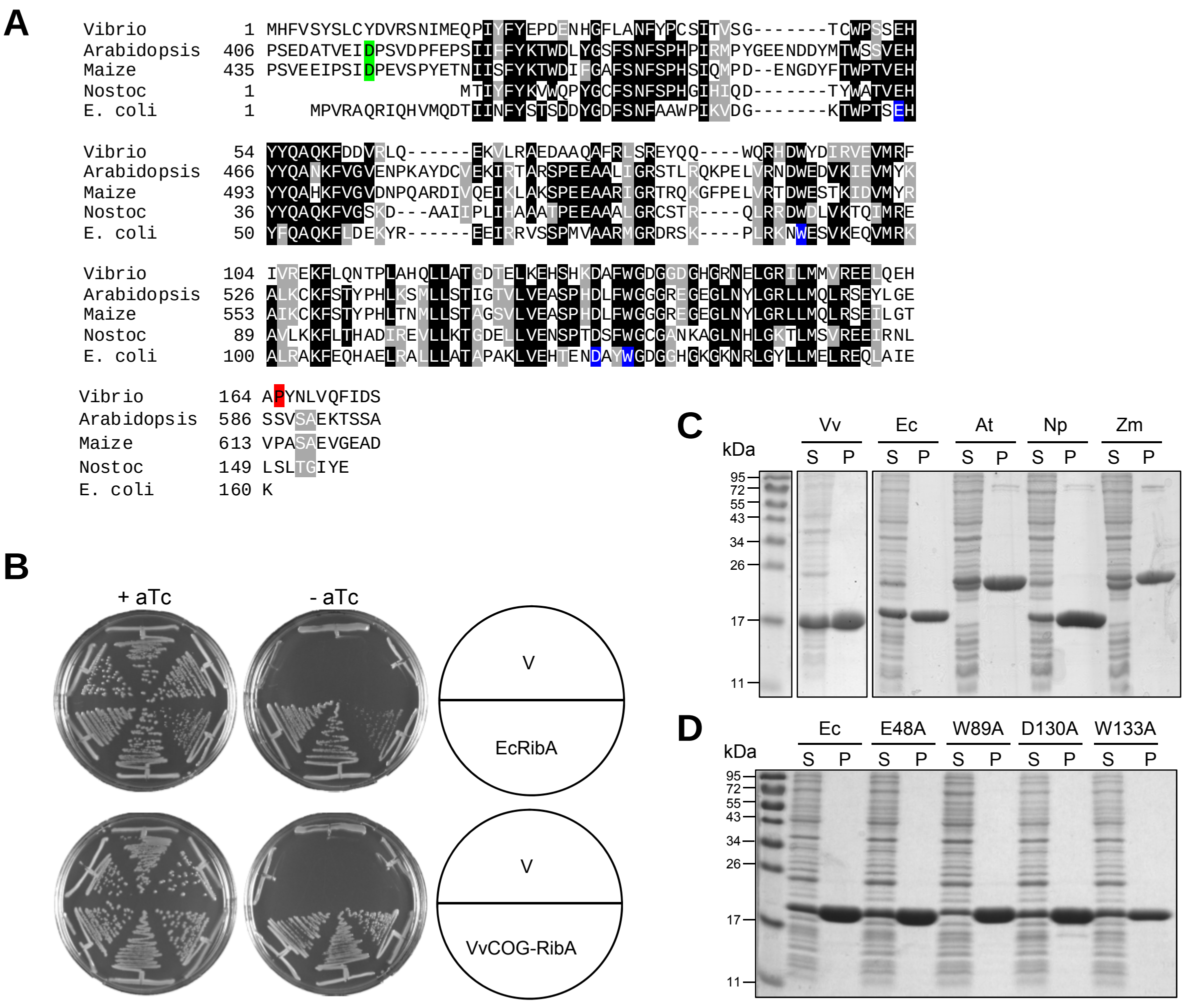


```
Vibrio         1 MHFVSYSLCYDVRSNIMEQPIYFYEPDENHGFLANFYPCSITVSG-------TCWPSSEH
Arabidopsis  406 PSEDATVEIDPSVDPFEPSIIFFYKTWDLYGSFSNFSPHPIRMPYGEENDDYMTWSSVEH
Maize        435 PSVEEIPSIDPEVSPYETNIISFYKTWDIFGAFSNFSPHSIQMPD--ENGDYFTWPTVEH
Nostoc         1                    MTIYFYKVWQPYGCFSNFSPHGIHIQD-------TYWATVEH
E. coli        1     MPVRAQRIQHVMQDTIINFYSTSDDYGDFSNFAAWPIKVDG-------KTWPTSEH

Vibrio        54 YYQAQKFDDVRLQ------EKVLRAEDAAQAFRLSREYQQ----WQRHDWYDIRVEVMRF
Arabidopsis  466 YYQANKFVGVENPKAYDCVEKIRTARSPEEAALIGRSTLRQKPELVRNDWEDVKIEVMYK
Maize        493 YYQAHKFVGVDNPQARDIVQEIKLAKSPEEAARIGRTRQKGFPELVRTDWESTKIDVMYR
Nostoc        36 YYQAQKFVGSKD---AAIIPLIHAAATPEEAAALGRCSTR----QLRRDWDLVKTQIMRE
E. coli       50 YFQAQKFLDEKYR------EEIRRVSSPMVAARMGRDRSK----PLRKNWESVKEQVMRK

Vibrio       104 IVREKFLQNTPLAHQLLATGDTELKEHSHKDAFWGDGGDGHGRNELGRILMMVREELQEH
Arabidopsis  526 ALKCKFSTYPHLKSMLLSTIGTVLVEASPHDLFWGGGREGEGLNYLGRLLMQLRSEYLGE
Maize        553 AIKCKFSTYPHLTNMLLSTAGSVLVEASPHDLFWGGGREGEGLNYLGRLLMQLRSEILGT
Nostoc        89 AVLKKFLTHADIREVLLKTGDELLVENSPTDSFWGCGANKAGLNHLGKTLMSVREEIRNL
E. coli      100 ALRAKFEQHAELRALLLATAPAKLVEHTENDAYWGDGGHGKGKNRLGYLLMELREQLAIE

Vibrio       164 APYNLVQFIDS
Arabidopsis  586 SSVSAEKTSSA
Maize        613 VPASAEVGEAD
Nostoc       149 LSLTGIYE
E. coli      160 K
```

Figure S1 Sequences of fused and free-standing COG3236 proteins, evidence for RibA activity of a RibA-COG3236 fusion, and purification of recombinant COG3236 proteins

(**A**) Alignment of COG3236 domains from the N-terminal region of the COG3236-RibA fusion of *Vibrio vulnificus* YJ016, the C-terminal region of the RIBR-COG3236 fusions of *Arabidopsis* and maize (At3g47390 and GRMZM2G090068), and the free-standing COG3236 proteins of *E. coli* (YbiA) and *Nostoc punctiforme* PCC 73102 (Npun_R5314). To express the fused COG3236 domains alone, the red highlighted residue in the *V. vulnificus* sequence was changed to a stop codon, and the green highlighted residues in the plant sequences were changed to start codons. The four conserved residues changed to alanine in *E. coli* YbiA are highlighted in blue.

(**B**) Complementation of the riboflavin auxotrophy of an *E. coli ribA*$^-$ strain by the COG3236-RibA fusion from *V. vulnificus*. The *E. coli* Ptet-*ribA* strain was transformed with vector alone (V) or encoding *V. vulnificus* COG3236-RibA or *E. coli* RibA (as a positive control). Three independent colonies per construct were plated on LB medium containing 50 µg/ml kanamycin, 100 µg/ml carbenicillin, and 0.2% (w/v) arabinose, plus or minus anhydrotetracycline (aTc) 1 µg/ml.

(**C**) Purification of the free-standing COG3236 proteins of *E. coli* (Ec) and *N. punctiforme* (Np), and the separated COG3236 domains of the *V. vulnificus* (Vv), *Arabidopsis* (At), and maize (Zm) fusion proteins. Proteins were isolated by Ni^{2+}-affinity chromatography, separated by SDS-PAGE (15% gel), and stained with Cooomassie blue. Lanes contained 10 μg of total soluble proteins (S) or purified protein (P).

(**D**) Purification of four single-residue mutant forms of *E. coli* COG3236, and the wild type protein (Ec). Isolation, separation, and visualization were as in **C**.

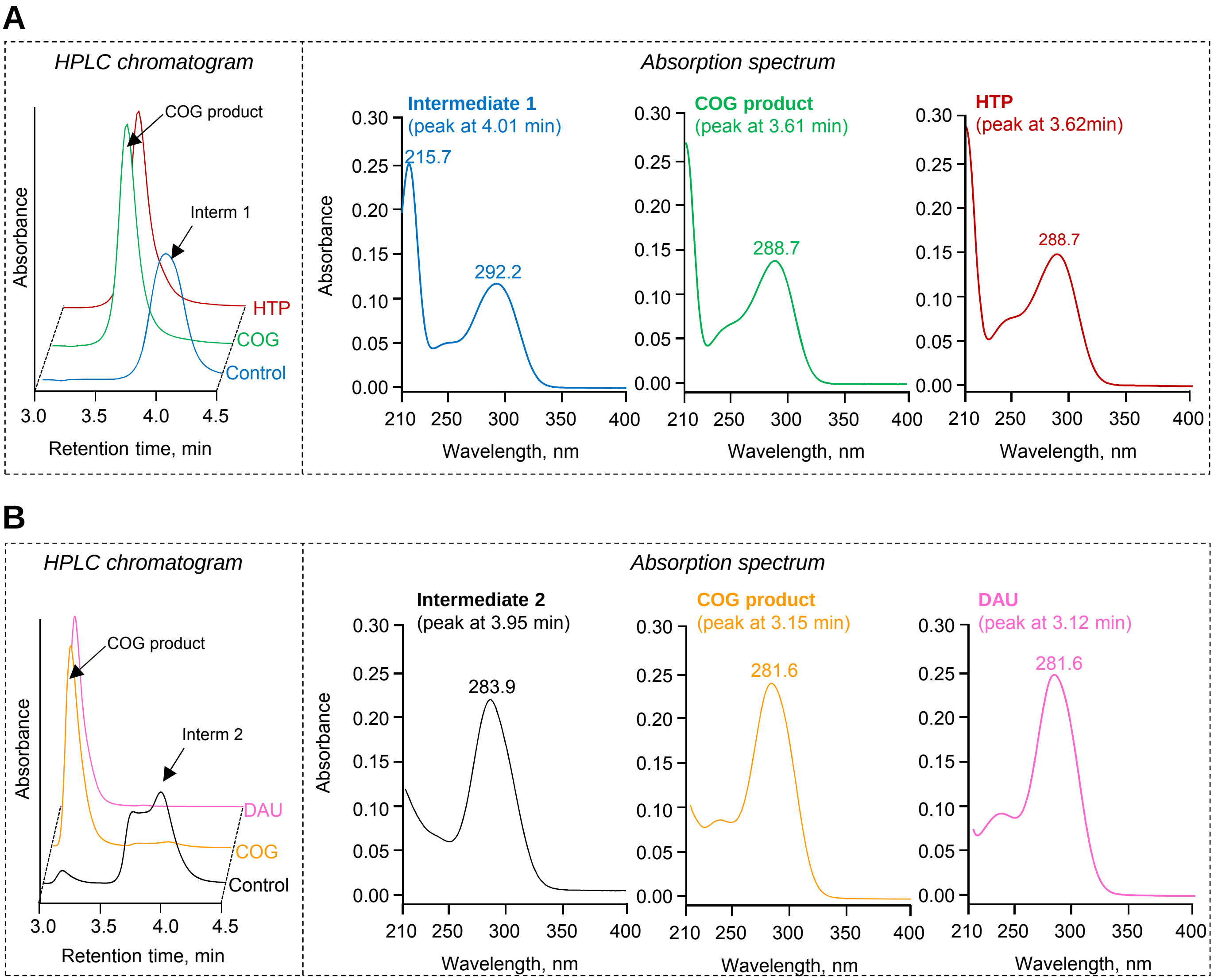


Figure S2 Chromatograms and absorption spectra of riboflavin synthesis intermediates and their COG3236-derived products

Intermediates 1 (**A**) and 2 (**B**) were incubated with *V. vulnificus* COG3236 (1 *µ*g and 20 *µ*g respectively) for 30 min at 22°C to obtain complete conversion to reaction products. HPLC chromatograms are shown in the leftmost boxes, and the spectra of the peaks are shown to their right. The products formed by other COG3236 proteins had identical retention times and spectra. The spectra of authentic 4-hydroxy-2,5,6-triamino-pyrimidine (HTP) and 5,6-diaminouracil (DAU) are shown for comparison with the COG3236 products from intermediates 1 and 2, respectively. Note that the reactions with intermediate 2 were separated on a different HPLC column (Spherisorb ODS-2 RP-18 column) to the one used for Figure 2, which resulted in different retention times. The shoulder on the intermediate 2 peak is due to partial resolution of the two anomeric forms.

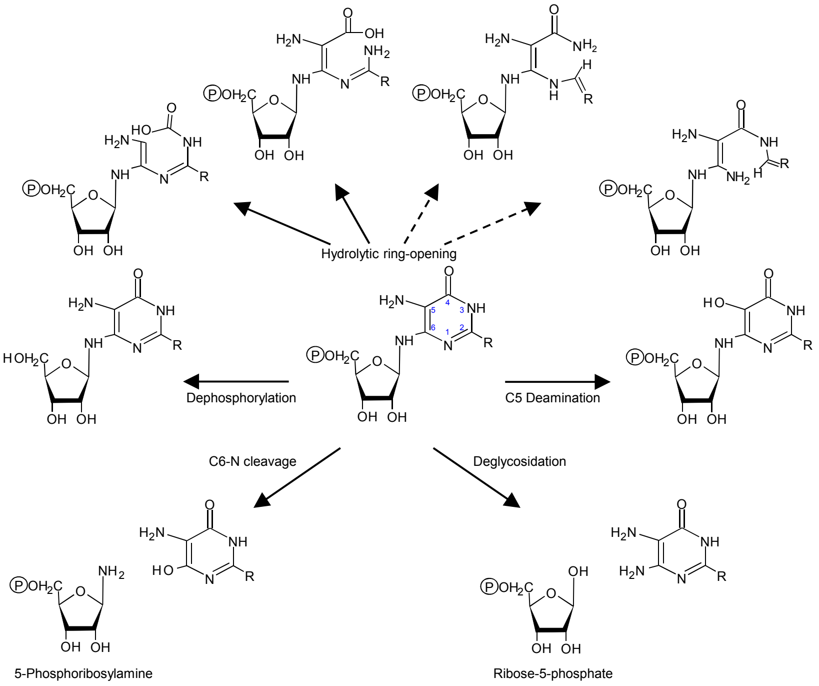


Figure S3 Products of predicted enzyme reactions undergone by intermediates of riboflavin synthesis

Enzyme reactions were predicted for the first and second intermediates of riboflavin synthesis as described in the text. R = NH_2 (intermediate 1) or OH (intermediate 2). Dashed arrows show reactions that could occur only for intermediate 2. The products from hydrolysis of the C4-C5, N3-C4, and N1-C2 bonds all contain a carbamate moiety (-NH-COOH) that is expected to decompose spontaneously to the corresponding amine, with loss of CO_2. The product from hydrolysis of the N3-C4 bond could tautomerize to give an imine group at C5 that spontaneously hydrolyzes to a keto group and NH_3. 5-Phosphoribosylamine formed by cleavage of the C6-N bond would hydrolyze spontaneously to ribose-5-phosphate and NH_3.

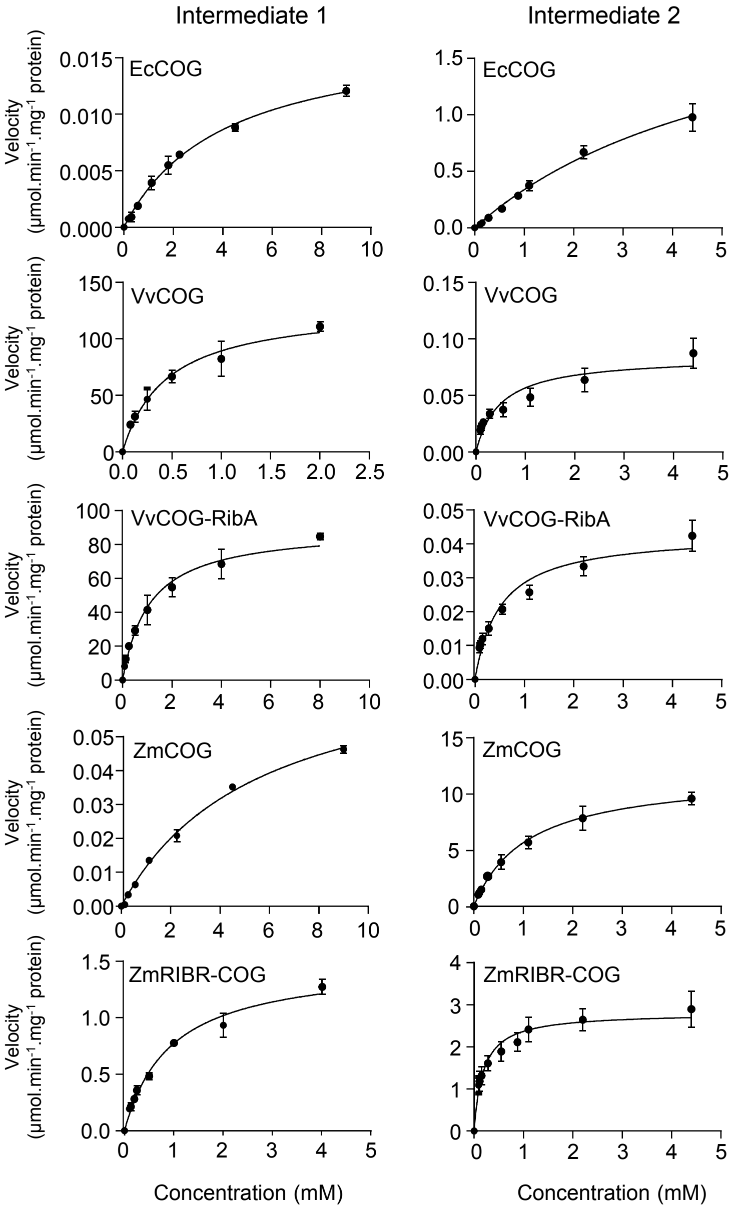


Figure S4 Velocity versus concentration plots for *E. coli* COG3236 (EcCOG), the isolated COG3236 domains from *V. vulnificus* (VvCOG) and maize (ZmCOG), and the full-length maize RIBR (ZmRIBR-COG) and *V. vulnificus* RibA (VvCOG-RibA) fusion proteins

Assays were run at 22°C for 10 min. Products were separated by HPLC and quantified by UV absorbance. Data are means ± S.E.M. for three independent replicates (except for ZmCOG with intermediate 1 where only duplicates were analyzed).

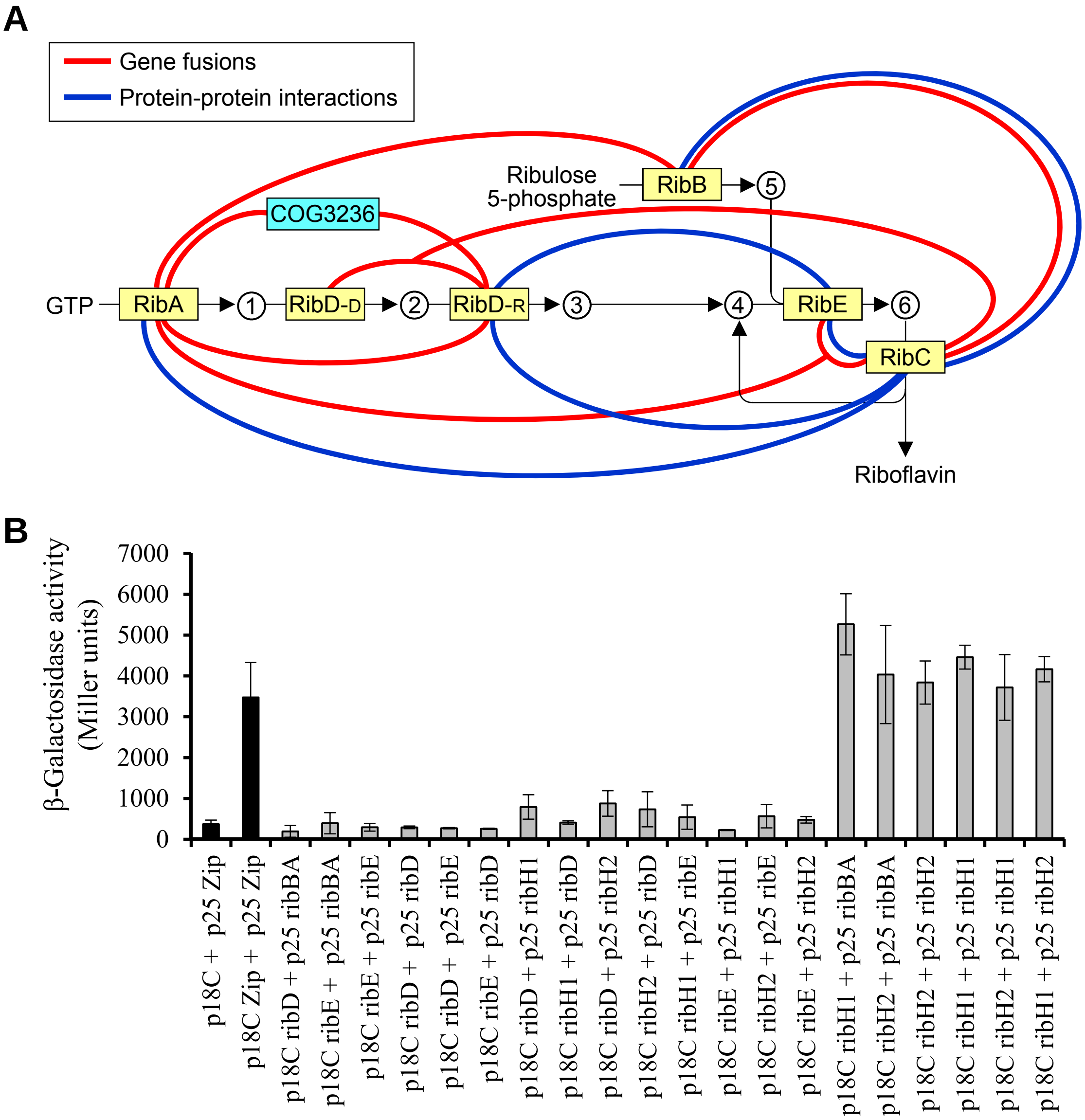


Figure S5 Evidence from gene fusions, protein-protein interaction data, and two-hybrid experiments for a multi-enzyme riboflavin biosynthesis complex

(**A**) The riboflavin pathway from plants and bacteria is shown in outline. Enzymes are designated by their names in *E. coli*; the deaminase and reductase domains of *E. coli* RibD are given –D and –R suffixes, respectively. Blue arcs link proteins for which interactions have been experimentally demonstrated. Red arcs link proteins whose genes are fused, pairwise or in threes, in some genomes. For simplicity, protein interactions with RibD are shown as links to its reductase domain. Pathway intermediate abbreviations: 1 – 2,5-diamino-6-ribosylamino-4(3*H*)-pyrimidinone 5′-phosphate); 2 – 5-amino-6-ribosylamino-2,4(1*H*,3*H*)-pyrimidinedione 5′-phosphate; 3 – 5-amino-6-ribitylamino-2,4(1*H*,3*H*)-pyrimidinedione phosphate; 4 – 5-amino-6-ribityl-amino-2,4(1*H*,3*H*)-pyrimidinedione; 5 – 3,4-dihydroxy-2-butanone-4-phosphate; 6 – 6,7-dimethyl-8-ribityllumazine.

(**B**) Bacterial two-hybrid analysis of interaction between enzymes of the *Sinorhizobium meliloti* riboflavin (Rib) synthesis pathway. Interactions are expressed in Miller units of LacZ activity; data are means ± S.D. for three independent replicates. Interaction between Rib-proteins is shown in gray and controls are shown in black. *S. meliloti* RibH (6,7-dimethyl-8-ribityllumazine synthase) corresponds to *E. coli* RibE, and *S. meliloti* RibE (riboflavin synthase) corresponds to *E. coli* RibC.

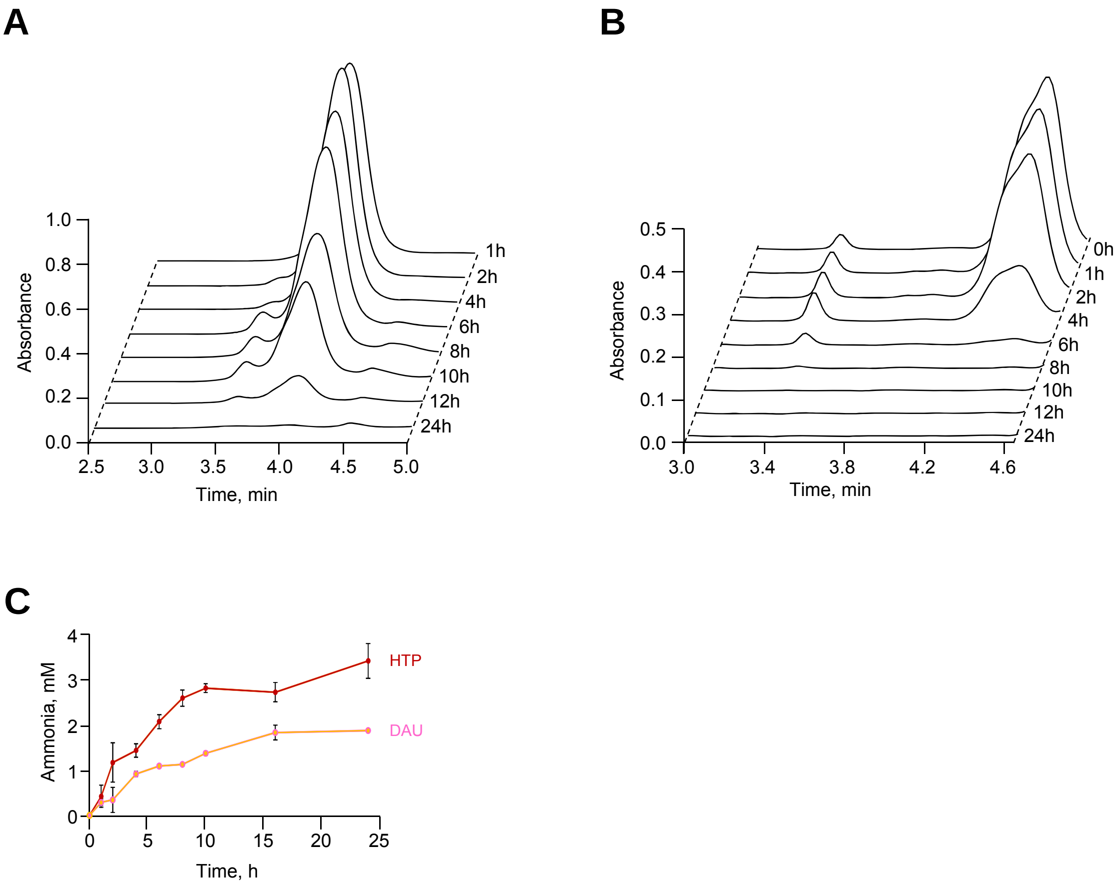


Figure S6 Time-dependent decomposition of riboflavin synthesis intermediates and related pyrimidines

(**A, B**) Chromatogram showing intermediates 1 (**A**) and 2 (**B**) that were incubated in assay buffer (pH 8.0) at 22°C for the times indicated, and analyzed by HPLC with UV detection at 293 nm (intermediate 1) or 284 nm (intermediate 2). The shoulder on the intermediate 2 peak is due to partial resolution of the two anomeric forms.
(**C**) Ammonia release from 4-hydroxy-2,5,6-triaminopyrimdine (HTP) and 5,6-diaminouracil (DAU). Data are means ± S.E.M. for three independent assays. Initial concentrations of HTP and DAU were 5 mM. Note that these are the products formed by cleavage of the *N*-glycosidic bond, not those formed by cleavage of the C_6-N bond.

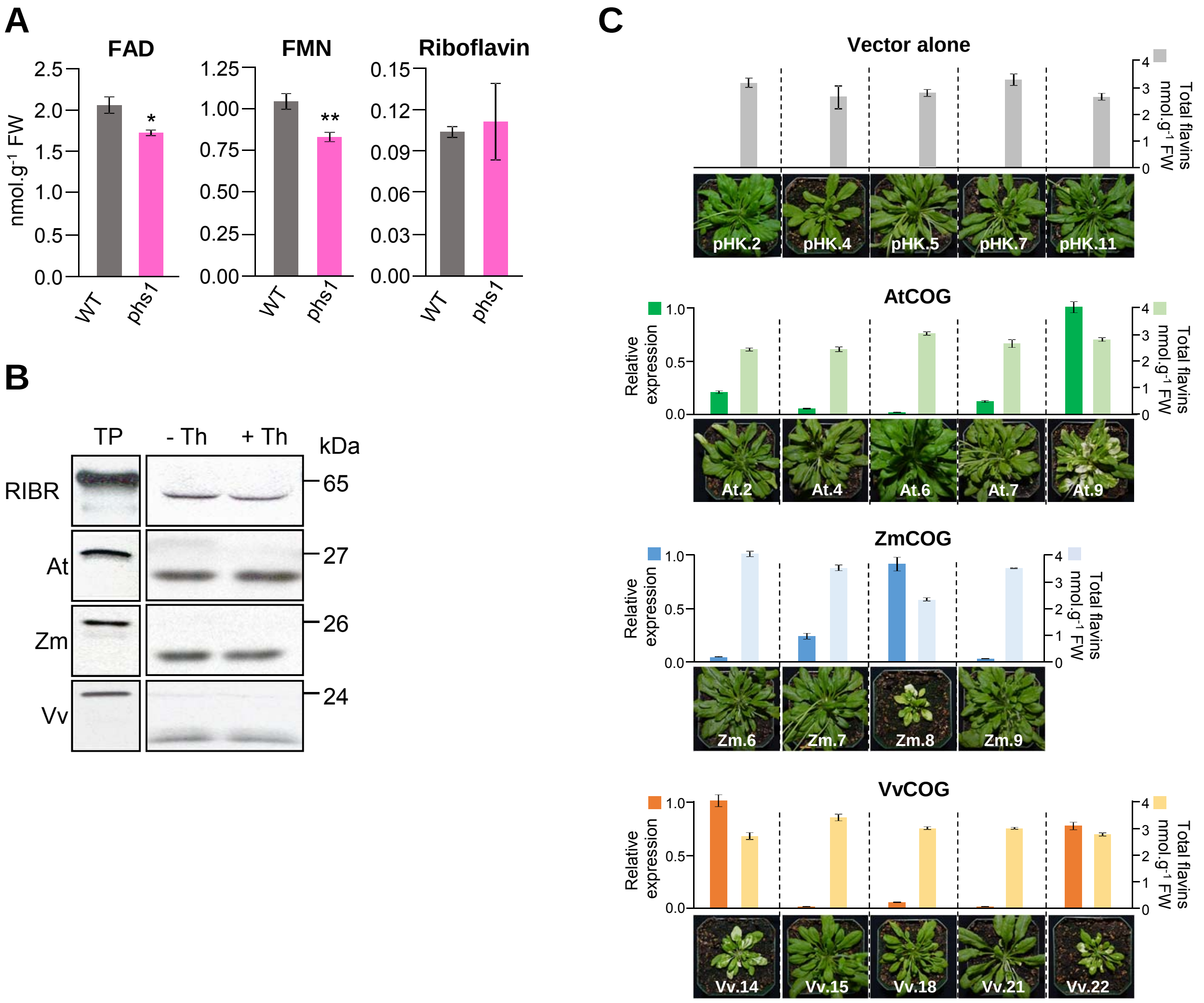


Figure S7 Flavin contents of the *Arabidopsis phs1* mutant and *Arabidopsis* plants transformed with COG3236 genes

(**A**) Contents of FAD, FMN, and riboflavin in leaf tissue from *phs1* mutant and wild type plants. Data are means ± S.E.M. for three replicates. Values that differ significantly between mutant and wild type are marked with one ($P < 0.05$) or two ($P < 0.01$) asterisks.

(**B**) *In vitro* assays to confirm chloroplast import of engineered COG3236 proteins. Sequences encoding the COG3236 domains of the *Arabidopsis* (At), maize (Zm), and *V. vulnificus* (Vv) fusion proteins, preceded by the pea Rubisco small subunit plastid targeting sequence, were transcribed and translated *in vitro* in the presence of $[^3H]$leucine. Translation products (TP) were incubated in the light for 20 min with purified pea chloroplasts, incubated with (+Th) or without (-Th) thermolysin, and repurified on a Percoll gradient. Proteins were separated by SDS-PAGE and visualized by fluorography, adjusting exposure times to give comparable band intensities. Full length *Arabidopsis* RIBR was included as a positive control.

(**C**) T_1 generation plants transformed with vector alone or encoding chloroplast-targeted COG3236 domains from *Arabidopsis* (AtCOG), maize (ZmCOG), or *V. vulnificus* (VvCOG) fusion proteins were analyzed for transgene expression and flavin content. Bars show transgene expression and total flavin content in leaves of individual plants, which are pictured beneath the bars. Plants with the highest transgene expression showed variable degrees of bleaching. Transgene expression is shown relative to the highest value observed for each transgene (= 1.0). Data are means ± S.E.M. for three technical replicates

Table S1 Primer sequences used in this study

Underlined is the restriction site; italicized is the Kozak sequence (for cloning in pGEM-4Z), and bold is the Shine-Dalgarno sequence (for cloning in pBAD33 and in pBSK).

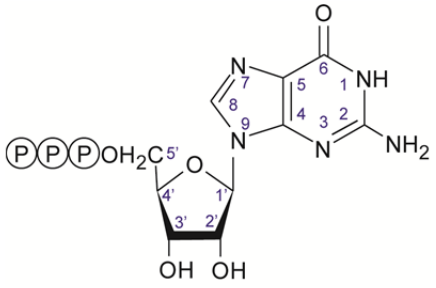
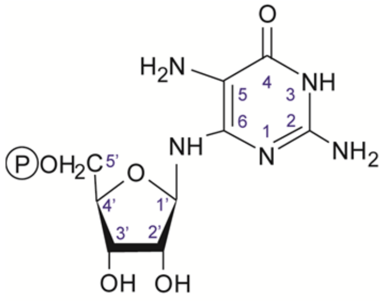
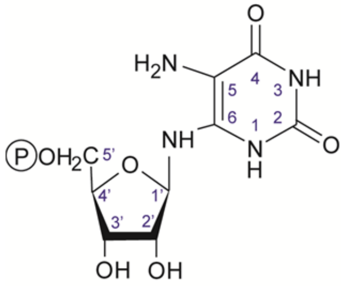
Name	Sequence 5' → 3'	Experiment
pBAD-Vv-F	CATATGGGTACC**AGGAGGAACAGCT**ATGGAGCAACCGATTTATTTCTAT	Functional complementation of *E. coli Ptet-ribA* mutant in pBAD33
pBAD-Vv-R	CATATGGCATGCCTAAGGTTTAACGAAGTTGCGATC	
pBAD-EcRibA-F	CATATGGGTACC**AGGAGGAACAGCT**ATGCAGCTTAAACGTGTGGCA	
pBAD-EcRibA-R	CATATGGCATGCTTATTTGTTCAGCAAATGGCCCA	
pET-YbiA-NdeF	GTGGCATATGCCCGTTCGAGCACAAAG	Cloning in pET28b for protein expression
pET-YbiA-EcoR	CAGTCAGAATTCTTACTTCTCTATAGCCAATTGTTC	
pET-FLRibA-NdeF	GTGGCATATGGAGCAACCGATTTATTTCTA	
pET-VvCOG-EcoR	CAGTCAGAATTCCTACGCGTGCTCTTGCAACTC	
pET-AtCOG-NheF	GTGGGCTAGCATGCCCAGTGTTGACCCCTTTG	
pET-AtCOG-EcoR	CAGTCAGAATTCTCAGGCCGAAGAAGTCTTTTC	
pET-ZmCOG-BamHIF	GTGGGGATCCAATGCCAGAAGTATCTCCATATGA	
pET-ZmCOG-HindR	CAGTCAAAGCTTTCAATCTGCCTCCCCTACTTC	
pET-NpCOG-NdeF	GTGGCATATGACAATCTACTTTTATAGCACTCG	
pET-NpCOG-EcoR	CAGTCAGAATTCTCATTTAACACTTCTACGATCAACA	
pET-YBIA-E48A-F	CTGGCCTACCTCAGCACACTATTTTCAGG	Site-directed mutagenesis of EcYbiA (COG3236) protein
pET-YBIA-E48A-R	CCTGAAAATAGTGTGCTGAGGTAGGCCAG	
pET-YBIA-W89A-F	CCTCTGCGTAAAAACGCGGAGTCGGTCAAAG	
pET-YBIA-W89A-R	CTTTGACCGACTCCGCGTTTTTACGCAGAGG	
pET-YBIA-D130A-F	GAGCATACGGAAAACGCTGCTTACTGGGGAGAC	
pET-YBIA-D130A-R	GTCTCCCCAGTAAGCAGCGTTTTCCGTATGCTC	
pET-YBIA-W133A-F	GGAAAACGATGCTTACGCGGGAGACGGTGGTCATG	
pET-YBIA-W133A-R	CATGACCACCGTCTCCCGCGTAAGCATCGTTTTCC	
SSU-PCR1-F	ATGGCTTCTATGATATCCTCT	SOeing the targeting peptide from the pea Rubisco small subunit (SSU) to COG3236 domains
SSUAt-PCR1-R	ACACTGGGCATTGGAGGCCACACCTGCAT	
SSUAtCOG-PCR2-F	GTGGCCTCCAATGCCCAGTGTTGACCCCTT	
AtCOG-PCR2-R	TCAGGCCGAAGAAGTCTTTT	
SSUZm-PCR1-R	ACTTCTGGCATTGGAGGCCACACCTGCAT	
ssuZmCOG-PCR2-F	TGTGGCCTCCAATGCCAGAAGTATCTCCATATGAAA	
ZmCOG-PCR2-R	TCAATCTGCCTCCCCTACTT	
SSUVv-R	TAAATCGGTTGCTCCATTGGAGGCCACACCTGCAT	
ssuVvCOGPCR2-F	GTGGCCTCCAATGGAGCAACCGATTTATTTCTATG	
VvCOGPCR2-R	CTACGCGTGCTCTTGCAACTC	
pGEM-SSU-COG-F	CAGTCAGAATTC*ACC*ATGGCTTCTATGATATCCTCT	Cloning the SSU-COG domain fusions in pGEM-4Z for chloroplast import assay
pGEM-FLAt3g47390-R	CAGTCAAAGCTTTCAGGCCGAAGAAGTCTTTT	
pGEM-SSUZmCOG-R	CAGTCAGAATTCTCAATCTGCCTCCCCTACTTC	
pGEM-SSUVvCOG-R	CAGTCAGGATCCCTACGCGTGCTCTTGCAACTC	
pFMV-SSU-F	CAGTCATCTAGAATGGCTTCTATGATATCCTCTTC	Cloning of the SSU-COG fusions to pFMV*
pFMV-FLAt3g-R	CAGTCAGGATCCTCAGGCCGAAGAAGTCTTTTC	
pFMV-ZmCOG-R	CAGTCAGGATCCTCAATCTGCCTCCCCTACTT	
AtPyrR-qPCR-F	GTAAAGTCGCCGGAGAAGGAT	RT-qPCR experiment
AtPyrR-qPCR-R	ACGCTTCCGCCGGTTT	
SSU-qPCR-F	GCGGTGGCTCCATTCG	
SSU-qPCR-R	TGACCTTCTTAACTGGGAATCCA	
UBC21-qPCR-F	TGGACGCTTCAGTCTGTGTGTA	
UBC21-qPCR-R	GGACTGTCCGGCTCAGGAT	
YSL8-qPCR-F	GCGTCTCGTCGTCATTCGTT	
YSL8-qPCR-R	CATCCATCTGCATACAGGTCTCA	
RibBA_BTH_F	ATGCGGATCCCATGTCCTACGACCAGAAGCG	For bacterial double-hybrid experiments
RibBA_BTH_R	ATGCGGTACCCGCAGAATCTCGGTAGCGGCGA	
RibD_BTH_F	ATGCGGATCCCATGGGCGAGCCGGCACGTGAAGA	

RibD_BTH_R	ATGCGGTACCCGGCTATCTCTCTCGTAATCCT	
RibE_BTH_F	ATGCGGATCCCATGTTTACAGGCATCATCAC	
RibE_BTH_R	ATGCGGTACCCGTTCCTGCTTGGGAGCCGGGA	
RibH1_BTH_F	ATGCGGATCCCATGGCGAAGATCAAGCCCGT	
RibH1_BTH_R	ATGCGGTACCCGCTTCTCTGCGCCCAATCTTT	
RibH2_BTH_F	ATGCGGATCCCATGTTCGCCCGACGGGATCA	
RibH2_BTH_R	ATGCGGTACCCGGGCGTCGACCAGCGCAAGCT	
pKT25-F	ATTCGGTGACCGATTACCTG	To sequence clones for double-hybrid experiments
pKT25-R	TTCCCAGTCACGACGTTGTA	
pUT18C-F	CAGTGGAACGCCACTGCAGG	
pUT18C-R	TGCGGCATACGAGCAGATTG	
OF_YbiAF	CAGCACCAGCGATGACTAC	To check the phenotype of *E. coli* deletant strain from the KEIO collection
OF_YbiAR	TGCCCTTACCATGACCACC	
OF_YbiAdw-R	ATCTGTTTTACCGCCCATAAGC	
OF_YbiAup-F	CGGCTCGCAATAACCACCA	
pBSK-YbiA-SacF	CTAGGAGCTCtTGA**AGGAAACAGCT**ATGCCCGTTCGAGCACAAAG	Cloning *ybiA* gene in pBSK vector†

*For the cloning of SSU-COG3236 from *V. vulnificus* into pFMV, the reverse primer pGEM-SSUVvCOG-R was used.

†The reverse primer for the cloning of *E. coli ybiA* in pBSK vector was pET-YbiA-EcoR.

Table S2 Chemical shifts from experimental and computed spectra of GTP, of intermediates 1 and 2 and their COG3236-mediated products, and of ribose 5-phosphate

Compound structures	Carbon position	Experimental chemical shifts (ppm)	Computed chemical shifts (ppm)
GTP 	C2	156.7	157.6
	C4	161.8	159.2
	C5	118.8	121.1
	C6	154.5	155.7
	C8	140.4	139.9
	C1’	89.2	87.1
	C2’	76.2	77.4
	C3’	73	74.7
	C4’	86.6	85.3
	C5’	67.8	66.9
Intermediate 1 	C2	153.7	155.4
	C4	162.8	164.2
	C5	101.4	99.7
	C6	162.3	161.7
	C1’	87.5	91.5
	C2’	76.2	79.3
	C3’	73.7	75.7
	C4’	85.5	87.1
	C5’	66.6	67.7
Intermediate 2 	C2	153.4	151.6
	C4	164.8	167
	C5	99	104.6
	C6	158.8	154.8
	C1’	87.7	92.8
	C2’	76.5	76.8
	C3’	73.9	75.1
	C4’	85.6	86.6
	C5’	66.6	67.8

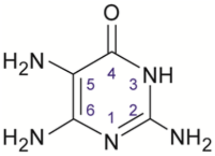
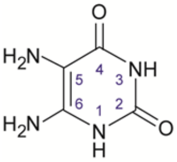
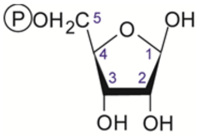
Intermediate 1 + COG3236 	C2	153.9	155.4
	C4	162.2	163
	C5	100.3	101.5
	C6	159.3	162
Intermediate 2 + COG3236 	C2	152.3	151.4
	C4	165.1	165.7
	C5	97.5	98.6
	C6	154.4	157
Ribose 5-phosphate 	C1	103.8	107.1
	C2	78.1	77.6
	C3	73.4	75.3
	C4	84.6	89.2
	C5	67.2	67.1

Table S3 Flavin contents of *E. coli* wild type, COG3236 deletant (Δ*ybiA*), and YbiA-over-expression (+ YbiA) strains

Values are means and S.E.M. of three biological replicates. **, mean value significantly different (*P* < 0.01) from the wild type.

Strain	Flavin content (pmol.mg^{-1} protein)			
	FAD	FMN	Riboflavin	Total
Wild type	242 ± 6	97.6 ± 2.3	23.5 ± 1.4	363 ± 16
Δ*ybiA*	211** ± 4	85.0** ± 2.2	21.2 ± 0.8	317** ± 9
Wild type + YbiA	236 ± 11	86.7 ± 5.1	17.0** ± 0.8	340 ± 16

Supplementary Materials and Methods

Bacterial two-hybrid analysis

To detect a riboflavin biosynthesis multi-enzyme complex, the bacterial adenylate cyclase two-hybrid system [1] was used to analyze enzymes from *Sinorhizobium meliloti*. Briefly, the coding regions of riboflavin biosynthetic enzymes were PCR-amplified and cloned into the BamHI/KpnI restriction sites of pUT18C and pKT25 plasmids. The constructs were verified by PCR and sequencing. The *cya*-deficient *E. coli* strain DHM1 was co-transformed with the pUT18C and pKT25 derivatives. Transformants were selected on LB plates supplemented with kanamycin (40 µg/ml), ampicillin (100 µg/ml), and nalidixic acid (15 µg/ml). A positive interaction between two tested proteins produced a functional adenylate cyclase leading to cAMP synthesis and *lacZ* expression. The β-galactosidase activities were determined according to the procedure by Bacterial Adenylate Cyclase Two-Hybrid System kit (Euromedex EUK001). Measurements of β-galactosidase activity were made in triplicate and the experiments were repeated at least twice.

Enzymatic determination of ammonia by coupled assay

A 100-µl reaction mixture containing 50 mM potassium phosphate, pH 8.0, 5 mM α-ketoglutarate, 0.1 mM NAPDH, 2.4 units of glutamate dehydrogenase (Sigma), and 10 µl of water (as a control) or 10 µl of an appropriately diluted sample was incubated for 5 min at room temperature. Ammonia concentration was calculated on the basis of NADPH oxidation, which was monitored at 340 nm. The detection limit was estimated as 1 µM.

In vitro chloroplast import assays

The COG3236 domains from the *V. vulnificus*, *Arabidopsis*, and maize proteins were fused to the pea Rubisco small subunit targeting peptide [2] by SOEing [3]. The fused DNA products were PCR-amplified using a forward primer specific to the targeting peptide sequence including a Kozak sequence (ACC) before the start codon and a reverse primer with a stop codon specific to each COG3236. Amplicons were cloned into pGEM-4Z (Invitrogen) digested with EcoRI-BamHI for the *V. vulnificus* sequence or with EcoRI-HindIII for the plant sequences. Constructs were checked by sequencing. Coupled in vitro transcription-translation was performed using a TnT® Coupled Wheat Germ Extract System (Promega) and [^{3}H]leucine (108 Ci/mmol) according to the manufacturer's protocol. pGEM-4Z-AtRIBR [13] was used as a positive control. Chloroplast purification from pea seedlings, import assays, and SDS-PAGE analysis were as described [4].

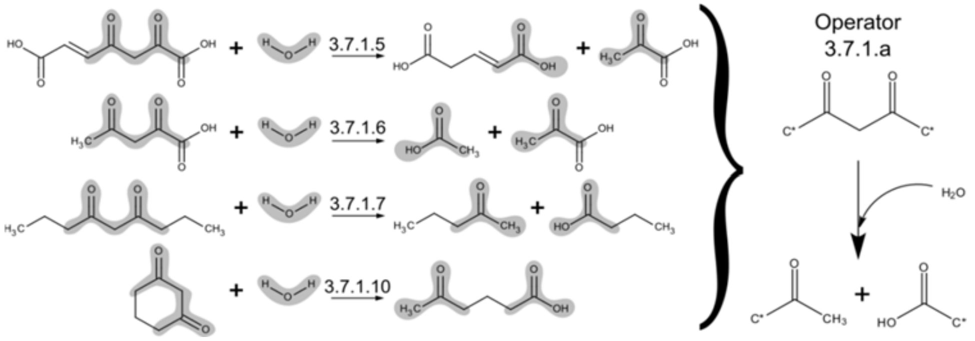
qRT-PCR analysis

Total RNA from *Arabidopsis* tissues (30 mg) was extracted using the RNeasy Plant Mini Kit (Qiagen), and treated with RNA-free DNase (Qiagen) according to the manufacturer's instructions. cDNAs were synthesized with the SuperScript® III First-Strand Synthesis System (Invitrogen) following the manufacturer's protocol using 2 µg of total RNA. Real-time PCR analyses were performed in 96-well plates using the Step One Plus System (Applied Biosystems) as described [5] except that the EvaGreen 2× qPCR MasterMix-ROX (abm) was used. Relative gene expression was calculated with the $2^{-\Delta\Delta CT}$ method [6]. For normalization of gene expression, At5g25760 (UBC21) and At5g08290 (YLS8) were used as internal standards.

Prediction of COG3236 action using BNICE.

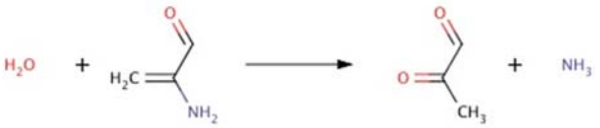
We applied the Biochemical Network Integrated Computational Explorer (BNICE) [24,25] to predict potential biochemical activities of the COG3236 protein given riboflavin synthesis intermediates 1 or 2 as substrate, and restricting co-substrates to water and oxygen alone. BNICE automatically generates novel chemical transformations based on reactions and chemical motifs abstracted from biochemical literature. Reactions are grouped using the first three divisions of the Enzyme Commission (EC) classification system. Then common chemical motifs in substrate structure and bond transformation are abstracted into generalized operators, which describe the chemistry performed by the EC subclass (Scheme 1).

Scheme 1. Abstraction of reactions into a general operator.

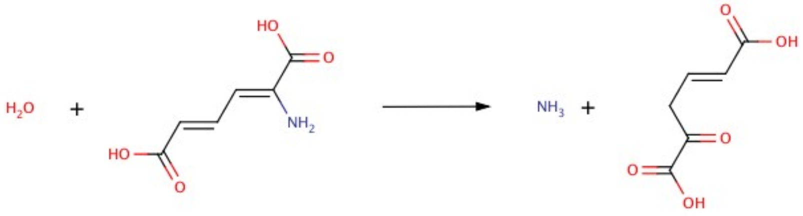


The operators used to predict enzymatic transformations and products that might result from the action of COG3236 are given below. Examples of matching reactions from the KEGG biochemistry database are included for comparison.

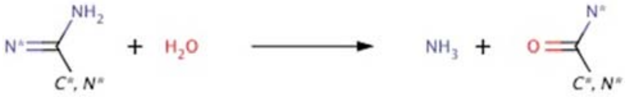
3.5.99a – Predicts the 5-deamination hydrolysis reaction



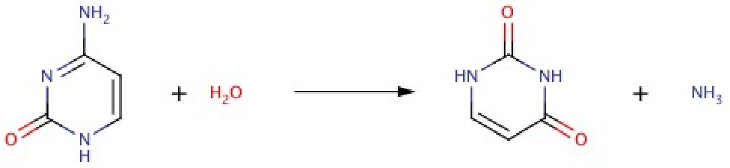
3.5.99.5: 2-Aminomuconate aminohydrolase



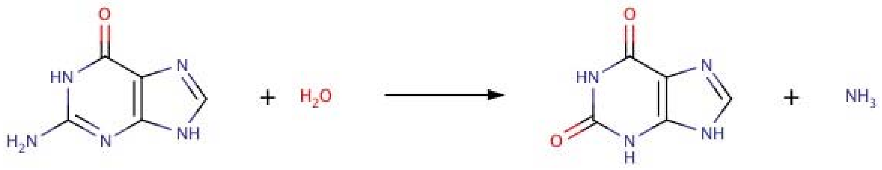
3.5.4a – Predicts the 6-hydrolytic cleavage reaction



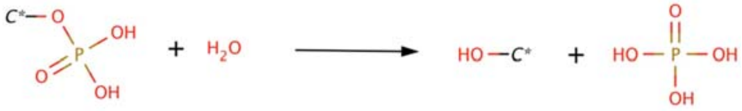
3.5.4.1: Cytosine aminohydrolase



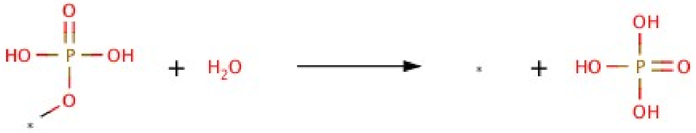
3.5.4.3: Guanine aminohydrolase



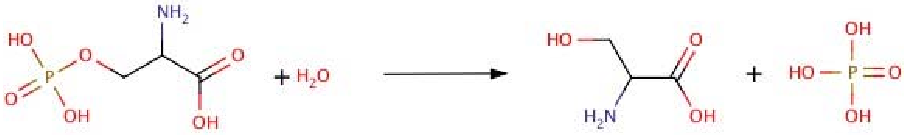
3.1.3a – Predicts the dephosphorylation reaction



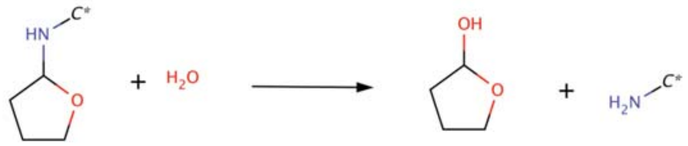
3.1.3.1: Alkaline phosphatase



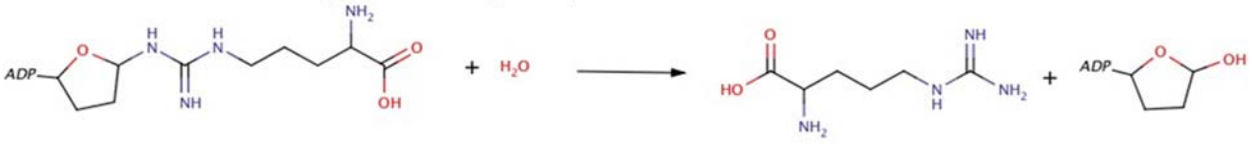
3.1.3.3: Phosphoserine phosphatase



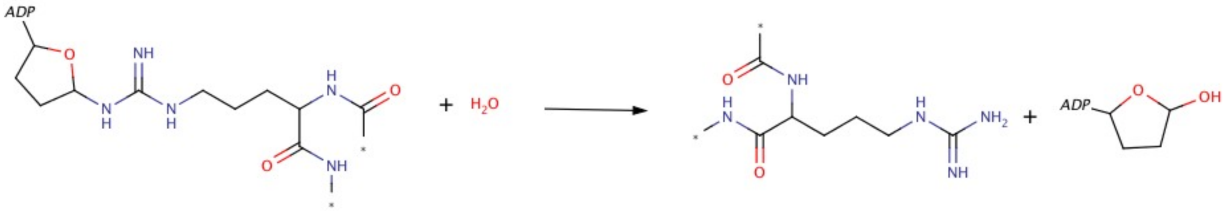
3.2.2b – Predicts the deglycosylation reaction



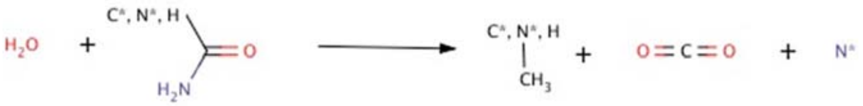
3.2.2.19: ADP-ribose-L-arginine cleavage enzyme



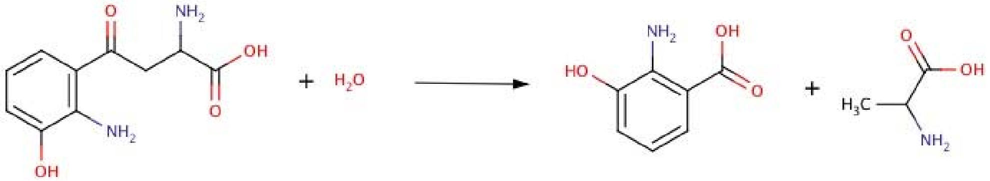
3.2.2.24: Azoferredoxin glycosidase



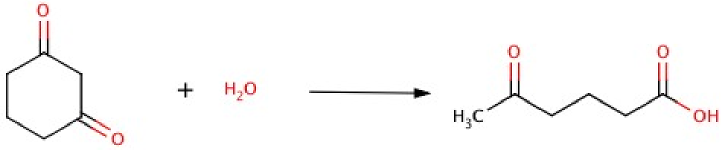
3.7.1d – Predicts the hydrolysis of 4,5 carbon-carbon bond, and subsequent spontaneous decarboxylation of the carbamate intermediate



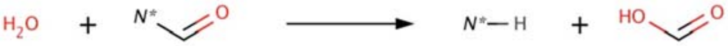
3.7.1.3: Hydroxy-L-kynurenine hydrolase



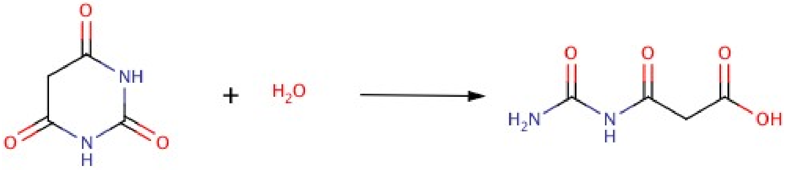
3.7.1.10: Cyclohexane-1,3-dione acylhydrolase (decyclizing)



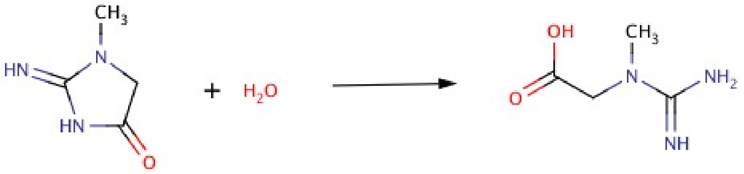
3.5.2a – Predicts the hydrolysis of 3,4 carbon-nitrogen bond



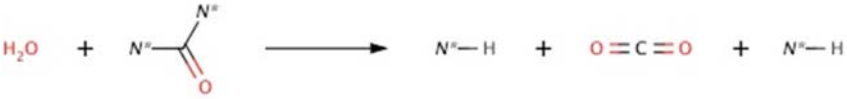
3.5.2.1: Barbiturate amidohydrolase



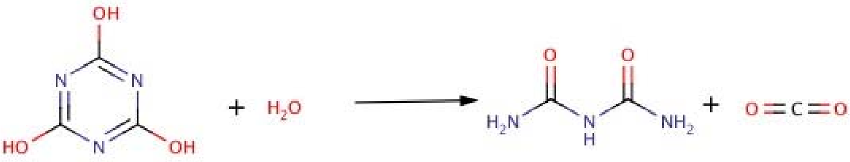
3.5.2.10: Creatinine amidohydrolase



3.5.2b – Predicts the hydrolysis of 1,2 and 2,3 carbon-nitrogen bonds in the second intermediate



3.5.2.15: Cyanuric acid amidohydrolase



NMR experiments

Intermediate 1 and 2 biosynthesis and COG3236 assays were performed as above but using 5 mM U-^{13}C, U-^{15}N-GTP (98-99%) (Cambridge Isotope Laboratories) as starting substrate. Sample volumes were typically 50 µl. NMR spectra were obtained using a custom-built 1.5-mm cryogenic high-temperature superconducting radio-frequency (RF) probe that is optimized for ^{13}C detection [56] on an Agilent VNMRS-600 spectrometer at 150.8 MHz. Data were acquired using VnmrJ 3.2 software over a spectral width of 251 ppm centered at 110 ppm. The acquisition time was 0.865 seconds for 32768 complex data points (Agilent/Varian, "np" = 65536 points) over that spectral width. A 45° RF-pulse was used with a 3.0 second relaxation delay, and composite-pulse (WALTZ-16) ^{1}H decoupling was used throughout the acquisition time and the post-acquisition delay. The time domain data were processed with NMRPipe [7] using a 2 Hz exponential line broadening, zero-filling 2x, and Fourier transformation. Depending upon solute concentrations, total times for spectral acquisition varied from approximately 50 min to 15.6 h, with typical times of about 4 h. ^{13}C NMR spectra were referenced to formate (173.916 ppm) from the BMRB [8].

Chemical shift calculations

All molecules were optimized at the B3LYP/6-31G* level [9]. Only the most stable tautomer for molecules undergoing tautomerizaion was calculated. NMR chemical shieldings were then calculated using gauge-including atomic orbitals (GIAO) [10] at the B3LYP/6-311++G** level, which has been shown to produce accurate NMR chemical shifts [11,12]. To account for solvent effects, we applied implicit solvent model IEFPCM [13]. All calculations were carried out using the Gaussian 09 package (Gaussian Inc.). Experimental and computational NMR data were compared in MATLAB using an interactive script we wrote that overlays computed and experimental spectra. Assignments of unstable compounds in Supplementary Table S1 with no database chemical shift values were made by comparing the computed chemical shifts with patterns of ^{13}C-^{13}C couplings, which clearly show ^{13}C nuclei with 0, 1, or 2 neighboring ^{13}C nuclei.

Acknowledgement

The 1.5-mm HTS NMR probe [56] was built with support from the NIH (R01EB009772-01 to A.S.E.) in partnership with the National High Magnetic Field Laboratory and Agilent Technologies.

References

1. Karimova, G.J., Pidoux, J., Ullmann, A. and Ladant, D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc. Natl. Acad. Sci. U.S.A. **95**, 5752-5756
2. Ma, X. and Cline, K. (2013) Mapping the signal peptide binding and oligomer contact sites of the core subunit of the pea twin arginine protein translocase. Plant Cell **25,** 999-1015
3. Heckman, K.L. and Pease, L.R. (2007) Gene splicing and mutagenesis by PCR-driven overlap extension. Nat. Protoc. **2**, 924-932
4. Cline, K. (1986) Import of proteins into chloroplasts. Membrane integration of a thylakoid precursor protein reconstituted in chloroplast lysates. J. Biol. Chem. **261**, 14804-14810
5. Frelin, O., Agrimi, G., Laera, V.L., Castegna, A., Richardson, L.G., Mullen, R.T., Lerma-Ortiz, C., Palmieri, F. and Hanson, A.D. (2012) Identification of mitochondrial thiamin diphosphate carriers from Arabidopsis and maize. Funct. Integr. Genomics **12**, 317-326
6. Schmittgen, T.D. and Livak, K.J. (2008) Analyzing real-time PCR data by the comparative CT method. Nat. Protoc. **3**, 1101-1108
7. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J. and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR **6**, 277-293

8. Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C.F., Tolmie, D.E., Kent Wenger, R., Yao, H. and Markley, J.L. (2008) BioMagResBank. Nucleic Acids Res. **36**, D402-D408
9. Becke, A.D. (1993) Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Physics **98**, 5648-5652
10. Wolinski, K., Hinton, J.F. and Pulay, P. (1990) Efficient implementation of the gauge-independent atomic orbital method for NMR chemical-shift calculations. J. Am. Chem. Soc. **112**, 8251-8260
11. Helgaker, T., Jaszunski, M. and Ruud, K. (1999) Ab initio methods for the calculation of NMR shielding and indirect spin-spin coupling constants. Chem. Rev. **99**, 293-352
12. Wang, B., Fleischer, U., Hinton, J.F. and Pulay, P. (2001) Accurate prediction of proton chemical shifts. I. Substituted aromatic hydrocarbons. J. Comput. Chem. **22**, 1887-1895
13. Mennucci, B. and Tomasi, J. (1997) Continuum solvation models: a new approach to the problem of solute's charge distribution and cavity boundaries. J. Chem. Physics **106**, 5151-5158