### **Supplementary Information**

## Mechanism of Substrate Recognition and Insight into Feedback Inhibition of Homocitrate Synthase from *Thermus thermophilus*

# Takuya Okada<sup>‡1</sup>, Takeo Tomita<sup>‡1</sup>, Asri Peni Wulandari<sup>‡</sup>, Tomohisa Kuzuyama<sup>‡</sup>, and Makoto Nishiyama<sup>‡§2</sup>

<sup>‡</sup>Biotechnology Research Center, Department of Biotechnology, The University of Tokyo, Tokyo 113-8657, Japan, <sup>§</sup>RIKEN SPring-8 Center, Hyogo 679-5148, Japan

Running title: Crystal structure of homocitrate synthase

<sup>1</sup>These authors contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed: Biotechnology Research Center, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan Tel.: +81-3-5841-3074; Fax: +81-3-5841-8030 E-mail: umanis@mail.ecc.u-tokyo.ac.jp

## Legends for Supplementary Figures

SUPPLEMENTARY FIGURE 1. Comparison of acetyl-CoA binding residues between TtHCS and LiCMS. A, Acetyl-CoA recognition of LiCMS is schematically illustrated. Residues responsible for acetyl-CoA recognition of LiCMS and residues located at the corresponding position in TtHCS (parentheses) are shown. Residues from adjacent subunit are indicated by asterisks. B. stereo view of the surfaces around the acetyl-CoA site of the LiCMS/Zn<sup>2+</sup>/Pyr/acetyl-CoA complex; C, same view of the TtHCS/Cu<sup>2+</sup>/ $\alpha$ -KG complex. Carbon, oxygen, nitrogen, and sulfur atoms are in gray, red, blue, and yellow, respectively. Bound acetyl-CoA in LiCMS is shown with yellow sticks for carbon atoms. Pyruvate in the LiCMS/Zn<sup>2+</sup>/Pyr/acetyl-CoA complex and  $\alpha$ -KG in the TtHCS/Cu<sup>2+</sup>/ $\alpha$ -KG complex are shown by yellow and orange sticks.  $Zn^{2+}$ the TtHCS/Cu<sup>2+</sup>/ $\alpha$ -KG  $Cu^{2+}$ and ions in complex and

LiCMS/Pyr/acetyl-CoA complex are shown as orange and gray spheres, respectively.

SUPPLEMENTARY FIGURE 2. Amino acid sequence alignment among lysine-sensitive HCS, lysine-insensitive HCS (NifV), CMS, and IPMS. Amino acid sequences were aligned by ClustalW (1) and alignment with secondary structures of TtHCS/Cu<sup>2+</sup>/α-KG complex (*TtHCS*-αKG) and TtHCS/Co<sup>2+</sup>/Lys complex (*TtHCS*-Lys) are drawn by ESPript (2). His72 of TtHCS is shown by vertical arrows. TtHCS, *T. thermophilus* HCS; SpHCS, *Schizosaccharomyces pombe* HCS; ScLys20, *Saccharomyces cerevisiae* HCS<sup>Lys20p</sup>; ScLys21, *Saccharomyces cerevisiae* HCS<sup>Lys21p</sup>; PcHCS, *Penicillium chrysogenum* HCS; AvNifV, *Azotobacter vinelandii* NifV; KpNifV, *Klebsiella pneumoniae* NifV; LiCMS, *L. interrogans* CMS; MjCMS, *M. jannaschii* CMS; TtIPMS, *T. thermophilus* IPMS; MtIPMS, *M. tuberculosis* IPMS.

#### References

- 1. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* 22, 4673-4680
- 2. Gouet, P., Courcelle, E., Stuart, D. I., and Mètoz, F. (1999) *Bioinformatics* 15, 305-308



Okada, T., et al., Supplementary Fig. 1

<i>TtHCS</i> -αKG			
TtHCS-Lys	1		
TtHCS SpHCS ScLys20 PcHCS AvNifV KpNifV LiCMS MjCMS TtIPMS MtIPMS	+ MR MSVSEA MTAAKP MSENNEF MVLLPPSLPVCQLKVTAPEFPSNFYLDGDHS MA ME MTKV MM MEKE MT β1	NGTETIKPPMNGNPYGPNPSDFLSRVNN    NPYAAKPGDYLSNVNN    QSVTESTTAPTTSNPYGPNPADYLSNVKN     GFVGIETRQNPHPSASRNPYGHDAGVTDFLSNVSR	WK. FS FQ FQ FQ VI VI VL LE VR IR SESPDAYTESFGAHTIVKPAGPPRVGQPSWNPQRASSMPVNRY α2 β3
<i>TtHCS</i> -αKG	β1	$ \begin{array}{c} \bullet & 0 \\ \eta^1 \\ \eta^$	$\begin{array}{cccc} & & & & \\ & & & & \\ & & & & \\ & & & & $
TtHCS-Lys TtHCS SpHCS ScLys20 ScLys21 PcHCS AvNifV KpNifV LiCMS MjCMS TtIPMS MtIPMS	TTDS TIES TIES LIDS LIDS LIDS LIDS TIES TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDDT TIDS TIDS	Q 2Q 3Q 4Q   TLREGE OF EKANFSTOD VEIAKA, ID EFGIEVIE   TLREGE OF ANAFDTEK, IQIAKA, IDDFGUDYIE   TLREGE OF ANAFDTEK, ILIARA, IDDFGUDYIE   TLREGE OF SANAPTAK, ILIARA, IAELGVELS   TLREGE OF SANAPTAK, ILIARA, IDAGITANG   TLREGE OSAGUAFNADE, IAIARA, IAELGVELS   TLREGE OSAGUAFNADE, INFOLONE   TLREGE OSAGUAFNADE, INFOLONE   TLREGE OSAGUAFNADE, INFOLONE   TLREGE OTRGVS FSTSE   TLREGE OTRGVS SAGUAFNADE, INFOLONE   TLREGE OTRGVSAGUAFNADE,	TT UDURUUUUUU 50 60 70 80 VTT PVA SPOSR KDAEVIAS LGLKAKVVTH I GCR. LDAA LTS PVA SEOSR KD CEA I CKLGLKAK LLTH I RCH. MDDA LTS PVA SEOSR KD CEA I CKLGLKAK LLTH I RCH. MDDA LTS PVA SEOSR KD CEA I CKLGLKAK LLTH I RCH. MDDA LTS PCASEOSR KD CEA I CKLGLKAK LLTH I RCH. MDDA IG I PSM GEE BR VMHA I AC LGLSSR LLAW CRLC. DVDL IG T PAM GDE I AR I QL VRR QL PDATLMTW CRLM. DVDI IAS ARV SKGBLETVQK I MEWAATEQ LTER IEI LGFVDG AG FPVIS GPMBFEAVR I IATEVKGF II I AALARTH. TLDI VG FPSASQTDFDFVRE I I EQGAI PDDVT I QVLTQ CRPE.LIER
<i>TtHCS</i> -αKG	$\alpha^{\alpha_3}$	α4	$\beta 5$ $\alpha 5$ $\eta^2$ $\beta 6$
TtHCS-Lys	$\alpha^{3}$ $\alpha^{3}$ $\beta^{2}$ $\beta^{3}$ $\gamma^{3}$ $\beta^{3}$ $\gamma^{3}$ $\gamma^{3}$	α4	$\beta^{\beta}$ $\gamma^{2}$ $\alpha^{\beta}$ $\eta^{3}$ $\beta^{\beta}$
TtHCS SpHCS ScLys20 ScLys21 PcHCS AvNifV KpNifV LiCMS MjCMS TtIPMS MtIPMS	KVAVETG.VQGTDLLFGTSKVLRA.     RVAVETG.VDGVDVVIGTSVLRKYLRA.     RVAVETG.VDGVDVVIGTSVLRQY     RVAVETG.VDGVDVVIGTSVLRQY     RVAVETG.VDGVDVVIGTSVLRQY     RIAVETG.VDGVDVVIGTSVLRQY     RIAVETG.VDGVDVVIGTSVLRQY     RIAVETG.VDGVDVVIGTSVLRQY     RIAVETG.VDGVDVVIGTSVLRQY     RIAVETG.VDGVDVVIGTSVLRQY     RQSADLG.IDWVDVISIPASDKLRQY     NKTVDWIKDSG.AKV.LINLIKSLHLEK     DAALECD.VDSVHLVVPTSPIHMKY     DQAA.KALEKAEKPRIHVFTSASKVHLQRY	PHGRDIPRIIEEAREVIAY.IREAAPHVEVR SHGKDMTYIIIDSATEVINF.VKSKG.IEVR SHGKDMTYIIAKSAVEVIEF.VKSKG.IEIR SHGKDMTYIIAKSAVEVIEF.VKSKG.IEIR SHGKDMTYIIKNAAIEVIEF.VKSKG.IEIR SHGKDMTYIIKNAAIEVIEF.VKSKG.IEIR CKLRRPLAVLLEVARLVGE.ARMAG.LEVC KLRRPLAVLLERLAMFIHL.AHTLG.LKVC QLGKTPKEFFTDVSFVIEY.AIKSG.LKIN KLRKTEDEVLETALKAVEY.AKEHG.LIVE MLRKTEEEVLEMADRMVRY.ARRYV.DDVE VFRANRAEVQAIATDGARKCVEQAAKYPGTQWR	F. SAED.   TPRSEEQDILLAVYEAVA.PYVDRVGLA     F. SSEDSPRSDLVDLLSVYKAVDKIGVNRVGIA     F. SSEDSPRSDLVDLLNIYKTVDKIGVNRVGIA     F. SSEDSPRSDLVDLLNIYKTVDKIGVNRVGIA     F. SSEDSPRSDLVDLLNIYKTVDKIGVNRVGIA     F. SSEDSPRSDLVDLLNIYKTVDKIGVNRVGIA     F. SSEDSPRSDLVDLLNIYKTVDKIGVNRVGIA     F. SSEDSPRSDLVDLLNIYKTVDKIGVNRVGIA     F. SSEDSPRSDLVDLLSIYSAVDKIGVNRVGIA     F. SSEDSPRSDLVDLLSIYSAVDKIGVNRVGIA     F. SSEDSPRSDLVDLLSIYSAVDKIGVNRVGIA     F. SSEDSPRSDLVDLLSIYSAVDKIGVNRVGIA     F. SSEDSPRSDLVDLLSIYSAVDKIGVNRVGIA     J. GCEDASRASGTLRAIAEVAQAAGARPLRFA     I. GCEDASRASGTLRAIAEVAQAAGARPLRFA     J. SAEDASRASGVIKALYEVASKEHIERIFLP     L. SAEDATRSDVNPLIKLVEVASKEHIERIFLP     L. SAEDATRSDVNPLIKLVEVAIEAGATTINIP     FEYSPESYTGTELEYAKQVCDAVGEVIAPTPERPILT
<i>TtHCS</i> -αKG	►TT <u>2000000000</u> <u>20</u> .0000	$\beta 7$ TT 000000000 $\beta 8$ $\eta$	3 α8 α9 η4 α10 2 ΤΤ <u>2020, 2020, 2020, 2020, 2020, 2020, 2020</u> 4 ΤΤ <u>2020, 2020, 2020, 2020, 2020</u>
TtHCS-Lys	►TT QUQUUQUU	$\rightarrow \qquad TT \qquad 2000 \qquad 210 \qquad 220$	2 TT 2020. 2020202 2 2020202 2 202020202
TtHCS SpHCS ScLys20 ScLys21 PcHCS AvNifV KpNifV LiCMS MjCMS TtIPMS MtIPMS	DTVGVATDRQVYALVREV.RRVV.GP DTVGCATDRQVYALVRTL.RGVV.S. DTVGCANDRQVYELIR.TL.KSVV.S. DAVGCANDRQVYELIR.TL.KSVV.S. DTVGCASDRQVYELVR.VL.RGVV.GD DTVGVMEPGMLDRFR.FLSRRL. DTVGVMEPGMLDRFR.FLSRRL. DTVGVLDPFTAAQIS.AL.RDVW. DTLGVLSDEFFQGVD.SLIQKYP. DTVGVLSDEFFQGVD.KLI.TENV. DTTGYGTDQKSQELFK.KI.TENV. DTTGYGTDQEYGALIR.RI.RDEVVRGR ATVEMTTNVYADSIEWMSRNL.ANRE.	VVDTEPHGHNDTGCAINNAYEAIEAGATHVDTTIL .CDLECHFHNDTGMAINAYCALEAGATHVDTTIL .CDLECHFHNDTGMAINAYCALEAGATHIDTSIL .CDLECHFHNDTGCAINNAYTALEGGARLIDVSVL .CDLECHFHNDTGCAINNAYTALEGGARLIDVSVL .CDLECHFHNDTGCAINNAFCALEAGATHIDTSVL DMELEVHAHDDFGLAINNAFCALEAGATHIDTSVL DMELEVHAHDDFGLATNTLAAVMGGATHINTTVL DIHPEFHGHNDYDLSWATANTLAAVMGGACSVGLHASIN NLPVSVHCHNDFGMATANTCSAVLGGAVQCHVTVN DVILSTHTHDDLGMATANTCSAVLGGAVQCHVTIN DVILSTHTHDDLGMATANTGSAVLGGAQQIECTIN SVILSLHPHNDRGTAVAAELGFAAGADRIEGCLF	GIGERNGITPEGGF.LARMYTLQP.EYVRRYKLEMEPEIDRM GIGERNGITPEGGF.LARMYVTDR.EYVRRYKLEMEPEIDRM GIGERNGITPEGGL.MARMYVTDR.EYVITHKYKLNQLRELENL GIGERNGITPEGGL.MARMIVAAP.DYVRSYKLEKIRDIENL GIGERNGITFEGGL.MARMIVAAP.DYVRSYKLEKIRDIENL GIGERNGITFEGL GECRAGNAALEEC.VIALKKLHGVETGVHFSALPALCQR GLGERAGNAALEEC.VIALKKLHGVETGVHFSALPALCQR GLGERAGNASLEEV.VNALKILYGYDTKIKMEKLYEVSRI GIGERAGNASLEEV.VMALKILYGYDTKIKMEKLYEVSRI GIGERAGNASLEEV.VMALKILYGYDTKIKMEKLYEVSRI GIGERAGNASLEEV.VMALKILYGYDTKIKMEKLYEVSRI GIGERAGNASLEEV.VMALKILYGYDTKIKMEKLYEVSRI
<i>TtHCS</i> -αKG	<u>β9</u> α11 • 00000	η <sup>5</sup> η <sup>5</sup> η <sup>6</sup> ος	β10
TtHCS-Lys	22222 TT		α11 200000000
TtHCS SpHCS ScLys20 ScLys21 PcHCS AvNifV KpNifV LiCMS MjCMS TtIPMS MtIPMS	VAR MÜGVET PFINNYITGETÄFSHKÄGMELKA VAD AVEVQTPFNNYITGETÄFSHKÄGMELKA VAD AVEVQTPFNNYITGETÄFSHKÄGMELKA VAD AVEVNIPFNNPITGPCAFTHKÄGTHÄKA VAD AVEVNIPFNNYITGPCAFTHKÄGTHÄKA VAD AVEVNIPFNNYITGPCAFTHKÄGTHÄKA VAE AVENNIPFNNYITGPCAFTHKÄGTHÄKA VER ASGROVAWOKSVVCAGVFTHEAGTHVDG VAE AAQRAIDPQQPLVGELVFTHESGVHVAA VEV FSGKRISANRPIVGEDVFTQTAGVHADG VSR LMKLPVPPNKAIVGDNAFAHESGTHVDG VER YTGMPVPPNKAIVGDNAFAHESGTHVDG VER YTGMPVPPNKAIVGDNAFAHESGTHODG	J SOQ     J SIQ       J YIN PE     AYEPYPE       I LAN PS     TYE ILLKFED       I LAN PS     TYE ILDFHD       I LNN PS     TYE ILDFHD       L KIR R     YA PILLKFEN       DKKGNL     YAN PILFER       L INN TE     TYE PILKPEM       VI KIR RA     TYE INDA&L       I NKG L DAMKLDADAAD CD V DDM LWQ V PY L PI DPR D	FGVERKLII I.ASELTG HAIKARAE.ELGL HYGEEELHEVTQ FGMERYVHV.GSRLTGWNAIKSRAE.QLNL.HLTDAQAKELTV FGMERYVHV.GSRLTGWNAIKSRAE.QLNL.HLTDAQAKELTV FGMERYVHF.ASELTGWNAIKSRVD.QLNL.NLTDDQIKEVTA FGMERYVHF.ASELTGWNAIKSRVD.QLNL.NLTDDQIKEVTA FGMERYVHF.ASELTGWNAIKSRAO.QLKLEMTTDQVIKEVTA FGMERYVHF.ASELTGWNAIKSRAO.QLKLEMTTDQIKEVTA FGMERYVHF.ASELTGWNAIKSRAO.QLKLEMTTQVKECTA GRESYRLV.LGKHSGANAVNGVFD.QMGY.HLNAAQINQLLP FGRERSYA.LGKLAGKASISENVK.QLGM.VLSDVVLQKVLE FGRERSYL.LGKLAGKASISENVK.QLGM.VLSDVVLQKVLE FGRERSYL.LGKLAGKASISENVK.QLGM.VLSDVVLQKVLE FGRERPAVLV.LGKHSGRAAFKKALE.DLGYKDLSEEEVKKLFA VGRERTYEAVIRVSQSGKGCVAVIMKTDHGL.SLPRRLQIEFSQ
TtHCS-Lys	α12 2020202 2020. 2020202		
TtHCS SpHCS ScLys20 ScLys21 PcHCS AvNifV KpNifV LiCMS MjCMS TtIPMS MtIPMS	360     370       HIKALADRGQ.LTLEEL.DRILREW.       RIKKLADVRT.LAMDDV.DRVLREYH       KIKKLGDVRS.LNIDDV.DSIIKNFH       KIKKLGDVRP.LNIDV.DSIIKNFH       KIKKLGDVRP.LNIDV.DSIIKNFH       KIKKLGDVRP.LNIDV.DSIIKNFH       KIKKLGDVRP.LNIDV.DSIIKNFH       KIKAMADIRP.LAVDA.DSIIRAYH       RIRAFSTRTKRRSPQ.PAELQDFY       RIRAFSTRTKRRSPC.PAELQDFY       RVIELGDQNKLVTPEDL.PFIIADVS       RVKEFGDLGKYISDADL.LAIVREVT       FKEIAEKKGPLSAEEL.QALVESER       VIQKIAEGTAGEGGGE.VSPKEMWDAFAEEYL	ADLSDAD. AEVSTPQVL AELSTPLK RNLKSGENK RQLCEQ. GRTGEKVLTI.KSCNIHSGIGIRPHAQIELEYQG. GKLVEEKIKL.DELTVVSGNKITPIASVKLHYKGE E.PTSHFFQL.EHVQFFSGSGLLPTATVKVKTP. A.PVRPLERIRQHVDAADDDGGTTSITATVKIN.	SAKKN PVN PVN PLLDL 
TtHCS_I ve			۵
TTHCS-LYS			Υ ΤΨλ
TTHES SpHCS ScLys20 ScLys21 PcHCS AvNifV KpNifV LiCMS MjCMS TTIPMS	.G.N. PGGKTDAL. VETRITWNKSLDLEE. GG.GTDAL. IEVVVKLRKGT. TG.STEAL. GQVTVKLRKGE.	. KKND . KGTD . TA . DQTFKTMGVH . EIVEVRKSD . LQAVGVGVS . LQAVGVGVS	RI