Supporting Information

Glyoxylate Carboligase-based Whole-Cell Biotransformation of Formaldehyde into Ethylene Glycol via Glycolaldehyde

Hye-Jin Jo,[a] Jun-Hong Kim,[b] Ye-Na Kim,[a] Pil-Won Seo,[b] Chae-Yun Kim,[a] Ji-Won Kim,[b] Hanna Yu,[a] Huijin Cheon,[a] Eun Yeol Lee,[c] Jeong-Sun Kim,*[b] and Jin-Byung Park*[a]

[a] Department of Food Science and Engineering, Ewha Womans University, Seoul 03760, Republic of Korea

[b] Department of Chemistry, Chonnam National University, Gwangju 61186, Republic of Korea

[c] Department of Chemical Engineering (Integrated Engineering), Kyung Hee University, Gyeonggi-do 17104, Republic of Korea

19 Pages, 4 Tables, 2 Schemes, and 7 Figures are presented.

Table of Contents

Table S1. Plasmids used in this study.

Plasmids	Description	Reference
pProEx Hta-HtALS	pBR322 origin, P_{Trc}, Amp^R	This study
p.LIC B3-VvALS	pBR322 origin, P_{T7}, Amp^R	This study
p.LIC B3-EcOCD	pBR322 origin, P_{T7}, Amp^R	This study
p.LIC B3-DmGCL	pBR322 origin, P_{T7}, Amp^R	This study
p.LIC B3-EcGCL	pBR322 origin, P_{T7}, Amp^R	This study
pET21b(+)-FLS	pBR322 origin, P_{T7}, Amp^R	[1]
pACYC-DhaT	p15A origin, P_{T7}, Chl^R	[2]
pACYC-FucO	p15A origin, P_{T7}, Chl^R	This study

***** HtALS: acetolactate synthase from *Hydrogenobacter thermophilus*, EcOCD: oxalyl-CoA decarboxylases from *E. coli*, DmGCL: glyoxylate carboligase from *Deinococcus metallilatus*, EcGCL: glyoxylate carboligase from *E. coli*, DhaT: 1,3-propanediol dehydrogenase from *Klebsiella pneumoniae*, FucO: 1,2-propanediol oxidoreductase from *E. coli*

Table S2. Oligonucleotide primers used for PCR.

Name	Primer sequence (5' - 3')
HtALS	(F)-GGCGGTGGTGGCGGCATGGCCCGTAAAGGGGCTGA
	(R)-GTTCTTCTCCTTTGCGCCCTTAACCAACTAAGTACATTGTGTCAG
VvALS	(F)-GGCGGTGGTGGCGGCATGACAGCAATGTTGTCAGGCG
	(R)-GTTCTTCTCCTTTGCGCCCTTAAGTTCTCTCCGTTTTGCTTAGC
EcOCD	(F)-GGCGGTGGTGGCGGCATGTCAGATCAACTTCAAATGACAG
	(R)-GTTCTTCTCCTTTGCGCCCTTAATTACCAGCGACTTGTTTTGG
DmGCL	(F)-GGCGGTGGTGGCGGCATGCCGAGAATGACCGCGGT
	(R)-GTTCTTCTCCTTTGCGCCCCTAGTCCAGCAACGCAATCGCG
EcGCL	(F)-GGCGGTGGTGGCGGCATGGCAAAAATGAGAGCCGTTG
	(R)-GTTCTTCTCCTTTGCGCCCTTATTCATAGTGCATGAAGCAGG
$EcGCL_{R484M}$	(F)-ATTCGTCAGTCACAAATGGCTTTTGACATGGAC
	(R)-GTCCATGTCAAAAGCCATTTGTGACTGACGAAT
$EcGCL_{R484MR108L}$	(F)-CCAGGCACCGCGCGCCCTTCTGCATAAAGAAGATTTT
	(R)-AAAATCTTCTTTATGCAGAAGGGCGCGCGGTGCCTGG
$EcGCL_{R484MQ258A}$	(F)-GGATGGTGGGTCTGGCAACCGCGCATCGTTAC
	(R)-GTAACGATGCGCGGTTGCCAGACCCACCATCC
$EcGCL_{R484MN283Q}$	(F)-ATCGGTAACCGTTTTGCTCAGCGTCATACCGGCTCG
	(R)-CGAGCCGGTATGACGCTGAGCAAAACGGTTACCGAT
$EcGCL_{R484MN284M}$	(F)-GGTAACCGTTTTGCTAACATGCATACCGGCTCGGTAGAG
	(R)-CTCTACCGAGCCGGTATGCATGTTAGCAAAACGGTTACC
$EcGCL_{R484MN283QA28S}$	(F)-GCCTTCGGTGTTCCGGGATCTGCAATCAATCCGTTCTAC
	(R)-GTAGAACGGATTGATTGCAGATCCCGGAACACCGAAGGC
$EcGCL_{R484MN283QF114L}$	(F)-CGTCTGCATAAAGAAGATTTACAGGCCGTAGATATTGAA
	(R)-TTCAATATCTACGGCCTGTAAATCTTCTTTATGCAGACG
$EcGCL_{R484MN283QL257M}$	(F)-ATGGCCGGGATGGTGGGTATGCAAACCGCGCATCGTTAC
	(R)-GTAACGATGCGCGGTTTGCATACCCACCATCCCGGCCAT
$EcGCL_{R484MN283QR284Q}$	(F)-GGTAACCGTTTTGCTCAGCAACATACCGGCTCGGTAGAG
	(R)-CTCTACCGAGCCGGTATGTTGCTGAGCAAAACGGTTACC
$EcGCL_{R484MN283QL478M}$	(F)-AACAACGCTTATCTGGGGATGATTCGTCAGTCACAACGC
	(R)-GCGTTGTGACTGACGAATCATCCCCAGATAAGCGTTGTT

$EcGCL_{R484MN283QA485F}$	(F)-ATTCGTCAGTCACAAATGTTCTTTGACATGGACTACTGC (R)-GCAGTAGTCCATGTCAAAGAACATTTGTGACTGACGAAT
FucO	(F)-ACCACAGCCAGGATCCGATGGCTAACAGAATGCTGGTGAAC (R)-ATGCGGCCGCAAGCTTTTACCAGGCGGTATGGTAAAGCTC

Table S3. Data collection and structure refinement statistics.

Data Collection	DmGCL
Space group	C2
Unit cell dimensions a, b, c (Å), a, b, g (°)	 148.06, 106.47, 89.09 90, 124.1, 90
Wavelength (Å)	0.9796
Resolution (Å)	50-2.10 (2.14-2.10)
R_{sym} (%)	8.6 (66.6)
I/s(I)	19.15 (1.92)
Completeness (%)	99.9 (97.0)
Redundancy	6.5 (6.0)
Refinement	
Resolution (Å)	44.78-2.10 (2.14-2.10)
No. of reflections	65709
R_{work} / R_{free} (%)	18.33 (27.21) / 19.82 (30.12)
No. atoms protein / water	 8590 / 375
RMSD bond lengths (Å) / angles (º)	 0.007 / 0.945
Average B-values (Å^2)	55.5
Ramachandran plot (%) favored / allowed / outliers	 95.86 / 3.33 / 0.81

Values in parentheses are for the highest-resolution shell.
4% of the randomly selected reflections were used for calculating R_{free} values.

Table S4. Homologous structures of FLS retrieved by DALI search

PDB ID	Z-score	rmsd	No. of aligned residues	No. of residues	% of identity	Function annotation
5k2o	40.4	2.3	533	585	26	Acetohydroxyacid synthase
2ji6	39.9	2.5	529	559	25	Oxalyl-CoA decarboxylase
2pan	39.5	2.0	535	592	25	Glyoxylate carboligase
5dx6	38.8	2.6	529	557	26	Acetolactate synthase
2pgn	38.1	2.4	536	587	24	Cyclohexane-1,2-dione hydrolase
2ihv	37.4	2.5	526	563	27	Carboxyethylarginine synthase
6a50	37.4	2.6	508	527	23	Benzoylformate decaroxylase
3ey9	36.8	2.8	528	571	21	Pyruvate dehydrogenase
4fee	35.5	3.3	532	586	21	Pyruvate oxidase
2vbf	34.6	2.8	498	546	22	Branched-chain alpha-ketoacid decarboxylase
1ovm	34.3	2.6	497	535	22	Indole-3-pyruvate decarboxylase
2q5l	33.6	3.2	492	538	24	Phenylpyruvate decarboxylase
2x7j	31.0	3.2	524	579	17	MenD

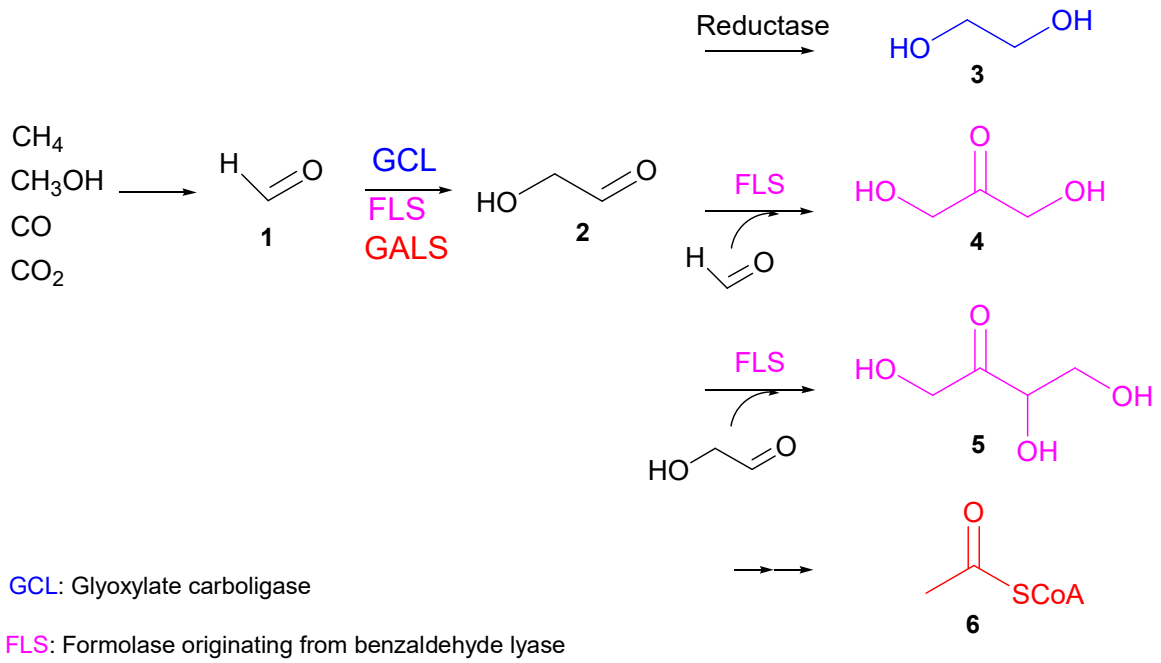
* Only the protein structures with Z-scores of more than 30 are listed.
* Homologous proteins with same annotation are represented by the enzyme with the highest Z-score value.

Table S5. Kinetic constants of EcGCL and its double variants

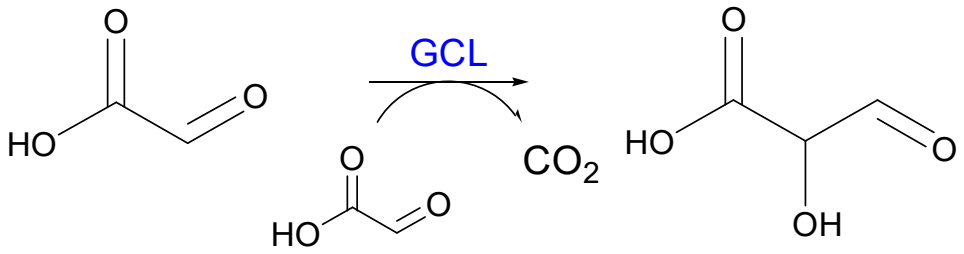
	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}\cdot s^{-1}$)
Wild type	62	0.1	1.6
R484MQ258A	31	0.08	2.5
R484MN283Q	24	0.08	3.5
R484MR284M	54	0.04	0.7
R484MR108L	66	0.02	0.4

Table S6. Kinetic constants of EcGCL and its triple variants

	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($M^{-1}\cdot s^{-1}$)
R484MN283QA28S	28	0.06	2.3
R484MN283QF114L	33	0.07	2.1
R484MN283QL478M	18	0.09	5.2
R484MN283QL257M	76	0.07	0.9
R484MN283QA485F	17	0.04	2.5
R484MN283QR284Q	28	0.05	1.7



Scheme S1. C1 biotransformation pathways. Formaldehyde (**1**) is enzymatically converted into glycolaldehyde (**2**), which can be transformed into ethylene glycol (**3**), dihydroxyacetone (**4**)[1] , erythruolse (**5**)[3], and acetyl CoA (**6**)[4].



Scheme S2. The GCL reaction scheme for the native substrate, glyoxylate[5].

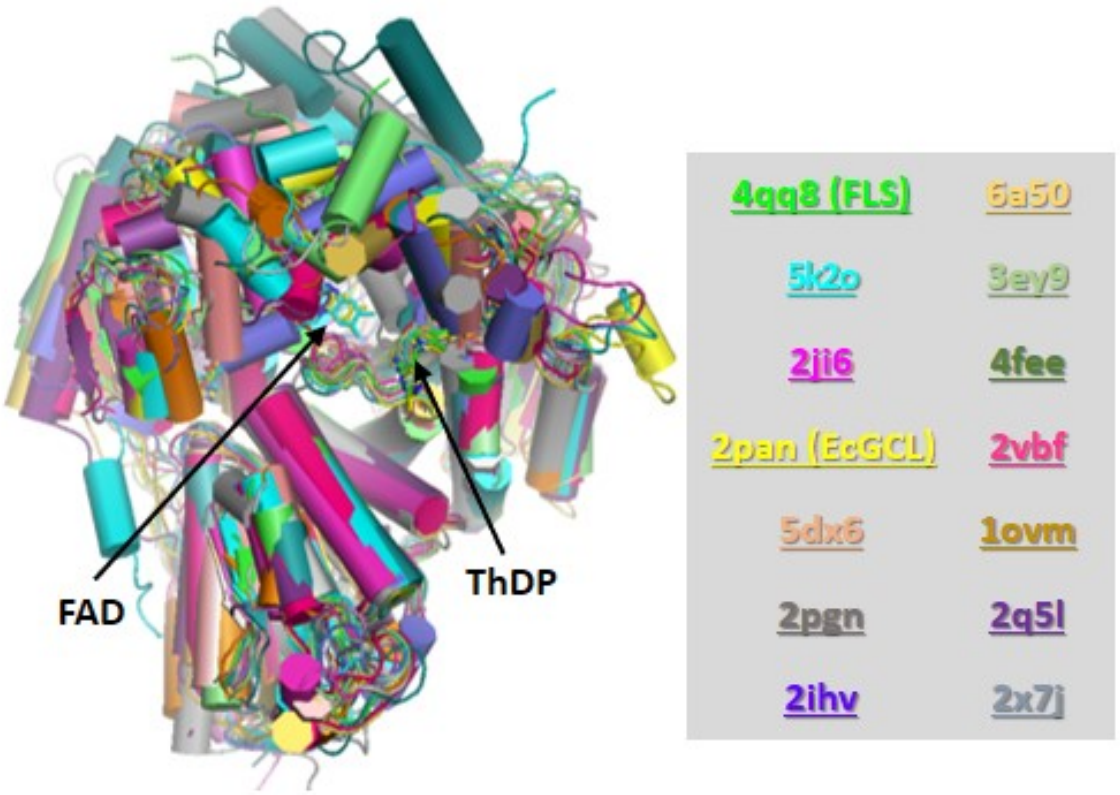


Figure S1. A ribbon diagram of the superposed structures. Each structure is differentiated by colors and α-helices are represented with cylinders. The active site is located, where the two cofactors of ThDP and FAD are bound.

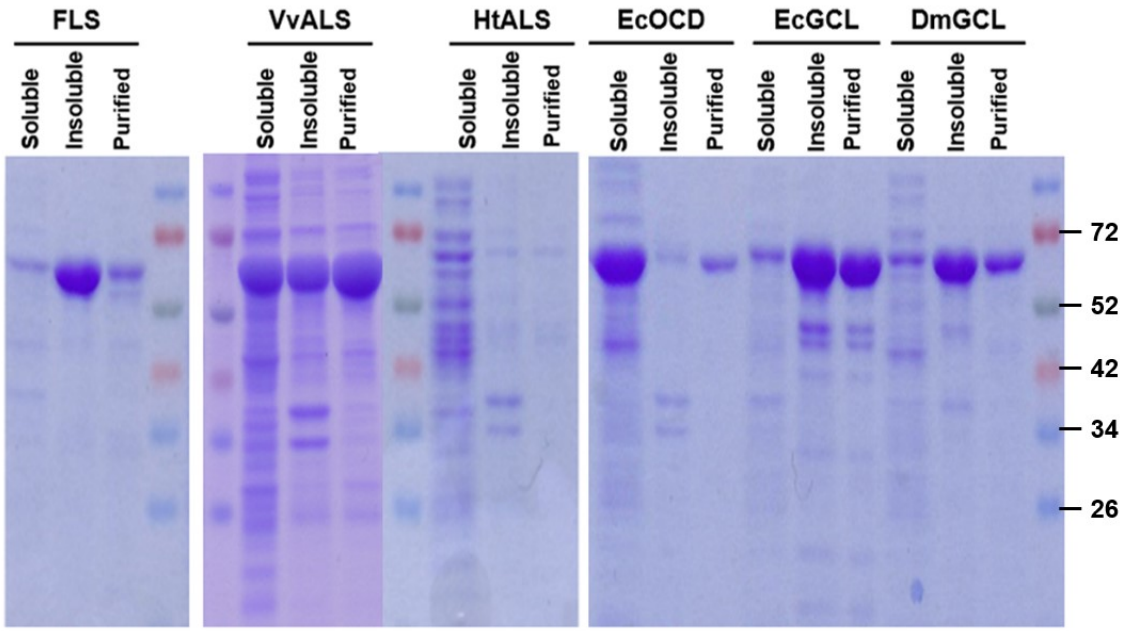
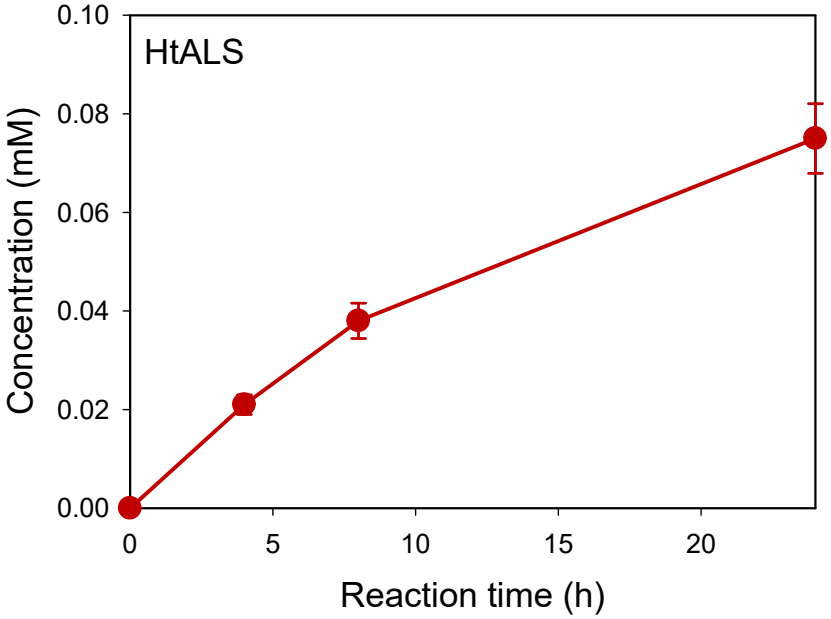
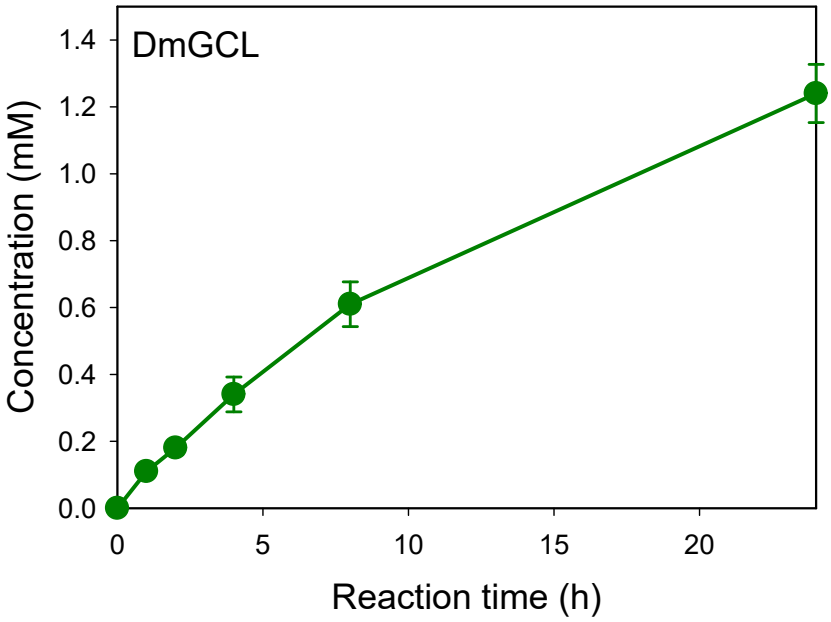


Figure S2. SDS-PAGE analysis of the enzymes. The enzymes (FLS, VvALS, HtALS, EcOCD, EcGCL, and DmGCL) were expressed in the recombinant *E. coli* BL21 (DE3) Star and purified by Ni-NTA chromatography. Soluble, soluble fraction; Insoluble, insoluble fraction; Purified, purified protein fraction. The molecular weight markers (kDa) are shown on the right of the panel.


HtALS
Concentration (mM)
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
Reaction time (h)


DmGCL
Concentration (mM)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
Reaction time (h)

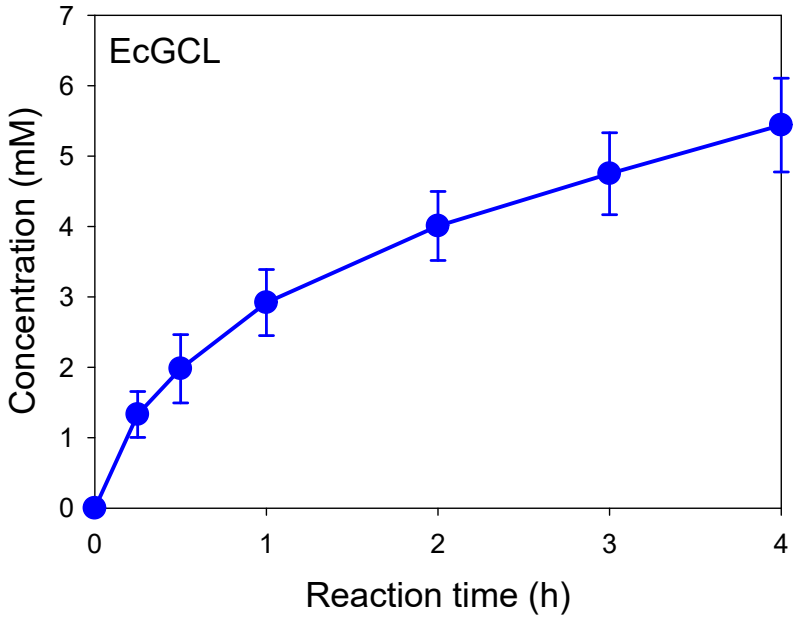


Figure S3. **Time course of formaldehyde biotransformation by the (A) HtALS, (B) DmGCL, and (C) EcGCL**. The enzymes were added to 1 mg/mL in the 50 mM potassium phosphate buffer (pH 6.0) containing 25 mM formaldehyde and sufficient amount of magnesium (II) ion, thiamine pyrophosphate (ThDP), and flavin adenine dinucleotide (FAD).

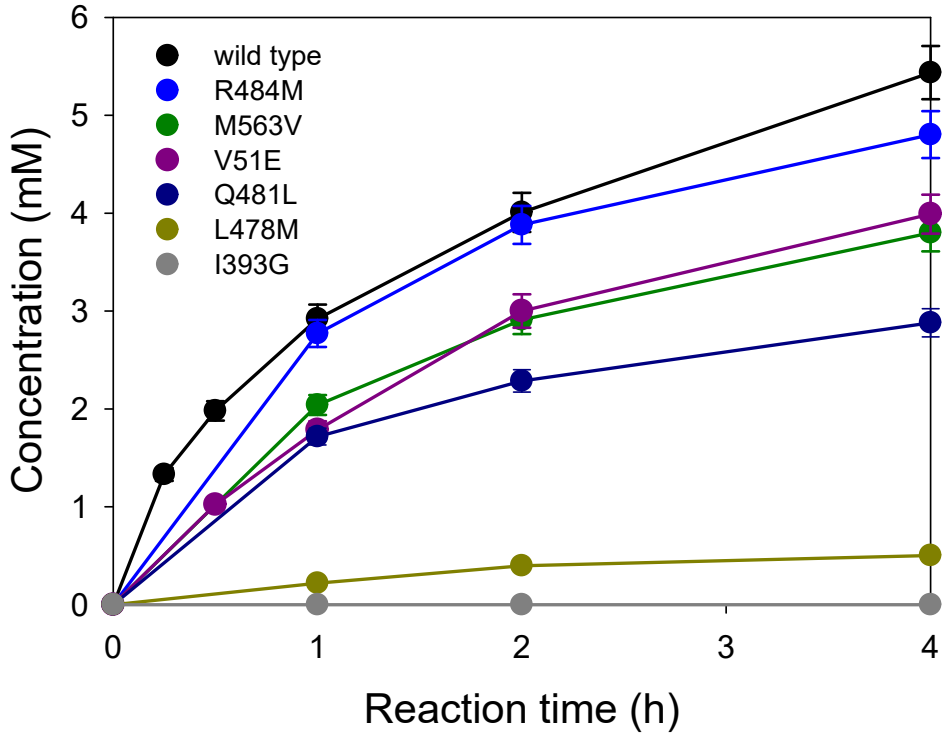


Figure S4. Time course of formaldehyde biotransformation by purified EcGCL and its variants. The biotransformation of formaldehyde into glycolaldehyde was initiated by adding 25 mM formaldehyde into the 50 mM potassium phosphate buffer (pH 6.0) containing 1 mg/mL EcGCL and its variants at 50 °C. The reaction buffer also included sufficient amount of magnesium (II) ion, ThDP, and FAD.

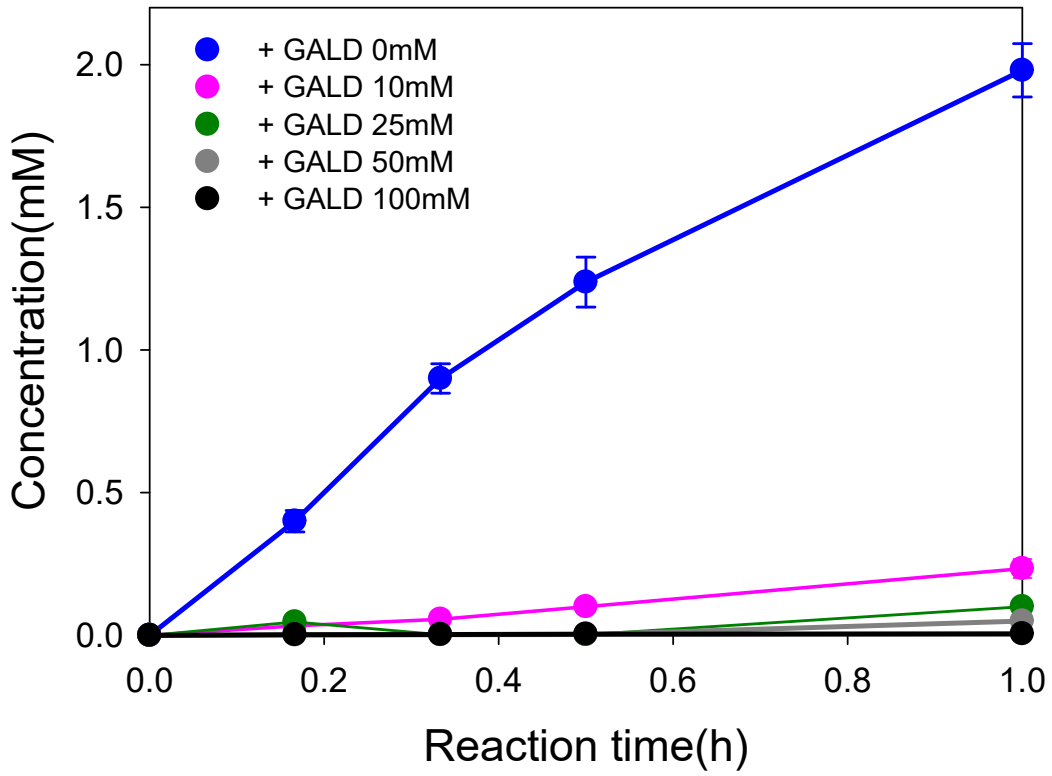


Figure S5. Toxic effects of glycolaldehyde on EcGCL. The toxic effects of glycolaldehyde on EcGCL were evaluated after incubation with zero to 100 mM glycolaldehyde for 1 h in the 50 mM potassium phosphate buffer (pH 6.0) (enzyme concentrations: 1 mg/mL). The residual enzyme activities were measured by adding 25 mM formaldehyde into the potassium phosphate buffer at 50 °C.

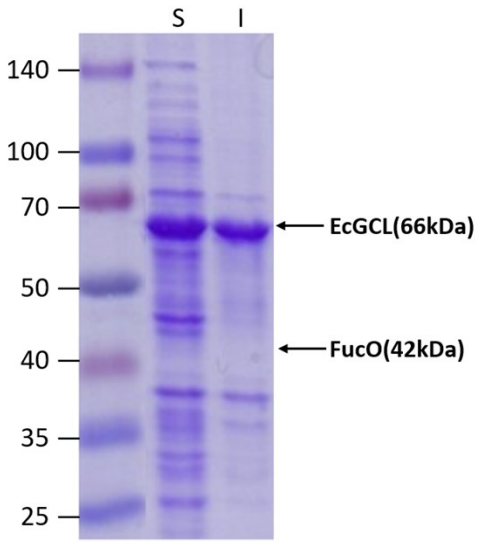


Figure S6. SDS-PAGE analysis of the proteins expressed in recombinant *E. coli* p.LIC B3-EcGCL$_{R484MN283QL478M}$/pACYC-FucO. S: soluble fraction, I: insoluble fraction. The molecular weight markers (kDa) are shown on the left of the panel.

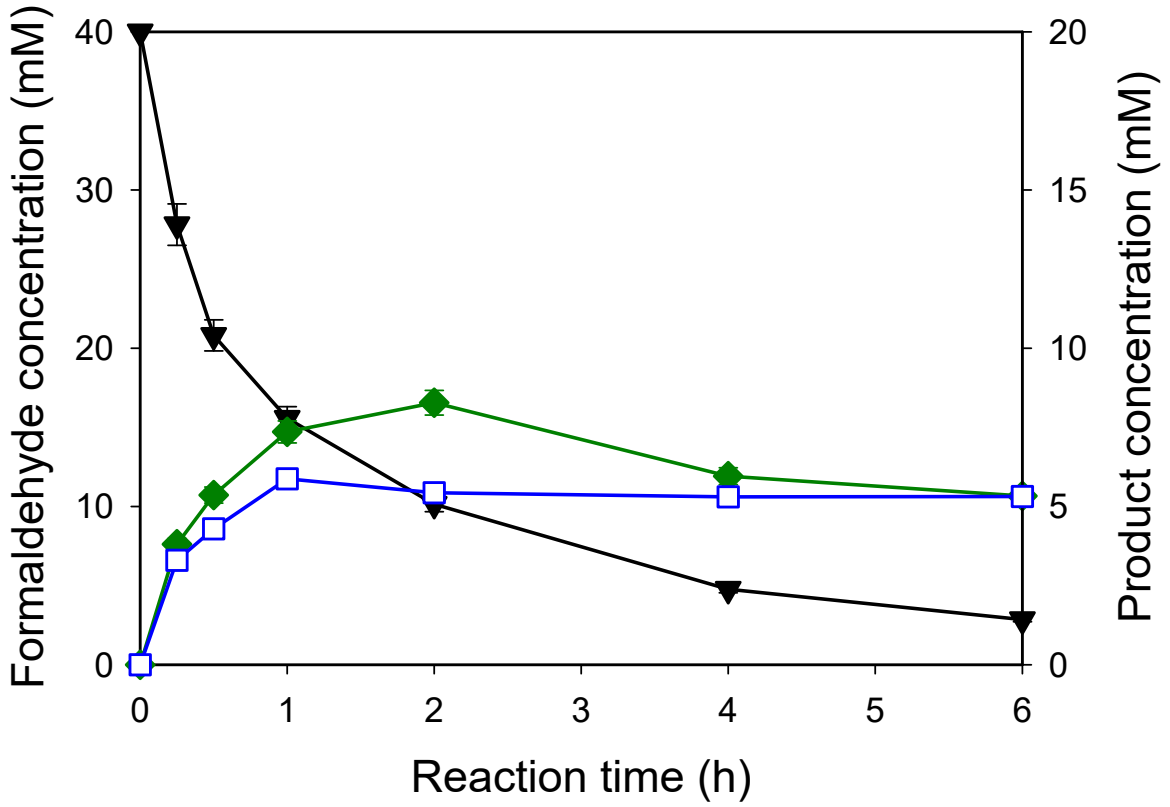


Figure S7. Time course of the cascade reaction of formaldehyde into ethylene glycol with high concentration. Biotransformation of formaldehyde into ethylene glycol was initiated by adding 40 mM formaldehyde at t=0 into the potassium phosphate buffer containing 11 g dry cells/L of recombinant *E. coli* expressing the $EcGCL_{R484MN283QL478M}$ and FucO. The symbols indicate the concentration of formaldehyde (▼), glycolaldehyde (◆) and ethylene glycol (□).

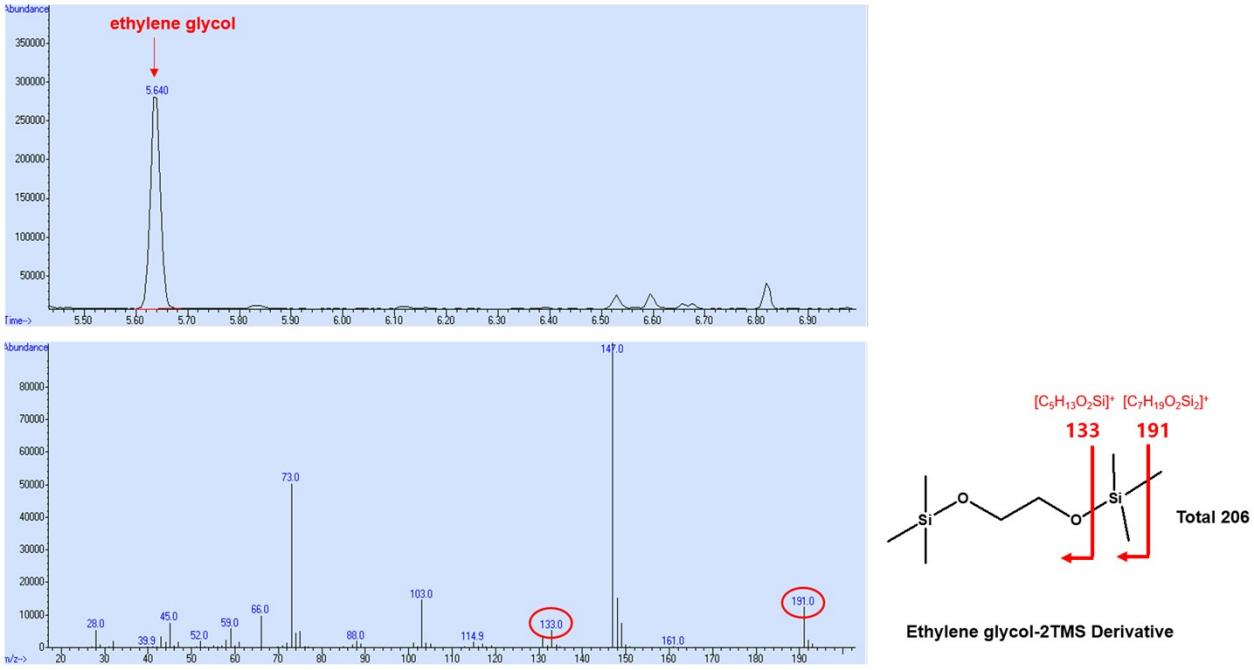


Figure S8. GC/MS analysis of the reaction products (3) purified. The reaction products were isolated from the reaction medium, shown in Fig. 5, via extraction with acetonitrile after *in vacuo* evaporation of the reaction broth. The resulting organic layer was evaporated to yield the product as a transparent liquid.

Reference

1. J. B. Siegel, A. L. Smith, S. Poust, A. J. Wargacki, A. Bar-Even, C. Louw, B. W. Shen, C. B. Eiben, H. M. Tran and E. Noor, *Proc. Natl. Acad. Sci.*, 2015, **112**, 3704-3709.
2. S. Lama, S. M. Ro, E. Seol, B. S. Sekar, S. K. Ainala, J. Thangappan, H. Song, D. Seung and S. Park, *Biotechnol. Bioprocess Eng.*, 2015, **20**, 971-979.
3. S. Güner, V. Wegat, A. Pick and V. Sieber, *Green Chem.*, 2021.
4. X. Lu, Y. Liu, Y. Yang, S. Wang, Q. Wang, X. Wang, Z. Yan, J. Cheng, C. Liu and X. Yang, *Nat. Commun.*, 2019, **10**, 1-10.
5. N. Nemeria, E. Binshtein, H. Patel, A. Balakrishnan, I. Vered, B. Shaanan, Z. e. Barak, D. Chipman and F. Jordan, *Biochemistry*, 2012, **51**, 7940-7952.