Electronic Supplementary Material (ESI) for Chemical Communications

Supplementary Information

Dual catalysis mode for the dicarbonyl reduction catalyzed by diketoreductase

Meiling Lu,‡[a] Yan Huang,‡[a] Mark A. White,‡[b] Xuri Wu,[a] Nan Liu,[a] Xiaodong Cheng, *[bc] Yijun Chen*[a]

*To whom correspondence should be addressed.

E-mail: yjchen@cpu.edu.cn, xcheng@utmb.edu

This file includes:

Materials and Methods

Scheme S1

Figure S1 to S9

Table S1 and S2

Full References

Materials and Methods

1. Chemicals.

The oligonucleotides used in site-directed mutagenesis were synthesized by Invitrogen Bio Inc., (Shanghai, China). Axy prep plasmid miniprep kit was from Axygene Biotech Ltd., (USA). *Escherichia coli* strains DH5α and BL21 (DE3) were obtained from Tiangen Biotech Co. Ltd. (Beijing, China) for the preparation of plasmids and gene expression, respectively. Chemicals were purchased from Sigma-Aldrich Chemical Company, Fluka and Merck. Crystallization kits were purchased from Hampton Research, Molecular Dimensions Limited and Molecular Dimensions. Ethyl 3, 5-diketo-6-benzyloxy hexanoate (**1**) was synthesized as described previously.[1] Intermediates **2**, **3** were purified from the reaction mixtures according to our published method. [2]

2. Methods

2.1 Site-directed mutagenesis

Mutagenesis was performed using a QuickChange site-directed mutagenesis kit (Stratagene) and an expression pET22b (+)-dkr was used as a template for mutagenesis according to the manufacturer's protocol. The mutagenic primers were designed to produce the mutant enzymes of DKR by replacing the respective codons in the cDNA with those for the replaced amino acids. Mutations were confirmed by DNA sequencing with ABI Genetic Analyzer 3730 (Invitrogen Bio Inc., Shanghai, China). The *dkr* gene from *Acinetobacter baylyi* ATCC 33305 (GenBank Accession No. EU273886) was used to construct the expression plasmid. The construction of the plasmid pET22b(+)-dkr was performed as previously described. [3, 4]

2.2 Protein expression and purification

Recombinant DKR and its mutants, used for crystallization, were overexpressed in *Escherichia*

coli BL21(DE3) cells (Novagen) and purified by column chromatography on DEAE-FF (GE Healthcare) and Superdex G-200 (GE Healthcare), as previously described. [3, 4] The purity of the protein samples was examined by 12% SDS-PAGE with Commassie blue staining.

Recombinant *E. coli* cells for human heart short chain L-3-hydroxyacyl-CoA dehydrogenase (HAD) were cultivated in LB media as described previously. [5] Briefly, after induction with IPTG, crude enzymes were obtained by high pressure cell disruption (Constant Systems, UK) and centrifugation. HAD was purified by affinity chromatography using a Nickel-NTA column (Novagen, USA) following the manufacturer's protocol. The homogeneity of the purified proteins was determined to be greater than 99 % on SDS-PAGE by Commassie blue staining.

2.3 Protein crystallization

Apo-DKR crystals were obtained using a reservoir solution of 2% (v/v) polyethylene glycol (PEG) 400, 0.2 M lithium sulfate, 1.6 M ammonium sulfate and 0.1 M sodium citrate (pH 5.6) and grown at 293K by the sitting-drop vapor diffusion method in which 1 μl of 6 mg ml^{-1} protein solution in 10 mM Tris buffer (pH 7.0) was mixed with an equal volume of the reservoir solution. Before flash-cooling in liquid nitrogen, all crystals were soaked briefly in a cryo-buffer consisting of reservoir solution plus 25% glycerol. Heavy-atom anion quick-soak salt was prepared by the addition of 500 mM KI to the cryo-buffer. For the DKR complex, 25 mM NAD^+ or NADH (final concentration) was added into the initial soaking solution and the crystals were soaked for 15 min before being directly flash-frozen in liquid nitrogen.

2.4 Crystallographic data collection and refinement

For Heavy atom phasing a 2.7Å native and a 2.5Å KI halide anion quick-soak derivative [6] data set were collected at 85K with a MacScience DIP2030 imaging plate detector on a M06HF-SRA Cu x-ray source equipped with Rigaku Blue confocal optics. All home-source data were

collected in the Sealy Center for Structural Biology and Molecular Biophysics X-ray Crystallography Laboratory at UTMB. The apo-DKR structure was solved using SIRAS heavy atom structure solution in phenix using the AutoSol routine. [7] This resulted in an incomplete model which was sufficient; however, to place a homology model, created using SWISS-MODEL [8] and the human HAD structures [9] 1F12, 1F14, and 1F17, into the electron density using COOT. [10] This model had previously failed to produce a successful Molecular Replacement solution, but fit the core regions of the experimental electron density for each individual domain well. Regions which did not fit the density, including large secondary structure elements, loops, and side-chains, were rebuilt using COOT's real space refinement before the model was refined in CNS [11] using PMB. [12-14] High resolution data for the apo-DKR structure were collected at the SSRL synchrotron on beamline 7.2. The NADH and NAD^+ complexes were each solved by rigid body refinement, using the apo model, with high resolution data collected on a Rigaku R-AXIS-IV^{++} equipped with an FRE^{++}DW generator and Varimax-Cu optics. During refinement the model was validated using COOT, PMB, reduce, [15] and Molprobity. [16, 17] The apo, NADH, and NAD^+ DKR models were each deposited with the PDB, and assigned ids 4E12, 4E13, and 4DYD respectively (Table S1).

2.5 Molecular docking

The pdb files of enol-, 3-keto, and 5-keto ligands were generated using Draw Molecule with JME using the PRODRG2 Server (http://davapc1.bioch.dundee.ac.uk/programs /prodrg) and minimized using the minimization program of Discovery Studio Client 2.5.5. Each substrate was docked individually into DKR by using AutoDock 4.2.3. [18] All water molecules were removed from the original Protein Data Bank file. Polar hydrogen atoms were added and Gasteiger charge, atomic solvation parameters, and fragmental volumes were assigned to the protein using

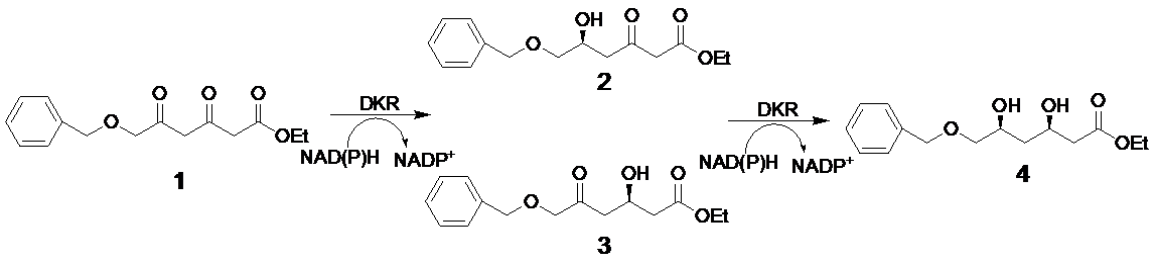
AutoDock Tools (ADT). The program AutoGrid was used to generate the grid maps. Previous site-directed mutation of Ser122, His143, and Glu155 of DKR showed that these residues are critical for catalysis, it is reasonable to assume that the binding site of DKR for these substrates should be around these residues. The grid size was set to be 60 ×60 ×60 with points separated by 0.375 Å. A conformational search was performed using the Lamarkian genetic algorithm (LGA). [19] The standard docking protocol consisted of 100 independent runs per ligand, using an initial population of 150 randomly placed individuals, with 2.5 ×106 energy evaluations, a maximum number of 27,000 iterations, a mutation rate of 0.02 and a crossover rate of 0.80, and an elitism value of 1. All the complex candidates were evaluated and ranked in terms of the binding energy by using the standard energy score function implemented in the docking program, and the best poses were analyzed for hydrogen bonding using PyMOL program.

2.6 Molecular dynamic simulation

Molecular dynamic (MD) simulations were carried out for the selected initial DKR-substrate complex using GROMACS 4.5.3 and the GROMOS-96(43a1) force field. [20] Solvation effects were simulated using the SPC water model. [21] Each system was inserted in an octahedron water box in which the edges were about 7.5 nm. 20 Na^+ ions were added to neutralize the negative charge on the system. A thorough energy minimization was performed before starting MD simulations, consisting of: 2000 steps using the steepest descent algorithm followed by 2000 steps of conjugate gradient minimization of the whole system. The minimized system was equilibrated for 300 ps at 310 K, then was subsequently submitted to 5 ns MD simulation, and snapshot structures were extracted every 1 ps. Temperature was kept constant by using the Berendsen coupling algorithm [22] with a time constant for heat-bath coupling of 0.1 ps. Long-range electrostatic calculations were performed by the particle mesh Ewald method, [23] with a

cutoff equal to 1.0nm. A cutoff for the van der Waals interactions was 1.4 nm. The lengths of covalent bonds involving hydrogen atoms were constrained with the P-LINCS algorithm. [24]

Scheme S1. Reaction catalyzed by diketoreductase



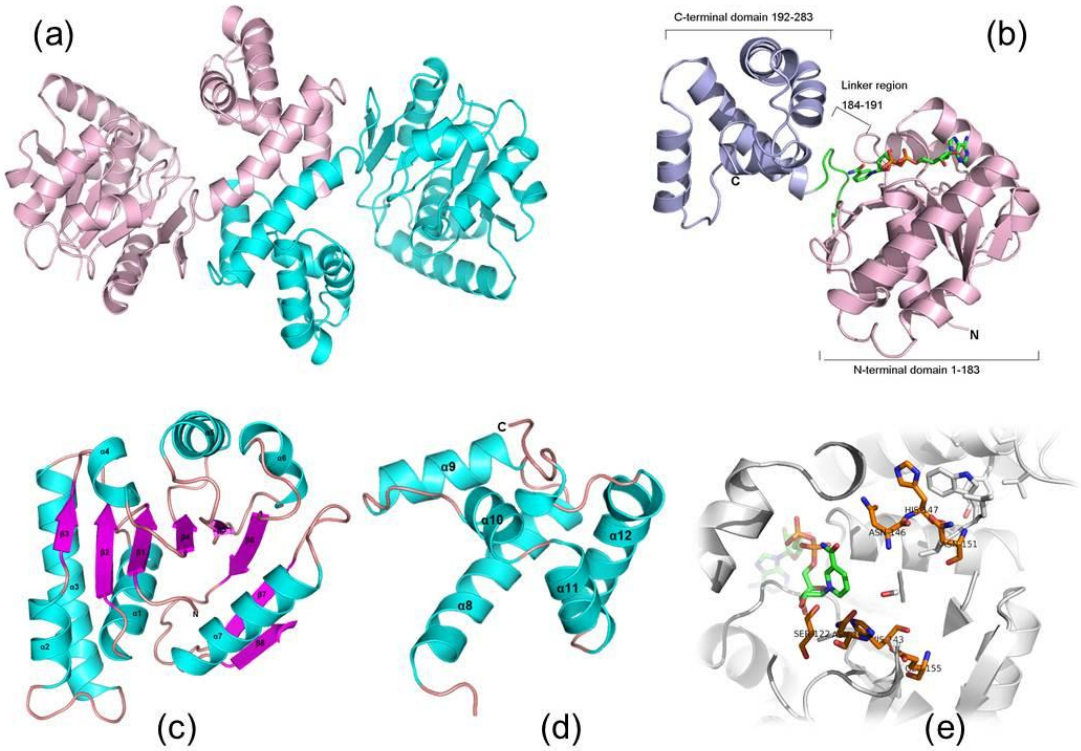


Figure S1. Structures of DKR.

(a) Cartoon drawing of DKR dimeric structure. Two subunits are shown in cyan and pink, respectively. (b) Cartoon drawing of NAD^+-bound DKR structure. N-terminal domain, linker region and C-terminal domain are shown in pink, green and light blue, respectively. The NAD^+ molecule is shown as a stick model. (c) Structure of N-terminal domain. The α-helix and β-sheet are shown in cyan and magenta, respectively. (d) Structure of C-terminal domain. The α-helix is shown in cyan. (e) Diagram of the active-site residues. The residues and the NADH molecular are depicted as stick models. NADH and key residues are shown in green and orange, respectively.

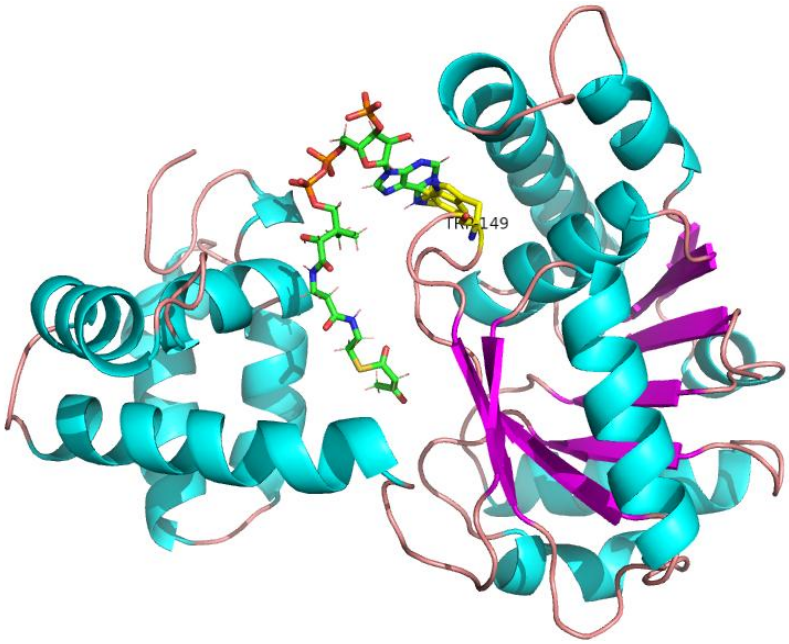


Figure S2. Superimposed structures of DKR and HAD substrate acetoacetyl-CoA.

The α-helix, β-sheet and random coil of DKR monomer are shown in cyan, magenta and pink, respectively. Trp149 is shown as a stick model in yellow and acetoacetyl-CoA molecule is shown as a stick model based on the binding with HAD.

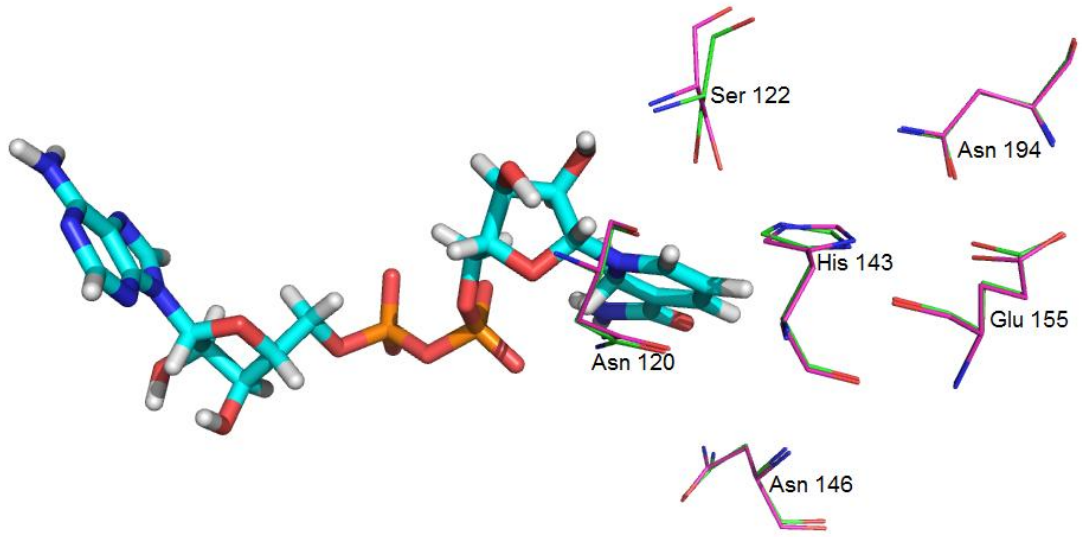


Figure S3. Conformational change of residue Ser122 upon NADH binding.

The active-site residues are superimposed: the apo-enzyme form (magenta) and the NADH-bound form (green). NADH molecule is colored in cyan. Ser122 undergoes a conformational change in the NADH-bound form. Hydrogen bonds are indicated by yellow dotted lines.

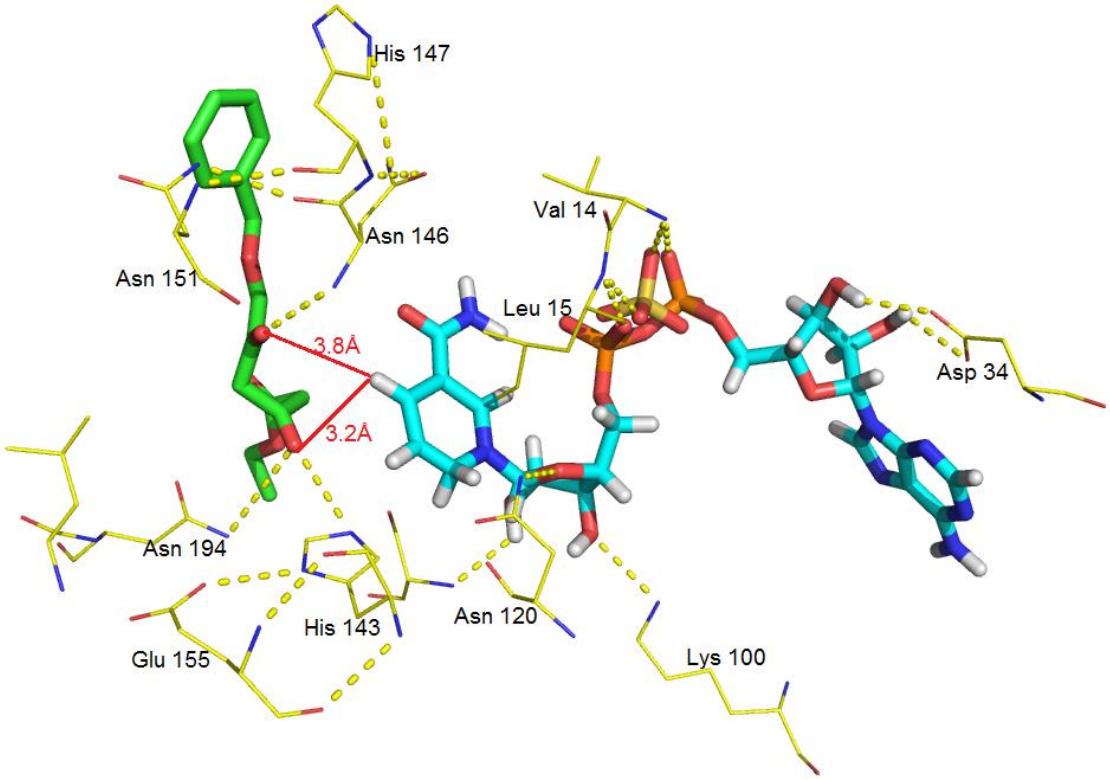


Figure S4. Molecular docking of DKR with substrate 1.

Close-up view of the modeled substrate **1** molecule (green), the surrounding protein residues and the NADH molecule (blue). The hydrogen-bonding network among the substrate and the protein residues is shown as yellow dotted lines. The distances from H of the NADH molecule to β- and δ-carbonyl group is shown as red lines.

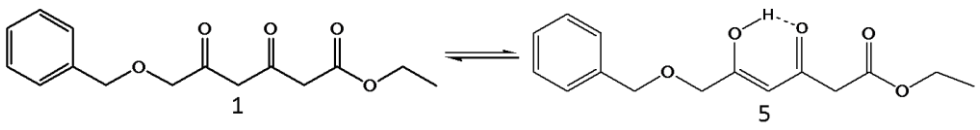


Figure S5. Keto-enol tautomerization of substrate 1

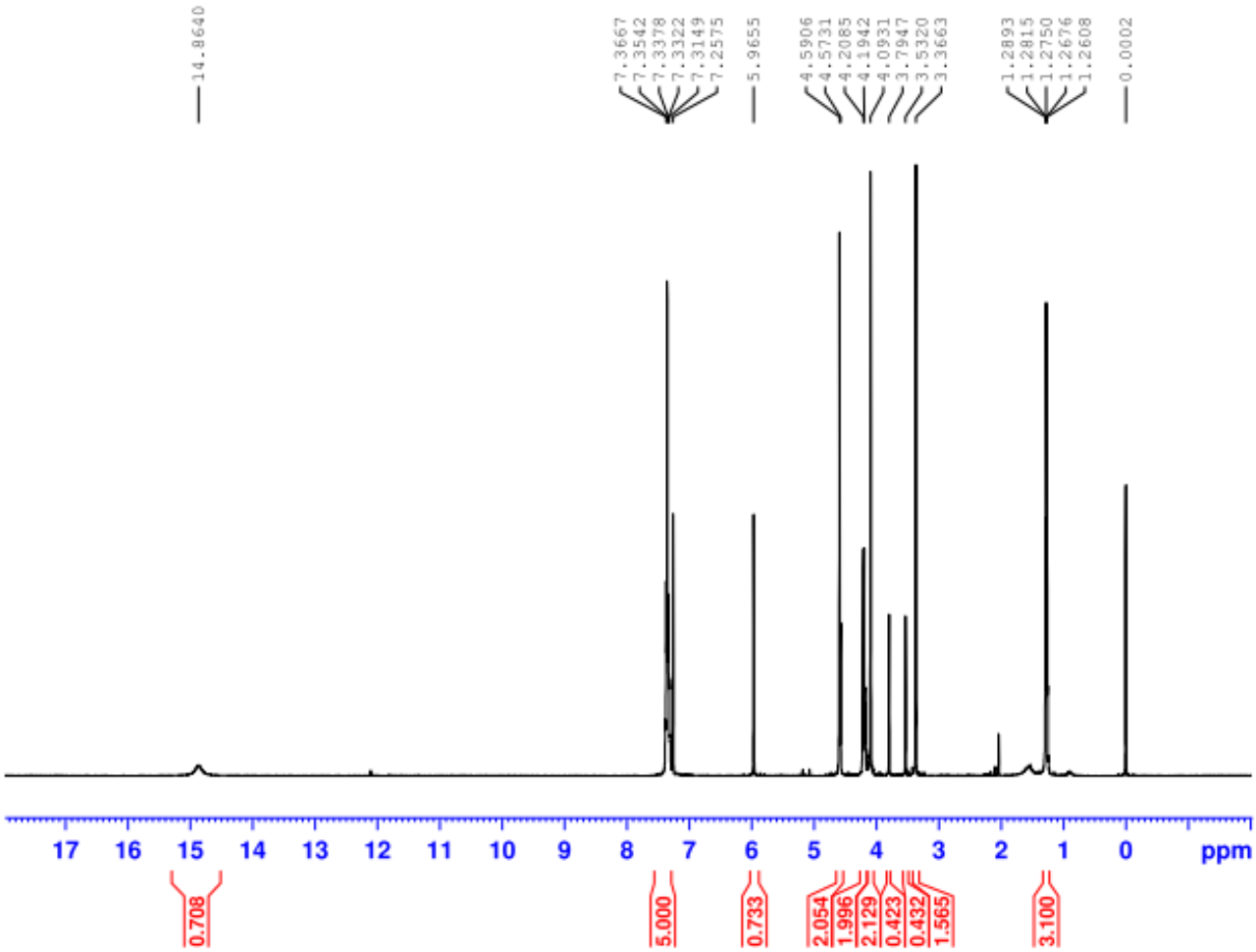


Figure S6. ^{1}H-NMR of substrate 1 (^{1}H: 500MHz**).**

Solvent peak is δ 7.26 ppm. The peaks represented the keto–enol form of **1** (87%) are δ 14.86 (s, 1H, OH) 7.31-7.37 (m, 5H, Ph), 5.97 (s, 1H, H-4), 4.59 (s, 2H, H-6), 4.20 (q, 2H, CO_2CH_2), 4.09 (s, 2H, Ph–CH_2), 3.37 (s, 2H, H-2), 1.28 (t, 3H, CH_2CH_3) ppm.

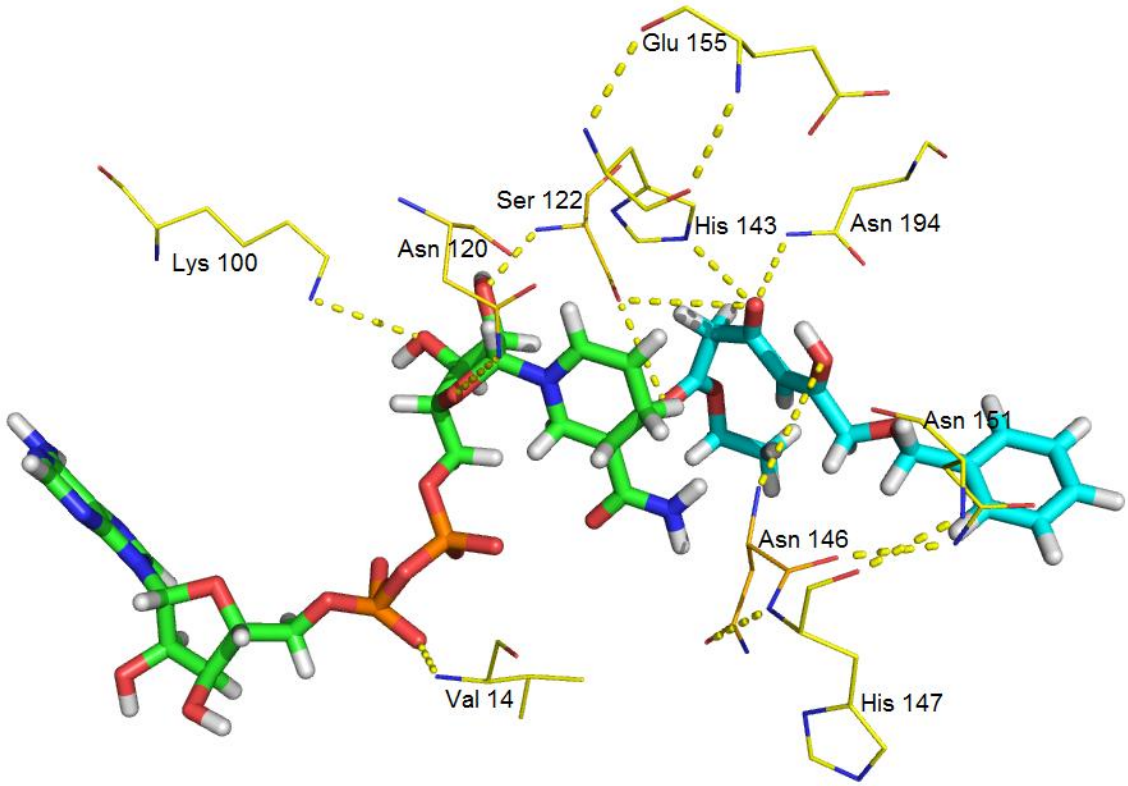


Figure S7. Hydrogen-bonding network of substrate **1** as keto-enol form **5** in DKR active site. The keto-enol form (**5**) of substrate **1** and NADH are colored in cyan and green, respectively. Hydrogen bonds are shown as yellow dashed lines.

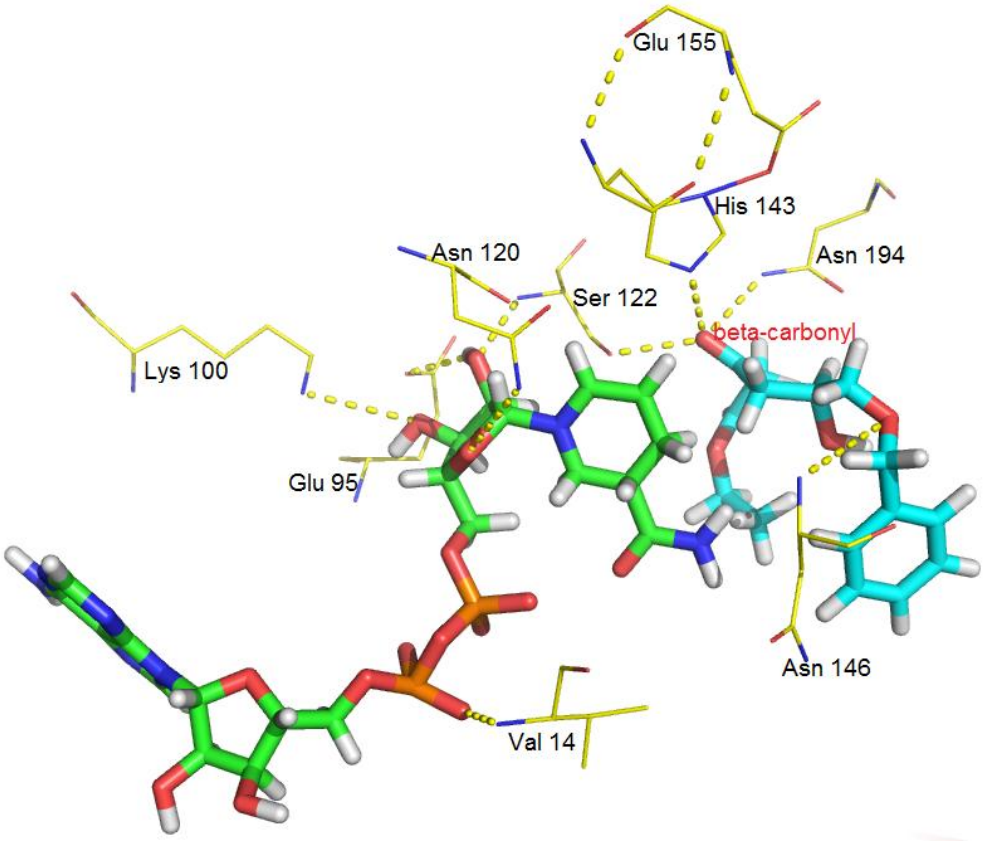


Figure S8. Hydrogen-bonding network of the intermediate **2** in DKR active site. Intermediate **2** and NADH are colored in cyan and green, respectively. Hydrogen bonds are shown as yellow dashed lines.

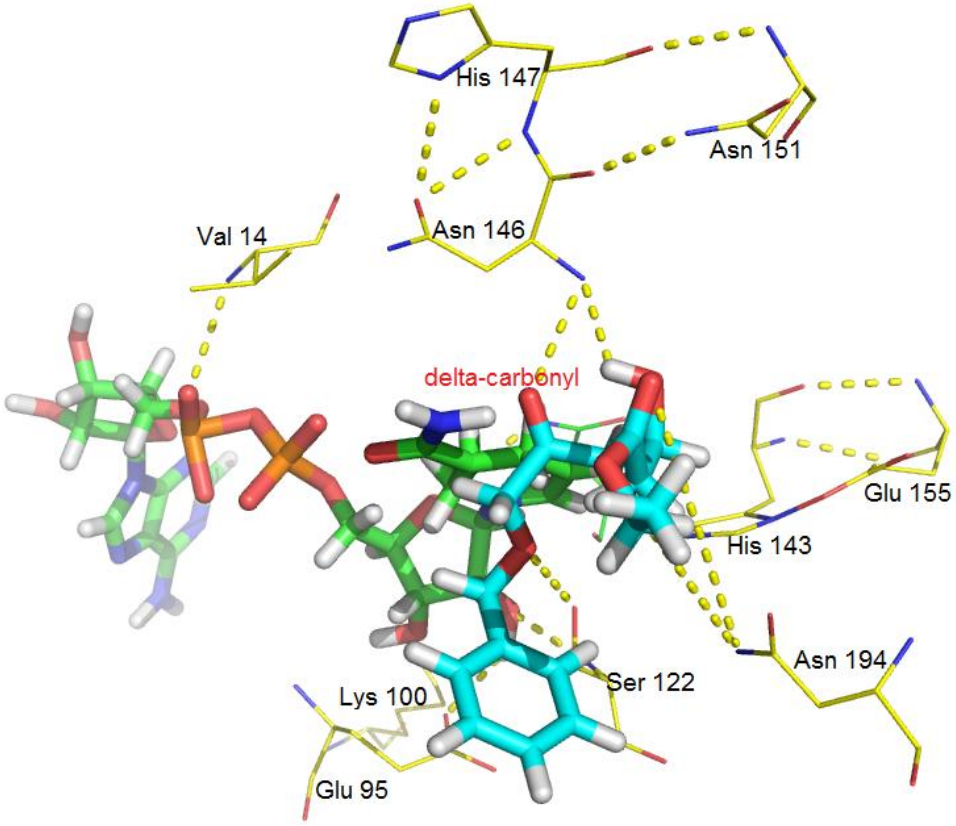


Figure S9. Hydrogen-bonding network of the intermediate **3** in the DKR active site. Intermediate **3** and NADH are colored in cyan and green, respectively. Hydrogen bonds are shown as yellow dashed lines.

Table S1. Data collection and refinement of the X-ray crystallography.

(A) SIRAS Data	Native DKR	KI Soak
Instrument	DIP2030	DIP2030
Space group	P $6_2$22	P $6_2$22
Cell Constants (Å)		
a, b	100.3	100.1
c	133.0	132.9
Resolution range (Å)	40 – 2.7	40 – 2.5
(high resolution bin)	(2.76 – 2.70)	(2.56 – 2.50)
Reflections	11,323 (728)	13,919 (914)
Multiplicity	8.4 (4.9)	8.9 (8.9)
Completeness (%)	99.9 (100)	98.0 (99.3)
I/sigma (I)	10 (2.1)	33 (5.5)
R_{merge} (%)	18.0 (58)	8.7 (46)
SIR Cross-X^2		45
No Sites		8

(B) Data Collection	Apo-DKR	DKR:NADH	DKR:NAD^+
Space group	P $6_2$22	P $6_2$22	P $6_2$22
Cell Constants (Å)			
a, b	99.6	100.8	100.5
c	131.9	132.1	131.9
Resolution range	50 – 1.93	40 – 2.08	40 – 1.95
(high resolution bin)	(1.96 – 1.93)	(2.12 – 2.08)	(1.98 – 1.95)
Reflections (All)	467,105	273,096	135,174
Reflections (Unique)	29,426 (1,362)	24,271 (1,051)	28,421 (987)
Multiplicity	16 (4.6)	11 (3.8)	4.8 (2.0)
Completeness (%)	99.4 (94.6)	99.0 (89.2)	96.2 (68.4)
I/sigma (I)	40 (4.0)	31 (2.1)	34 (1.8)
R_{merge} (%)	6.0 (28)	7.8 (61)	6.1 (58)

(C) Refinement	Apo-DKR	DKR:NADH	DKR:NAD^+
PDB code	4E12	4E13	4DYD
Resolution (Å)	1.93	2.08	1.95
Reflections	29,394	23,105	26,934
Working Set Reflections	27,989	23,089	26,335
Test Set Reflections	1,405	1,133	1,462
R_{factor}	21.3	22.7	21.7
R_{free}	23.3	25.0	24.2
Number of atoms:			
Protein atoms	2,136	2,136	2,136
Solvent molecules	126	80	117
NAD(H)	-	44	44
Average B-factors ($Å^2$)			
Protein	40	46	47
Solvent	60	67	57
Ligand	-	58	73
Bond RMSD (Å)	0.007	0.006	0.004
Ramachandran Plot			
Most Favored (%)	96.4	96.4	96.8
Allowed Regions (%)	3.6	3.6	3.2
Restricted (%)	0	0	0

Table S2. Enzyme activity for wild-type DKR and mutants. [a]

Mutant	Substrate **1**		Intermediate **2**		Intermediate **3**	
	k_{cat} (s^{-1})	K_m (μM)	k_{cat} (s^{-1})	K_m (μM)	k_{cat} (s^{-1})	K_m (μM)
WT	8.30 ± 0.21	267.5 ± 16.7	3.43 ± 0.16	204.2 ± 29.0	2.16 ± 0.06	139.4 ± 13.1
S122K	ND[b]	--	ND	--	ND	--
S122T	ND	--	ND	--	ND	--
H143N	ND	--	ND	--	ND	--
H143Q	ND	--	ND	--	ND	--
E155D	ND	--	ND	--	ND	--
E155S	ND	--	ND	--	ND	--
H147A	2.13 ± 0.06	231.8 ± 16.0	0.69 ± 0.05	200.3 ± 37.4	0.62 ± 0.02	218.7 ± 20.7
H147K	2.00 ± 0.11	228.8 ± 35.4	0.62 ± 0.02	177.5 ± 18.2	0.57 ± 0.01	205.90 ± 9.98
N151Q	5.68 ± 0.34	247.0 ± 42.9	2.82 ± 0.07	296.4 ± 19.7	1.62 ± 0.11	354.8 ± 52.8
N151A	0.72 ± 0.04	287.5 ± 42.5	0.26 ± 0.02	398.0 ± 47.9	0.56 ± 0.04	278.4 ± 49.9
N194A	5.62 ± 0.22	229.4 ± 33.5	0.66 ± 0.01	299.4 ± 40.8	0.57 ± 0.03	192.4 ± 24.4
N146T	0.25 ± 0.02	210.0 ± 53.5	2.55 ± 0.06	206.9 ± 13.7	ND	--
N146A	0.16 ± 0.02	153.2 ± 30.1	3.86 ± 0.13	256.3 ± 20.6	ND	--

[a]The data are an average of three independent experiments with standard deviations. K_m and k_{cat} values were calculated using the Michaelis-Menten model.
[b]ND, not detectable.

References

1. Z. Guo, Y. Chen, A. Goswami, R. L. Hanson, R. N. Patel, Synthesis of ethyl and t-butyl (3R, 5S)-dihydroxy-6-benzyloxy hexanoates via diastereo- and enantioselective microbial reduction, *Tetrahedron: Asymmetry* **2006**, 17, 1589.
2. X. Wu, C. Chen, N. Liu, Y. Chen, Two step process for diketo-reduction catalyzed by diketoreductase, *J. China Pharm. Univ.* **2010**, 41, 408.
3. X. Wu, N. Liu, Y. He, Y. Chen, Cloning, expression, and characterization of a novel diketoreductase from *Acinetobacter baylyi*. *Acta Biochim. Biophys. Sin.*, **2009**, 41, 163.
4. Y. Huang, Z. Lu, N. Liu, Y. Chen, Identification of important residues in diketoreductase from *Acinetobacter baylyi* by molecular modeling and site-directed mutagenesis. *Biochimie*, **2012**, 94, 471.
5. J. J. Barycki *et al*., Biochemical characterization and crystal structure determination of human heart short chain L-3-hydroxyacyl-CoA dehydrogenase provide insight into catalytic mechanism. *Biochemistry*, **1999**, 38, 5786.
6. M. A. White *et al*., Use of complementary cation and anion heavy-atom salt derivatives to solve the structure of cytochrome P450 46A1. *Acta Crystallogr. D Biol. Crystallogr*., **2008**, 64, 487.
7. P. D. Adams *et al*., The Phenix software for automated determination of macromolecular structures. *Methods*, **2011**, 55, 94.
8. L. Bordoli, T. Schwede, Automated Protein Structure Modeling with SWISS-MODEL Workspace and the Protein Model Portal. *Methods Mol. Biol*., **2012**, 857, 107.
9. J. J. Barycki, L. K. O'Brien, A. W. Strauss, L. J. Banaszak, Sequestration of the active site by interdomain shifting. Crystallographic and spectroscopic evidence for distinct

conformations of L-3-hydroxyacyl-CoA dehydrogenase. *J. Biol. Chem.*, **2000**, 275, 27186.

10. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr*., **2010**, 66, 486.
11. A. T. Brünger *et al*., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **1998**, 54, 905.
12. R. Singh *et al*., Structure of a glutathione conjugate bound to the active site of aldose reductase. *Proteins*, **2006**, 64, 101.
13. E. E. Scott *et al*. Structure of mammalian cytochrome P450 2B4 complexed with 4-(4-chlorophenyl) imidazole at 1.9-A resolution: insight into the range of P450 conformations and the coordination of redox partner binding. *J. Biol. Chem.*, **2004**, 279, 27294.
14. A. T. Russo, R. D. Malmstrom, M. A. White, S. J. Watowich, Structural basis for substrate specificity of alphavirus nsP2 proteases. *J. Mol. Graph. Model*., **2010**, 29, 46.
15. J. M. Word, S. C. Lovell, J. S. Richardson, D. C. Richardson, Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.*, **1999**, 285, 1735.
16. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **2010**, 66, 12.
17. O. Carugo, Correlation between occupancy and B factor of water molecules in protein crystal structures. *Protein Eng*., **1999**, 12, 1021.
18. G. M. Morris *et al*., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem*., **2009**, 30, 2785.
19. G. M. Morris *et al*., Automated docking using a Lamarckian genetic algorithm and

empirical binding free energy function. *J. Comput. Chem*., **1998**, 19, 1639.

20. C. Kutzner, J. Czub, H. Grubmüller, Keep it flexible: driving macromolecular rotary motions in atomistic simulations with GROMACS. *J. Chem. Theory Comput*., **2011**, 7, 1381.
21. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans, *Interaction Models for Water in Relation to Protein Hydration* (Reidel Publishing Company, Dordrecht, 1981).
22. H. J. C. Berendsen, J. P. M. Postma, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys*., **1984**, 81, 3684.
23. T. Darden, D. York, L. Pedersen, Particle mesh Ewald—an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **1993**, 98, 10089.
24. H. Berk, P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.*, **2008**, 4, 116.