Supporting Information

Engineered Aldolases Catalyzing Stereoselective Aldol Reactions Between Aryl-Substituted Ketones and Aldehydes

Eugenia Chukwu Cornelius,† Michael Bartl,† Louise J Persson, Ruisheng Xiong, Daniela Cederfelt, Farshid Mashayekhy Rad, Thomas Norberg, Sarah Engel, Erik G Marklund, Doreen Dobritzsch and Mikael Widersten*

Department of Chemistry – BMC, Uppsala University, Box 576, SE-751 23 Uppsala, Sweden

*Corresponding author: mikael.widersten@kemi.uu.se; † These authors contributed equally to the work.

Index

Experimental Details S2

Supporting Tables S6

Table S1. *Formation of aldol* ***6*** *– integrated peak areas* S6

Table S2. *Formation of aldol* ***9a*** *– integrated peak areas and catalyzed reaction rates* S7

Table S3. *Steady state kinetic parameters of $VGGC_{107}C_{163}$ catalyzed formation of* ***9a*** S9

Table S4. *Sequences of deoxyoligonucleotides used in gene library construction* S9

Table S5. *ESI-MS/MS parameters for target analytes* S9

Table S6. *Data collection and refinement statistics FSA VG* S10

Table S7. *Distances between the sulfurs of C107 and C163 in the $VGGC_{107}C_{163}$ variant* S10

Supporting Figures S11

Chart S1. *Compounds studied* S11

Figure S1. *Crystal structure of the FSA VG variant* S11

Figure S2. *Production of aldol* ***6*** *in the presence of selected $VGGX_{107}X_{163}$ variants* S12

Figure S3. *Chiral separation of aldols and reaction mixtures* S12

Figure S4. *Production of aldol* ***8*** *of selected aldolase variants* S13

Figure S5. *Production of aldol* ***7*** *of selected aldolase variants* S13

Figure S6. *Frequency of water being present* S13

Figure S7. *^{1}H-NMR of* ***9a*** S14

Figure S8. *Analysis of distillation-purified phenylacetaldehyde* S14

Figure S9. *Stepwise assembly of gene fragments* S15

Figure S10. *SDS-PAGE of purified FSA $VGGX_{107}X_{163}$ variants* S15

Figure S11. *Steady state saturation curves of* ***9a*** *formation and standard curve for* ***6*** *and* ***9a*** *quantification* S16

Figure S12. *ESI MS/MS product ions of aldols* ***7*** *and* ***9a***-***9d*** S17

Figure S13. *Parameterization compounds* S18

Figure S14. *Number of structures per cluster* S18

Figure S15. *Comparison of modeled structures of (R)- and(S)-carbinolamine in the active site of $VGGGC_{107}C_{163}$* S19

Supporting References S19

Experimental Details

Chemicals and Reagents. Compounds **1**–**5a** (Chart S1) were purchased from Aldrich (#138185, #R277614, #D107204, #A10701 and #107295, respectively). Phenylacetaldehyde (**5a**) was re-distilled prior to use (Sigma-Aldrich #107295-100, Sigma batch # MKBN8351V, our batch number: 135-009-0, amount: approximately 10 mL) was distilled under oil pump vacuum (1 mbar), bath temperature 55–85 °C. Two fractions were collected: Fraction 1, bp <31 °C (135-009-1, 0.3 g), Fr 2, bp 31-40 °C (135-009-2, 2.6 g). The fractions were analyzed by TLC (Silica gel 60 F254, toluene-ethyl acetate 9:1, UV detection) and ^{1}H-NMR. As can be seen from the TLC analysis (Figure S8A), the starting material was >50% degraded (slower migrating spots), but both the distillate fractions appeared to be reasonably pure. The difference between the starting material and the distilled fractions was also evident from the different viscosities. The starting material was a colorless, rather viscous liquid (even at room temperature), whereas both the distilled fractions (also colorless) had a more water-like viscosity. The ^{1}H-NMR analysis of the starting material (135-009-0) showed signals from phenylacetaldehyde but also from many other compounds, probably polymers self-condensation products and/or oxidation products. The main distillate fraction (135-009-2) appeared to be reasonably pure, with dominant signals from phenylacetaldehyde (purity 97 % by ^{1}H-NMR) (Figure S1B). Fractions 1 and 2 were stored refrigerated under nitrogen. ^{1}H-NMR spectra were recorded on Varian 400 MHz instruments. Chemical shifts (ppm) are reported referenced to the internal signal of residual protic solvent.

The synthesis of substituted phenylacetaldehydes (**5b**–**d**) and aldols **6** and **9a**–**b** has been described earlier.[S1,S2] Custom synthesized deoxyoligonucleotides were purchased from Thermo Fisher Scientific GmbH (Ulm, Germany). Protein extraction reagent (B-PER), DNA polymerases, restriction endonucleases and other molecular biology reagents were purchased from Thermo Fisher Scientific unless otherwise stated.

Construction of Library of FSA Variants. The GeneArt Site-Directed Mutageneis PLUS Kit was used with a pair of mutagenesis primers to generate an A129G mutation, using the gene of an R134V/S166G mutant of FSA created earlier,[S3] herein referred to as the VG mutant, as DNA template. This 'VGG' gene was subsequently applied as template to generate a gene library containing unbiased saturation mutagenesis[S4] at codons L107 and L163. The mutated genes were step-wise assembled by overlapping PCR reactions (Figure S9). Deoxyoligonucleotide sequences are given in Table S4.

Protein Production. A plasmid encoding for $(His)_5$-tagged FSA-VGG or the mixture of plasmids encoding for C-terminally $(His)_5$-tagged FSA-VGG$X_{107}X_{163}$ were transformed in parallel into *Escherichia coli* strain BL21-AI (Thermo Fisher Scientific) [pREP4-GroEL-ES].[S5] The bacteria were plated on LB-plates (LB: 10 g/l tryptone, 5 g/l yeast extract and 10 g/l NaCl) supplemented with 100 µg/mL ampicillin and 50 µg/mL kanamycin, and grown overnight at 37 °C. Resulting plates with transformants were stored at 8 °C until further use. Randomly picked transformant colonies were picked for overnight cultures in 2TY medium (2TY: 16 g/l tryptone, 10 g/l yeast extract and 5 g/l NaCl) supplemented with 100 µg/mL ampicillin and 30 µg/mL kanamycin. These overnight cultures were used for the inoculation of 100 mL of 2TY medium with 50 µg/mL ampicillin and 15 µg/mL kanamycin, aiming for an initial OD_{600} of 0.05. As OD_{600} reached ~0.8, the cultures were induced by addition of L-arabinose and isopropyl β-D-thiogalactoside to final concentrations of 0.04 % (w/v) and 0.5 mM, respectively. Following overnight incubation (30 °C, 220 rpm), the cells were harvested (6,000 *g*, 20 min, 4 °C) and the resulting bacterial pellets were stored at -80 °C awaiting further processing.

The cell pellets were resuspended in 10 mL lysis buffer (20 mM bicine, 500 mM NaCl, 20 mM imidazole, 0.1 mg/mL lysozyme, 10 µg/mL DNaseI, EDTA-free protease inhibitor, pH 8.35) and lysed using ultrasonication (5 min, 15 s on/off, 70 % amplitude). Insoluble components were separated by centrifugation (16,000 *g*, 60 min, 4 °C). The supernatant was subsequently incubated with 1 mL pre-equilibrated Ni^{2+}-IMAC-resin overnight at 8 °C while shaking. The resin was washed with 20 mL washing buffer (40 mM bicine, 500 mM NaCl, 40 mM imidazole, pH 8.35), and His-tagged aldolase was eluted with 3×1.5 mL elution buffer (40 mM bicine, 500 mM NaCl, 500 mM imidazole, pH 8.35). 2.5 mL of the combined protein-containing fractions were desalted using pre-packed Sephadex G-25 (PD-10) columns, pre-equilibrated with desalting buffer (60 mM bicine, pH 8.35). Potentially formed precipitates in the elution fractions were removed by centrifugation (3,000 *g*, 30 min, 4 °C), before applying them to the PD-10 columns. See Figure S10 for SDS-PAGE separation following purification of a selection of isolated variants.

Protein concentration was obtained from intrinsic protein absorbance at 280 nm with the use of theoretical absorption coefficients, as provided by the ProtParam service (https://www.expasy.org/resources/protparam), $VGGX_{107}X_{163}$, 13,075 $M^{-1}cm^{-1}$, $VGGX_{107}Y_{163}$, 14,565 $M^{-1}cm^{-1}$ or $VGGX_{107}W_{163}$, 18,575 $M^{-1}cm^{-1}$. If no protein was obtained after expression and purification, the possible existence of inclusion bodies or the lack of a His-tag due to a frame-shift mutation was established following SDS-PAGE stained with Coomassie Brilliant Blue R-250.

Sequencing of Aldolase Genes. Aliquots of overnight cultures were used for DNA preparation using the GeneJET Plasmid Miniprep Kit. DNA sequencing services were provided by Eurofins. Forward and reverse primer for the plasmid were used independently and the obtained sequences were aligned. In case of unclear sequencing results, the samples were interpreted to be contaminated and they were not used for further analysis.

Screening for Aldolase Activity in Enzyme Library. Samples were set up in 58.5 mM bicine, pH 8.35 with 2.5 % (v/v) acetonitrile by addition of 10 µM (2 nmol) purified aldolase, 5 mM of **5a** and 20 mM ketone **1** or **4**, respectively, in a total volume of 200 µl. The reaction mixtures were shaken in the dark at 30 °C for 22 h. The reactions were terminated by the addition of methanol to a final concentration of 50 % (v/v). Protein was removed by centrifugation (17,000 *g*, 30 min).

The supernatant was analyzed by reversed phase HPLC over an Ascentis C-18, (25×0.46 cm, 5 µm bead size) column using Shimadzu Providence system equipped with a LC-20AD pump. The solvent system consisted of methanol and 50 mM sodium phosphate, pH 3.0. Samples (20 µl) were injected with a SIL-20A autosampler and eluted with a gradient of 40−90 % methanol over 50 min at a flow rate of 0.5 mL/min. Eluted components were detected at 212 nm or 244 nm using an SPD-M20A diode array detector. Reference compounds (**6** and *syn*- or *anti*-**9a** and **9b**) were included to establish relevant chromatographic retention times and also allowed for direct analysis of the ratios of the diastereomers of **9a**−**d**. Identification of the product peaks of **8** was achieved from the similar elution profile to that of **9a** together with the characteristic UV absorbance maximum of the phenone moiety at ~244 nm. Obtained product amounts were determined from the correlation of HPLC peak areas of different standardized amounts of **9a** with 1 mM benzylalcohol as internal standard (Figure S11C). Quantification of aldol **6** was similarly performed after calibration against a standard curve of peak area @212 nm versus amounts of **6** (Figure S11D).

Blank reactions (without added aldolase) were treated identically, and the corresponding aldol peak area was subtracted from sample peak areas of the same HPLC run. For multiple runs the mean of individual samples was formed. Errors were estimated to be the quotient of standard deviation and mean of multiple runs of an example reaction or the standards, respectively. The addition of blank- and sample-error resulted in the total error. A total error of 12.8 % was estimated for reactions in the presence of **3** as ketone substrate, based on the typical errors of other reactions. Diastereomeric ratios were provided as quotients of the peak integrals of the *syn*- and *anti*-diastereomer.

Possible Catalytic Contributions From Buffer Components. The use of sodium phosphate buffer instead of bicine, both at pH 8.35, showed a similar product formation (1.31×10^6 mA×min for the phosphate buffer, 1.41×10^6 mA×min for the bicine-system) with similar ratios of formed *syn*- and *anti* diastereomers (*anti*:*syn*=1:3.2 and 1:3.4, respectively). This suggests that bicine, although containing a tertiary amine that could potentially contribute electrophilic catalysis via hydrogen bonding, is not affecting the diastereomeric ratio or amount of product formation.

Steady State Kinetics. Initial reaction velocities were determined in the presence of either 20 mM **4** and varying concentrations (0.4−4 mM) of aldehyde **5a** and in the presence of 5 µM FSA $VGGC_{107}C_{163}$, or in the presence of 10 mM **5a** and varied concentrations (1−12 mM) of **4**, in the presence of 10 µM FSA $VGGC_{107}C_{163}$, at 30 °C and in 40 mM bicine, pH 8.35. Reactions were set up in triplicates with aliquots removed at three time points and the reactions were stopped by mixing with 1 volume of methanol (containing 2 mM benzylalcohol as internal standard). The samples were centrifuged at 17,000 *g* for 30 min. Cleared reaction mixtures were subsequently separated by reversed phase HPLC, as described above. Product peaks of formed **9a** were integrated and normalized to the internal standard. Product amounts were calculated from calibration against a standard curve. (Figure S11C) Initial velocities were determined from linear regression of the three time points. Kinetic parameters were estimated following non-linear regression of the aldehyde concentration dependencies of the initial velocities fitting the Michaelis-Menten equation using program MMFIT in the SIMFIT package (www.simfit.org.uk) (Figures S11A and B).

Synthetic Reactions. (a) Synthesis of **7**: Reactions were set up in 40 mM of bicine, pH 8.35 with 2.5 % (v/v) acetonitrile by addition 12 mM of **5a**, 18 mM of **2** and 24 µM (0.26 µmol) of aldolase ($VGGC_{107}C_{163}$) in a total volume of 11 mL. (b) Synthesis of **9a**–**9d**: Reactions were set up in 58.5 mM of bicine, pH 8.35 with 2.5 % (v/v) acetonitrile by addition 10 mM (10 µmol) of each phenylacetaldehyde derivative (**5a**–**d**), 10 mM (10 µmol) ketone (**3** or **4**, respectively) and 200 µM (0.2 µmol) of aldolase ($VGGC_{107}V_{163}$ in the reactions with **3**, $VGGC_{107}C_{163}$ in the reactions with **4**) in a total volume of 1 mL.

Reactions were stirred in the dark at room temperature for 23 h. Reactions were stopped by the addition of methanol to a final concentration of 50 % (v/v) and protein was removed by centrifugation (17,000 *g*, 30 min) and filtration through a PVDF filter (cutoff 0.45 µm). The solvent was evaporated *in vacuo* at 50 °C, and the residual was dissolved in ethyl acetate or methanol.

Dissolved **9a** was filtered and concentrated under reduced pressure. The obtained white solid was purified by preparative TLC (CH_2Cl_2/MeOH: 100/2, Rf: 0.17) to give a white solid (0.9 mg, non-optimized yield: 35 %).

Product Analysis. Aldol products **9a–d** were purified by reversed phase HPLC and molecular masses and fractionation patterns were determined by LC-MS tandem mass spectrometry. A 15 µl aliquot of each analyte was separately injected and pumped to mass spectrometer on 1200 series HPLC (Agilent technologies, Santa Clara, CA, US) with a 200 µl/min isocratic flow rate of mobile phase, composed of water/acetonitrile 50/50. The fragmentations of analytes were achieved using a 6410 Triple-Quad mass spectrometer (Agilent technologies) equipped with an electrospray ionization interface (ESI) working in negative polarity mode. MS/MS acquisitions were performed in Product ion scan mode with the deprotonated molecule $[M-H]^-$. The ESI parameters, detailed collision energies (CE) and scanned mass range are presented in Table S5 and product ions are shown in Figure S12. Obtained data were processed and illustrated by MassHunter workstation (Data Acquisition and Qualitative) software packages (Agilent).

9a was further analyzed by ^{1}H-NMR (Figure S7). The ^{1}H-NMR spectrum of **9a** was compared with previously published data on the same compound.[S2,S6]

Stereoconfigurations were analyzed by chromatography and ^{1}H-NMR: diastereomeric ratios *anti*:*syn* were determined from the integrated peak areas following reversed phase chromatography after parallel runs of reference compounds.[S2] For compound **9a** the diasteromer ratios were also estimated from the H2-H3 integrals extracted from the ^{1}H-NMR spectrum. Enantiomeric ratios of aldols **9a**–**d** were determined from the integrated areas of assigned product peaks following chiral HPLC (Chiralpak AS-H) with a mobile phase of 25:75 2-propanol:*n*-hexane (Figure S3). Peak assignments were based on comparisons with reference compounds synthesized and characterized previously.[S2]

Protein Crystallization, Data Collection and Structure Refinement. The VG variant of FSA was crystallized by hanging drop vapor diffusion at 20 °C (293 K) (Figure S1). The drop consisted of 1 µl protein solution (11.3 mg/mL in 30 mM Tris-HCl pH 8.0, 12 % (v/v) glycerol) and 1 µl reservoir solution, and was equilibrated against 1 mL of the reservoir solution comprised of 1.7 M $(NH_4)_2SO_4$, 0.1 M bicine pH 8.75, and 150 mM NaCl. Crystallization was enhanced by introducing microcrystals of FSA VG into the drop by streak seeding. Co-crystallization with hydroxyacetone (compound **1**) yielded crystals from a drop comprised of 1.1 µl protein-ketone solution (10.2 mg/mL in 30 mM Tris-HCl pH 8.0, 10.8 % glycerol (v/v) and 20 mM ketone **1**) and 0.9 µl reservoir solution containing 1.8 M $(NH_4)_2SO_4$, 0.1 M bicine pH 8.75, 150 mM NaCl, streak-seeded with microcrystals of FSA VG and equilibrated against 1 mL of the reservoir solution. Crystals were flash frozen in liquid nitrogen and prior to X-ray diffraction analysis at the ESRF (Grenoble, France), beamline ID29.

Crystallographic data was automatically processed on site.[S7] The structure of the FSA VG variant was solved by molecular replacement using PHASER[S8] with the wild-type FSA subunit structure (pdb ID: 1L6W[S9]) as search model. Manual model building, addition of water molecules and ligand building was performed with COOT[S10], and alternated with restrained refinement using REFMAC5 until R-factors converged.[S11] A set of randomly selected reflections (~5 %) was set aside to monitor R_{free}. The final model contains 15 FSA VG subunits, each containing a covalent adduct of ketone **1** and glyceraldehyde with K85, and 177 water molecules. The structure has been deposited in the Protein Data Bank with the ID 7QXF. Data collection and refinement statistics are given in Table S6.

Parametrization of the Carbinolamine. GAFF parameters were produced for the carbinolamine, using acpype.[S13] The initial parameterization was done separately for the side chain, where the peptide backbone had been replaced with a methyl group, in an ideal geometry (Figure S13A). We opted for empirical

BCC charges because for biomolecular simulations they have proven to be of similar quality to the more demanding RESP-derived charges.[S14] The resulting topology was then merged with the parameters for lysine from the Amber ff99SB-ILDN,[S15] where lysine parameters were used for the backbone and the proximal part of the side chain up to and including the C_β atom (Figure S13B). In order to achieve a charge-neutral species, the charge of the two hydrogens attached to the C_γ atom were each reduced by 0.0175 *e*, accounting for minor differences in the charge distribution along the molecule/residue in the original force field and the GAFF parameters. Bonded parameters for interactions spanning the "joint" between the lysine and carbinol intermediate parts were taken from the original lysine parameters if at least half of the atoms involved were originally lysine atoms, otherwise from the GAFF parameters. Since the joint connects two chemically similar parts of the aliphatic chains, we expect our choice of bridging parameters to have a negligible effect on the simulations. In order to enable 5-fs timesteps in the simulations we also used MkVsites[S16] to obtain angle constraints for the alcohols in the carbinolamine. Since no primary amines or alkyl groups using the GAFF atom types were present, no new virtual site parameters were needed.

All new parameters were merged into the Amber ff99SB-ILDN force field files, which were used for the subsequent simulations. The force field files are available for download (see Supplementary Data). Notably, because of the symmetry in the functional form of the parameters, the (*R*)- and (*S*)-enantiomer with respect to the stereocenter at the X atom will be determined by the conformation, hence the same parameters can be used regardless of enantiomer.

Creating Structure Models. Structures for molecular dynamics simulations were created with PyMOL 2.5.2.[S12] The carbinolamine intermediate was created with the Builder tool, and other amino acid mutations were made with the Mutagenesis Wizard. The Sculpting Wizard was subsequently used to relieve strains in and around modified residues.

Molecular Dynamics Simulations. Molecular dynamics simulations were performed with the Gromacs simulation package, version 2021.3.[S17] The decameric proteins were solvated in TIP3P water with 0.154 M NaCl, and modelled with the Amber ff99SB-ILDN force field[S18] with parameters added for the carbinolamine intermediate. Equilibration was carried out in five steps, beginning with steepest-descent energy minimization, followed by a 100-ps *NVT* simulation with position restraints on the heavy atoms of the protein, a 1-ns *NVT* simulation without position restraints, a 1-ns *NPT* simulation employing the Berendsen barostat[S19] set to 1 atm, and finally an *NVT* simulation of 10 ns. In the production simulations that followed, Parrinello-Rahman[S20] was used for pressure coupling. All simulations had the temperature set to 300 K with the velocity rescaling thermostat by Bussi *et al*..[S21] Hydrogens were simulated as virtual sites,[S22] and the LINCS algorithm[S23] with 6th order expansion was used to constrain the length of all bonds. The fast smooth Particle-Mesh Ewald method[S24] with a 0.12-nm Fourier spacing was used to calculate the electrostatic interactions, and the cut-off for non-bonded interactions was set to 1.0 nm with the Verlet scheme. Each model was simulated for 200 ns with a 5-fs time step, and system sizes ranged between 216 000 and 235 000 atoms.

Analyses of Simulation Trajectories. Cluster analyses of the molecular dynamics trajectories were performed (Figure S14) with *gmx cluster* from the Gromacs simulation package, using the gromos clustering algorithm[S25] with a 0.075-nm cut-off. The analyses considered the residues Q59, T109, Y131, and the carbinolamine together with K85, excluding hydrogen atoms. Only the final 100 ns of the simulations were included in the cluster analysis, which amounted to 1001 frames and hence 10010 structural conformations over the ten protein chains. Inter-atomic distances were measured with MDAnalysis 2.0.0.[S26]

Supporting Tables

Table S1. Formation of aldol **6** – integrated peak areas

Codon			**6** Peak area	
L107	L163	sample code	@212 nm (mA×min)	error
A	C	average	5091730	981837
		B03	5231506	
		D09	6974267	
		F12	3069417	
A	C	T185N	434538	86908
A	D	F11	741060	142899
A	E	average	1993713	384447
		C06	1163360	
		D05	2824066	
A	G	average	53464	10310
		A03 03.26.	10662	
		C05 04.09.	-42312	
		D12	-401281	
		G03	197742	
		G11	502511	
A	I	A10	11548927	2226978
A	K	average	3465399	668232
		E15	27895212	
		F03	4141277	
A	L	A11	12557549	2421470
A	M	A08	11163998	2152752
A	S	D13	1359839	262218
A	V	average	9303299	1793954
		B14	10556056	
		F04	8050542	
A	W	F15	827228	159514
A	Y	F07	1608354	310139
C	A	C04	5541986	1068660
C	C	average	5987343	1154538
		C11	8090846	
		F08	3883840	
C	D	C14	691651	133371
C	E	F10	5333355	1028430
C	G	average	-14380	-2773
		E05	522114	
		G02	-550875	
C	R	B10	-657349	-126757
C	V	average	9166422	1767560
		E02	9210056	
		G07	9122789	
C	W	B02	-64131	-12366
D	D	E06	-63876	-12317
E	E	F05	713079	137503
F	C	average	10978257	2116936
		F14	11435845	
		G05	10520670	
F	E	D02	390113	75225
F	G	G09	10853087	2092799
F	V	average	9149457	1764288
		D04	11162208	
		D06	9975714	
		E11	8383068	
		G14	7076839	
G	C	G12	6758743	1303287
G	G	B11	1268701	244643
G	K	E13	3732150	719670
H	L	D03	10132643	1953876
L	L	average	10201738	1967199
		D01	11287551	
		E01	9275848	
		F01	10041815	
M	G	average	11234151	2166279
		D08	11057952	
		D10	11410349	
M	K	F13	10182974	1963581
N	C	C03	9754808	1881018
Q	G	C10	11409686	2200128
S	G	E08	480478	92650
S	W	B08	3293307	635048
V	C	average	7711102	1486930
		G06	6314118	
		G13	9108086	
V	G	E14	-68831	-13273
V	V	A05	11693900	2254932
V	V	P187Q	1120157	224103
V	W	average	-389013	-75013
		C09	-680946	
		D15	-97081	
W	C	E10	1224297	236081
W	G	G01	10450170	2015104
W	I	F06	3280541.0	632586.2
W	M	C12	-561963	-108363
Y	G	E07	9818764	1893351
Y	M	B04	6870124	1324765
Y	Q	B07	7162028	1381053

Codon	sample code	**6** Peak area	error

Table S2. Formation of aldol **9a** – integrated peak Areas and catalyzed reaction rates

Codon			*anti*-**9a**				*syn*-**9a**				
L107	L163	sample code	Peak area @244 nm (mA×min)	error	rate (h^{-1})[a]	error	Peak area @244 nm (mA×min)	error	rate (h^{-1})	error	dr **9a** *anti*:*syn* = 1:x
A	C	average	142650	18066	0.11	0.01	988577	127063	0.74	0.1	6.9
		B03	139605				1376805				
		D09	131870				571570				
		F12	156476				1017356				
A	C	T185N	35979	4605	0.027	0.003	168167	21525	0.13	0.02	4.7
A	D	F11	72385	9167	0.054	0.007	248163	31897	0.19	0.02	3.4
A	E	average	80252	10164	0.060	0.008	22957	2951	0.017	0.002	0.3
		C06	31035				99971				
		D05	129469				-54057				
A	G	average	27061	3427	0.020	0.003	25705	3304	0.019	0.002	0.9
		A03 03.26.	29681				82369				
		C05 04.09.	42416				105762				
		D12	6019				-72015				
		G03	9716				-45550				
		G11	47474				57959				
A	I	A10	117020	14820	0.088	0.01	899273	115584	0.67	0.09	7.7
A	K	average	65980	8356	0.050	0.006	278403	35783	0.21	0.03	4.2
		E15	52031				211461				
		F03	79929				345346				
A	L	A11	105662	13382	0.079	0.01	656514	84382	0.49	0.06	6.2
A	M	A08	96466	12217	0.072	0.009	596568	76677	0.45	0.06	6.2
A	S	D13	-266236	-33718	0	0	-805541	-103537	0	0	n.d.
A	V	average	109733	13897	0.082	0.01	568230	73035	0.43	0.02	5.2
		B14	125198				624208				
		F04	94269				512253				
A	W	F15	100792	12765	0.076	0.01	314198	40384	0.24	0.03	3.1
A	Y	F07	123824	15682	0.093	0.01	483650	62164	0.36	0.05	3.9
C	A	C04	-44955	-5693	0	0	73663	9468	0.055	0.007	n.d.
C	C	average	103414	13097	0.078	0.01	2381835	306139	1.8	0.2	23
		C11	75766				3109238				
		F08	131062				1654433				
C	D	C14	44453	5630	0.033	0.004	219768	28247	0.16	0.02	4.9
C	E	F10	98198	12437	0.074	0.009	448142	57600	0.34	0.04	4.6
C	G	average	48763	6176	0.037	0.005	144659	18593	0.11	0.01	3.0
		E05	24942				101005				
		G02	72583				188313				
C	R	B10	154882	19615	0.12	0.01	529441	68049	0.40	0.05	3.4
C	V	average	61679	7811	0.046	0.006	679292	8731	0.51	0.07	11
		E02	75273				730181				
		G07	48084				628404				
C	W	B02	80661	10216	0.061	0.008	229949	29556	0.17	0.02	2.9
D	D	E06	-8945	-1133	0	0	-23510	-3022	0	0	n.d.
E	E	F05	96802	12260	0.073	0.009	400908	51529	0.30	0.04	4.1
F	C	average	2652	336	0.014	0.002	-82976	-10665	0	0	n.d.

		F14	-13510				-139043				
		G05	18814				-26910				
F	E	D02	98252	12443	0.074	0.009	-14766	-1898	0	0	n.d.
F	G	G09	177318	22457	0.13	0.02	87023	11185	0.065	0.008	0.5
F	V	average	18215	2307	0.014	0.002	-225906	-29036	0	0	n.d.
		D04	-42855				-521390				
		D06	45129				-444589				
		E11	50444				119662				
		G14	20143				-57305				
G	C	G12	87201	11044	0.065	0.008	486919	62584	0.37	0.05	5.6
G	G	B11	107067	13560	0.080	0.01	323887	41629	0.24	0.03	3.0
G	K	E13	1229	156	0.0009	0.0001	-49597	-6375	0	0	n.d.
H	L	D03	13829	1751	0.010	0.001	-127704	-16414	0	0	n.d.
L	L	average	38355	4858	0.029	0.004	177911	22867	0.13	0.02	4.6
		D01	-15545				-28886				
		E01	37632				176365				
		F01	92978				386255				
M	G	average	154361	19549	0.12	0.01	84414	10850	0.063	0.008	0.5
		D08	145615				158544				
		D10	163108				10285				
M	K	F13	81906	10373	0.061	0.008	283839	36482	0.21	0.03	3.5
N	C	C03	28803	3648	0.022	0.003	888051	114142	0.67	0.09	31
Q	G	C10	67788	8585	0.051	0.006	252109	32404	0.19	0.02	3.7
S	G	E08	55859	7074	0.042	0.005	193175	24829	0.14	0.02	3.5
S	W	B08	203864	25819	0.15	0.02	933934	120039	0.70	0.09	4.6
V	C	average	49161	6226	0.037	0.005	2800052	359893	2.1	0.3	57
		G06	60376				3234388				
		G13	37946				2365716				
V	G	E14	266008	33689	0.20	0.03	202166	25984	0.15	0.02	0.8
V	V	A05	9441	1196	0.0071	0.001	444171	57090	0.33	0.04	47
V	V	P187Q	231007	29569	0.17	0.02	413808	52967	0.31	0.04	1.8
V	W	average	244838	31008	0.18	0.02	455937	58602	0.34	0.04	1.9
		C09	200485				710207				
		D15	289190				201668				
W	C	E10	49642	6287	0.037	0.005	150329	19322	0.11	0.01	3.0
W	G	G01	-3795	-481	0	0	-79387	-10204	0	0	n.d.
W	I	F06	58109	7359	0.044	0.006	197020	25323	0.15	0.02	3.4
W	M	C12	103352	13089	0.078	0.01	283889	36488	0.21	0.03	2.7
Y	G	E07	4448	563	0.0033	0.0004	20322	2612	0.015	0.002	4.6
Y	M	B04	282637	35795	0.21	0.03	455764	58580	0.34	0.04	1.6
Y	Q	B07	229092	29014	0.17	0.02	817935	105130	0.61	0.08	3.6

[a] Reaction rates were calculated from rate (h^{-1}) = (integrated product peak area) /(605759 × reaction time × 0.1), where 605759 is the peak area corresponding to 1 nmol of product (**Figure S11C**) and 0.1 is the amount of enzyme, in nmol, resulting in the obtained product peaks.

Table S3. Steady state kinetic parameters of the $VGGC_{107}C_{163}$ catalyzed formation of **9a**

Varied substrate	k_{cat}^{app} (min^{-1})	K_M (mM)	k_{cat}/K_m^{app} ($min^{-1}\times M^{-1}$)
5a[a]	0.021±0.004 (102±12)[c]	3.7±1 (2.7±1)[c]	5.8±2
4[b]	0.067±0.008	17±3	4.2±0.3

[a] Determined in the presence of 20 mM (1.2×K_M) **4** and varying concentrations of **5a**. [b] Determined in the presence of 10 mM (2.7×K_M) **5a** and varying concentrations of **4**. [c] Values in brackets refer to the FSA VGG catalyzed aldol reaction with saturating concentration of **1** and varied concentrations of **5a**. Data adapted from ref. S3. Fits of the Michaelis-Menten equation to the substrate concentration dependency of initial velocities are shown in Figures S11A and B.

Table S4. Sequences of deoxyoligonucleotides used in gene library construction

Name	**Sequence 5´→3´**
A129G_Forward	GCGCTGGCAGGTGCGGAATATGTTGGGCCTTACGTTAATGTTATTGATGCT
A129G mutamer 2	CCCTGAGCATCAATAACATTAACGTAAGGCCCAACATATTCCGCACCTGCCAGC
pUC_Forward	CCCAGTCACGACGTTGTAAAACG
VGMut_107_Reverse	CGTCGGAATCCCTTCCGC
VGMut_107_Forward_NDT	GCGGAAGGGATTCCGACG**NDT**GGAACCGCGGTATATGGCG[a,b]
VGMut_107_Forward_VMA	GCGGAAGGGATTCCGACG**VMA**GGAACCGCGGTATATGGCG
VGMut_107_Forward_ATG	GCGGAAGGGATTCCGACG**ATG**GGAACCGCGGTATATGGCG
VGMut_107_Forward_TGG	GCGGAAGGGATTCCGACG**TGG**GGAACCGCGGTATATGGCG
VGMut_163_Reverse	CACTTTCGCCTGCGGCG
VGMut_163_Forward_NDT	CGCCGCAGGCGAAAGTG**NDT**GCAGCGGGTTTCAAAACCC
VGMut_163_Forward_VMA	CGCCGCAGGCGAAAGTG**VMA**GCAGCGGGTTTCAAAACCC
VGMut_163_Forward_ATG	CGCCGCAGGCGAAAGTG**ATG**GCAGCGGGTTTCAAAACCC
VGMut_163_Forward_TGG	CGCCGCAGGCGAAAGTG**TGG**GCAGCGGGTTTCAAAACCC
pUC_Reverse	AGCGGATAACAATTTCACACAGG

[a] Mutated positions are written in bold and underlined. [b] N = A/C/G/T; D = A/G/T; V = A/C/G; M = A/C.

Table S5. ESI-MS/MS parameters for target analytes

		MS/MS setting for target analytes			
Parameters		Analytes (aldol)	Parent ion $(M-H)^-$	Scanned range (m/z)	CE (V)
Polarity	Negative	**9a**	255.2	50-260	5, 10, 15, 20
Spray voltage (V)	3000	**9b**	285.2	50-290	5, 10, 15
Desolvation temperature (°C)	350	**9c**	285.2	50-290	5, 10
Desolvation gas flow (l/min)	10	**9d**	315.1	50-320	5, 8
Nebulizer gas (psi)	30	**7**	207.1	50-210	5, 10, 15, 20
Dwell time (ms)	200				

Table S6. Data collection and refinement statistics FSA R134V/S166G

Data Collection		**Refinement**	
Space group	C2	No. of reflection in work set	107028
a, *b*, *c* (Å)	182.2, 176.9, 120.1	No. of reflection in free set	5735
α, β, γ (°)	90, 91.5, 90	R_{work}	0.2195
Molecules in a. u.	15	R_{free}	0.2621
Wavelength (Å)	1.0722	No. of non-hydrogen atoms	
Resolution (Å)	49.67-2.62 (2.71-2.62)[a]	Protein	24334
Total reflections	383607 (33915)	Solvent	177
Unique reflections	112831 (10827)	Average B factor (Å^2)	
Multiplicity	3.4 (3.1)	All atoms	64.1
Completeness (%)	99.1 (96.8)	Solvent	40.8
< I/σ(*I*) >	7.2 (1.0)	r.m.s.d. bonds (Å)	0.0074
Wilson B factor (Å^2)	64.3	r.m.s.d. angles (°)	1.529
R_{merge}	0.121 (0.866)	Residues in Ramachandran plot regions (%)	
R_{meas}	0.144 (1.043)	favored	97.7
R_{pim}	0.077 (0.575)	allowed	2.3
$CC_{1/2}$	0.993 (0.650)	outliers	0.0
		PDB-ID	7QXF

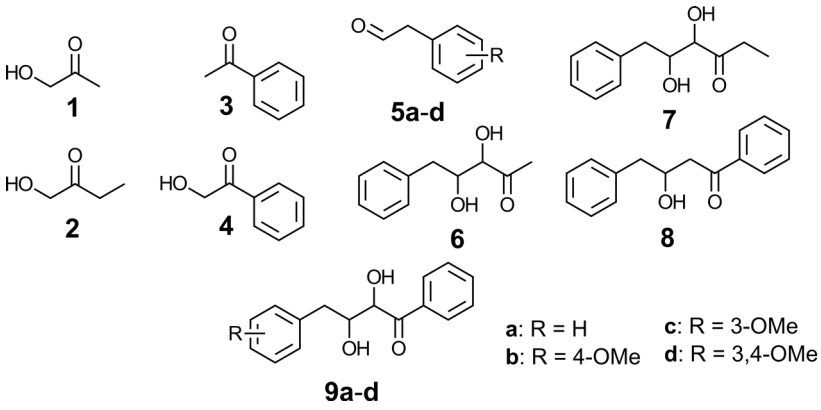
[a] Values in parentheses are for the highest resolution shell.

Table S7. Distances between the sulfurs of C107 and C163 in the VGGC$_{107}$C$_{163}$ variant, during 200-ns of molecular dynamics simulations

	0-200 ns		100-200 ns	
Enantiomer	Average (Å)	Minimum (Å)	Average (Å)	Minimum (Å)
(*R*)	5.76	3.29	5.68	3.34
(*S*)	6.21	3.21	6.25	3.32

Supporting Figures

Chart S1. Compounds studied



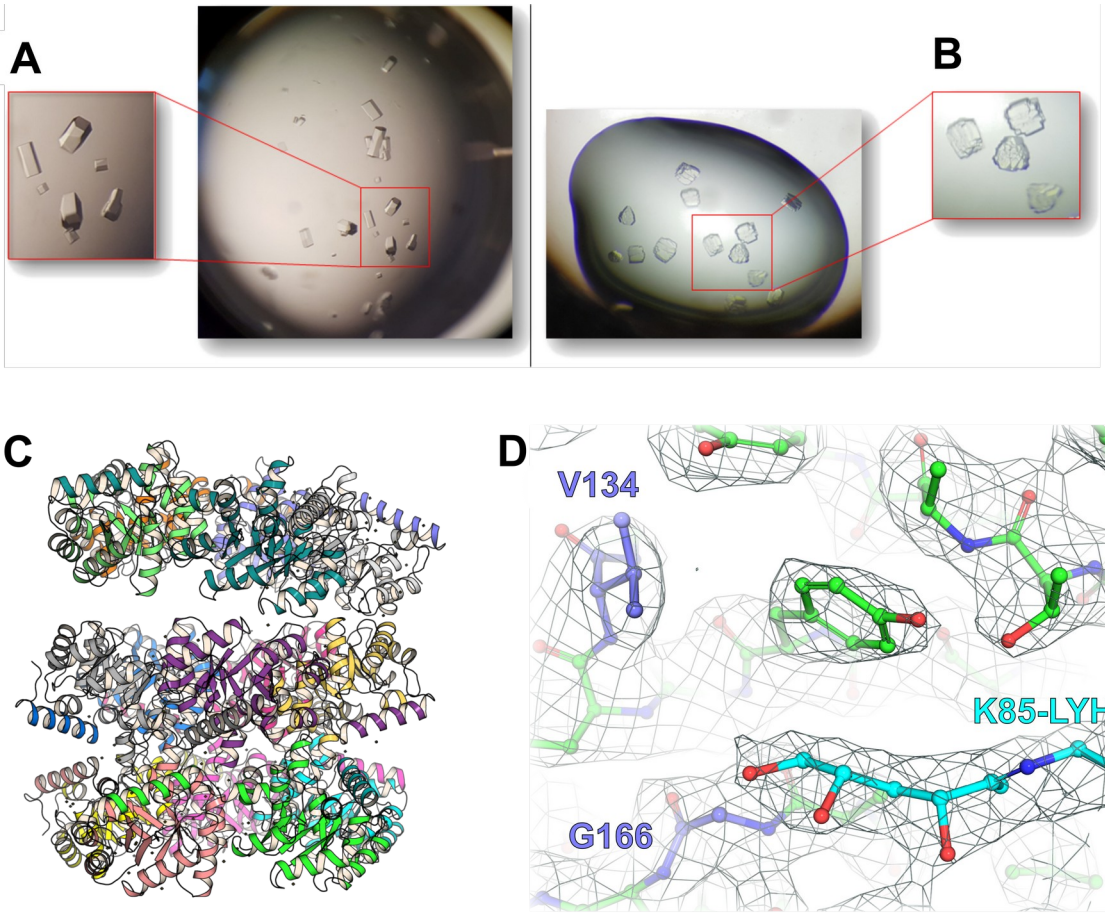


Figure S1. X-ray crystallography of the FSA VG variant. (**A**) and (**B**) Crystals of FSA VG obtained in absence (**A**) and presence of 20 mM **1** (**B**) upon seeding with FSA VG microcrystals. (**C**) Packing of the 15 subunits of the VG variant in the crystal asymmetric unit, forming one complete homodecamer as well as one pentameric ring ofanother that is assembled by crystallographic symmetry. Coloring is by subunit chain. (**D**) Close-up of the active site of VG, chain A. Clear electron density for a covalent adduct at K85 was observed and interpreted as belonging to the imine of (3*S*,5*R*,6)-tri-hydroxyhexan-2-one (LYH). This was based on the rationale that the protein had been co-crystallized with ketone **1** which reacted with K85. This original adduct reacted then further with glyceraldehyde formed from glycerol, which was abundant in the crystallization liquor. The mutated residues R134V and S166G are colored in purple. The $2F_oF_c$ map is contoured at 1σ. Figures **C** and **D** were constructed in PyMol ver. 2.5[S12] from the atomic coordinates in PDB entry 7QXF.

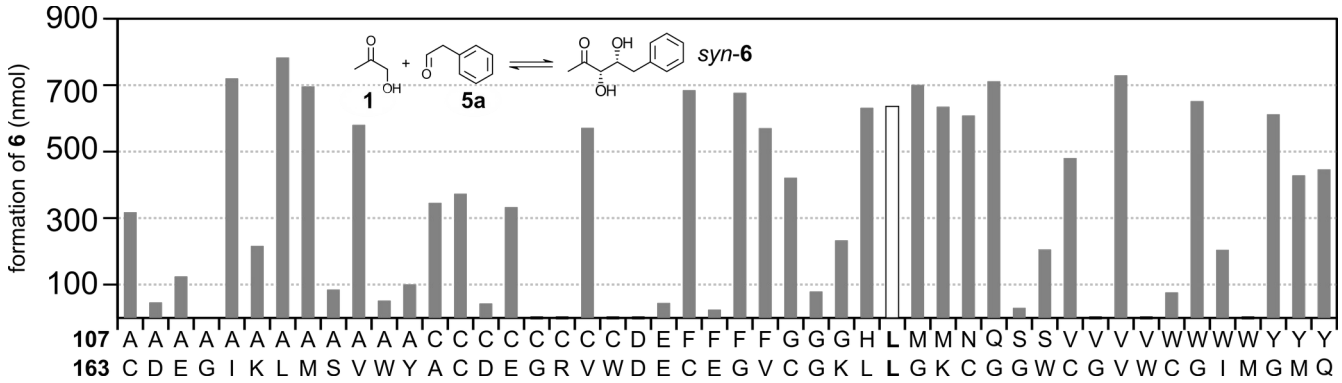


Figure S2. Formation of aldol **6** following 22 h reaction in the presence of 10 μM (2 nmol) of the different FSA $VGGX_{107}X_{163}$ variants. Bars represent integrated peak areas after reversed phase HPLC separation and calibration against a standard curve (Figure S11D). The value for the parent enzyme $VGGL_{107}L_{163}$ is shown as a white bar. The standard deviation of the values are ±13 %.

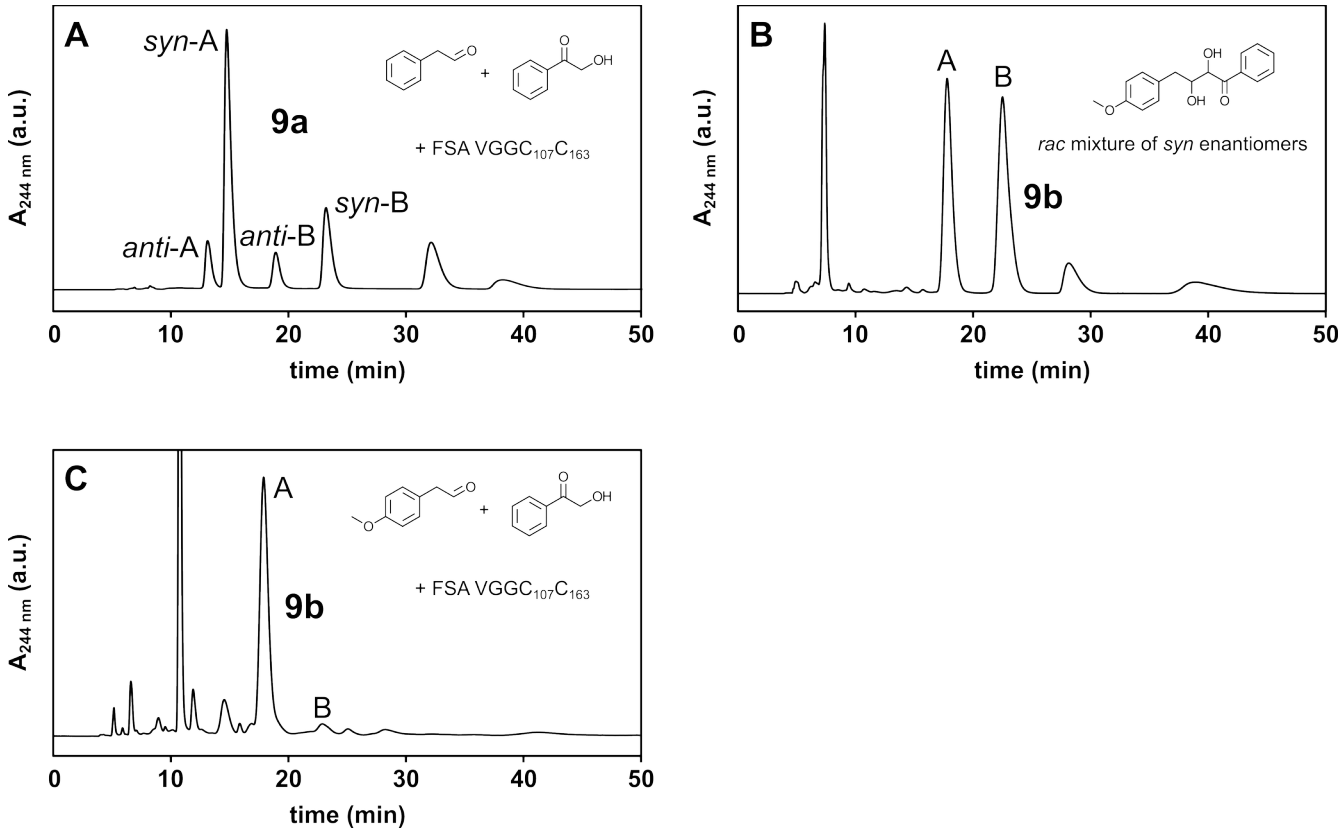


Figure S3. Chiral separation of aldols and FSA-containing reaction mixtures. Samples were chromatographed through a Chiralpak AS-H column with a mobile phase consisting of 25:75 2-propanol:*n*-hexane. (**A**) Product mixture after 23 h incubation of 40 mM (40 μmol) **4** and 20 mM (20 μmol) **5a** and 200 μM (0.2 μmol) $VGGC_{107}C_{163}$ at 30 °C. The *syn* and *anti* diastereomers were purified by preparative TLC and subsequently analyzed by chiral HPLC for resolution of individual enantiomers. (**B**) Reference aldol, *syn* diastereomer of **9b** synthesized using SmI_2.[S2] (**C**) Crude reaction mixture after 23 h incubation of 10 mM (10 μmol) **4** and 10 mM (10 μmol) **5b**, in the presence of 200 μM (0.2 μmol) FSA $VGGC_{107}C_{163}$ at room temperature. Peak A in (**C**) fully agrees with Peak A in the racemic mixture of *syn*-**9b** in (**B**) regarding retention time in the chromatography as well as in λ_{max}.

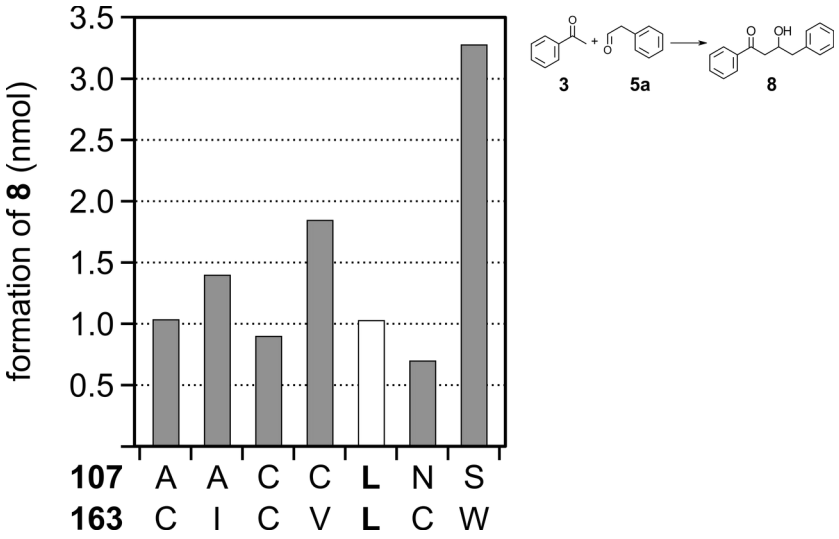


Figure S4. Formation of aldol **8** in the presence of 10 μM (2 nmol) selected aldolase variants with acetophenone (**3**) as ketone donor. The parent enzyme $VGGL_{107}L_{163}$ is shown as a white bar. The standard deviation of the values are ±13 %.

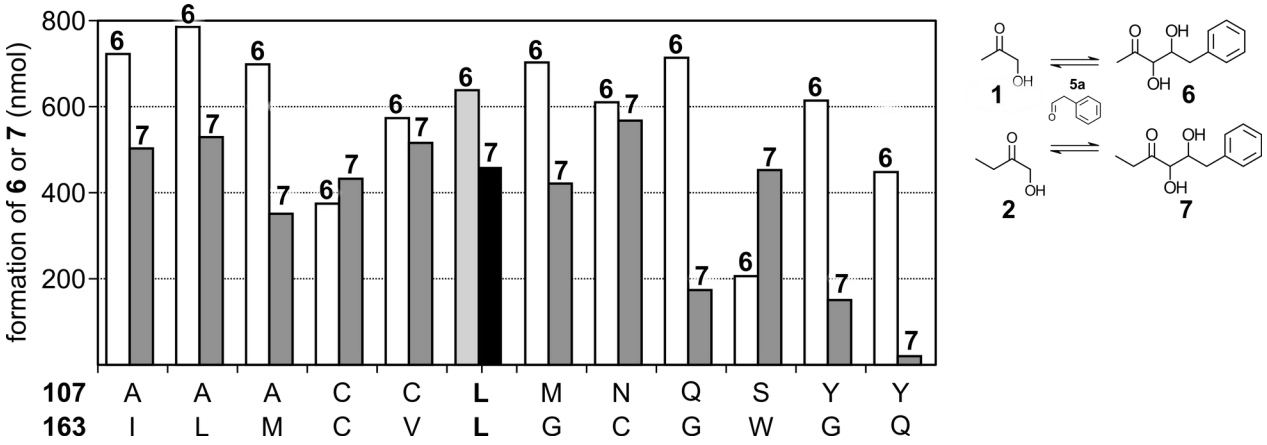


Figure S5. Formation of aldol **7** (gray bars) as ketone substrate of selected $VGGX_{107}X_{163}$ variants, as compared to aldol **6** (white bars) from reactions with either **2** or **1**, respectively together with aldehyde **5a**. The $VGGL_{107}L_{163}$ parent is included with light gray (**6**) or black (**7**) bars.

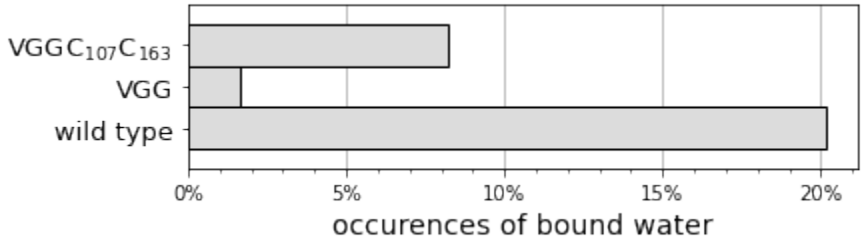


Figure S6. Frequency of water being present in a pocket defined as within 4 Å of the side-chain carbonyl oxygen of Q59, and the hydroxyl oxygens of Y131 and T109, during the final 100 ns of simulation.

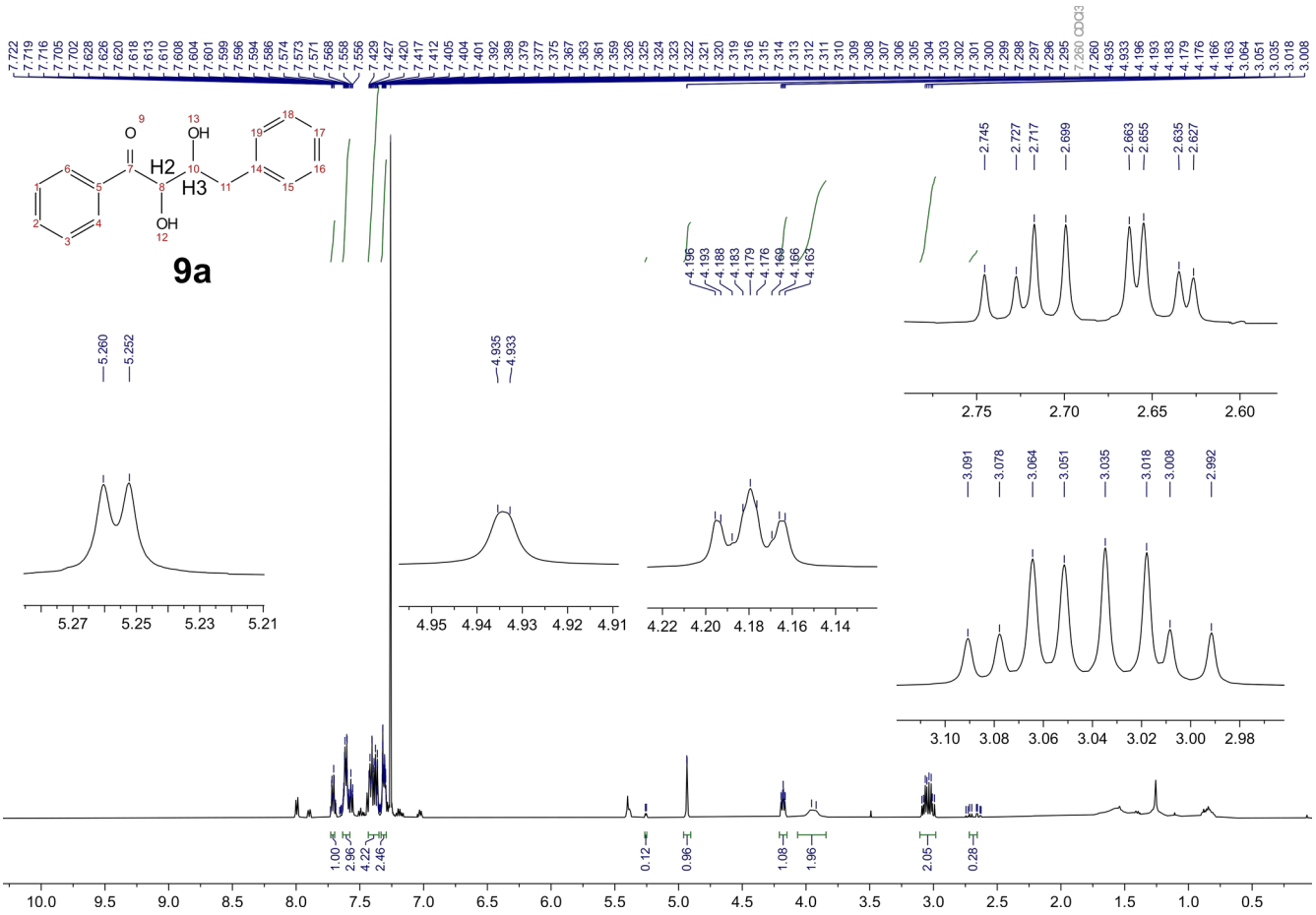


Figure S7. ^{1}H-NMR (500 MHz, $CDCl_3$) δ 7.73-7.69 (m, f1H), 7.65-7.54 (m, 3H), 7.43-7.34 (m, 4H), 7.34-7.29 (m, 2H), 5.26 (d, *J* = 4.1 Hz, 0.12H, *anti*-isomer), 4.93 (d, *J* = 1.3 Hz, 1H), 4.18 (ddd, *J* = 8.0, 6.5, 1.3 Hz, 1H), 4.05–3.85 (br, 2H, OH), 3.07, 3.01 (dABq, *J* = 6.5 Hz, J_{AB} = 13.2 Hz, 2H), 2.72, 2.65 (dAbq, *J* = 9.0 Hz, *J* = 4.1 Hz, J_{AB} = 14.0 Hz, 0.24H, *anti*-isomer). For clarity, only aliphatic protons of the *anti*-isomer are listed.

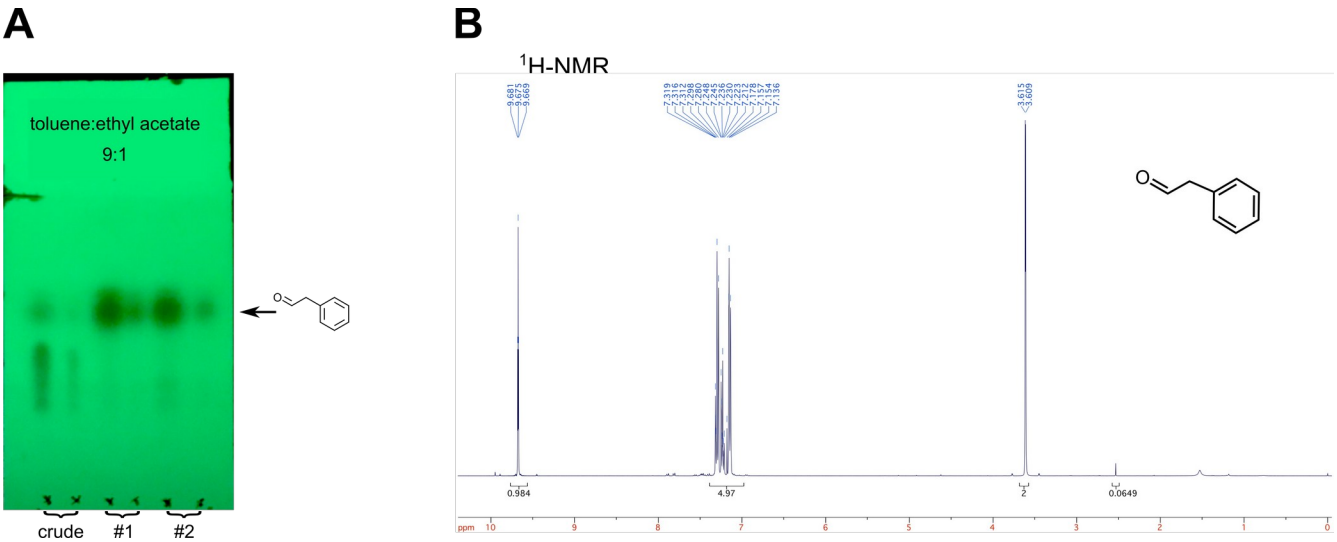


Figure S8. Analysis of distillation-purified phenyl acetaldehyde. (**A**) TLC, "crude", starting material, "#1" and "#2", purified fractions. (**B**) ^{1}H-NMR of fraction #2.

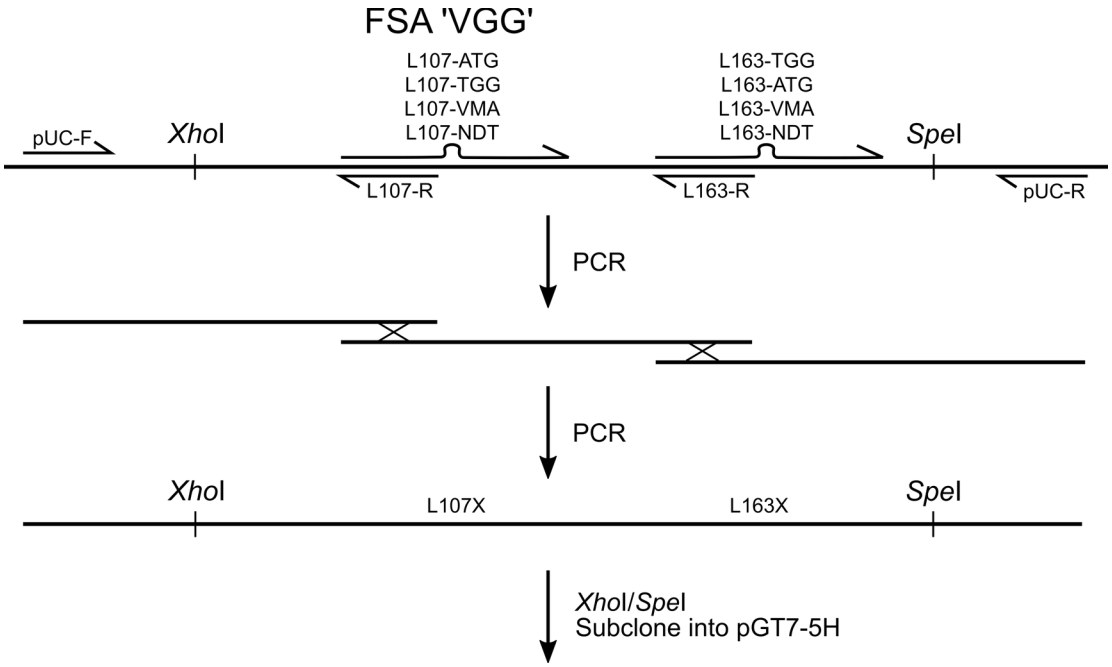


Figure S9. Scheme describing the stepwise assembly of PCR-generated gene fragments encoding the FSA VGG$X_{107}X_{163}$ enzyme library.

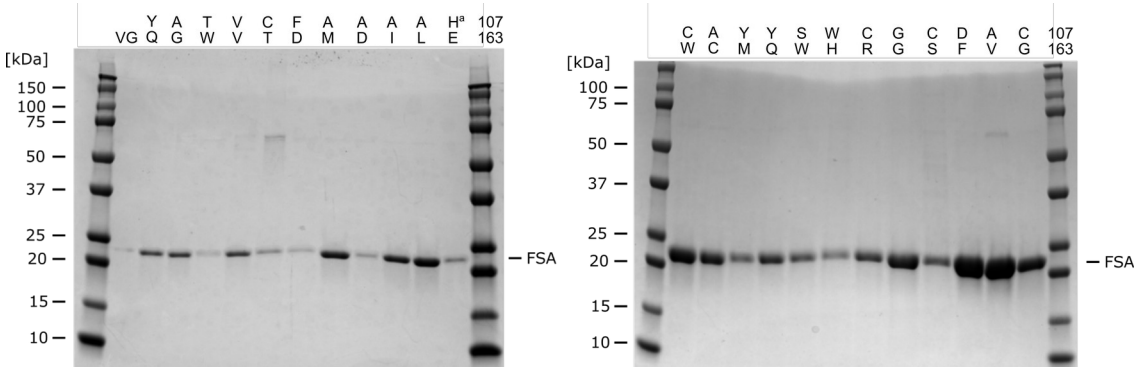


Figure S10 SDS-PAGE of purified FSA VGG variants. Samples are desalted proteins following Ni-IMAC purification as described in the Experimental Details section.

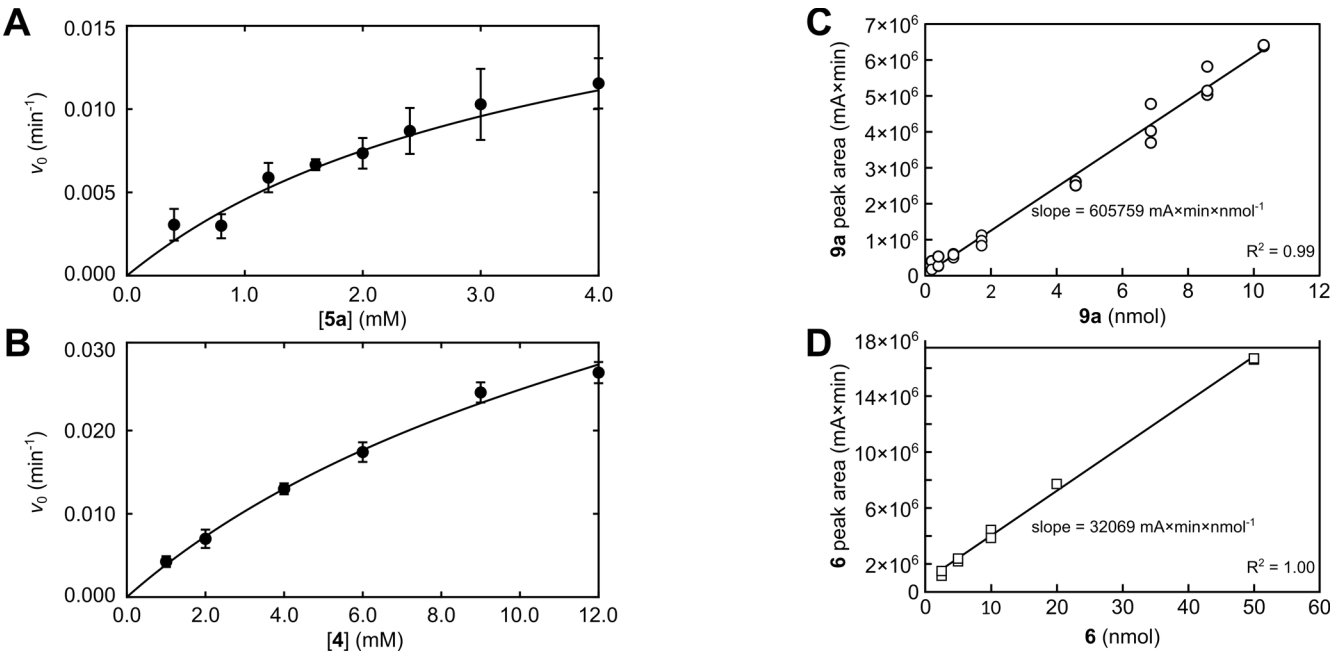


Figure S11. (**A**) and (**B**) Steady state reaction velocities of formation of **9a** in the presence of (**A**), 5 μM FSA $VGGC_{107}C_{163}$ and 20 mM **4**, and varying concentrations of **5a**, and (**B**), 10 μM FSA $VGGC_{107}C_{163}$ and 10 mM **5a** and varying concentrations of **4**. Error bars represent standard deviations of three parallel reactions. Solid lines represent non-linear regression fits of the Michaelis-Menten equation. See the Experimental sections for details. (**C**) Standard curve of peak areas following reversed phase HPLC of aldol **9a**. (**D**) Standard curve of peak areas following reversed phase HPLC of aldol **6**.

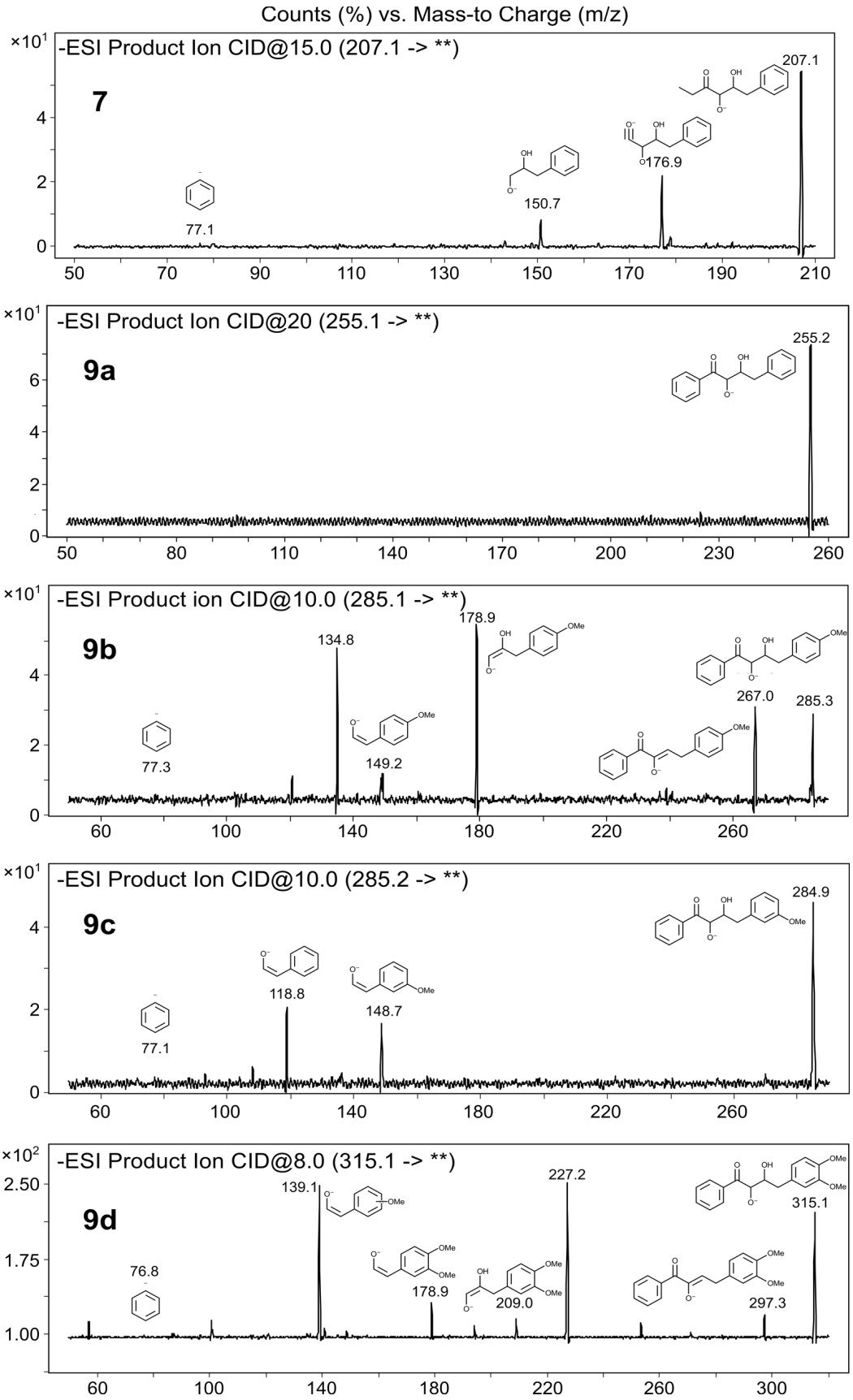


Figure S12. ESI MS/MS product ions of aldols **7** and **9a**-**9d**.

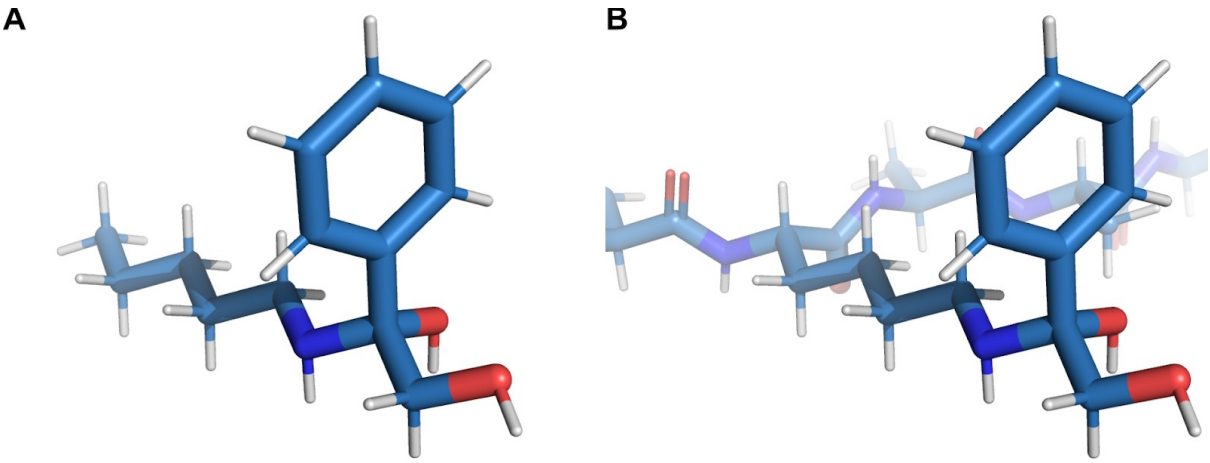


Figure S13. Parameterization compounds. (**A**) The molecule used for initial parameterization with ACPYPE. (**B**) Illustration of the complete carbinol intermediate incorporated into a peptide backbone.

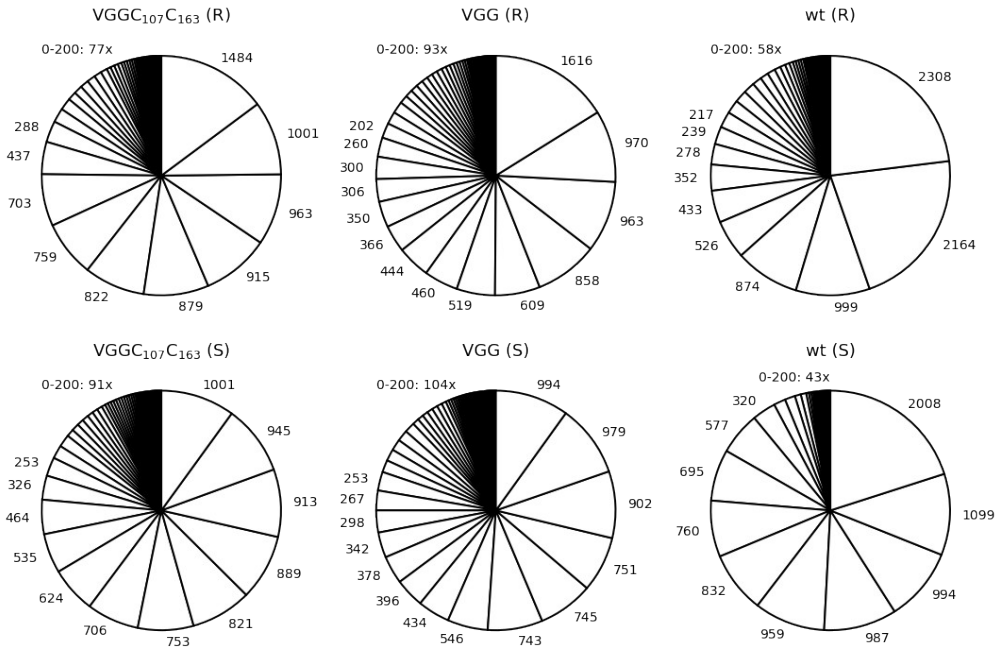


Figure S14. Number of structures per cluster that resulted from cluster analyses of the simulation trajectories.

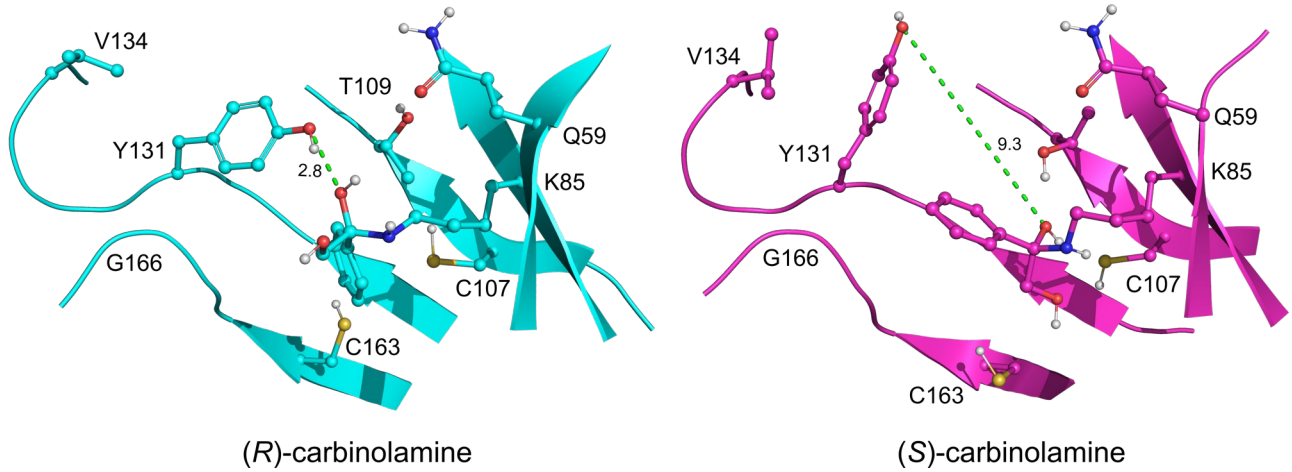


Figure S15. Comparison of the highest populated structure clusters following molecular dynamics simulations, modeling either the (*R*)-carbinolamine (left) or (*S*)-carbinolamine (right) formed after reaction between ketone **4** and the active-site K85. In the (*S*)-carbinolamine the leaving group hydroxyl is pointing away from the proposed catalytic acid Y131 and the phenyl substituent of the adduct is not occupying the cleft formed between cysteines 107 and 163.

Supporting References

S1 D. Al-Smadi, T. R. Enugala, T. Norberg, J. Kihlberg and M. Widersten, *Synlett* 2018, **29**, 1187-1190.

S2 D. Al-Smadi, T. R. Enugala, V. Kessler, A. R. Mhasal, S. C L. Kamerlin, J. Kihlberg, T. Norberg, and M Widersten, *J. Org. Chem.* 2019, **84**, 6982-6991.

S3 H. Ma, S. Engel, T. R. Enugala, D. Al-Smadi, C. Gautier and M.; Widersten, *Biochemistry* 2018, **57**, 5877-5885.

S4 L. Tang, H. Gao, X. Zhu, X. Wang and M. Zhou, *BioTechniques* 2012, **52**, 149-157.

S5 G. E. Dale, H.-J. Schönfeld, H. Langen and M. Stieger, *Prot. Engin.* 1994, **7**, 925-931.

S6 N. Miyoshi, S. Takeuchi and Y. Ohgo, Y. *Chem. Lett.* 1993, 2129-2132.

S7 S. Monaco, E. Gordon, M. W. Bowler, S. Delageniere, M. Guijarro, D. Spruce, O, Svensson, S. M. McSweeney, A. A. McCarthy, G. Leonard and M. H. Nanao, *J. Appl. Cryst.* 2013, **46**, 804-810.

S8 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Cryst.* 2007, **40**, 658-674.

S9 S. Thorell, M. Schürmann, G. A. Sprenger and G. Schneider, *J. Mol. Biol.* 2002, **319**, 161-171.

S10 P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Cryst.* 2010, **D66**, 486-501.

S11 G. N. Murshudov, A. A. Vagin and E. J. Dodson, *Acta Cryst.* 1997, **D53**, 240-255.

S12 The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

S13 A. W. Sousa da Silva and W. F. Vranken, *BMC Res. Notes* 2012, **5**, 367.

S14 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins* 2010, **78**, 1950-1958.

S15 P. G. A. Aronica, S. J. Fox and C. S. Verma, *ACS Omega* 2018, **3**, 4664-4673.

S16 P. Larsson, R. C. Kneiszl and E. G. Marklund, *J. Comp. Chem.* 2020, **41**, 1564-1569.

S17 M. J. Abraham, R. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX* 2015, **1-2**, 19-25.

S18 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins Struct. Funct. Bioinforma.* 2010, **78**, 1950-1958.

S19 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.* 1984, **81**, 3684-3690.

S20 M. Parrinello and A. Rahman, *J. Appl. Phys.* 1981, **52**, 7182-7190.

S21 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.* 2007, **126**, 014101.

S22 K. A. Feenstra, B. Hess and J. C. Berendsen, *J. Comput. Chem.* 1999, **20**, 786-798.

S23 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.* 1997, **18**, 1463-1472.

S24 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.* 1995, **103**, 8577-8593.

S25 X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *Angew. Chemie - Int. Ed.* 1999, **38**, 236-240.

S26 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *J. Comput. Chem.* 2011, **32**, 2319-2327.