Supporting Information for

Controlling Protein Crystallization by Free Energy Guided Design of Interactions at Crystal Contacts

Johannes Hermann[a], Daniel Bischoff[a], Phillip Grob[a], Robert Janowski[b], Dariusch Hekmat[a], Dierk Niessing[b,c], Martin Zacharias[d], Dirk Weuster-Botz*[a]

[a] J. Hermann, D. Bischoff, P. Grob, Dr. D. Hekmat, Prof. Dr.-Ing. D. Weuster-Botz
Technical University of Munich, Institute of Biochemical Engineering,
Boltzmannstraße 15, 85748 Garching, Germany, *E-Mail: johannes.hermann@tum.de

[b] Dr. R. Janowski, Prof. Dr. D. Niessing
Helmholtz Zentrum München, Institute of Structural Biology,
Ingolstädter Landstraße 1,85764 Neuherberg, Germany

[c] Prof. Dr. D. Niessing
Ulm University, Institute of Pharmaceutical Biotechnology,
James Franck Ring N27, 89081 Ulm, Germany

[d] Prof. Dr. M. Zacharias
Technical University of Munich, Physics Department - T38,
James-Franck-Str. 1, 85748 Garching, Germany

Contents

S1 Primer Design

Table S1: Primer design (forward and reverse) for *Lb*ADH mutants Q207D, D54A, K45A, and H39A.

mutant	forward primer	reverse primer
Q207D	GATGTCAGATCGGACCAAGACGCCAATG	CTTGGTCCGATCTGACATCGCTTCTTCG
D54A	CACTCCTGCTCAGATTCAATTTTTCC	GAATCTGAGCAGGAGTGCCGACACTC
K45A	GCAGCTGCGAGTGTCGGCACTC	AGCCACTCGCAGCTGCTTTTTC
H39A	GGCCGGGCCAGCGATGTTGGTGAAAAAG	CAACATCGCTGGCCCGGCCGGTAATCATG

S2 Enzymatic Activity

Enzymatic activity assays were performed for non-crystallizable mutants (K45A, D54A, H39A) as described previously [1]. For spectrophotometric detection of the NADPH oxidation to $NADP^+$ during reduction of acetophenone to 1-phenylethanol in microtiter plates (Multiskan FC Microplate Photometer, Thermo Fisher Scientific, Darmstadt, Germany), 20 µL of *Lb*ADH solution (6 mg L^{-1} *Lb*ADH in the protein buffer) was added to 180 µL of protein buffer containing 10 mM acetophenone and 0.5 mM NADPH at pH 7.0 and at 25 °C. The absorption was measured at 340 nm (absorption maximum of NADPH) at 6 s intervals for 10 min. The resulting maximum rate of the enzymatic reaction was compared to maximum enzymatic rate of wild type *Lb*ADH $V_{max} = 24.9 \pm 3.0$ U mg^{-1}. Figure S1 reports the relative maximum enzymatic activity.

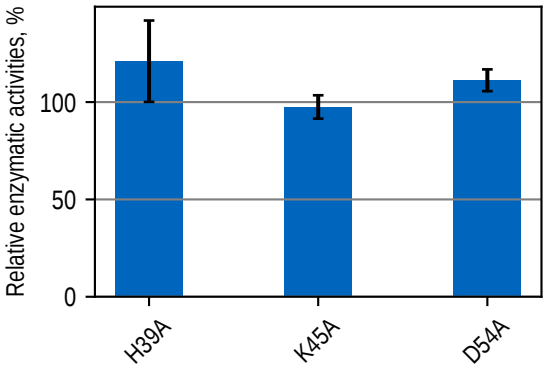

Figure S1: Relative enzymatic activity of *Lb*ADH mutants H39A, K45A, and D54A compared to the maximum enzymatic activity of wild type *Lb*ADH $V_{max} = 24.9 \pm 3.0$ U mg^{-1}. Error bars correspond to standard deviations of 3 biological replicates (each triplicate measurements).

S3 Method Consistency Evaluation

In order to evaluate robustness and consistency of the method and simulation system, simulations were set-up which should yield the same free energy change. The crystal bound state as fully periodically interacting UC setup (bound version (a) described above) of mutant Q126H was employed. After equilibration we prolonged the simulation to extract three independent configurations as starting structure from the system phase space. In this setup, the free energy change in each simulation is calculated for eight mutations in a UC at once. The values were in close agreement with a low average standard deviation of 0.58 kJ mol^{-1} (see Table S2) and were within the calculated error. The low standard deviation demonstrates the robustness of method and setup.

S4 System Reduction Procedure

Direct translation of the bound crystal state into our model system results in the simulation of a crystallographic UC with PBC. In order to be able to use the described charge correction scheme, the simulation system needs to exhibit enough solvent volume for solvation of an alchemical counter ion. In case of the bound crystal state this requirement is not fulfilled for bound version (a) described above (a fully periodically interacting UC setup) but

Table S2: Three independent simulations of the bound UC setup with PBC for mutation Q126H in the crystal packing of Q126H. The starting frames were taken by prolonging the simulation after equilibration for 0 ns (simulation 1), 15 ns (simulation 2) and 35 ns (simulation 3). The values are for a whole UC simulation, i. e. eight mutations at once. All values in kJ mol^{-1}. The error of each individual simulation is calculated from block averaging.

simulation 1	simulation 2	simulation 3
637.66±0.61	637.94±0.84	638.77±0.48

fulfilled in bound version (b) (a reduced setup where only the interacting monomers of the respective contact are placed in a simulation box).

Furthermore, the unbound state biologically consists of a solvated tetramer. In order to save computer power, we inspected if the unbound solvated state of Q126H as tetramer (unbound version (a) described above) can be reduced to two monomers (unbound version (b)). For all set-ups we calculated the free energy change for wild type and mutant system to check if the system is reducible without loss of information. Furthermore, each free energy calculations was started from two different starting structures where one structure was equilibrated as wild type and and the other as mutant. The low standard deviation of all calculations (below 0.44 kJ mol^{-1}) showed that each system could be reduced enabling investigation of any mutation (see Table S3).

Table S3: Individual free energy change upon mutation ΔG for different systems (mutant crystal system, WT crystal system, and unbound state) and different setups (bound UC setup with PBC, reduced setup isolated crystal contact, solvated unbound tetramer, solvated unbound dimer). The values are calculated for mutation Q126H. For each system and setup two simulations were performed where one equilibration was performed with wild type parameters (EQ WT) and the other on with mutant parameters (EQ Mut). All values in kJ mol^{-1}.

		EQ WT	EQ Mut	mean(ΔG)	SD(ΔG)
Mutant Crystal	UC	79.77±0.05	79.67±0.13	79.85±0.18	0.17
	reduced	79.91±0.62	80.07±0.39		
WT crystal	UC	82.49 ±0.06	82.21±0.13	82.71±0.17	0.44
	reduced	83.15±0.36	83.01±0.58		
unbound	Tetramer	83.92 ±0.15	83.55±0.17	83.56±0.14	0.25
	reduced	83.40±0.27	83.37±0.42		

S5 Thermodynamic Cycle and Free Energy Difference Calculation

In order to quantify the effect of a mutation we calculated the energy difference between crystallizing WT and mutant protein. Therefore, we constructed a thermodynamic path from the unbound solvated crystal building blocks to the bound crystal state for WT and mutant (see Figure 3a). Each corner represents a thermodynamic state: The horizontal paths represent the crystallization process of WT and mutant with associated free energy differences $\Delta G^{\mathrm{WT}}_{\mathrm{crystallization}}$and $\Delta G^{\mathrm{Mut}}_{\mathrm{crystallization}}$. The free energy difference

$$\Delta\Delta G \equiv \Delta\Delta G^{\mathrm{WT}\,\rightarrow\,\mathrm{Mut}}_{\mathrm{crystallization}} = \Delta G^{\mathrm{Mut}}_{\mathrm{crystallization}} - \Delta G^{\mathrm{WT}}_{\mathrm{crystallization}} \tag{1}$$

defines which process has a lower free energy minimum, which process is more likely to happen, and which crystals are more stable. In the definition of Eq. (1) a negative $\Delta\Delta G$ indicates a thermodynamic favor of mutant crystallization over WT crystallization. However, the individual free energy $\Delta G^{\mathrm{WT}}_{\mathrm{crystallization}}$and $\Delta G^{\mathrm{Mut}}_{\mathrm{crystallization}}$are not accessible in simulations as this involves the lengthy non-equilibrium process of crystallization. In a closed thermodynamic cycle the individual free energy differences sum up to zero

$$\Delta G^{\mathrm{Mut}}_{\mathrm{crystallization}} - \Delta G^{\mathrm{WT}}_{\mathrm{crystallization}} - \Delta G^{\mathrm{WT}\,\rightarrow\,\mathrm{Mut}}_{\mathrm{crystal}} + \Delta G^{\mathrm{WT}\,\rightarrow\,\mathrm{Mut}}_{\mathrm{solvated}} = 0. \tag{2}$$

The free energy changes upon mutation in both state $\Delta G^{\mathrm{WT}\,\rightarrow\,\mathrm{Mut}}_{\mathrm{crystal}}$ and $\Delta G^{\mathrm{WT}\,\rightarrow\,\mathrm{Mut}}_{\mathrm{solvated}}$ (vertical paths in Figure 3a) can be calculated by alchemically transforming the residues (transforming the force field parameters of the residues in the MD simulations) and monitoring the free energy change $\Delta G^{\mathrm{WT}\,\rightarrow\,\mathrm{Mut}}_{\mathrm{crystal}}$and $\Delta G^{\mathrm{WT}\,\rightarrow\,\mathrm{Mut}}_{\mathrm{solvated}}$. Therefore the free energy difference Eq. (1) can be calculated as

$$\Delta\Delta G = \Delta G^{\mathrm{WT}\,\rightarrow\,\mathrm{Mut}}_{\mathrm{crystal}} - \Delta G^{\mathrm{WT}\,\rightarrow\,\mathrm{Mut}}_{\mathrm{solvated}}. \tag{3}$$

S6 Experimental Crystallization Behavior of Q207D

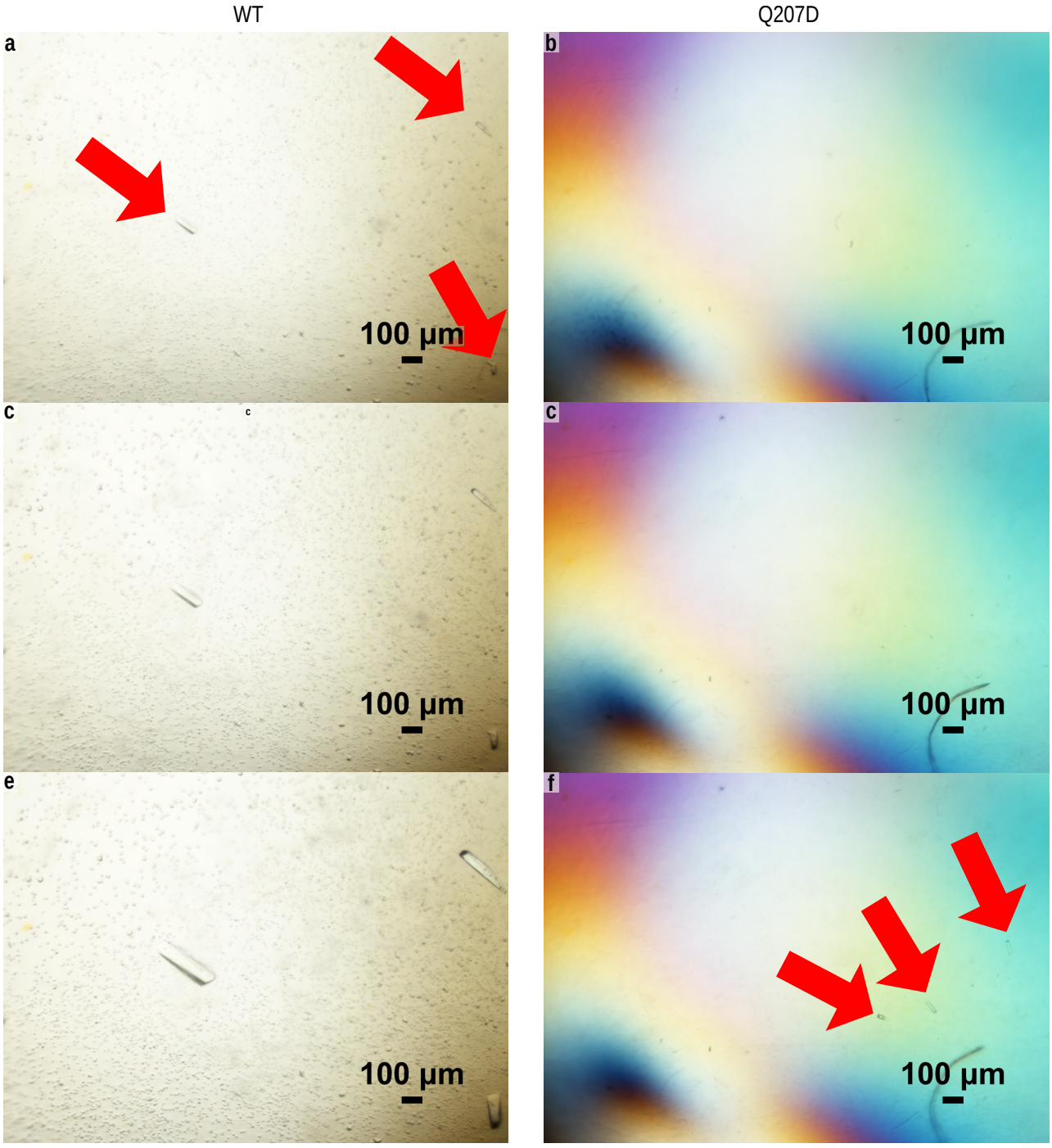

Figure S2: Crystallization experiments of *Lb*ADH wild type (WT) and Q207D at standard conditions (10 $g\,L^{-1}$ *Lb*ADH and 100 $g\,L^{-1}$ PEG 550 MME at 20 °C). Photographs were taken after 1 h (**a** and **b**), 2 h (**c** and **d**) and 5 h (**e** and **f**). *Lb*ADH WT started to crystallize after 1 h while no crystals could be observed for mutant Q207D. The onset of crystallization for mutant Q207D was observed after 5 h. Comparing to Figure 2a and b, which show the crystallization state after 24h, the lowered crystallization kinetics of variant Q207D compared to WT *Lb*ADH can be variefied. The photographs are representative for ten crystallization droplets for each variant.

S7 Investigated *Lb*ADH Variants

Table S4: Investigated protein variants of *Lb*ADH. Targeted crystal contact is given. For crystallizable mutants, space group (SG), UC parameters, and root mean square deviation (rmsd) of a single monomer after alignment on wild type *Lb*ADH C_α atoms are reported.

	targeted contact	SG	UC vectors, Å	Cα rmsd, Å
WT (6h07)	-	$P2_122_1$	56.03, 83.31, 114.38	-
K32A (6hlf)	edge contact	$I222$	55.58, 81.78, 114.89	0.15
D54F (6y1c)	edge contact	$P2_122_1$	55.76, 81.09, 113.13	0.16
Q126H (6y10)	corner contact	$I222$	56.05, 80.57, 113.47	0.19
Q126K (6y0z)	corner contact	$I222$	55.77, 84.23, 113.56	0.17
T102E (6y0s)	corner contact	$P2_122_1$	55.62, 81.15, 115.56	0.13
Q207D (7a2b)	side contact	$I222$	55.93, 80.70, 115.27	0.17
D54A	edge contact		*non-crystallizable mutant*	
K45A	corner contact		*non-crystallizable mutant*	
H39A	corner contact		*non-crystallizable mutant*	

S8 Detailed System Description

Table S5: Detailed simulation set-up description for unbound solvated state which is simulated as tetramer or reduced (as described in Computational details). Depending on the position of the mutation the tetramer can be reduced to one monomer or two monomers thereof. Mutation Q126H is simulated in both set-ups.

mutation	number of monomers	simulation box, $Å^3$
Q126H	4 (tetramer)	$88.00 \times 98.00 \times 93.00$
	2	$86.71 \times 107.88 \times 69.63$
Q126K	2	$93.88 \times 114.60 \times 76.44$
K32A	1	$90.56 \times 90.06 \times 82.02$
D54F	1	$90.56 \times 90.06 \times 82.02$
D54A	1	$90.56 \times 90.06 \times 82.02$
T102E	2	$93.72 \times 114.41 \times 76.30$
K45A	1	$90.56 \times 90.06 \times 82.02$
H39A	1	$90.56 \times 90.06 \times 82.02$
Q207D	4 (tetramer)	$88.19 \times 98.21 \times 93.20$

Table S6: Detailed simulation set-up description for all simulated crystal bound states for the reduced contact and UC with PBC. For each mutation and crystal system (cryst. sys.), symmetry operators of space group $P2_122_1$ (WT (6h07), D54F (6y1c), T102E (6y0s)) and $I222$ (K32A (6hlf), Q126H (hy10), Q126K (6y0z), Q207D (7a2b)) and unit cell translations to be applied on the asymmetric unit are displayed. The monomers which are extracted of the respective symmetry mate are given by indicating the chain-IDs (A, B) from the corresponding PBD-entry. The mutated monomers in the set-up are printed **bold**. Notice: in crystal system $P2_122_1$ two corner contacts (denoted by contact 1 and contact 2) exist.

mutation	cryst. sys.	symmetry operation and extracted monomers of asymmetric unit	simulation box size, Å^3
Q126H	WT: $P2_122_1$	UC with PBC: **A**+**B** of (x, y, z) + **A**+**B** of $(x + 1/2, -y, -z + 1/2)$ + **A**+**B** of $(-x, y, -z) + (1\ 0\ 0)\}$ + **A**+**B** of $(-x + 1/2, -y, z + 1/2) + (0\ 1\ 0)\}$	$56.03 \times 83.31 \times 114.38$
		contact 1: **A**+B of (x, y, z) & A of $(x + 1/2, -y, -z + 1/2)$	$114.01 \times 125.49 \times 100.88$
		contact 2: A+**B** of (x, y, z) & B of $\{(x + 1/2, -y, -z + 1/2) + (-1\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
	mutant: $I222$	UC with PBC: **A**+ of (x, y, z) + **A** of $(-x, -y, z)$ + **A** of $\{(x + 1/2, -y + 1/2, -z + 1/2)$ + **A** of $\{(-x + 1/2, y + 1/2, -z + 1/2)$ + **A** of $\{(-x + 1/2, -y + 1/2, z + 1/2)$ + **A** of $\{(x + 1/2, y + 1/2, z + 1/2)$	$56.05 \times 80.57 \times 113.47$
		contact: **A** of (x, y, z) + A of $(-x, -y, z)$ & A of $\{(x + 1/2, -y + 1/2, -z + 1/2) + (-1\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
Q126K	WT: $P2_122_1$	contact 1: **A**+B of (x, y, z) & A of $(x + 1/2, -y, -z + 1/2)$	$114.01 \times 125.49 \times 100.88$
		contact 2: **A**+B of (x, y, z) & B of $\{(x + 1/2, -y, -z + 1/2) + (-1\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
	mutant: $I222$	contact: **A** of (x, y, z) + A of $(-x, -y, z)$ & A & A of $\{(x + 1/2, -y + 1/2, -z + 1/2) + (-1\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
K32A	WT: $P2_122_1$	contact: **A** of (x, y, z) & **B** of $(x, y, z) + (0\ -1\ 0)$	$97.44 \times 142.46 \times 81.86$
	mutant: $I222$	contact: **A** of (x, y, z) & **A** of $(-x, -y, z) + (0\ 1\ 0)$	$97.44 \times 142.46 \times 81.86$
D54F	WT: $P2_122_1$	contact: **A** of (x, y, z) & **B** of $(x, y, z) + (0\ -1\ 0)$	$97.44 \times 142.46 \times 81.86$
	mutant: $P2_122_1$	contact: **A** of (x, y, z) & **B** of $(x, y, z) + (0\ 1\ 0)$	$97.44 \times 142.46 \times 81.86$
D54A	WT: $P2_122_1$	contact: **A** of (x, y, z) & **B** of $(x, y, z) + (0\ -1\ 0)$	$97.44 \times 142.46 \times 81.86$
T102E	WT: $P2_122_1$	contact 1: **A**+B of (x, y, z) & B of $(x + 1/2, -y, -z + 1/2) + (-1\ 1\ 0)$	$114.01 \times 125.49 \times 100.88$
		contact 2: A+**B** o (x, y, z) & A of $(x + 1/2, -y, -z + 1/2)$	$114.01 \times 125.49 \times 100.88$
	mutant: $P2_122_1$	contact 1: **A**+B of (x, y, z) & B of $(x + 1/2, -y, -z + 1/2)$	$114.01 \times 125.49 \times 100.88$
		contact 2: A+**B** of (x, y, z) & A of $(x + 1/2, -y, -z + 1/2) + (-1\ 1\ 0)$	$114.01 \times 125.49 \times 100.88$
K45A	WT: $P2_122_1$	contact 1: **A** of (x, y, z) & A+B of $(x + 1/2, -y, -z + 1/2) + (-1\ 0\ 0)$	$114.01 \times 125.49 \times 100.88$
		contact 2: **B** of (x, y, z) & A+B of $\{(x + 1/2, -y, -z + 1/2) + (0\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
H39A	WT: $P2_122_1$	contact 1: **A** of (x, y, z) & A+B of $(x + 1/2, -y, -z + 1/2) + (-1\ 0\ 0)$	$114.01 \times 125.49 \times 100.88$
		contact 2: **B** of (x, y, z) & A+B of $\{(x + 1/2, -y, -z + 1/2) + (0\ 1\ 0)\}$	$114.01 \times 125.49 \times 100.88$
Q207D	WT: $P2_122_1$	contact: **A**+B of (x, y, z) + A+**B** of $\{(-x, y, -z) + (1\ 0\ 0)\}$ & **A**+B of $(-x, y, -z)$ + A+**B** of $\{(x, y, z) + (-1\ 0\ 0)\}$	$141.29 \times 105.26 \times 98.09$
	mutant: $I222$	contact: **A** of (x, y, z) + **A** of $\{(x, -y, -z) + (0\ 1\ 0)\}$ + A of $\{(-x, y, -z) + (1\ 0\ 0)\}$ + A of $\{(-x, -y, z) + (1\ 1\ 0)\}$ & **A** of $(-x, y, -z)$ + **A** of $\{(-x, -y, z) + (0\ 1\ 0)\}$+ A of $\{(x, y, z) + (-1\ 0\ 0)\}$ + A of $\{(x, -y, -z) + (-1\ 1\ 0)\}$	$141.29 \times 105.26 \times 98.09$

S9 Crystal Structure and MD Simulation Analysis

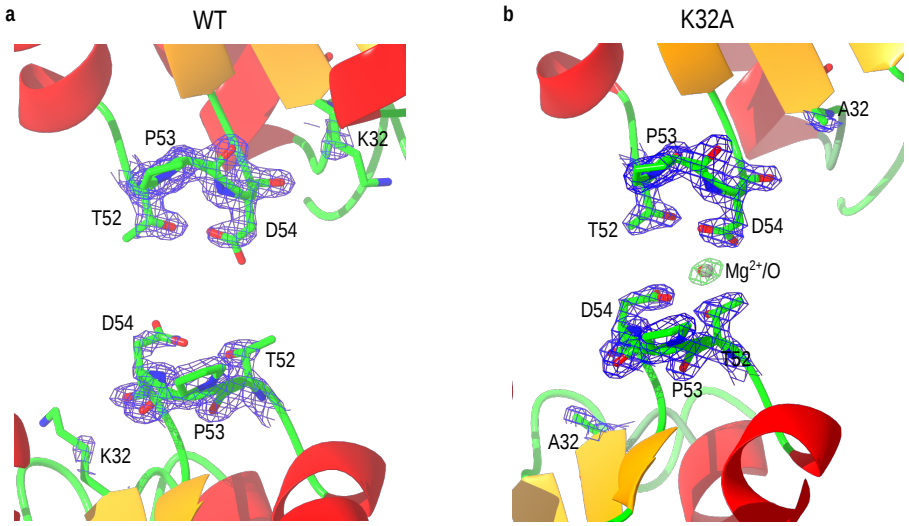

Figure S3: **Crystal structure analysis of edge contact of wild type and K32A crystals. a** X-ray crystal structure of the wild type *Lb*ADH with electron density map displayed in blue for selected residues. No additional electron density between the two interfacing aspartates (D54) was observed. Electron density maps were calculated with structure factor amplitude difference $2F_o - F_c$ and displayed at 1.0σ. **b** X-ray crystal structure of K32A mutant with electron density map displayed in blue for selected residues. Additionally an omit map for Mg^{2+}/O was calculated and displayed in green (simulated annealing omit map calculated with phenix [2]).

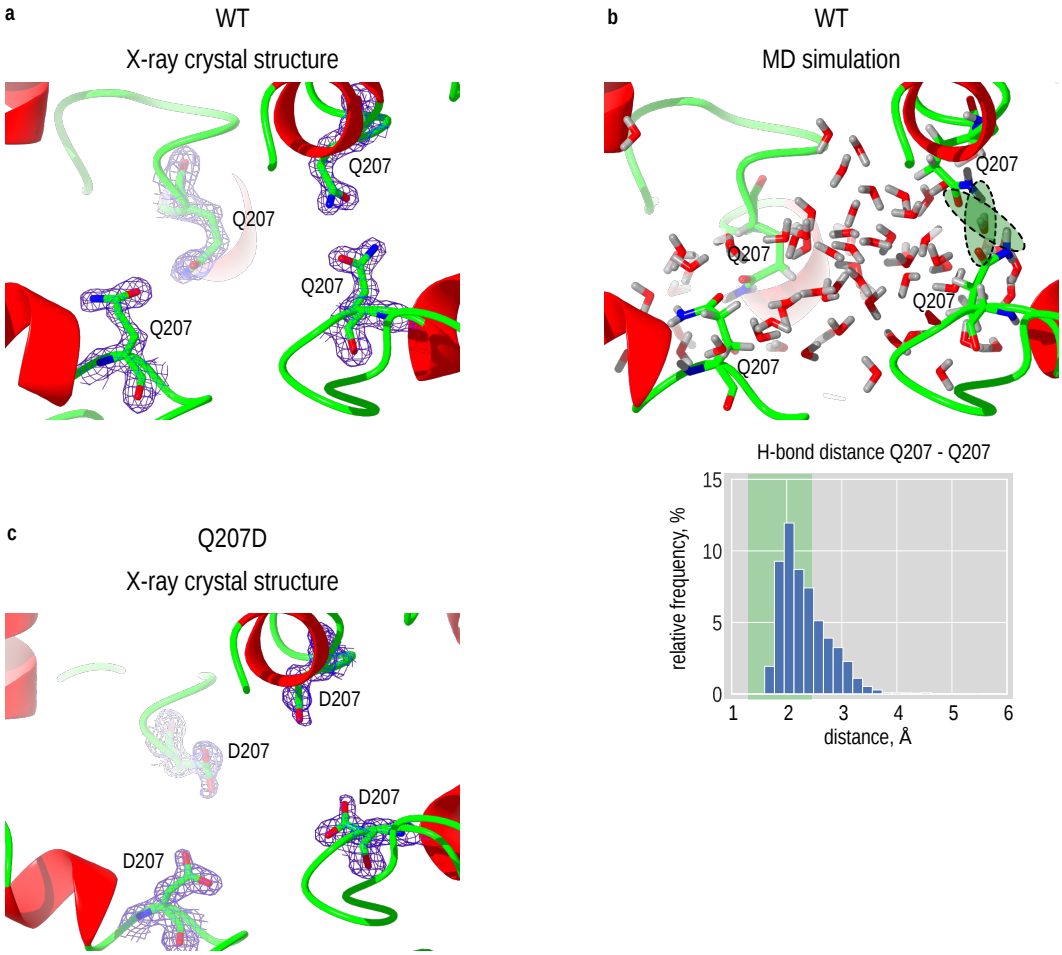

Figure S4: **Crystal structure and MD simulation analysis of side contact of wild type and Q207D crystals. a** X-ray structure of wild type crystal with electron density contoured in blue for displayed residues. **b** MD simulation and analysis of H-bond interactions between interfacing Q207 residues showed an interaction in $48.2 \pm 8.8\%$ of the simulation time. H-bond interactions (distance < 2.5 Å) are displayed in green. **c** X-ray structure of Q207D crystal variant with electron density contoured in blue for displayed residues. D207 residues are too far apart from each other to participate in an interaction. Electron density maps are calculated with structure factor amplitude difference $2F_o - F_c$ and displayed at 1.0σ.

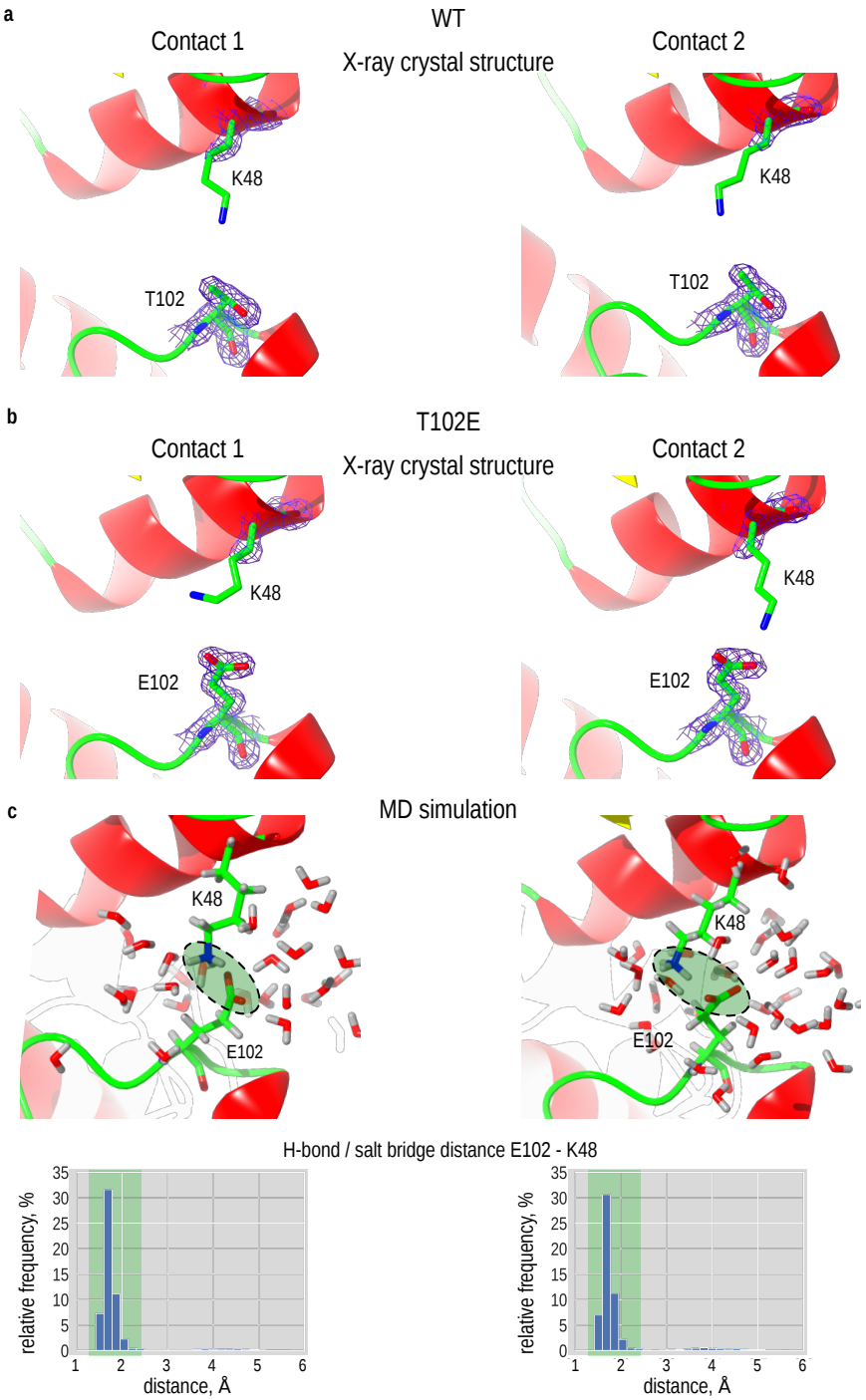

Figure S5: **Crystal structure and MD simulation analysis of corner contact 1 and 2 of wild type and T102E crystals. a** X-ray structure of wild type crystal with measured electron density contoured in blue for displayed residues. **b** X-ray structure of T102E crystal. **c** MD simulation and analysis of contact 1 and 2 revealed a highly stable H-bond between K48 and E102 in $92.0 \pm 5.1\%$ and $91.5 \pm 7.3\%$ of the time, respectively. Electron density maps are calculated with structure factor amplitude difference $2\mathrm{F}_o - \mathrm{F}_c$ and displayed at 1.0σ.

S10 Data Collection and Refinement Statistics

Table S7: Crystallographic data collection and refinement statistics. Values in parentheses are for highest-resolution shell.

Data collection	Q126H	Q126K	D54F	Q207D
PDB ID	6y10	6y0z	6y1c	7a2b
Beamline	SLS PXIII X06DA	SLS PXIII X06DA	SLS PXI X06SA	SLS PXIII X06DA
Wavelength	0.966	0.966	1.000	1.000
Space group	$I222$	$I222$	$P2_122_1$	$I222$
Cell dimensions, Å	56.05, 80.57, 113.47	55.77, 84.23, 113.56	55.76, 81.09, 113.13	55.82, 80.52, 115.10
No. of molecules per asymmetric unit	1	1	2	1
Resolution, Å	50-1.22(1.25-1.22)	50-1.21(1.28-1.21)	46.83-1.41(1.44-1.41)	46.83-1.40(1.40-1.49)
$I/\sigma(I)$	13.5 (1.4)	10.4 (0.8)	15.7 (1.7)	3.04 (3.8)
CC (1/2)	99.9 (70.9)	99.9 (63.8)	99.9 (60.6)	99.9 (96.9)
Completeness, %	99.1 (99.4)	98.9 (97.7)	98.0 (94.5)	99.6 (98.6)
Redundancy	7.0 (6.3)	5.07 (4.4)	13.4 (12.3)	4.0 (3.91)
R_{meas}, %	8.7 (157.0)	8.6 (191.1)	9.7 (142.1)	7.4 (19.4)
R_{merge}, %	8.3 (144.4)	7.7 (168.5)	9.4 (136.2)	7.0 (18.1)
R_{pim}, %	3.3 (60.0)	3.7 (88.3)	2.7 (39.9)	2.6 (6.9)
Refinement				
Resolution, Å	1.22	1.21	1.41	1.40
No. unique reflections	72057 (5601)	77296 (12811)	93011 (6907)	98425 (15718)
R_{work} / R_{free}	13.8/16.4	14.9/17.2	12.7/16.2	14.4/17.5
No. atoms				
Protein	2077	2079	4053	2010
Water	407	286	589	318
Other	36	52	81	1
B-factors				
overall	14.3	19.0	18.0	16.5
protein main chain	12.0	16.3	14.3	15.3
protein side chain	14.0	17.4	17.7	18.0
R.m.s. deviations				
Bond lengths, Å	0.008	0.010	0.007	0.009
Bond angles, °	1.49	1.59	1.38	1.48
Ramachandran plot				
Most favored, %	96	97	98	98
Additional allowed, %	4	2	2	2

References

[1] P. Grob, M. Huber, B. Walla, J. Hermann, R. Janowski, D. Niessing, D. Hekmat, and D. Weuster-Botz. Crystal contact engineering enables efficient capture and purification of an oxidoreductase by technical crystallization. *Biotechnology Journal*, page 2000010, 2020.

[2] D. Liebschner, P. V. Afonine, N. W. Moriarty, B. K. Poon, O. V. Sobolev, T. C. Terwilliger, and P. D. Adams. Polder maps: improving omit maps by excluding bulk solvent. *Acta Crystallographica Section D: Structural Biology*, 73(2):148–157, 2017.