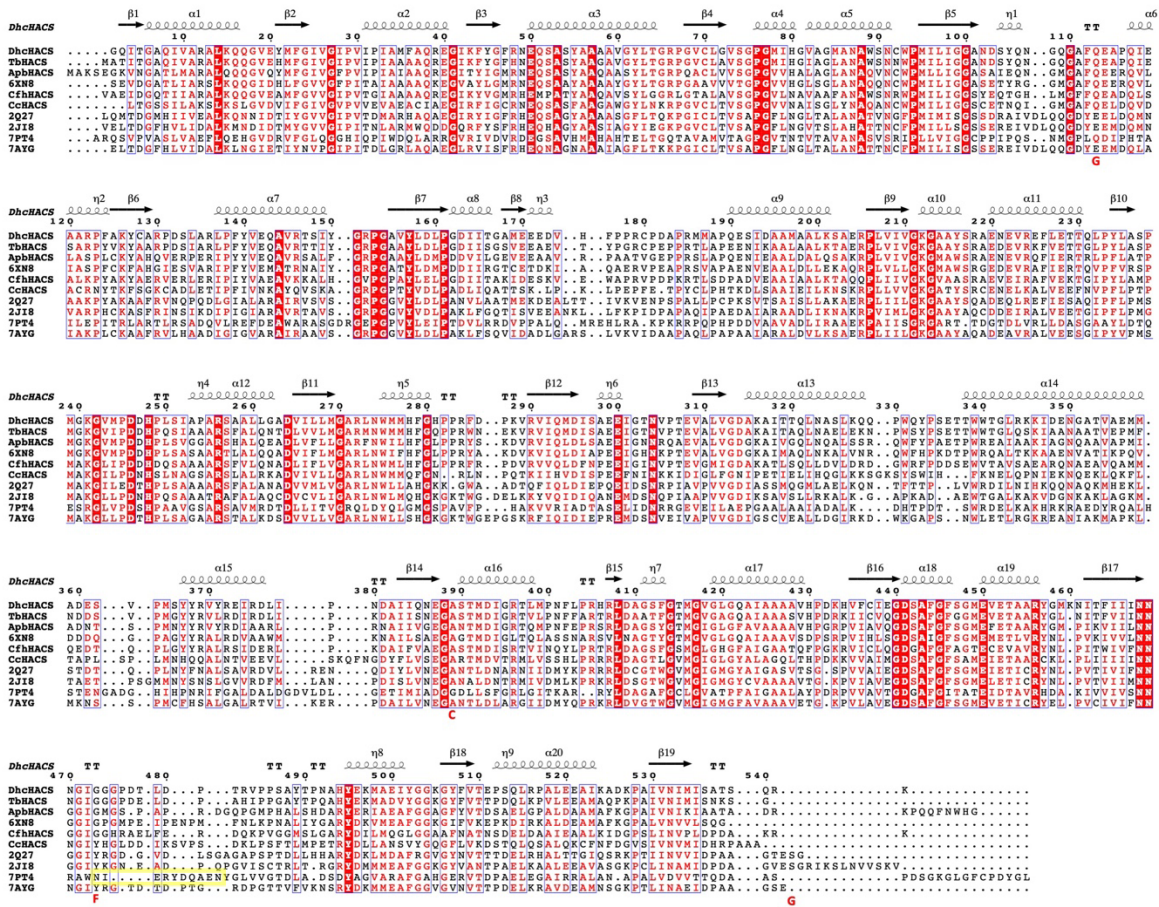
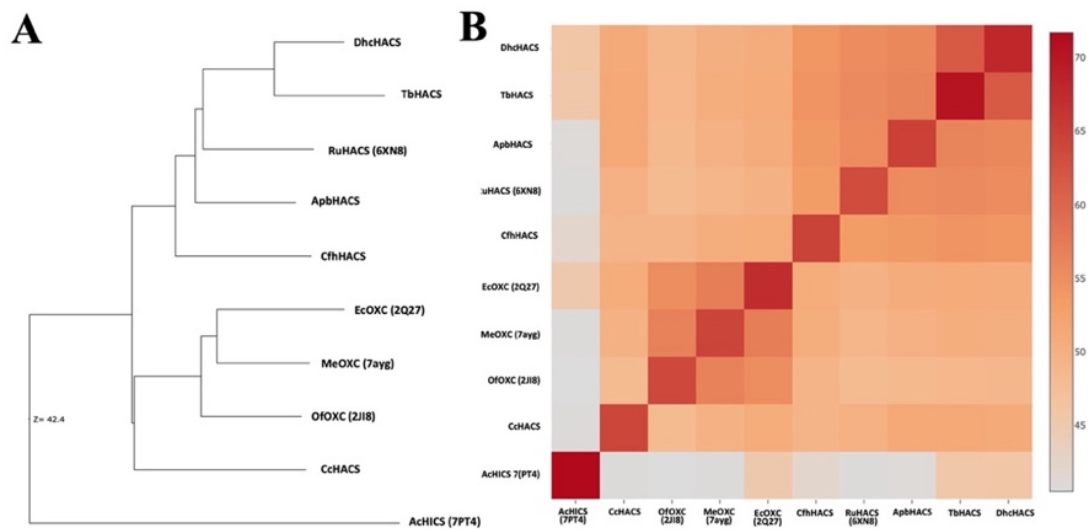


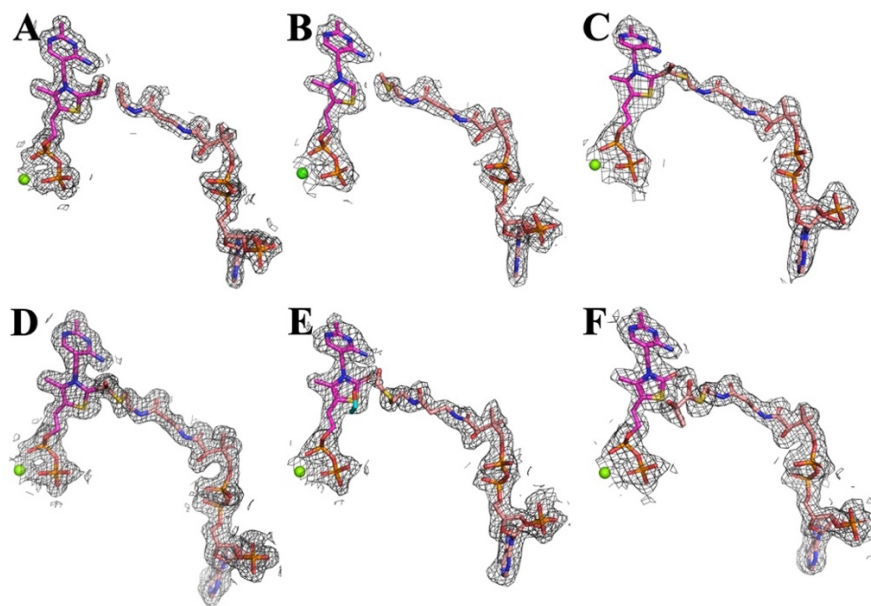
Supplementary figures



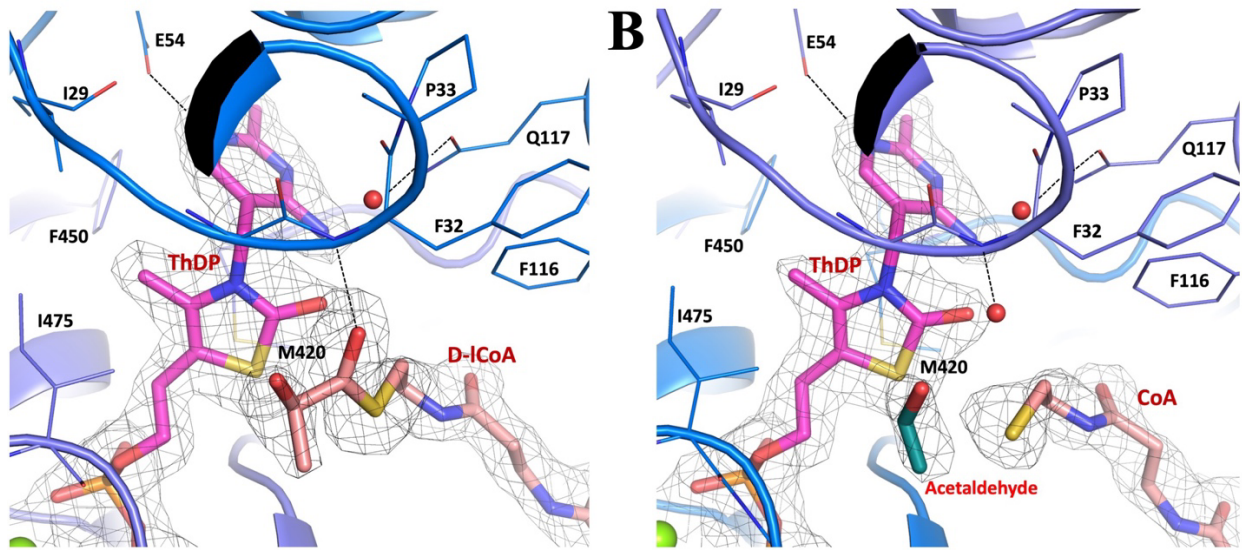
Supplementary Figure S1. Sequence alignment of HACL/S including RuHACL/S, OXC and AcHACL/S sub-families. Y497F, S568G, E135G, and A415C mutations from MeOXC (PDB id: 7ayg) to MeOXC4 (PDB id: 7b2e)¹ are indicated in red just below 7ayg sequence. Yellow box indicates the α -helix in AcHACL/S (PDB id: 7pt4) distinct from other enzymes.



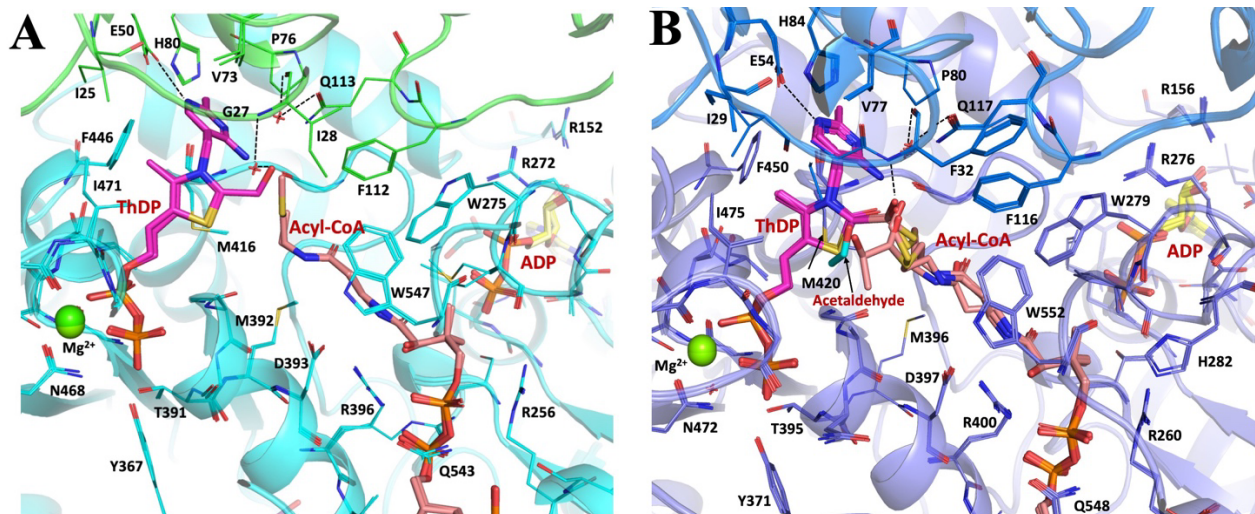
Supplementary Figure S2. Structural similarity tree (A) and dendrogram (B). The dendrogram is derived by average linkage clustering of the structural similarity matrix (Dali Z-scores)².



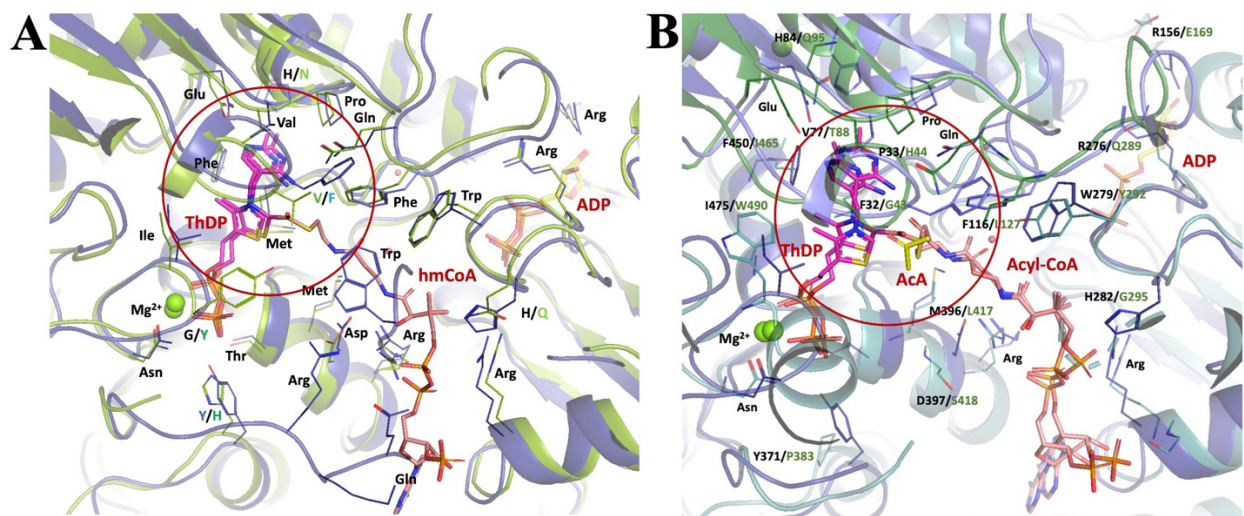
Supplementary Figure S3. Electron density for ThDP and acyl-CoA molecules found in the TbHACL/S-fCoA (A), TbHACL/S-hCoA (B), ApbHACL/S-CoA (C), ApbHACL/S-fCoA (D), ApbHACL/S-L-ICoA (E), and ApbHACL/S-D-ICoA (F) complexes. Magnesium ion is shown as green sphere. The electron density (composite annealed map, 2DFo-mFc) for the ligands is contoured at 1.0 σ level.



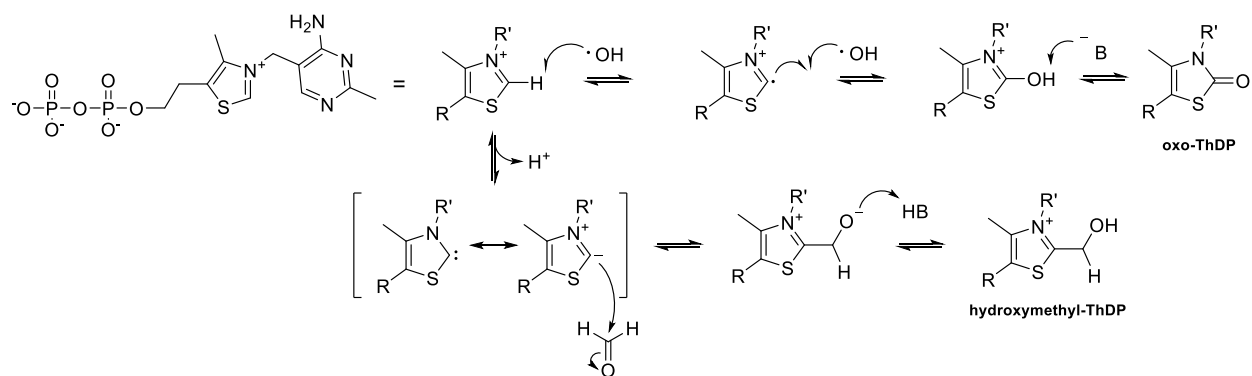
Supplementary Figure S4. Active sites of ApbHACL/S-D-lCoA contain two different species of acyl-CoA, D-lCoA (A) and hydrolyzed to acyl-CoA plus acetaldehyde (B). ThDP and acyl-CoA molecules shown in magenta and orange sticks, respectively and acetaldehyde in dark green. The electron density (composite annealed map, 2DFo-mFc) for the ligands is contoured at 1.0 σ level.



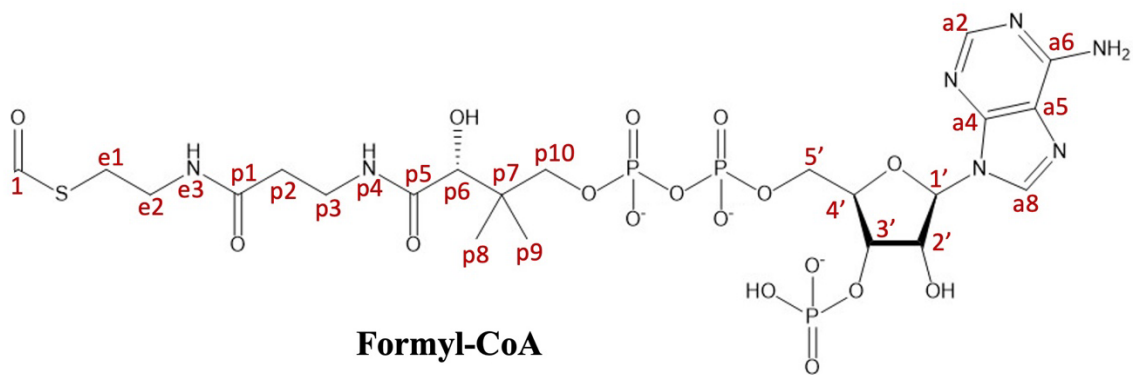
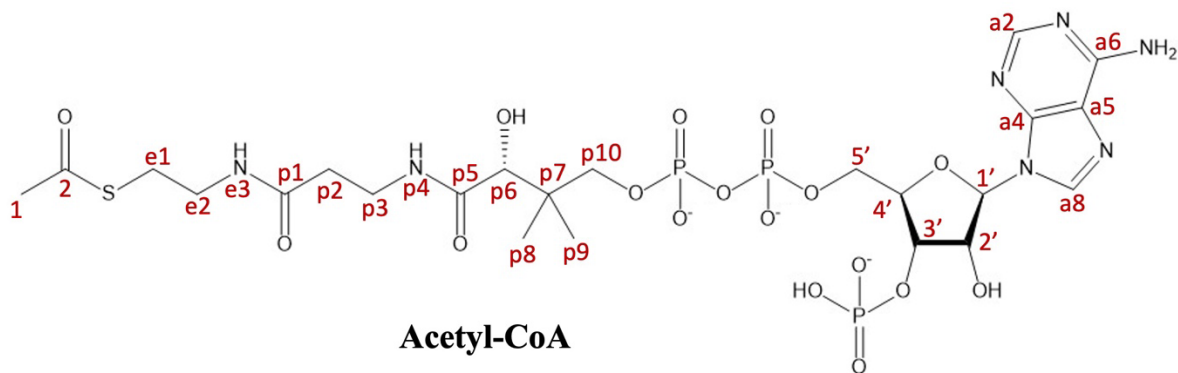
Supplementary Figure S5. Structure of active sites of TbHACL/S complex (A) and ApbHACL/S complex (B) showing interaction of enzymes with ThDP and acyl-CoAs. TbHACL/S in cyan and green cartoon drawing, ApbHACL/S in fade blue and dark blue, ThDP in magenta sticks and acyl-CoAs in orange.



Supplementary Figure S6. Differences between active sites in enzymes (A) CcHACL/S (light green) and ApbHACL/S (dark blue) and (B) ApbHACL/S (dark blue) and AchHACL/S (pale green). Active sites in circles. ThDP in magenta sticks and acyl-CoAs in orange sticks.



Supplementary Figure S7. Possible mechanisms for generation of “oxo-ThDP” and “hydroxymethyl-ThDP” via a ROS mechanism (an example is provided with hydroxyl radical) or an attack of formaldehyde, respectively.



Supplementary Figure S8. Carbon atom numbering schemes for acetyl-CoA and formyl-CoA

Supplementary tables

Supplementary Table S1. Crystal data and refinement statistics

	<i>DhcHACL/S</i>	<i>CcHACL/S</i>	<i>TbHACL/S</i>	<i>CfhHACL/S</i>
<i>Data Collection</i>	19-ID SBC	19-ID SBC	19-ID SBC	19-ID SBC
<i>Space group</i>	<i>C</i> 2	<i>P</i> 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 3 ₁ 2 ₁
<i>Cell dimensions</i>				
<i>a, b, c (Å)</i>	324.22, 77.84, 229.55, $\alpha=\gamma=90$, $\beta=134.5$	64.74, 179.29, 96.83, $\alpha=\gamma=90$, $\beta=108.2$	120.59, 134.01, 173.39, $\alpha=\beta=\gamma=90$	<i>a</i> = <i>b</i> = 111.33, <i>c</i> =193.16, $\alpha=\beta=90$, $\gamma=120$
<i>Resolution range (Å)^a</i>	2.70-50.00 (2.70-2.75)	1.85-50.00 (1.85-1.88)	2.25-50.00 (2.25-2.29)	2.65-50.00 (2.65-2.70)
<i>Unique reflections^a</i>	112,373 (5,449)	167,273 (6,685)	133,571 (6,589)	39,660 (1,349)
<i>R-merge^b</i>	0.175 (1.686)	0.107 (0.818)	0.159 (1.196)	0.162 (0.913)
<i>Mean I/σ(I)</i>	10.1 (1.1)	14.7 (1.3)	10.3 (1.6)	13.0 (1.0)
<i>CC_{1/2}^c</i>	0.996 (0.423)	0.996 (0.791)	0.983 (0.659)	1.0 (0.837)
<i>Completeness (%)</i>	99.7 (97.8)	94.1 (76.1)	99.9 (100)	97.1 (68.4)
<i>Redundancy</i>	8.5 (7.0)	7.8 (5.8)	5.2 (5.0)	14.2 (8.7)
<i>Wilson B-factor (Å²)</i>	66.0	22.4	27.6	59.2
<i>Refinement</i>				
<i>Resolution range (Å)</i>	2.70-40.53 (2.70-2.73)	1.85-44.82 (1.85-1.87)	2.25-43.39 (2.25-2.27)	2.65-48.23 (2.65-2.72)
<i>Reflections work/test</i>	106,761/5,492	158,800/8,345	126,194/6,707	37,562/1,913
<i>R_{work}/R_{free}</i>	0.194/0.235	0.156/0.196	0.160/0.195	0.208/0.244
<i>Number atoms</i>	33287	18514	17416	8326
<i>Protein</i>	32822	16813	16299	8195
<i>Ligand/ion</i>	460	298	297	127
<i>Water</i>	5	1403	820	4
<i>Protein residues</i>	4317	2234	2175	1092
<i>RMSD (bonds) (Å)</i>	0.002	0.006	0.004	0.002
<i>RMSD (angles) (°)</i>	0.513	0.876	0.690	0.481
<i>Ramachandran favored^c (%)</i>	96.96	97.72	98.23	97.22
<i>allowed (%)</i>	3.01	2.19	1.77	2.59
<i>outliers (%)</i>	0.02	0.09	0.0	0.19

<i>Rotamer outliers (%)^d</i>	1.81	0.87	0.72	2.67
<i>Clashscore</i>	4.73	2.64	1.64	3.25
<i>Average B-factor (Å²)</i>	77.2	29.7	32.5	69.6
<i>Protein</i>	77.3	29.6	32.5	69.8
<i>Ligand/ion</i>	66.3	26.5	32.0	62.3
<i>Water</i>	74.2	33.6	33.0	54.6
<i>Number of TLS groups</i>	53	31	23	12
<i>PDB ID</i>	8vzj	8vzi	8vzh	8vzk

	<i>TbHACL/S-fCoA</i>	<i>TbHACL/S-hCoA</i>	<i>ApbHACL/S-CoA</i>	<i>ApbHACL/S-fCoA</i>	<i>ApbHACL/S-L-lCoA</i>	<i>ApbHACL/S-D-lCoA</i>
<i>Data Collection</i>	23-IDB GMCA	19-ID SBC	19-ID SBC	19-ID SBC	19-ID SBC	23-IDB GMCA
<i>Space group</i>	C222 ₁	C222 ₁	P2 ₁ 2 ₁ 2 ₁	P1	P1	P2 ₁ 2 ₁ 2 ₁
<i>Cell dimensions</i>						
<i>a, b, c (Å)</i> ,	100.17, 182.51, 123.92,	100.62, 182.77, 123.84,	69.80, 178.13, 196.48,	69.08, 102.44, 103.98, 113.8,	69.41, 102.68, 104.34, 114.1,	69.71, 178.00, 196.11,
<i>α, β, γ (°)</i>	α=β=γ=90	α=β=γ=90	α=β=γ=90	96.8, 105.5	96.7, 105.2	α=β=γ=90
<i>Resolution range (Å)^a</i>	2.20-50.00 (2.20-2.24)	1.70-50.00 (1.70-1.73)	2.36-50.00 (2.36-2.39)	1.72-50.00 (1.72-1.76)	2.05-50.00 (2.05-2.09)	1.98-50.00 (1.98-2.08)
<i>Unique reflections^a</i>	56,818 (2,390)	123,996 (5,095)	99,068 (4,335)	23,8150 (1,487)	144,733 (5,583)	165,964 (20,426)
<i>R-merge^b</i>	0.233 (0.678)	0.188 (0.850)	0.263 (1.126)	0.052 (0.338)	0.189 (0.671)	0.275 (2.039)
<i>Mean I/σ(I)</i>	6.8 (1.0)	14.0 (1.2)	6.2 (1.1)	7.2 (1.6)	7.9 (1.4)	7.8 (0.9)
<i>CC_{1/2}^c</i>	0.977 (0.576)	0.984 (0.506)	0.960 (0.548)	0.996 (0.806)	0.928 (0.490)	0.995 (0.435)
<i>Completeness (%)</i>	98.1 (83.7)	98.3 (81.4)	97.2 (85.6)	92.3 (59.9)	94.9 (73.2)	97.3 (83.9)
<i>Redundancy</i>	8.2 (3.0)	8.0 (3.8)	5.8 (4.9)	2.7 (2.6)	3.5 (2.7)	12.0 (7.2)
<i>Wilson B-factor (Å²)</i>	21.6	14.9	32.7	18.4	25.4	25.7
Refinement						
<i>Resolution range (Å)</i>	2.20-37.41 (2.20-2.24)	1.70-48.04 (1.70-1.71)	47.35-2.36 (2.36-2.38)	1.72-38.63 (1.72-1.74)	2.05-49.52 (2.05-2.08)	1.98-47.27 (1.98-2.00)
<i>Reflections work/test</i>	53,986/2,797	117,122/6,044	93,957/4,893	226,339/11,600	137,400/7,055	156,814/8,175
<i>R_{work}/R_{free}</i>	0.198/0.234	0.145/0.175	0.208/0.246	0.160/0.199	0.209/0.243	0.182/0.219
<i>Number atoms</i>	9071	9574	17547	18812	17923	18742
<i>Protein</i>	8313	8463	16756	16734	16695	16745
<i>Ligand/ion</i>	245	173	454	344	438	514

<i>Water</i>	513	938	337	2078	790	1483
<i>Protein residues</i>	1103	1135	2226	2228	2219	2227
<i>RMSD (bonds) (Å)</i>	0.002	0.011	0.002	0.008	0.002	0.002
<i>RMSD (angles) (°)</i>	0.561	1.138	0.696	1.048	0.666	0.692
<i>Ramachandran favored^c (%)</i>	97.81	98.44	97.25	98.05	97.73	97.69
<i>allowed (%)</i>	2.19	1.56	2.75	1.95	2.23	2.31
<i>outliers (%)</i>	0.00	0.00	0.0	0.00	0.05	0.00
<i>Rotamer outliers (%)^d</i>	1.53	0.82	1.08	0.84	0.85	0.60
<i>Clashscore</i>	2.01	1.06	2.62	1.21	2.07	1.35
<i>Average B-factor (Å²)</i>	23.3	17.5	36.2	22.7	25.1	26.6
<i>Protein</i>	23.3	16.6	36.2	21.7	25.1	26.2
<i>Ligand/ion</i>	23.6	20.0	39.3	26.6	24.1	30.7
<i>Water</i>	23.4	25.7	32.3	30.5	26.7	30.4
<i>Number of TLS groups</i>	13	10	24	23	27	30
<i>PDB ID</i>	8vzd	8vze	8vzf	8vzc	8vza	8vzb

^a Values in parentheses correspond to the highest resolution shell.

^b $R_{\text{merge}} = \frac{\sum_h \sum_j |I_{hj} - \langle I_h \rangle|}{\sum_h \sum_j I_{hj}}$, where I_{hj} is the intensity of observation j of reflection h .

^c As defined by Karplus and Diederichs³.

^d As defined by Molprobity⁴.

Supplementary Table S2: Pairwise Comparisons of RMSD and sequence identity.

The numbers are shown in the following order: pairwise sequence identities (%), RMSDs of /540 superimposed C α atom positions of active site residues/residues binding to ThDP, ADP, or acyl-CoA/residues binding to ThDP or acyl-CoA/residues binding to ThDP. The residues in active sites, binding to ThDP and ADP and used for calculations are listed in Supplementary Table S3.

Name	CcHACL/S	TbHACL/S	CfhHACL/S	ApbHACL/S	RuHACL/S	OfOXC (2ji8)	EcOXC (2q27)	MeOXC (7ayg)	AcHICL/S (7pt4)
DhcHACL/S	43.2/1.33 0.56/0.55 0.51	79.0/0.47 0.23/0.29 0.28	51.5/1.00 0.54/0.51 0.52	61.7/0.74 0.35/0.42 0.40	53.7/0.87 0.48/0.40 0.37	40.4/1.57 0.60/0.62 0.63	41.8/1.23 0.65/0.62 0.66	41.4/1.31 0.74/0.69 0.75	25.5/1.92 1.49/1.45 1.47
CcHACL/S		41.7/1.28 0.62/0.49 0.44	39.0/1.53 0.80/0.61 0.58	42.2/1.25 0.57/0.55 0.49	40.4/1.25 0.63/0.56 0.49	40.2/1.55 0.57/0.72 0.72	41.9/1.36 0.56/0.58 0.55	37.8/1.48 0.58/0.80 0.79	23.5/1.71 1.37/1.35 1.38
TbHACL/S			52.5/0.99 0.54/0.44 0.44	60.7/0.70 0.36/0.29 0.29	55.3/0.78 0.45/0.28 0.25	40.0/1.60 0.67/0.67 0.67	41.1/1.24 0.70/0.55 0.55	41.2/1.35 0.76/0.78 0.76	24.5/1.92 1.52/1.35 1.36
CfhHACL/S				51.1/0.99 0.58/0.52 0.53	47.9/1.23 0.67/0.41 0.35	38.1/1.44 0.83/0.78 0.79	42.5/1.33 0.84/0.73 0.74	39.8/1.38 0.92/0.90 0.90	29.6/1.85 1.46/1.35 1.34
ApbHACL/S					62.1/0.73 0.49/0.39 0.37	37.7/1.49 0.65/0.69 0.69	39.6/1.24 0.65/0.59 0.59	39.8/1.40 0.73/0.79 0.77	29.6/1.82 1.49/1.41 1.40
RuHACL/S (6xn8)						38.6/1.53 0.71/0.77 0.75	41.2/1.33 0.78/0.67 0.67	39.1/1.51 0.85/0.88 0.85	27.3/1.76 1.49/1.46 1.45
OfOXC (2ji8)							55.7/0.82 0.43/0.42 0.43	62.0/0.79 0.35/0.38 0.37	22.9/1.85 1.32/1.27 1.29
EcOXC (2q27)								61.3/0.74 0.43/0.43 0.42	24.6/1.85 1.26/1.13 1.14
MeOXC (7ayg)									26.2/1.88 1.27/1.15 1.17

Supplementary Table S3: Residues involved in interacting with ThDP, acyl-CoA, and ADP

Name	Catalytic site-ThDP	Pyrophosphate-ThDP	Acyl-CoA	ADP
DhcHACS	G27-P29, E50, F112, Q113, G414, M416,	A389-T391, Y367, I472-G473, D441-A443, N469,	R256, R396,	R152, R272, K212, D295-I296, S237, A315,
CcHACS	G29-P31, E52, F114, Q115, G416, M418,	A391-T393, H365, I473-Y474, D443-A445, N470),	R256, R398,	R154, R272, K212, D293-I294, T237, I313,
TbHACS	G27-P29, E50, F112, Q113, G414, M416,	A389-T391, Y367, I471-G472, D441-A443, (N468	R256, R396,	R152, R272, K212, D295-I296, S237), A315,
CfhHACS	G34-P36, E57, F119, Q120, G421, M423,	A396-T398, Y374, I478-G479, D448-A450, N475	R263, R403,	V159, R279, K219, D302-F303, M244, A322,
ApbHASL	G31-P33, E54, F116, Q117, G418, M420,	A393-T395, I475-G476; Y371, D445-A447, N472),	R260, R400,	R156, R276, K216, D299-L300, T241, G319,
RuHACS (6xn8)	G27-P29, E50, F112, Q113, G414, M416,	A389-T391, Y367, I471-G472; D441-A443, N468),	R272, L396,	R152, R256, K212, D295-I296, S237, G315,
OfOXC (2ji8)	G33-P35, E56, Y120, E121, G426, M428,	A401-A403, Y377, I482-Y483; D452-A454, N479),	R282, R408,	R160, R266, K222, D306-I307, M247, I326,
AcHICS (7pt4)	G42-H44, E65, L127, Q128, G433, L435,	G410- L412, P383, W490-N491, D460-A462, N487),	R273, R417,	E169, R288, Q289, R230, D311-D312, T254, P331,
EcHACS (2q27)	G31-P33, E54, Y118, E119, M421, M423	A396-T398, Y372, I477-Y478, D447-A449, N470),	R264, R403,	R158, R280, K220, D302-I303, M245, I322,

Residues highlighted in green are from the second subunit (protomer)

Supplementary Table S4. Kinetic parameters of HAACL/S variants determined for HAACL/S selected for structural analysis.

Substrate	$k_{cat,app}$ (s^{-1})	$K_{m,app}$ (mM)	$k_{cat,app}/K_{m,app}$ ($mM^{-1} s^{-1}$)
RuHAACL/S			
Formaldehyde	3.72 ± 0.66	7.40 ± 0.97	5.03×10^2
Acetaldehyde	3.63 ± 0.58	4.11 ± 0.41	8.82×10^2
Propionaldehyde	3.36 ± 0.17	17.17 ± 1.11	1.96×10^2
ApbHAACL/S			
Formaldehyde	11.27 ± 0.31	2.028 ± 0.34	5.56×10^3
Acetaldehyde	34.89 ± 2.30	0.41 ± 0.06	8.51×10^4
Propionaldehyde	10.17 ± 1.09	1.65 ± 0.12	6.17×10^3
DhcHAACL/S			
Formaldehyde	12.27 ± 1.11	2.12 ± 0.37	5.79×10^3
Acetaldehyde	41.35 ± 4.00	0.97 ± 0.10	4.26×10^4
Propionaldehyde	20.58 ± 1.88	1.071 ± 0.10	1.92×10^4
CcHAACL/S			
Formaldehyde	N.D	N.D.	N.D.
Acetaldehyde	26.71 ± 4.23	4.76 ± 0.58	5.61×10^3
Propionaldehyde	31.65 ± 2.46	3.497 ± 0.33	9.05×10^3
TbHAACL/S			
Formaldehyde	9.681 ± 0.31	3.73 ± 0.12	2.60×10^3
Acetaldehyde	28.63 ± 3.70	2.94 ± 0.49	9.73×10^3
Propionaldehyde	24.91 ± 2.55	2.21 ± 0.31	1.13×10^4
CoHAACL/S			

Formaldehyde	10.63 ± 1.47	9.47 ± 1.62	1.12 x 10 ³
Acetaldehyde	58.43 ± 9.11	3.80 ± 0.29	1.53 x 10 ⁴
Propionaldehyde	13.08 ± 1.89	1.37 ± 0.22	9.57 x 10 ²
CfhHACL/S			
Formaldehyde	16.60 ± 1.60	2.47 ± 0.33	6.73 x 10 ³
Acetaldehyde	59.23 ± 5.47	8.83 ± 1.09	6.70 x 10 ³
Propionaldehyde	5.69 ± 0.63	3.61 ± 0.42	1.58 x 10 ³

Supplementary Table S5. Kinetic parameters of ApbHACL/S on acetaldehyde with and without 0.5 mM ADP. While apparent k_{cat} and catalytic efficiency ($k_{\text{cat}}/K_{\text{M}}$) are slightly higher with ADP, the differences are not statistically significant ($p = 0.55$).

Substrate	ADP (mM)	$k_{\text{cat,app}}$ (s⁻¹)	$K_{\text{m,app}}$ (mM)	$k_{\text{cat,app}}/K_{\text{m,app}}$ (mM⁻¹ s⁻¹)
Acetaldehyde	0.5	34.51 ± 4.02	0.49 ± 0.10	7.04 x 10 ⁴
Acetaldehyde	0	29.01 ± 6.51	0.51 ± 0.11	5.69 x 10 ⁴

Supplementary References

1. Nattermann, M. *et al.* Engineering a Highly Efficient Carborigase for Synthetic One-Carbon Metabolism. *ACS Catalysis* **11**, 5396-5404, doi:10.1021/acscatal.1c01237 (2021).
2. Holm, L., Laiho, A., Törönen, P. & Salgado, M. DALI shines a light on remote homologs: One hundred discoveries. *Protein Science* **32**, e4519, doi:<https://doi.org/10.1002/pro.4519> (2023).
3. Karplus, P. A. & Diederichs, K. Linking Crystallographic Model and Data Quality. *Science* **336**, 1030-1033, doi:10.1126/science.1218231 (2012).
4. Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **32**, W615-619, doi:10.1093/nar/gkh39832/suppl_2/W615 [pii] (2004).