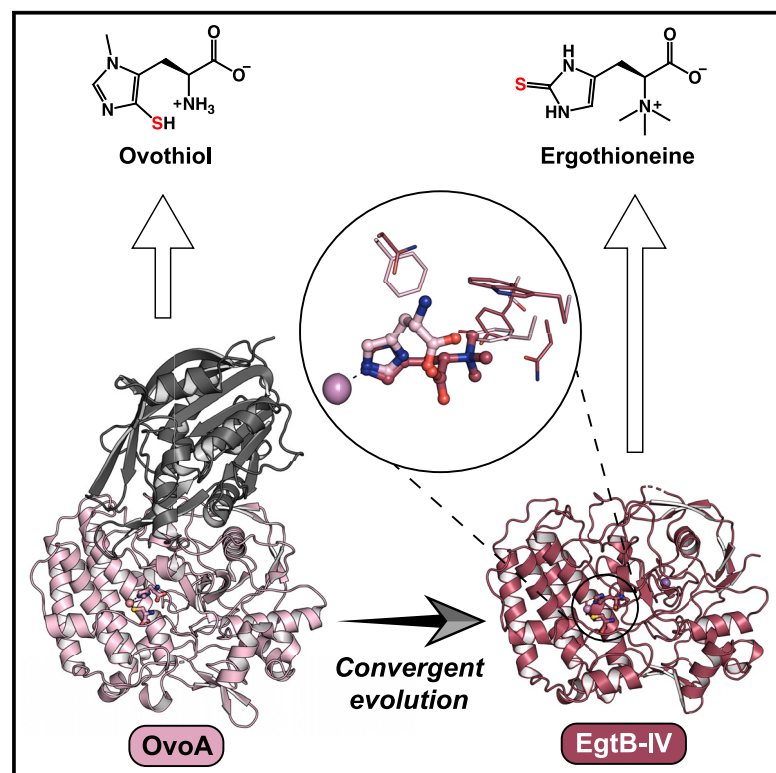


Structure

Structural insights into the convergent evolution of sulfoxide synthase EgtB-IV, an ergothioneine-biosynthetic homolog of ovothiol synthase OvoA

Graphical abstract



Authors

Kendra A. Ireland, Chase M. Kayrouz, Marissa L. Abbott, Mohammad R. Seyedsayamdost, Katherine M. Davis

Correspondence

katherine.davis@emory.edu

In brief

Ireland et al. report X-ray crystal structures of type IV EgtB enzymes involved in ergothioneine biosynthesis. The results shed light on the convergent evolution of sulfoxide synthase EgtB-IV, which features closest homology to ovothiol-biosynthetic OvoA. Additionally, they clarify the structural basis for EgtB-IV's specificity for cysteine as the S-donor.

Highlights

- Hercynine-bound structures offer insights into the convergent evolution of EgtB-IV
- EgtB-IV cannot be converted into OvoA through simple active site mutations
- An active site lid-loop dictates cysteine specificity in OvoA-like enzymes
- A potential evolutionary intermediate between OvoA and EgtB-IV is presented

Article

Structural insights into the convergent evolution of sulfoxide synthase EgtB-IV, an ergothioneine-biosynthetic homolog of ovothiol synthase OvoA

Kendra A. Ireland,^{1,4} Chase M. Kayrouz,^{2,4} Marissa L. Abbott,² Mohammad R. Seyedsayamdost,^{2,3} and Katherine M. Davis^{1,5,*}

¹Department of Chemistry, Emory University, Atlanta, GA 30322, USA

²Department of Chemistry, Princeton University, Princeton, NJ 08544, USA

³Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

⁴These authors contributed equally

⁵Lead contact

*Correspondence: katherine.davis@emory.edu

<https://doi.org/10.1016/j.str.2024.08.006>

SUMMARY

Non-heme iron-dependent sulfoxide/selenoxide synthases (NHISS) constitute a unique metalloenzyme class capable of installing a C–S/Se bond onto histidine to generate thio/selenoimidazole antioxidants, such as ergothioneine and ovothiol. These natural products are increasingly recognized for their health benefits. Among associated ergothioneine-biosynthetic enzymes, type IV EgtBs stand out, as they exhibit low sequence similarity with other EgtB subfamilies due to their recent divergence from the ovothiol-biosynthetic enzyme OvoA. Herein, we present crystal structures of two representative EgtB-IV enzymes, offering insights into the basis for this evolutionary convergence and enhancing our understanding of NHISS active site organization more broadly. The ability to interpret how key residues modulate substrate specificity and regioselectivity has implications for downstream identification of divergent reactivity within the NHISS family. To this end, we identify a previously unclassified clade of OvoA-like enzymes with a seemingly hybrid set of characteristics, suggesting they may represent an evolutionary intermediate between OvoA and EgtB-IV.

INTRODUCTION

Sulfur (S) and selenium (Se) are essential for maintaining cellular life, fulfilling critical roles in both primary and secondary metabolism. Furthermore, the natural products that contain these elements often exhibit potent therapeutic properties, and great strides have been made toward unraveling the diverse strategies enzymes employ to forge C–S/Se bonds.^{1,2} Thio/selenoimidazoles (TSIs), for example, are a redox-active class of histidine-derived biomolecules, primarily produced by bacteria and fungi, that have been studied extensively for their promising antioxidant, anti-inflammatory, and cytoprotective properties.^{3–8} Four members of the TSI family have been identified to date: ergothioneine, ovothiol, selenoneine, and the recently discovered metabolite ovoselenol (Figure 1A).⁹

In each case, the characteristic C–S/Se bond is installed by an O₂-dependent, mononuclear non-heme iron sulfoxide/selenoxide synthase (NHISS, Figures 1A and 1B). The sulfoxide synthases (OvoA and EgtB) utilize cysteine derivatives to catalyze oxidative C–S bond formation in the biosynthesis of ovothiol and ergothioneine, respectively (Figure S1), while the selenoxide synthases (OvsA and SenA) exploit the concomitant production of selenosugars to generate C–Se bonds in ovoselenol and sele-

noneine. The regioselectivity of these reactions is dictated by the identity of the histidine-derived substrate. OvoA and OvsA catalyze C–S/Se bond formation at the imidazole 5' carbon of histidine,^{9,10} whereas EgtB and SenA modify the 2' position of the trimethylated amino acid hercynine.^{11,12}

Phylogenetic analysis of biochemically characterized bacterial NHISS enzymes illustrates an early divergence in their evolution, which led to differences in regioselectivity and separated the EgtB-like and OvoA-like enzymes into two main branches (Figure 1C). Several other instances of active site remodeling further differentiated the family into six distinct bacterial subfamilies: EgtB types I, II, and IV, as well as OvoA, SenA, and OvsA. Intriguingly, although type IV EgtBs share high sequence identity with the OvoA subfamily, suggesting they evolved from an ancestral ovothiol-producing OvoA enzyme, these enzymes exhibit ergothioneine-biosynthetic activity.¹³ Likewise, OvsA shares comparatively little sequence homology with OvoA yet catalyzes a similar reaction. We recently demonstrated that condensing the active site residues involved in substrate recognition and catalysis into composite motifs (Figures 1B and 1C) can aid in classifying uncharacterized NHISS subfamilies irrespective of their overall sequence homology.⁹ However, structure elucidation is necessary to fully understand the intricacies

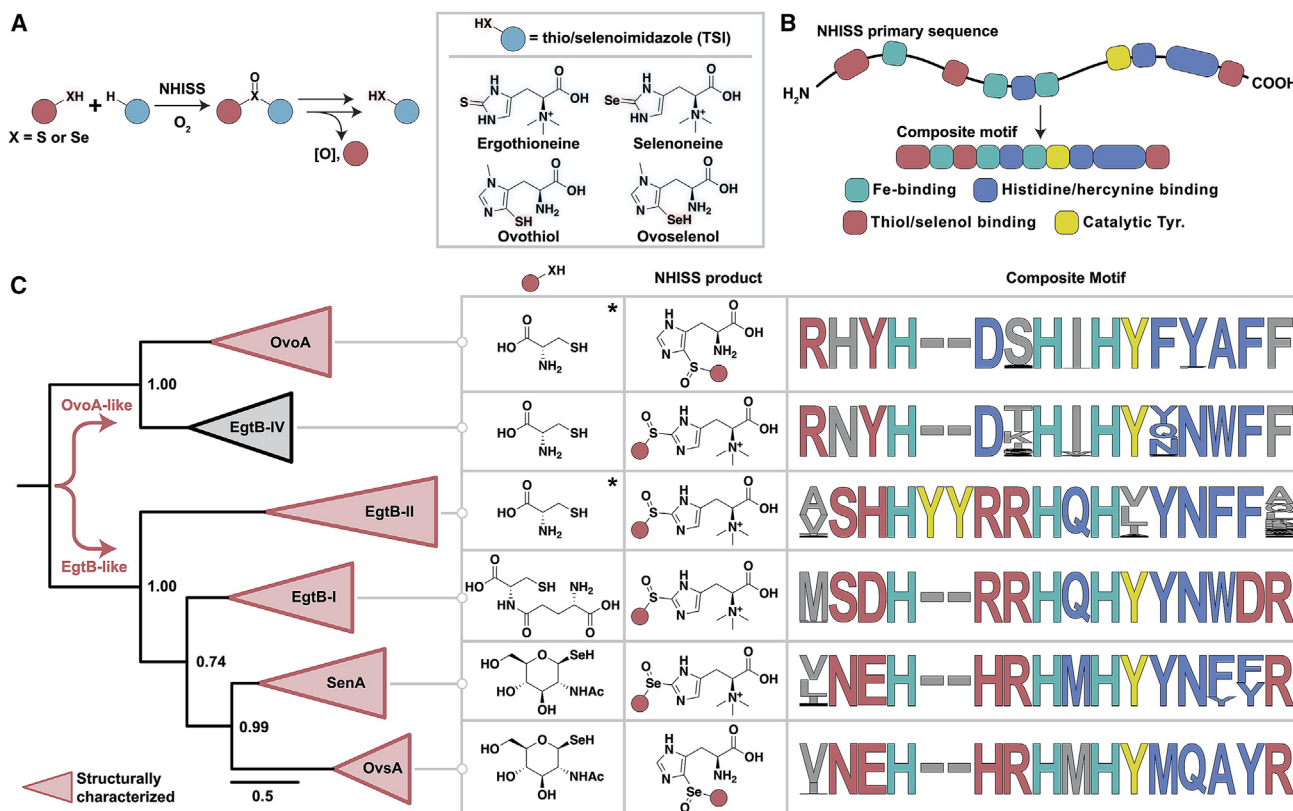


Figure 1. NHISS enzymes involved in the biosynthesis of S- and Se-containing histidine-derived antioxidants

(A) General reaction catalyzed by non-heme iron-dependent sulfoxide/selenoxide synthases. There are four known final products (TSIs) of NHISS-containing biosynthetic pathways.

(B) Compression of active site residues involved in iron binding, substrate recognition, and catalysis into composite motifs.

(C) Phylogenetic analysis for the biochemically characterized bacterial NHISS subfamilies with the preferred substrate, NHISS product, and composite motif shown for each subfamily. The initial bifurcation between OvoA-like and EgtB-like branches is emphasized in red. Bootstrap values (1,000 samples) are displayed at key nodes. The asterisk for OvoA and EgtB-II indicates that some homologs can accept γ -Glu-Cys in addition to their preferred substrate, cysteine.

of ligand recognition within each subfamily. Here, we report X-ray crystal structures of two representative EgtB-IV enzymes and offer insights into the convergent evolution of ergothioneine-biosynthetic reactivity in OvoA-like enzymes.

RESULTS

Global structure of EgtB-IV

We began our investigation by recombinantly expressing the EgtB-IV from *Crocospaera subtropica* ATCC 51142 (EgtB-IV_{Csu}) in *Escherichia coli*, purifying it using standard methods, and crystallizing it via the sitting drop vapor diffusion approach. The structure of the enzyme was subsequently solved to 1.96 Å resolution (PDB: 8VIK; Table 1). The overall architecture of EgtB-IV_{Csu} closely resembles the sulfoxide/selenoxide synthase domains of other NHISS subfamilies (Table S1; Figure S2). More specifically, the enzyme adopts a DinB-like fold¹⁴ characterized by a four-helix bundle at the N-terminus (residues 1–174), while the C-terminus (residues 175–448) is primarily composed of loops and structurally reminiscent of the formylglycine-generating enzyme (FGE) subclass of C-type lectins (Figure 2A).^{15,16} The interface between these domains forms the enzyme's active

site, with the mononuclear iron cofactor (Mn²⁺ in the ternary complex, see the [method details](#) section for more information) coordinated by residues H61, H154, and H158. Although several putative NHISS subfamilies appear to deviate from this motif,⁹ the three-His facial triad is conserved in all six subfamilies characterized to date, as well as in most members of the DinB superfamily.

Hercynine binding elucidates the evolution of an OvoA-like enzyme for ergothioneine biosynthesis

To investigate the basis for EgtB-IV's somewhat unexpected substrate specificity—that is, its preference for hercynine rather than histidine,¹³ we also determined structures of the holoenzyme complexed with hercynine (1.55 Å, PDB: 8VIH; Figure S3A) and a ternary complex consisting of EgtB-IV_{Csu}·Mn·hercynine·cysteine (1.62 Å, PDB: 8VII). In both models, hercynine adopts a configuration wherein its imidazole 2' carbon is oriented toward the side chain of Y397 (Figures 2B and 2C). Previous biochemical investigations on OvoA from *Erwinia tasmaniensis* (OvoA_{Eta}) and EgtB-I from *Mycobacterium thermoresistibile* (EgtB-I_{Mth}) indicate that this conserved Tyr residue (Figure 1C) is necessary for sulfoxidation activity.^{17,18} The observed proximity of Y397 to the

Table 1. Crystallographic data processing and refinement statistics for EgtB-IV structures

PDB ID	8VIK (EgtB-IV _{CSU} + Fe)	8VIH (EgtB-IV _{CSU} + Fe + hercynine)	8VII (EgtB-IV _{CSU} + Mn + hercynine + cysteine)	8VIL (EgtB-IV _{CSU} + Fe + DMH)	8VIG (EgtB-IV _{GEM} + Fe + hercynine)
Data collection^a					
Space group	P 1 2 ₁ 1	P 1 2 ₁ 1	P 1 2 ₁ 1	P 1 2 ₁ 1	P 2 ₁ 2 ₁ 2 ₁
Unit cell (Å)	a = 62.197, b = 94.511, c = 83.747 α = γ = 90, β = 100.162	a = 62.197, b = 94.511, c = 83.747 α = γ = 90, β = 100.162	a = 62.197, b = 94.511, c = 83.747 α = γ = 90, β = 100.162	a = 62.197, b = 94.511, c = 83.747 α = γ = 90, β = 100.162	a = 47.177, b = 95.639, c = 101.392 α = β = γ = 90
Wavelength (Å)	0.9686	0.9537	0.9686	0.9686	0.9537
Resolution range (Å)	29.43–1.961 (2.031–1.961)	47.26–1.55 (1.605–1.55)	29.43–1.621 (1.678–1.621)	29.43–1.95 (2.02–1.95)	47.82–1.6 (1.657–1.6)
Total observations	133398 (10433)	264471 (24971)	221033 (16982)	132129 (11643)	122367 (11940)
Total unique observations	66735 (5237)	136888 (13254)	116826 (9442)	68198 (6103)	61246 (5990)
CC _{1/2}	0.998 (0.549)	0.996 (0.506)	0.993 (0.551)	0.997 (0.569)	0.996 (0.509)
I/σ ₁	9.17 (1.10)	7.38 (1.13)	6.70 (1.25)	10.10 (1.25)	7.45 (1.02)
Completeness (%)	97.37 (76.57)	99.12 (96.69)	96.67 (78.18)	97.86 (86.03)	99.81 (99.40)
R _{merge}	0.0487 (0.5962)	0.0513 (0.5944)	0.06398 (0.5344)	0.05831 (0.646)	0.05404 (0.677)
R _{pim}	0.0487 (0.5962)	0.0513 (0.5944)	0.06398 (0.5344)	0.05831 (0.646)	0.05404 (0.677)
Redundancy	2.0 (2.0)	1.9 (1.9)	1.9 (1.8)	1.9 (1.9)	2.0 (2.0)
Refinement statistics^a					
Resolution range (Å)	29.43–1.961	47.26–1.55	29.43–1.621	29.43–1.95	47.82–1.6
Reflections (total)	66565 (5219)	136711 (13221)	116800 (9435)	67983 (5928)	61241 (5987)
Reflections (test)	972 (75)	1994 (193)	1707 (137)	993 (86)	1997 (187)
Total atoms refined	7403	8006	8250	7455	4053
Solvent	497	854	997	516	444
R _{work} /R _{free}	0.1899/0.2194	0.1794/0.1982	0.1662/0.1932	0.2010/0.2153	0.1861/0.2072
RMSDs					
Bond lengths (Å)/angles (°)	0.004/0.67	0.016/1.31	0.008/0.97	0.003/0.61	0.007/0.93
Ramachandran plot					
Favored/allowed (%)	97.99/2.01	98.38/1.62	98.41/1.59	97.29/2.71	97.92/2.08
Mean B values (Å²)					
Protein	35.57	20.40	21.79	32.33	19.99
Ligands	40.77	22.04	19.56	41.79	27.77
Solvent	39.29	29.42	31.18	37.14	30.17

^aValues in parentheses refer to the high-resolution shell.

2' carbon involved in C–S bond formation and the site of O₂ binding to the iron cofactor support a similar role for Y397 in EgtB-IV. Furthermore, the observed spatial arrangement facilitates the 2' regioselectivity necessary for ergothioneine biosynthesis.

Analogous to the other hercynine-accepting subfamilies,^{19–21} EgtB-IV_{CSU} binds the substrate through water-mediated hydrogen bonds (H-bonds) and dipolar interactions (Figures 2B–2D). These include dipolar contacts and cation–π interactions between hercynine's trimethylammonium group and the Y428/N430 and W431/F432 side chains, respectively (Figure 2D). Additionally, R52 and T161 interact with the carboxylate moiety via a water molecule, while the backbone carbonyl and the side chain of N400 form a water-mediated H-bond with the imidazole π-nitrogen. N400 and D93 also weakly coordinate to hercynine's carboxylate moiety through two connected water molecules (Figure 2C). OvoA from *Hydrogenimonas thermophila* (OvoA_{TH2}) features the same

conserved Asp residue (D103), which instead forms a direct H-bond to the histidine carboxylate.

Particularly notable, however, among EgtB-IV_{CSU}'s hercynine-binding contacts are those between the trimethylammonium group of hercynine and the side chains of N430 and W431. The presence of an Asn residue at this position immediately followed by an aromatic residue (Trp, Phe, or Tyr) is strictly conserved among hercynine-binding NHISS enzymes (Figure 1C). Furthermore, Liao and Seebeck demonstrated that the introduction of this motif (N-W) into OvoA_{ETA} engendered modest selectivity for hercynine.¹³ Similarly, we showed that OvsA's substrate specificity can be converted to hercynine by transplanting this motif from SenA (N-F) into OvsA from *Halomonas utahensis* (in combination with an M401Y mutation: OvsA_{HUT}-M401Y/Q430N/A431F).⁹ In the absence of this N-W/F/Y motif, histidine is the preferred substrate and adopts an alternative binding mode in

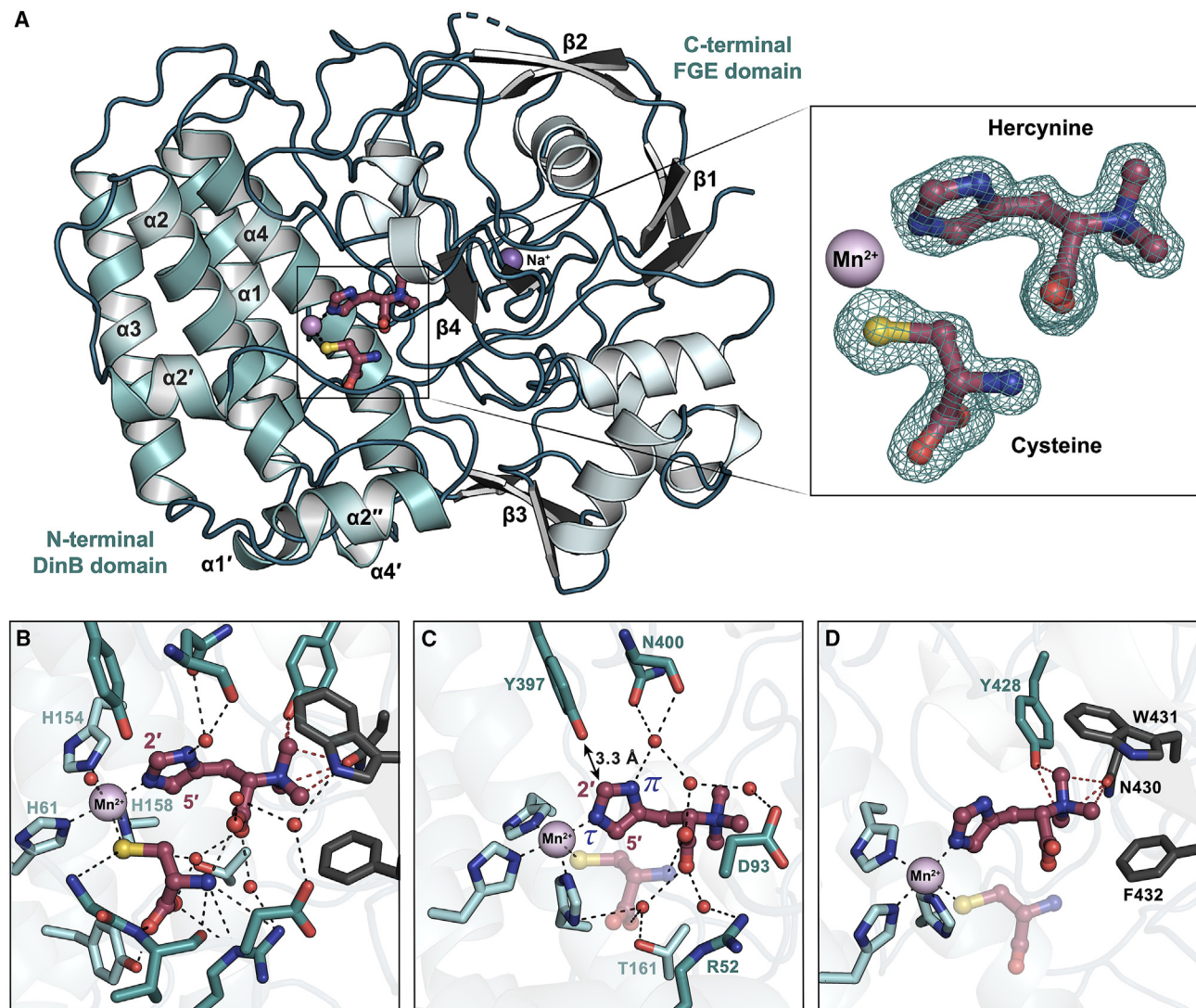


Figure 2. Structural analysis of EgtB-IV_{Csu} in complex with hercynine and cysteine

(A) Overall architecture of EgtB-IV_{Csu} (PDB: 8VII). Domains and secondary structure elements are labeled. The DinB domain contains four minor α -helices ($\alpha 1'$, $\alpha 2'$, $\alpha 2''$, and $\alpha 4'$) in addition to the primary four-helix bundle ($\alpha 1$ – $\alpha 4$). The loop-rich FGE domain comprises four antiparallel β -sheets ($\beta 1$ – $\beta 4$) and nine short α -helices. Hercynine and cysteine are shown in maroon sticks and Mn^{2+} is shown as a purple sphere. Inset: composite omit map ($F_o - F_c$) contoured at 1.0σ for hercynine and cysteine in the EgtB-IV_{Csu} ternary complex.

(B) H-bonds (black dashed lines) and dipolar interactions (red dashed lines) for substrate binding in EgtB-IV_{Csu}. Corresponding to (A), residues located within the DinB domain α -helices are shown in light blue, residues located within loops are shown in dark teal, and residues present in β -sheets within the FGE domain are shown in dark gray. In the absence of O_2 , a water molecule binds to the Mn^{2+} cofactor to complete the octahedral arrangement.

(C) H-bonds between hercynine and EgtB-IV_{Csu}.

(D) Dipolar interactions (red dashed lines) and cation- π interactions (mediated by W431 and F432) with the trimethylammonium group of hercynine.

which the imidazole 5' carbon faces the catalytic Tyr (Figure 3). This regioselectivity-modulating reorientation is facilitated by an H-bonding interaction between the now free substrate α -amino group and the backbone carbonyl of a conserved active site residue (F409 in OvoA_{Th2}, M401 in OvsA_{Hut}; Figures 3 and S4).

To investigate whether the dipolar contacts with N430 and cation- π interactions with W431 are imperative for conferring hercynine specificity in EgtB-IV, we generated N430A-, W431A-, and N430A/W431A-EgtB-IV_{Csu} variants. Despite the clear correlation between the presence of the N-W/F/Y motif and native reactivity

with hercynine, these mutations had negligible effect on the enzyme's ability to sulfoxidize the methylated substrate as enzymatic activity assays showed similar turnover as the wild-type (Figure S5A). We subsequently generated a quintuple mutant consisting of N400F/F404T/Y428S/N430Y/W431A designed to convert the EgtB-IV_{Csu} active site to that of OvoA (Figures 3 and S5B). Remarkably, this variant retained significant EgtB-like activity (Figure S5A), indicating the interactions governing substrate specificity in EgtB-IV are substantially more nuanced than they initially appear. One potential explanation is that during the evolutionary transition from an ancestral OvoA to the extant EgtB-IV

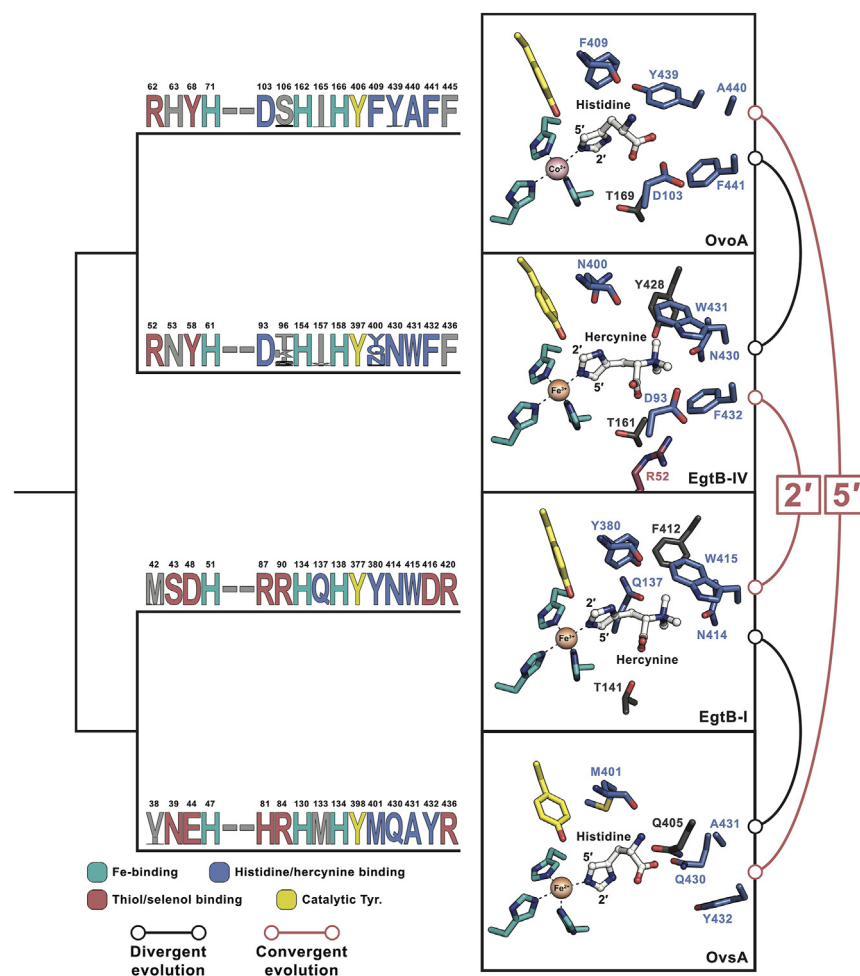


Figure 3. Comparison of histidine/hercynine-binding residues illustrates two examples of convergent evolution in the NHISS family

Active site residues for OvoA_{Th2} (PDB: 8KHQ), EgtB-IV_{Csu} (PDB: 8VIH), EgtB-I_{Mth} (PDB: 4X8E), and OvsA_{Hut} (PDB: 8U41) are colored according to the composite motifs. Substrate-binding residues not included in the composite motifs are shown in dark gray. In addition to serving as a thiol-binding residue, R52 in EgtB-IV_{Csu} forms a water-mediated H-bond with the carboxylate of hercynine.

tween the two enzymes, while additional variations in sequence are evenly distributed throughout the rest of the enzyme (Figure S6). It was not too surprising, therefore, that the observed orientations of hercynine are nearly identical between the homologous structures, albeit with slightly different interactions. More specifically, EgtB-IV_{Gem} lacks the water-mediated H-bond between the N400 side chain and the π -nitrogen of the substrate imidazole ring. Instead, the hydroxyl group of Y395 appears to form an additional dipolar contact with the trimethylammonium group in EgtB-IV_{Gem} (Figure S6). The EgtB-like branch seems to have evolved with Tyr at this position to aid in hercynine binding, suggesting a functional adaptation. However, it is evident from our analysis of type IV EgtB orthologs that the associated arrange-

ment of the active site is not the sole configuration capable of conferring hercynine specificity. Evolution of an amide group at the same position, for example, facilitates interaction with a disparate moiety of the substrate to yield an alternative substrate binding mode. We conclude that the precise identity of this residue is not crucial so long as strong association with the substrate is maintained, as is the case with EgtB-IV_{Gem}.

subfamily, EgtB-IV's overall scaffold became preorganized for hercynine binding, irrespective of the exact active site composition. EgtB-IV's active site may also possess more inherent flexibility, allowing it to bind hercynine even in the absence of N430 and W431. For example, a potential mode of hercynine binding to the EgtB-IV_{Csu} quintuple mutant can be envisioned where side-chain rotation of the newly introduced Thr and Tyr residues (F404T and N430Y) facilitate additional H-bonding and cation- π interactions (Figure S5C).

Further inspection of the NHISS composite motifs reveals an irregularity in the hercynine-binding residue immediately following the catalytic Tyr of EgtB-IV (Figure 1C). Notably, type IV EgtBs are the sole hercynine-binding subfamily lacking a fully conserved Tyr at this position, with Gln and Asn appearing most frequently. Although H-bonding predominantly involves the backbone carbonyl of this residue, we questioned whether this unusual heterogeneity in the substrate-binding motif might offer insights into the basis for EgtB-IV's unusually strong preference for hercynine. To complement our structures of EgtB-IV_{Csu}, which features Asn (N400) at this position, we also crystallized a Tyr-containing (Y395) homolog from *Geminocystis* sp. isolate SKYG4 (EgtB-IV_{Gem}) in the presence of hercynine (PDB: 8VIG; 1.60 Å resolution; Figure S3B). The active site is otherwise fully conserved, despite moderate sequence identity of 55% be-

ment of the active site is not the sole configuration capable of conferring hercynine specificity. Evolution of an amide group at the same position, for example, facilitates interaction with a disparate moiety of the substrate to yield an alternative substrate binding mode. We conclude that the precise identity of this residue is not crucial so long as strong association with the substrate is maintained, as is the case with EgtB-IV_{Gem}.

Regioselectivity of EgtB-IV-catalyzed bond formation is unaltered by the N- α -methylation pattern

While the aforementioned conservation analysis provided compelling insights into the convergent evolution of ergothioneine biosynthesis, the basis for EgtB-IV's seemingly immutable preference for hercynine remains elusive. Equally perplexing is the lack of activity for any of the tested EgtB-IV_{Csu} variants when presented with histidine (Figure S5A). To further explore the structural basis for reactivity with non-native substrates in EgtB-IV, we solved a crystal structure of EgtB-IV_{Csu} in complex with N- α -dimethyl-histidine (DMH) to 1.95 Å resolution (PDB: 8VIL; Figure S7A). Previous structural characterization of other NHISS subfamilies demonstrated that the N- α -methylation pattern of the presented substrate influences its orientation within NHISS active sites, pivotal for determining the regioselectivity of C-S/Se bond formation.^{9,19–23}

Despite the loss of a methyl group, the mode of DMH binding we observe in EgtB-IV_{Csu} (Figure S7B) closely resembles that of hercynine, indicating conservation of EgtB-like reactivity. This is consistent with the previously reported retention of 2' regioselectivity, albeit in poor yield, when histidine is presented as a substrate to EgtB-IV from *Microcystis aeruginosa*.¹³ Furthermore, these data align with the general observation that DMH typically exhibits regioselectivity following that of the NHISS enzyme's native substrate. For example, DMH binds to EgtB-I_{Mth} with the imidazole 2' carbon facing its catalytic Tyr residue (PDB: 4X8D), while OvsA catalyzes C–Se bond formation onto DMH with 5' regioselectivity (Figure S8). By contrast, OvoA_{Eta} displays a peculiar proclivity to sulfoxidize methylated histidine derivatives with non-native, EgtB-like reactivity: *N*- α -monomethyl-histidine with a 3:2 ratio of 5':2' regioselectivity, and DMH/hercynine with 2' regioselectivity.²⁴ As previously suggested by Liao and Seebeck, this unusual promiscuity likely aided in OvoA's evolutionary transition away from ovothiol and toward ergothioneine biosynthesis.¹³ Altogether, it seems likely that EgtB-IV's evolutionary divergence from an ancestral ovothiol-producing OvoA enzyme not only led to a pronounced preference for methylated histidine derivatives but also facilitated exclusive 2' sulfoxidation, regardless of the *N*- α -methylation pattern.

EgtB-IV features the OvoA-like cysteine-binding motif

We next explored the mode of cysteine binding and noticed that recognition of this amino acid in the EgtB-IV_{Csu} ternary complex is divergent from other EgtB subfamilies. Broadly speaking, NHISS enzymes primarily bind their respective S/Se-donors through direct H-bonds with active site loops linking α -helices 1–3 of the DinB domain (α 1-2 and α 2-3; Figure S9). EgtB-IV_{Csu} is no exception. Our structure depicts the α -amino and carboxylate moieties of the substrate cysteine interacting with the side chain of R52 from α 1-2 and the backbone of V92 in α 2-3 (Figures 4A and 4B), while G91 forms an H-bonding interaction with the cysteine thiolate. Additional H-bonding contacts are observed between the cysteine carboxylate and the side chains of Y58 and T161, the latter of which is facilitated by a bridging water molecule. Both R52 and Y58 are highlighted by our composite motifs as key residues that are fully conserved within the subfamily along with other OvoA-like enzymes. However, this R–Y motif is radically different from the analogous S–D/H pattern observed in the EgtB-like branch (Figure 1C). To probe the importance of this motif in cysteine binding by EgtB-IV_{Csu}, we generated both single and double mutants that would be incapable of maintaining the associated substrate binding interactions: R52A-, Y58F-, and R52A/Y58F-EgtB-IV_{Csu}. The individual mutations led to nearly complete loss of activity, while the double mutant was fully inactive (Figure S5A).

Although the R–Y motif is clearly indispensable for cysteine binding by EgtB-IV_{Csu}, comparison of the EgtB-IV_{Csu}/OvoA_{Th2} and EgtB-I_{Mth}/EgtB-II_{Cth} (EgtB-II from *Chloracidobacterium thermophilum*) active sites highlights the diversity in S-donor recognition between the two primary NHISS branches (Figures 4C and S10). Interactions with the carboxylate moiety of cysteine-derived substrates are maintained in either active site configuration. However, in enzymes containing the S–D/H motif (EgtB-I/II), an Arg dyad is required to replace interactions with the substrate amino group. Analysis of sequence alignment data (Figure 1C)

reveals that this EgtB-like Arg dyad substitutes for the hercynine/histidine-binding Asp residue of the OvoA-like enzymes (D93 in EgtB-IV_{Csu}, D103 in OvoA_{Th2}). In fact, there is only one apparent commonality between the modes of thiol binding in OvoA-like and EgtB-like enzymes. Namely, a Thr residue (T161 in EgtB-IV_{Csu}) forms a water-mediated H-bond with the cysteinyl/cysteine carboxylate in type I and IV EgtBs that is also hinted at in the lower resolution structure of OvoA (Figures 4C and S10).

The α 2-3 loop serves as a flexible lid for the active site and dictates cysteine specificity

In spite of this diversity and the loop-rich nature of NHISS enzymes, previous structural investigations have demonstrated remarkable conservation of the fold across subfamilies and even upon substrate binding. All substrate-free and substrate-bound structures reported to date closely align, with root-mean-square deviations (RMSDs) of the peptide backbone ranging from 0.07 to 0.26 Å. In line with this trend, superposition of the ternary complex of EgtB-IV_{Csu} with the free enzyme yields an RMSD of 0.10 Å (over 1,564 atoms); however, the α 2-3 loop displays marked flexibility in the absence of cysteine (Figures S11A–S11C). We observe 50% higher average B-factors (B_{avg}) in this loop compared to the enzyme at-large, with many residues excluded from the model due to insufficient electron density. By contrast, the entire α 2-3 loop assumes a significantly stabilized, closed conformation that extends across the active site upon cysteine binding (Figure S11D). It, therefore, seems reasonable to suggest that the loop functions as a flexible lid controlling thiol access to the active site.

Supporting this hypothesis, the extended conformation of the α 2-3 loop in EgtB-IV_{Csu}/OvoA_{Th2} results in a smaller binding pocket, effectively excluding γ -Glu-Cys (Figures 4D–4F). Volume calculations using the CavitOmiX PyMOL plugin indicate that the active site of EgtB-IV_{Csu} is approximately 35% smaller than that of EgtB-I_{Mth}, which exclusively accepts γ -Glu-Cys (Figure 4F).¹¹ Correspondingly, no EgtB-IV_{Csu} or OvoA_{Th2} activity was observed with γ -Glu-Cys (Figure S12). A similar rationale can also be applied to explain the promiscuity of EgtB-II_{Cth}, which harbors a binding pocket large enough to accommodate either cysteine or γ -Glu-Cys (Figure 4F).²² Clearly, the α 2-3 loop in EgtB-IV_{Csu}/OvoA_{Th2} is essential for imparting cysteine specificity (Figure S13), and the structural and mutagenesis data define features that are important for binding this amino acid substrate in EgtB-IV_{Csu}.

Identification of a previously unclassified clade of OvoA-like enzymes

Our structural and mutagenesis results indicate that a significant amount of evolutionary remodeling must have occurred to facilitate the transition from histidine to hercynine specificity exhibited by OvoA and EgtB-IV, respectively. The evolutionary distance between the two subfamilies supports this conclusion. While not fully captured by our composite motifs, this divergence is reflected by both their domain structures and taxonomic distribution. OvoA harbors an additional C-terminal methyltransferase module responsible for installation of the π -N-methyl group during ovothiol biosynthesis that is unnecessary for ergothioneine production. Additionally, *egtB-IV* genes are found exclusively

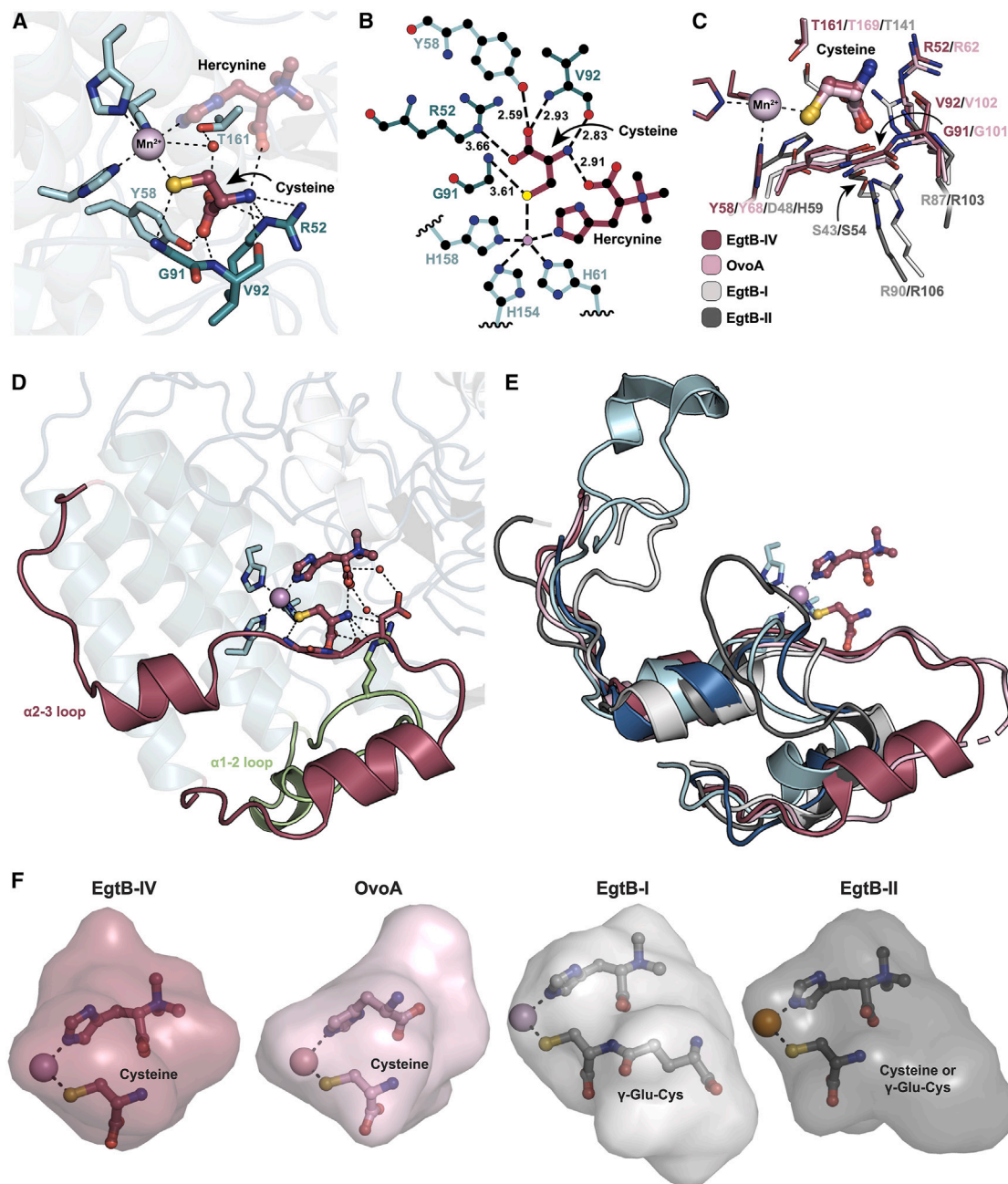


Figure 4. Active site loops in EgtB-IV facilitate substrate recognition and cysteine specificity

(A and B) H-bonding interactions for cysteine binding in EgtB-IV_{Csu}, shown as the (A) crystal structure of the EgtB-IV_{Csu} ternary complex and (B) 2D interaction diagram (visualizing only the direct H-bonds).
(C) Overlay of cysteine-bound EgtB-IV_{Csu} and OvoA_{Th2} with EgtB-I_{Mth} and EgtB-II_{Cth} (PDB: 6QKJ). EgtB-IV and OvoA employ an alternative cysteine binding mode characterized by a conserved R-Y motif, in place of an S-D/H motif and Arg dyad in EgtB-I/II.
(D) Both active site loops in EgtB-IV_{Csu} interact with hercynine and cysteine through direct and water-mediated H-bonds.
(E) Comparison of the α 2-3 loop in EgtB-IV_{Csu} (maroon), OvoA_{Th2} (light pink), EgtB-I_{Mth} (light gray), EgtB-II_{Cth} (dark gray), OvoA_{Hut} (dark blue), and SenA from *Variovorax paradoxus* (SenA_{Vpai}; light blue, PDB: 8K5I). The α 2-3 loop in EgtB-IV_{Csu} and OvoA_{Th2} adopts an extended conformation, functioning as a lid for the substrate-binding pocket.
(F) Surface representations of the active site pockets for EgtB-IV_{Csu}, OvoA_{Th2}, EgtB-I_{Mth}, and EgtB-II_{Cth}. The smaller binding pockets in EgtB-IV_{Csu} and OvoA_{Th2} exclude γ -Glu-Cys from binding.

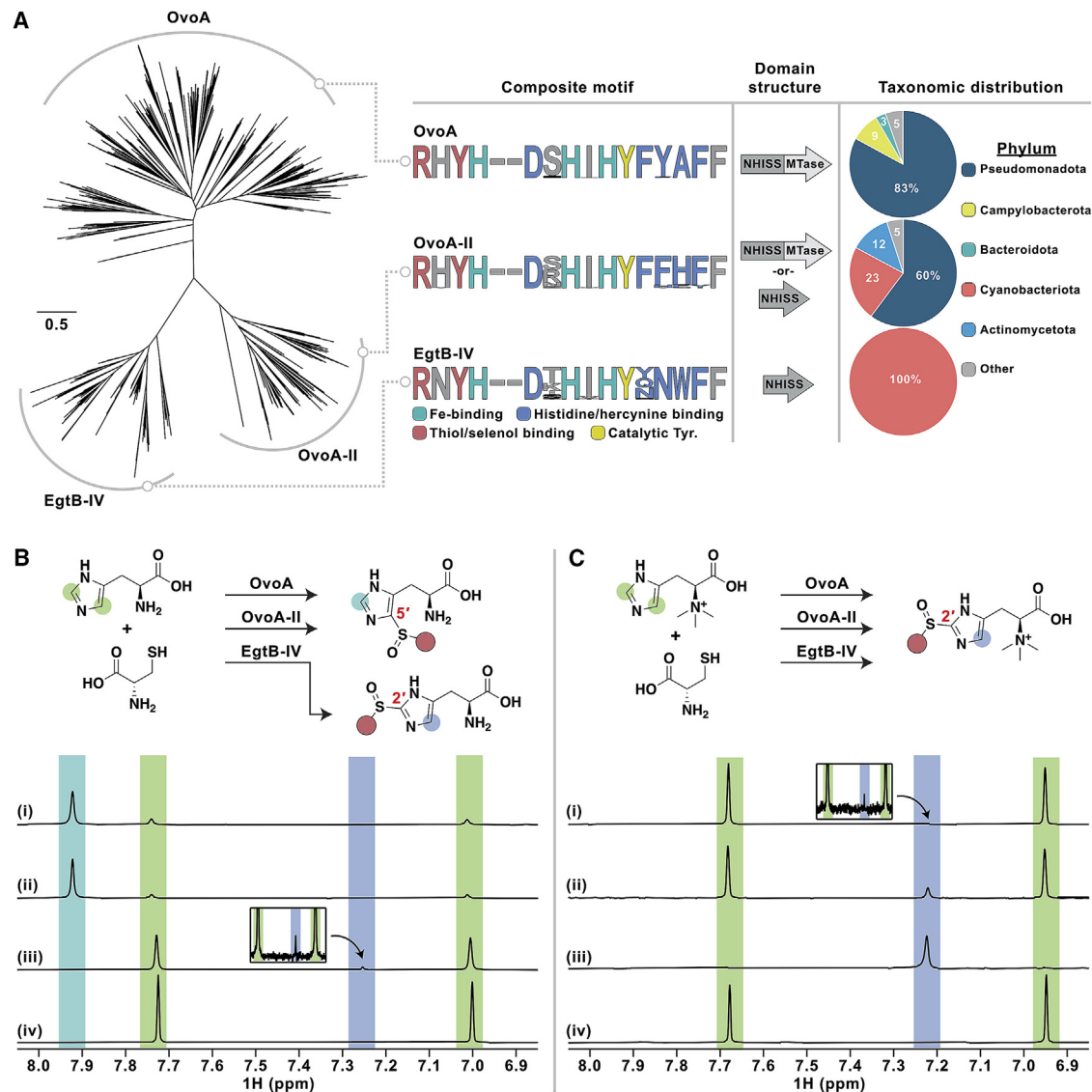


Figure 5. A previously unclassified clade of OvoA-like enzymes

(A) Left: unrooted phylogenetic tree of the OvoA-like branch of NHISS enzymes, now displaying an uncharacterized clade (provisionally termed OvoA-II) with similarities to OvoA and EgtB-IV. Right: characteristics of the three clades. The composite motif and taxonomic distribution of OvoA-II bears more resemblance to EgtB-IV, while its domain structure suggests OvoA activity.

(B) Product distributions for reactions of cysteine and histidine with (i) OvoA_{Th2}, (ii) OvoA-II_{Rar}, (iii) EgtB-IV_{Csu}, or (iv) no enzyme, monitored by ¹H NMR. Substrate imidazole protons are labeled green, while product imidazole protons are labeled teal or blue (2' or 5', respectively).

(C) Same as (B), except with mercaptine in place of histidine. Zoomed, inset spectra are shown for low intensity product resonances.

among cyanobacteria; no members of this phylum harbor an *ovoA* gene, as defined by our composite motifs.

Examining the phylogeny of as-yet-uncharacterized NHISS members within the OvoA-like branch yields additional insights into the evolution of this fascinating subfamily. In particular, we identify a previously unexamined clade that features a seemingly hybrid set of characteristics (Figure 5A). Most members of this group retain the OvoA dual-domain structure, suggesting OvoA activity, while exhibiting a composite motif and taxonomic distribution more similar to EgtB-IV. This observation suggests an early, cross-phylum transfer of an ancestral *ovoA*-like gene from a proteobacterium to a cyanobacterium, followed by diver-

gence over time into the extant clades. To gain more insight into this subfamily (provisionally termed type II OvoA), we expressed, purified, and assayed the activity of one of its members, harbored by *Rugosibacter aromaticivorans* (OvoA-II_{Rar}). As expected, upon incubation with histidine and cysteine, OvoA-II_{Rar} functions like a typical OvoA, furnishing the 5' sulfoxide product (Figure 5B). However, when presented with mercaptine, OvoA-II_{Rar} behaves more similarly to EgtB-IV, catalyzing 2' sulfoxidation much more efficiently than OvoA_{Th2} (Figure 5C). This heightened regiochemical flexibility is also consistent with the hypothesis that the EgtB-IV subfamily shares a more recent common ancestor with OvoA-II than with OvoA. While these

observations do not resolve the specific determinants of EgtB-IV regioselectivity, they imply that the answer lies somewhere along the complicated NHISS evolutionary trajectory.

DISCUSSION

In this study, we provide high-resolution crystal structures of two type IV EgtBs, ergothioneine-biosynthetic enzymes which are constituent members of the broader class of NHISS enzymes known for their production of redox-active thio/selenoimidazoles. Despite closest homology to the ovothiol-biosynthetic OvoA subfamily, EgtB-IV enzymes catalyze highly selective sulfoxidation at the imidazole 2' carbon of hercynine. Presented structures comparing the active site configurations of EgtB-IV with OvoA and EgtB-I/II showcase a striking instance of convergent evolution and illuminate the structural adaptations required for ergothioneine biosynthesis. While the overall architecture and active site of EgtB-IV bears a close resemblance to OvoA, it features the characteristic EgtB-like N-W/F/Y motif associated with hercynine binding. Mutation of these residues, however, has minimal impact on the enzyme's catalytic activity, as did our attempts to fully convert the EgtB-IV active site to that of OvoA. As neither bioinformatic analysis nor structure comparison provide a readily apparent explanation for this behavior, we hypothesize that the structural basis for substrate specificity in EgtB-IV may be unique among hercynine-binding NHISS enzymes.

Despite their different preferences for the imidazole-derived substrate, the EgtB-IV and OvoA subfamilies share a conserved mode of recognition for the S-donor, featuring an R-Y cysteine-binding motif unique to the OvoA-like sulfoxide synthases. Mutagenesis experiments validated these residues (R52 and Y58) as indispensable for cysteine binding in EgtB-IV_{Csu}. Moreover, the EgtB-IV homologs examined herein feature the OvoA-like active site lid loop that assumes a stabilized, extended conformation upon cysteine binding. Such a conformation facilitates both substrate recognition and cysteine specificity by forming a smaller binding pocket that excludes binding of γ -Glu-Cys. We propose that increased flexibility in this region that enables the sampling of different loop conformations may correlate with thiol substrate promiscuity.

This study fills a crucial gap in understanding ligand recognition and reactivity within NHISS members. The family's recent expansion to include selenoxide synthases suggests that we have only begun to uncover the diversity and complexity within this enzyme family. To this end, a comprehensive phylogenetic analysis revealed several unexplored subfamilies with divergent iron- and substrate-binding motifs, indicating the potential for unique reactivity and presenting exciting opportunities for future investigations.⁹ Furthermore, the catalytic mechanism of the recently discovered selenoxide synthases and the precise role of the catalytic Tyr residue remain open questions. We hope that our work stimulates further characterization of this emerging and intriguing class of enzymes along with their associated TSI antioxidants.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

● KEY RESOURCES TABLE

● RESOURCE AVAILABILITY

- Lead contact
- Materials availability
- Data and code availability

● EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

- Bacterial strains

● METHOD DETAILS

- Bioinformatics
- General experimental procedures
- EgtB-IV_{Csu} protein sequence
- EgtB-IV_{Gem} protein sequence
- OvoA-II_{Rar} protein sequence
- Protein expression and purification
- Enzyme activity assays
- Crystallization of EgtB-IV_{Csu}
- Crystallization of EgtB-IV_{Gem}
- X-ray data collection and processing

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.str.2024.08.006>.

ACKNOWLEDGMENTS

We thank the National Science Foundation (Graduate Research Fellowship Program No. 1937971 to K.A.I. and NSF CAREER Award No. 1847932 to M.R.S.), the Eli Lilly-Edward C. Taylor Fellowship in Chemistry (to C.M.K.), and the National Institutes of Health (grants R01 GM129496 to M.R.S. and R35 GM147557 to K.M.D.) for financial support. This research used resources of the Canadian Light Source, a national research facility of the University of Saskatchewan, which is supported by the Canada Foundation for Innovation (CFI), the Natural Sciences and Engineering Research Council (NSERC), the National Research Council (NRC), the Canadian Institutes of Health Research (CIHR), the Government of Saskatchewan, and the University of Saskatchewan. This work also includes research conducted at the Center for High-Energy X-ray Sciences (CHEXS), which is supported by the NSF (BIO, ENG, and MPS Directorates) under award DMR-1829070, and the Macromolecular Diffraction at CHESS (MacCHESS) facility, which is supported by award 1-P30-GM124166-01A1 from the National Institute of General Medical Sciences, NIH and by New York State's Empire State Development Corporation (NYSTAR).

AUTHOR CONTRIBUTIONS

K.A.I. and C.M.K. designed the research; K.A.I., C.M.K., and M.L.A. conducted the experiments; K.A.I., C.M.K., and M.L.A. analyzed the data; K.A.I., C.M.K., M.R.S., and K.M.D. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 21, 2024

Revised: July 16, 2024

Accepted: August 6, 2024

Published: August 30, 2024

REFERENCES

1. Dunbar, K.L., Scharf, D.H., Litomska, A., and Hertweck, C. (2017). Enzymatic carbon-sulfur bond formation in natural product biosynthesis. *Chem. Rev.* **117**, 5521–5577.
2. Steele, A.D., Kiefer, A.F., and Shen, B. (2023). The many facets of sulfur incorporation in natural product biosynthesis. *Curr. Opin. Chem. Biol.* **76**, 102366–102374.

3. Halliwell, B., Cheah, I.K., and Tang, R.M.Y. (2018). Ergothioneine - a diet-derived antioxidant with therapeutic potential. *FEBS Lett.* **592**, 3357–3366.
4. Cheah, I.K., and Halliwell, B. (2012). Ergothioneine; antioxidant potential, physiological function and role in disease. *Biochim. Biophys. Acta* **1822**, 784–793.
5. Fu, T.T., and Shen, L. (2022). Ergothioneine as a natural antioxidant against oxidative stress-related diseases. *Front. Pharmacol.* **13**, 850813.
6. Cordell, G.A., and Lamahewage, S.N.S. (2022). Ergothioneine, ovothiol A, and selenoneine—histidine-derived, biologically significant, trace global alkaloids. *Molecules* **27**, 2673.
7. Castellano, I., and Seebeck, F.P. (2018). On ovothiol biosynthesis and biological roles: from life in the ocean to therapeutic potential. *Nat. Prod. Rep.* **35**, 1241–1250.
8. Alhasan, R., Nasim, M.J., Jacob, C., and Gaucher, C. (2019). Selenoneine: a unique reactive selenium species from the blood of tuna with implications for human diseases. *Curr. Pharmacol. Rep.* **5**, 163–173.
9. Kayrouz, C.M., Ireland, K.A., Ying, V.Y., Davis, K.M., and Seyedsayamdost, M.R. (2024). Discovery of the selenium-containing antioxidant ovoselenol derived from convergent evolution. *Nat. Chem.* <https://doi.org/10.1038/s41557-024-01600-2>.
10. Braunschauen, A., and Seebeck, F.P. (2011). Identification and characterization of the first ovothiol biosynthetic enzyme. *J. Am. Chem. Soc.* **133**, 1757–1759.
11. Seebeck, F.P. (2010). In vitro reconstitution of mycobacterial ergothioneine biosynthesis. *J. Am. Chem. Soc.* **132**, 6632–6633.
12. Kayrouz, C.M., Huang, J., Hauser, N., and Seyedsayamdost, M.R. (2022). Biosynthesis of selenium-containing small molecules in diverse microorganisms. *Nature* **670**, 199–204.
13. Liao, C., and Seebeck, F.P. (2017). Convergent evolution of ergothioneine biosynthesis in cyanobacteria. *ChemBiochem* **18**, 2115–2118.
14. Cooper, D.R., Grelewski, K., Kim, C.Y., Joachimiak, A., and Derewenda, Z.S. (2010). The structure of DinB from *Geobacillus stearothermophilus*: a representative of a unique four-helix-bundle superfamily. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 219–224.
15. McMahon, S.A., Miller, J.L., Lawton, J.A., Kerkow, D.E., Hodes, A., Marti-Renom, M.A., Doulatov, S., Narayanan, E., Sali, A., Miller, J.F., and Ghosh, P. (2005). The C-type lectin fold as an evolutionary solution for massive sequence variation. *Nat. Struct. Mol. Biol.* **12**, 886–892.
16. Carlson, B.L., Ballister, E.R., Skordalakes, E., King, D.S., Breidenbach, M.A., Gilmore, S.A., Berger, J.M., and Bertozzi, C.R. (2008). Function and structure of a prokaryotic formylglycine-generating enzyme. *J. Biol. Chem.* **283**, 20117–20125.
17. Chen, L., Naowarajna, N., Song, H., Wang, S., Wang, J., Deng, Z., Zhao, C., and Liu, P. (2018). Use of a tyrosine analogue to modulate the two activities of a nonheme iron enzyme OvoA in ovothiol biosynthesis, cysteine oxidation versus oxidative C–S bond formation. *J. Am. Chem. Soc.* **140**, 4604–4612.
18. Goncharenko, K.V., and Seebeck, F.P. (2016). Conversion of a non-heme iron-dependent sulfoxide synthase into a thiol dioxygenase by a single point mutation. *Chem. Commun.* **52**, 1945–1948.
19. Goncharenko, K.V., Vit, A., Blankenfeldt, W., and Seebeck, F.P. (2015). Structure of the sulfoxide synthase EgtB from the ergothioneine biosynthetic pathway. *Angew. Chem. Int. Ed. Engl.* **54**, 2821–2824.
20. Stampfli, A.R., Goncharenko, K.V., Meury, M., Dubey, B.N., Schirmer, T., and Seebeck, F.P. (2019). An alternative active site architecture for O₂ activation in the ergothioneine biosynthetic EgtB from *chloracidobacterium thermophilum*. *J. Am. Chem. Soc.* **141**, 5275–5285.
21. Liu, M., Yang, Y., Huang, J.W., Dai, L., Zheng, Y., Cheng, S., He, H., Chen, C.C., and Guo, R.T. (2024). Structural insights into a novel nonheme iron-dependent oxygenase in selenoneine biosynthesis. *Int. J. Biol. Macromol.* **256**, 128428–128434.
22. Naowarajna, N., Irani, S., Hu, W., Cheng, R., Zhang, L., Li, X., Chen, J., Zhang, Y.J., and Liu, P. (2019). Crystal structure of the ergothioneine sulfide synthase from *candidatus chloracidobacterium thermophilum* and structure-guided engineering to modulate its substrate selectivity. *ACS Catal.* **9**, 6955–6961.
23. Wang, X., Hu, S., Wang, J., Zhang, T., Ye, K., Wen, A., Zhu, G., Vegas, A., Zhang, L., Yan, W., et al. (2023). Biochemical and structural characterization of OvoA(Th2): a mononuclear nonheme iron enzyme from *Hydrogenimonas thermophila* for ovothiol biosynthesis. *ACS Catal.* **13**, 15417–15426.
24. Song, H., Leninger, M., Lee, N., and Liu, P. (2013). Regioselectivity of the oxidative C–S bond formation in ergothioneine and ovothiol biosyntheses. *Org. Lett.* **15**, 4854–4857.
25. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2 — approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490.
26. Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132.
27. Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214.
28. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589.
29. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501.
30. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221.
31. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315.
32. Schrodinger, L.L.C. (2023). The PyMOL Molecular Graphics System, version 2.5.5.
33. Laskowski, R.A., and Swindells, M.B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786.
34. Innophore GmbH (2022). CavitOmiX V. 1.0.
35. Naowarajna, N., Huang, P., Cai, Y., Song, H., Wu, L., Cheng, R., Li, Y., Wang, S., Lyu, H., Zhang, L., et al. (2018). In vitro reconstitution of the remaining steps in ovothiol A biosynthesis: C–S lyase and methyltransferase reactions. *Org. Lett.* **20**, 5427–5430.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
<i>Escherichia coli</i> DH5 α	New England Biolabs	Cat# C2987I
<i>Escherichia coli</i> BL21(DE3)	New England Biolabs	Cat# C2527I
Chemicals, peptides, and recombinant proteins		
NdeI restriction endonuclease	New England Biolabs	Cat# R0111S
XhoI restriction endonuclease	New England Biolabs	Cat# R0146S
HiFi DNA Assembly Master Mix	New England Biolabs	Cat# E2621L
His-tagged-EgtB-IV _{Csu}	This study	N/A
His-tagged-EgtB-IV _{Gem}	This study	N/A
His-tagged-EgtB-IV _{Csu} variants	This study	N/A
His-tagged-OvoA _{Th2}	This study	N/A
His-tagged-OvoA-Il _{Rar}	This study	N/A
His-tagged-OvsA _{Hut}	Kayrouz et al. ⁹	N/A
His-tagged-OvoB _{Eta}	This study	N/A
Histidine	Fisher Scientific	Cat# AAA1041322
Hercynine	Ambeed	Cat# A1463528
Cysteine	Fisher Scientific	Cat# AAA1038922
γ -glutamyl-cysteine	MedChemExpress	Cat# HY-113402
<i>N</i> -acetyl-1-seleno- β -D-glucosamine	Kayrouz et al. ⁹	N/A
Monobromobimane	MilliporeSigma	Cat# 596105
<i>N</i> - α -dimethyl-histidine	Chem-Impex	Cat# 09542
PEG 3350	Fisher Scientific	Cat# 18-605-319
Ammonium acetate	Sigma-Aldrich	Cat# 631-61-8
Deposited data		
EgtB-IV _{Csu} ·Fe crystal structure	This study	PDB: 8VIK
EgtB-IV _{Csu} ·Fe·hercynine crystal structure	This study	PDB: 8VIH
EgtB-IV _{Csu} ·Mn·hercynine·cysteine crystal structure	This study	PDB: 8VII
EgtB-IV _{Csu} ·Fe· <i>N</i> - α -dimethyl-histidine crystal structure	This study	PDB: 8VIL
EgtB-IV _{Gem} ·Fe·hercynine crystal structure	This study	PDB: 8VIG
OvsA _{Hut} ·Fe·histidine crystal structure	Kayrouz et al. ⁹	PDB: 8U41
EgtB-Il _{Mth} ·Fe·hercynine crystal structure	Goncharenko et al. ¹⁹	PDB: 4X8E
EgtB-Il _{Cth} ·Fe·hercynine crystal structure	Stampfli et al. ²⁰	PDB: 6QKJ
SenA _{Vpa} ·Fe·hercynine·thioglucose crystal structure	Liu et al. ²¹	PDB: 8K5I
OvoA _{Th2} ·Co·histidine·cysteine crystal structure	Wang et al. ²³	PDB: 8KHQ
Recombinant DNA		
pET28b vector	MilliporeSigma	Cat# 69865
pET28b-His6x-EgtB-IV _{Csu}	This study	N/A
pET28b-His6x-EgtB-IV _{Gem}	This study	N/A
pET28b-His6x-EgtB-IV _{Csu} variants	This study	N/A
pET28b-His6x-OvoA _{Th2}	This study	N/A
pET28b-His6x-OvoA-Il _{Rar}	This study	N/A
pET28b-His6x-OvsA _{Hut}	Kayrouz et al. ⁹	N/A
pET28b-His6x-OvoB _{Eta}	This study	N/A
Software and algorithms		
FastTree	Price et al. ²⁵	http://www.microbesonline.org/fasttree/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FigTree	Rambaut Lab, University of Edinburgh	http://tree.bio.ed.ac.uk/software/figtree
MassHunter Software Suite	Agilent	https://www.agilent.com/en/product/software-informatics/mass-spectrometry-software
MestReNova	Mestrelab Research	https://mestrelab.com/download/mnova/
XDS	Kabsch ²⁶	https://xds.mr.mpg.de
AIMLESS	Evans and Murshudov ²⁷	https://www.ccp4.ac.uk/html/aimless.html
AlphaFold2	Jumper et al. ²⁸	https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
Coot	Emsley et al. ²⁹	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Phenix	Adams et al. ³⁰	https://phenix-online.org/
Molprobit	Williams et al. ³¹	http://molprobit.biochem.duke.edu/
PyMOL	Schrodinger, LLC ³²	http://www.pymol.org
LigPlot+	Laskowski and Swindells ³³	https://www.ebi.ac.uk/thornton-srv/software/LigPlus/
CavitOmiX	Innophore GmbH ³⁴	https://innophore.com/cavtomix/
Other		
HisPur Ni-NTA resin	ThermoFisher	Cat# 88222
Sephadex G-25 medium resin	Cytiva	Cat# 17003301
HiPrep 16/60 Sephacryl S-200 HR column	Cytiva	Cat# 17116601
Cary 60 UV-Visible Spectrophotometer	Agilent	Cat# G6860A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Katherine M. Davis (katherine.davis@emory.edu).

Materials availability

Protein expression plasmids are available upon request. All unique/stable reagents generated in this study are available from the [lead contact](#) with a completed Materials Transfer Agreement.

Data and code availability

The atomic coordinates and structure factors for the EgtB-IV structures have been deposited in the Protein DataBank and are publicly available as of the date of publication. The accession numbers for the structures are PDB: 8VIG, 8VIH, 8VII, 8VIK, 8VIL, which are also listed and linked in the [key resources table](#). This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Bacterial strains

Escherichia coli (E. coli) BL21 (DE3) cells were used for producing the recombinant proteins in this study. The cells were cultured in standard media (LB broth) at 37°C. The [method details](#) section gives more comprehensive details about protein expression and purification.

METHOD DETAILS

Bioinformatics

NHISS family protein sequences were retrieved and composite motifs were compiled as previously described.⁹ Phylogenetic trees were constructed by first performing multiple sequence alignments, which were then used as input to FastTree²⁵ for maximum-likelihood phylogeny approximation with parameters -gamma, -mlacc 2, and -slow. Trees were visualized using FigTree (<http://tree.bio.ed.ac.uk/software/figtree>). The tree in [Figure 1C](#) consists of five random homologs from each of the six characterized subfamilies.

The initial bifurcation between the OvoA-like (OvoA, EgtB-IV) and EgtB-like (EgtB-I, EgtB-II, SenA, OvsA) branches was selected as the tree root. The tree in Figure 5 is unrooted and encompasses all bacterial OvoA and EgtB-IV sequences, as well as 181 OvoA-like sequences that feature a unique composite motif, provisionally termed the OvoA-II subfamily.

General experimental procedures

High-performance liquid chromatography-coupled mass spectrometry (HPLC-MS) was performed on an Agilent instrument equipped with a 1260 Infinity Series HPLC, an automated liquid sampler, a photodiode array detector, a JetStream ESI source, and the 6540 Series Q-tof mass spectrometer. HPLC-MS data were acquired and analyzed with Agilent MassHunter software. Solvents for all HPLC-MS experiments were water +0.1% formic acid (Solvent A) and MeCN +0.1% formic acid (Solvent B). Analytes were separated on a Synergi Fusion-RP column (Phenomenex, 100 × 4.6 mm, 4 μm) with a flow rate of 0.5 mL/min and an elution program consisting of 5% solvent B for 3 min, followed by a gradient of 5–95% solvent B over 7 min. NMR spectra were collected at the Princeton University Department of Chemistry's NMR facility on a Bruker Avance III 500 MHz NMR spectrometer equipped with a DCH double resonance cryoprobe. NMR data were acquired using Bruker Topspin software and processed using MestReNova software.

EgtB-IV_{Csu} protein sequence

NCBI identifier: WP_009547064.1

MGSSHHHHHHSSGLVPRGSHMSFIKSQLPILNCTQDSVINYFQNSWELENILMRSIIDETFYINPDPLRNPLIFYLGHSAAFYINKLIR
VELLEKGINSDYEILFEFGVDPENAEELNQAIAHINWPDVRQVWDYRNKAYEVILEVIKNTTFDLPIHASHPLWALMMGMEHQRIHFETSS
MLLRQLPTEKVEKPQGWQYAPSQGVNTNKMILVEGGTVTLGAKDNPLYGWDCEYGDRLVKVDSFFASQYLVNNGEFLFINRKGYE
TQSYWNEKSWQWKEENKVKNPFWQFNNGKYSYRAMFDEIPLDWPVEVNYEAMAYCGWKKGKTRLMSEAENLAAYGSNDN
YQVDIEKVNDYNLNLKFGSPSPVGLVKTAQSHSGLWDLRGNVWEWLDENFHLPLPGFEPHFLYEDNSAPFFDNNHKMMLGGAWVTQG
TETLKYYRNWFRPNFYQHAGFRIVTNH

EgtB-IV_{Gem} protein sequence

NCBI identifier: MCS6943588.1

MGSSHHHHHHSSGLVPRGSHMTMEYSLPLNSCDREQILSYFEESWWKEDCLFNSIKKEEIFYTNPDPLRNPLIFYLGHSVAFYINKMRR
AGMIKESINEGYEEMAYAVGVDPENAEQLREKIVKIKWDRVEEVWDYRKRAYEKIREAIENTSLDLPITEENPWWSVIMGIEHQRIHIETSSM
LIRQVEEKWLEKPSGWYASTRGVNPSQEMVKVEGGRVIRGRDRNDNYGWDVDFGKKEVEVKDFWVSKYLVNNGEFLRFVEEGGYE
NPEYWHEEGWIWKEENGKVKHPKFWGKRGEEGYRYRLMFEEVELPLDFPVEVSLYEAMAYCRYLGGRDGCNYRLMTEGEWHLASRK
EGEKGEDYNLNFYHSPTVPGSMREARSDSGVYDCRGNVWEWLGEKLPLEGFTTHYLYEDYSAPFFDDNHLLIGGSWASSGHSAS
RFYRNWFRPFYFYQHAGFRLVLAEL

OvoA-II_{Rar} protein sequence

NCBI identifier: WP_237218172.1

MGSSHHHHHHSSGLVPRGSHMNPPSTIAVRENEPWRANLRGNQTALAGARDETWWTGKSPVACPGLTAEGTLTALPLPNLATCTRR
EVQDYFDSGWTLTETLFAGLKGEEAFFRPPYHGLRHPMVFYFCHPPVFTYTNKLVRAGLISAPINPYFEKLFVGVDEMWRDDMSKNEM
LWPSVDEVRAAYAAVYTTVSEIATHPGLAEGHALITLNDPLWALFMGFEHERIHLETSVLIHELPLDYVQRPAAWPALYPAAPHDASTA
LRPRPGTDYPVLALRDVAATRTLKGKPADWPSYGDWNEYGSRSVRVEAFRASTQLISNGAFHEFVAAGGYREERFWSPTGWAWRSYR
NIKWPFTFWVPDGPAGLHRYKLRTLFEVIDMPWNWPAEVNGHEAYAYCQWKSEQDGTTYRLPSENEHHALRDPVHRVTDLALQCDG
LAFRQRGINLNLAYGSSQPVDAHITSAGFHDVFGNVWQWSADDFNPLPGSEVHPLYDDFSSPCYDGGHQMMMLGGAWVSTGEEAGP
WARFHFRPHFLFHAGFRLVEAAHDGGAHLNSDDAAHQVYEDRQMVNDYMLLHYGSPADQMPHAFGPQAATEFPLRCAGCSMVRQ
SLVLALSVRWKLVALSAGQVSSWHVITK.

Protein expression and purification

Genes encoding EgtB-IV_{Csu}, EgtB-IV_{Gem}, EgtB-IV_{Csu} variants, OvoA_{Th2},²³ OvoA-II_{Rar}, OvsA_{Hut},⁹ and OvoB from *Erwinia tasmaniensis* (OvoB_{Eta})³⁵ were obtained as synthetic DNA fragments, codon-optimized for expression in *E. coli* with overhangs to allow assembly into pET28b(+). Protein expression plasmids were assembled from gene fragments and vector pET28b(+), linearized with NdeI and XhoI (NEB), using HiFi DNA Assembly Master Mix (NEB) following the manufacturer's instructions. Ligation mixtures were transformed into chemically competent *E. coli* DH5α by heat-shock and plated onto LB agar containing 50 mg/L kanamycin. After confirmation by Sanger sequencing, assembled plasmids were transformed into *E. coli* BL21(DE3) for protein expression.

For protein expression, starter cultures were prepared by inoculating 15 mL of LB medium containing 50 mg/L kanamycin with a single colony of *E. coli* BL21(DE3) carrying the desired plasmid. After overnight growth at 37°C/200 rpm, starters were used to inoculate 2.8 L baffled Fernbach flasks containing 1.5 L Luria-Bertani (LB) broth supplemented with 50 mg/L kanamycin (1% inoculum) and incubated at 37°C/200 rpm. At OD₆₀₀ = 0.5–0.6, protein expression was induced with 0.2 mM isopropyl β-D-1-thiogalactopyranoside, and cultures were incubated at 18°C/200 rpm for an additional 12–24 h. Cells were pelleted by centrifugation (8,000 g, 15 min, 4°C). Cell paste was stored at –80°C until purification.

All purification steps were carried out in a cold room at 4°C. Cells were resuspended in lysis buffer (5 mL/g cell paste), which consisted of 25 mM Tris-HCl (pH 8), 500 mM NaCl, 10 mM imidazole, 10% glycerol, supplemented with 1 mM phenylmethylsulfonyl fluoride. Once homogeneous, 0.1 mg/mL deoxyribonuclease I (Alfa Aesar) was added, and the cells were lysed by the addition of

5 mg/mL lysozyme followed by sonication using 30% power (~150 W) in 15 s on/15 s off cycles for a total of 4 min. This process was repeated three times. The lysate was then clarified by centrifugation (30,000 g, 1 h, 4°C) and loaded onto a 5 mL Ni-NTA column pre-equilibrated in lysis buffer. The column was washed with lysis buffer and His-tagged proteins were eluted with elution buffer consisting of 25 mM Tris-HCl (pH 8), 500 mM NaCl, 300 mM imidazole, 10% glycerol. Eluted proteins were then buffer-exchanged using a 50 mL column of Sephadex G-25 (Cytiva) into storage buffer consisting of 50 mM Tris-HCl (pH 8), 150 mM NaCl. For crystallography, EgtB-IV_{Csu} and EgtB-IV_{Gem} were further purified by fractionation on a Sephacryl S-200 HR HiPrep 16/60 column (Cytiva) with a running buffer consisting of 50 mM Tris-HCl (pH 8), 150 mM NaCl. Purified proteins were stored at –80°C. Protein concentrations were determined spectrophotometrically on a Cary 60 UV-visible spectrophotometer (Agilent) using calculated molar extinction coefficients at 280 nm.

Enzyme activity assays

Assays were carried out in 100 μL reactions containing 50 mM sodium phosphate buffer (pH 8), 4 mM dithiothreitol (DTT), and 2 mM sodium ascorbate. Histidine, hercynine, thiol (cysteine or γ-Glu-Cys), and/or selenol (*N*-acetyl-1-seleno-β-D-glucosamine, SeGlcNAc)¹² substrates were added to a final concentration of 1 mM each. Enzymes were added to a final concentration of 10 μM. Reactions were incubated at room temperature.

For variant activity assays, reactions were quenched after 1 h by 30 kDa MWCO centrifuge filter to remove the enzyme. OvoB_{Eta} (C-S lyase) was then added to a final concentration of 10 μM to facilitate conversion to ergothioneine or 5-thiohistidine. After 2 h, samples were derivatized with 50 μL of 10 mM monobromobimane (mBBR) in MeCN. Reactions were incubated for an additional 30 min at room temperature in the dark to allow for complete derivatization with mBBR. The samples were then filtered and analyzed by HPLC-MS. Extracted ion counts of mBBR-ergothioneine (*m/z* = 420.2 Da) or mBBR-5-thiohistidine (*m/z* = 378.1 Da) were recorded.

For EgtB-IV_{Csu} and OvoA_{Th2} thiol preference assays, reactions were quenched after 1 h by addition of an equal volume of MeOH. The samples were then filtered and analyzed by HPLC-MS. Extracted ion counts of sulfoxide products were recorded.

HPLC-MS assays of the OvsA reaction with *N*-α-dimethyl-histidine (DMH) were carried out as described above, except mBBR was added directly following the 1 h reaction period, without OvoB_{Eta} addition. To confirm the regioselectivity, a 700 μL reaction was incubated for 1 h, followed by enzyme removal by 30 kDa MWCO centrifuge filter, addition of 10% D₂O, and analysis by ¹H NMR.

To determine the activity of OvoA-II_{Rar}, 700 μL reactions were prepared as described above and incubated for 6 h, followed by enzyme removal by 30 kDa MWCO centrifuge filter. For each reaction, 600 μL of the flowthrough was lyophilized, redissolved in 120 μL D₂O, and analyzed by ¹H NMR. Control reactions with either EgtB-IV_{Csu}, OvoA_{Th2}, or no enzyme were prepared and analyzed in the same fashion.

Crystallization of EgtB-IV_{Csu}

For crystallization, enzymes were stored in 50 mM Tris (pH 8), 150 mM NaCl. To generate the holo-form of each enzyme, two molar equivalents (eq.) of (NH₄)₂Fe(SO₄)₂ were added prior to all crystallization trials, excluding those of the unreactive Mn-bound ternary complex. Crystals of EgtB-IV_{Csu} were grown using the sitting drop vapor diffusion method at room temperature. In the absence of substrate, EgtB-IV_{Csu} (10.5 mg/mL) was mixed 1:1 with a solution of 0.1 M Bis-Tris (pH 6.1), 25% PEG 3350, 0.18 M ammonium acetate to induce nucleation. To obtain a structure of the enzyme bound to DMH, 30 mM DMH was added prior to mixing with precipitant. By contrast, crystals of EgtB-IV_{Csu} bound to hercynine required slightly modified conditions in which 10 mM hercynine was added to EgtB-IV_{Csu} at 11 mg/mL, before mixing 1:1 with a precipitant solution of 0.1 M Bis-Tris (pH 5.9), 23% PEG 3350, 0.2 M ammonium acetate. Structures containing both cysteine and hercynine were acquired by first removing any bound iron by incubating the protein solution overnight with 5 mM EDTA, followed by buffer exchange to remove the chelating agent. The protein solution was then degassed in an ultrasonic water bath for 15 min before being brought into an anaerobic chamber. The sample was subsequently left to exchange for an additional 6 h during which 10 mM sodium dithionite was added to consume any residual O₂. Manganese was then introduced (5 mM MnCl₂) in place of iron to prevent unintended turnover, along with 30 mM cysteine and 10 mM hercynine before mixing 1:1 with a precipitant solution of 0.1 M Bis-Tris (pH 6.1), 21% PEG 3350, 0.2 M ammonium acetate.

In all cases, the production of large, single crystals required subsequent streak seeding, after which thin plates appeared within 4 days and were fully formed within 2–4 weeks. For harvesting, crystals were looped and briefly transferred into cryoprotectant before flash freezing in liquid N₂. Cryoprotectants were composed of the precipitant and 20–25% (v/v) ethylene glycol, along with any presumed ligands (10–30 mM hercynine, 30 mM DMH, and 30 mM cysteine).

Crystallization of EgtB-IV_{Gem}

Crystals of EgtB-IV_{Gem} were grown via similar protocols. To obtain crystals of EgtB-IV_{Gem} bound to hercynine, 20 mM of the substrate was added to 21 mg/mL of EgtB-IV_{Gem}. The protein was then mixed 1:1 with a precipitant solution of 0.1 M Bis-Tris (pH 6.7), 23% PEG 3350, 0.2 M MgCl₂. Thin plates appeared within 4 days and were fully formed within 1–2 weeks. For crystal harvesting, the crystals were looped and briefly transferred into cryoprotectant (composed of the precipitant with 30% (v/v) ethylene glycol) before flash freezing in liquid N₂.

X-ray data collection and processing

All crystals were maintained at 100 K to minimize X-ray-induced damage during data collection. Data for hercynine-bound EgtB-IV_{Csu} and EgtB-IV_{Gem} were collected sequentially ($\Delta\phi = 0.2^\circ$) with an incident wavelength of 0.9537 Å using an Eiger X 9M (Dectris) at the

CMCF beamline 08ID of the Canadian Light Source. By contrast, diffraction images for substrate-free and DMH-bound EgtB-IV_{Csu}, as well as the associated hercynine·cysteine ternary complex were collected sequentially ($\Delta\phi = 0.25^\circ$) with an incident wavelength of 0.9686 Å using an Eiger2 16M at beamline ID7B2 of the Cornell High Energy Synchrotron Source. All data were indexed, integrated, and scaled with XDS before merging with *AIMLESS*.^{26,27} Structures of hercynine-bound EgtB-IV_{Csu} and EgtB-IV_{Gem} were solved first via molecular replacement implemented in *PHASER*. The AlphaFold2-predicted structure of EgtB-IV_{Csu} was used as the search model,²⁸ following removal of all residues with per-residue confidence scores (pLDDT) below 90%. All subsequent structures were solved via isomorphous replacement after transferring and truncating the R_{free} flags from the hercynine-bound EgtB-IV_{Csu} structure. Additional model building was conducted in Coot,²⁹ the structures refined in Phenix,³⁰ and model quality assessed using Molprobity.³¹ All EgtB-IV_{Csu} crystals were in the monoclinic P 1 2₁ 1 space group and contain two molecules in the asymmetric unit. Crystals of EgtB-IV_{Gem}, by contrast, formed in the orthorhombic P 2₁ 2₁ 2₁ space group and contain one molecule in the asymmetric unit. Figures depicting the structures were generated with PyMOL and visualize chain A from all structures.³² The 2-D interaction diagram for cysteine binding in Figure 4B was generated with LigPlot+.³³ The substrate-binding pockets in Figures 4F and S13 were calculated using CavitOmiX.³⁴

QUANTIFICATION AND STATISTICAL ANALYSIS

Data processing and refinement statistics for structural data reported in this article are summarized in Table 1.