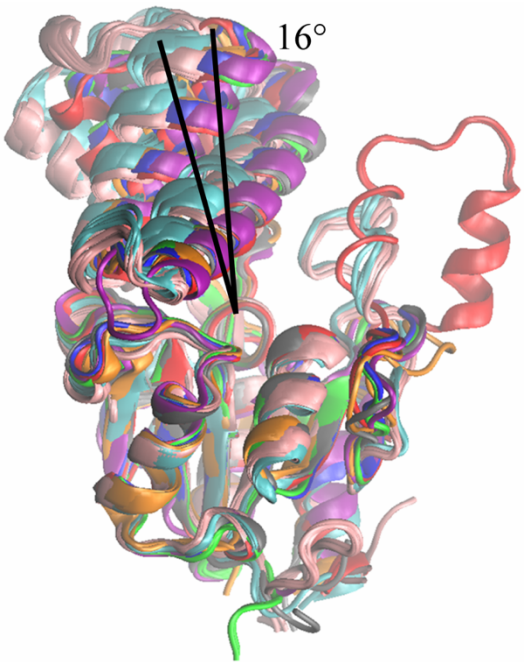

Figure S2. Relative displacement between the cap and core domains in HADs. The structures are compared by superposition of the Cα atoms of core domain residues. The number of monomers in the crystals' asymmetric units is listed as n. L-Dex YL (PDB code 1ZRM, n=1, blue), DehlB (1QQ5, n=2, red), DehIVa (2NO4, n=2, grey), ST2570 (2W43, n=2, orange), Bpro0530 (3UM9, n=2, green), Rha0230 (3UMG, n=8, pink), PA0810 (3UMC, n=4, cyan) and RSc1362 (3UMB, (n=1, purple). The maximum rotation angle observed between these examples of HADs is 16°.

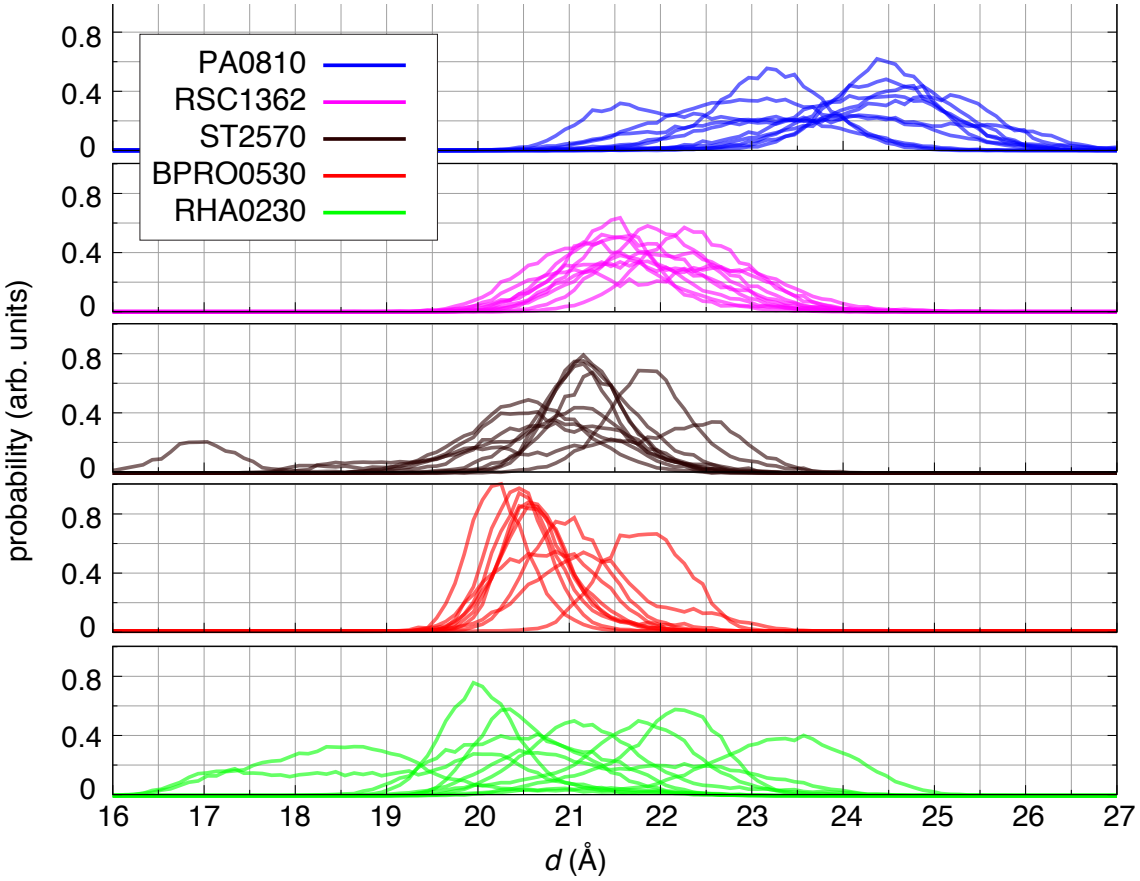

Figure S3. Probability distribution of the Euclidian distance *d* between the C_α-atoms of residues D10-R41-N119 (Bpro0530 numbering; corresponding residues for the other four enzymes) from ten independent 300 ns MD simulations of each enzyme at 300 K.

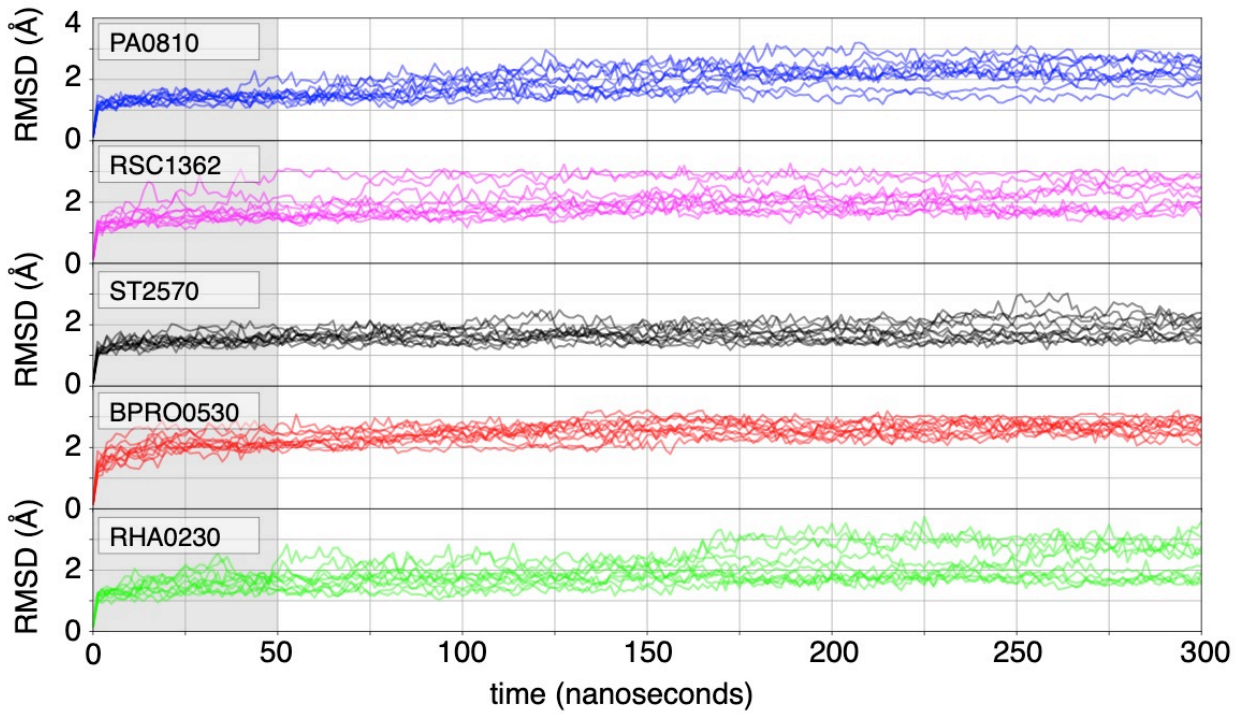

Figure S4. Structural fluctuations of dehalogenases in molecular simulations. Root-mean-square deviations of backbone atoms ('N', 'CA', 'C', 'O') were computed with respect to initial crystallographic models as a function of simulation time, for all ten independent simulation repeats for each of the five dehalogenases, Bpro0530 (PDB-ID: 3UM9), Rha0230 (PDB-ID: 3UMG), PA0810 (PDB-ID: 3UMC), RSc1362 (PDB-ID: 3UMB), and ST2570 (PDB-ID: 2W43). At a simulation time of t=300 ns, average RMSD across all simulation repeats was 2.2 ± 0.2, 2.2 ± 0.2, 1.9 ± 0.1, 2.7 ± 0.1, 2.4 ± 0.2, for PA0810, RSc1362, ST2570, Bpro0530, Rha0230, respectively. Here, error is reported as the standard error of mean over the RMSD values of the ten simulation repeats at time 300 ns.

Table S1. X-ray diffraction data and refinement statistics

Enzyme	PA0810	RSc1362	Bpro0530	Rha0230*
Modification	Se-Met	Se-Met		
PDB code	3UMC	3UMB	3UM9	3UMG
X-ray source	19ID, APS	FR-E	MicroMax-007	14BM-C, APS
Wavelength (Å)	0.979	1.54	1.54	0.980
Resolution limit (Å)	28.4-2.15 (2.27-2.15)	25.0-2.20 (2.32-2.20)	19.0-2.20 (2.25-2.20)	38.5-2.25 (2.31-2.25)
$\langle I \rangle / \langle \sigma(I) \rangle$	10.4 (2.4)	13.0 (3.5)	32.8 (3.8)	19.6 (3.9)
Space group	C2	C222	$P6_5$	$P2_12_12_1$
No. molecules in ASU	4	1	2	8
Unit cell parameters				
a (Å) *b* *c*	73.4 123.9 125.6	66.4 129.6 54.3	96.6 96.6 92.0	102.2 148.7 152.6
α (°) β γ	90 90.97 90	90 90 90	90 90 120	90 90 90
R_{sym} (%)	7.2 (30.1)	5.0 (21.4)	6.1 (42.9)	6.9 (46.8)
Completeness (%)	99.9 (100.0)	99.9 (100)	99.1 (89.0)	96.6 (97.8)
Bond lengths rmsd (Å)	0.008	0.010	0.020	0.018
Bond angles rmsd (°)	1.02	1.17	1.818	1.613
R_{crys} (%)	18.4	22.9	14.8	21.5
R_{free} (%)	22.1	27.2	19.8	27.2
Ramachandran plot (%)				
Favored	91.2	91.1	92.8	91.1
Allowed	8.4	8.3	6.8	8.7
Generously allowed	0.4	0.0	0.5	0.1
Outlier	0.0	0.6	0.0	0.0

$$\mathbf{R}_{sym} = \Sigma_{hkl} \Sigma_j |\boldsymbol{I}_{hkl,j} - \langle \boldsymbol{I}_{hkl} \rangle| / \Sigma_{hkl} \Sigma_j \boldsymbol{I}_{hkl,j}$$

Where $\langle \boldsymbol{I}_{hkl} \rangle$ is the average of symmetry- (or Friedel-) related observations of a unique reflection.

$$\mathbf{R}_{crys} = \Sigma ||\mathbf{F}_o| - |\mathbf{F}_c|| / \Sigma |\mathbf{F}_o|$$

$\mathbf{R}_{free}$ is the same as $\mathbf{R}_{crys}$, but for a subset of 5% of all reflections never used in crystallographic refinement.

Ramachandran analysis performed by MolProbity (22).

*Numbers in parentheses represent those in the highest resolution bin. *A dataset collected at X8C (National Synchrotron Light Source, Brookhaven National Laboratory, Brookhaven, USA) on a Se-Met derivatized crystal was used for phase determination by single wavelength anomalous dispersion, and for building the initial atomic model. Data included in this column refer to the native data set collected later at beamline 14-BM-C, BioCARS, Advanced Photon Source, Argonne National Laboratory, Lemont, IL, USA) and used in refinement. Names of defluorinating enzymes are on green backgrounds.*

Table S2. RMSD values for C_α-atoms (in Å) when the structures of the various HADs are superimposed with help of the *cealign* algorithm of the program PyMOL.[16] Numbers above the diagonal refer to values for the cap domains and those below the diagonal to values for the core domains. The three enzymes with the capability to break carbon-fluorine bonds are marked with a green background.

	LDexYL	DhlB	DehlVa	RSc1362	PA0810	Rha0230	Bpro0530	ST2570
LDexYL	-	1.37	1.62	1.28	4.35	2.35	0.68	2.29
DhlB	2.64	-	2.64	1.98	3.20	3.14	1.42	2.66
DehlVa	1.21	1.71	-	1.62	2.49	4.25	1.61	3.13
RSc1362	1.05	1.64	1.30	-	2.62	3.96	1.96	3.51
PA0810	2.49	3.76	2.56	2.43	-	2.25	3.59	3.17
Rha0230	2.55	3.71	2.63	2.54	0.70	-	3.34	3.13
Bpro0530	1.45	2.27	1.63	1.45	2.6	2.70	-	2.79
ST2570	1.76	2.24	1.81	1.83	2.68	2.96	1.69	-

Table S3. Interatomic distances in active sites of crystalline HADs.

Enzyme	PDB	chain	Active site content	Asp(Oδ1)-Asn(Nδ2)	Asp(Oδ1)-Arg(Nη2)	Asn(Nδ2)-Arg(Nη1)	Asn(Nδ2)-Arg(Nη2)	Asp(Oδ1)-Tyr(Oη)
L-Dex YL	1 JUD	A	-	7.26	9.47	11.03	10.45	-
	1ZRM	A	covalent intermediate	6.23	6.58	7.59	7.81	-
	1ZRN	A	covalent intermediate	a	a	8.42	9.94	-
	1QH9	A	disrupted Asp and Asn conformation; exclude from analysis					
DhlB	IAQ6	A	formate	6.53	6.08	6.11	6.86	-
		B	formate	6.19	6.11	6.19	6.72	-
	IQQ5	A	formate	5.92	6.08	6.22	6.61	-
		B	formate	5.89	6.13	6.15	6.56	-
	1QQ6	A	cov int, Cl in pocket	6.01	6.25	6.33	6.48	-
		B	cov int, Cl in pocket	6.12	6.23	6.25	6.43	-
	1QQ7	A	cov int, Cl in pocket	6.03	6.05	6.29	6.34	-
		B	cov int, Cl in pocket	6.12	6.10	6.26	6.33	-
DehIVa	2NO4	A	sulphate	6.09	6.35	6.74	7.35	-
		B	sulphate	6.09	6.32	6.81	7.35	-
	2NO5	A	covalent intermediate	6.05	6.00	7.08	7.54	-
		B	covalent intermediate	6.01	5.82	7.05	7.20	-
RSc1362	3UMB	A	-	6.56	6.46	6.16	6.32	-
PA0810	3UMC	A	chloride d	6.00	6.50	6.92	6.79	9.20
		B	chloride	6.00	6.67	7.22	6.77	5.74
		C	chloride	6.18	7.34	7.71	7.24	9.70
		D	chloride	6.43	7.21	6.62	8.07	9.21
Rha0230	3UMG	A	chloride d	6.04	6.11	6.75	6.37	5.47
		B	chloride	5.90	6.09	7.56	6.93	5.20
		C	chloride	6.30	6.16	8.06	7.47	5.42
		D	chloride	6.02	6.11	6.97	6.51	5.54
		E	chloride	6.17	6.33	7.13	6.77	5.48
		F	chloride	5.87	6.76	7.62	6.55	5.77
		G	chloride	5.80	6.31	7.45	7.14	5.15
		H	chloride	6.01	5.46	7.00	6.55	5.09
Bpro00530	3UM9	A	sulphate	6.33c	5.68c	6.00	6.80	-
		B	sulphate	6.21c	6.15c	6.24	6.70	-
ST2570	2W11	A	lactate	6.01	5.70	5.56	5.97	-
		B	lactate	6.08	5.64	5.45	5.99	-
	2W43	A	phosphate	6.21	5.88	5.84	6.23	-
		B	phosphate	6.20	7.21b	6.01b	7.00b	-

The active sites of all HADs crystal structures are analyzed. Defluorinating enzymes are marked by a green background.
a *- Distances to the aspartate nucleophile are not measured because of its unusual conformation.* ***b*** *- The halide binding arginine adopts two conformations, and the more deeply buried conformer was measured.* ***c*** *- Disordered aspartate nucleophile; active site distances are estimated from the superposed aspartate from L-Dex YL (1ZRM).* ***d*** *- The chloride ion is not bound in the halide pocket, but to the serine residue of the carboxylate recognition pocket instead.*

Table S4. Simulation cell composition for all molecular simulation systems.

System	Protein structure*	# Na+ ions	# Cl^- ions	# water molecules	cell size (in nm)
Bpro0530	3UM9	6	0	20,181	8.17 x 9.75 x 7.99
Rha0230	3UMG	11	0	21,855	7.94 x 10.54 x 8.27
PA0810	3UMC	7	0	17,849	9.86 x 7.75 x 7.56
RSc1362	3UMB	2	0	21,006	9.40 x 8.60 x 8.30
ST2570	2W43	0	1	21,067	8.70 x 8.76 x 8.77

1 Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr Section D Biol. Crystallogr.* **66**, 133-144
2 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674
3 Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics (Oxford, England)* **22**, 195-201
4 Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. Section D Biol. Crystallogr.* **67**, 355-367
5 Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Section D Biol. Crystallogr.* **66**, 486-501
6 Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. Section D Biol. Crystallogr.* **67**, 235-242
7 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, Carter, C. W. and Sweet, R. M., eds., Academic Press. pp 307-326
8 Terwilliger, T C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. Section D Biol. Crystallogr.* **55**, 849-861
9 Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171-1179
10 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Section D Biol. Crystallogr.* **66**, 213-221
11 Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta Crystallogr. Section D Biol. Crystallogr.* **66**, 22-25
12 Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on the conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236
13 Hess, B., Kutzner, C., van der Spoel, D., and Lindhal, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalabale molecular simulations. *J. Chem.Theory Comput.* **4**, 435–447
14 Kaminski, G .A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B.* **105**, 6474–6487
15 Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R., and Klein, M. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935
16 The PyMOL Molecular Graphics System, Version 2.2.2, Schrödinger, LLC.

17 Darden. T., York, D., and Pedersen, L. G. (1993) Particle mesh Ewald: An *N log(N)* method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092
18 Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Perdersen, L. G. (1995) A smooth particle mesh Ewald potential. *J. Chem. Phys.* **103**, 8577–8593
19 Berendsen, H. J. C., Postma, J. P. M., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690
20 van Gunsteren, W.F., and Berendesen, H. J. C. (1988) A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* **1**, 173–185
21 Miyamoto, S., and Kollman, P. A. (1992) SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comput. Chem.* **13**, 952–962
22 Hess, B. (2008) P-LINCS: A parallel linear constraint solver for molecular simulations. *J. Chem. Theory Comput.* **4**, 116–122
23 Chen, W. B., Arendall III, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) *MolProbity*: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Section D Biol. Crystallogr.* **66**, 12-21