Supporting Information

Structural basis for branched substrate selectivity in a ketoreductase from *Ascaris suum*

Hongjun Dong[1,4] and Michelle C. Y. Chang[1,2,3]*

[1]*Department of Chemistry, University of California, Berkeley, Berkeley, CA 94720-1460.*
[2]*Department of Molecular & Cell Biology, University of California, Berkeley, CA 94720-3200.*
[3]*Department of Chemical & Biomolecular Engineering, University of California, Berkeley, CA 94720-1462.*
[4]*Current Institution: Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, 300308, People's Republic of China*

[*] *To whom correspondence should be addressed: mcchang@berkeley.edu*

Materials and Methods

Supplementary Data

References

Materials and Methods

Commercial materials. Luria-Bertani (LB) broth/agar, Terrific Broth (TB), and chloramphenicol (Cm) were purchased from EMD Biosciences (Darmstadt, Germany). Lysozyme, 2-mercaptoethanol (BME), bovine serum albumin (BSA), acetoacetyl coenzyme A (3-oxobutyryl-CoA, OB-CoA), nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), nicotinamide adenine dinucleotide hydrate (NAD^+), formic acid, adipic acid, sodium 3-hydroxybutyrate, thiophenol, 4-dimethylaminopyridine (DMAP), ethyl 3-methyl-2-oxobutyrate, and N,N'-dicyclohexylcarbodiimide (DCC) were purchased from Sigma-Aldrich (St. Louis, MO). 2-Methyl-3-hydroxybutyric acid was purchased from Santa Cruz Biotechnology (Dallas, Texas). Carbenicillin (Cb), kanamycin (Km), isopropyl-β-D-thiogalactopyranoside (IPTG), dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), acetonitrile, and EMD Millipore MultiScreenHTS 96-Well Filter Plates were purchased from Fisher Scientific (Pittsburgh, PA). Rectangular 24-deep well plates were purchased from VWR (Radnor, PA). Aeraseal Sealing Film was purchased from Research Products International (Mount Prospect, IL). Restriction enzymes and Phusion DNA polymerase were purchased from NEB (Ipswich, MA). DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). Oligonucleotides were purchased from IDT (Coralville, IA), resuspended at a stock concentration of 100 μM in water, and stored at -20 °C for long-term storage. dNTPs were purchased from Invitrogen (Carlsbad, CA). Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED), and ammonium persulfate, and Bio-Rad Protein Assay reagent were purchased from Bio-Rad (Hercules, CA). Imidazole was purchased from Acros Organics (Morris Plains, NJ). cOmplete EDTA-free protease inhibitor was purchased from Roche (Penzberg, Germany). Amicon Ultraspin concentrators were purchased from Merck Millipore (Cork, Ireland). Tris(2-carboxyethyl)phosphine (TCEP) was purchased from Biosynth, Inc. (Itasca, IL). PageRuler™ Plus prestained protein ladder was purchased from Fermentas (Glen Burnie, Maryland).

Bacterial strains. The T1 bacteriophage-resistant strain *E. coli* $DH10B^{T1R}$ was used for plasmid construction. *E. coli* BAP1 was used for *in vivo* screening of the AsHadh2 library and *E. coli* $BL21(DE3)^{T1R}$ was used for heterologous protein production.

Plasmid construction. Plasmid construction was carried out using standard molecular biology techniques and the Gibson protocol [*1*]. PCR amplifications were carried out with Phusion DNA polymerase, following manufacturer instructions. Primer sequences are listed in Table S1. Constructs were verified by sequencing.

AsHadh2 expression vectors: The AsHadh2 expression vector pET23a-AsHadh2-PreSC was constructed by inverse PCR of pET23a-AsHadh2 (constructed by Dr. Michael R. Blaisse) using primers 156-PreSC-AsHadh2-R and 157-PreSC-AsHadh2-F, to generate pET23a-AsHadh2-PreSC containing AsHadh2 with a N-terminal PreScission protease site and His_6 tag. The mutations identified from the *in vivo* screen were introduced into pET23a-AsHadh2-PreSC by same inverse PCR strategy using primers 461-468 (see primer table).

AsHadh2 libraries: The plasmid 1 pAsAcat3-AsHadh2 [*2*] was used as template for reverse PCR using primers Hadh-X-A and Hadh-X-B primers (see primer table) for each site saturation mutation.

Expression and purification of His_6-AsHadh2 variants. TB (1L) containing carbenicillin (50 µg/mL) in a 2.5 L Ultra yield baffled shake flasks (Thomson Instrument Company, Oceanside, CA) was inoculated to OD_{600} = 0.05 with an overnight TB culture of freshly-transformed *E. coli* BL21(DE3)^{T1R} containing the expression plasmid. The cultures were grown at 37°C at 200 rpm to OD_{600} = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with IPTG (0.4 mM), and overnight growth at 16°C. Cell pellets were harvested by centrifugation at 8,000 × *g* for 5 min at 4ºC and stored at -80ºC for protein purification.

Frozen cell pellets were thawed and resuspended at 3 mL per gram cell paste with Wash Buffer (50 mM HEPES, 10 mM imidazole, 300 mM NaCl, 5% *v/v* glycerol, pH 7.5) supplemented with 8 mM BME, 1 mg/mL lysozyme, 0.04 µL/ml benzonase, and one cOmplete EDTA-free protease inhibitor cocktail tablet per 25 mL. The cells were lysed by sonication using a Misonix Sonicator 3000, and lysate was centrifuged at 10,000 × *g* for 30 min at 4°C to separate soluble and insoluble fractions. The soluble lysate was loaded onto Ni-NTA resin column (Qiagen, 1.5 mL, 50% slurry in 30% ethanol) by gravity flow. The column was washed with 40 mL Wash Buffer with 5 mM BME and 50 mL of 90% Wash Buffer: 10% Elution Buffer (50 mM HEPES, 250 mM imidazole, 300 mM NaCl, 5% glycerol, pH 7.5) with 5 mM BME.

Protein was eluted with 10 mL Elution Buffer containing 5 mM BME. Fractions containing the target protein were pooled by $A_{280\ nm}$ and concentrated using a 10 kDa Amicon Ultraspin concentrator. Finally, protein was exchanged into Storage Buffer (20 mM HEPES, 100 mM NaCl, 10% *v/v* glycerol, pH 7.5) with 0.2 mM TCEP using a 10 kDa Amicon Ultraspin concentrator. Protein concentration was estimated using the Bio-Rad Bradford Protein Assay Kit using BSA as standard. Aliquots were frozen in liquid nitrogen and stored at -80°C.

Crystallization and determination of the His_6-AsHadh2-NAD^+ crystal structure. Crystals of His_6-AsHadh2-NAD^+ complex were obtained by hanging drop vapor diffusion by combining 800 nL of protein solution (3 mg/mL AsHadh2, 1 mM NAD^+) and reservoir solution (35% PEG550 MME, 144 mM potassium sodium tartrate, 100 mM MES, pH6.5 2 M sodium malonate, pH 6.5). Crystals were observed within 24 h, and crystals were fished out with a loop and flash-frozen by immersion in liquid nitrogen. Data were collected at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory). Data were processed with XDS [*3*] and scaled and merged with Aimless [*4*]. Molecular replacement using BALBES [*5*] was used for phase determination. The molecular replacement model was one monomer of rat brain type II 3-hydroxyacyl-CoA dehydrogenase (rHadh2, 1E3S) based on sequence alignment with AsHadh2. Any missing parts of the model were manually built using WinCoot [*6*]. Iterative cycles using Phenix.refine [*7*] and manual refinement in WinCoot were used to generate the final model, with Phenix.refine strategies based on XYZ coordinates, real-space refinement, atomic occupancies, and individual B-factors. Chimera [*8*] was used for visualization of final structures. The data statistics are summarized in Table S2. The refined His_6-AsHadh2-NAD^+ complex models and structure factors were deposited in the Protein Data Bank as the PDB ID 7N09.

Docking of OMB-CoA to the His_6-AsHadh2-NAD^+ crystal structure. Docking of OMB-CoA to His_6-AsHadh2-NAD^+ was performed using Glide SP [*9*,*10*] in the Maestro molecular visualization environment developed by SCHRODINGER (http://www.schrodinger.com). The OMB-CoA ligand was generated from acetoacetyl-coenzyme A (CAA) ligand from PDB database by adding one (*S*)-methyl group at the C2 of acetoacetyl moiety. The chain A structure of His_6-AsHadh2-NAD^+ complex was processed by Protein Preparation Wizard of Maestro for docking. To define the grid for docking, the chain A was aligned with RePhaB-$NADP^+$-CAA (PDB ID: 3VZS), and the space occupied by CAA was set as the receptor grid. Finally, OMB-

CoA was docked as a flexible ligand into prepared AsHadh2 by Glide program, and top eight poses were selected for further analysis.

AsHadh2 library transformation and *in vivo* screening. *E. coli* BAP1 was transformed with pTrc33-TesB and pAsAcat3-AsHadh2* generated by NNK saturation mutagenesis. 24 colonies were picked for each positional library for a total of 336 variants. Strains were cultured in 24-deep well plates for 2 d at 30°C in TB broth containing 2% (*w/v*) glucose, 0.2% (*w/v*) sodium propionate, and 0.418% (*w/v*) MOPS. 3-Hydroxyacids from cell culture were quantified by liquid chromatography-triple quadrupole mass spectrometry as previously reported [*2*]. The sum of the % increase in HMB and % decrease in HB (ΔHMB%-ΔHB%) was used to variant selectivity compared to wild-type with $\Delta\text{HB\% or } \Delta\text{HMB\%} = \frac{\text{titer of mutant} - \text{titer of wildtype}}{\text{titer of wildtype}}$. The AsHadh2 variants corresponding to higher selectivity for either HMB or HB were regrown for DNA sequencing. When the site of mutation was identified, ~100 colonies were picked and sequenced for positional library of interest in order to obtain maximal portion of 19 variants at that position. The missed mutants were generated by inverse PCR using primers 292-328 (see primer table). These saturation libraries for I55, M205, P209, and R213 were then analyzed similarly.

Steady-state kinetic characterization of AsHadh2 variants. AsHadh2 variants were assayed by monitoring the initial consumption of NADPH at 340 nm on a Beckman DU-800 spectrophotometer (Beckman, USA). Assays were performed at room temperature in a 1 cm-pathlength quartz cuvette in a total volume of 100 μL in HEPES buffer (10 mM HEPES, 150 mM KCl, 10 mM KH_2PO_4, pH 7.5) with NADH (0.15 mM), varying amounts of 3-oxoacyl-CoA substrate (either OB-CoA or OMB-CoA), and enzyme. Synthetic 3-oxo-2-methylbutyryl-CoA was prepared and quantified as previously reported[*2*]. The reaction was initiated by addition of 3-oxoacyl-CoA substrate and mixing rapidly by pipette. The concentrations of AsHadh2 for OB-CoA and OMB-CoA are listed.

AsHadh2 (nM)	OB-CoA	OMB-CoA
WT	200	40
I155A	696	40
I155T	682	40
M205F	136	40
R213S	68	40
R213L	140	40

The data were fit to different equations according to the performance of each enzyme: I155T/OB-CoA was fit to standard Michaelis–Menten equation (Eq. 1):

$$v_0 = \frac{V_{max}[S]_0}{K_M + [S]_0} \quad \text{Eq. 1}$$

Wild type/OB-CoA, I155A/OB-CoA, M205F/OB-CoA, R213L/OB-CoA, R213S/OB-CoA, I155A/OMB-CoA, and I155T/OMB-CoA were fit to a modified Michaelis-Menten equation with substrate inhibition (Eq. 2):

$$v_0 = \frac{V_{max}[S]_0}{K_M + [S]_0\left(1 + \frac{[S]_0}{K_I^n}\right)} \quad \text{Eq. 2}$$

The remaining combinations that could not be fit well to either of these equations were fit to the Hill equation without (Eq. 3, wild type/OMB-CoA and M205F/OMB-CoA) or with (Eq. 4, R213L/OMB-CoA and R213S/OMB-CoA) substrate inhibition:

$$v_0 = \frac{V_{max}[S]_0^n}{K_M^n + [S]_0^n} \quad \text{Eq. 3}$$

$$v_0 = \frac{V_{max}[S]_0^n}{K_M^n + [S]_0^n(1 + \frac{[S]_0^n}{K_I^n})} \quad \text{Eq. 4}$$

Molecular dynamics simulations. The initial coordinate used for molecular dynamics simulations was the OMB-CoA docked structure of AsHadh2-NADH. The complex structure of AsHadh2-NADH-OMB-CoA was also used to generate AsHadh2-NADH-OB-CoA structure by deleting the alpha-methyl group of acyl moiety of OMB-CoA. The initial unit cells were built by using the SPC solvent model with the OPLS3e force field in the system builder panel of Desmond (Schrödinger Release 2019-3). For molecular dynamics simulation, the total simulation time was set to 50 ns, with 5 ps trajectory recording intervals, to reach the equilibrium system. The system energy was set to 1.2, and the ensemble class used was NPT. Simulations were set to run at 300.0 K and at 1.01325 bar.

Supplementary Data

Table S1. Strains, plasmids, and primers used in this study. (A) Strains and plasmids. (B) Primers.

A. Strains and plasmids

Strain	Genotype	Source
DH10B^{T1R}	F$^-$ *end*A1 *rec*A1 *gal*E15 *gal*K16 *nup*G *rps*L *fhuA* Δ*lac*X74 Φ80d*lac*ZΔM15 *ara*D139 Δ(*ara,leu*)*7697 mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) *tonA*	Invitrogen
BL21(DE3)^{T1R}	F$^-$ *omp*T *gal dcm lon hsd*S_B($r_B$$^-$ $m_B$$^-$) λ(DE3 [*lac*I *lac*UV5-T7 *gene 1 ind*1 *sam*7 *nin*5]) *tonA*	Novagen
BAP1	F$^-$ *omp*T *gal dcm lon hsd*S_B($r_B$$^-$ $m_B$$^-$) λ(DE3 [*lac*I *lac*UV5-T7 *gene 1 ind*1 *sam*7 *nin*5]) Δ*prpRBCDE* (*sfp* (T7)*, prpE* (T7))]	Ref. 11

Plasmid	Description	Source
pET23a	T7, *lacI*, Cb^r*,* pBR322	Novagen
pET23a-AsHadh2	*His$_{10}$-Xa-AsHadh* (T7), *lacI*, Cb^r*,* pBR322	M. Blaisse
pET23a-AsHadh2-PreSC	*AsHadh2-PreSC-His$_{10}$* (T7), *lacI*, Cb^r*,* pBR322	This study
pET23a-AsHadh2 I155A-PreSC	*AsHadh2 I155A-PreSC-His$_{10}$* (T7), *lacI*, Cb^r*,* pBR322	This study
pET23a-AsHadh2 I155T-PreSC	*AsHadh2 I155T-PreSC-His$_{10}$* (T7), *lacI*, Cb^r*,* pBR322	This study
pET23a-AsHadh2 M205F-PreSC	*AsHadh2 M205F-PreSC-His$_{10}$* (T7), *lacI*, Cb^r*,* pBR322	This study
pET23a-AsHadh2 R213S-PreSC	*AsHadh2 R213S-PreSC-His$_{10}$* (T7), *lacI*, Cb^r*,* pBR322	This study
pET23a-AsHadh2 R213L-PreSC	*AsHadh2 R213L-PreSC-His$_{10}$* (T7), *lacI*, Cb^r*,* pBR322	This study
pAsAcat3-AsHadh2	*AsAcat3-AsHadh2* (Trc), *lacI*, Cb^r*,* pBR322	Ref. 2
pAsAcat3-AsHadh2 library	*AsAcat3-AsHadh2 variant* (Trc), *lacI*, Cb^r*,* pBR322	This study
pTrc33-TesB	*TesB*(9.2k) (Trc), Cm^r*,* p15a	Ref. 12

B. Primers

Primer	Sequence
156-PreSC-AsHadh2-R	*CGGGCCCTGAAAAAGCACTTCTAG*ATGATGATGATGATGATGGCTGCTAGCCATATG
157-PreSC-AsHadh2-F	*CTAGAAGTGCTTTTTCAGGGCCCG*TCCGCACTGCGCTCCACGAAAGGTCTGGTCGCGCTG
206-Hadh-A94-A	GCGCGGGTATTNNKTACAGCTTTAAACTGTTCAACGTGAAGAAGAAAAAA
207-Hadh-A94-B	TTAAAGCTGTAMNNAATACCCGCGCAGTTCACCGCAACGTCCAAACGACC
210-Hadh-A153-A	TTATTAATACCNNKAGCATTGCAGCGTTCGACGGTCAAGCCGGCCAAAGC
211-Hadh-A153-B	GCTGCAATGCTMNNGGTATTAATAATAACACCGCGTTGGCCCATCTCGTC
214-Hadh-I155-A	ATACCGCAAGCNNKGCAGCGTTCGACGGTCAAGCCGGCCAAAGCGCGTAT

215-Hadh-I155-B	TCGAACGCTGCMNNGCTTGCGGTATTAATAATAACACCGCGTTGGCCCAT
216-Hadh-A156-A	CCGCAAGCATTNNKGCGTTCGACGGTCAAGCCGGCCAAAGCGCGTATAGC
217-Hadh-A156-B	CCGTCGAACGCMNNAATGCTTGCGGTATTAATAATAACACCGCGTTGGCC
218-Hadh-Q164-A	GTCAAGCCGGCNNKAGCGCGTATAGCGCCTCCAAGGGCGCGATTGTTGGC
219-Hadh-Q164-B	CTATACGCGCTMNNGCCGGCTTGACCGTCGAACGCTGCAATGCTTGCGGT
226-Hadh-P197-A	TGACCATCGCTNNKGGCATCTTTGACACCCCGATGATGGCGTCCTTCCCG
227-Hadh-P197-B	TCAAAGATGCCMNNAGCGATGGTCACAACGCGGATGCCATCGTCGGCGAA
228-Hadh-G198-A	CCATCGCTCCGNNKATCTTTGACACCCCGATGATGGCGTCCTTCCCGGAC
229-Hadh-G198-B	GTGTCAAAGATMNNCGGAGCGATGGTCACAACGCGGATGCCATCGTCGGC
230-Hadh-I199-A	TCGCTCCGGGCNNKTTTGACACCCCGATGATGGCGTCCTTCCCGGACAAA
231-Hadh-I199-B	GGGGTGTCAAAMNNGCCCGGAGCGATGGTCACAACGCGGATGCCATCGTC
232-Hadh-F200-A	CTCCGGGCATCNNKGACACCCCGATGATGGCGTCCTTCCCGGACAAAGTT
233-Hadh-F200-B	ATCGGGGTGTCMNNGATGCCCGGAGCGATGGTCACAACGCGGATGCCATC
238-Hadh-M204-A	TTGACACCCCGNNKATGGCGTCCTTCCCGGACAAAGTTCGTAACTTCCTG
239-Hadh-M204-B	AAGGACGCCATMNNCGGGGTGTCAAAGATGCCCGGAGCGATGGTCACAAC
240-Hadh-M205-A	ACACCCCGATGNNKGCGTCCTTCCCGGACAAAGTTCGTAACTTCCTGATT
241-Hadh-M205-B	GGGAAGGACGCMNNCATCGGGGTGTCAAAGATGCCCGGAGCGATGGTCAC
244-Hadh-F208-A	TGATGGCGTCCNNKCCGGACAAAGTTCGTAACTTCCTGATTGGTTTGGTG
245-Hadh-F208-B	ACTTTGTCCGGMNNGGACGCCATCATCGGGGTGTCAAAGATGCCCGGAGC
246-Hadh-P209-A	TGGCGTCCTTCNNKGACAAAGTTCGTAACTTCCTGATTGGTTTGGTGCCG
247-Hadh-P209-B	CGAACTTTGTCMNNGAAGGACGCCATCATCGGGGTGTCAAAGATGCCCGG
248-Hadh-R213-A	CGGACAAAGTTNNKAACTTCCTGATTGGTTTGGTGCCGAATCCGAAGCGT
249-Hadh-R213-B	ATCAGGAAGTTMNNAACTTTGTCCGGGAAGGACGCCATCATCGGGGTGTC
292-R213E-F	CGGACAAAGTTGAAAACTTCCTGATTGGTTTGGTGCCGAATCCGAAGCGT
293-R213E-R	ATCAGGAAGTTTTCAACTTTGTCCGGGAAGGACGCCATCATCGGGGTGTC
294-I155D-F	ATACCGCAAGCGATGCAGCGTTCGACGGTCAAGCCGGCCAAAGCGCGTAT
295-I155D-R	TCGAACGCTGCATCGCTTGCGGTATTAATAATAACACCGCGTTGGCCCAT
296-I155G-F	ATACCGCAAGCGGCGCAGCGTTCGACGGTCAAGCCGGCCAAAGCGCGTAT
297-I155G-R	TCGAACGCTGCGCCGCTTGCGGTATTAATAATAACACCGCGTTGGCCCAT
298-I155V-F	ATACCGCAAGCGTGGCAGCGTTCGACGGTCAAGCCGGCCAAAGCGCGTAT
299-I155V-R	TCGAACGCTGCCACGCTTGCGGTATTAATAATAACACCGCGTTGGCCCAT
300-I155W-F	ATACCGCAAGCTGGGCAGCGTTCGACGGTCAAGCCGGCCAAAGCGCGTAT
301-I155W-R	TCGAACGCTGCCCAGCTTGCGGTATTAATAATAACACCGCGTTGGCCCAT
302-M205E-F	ACACCCCGATGGAAGCGTCCTTCCCGGACAAAGTTCGTAACTTCCTGATT
303-M205E-R	GGGAAGGACGCTTCCATCGGGGTGTCAAAGATGCCCGGAGCGATGGTCAC
304-M205N-F	ACACCCCGATGAACGCGTCCTTCCCGGACAAAGTTCGTAACTTCCTGATT
305-M205N-R	GGGAAGGACGCGTTCATCGGGGTGTCAAAGATGCCCGGAGCGATGGTCAC
306-M205Y-F	ACACCCCGATGTATGCGTCCTTCCCGGACAAAGTTCGTAACTTCCTGATT
307-M205Y-R	GGGAAGGACGCATACATCGGGGTGTCAAAGATGCCCGGAGCGATGGTCAC
308-P209D-F	TGGCGTCCTTCGATGACAAAGTTCGTAACTTCCTGATTGGTTTGGTGCCG
309-P209D-R	CGAACTTTGTCATCGAAGGACGCCATCATCGGGGTGTCAAAGATGCCCGG
310-P209I-F	TGGCGTCCTTCATTGACAAAGTTCGTAACTTCCTGATTGGTTTGGTGCCG
311-P209I-R	CGAACTTTGTCAATGAAGGACGCCATCATCGGGGTGTCAAAGATGCCCGG
312-P209N-F	TGGCGTCCTTCAACGACAAAGTTCGTAACTTCCTGATTGGTTTGGTGCCG
313-P209N-R	CGAACTTTGTCGTTGAAGGACGCCATCATCGGGGTGTCAAAGATGCCCGG
314-R213E-F	CCTTCCCGGACAAAGTTGAAAACTTCCTGATTGGTTTG
315-R213	AACTTTGTCCGGGAAGGACGCCATCATCG
316-I155D-F	TGTTATTATTAATACCGCAAGCGATGCAGCGTTCGACG

317-I155	GCTTGCGGTATTAATAATAACACCGCGTTGGCC
318-I155G-F	TGTTATTATTAATACCGCAAGCGGCGCAGCGTTCGACG
319-I155V-F	TGTTATTATTAATACCGCAAGCGTGGCAGCGTTCGACG
320-I155W-F	TGTTATTATTAATACCGCAAGCTGGGCAGCGTTCGACG
321-M205E-F	CTTTGACACCCCGATGGAAGCGTCCTTCC
322-M205	CATCGGGGTGTCAAAGATGCCCGGAGC
323-M205N-F	CTTTGACACCCCGATGAACGCGTCCTTCC
324-M205Y-F	CTTTGACACCCCGATGTATGCGTCCTTCC
325-P209D-F	GATGATGGCGTCCTTCGATGACAAAGTTCG
326-P209	GAAGGACGCCATCATCGGGGTGTCAAAGA
327-P209I-F	GATGATGGCGTCCTTCATTGACAAAGTTCG
328-P209N-F	GATGATGGCGTCCTTCAACGACAAAGTTCG
461-I155A-P1	GTGTTATTATTAATACCGCAAGCGCGGCAGCGTTCGACG
462-I155T-P1	GTGTTATTATTAATACCGCAAGCACCGCAGCGTTCGACG
463-M205F-P1	TCTTTGACACCCCGATGTTTGCGTCCTT
464-R213S-P1	CCTTCCCGGACAAAGTTAGCAACTTCCTGATTGG
465-R213L-P1	CCTTCCCGGACAAAGTTCTGAACTTCCTGATTGG
466-I155-P2	GCTTGCGGTATTAATAATAACACCGCGTTGGCC
467-M205-P2	CATCGGGGTGTCAAAGATGCCCGGAGC
468-R213-P2	AACTTTGTCCGGGAAGGACGCCATCATCG

Table S2. Data collection and refinement statistics for the AsHadh2-NAD$^+$ structure.

	AsHadh2, NAD$^+$ bound
Data collection	
Wavelength (Å)	1.116
Space group	C 1 2 1
Unit cell	
a, b, c (Å)	128.5 54.9 89.3
α, β, γ (°)	90 131 90
Resolution (Å)	64.050 - 2.0 (2.072 - 2.0)*
Total reflections	63882 (6307)
Unique reflections	31971 (3160)
Multiplicity	2.0 (2.0)
Completeness (%)	99.86 (99.15)
Mean I/sigma(I)	12.53 (1.95)
Wilson B-factor	31.22
R_{merge}	0.02678 (0.3468)
R_{meas}	0.03787 (0.4905)
R_{pim}	0.02678 (0.3468)
$CC_{1/2}$	0.999 (0.81)
CC*	1 (0.946)
Refinement	
Resolution (Å)	64.050 - 2.0
Reflections used in refinement	31959 (3160)
Reflections used for R_{free}	1448 (148)
R_{work}/R_{free}	0.1768 (0.2896)/0.2191 (0.3350)
CC_{work}/CC_{free}	0.967 (0.854)/0.973 (0.722)
Number of non-hydrogen atoms	4261
Macromolecules	3798
Ligands	88
Solvent	375
Protein residues	518
RMSD	
Bond lengths (Å)	0.007
Bond angles (°)	0.95
Ramachandran	
Favored (%)	96.89
Allowed (%)	3.11
Outliers (%)	0.00
Rotamer outliers (%)	0.52
Clashscore	5.00
Average B-factor	35.99
Macromolecules	35.29
Ligands	40.19
Solvent	42.07

* Values for the highest-resolution shell are shown in parentheses.

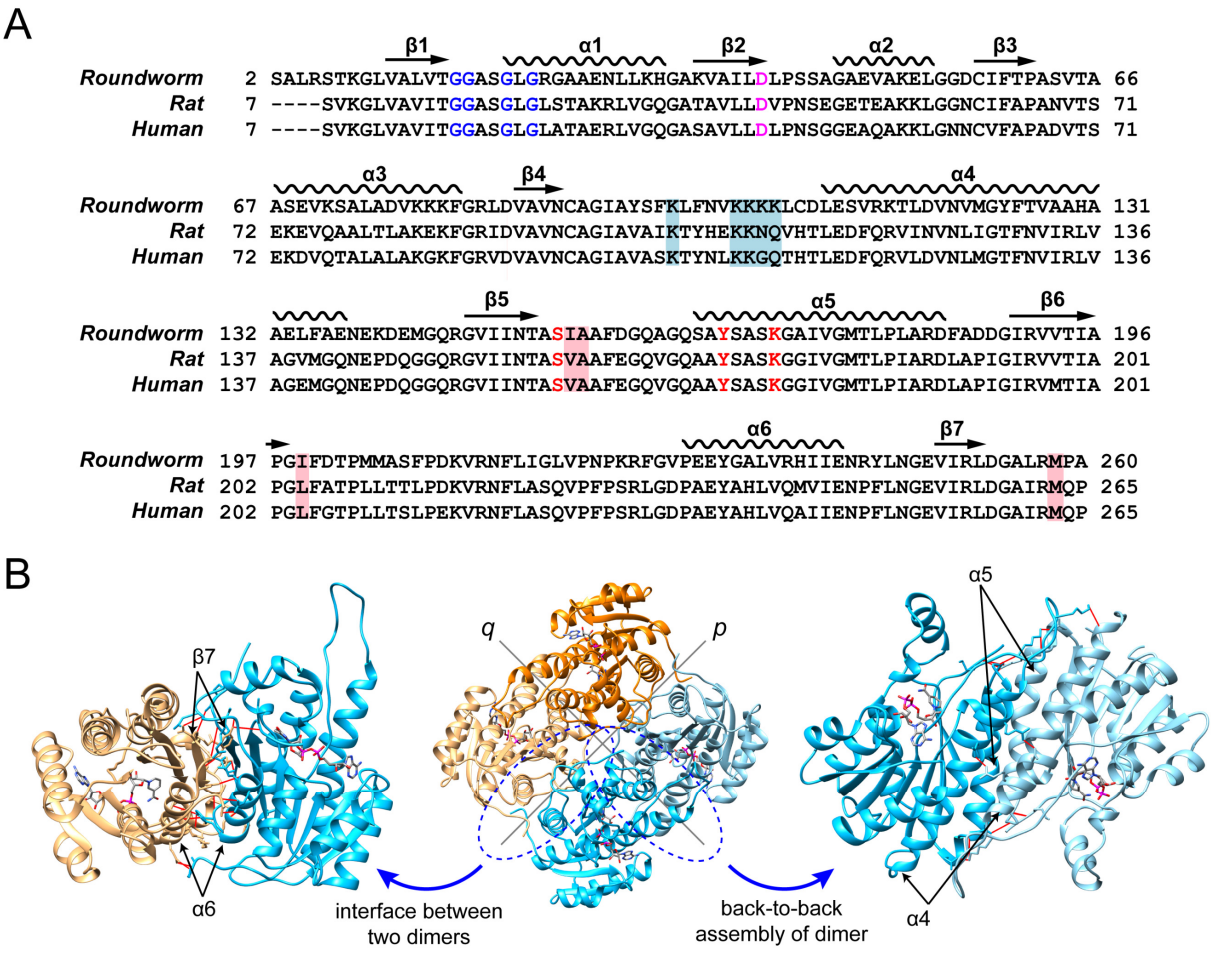


Figure S1. Crystal structure of AsHadh2. (A) Sequence alignment of AsHadh2, rHadh2, and hHadh2. Secondary structure elements are drawn based on the AsHadh2 structure. Catalytic residues (red), D40 involved in NAD^+ selectivity (magenta), and conserved GGXXGXG Rossmann motif for dinucleotide binding (blue) are highlighted in color. Residues proposed to be involved in acyl-CoA binding are shaded in a box (α-methyl group binding pocket, pink; diphosphate linkage, light blue). (AsHadh2, *A. suum*; rHadh2, *Rattus norvegicus*; hHadh2, *Homo sapiens*) (B) The dimer interface forming the back-to-back assembly and the interface between two dimers forming the tetramer. Structure of the tetrameric AsHadh2 assembly as a ribbon diagram with NAD^+ (stick) showing the location of the active site. Back-to-back dimers are shown in blue and orange. Within the dimer, each monomer unit is shaded in dark or light. The two dimers are related by a *p* axis of symmetry, whereas the two subunits in the dimers are related by a *q* axis of symmetry. Hydrogen bonds between two monomers are indicated by solid red lines. The key secondary structure elements involved in the interface are labelled.

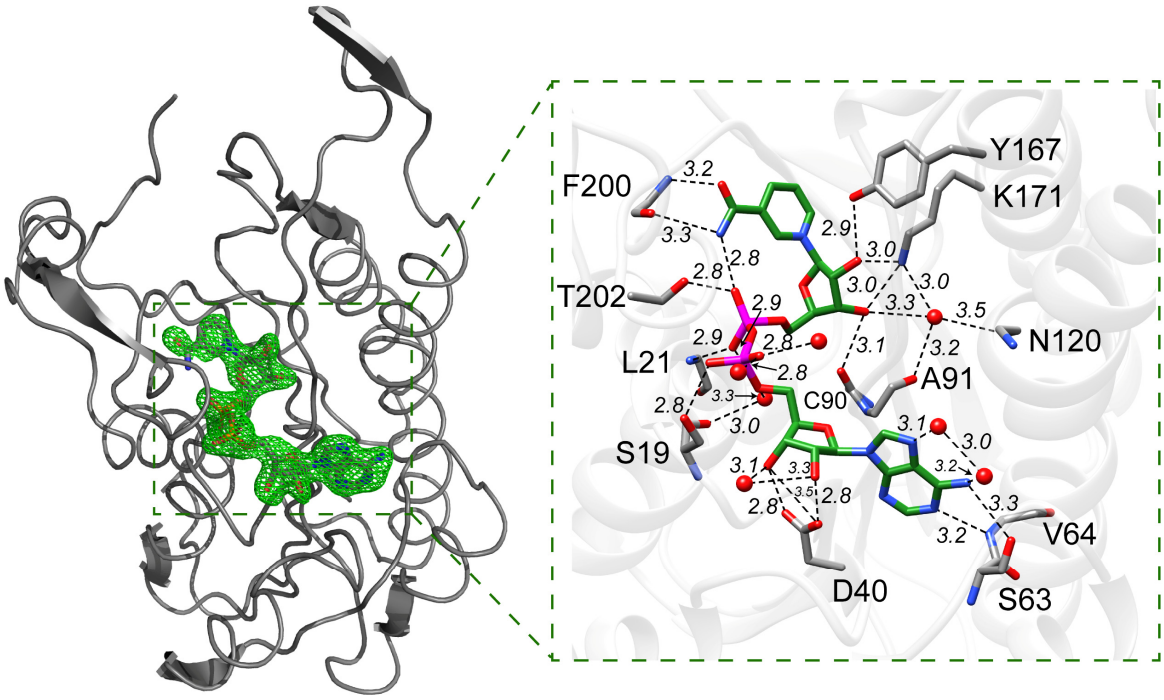


Figure S2. AsHadh2 bound to NAD$^+$. The $F_o - F_c$ omit map for the ligand NAD$^+$ (green) was calculated by removing ligands before refinement and is contoured at ± 3σ. The bound NAD$^+$ cofactor is shown in stick with carbons colored in green. Dashed lines indicate hydrogen bond contacts with distances (Å) between two atoms noted. Side chains or main chains interacting with the NAD$^+$ are show in stick with carbons colored in gray. Structured waters are shown as red spheres. (O, red; N, blue; P, fuchsia)

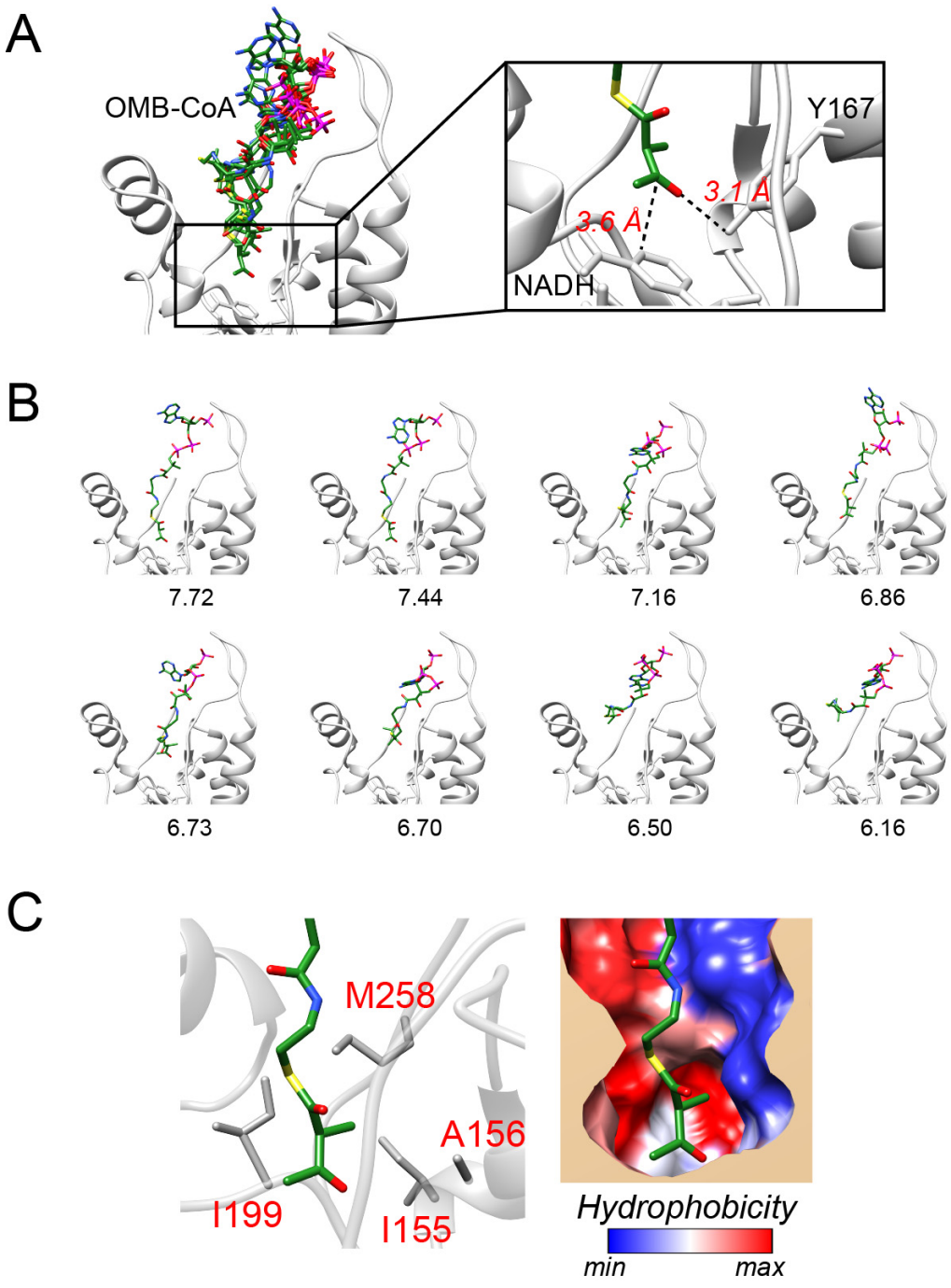
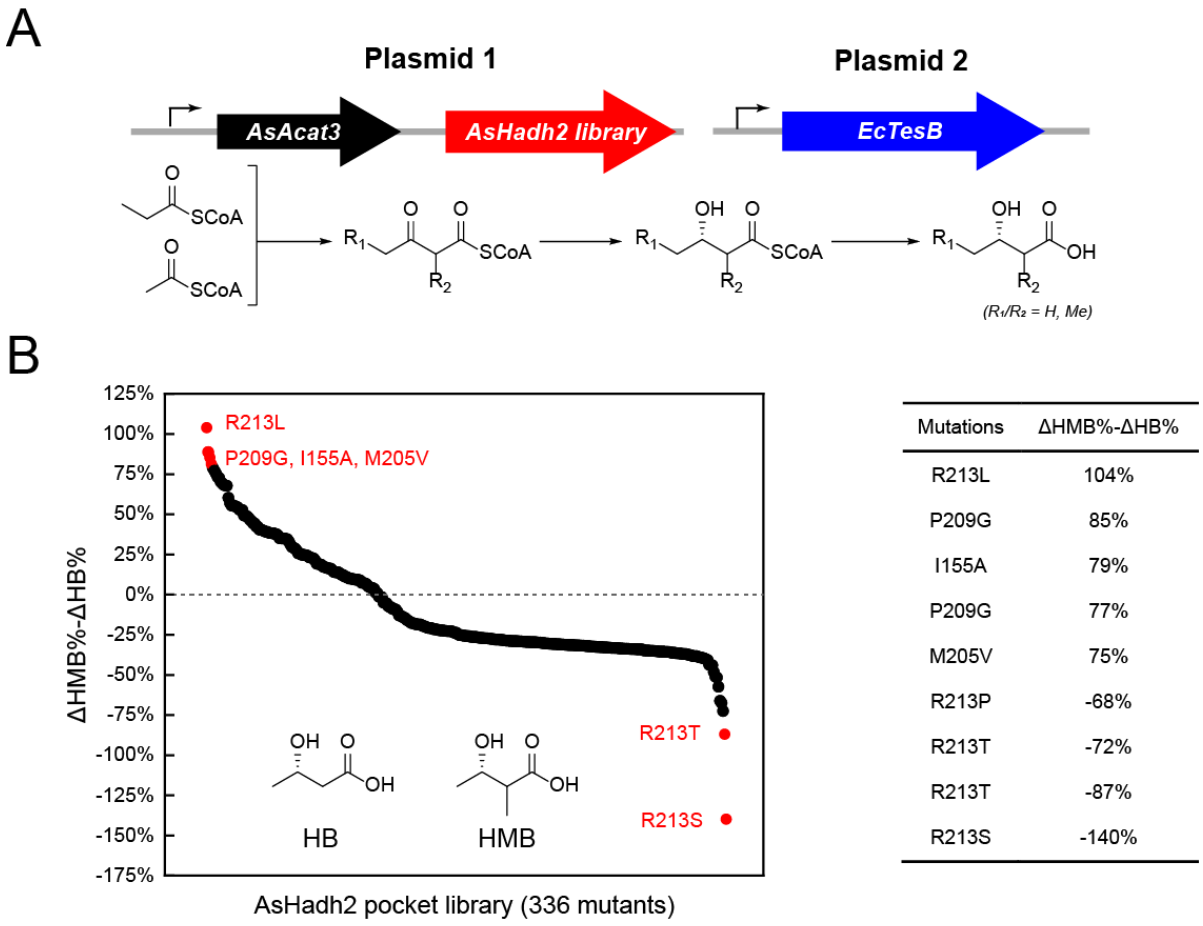


Figure S3. Docking of (*S*)-3-oxo-2-mehthylbutyryl-CoA into AsHadh2. (A) Eight given poses of ligand are shown here of the OMB-CoA (C, green) docked into the AsHadh2-NADH crystal structure (*left*) with a zoomed in view of the pose with highest docking score (*right*). (B) The eight poses shown overlaid in Figure S3A with scores. (C) Structural features involved in binding α-methyl group of substrate. A surface representation of the hydrophobic pocket formed by I155, A156, I199, and M258 is shown colored by hydrophobicity. The key residues for the hydrophobic interaction with α-methyl group of OMB-CoA (C, green) are show in stick model. (C, grey; O, red; N, blue; S, yellow; P, fuchsia)



Mutations	ΔHMB%-ΔHB%
R213L	104%
P209G	85%
I155A	79%
P209G	77%
M205V	75%
R213P	-68%
R213T	-72%
R213T	-87%
R213S	-140%

Figure S4. Generation and screening of an AsHadh2 pocket library. The random mutation of these pocket-residues was performed using NNK degenerate codon mutagenesis to form 14 AsHadh2 libraries (A94, A153, I155, A156, Q164, P197, G198, I199, F200, M204, M205, F208, P209, and R213L). 24 colonies were picked from each positional library for preliminary screening of 336 total variants. (A) AsHadh2 variants were analyzed using an *in vivo* screen in *E. coli* bearing a pathway for the production of 3-hydroxy acids. (B) *E. coli* BAP1 was transformed with pTrc33-TesB pAsAcat3-AsHadh2*. After 2 d, cell cultures were analyzed for production of branched (HMB) vs linear (HB) product. Data are plotted in terms of the sum of the percent increase in HMB and percent decrease in HB compared to wild-type. (AsAcat3, thiolase from *A. suum*, AsHadh2, variant from library; TesB, thioesterase from *E. coli*). Based on these initial results, positions I155, M205, P209 and R213 (red) were selected for more detailed analysis.

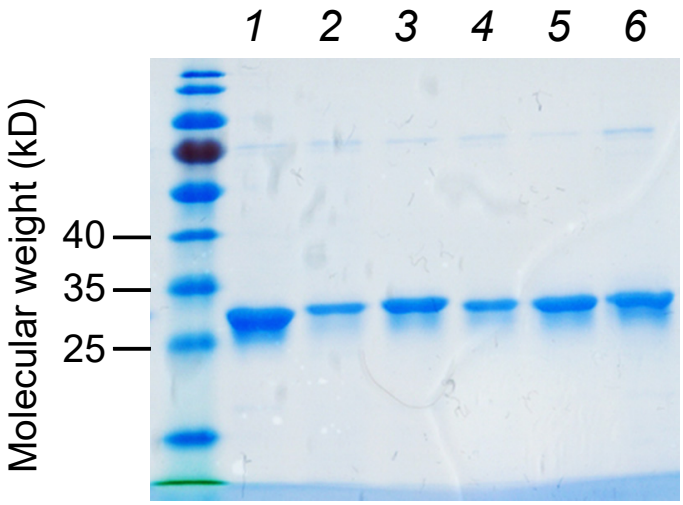


Figure S5. SDS-PAGE of purified Hadh2 variants. Lane 1, wild type; lane 2, I155A; lane 3, I155T; lane 4, M205F; lane 5, Hadh2, R213S; lane 6, R213L.

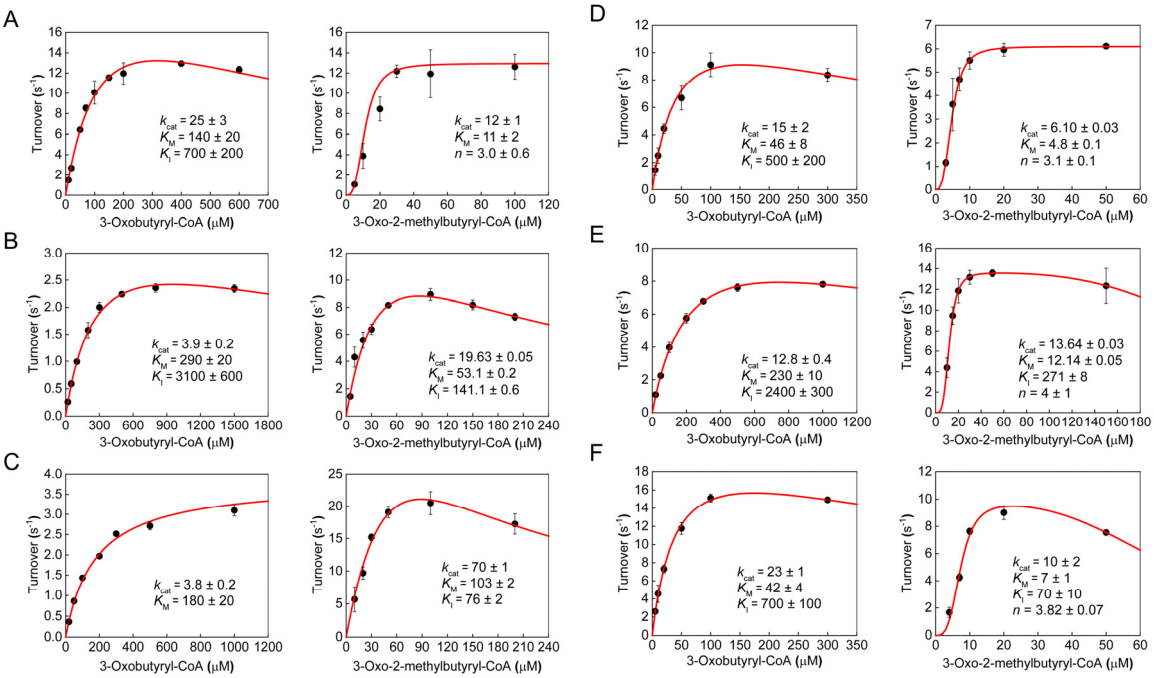


Figure S6. Steady-state kinetic analysis for Hadh2 variants with OMB-CoA and OB-CoA. Initial rate data for Hadh2 variants for the reduction of branched (OMB-CoA) or linear (OB-CoA) were collected as a function of substrate concentration by monitoring oxidation of NADH. Data were fit to the standard Michaelis-Menten equation (Eq. 1), the Michaelis-Menten equation with substrate inhibition (Eq. 2), the Hill equation (Eq. 3), or the Hill equation with substrate inhibition (Eq. 4) as noted in the method. Data are mean ± standard error (n = 3). (A) Wild-type AsHadh2. (B) AsHadh2 I155A. (C) AsHadh2 I155T. (D) AsHadh2 M205F. (E) AsHadh2 R213L. (F) AsHadh2 R213S.

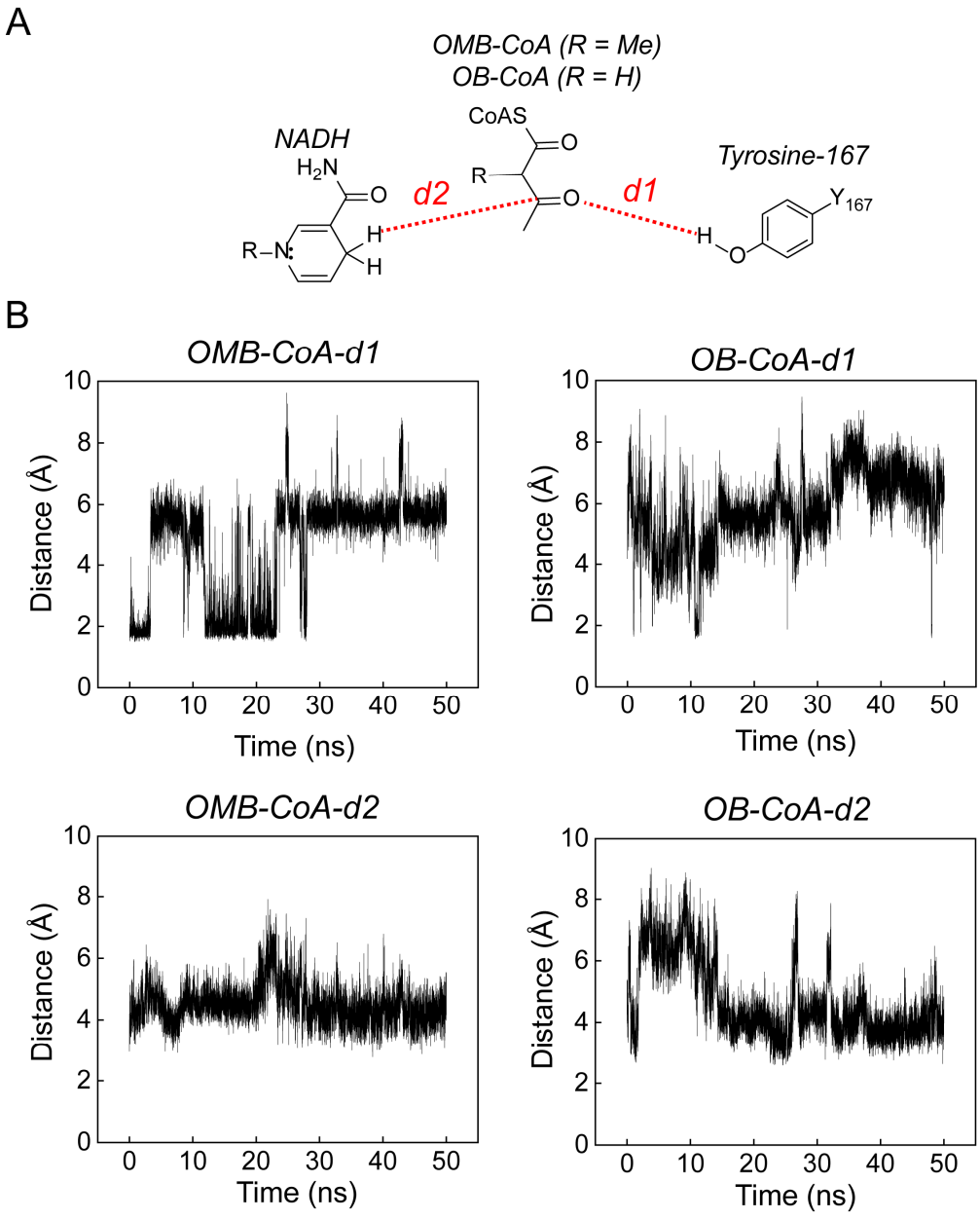


Figure S7. Molecular dynamics simulations of AsHadh2. The structure of AsHadh2-NAD^+ with either OMB-CoA (branched) or OB-CoA (unbranched) docked as a ligand was used for molecular dynamics simulations. The distances between the keto group and either Y167 (*d1*) or NADH (*d2*) were plotted to monitor the distances of the reacting centers as they are considered to serve as a minimal metric for monitoring formation of an 'active conformation' for hydride transfer [*13*,*14*]. (A) Distances between reacting centers where *d1* is $O3_{O(M)BCoA}$-HH_{Y167} and *d2* is $C3_{O(M)BCoA}$-$H4N1_{NADH}$. (B) The profiles of *d1* and *d2* for ligands OMB-CoA and OB-CoA in 50 ns simulation process. The active conformation was defined as *d1* ≤ 3.4 Å and *d2* ≤ 4.5 Å [*13*]. The proportions of pre-reaction conformations of AsHadh2 were 11.8% for OMB-CoA, and 0.4% for OB-CoA.

References

1. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Protoc.* **2009**, 343-345.
2. Dong, H.; Liffland, S.; Hillmyer, M. A.; Chang, M. C., Engineering *in vivo* production of α-branched polyesters. *J. Am. Chem. Soc.* **2019**, *141*, 16877-16883.
3. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 125-132.
4. Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.* **2013**, *69*, 1204-1214.
5. Long, F.; Vagin, A. A.; Young, P.; Murshudov, G. N. BALBES: A molecular-replacement pipeline. *Acta Cryst. D. Biol. Crystallogr.* **2008**, *64*, 125-132.
6. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486–501.
7. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 213–221.
8. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
9. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739-1749.
10. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750-179.
11. Pfeifer, B. A.; Admiraal, S. J.; Gramajo, H.; Cane, D. E.; Khosla, C., Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **2001**, *291*, 1790-1792.
12. Blaisse, M. R.; Dong, H.; Fu, B.; Chang, M. C., Discovery and engineering of pathways for production of α-branched organic acids. *J. Am. Chem. Soc.* **2017**, *139*, 14526-14532.
13. Zhou, J.; Wa,ng, Y.; Xu, G.; Wu, L.; Han, R.; Schwaneberg, U.; Rao, Y.; Zhao, Y. L.; Zhou, J.; Ni Y. Structural insight into enantioselective inversion of an alcohol dehydrogenase reveals a "Polar Gate" in stereorecognition of diaryl ketones. *J. Am. Chem. Soc.* **2018**, *140*, 12645−12654.

14. Mugnai , M. L.; Shi, Y.; Keatinge-Clay , A. T.; Elber, R. Molecular dynamics studies of modular polyketide synthase ketoreductase stereospecificity. *Biochemistry* **2015**, *54*, 2346−2359.