

SUPPORTING INFORMATION

A new class of phosphoribosyltransferase involved in cobamide biosynthesis is found in methanogenic archaea and cyanobacteria

Victoria L. Jeter¹, Theodor A. Mattes^{1§}, Nathaniel R. Beattie², and Jorge C. Escalante-Semerena^{1*}

¹Department of Microbiology, ²Department of Biochemistry & Molecular Biology University of Georgia, Athens, GA 30602

[§]Present address: Department of Genetics, University of Georgia, Athens, GA 30602

*Corresponding author: Department of Microbiology, University of Georgia, 212C Biological Sciences Building, 120 Cedar Street, Athens, GA 30602, USA; Tel: +1 (706) 542-2651, Fax: (+1 (706) 542-2815, Email: jcescala@uga.edu, www.escalab.com

Running title: cyanobacteria have a CobT enzyme of archaeal origin

KEYWORDS: a new class of CobT phosphoribosyltransferase, cyanobacteria, B₁₂ biosynthesis, cobamides, *Methanocaldococcus jannaschii*, archaeal PRTase

Table S1. Strains used in this study.		
Strains	Genotype	Source
<i>S. enterica</i>¹		
JE7088	$\Delta metE2702 ara-9$	Laboratory strain collection
Derivatives of JE7088		
JE11685	/ pBAD24	
JE12872	<i>cobT1379::kan⁺</i>	
JE12893	<i>cobT1379::kan⁺ cobB1374::cat⁺</i>	
JE17824	<i>cobT1379::kan⁺ / pCOBT140</i>	Laboratory strain collection
JE18113	<i>cobT1379::kan⁺ / pMjCobT1</i>	This study
JE18114	<i>cobT1379::kan⁺ cobB1374::cat⁺ / pMjCobT1</i>	This study
JE18115	<i>cobT1379::kan⁺ cobB1374::cat⁺ / pCOBT140</i>	This study
JE22263	/ pCV1	
JE24014	<i>cobT1379::kan⁺ cobB1374::cat⁺ / pMjCobT3</i>	This study
JE24016	<i>cobT1379::kan⁺ cobB1374::cat⁺ / pMjCobT4</i>	This study
JE24494	<i>cobT1379::kan⁺ cobB1374::cat⁺ / pMjCobT5</i>	This study
JE24636	<i>cobT1379::kan⁺ cobB1374::cat⁺ / pSynCobT2</i>	This study
<i>E. coli</i> strains		
JE21547	F ⁻ f80d <i>lacD(lacZ)M15</i> $\Delta(lacZYA-argF)$ U169 <i>deoR recA1 endA1 hsdR17</i> (rK,	Laboratory strain collection

	mK ⁺) <i>phoA supE44 l- thi1 gyrA96 relA1</i>	
JE6663 C41(λDE3)	F ⁻ <i>ompT hsdSB (rB⁻ mB⁻) gal dcm</i> (λDE3)	Avidis
Plasmids		
pTEV5	Overexpression vector that fuses the N-terminus of the protein of interest to a H ₆ tag, which can be removed by rTEV protease, <i>bla</i> ⁺ (Ap ^R)	{Rocco, 2008 #11866}
pBAD24	complementation vector, P _{araBAD} , L-(+)-arabinose inducible promoter, <i>bla</i> ⁺ (Ap ^R)	{Guzman, 1995 #25015}
pCOBT140	<i>S. enterica cobT</i> ⁺	{Mattes, 2017 #26254}
pMjCobT1	MJ_RS08515 <i>cobT</i> ⁺ <i>bla</i> ⁺ (Ap ^R) cloned into pBAD24	This study
pMjCobT2	MJ_RS08515 <i>cobT</i> ⁺ <i>bla</i> ⁺ (Ap ^R) cloned into pTEV5	This study
pMjCobT3	MJ_RS08515 <i>cobT</i> ⁺ (CobT ^{E315A}) <i>bla</i> ⁺ (Ap ^R) cloned into pBAD24	This study
pMjCobT4	MJ_RS08515 <i>cobT</i> ⁺ (CobT ^{E150A E315A}) <i>bla</i> ⁺ (Ap ^R) cloned into pBAD24	This study
pMjCobT5	MJ_RS08515 <i>cobT</i> ⁺ (CobT ^{E150A}) <i>bla</i> ⁺ (Ap ^R) cloned into pBAD24	This study
pSynCobT2	SYNWH7803_RS09275 ⁺ <i>bla</i> ⁺ (Ap ^R) cloned into pBAD24	This study

¹All strains were derivatives of *Salmonella enterica enterica* sv Typhimurium LT2

Table S2. Primers used in this study.	
Primer Name	Primer Sequence (5'-3')
Mj_1598_cobT_PIPE_f	CAG GAG GAA TTC ACC ATG AGC ATA ATA GCA AT
Mj_1598_cobT_PIPE_rev	TTG CAT GCC TGC AGG TCG GCT TTA TTC TTT ATA CCA C
pBAD24-PIPE-5'	CAT GGT GAA TTC CTC CTG CTA GCC CA
pBAD24-PIPE-3'	CGA CCT GCA GGC ATG CAA GCT T
Mj_1598_TEV5_PIPE_f	AAC CTG TAT TTT CAG GGC ATG AGC ATA ATA GCA AT
Mj_1598_TEV5_PIPE_rev	AGC TCG AGA ATT CCA TGG GCT TTA TTC TTT ATA CCA C
pTEV5-PIPE-5'	CAT GGT GAA TTC CTC CTG CTA GCC CA
pTEV5-XhoI-3'	CGA CCT GCA GGC ATG CAA GCT T
MjCobT_sdm_E150A_F	GTCCCACCAGGAACACTCGCCCCAACAATTAATAATT
MjCobT_sdm_E150A_R	AATTATTAATTGTTGGGGCGAGTGTTCTTGGTGGGAC
MjCobT_sdm_E315A_F	CAGCCCCTACTCCTGCTTTTACTGAACCTTTGCA
MjCobT_sdm_E315A_R	TGCAAAGGTTTCAGTAAAAGCAGGAGTAGGGGCTG
SynCobT pBAD F	NNGCTCTTCNTTCATGGTGTGGCTGCCGAGC
SynCobT pBAD R	NNGCTCTTCNTAATTACGGCGCCGGCGGACG

Table S3. Methanogen CobT homologues			
Query	<i>M. jannaschii</i> CobT		
Homology			
Methanogen Class I			
Organism	% Identity	% Positives	E-value
<i>Methanocaldococcus</i> sp. FS406-22	94	97	0.0
<i>Methanocaldococcus bathoardescens</i>	93	97	0.0
<i>Methanocaldococcus fervens</i>	89	94	0.0
<i>Methanocaldococcus fervens</i> AG86	88	94	0.0
<i>Methanocaldococcus vulcanius</i>	85	92	0.0
<i>Methanotorris igneus</i>	73	84	0.0
<i>Methanocaldococcus villosus</i>	73	86	0.0
<i>Methanotorris formicicus</i>	70	84	0.0
<i>Methanothermococcus okinawensis</i>	66	81	2e-162
<i>Methanococcus vannielii</i>	66	79	7e-159
<i>Methanothermococcus thermolithotrophicus</i>	65	78	3e-158
<i>Methanocaldococcus infernus</i>	69	82	3e-158
<i>Methanocaldococcus infernus</i> ME	69	82	5e-158
<i>Methanococcus maripaludis</i>	62	78	5e-154
<i>Methanococcus maripaludis</i> C6	62	77	1e-150
<i>Methanococcus maripaludis</i> C7	61	75	1e-148
<i>Methanococcus aeolicus</i>	61	78	6e-148
<i>Methanococcus voltae</i>	58	73	2e-133
<i>Methanothermus fervidus</i>	48	64	5e-90
<i>Methanothermobacter tenebrarum</i>	43	61	1e-83
Methanogen Class II			
Organism	% Identity	% Positives	E-value
<i>Methanocella arvoryzae</i>	39	56	8e-65
<i>Methanocella paludicola</i> SANA E *	38	54	4e-64
<i>Methanosarcina mazei</i>	37	56	1e-61
<i>Methanosarcina</i> sp. MTP4	38	57	2e-61
<i>Methanohalobium evestigatum</i>	38	59	3e-61
<i>Methanosarcina</i> sp. Kolksee	36	57	4e-61
<i>Methanosarcina thermophila</i>	37	57	1e-60
<i>Methanosarcina vacuolata</i>	36	57	2e-60
<i>Methanosarcina spelaei</i>	37	56	2e-59
<i>Methanosarcina flavescens</i>	38	58	2e-59
<i>Methanosalsum zhilinae</i>	36	58	5e-59
<i>Methanosarcina barkeri</i>	36	56	2e-58
<i>Methanosarcina horonobensis</i>	37	55	2e-58
<i>Methanosarcina lacustris</i>	37	54	1e-57
<i>Methanohalophilus halophilus</i>	39	57	1e-57
<i>Methanococcoides vulcani</i>	37	57	8e-57
<i>Methanosarcina siciliae</i>	36	56	2e-56
<i>Methanosarcina acetivorans</i>	36	56	4e-56

<i>Methanohalophilus mahii</i>	38	57	4e-56
<i>Methanocella conradii</i>	38	54	3e-55

Results from a protein homology search using NCBI BLASTP are shown. *M. jannaschii* CobT coding sequence was used as the query protein to search for homologues in Methanogen Class I (taxid: 2283794) and Methanogen Class II (taxid:224756). The organism name is provided with the identity (%) and similarity (% positives) to the query sequence as well as the E-value. Results shown in bold typeface are annotated as nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferases and results with an asterisk (*) are annotated as hypothetical or putative proteins. All other hits are annotated as TIGR00303 family proteins. 20 unique results with the highest bit scores are shown for brevity. Only hits with query coverage > 85% are included in the table.

Table S4. Cyanobacterial CobT homologues to <i>S. enterica</i> LT2 CobT, <i>B. megaterium</i> DSM 319 CobT, and <i>P. denitrificans</i> ATCC 13867 CobT			
Homology			
Query	<i>S. Typhimurium</i>¹ CobT	% Identity	% Positives
<i>Prochlorothrix hollandica</i> PCC 9006		38	54
<i>Nostoc</i> sp. 3335mG		34	50
Query	<i>B. megaterium</i> DSM 319 CobT	% Identity	% Positives
<i>Prochlorothrix hollandica</i> PCC 9006		45	61
<i>Nostoc</i> sp. 3335mG		40	52
Query	<i>P. denitrificans</i> ATCC 13867 CobT	% Identity	% Positives
<i>Prochlorothrix hollandica</i> PCC 9006		72	79
<i>Nostoc</i> sp. 3335mG		48	61

Homologues from a protein homology search using NCBI BLASTP are shown. *S. enterica enterica* sv Typhimurium LT2 (*S. Typhimurium*) CobT, *Bacillus megaterium* DSM 319 CobT, and *Pseudomonas denitrificans* ATCC 13867 CobT coding sequences were used as queries to search for homologues in cyanobacteria (taxid:1117). The species name is provided with the identity (%) and similarity (% positives) to the query sequence as well as the E-value. All hits were annotated as nicotinate-nucleotide-dimethylbenzimidazole PRTase. Only hits with query coverage > 90% are included in the table.

Table S5. Cyanobacterial CobT homologues			
Query	<i>M. jannaschii</i> CobT	Homology	
Bacterium	% Identity	% Positives	E-value
<i>Synechococcus</i> sp. PCC 8807	39	56	3e-59
<i>Pleurocapsa minor</i>	40	55	1e-58
<i>Synechococcus</i> sp. NKBG042902	38	56	4e-58
<i>Synechococcus</i> sp. BDU 130192	38	56	1e-57
<i>Synechococcus</i> sp. PCC 73109	38	55	1e-57
<i>Crocospaera watsonii</i>	40	54	4e-57
<i>Synechococcus</i> sp. NIES-970	38	56	4e-57
<i>Synechococcus</i> sp. PCC 7117	38	56	5e-57
<i>Roseofilum reptotaenium</i> AO1-A	36	56	2e-56
<i>Synechococcus</i> sp. PCC 7003	37	57	3e-56
<i>Geminocystis</i> sp. NIES-3708	38	55	8e-56
<i>Gloeobacter kilaueensis</i>	34	56	3e-55
<i>Microcystis aeruginosa</i>	39	56	6e-55
<i>Phormidesmis priestleyi</i>	38	55	1e-54
<i>Leptolyngbya frigida</i>	37	54	1e-54
<i>Trichodesmium erythraeum</i>	35	56	2e-54
<i>Lyngbya aestuarii</i>	35	53	2e-54
<i>Limnothrix rosea</i>	36	54	3e-54
<i>Cyanothece</i> sp. PCC 7424	39	54	3e-54
<i>Limnoraphis robusta</i> CS-951	36	54	4e-54
<i>Hydrococcus rivularis</i>	39	53	7e-54
<i>Limnoraphis robusta</i>	36	54	8e-54
<i>Lyngbya</i> sp. PCC8106	34	56	1e-53
<i>Microcystis</i> sp. 0824	39	56	1e-53
<i>Leptolyngbya</i> sp. PCC 7376	35	55	2e-53
<i>Merismopedia glauca</i>	37	52	3e-53
<i>Pseudanabaena</i> sp. ABRG5-3	37	55	5e-53
<i>Leptolyngbya</i> sp. 'hensonii'	36	53	5e-53
<i>Microcystis</i> sp. MC19	39	55	5e-53
<i>Cyanothece</i> sp. CCY0110	37	53	6e-53
<i>Cyanothece</i> sp. BG0011	38	53	1e-52
<i>Microcystis</i> sp. T1-4	39	55	1e-52
<i>Synechocystis</i> sp. PCC 7509	37	53	1e-52
<i>Nostoc</i> sp. PCC7524	37	52	5e-52
<i>Chamaesiphon minutus</i>	34	53	6e-52
<i>Planktothricoides</i> sp. SR001	34	51	9e-52
<i>Phormidium ambiguum</i>	36	53	1e-51
<i>Rivularia</i> sp. PCC 7116	34	51	1e-51
<i>Dactylococcopsis salina</i>	37	53	2e-51
<i>Microcystis panniformis</i> FACHB-1757	39	55	2e-51
<i>Chroococciopsis cubana</i>	36	53	2e-51
<i>Euhalothece</i> sp. KZN 001	37	53	2e-51
<i>Anabaena</i> sp. PCC 7108	36	50	3e-51
<i>Chroococciopsis thermalis</i>	36	53	3e-51

<i>Calothrix</i> sp. 336/3	37	53	3e-51
<i>Chroococciopsis</i> sp. CCALA 051	36	53	3e-51
<i>Aphanothece hegwaldii</i>	36	53	4e-51
<i>Cyanobacterium</i> sp. HL-69	36	52	5e-51
<i>Nodosilinea nodulosa</i>	36	53	6e-51
<i>Cyanbacterium</i> TDX16	36	53	7e-51
<i>Phormidium tenue</i>	37	55	1e-50
<i>Cyanothece</i> sp. PCC 8802	39	52	1e-50
<i>Leptolyngbya</i> sp. NIES-3755	38	54	1e-50
<i>Oscillatoriales cyanobacterium</i> JSC-12	37	53	1e-50
<i>Microcystis aeruginosa</i> PCC 9443	37	54	1e-50
<i>Hydrocoleum</i> sp. CS-953	35	52	1e-50
<i>Stanieria</i> sp. NIES-3757	36	53	2e-50
<i>Desertifilum</i> sp. IPPAS B-1220	34	54	2e-50
* <i>Nostoc</i> sp. HK-01	36	53	3e-50
<i>Pleurocapsa</i> sp. PCC 7319	34	52	3e-50
<i>Leptolyngbya</i> sp. NIES-2104	39	55	3e-50
<i>Leptolyngbya ohadii</i>	36	52	4e-50
<i>Calothrix elsteri</i>	35	53	4e-50
<i>Aliterella atlantica</i>	37	52	5e-50
<i>Anabaenopsis circularis</i>	37	52	5e-50

Results from a protein homology search using NCBI BLASTP are shown. *M. jannaschii* CobT coding sequence was used as the query protein to search for homologues in cyanobacteria (taxid:1117). The species name is provided with the identity (%) and similarity (% positives) to the query sequence as well as the E-value. Homologues shown in bold typeface are annotated as nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferases and homologues with an asterisk (*) are annotated as hypothetical proteins. All other proteins are annotated as TIGR00303 family proteins. Only homologues with query coverage >90% are included in the table.

Table S6. Thaumarchaeota CobT homologues			
Query	<i>M. jannaschii</i> CobT	Homology	
Organism	% Identity	% Positives	E-value
<i>Candidatus Nitrososphaera evergladensis</i> SR1	42	60	5e-74
<i>Candidatus Nitrososphaera gargensis</i> Ga9.2 *	42	58	6e-73
<i>Candidatus Nitrosocaldus islandicus</i>	39	58	2e-69
<i>Candidatus Nitrosocaldus cavascurensis</i>	39	58	2e-69
<i>Candidatus Nitrocosmicus oleophilus</i> *	39	58	3e-69
<i>Candidatus Nitrosoarchaeum</i> sp. *	40	60	1e-65
<i>Thaumarchaeota archaeon</i> CSP1-1 *	39	59	3e-64
<i>Thaumarchaeota archaeon</i> SCGC AC-337_F14	39	58	7e-64
<i>Nitrosopumilales archaeon</i>	39	57	3e-63
<i>Nitrosoarchaeum koreense</i>	39	58	6e-63
<i>Candidatus Nitrosotenus chungbukensis</i>	38	59	5e-62
<i>Nitrosopumilus</i> sp. PRT-SC01	37	58	1e-61
<i>Candidatus Nitrosopumilus salaria</i>	37	57	2e-61
<i>Candidatus Nitrosopumilus salaria</i> BD31 *	37	57	3e-61
<i>Candidatus Nitrosoarchaeum limnia</i>	39	59	3e-61
<i>Candidatus Nitrosopumilus adriaticus</i>	39	58	4e-61
<i>Candidatus Nitrosopumilus piranensis</i> *	39	59	1e-60
<i>Candidatus Nitrosoarchaeum limnia</i> SFB1 *	39	59	3e-60
<i>Thaumarchaeota archaeon</i> N4	34	56	3e-59
<i>Nitrosopumilus maritimus</i> SCM1	38	58	2e-58
<i>Candidatus Nitrosopumilus</i> sp. NM25	35	57	2e-58
<i>Candidatus Nitrosotenus aquarius</i>	34	56	8e-57
<i>Candidatus Nitrosomarinus</i> sp.	37	56	2e-56
<i>Nitrosopumilus</i> sp. Nsub	37	56	1e-55
<i>Candidatus Nitrosotenus cloacae</i>	35	57	2e-55
<i>Candidatus Nitrosomarinus Catalina</i>	37	56	9e-55
<i>Candidatus Nitrosopelagicus</i> sp.	38	54	7e-53
<i>Candidatus Nitrosopelagicus brevis</i>	35	52	2e-50

Results from a protein homology search using NCBI BLASTP are shown. *M. jannaschii* CobT coding sequence was used as the query protein to search for homologues in Thaumarchaeota (taxid: 651137). The organism name is provided with the identity (%) and similarity (% positives) to the query sequence as well as the E-value. Results shown in bold typeface are annotated as nicotinate-nucleotide-dimethylbenzimidazole phosphoribosyltransferases and results with an asterisk (*) are annotated as hypothetical or putative proteins. All other hits are annotated as TIGR00303 family proteins. Only hits with query coverage > 90% are included in the table. Results from organisms listed as “uncultured marine thaumarchaeote” were omitted from the table.

1 10 20 30 40 50 60
Native ATGGTCTGGTTGGCCCTCGATCCCCTCTGAGCTCCCTGCCGGCTGTCGCGGTGACAGGGGCAT
Optimized ATGGTGTGGCTGGCCGAGCATTCCGAGCGAAGTCCGCCGGCGGCTGTCGTCGTCGACCCGGTGC

70 80 90 100 110 120
Native CCCCAGGAGCCGGAGCCACAGCCCTACGCAACCGCTGGCAATCCAGGCCGTACCACGCCGGA
Optimized GCCCAGGTGCGGTGCAGCCGCGCTGCGTAACCGTGGCAGAGCCAGAGCGCTGCCGACC

130 140 150 160 170 180 190
Native CTGTTTGCTGGTGGCTCGCCTCACCCCGCTCCGCTGAGGTTCCAGGGATCTCCGCTGCTGGATGC
Optimized CTGCCCTGCTGGTGGCTGGCGAGCACCCGTAGCGCGGAAGTCCCGGTTATTAGCGCGCCGGTGC

200 210 220 230 240 250
Native ACGGAGCAGCGCGTCCGACCACGGCTCTCGCCGATGCCGAGCTGCTGATTCATGGCCCGGGTG
Optimized ACCGCGGCGCGCGTCCGACCACCGCGCTGGCGGATGCCGAACTGCTGATTCATGGTCCCGGCG

260 270 280 290 300 310 320
Native TGCCCTCCGCGCTGGCCGCTGCCCCCGCTCCCTGCAAGCGCTGTCAACAGCGTTGATCAGCAGGGT
Optimized TGCCGCCGCGCTGGCCGCTGCCGCCGCTGCCGGCGGGTGTAGCCCGCGCTGATCAGCCGTGT

330 340 350 360 370 380
Native TGTTCGCGAAACAATCCCGCTCAATCTCCAGGTGGCAGCCCTTGGCTTGGCATGAGCCCTCC
Optimized GGTTCGCGAAACATCCCGCTGAACTCCAGGTGGCGCCCTGGTCTGGCATGAGCCGCC

390 400 410 420 430 440
Native TTCCCCTATCTTCGCTTTCGAGGCCTCCGCGCCATGGTCCGCGCGCTGCATCAGCGGTGGTGC
Optimized TTCCCGCATCTGCTTTGAGGCGCCGCTCATGGTCCGCCGCTTGCATTAGCGGTGGCGCG

450 460 470 480 490 500 510
Native CGATGGACCCCGAGCGAGTGGACCACTGATCCACCGGGGGACACCGCATGGGCCGCAACCTGCG
Optimized CGATGGACCCGGAGCGTGTGGATCAACTGATTCACCGTGGCACCGTATGGGCCGTAACCTGCG

520 530 540 550 560 570
Native CAGGCCCTGGTTTGGCGGAGTGGCTGCCGGGTGGGACCACACGGCGCTGGCGGTGCTCAG
Optimized TCGTCCGCTGGTGCAGCGGAATGGCTTCCGGGTGGCACCACACCGCGCTGGCGGTCTGAC

580 590 600 610 620 630 640
Native GGCCTCGGCTTCTTGTGACCCAACTCGTGAGTGGAGCGCTCTGCATCCACCGATGGCGCTCA
Optimized GGTCTGGGCTGCGGGTGAACCAACTGGTAGCGGTAGCGCTCTGCACCCGCCGATGGCGCTGA

650 660 670 680 690 700
Native AGCAGGAGCTTGTCCGAGAGGGCTGTCGGCGATCGCGCTGGCACCCTGATCCGCAGGCGCT
Optimized AGCAAGAGCTGTTGCAGGAAGTCTGAGCGGATTCGCCCGGGCACCACCGACCCGCAGGCGCT

710 720 730 740 750 760

Native GCTGGCGGCGGTTGGGATCCTTTTCAGGCGCTCACGATGGGTGTGTGCTGGGGCGGTTGAG
 Optimized GCTGGCGGCGGTTGGTGATCCGTTCCAAAGCGCTGACCATGGGTGTGCTGCTGGGTGCGGTTGAG

770 780 790 800 810 820 830

Native AGGATCAACCATCCTTCTGGCCGGAGGCAGCCAGATGGTGGCGGTCTGGCTCTTGCCTGG
 Optimized CGTGACCAAGCCGATCTGCTGGCCGGTGGCAGCCAAATGGTGGCGGTCTGGCCGCTGGCGCTGG

840 850 860 870 880 890

Native CAGCGCTGCCAGCACACCAGCGCAACTTCTCTGCCACAAGGTGATGCTGGGACCACGGCATG
 Optimized CGGCGCTGCCGGCGCACCAGCGTCAACTGTGTGCCACCGTGTATGCTGGGACCACCGCGTG

900 910 920 930 940 950

Native GTTGGCGACGGAATCGCTGGAGTCAAGCGACCGGTTCTCATC..CCTCGAAGCCCTGCTGCTTC
 Optimized GCTGGCGACCGAGAGCCTGGAAAGCGCGACCGG..CAGCAGCAGCCCTGGAGGCGCTGCTGCTGC

960 970 980 990 1000 1010 1020

Native GCCTTGAACAGCACTTCGGCATTGCCCTTGGAGGCCATATGCAGCGGGGTGCGGTTCTCCACAG
 Optimized GTCTTGAACAGCACTTTGGTATTGCCCTTGGAAAGCCTATCGCGCGCGGGCCCTGCGTTCTAGCCACAG

1030 1040 1050 1060 1070 1080

Native TCGCATCAACAATTCGCGGACTTTGAATCCGGTTACGTGAAGGAAGGGTGGGTGCCGGGGGG
 Optimized CCGTCAACAGCAACTTCGCGTGAATTTGAGAGCGGTTATGTGAAGAAGGTGTGGGTGCCGGTGGC

1090 1100 1110 1120 1130 1140 1150

Native CTGGCCCTGCTGGCGGCCTTGCCCGGGGTGGACCACGGCACGCTGCTGCAGGGGTGCTGACGATG
 Optimized CTGGCGCTGCTGGCGGCGCTTGCCCTGGTGTGATCACGGCACCTGCTGCAGGGCTGCTGATGATG

1160 1170 1180 1190 1200

Native CGATGGATCGGTGCTGCAGGCGCAGGGTCCGTCCTCCGCCCGCCCTTAG
 Optimized CGATGGACCGTCTGCTGCAGAGCGCAAGGTTCGTCCTCCGCCGCGCCGTA

Figure S1. Nucleotide differences between the native nucleotide sequence of *Synechococcus sp.* WH7803 *cobT* and the nucleotide sequence optimized for expression in *E. coli*. Changes made for optimal expression in *E. coli* are shown in red bold typeface.