Supporting Information

Crystal Structure of the Ergothioneine Sulfoxide Synthase from *Candidatus Chloracidobacterium* thermophilum and Structure-Guided Engineering to Modulate Its Substrate Selectivity

Nathchar Naowarojna,[†,⊥] Seema Irani,[‡,⊥] Weiyao Hu,[†,§,⊥] Ronghai Cheng,[†] Li Zhang,[†] Xinhao Li,[§] Jiesheng Chen,[§,*] Yan Jessie Zhang,[‡,*] and Pinghua Liu[†,*]

[†]Department of Chemistry, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts, 02215, United States

[‡]Department of Molecular Biosciences and Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas, 78712, United States

[§]School of Chemistry and Chemical Engineering, Shanghai JiaoTong University, Shanghai, 200240, China

*E-mail: pinghua@bu.edu, jzhang@cm.utexas.edu, and chemcj@sjtu.edu.cn

Table of Contents

Experimental Procedures **3**

Protein similarity network analysis 3

Protein overexpression and purification 3

$EgtB_{Cth}$ iron content determination using atomic emission spectroscopy 3

Activity assay and product purification 3

Steady-state kinetic analysis 4

Crystallization and structure determination 4

$EgtB_{Cth}$ substrate complex 4

Structure refinement 4

Modeling of $EgtB_{Cth}$ bound to both hercynine and γ-Glu-Cys 5

Protein engineering using Rosetta Enzyme Design 5

Figures **6**

Figure S1. Characterization of $EgtB_{Cth}$ by SDS-PAGE analysis and iron content determination 6

Figure S2. Characterization of hercynyl-γ-Glu-Cys sulfoxide from $EgtB_{Cth}$ wild-type reaction 7

Figure S3. Characterization of hercynyl-Cys sulfoxide from $EgtB_{Cth}$ wild-type reaction 8

Figure S4. Kinetic analysis of $EgtB_{Cth}$ wild-type with hercynine and γ-Glu-Cys 9

Figure S5. Kinetic analysis of $EgtB_{Cth}$ wild-type with hercynine and L-Cys 9

Figure S6. ^{1}H-NMR spectrum of $EgtB_{Cth}$ wild-type reaction using hercynine and γ-Glu-Cys as the substrates for coupling product quantification 10

Figure S7. ^{1}H-NMR spectrum of $EgtB_{Cth}$ wild-type reaction using hercynine and L-Cys as the substrates for coupling product quantification 11

Figure S8. Comparing the structure of $EgtB_{Cth}$ and $EgtB_{Cth}$ •hercynine. 12

Figure S9. Structural comparison among $EgtB_{Cth}$, $EgtB_{Mth}$, and Egt1 12

Figure S10. Kinetic analysis of $EgtB_{Cth}$ A420Y mutant with hercynine and γ-Glu-Cys as the substrates 13

Figure S11. ^{1}H-NMR analysis of $EgtB_{Cth}$ A420Y reactions 13

Figure S12. Kinetic analysis of $EgtB_{Cth}$ A420Y mutant with hercynine and L-Cys as the substrates 14

Figure S13. ^{1}H-NMR spectrum of $EgtB_{Cth}$ A420Y reaction using hercynine and γ-Glu-Cys as the substrates for coupling product quantification 14

Figure S14. ^{1}H-NMR spectrum of $EgtB_{Cth}$ A420Y reaction using hercynine and L-Cys as the substrates for coupling product quantification 15

Figure S15. ^{1}H-NMR analysis of $EgtB_{Cth}$ D52L reactions 16

Figure S16. Kinetic analysis of $EgtB_{Cth}$ D52L mutant with hercynine and γ-Glu-Cys as the substrates 16

Figure S17. Kinetic analysis of $EgtB_{Cth}$ D52L mutant with hercynine and L-Cys as the substrates 17

Figure S18. ^{1}H-NMR spectrum of $EgtB_{Cth}$ D52L reaction using hercynine and γ–Glu-Cys as the substrates for coupling product quantification 17

Figure S19. ^{1}H-NMR spectrum of $EgtB_{Cth}$ D52L reaction using hercynine and L-Cys as the substrates for coupling product quantification 18

Figure S20. ^{1}H-NMR analysis of $EgtB_{Cth}$ D52L/A420Y reactions 19

Figure S21. Kinetic analysis of $EgtB_{Cth}$ D52L/A420Y mutant with hercynine and γ-Glu-Cys as the substrates 19

Figure S22. ^{1}H-NMR spectrum of EgtB$_{Cth}$ D52L/A420Y reaction using hercynine and γ–Glu-Cys as the substrates for coupling product quantification 20

Figure S23. Kinetic analysis of EgtB$_{Cth}$ D52L/A420Y mutant with hercynine and L-Cys as the substrates 20

Figure S24. ^{1}H-NMR spectrum of EgtB$_{Cth}$ D52L/A420Y reaction using hercynine and L-Cys as the substrates for coupling product quantification 21

Tables **21**

Table S1. List of EgtB$_{Cth}$ homologs retrieved from sequence similarity network analysis 21

Table S2. Data collection and statistic of EgtB$_{Cth}$ structure 22

Table S3. List of top 20 EgtB$_{Cth}$ mutants predicted from Rosetta Enzyme Design 23

References 23

Experimental Procedures

Reagents were purchased from Sigma-Aldrich and Fisher Scientific unless otherwise indicated. Hercynine and γ-Glu-Cys were synthesized following reported procedures.[1] Proton nuclear magnetic resonance (^{1}H NMR) spectra were recorded using Agilent 500 (500 MHz VNMRS). Mass spectrometric analysis was performed using LC-MS on an LTQ-FT-ICR mass spectrometer (Thermo Scientific).

Protein similarity network analysis

In order to systematically analyze the sulfoxide synthase candidates, 21475 sequences retrieved from pfam database with either DinB_2 domain (Pfam ID: PF12867) or FGE sulfatase domain (Pfam ID: PF03781) were retrieved from Pfam protein family database.[2] Among these, 3000 sequences were randomly selected for protein similarity network analysis. The protein similarity network was constructed by all-to-all blast of these 3000 sequences. The cut-off of the E value was 10^{-50}. The protein similarity network was visualized by Cytoscape.

Protein overexpression and purification

The gene Cabther_A1318 (EgtB$_{Cth}$) was codon-optimized for *Escherichia coli* overexpression system by Genscript. The gene was then sub-cloned into pASK-IBA3plus expression vector using EcoRI and XhoI restriction sites. The plasmid was then transformed into BL21(DE3) and one colony was inoculated in LB media supplemented with 100 μg/mL of ampicillin at 37 °C overnight. Then, 10 mL of this overnight culture was then transferred to 1 L LB medium supplemented with 100 μg/mL of ampicillin and 0.1 mM ferrous ammonium sulfate. Cells were culture at 37 °C until OD_{600} reached 0.8, and protein overexpression was induced with the addition of anhydrotetracycline to 500 μg/L final concentration. Cells were cultured for 14 hrs at 25 °C before harvest by centrifugation.

The following procedure was carried out under anaerobic condition at 4 °C unless otherwise specified. 25 g cell was resuspended in 100 mL of lysis buffer (100 mM Tris-HCl, 50 mM NaCl buffer pH 8.0). Then, lysozyme was added to 1 mg/mL final concentration and the mixture was incubated on ice for 30 mins. Ferrous ammonium sulfate and sodium ascorbate was added to 0.1 mM, and 0.2 mM final concentration, respectively. Cells were lysed by sonication (18 cycles of 8 sec on, 59 sec off) and cell debris was removed by centrifugation at 20,000 g for 45 mins. The supernatant was then incubated with 15 mL Streptactin resin (IBA) on ice for 30 mins with gentle agitation. The column was then washed with lysis buffer until $OD_{280} < 0.05$. Subsequently, the desired protein was eluted using 2.5 mM D-desthiobiotin in lysis buffer. The elution fractions were combined and concentrated using ultrafiltrator (Millipore). The protein was flash frozen and stored in -80 °C. For crystallographic studies, protein was further purified using size-exclusion chromatography to near homogeneity.

EgtB$_{Cth}$ selenomethionine derivative was obtained using minimal medium supplemented with amino acid mixture for cell culture. Protein overexpression was induced following the protocol for that of wild-type.

EgtB$_{Cth}$ mutants were generated using Q5 Site-directed mutagenesis kit following recommended protocol (New England Biolabs). The protein was overexpressed and purified following the procedure for wild-type EgtB$_{Cth}$.

EgtB$_{Cth}$ iron content determination using atomic emission spectroscopy

The iron content of EgtB$_{Cth}$ after purification was carried out using Agilent 4200 MP-AES to measure the intensity. To release iron cofactor into solution, EgtB$_{Cth}$ protein was diluted using 1% HNO_3 to 1.0 mg/mL final concentration. Then, precipitated protein was removed by centrifugation. The sample was prepared in triplicated. The iron standard curve was constructed by measuring intensity at 259.940 nm. The average intensity for the sample was 9639 ± 283 and the iron concentration of the sample derived from the standard curve was 17.4 ± 0.5 µM. This result suggested that there was 0.92 iron per one EgtB$_{Cth}$ monomer.

Activity assay and product purification

A typical 1-mL reaction containing 1 mM hercynine, 1 mM L-Cys (or γ–Glu-Cys), 1 mM TCEP (Tris(2-carboxyethyl)phosphine hydrochloride), 0.2 mM ascorbate, and 10 µM EgtB$_{Cth}$ in 50 mM KPi pH 8.0, was set up at 28 °C for 2 hrs. The protein was then removed using centrifugal filter column (Amicon). The reaction mixture was lyophilized and analyzed by ^{1}H NMR.

For product characterization, a reaction similar to the above condition was set up, however, the reaction volume was scaled to 10 mL. The product was isolated using Dowex 50WX8-100 ion-exchange resin following reported procedure.[3] Then, the product was characterized by ^{1}H NMR and high-resolution mass spectrometry.

Hercynyl-γ-Glu-Cys sulfoxide **3**: ^{1}H NMR δ 2.06-2.15 (m, 2H), δ 2.43 (m, 2H), δ 3.21 (s, 9H), δ 3.31 (dd, J=13.8, 11.7 Hz 1H), δ 3.40 (dd, J=14.0, 3.4 Hz 1H), δ 3.62 (dd, J=13.6, 9.2 Hz, 1H), δ 3.85 (dd, J=14.0, 4.3 Hz, 1H), δ 3.98 (t, J=6.6 Hz, 1H), δ 4.10 (dd, J=11.9, 4.0 Hz, 1H), δ 4.74 (dd, J=9.5, 4.3 Hz, 1H), δ 7.34 (s, 1H). HRMS (ESI): Calculated value for compound **4** as $[M-H]^-$ (negative mode) form was *m/z* 460.1508 and found *m/z* 460.1505.

Hercynyl-cysteine sulfoxide **4**: ^{1}H NMR (500 MHz, D_2O): δ 3.13 (m, 11 H), δ 3.61 (dd, J = 14.4, 8.2 Hz, 1H), δ 3.69 (dd, J = 14.4, 4.4 Hz, 1H), δ 3.78 (dd, J = 11.7, 3.9 Hz, 1H), δ 3.97 (dd, J = 8.2, 4.4 Hz, 1H), δ 7.15 (s, 1H). HRMS (ESI): Calculated value for compound **5** as $[M-H]^-$ (negative mode) form was *m/z* 331.1082 and found *m/z* 331.1078.

Steady-state kinetic analysis

The kinetic parameter of EgtB$_{Cth}$ was monitored using the NeoFox oxygen probe (Ocean Optics). For hercynine concentration dependent, a 1-mL reaction contained 2 mM L-Cys, 2 mM TCEP, 0.2 mM sodium ascorbate, 1.5 µM EgtB$_{Cth}$, and varied concentration of hercynine (0.03 – 4 mM) in 50 mM KPi pH 8.0. A similar reaction was setup for L-Cys concentration dependent, except hercynine concentration was kept at 2 mM and L-Cys concentration was varied (0.05 – 4 mM). EgtB$_{Cth}$ exhibited a K_M for hercynine of 87.7 ± 7.6 µM and a K_M for L-Cys of 205 ± 18 µM with a k_{cat} of 26.6 ± 0.7 min^{-1}.

Similar experimental set up was carried out for reaction using hercynine and γ–Glu-Cys pair of substrates, except the enzyme concentration was increased to 2 µM. Using hercynine and γ-Glu-Cys as the substrates, EgtB$_{Cth}$ exhibited a K_M for hercynine of 41.4 ± 3.5 µM and a K_M for γ-Glu-Cys of 5.9 ± 0.9 mM with a k_{cat} of 17.5 ± 0.4 min^{-1}.

For EgtB$_{Cth}$ mutants, the experimental condition for kinetic analysis remained largely unchanged. Using hercynine and L-Cys as the substrates, EgtB$_{Cth}$ A420Y exhibited a K_M for hercynine of 39.2 ± 1.7 µM and a K_M for L-Cys of 28.1 ± 1.6 µM with a k_{cat} of 26.4 ± 0.3 min^{-1}. Using hercynine and γ-Glu-Cys as the substrates, EgtB$_{Cth}$ A420Y exhibited a K_M for hercynine of 13.2 ± 1.3 µM and a K_M for γ-Glu-Cys of 3.1 ± 0.3 mM with a k_{cat} of 17.4 ± 0.3 min^{-1}.

For EgtB$_{Cth}$ D52L variant, the enzyme concentration in the assay was adjusted to 2 µM. Using hercynine and L-Cys as the substrates, EgtB$_{Cth}$ D52L exhibited a K_M for hercynine of 13.9 ± 0.8 µM and a K_M for L-Cys of 47.3 ± 2.4 µM with a k_{cat} of 26.9 ± 0.3 min^{-1}. Using hercynine and γ-Glu-Cys as the substrates, EgtB$_{Cth}$ D52L mutant exhibited a K_M for hercynine of 19.6 ± 1.4 µM and a K_M for γ-Glu-Cys of 3.9 ± 0.3 mM with a k_{cat} of 12.9 ± 0.2 min^{-1}.

For EgtB$_{Cth}$ D52L/A420Y variant, the enzyme concentration in the assay was 1.5 µM. Using hercynine and L-Cys as the substrates, EgtB$_{Cth}$ D52L/A420Y exhibited a K_M for hercynine of 23.3 ± 1.2 µM and a K_M for L-Cys of 48.7 ± 3.5 µM with a k_{cat} of 32.7 ± 0.3 min^{-1}. Using hercynine and γ-Glu-Cys as the substrates, EgtB$_{Cth}$ D52L/A420Y mutant exhibited a K_M for hercynine of 15.1 ± 1.3 µM and a K_M for γ-Glu-Cys of 4.2 ± 0.5 mM with a k_{cat} of 18.6 ± 0.4 min^{-1}.

Crystallization, structure determination and refinement

Crystallization screening trays with EgtB$_{Cth}$ at a concentration ~20mg/ml were set up using the Phenix crystallization robotic system (Art Robinson Instruments). Several hits were identified and optimized using vapor diffusion in 24-well trays, eventually leading to cube like crystals in a condition containing 0.1-0.2M sodium acetate, 0.1M Tris-acetate pH 6.0-6.5 and 3-8% PEG 4000 at 4 ºC, with 1:2 ratio of protein to mother liqour.

To obtain the ligand incorporated structures, both co-crystallization and soaking were used with identical results. For co-crystallization, 10mM ligand hercynine was incubated with EgtB$_{Cth}$ at ~20 mg/ml for 1 hour on ice before setup and crystallized under

identical conditions like apo protein. For soaking, pre-formed crystals were soaked in mother liquor containing 10mM hercynine for 12 hrs at 4 ºC. The crystals were subsequently cryo-protected in a cryogenic solution containing the same mother liquor with 30% glycerol, before vitrification in liquid nitrogen.

Selenomethionine-incorporated EgtB$_{Cth}$ protein was crystalized under similar conditions as wild type enzyme and phasing was performed using single wavelength anomalous dispersion (SAD) data and Phenix Autosol.[4]

Diffraction data were collected at the beamlines of Advance Light Source Berkley ALS BL 5.0.3 and Advanced Photon Source APS BL 23-ID-D. Diffraction data sets were integrated and scaled using HKL2000.[5] Several rounds of structure building in Coot[6] and refinement using Phenix Refine were performed to obtain a final structure for EgtB$_{Cth}$ with a R_{work}/R_{free} of 18.5%/22.6% . The structure shows good geometry with 97.9% of the phi/psi angles being Ramachandran favoured, and a MolProbity score of 1.14 (100th percentile).[7] For complex structure with hercynine, jLigand[8] was used to model the ligand in the positive density. The final R_{work}/R_{free} of the EgtB$_{Cth}$•hercynine binary complex structure is 18.8%/22.4%, the structure shows good geometry with a MolProbity score of 1.19 (100th percentile) and 97.6% of Ramachandran favoured phi/psi angles. The statistics for data collection and refinement is summarized in Table S2.

Modeling of EgtB$_{Cth}$ bound to both hercynine and γ–Glu-Cys

The model for EgtB$_{Cth}$ bound to both substrates was obtained by superimposing the complex structure of EgtB$_{Cth}$ bound to hercynine and tertiary structure of EgtB$_{Mth}$ bound to dimethyl histidine and γ–Glu-Cys (PDB ID: 4X8D). Residues that coordinated the cysteine portion of the γ–Glu-Cys substrate remained conserved in both structures, whereas the residues interacting with the glutamyl end of the substrate appeared to be flipped. The substrate position from the previously obtained pdb ID 4X8D was therefore manually adjusted to form the corresponding interactions in the EgtB$_{Cth}$ structure.

Protein engineering using Rosetta Enzyme Design

The EgtB$_{Cth}$ structure with hercynine and L-Cys complex was used as a template for mutagenesis. Iterative mutagenesis of three active site residues (D52, F416, and A420) of EgtB$_{Cth}$ was performed with a combination of Pymol mutagenesis wizard and custom Python scripts. In order to find the valid rotamer for each specific mutation, the “minimizer” algorithm from Rosetta 3.0 was used to globally minimize the energy of each rotamers. The specific parameters for “minimizer: are “-run:min_type lbfgs_armijo_nonmonotone, -run:min_tolerance 0.005”. The model with the smallest global energy was used as the most valid rotamer for each of specific mutants, which favours the binding of L-Cys substrate. Global energies for all mutants were calculated and ranked by Rosetta 3.0 scoring function “score_jd2”. Default parameters are used for “score_jd2”. The variant structures which have lower score were then inspected manually in conjunction with structural results to identify variants to be validated by biochemical studies.

Figures

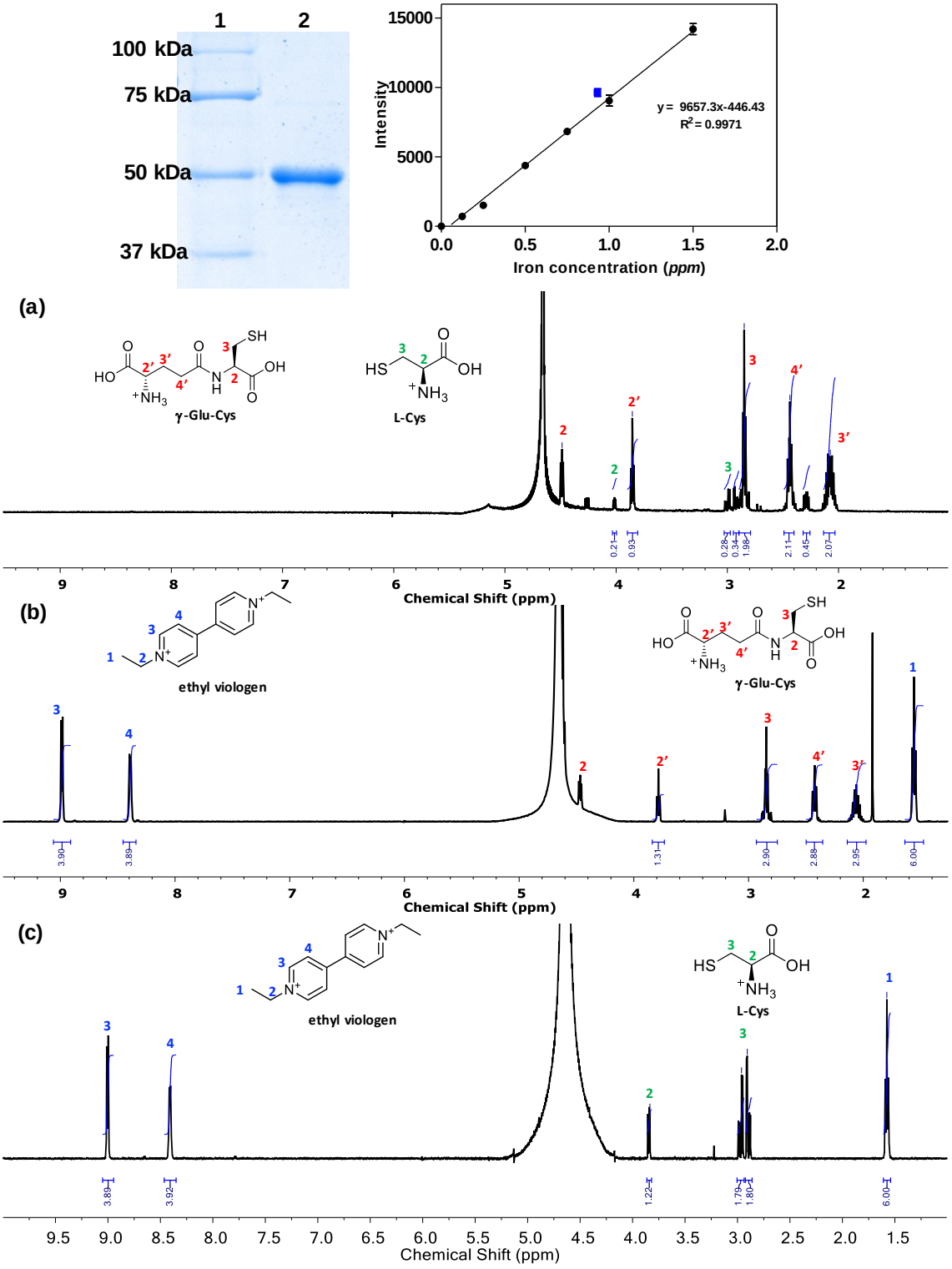


Figure S1. Characterization of purified EgtB$_{Cth}$ and the substrates for the reaction. **Top Left**: SDS-PAGE analysis of purified EgtB$_{Cth}$. Lane 1 is protein marker and lane 2 is anaerobically purified EgtB$_{Cth}$. **Top Right**: The iron standard curve. The average intensity for the sample was 9639 ± 283 (blue square) and the iron **Bottom**: (a) ^{1}H NMR spectrum of γ-Glu-Cys substrate contaminated with L-Cys. (b) ^{1}H NMR spectrum of γ-Glu-Cys substrate used in biochemical characterization (500 MHz, D_2O): δ 2.01-2.11 (m, 2H), δ 2.35-2.50 (m, 2H), δ 2.93-2.75 (m, 2H), δ 3.79 (t, J=6.4 Hz, 1H), δ 4.47 (t, 6.5 Hz, 1H). (c) ^{1}H NMR spectrum of L-Cys substrate used for biochemical characterization (500 MHz, D_2O): δ 2.87 (dd, J=15.0, 4.1 Hz, 2H), δ 2.95 (dd, J= 15.0, 5.7 Hz, 2H), δ 3.83 (dd, J=5.7, 4.1 Hz, 2H).

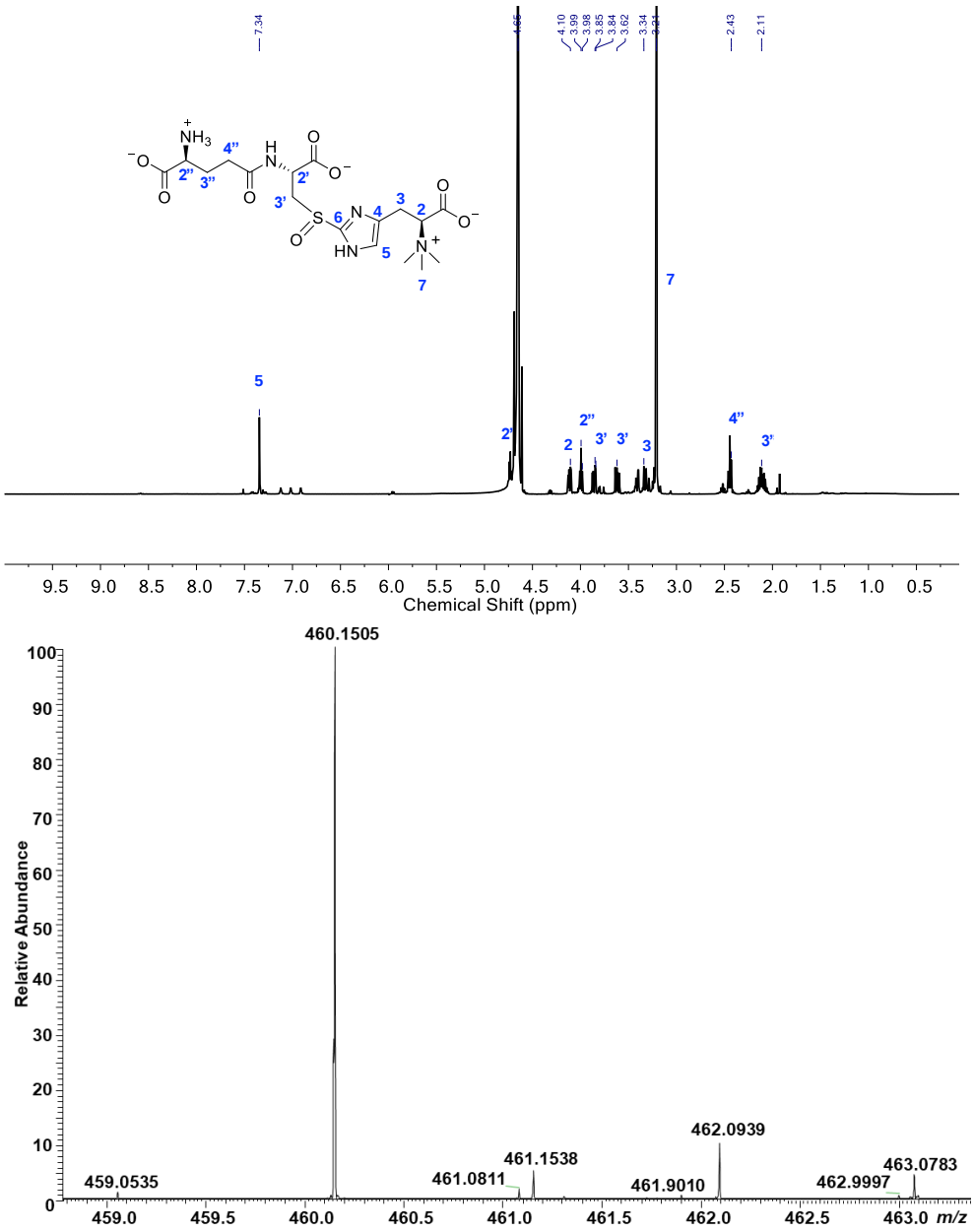


Figure S2. Characterization of coupling product from EgtB$_{Cth}$ wild-type reaction using hercynine and γ-Glu-Cys as the substrates. **Top**: ^{1}H NMR spectrum (500 MHz, D_2O): δ 2.06-2.15 (m, 2H), δ 2.43 (m, 2H), δ 3.21 (s, 9H), δ 3.31 (dd, J=13.8, 11.7 Hz, 1H), δ 3.40 (dd, J=14.0, 3.4 Hz, 1H), δ 3.62 (dd, J=13.6, 9.2 Hz, 1H), δ 3.85 (dd, J=14.0, 4.3 Hz, 1H), δ 3.98 (t, J=6.6 Hz, 1H), δ 4.10 (dd, J=11.9, 4.0 Hz, 1H), δ 4.74 (dd, J=9.5, 4.3 Hz, 1H), δ 7.34 (s, 1H). **Bottom**: High resolution mass spectrum of sulfoxide **3**. Calculated value for compound **3** as $[M-H]^-$ (negative mode) form was *m/z* 460.1508 and found *m/z* 460.1505.

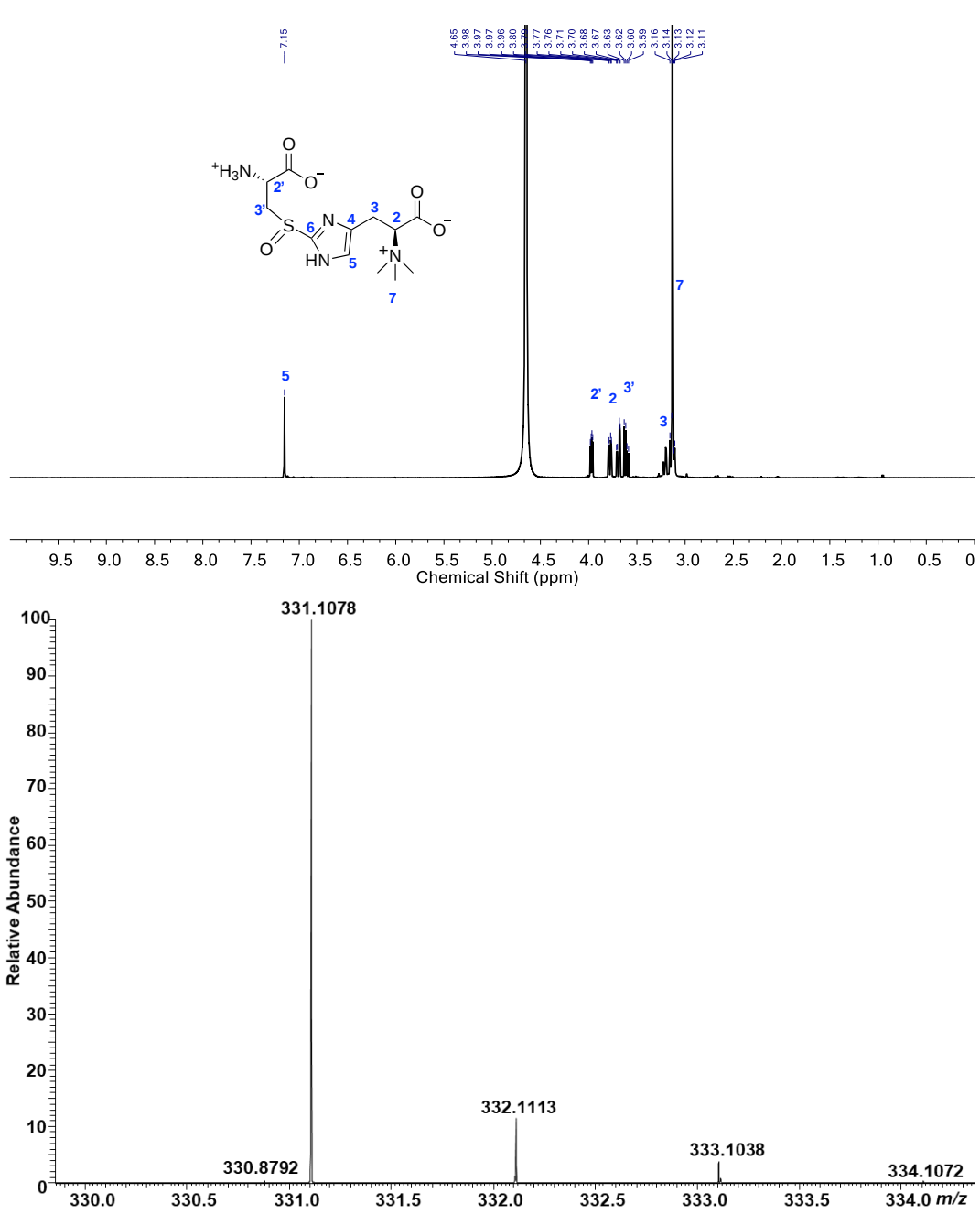
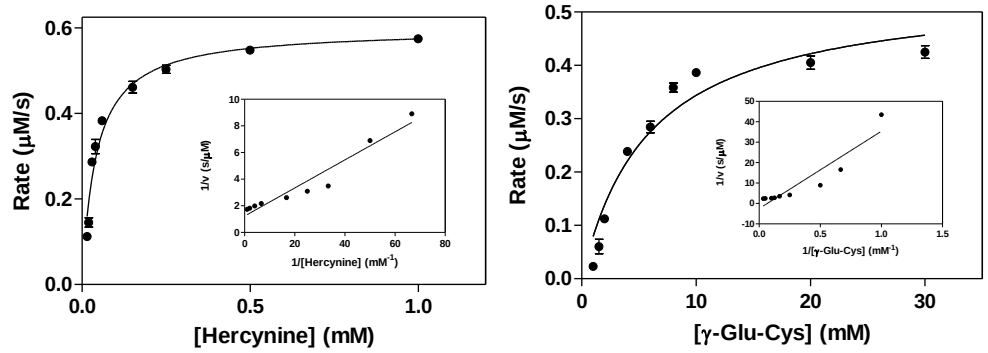


Figure S3. Characterization of of coupling product from EgtB$_{Cth}$ wild-type reaction using hercynine and L-Cys as the substrates. **Top:** ^{1}H NMR spectrum (500 MHz, D_2O): δ 3.13 (m, 11 H), δ 3.61 (dd, J = 14.4, 8.2 Hz, 1H), δ 3.69 (dd, J = 14.4, 4.4 Hz, 1H), δ 3.78 (dd, J = 11.7, 3.9 Hz, 1H), δ 3.97 (dd, J = 8.2, 4.4 Hz, 1H), δ 7.15 (s, 1H). **Bottom**: High resolution mass spectrum of the hercynyl-cysteine coupling product. Calculated value for compound **4** as $[M-H]^-$ (negative mode) form was *m/z* 331.1082 and found *m/z* 331.1078.



5

Figure S4. Kinetic analysis of EgtB*Cth* wild-type with hercynine and γ-Glu-Cys. For hercynine concentration dependent, a 1-mL reaction contained 15 mM γ-Glu-Cys, 2 mM TCEP, 0.2 mM sodium ascorbate, 2 μM EgtB*Cth*, and varied concentration of hercynine (0.015 – 1.0 mM) in 50 mM KPi pH 8.0. A similar reaction was setup for γ-Glu-Cys concentration dependent studies, except hercynine concentration was kept at 1 mM and γ-Glu-Cys concentration was varied (1 – 30 mM). EgtB*Cth* exhibited a K_M for hercynine of 41.4 ± 3.5 μM and a K_M for γ-Glu-Cys of 5.9 ± 0.9 mM with a k_{cat} of 17.5 ± 0.4 min^{-1}.

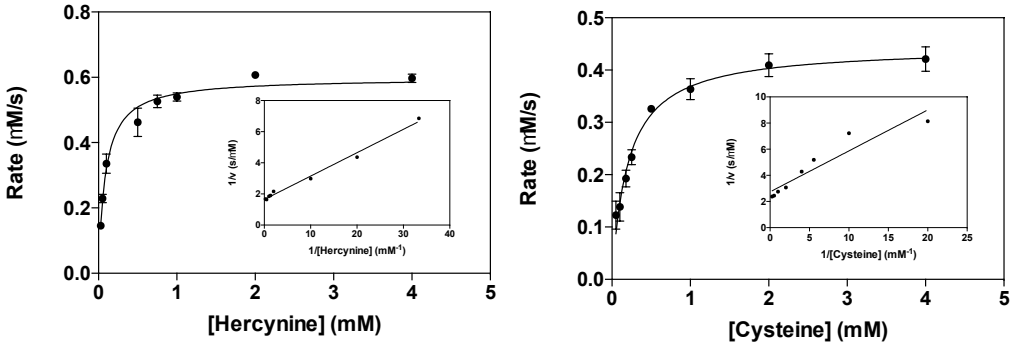


Figure S5. Kinetic analysis of EgtB*Cth* wild-type with hercynine and L-Cys. For hercynine concentration dependent, a 1-mL reaction contained 2 mM L-Cys, 2 mM TCEP, 0.2 mM sodium ascorbate, 1.5 μM EgtB*Cth*, and varied concentration of hercynine (0.03 – 4 mM) in 50 mM KPi pH 8.0. A similar reaction was setup for L-Cys concentration dependent, except hercynine concentration was kept at 2 mM and L-Cys concentration was varied (0.05 – 4 mM). EgtB*Cth* exhibited a K_M for hercynine of 87.7 ± 7.6 μM and a K_M for L-Cys of 205 ± 18 μM with a k_{cat} of 26.6 ± 0.7 min^{-1}.

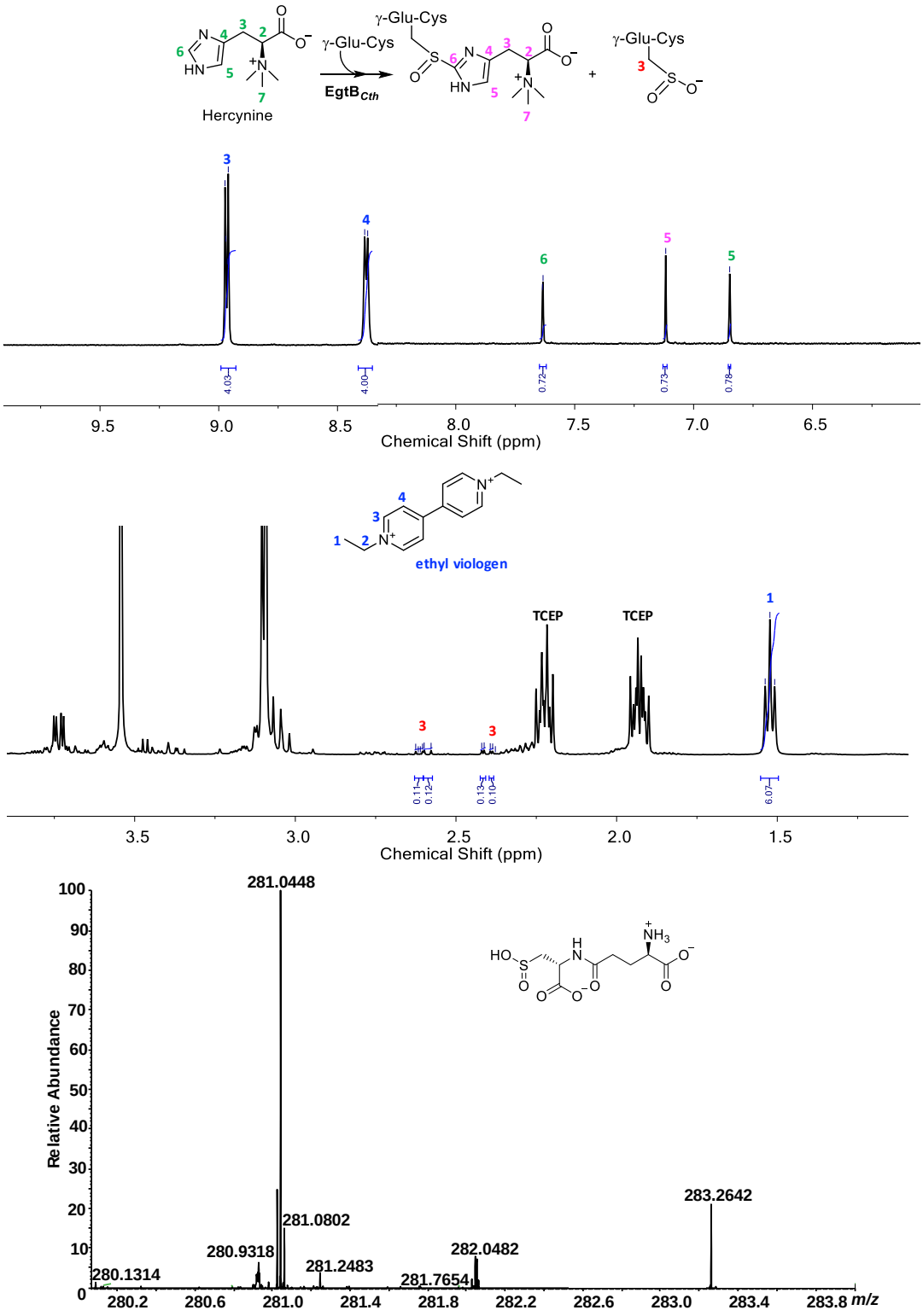


Figure S6. ^{1}H-NMR spectrum of EgtB$_{Cth}$ wild-type reaction using hercynine and γ-Glu-Cys as the substrates for coupling product quantification. **Top**: Expanded regions between 6.5 ppm - 9.5 ppm. **Middle**: Expanded regions between 1.0 – 4.0 ppm regions. The ratio between products from sulfoxide synthase and cysteine dioxygenase activities was quantified using ethyl viologen as an internal standard. The ratio between the proton of coupling product (labelled 5 in pink) and aromatic proton of ethyl viologen (labelled 4 in blue), and the ratio between β-hydrogen of γ-Glu-Cys sulfinic acid (labelled 3 in red) and the ethyl group of ethyl viologen (labelled 1 in blue) were used to quantify the ratio of the two products. In EgtB$_{Cth}$ wild-type reaction, the ratio between hercynyl-γ-Glu-Cys sulfoxide and γ-Glu-Cys sulfinic acid product was 3.17: 1. This ratio suggests that there was 75 % of coupling product in the reaction mediated by EgtB$_{Cth}$ wild-type. **Bottom**: High resolution mass spectrum of γ-Glu-Cys sulfinic acid. Calculated value for this compound as $[M-H]^-$ (negative mode) form was *m/z* 281.0449 and found *m/z* 281.0448.

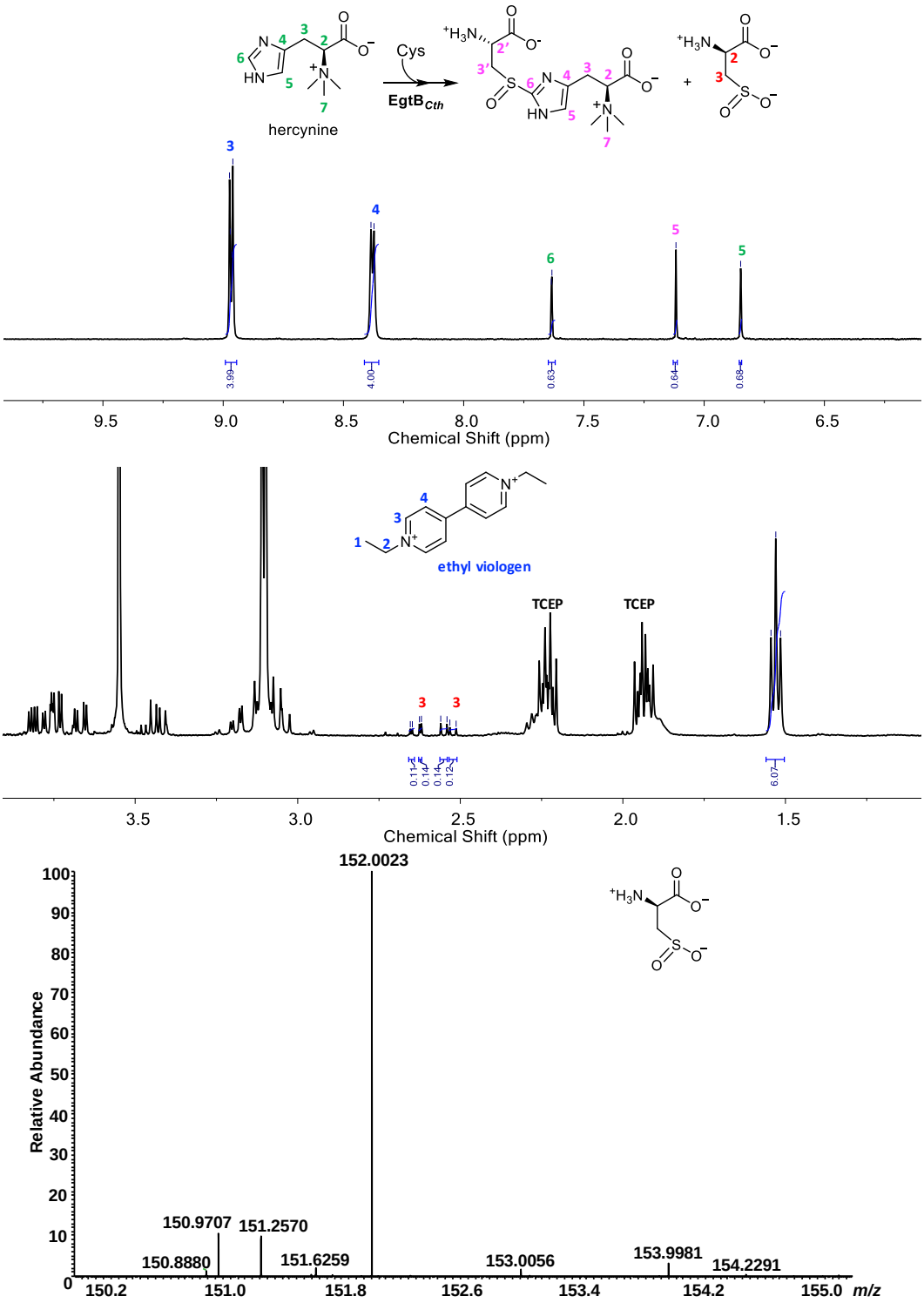


Figure S7. ^{1}H-NMR spectrum of EgtB$_{Cth}$ wild-type reaction using hercynine and L-Cys as the substrates for coupling product quantification. **Top**: Expanded regions between 6.5 ppm - 9.5 ppm. **Middle**: Expanded regions between 1.0 – 4.0 ppm regions. The ratio between products from sulfoxide synthase and cysteine dioxygenase activities was quantified using ethyl viologen as an internal standard. The ratio between the proton of coupling product (labelled 5 in pink) and aromatic proton of ethyl viologen (labelled 4 in blue), and the ratio between β-hydrogen of cysteine sulfinic acid (labelled 3 in red) and the ethyl group of ethyl viologen (labelled 1 in blue) were used to quantify the ratio of the two products. In EgtB$_{Cth}$ wild-type reaction, the ratio between hercynyl-cysteine sulfoxide and cysteine sulfinic acid product was 2.56: 1. This ratio suggests that there was 72 % of coupling product in the reaction mediated by EgtB$_{Cth}$ wild-type. **Bottom**: High resolution mass spectrum of cysteine sulfinic acid. Calculated value for this compound as $[M\text{-}H]^-$ (negative mode) form was *m/z* 152.0023 and found *m/z* 152.0023.

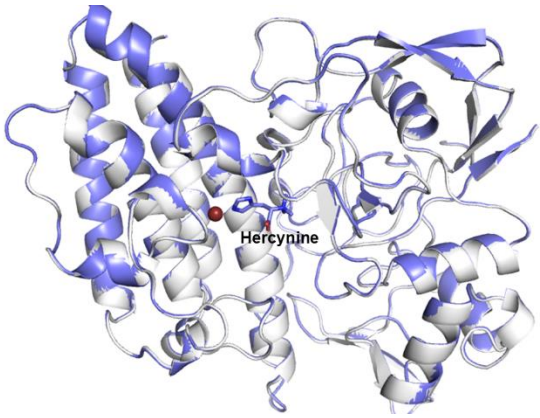


Figure S8. Comparing the structure of EgtB$_{Cth}$ and EgtB$_{Cth}$ •hercynine. Superimposition of the structure of EgtB$_{Cth}$ (shown in white) and EgtB$_{Cth}$•hercynine (shown in blue, with the metal ion represented as a sphere and the hercynine as sticks).

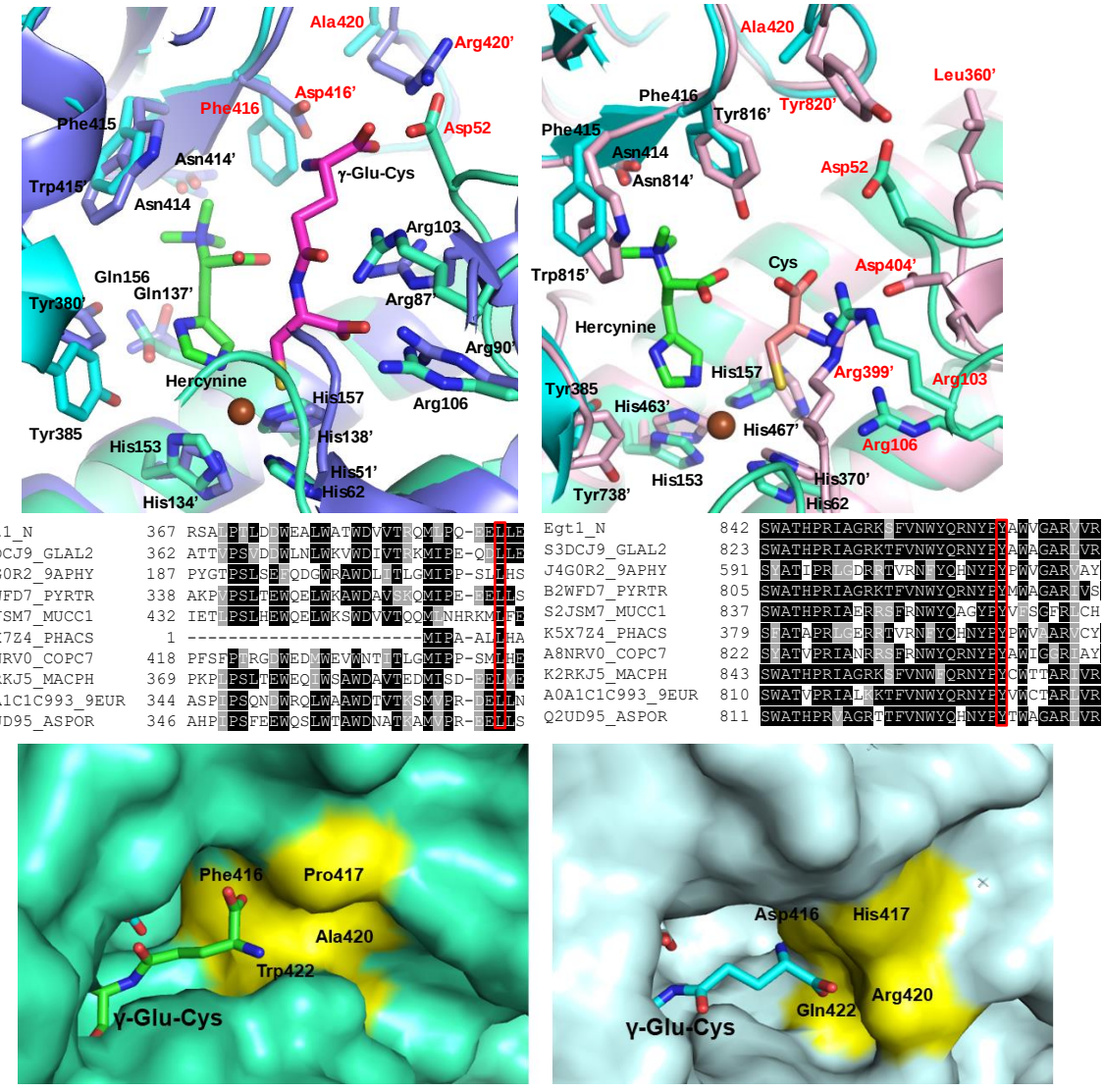


Figure S9. Structural comparison of EgtB$_{Cth}$, EgtB$_{Mth}$, and Egt1. **Top Left:** Structure alignment between EgtB$_{Cth}$ (in light blue/teal) and EgtB$_{Mth}$ (blue, PDB: 4X8D). The non-conserved residues involved in interacting with sulfur donor are labelled in red. The iron center is in brown sphere. **Top Right**: Structural comparison between EgtB$_{Cth}$ (in light blue/teal) and Egt1 (in light pink) shows the non-conserved residue potentially for controlling the sulfur donor selectivity (labelled in red). For all these enzymes, the conserved residues for interacting with hercynine (green stick) are labelled in black. **Middle**: Partial sequence alignment of some randomly-selected Egt1 homologs retrieved from Pfam database show a conserved Leu and Tyr (red box). **Bottom Left:** Surface representation of the active site of EgtB$_{Cth}$ with the modelled γ-Glu-Cys, residues that are located near the glutamyl end shown in yellow. **Bottom Right:** The active site of EgtB$_{Mth}$ complexed with γ-Glu-Cys (PDB ID: 4X8D) with corresponding residues at the active site highlighted in yellow.

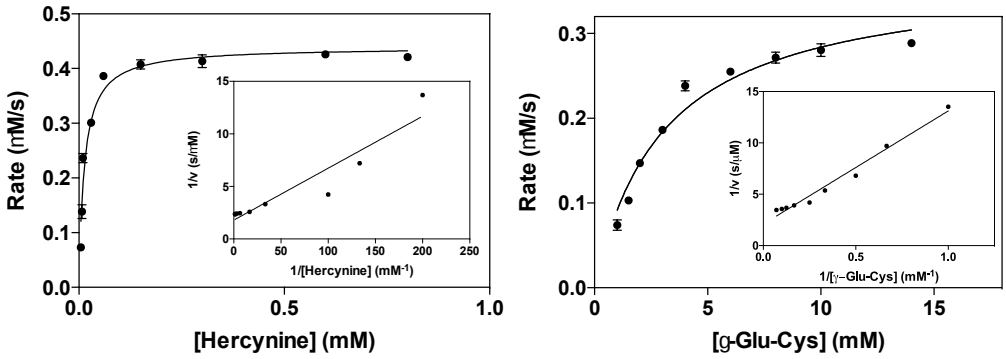


Figure S10. Kinetic analysis of EgtB$_{Cth}$ A420Y mutant with hercynine and γ-Glu-Cys as the substrates. For hercynine concentration dependent, a 1-mL reaction contained 10 mM γ-Glu-Cys, 2 mM TCEP, 0.2 mM sodium ascorbate, 1.5 μM EgtB$_{Cth}$ and varied concentration of hercynine (0.005 – 0.8 mM) in 50 mM KPi pH 8.0. A similar reaction was setup for γ-Glu-Cys concentration dependent, except hercynine concentration was kept at 1 mM and γ-Glu-Cys concentration was varied (1 – 14 mM). EgtB$_{Cth}$ A420Y exhibited a K_M for hercynine of 13.2 ± 1.3 μM and a K_M for γ-Glu-Cys of 3.1 ± 0.3 mM with a k_{cat} of 17.4 ± 0.3 min^{-1}.

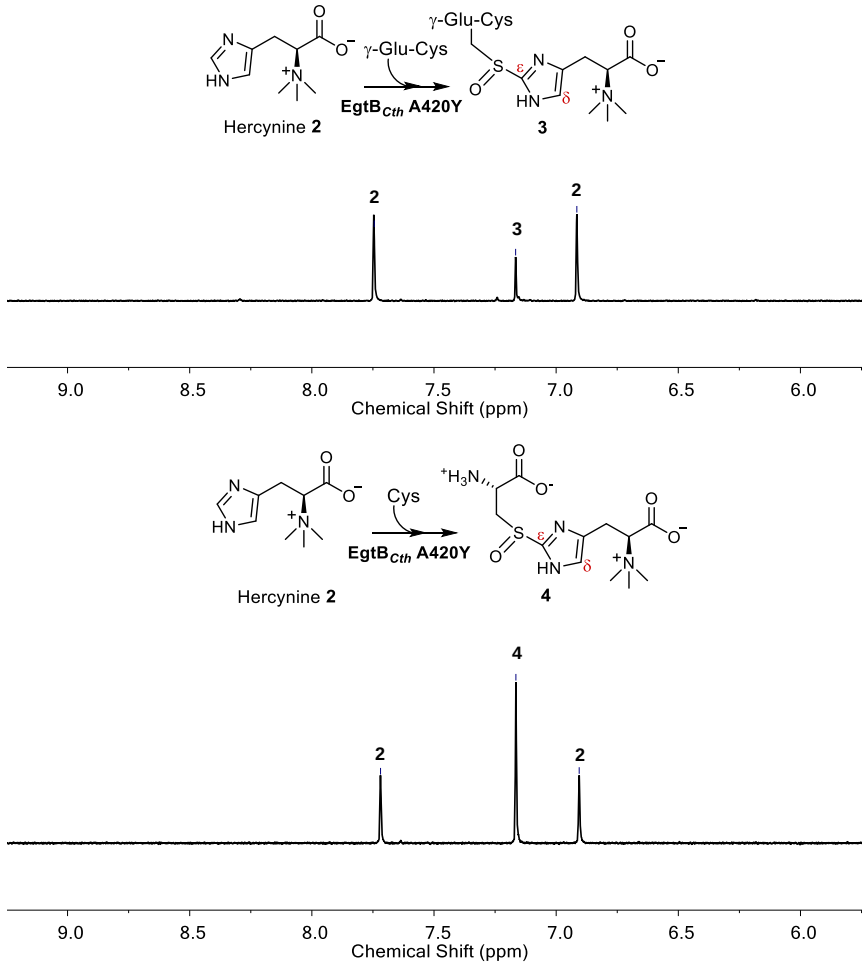


Figure S11. ^{1}H-NMR analysis of EgtB$_{Cth}$ A420Y reactions. **Top**: Reaction of EgtB$_{Cth}$ A420Y using hercynine and γ-Glu-Cys as the substrates shows the hercynyl-γ-Glu-Cys sulfoxide at δ 7.17. **Bottom**: Reaction of EgtB$_{Cth}$ A420Y using hercynine and L-Cys as the substrates shows the coupling product at δ 7.16.

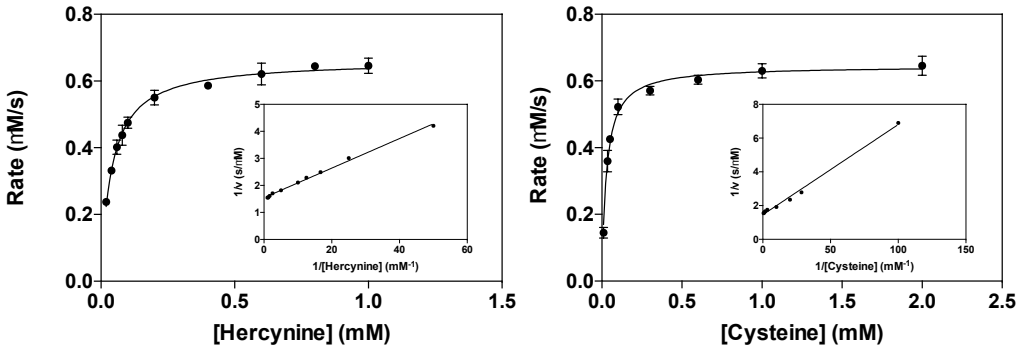


Figure S12. Kinetic analysis of EgtB$_{Cth}$ A420Y mutant with hercynine and L-Cys as the substrates. For hercynine concentration dependent, a 1-mL reaction contained 2 mM L-Cys, 2 mM TCEP, 0.2 mM sodium ascorbate, 1.5 μM EgtB$_{Cth}$, and varied concentration of hercynine (0.03 – 1 mM) in 50 mM KPi pH 8.0. A similar reaction was setup for L-Cys concentration dependent, except hercynine concentration was kept at 2 mM and L-Cys concentration was varied (0.05 – 2 mM). EgtB$_{Cth}$ A420Y exhibited a K_M for hercynine of 39.2 ± 1.6 μM and a K_M for L-Cys of 28.1 ± 1.8 μM with a k_{cat} of 27.6 ± 0.5 min^{-1}.

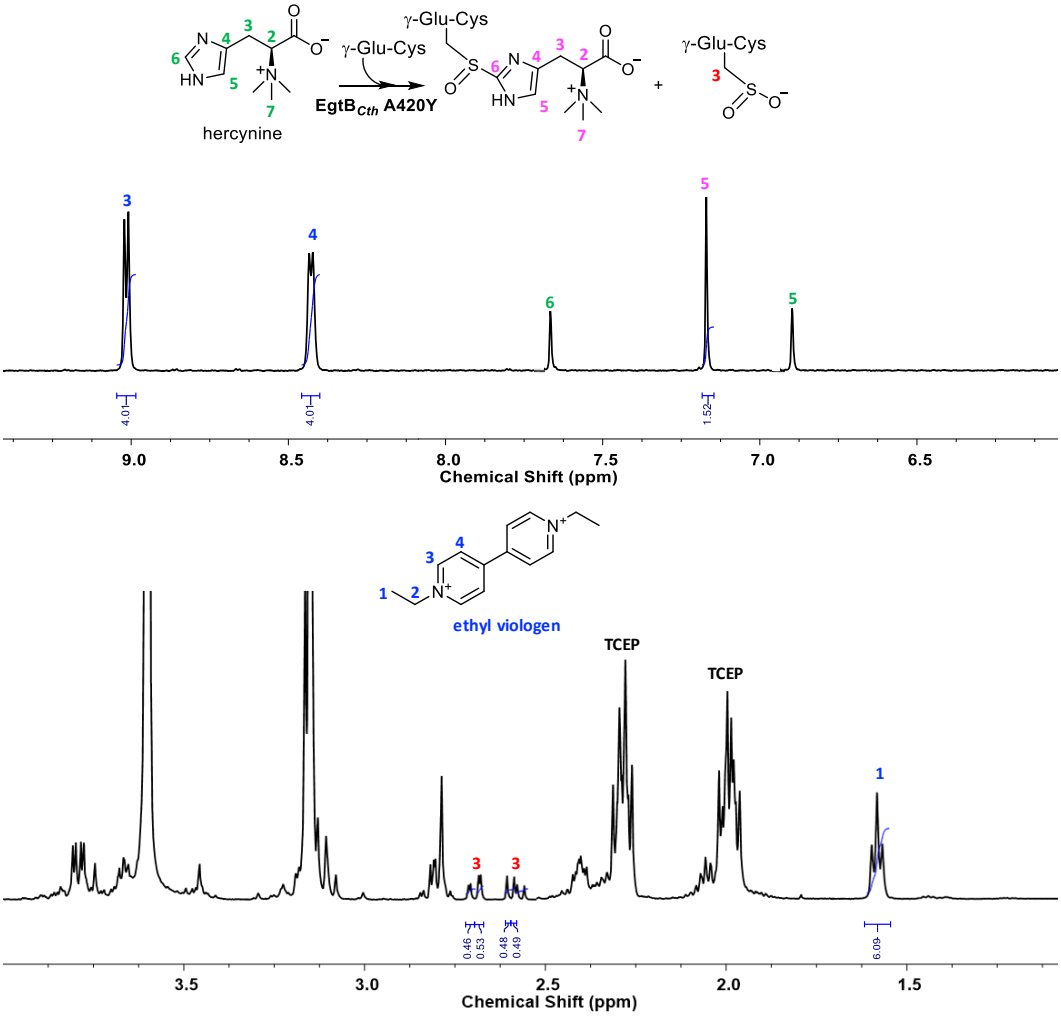


Figure S13. ^{1}H-NMR spectrum of EgtB$_{Cth}$ A420Y reaction using hercynine and γ-Glu-Cys as the substrates for coupling product quantification. **Top**: Expanded regions between 6.5 ppm - 9.5 ppm. **Bottom**: Expanded regions between 1.0 – 4.0 ppm regions. The ratio between products from sulfoxide synthase and cysteine dioxygenase activities was quantified using ethyl viologen as an internal standard. The ratio between the proton of coupling product (labelled 5 in pink) and aromatic proton of ethyl viologen (labelled 4 in blue), and the ratio between β-hydrogen of γ-Glu-Cys sulfinic acid (labelled 3 in red) and the ethyl group of ethyl viologen (labelled 1 in blue) were used to quantify the ratio of the two products. In EgtB$_{Cth}$ A420Y reaction, the ratio between hercynyl-γ-Glu-Cys sulfoxide and γ-Glu-Cys sulfinic acid product was 1.54: 1. This ratio suggests that there was 61 % of coupling product in the reaction mediated by EgtB$_{Cth}$ A420Y variant.

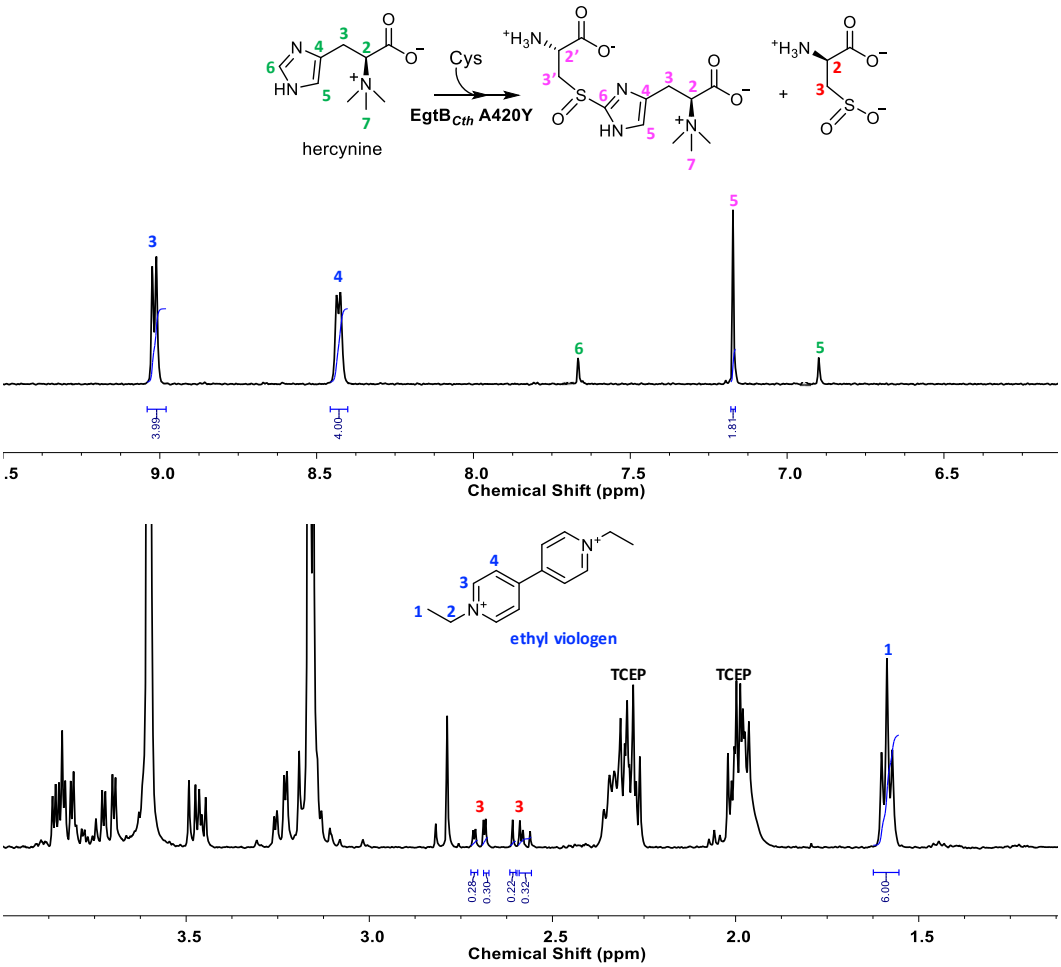


Figure S14. ^{1}H-NMR spectrum of EgtB$_{Cth}$ A420Y reaction using hercynine and L-Cys as the substrates for coupling product quantification. **Top**: Expanded regions between 6.5 ppm - 9.5 ppm. **Bottom**: Expanded regions between 1.0 – 4.0 ppm regions. The ratio between products from sulfoxide synthase and cysteine dioxygenase activities was quantified using ethyl viologen as an internal standard. The ratio between the proton of coupling product (labelled 5 in pink) and aromatic proton of ethyl viologen (labelled 4 in blue), and the ratio between β-hydrogen of cysteine sulfinic acid (labelled 3 in red) and the ethyl group of ethyl viologen (labelled 1 in blue) were used to quantify the ratio of the two products. In EgtB$_{Cth}$ A420Y reaction, the ratio between hercynyl-cysteine sulfoxide and cysteine sulfinic acid product was 3.21: 1. This ratio suggests that there was 76 % of coupling product in the reaction mediated by EgtB$_{Cth}$ A420Y variant.

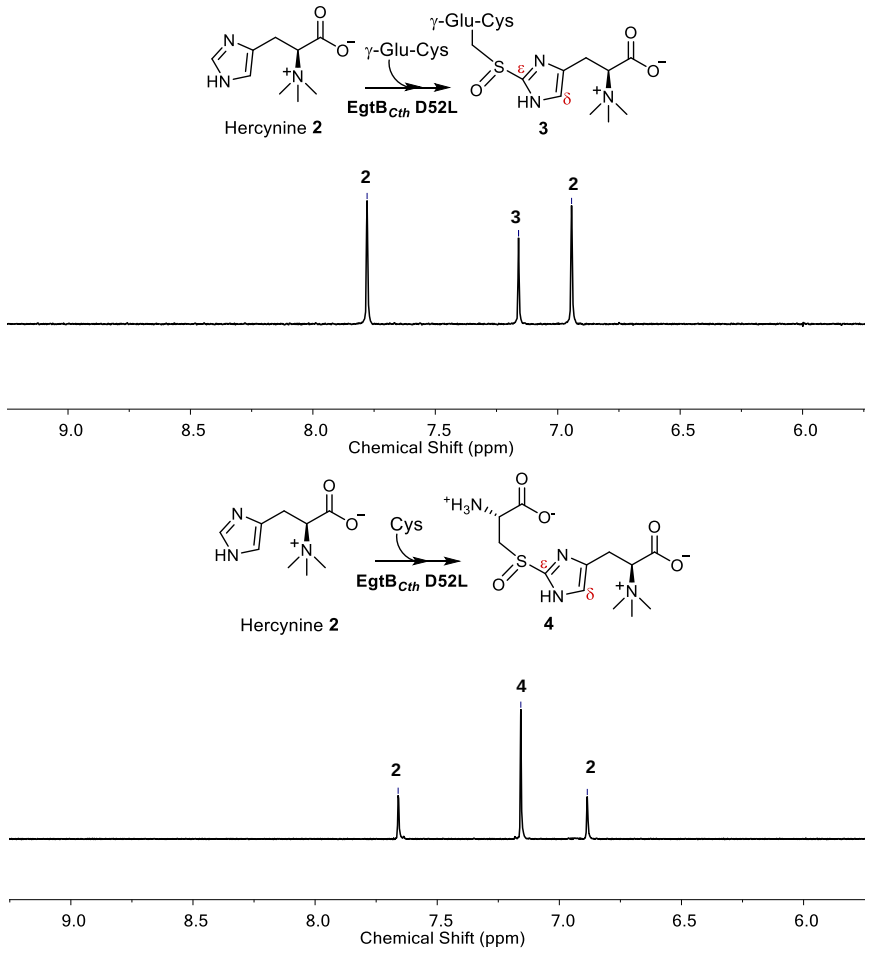


Figure S15. ^{1}H-NMR analysis of EgtB$_{Cth}$ D52L reactions. **Top**: Reaction of EgtB$_{Cth}$ D52L using hercynine and γ-Glu-Cys as the substrates shows the coupling product at δ 7.16. **Bottom**: Reaction of EgtB$_{Cth}$ D52L using hercynine and L-Cys as the substrates shows the coupling product at δ 7.16.

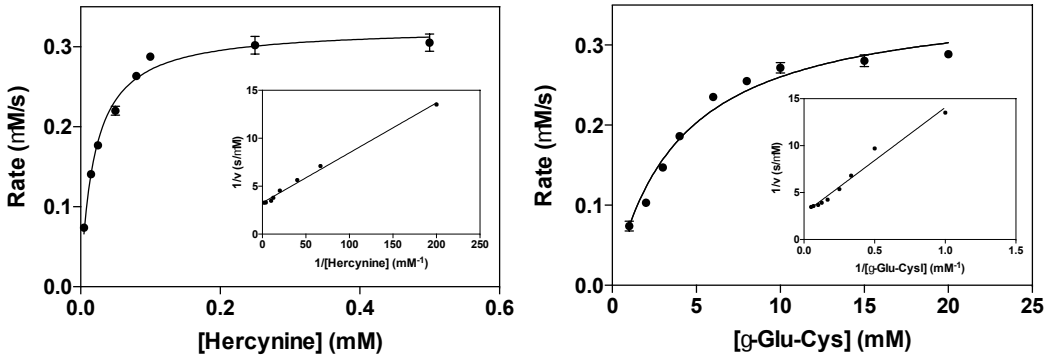


Figure S16. Kinetic analysis of EgtB$_{Cth}$ D52L mutant with hercynine and γ-Glu-Cys as the substrates. For hercynine concentration dependent, a 1-mL reaction contained 10 mM γ-Glu-Cys, 2 mM TCEP, 0.2 mM sodium ascorbate, 1.5 µM EgtB$_{Cth}$, and varied concentration of hercynine (0.005 – 0.5 mM) in 50 mM KPi pH 8.0. A similar reaction was setup for γ-Glu-Cys concentration dependent, except hercynine concentration was kept at 1 mM and γ-Glu-Cys concentration was varied (1 – 20 mM). EgtB$_{Cth}$ D52L mutant exhibited a K_M for hercynine of 19.6 ± 1.4 µM and a K_M for γ-Glu-Cys of 3.9 ± 0.3 mM with a k_{cat} of 12.9 ± 0.2 min^{-1}.

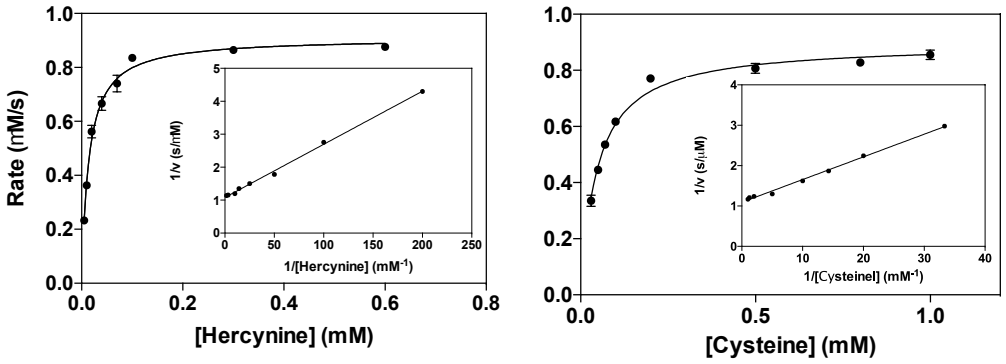


Figure S17. Kinetic analysis of EgtB$_{Cth}$ D52L mutant with hercynine and L-Cys as the substrates. For hercynine concentration dependent, a 1-mL reaction contained 2 mM L-Cys, 2 mM TCEP, 0.2 mM sodium ascorbate, 1.5 μM EgtB$_{Cth}$, and varied concentration of hercynine (0.005 – 0.6 mM) in 50 mM KPi pH 8.0. A similar reaction was setup for L-Cys concentration dependent, except hercynine concentration was kept at 2 mM and L-Cys concentration was varied (0.1 – 2 mM). EgtB$_{Cth}$ D52L mutant exhibited a K_M for hercynine of 13.9 ± 0.9 μM and a K_M for L-Cys of 47.3 ± 2.4 μM with a k_{cat} of 27.6 ± 0.5 min^{-1}.

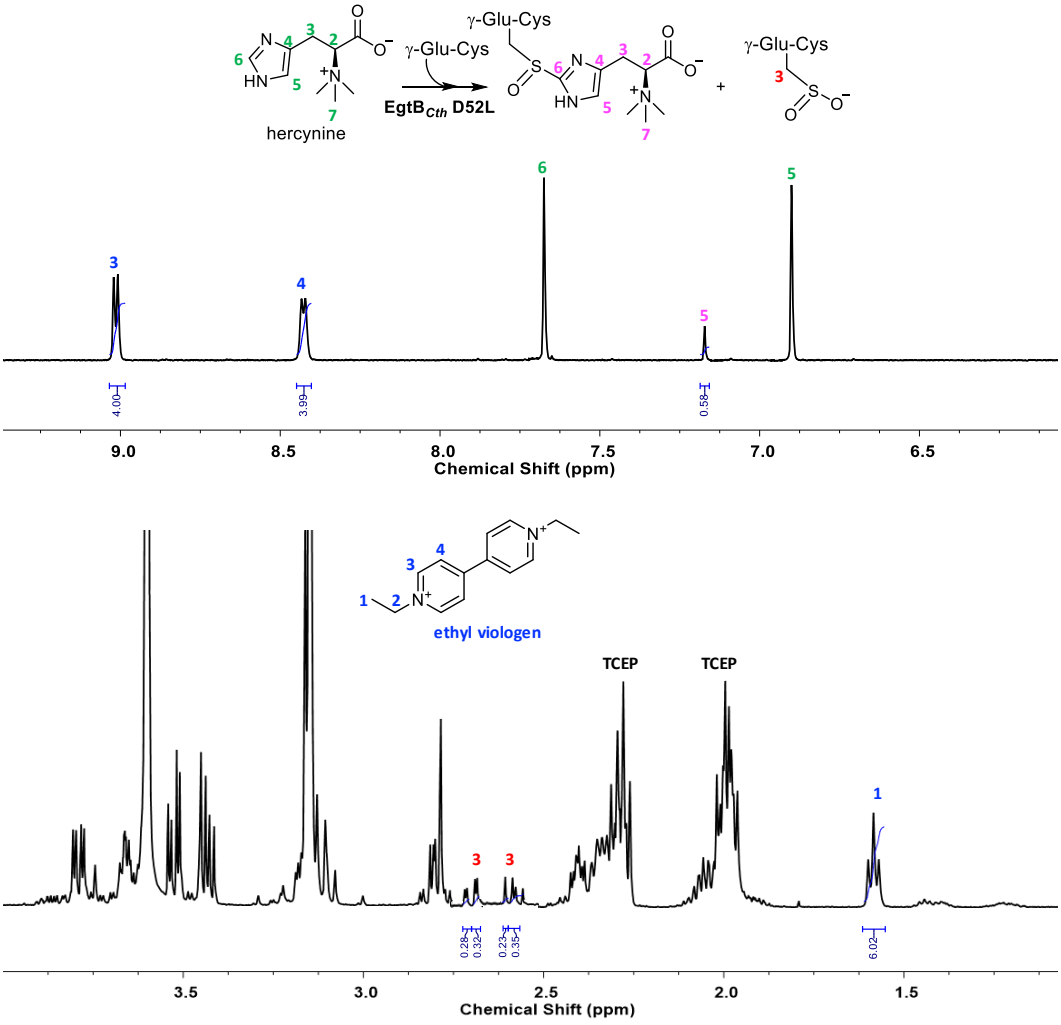


Figure S18. ^{1}H-NMR spectrum of EgtB$_{Cth}$ D52L reaction using hercynine and γ-Glu-Cys as the substrates for coupling product quantification. **Top**: Expanded regions between 6.5 ppm - 9.5 ppm. **Bottom**: Expanded regions between 1.0 – 4.0 ppm regions. The ratio between products from sulfoxide synthase and cysteine dioxygenase activities was quantified using ethyl viologen as an internal standard. The ratio between the proton of coupling product (labelled 5 in pink) and aromatic proton of ethyl viologen (labelled 4 in blue), and the ratio between β-hydrogen of γ-Glu-Cys sulfinic acid (labelled 3 in red) and the ethyl group of ethyl viologen (labelled 1 in blue) were used to quantify the ratio of the two products. In EgtB$_{Cth}$ D52L reaction, the ratio between hercynyl-γ-Glu-Cys sulfoxide and γ-Glu-Cys sulfinic acid product was 1: 1. This ratio suggests that there was 50 % of coupling product in the reaction mediated by EgtB$_{Cth}$ D52L variant.

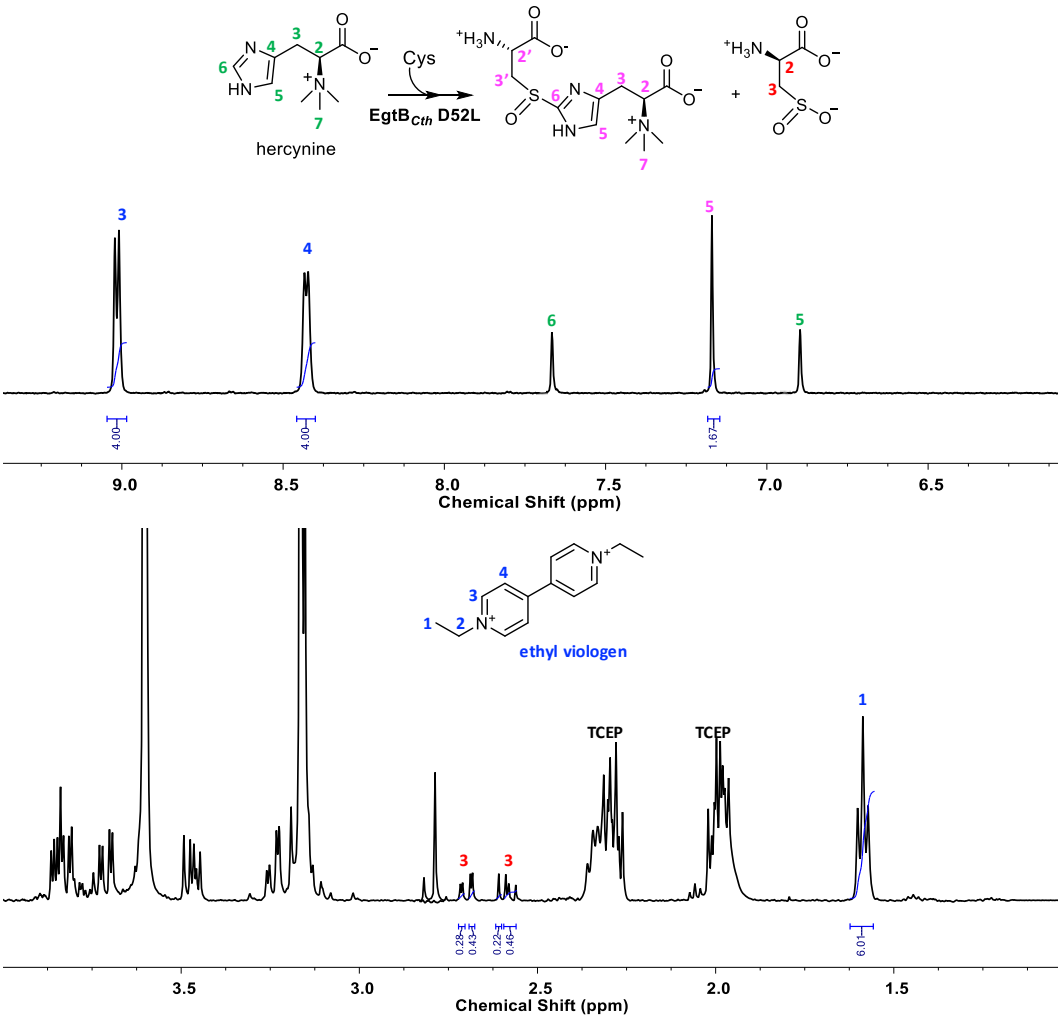


Figure S19. ^{1}H-NMR spectrum of EgtB$_{Cth}$ D52L reaction using hercynine and L-Cys as the substrates for coupling product quantification. **Top**: Expanded regions between 6.5 ppm - 9.5 ppm. **Bottom**: Expanded regions between 1.0 – 4.0 ppm regions. The ratio between products from sulfoxide synthase and cysteine dioxygenase activities was quantified using ethyl viologen as an internal standard. The ratio between the proton of coupling product (labelled 5 in pink) and aromatic proton of ethyl viologen (labelled 4 in blue), and the ratio between β-hydrogen of cysteine sulfinic acid (labelled 3 in red) and the ethyl group of ethyl viologen (labelled 1 in blue) were used to quantify the ratio of the two products. In EgtB$_{Cth}$ D52L reaction, the ratio between hercynyl-cysteine sulfoxide and cysteine sulfinic acid product was 2.24: 1. This ratio suggests that there was 69 % of coupling product in the reaction mediated by EgtB$_{Cth}$ D52L variant.

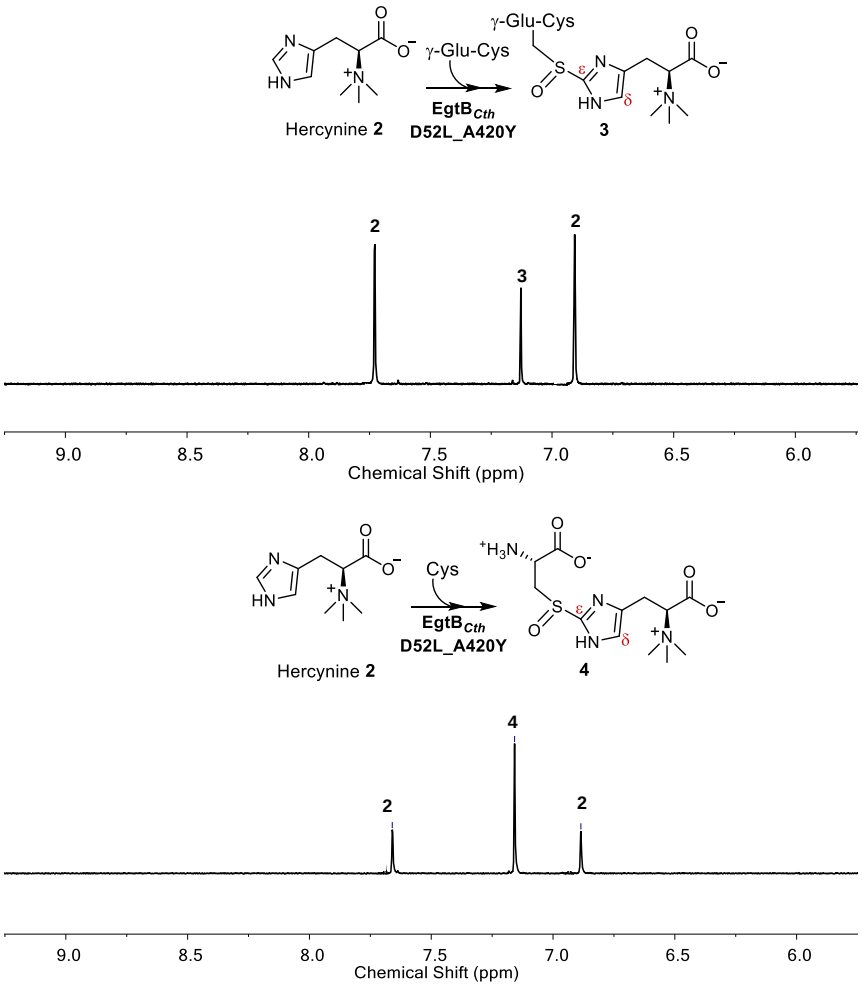


Figure S20. ^{1}H-NMR analysis of EgtB$_{Cth}$ D52L/A420Y reactions. **Top**: Reaction of EgtB$_{Cth}$ D52L/A420Y using hercynine and γ-Glu-Cys as the substrates shows the coupling product at δ 7.13. **Bottom**: Reaction of EgtB$_{Cth}$ D52L using hercynine and L-Cys as the substrates shows the coupling product at δ 7.16.

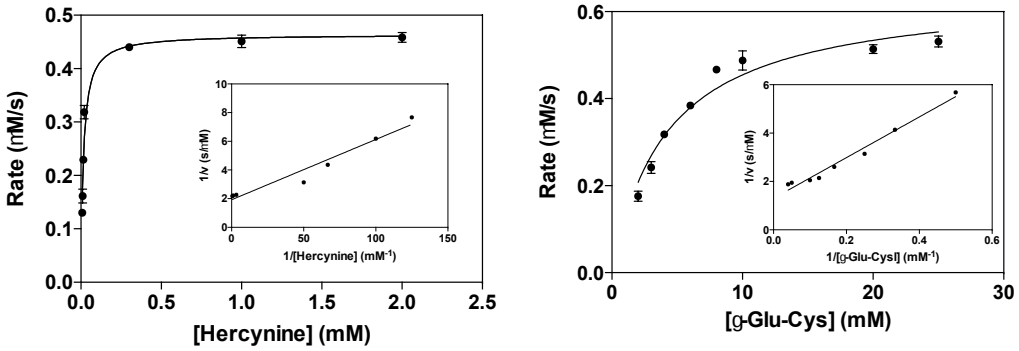


Figure S21. Kinetic analysis of EgtB$_{Cth}$ D52L/A420Y mutant with hercynine and γ-Glu-Cys as the substrates. For hercynine concentration dependent, a 1-mL reaction contained 10 mM γ-Glu-Cys, 2 mM TCEP, 0.2 mM sodium ascorbate, 1.5 μM EgtB$_{Cth}$ D52L/A420Y, and varied concentration of hercynine (0.08 – 2 mM) in 50 mM KPi pH 8.0. A similar reaction was setup for γ-Glu-Cys concentration dependent, except hercynine concentration was kept at 1 mM and γ-Glu-Cys concentration was varied (2 – 25 mM). EgtB$_{Cth}$ D52L/A420Y mutant exhibited a K_M for hercynine of 15.1 ± 1.3 μM and a K_M for γ-Glu-Cys of 4.2 ± 0.5 mM with a k_{cat} of 18.6 ± 0.4 min^{-1}.

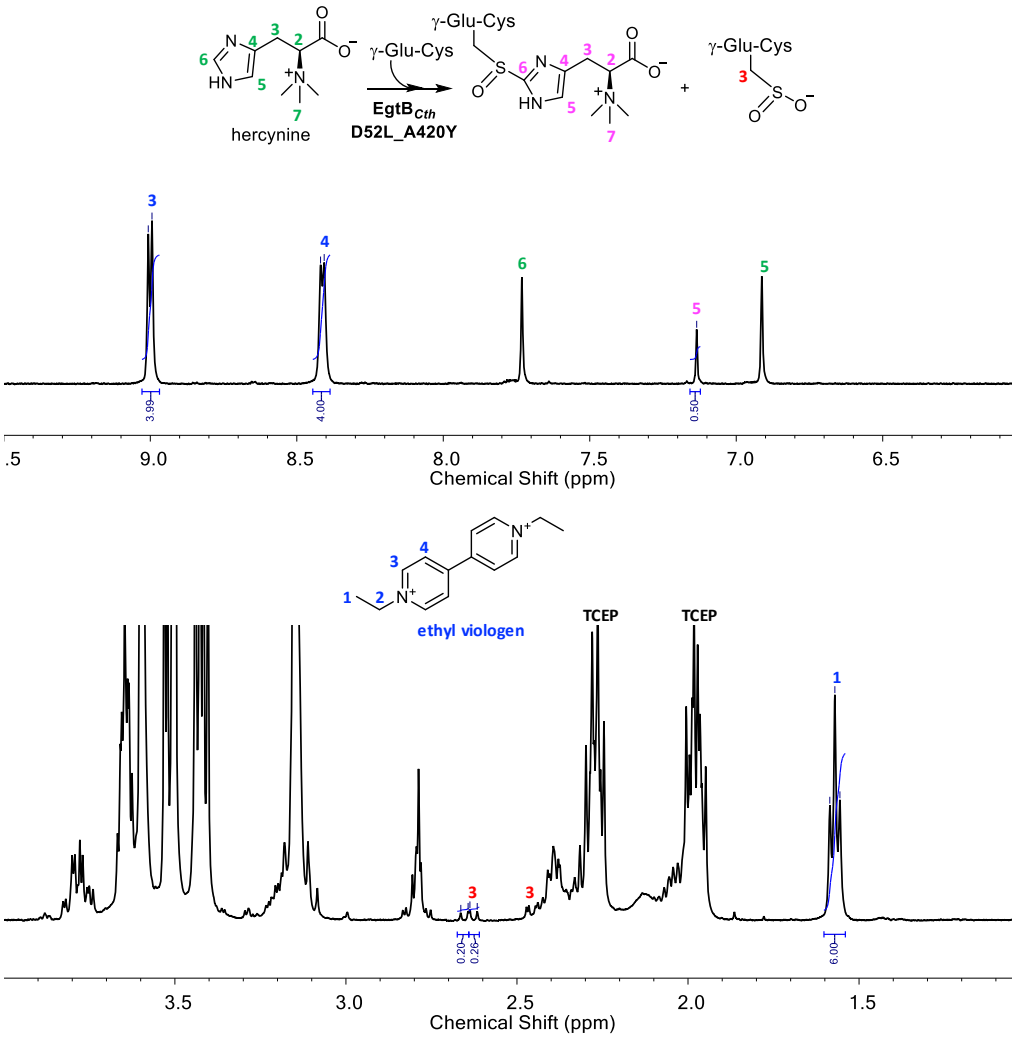


Figure S22. ^{1}H-NMR spectrum of EgtB$_{Cth}$ D52L/A420Y reaction using hercynine and γ-Glu-Cys as the substrates for coupling product quantification. **Top**: Expanded regions between 6.5 ppm - 9.5 ppm. **Bottom**: Expanded regions between 1.0 – 4.0 ppm regions. The ratio between products from sulfoxide synthase and cysteine dioxygenase activities was quantified using ethyl viologen as an internal standard. The ratio between the proton of coupling product (labelled 5 in pink) and aromatic proton of ethyl viologen (labelled 4 in blue), and the ratio between β-hydrogen of γ-Glu-Cys sulfinic acid (labelled 3 in red) and the ethyl group of ethyl viologen (labelled 1 in blue) were used to quantify the ratio of the two products. In EgtB$_{Cth}$ D52L/A420Y reaction, the ratio between hercynyl-γ-Glu-Cys sulfoxide and γ-Glu-Cys sulfinic acid product was 1.1: 1. This ratio suggests that there was 52 % of coupling product in the reaction mediated by EgtB$_{Cth}$ D52L/A420Y variant.

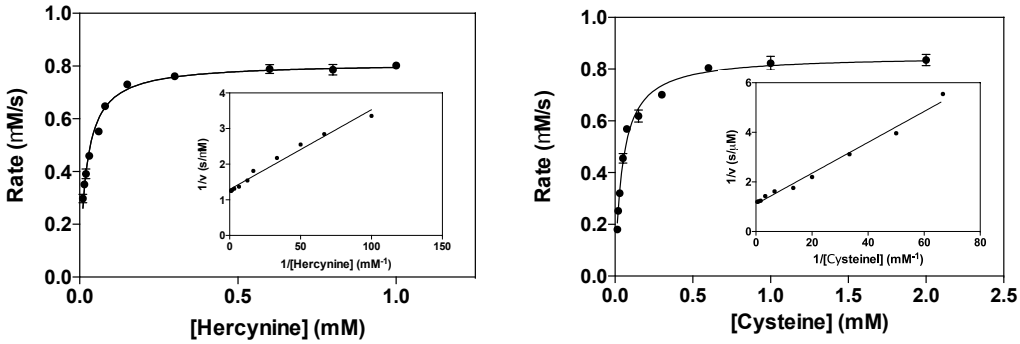


Figure S23. Kinetic analysis of EgtB$_{Cth}$ D52L/A420Y mutant with hercynine and L-Cys as the substrates. For hercynine concentration dependent, a 1-mL reaction contained 2 mM L-Cys, 2 mM TCEP, 0.2 mM sodium ascorbate, 1.5 μM EgtB$_{Cth}$ variant, and varied concentration of hercynine (0.01 – 1 mM) in 50 mM KPi pH 8.0. A similar reaction was setup for L-Cys concentration dependent, except hercynine concentration was kept at 2 mM and L-Cys concentration was varied (0.02 – 2 mM). EgtB$_{Cth}$ D52L/A420Y mutant exhibited a K_M for hercynine of 23.3 ± 1.2 μM and a K_M for L-Cys of 48.7 ± 3.5 μM with a k_{cat} of 32.7 ± 0.3 min^{-1}.

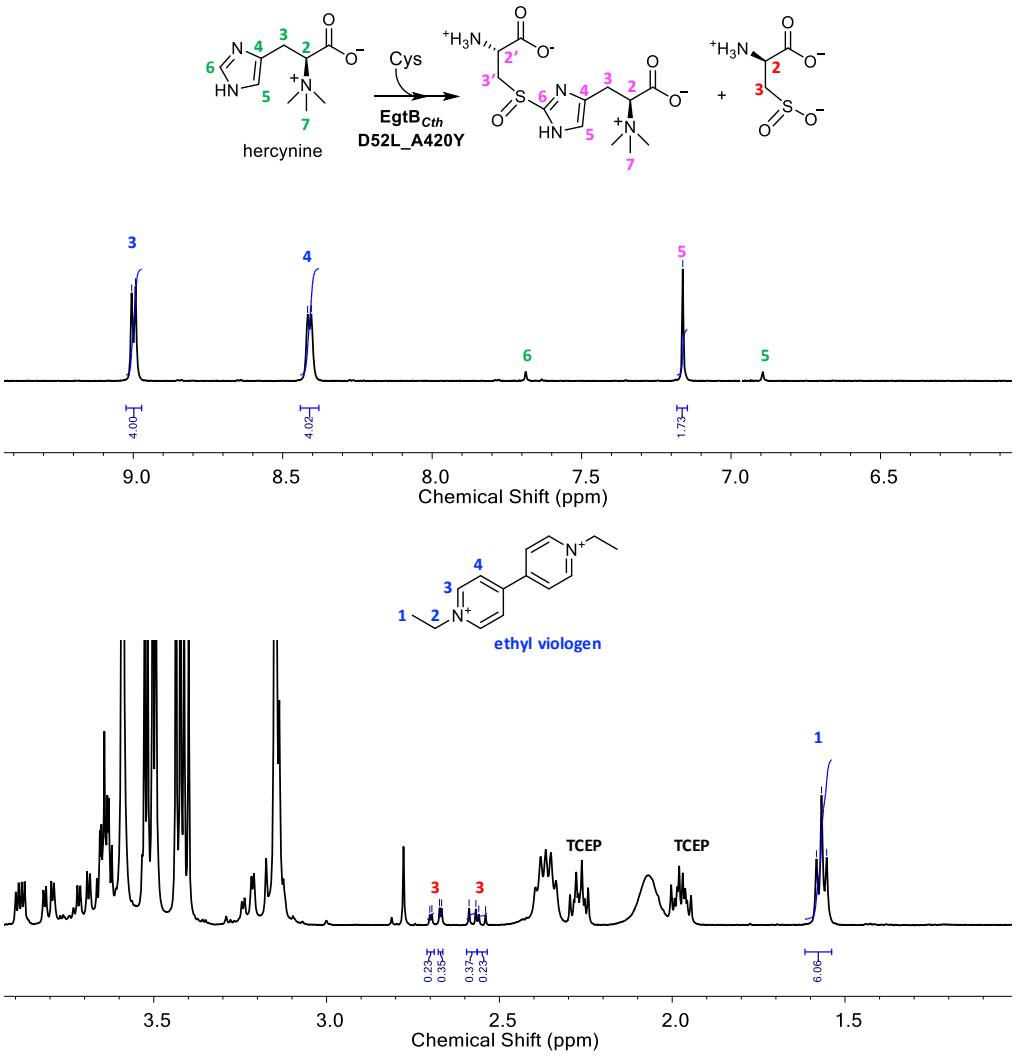


Figure S24. ^{1}H-NMR spectrum of EgtB$_{Cth}$ D52L/A420Y reaction using hercynine and L-Cys as the substrates for coupling product quantification. **Top**: Expanded regions between 6.5 ppm - 9.5 ppm. **Bottom**: Expanded regions between 1.0 – 4.0 ppm regions. The ratio between products from sulfoxide synthase and cysteine dioxygenase activities was quantified using ethyl viologen as an internal standard. The ratio between the proton of coupling product (labelled 5 in pink) and aromatic proton of ethyl viologen (labelled 4 in blue), and the ratio between β-hydrogen of cysteine sulfinic acid (labelled 3 in red) and the ethyl group of ethyl viologen (labelled 1 in blue) were used to quantify the ratio of the two products. In EgtB$_{Cth}$ D52L/A420Y reaction, the ratio between hercynyl-cysteine sulfoxide and cysteine sulfinic acid product was 2.93: 1. This ratio suggests that there was 75 % of coupling product in the reaction mediated by EgtB$_{Cth}$ D52L/A420Y variant.

Tables

Table S1. List of other genes from the same node of EgtB$_{Cth}$ which were retrieved from sequence similarity network analysis

#	Accession Code	Organism	Annotation
1	CAN97242	*Sorangium cellulosum*	5-histidylcysteine sulfoxide synthase
2	WP_062916112	*Paraburkholderia caribensis*	ergothioneine biosynthesis protein EgtB
3	WP_046312457	*Pontibacter korlensis*	ergothioneine biosynthesis protein EgtB

Table S2. Data collection and refinement statistics for $EgtB_{Cth}$ structure

	$EgtB_{Cth}$ Se Met	$EgtB_{Cth}$	$EgtB_{Cth}$ with Hercynine
Accession Number		6O6M	6O6L
Space group	P 1 2_1 1	P 1 2_1 1	P 1 2_1 1
Cell dimensions			
a, b, c (Å)	85.3, 137.1, 85.4	85.3,137.3, 85.4	85.2, 137.5, 85.2
α, β, γ (°)	90.0, 92.4, 90.0	90.0, 92.43, 90.0	90.0, 92.37, 90.0
Resolution (Å)	50.00 - 2.30 (2.34 - 2.30)[a]	50 - 2.5 (2.54 - 2.50)	50 - 2.25 (2.29 - 2.25)
R_{sym}	0.087 (0.334)	0.170 (0.604)	0.139 (0.585)
$I/\sigma I$	26.6 (5.2)	7.5 (1.8)	8.6(1.5)
Completeness (%)	100(100)	100 (99.8)	99.6 (95.4)
Redundancy	6.9 (7.0)	3.8 (3.7)	4.0 (3.7)
Wavelength	0.97942	0.97648	0.97648
Refinement			
Resolution (Å)		45.86 - 2.51	45.83 - 2.25
No. reflections		67215	92239
R_{work}/ R_{free}[b]		0.1853/0.2257	0.1877/0.2241
No. atoms			
Protein		13089	13062
Ligand		-	56
Ligand (Fe)		4	4
Solvent		504	651
B-factors (Å^2)			
Protein		17.8	20.9
Ligand		-	14.4
Ligand (Fe)		19.7	26.6
Solvent		16.6	19.7
R.m.s deviations			
Bond lengths (Å)		0.003	0.012
Bond angles (°)		0.656	1.086
Ramachandran Favoured (%)		97.8	97.6
Ramachandran Outliers (%)		0.0	0.0

[a] Represents statistics for the outmost shell

[b] R_{free} test set comprised of 2.1% of the unique reflections (max of 2000 spots)

Table S3. List of top 20 EgtB$_{Cth}$ mutants predicted from Rosetta Enzyme Design

#	Total Score	Mutations	#	Total Score	Mutations
1	-1170.13	D52HIS_F416GLU_A420ASP	11	-1162.81	D52ILE_F416GLU_A420CYS
2	-1169.53	D52HIS_F416ILE_A420ALA	12	-1162.78	D52ILE_F416GLN_A420ALA
3	-1167.37	D52LEU_F416GLY_A420ALA	13	-1162.76	D52ILE_F416GLY_A420TRP
4	-1166.03	D52ILE_F416LYS_A420PHE	14	-1162.67	D52ILE_F416ALA_A420THR
5	-1165.05	D52HIS_F416LEU_A420VAL	15	-1162.62	D52TRP_F416SER_A420HIS
6	-1164.86	D52ILE_F416ALA_A420CYS	16	-1162.3	D52ILE_F416GLU_A420THR
7	-1164.83	D52TRP_F416GLN_A420GLY	17	-1162.08	D52ILE_F416LYS_A420SER
8	-1164.54	D52ILE_F416GLY_A420LYS	18	-1161.81	D52GLY_F416THR_A420TYR
9	-1164.3	D52ILE_F416ALA_A420SER	19	-1161.75	D52HIS_F416CYS_A420ALA
10	-1163.97	D52HIS_F416THR_A420VAL	20	-1161.58	D52ILE_F416ALA_A420PHE

References

1. Song, H.; Her, A. S.; Raso, F.; Zhen, Z.; Huo, Y.; Liu, P., Cysteine Oxidation Reactions Catalyzed by a Mononuclear Non-heme Iron Enzyme (Ovoa) in Ovothiol Biosynthesis. *Org. Lett.* **2014,** *16*, 2122-2125.
2. Finn, R. D.; Coggill, P.; Eberhardt, R. Y.; Eddy, S. R.; Mistry, J.; Mitchell, A. L.; Potter, S. C.; Punta, M.; Qureshi, M.; Sangrador-Vegas, A., The Pfam Protein Families Database: Towards a More Sustainable Future. *Nucleic Acids Res.* **2016,** *44*, D279-D285.
3. Hu, W.; Song, H.; Her, A. S.; Bak, D. W.; Naowarojna, N.; Elliott, S. J.; Qin, L.; Chen, X.; Liu, P., Bioinformatic and Biochemical Characterizations of C-S Bond Formation and Cleavage Enzymes in the Fungus *Neurospora Crassa* Ergothioneine Biosynthetic Pathway. *Org. Lett.* **2014,** *16*, 5382-5385.
4. Terwilliger, T. C.; Adams, P. D.; Read, R. J.; McCoy, A. J.; Moriarty, N. W.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Zwart, P. H.; Hung, L.-W., Decision-making in Structure Solution Using Bayesian Estimates of Map Quality: The Phenix Autosol Wizard. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2009,** *65*, 582-601.
5. Otwinowski, Z.; Minor, W., Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Macromol. Crystallogr. A* **1997,** *276*, 307–326.
6. Emsley, P.; Cowtan, K., Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2004,** *60*, 2126-2132.
7. Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MolProbity: All-atom Structure Validation for Macromolecular Crystallography. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2010,** *66*, 12-21.
8. Lebedev, A. A.; Young, P.; Isupov, M. N.; Moroz, O. V.; Vagin, A. A.; Murshudov, G. N., JLigand: a Graphical Tool for the CCP4 Template-Restraint Library. *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2012,** *68*, 431-440.