

Supporting Information

Computational Redesign of an ω -Transaminase from *Pseudomonas jessenii* for Asymmetric Synthesis of Enantiopure Bulky Amines

Qinglong Meng^{1,4}, Carlos Ramírez-Palacios^{1,2,4}, Nikolas Capra¹, Mattijs E. Hooghwinkel¹, Sebastian Thallmair^{2,3}, Henriëtte J. Rozeboom¹, Andy-Mark W. H. Thunnissen¹, Hein J. Wijma¹, Siewert J. Marrink², Dick B. Janssen^{1*}

¹Biotransformation and Biocatalysis, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 4, 9747 AG Groningen, Groningen, The Netherlands

²Molecular Dynamics, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 7, 9747 AG Groningen, Groningen, The Netherlands

³Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany

⁴These authors contributed equally: Qinglong Meng, Carlos Ramírez-Palacios.

* Corresponding author

Groningen Biomolecular Sciences and Biotechnology Institute
University of Groningen
Nijenborgh 4
9747 AG Groningen
The Netherlands

E-mail: d.b.janssen@rug.nl

Contents

Supplementary Figures S1-S14.....	3
Supplementary Table S1	13
Description of trends and outliers shown in Figure 1	15
Water displacement approach, Figures S14-S17.....	17
Thermostability, Table S2.....	21
Crystallographic details, Table S3	22
References	23

Supplementary Figures S1-S14

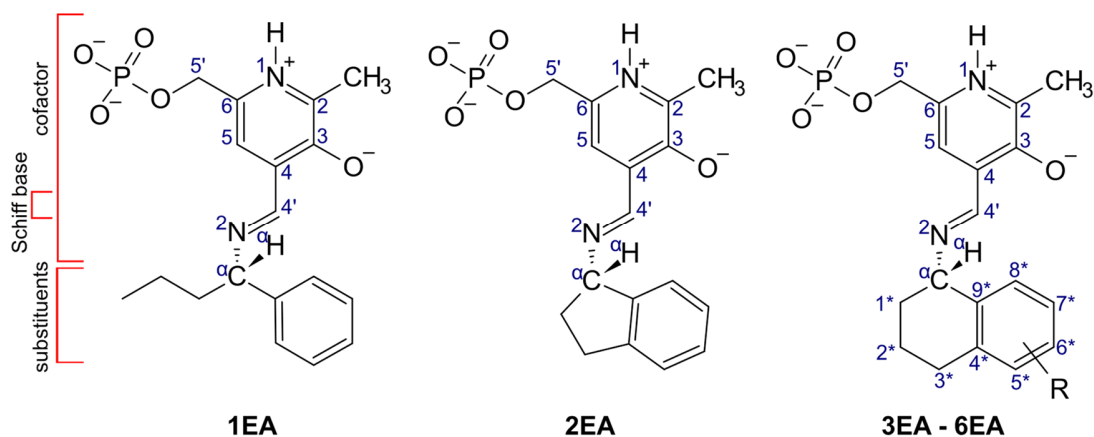


Figure S1. External aldimines used in this study. The ligands **1EA–6EA** correspond to the external aldimine form of substrates **1a–6a**, respectively. The ligand structures were made by adding a Schiff base between the substrate and the cofactor atoms. The cofactor atoms were kept fixed during the rotamer generation stage. R = H for **3EA**, and -OMe for **4EA**, **5EA** and **6EA**, in positions 5*, 6* and 7* of the tetralin moiety, respectively.

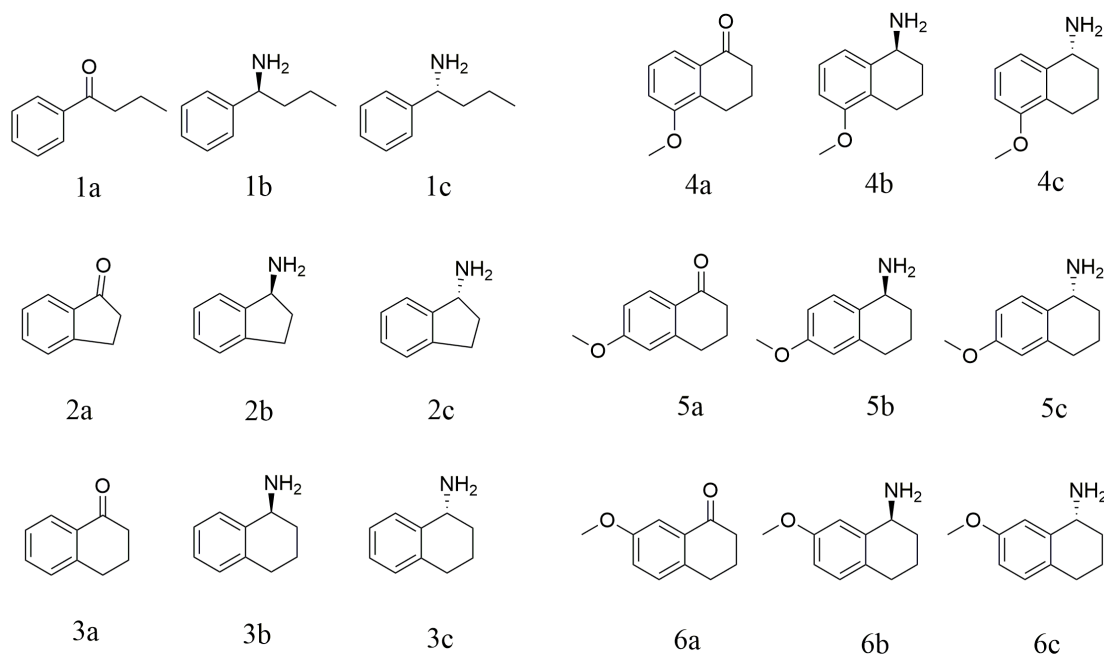


Figure S2. Ketones and chiral amines analyzed by chiral GC or GC-MS. a, ketones; b, (*S*)-enantiomers; c, (*R*)-enantiomers.

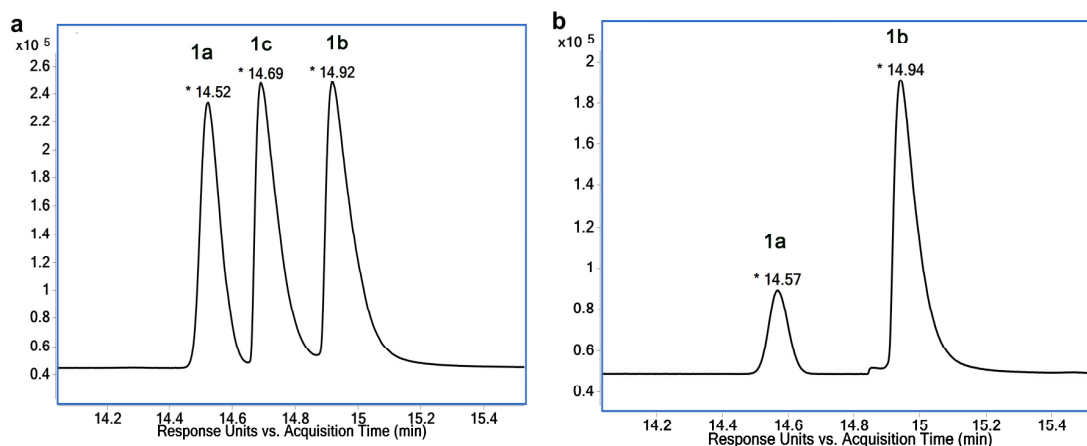


Figure S3. Chiral GC of **1a**, **1b** and **1c** using method 1. (a) 20 mM **1a**, **1b** and **1c** without enzyme; (b) Reaction product formed from 20 mM **1a** with *PjTA*-R6 mutant W58M+F86L+R417L. Retention times: **1a** 14.5 min, **1b** 14.9 min, **1c** 14.7 min. Column: CP Chiralsil Dex CB. GC settings for method 1¹: injector 220 °C, helium flow 1.65 mL/min, start at 100 °C, increase at 1.75 °C/min to 145 °C, hold for 5 min, increase at 20 °C/min to 180 °C, hold for 6.5 min.

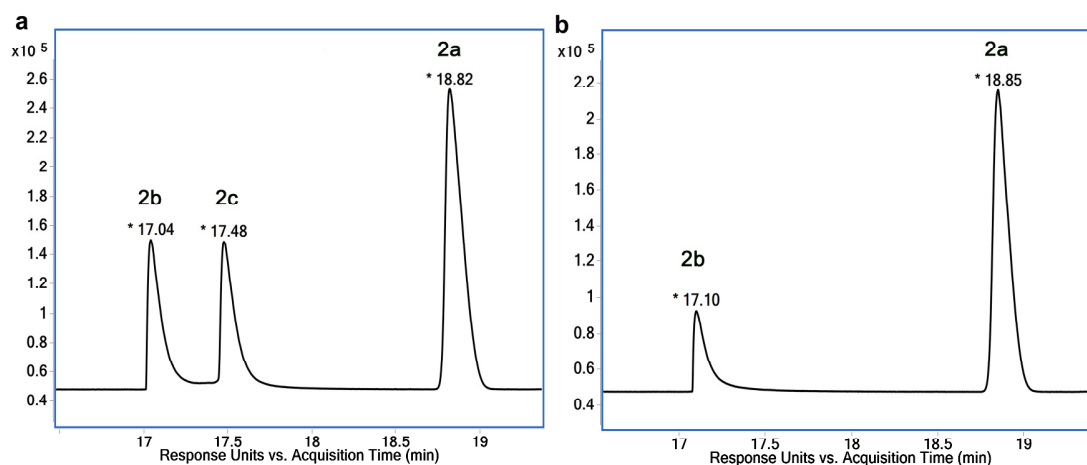


Figure S4. Chiral GC of **2a**, **2b** and **2c** using method 2. (a) 20 mM **2a**, **2b** and **2c** without enzyme; (b) Conversion of 20 mM **2a** with *PjTA*-R6 mutant W58G. Retention times: **2a** 18.8 min, **2b** 17 min, **2c** 17.5 min. Column: CP Chiralsil Dex CB. GC settings for method 2²: injector 200 °C, helium flow 1.7 mL/min, start at 100 °C, hold for 2 min, increase at 2 °C/min to 130 °C, hold for 5 min, increase at 10 °C/min to 180 °C, hold for 9 min.

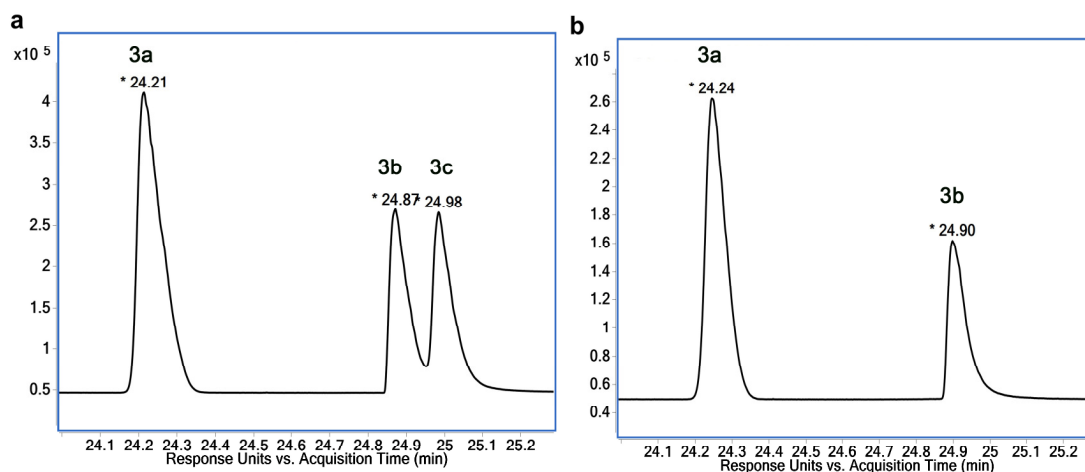


Figure S5. Chiral GC of **3a**, **3b** and **3c** using method 2. (a) 20 mM **3a**, **3b** and **3c** without enzyme; (b) Conversion of 20 mM **3a** with *PjTA*-R6 mutant W58G. Column: CP Chiralsil Dex CB. GC method: see Figure S4. Retention times: **3a** 24.2 min, **3b** 24.9 min, **3c** 25 min.

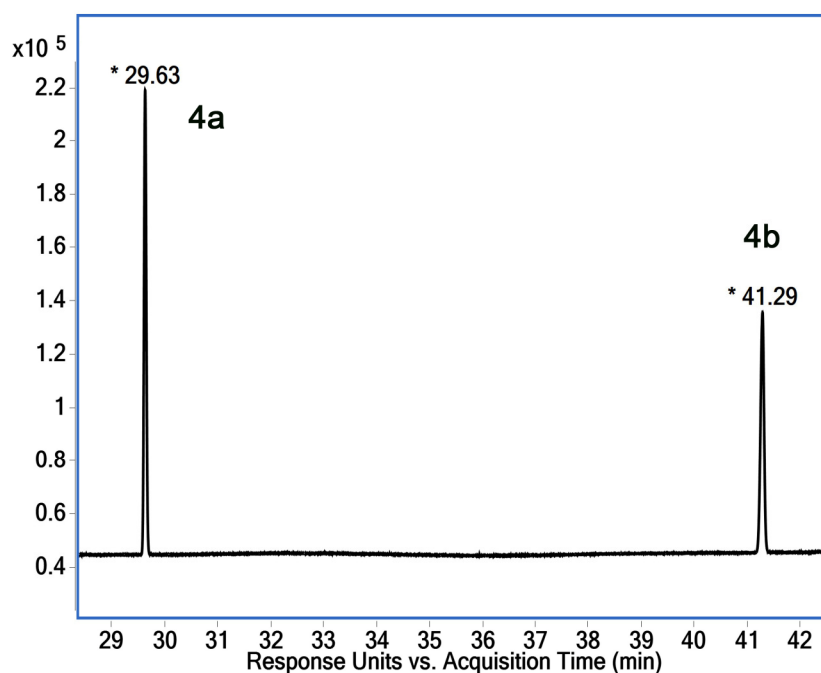


Figure S6. Chiral GC of reaction product formed from 20 mM **4a** with *PjTA*-R6 mutant W58G using method 3. GC settings for method 3²: injector 200 °C, helium flow 1.7 mL/min, start at 100 °C, hold for 2 min, increase at 2 °C/min to 130 °C, hold for 5 min, increase at 10 °C/min to 180 °C, hold for 9 min, increase at 10 °C/min to 200 °C, hold for 10 min. Retention times: **4a** 29.6 min, **4b** 41.3 min (**4b** was derivatized as acetamide). Column: CP Chiralsil Dex CB.

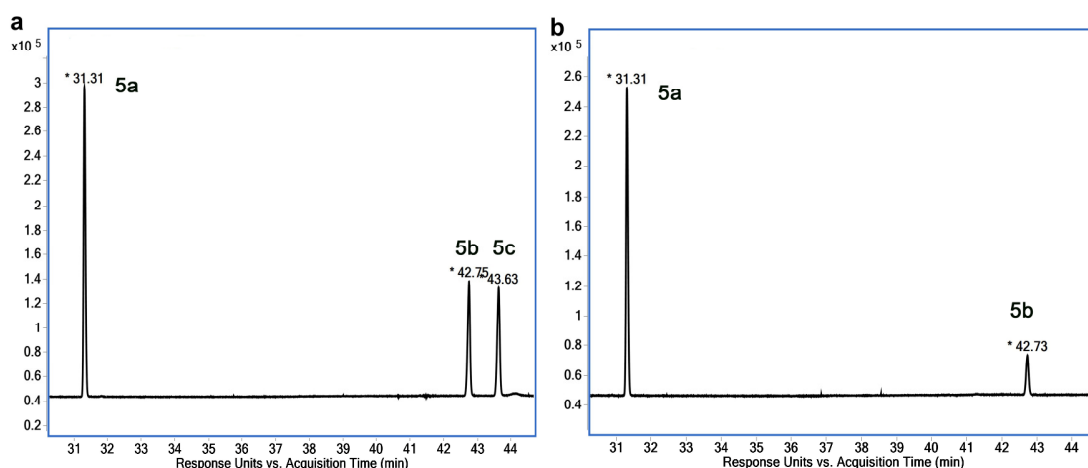


Figure S7. Chiral GC of **5a**, **5b** and **5c** using method 3. (a) 20 mM **5a**, **5b** and **5c** without enzyme; (b) Conversion of 20 mM **5a** with *PjTA*-R6 mutant W58G. Column: CP Chiralsil Dex CB. GC method: see Figure S6. Retention times: **5a** 31.3 min, **5b** 42.7 min, **5c** 43.6 min (**5b** and **5c** were derivatized as acetamide).

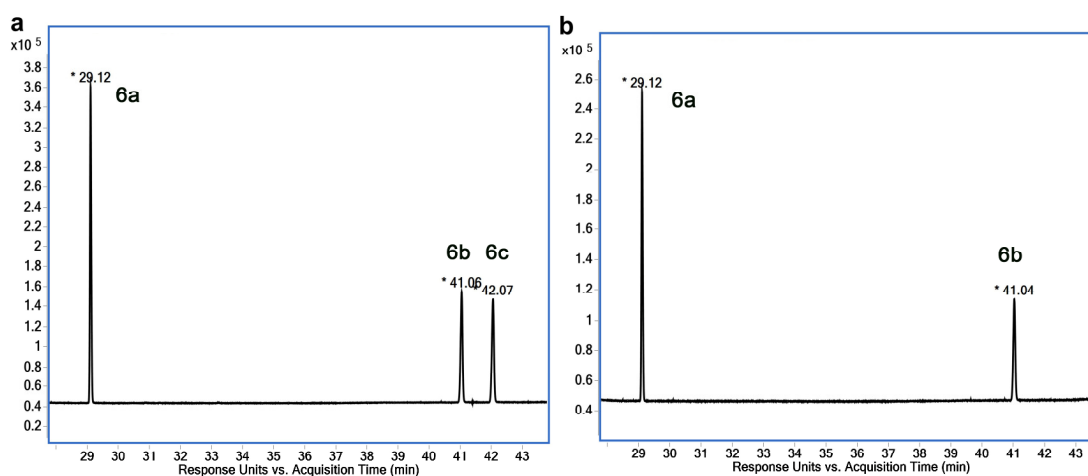


Figure S8. Chiral GC of **6a**, **6b** and **6c** using method 3. (a) 20 mM **6a**, **6b** and **6c** without enzyme; (b) Conversion of 20 mM **6a** with *PjTA*-R6 mutant W58G. Column: CP Chiralsil Dex CB. GC method: see Figure S6. Retention times: **6a** 29.1 min, **6b** 41.1 min, **6c** 42.1 min (**6b** and **6c** were derivatized as acetamide).

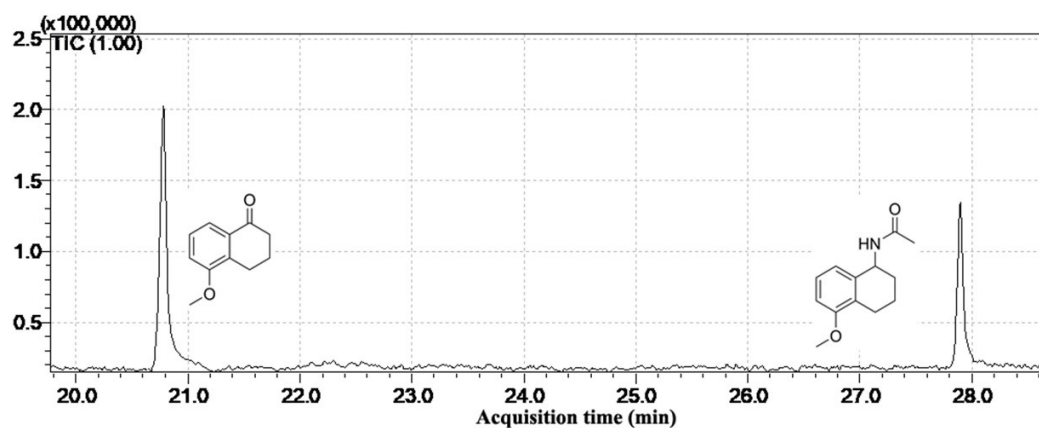


Figure S9. GC-MS of reaction product formed from 20 mM **4a** with *Pj*TA-R6 mutant W58G using method 4. GC-MS settings for method 4³: injector 250 °C, start at 80 °C, hold for 6.5 min, increase at 10 °C/min to 160 °C, hold for 5 min, increase at 20 °C/min to 200 °C, hold for 2 min, increase at 20 °C/min to 280 °C, hold for 1 min. Column: HP-1MS. Retention times: **4a** 20.8 min, **4b** 27.9 min (**4b** was derivatized as acetamide).

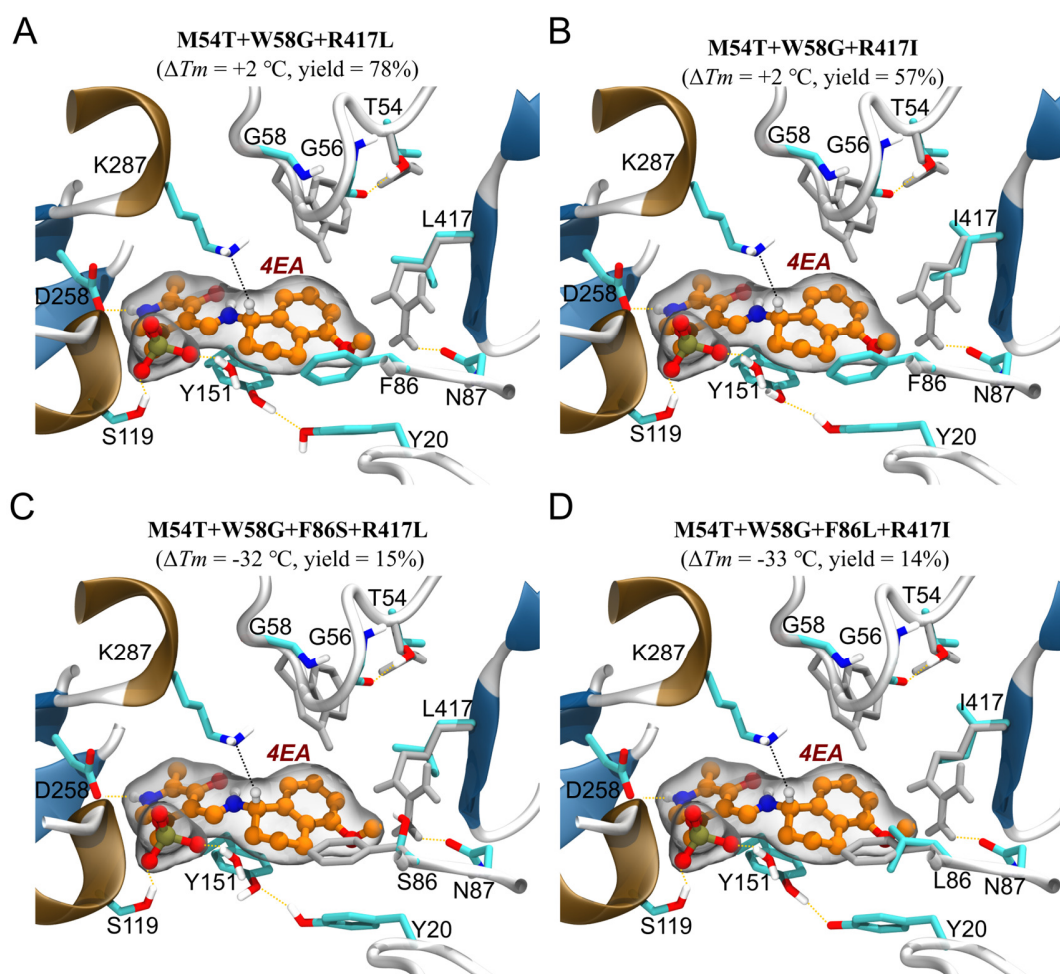


Figure S10. Predicted structures of enzyme-external aldimine complexes of variants designed for the asymmetric synthesis of **4b.** Residues surrounding the binding site for the external aldimine in the design structures and the original *PjTA*-R6 crystal structure are shown in cyan and gray, respectively. Ligand atoms are shown in CPK (ball and stick) representation. **A)** Variant M54T+W58G+R417L is capable of converting **4a** to **4b**, even though the structure shows steric clashes between the tetralin moiety of **4EA** and residue Phe86. **B)** Similar clashes are visible in variant M54T+W58G+R417I. **C)** Adding the F86S mutation to variant M54T+W58G+R417L alleviates the steric clashes but gives a variant that gives a low yield of **4b** and shows poor thermostability. The cause of the loss of stability is unclear. **D)** The F86S mutation also removes steric clashes in M54T+W58G+R417L but the effect on stability and yield is also detrimental.

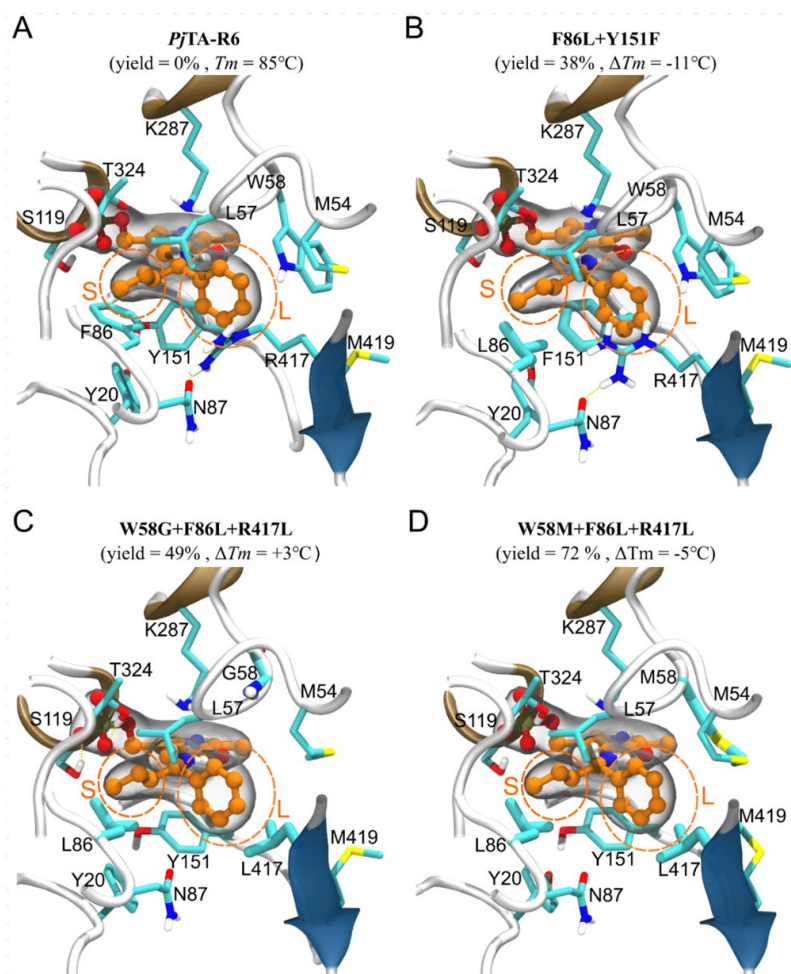


Figure S11. Docked structures of the best variants for the synthesis of **1b.** Residues surrounding the binding site for the external aldimine in the design structures and the original crystal structure are shown in cyan and gray, respectively. Ligand atoms are shown in CPK representation. **A)** The alkyl substituent of substrate **1a** does not fit in the small binding pocket due to steric clashes with the sidechain of Phe86, explaining the lack of activity of *PjTA*-R6 with this substrate. Also, a water bridge between Y151 and the phosphate group of the cofactor may hinder substrate binding. **B)** Mutations F86L and Y151F create more space for binding the alkyl substituent of **1a**. The Y151 mutation leads to loss of a water bridge, making the mutant more prone to loss of the PLP cofactor and lowering of the T_m^{app} . Concomitant introduction of L57D, which is close to the reactive lysine, suppresses the loss of stability but reduces activity. **C)** Mutations in variant *PjTA*-R6- W58G+F86L+R417L alleviate steric clashes in the small pocket (F86L), while opening up more space in the large pocket (W58G). Mutation R417L makes the binding site more hydrophobic. The use of IPA as amino donor eliminates the need for a flipping arginine (R417). **D)** The docked structure of the best variant for asymmetric synthesis of **1b** showing that the W58M mutation contributes to a more spacious hydrophobic binding site.

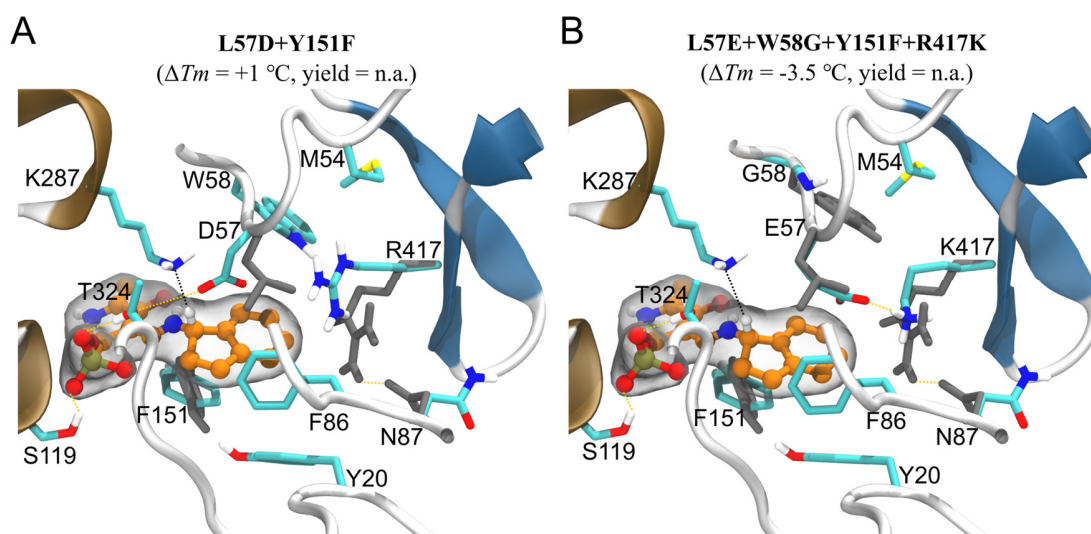


Figure S12. Docking structures of two inactive mutants designed for the synthesis of **2b.** The structures are shown in colored licorice representation overlaid to the template crystal structure (gray). Ligand atoms are shown in CPK (ball and stick) representation. **A)** Variant *PjTA*-R6-L57D+Y151F. The L57D mutation introduces an Asp57 - Thr324 hydrogen bond, disturbing the catalytically important interaction of the latter with Lys287. The Y151F mutation can reduce the thermostability by eliminating a water bridge between Tyr151 and the cofactor's phosphate. The observed $\Delta T_m^{\text{app}} = +1^\circ\text{C}$ can be explained by a newly-formed H-bond network (PLP-OPO₃²⁻ :: Thr324 :: Asp57) replacing the water bridge in *PjTA*-R6. **B)** Despite being able to accommodate **2EA** (thanks to mutation W58G), the mutant *PjTA*-R6-L57E+W58G+Y151F+R417K is inactive because the indane moiety is pushed downwards (in the direction of Phe151) by the newly-formed salt bridge between Glu57 and Lys417, causing H_a (the external aldimine hydrogen that is abstracted by the catalytic lysine) ([Figure S1](#)) to move farther apart from the catalytic Lys287.

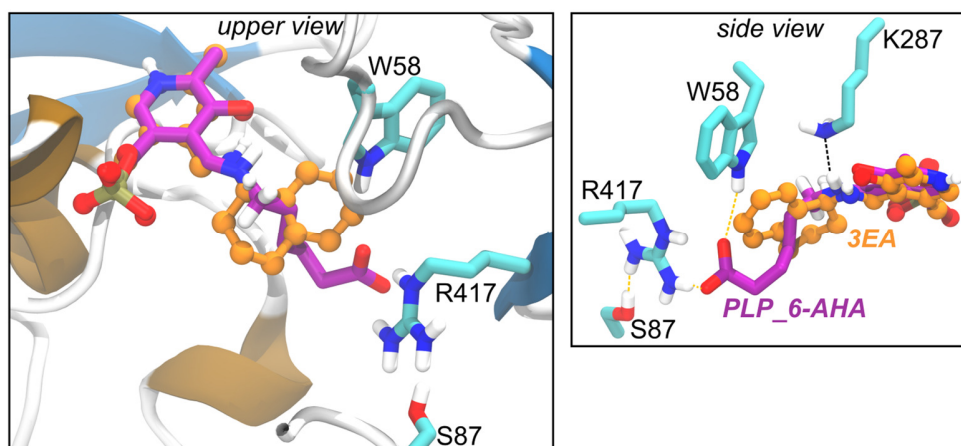


Figure S13. Comparison of the external aldimine structure of the native substrate of *PjTA* and **3EA.** The external aldimine form of 6-aminohexanoic acid (6-AHA) is shown in magenta licorice, and **3EA** is shown in orange CPK representation. The external aldimine form of 6-AHA (PLP_6-AHA) was obtained from the crystal structure (PDB 6G4E). A salt bridge between the distal carboxylate of 6-AHA and Arg417, which in turn is assisted by a hydrogen bond to Ser87, gives stability to the structure. Additional stability is conferred by a second salt bridge between the other oxygen of the distal carboxylate of 6-AHA and Trp58. For **3EA** (or **2EA**, **4EA**, **5EA**, **6EA**) residue Trp58 produces steric hindrance preventing the original template (*PjTA*-R6) from accepting the indane or tetralin moieties. In *PjTA*-R6, the wild-type Ser87 is replaced by Asn87 to increase the enzyme's thermostability.

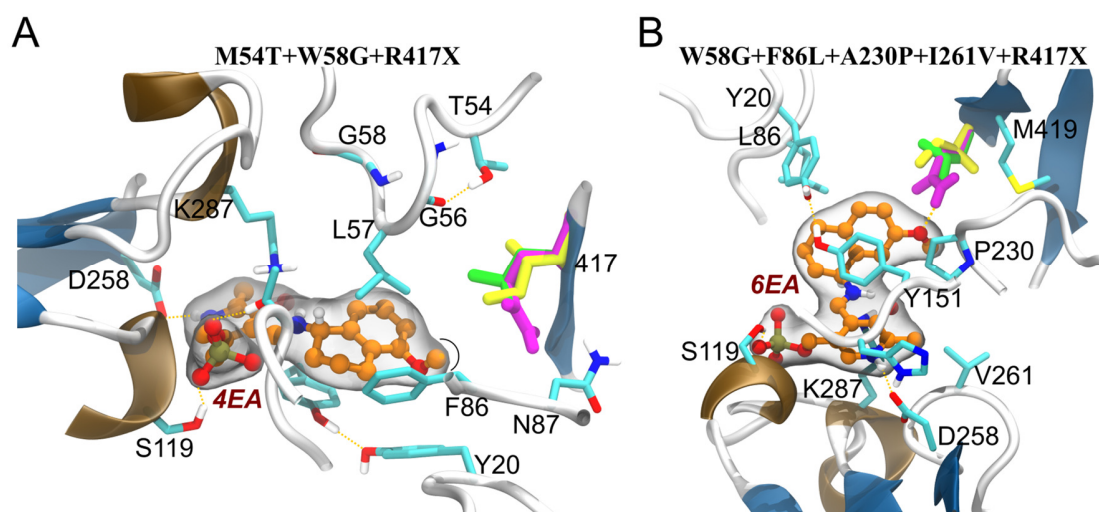


Figure S14. Rosetta docking structures of variants containing mutation R417X. The relative position of the -OMet substituent on the tetralin ring (5* and 7* for **4EA** and **6EA**, respectively) influences the interactions of the ligand in R417X variants. **A)** The M54T+W58G+R417X series for the synthesis of **4b**, where X = L (green), I (yellow), and Q (magenta), showed a yield of 78%, 57%, and 1% respectively. The position of the -OMet substituent in the tetralin ring does not allow hydrogen bond formation with Q417. **B)** W58G+F86L+A230P+I261V+R417X variants for the synthesis of **6b**, where X = L (green), Q (magenta), and I (yellow). Mutation of Arg417 to Gln increased the yield from 4% to 11% (W58G+F86L+A230P+I261V and W58G+F86L+A230P+I261V+R417Q, respectively). Gln417 can make hydrogen bond contacts with the oxygen lone pairs of the -OMet substituent. The interaction is only possible if the -OMet substituent is at position 7* of the tetralin moiety. Mutation of Arg417 to Leu also increased the experimental yield from 4% to 12% (W58G+F86L+A230P+I261V and W58G+F86L+A230P+I261V+R417L, respectively). Leu417 can make favorable hydrophobic interactions with the methyl group of **6EA**, via either of the two distal methyl groups of Leu. On the other hand, mutation of Arg417 to Ile kept the experimental yield at 4% (W58G+F86L+A230P+I261V and W58G+F86L+A230P+I261V+R417I, respectively).

Supplementary Table S1

Table S1. Search space for library design. In parenthesis the number of unique mutations produced in each run. Runs are not consecutive. In some cases, the search space includes the scaffold residue itself to allow Rosetta to leave the position intact.

1EA					
Run 1 (324)		Run 2 (54)		Run 3 (25)	
57	ACFGILMPVWY	58	ACFGILMPVWY	86	SDNFLY
58	ACFGILMPVWY	230	ACFGILMPVWY	20	FSY
230	ACFGILMPVWY	417	ACFGILMPVWY	57	LA
260	ACFGILMPVWY	20	ACFGILMPVWY	58	WG
261	ACFGILMPVWY	57	ACFGILMPVWY	154	IVA
417	ACFGILMPVWY	86	ACFGILMPVWY	230	AG
419	ACFGILMPVWY			261	IVA
20	ACFGILMPVWY			417	RKFL
87	ACFGILMPVWY				
89	ACFGILMPVWY				
151	ACFGILMPVWY				
153	ACFGILMPVWY				
321	ACFGILMPVWY				
163	ACFGILMPVWY				
86	ACFGILMPVWY				
2EA					
Run 1 (71)		Run 2 (20)			
57	ADEFGHILNPVWY	20	FSY		
58	WG	57	LA		
150	FY	58	WG		
261	AFGILVWY	154	IVA		
417	RKL	230	AG		
		261	IVA		
		417	RKLF		
		86	SNDFL		
3EA					
Run 1 (18)		Run 2 (9)			
20	FSY	419	ACDEFGHIKLMNPQRSTVWY		
57	LA	261	ACDEFGHIKLMNPQRSTVWY		
58	WG	54	ACDEFGHIKLMNPQRSTVWY		
154	IVA	230	ACDEFGHIKLMNPQRSTVWY		
230	AG	426	ACDEFGHIKLMNPQRSTVWY		
261	IVA				
417	RKLF				
86	SNDFL				
4EA					
Run 1 (9)		Run 2 (27)		Run 3 (38)	
20	FSY	58	G	54	ACDEFGHIKLMNPQRSTVWY
54	T	154	IVA	58	G
57	LA	230	AP	154	IVA
58	WG	261	ADEGHINPQSTV	230	AP
154	IVA	417	ACDEFGHIKLMNPQRSTVWY	261	ADEGHINPQSTV
230	AG	86	SNFL	417	ACDEFGHIKLMNPQRSTVWY
261	IVA			86	SNFL
417	RKLF				
86	SNDFL				
Run 4 (14)		Run 5 (43)		Run 6 (8)	
20	FSY	58	G	58	G
57	LA	154	IVA	54	ACDEFGHIKLMNPQRSTVWY

58	WG	230	AP	417	ACDEFGHIKLM NPQRSTVWY
154	IVA	261	ADEGHINPQSTV		
230	AG	417	ACDEFGHIKLMNPQRSTVWY		
261	IVA	86	SNFL		
417	RKLF				
86	SNDFL				
5EA					
Run 1 (18)		Run 2 (43)		Run 3 (56)	
20	FSY	58	G	54	ACDEFGHIKLM NPQRSTVWY
54	T	154	IVA	58	G
57	LA	230	AP	154	IVA
58	WG	261	ADEGHINPQSTV	230	AP
154	IVA	417	ACDEFGHIKLM NPQRSTVWY	261	ADEGHINPQSTV
230	AG	86	SNFL	417	ACDEFGHIKLM NPQRSTVWY
261	IVA			86	SNFL
417	RKLF				
86	SNDFL				
Run 5 (38)		Run 6 (51)		Run 7 (42)	
58	G	54	ACDEFGHIKLM NPQRSTVWY	58	G
154	IVA	58	G	54	ACDEFGHIKLMNPQRSTVWY
230	AP	154	IVA	261	ACDEFGHIKLMNPQRSTVWY
261	ADEGHINPQSTV	230	AP	417	ACDEFGHIKLMNPQRSTVWY
417	ACDEFGHIKLM NPQRSTVWY	261	ADEGHINPQSTV		
86	SNFL	417	ACDEFGHIKLMNPQRSTVWY		
		86	SNFL		
6EA					
Run 1 (11)		Run 2 (27)		Run 3 (40)	
20	FSY	58	G	54	ACDEFGHIKLM NPQRSTVWY
54	T	154	IVA	58	G
57	LA	230	AP	154	IVA
58	WG	261	ADEGHINPQSTV	230	AP
154	IVA	417	ACDEFGHIKLM NPQRSTVWY	261	ADEGHINPQSTV
230	AG	86	SNFL	417	ACDEFGHIKLM NPQRSTVWY
261	IVA			86	SNFL
417	RKLF				
86	SNDFL				
Run 5 (34)		Run 6 (24)		Run 7 (17)	
58	G	58	G	58	G
154	IVA	154	IVA	54	ACDEFGHIKLMNPQRSTVWY
230	AP	230	AP	261	ACDEFGHIKLMNPQRSTVWY
261	ADEGHINPQSTV	261	ACDEFGHIKLM NPQRSTVWY		
417	ACDEFGHIKLM NPQRSTVWY	417	ACDEFGHIKLMNPQRSTVWY		
86	SNFL	86	SNFL		
		380	ACDEFGHIKLMNPQRSTVWY		

Description of trends and outliers shown in Figure 1

The computational procedure described in this work was aimed at producing *PjTA*-R6 variants with new activities toward bulky amines. The procedure yielded the desired variants with an excellent hit rate. In search for a convenient computational parameter that predicts catalytic performance of a series of designs we examined if the Rosetta Interface Energies (docking scores) correlate with experimental yields in amine synthesis (Figure 1). Both linear ($y = mx + b$) and exponential ($y = e^x + b$) regression lines were plotted. The outcome suggests that interface energies offer a reasonable predictor of yield in experimental amine synthesis. In some cases, the trend approaches an exponential relation (**2EA** and **6EA**), while in others the trend is arguably linear (**3EA** and **5EA**). However, a few enzyme variants were found to not follow the trend and are labeled as outliers in Figure 1 (colored red). Individual substrates and trends are discussed below. It should be noted that no attempt was made to fully justify each and every outlier due to the numerous variables and factors that may influence synthetic yields that are not accounted for by calculating binding energies with a single reaction intermediate.

Ligand 1EA. Almost all enzyme variants follow the expected trend (the lower the Rosetta score, the higher the yield), with the best variant being correctly identified by the Rosetta score (*mut2*). There is but one mutant that does not seem to follow the trend: *mut5*. The expected yield for *mut5* (38% yield) would be closer to the yield of *mut2* if it followed the same trend as the other mutants. The only characteristic from *mut5* is that it has a $\Delta T_m^{\text{app}} = -11^\circ\text{C}$, which may contribute to the lower yield.

Ligand 2EA. In general, the yields for this dataset were low (*mut2* had a yield of 51%, other variants have a yield $\ll 20\%$), and the dataset is small, which makes it difficult to find correlations even in case of a perfect predictor. The dataset has only five active variants, all of which follow the trend: *mut2* \gg *mut7* $>$ *mut12* $>$ *mut10* $>$ *mut27*. The best variant, *mut2*, is correctly predicted by Rosetta to be the best in the dataset. There are two variants with no observed enzymatic activity toward compound 2: *mut6* and *mut26*. An explanation for the lack of enzymatic activity is given in Figure S12, suggesting that the enzyme-intermediate complexes do not adopt catalytically competent poses. *Mut6* and *mut26* were therefore both marked as outliers (*red squares* in Figure 1) regardless of whether they follow the overall trend.

Ligand 3EA. There are three mutants in the dataset that appear to be out of place: *mut10*, *mut12*, and *mut13*. The thermostability of these variants is not very different from that of *PjTA*-R6 ($\Delta T_m^{\text{app}} = +4.5$, -2 , and -3°C for *mut10*, *mut12*, and *mut13*, respectively) and no structural explanation for the lower-than-expected yields could be found. The three variants are triple mutants at positions 58, 86, and 417. For cases like these, wherein rationalization proved too challenging, we examined if water displacement could play a role (presented

below). If it is assumed that *mut2* is the outlier, the correlation would be $r^2_{\text{linear}} = 0.52$ and $r^2_{\text{exp.}} = 0.57$. In either case, the good variants for **3EA** exhibit binding energies better than -22.25 kcal/mol.

Ligand 4EA. The main feature in the dataset is the presence of two clusters or regions (one in the upper left corner of the plot and the other in the lower right corner of the same plot). From the upper left region, *mut2*, *mut17*, *mut7*, *mut15*, and *mut12* exactly follow the expected trend. Furthermore, the outliers *mut19*, *mut32*, and *mut33* show strongly reduced stability (T_m^{app} reduced by 34, 32, and 33°C, respectively). For *mut11*, the loss of stability is smaller but it may still be significant ($\Delta T_m^{\text{app}} = -8^\circ\text{C}$). Variants *mut3* and *mut11* give less prominent outliers (e.g., r^2 increases from 0.75 to 0.83 when omitting *mut3*) and the yields deviate only slightly from expected values.

Ligand 5EA. The Rosetta Interface Energy correctly predicts *mut2* as the best variant, followed by *mut14*, *mut13*, *mut12*, *mut38*, and *mut35*. There are four variants, however, that do not follow the described trend: *mut20*, *mut21*, *mut22*, and *mut34*. No structural explanation could be found to rationalize these four mutants. The ΔT_m^{app} values were -8 , -3 , -11 , and -5°C for *mut20*, *mut21*, *mut22*, and *mut34* respectively.

Ligand 6EA. Even though no variant was labeled as outlier in this plot, it could be argued that *mut2* is out of place respect to the other variants. In that case, the correlation would increase to $r^2_{\text{linear}} = 0.77$ and $r^2_{\text{exp.}} = 0.88$. Be that as it may, the results do suggest that the Rosetta Interface Energy taken from the external aldimine complex can generally predict the outcome of the transamination reaction.

Water displacement approach, Figures S14-S17

The water displacement approach was performed to further rationalize the unexplained variants from the datasets of ligands **3EA**, **4EA**, and **5EA** (*red circles* in [Figure 1](#)). MD simulations were not performed on the **1EA**, **2EA**, and **6EA** datasets. We wanted to investigate whether specific mutations increased the number of waters that would need to be displaced upon substrate binding and if this could explain yields that were lower than expected from Interface Energies. The water displacement that would need to happen for substrate binding to occur was calculated because one of the main pitfalls of docking algorithms, in general, is the lack of explicit water molecules. The energy penalty of water displacement upon substrate binding has been studied in other systems,^{4,5} but it does not appear to have been studied in ω -TAs despite most mutagenesis efforts aimed at broadening the substrate range do so by increasing the size of the binding pockets. A larger binding site means that in absence of the ligand, the binding site might be occupied by water molecules that need to be removed for substrate binding to occur with an associated energy penalty.^{6,7} We hypothesized that if some of the designed variants had binding sites too enlarged (otherwise necessary to accommodate the bulky compounds), then the additional energy penalty could explain the lower-than-expected yields of some variants in the **3EA**, **4EA**, and **5EA** datasets.

The water displacement that would be needed for the substrate to bind could be estimated by comparing the volume of the binding site of PMP-bound and LIG-bound structures using, for example, Voronoi-based methods.⁸ However, the high hydrophobicity of the binding site of ω -TAs creates an unfavorable environment for water molecules to actually occupy the available space ([Figure S15A](#)) that the tetralin moiety would otherwise occupy ([Figure S15B](#)). Therefore, we performed MD simulations of the LIG-bound and PMP-bound systems, and counted the average number of water molecules around a 1.4 nm radius from a central atom in both the PMP-bound and in LIG-bound simulations. To prepare the initial simulation frame, water molecules in enough quantity to fill the simulation cell were added. These water molecules were added from a pre-equilibrated simulation of pure water ([see below](#)), and water molecules sterically overlapping with the solute were removed which can result in small vacuum bubbles forming near the solute atoms. The MD simulations can remove some of the bubbles and smooth the water density around the binding site.⁹ The simulations are not intended to fully hydrate the protein, which could take from the nanosecond to the microsecond time scales.^{10,11} Conversely, water molecules in the hydration layer (diffusive waters), which are involved in molecular recognition of ligands, have dynamics in the picosecond to nanosecond time scales.¹² Simulations in the nanosecond time scale (1 – 10 ns) have been previously used to estimate thermodynamic properties that involve binding site solvation.^{7,9,13,14}

MD simulations procedure: Two sets of MD simulations were performed, one containing an external aldimine intermediate in the binding site (**3EA**, **4EA**, **5EA**), named “LIG-bound”, and the other containing the PMP cofactor, named “PMP-bound”. Initial atom configurations were obtained from Rosetta docking calculations (ENZ:PMP or ENZ:LIG).

MD simulations were run using the GROMACS 2020 software (<http://www.gromacs.org>).¹⁵ Topology parameters were generated in the AMBER99SB force field¹⁶ for protein residues, and the GAFF force field for the ligands.¹⁷ Parameters for the ligand (PMP or external aldimine) were determined using the AMBER's *antechamber* program¹⁷ with AM1-BCC partial charges.¹⁸ The initial complex was placed in a cubic simulation box of 10×10×10 nm³ in size with periodic boundary conditions. Simulations used explicit TIP3P water molecules (~26,000 molecules per box).¹⁹ Crystallographic water molecules were kept in the original position (PDB: 6TB1), and the rest of the water molecules necessary to fill the entire simulation box were added from a pre-equilibrated box containing pure TIP3P water (5×5×5 box, simulated for 100 ns): `gmx solvate -cp complex.gro -cs water.gro -p prot.top -o complex_solvated.gro -box 10 10 10 -radius 0.105 -scale 0.57`. Na⁺ and Cl⁻ ions were added to the simulation cell to a concentration of 0.15 M. Additional Na⁺ ions were included to neutralize the system. The system now containing protein, ligand, water, and ions was energy-minimized with 5,000 steps of steepest descent minimization, followed by two short equilibrations of 200 ps (timestep = 1.0 fs) and 300 ps (timestep = 1.5 fs). Equilibration and production simulations were carried out with harmonic position restraints on all non-solvent non-hydrogen atoms (force constant $\kappa_{pr} = 1,000 \text{ kJ/mol}^{-1}\text{nm}^{-2}$). The position restraints are necessary to prevent the ligand (external aldimine) from moving the C $_{\alpha}$ -H $_{\alpha}$ bond away from the initial configuration (Figure S1), where the H $_{\alpha}$ is closest to the catalytic Lys287-NH₂. Simulations were run at 298K and 1 bar using a velocity rescale thermostat (time constant $\tau_T = 0.1 \text{ ps}$)²⁰ and a Parrinello-Rahman barostat (relaxation time constant $\tau_p = 0.5 \text{ ps}$ and compressibility $\kappa_T = 4.5 \times 10^{-5} \text{ bar}$).²¹ Long range electrostatics were calculated with the Particle Mesh Ewald (PME) method.²² Lennard-Jones interactions were truncated at 1.0 nm, by a force switch smoothing function (*force-switch* in GROMACS) from 0.9 and 1.0 nm. The LINCS algorithm was applied to constraint all bond lengths, with 2 iterations of constraint per time step and a 4th order expansion of the constraint coupling matrix.²³ Production simulations were run with an integration timestep of 2.0 fs for 1.0 ns (3 replicas) or 10.0 ns (1 replica). The frames were saved every 10 ps. The number of water molecules near the binding site was estimated by counting the number of waters around a radius of 1.4 nm from atom N2 (Figure S1) of the external aldimine (LIG-bound) or from the equivalent N4' atom of the cofactor (PMP-bound) (Figure S15). Both water counting and figure generation were done in the VMD software (<http://www.ks.uiuc.edu/Research/vmd/>).

Results of the water displacement approach. The water displacement approach aimed to explain a few variants in the 3EA, 4EA, and 5EA datasets that did not produce the expected yield based on the Interface Energy of the docked complex (Figure 1). The approach was unsuccessful in helping us explain outliers. This means that mutants *mut10*, *mut12*, *mut13* (3EA), *mut11* (4EA), and *mut20*, *mut21*, *mut34*, *mut22* (5EA) do not stand out from all the other mutants by, for example, needing to displace much more water molecules than the rest. However, as shown in Figure S17, when considering all variants from the 3EA, 4EA, and 5EA datasets (outliers and non-outliers) a weak trend can be observed: the variants with better yields tended to need to displace fewer water molecules

for the substrate to bind (here measured as the difference between the PMP-bound and LIG-bound water count, as shown in [Figures S15 and S16](#)). The trend is weak and notably *mut2* in **3EA** and **5EA** does not follow it. [Figure S17](#) is not intended to be taken as any correlation, but only to show that there might be a trend between water displacement and experimental yield.

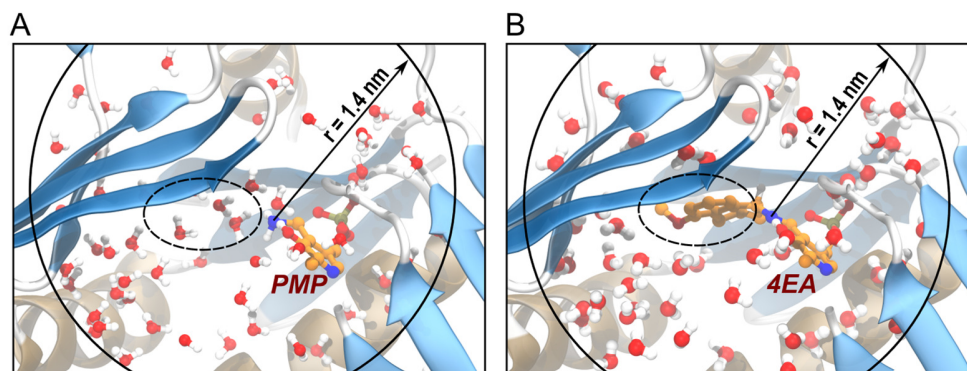


Figure S15. Number of water molecules around the binding site of *Pj*TA-R6 mutant M54T+W58G+R417I. The difference between the number of water molecules surrounding the binding site of the PMP-bound (**A**) and LIG-bound (**B**) in 1.0 or 10.0 ns simulations served as an estimation of the number of water molecules that need to be displaced for **4a** to fit into the active site. Because it is difficult to precisely define the binding site boundaries, we estimated the number of water molecules around the binding site by defining a sphere of radius 1.4 nm around an arbitrarily-chosen central atom. The absolute values are meaningless but the differences should correspond to actual changes in the number of water molecules around the binding site between the LIG-bound and PMP-bound structures.

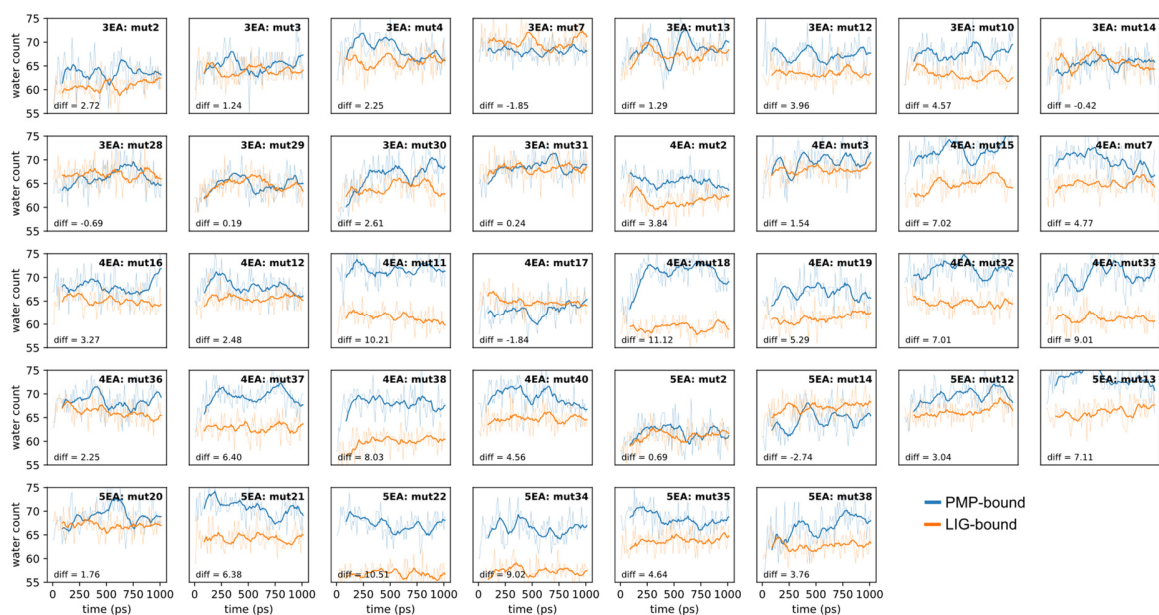


Figure S16. Time evolution of the 1.0 ns MD simulations. The number of water molecules found within 1.4 nm of the arbitrarily-chosen central atom (y-axis) is plotted across the simulation time (x-axis). The figure annotation “diff” refers to the average difference between the number of water molecules of the LIG-bound and the PMP-bound simulations. In all subplots, PMP-bound is colored *blue* and LIG-bound is colored *orange*.

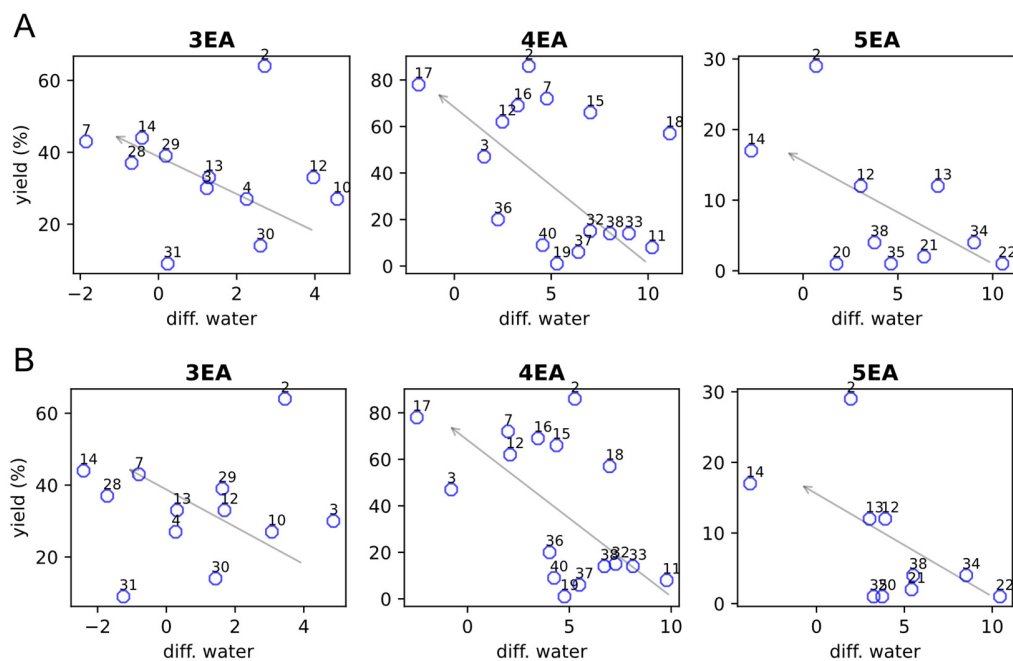


Figure S17. Difference in the number of water molecules between the PMP-bound and LIG-bound simulations (*x-axis*) plotted against the experimental yield (*y-axis*). The difference in the number of water molecules is averaged over the **A**) 1.0 ns (3 replicas) or **B**) 10.0 ns (1 replica) simulations. The hand-drawn arrows show what visually appears to be the trend: mutants that needed to displace fewer water molecules also had better yields. The trend is weak.

Thermostability, Table S2

Table S2. Thermal shift assays under different conditions.

Variant	Features	Standard conditions	ΔT_m^{app} (°C) ^a		
			1 h with 1 M IPA	1 h with 20% DMSO	24 h with 1 M IPA and 20% DMSO at 56 °C
Variant 8 (W58M+F86L+ R417L)	Retained stability, good yield	−5	−9	−6	−55 ^b
Variant 19 (M54T+W58G+ R417Q)	Drastically reduced stability, low yield	−34 ^b	−54 ^b	−54 ^b	−64 ^b
Variant 20 (M54T+W58G+ F86S)	Somewhat reduced stability, low yield	−8	−10	−42	−55 ^b

^a Values represent the difference in T_m^{app} between *PjTA*-R6 ($T_m^{\text{app}} = 85^\circ\text{C}$ under standard conditions) and mutants under the conditions shown.

^b Reduced peak in thermal shift assays.

Crystallographic details, Table S3

Table S3. Crystallographic data collection and refinement.

	R6+W58G	R6+W58M+F86L+R417L
Data collection		
Beamline	DLS/I24	DLS/I04
Wavelength (Å)	0.9686	0.9795
Space group	P4 ₃	P4 ₃
Unit cell dimensions a,b,c, (Å)	97.8, 97.8, 119.9	97.8, 97.8, 118.7
Resolution (Å) ^A	97.8 – 1.70 (1.73-1.70)	45.74 – 1.90 (1.93-1.90)
Total observations	677080 (33031)	430018 (22028)
Unique reflections	123358 (6058)	87797 (4459)
$\langle I/\sigma \rangle$ ^A	17.4 (1.9)	15.1 (1.8)
CC _(1/2)	0.999 (0.863)	0.999 (0.700)
Completeness (%) ^A	99.9 (99.4)	100.0 (100.0)
R _{merge} (%) ^{A, B}	3.8 (45.8)	4.7 (65.0)
Refinement		
R _{work} /R _{free} (%) ^B	17.3 / 19.1	17.7 / 20.7
Number of non-H atoms		
Protein	6946 (2 chains)	6924 (2 chains)
Solvent	435	398
Cofactor	38 (2 PLP)	40 (2 PLP)
Average B (Å ²)	32.9	36.0
RMSD		
Bond lengths (Å)	0.007	0.008
Bond angles (°)	1.4	1.4
Ramachandran plot % favoured, outliers	93.99, 1.02	93.36, 1.01
PDB entry	7B4I	7B4J

^A Values in parentheses correspond to highest resolution shell.

^B R_{free} is calculated as R_{work} using 5% of all reflections randomly chosen, which were excluded from structure refinement.

References

- (1) Genz, M.; Melse, O.; Schmidt, S.; Vickers, C.; Dörr, M.; van den Bergh, T.; Joosten, H. J.; Bornscheuer, U. T. Engineering the Amine Transaminase from *Vibrio Fluvialis* towards Branched-Chain Substrates. *ChemCatChem*. **2016**, *8*, 3199–3202.
- (2) Pressnitz, D.; Fuchs, C. S.; Sattler, J. H.; Knaus, T.; Macheroux, P.; Mutti, F. G.; Kroutil, W. Asymmetric Amination of Tetralone and Chromanone Derivatives Employing ω -Transaminases. *ACS Catal.* **2013**, *3*, 555–559.
- (3) Mutti, F. G.; Knaus, T.; Scrutton, N. S.; Breuer, M.; Turner, N. J. Conversion of Alcohols to Enantiopure Amines through Dual-Enzyme Hydrogen-Borrowing Cascades. *Science*. **2015**, *349*, 1525–1529.
- (4) Maurer, M.; Oostenbrink, C. Water in Protein Hydration and Ligand Recognition. *J. Mol. Recognit.* **2019**, *32*, e2810.
- (5) Wahl, J.; Smieško, M. Thermodynamic Insight into the Effects of Water Displacement and Rearrangement upon Ligand Modifications Using Molecular Dynamics Simulations. *ChemMedChem*. **2018**, *13*, 1325–1335.
- (6) Ladbury, J. E. Just Add Water! The Effect of Water on the Specificity of Protein-Ligand Binding Sites and Its Potential Application to Drug Design. *Chem. Biol.* **1996**, *3*, 973–980.
- (7) Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the Active-Site Solvent in the Thermodynamics of Factor Xa Ligand Binding. *J. Am. Chem. Soc.* **2008**, *130*, 2817–2831.
- (8) Poupon, A. Voronoi and Voronoi-Related Tessellations in Studies of Protein Structure and Interaction. *Curr. Opin. Struct. Biol.* **2004**, *14*, 233–241.
- (9) Masters, M. R.; Mahmoud, A. H.; Yang, Y.; Lill, M. A. Efficient and Accurate Hydration Site Profiling for Enclosed Binding Sites. *J. Chem. Inf. Model.* **2018**, *58*, 2183–2188.
- (10) Persson, F.; Söderhjelm, P.; Halle, B. How Proteins Modify Water Dynamics. *J. Chem. Phys.* **2018**, *148*, 215103.
- (11) Barnes, R.; Sun, S.; Fichou, Y.; Dahlquist, F. W.; Heyden, M.; Han, S. Spatially Heterogeneous Surface Water Diffusivity around Structured Protein Surfaces at Equilibrium. *J. Am. Chem. Soc.* **2017**, *139*, 17890–17901.
- (12) Mondal, S.; Mukherjee, S.; Bagchi, B. Origin of Diverse Time Scales in the Protein Hydration Layer Solvation Dynamics: A Simulation Study. *J. Chem. Phys.* **2017**, *147*, 154901.
- (13) Beuming, T.; Farid, R.; Sherman, W. High-Energy Water Sites Determine Peptide Binding Affinity and Specificity of PDZ Domains. *Protein Sci.* **2009**, *18*, 1609–1619.
- (14) Higo, J.; Nakasako, M. Hydration Structure of Human Lysozyme Investigated by Molecular Dynamics Simulation and Cryogenic X-Ray Crystal Structure Analyses: On the Correlation between Crystal Water Sites, Solvent Density, and Solvent Dipole. *J. Comput. Chem.* **2002**, *23*, 1323–1336.
- (15) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX*. **2015**, *1–2*, 19–25.
- (16) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *PROTEINS Struct. Funct. Bioinforma.* **2006**, *65*, 712–725.
- (17) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (18) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.

- (19) 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*, 926–935.
- (20) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (21) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (22) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (23) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.