SUPPLEMENTARY INFORMATION

Mechanism-Based Strategy for Optimizing HaloTag Protein Labeling

Sérgio M. Marques[a,b‡], Michaela Slanska[a‡], Klaudia Chmelova[a,b], Radka Chaloupkova[a,c], Martin Marek[a,b], Spencer Clark[d], Jiri Damborsky[a,b], Eric T. Kool[d], David Bednar[a]*, Zbynek Prokop[a,b]*

[a]*Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, 625 00 Brno, Czech Republic*
[b]*International Clinical Research Center, St. Anne's University Hospital, 656 91 Brno, Czech Republic*
[c]*Enantis Ltd., Biotechnology Incubator INBIT, 625 00 Brno, Czech Republic*
[d]*Department of Chemistry, Stanford University, Stanford, California 94305, USA*

‡These authors contributed equally to this study.

*Corresponding authors: David Bednar, 222755@mail.muni.cz; Zbynek Prokop, zbynek@chemi.muni.cz

Content

Section I: Protein expression and purification

Methods

The haloalkane dehalogenases genes *linBH272F*, *dhaAH272F*, *dhaAHT* and *dmmAH315F* were cloned into pAQN, pET21b or pET24a vectors, were transformed into chemocompetent *Escherichia coli* BL21 or BL21(DE3) cells (**Supplementary Table S1**). The cells were grown at 37 °C until they reached an optical density of about 0.6 at 600 nm in Luria-Bertani (LB) medium containing ampicillin (100 $\mu g \cdot ml^{-1}$) or kanamycin (30 $\mu g \cdot ml^{-1}$). The protein expression was induced by addition of 0.5 mM isopropyl β-D-1-thiogalacto-pyranoside and cells were further incubated at 20 °C. The harvested cell biomass was homogenized by Ultrasonic processor (Hielscher, Germany) and clarified by centrifugation at 21 000 g for 1 h (4 °C). Soluble proteins were purified by metal-affinity chromatography using 5 ml Ni-NTA Superflow column (Qiagen, Germany) attached to FPLC purification system (Bio-Rad, USA) which was equilibrated with purification buffer A (500 mM NaCl, 20 mM K_2HPO_4/KH_2PO_4 (pH 7.5), 10 mM imidazole). The His-tagged proteins were eluted using purification buffer B containing 300 mM imidazole (**Supplementary Figure S1A**). The eluted proteins were dialyzed against phosphate-buffer saline (PBS) (137 mM NaCl; 2.7 mM KCl; 10 mM Na_2HPO_4; 1.8 mM KH_2PO_4; pH 7.4) with 0.01% (w/v) CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate) either by the dialysation membrane or using gel permeation chromatography (GPC) equipped with preparative HiLoadTM 16/600 SuperdexTM 200 column (GE Healthcare, Sweden), pre-equilibrated by PBS buffer, attached to Äkta FPLC (GE Healthcare, Sweden). The protein purity was confirmed by SDS-polyacrylamide gel electrophoresis (**Supplementary Figure S1B**) using 15% polyacrylamide gels, stained with Coomassie brilliant blue R-250 dye (Fluka, Buchs, Switzerland). The protein concentration was determined spectrophotometrically using DeNovix DS-11 spectrophotometer by measuring absorbance at 280 nm with a known extinction coefficient calculated from the protein sequence using the ProtParam tool provided by ExPASy[1].

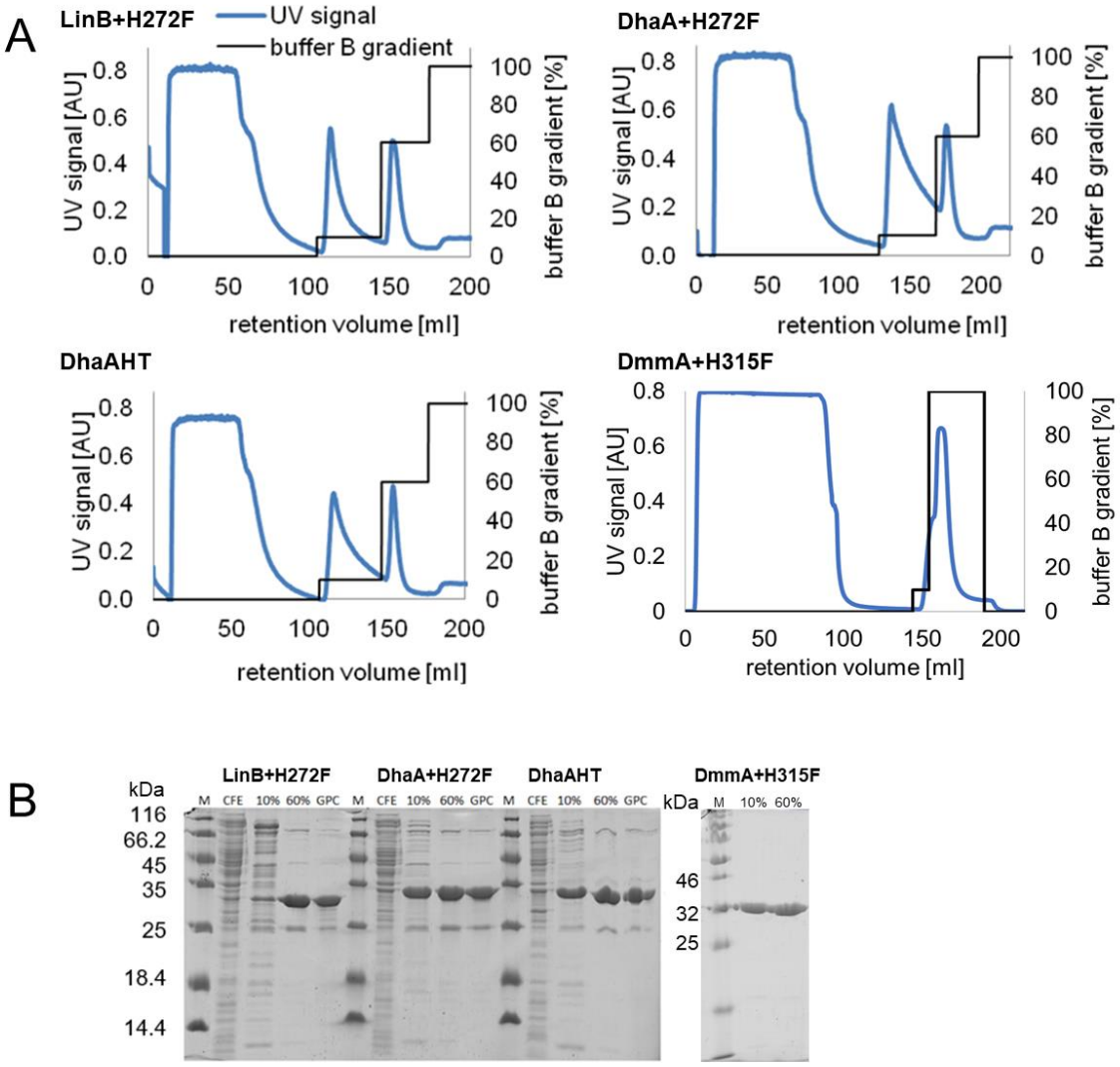
Supplementary Table S1. Haloalkane dehalogenases used for fluorescence polarization experiments.

Enzymes	Mutations	Expression vectors
DhaAHT[a]	K175M, C176G, V197I, H272F, Y273L, A292G	pET21b
DhaAH272F	H272F	pAQN
LinBH272F	H272F	pAQN
DmmAH315F	H315F	pET24a

[a] The DhaAHT used in this study carries two stabilization mutations (V192I, A282G) as compared to the one formerly published DhaAHT[2] and to the variant generally referred to as HaloTag®[3].

Results

The enzymes were overexpressed in *E. coli* BL21 and BL21(DE3) cells and purified by metal-affinity chromatography (**Supplementary Figure S1**). After affinity purification proteins were dialyzed against PBS buffer (pH 7.4, 0.01% CHAPS). The purity of proteins was analyzed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE; **Supplementary Figure S1**).



Supplementary Figure S1. (A) Chromatograms from purification of LinBH272F, DhaAH272F, DhaAHT and DmmAH315F by metal-affinity chromatography. The target proteins were eluted by 10 and 60% gradient of purification buffer B containing 50 mM and 300 mM imidazole, respectively. (B) SDS-PAGE of enzymes before and after purification and after GPC separation, here used as dialysis. Lane M: molecular weight marker; lane CFE: cell-free extract; lane 10%: protein eluted from the column by 10% gradient of buffer B; lane 60%: protein eluted by 60% gradient of buffer B; lane GPC: protein after GPC separation. The overexpressed target protein corresponds to the most distinct band.

Section II: Kinetic analysis

Methods

Fluorescence intensity and anisotropy measurements

The fluorescence anisotropy and fluorescence intensity measurements with the selected fluorescent ligands (E)-1-(5-chloropentyl)-4-(4-(dimethylamino)styryl)pyridin-1-ium iodide (**1B**), (E)-1-(7-chloroheptyl)-4-(4-(dimethylamino)styryl)pyridin-1-ium chloride (**1D**), (E)-1-(8-chlorooctyl)-4-(4-(dimethylamino) styryl)pyridin-1-ium chloride (**1E**) and 4-((2-(2-((6-chlorohexyl)oxy)ethoxy)ethyl)carbamoyl)-2-(6-(dimethylamino)-3-(dimethyliminio)-3H-xanthen-9-yl)benzoate (**TMR**) were performed in black coloured 96-well polypropylene plates. The reaction mixture, with the total volume of 200 µl, consisted of PBS buffer, CHAPS (3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate) to a final concentration of 0.01 % (w/v), 0.001 – 0.1 µM of the fluorescent ligand and 0.001 – 8 µM of the enzyme. The concentration of the enzyme was kept equal or in access over the concentration of ligand to reduce the background effects of the unbound ligand in anisotropy measurements. The time course of the signal was monitored using Infinite F500 plate reader (Tecan, Switzerland) equipped with polarization filters with excitation/emission wavelengths 544 nm/620 nm or 544 nm/580 nm at 30° C. The signal of the enzyme-free incubation of the ligands in the reaction buffer was monitored over time as a negative control. The kinetic experiments were performed in two to three independent replicates.

Kinetic data analysis and statistics

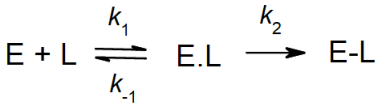
The conventional analysis was performed by fitting the kinetic data by nonlinear regression using exponential functions (**Supplementary Equations S1-3**) built into KinTek Explorer software (KinTek, USA). The concentration dependence of the observed rate of the fast phase and slow phase was then fit to **Supplementary Equations S4-6** by nonlinear regression based on the Levenberg-Marquardt algorithm using Origin 6.0 software (OriginLab, USA). The standard error estimates in fitted parameters k_1 and k_{-1} were propagated to yield error estimates in the calculated value of the equilibrium dissociation constant $K_D=k_{-1}/k_1$. Alternatively, the apparent rate constants for labeling was calculated according to Los et al. 2008 by dividing the determined single exponential rate constants by the concentration of the reactant in excess.

The global analysis was performed using the KinTek Explorer software (KinTek, USA), a dynamic kinetic simulation program that allowed multiple data sets to be fit simultaneously to a single model. Data fitting used numerical integration of rate equations from an input model searching a set of parameters using the Bulirsch–Stoer algorithm with adaptive step size that produces a minimum $\chi 2$ value calculated by using nonlinear regression based on the Levenberg-Marquardt method[4]. Residuals were normalized by sigma value for each data point. To account for slight variations in the data, enzyme concentrations were allowed to vary within an interval of ±10% to make the best fits possible. The standard error was calculated from the covariance matrix during nonlinear regression. In addition to standard error values, more rigorous analysis of the variation of the kinetic parameters was accomplished by confidence contour analysis by using FitSpace Explorer (KinTek, USA)[5]. In this analysis, the lower and upper limits for each parameter were derived from the confidence contours for $\chi 2$ threshold at boundary 0.95.

Results

Conventional fitting of kinetic data

First, the obtained anisotropy data were analyzed by exponential fitting using nonlinear regression. To compare the consistency of the data with earlier published results, the apparent second-order rate constants were calculated by the same procedure used by Los and co-workers for **TMR** interaction with DhaAHT. The constants were calculated for each experiment by dividing the observed rate (k_{obs1}) calculated from single-exponential fit (**Supplementary Equation S1**) by the respective concentration of the enzyme, the component in excess. The obtained values for the incorporation of **TMR** into DhaAHT ranging between 0.8 and 2.3 × 10^6 $M^{-1}{\cdot}s^{-1}$ for various enzyme concentrations correspond well with the value 2.7 × 10^6 $M^{-1}{\cdot}s^{-1}$ published by Los and co-workers[6]. This quick analysis served solely for a basic comparison of the new variants with the previously reported results; however, the obtained constant has no mechanistic meaning. To derive mechanistic information, the concentration dependence of the rates obtained by exponential fitting was then analyzed systematically. Although the single-exponential fit (**Supplementary Equation S1**) indicated satisfying statistics for fitting kinetic traces obtained for reaction of **TMR** with DhaAHT (χ2/DoF = 2.29; *p*-value = 0.28), the use of a double-exponential function (**Supplementary Equation S2**) provided significantly improved goodness of fit (χ2/DoF = 1.21; *p*-value = 0.43) and distinguished two separate, fast and slow, kinetic phases. The two phases indicated in the anisotropic data are well consistent with the two-step mechanism expected for the reaction, which includes ligand binding and subsequent chemical conversion leading to the formation of the final covalent complex (**Supplementary Scheme S1**).



Supplementary Scheme 1

The concentration dependence of the observed rates was used for secondary fitting to approximate rate equations (**Supplementary Equations S4-6**) derived for a proposed two-step kinetic model (**Supplementary Scheme S1**). More specifically, the concentration dependence of observed rate for fast (k_{obs1}) and slow phase (k_{obs2}) were fitted to **Supplementary Equations S4** and **S5**, respectively. The analysis provided estimates of the rate constant for association (k_1) and dissociation of enzyme-ligand complex (k_{-1}) and the rate constant for the subsequent chemical step (k_2) (**Supplementary Table S2**).

$$y = A_1.e^{-k_{obs1}.t} + C \qquad \textit{Supplementary Equation S1}$$

$$y = A_1.e^{-k_{obs1}.t} + A_2.e^{-k_{obs2}.t} + C \qquad \textit{Supplementary Equation S2}$$

$$y = A_1.e^{-k_{obs1}.t} + C + D.t \qquad \textit{Supplementary Equation S3}$$

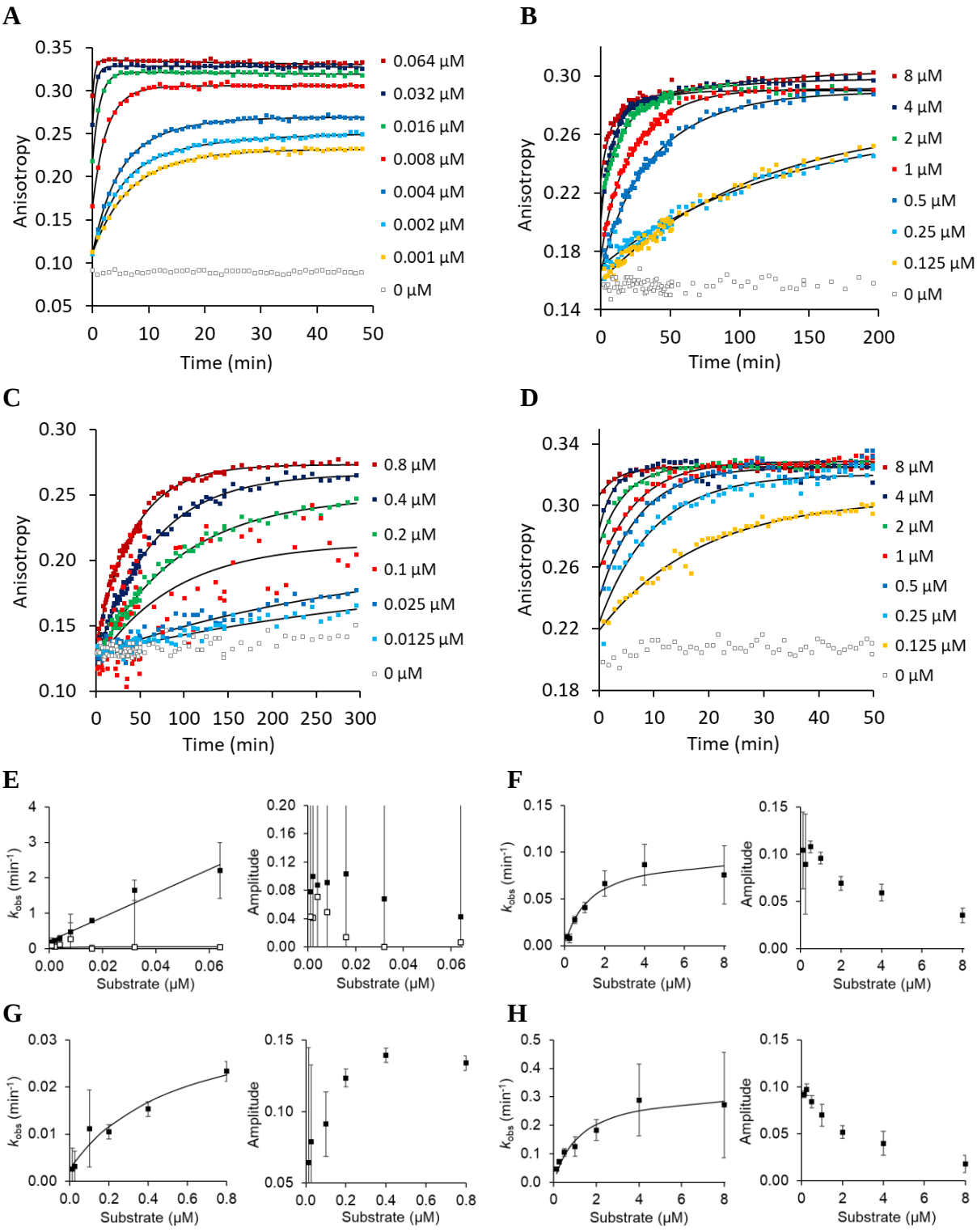
$$k_{obs1} = k_1.[E] + k_{-1} + k_2 \qquad \textit{Supplementary Equation S4}$$

$$k_{obs2} = \frac{k_1.k_2.[E]+k_{-1}}{k_1.[E]+k_{-1}+k_2}$$ *Supplementary Equation S5*

$$k_{obs1} = \frac{k_2.[E]}{K_D+[E]}$$ *Supplementary Equation S6*

In the case of 4-stilbazolium-based ligands, the double-exponential fit did not provide reasonable estimates and the rates and amplitudes were obtained by fitting the kinetic data (**Supplementary Figures S1B-D**) using single-exponential (**Supplementary Equation S1**) or combination of single-exponential and linear (**Supplementary Equation S3**) functions. The concentration dependence of the rate (k_{obs1}) and amplitude (A_1) indicated that the binding phase gradually disappears in a dead-time of the measurement with increasing concentration of the enzyme (**Supplementary Figures S1F-H**). Moreover, the hyperbolic shape of the concentration dependence of the rate also implies that the observed kinetic phase is significantly affected by the velocity of the chemical step and not by ligand binding. The dependence of the rate on concentration was fitted to the simplified **Supplementary Equation S6**, which allowed to define estimates of the equilibrium dissociation constant of enzyme-ligand complex ($K_D = k_{-1}/k_1$) and the rate of consecutive the chemical step (k_2) (**Supplementary Table S2**).

Although this is the currently accepted method for data analysis, there are several limitations related to the conventional fitting. First, the fitting curves to multi-exponential functions is error-prone because it disregards important relationships between rates and amplitudes[7,8]. The resulting errors from the weak definition of the individual amplitudes propagate to errors in rates, tending to overestimate both rates and both amplitudes as can be seen in **Supplementary Figure S2E**. Moreover, when fitting the data using equations, the large number of independent parameters needed to fit the data contributes to increased uncertainty[7,8]. In the example of DhaAHT reaction with TMR, fitting seven traces to a double-exponential function, a total of 35 independent parameters are derived for the specific dataset to extract only three relevant kinetic parameters (k_1, k_{-1} and k_2) in following fit to the rate functions (**Supplementary Equations S4** and **S5**). The transfer of individual parameters from one fitting step to another and the inability to maintain the relationship between rates and amplitudes means a substantial loss of information contained in the original raw data. To overcome these limitations, the global fitting of an entire raw concentration dependence data was performed using numerical integration of the rate equations. The initial estimates for rate and equilibrium constants obtained by the conventional analysis were used as starting values for the numerical integration in the analysis of all tested enzyme variants and ligands.



Supplementary Figure S2. Conventional analysis of DhaAHT reaction with selected ligands. Anisotropy kinetic traces obtained upon mixing of 0.001 µM **TMR** (A), 0.1 µM **1B** (B), 0.01 µM **1D** (C) and 0.1 µM **1E** (D) with DhaAHT in concentration range specified in individual legends. The experiments were performed at 30° C in PBS with 0.01 % (w/v) CHAPS. The solid lines represent the best exponential fit to the data. The concentration dependence of the rates (A, left) and amplitudes (A, left) of the fast (black

squares) and slow phases (white squares) derived from doble exponential fitting. Only the single exponential fit provided reasonable estimates for **1B** (F), **1D** (G) and **1E** (H) providing the concentration dependence of the rates (left) and amplitudes (right). The error bars show the standard error of the fitted parameter. The solid lines represent the best fit of the concentration dependences to approximate rate equations derived for a two-step kinetic model (**Supplementary Equations S4-6**). The concentration dependence of the amplitude cannot be used for reliable parameter estimation.

Supplementary Table S2. Initial estimates of the kinetic constants obtained by conventional analysis. The parameters for reaction of DhaAHT with all tested probes are shown with standard errors derived by nonlinear regression. All experiments performed at 30° C in PBS (pH 7.4) with 0.01 % (w/v) CHAPS.

	k_1 / $\mu M^{-1} \cdot min^{-1}$	k_{-1} / min^{-1}	K_D / μM	k_2 / min^{-1}
TMR	41 ± 4	0.08 ± 0.04	0.002 ± 0.001 [a]	0.06 ± 0.04
1B	- [b]	- [b]	1.3 ± 0.4	0.10 ± 0.01
1D	- [b]	- [b]	0.6 ± 0.5	0.03 ± 0.01
1E	- [b]	- [b]	1.2 ± 0.4	0.33 ± 0.03

[a] The standard errors in fitted parameters were propagated to yield error estimates in calculated value K_D.
[b] The individual rate constants k_1 and k_{-1} cannot be derived, only the equilibrium constant for enzyme-ligand dissociation $K_D = k_{-1}/ k_1$ could be estimated by the conventional fitting.

Fitting by numerical integration of rate equations

The concentration dependence data were globally fitted in their raw form (**Supplementary Figure S3**) to the proposed kinetic model (**Supplementary Scheme S1**) by using Global Kinetic Explorer (KinTek, USA)[4] providing a unique set of related kinetic parameters (**Supplementary Table S3**). The observable anisotropy signal (*y*) was defined as the sum of the contributions of each species to the total signal with scaling factors for each species. Factor *a* scales the signal to the concentration of free ligand (L), factors *b* and *c* reflect the change of the signal corresponding to the formation of enzyme-ligand complex (E.L) and covalent alkyl-enzyme complex (E-L), respectively (**Supplementary Equation S7**). The scaling factors are summarized in **Supplementary Table S4**.

$$y = a * (L + b * E.L + c * E - L) \qquad \textit{Supplementary Equation S7}$$

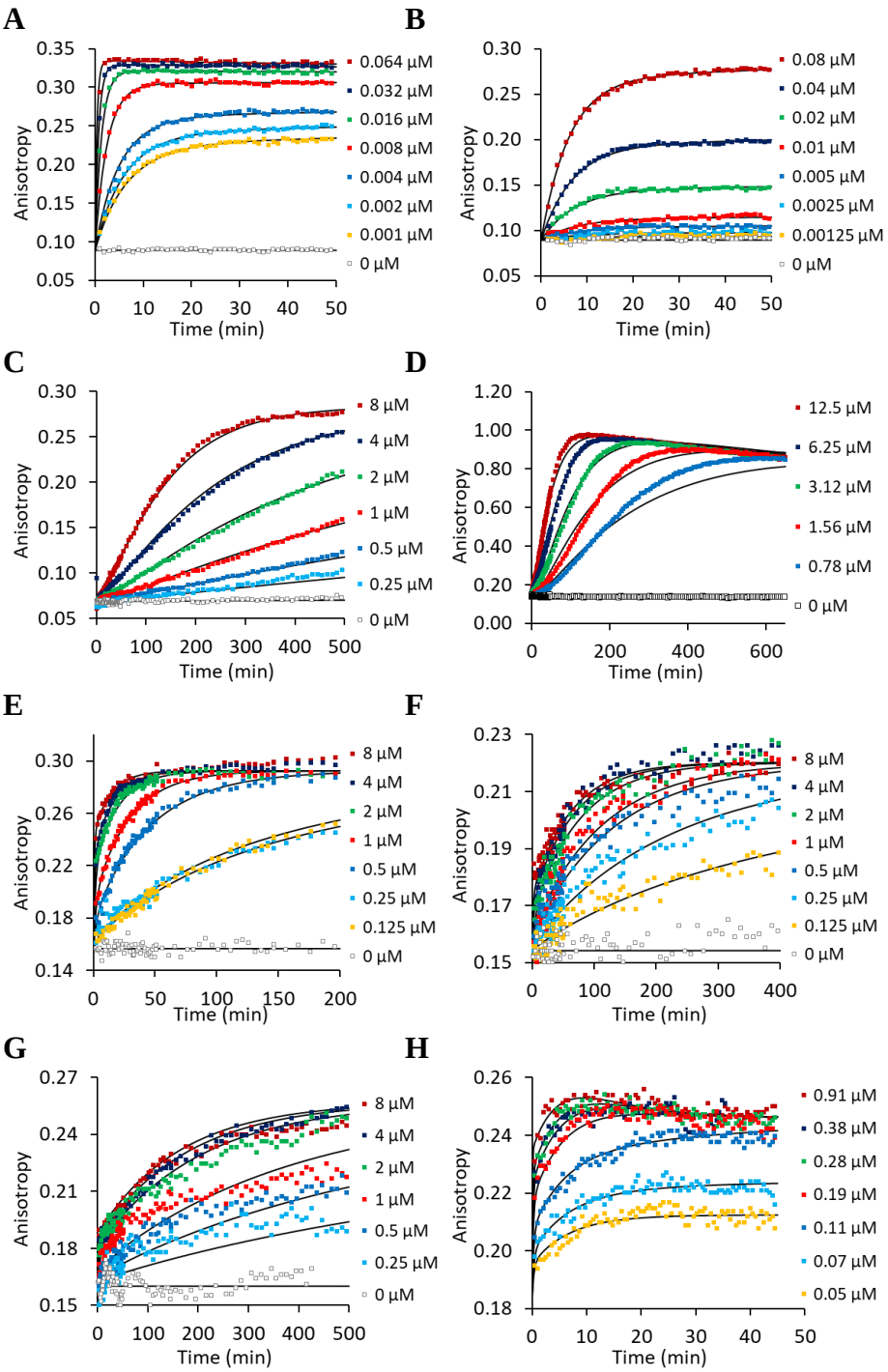
The values are comparable to those obtained by conventional data fitting with standard errors significantly smaller with the global data fitting using only six parameters to fit complete data set in comparison to 35 required for conventional analysis. Additional to standard errors, more rigorous analysis of the variation of the kinetic parameters was accomplished by confidence contour analysis by using FitSpace Explorer (KinTek, USA)[5]. In this analysis, the lower and upper limits for each parameter were derived from the confidence contours for χ^2 threshold at boundary 0.95 (**Supplementary Table S5**).

The free energy of activation ($\Delta G^{\ddagger}$) was calculated for each reaction step from the globally fitted rate constants (k_i) using Eyring equation (**Supplementary Equation 8**), where T is the temperature in Kelvin, R is gas constant, h is Planck´s constant and k_B is Boltzmann constant.

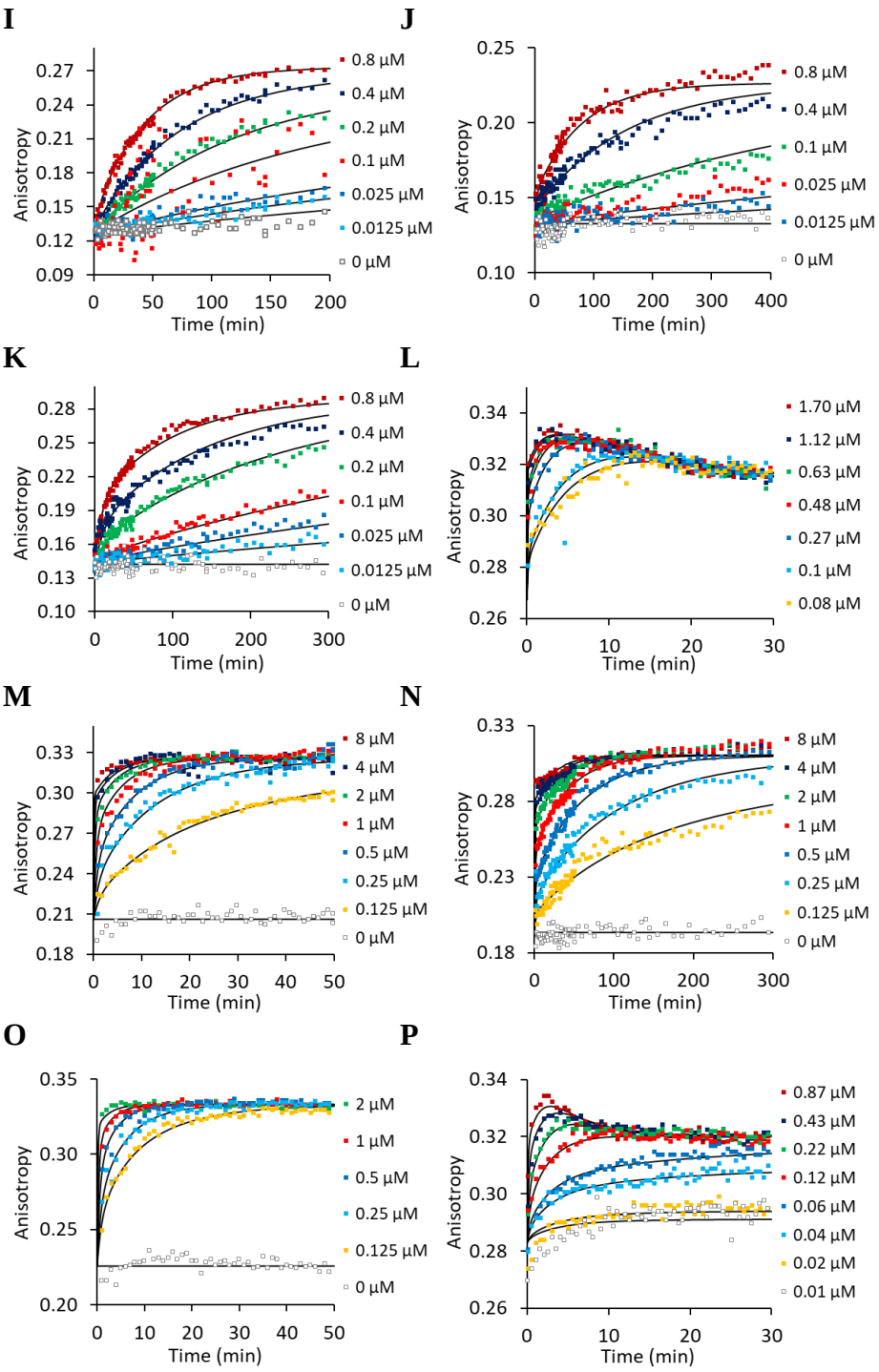
$$\Delta G^{\ddagger} = -R.T.ln\left(\frac{k_i.h}{k_B T}\right) \qquad \textit{Supplementary Equation S8}$$

The free energy of ground state (ΔG^0) was calculated from the equilibrium dissociation constant for enzyme-ligand complex (K_D) using **Supplementary Equation 9** at [E] = 10 μM (the component in excess) reflecting better the real reaction condition rather than the 1 M standard state. All thermodynamic parameters calculated at the reference temperature 30 °C (303.15 K) are summarized in **Supplementary Table S6** and **Figure S4**.

$$\Delta G^0 = -R.T.ln\left(\frac{[E]}{K_D}\right) \qquad \textit{Supplementary Equation S9}$$



Supplementary Figure S3. **Global data analysis using numerical integration.** Anisotropy kinetic traces obtained upon mixing of 0.001 µM **TMR** with DhaAHT (A) and DhaAH272F (B), 0.01 µM **TMR** with LinBH272F (C) and DmmAH315F (D), 0.1 µM **1B** with DhaAHT (E), DhaAH272F (F), LinBH272F (G), DmmAH315F (H), 0.01 µM **1D** with DhaAHT (I), DhaAH272F (J), LinBH272F (K), and 0.025 µM **1D** with DmmAH315F (L), and 0.1 µM **1E** with DhaAHT (M), DhaAH272F (N), LinBH272F (O) and DmmAH315F (P). The experiments were performed at 30°C in PBS (pH 7.4) with 0.01 % (w/v) CHAPS. The solid lines represent the best global fit.



Supplementary Figure S3. (Continued)

Supplementary Table S3. **Kinetic constants obtained by global fitting.** The parameters are shown with standard errors derived in globally fitting the data by nonlinear regression. All experiments were performed at 30° C in PBS (pH 7.4) with 0.01 % (w/v) CHAPS. The standard errors of the fitted parameters k_1 and k_{-1} were propagated to yield the error estimates in the calculated value K_D.[a]

		k_1 / $\mu M^{-1} \cdot min^{-1}$	k_{-1} / min^{-1}	k_2 / min^{-1}	K_D / μM
TMR	DhaAHT	39.7 ± 0.6	0.051 ± 0.003	0.004 ± 0.003	$(1.28 \pm 0.08) \times 10^{-3}$
	DhaAH272F	0.40 ± 0.02	0.127 ± 0.004	0.0014 ± 0.0009	0.32 ± 0.02
	LinBH272F	0.0013 ± 0.0001	0.003 ± 0.001	0.021 ± 0.002	2.34 ± 0.8
	DmmAH315F	0.0042 ± 0.0005	n.d.	0.027 ± 0.004	n.d.
1B	DhaAHT	0.17 ± 0.05	0.2 ± 0.1	0.08 ± 0.04	1.2 ± 0.7
	DhaAH272F	0.07 ± 0.01	0.04 ± 0.01	0.017 ± 0.008	0.6 ± 0.2
	LinBH272F	0.07 ± 0.02	0.04 ± 0.02	0.007 ± 0.006	0.6 ± 0.3
	DmmAH315F	16 ± 2	1.1 ± 0.2	0.19 ± 0.01	0.07 ± 0.02
1D	DhaAHT	0.10 ± 0.02	0.07 ± 0.06	0.05 ± 0.04	0.7 ± 0.6
	DhaAH272F	0.039 ± 0.008	0.08 ± 0.04	0.04 ± 0.03	2 ± 1
	LinBH272F	0.091 ± 0.009	0.04 ± 0.02	0.014 ± 0.009	0.4 ± 0.2
	DmmAH315F	25 ± 3	4.6 ± 0.6	0.67 ± 0.05	0.18 ± 0.03
1E	DhaAHT	1.5 ± 0.3	0.5 ± 0.2	0.25 ± 0.04	0.3 ± 0.1
	DhaAH272F	0.7 ± 0.2	0.4 ± 0.1	0.04 ± 0.01	0.6 ± 0.2
	LinBH272F	3.0 ± 0.6	0.4 ± 0.3	0.4 ± 0.1	0.13 ± 0.10
	DmmAH315F	9.9 ± 0.8	1.6 ± 0.3	0.68 ± 0.07	0.16 ± 0.03

[a]n.d. means not determined.

Supplementary Table S4. Scaling factors of the observed signal included in global fit. Factor *a* scales the observable signal to the concentration of the free ligand, factors *b* and *c* reflect the change of the signal corresponding to the formation of the binding complex (E.L) and the final covalent alkyl-enzyme complex (E-L), respectively.

		a	*b*	*c*
TMR	DhaAHT	89.7 ± 0.3	3.81 ± 0.01	3.49 ± 0.04
	DhaAH272F	89.7 ± 0.1	9.8 ± 0.2[a]	9.8 ± 0.2[a]
	LinBH272F	6.86 ± 0.02	2.4 ± 0.2	4.12 ± 0.02
	DmmAH315F	2.1 ± 0.2	7.3 ± 0.1	3.4 ± 0.1
1B	DhaAHT	1.568 ± 0.005	1.67 ± 0.02	1.865 ± 0.007
	DhaAH272F	1.542 ± 0.004	1.146 ± 0.009	1.428 ± 0.005
	LinBH272F	1.623 ± 0.005	1.15 ± 0.01	1.499 ± 0.007
	DmmAH315F	1.97 ± 0.01	1.30 ± 0.01	1.36 ± 0.01
1D	DhaAHT	12.54 ± 0.07	1.6 ± 0.6	2.21 ± 0.02
	DhaAH272F	13.28 ± 0.04	1.8 ± 0.1	1.71 ± 0.01
	LinBH272F	14.17 ± 0.03	1.71 ± 0.02	2.03 ± 0.02
	DmmAH315F	10.7 ± 0.02	1.22 ± 0.01	1.26 ± 0.01
1E	DhaAHT	2.069 ± 0.005	1.46 ± 0.01	1.575 ± 0.004
	DhaAH272F	1.935 ± 0.004	1.483 ± 0.006	1.604 ± 0.003
	LinBH272F	2.268 ± 0.004	1.43 ± 0.01	1.468 ± 0.003
	DmmAH315F	2.425 ± 0.007	1.139 ± 0.004	1.187 ± 0.006

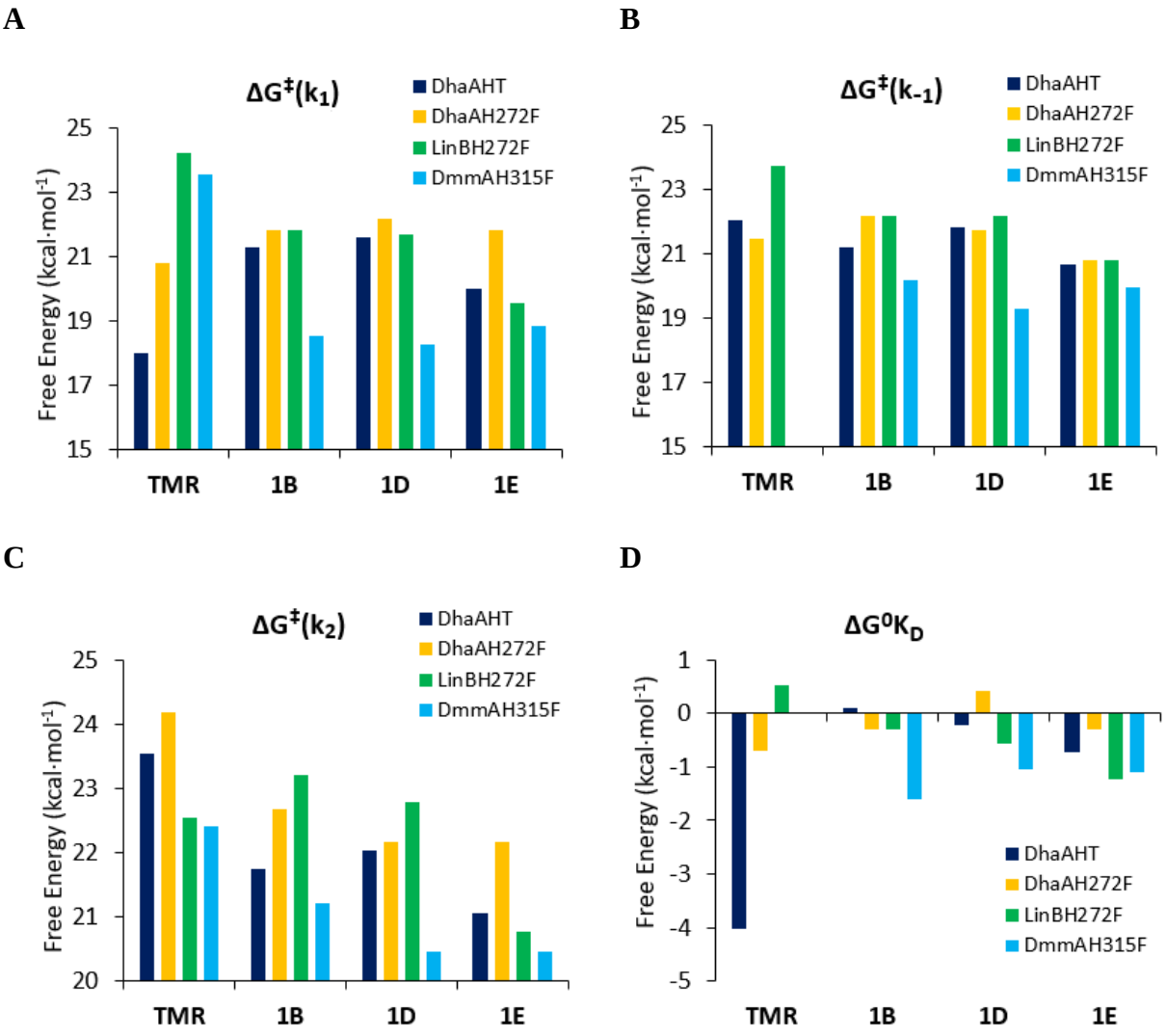
[a] Common scaling factor used for the enzyme-ligand and covalent alkyl-enzyme complex

Supplementary Table S5. Confidence contour analysis for global data fitting. The lower and upper limits for each rate constant (Scheme 1) derived from the confidence contours for χ^2 threshold at boundary 0.95.

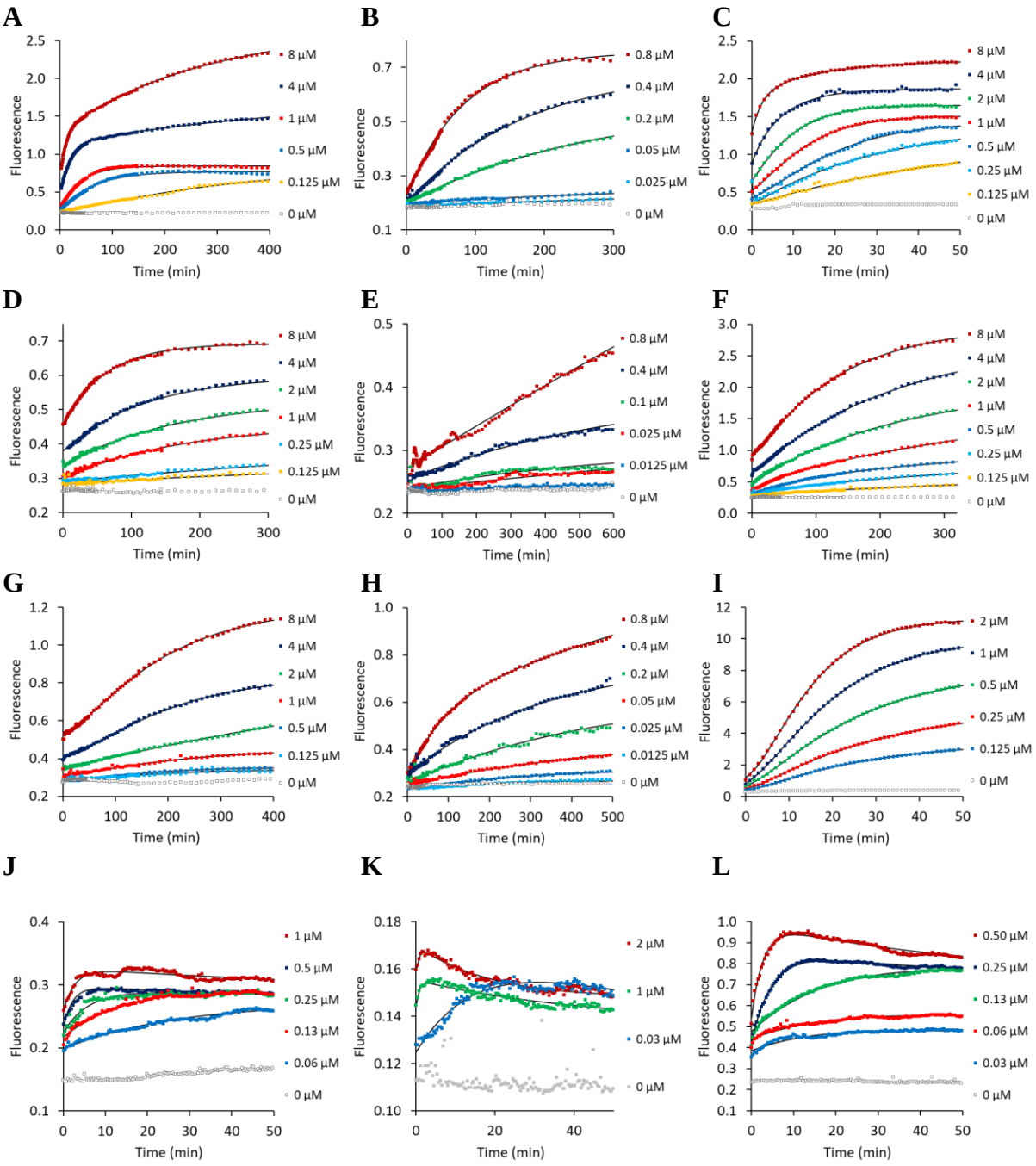
		k_1 / $\mu M^{-1} \cdot min^{-1}$	k_{-1} / min^{-1}	k_2 / min^{-1}
TMR	DhaAHT	38 - 42	0.045 - 0.058	< 0.009
	DhaAH272F	0.36 - 0.45	0.118 - 0.136	0.0003 - 0.003
	LinBH272F	0.0011 - 0.0014	< 0.007	0.016 - 0.032
	DmmAH315F	0.0046 - 0.0052	< 0.0001	0.019 - 0.027
1B	DhaAHT	0.1 - 2.1	0.05 - 5	0.05 - 0.12
	DhaAH272F	0.05 - 0.11	0.01 - 0.10	0.012 - 0.024
	LinBH272F	0.07 - 0.25	0.01 - 0.13	0.002 - 0.011
	DmmAH315F	> 7.6	> 0.38	0.12 - 0.37
1D	DhaAHT	0.03 - 0.59	< 0.6	0.01 - 1
	DhaAH272F	0.04 - 0.13	< 1.5	0.003 - 0.8
	LinBH272F	0.08 - 0.15	0.02 - 0.12	0.01 - 0.05
	DmmAH315F	> 16	> 1.9	0.43 - 1.31
1E	DhaAHT	1.2 - 10	0.2 - 6	0.19 - 0.39
	DhaAH272F	0.3 - 1.4	0.08 - 0.9	0.02 - 0.05
	LinBH272F	2.4 - 10	0.1 - 3.6	0.2 - 0.9
	DmmAH315F	6.3 - 24.1	0.57 - 8.57	0.38 - 2.21

Supplementary Table S6. **Thermodynamic parameters derived from experimental measurements**. The free energies of ground state ΔG^0 and free energy of activation $\Delta G^{\ddagger}$ calculated from the fitted parameters at reference temperature 30 °C (303.15 K) and [E] = 10 μM reflecting the real reaction condition.

		$\Delta G^{\ddagger}(k_1)$ / kcal·mol^{-1}	$\Delta G^{\ddagger}(k_{-1})$ / kcal·mol^{-1}	$\Delta G^{\ddagger}(k_2)$ / kcal·mol^{-1}	$\Delta G^0_{K_D}$ / kcal·mol^{-1}
TMR	DhaAHT	18.0	22.0	23.5	-4.0
	DhaAH272F	20.8	21.5	24.2	-0.7
	LinBH272F	24.2	23.7	22.5	0.5
	DmmAH315F	23.5	n.d.	22.4	n.d.
1B	DhaAHT	21.3	21.2	21.7	0.1
	DhaAH272F	21.8	22.2	22.7	-0.3
	LinBH272F	21.8	22.2	23.2	-0.3
	DmmAH315F	18.6	20.2	21.2	-1.6
1D	DhaAHT	21.6	21.8	22.0	-0.2
	DhaAH272F	22.2	21.7	22.2	0.4
	LinBH272F	21.7	22.2	22.8	-0.6
	DmmAH315F	18.3	19.3	20.5	-1.0
1E	DhaAHT	20.0	20.6	21.1	-0.7
	DhaAH272F	21.8	20.8	22.2	-0.3
	LinBH272F	19.6	20.8	20.8	-1.2
	DmmAH315F	18.8	19.9	20.5	-1.1



Supplementary Figure S4. Thermodynamic parameters derived from experimental measurements. The energy of activation $\Delta G^{\ddagger}$ for individual kinetic steps (A) k_1, (B) k_{-1} and (C) k_2 and (D) free energy of ligand binding ΔG^0 ($1/K_A = K_D = k_{-1}/k_{-1}$) calculated from the fitted parameters at reference temperature 30 °C (303.15 K) and [E] = 10 μM reflecting the real reaction condition.



Supplementary Figure S5. Fluorescence intensity measurements. Fluorescence intensity signal recorded upon mixing of different concentration of DhaAHT with 0.1 µM **1B** (A), 0.01 µM **1D** (B) and 0.1 µM **1E** (C), different concentration of DhaAH272F with 0.1 µM **1B** (D), 0.01 µM **1D** (E) and 0.1 µM **1E** (F), different concentration of LinBH272F with 0.1 µM **1B** (G), 0.01 µM **1D** (H) and 0.1 µM **1E** (I) and different concentration of DmmAH315F with 0.1 µM **1B** (J), 0.025 µM **1D** (K) and 0.1 µM **1E** (L). The experiments were performed at 30° C in PBS with 0.01 % (w/v) CHAPS. The solid lines represent the best fit.

Non-specific ligand binding

Methods

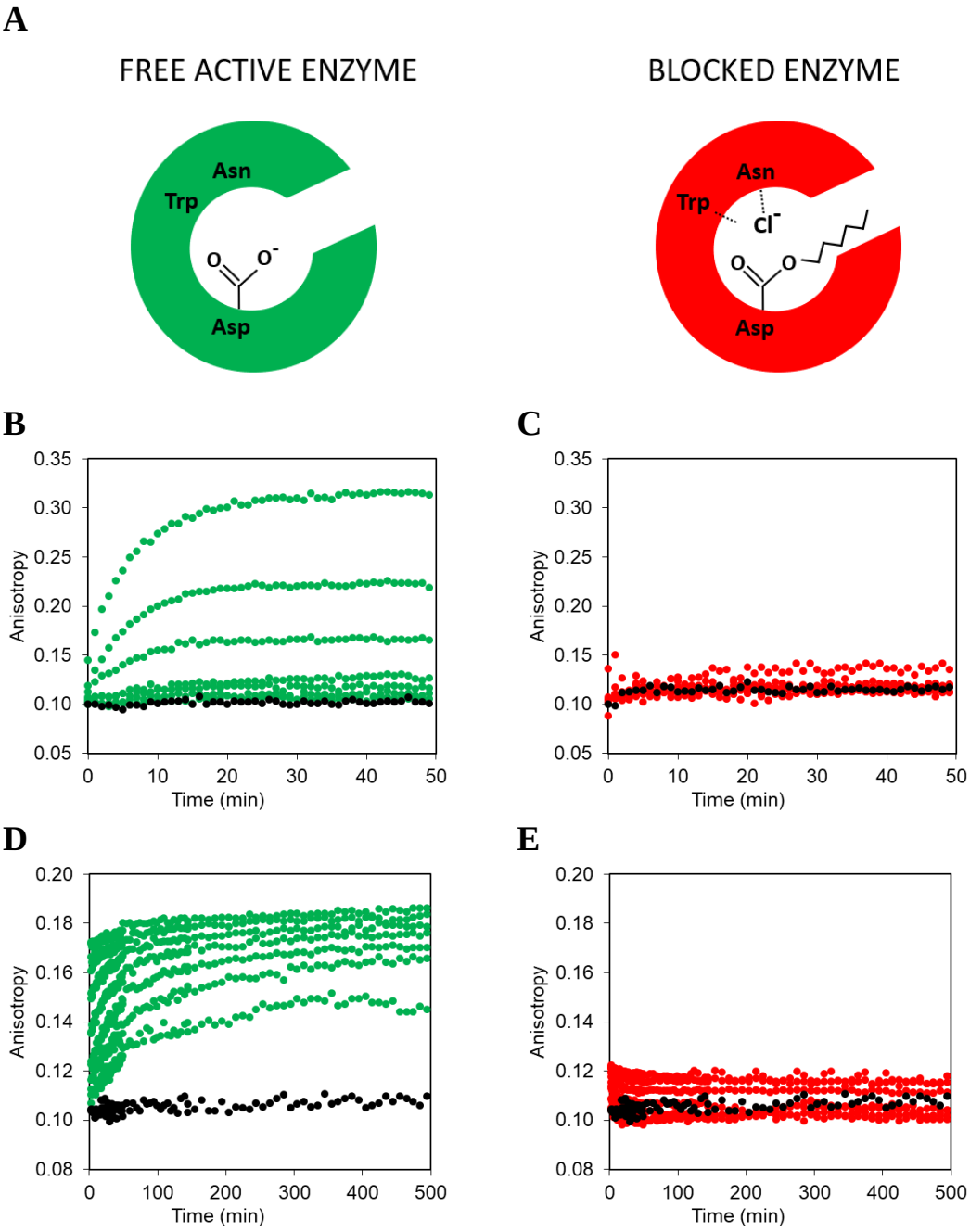
Non-specific interactions of **TMR** and **1E** ligands were measured by monitoring the changes in anisotropy upon mixing with DhaAH272F in the active form and after blocking the active sites with conventional non-fluorescent substrate 1-chlorohexane. For blocking the active site, the enzyme was pre-incubated with excess 1-chlorohexane (solubility limit ~0.75 mM) at 30 °C and 150 rpm for >1 hour to ensure complete formation of the covalent alkyl-enzyme complex. Fluorescence anisotropy was measured in black coloured 96-well polypropylene plates. The reaction mixture, with a total volume of 200 µl, consisted of PBS buffer, CHAPS to a final concentration of 0.01 % (w/v), the fluorescent ligand and blocked enzyme at concentrations corresponding to the experiments in **Section II**. The time course of the signal was monitored using Infinite F500 plate reader (Tecan, Switzerland) equipped with polarization filters with excitation/emission wavelengths 544 nm/580 nm or 544 nm/620 nm at 30 °C. The signal of the enzyme-free incubation of the ligands in the reaction buffer was monitored over time as a negative control.

Results

The effects of non-specific interactions of the two types of ligands, **TMR** and **1E**, were studied by monitoring the anisotropy upon the interaction of ligands with the free enzyme and with the enzyme with blocked active site after the reaction with 1-chlorohexane (**Supplementary Figure S6A**). The experiments were designed to analyse the effect of non-specific interactions during both fast (DhaAH272F+**TMR**) and slow labelling reaction (DhaA272F+**1E**).

A negligible non-specific binding was observed for DhaAH272F+**TMR** pair, which has high reactivity (**Supplementary Figures S6B and 6C**). For the slower non-optimal reaction of DhaAH272F with **1E**, when the reaction needs to be performed at 100-times higher ligand concentration, the non-specific binding was more pronounced. However, the specific binding was dominant in both cases, and non-specific interactions do not represent a significant bias in the kinetic data. Weak and reversible non-specific interactions are also very unlikely to affect the efficiency of the ligand incorporation into the enzyme, i.e., they compete against stronger specific binding and, more importantly, the irreversible formation of the final covalent complex.

In case the non-specific binding might cause adverse effects in the target applications of the HaloTag technology, it is desirable to examine their extent in this simple way and to consider the result of such an analysis in the final selection of the optimal enzyme-ligand pair.



Supplementary Figure S6: Effect of non-specific interactions of TMR and 1E. Scheme of the HaloTag enzyme in the active form and after blocking the active sites with conventional non-fluorescent substrate 1-chlorohexane (A). Anisotropy kinetic traces obtained upon mixing 0.001 μM **TMR** with 0.00125-0.08 μM DhaAH272F in free active form (B) and after active site blocking (C). Anisotropy kinetic traces obtained upon mixing 0.1 μM **1E** with 0.125-8 μM DhaAH272F in free active form (D) and after active site blocking (E). The green and red curves show interaction with active and blocked dehalogenases, respectively. The black signal represents a negative control, the incubation of the ligands in reaction buffer without enzyme.

Section III: MALDI-TOF MS analysis

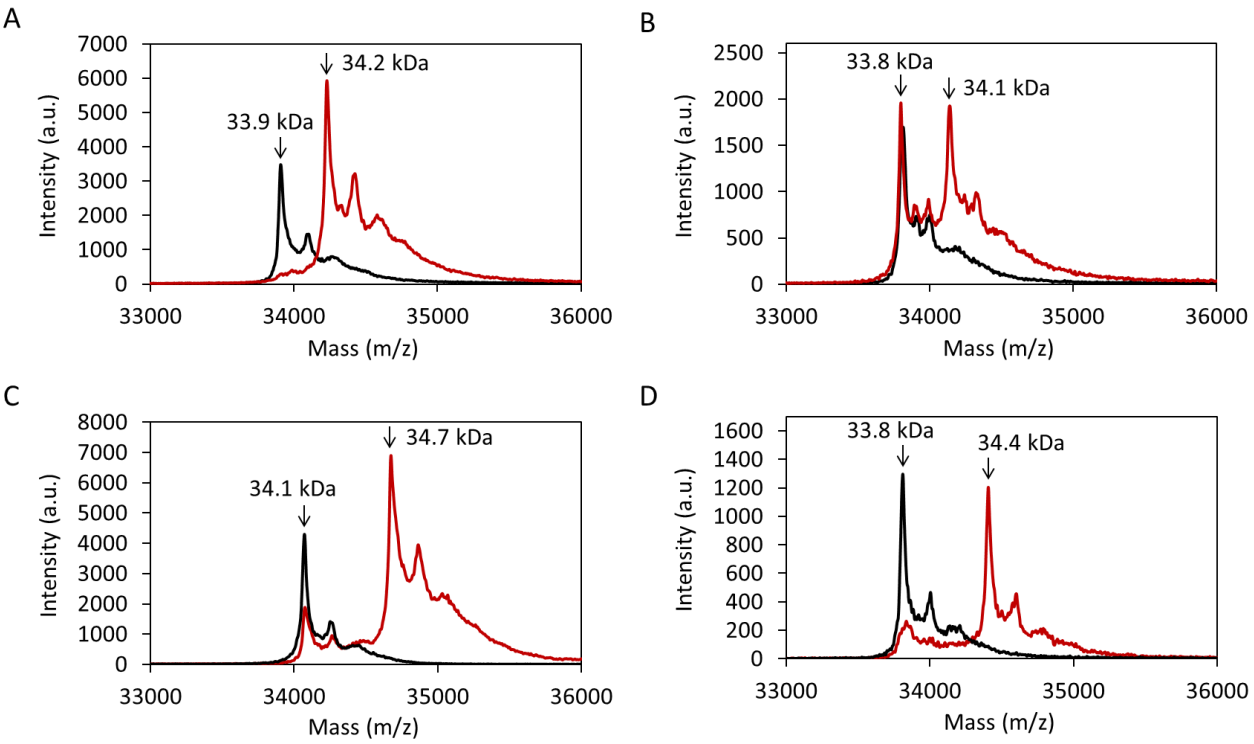
The matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) experiments were performed to verify the formation of covalent complexes between the HaloTag proteins and fluorescent ligands.

Methods

The phosphate buffer used for the purification of haloalkane dehalogenases DhaAHT, DhaA31H272F, and LinBH272F was exchanged with glycine buffer (20 mM, pH 8.6) by overnight dialysis. The protein-ligand reaction proceeded in a volume of 100-200 µl at 4 or 25 °C. The concentration of the enzyme was 1.5 $mg \cdot ml^{-1}$ and the concentration of the **1D** and **TMR** ligands was adjusted to achieve 100% saturation of the enzyme. MALDI-TOF mass spectra were recorded on an Ultraflextreme instrument (Bruker Daltonics, Billerica, Germany) operated in linear mode for detecting positive ions. Ferulic acid (12.5 $mg \cdot ml^{-1}$ in a mixture of water:acetonitrile:formic acid 50:33:17, v/v/v) was used as the MALDI matrix. 0.6 µl of the enzyme sample was mixed with 2.4 µl of matrix solution, after which 0.6 µl of the mixture was deposited onto a stainless steel MALDI target. Protein Calibration mixture II (Bruker Daltonics, Billerica, Germany) was used for external calibration of the mass spectra, which were processed with FlexAnalysis 3.4 software (Bruker Daltonics, Billerica, Germany).

Results

MALDI-TOF MS spectra of DhaAHT and LinBH272F were determined before and after interaction with the selected 4-stilbazolium-based ligand **1D** (**Supplementary Figure S7A** and **7B**). The analysis confirmed the formation of covalent complexes between the HaloTag proteins and the fluorescent ligands. Similarly, we have previously confirmed the formation of a covalent complex between the HaloTag enzymes and the **TMR** ligand. In these reactions, LinBH272F and the mutated variant DhaA31H272F were analyzed (**Supplementary Figure 7C** and **7D**).



Supplementary Figure S7. Comparison of MALDI-TOF MS spectra of HaloTag proteins in a free form and upon their interaction with HaloTag ligands. (A) MALDI-TOF MS spectra of DhaAHT before (black line) and after interaction (red line) with **1D** ligand. The peaks at 33.9 and 34.2 kDa corresponds to the free and the complexed forms of the protein, respectively. (B) MALDI-TOF MS spectra of LinBH272F before and after interaction with **1D** ligand. The peaks at 33.8 and 34.1 kDa corresponds to the free and the complexed forms of the protein, respectively. (C) MALDI-TOF MS spectra of DhaA31H272F before (black line) and after interaction (red line) with **TMR** ligand.[9] The peaks at 34.1 and 34.7 kDa corresponds to the free and the complexed forms of the protein, respectively. (D) MALDI-TOF MS spectra of LinB 57 before (black line) and after interaction (red line) with **TMR** ligand.[9] The peaks at 33.8 and 34.4 kDa corresponds to the free and the complexed forms of the protein, respectively. Experiments were performed in 20 mM glycine buffer pH 8.6 at 4 °C (binding experiments with **1D** ligand) and 25 °C (binding experiments with **TMR** ligand).

Section IV: Computational analysis

Methods

In silico mutagenesis

The models of DhaAHT, DhaAH272F, LinBH272F and DmmAH315F were constructed by *in silico* mutagenesis performed on the crystal structures of the template enzymes, obtained from the RCSB Protein Data Bank[10] (PDB entry 4E46 for DhaA, 1MJ5 for LinB, and 3U1T for DmmA). The water molecules and ions were removed, and the protein chain was renumbered in order to start from position 1. The resulting structure was minimized by Rosetta using the *minimize_with_cst* module. The minimization was performed according to Kellog *et al*[11]. Both backbone and side chain optimization was enabled (*sc_min_only false*), the distance for full atom pair potential was set to 9 Å (*fa_max_dis 9.0*), the standard weights for the individual terms in the energy function were used with a constraint weight 1 (*constraint_weight* 1.0). The output from the minimization step was used by the script *convert_to_cst_file.sh* for creation of the constraints file. In order to calculate the most stable conformers of each mutant, Protocol 16 was followed. For that, *ddg_monomer* module of Rosetta was used according to Kellog *et al*[11], incorporating the backbone flexibility. The soft-repulsive design energy function (*soft_rep_design weights*) was used for side chains repacking and backbone minimization (*sc_min_only false*). The optimization was performed on the whole protein without distance restriction (*local_opt_only false*). The previously generated constraints *cst* file was used during minimization (*min_cst true*) to impose a restraint of 0.5 Å on the C_α atoms. The optimization was performed in three rounds with increasing weight on the repulsive term (*ramp_repulsive* true). The structure with lowest energy (*mean false, min true*) was selected from the 50-iteration cycle (*iterations 50*), and it was used as the final result to obtain the minimized model of the mutant. All calculations used the *talaris2014*[12,13] force field.

Tunnel calculation

CAVER 3.02[14] was used to identify the tunnels in the predicted models of DhaAHT, DhaAH272F, LinBH272F and DmmAH315F. The tunnels were calculated using the default settings and a probe radius of 0.8 Å, a shell radius of 5 Å and shell depth 4 Å. The starting point for the tunnel calculation, located in the active site, was defined by the center of mass of the atoms OD1 and OD2 of D106, for the DhaA variants, or D108 for the LinB variants.

Ligand preparation

The structures of **TMR** and **1E** ligands were constructed and minimized using Avogadro 2[15]. The minimization step was performed by the Auto Optimization Tool of Avogadro, using the UFF force field[16] with steepest descent algorithm. The resulting structures were then submitted to further optimization and calculation of their partial atomic charges using Gaussian 09[17], with the Hartree-Fock method and 6-31G(d) basis set in vacuum. These settings were similar to those previously used to parameterize other ligands[18]. For the **TMR** ligand, however, the final structure showed the closing of the carboxylate with the aromatic system to form a lactone ring. To test the validity of this result, a more robust calculation was performed with B3LYP and the 6-311+G(d,p) basis set and implicit solvent (the Polarizable Continuum Model). Finally, the latter conditions were used with both fluorescent ligands. The *antechamber* module of AmberTools 16[19] was used to extract the RESP charges for the ligands from the Gaussian output files,

which were compiled into the PREPI and FRCMOD parameters files using the atom types of the General Amber force field (GAFF).

System preparation and equilibration

The hydrogen atoms were removed from the model structures of DhaAHT and LinBH272F, and added using the H++ server[20], calculated in implicit solvent at pH 7.5, 0.1 M salinity, internal dielectric constant of 10 and external of 80. The water molecules from the original crystal structure of DhaA that did not overlap with any of the protein atoms were added to those structures. The **TMR** or **1E** ligands were randomly placed at least 5 Å from the protein by the High Throughput Molecular Dynamics (HTMD)[21] python script. The systems were solvated in a cubic water box of TIP3P[22] water molecules with the edges at least 10 Å away from the protein by the *solvate* module of HTMD. Cl^- and Na^+ ions were added to neutralize the charge of the protein and get a final salt concentration of 0.1 M. The systems topology was built, using the *amberbuild* module of HTMD, with the ff14SB[23] Amber force field and the previously compiled PREPI parameters file for the ligands.
The systems were equilibrated using the *Equilibration_v2* module of HTMD[21]. The system was first minimized using conjugate-gradient method for 500 steps. Then the system was heated and minimized as follows: (I) 500 steps (2 ps) of NVT thermalization with the Berendsen barostat to 300 K with constraints on all heavy atoms of the protein, (II) 625 000 steps (2.5 ns) of NPT equilibration with Langevin thermostat with 1 $kcal \cdot mol^{-1} \cdot Å^{-2}$ constraints on all heavy atoms of the protein and (III) 625 000 steps (2.5 ns) of NPT equilibration with the Langevin thermostat without constraints. During the equilibration simulations, holonomic constraints were applied on all hydrogen-heavy atom bond terms and the mass of the hydrogen atoms was scaled with factor 4, enabling 4 fs time step[24–27]. The simulations employed periodic boundary conditions, using the particle mesh Ewald method for treatment of interactions beyond 9 Å cut-off, electrostatic interactions suppressed for more than 4 bond terms away from each other and the smoothing and switching van der Waals and electrostatic interaction cut-off at 7.5 Å[25].

Adaptive sampling

HTMD was used to perform adaptive sampling of the binding of the ligand's reactive end to the protein's active site. The 40 ns production MD runs were started with the system that resulted from the equilibration cycle and employed the same settings as the last step of the equilibration. The trajectories were saved every 0.1 ns. Adaptive sampling was performed using the distance between the reactive carbon atom of the ligand and the carboxylic carbon atom (Cγ) of the catalytic nucleophile D106 as the reaction coordinate, and time-based independent component analysis (tICA)[28] in 1 dimension. 50 epochs of 10 MDs each were performed for every system, corresponding to a cumulative simulation time of 20 μs.

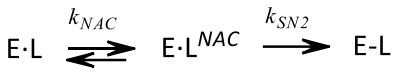
Markov state model construction

The simulations were made into a simulation list using HTMD[29], the water was filtered out, and unsuccessful simulations with length less than 40 ns were omitted. This resulted in 20 μs of cumulative simulation time (500 × 40 ns). The binding dynamics was studied by the same metric used in the adaptive sampling: the distance between the ligand's reactive carbon atom and the Cγ atom of D106 (DhaA numeration). The data was clustered using MiniBatchKmeans algorithm to 1000 clusters. A lag time of 25 ns was used in the models to construct 4 Markov states, and the Chapman-Kolmogorov test was performed to assess the quality of the constructed states. The resulting Markov models allowed us to estimate the transition mean times between the most-unbound and most-bound states, kinetic rates of binding and

unbinding, the equilibrium constants, energies and the population fractions of the bound state (f_{bound}). A bootstrapping calculation was performed with 80% of the data and repeated 100 times to estimate the errors in the calculated kinetic and population parameters.

Pre-reactive complexes

The pre-reactive complexes were analyzed in order to find the potentially reactive configurations to undergo the S_N2 reaction. The *cpptraj*[30] module of AmberTools 16[19] was used to calculate distances and angles between the nucleophile and substrate atoms, and an in-house *python* script evaluated them according to Hur *et al.*[31]: the distance between one of the nucleophile's carboxyl oxygen atoms and the halogen-bound carbon atom had to be ≤ 3.41 Å, and simultaneously the angle formed by that oxygen, carbon and the halide atoms must be ≥ 157 °. We also required at least a weak hydrogen bond between the reactive chlorine atom and the halide-stabilizing residues, defined by the distance between the halide and the indole hydrogen of W107, or one of the side chain NH hydrogens of N41, to be ≤ 3.0 Å.
To calculate further the chemical step, we took into account the kinetic scheme in the above Supplementary Scheme 1, and we dissected it into two parts, described in the Supplementary Scheme 2:



Supplementary Scheme 2

Based on this scheme, we can derive **Supplementary Equations S10–S12**:

$$\Delta G^{\ddagger}_{2} = \Delta G_{NAC} + \Delta G^{\ddagger}_{SN2}$$

Supplementary Equation S10

$$\Delta G_{NAC} = -R.T.lnK_{NAC}$$

Supplementary Equation S11

Where the equilibrium constant for the NAC formation is guiven by:

$$K_{NAC} = \frac{[NAC]}{[Bound] - [NAC]}$$

Supplementary Equation S12

According to **Supplementary Equation S12**, we can estimate K_{NAC} based on the number of snapshots with identified *NAC*, [*NAC*], the probability –or fraction– of snapshots in the bound state, P_{bound}, and the total number of snapshots (2×10^5), as: $K_{NAC} = [NAC]/(P_{bound} \times 2 \times 10^5 - [NAC])$.

QM/MM adiabatic mapping

To evaluate and compare the energetics of the S_N2 reaction of the **TMR** and **1E** ligands with DhaAHT and LinBH272F (**Figures 3C–3D** in the main text) and calculate the respective energy barrier, $\Delta G^{\ddagger}$, we studied

the potential energy surface (PES) along the reaction coordinate using a quantum mechanics/molecular mechanics (QM/MM) hybrid approach[32,33]. For that, all the snapshots of the pre-reactive complex, identified as described above, were subjected to QM/MM calculations[34], as implemented in AMBER. The topology of each structure was prepared by tLEaP module using the ff14SB force field for the proteins and the PREPI parameters for ligand. The complexes were minimized in vacuum (*igb=6*). Five rounds of optimization, each one consisting of 500 cycles of steepest descent followed by 500 conjugate gradient cycles, were performed as: (i) one step with all heavy atoms restrained with 500 kcal·mol^{-1}·Å^{-2} harmonic force constant, and (ii) four steps with decreasing restraints on the protein backbone atoms with 500, 125, 25 and 1 kcal·mol^{-1}·Å^{-2} force constant. Adiabatic mapping along the reaction coordinate was performed by the sander module of AMBER 16[19]. The QM part of the system contained the ligand molecule, the side-chains of the halide stabilizing residues (N41 and W107) and the catalytic aspartate (D106), and had charge -1 (for the **TMR** ligand) or 0 (for **1E**). The semi-empirical PM6 Hamiltonian was used to treat the QM part of the system[35] and the ff14SB force field to treat the MM part. The QM/MM boundary was treated through explicit link atoms and the cutoff for the QM/MM charge interactions was set to 999 Å. The backbone was constrained with a force constant of 1.0 kcal·mol^{-1}·Å^{-2}. The reaction coordinate was defined as the distance between the nearest OD atom of D106 and the C-atom of the ligand under attack. The tracking along the reaction coordinate was performed in decrements of 0.05 Å, each involving 1,000 minimization steps of the limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm[36]. The total potential energy of the system was extracted from the AMBER output files for each step. The energy barrier $\Delta G^{\ddagger}$ was calculated as the difference between the lowest energy of the ground state and the energy of the transition state, and corresponds to $\Delta G^{\ddagger}_{SN2}$. Combining **Supplementary Equations S8** and **S9** we derive **Supplementary Equation S13**:

$$\Delta G^{\ddagger}_{2} = \Delta G_{NAC} + \Delta G^{\ddagger}_{SN2} = -RT.lnK_{NAC} + \Delta G^{\ddagger}_{SN2}$$

Supplementary Equation S13

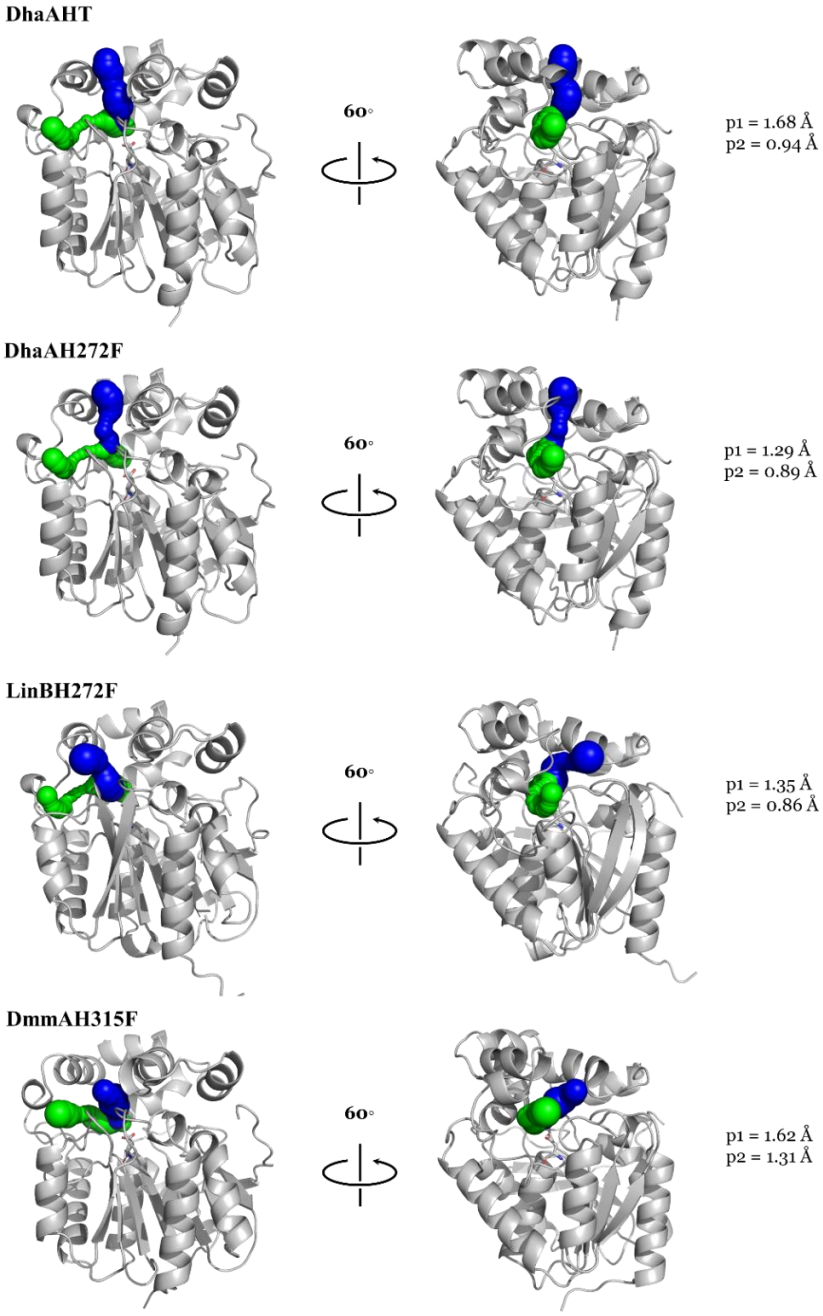
which allows us to estimate $\Delta G^{\ddagger}_{2}$, using the previously determined K_{NAC} (from the previous section) and the newly determined $\Delta G^{\ddagger}_{SN2}$, and to compare it with the experimentally determined $\Delta G^{\ddagger}_{2}$.

CaverDock simulations

The p1 tunnels previously identified with CAVER on the modeled structures of DhaAHT, DhaAH272F, LinBH272F and DmmAH315F were used for consecutive docking of the fluorescent ligands **TMR**, **1B**, **1D** and **1E**, using the standalone version of CaverDock[37,38]. The pdb files of the tunnels were discretized in discs spaced by 0.3 Å and subsequently extended by 5 Å towards the outside. The ligands were docked from the outside inwards, by constraining the position of the reacting carbon atom of the ligands to the consecutive discs. The trajectories and binding energy profiles were calculated and analyzed for all the ligands and protein systems.

Results

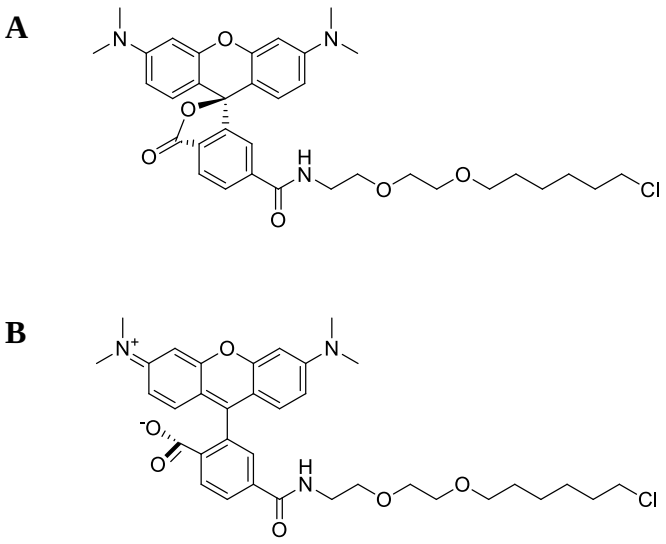
Molecular access tunnels



Supplementary Figure S8. Access tunnels found in the models of DhaAHT, DhaAH272F, LinBH272F and DmmAH315F. The main tunnel (p1, blue) and slot tunnel (p2, green), calculated using CAVER 3.02 [14] with a ligand of 0.8 Å radius, using the origin at the carboxyl oxygen atoms of the catalytic aspartate (D106 in DhaA numeration, shown as sticks); the respective bottleneck radii are indicated.

Density Functional Theory modelling of ligands

The fluorescent ligands **TMR** and **1E** were simulated here as the ligands that will bind to the active site of the HLDs under study. *Ab initio* calculations were performed to minimize and calculate the partial charges of those ligands, and were carried out in vacuum with the Hartree-Fock method and the 6-31G(d) basis set, as previously[18]. However, for the **TMR** those settings resulted in the closure of the carboxylic acid with the aromatic system to form a lactone ring (**Supplementary Figure S9A**). This was surprising but not unprecedented, since the tautomeric equilibrium of similar structures has been reported before[39]. To confirm this result, we performed higher level Density Functional Theory (DFT) calculations with the B3LYP theory, the 6-311+G(d,p) basis set, and implicit solvent. This combination can be more suitable to treat ionized species, since it incorporates multiple polarization and diffuse functions. The structure with the open ring was obtained (**Supplementary Figure S9B**), as initially expected. The explanation for this behavior can be that, while in vacuum the preferred structure is the one with a closed lactone ring without charge separation, in water the **TMR** exists preferentially in the conformation with free carboxylate due to the stabilization of the charges provided by the solvent.



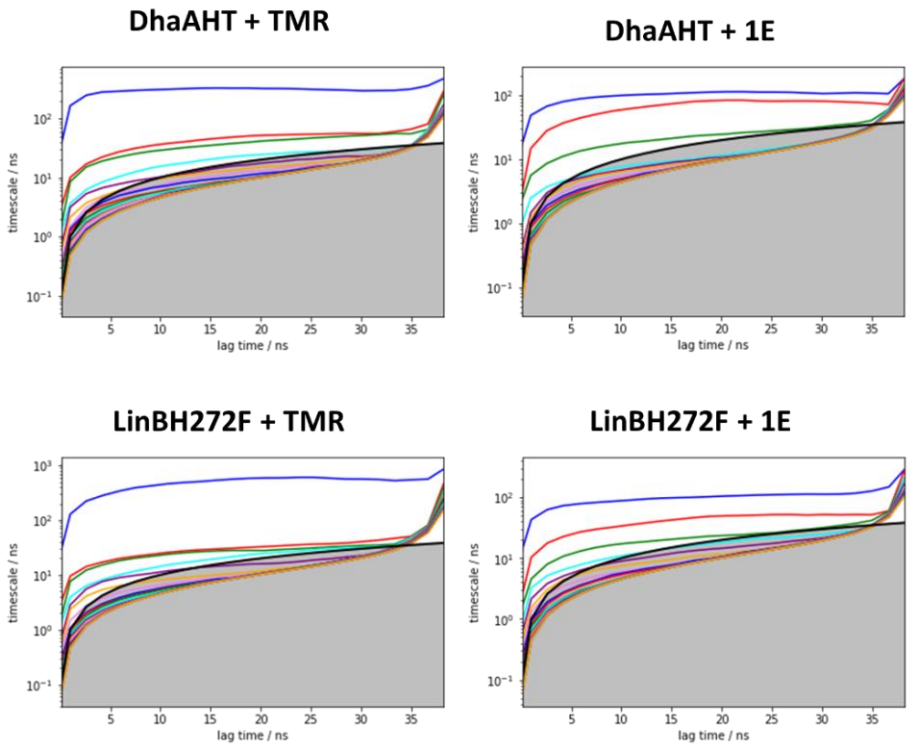
Supplementary Figure S9. Structures obtained from the modelling studies of the TMR ligand. A) TMR with the closed lactone ring, resulting from the initial quantum mechanical calculations in vacuum; and B) with the free carboxylic acid, which was confirmed with the higher level DFT calculations with implicit solvent.

Adaptive sampling and ligand binding

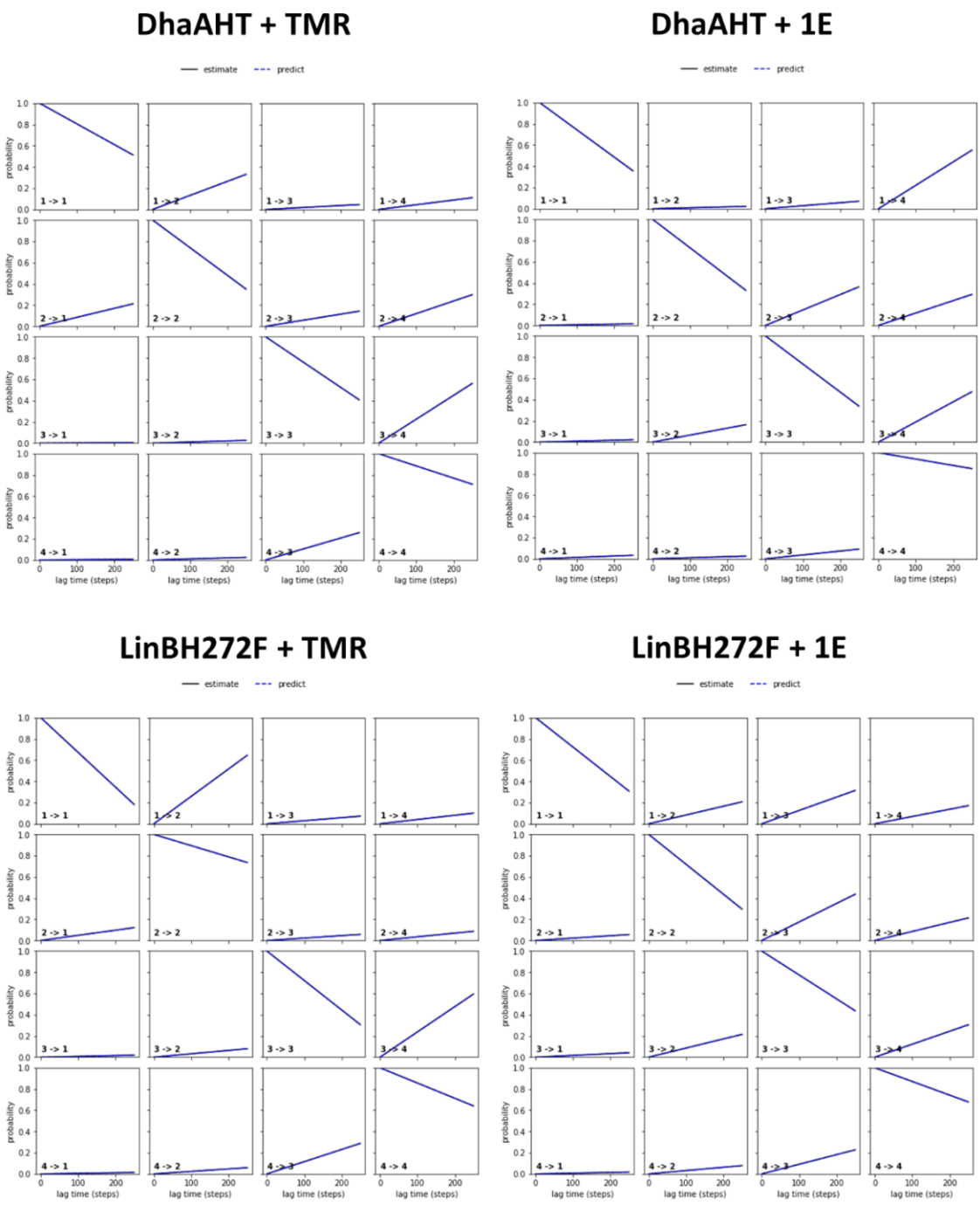
The adaptive sampling simulations were performed based on a metric defined by the distance of the reacting groups: the carbon atom in the ligand and the carboxylic carbon of the nucleophile residue, D106. The resulting simulations (with a total cumulative time of 20 μs) were analyzed by the Markov state model (MSM) method. For all the systems, several long time-scale transitions have achieved convergence of their relaxation timescales after reasonable lag times (< 25 ns), as can be observed by the implied time scale plots (ITS; **Supplementary Figure S10**). The steep rise of the ITS curves observed in those plots for high lag time values is not uncommon or surprising. The reason is that, when increasing the lag time to values near

the length the individual trajectories (40 ns, in this case), we will find very few transition counts in the transition matrix due to the lack of sampling. Therefore, the statistics become inaccurate and thus we obtain mostly noise in the plot. This means that this region of the plot is not very relevant, and the ITSs obtained here are well converged. The ITS plots also show a separation between 2 to 4 slow implied timescales and the fast ones. Here we chose to group all the microstates into 4 macrostates. The Chapman-Kolmogorov test showed that our systems are Markovian for the chosen lag time of 25 ns (**Supplementary Figure S11**). In all cases, the binding process could be best described by 4 Markov states: a fully bound state having the lowest distances between the reacting carbon atom of the ligand and the nucleophile carbonyl group; 2 intermediate states where the ligand's linker chain is located in the tunnel or near the tunnel mouth; and finally, a fully unbound state with the ligand located on the protein's surface or in the bulk solvent (**Supplementary Figure S12**). The spatial distribution of the MSM states was quite similar for all the systems, with the same high-density regions of the ligand at the protein surface. The main differences corresponded to the fully bound states, for which the aromatic part of ligand **1E** was more constrained and partially inserted in the tunnel, while the conjugated system of **TMR** was fully outside (**Supplementary Figure S13**). Besides the number of slow implied time scales (Supplementary Figure S7), a reason for selecting 4 macrostates and not fewer was to have a good conformational resolution for further analysis. Since we will use the most bound state to equilibrium populations and energies, we need at least one state where the reactive end of the probe is well inserted in the catalytic site and not too far from the nucleophile. We found that this condition could not be satisfied by specifying less than 4 macrostates in our model, otherwise the state with closest distances would have the probe already too disperse over the active site and the tunnel.

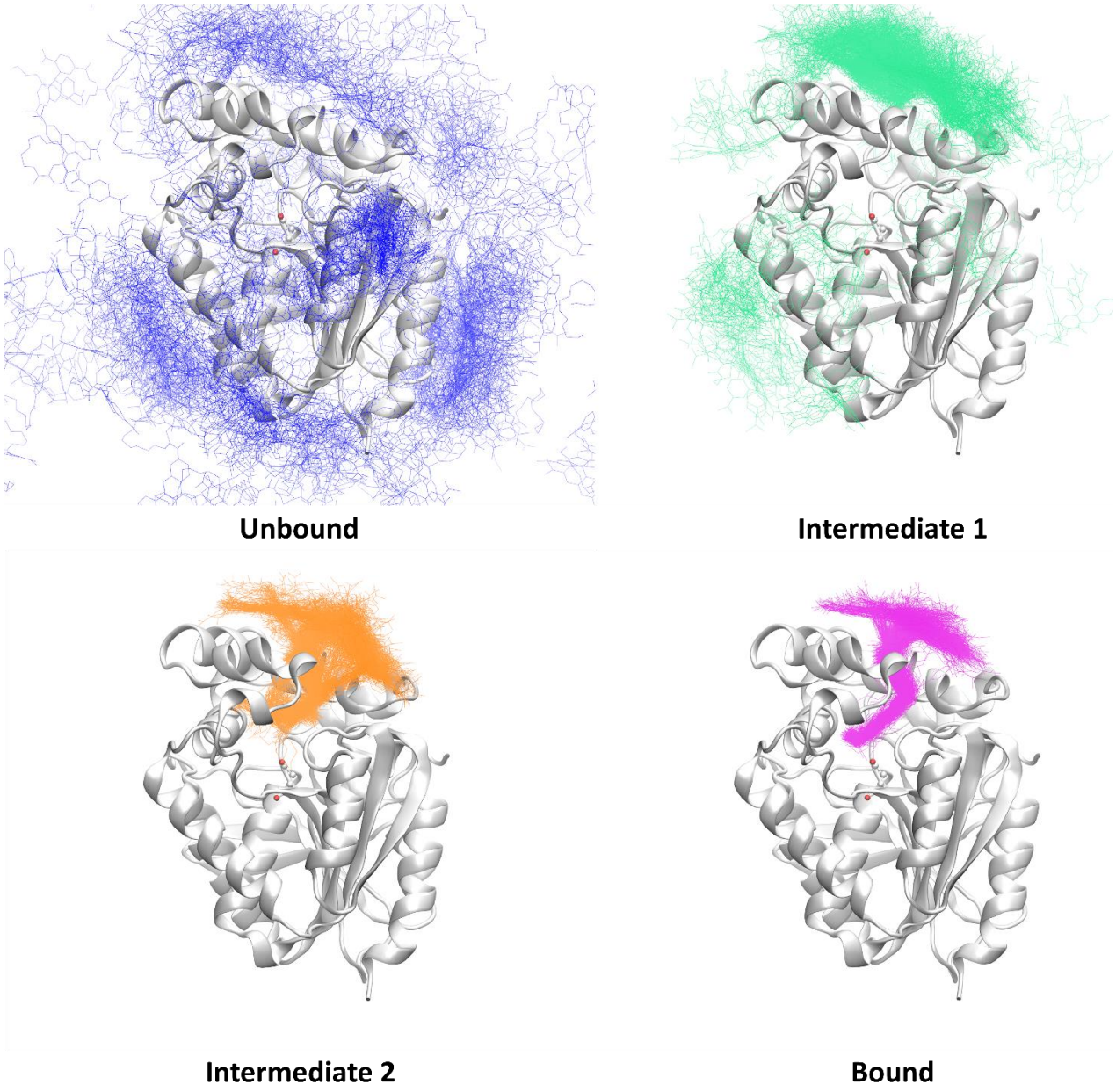
The kinetic parameters were calculated from the 4-state MSM, and the binding was described by the transition from the most unbound state to the fully bound state (**Supplementary Figure S12**). The results showed that the calculated binding rates (k_1) follow the same order as the experimental k_1 (**Supplementary Tables S3** and **S7**). The discrepancy is discussed in the main text. Interestingly, the estimated dissociation constants (K_d) showed much higher affinity for DhaAHT with **TMR** ($5.7 \pm 1.9 \times 10^{-4}$ M) than with **1E** ($5.9 \pm 2.5 \times 10^{-3}$ M), which corresponds well with the experimental binding rates (**Supplementary Tables S3** and **S7**).



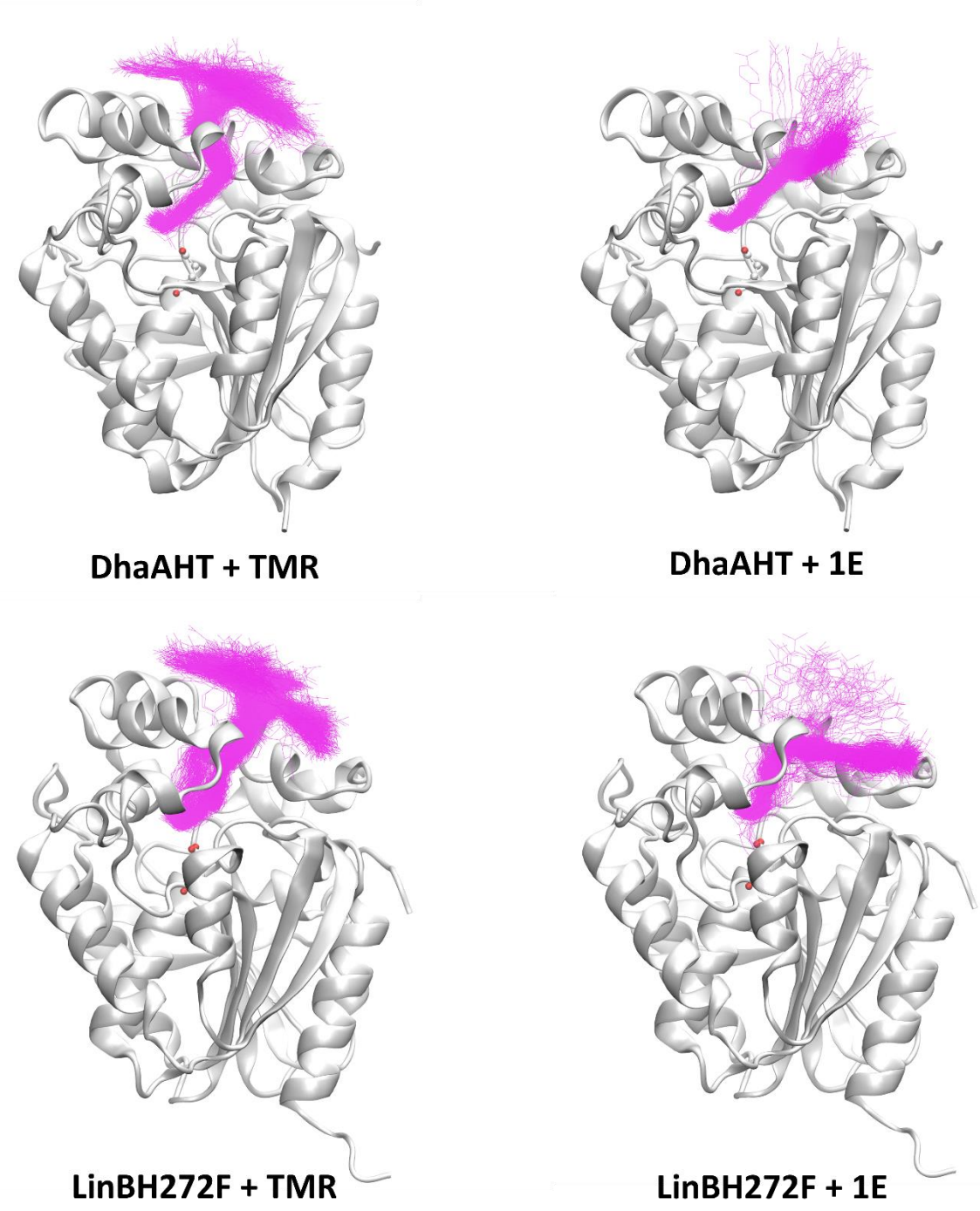
Supplementary Figure S10. Implied time scales plots versus the lag times, for the binding of the ligands with the enzymes, for all the studied systems. The plots show the convergence of the implied time scales (a parameter in the transition matrix in the Markov theory) of the slowest processes for lag times of 25 ns or higher. The implied time-scales plot is used to estimate whether the simulations display Markovian behavior (the convergence of the transitions), the suitable lag time for the calculations, and estimate the minimum number of well-differentiated transitions[21].



Supplementary Figure S11. Chapman-Kolmogorov tests for the 4 Markov states obtained from the binding simulations, for all the simulated systems. The perfect superimposition of the "estimate" with "predict" transition probabilities shows that the models are Markovian for the selected lag times. The Chapman-Kolmogorov test is used to assess the accuracy of the generated MSM. Basically, it predicts the transitions probability based on the model (*predict*, dashed lines) and compares it with direct estimates from the simulation data (*estimate*, full lines). A well-predictive model will have the two lines overlapping in the plot[21], which is our case.



Supplementary Figure S12. Individual Markov states describing the binding of the fluorescent ligand TMR to DhaAHT. The enzyme is represented by the white cartoons, the catalytic D106 as ball-and-sticks, and the **TMR** as the superimposed lines of different colors: blue – fully unbound state; green – intermediate state with the ligand at the tunnel mouth and other shallow cavities; orange – intermediate state with linker partially inserted in the access tunnel; red – fully bound state with the linker reaching the active site.



Supplementary Figure S13. Representative structures of the bound states of the fluorescent ligands TMR and 1E to DhaAHT and LinBH272F. The enzymes are represented by the white cartoons, the catalytic D106 (DhaAHT) or D108 (LinBH272F) as ball-and-sticks, and the ligands as the superimposed magenta lines.

Supplementary Table S7. Kinetic rates obtained from the adaptive MD simulations of binding of the ligands with the HLDs, the calculated free energy of activation for the S_N2 reaction, and respective experimental values.

Parameter[a]	DhaAHT + TMR	LinBH272F + TMR	DhaAHT + 1E	LinBH272F + 1E
Theoretical[b]				
τ_1 (ns)	$8.8 \pm 1.3 \times 10^2$	$3.6 \pm 1.1 \times 10^2$	$2.57 \pm 0.85 \times 10^3$	$2.20 \pm 0.44 \times 10^3$
τ_{-1} (ns)	$6.0 \pm 1.5 \times 10^3$	$1.07 \pm 0.15 \times 10^3$	$1.44 \pm 0.29 \times 10^3$	$2.88 \pm 0.24 \times 10^2$
k_1 ($M^{-1} \cdot s^{-1}$)	$2.99 \pm 0.45 \times 10^8$	$7.6 \pm 2.5 \times 10^7$	$1.09 \pm 0.23 \times 10^8$	$1.22 \pm 0.21 \times 10^8$
k_{-1} (s^{-1})	$1.71 \pm 0.38 \times 10^5$	$9.7 \pm 1.8 \times 10^5$	$7.3 \pm 1.2 \times 10^5$	$3.56 \pm 0.28 \times 10^6$
K_D (M)	$5.7 \pm 1.9 \times 10^{-4}$	$2.8 \pm 1.4 \times 10^{-2}$	$5.9 \pm 2.5 \times 10^{-3}$	$3.95 \pm 0.91 \times 10^{-2}$
ΔG^0_{eq} (kcal·mol^{-1})	-4.49 ± 0.20	-2.20 ± 0.27	-3.10 ± 0.22	-1.94 ± 0.13
P_{bound}	0.260 ± 0.043	0.068 ± 0.026	0.054 ± 0.012	0.047 ± 0.008
[*NAC*] (per 20 µs)	13	99	107	183
K_{NAC}	2.53×10^{-4}	7.37×10^{-3}	1.00×10^{-2}	1.96×10^{-2}
$\Delta G^{\ddagger}_{SN2}$ (kcal·mol^{-1})	15.5 ± 1.3	12.1 ± 1.9	15.3 ± 1.7	13.8 ± 1.8
$\Delta G^{\ddagger}_2$ (kcal·mol^{-1})	20.6	15.2	18.1	16.2
Experimental[c]				
k_1 ($M^{-1} \cdot s^{-1}$)	$6.62 \pm 0.10 \times 10^5$	21.7 ± 1.7	$2.50 \pm 0.50 \times 10^4$	$5.0 \pm 1.0 \times 10^4$
k_{-1} (s^{-1})	$8.50 \pm 0.50 \times 10^{-4}$	$5.0 \pm 1.7 \times 10^{-5}$	$8.3 \pm 3.3 \times 10^{-3}$	$6.7 \pm 5.0 \times 10^{-3}$
K_D (M)	$1.28 \pm 0.08 \times 10^{-9}$	$2.34 \pm 0.8 \times 10^{-6}$	$0.3 \pm 0.1 \times 10^{-6}$	$0.13 \pm 0.10 \times 10^{-6}$
k_2 (s^{-1})	66 ± 50	$3.50 \pm 0.33 \times 10^2$	$4.17 \pm 0.68 \times 10^3$	$6.7 \pm 1.7 \times 10^3$
ΔG^0_{eq} (kcal·mol^{-1})	-4.0	0.5	-0.7	-1.2
$\Delta G^{\ddagger}_2$ (kcal·mol^{-1})	23.5	22.5	21.1	20.8

[a] τ_1, mean transition association time; τ_{-1}, mean transition dissociation time; k_1, association rate, often also called k_{on}; k_{-1}, dissociation rate, often called k_{off}; K_D, equilibrium dissociation constant; ΔG^0_{eq}, free energy difference between bound and unbound states; P_{bound}, probability of the bound state, equivalent to the respective fraction of population in equilibrium obtained from the MSM analysis; [*NAC*], the total number of snapshots in pre-reactive conformation, also known as near-attack conformation, per 20 µs of simulation; K_{NAC}, equilibrium constant for reaching a *NAC* from the bound state; $\Delta G^{\ddagger}_{SN2}$, free energy of activation for the S_N2 chemical reaction; k_2, kinetic rate of step 2, which converts the bound state into the covalent complex; $\Delta G^{\ddagger}_2$, free energy of activation of step 2. [b]The errors correspond to the standard deviations, either obtained from a bootstrap analysis of 100 sub-samples extracted from the original MD ensembles (for parameters obtained from MSM analysis), or from the number of pre-reactive complexes analyzed (for $\Delta G^{\ddagger}$ values). [c]These values were obtained from Supplementary Tables S3-S6 and converted to the same standard units used for the theoretical parameters; the errors correspond to the standard errors derived from the nonlinear regression.

Chemical step

The adaptive MD simulations were analyzed to detect the formation of pre-reactive complexes that may lead to the S_N2 reaction. To identify them, geometric conditions were applied as previously defined [40]. These regarded the distance between the reacting atoms (the nearest carboxyl oxygen atom and the ligand's carbon atom adjacent to the chlorine), the O-C-Cl angle, and the formation of a hydrogen bond between the

Cl-atom and the residues W107 or N41. In total, a number of pre-reactive complexes (also called "near-attack conformations", *NAC*) were found for the studied systems (**Supplementary Table S7**). The energy barrier for the S_N2 reaction between the fluorescent ligands and the D106 nucleophile (**Figures 3C–D** in the main text) was calculated by QM/MM adiabatic mapping (**Figure 4C** in the main text), and few outlier results were discarded in both systems.

CaverDock simulations

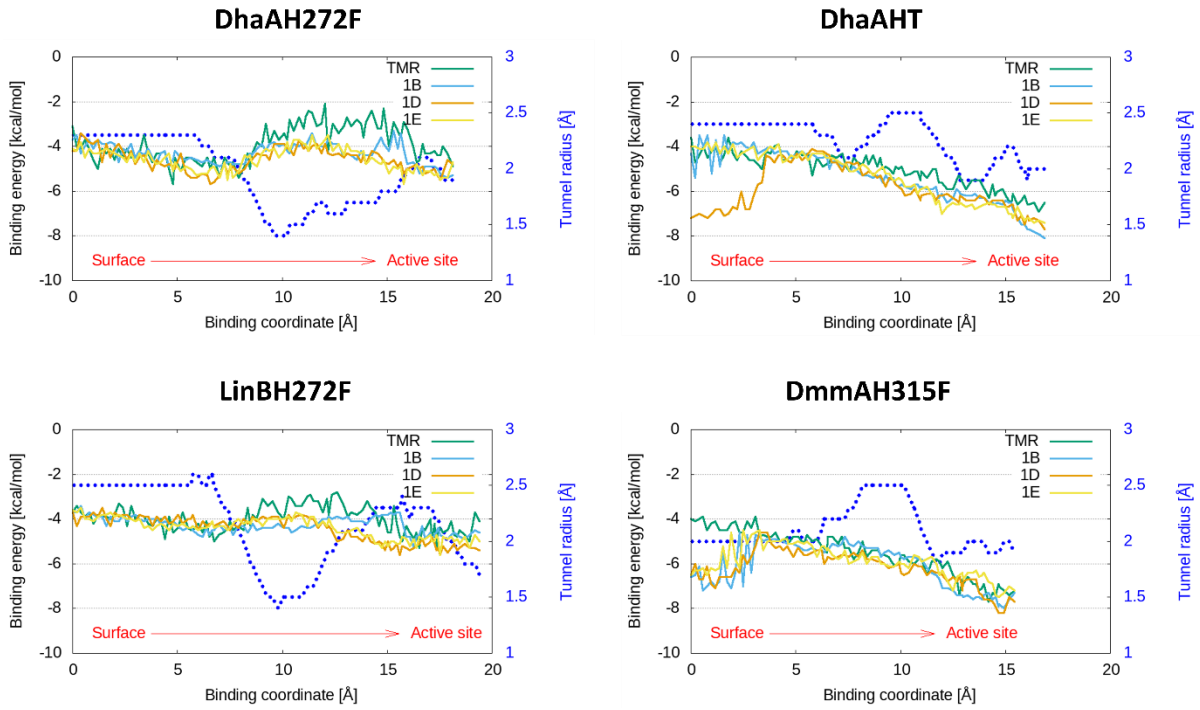
To test a much simpler and faster computational approach, we also used CaverDock[37,38] binding of all the fluorescent ligands to the proteins studied here to predict the thermodynamics of binding of all the ligands (**TMR**, **1B**, **1D**, and **1E**) to the different studied proteins studied here (DhaAH272F, DhaAHT, LinBH272F, and DmmAH315F). CaverDock is a recent method based on molecular docking, developed in our group, which has proven useful for quickly assessing thermodynamic and kinetic information on the (un)binding of ligands through molecular tunnels[38,41,42]. Its features can make it valuable for different practical purposes involving proteins containing deep or buried binding pockets, such as: 1) virtual screening of new drugs, 2) predicting kinetic barriers for binding/unbinding of substrates or products, or 3) assess the impact of protein mutagenesis on (un)binding kinetics. In principle, one can use CaverDock to estimate (un)binding energy barriers ($\Delta E_{barrier}$), which are related to the respective kinetic constants (k_1 or k_{-1}), and the energy differences between the bound state and unbound state (ΔE_{bs}), which is related to the binding affinity (K_D).

The results obtained for the studied systems (**Supplementary Table S8** and **Figure S14**) reveal clear differences in the energy profiles which are directly related to the tunnel geometry: the proteins with narrower tunnels (DhaAH272F and LinBH272F) have clear energy barriers, located at the respective tunnel bottlenecks (where the tunnel has its lowest radius), while for the proteins with wider tunnels (DhaAHT and DmmAH315F) the barrier is almost inexistent or less pronounced. Moreover, the energy of the fully bound state is always more negative than at the beginning of the trajectory, indicating the thermodynamic propensity of all these ligands for binding into the active sites of all these proteins. However, when we compare the energy barriers and the energies of the bound states for the different ligands and proteins, we cannot find a correlation with the experimental bind rates (k_1) that might help us predict, in such a quick and inexpensive manner, which ligands could be more suitable for each protein. We can conclude that this task is beyond the current limitations of CaverDock. This is most likely due to the fact that currently CaverDock does not take into account the protein flexibility and dynamic ensembles, which is necessarily important when accommodating large ligands such as the ligands investigated here. We are presently working on the next version that will incorporate protein flexibility in smart ways that will improve the accuracy and still avoid lengthy simulation times.

Supplementary Table S8. Energy parameters predicted by CaverDock for the binding of **TMR**, **1B**, **1D** and **1E** to DhaAH272F, DhaAHT, LinBH272F, and DmmAH315F along the p1 tunnel.[a]

	DhaAH272F + TMR	**DhaAH272F + 1B**	**DhaAH272F + 1D**	**DhaAH272F + 1E**
E_{bound}	-5.6	-5.4	-5.3	-5.6
$\Delta E_{barrier}$	3.4	1.4	1.6	1.7
ΔE_{bs}	0.0	-0.3	0.5	-0.2
	DhaAHT+ TMR	**DhaaHT + 1B**	**DhaAHT+ 1D**	**DhaAHT+ 1E**
E_{bound}	-6.9	-7.9	-7.5	-7.3
$\Delta E_{barrier}$	0.7	1.5	3.5	0.5
ΔE_{bs}	-1.9	-2.1	-0.1	-2.4
	LinBH272F + TMR	**LinBH272F + 1B**	**LinBH272F + 1D**	**LinBH272F + 1E**
E_{bound}	-6.0	-5.3	-5.7	-5.4
$\Delta E_{barrier}$	2.3	0.7	0.8	1.0
ΔE_{bs}	-1.0	-0.7	-1.1	-0.6
	DmmAH315F + TMR	**DmmAH315F + 1B**	**DmmAH315F + 1D**	**DmmAH315F + 1E**
E_{bound}	-7.7	-8.0	-8.2	-7.5
$\Delta E_{barrier}$	0.5	2.4	2.0	1.6
ΔE_{bs}	-2.6	-0.8	-1.1	-1.0

[a] Binding energy parameters (in kcal·mol^{-1}): E_{bound}, energy of the bound state, which is the lowest energy in the active site; $\Delta E_{barrier}$, activation energy of binding, is the difference between the energy maximum (at the tunnel bottleneck) and the minimum before the barrier (closer to the surface); ΔE_{bs}, energy difference between the bound state (in the active site) and the unbound state (at the tunnel mouth).



Supplementary Figure S14. Energy profiles predicted by CaverDock for the binding of TMR, 1B, 1D and 1E to DhaAH272F, DhaAHT, LinBH272F, and DmmAH315F along the p1 tunnel. The binding energy of the ligands (respective axis on the left) are represented by the solid lines and the tunnel radius along its length (respective axis on the right) is represented by the dotted blue line. The ligand was pulled by the reacting carbon atom from the surface (0 Å of the x-axis) to the active site; the tunnel was extended

by 5 Å towards the solvent, so first 5 Å of the trajectory correspond to the ligand located in the solvent/tunnel mouth.

References

(1) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*; Walker, J. M., Ed.; Springer Protocols Handbooks; Humana Press: Totowa, NJ, 2005; pp 571–607. https://doi.org/10.1385/1-59259-890-0:571.

(2) Kaushik, S.; Prokop, Z.; Damborsky, J.; Chaloupkova, R. Kinetics of Binding of Fluorescent Ligands to Enzymes with Engineered Access Tunnels. *FEBS J.* **2017**, *284*, 134–148. https://doi.org/10.1111/febs.13957.

(3) Encell, L. P.; Ohana, R. F.; Zimmerman, K.; Otto, P.; Wood, M. G.; Los, G. V; Mcdougall, M. G.; Zimprich, C.; Karassina, N.; Learish, R. D.; Hurst, R.; Hartnett, J.; Wheeler, S.; Stecha, P.; English, J.; Zhao, K.; Mendez, J.; Benink, H. A.; Murphy, N.; Daniels, D. L.; Slater, M. R.; Urh, M.; Darzins, A.; Klaubert, D. H.; Bulleit, R. F.; Wood, K. V. Development of a Dehalogenase-Based Protein Fusion Tag Capable of Rapid , Selective and Covalent Attachment to Customizable Ligands. *Curr. Chem. Genomics* **2012**, *6*, 55–71.

(4) Johnson, K. A.; Simpson, Z. B.; Blom, T. Global Kinetic Explorer: A New Computer Program for Dynamic Simulation and Fitting of Kinetic Data. *Anal. Biochem.* **2009**, *387* (1), 20–29. https://doi.org/10.1016/j.ab.2008.12.024.

(5) Johnson, K. A.; Simpson, Z. B.; Blom, T. FitSpace Explorer: An Algorithm to Evaluate Multidimensional Parameter Space in Fitting Kinetic Data. *Anal. Biochem.* **2009**, *387* (1), 30–41. https://doi.org/10.1016/j.ab.2008.12.025.

(6) Los, G. V.; Encell, L. P.; McDougall, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; Simpson, D.; Mendez, J.; Zimmerman, K.; Otto, P.; Vidugiris, G.; Zhu, J.; Darzins, A.; Klaubert, D. H.; Bulleit, R. F.; Wood, K. V. HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *ACS Chem. Biol.* **2008**, *3* (6), 373–382. https://doi.org/10.1021/cb800025k.

(7) Johnson, K. A. Fitting Enzyme Kinetic Data with KinTek Global Kinetic Explorer. *Methods Enzymol.* **2009**, *467*, 601–626. https://doi.org/10.1016/S0076-6879(09)67023-3.

(8) Li, A.; Ziehr, J. L.; Johnson, K. A. A New General Method for Simultaneous Fitting of Temperature and Concentration Dependence of Reaction Rates Yields Kinetic and Thermodynamic Parameters for HIV Reverse Transcriptase Specificity. *J. Biol. Chem.* **2017**, *292* (16), 6695–6702. https://doi.org/10.1074/jbc.M116.760827.

(9) Kaushik, S.; Prokop, Z.; Damborsky, J.; Chaloupkova, R. Kinetics of Binding of Fluorescent Ligands to Enzymes with Engineered Access Tunnels. *FEBS J.* **2017**, *284* (1), 134–148. https://doi.org/10.1111/febs.13957.

(10) Rose, P. W.; Bi, C.; Bluhm, W. F.; Christie, C. H.; Dimitropoulos, D.; Dutta, S.; Green, R. K.; Goodsell, D. S.; Prlić, A.; Quesada, M.; Quinn, G. B.; Ramos, A. G.; Westbrook, J. D.; Young, J.; Zardecki, C.; Berman, H. M.; Bourne, P. E. The RCSB Protein Data Bank: New Resources for Research and Education. *Nucleic Acids Res.* **2013**, *41* (D1), D475–D482. https://doi.org/10.1093/nar/gks1200.

(11) Kellogg, E. H.; Leaver-Fay, A.; Baker, D. Role of Conformational Sampling in Computing Mutation-Induced Changes in Protein Structure and Stability. *Proteins* **2011**, *79* (3), 830–838. https://doi.org/10.1002/prot.22921.

(12) Song, Y.; Tyka, M.; Leaver-Fay, A.; Thompson, J.; Baker, D. Structure-Guided Forcefield Optimization. *Proteins* **2011**, *79* (6), 1898–1909. https://doi.org/10.1002/prot.23013.

(13) O'Meara, M. J.; Leaver-Fay, A.; Tyka, M. D.; Stein, A.; Houlihan, K.; DiMaio, F.; Bradley, P.; Kortemme, T.; Baker, D.; Snoeyink, J.; Kuhlman, B. Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. *J. Chem. Theory Comput.* **2015**, *11* (2), 609–622. https://doi.org/10.1021/ct500864r.

(14) Chovancová, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **2012**, *8* (10), e1002708. https://doi.org/10.1371/journal.pcbi.1002708.

(15) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminformatics* **2012**, *4* (1), 17. https://doi.org/10.1186/1758-2946-4-17.

(16) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114* (25), 10024–10035. https://doi.org/10.1021/ja00051a040.

(17) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision E.01*; Gaussian, Inc.: Wallingford, CT, 2009.

(18) Marques, S. M.; Dunajova, Z.; Prokop, Z.; Chaloupkova, R.; Brezovsky, J.; Damborsky, J. Catalytic Cycle of Haloalkane Dehalogenases Toward Unnatural Substrates Explored by Computational Modeling. *J. Chem. Inf. Model.* **2017**, *57* (8), 1970–1989. https://doi.org/10.1021/acs.jcim.7b00070.

(19) Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, III, T. E.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; Kollman, P. A. *AMBER 16*; University of California: San Francisco, 2016.

(20) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A Server for Estimating PKas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33* (suppl 2), W368–W371. https://doi.org/10.1093/nar/gki464.

(21) Doerr, S.; Harvey, M. J.; Noé, F.; De Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J. Chem. Theory Comput.* **2016**, *12* (4), 1845–1852. https://doi.org/10.1021/acs.jctc.6b00049.

(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(23) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.

(24) Feenstra K. Anton; Hess Berk; Berendsen Herman J. C. Improving Efficiency of Large Time-scale Molecular Dynamics Simulations of Hydrogen-rich Systems. *J. Comput. Chem.* **1999**, *20* (8), 786–798. https://doi.org/10.1002/(SICI)1096-987X(199906)20:8<786::AID-JCC5>3.0.CO;2-B.

(25) Harvey, M. J.; De Fabritiis, G. An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. *J. Chem. Theory Comput.* **2009**, *5* (9), 2371–2377. https://doi.org/10.1021/ct900275y.
(26) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5* (6), 1632–1639. https://doi.org/10.1021/ct9000685.
(27) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11* (4), 1864–1874. https://doi.org/10.1021/ct5010406.
(28) Naritomi, Y.; Fuchigami, S. Slow Dynamics in Protein Fluctuations Revealed by Time-Structure Based Independent Component Analysis: The Case of Domain Motions. *J. Chem. Phys.* **2011**, *134* (6), 065101. https://doi.org/10.1063/1.3554380.
(29) Doerr, S.; Harvey, M. J.; Noé, F.; De Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J. Chem. Theory Comput.* **2016**, *12* (4), 1845–1852. https://doi.org/10.1021/acs.jctc.6b00049.
(30) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.
(31) Hur, S.; Kahn, K.; Bruice, T. C. Comparison of Formation of Reactive Conformers for the S_N2 Displacements by CH_3CO^- in Water and by Asp124-CO^- in a Haloalkane Dehalogenase. *Proc. Natl. Acad. Sci.* **2003**, *100* (5), 2215–2219. https://doi.org/10.1073/pnas.242721799.
(32) Ranaghan, K. E.; Mulholland, A. J. Investigations of Enzyme-Catalysed Reactions with Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods. *Int. Rev. Phys. Chem.* **2010**, *29* (1), 65–133. https://doi.org/10.1080/01442350903495417.
(33) Lonsdale, R.; N. Harvey, J.; J. Mulholland, A. A Practical Guide to Modelling Enzyme-Catalysed Reactions. *Chem. Soc. Rev.* **2012**, *41* (8), 3025–3038. https://doi.org/10.1039/C2CS15297E.
(34) Walker, R. C.; Crowley, M. F.; Case, D. A. The Implementation of a Fast and Accurate QM/MM Potential Method in Amber. *J. Comput. Chem.* **2008**, *29* (7), 1019–1031. https://doi.org/10.1002/jcc.20857.
(35) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13* (12), 1173–1213. https://doi.org/10.1007/s00894-007-0233-4.
(36) Zhu, C.; Byrd, R. H.; Lu, P.; Nocedal, J. Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization. *ACM Trans Math Softw* **1997**, *23* (4), 550–560. https://doi.org/10.1145/279232.279236.
(37) Filipovič, J.; Vávra, O.; Plhák, J.; Bednář, D.; Marques, S. M.; Brezovský, J.; Matyska, L.; Damborský, J. CaverDock: A Novel Method for the Fast Analysis of Ligand Transport. **2018**.
(38) Vavra, O.; Filipovic, J.; Plhak, J.; Bednar, D.; Marques, S. M.; Brezovsky, J.; Stourac, J.; Matyska, L.; Damborsky, J. CaverDock: A Molecular Docking-Based Tool to Analyse Ligand Transport through Protein Tunnels and Channels. *Bioinformatics* **2019**, *35* (23), 4986–4993. https://doi.org/10.1093/bioinformatics/btz386.
(39) Mchedlov-Petrossyan, N. O.; Cheipesh, T. A.; Shekhovtsov, S. V.; Redko, A. N.; Rybachenko, V. I.; Omelchenko, I. V.; Shishkin, O. V. Ionization and Tautomerism of Methyl Fluorescein and Related Dyes. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **2015**, *150*, 151–161. https://doi.org/10.1016/j.saa.2015.05.037.
(40) Hur, S.; Kahn, K.; Bruice, T. C. Comparison of Formation of Reactive Conformers for the SN2 Displacements by CH3CO2- in Water and by Asp124-CO2- in a Haloalkane Dehalogenase. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (5), 2215–2219. https://doi.org/10.1073/pnas.242721799.
(41) Pinto, G. P.; Vavra, O.; Filipovic, J.; Stourac, J.; Bednar, D.; Damborsky, J. Fast Screening of Inhibitor Binding/Unbinding Using Novel Software Tool CaverDock. *Front. Chem.* **2019**, *7*, 709. https://doi.org/10.3389/fchem.2019.00709.

(42) Pinto, G. P.; Vavra, O.; Marques, S. M.; Filipovic, J.; Bednar, D.; Damborsky, J. Screening of World Approved Drugs against Highly Dynamical Spike Glycoprotein of SARS-CoV-2 Using CaverDock and Machine Learning. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 3187–3197. https://doi.org/10.1016/j.csbj.2021.05.043.