

The Diversity of the Iron-dependent Sulfoxide Synthases in Ergothioneine Biosynthesis

A comparative approach to disentangle the biology and evolution of ergothioneine.

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Anja Renée Stampfli

von Flumenthal und Balm bei Günsberg, SO und Neuseeland

2022

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel

edoc.unibas.ch

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Seebeck, Florian
Prof. Dr. Blankenfeldt, Wulf

Basel, 17 Dezember 2019

Prof. Dr. Martin Spiess

He rises and begins to round,
He drops the silver chain of sound,
Of many links without a break,
In chirrup, whistle, slur and shake...

For singing till his heaven fills,
'Tis love of earth that he instils,
And ever winging up and up,
Our valley is his golden cup,
And he the wine which overflows
To lift us with him as he goes...

Till lost on his aerial rings
In light, and then the fancy sings.

- George Meredith

I. Abstract

The understanding of any biological system must start with a detailed understanding of the chemistry of the individual enzymes, their catalytic mechanisms and their structures. Ergothioneine, a small molecular weight thiol, is an ideal target for molecular dissection, due its ubiquity in life and complex and enigmatic physiological function. EgtB, an iron dependent sulfoxide synthase, was discovered through this approach, and represents an entirely new catalyst type, that is distinct in both reactivity and structure from other iron oxygenases. This makes dissection of its mechanistic details an attractive endeavor. As EgtB is also the central and characteristic enzyme of the oxidative pathway for ergothioneine biosynthesis, exploration of its reactivity and evolutionary history may shed light on questions pertaining to the emergence of ergothioneine.

EgtB catalyzes oxidative carbon sulfur bond formation between the C2 carbon of *N,N,N*- α -Trimethyl-histidine and the sulfhydryl group of cysteine or γ -glutamyl-cysteine to form a hercynine-(γ -glutamyl)cysteine-sulfoxide conjugate in a four electron oxidation that is coupled to the reduction of molecule oxygen to water. Despite having been the focus of numerous biochemical, bioinformatic and computational studies, the evolutionary history and catalytic mechanism of EgtB are respectively unknown and disputed.

This thesis tackles both questions through an approach which involves the characterization of divergent EgtB homologues. Crystal structures of these homologues, compounded with kinetic and/or bioinformatic characterization revealed that the EgtB family is characterized by extreme active site diversity. This diversity manifests itself in changes in catalytic residues, substrates and perhaps even reactivity, all of which can be assigned to a particular sequence motif. These differences provide a platform to explore the mechanism of EgtB *via* comparative enzymology, and allowed us to explore possible evolutionary routes to the diverse EgtBs. The crystal structures alone provide a valuable test of any mechanistic proposal. We envisage this work will drive mechanistic discussions and further exploration of the EgtB sequence space to capture the full diversity of this family and consequently ergothioneine biosynthesis.

An identical strategy was leveraged in the discovery, characterization and identification of substrate determinants of solute-binding proteins involved in an ergothioneine transport system. This molecular basis was compounded with a bioinformatic approach to reveal several key insights into ergothioneine utilisation and the evolutionary history of the ergothioneine transporter system. This allows us to identify organisms that utilize ergothioneine that do not produce or degrade it, dramatically increasing the number of organisms to which we know ergothioneine plays a role in. This is yet another example of how unravelling the molecular basis of ergothioneine reveals that ergothioneine is more prevalent in nature than previously thought, and is likely an important molecule to many life forms.

II. Table of Contents

I. Abstract	vii
II. Table of Contents	ix
III. Abbreviations	xi
1. Introduction: Diversity in Ergothioneine Biosynthesis	1
Aim of this thesis	27
Synopsis	29
2. An Alternative Active Site Architecture for O ₂ Activation in the Ergothioneine Biosynthetic EgtB from <i>Chloracidobacterium thermophilum</i>	33
Addendum on <i>CthEgtB</i> Substrate Selectivity	51
Supplementary Information	57
3. The Role of Oligomerisation and Loop Folding in a type II EgtB	77
Addendum on <i>CthEgtB</i> Ser92	93
Supplementary Information	98
4. The Structure of a Sulfoxide Synthase Homologue implicated in Carbon-Selenium Bond Formation reveals insights into Selenium Activation	123
Supplementary Information	139
5. The Structural Characterization of Ergothioneine Solute Binding Proteins	149
Supplementary Information	185
Conclusion and Outlook	209
References	213
Acknowledgements	233
Curriculum vitae	237

III. Abbreviations

γ -GC	γ -glutamyl cysteine
•OH	Hydroxyl radical
ABC	ATP-binding cassettes
ADP	Adenosine diphosphate
AfuEgt-1	EgtB from <i>Aspergillus fumigatus</i>
ATP	Adenosine triphosphate
CDO	Cysteine dioxygenase
Cth-Egt1	EgtB from <i>Chaetomium thermophilum</i>
CthEgtB	EgtB from <i>Chloracidobacterium thermophilum</i>
CthEgtB.CF	Cysteine-free EgtB from <i>Chloracidobacterium thermophilum</i> (CthEgtB.C37A.C231F.C303S.C391S.C401A)
Cys	Cysteine
DMH	<i>N,N</i> - α -diimethylhistidine
DSF	Differential scanning fluorimetry
EDTA	Ethylenediaminetetraacetic acid
EGT	Ergothioneine
EGTSO ₃ ⁻	Ergothioneine sulfonic acid
esd	Estimated standard deviation
EtaOvoA	OvoA from <i>Erwinia tasmaniensis</i>
FGE	Formylglycine-generating enzyme
G1-SP	Glucose-1-thiophosphate
GB	Glycine betaine
GDP	Guanidine diphosphate
GNT	Genome Neighborhood Tool
GOE	Great oxidation event
GT	Glycosyl transferase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOCl	Hypochlorous acid
HRMS	High resolution mass spectrometry
IPR007210	Interpro family 007210 - ABC-type glycine betaine transport system - substrate-binding domain
ITC	Isothermal titration calorimetry
KIE	Kinetic isotope effect
KSIE	Kinetic solvent isotope effect
LB	Lysogeny broth
LMCT	Ligand-to-metal charge transfer
MaeEgtB	EgtB from <i>Microcystis aeruginosa</i>

MeEGT	S-methyl Ergothioneine
MeTyr	3-methoxytyrosine
MthEgtB	EgtB from <i>Mycobacterium thermoresisibile</i>
MtTyr	2-amino-3-(4-hydroxy-3-(methylthio)phenyl) propanoic acid
MWM	Molecular Weight Marker
NcaEgtB	EgtB from <i>Neurospora crassa</i>
OCNT1	Human ergothioneine transporter
ONOO	Peroxynitrite
PB	Proline Betaine
PCET	Proton-coupled electron transfer
PDB	Protein data bank
PLP	Pyridoxal phosphate
ProX _{archeal}	ProX from <i>Archeoglobus fulgidus</i>
ProX _{bacterial}	ProX from E coli
rmsd	root-mean-square deviation
SAM	S-Adenosyl methionine
SBP	Solute binding proteins
SBP13	Solute binding protein from <i>Cohnella kolymensis</i> (WP_041060754.1)
SBP14	Solute binding protein from <i>Trabulsiella odontotermitis</i> (WP_049857012.1)
SBP4	Solute binding protein from <i>Bacillus sp. FJAT-27264</i> (WP_066368978)
SBP5	Solute binding protein from <i>Marteella endophytica</i> WP_045679504.1
SDS	Sodium dodecyl sulfat
SEC	Size Exclusion Chromatography
SeCys	Selenocysteine
SHE	Standard hydrogen electrode
SpoEgtB	EgtB from <i>Schizosaccharomyces pombe</i>
SSN	Sequence similarity network
SvaGT	Glycosyl transferase from <i>Sulfurifustis variabilis</i>
TCEP	Tris(2-carboxyethyl)phosphine
TcuOvoA	OvoA from <i>Trypanosoma cruzi</i>
TdeErgothionase	Ergothionase from <i>Treponema denticola</i>
TelEgtB	EgtB from <i>Thermosynechococcus elongatus</i>
TMD	Transmembrane domain
TMH	N,N,N- α -trimethylhistidine
UDP	Iridine diphosphate
UniRef	UniProt Reference Clusters
Vpa2032	Selenokinase from <i>Variovorax paradoxus</i> (WP_012747182.1)
Vpa2053	Glycosyl transferase from <i>Variovorax paradoxus</i> (ACS18697.1)
Vpa2054	Sulfoxide synthase from <i>Variovorax paradoxus</i> (WP_012747184.1)

1. Introduction: Diversity in Ergothioneine Biosynthesis

Enzymes and Life

Enzymes are nature's catalysts, and are essential for life. A single cell contains thousands of molecules, which are rapidly transformed in a series of chemical reactions to produce energy and to build functionally-important molecules.¹⁻² Each chemical step is facilitated by a specific enzyme. Enzymes are extraordinary catalysts, speeding up chemical reactions by orders of 5 to 17 in magnitude, and doing so with astonishing specificity, readily discriminating between substrates with similar structures.³ Without these rate enhancements life could not exist. For instance, a key reaction in heme and chlorophyll biosynthesis is the decarboxylation of uroporphyrinogen. In the absence of an enzyme, the decarboxylation reaction proceeds with a half-life of 2.3 billion years, at 25 °C.⁴ The human enzyme uroporphyrinogen decarboxylase, catalyses this reaction with a k_{cat} of 0.16 s^{-1} , providing a rate enhancement by a factor of 1.2×10^{17} .⁴⁻⁵ While this is one of the largest rate enhancements known for an enzyme catalysed reaction, it clearly demonstrates the tremendous catalytic abilities of enzymes and that without enzymes, the time scale at which uncatalyzed reactions occur could not support life as we know it.

The structures, functions, and mechanisms of enzymes are the products of millions of years of evolution. The molecular mechanisms by which an enzyme performs these remarkable rate enhancements involves a precisely orchestrated sequence of steps. These systems have also evolved to be carefully controlled, with numerous modes of regulation existing to facilitate the rapid response to metabolic flux and changing cellular conditions. Evolution of nature's catalysts has directly determined which chemical moieties are present and abundant in nature. Furthermore, a given chemical function group is only present and relevant in life if an enzyme evolved to catalyse its formation and degradation, thereby cementing the importance and centrality of enzymes to the evolution of life and life itself.

Enzymatic activity has been recognised in living cells, such as the fermentation process in yeast cells, since the 1700's.⁶ However, it was not until 1897 that Buchner was able to show that enzymatic activity can occur independently of whole cells and the protoplasm, establishing the foundations for the field of enzymology.⁷ The biochemical identity of enzymes was established by Sumner, who in 1926 isolated and crystallised urease, demonstrating that enzymes are proteins, and can be purely protein, a notion which was at the time controversial.⁸ RNA was later discovered to also have catalytic capabilities.⁹ A fundamental realisation in the transformation of enzymology into a rigorous science was the recognition that fundamental chemical principles apply to enzymatic catalysis. Key breakthroughs include the derivation of a kinetic model by Michaelis and Menten, and the application of transition state theory to enzyme catalysis by Pauling.¹⁰⁻¹¹ These major pieces of work demonstrate the fundamental importance of chemical principles in the dissection of enzymology and biology. As Silverman states, "Enzymes are highly efficient organic chemists".¹² Any mechanistic proposal requires a theoretical foundation, highlighting the value of chemical proficiency in the study of enzymes.

A chemical perspective. Enzyme function can be viewed and studied from either a chemical or biological perspective. For example, to a chemist, a protease catalyzes the hydrolytic cleavage of an amide bond either through catalytic Bronsted or Lewis acid mechanisms, while to a biologist, a protease enables the digestion of food and regulates processes such as blood clotting or apoptosis.³ While both viewpoints are incredibly valuable, this thesis explores biology and enzymology from a chemical perspective. As described hereafter, the application of chemical intuition to a biological system allows us to make advancements in three key facets; (1) biological or medicinal relevance (2) evolution (3) applications to engineering biocatalysts.

Enzymes and biological function. At the heart of all biological processes are chemical reactions. Understanding the biology of a given molecule is highly dependent on a profound understanding of the involved enzymology; not only the discovery of the responsible enzymes, but disentanglement of their structures and mechanistic details. The identification of components and interacting partners alone is not sufficient to fully describe and understand the processes of complex biology. Understanding of any biochemical process must start with a detailed understanding of the chemistry, of the individual enzymes, their mechanistic modes of action and their structures. Incorporation of this detailed molecular basis into a larger framework, which includes interconnected metabolic pathways and whole organisms, can provide a mechanistic understanding of cells, organisms, their regulation and at a very basic level, even life.

Enzymes and evolution. The ability of an organism to adapt to environmental conditions is essential to guarantee their survival. At a metabolic level, this process of adaptation originates from the ability of an enzyme to evolve beneficial functions in response to an environment of changing chemical conditions.¹³ To understand how such a change has come about, we must explore evolution. Fundamental questions regarding the evolution of a biological system can be addressed at the molecular level through enzymology. Such an approach is indispensable to connect changes at the molecular level of nucleic acids and proteins to changes within a biochemical system and the underlying chemical and physical origins of this change. Study of enzyme evolution may provide insight into selection pressures that leads to a particular reactivity or feature and may provide even wider lessons on the fundamental evolutionary processes that have shaped life.

Enzymes and biotechnological applications. Understanding the evolutionary processes that produce new enzymes is not only important to probe molecular function and biological function, but also for the generation of tailor-made enzymes. However, the ability to manipulate enzyme function requires a thorough mechanistic understanding of the chemistry occurring and an understanding of sequence-structure-function relationships. Principles learnt about control of reactivity of an enzyme can be applied to purposes in the biotechnology or pharmaceutical industries. Greater lessons and rules for how nature utilises and manipulates chemistry can be learnt.

Ergothioneine

A combination of basic chemical principles and mechanistic enzymology can provide important insights into the behavior of biological systems. The small molecular weight thiol, ergothioneine (EGT) (**1**), is an ideal target to which mechanistic enzymology can be applied to learn more about the biology, evolution and biotechnological applications related to this molecule and its biosynthesis. The *in vivo* role of EGT is complex, and still under investigation, and as will become apparent, the enzymes involved in ergothioneine biosynthesis have a rich and complex evolutionary history. Additionally, some of the EGT biosynthetic enzymes represent the discovery of enzymes with unique reactivities, highlighting novel approaches utilized by nature to perform complex chemistry. Therefore the application of chemical intuition to the biology of ergothioneine will provide us with a wealth of information.

Ergothioneine. EGT (**1**) is a thiourea derivative of histidine betaine (Figure 1). Since its discovery in 1909, EGT has been the subject of hundreds of studies regarding its chemical and physiological properties, abundance in nature and potential role in human health.¹⁴⁻¹⁵ Considerable antioxidant and cryoprotective properties of EGT have been demonstrated *in vitro*, yet an *in vivo* role is yet to be established.¹⁶

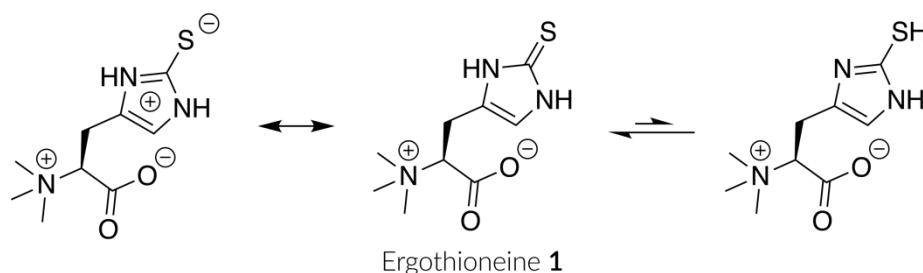


Figure 1. Structure of EGT, and the tautomeric equilibrium between the predominant thione and thiol form. A zwitterionic resonance form of EGT, with a negative charge on the thiol group and positive charge delocalized on the imidazolium ring, is also shown.

Chemical Properties. Two unique properties distinguish EGT from other low molecular weight thiols. The first is that of two tautomers that can exist, the thione, rather than the thiol, is the predominant structure under physiological conditions.¹⁷ EGT also has a higher standard redox potential (-0.06 V) when compared to other naturally occurring thiols, which typically range between -0.2 and -0.32 V (SHE).¹⁸ These two features confer a high stability towards oxidation to EGT compared to other low molecule weight thiols.¹⁴

Ergothioneine and Humans. EGT is ubiquitous in nature, appearing in the cells and tissues of most plants and mammals, despite their inability to synthesize EGT.¹⁹ In humans, a EGT specific transporter (OCNT1) has been found to facilitate uptake and assimilation via diet.²⁰ Ergothioneine has been demonstrated to accumulate at high concentrations in cells and tissues, between 100 μ M and 2 mM, with particularly high concentrations found in erythrocytes, bone marrow, liver, kidney, seminal fluid and the lens and cornea of eyes.^{14, 21-24}

Antioxidant properties. EGT has been demonstrated to be a powerful scavenger of hydroxyl radicals ($\bullet\text{OH}$), hypochlorous acid (HOCl) and peroxynitrite (ONOO) in *in vitro studies*, favoring a role as an antioxidant.²⁵⁻²⁸ Many have noted that the distribution of ergothioneine in human tissues and cells advocates for an antioxidant role, with a predisposition to accumulate in locations with high levels of oxidative stress.¹⁴ *In vivo* antioxidant properties have also been reported. The silencing of OCTN1 has been shown to increase oxidative burden in mammalian cells.²⁹ One particular study demonstrated that EGT decreased the amount of oxidative damage in rats following administration with ferric-nitrilotriacetate, a Fenton chemistry catalyst.³⁰ It should however be noted that the oxidative stress imposed on the rats was severe, and is not representative of normal stress conditions. While this shows the potential of EGT to respond in such a way, further research is required to further illuminate the physiological role of EGT. Despite this mounting evidence, the *in vivo* role still remains a complex question.

To contribute to the understanding of EGT, we are closely analyzing the biochemistry of its biosynthesis. We aim to have a detailed chemical understanding of the individual enzymes, their mechanistic modes of action and their structures. What can we learn from analysing the distribution of ergothioneine production among life? Which species produce EGT? How do these species produce EGT and what are the molecular details of how these enzymes work? How are these enzymes and pathways regulated? Can we shed light on the evolutionary emergence of EGT biosynthesis by studying the enzymes and genes involved? What is the evolutionary history and emergence of the individual enzymes? What selection pressures influenced their evolution? For organisms which do not appear to produce ergothioneine - can they obtain ergothioneine by other means? What happens in organisms that lack EGT? We aim to leverage mechanism, structures and distribution to learn about the biology, evolution and biotechnological applications of EGT and its biochemistry. This chemical perspective and mechanistic enzymology basis will enable the incisive dissection of the biology of EGT.

Ergothioneine Biosynthesis

To date, several EGT biosynthetic pathways have been identified and characterized.³¹⁻³³ Exploration of a fungal biosynthetic pathway began in the 1970's, which led to the identification of pathway intermediates and hypotheses about the responsible enzymes.³⁴⁻³⁶ This provided a platform for the identification of the mycobacterial pathway, which enabled the discovery of alternative pathways in other prokaryotes and in fungi through comparative genomics.^{32, 37-38}

Initial Findings. The key elements of EGT biosynthesis were first identified in *Neurospora crassa* cell-free extracts in the 1970's. Preliminary work with *N. crassa* cells determined that *N,N,N*- α -trimethyl histidine (TMH) (**2**) is a key intermediate in EGT biosynthesis.³⁴ Cell-free extracts of *N. crassa* were then further demonstrated to convert histidine to TMH in the presence of *S*-Adenosyl methionine (SAM). Purification of the associated enzyme provided evidence that a single enzyme is responsible for all three methylation events to form TMH.³⁵ In the presence of iron(II) and molecular oxygen a sulfoxide intermediate (**4**) was identified and characterized from cell-free extracts, implicating an iron-dependent oxygenase that utilizes TMH and *L*-cysteine (cys). The sulfoxide was then converted to EGT (**1**) and pyruvate by a pyridoxal phosphate (PLP) dependent enzyme.³⁶ These results implicated a SAM-dependent methyltransferase, an iron oxygenase, and PLP dependent β -lyase as ergothioneine biosynthetic enzymes. However no genetic assignment was accomplished at the time.

A five-gene cluster in Mycobacteria. In 2010 the genes for EGT production in mycobacteria were identified, corresponding to a five-gene cluster designated *egtABCDE* (Figure 2A).³¹ EgtD encodes for a SAM-dependent methyl transferase, catalyzing trimethylation of the α -amino group of histidine.³⁹ No functional description was available for *egtB*, yet the gene sequence indicated similarity to a formylglycine-generating enzyme (FGE) and a DinB_2 domain. Upon reconstitution with iron(II), a TMH-sulfoxide conjugate (**3**) was formed with a γ -glutamyl cysteine (γ -GC) sulfur donor (**5**), instead of cys as predicted by Melville.³⁶ EgtA corresponds to an adenosine triphosphate (ATP) dependent ligase producing the γ -GC used by EgtB. The glutamyl-moiety is cleaved off the EgtB sulfoxide product (**4**) by EgtC, which in a final step is converted to EGT by the PLP dependent EgtE (Figure 2C).³¹

Fungal Biosynthesis. Assignment of the ergothioneine biosynthetic genes in mycobacteria allowed for identification of the first fungal biosynthetic gene, Egt-1, in the filamentous fungi *N. crassa* and fission yeast *Schizosaccharomyces pombe* through sequence homology.^{37, 40} In both *N. crassa* and *S. pombe* deletion of Egt-1 was found to abolish ergothioneine production. Egt-1 corresponds to the fusion of a mycobacterial-like EgtD with an EgtB, indicating that Egt-1 is a bifunctional enzyme catalyzing the trimethylation of histidine and subsequent formation of a sulfoxide using cys as the sulfur source (Figure 2B). Use of cys is consistent with preliminary work by Melville and eliminates the need for EgtA or EgtC, reducing EGT biosynthesis to a three step - two gene pathway (Figure 2C, blue enzymes).³⁷ This premise was corroborated through

recombinant production and biochemical characterization of the *N. crassa* Egt-1 gene, which was shown to have both EgtD and EgtB activity.³³ The sulfoxide synthase activity of Egt-1 from another filamentous fungi, *Chaetomium thermophilum*, has also recently been demonstrated.⁴¹ The second gene in the fungal EGT biosynthetic pathway, Egt-2, was identified as a PLP-dependent C-S lyase, responsible for transforming the sulfoxide intermediate to EGT, which in *N. crassa* is encoded independently of Egt-1.³³ Deletion of Egt-2 in *S. pombe* resulted in the detection of small amounts of EGT. This is consistent with previous reports that the sulfoxide intermediate can spontaneously convert to ergothioneine.^{31, 40} The distribution of Egt-1 in fungi was analyzed, and was found to be present in over 400 diverse prokaryotic species. In a further survey of 100 fungal genomes, Egt-1 is present in all phyla with the exception of a subphyla of Saccharomycotina.⁴² This bioinformatic analysis is valuable as it demonstrates that the ability of fungi to produce EGT is much more prevalent than previously reported.⁴²⁻⁴³

Three step Ergothioneine Biosynthesis in Bacteria. As with fungi, the prevalence and distribution of EGT biosynthesis in bacteria was extremely underestimated before the annotation of the mycobacterial EGT biosynthetic genes. Earlier reports stated that only bacteria belonging to the order Actinomycetales could produce EGT.⁴³ While the five-gene EGT cluster identified in mycobacteria is limited to actinobacteria, numerous other bacteria are now known to produce EGT, instead utilizing a shorter three-step three-gene biosynthetic pathway.^{42, 44} This shorter pathway, like the fungal, employs a cys-utilizing EgtB.⁴⁴ The bacterial and fungal cys-utilizing EgtBs are, however, not closely related, having different evolutionary origins. The evolutionary relationship of EgtB homologues will be discussed in detail in chapter two, in conjunction with the structural and biochemical characterization of a cys-utilizing EgtB.

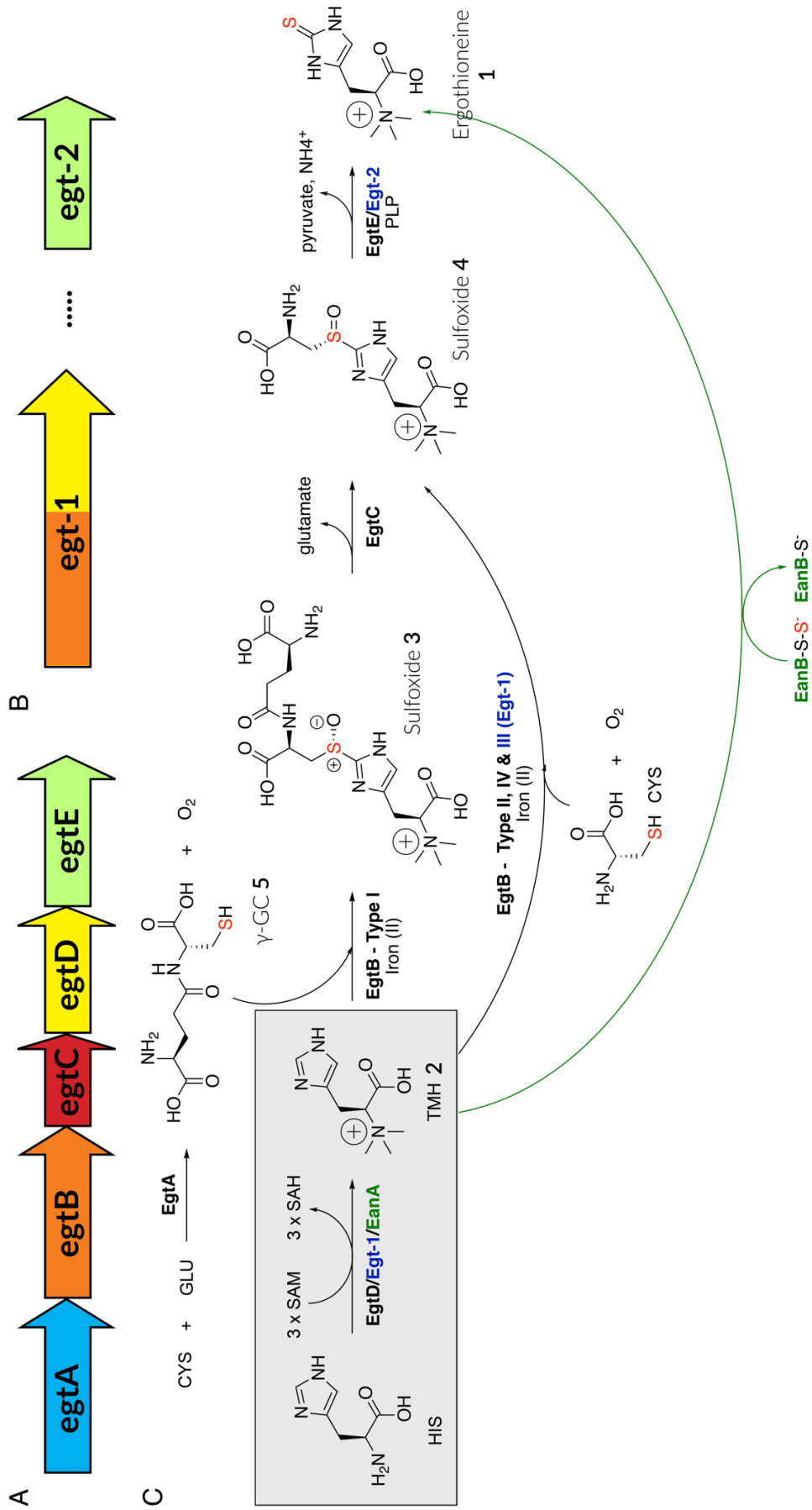


Figure 2. A. The mycobacterial EGT gene cluster codes for a γ-GC synthetase (EgtA), an FGE-like protein (EgtB), a glutamine amidotransferase (EgtC), a methyltransferase (EgtD), and a PLP-binding protein (EgtE). B. Fungal EGT biosynthesis is encoded for by two genes, Egt-1, an EgtB-EgtD fusion, and Egt-2, a C-S lyase. C. Reaction sequence of the various EGT biosynthetic pathways. The enzymes involved in fungal biosynthesis are blue, while the anaerobic pathway and enzymes are in green

Anaerobic Ergothioneine Biosynthesis. In the ergothioneine biosynthesis pathways discussed thus far, EgtD is commonly co-encoded with, if not fused together, with an EgtB.^{31,42} An EgtD homologue (Clim_1148) was identified in the anaerobic green sulfur bacterium *Chlorobium limicola*, yet, this genome does not encode for any EgtB homologue. EgtD like activity of Clim_1148 (EanA) was verified, inferring involvement in a distinct metabolic pathway. Clim_1148 (EanA) is co-encoded with a rhodanese-like sulfur transferase (EanB). EanB was found to have sulfur transfer reactivity transferring a sulfur atom from a cys derived persulfide onto the C2 position of the imidazolium ring of TMH to form ergothioneine (Figure 2C, green pathway).³² Structural and kinetic characterization of EanB revealed that C-S bond formation is mediated by the nucleophilic attack of an active-site persulfide anion onto the TMH imidazolium ring.⁴⁵ This is first example of a rhodanese-like enzyme that can substitute a C-H bond with a C-S bond, presenting a novel catalytic strategy for enzymatic C-S bond formation. This also represents the discovery of an entirely novel EGT biosynthetic that is independent of oxygen. The evolution of an alternative pathway alludes to the importance of EGT across numerous life forms. In addition, the discovery of an anaerobic biosynthesis for EGT that appears in strict anaerobes contradicts the general hypothesis that EGT is a physiologically-relevant antioxidant. This works suggests that EGT may have several functions and that the current view of EGT physiology is too narrow.³²

A second insight comes from the discovery of this pathway, with respect to the evolutionary history of the enzymes involved and emergence of EGT. Planetary life originated in an anoxic environment. Approximately 2.4 billion years ago, oxygenic photosynthesis emerged in ancestral cyanobacteria, which as a byproduct of phototrophic water oxidation, began to produce substantial amounts of O₂. The appearance of O₂ must have led to significant oxidative stress for the anaerobic organisms. How these organisms, particularly those that were the first to experience intracellular dioxygen, were able to overcome the imposing oxidative stress is of great interest. It has been postulated that these early oxygenic organisms likely adapted to oxidative stress by adopting pre-existing systems with serendipitous antioxidant properties.⁴⁶ EGT may have been such a molecule. Phylogenetic analysis of the EgtD family supports this hypothesis by indicating that ergothioneine may be an ancient molecule. EgtD is the common factor in the anaerobic and aerobic biosynthetic pathways, indicating that one is likely to have evolved from the other. Phylogenetic analysis of numerous EgtD and EanB enzymes found that the two types share little sequence diversity and sort into two distinct clades. The distinct separation of methyl transferases involved in the anoxic pathway from those utilized in the aerobic pathway indicates that the divergence between the types was not a recent evolutionary event, indicating that EGT is an old molecule.³² It is possible that ergothioneine is so old that its first appearance may predate the great oxidation event (GOE). This suggests that EGT plays a role in anoxic biochemistry, perhaps providing defense mechanisms against physical and chemical stresses associated with anaerobic life. Following the GOE, EGT proved beneficial against oxidative stress and was repurposed as an anti-oxidant. Such an evolutionary history is consistent with the broad distribution of ergothioneine in many ecological niches, including both anaerobic and aerobic life forms. These findings suggest that EGT can support cellular life in numerous ways, by virtue of having numerous functions.

Diversity in Ergothioneine Biosynthesis. One remarkable feature of EGT biosynthesis is the huge amount of diversity observed in its biosynthetic pathway. To date, three different EGT biosynthetic pathways are

known, which have emerged in evolution at least three times independently.^{31-32,47} As will become apparent throughout this thesis, many of the biosynthetic enzymes also exhibit great diversity, more than is typically observed for metabolic enzymes. In particular, the EgtB scaffold from aerobic ergothioneine appears to permit a wide range of perturbations for the bifurcation of substrates, active site, catalytic residues and perhaps even reactivity, as will be outlined.

The multiple emergences of EGT in the evolutionary history of life, and the diversity of these pathways indicates that 1) pathways for EGT biosynthesis must evolve easily and 2) EGT must serve an important physiological role. This observation raises the intriguing question as to the origin of this extreme diversity. One possible explanation is that this diversity stems from the response to various evolutionary pressures. Theoretical reasons for adaptation include (1) a need to overcome a limited substrate, for instance O₂ or thiol donor (2) a change in environmental conditions or (3) a change in product. An alternative theory to explain the extreme diversity is neutral drift or convergent evolution, the latter of which is discussed in chapter two. These reasons would only be plausible if the biosynthetic pathways have a long evolutionary history to allow enough time to accumulate diversity via neutral drift. This is consistent with the suggestion that EGT is an ancient molecule, the pathways of which have been modified and repurposed numerous times in the domains of life.

EgtB

The discovery of the various EGT biosynthetic pathways revealed numerous uncharacterized enzymes with fascinating chemistry. Of these enzymes, EgtB is particularly exciting as its discovery represents an entirely new catalyst type that is distinct in both reactivity and structure from other iron oxygenases. This makes dissection of its mechanistic details an attractive endeavor to understand general principles of reactivity and for the generation of novel biocatalysts. EgtB is also the central and characteristic enzyme of the oxidative pathway, therefore exploration of its reactivity and evolutionary history may shed light on questions pertaining to the emergence of ergothioneine.

EgtB catalyzes oxidative carbon sulfur bond formation between the C2 carbon of TMH and the sulfhydryl group of cysteine or γ -GC to form a hercynine-(γ -glutamyl)cysteine-sulfoxide conjugate in a four electron oxidation coupled to the reduction of molecule oxygen to water. The stereochemistry of the bacterial sulfoxide product was determined to be the *S*-isomer in a crystal structure of the preceding enzyme, EgtC.⁴⁸ EgtB was only active upon reconstitution with iron(II), while incubation with ethylenediaminetetraacetic acid (EDTA) absolved any activity, establishing the iron dependence of EgtB.³¹ The discovery of EgtB and functional assignment as a sulfoxide synthase marks the discovery of a novel class of iron mono-oxygenases. This new class is distinct from other known oxygenases in both reactivity and structure, raising questions as to the catalytic mechanism of EgtB. The evolutionary emergence of EgtB, from the fusion of an FGE-like domain with a DinB_2 fold is also an intriguing facet, which will be explored in this thesis.

Since the discovery of EgtB in 2010, this peculiar sulfoxide synthases class has been the focus of numerous biochemical, bioinformatic and computational studies in efforts to lineate the evolutionary history and catalytic mechanism. The empirical biochemical observations will be presented herein first. These results, together with computational considerations will be discussed in the context of mechanistic proposals, in the following section.

***MthEgtB* Crystal Structure.** The first EgtB crystal structure was solved in 2015 of the EgtB from *Mycobacterium thermoresistibile* (*MthEgtB*).⁴⁹ Structures revealed that EgtB is comprised of two domains connected by a long linker region (Figure 3A). The N terminal domain is comprised of a DinB_2 like fold: a four-helix bundle arranged in an up-down-down-up conformation, with long linkers between each of the helices. An extended two-stranded beta sheet commences the second C-terminal domain, wrapping around a C-type lectin fold, which together make up an FGE-like fold. This fold lacks many secondary structural elements, containing numerous loops that are stabilized by buried ionic interactions. The active site is located at the interface of the two domains in a deep and wide active site tunnel. At the bottom of the tunnel, three histidine residues from the DinB domain co-ordinate the catalytic iron in a facial coordination mode. In the native crystal structure three water molecules are also bound to Iron to give a six-coordinate, octahedral geometry (Figure 3B).⁴⁹

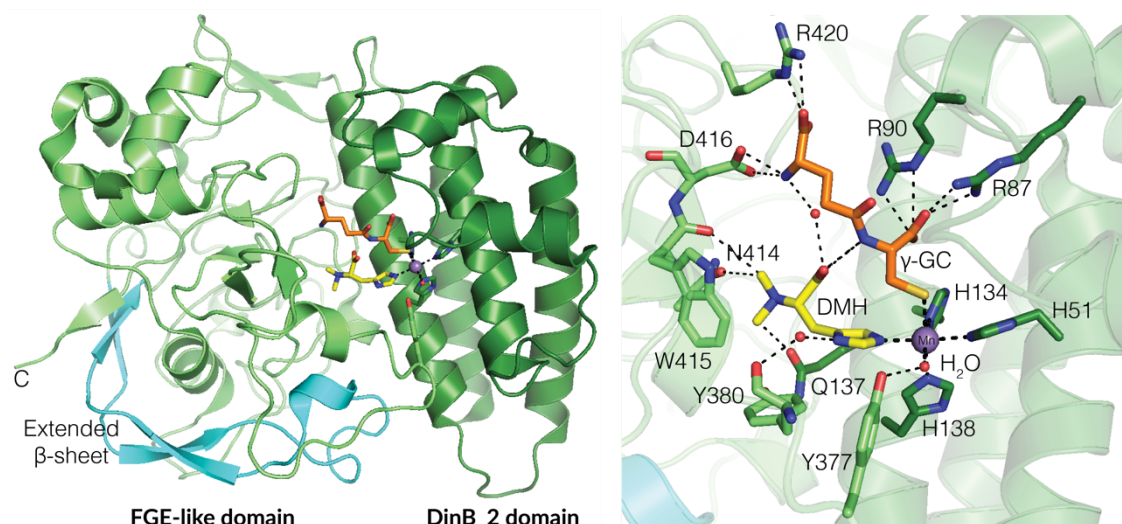


Figure 3. *MthEgtB* crystal structure containing Mn, *N,N*- α -dimethylhistidine (DMH) and γ -GC (PDB: 4X8D).⁴⁹ A. Monomeric structure, shown in cartoon form with the DinB_2 domain in dark green, extended β -sheet in blue and C-lectin fold in light green. The active site is shown as sticks. B. *MthEgtB* Active Site, with substrates DMH and γ -GC, and important substrate binding residues shown as sticks. Proposed interactions are indicated by the dotted lines.

Substrate Binding. In the ternary complex structure of *MthEgtB* containing iron and TMH, TMH coordinates directly to the catalytic iron, replacing one of the water molecules in the native structure, binding via the $N\tau$ of the imidazole ring. The other nitrogen of the substrate imidazole ring, $N\pi$, forms a water-mediated hydrogen bond to Tyr380. Two of the *N*- α -methyl groups from the quaternary amine stack against Trp415, while the third methyl group makes dipolar interactions to the side chains of Gln137 and Asn414. To probe the γ -GC binding, crystals were grown with Mn and soaked with *N,N*- α -dimethylhistidine (DMH) and γ -GC. DMH sits in superimposable position to TMH. γ -GC also directly co-ordinates to iron via its sulfhydryl group, replacing a second water molecule in the native structure. The amide group of γ -GC hydrogen bonds to the carboxylate of TMH, while the carboxylate group of the cysteine moiety can form salt bridges to Arg87 and Arg90, while the amino and carboxylic group of the glutamyl moiety form salt bridges to Asp416 and Arg420 respectively. The former interaction was probed by mutation of Asp416 to an Asparagine. The K_M for γ -GC increased 200-fold, while the K_M for TMH or k_{cat} did not significantly change, providing evidence that Asp416 is important for substrate recognition. In this quaternary structure, one water molecule is left coordinating to iron. This is likely the site of oxygen binding. A narrow tunnel connects the EgtB exterior to the proposed oxygen-binding site, which has been suggested as a tunnel for oxygen transport. This water molecule hydrogen bonds to the phenolic side chain of Tyr377, which points into the active site from the C-terminal domain.⁴⁹

***MthEgtB.Y377F*, a dioxygenase.** A single point mutation of Tyr377 to phenylalanine converted the *MthEgtB* into a γ -GC-dioxygenase, switching the dominant activity from oxidative C-S bond formation to di-oxygenation.⁵⁰ The efficiency of *MthEgtB.Y377F* at catalyzing di-oxygenation rivals that of native oxygenases.⁵¹⁻⁵² The presence of TMH is required for di-oxygenation. The K_M for TMH and γ -GC and rate of

γ -GC consumption were unchanged by the mutation, indicating that reduced sulfoxide synthase activity arises solely due to a reduced k_{cat} and not from impaired substrate binding or oxygen activation. This highlights Tyr377 as a key catalytic residue for oxidative C-S bond formation. A kinetic isotope effect (KIE) near unity was measured for both WT and the Y377F mutant when C2-deuterated TMH was used, indicating that hydrogen or proton removal is not an essential function of Tyr377. The remaining sulfoxide synthase activity of the *MthEgtB*.Y337F mutant displays a kinetic solvent isotope effect (KSIE) of 1.9 ± 0.1 and pH dependence, while the di-oxygenation and WT sulfoxidation pathways have a KSIE near unity. This indicates that the transfer of at least one proton is important for sulfoxidation, and this step becomes rate-limiting upon mutation of Tyr377 to phenylalanine. However, protonation is not important for the di-oxygenation pathway.⁵⁰

Selenocysteine, an EgtB substrate and inhibitor. Enzymes involved in sulfur metabolism do not typically discriminate against selenium. However, another cys-utilising iron oxygenase, cysteine dioxygenase (CDO), does not tolerate the substitution of sulfur with selenium and is unable to oxidise selenocysteine.⁵³⁻⁵⁶ This suggests that iron oxygenases are perhaps more sensitive to Se-S substitution due to their requirement for finely-tuned electronics to mediate oxygen activation at an iron centre. We were therefore interested to use selenocysteine (Secys), a substrate analogue of cysteine, as a chemical probe of the bacterial, cysteine-utilising EgtB from *Chloracidobacterium thermophilum* (*CthEgtB*) (see chapter two for the full structural and functional characterization of *CthEgtB*). Secys was found to be a poor substrate yet an excellent mechanistic probe.⁴¹ Substitution of cysteine for selenocysteine resulted in the enzyme catalyzed O_2 activation being at least ten times slower, however only the selenoether product could be detected (Figure 4). The most likely explanation is that the selenoxide is produced in a similar way to the sulfoxide, yet is rapidly reduced in the presence of selenols, ascorbate or tris(2-carboxyethyl)phosphine (TCEP), all of which are present in great excess under the reaction conditions.^{41, 57-58} An inhibitory constant (K_i) of $17 \pm 4 \mu\text{M}$ indicates that this reduced reactivity does not arise from impaired binding of SeCys.

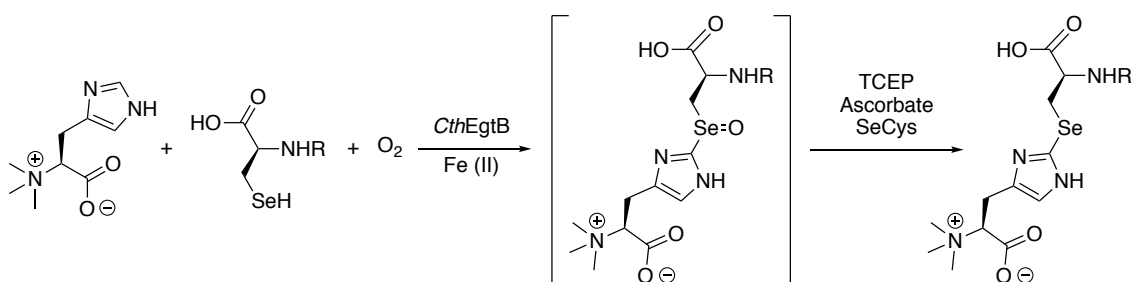


Figure 4. Proposed Scheme for production of selenoether, under the *CthEgtB* + Secys reaction conditions.

The inhibitory properties of SeCys were exploited to determine the substrate binding sequence. Inhibitory kinetic data showed that (i) selenocysteine acts as a competitive inhibitor with respect to cysteine and (ii) an uncompetitive inhibitor with respect to TMH. This means SeCys can only bind to *CthEgtB* in the presence of bound TMH, indicating an ordered sequential binding mechanism in which TMH binds first to the enzyme, presumably followed by cysteine and then oxygen.⁴¹ Despite being a poor substrate, the use of SeCys elicited

further effects on catalysis. The combination of SeCys with C2-deuterated TMH is characterized by a large substrate isotope effect of 3.1 ± 0.1 , which can be attributed to a primary kinetic isotope effect (KIE) associated with C-D bond cleavage. For such an effect to be observed, the C-Se bond must be formed prior to removal of the C2 deuterium. This indicates that deprotonation of C2 of imidazole of TMH takes place after C-Se bond formation. A model system showed that the protons α to a sulfoxide are more acidic than those α to selenoxide, which would explain the emergence of a KIE upon substitution of S with Se.⁴¹

The use of SeCys as a mechanistic probe also exposed fundamental differences in reactivity between the EgtBs of different pathways. In contrast to the bacterial system, efficient selenoether formation was catalyzed by the fungal EgtB from *Chaetomium thermophilum* (Cth-Egt1). Under saturating selenocysteine and TMH conditions, the rate of selenoether production was only 1.7 times slower than the reaction rate of sulfoxide production using cysteine as a substrate. These differences in SeCys tolerance are likely to arise from significant differences in O₂ binding and activation abilities between the fungal and bacterial enzymes. However lack of a fungal crystal structure hinders identification of the contributing structural features.^{41,59} This topic will be explored further in chapter five.

MthEgtB.A82S modulates iron-thiolate reactivity. The rational design of enzyme mutations is one approach by which an enzyme can be engineered for biocatalytic purposes. Alternatively, active site engineering is a powerful tool to assess mechanistic hypotheses. In *MthEgtB* a hydrogen bond was introduced to the sulfur of the iron-thiolate formed upon substrate binding. This was achieved through mutation of Ala82 to a serine. A blue shift in the ligand-to-metal charge transfer (LMCT) band corresponding to an energy difference of ~ 1 kcal/mol was observed upon mutation, which is consistent with the introduction of a hydrogen bond between the substrate thiol and A82S. A crystal structure thereof confirms this notion. In regards to reactivity, a 5-fold drop in k_{cat} was observed, indicating that a mechanistic step involving the thiol becomes rate-limiting when the reduction potential is increased. It is well established that hydrogen bonding to a metal thiolate modulates the covalency of the bond, leading to a more positive redox potential.⁶⁰⁻⁶¹ A *MthEgtB*.Y337F.A82S double mutant was subject to a proton inventory experiment that indicated that for this system two protons are involved in a rate-determining step, providing evidence for proton coupled electron transfer.⁶² Mechanistic proposals must therefore be able to account for the latter two observations, and will be discussed hereafter.

OvoA

The assignment of EGT biosynthetic genes in mycobacteria not only led to the discovery of alternative ergothioneine pathways, but also to the discovery of another sulfoxide synthase, OvoA. While OvoA has many commonalities with EgtB, several features distinguish it from EgtB, in particular differences in substrate- and regio- selectivity. These differences are advantageous to us as (1) it provides a fantastic opportunity for the use of comparative enzymology as a means to dissect the catalytic mechanism of the sulfoxide synthases. (2) The differences in substrate- and regio- selectivity highlight the potential of this scaffold for tailoring to accept other substrates or to modulate regioselectivity. Finally, the distribution and chemistry of the biosynthetic enzymes may reveal common themes surrounding sulfhydryl imidazole compounds in biology or may highlight differences which aid determination of their biological roles. Finally, the evolutionary relationship of Ovo with EgtB provides further insight into the evolution of the unique scaffold and both compounds.

OvoA Discovery. OvoA, is involved in the biosynthesis of Ovothiol A (Figure 5).⁶³ Ovothiols are N π -methyl-5-thiohistidines that accumulate to high concentrations in the eggs of marine invertebrates and are believed to protect the DNA of these eggs from oxidative stress.⁶⁴ Like EGT, the exact *in vivo* role of the ovothiols requires further investigation. The ovothiols are characterized by a very acidic thiol group (pKa = 1.4) and a redox potential higher than most thiols (-0.09 V vs SHE), which may facilitate its efficiency in scavenging peroxides.^{63, 65-66} In contrast to ergothioneine the predominant tautomer of Ovothiol A is a thiol.

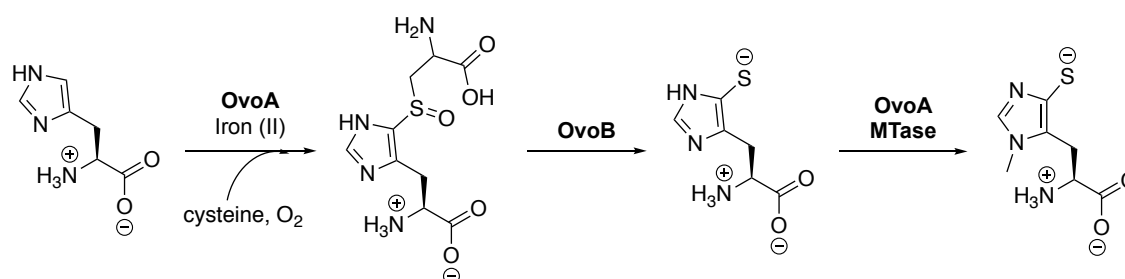


Figure 5. Reaction catalyzed by OvoA and OvoB in the biosynthesis of Ovothiol A.⁶⁷

Key Differences. The gene encoding for OvoA corresponds to a N-terminal sulfoxide synthase and a SAM-dependent methyltransferase at the C-terminus, which is responsible for methylation of N π in the ultimate step of Ovothiol A biosynthesis.^{63, 68} Characterization of the OvoAs from *Erwinia tasmaniensis* (*EtaOvoA*) and *Trypanosoma cruzi* (*TcrOvoA*) demonstrated their ability to catalyze oxidative coupling, yet highlighted three key differences between OvoA and EgtB despite utilizing the same fold. (i) OvoA prefers histidine as a substrate, and uses cysteine as the thiol donor (ii) sulfurization occurs at the C5 position of histidine (iii) OvoA is a bifunctional enzyme.⁶³

Substrate- and Regio-selectivity. OvoA was found to be highly specific for cysteine, while alternative thiol donors such as glutathione or γ -GC were not accepted as substrates.⁶³ In contrast, the substrate scope for

the acceptor is much broader; various imidazole- containing variants were tolerated as alternative substrates by OvoA. Two substrate analogues are of particular significance. Firstly, 2-fluoro-L-histidine, which was turned over with a similar efficiency to histidine. The implications of this on mechanistic proposal is discussed later. Secondly, D-Histidine, which was also tolerated as a substrate.⁶⁹ The use of D-histidine presented a mixture of two regioisomers conjugated at either the C2 or C5 position of histidine. This result indicated that the binding mode dictates regioselectivity, providing an explanation for the differing regioselectivities of EgtB and OvoA.⁶⁹ OvoA was also found to accept mono-, di- and tri-methylated histidine as substrates, each subsequent methylation increased the proportion of the C5- conjugate produced, with the di- and tri-methylated histidines producing almost solely the C5-conjugated product. This again demonstrates that the OvoA substrate binding pocket plays a key role in substrate binding, which consequently determines regioselectivity of oxidative C-S bond formation.⁷⁰ Upon modification of the acceptor substrate, reaction uncoupling was also observed, resulting in cysteine oxidation to cysteine sulfinic acid and cystine. In the presence of native substrates, OvoA already produces 11 % sulfinic acid. Cysteine oxidation became prominent with increased perturbation of acceptor binding, with use of TMH as a substrate resulting in approximately 60 % of the sulfinic acid product. This shows that substrate positioning is not only important for regioselectivity but also for avoiding alternative oxidation events. Any changes to the precise alignment of substrates, even minimal perturbations, facilitates the easier dioxygenation pathway, leading to a change in product ratios.⁷¹ This provides us with an excellent tool for probing active site geometry.

Y417F and Variants. As both EgtB and OvoA utilize the same fold, there is an expectation that OvoA also requires a catalytic acid. Bioinformatic analysis and modelling of OvoA suggested that Tyr417 is the OvoA counterpart to Tyr377 in *MthEgtB*. Functional characterization of a Tyr417F mutant confirmed this, with OvoA being converted to a cysteine dioxygenase with no detectable sulfoxide synthase activity. Tyr417 was also substituted with a non-natural tyrosine derivative, 2-amino-3-(4-hydroxy-3-(methylthio)phenyl) propanoic acid (MtTyr). Tyr417-MtTyr decreased sulfoxide activity compared to the WT. Use of D₂O also further decreased productive reactivity, reaffirming the sensitivity of this enzyme to minimal perturbations that lead to uncoupling.⁷² Tyr417 was also replaced with 3-methoxytyrosine (MeOTyr), which again modulated the partitioning between the mono- and di-oxygenation pathways.⁷³ The authors argue that tyrosine acts as a redox mediator, and that the tyrosine analogues modulate the redox potential, thereby altering the ratio of the two competing pathways.⁷²⁻⁷³ However, (1) this work does not demonstrate this and (2) a redox role is not necessary to explain the biochemical observations.

An OvoA like EgtB - Convergent Evolution in Ergothioneine biosynthesis. Exploration of the OvoA bioinformatic space identified a cyanobacteria with an OvoA-like enzyme that lacks a C-terminal methyltransferase domain, and is instead co-encoded with an EgtD. The short OvoA from *Microcystis aeruginosa* (*MaeEgtB*) was recombinantly produced and characterized. Histidine was found to be a poor substrate; instead *MaeEgtB* was proficient in catalyzing EgtB activity. The switch in substrate selectivity was hypothesized to correspond to the presence of a Asn-Trp motif in TMH acceptors. To test this hypothesis, this Asn-Trp motif was installed in *EtaOvoA*. While the mutant is less active than the WT, the substrate

specificity was dramatically altered by over 100 fold to favor TMH as a substrate. While the two single mutations were not sufficient to completely alter substrate selectivity, they were found to make a large contribution. This work reports the adaption of a cyanobacterial OvoA-like sulfoxide synthase to catalyze ergothioneine production. The existence of two distinct families of cyanobacterial sulfoxide synthases (EgtB-like and OvoA-like) indicates that aerobic ergothioneine production has emerged at least twice through independent evolutionary pathways, unravelling even more diversity within the biosynthetic pathways and highlighting an evolutionary route for adaptation.⁴⁷

1.6 Cysteine Dioxygenase

The diversity of the sulfoxide synthases enables the use of comparative enzymology as a tool to explore the mechanism and evolution of this unique enzyme. While OvoA is different enough to allow the use of this approach, OvoA and EgtB utilize a very similar co-ordination sphere and the same protein scaffold. A catalyst with identical reactivity, yet no evolutionary relation, and complete structural diversity would provide an even more powerful comparison. Cysteine dioxygenase (CDO) provides this opportunity. While CDO has no sulfoxide synthase activity, it catalyzes cysteine di-oxidation, which is the side reactivity of most sulfoxide synthases and prominent reactivity of several mutants. Comparison to CDO makes for a fascinating parallel.

Cysteine dioxygenase (CDO) is a thiol oxygenase with a cupin fold, that catalyzes the oxidation *L*-cysteine to cysteine sulfinic acid in the first step of cys catabolism in mammals.⁷⁴ While CDO has no evolutionary relationship to the sulfoxide synthases, there are resemblances between the two enzymes, particularly in the first co-ordination sphere of iron. In CDO, iron is also coordinated by a 3-His facial triad, deviating from the 2-His-1-carboxylate archetype. The thiol and amine groups of cysteine coordinate directly to iron, leaving a sixth site free for oxygen binding (Figure 6).⁷⁵⁻⁷⁶ A tyrosine (Tyr157) hydrogen bonds to the substrate cysteine, playing a role in substrate binding and positioning. It has also been proposed that Tyr157 forms a hydrogen-bonding interaction to distal oxygen during catalysis.⁷⁷ In mammalian CDO, this tyrosine is crosslinked to a cys (Cys93), in a cysteinyltyrosine post-translational modification near the active site. The nature and role of this crosslink has been heavily investigated, yet its exact role is not clear. Mutation of Tyr157 to phenylalanine greatly reduced activity to 5-8 % of the WT, indicating Tyr157 is important, but not essential for reactivity. While cys cannot form any direct interactions to oxygen or cys, its mutation decreased reactivity to 50-57 % of that of the WT, showing the importance of not only Tyr157 but also the post-translational modification for activity.^{75,77} This crosslink is absent in prokaryotic CDOs.⁷⁵

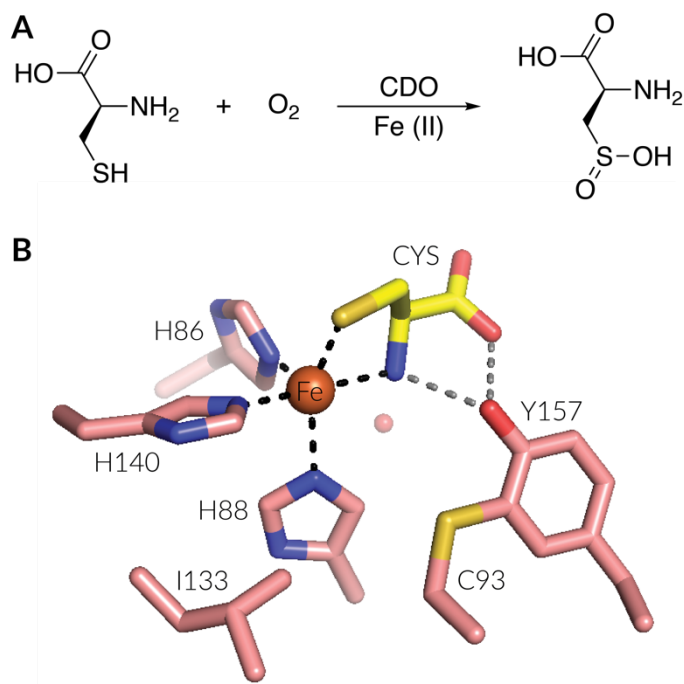


Figure 6. Crystal Structure of rat CDO active site with cysteine bound (PDB:4IEV).⁷⁶

Details within the catalytic mechanism (Figure 7) of CDO are still disputed, in particular the identity of a key intermediate. Two alternative mechanisms have been proposed on the basis of spectroscopic, structural and computational studies. In the first step, cysteine binds to the resting iron(II) enzyme (**A** & **A'**), which is followed by dioxygen binding to form an iron(III) superoxo species (**B**). The distal oxygen attacks the sulfur of cys, forming a four-membered, cyclic peroxo intermediate (**C**). The following step involves scission of the O-O bond to form a metal bound, activated oxygen species and S-O.⁷⁷ However, whether homolytic or heterolytic cleavage occurs and the exact identity of formed species are both disputed. Heterolytic cleavage forms an iron(IV)-oxo intermediate and sulfenate (**D**).⁷⁸ Alternatively, the work of the Brunold group advocates for homolytic cleavage to form iron(III) bound to O^{•-}.^{56, 79} While this intermediate is a resonance form of a iron(IV)-oxo-like species, the 3-His iron co-ordination sphere of CDO is believed to be relevant as it has been suggested it disfavors formation of a high valent iron species.⁵⁶ This may be an important parallel for consideration of the EgtB mechanism. Reorganization of the formed sulfenate has been suggested, occurring either by rotation or dissociation to co-ordinate to iron via the oxygen atom (**D'**). The proximal oxygen and sulfur recombine to form the final S-O bond to form and then release the product (**E**).⁷⁵

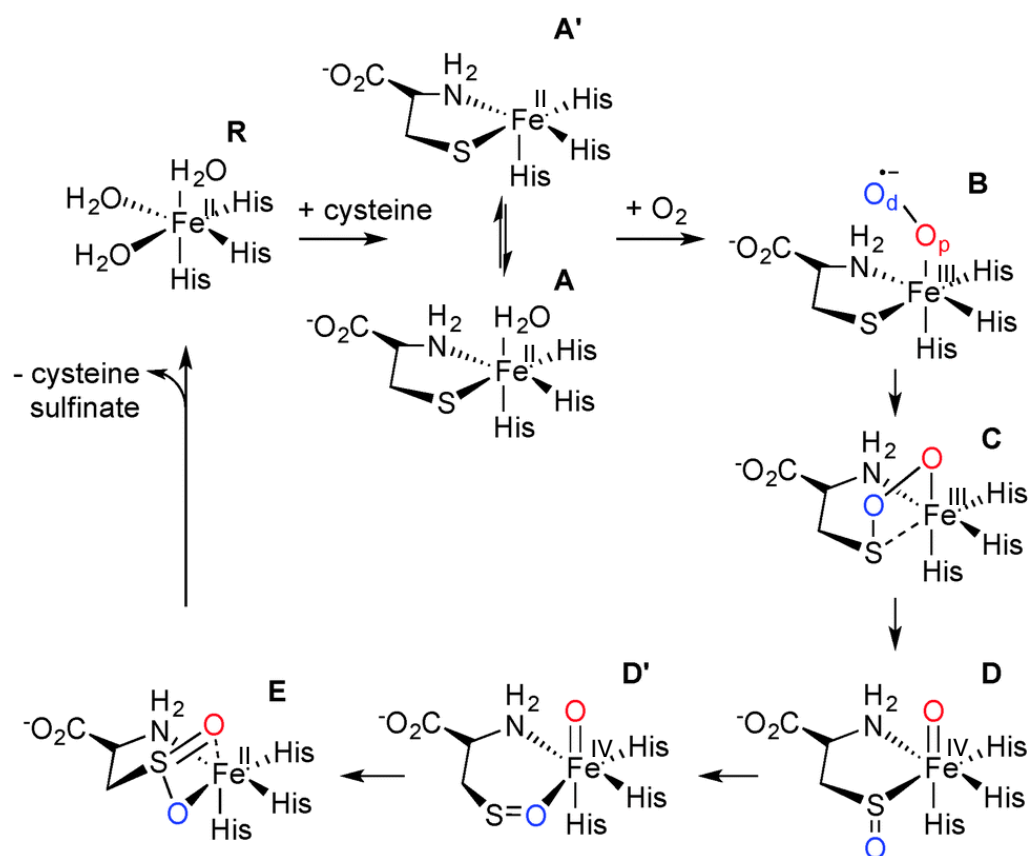


Figure 7. Proposed intermediates in the catalytic mechanism of CDO. Figure has been adapted from ⁷⁸ to include iron oxidation states.

Mechanism of oxidative C-S bond formation

The mechanism of sulfoxide synthase activity in EgtB and OvoA is intriguing due to its unprecedented reactivity, unique structure and dissimilarity to any other known iron-oxygenase. Despite differences in substrate- and regio-selectivity, OvoA and EgtB reactivity is likely to proceed via a similar mechanism. The facile transition between EgtB and OvoA-like activity, with minimal mutations required for functional adaption, advocates for a conserved mechanism.⁴⁷ The catalytic mechanism remains controversial, with several mechanisms based on experimental and computational grounds having been proposed.⁸⁰⁻⁸³ The contention over the mechanism deduces to two steps. The computational proposals dispute the order of key events: does carbon-sulfur bond formation or oxidation of sulfur to a sulfoxide occur first? While the biochemical groups do not agree on the role of the catalytic tyrosine. While the importance of Tyr377/Tyr417 is accepted, how it mediates its role, as a redox agent, or as an acid-base catalyst, is arguable.

Initial mechanistic proposals. Initial studies on OvoA established three key observations, which all proposed mechanisms must account for: (i) No KIE was observed for C5-deuterated histidine. (ii) No KSIE was observed when reactions were carried out in D₂O and (iii) the reaction efficiency is impartial to use of the electron poor 2-fluoro-histidine.⁶⁹

Mashabela. and Seebeck proposed four reaction mechanisms, which were assessed according to the above criteria (Figure 8). In mechanisms 1-3, sulfoxidation precedes C-S bond formation, proceeding via a high valent iron(IV) oxo species formed upon oxidation of cys to an iron bound - sulfenic acid (Figure 8, step a).⁶⁹ The iron(IV) oxo species has been observed as an intermediate in numerous oxygenases.⁸⁴⁻⁸⁵ Following formation of the ferryl-oxo species, both two and one electron chemistry could be premised. In mechanism 1 the iron(IV) oxo species abstracts a hydrogen atom from the imidazole C5 carbon, reminiscent of the consensus mechanism for hydroxylation in α -KG dependent enzymes (Figure 8, step b).⁸⁶ Homolytic C2-H cleavage would generate an sp²-centered radical on the imidazole ring, which could recombine with sulfur to form the C-S bond and subsequently reduce ferric iron back to its ferrous resting state. Sp² radicals are, however, incredibly unstable. Consequently, hydrogen atom transfer would likely be rate determining and accompanied by a primary KIE. A computational assessment of this mechanism predicts a KIE of 5.7.⁸³ However, no primary isotope was detected: **observation (i)**, a KIE of near unity, confutes this mechanism.

Alternatively, the hydrogen atom from the imidazole amine could be abstracted, facilitating the one electron oxidation of the imidazole ring, as in mechanism 2, pathway c (Figure 8). This step is also likely to be rate limiting, yet the absence of a significant solvent isotope effect **observation (ii)** and **observation (iii)** that 2-fluoro-histidine is an efficient substrate, indicates this step either does not occur or is rate limiting. Following formation of the iron(IV) species, a two electron reaction is also possible, in which the imidazole ring could act as a nucleophile and directly attack the iron bound sulfenic acid (Figure 8, Mechanism 3, step d). This mechanism is, however, inconsistent with several biochemical observations. Firstly, introduction of a strong electron withdrawing group onto the histidine ring, such as fluoride, would disfavor nucleophilic attack of the imidazole ring. However, no such effect is observed: **observation (iii)** 2-fluoro-histidine is an efficient substrate. This step is either not rate-limiting or does not occur. If nucleophilic attack is not rate-limiting, then C-H cleavage would be the most likely rate-determining step. However **observation (1)** a KIE near unity contradicts this postulate. Y377 had been suggested as the catalytic base for this step, however both *MthEgtB.WT* and *MthEgtB.Y377F* displayed a KIE of near unity when the C2-deuterated substrate was utilized, indicating that hydrogen or proton removal from TMH is not an essential function of Y377, thereby refuting this suggestion.^{50, 87}

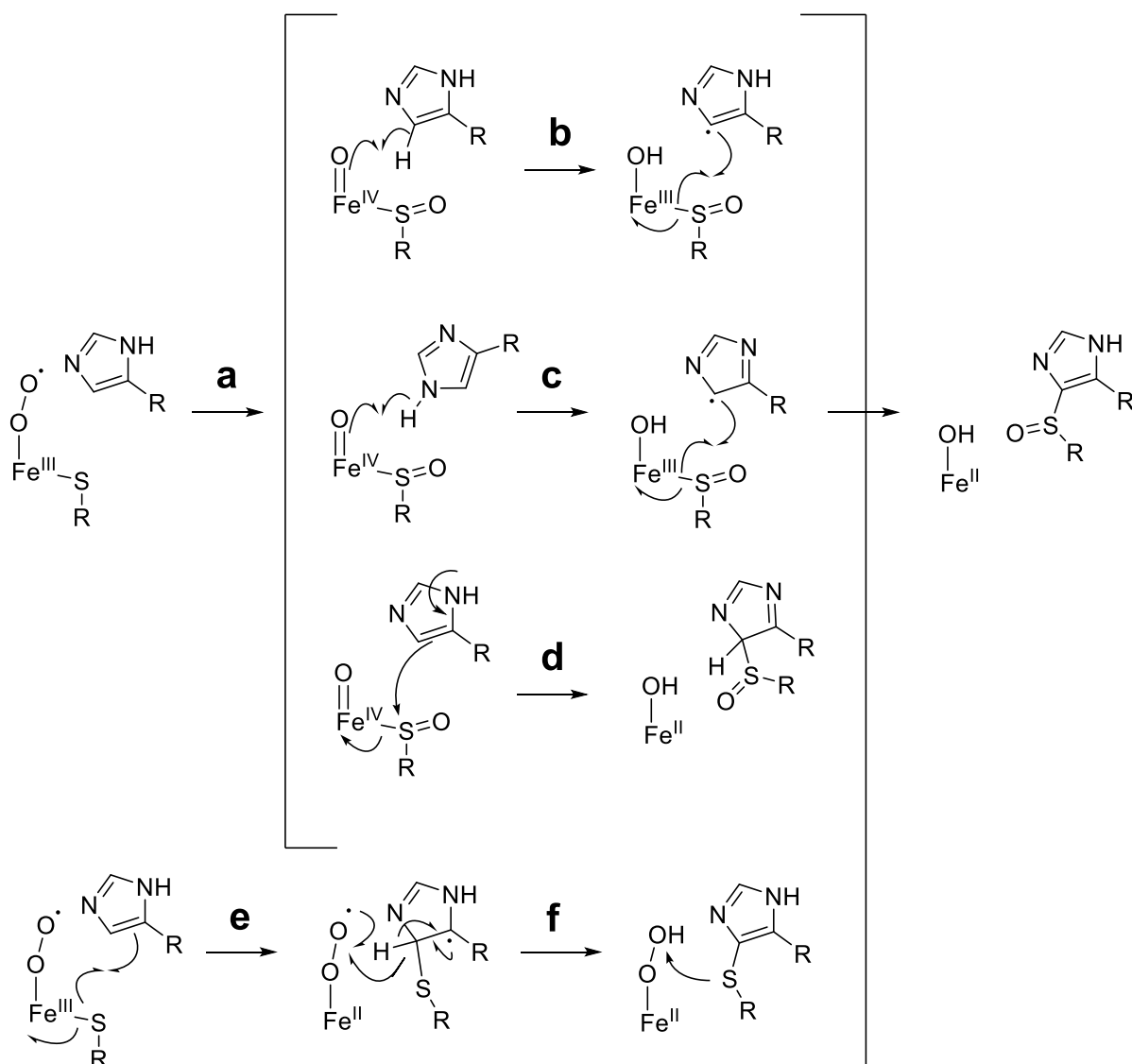


Figure 8. Initial mechanistic proposals for oxidative carbon-sulfur bond formation in OvoA. Adapted from ⁴⁹.

A fourth mechanism, in which C-S bond formation precedes sulfoxidation, is consistent with **observations (i-iii)** (Figure 8, Mechanism 4). Oxygen binding leads to formation of an iron(III) superoxo species, a thiyl radical attacks the imidazole ring (Figure 8, step e) followed by hydrogen atom abstraction and re-aromatization of the imidazole ring (Figure 8, step f) to form a thioether, which can be oxidized. In this mechanism, the substrate imidazole ring acts as electrophile rather than a nucleophile, which would explain the ability to tolerate an electron poor substrate. This approach to C-S bond formation is akin to a thiol-ene mechanism for formation of a thioether from a thiol and an alkene (Figure 9).⁸⁸ The latter mechanism is consistent with the initial biochemical observations and has been revised and updated as further biochemical insights have been gained. The current mechanistic proposal is shown in Figure 10.

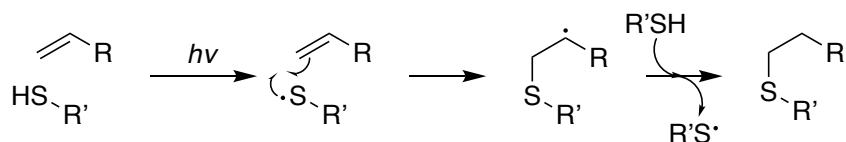


Figure 9. Thiol-ene mechanism for thioether formation.⁸⁸

The *MthEgtB* structure revealed that TMH, γ -GC, and conceivably oxygen, directly co-ordinate to the iron centre.⁴⁹ In the proposed mechanism, TMH binds first to the ferrous catalytic center (Figure 10, step a). Studies with SeCys demonstrated that TMH binding precedes cysteine and oxygen binding.⁴⁴ In the following step, (step b) cysteine and oxygen bind, presumably in this order, as oxygen is typically the last substrate to bind in iron oxygenases.⁸⁹

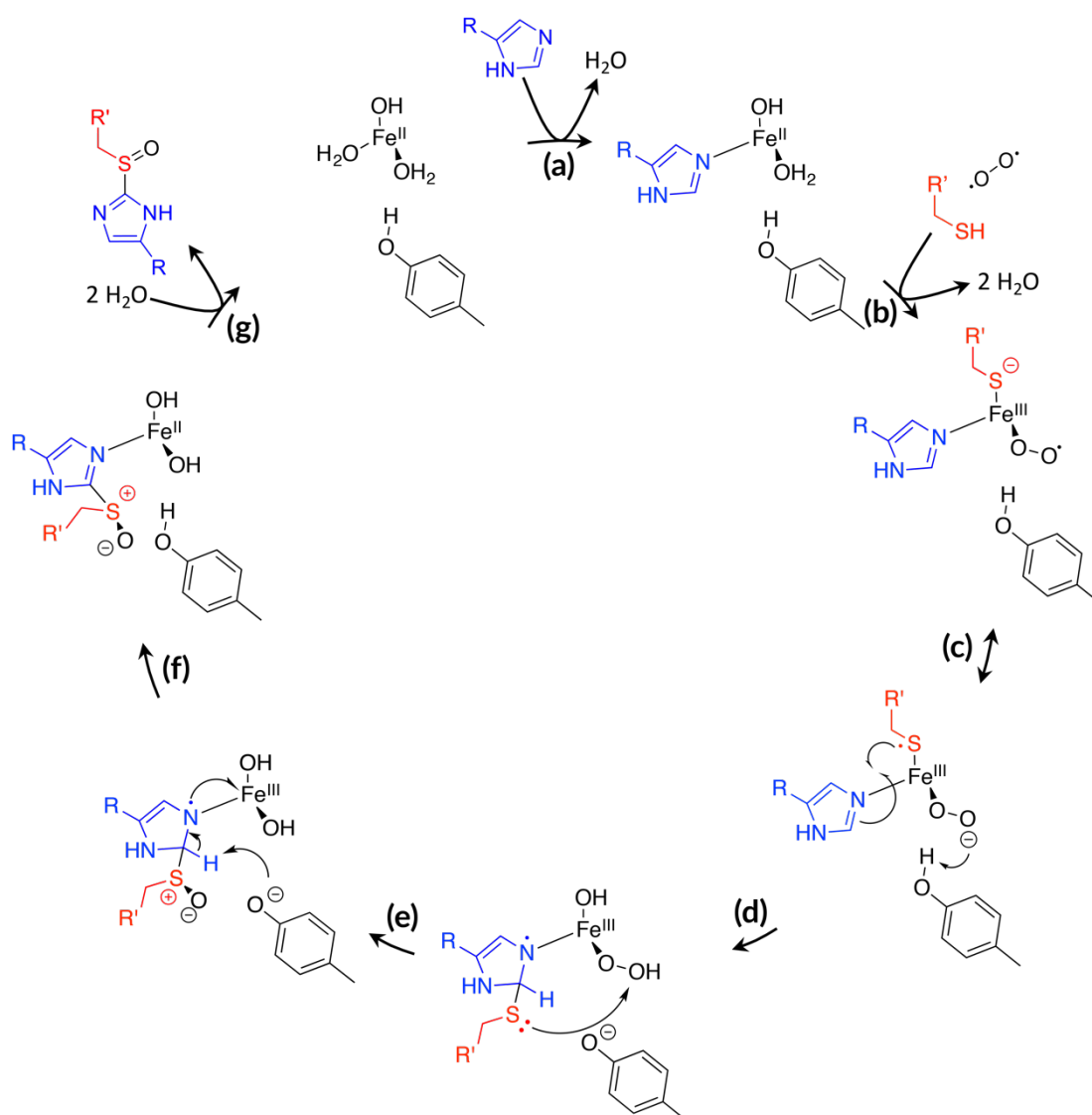


Figure 10. Proposed mechanism for sulfoxide synthase reactivity. The iron-coordinating 3-His triad has been omitted for clarity.

A thiyl radical facilitates C-S bond formation. Upon oxygen binding, we propose that a ferric-superoxo species is formed. While this transient species has rarely been detected, growing evidence suggests that this is a common intermediate for iron dependent oxygenases.^{85, 90-94} Transfer of an electron from the thiol substrate to the superoxo species would form an iron(III) hydroperoxide and a thiyl radical (step c). This thiyl radical can attack the imidazole ring to form the C-S bond (step d). These preliminary steps are consistent with the first mechanistic restraints set from the preliminary OvoA work.⁹⁵ The observation of reduced k_{cat} for the *MthEgtB*.A82S mutant is also consistent with formation of a thiyl radical. The introduction of hydrogen bond to the thiol donor increases the reduction potential of the system, which would disfavor formation of the thiyl radical and would account for the reduction of k_{cat} . This result also suggests that thiyl formation becomes rate-limiting in this system.⁶²

Tyr377/417 is a catalytic acid. Concurrent to electron transfer, it is proposed that a catalytic tyrosine (Tyr377_{*MthEgtB*} or Tyr417_{*OvoA*}) protonates the ferric-peroxo species. This protonation step is essential for sulfoxide synthase activity. In the absence of a phenylalanine, protonation becomes the rate-determining step for oxidative C-S bond formation, and gives rise to a pH dependence and KSIE. These mechanistic steps are consistent with the Tyr377F phenotype.⁵⁰ The KSIE and proton inventory experiments with *MthEgtB*.Y377F.A82S indicate that this protonation and electron transfer to form the thiyl radical are likely to be coupled (proton-coupled electron transfer (PCET)). For this mutant, two protons are involved in the transition state of a rate-determining step. This solvent isotope-sensitive step is most consistent with protonation of the superoxo-intermediate. PCET is likely important to increase thiyl radical character on the metal thiolate to facilitate carbon-sulfur bond formation.⁶²

A common intermediate. The availability of a proton from Tyr377_{*MthEgtB*} or Tyr417_{*OvoA*} determines if mono- or di-oxygenation occurs.^{50, 72} This implicates the ferric superoxo species as a common intermediate and branch point for both competing reactivities. This is supported by the finding that TMH is required for di-oxygenation, indicating both reactivities occur via the same reactive complex.⁵⁰ This branch point is very sensitive and can easily be modulated through essentially any perturbation to the active site. Use of substrate analogues, catalytic tyrosine analogues/variants and even use of D₂O have altered the ratio of mono- to di-oxygenation products.^{50, 71-73}

In the described mechanistic proposal, C-S precedes S-O bond formation. Studies of *CthEgtB* with SeCys have been useful in delineating the order of steps, with the appearance of a KIE upon use of SeCys and C2-deuterated TMH indicating that C-S bond formation precedes C-H bond cleavage.⁴¹ This result provides further evidence to confute the first mechanistic proposal, and those warranted by the computational studies. Ensuing C-S bond formation, C-H bond cleavage, and S-O bond formation remain to be catalyzed. We believe that sulfoxidation occurs first, facilitated by attack of the thioether onto the distal oxygen. This is followed by abstraction of the C2 proton by the catalytic tyrosine, coupled to aromatization of imidazole ring. This order of steps is rationalized as protons α to a sulfoxide are more acidic than those α to a thioether, thereby reducing difficulty of the C-H cleavage step.⁹⁶⁻⁹⁷ Work is ongoing to determine if thioether is an

inhibitor or a viable intermediate. This catalytic mechanism is also consistent within the geometrical constraints imposed by the *MthEgtB* crystal structures to produce the correct sulfoxide stereochemistry, the *S*-isomer.

This proposed mechanism has thus far survived all experimental scrutiny. Three computational studies challenge this mechanism, each with an alternative mechanistic proposition. These mechanisms and their feasibility are discussed in chapter two.

Contributions of Mechanistic Enzymology

The careful chemical dissection of the biochemistry of EGT over the past 10 years has been incredibly fruitful. A chemical perspective has enabled the discovery of numerous biosynthetic pathways, identification of the individual enzymes, resolution of their structures and exploration of their mechanistic modes of action. Each of these facets has illuminated several fascinating features regarding the biology, evolution and biocatalytic applications of EGT.

Before the discovery of the mycobacterial EGT synthetic pathway, the only organisms reported to synthesize ergothioneine were bacteria belonging to the order Actinomycetales (for example, mycobacteria) and non-yeast-like fungi, including members of the division Basidiomycota and Ascomycota.⁴³ The assignment of biosynthetic genes enabled the distribution of these genes to be studied, revealing that EGT production is much more ubiquitous than previously thought, being produced in numerous bacteria, fungi and even archaea. The genetic annotation is powerful as it allows us access to organisms which cannot be cultured or are not yet amenable to genetic modifications to determine the prevalence of EGT biosynthesis in nature. Previously, EGT production was determined through growth of the organism, isolation and quantification of ergothioneine; a laborious process. The ubiquity of organisms we now know to produce EGT, indicates EGT plays an important physiological role.

The assignment of the EGT biosynthetic genes also provided a platform for *in vivo* studies in which the role of EGT could be studied through silencing of the biosynthetic genes. This led to the flourishing of *in vivo* studies which were not possible ten years ago.^{37-38, 40, 44, 98-103} Similar studies have also been carried out by silencing the ergothioneine transporter.¹⁰⁴⁻¹⁰⁷ These studies allow us to gain insight into the physiological role of EGT, enabled by the chemocentric view. The identification of biosynthetic groups has also been capitalized upon by numerous research groups and patents as means to produce EGT.⁸⁹⁻¹⁰⁸⁻¹¹⁰ This goal has driven further exploration of homologues that are amenable for overexpression, and focused efforts into the biocatalytic engineering of EGT biosynthetic enzymes.

The multiple biosynthetic pathways, and the diversity of the pathways and enzymes also advocate for a fundamental role in life. The finding that anaerobic bacteria produce EGT suggests that the current research efforts that focus on an antioxidant role may be too narrow, and that alternative roles should be considered. The latter is supported by the recent suggestion that EGT is an ancient molecule that may predate the GOE, suggesting that EGT first played a role in anoxic biochemistry before being co-opted as an antioxidant after the emergence of molecular oxygen.

A third aspect which we can learn about from the chemistry of ergothioneine is the chemical insight into C-S bond formation, gained upon the discovery of two new catalysts for oxidative carbon-sulfur bond formation. A comprehensive discussion has been described in the theses of Reto Burn and Sebastian Flückiger.^{59, 111}

A wealth of knowledge regarding the chemical basis of biosynthesis, on biology, evolution and biocatalysts of EGT has been uncovered. However, much remains to be discovered. With regards to the biosynthesis of EGT, the mechanistic details and evolutionary history of EgtB, the key C-S forming enzyme in the aerobic pathway, remains a complex question that requires further exploration. We can competently predict which organisms can and cannot produce ergothioneine. However, we cannot yet determine prokaryotes that can utilize EGT without synthesizing it themselves. It would be incredibly powerful to have a molecular basis for organisms that take up, utilize or degrade ergothioneine without synthesizing it themselves. We hypothesize the ability to identify such organisms would further establish the ubiquity and importance of EGT in life.

Aim of this thesis

The aim of this thesis is to further explore the diversity of proteins involved in ergothioneine (EGT) production and utilization. Doing so will provide valuable insights into the biology and evolutionary emergence of ergothioneine. An approach will be utilized that involves the structural characterization of enzymes or proteins likely to be divergent from characterized homologues. These crystal structures will then be leveraged to either inform mutational studies or, in a comparative analysis, to tease out the molecular determinants of a particular structural or functional element such as substrate binding or oxygen activation. This approach will first be applied to the EgtB family and then to novel solute-binding proteins that we show are involved in EGT transport.

Synopsis

This thesis explores the diversity of ergothioneine (EGT) from a chemical perspective. This enables insight into the biology and evolutionary history of EGT. Protein crystallography is an invaluable technique for investigating enzyme structure and function, and plays a central role in this thesis. In conjunction with mechanistic and bioinformatic studies, sequence motifs for ergothioneine-related chemistry can be identified. Chapters two - four explore the diversity of a unique enzyme class involved in anaerobic ergothioneine biosynthesis, the sulfoxide synthases. Two overarching enigmas surround the sulfoxide synthase family; firstly, the catalytic mechanism for sulfoxide synthase dependent oxidative C-S bond formation and secondly, the evolutionary emergence of the sulfoxide synthase family. We tackle these topics using an approach that begins with the structural characterization of divergent sulfoxide synthases. The insight gained from crystal structures informs mutational studies to dissect key residues that indicate a particular reactivity or substrate selectivity. A similar approach is utilised in chapter five in the characterisation of a component of EGT transport system. However the comparison of the structures of divergent proteins, rather than mutational studies, allows for the dissection of a fingerprint for substrate selectivities. An understanding of the complex molecular details, such as sulfoxide synthase mechanism, or sequence motifs for substrate binding may shed light on the evolutionary emergence and physiological role of ergothioneine.

Chapter one provides an introduction to this thesis, which begins with a description of the central and essential role that enzymes play in life and how a chemical approach to studying enzymes and nature is incredibly valuable. This is followed by a brief overview of the possible roles and relevance of ergothioneine to life. The various biosynthetic pathways are then described, noting the key biochemical findings of their discoveries. This leads to a comprehensive report of the biochemical characterization of the sulfoxide synthases, the implications of which are applied to a discussion of a mechanistic proposal. Finally the key contributions provided thus far by a chemocentric view to understand the biology and evolution of EGT are summarized.

The introduction highlights how characterization of distant homologues (EgtB and OvoA) is an incredibly valuable approach to understanding the evolutionary emergence of the sulfoxide synthases, and to refine current mechanistic proposals. A bacterial EgtB from *Chloracidobacterium thermophilum* (CthEgtB) was identified as being divergent from other characterized EgtBs due to an apparent lack of residues for γ -GC binding and our inability to identify a counterpart to the vital catalytic acid, Tyr377_{MthEgtB}/Tyr417_{EtaOvoA}. The latter observation is particularly intriguing due to the expectation that important functional residues are conserved across a family of proteins.¹¹² Tyr377_{MthEgtB}/Tyr417_{EtaOvoA} are expected to be subject to severe evolutionary constraints, and loss thereof during evolution should be deleterious to function.¹¹²

Synopsis

The structural and kinetic characterization of *CthEgtB*, is presented in **Chapter two**, and shows that this enzyme supports ergothioneine biosynthesis. The structure revealed a completely different configuration of active site residues that are involved in oxygen binding and activation. This change in active site residues is also coupled to the complete re-modelling of the active site. These drastic active site differences provide us with a powerful and unique approach to tease out mechanistic features and to isolate the core catalytic principles of these catalysts via comparative enzymology. Re-evaluation of the contentious mechanistic proposals in view of the structure of *CthEgtB* provides a new test that may help to distinguish between different models. We believe that these additional constraints will be of significant help in the elimination or validation of mechanistic proposals for the sulfoxide synthase-catalyzed reaction.

Comparison of the *CthEgtB* structure to characterized homologues enabled a classification of all ergothioneine biosynthetic EgtBs into five subtypes, each characterized by unique active-site features. This allowed us to hypothesize which features an ancestral EgtB may have possessed, and an explanation for the diversity observed. This approach additionally led to the identification of another divergent homologue, the uncharacterized EgtB from *Thermosynechococcus elongatus* (*TelEgtB*), again highlighting the diversity among the sulfoxide synthase family, and the strength of our sequence motifs for predicting features. An addendum discusses the substrate selectivity of *CthEgtB* and confutes the published findings of another group on the same enzyme.¹¹³

The characterisation of *CthEgtB* in chapter two highlights several differences compared to other sulfoxide synthase types. These differences pertain to the introduction of tetramerization and two flexible active site loops. This observation alone raises interesting questions in regards to the role of quaternary structure and the timing and trigger of loop folding to form the active site, and in particular the oxygen binding site. Our interest in understanding these components was heightened by the challenge associated with gaining crystal structures of the two active sites in a closed, catalytically relevant conformation or in complex with a second substrate cysteine, cysteine analogue or inhibitor.

Chapter three address these questions and explores features unique to *CthEgtB*, through a thorough biophysical characterization of *CthEgtB* and several loop mutants that destabilize the active site loops and/or tetramerization. This work reveals a complex interplay between TMH binding, loop folding and quaternary structure. This contributes to the discussion on homo-oligomerization by providing a unique example in which tetramerization is an essential strategy to stabilize active enzyme form and to optimize enzymatic performance, while other EgtBs are monomeric. Together with chapter two, a complete functional assignment of the four conserved residues on an a catalytic active site loop is provided. This touches on the concept of how hydrogen bonding may modulate reactivity, which is a common theme in several chapters. We also propose that oligomerization enabled divergent evolution of the EgtB family, providing an explanation for the extreme active site plasticity and divergence seen in *CthEgtB*.

In chapter two, *CthEgtB* was identified as a distant homologue through sequence comparisons. An alternative approach to identifying divergent enzymes is through exploration of genomic environment to identify sulfoxide synthases conserved in unique gene clusters. The genome of *Variovorax paradoxus* was found to contain two EgtB homologues. The first homologue is encoded with archetypal ergothioneine biosynthetic proteins, while the second homologue (*Vpa2054*) is found within a unique genomic environment conserved among many β -proteobacteria. The genetic annotations of this gene cluster implicate a selenium-based molecule as a substrate for *Vpa2054*. This hypothesis is captivating as 1) enzymes with dedicated selenium reactivities are incredibly rare and 2) this conserved cluster provides a possible route for the biosynthesis of selenoneine, an EGT isolog with ramifications for human health.

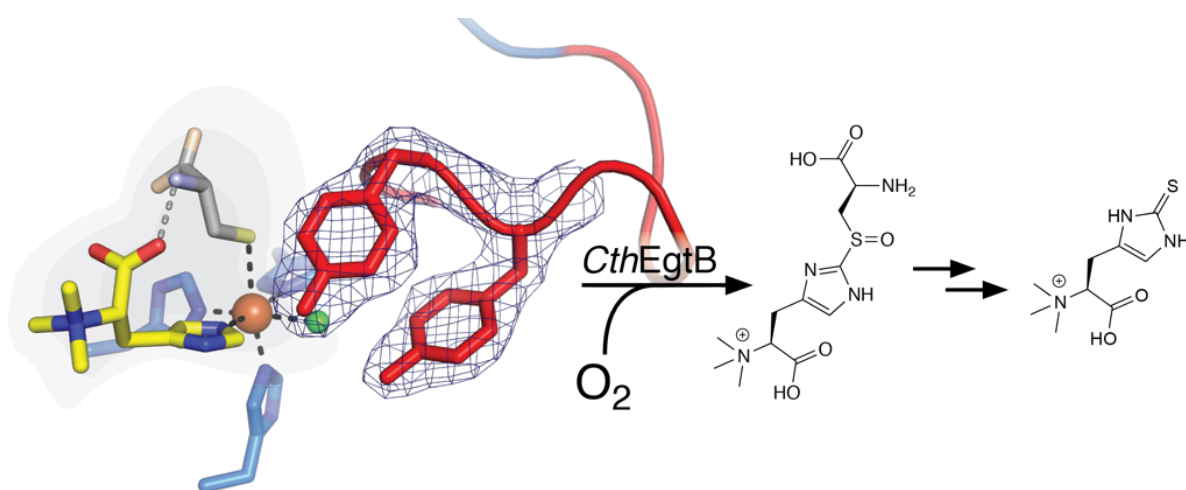
The structural characterization of *Vpa2054* is discussed in **Chapter four**. The protein crystal structure and thermodynamic measurements provides a platform for the dissection of possible substrates and identification of structural features that may facilitate Se-dependent oxygen activation. Modulation of the metal-thiolate bond through hydrogen bonding appears again, providing a parallel to *MthEgtB.A82S* (introduction) and *CthEgtB.S92A* (chapter three). The structure provides another example of the diversity of the sulfoxide synthase fold and highlights the amenability of the sulfoxide synthase scaffold for evolutionary divergence, in particular, the variance of the $\alpha 1$ - $\alpha 2$ loop to generate plasticity in the EgtB active site.

Chapter five moves away from the sulfoxide synthases and ergothioneine biosynthesis and focuses on an aspect of ergothioneine utilization and degradation. A discrepancy in the organisms that can produce ergothioneine and degrade ergothioneine (enabled through a profound understanding of the involved enzymology) indicated that a transport system for ergothioneine and its derivatives must exist. Again a structural approach was essential to characterize and assign a function to several solute-binding proteins as either EGT or ergothioneine-sulfonic acid (EGTSO₃⁻) specific transporters. The comparison of several atomic resolution structures allowed for the recognition of the residues important for substrate selectivity, which in combination with a bioinformatic analysis provides a fingerprint for EGT and EGTSO₃⁻ transporters. These studies provide the first solid evidence that the oxidized analogue, ergothioneine-sulfonic acid is indeed a relevant, degradation product in nature. Additionally, this molecular fingerprint allows us to identify organisms that uptake ergothioneine, yet cannot produce or degrade it. This dramatically increases the number of organisms to which we now know EGT is relevant, and may play a physiological role. This shows that EGT chemistry is much more prevalent in nature than originally hypothesized.

2. An Alternative Active Site Architecture for O₂ Activation in the Ergothioneine Biosynthetic EgtB from *Chloracidobacterium thermophilum*

This research was originally published in J. Am. Chem. Soc.

Anja R. Stampfli, Kristina V. Goncharenko, Marcel Meury, Badri N. Dubey, Tilman Schirmer and Florian P. Seebeck. *J. Am. Chem. Soc.* 141, 13, 5275-5285 (2019)



Sulfoxide synthases are nonheme iron enzymes that catalyze oxidative carbon–sulfur bond formation between cysteine derivatives and *N,N,N*- α -trimethylhistidine as a key step in the biosynthesis of thiohistidines. The complex catalytic mechanism of this enzyme reaction has emerged as the controversial subject of several biochemical and computational studies. These studies all used the structure of the γ -glutamyl cysteine utilizing sulfoxide synthase, *MthEgtB* from *Mycobacterium thermophilum* (EC 1.14.99.50), as a structural basis. To provide an alternative model system, we have solved the crystal structure of *CthEgtB* from *Chloracidobacterium thermophilum* (EC 1.14.99.51) that utilizes cysteine as a sulfur donor. This structure reveals a completely different configuration of active site residues that are involved in oxygen binding and activation. Furthermore, comparison of the two EgtB structures enables a classification of all ergothioneine biosynthetic EgtBs into five subtypes, each characterized by unique active-site features. This active site diversity provides an excellent platform to examine the catalytic mechanism of sulfoxide synthases by comparative enzymology, but also raises the question as to why so many different solutions to the same biosynthetic problem have emerged.

Introduction

EgtB is a non-heme iron-dependent sulfoxide synthase that participates in the biosynthesis of ergothioneine (**1**, Figure 1).^{31,49,114} EgtB catalyses oxidative carbon–sulfur (C–S) bond formation between the imidazole ring of *N,N,N*- α -trimethyl histidine (TMH, **2**, Figure 1) and γ -glutamyl cysteine (γ -GC) or cysteine.^{31,33,38,40,47,115-116} Subsequent oxidation of the sulfur atom concludes the four-electron reduction of molecular oxygen (Figure 1) and produces a histidinyl- γ -GC sulfoxide conjugate (**3**, Figure 1) or the histidinyl-cysteine sulfoxide conjugate (**4**). Discovery of EgtB and other ergothioneine biosynthetic enzymes opened new avenues in deciphering the complex biological function of this sulfur compound. For example, studies on *Mycobacterium tuberculosis*, *Burkholderia thailandensis* and *Aspergillus fumigates* showed that deletion of ergothioneine biosynthetic genes reduces tolerance against oxidative stress, suggesting that ergothioneine plays an important role in the redox homeostasis in these pathogenic microorganisms.^{44,103,117-118}

From a chemical perspective, EgtB is of particular interest because this enzyme represents an entirely new catalyst type, distinguishable from any other oxygen-utilizing enzyme by its reactivity and structure. For example, unlike γ -ketoglutarate-dependent oxygenases which adopt a jelly-roll fold,^{114,119} sulfoxide synthases adopt a two-domain structure containing a C-terminal domain related to the copper-dependent formylglycine generating enzyme (FGE-like domain)¹²⁰⁻¹²¹ and a N-terminal domain that is most homologous to zinc-dependent thiol-S transferases (DinB_2 domain).¹²² The crystal structure of EgtB from *Mycobacterium thermoresistibile* (*MthEgtB*) revealed that the active site is located at the interface between the two domains.^{49-50,63} The active site hosts a three-histidine facial triad as a metal binding motif, and several residues that are essential for TMH- and γ -GC-binding. Subsequent biochemical analysis implicated Tyr377 as an essential catalytic residue.^{50,123} Mutating Tyr377 to Phe resulted in an enzyme that produces γ -GC dioxide instead of sulfoxide **4**.⁵⁰ Analogous experiments with the ovothiol biosynthetic sulfoxide synthase OvoA showed that this distant relative of *MthEgtB* uses an equivalent Tyr residue (Tyr417) in the same way.¹²³ Both studies concluded that an active site tyrosine plays a pivotal role in steering an early catalytic intermediate towards sulfoxide production and away from thiol dioxygenation. However, the mechanism by which Tyr377 or Tyr417 influence reactivity remains disputed.^{50,81-82,124}

Given the functional importance of Tyr377 we were surprised to find that a large sub-class of EgtB homologs lack this residue.¹²⁵ Most of these enzymes are encoded by Proteobacteria but also occur in several species from other phyla (SI Table 1). To understand how these EgtB homologs could support ergothioneine production, we examined the crystal structure and kinetic behavior of EgtB from *Chloracidobacterium thermophilum* (*CthEgtB*). This study highlighted five important differences between *MthEgtB* and *CthEgtB*. First, *CthEgtB* uses Cys instead of γ -GC as a sulfur donor; secondly, *CthEgtB* utilizes a tyrosine from the N-terminal domain for the exact same function as Tyr377 in *MthEgtB*; third, the *CthEgtB* active site employs a second active site tyrosine; fourth, a large section of the rigid active site observed in *MthEgtB* is replaced in *CthEgtB* by two mobile active site loops that fold in a substrate-dependent fashion; finally, all *CthEgtB*-type EgtBs appear to adopt a D_2 -symmetric tetrameric quaternary structure. Because of these structural

differences to *MthEgtB*, *CthEgtB* provides a complementary study system that may help to solve the emerging controversy about the catalytic mechanism of sulfoxide synthases. Furthermore, comparison between the structures of *MthEgtB* and *CthEgtB*, compounded by phylogenetic analysis of other homologs illuminates the evolutionary origin of sulfoxide synthases and raises the question as to what evolutionary pressures may have given rise to the observed active site diversity within the enzyme class of sulfoxide synthases.

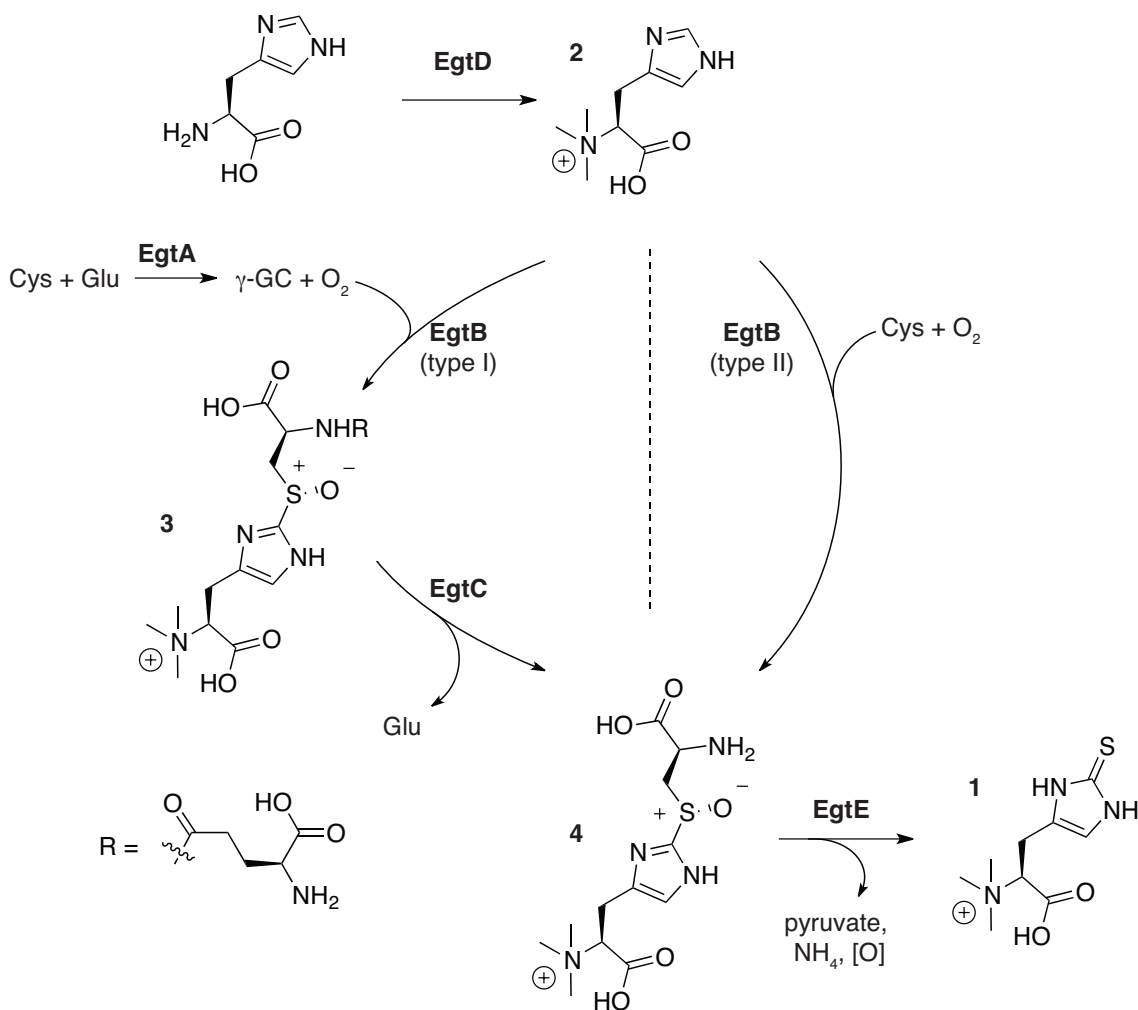


Figure 1. *Mycobacterial* ergothioneine (1) biosynthesis starts with histidine and is catalyzed by the enzymes EgtABCDE (left).³¹ Three key reaction, oxidative sulfurization of TMH (2) is catalyzed by a type I EgtB. Other bacteria, such as *Chloracidobacterium thermophilum* and most fungi utilize a cysteine specific type III or IV sulfoxide synthase, in a three-step pathway.^{40, 47, 115-116, 118, 126}

Results

Crystal structure determination of *CthEgtB*. The structure of *CthEgtB* was determined by X-ray crystallography. *CthEgtB* was crystallized as the native protein in complex with iron (II). The crystal diffracted to a resolution of 2.0 Å and belonged to space group P2₁ with four monomers in the asymmetric unit. For data collection and refinement statistics, see Tables S2 and S3, respectively. The crystal structure was solved by molecular replacement using a homology model built from the native structure of *MthEgtB* (PDB: 4X8E) as a search model. *MthEgtB* and *CthEgtB* share 32 % sequence identity. The electron density revealed a continuous polypeptide chain from residues 17 to 434 with the exception of segments 93-98, 183-193, and 377-384 (Figure S1). In presence of TMH, *CthEgtB* was crystallized in space group C222₁ with cell constants a, c = 108 Å, b = 200 Å. *CthEgtB* crystals containing TMH were obtained through co-crystallization. The structure of the *CthEgtB*/Fe/TMH complex was solved to 2.2 Å by molecular replacement with the native *CthEgtB* model. The (2Fo-Fc) electron density map showed well-defined density into which two of the previously missing regions (loop 1: residues 93-98 and loop 2: 378-384) could be fitted resulting in a model comprising all residues from 7 to 434 aside from residues 183-193 of the linker between the N- and C-terminal domain (Figure 2A, Figure S2). The TMH ligand could be unambiguously modeled into the (Fo-Fc) difference density map (Figure S3). Data collection and refinement statistics are summarized in the supporting information as Table 2 and Table 3, respectively.

***CthEgtB* forms a stable tetramer.** Overall, the tertiary structure of *CthEgtB* (Fig. 2A) is strikingly similar to that of *MthEgtB* (Fig. S4; r.s.m.d. = 1.2 Å for 1709 of 2341 aligned), despite the moderate sequence. The most important differences between the two homologs are their quaternary structures and the organization of the active sites. In the native *CthEgtB* crystals, the four molecules of the asymmetric unit form a tetramer of D2 symmetry, assembled *via* two interfaces, each with a two-fold axis (Figure 2B). The same tetramer is observed in the *CthEgtB*/TMH crystal, but with one of the molecular dyads coincident with a crystallographic 2-fold axis (running along the C-terminal interface). Also in solution, as evidenced by size exclusion chromatography, *CthEgtB* forms a tetramer, in contrast to monomeric *MthEgtB*⁴⁹ (Fig. S5). The N-terminal interface (Fig. 2D) occurs between ends of the α -helix bundle N-terminal domain with an interface area of 1310 Å². Two salt bridges are formed by Arg111 and Asp44, while the following residue pairs form hydrogen bonds: Val114 (backbone) and Glu43, Arg111 and Asp44, Thr110 and Asp444, Thr171 and Leu108 and Asn172 and Arg111 (backbone). The C-terminal interface (Fig. 2C). spans a slightly smaller area (1094 Å²), and the interacting residues are fewer in number and less conserved. Arg376 and Asp 432 form a salt bridge while the residue pairs Ser260 and Lys 388, Tyr308 and Asn392, and two symmetry-related Asn392 residues form hydrogen bonds. Residues Thr/Ser110, Arg111, Asn172, Glu43 are conserved in all type II EgtBs (see discussion) suggesting that in general type II EgtB might be tetrameric.

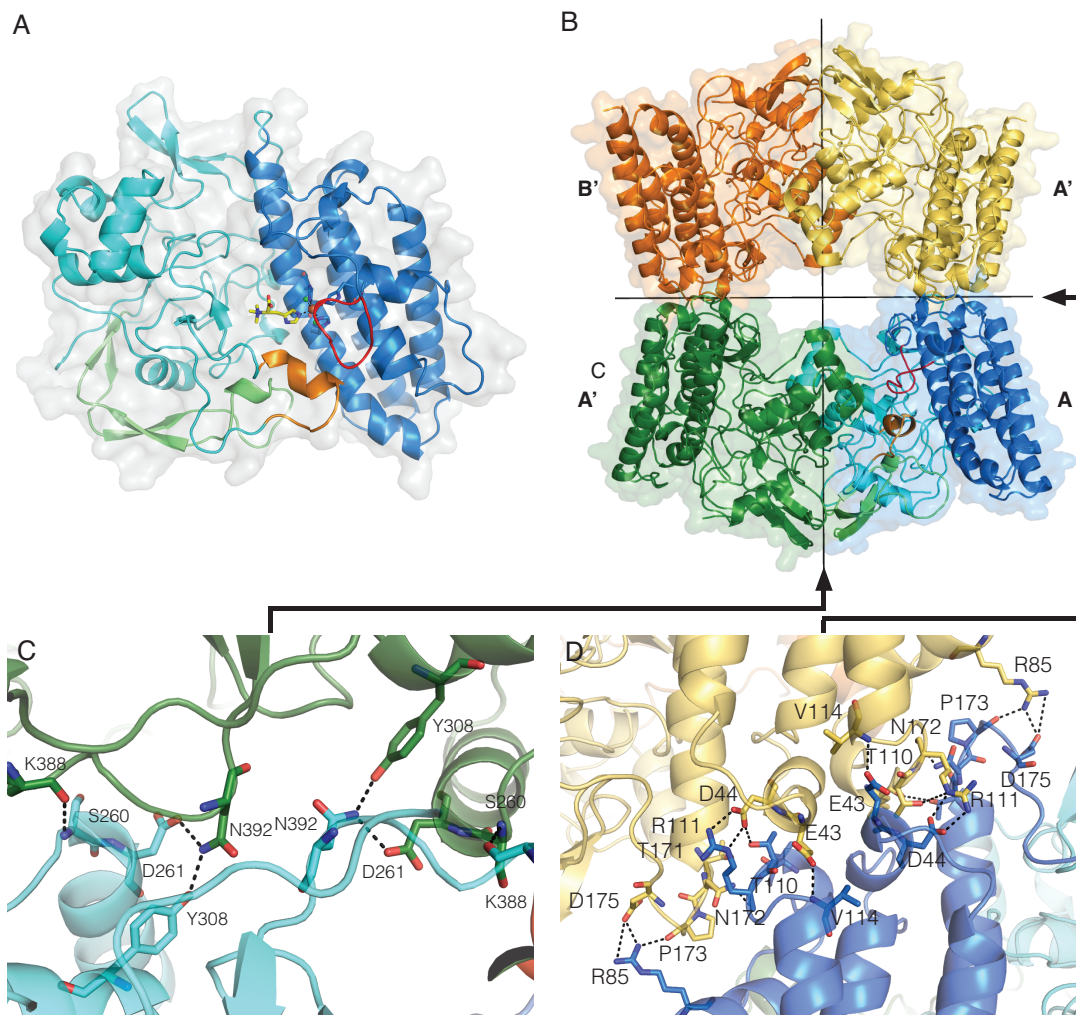


Figure 2: Structural and Oligomeric Analysis of *CthEgtB* in complex with TMH A: Cartoon of the monomeric *CthEgtB*/Fe/TMH structure. The N-terminal DinB_2-like domain (residues 17-176) is shown in dark blue while the C-terminal FGE-like domain is shown in green (residues 198-234) and light blue (residues 235-434). Active site loop 1 (93-99) is in red, and active site loop 2 (378-386) is in orange, close over the active site. The substrate TMH (yellow), iron (orange), chloride ligands (green), iron (brown), and the metal coordinating histidines (blue) are shown to indicate the location of the active site. B: View of the tetrameric structure, with two of the three 2-fold axes shown indicating the two different interfaces, the N-terminal & C-terminal interface. The third dyad is perpendicular to the page. C, D: Close-up views onto the C-terminal (C) and N-terminal (D) interfaces with interacting residues shown in full. Dashed lines indicate H-bonds.

Active site of *CthEgtB*. The structure of *CthEgtB* in complex with TMH shows that the imidazole rings of His62, His153 and His157 form an iron-binding three histidine facial triad, with the N π of the TMH imidazole ring joining as the fourth. The N π of this imidazole ring hydrogen bonds with the backbone carbonyl 385 via a bridging water molecule (Figure 3A). In the native structure loop 2 containing Tyr385 is unresolved, suggesting that the presence of TMH immobilizes this loop. The N- γ -trimethylamino moiety of TMH is boxed in by the aromatic rings of Phe415 and Phe416. Furthermore, each of the three N-methyl groups also interact with the carbonyl groups of either Phe415 (3.2 Å, backbone), Gln156 (3.2 Å, side chain) or Asn414 (3.4 Å, side chain). Comparison to *MthEgtB* shows that both sulfoxide synthases recognize TMH by an analogous set of interactions (Figure 3A & E).

The remaining two coordination sites in the *CthEgtB*/TMH complex are occupied by residual electron density that is best modeled by two chloride ligands: the axial chloride, binding opposite to His156 (Cl_{ax}), and the equatorial chloride, binding opposite to His152 (Cl_{eq}). Modelling the two features as water molecules did not sufficiently reduce the residual electron density. While the modelled iron-Cl distances are slightly longer than expected for iron-Cl bonds ($\text{Fe}-\text{Cl}_{\text{ax}}$, $\text{Fe}-\text{Cl}_{\text{eq}}$: 2.5 Å), it has also been noted that hydrogen bonding interactions to Cl may be involved in lengthening the iron-Cl bond.¹²⁷⁻¹²⁸ During catalysis one of these two sites must coordinate the second substrate cysteine. However, for reasons that may be related to the flexibility of loop 1 and 2 we were unable to obtain crystals of *CthEgtB* with bound cysteine despite an extensive search for appropriate crystallization and soaking conditions. As an alternative, we modelled the *CthEgtB*/TMH/Cys complex (Figure 3C) using the structure of *MthEgtB* in complex with manganese (II), *N*- α -trimethyl histidine (DMH) and γ -GC as a template (Figure 3E). In *MthEgtB* the sulfur atom of γ -GC occupies the axial coordination site (Figure 3E). The cysteinyl carboxylate interacts with Arg87 and Arg90 and the glutamyl moiety salt bridges with Asp416 and Arg420. Finally, the amide function of γ -GC hydrogen bonds with the carboxylate of TMH. The model of the *CthEgtB*/TMH/Cys complex shows that most of these interactions are conserved. The cysteine carboxylate makes a similar interaction with Arg103 and Arg106, and the γ -amino group interacts with the carboxylate of TMH (Figure 3C). As a key difference, the *MthEgtB* residues that are responsible for binding the glutamyl moiety of γ -GC, are replaced by Phe416 and Ala420 in *CthEgtB*. These two mutations provide a clear structural explanation for the distinct substrate specificity among type I and II EgtBs.

The most intriguing structural differences between *MthEgtB* and *CthEgtB* map to the presumed oxygen binding site, which is occupied by Cl_{eq} in the *CthEgtB*/TMH complex (Figure 3B).^{49, 81-82, 124} The side chain hydroxyl-groups of Ser92, Tyr93, and Tyr94 of loop 1 point towards the presumed oxygen binding site. The γ -hydroxyl side chain of Ser92 was resolved in two conformations in which the γ -hydroxyl side chain hydrogen bonds either with the axial or the equatorial chloride ligands ($\text{O}-\text{Cl}_{\text{ax}}$: 2.9 Å, and $\text{O}-\text{Cl}_{\text{eq}}$: 3.1 Å) (Figure 3D). The hydroxyl group of Tyr93 makes no direct contact with Cl_{eq} ($\text{Cl}_{\text{eq}}-\text{O}$: 4.0 Å), but instead hydrogen bonds to the backbone carbonyl of Tyr385 (2.9 Å), and packs against the imidazole ring of TMH ($\text{O}-\text{C}_2$: 3.1 Å). The aromatic ring of Tyr94 makes π - π stacking interaction with Tyr93, and the hydroxyl group of Tyr94 is juxtaposed with the Cl_{eq} ($\text{Cl}_{\text{eq}}-\text{O}$: 3.4 Å). Although this is a long distance for efficient hydrogen bonding, we note that the Tyr94 hydroxyl group is embedded in a largely hydrophobic environment formed by Phe66, Leu382 and Tyr385 providing no alternative hydrogen bonding partners. Hence, slight structural adjustments in the reactive complex would allow strong interactions between Tyr94 and iron-bound oxygen.

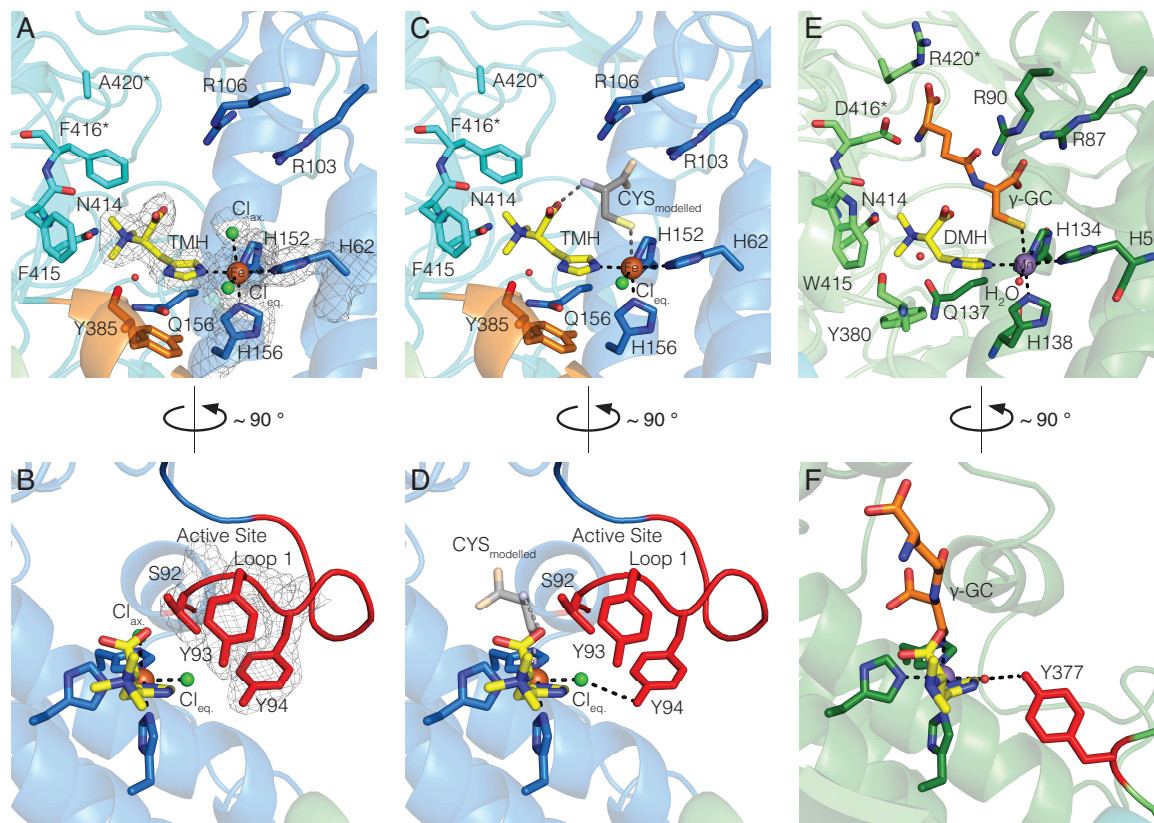


Figure 3. Structural Comparison of the active site of *CthEgtB* with *MthEgtB*. Important metal or substrate binding residues are shown as sticks and are colored according domain or segment. iron is shown in brown, Mn in purple, DMH in yellow, modelled CYS in gray, γ -GC in orange, Cl ions and water as spheres in green and red respectively. Upper Panel: A front on view of the active site, focusing on the metal co-ordination sphere and TMH and γ -GC/Cys binding sites. Active site loop 1 has been omitted for clarity. Lower panel: This view focuses on the oxygen binding site and is rotated by approximately 90° to the left from the front view. Active site loop 2 has been omitted for clarity. A & B: Active site of *CthEgtB* in complex with iron (II), and TMH. The 2m|Fo|-D|Fc| omit map for iron, the three histidine ligands, TMH, S92, Y93 & Y94 (red) and the axial (Cl_{ax}) and equatorial chloride (Cl_{eq}) is contoured at σ -level = 1. C & D: Cysteine is modelled into to the active site of *CthEgtB* based on the location of the cysteinyl moiety of γ -GC in *MthEgtB*. Proposed interactions are shown by dashed lines. E & F: Active site of *MthEgtB* (4X8D) with Y377 pointing towards the proposed oxygen binding site occupied by a water molecule (red sphere).

The apparent active site geometry of *CthEgtB* raises two important propositions. Firstly, the presence of Tyr93 and Tyr94 close to the presumed oxygen-binding site is highly suggestive of a catalytic role for one or both of these residues. This is interesting because the oxygen-binding pocket of *MthEgtB* contains only one tyrosine (Tyr377). Superposition of the *MthEgtB* with *CthEgtB* structures shows that the phenol function of Tyr377 is positioned roughly between the phenol functions of Tyr93 and Tyr94 (Figure 4A). Therefore, based on structural comparisons alone it is impossible to decide which of the two Tyr residues in *CthEgtB* could assume the same catalytic role as Tyr377 in *MthEgtB*. Secondly, comparison of the *CthEgtB* native structure and that of the *CthEgtB*/TMH complex show that TMH-induced loop-folding converts a wide-open crevice to a tightly closed pocket. In the closed structure the presumed oxygen-binding pocket is completely engulfed by the metal center, the substrates and the side chains of Ser92, Tyr93 and Tyr94, suggesting that efficient oxygen-binding may require unfolding of loop 2. Hence, the two *CthEgtB* structures provide evidence that

loops 1 and 2 are flexible, and that their folding and unfolding may be obligatory steps in each catalytic cycle to allow substrate binding and product release. In *MthEgtB* oxygen binds to the equivalent equatorial coordination site,^{49, 81-83} but the shape of this pocket is different. This pocket is connected to the protein exterior by a narrow water-filled tunnel. Crystal structures of *MthEgtB* in native form and in complex with TMH or with TMH and γ -GC show that this tunnel does not change shape upon substrate binding, suggesting that oxygen uptake does not require large-scale conformational change.⁵⁰

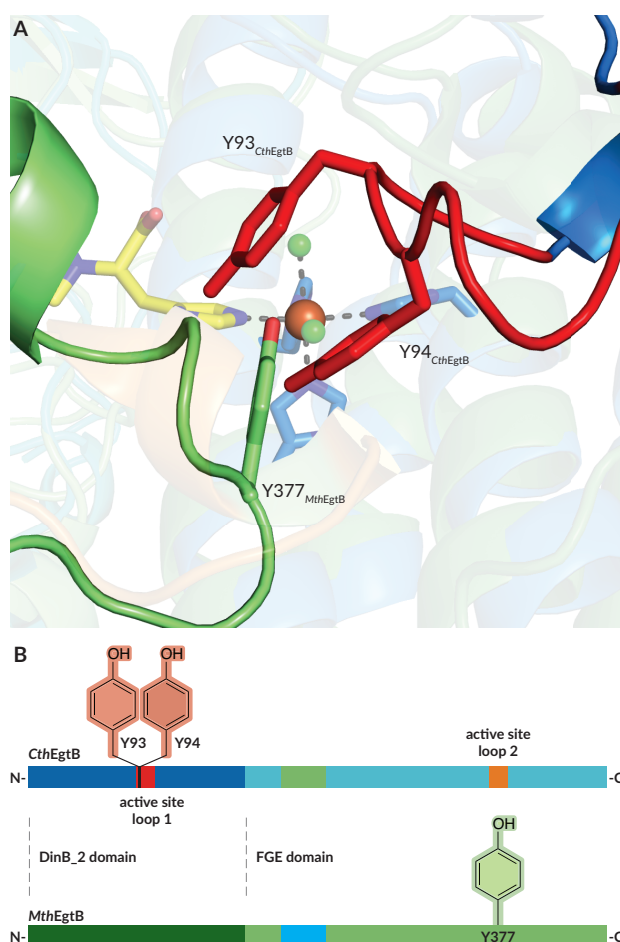


Figure 4. Comparison of the catalytic residues of *CthEgtB* and *MthEgtB* A): Superposition of the *MthEgtB* (green) and *CthEgtB* (red) active sites. The iron coordination sphere of iron in the *CthEgtB* is shown faintly in the background. Tyrosine residues from *MthEgtB* (green) and *CthEgtB* (red) with an assigned catalytic function are shown in full. B): Schematic representation of the architecture of *MthEgtB* & *CthEgtB*.

Catalytic activity of *CthEgtB* and variants thereof. We characterized the catalytic activity of *CthEgtB* using the same HPLC-based kinetic assay as previously developed for *MthEgtB* (Table 1).⁴⁹⁻⁵⁰ In brief, reactions containing TMH, cysteine, Fe sulfate, tris(2-carboxyethyl)phosphine (TCEP) and ascorbate were initiated by addition of purified *CthEgtB*. Reaction products were analyzed by ¹H NMR and RP-HPLC. An initial experiment showed that *CthEgtB* accepts cysteine to make sulfoxide **4** (Figure 1) but cannot turnover γ -GC. By monitoring time-dependent production of **4** we determined k_{cat} for sulfoxide production ($k_{cat,4}$) and K_M for TMH ($K_{M,TMH}$) in the presence of 0.5 mM cysteine (Figure S8a). We also monitored the consumption of

cysteine to determine $k_{\text{cat,cys}}$ and K_M for cysteine ($K_{M,\text{cys}}$) in the presence of 0.2 mM TMH (Figure S8b). ¹H NMR analysis of reactions that initially contained 1 mM TMH and 0.5 mM cysteine showed that wild type *CthEgtB* oxidizes about 20 % of cysteine to cysteine sulfinic acid, and 80 % to sulfoxide **4** (Figure S5 & S6). The same side reactivity has been reported for *EgtB* from *Burkholderia thailandensis*,¹²⁵ and *OvoA* from *Erwinia tasmaniensis*.^{123, 129} Mechanistic investigations on *MthEgtB* and on *OvoA* from *Erwinia tasmaniensis* showed that this side activity occurs because an early reaction intermediate can react either to form the sulfoxide or to form thiol dioxide.^{50, 123}

Table 1^[a]

Enzyme	$k_{\text{cat,cys}}$ ^[b] [s ⁻¹]	$K_{M,\text{cysteine}}$ [10 ⁻⁶ x M ⁻¹]	$k_{\text{cat,4}}$ ^[c] [s ⁻¹]	$K_{M,\text{TMH}}$ [10 ⁻⁶ x M ⁻¹]	$k_{\text{cat,cys}}/k_{\text{cat,4}}$
<i>CthEgtB</i> _{wt}	0.14 ± 0.05	27 ± 3	0.2 ± 0.01	65 ± 2	1
<i>CthEgtB</i> _{Y93F}	0.16 ± 0.03	120 ± 30	0.0004 ± 0.0001	5 ± 1	400
<i>CthEgtB</i> _{Y94F}	0.021 ± 0.004	34 ± 12	0.003 ± 0.001	35 ± 10	7
<i>CthEgtB</i> _{Y93F, Y94F}	0.049 ± 0.003	90 ± 15	< 0.0001	n.d. ^[d]	> 500

^[a]Kinetic parameters were determined in the presence of 4 μM FeSO₄, 2 mM TCEP, 2 mM ascorbate, 100 mM NaCl, 100 mM HEPES, pH 8.0 at 26 °C. Rate determined by monitoring the consumption of cysteine^[b], or the production of sulfoxide **4**^[c]. ^[d]n.d. = not determined.

To probe the catalytic contributions of Tyr93 and Tyr94 we mutated either, or both residues to Phe. ¹H NMR analysis of reaction mixtures showed that *CthEgtB*_{Y93F} and *CthEgtB*_{Y94F} both produced cysteine sulfinic acid as the main product, and almost no sulfoxide **4** (Figure S10 & S11). Subsequent determination of the Michaelis-Menten parameters showed that the Tyr93Phe mutation slightly increased the apparent K_M for cysteine and reduced the apparent K_M for TMH by ten-fold (Table 1). The phenol function of Tyr93 makes direct contact with C₂ of the TMH imidazole ring (3.1 Å). This interaction may be unfavorable and hence deletion of the hydroxyl group in *CthEgtB*_{Y93F} could stabilize the enzyme/substrate complex. Alternatively it is possible that $K_{M,\text{TMH}}$ is lowered due to a kinetic effect. Because the mutation reduced k_{cat} (see below) relative to the rates of binding and unbinding of TMH the value of $K_{M,\text{TMH}}$ could approach that of a true binding constant ($K_{D,\text{TMH}}$). Mutation of Tyr94 reduced $K_{M,\text{TMH}}$ only by two-fold, and caused no significant effect on $K_{M,\text{Cys}}$. Hence, the observed reduction of sulfoxide synthase activity is not due to defects in cysteine- or TMH-binding.

On the other hand, the turnover rates ($k_{\text{cat,4}}$) for sulfoxide production were strongly affected in all variants. Mutating Tyr93 reduced $k_{\text{cat,4}}$ by 500-fold, but left the rate of cysteine consumption ($k_{\text{cat,cys}}$) essentially unchanged. As evidenced by the NMR analysis discussed above, *CthEgtB*_{Y93F} produces almost exclusively cysteine sulfinic acid, instead of sulfoxide **4**. Mutating Tyr94 reduced sulfoxide synthesis by 70-fold and cysteine consumption by 7-fold. The double mutant showed no detectable sulfoxide production, and a cysteine consumption activity only 3-fold less than that of wild type. Summarizing these results we arrive at the following conclusions: a) the phenol functions of both active site tyrosines are important for catalysis but

are unimportant for substrate binding; b) Tyr93 is more important in determining the product specificity than Tyr94, c) Tyr93 is entirely dispensable for oxidative cysteine consumption, d) Tyr94 plays a significant role in oxidative cysteine consumption, e) in the absence of Tyr93, Tyr94 supports sulfoxide production at a low but observable rate.

Further evidence that both tyrosines are involved in catalyzing sulfoxide production comes from the fact that both mutations induce a significant kinetic solvent isotope effect (KSIE). Measuring sulfoxide production rates in the presence of saturating substrate concentrations in either H₂O or D₂O revealed a KSIE of 1.3 ± 0.2 for the wild type, 2.8 ± 0.2 for *CthEgtB*_{Y93F} and 3.9 ± 0.2 for *CthEgtB*_{Y94F} (Figure S12) The effects of the Tyr93Phe mutation in catalysis is very similar to those observed for the mutation of Tyr377 in *MthEgtB*, and Tyr417 in *OvoA*.^{50, 72} In both cases elimination of this catalytic acid reduced sulfoxide synthase activity dramatically, induced a significant KSIE, but did not affect the ability to use oxygen and oxidize thiols. However, the presence and catalytic importance of a second tyrosine in *CthEgtB* has no correspondence in previously characterized sulfoxide synthases.

Discussion

CthEgtB and *MthEgtB* catalyze almost identical reactions. Therefore, the structural differences between the two active sites are both surprising and informative. The current proposals explaining the catalytic mechanism of sulfoxide synthases were all developed based on the structure of *MthEgtB*. As discussed below, reevaluation of these proposals in view of the structure of *CthEgtB* provides a new test that may help to distinguish between different models. In the second part of the discussion we introduce a classification of all known ergothioneine biosynthesis sulfoxide synthase into five types, and we propose a possible evolutionary trajectory by which this diversity among extant sulfoxide synthases may have emerged.

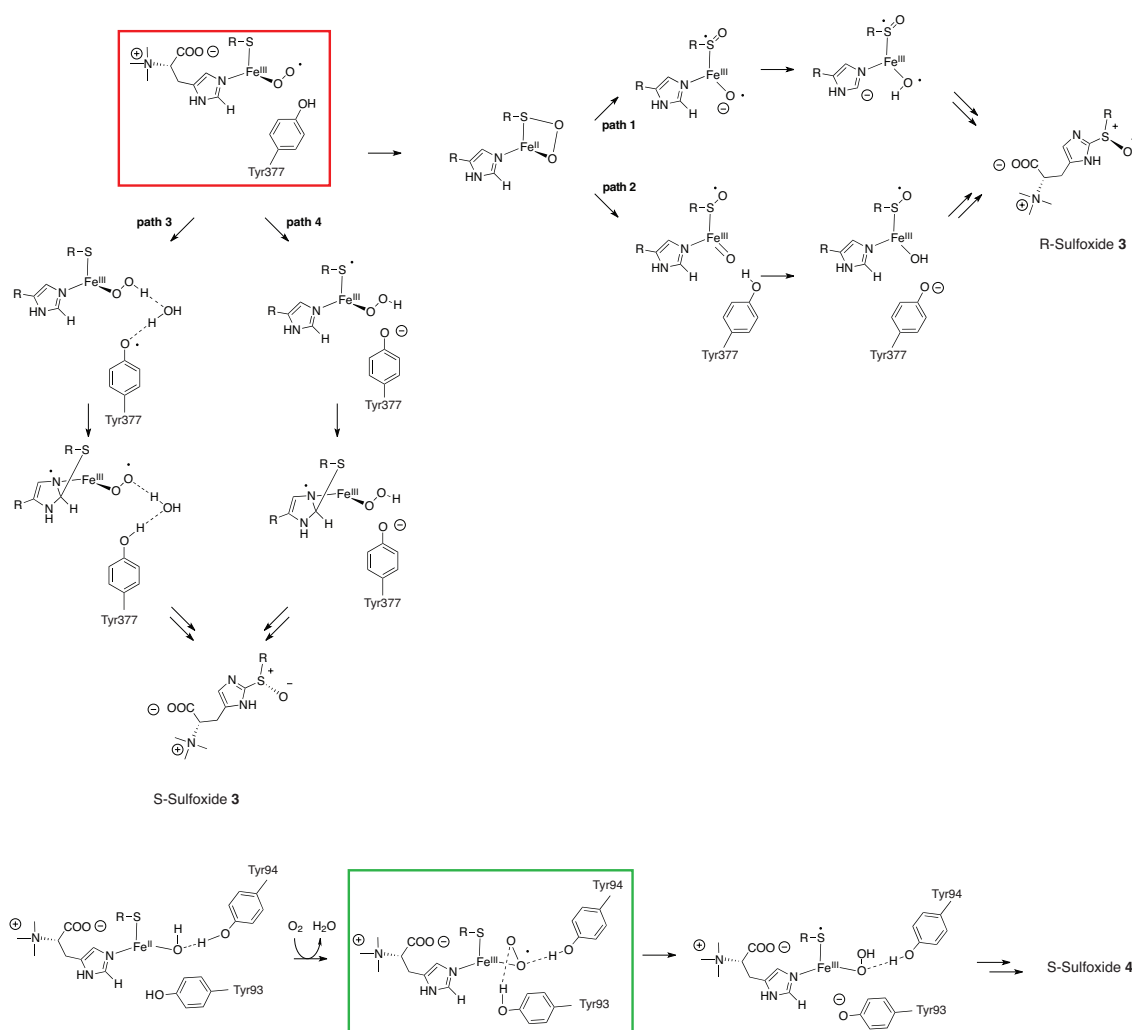


Figure 5. Top: Catalytic mechanisms proposed for *MthEgtB*, adapted from Ref. 1, 20 – 22. Two computational studies propose that the reactive enzyme/substrate complex (*MthEgtB*/Fe/TMH/ γ -CG/O₂ complex, red box) first reacts via oxygen atom transfer to the sulfur atom of γ -GC (paths 1 & 2). Both models predict the formation of sulfoxide **3** with R-configuration at the sulfur atom.^{82,124} Two alternative models suggest that the reactive complex first reacts via reduction of the superoxide ligand, followed by C-S bond formation between TMH and γ -GC (paths 3 & 4). This trajectory leads to formation of the sulfoxide **3** with S-configuration.^{49,81} Bottom: proposed roles of Tyr93 and 94 in the reactive complex of *CthEgtB* (*CthEgtB*/Fe/TMHCys/O₂ complex, green box, see text).

***MthEgtB*-based proposals for the catalytic mechanism of sulfoxide synthases.** The four main proposals for the mechanism of *MthEgtB* catalyzed C-S bond formation (paths 1 – 4, Figure 5) all assume that the enzyme combines with the substrates TMH, γ -GC and oxygen to form an iron (III) superoxide complex as a first intermediate (red box, Figure 5).^{49, 81-82, 124} Even though enzyme-bound iron (III) superoxides are notoriously difficult to detect, growing evidence suggests that these species form as early intermediates in many non-heme iron enzymes.^{85, 91-93, 130-131} The disagreement among the four models starts with the next step. Two computational studies concluded that oxygen transfer to the sulfur atom of γ -GC occurs first (paths 1 & 2, Figure 5).^{82, 124} Both models suggest that Tyr377 serves as a proton donor to the iron coordinated oxygen atom (proximal oxygen), after cleavage of the O-O bond. Subsequent C-S bond formation between γ -GC and TMH produces a sulfoxide with *R*-configuration (*R*-sulfoxide **3**). This prediction is the weakest aspect of these models because the enzymatic product has been determined to be the *S*-sulfoxide **3**.¹³² The remaining two models, one based on structural and biochemical analysis (path 4), the other based on QM/MM calculations (path 3) favor a reaction path where the iron (III) superoxide is first reduced by accepting a proton and an electron. In these models C-S bond formation precedes S-O bond formation which leads to the correct sulfoxide stereochemistry. The two models disagree over the specific role of Tyr377. The computational analysis identified a low-energy pathway along which the iron (III) superoxide is reduced by hydrogen atom transfer from Tyr377 (path 3, Figure 5). However, the calculations also showed that the radical character transfers from Tyr377 to the sulfur atom of γ -GC en route to the transition state of the C-S bond forming step. Hence, it remains an open question as to whether oxidation of Tyr377 represents a necessary – on the pathway – step, or whether γ -GC and Tyr377 share radical character in an equilibrium prior to C-S bond formation. An important feature of this computational model is that in the reactive complex Tyr377 retracts from the position observed in the crystal structure and only interacts with the iron:oxygen species via an intervening water molecule (Figure 5, path 3). There is no empirical evidence for such a movement. In fact, the active site geometry of *MthEgtB*, and specifically the position of Tyr377 were found invariant among different complexes of *MthEgtB*.⁴⁹ The last model (path 4, Figure 5) suggests that the iron (III) superoxide is reduced by proton transfer from Tyr377 and electron transfer from the iron-coordinated γ -GC.⁵⁰ In this way γ -GC is directly activated as a radical that can attack the imidazole ring of TMH.

Importantly, the four models assign different functions to Tyr377 in the early reaction stage, depending on whether its phenol ring interacts with the proximal or the distal oxygen of the iron (II) superoxide species, or whether there is no direct interaction at all. The position of Tyr377 as determined by crystallography does not allow a distinction between these models. As discussed below *CthEgtB* provides a related system featuring a tyrosine (Tyr93) residue with the same catalytic functions as Tyr377 that cannot interact with the proximal oxygen atom.

Catalytic roles of Tyr93 and Tyr94 in *CthEgtB*. Our Structural and kinetic observations implicate Tyr93 and Tyr94 as important catalytic residues. The phenotypes of the Tyr93Phe mutation – i) dramatic reduction of sulfoxide synthase activity, ii) retainment of cysteine oxidation activity, and iii) introduction of a significant KSIE on the residual sulfoxide synthase activity – match the phenotypes induced by the Tyr377 mutation in

MthEgtB, and the Tyr417 mutation in *OvoA*.^{50,72} This evidence supports the conclusion that Tyr93 serves the same catalytic function as Tyr377 and Tyr417 despite their different location in the primary sequence, and the different orientation in the active site (Figure 3C & D).

The second tyrosine in the oxygen binding pocket of *CthEgtB*, Tyr94, introduces a new aspect to the catalytic mechanism of sulfoxide synthases. Tyr94 is juxtaposed to the coordination site that is either occupied by water, oxygen or, as seen in the crystal structure, by Cl_{eq} (Figure 3D). This position strongly implicates Tyr94 as a hydrogen bonding partner to iron (II)-bound water or to the proximal oxygen of iron (III)-bound superoxide (green box, Figure 5). This interaction is likely strong, because Tyr94 has no apparent alternative hydrogen bonding partner. Mutation of this residue affected sulfoxide production by 70-fold, but also reduced oxidative cysteine consumption by 7-fold, suggesting that Tyr94 assists the catalytic cycle in two independent steps. First, Tyr94 plays a supporting role in guiding the iron (III) superoxide towards the sulfoxide production pathway. One way how Tyr94 could do so is by hydrogen bonding with the proximal oxygen of the iron (III) superoxide species (Figure 5). This acidic interaction would certainly increase the electron affinity of the iron (III) superoxide species and thereby facilitate its reduction in the first catalytic step. The observation of a sizable KSIEs in the Tyr94Phe mutant is consistent with this interpretation: this mutation could make the iron (III) superoxide species less oxidative, rendering proton-coupled electron transfer to this species rate limiting. An alternative interpretation of the *CthEgtB*_{Y94F} phenotype would be that Tyr94 acts as a hydrogen bond donor to Tyr93 to activate the later as a catalytic acid. However, in the crystal structure the two phenol groups of Tyr93 and Tyr94 are separated by 4 Å, and in the conformation given, do not possess a geometry that could facilitate hydrogen bonding. Direct hydrogen bond between the two residues might be weak at best.

Reduced cysteine consumption activity in *CthEgtB*_{Y94F} indicates that Tyr94 is also involved in oxygen binding and activation. The Tyr94Phe mutation could affect this activity in two ways: the lack of a proton donor could slow down protonation and removal of the iron (II) coordinated hydroxide and thereby slow down oxygen binding (bottom, Figure 5). Alternatively, the lack of a hydrogen bond could destabilize the iron (III) superoxide. A similar interaction has been observed in human and murine cysteine dioxygenase (CDO, EC 1.13.11.20).⁷⁵ Although CDOs are entirely unrelated to sulfoxide synthases, the local geometries around the catalytic iron center are remarkably similar.⁵⁰ CDO also coordinate iron (II) by three-his facial triads. In the reactive complex the remaining coordination sites are filled by amine- and thiolate-ligands from the substrate cysteine. Finally, addition of oxygen as the last ligand gives rise to an iron (III) superoxide. CDO also contains a second-coordination sphere tyrosine (Tyr157) that hydrogen bonds with the iron-coordinated oxygen.¹³³⁻¹³⁴ Mutation of Tyr157 to Phe resulted in 8 - 20-fold reduction of dioxygenase activity, showing that Tyr157 – similar to Tyr94 in *CthEgtB* – is important but not crucial for the oxygen activation by CDO.^{75,135} The double mutant *CthEgtB*_{Y93F, Y94F} lacks any measurable sulfoxide synthase activity, and oxidizes cysteine 3-fold less efficiently than wild type or *CthEgtB*_{Y93F} (Table 1). This activity pattern shows that the Tyr93Phe and Tyr94Phe mutations lead to additive effects, and that the two Tyr residues have limited capacity to compensate for the absence of each other. Hence the two tyrosines must play complementary roles that are related to their specific position relative to the oxygen binding site.

A tyrosine residue near the oxygen binding site of non-heme iron enzyme automatically raise the question as to whether this residue participates in the redox chemistry. Oxygen can react with ferrous iron to produce highly reactive species that could oxidize the comparatively electron-rich tyrosine side chain. Oxidation of active site residues could either be part of the catalytic mechanism, or could lead to maturing or deactivating automodifications.¹³⁶ The endoperoxide forming enzyme FtmOx1 provides an example of redox active tyrosine involved in the catalytic mechanism.¹³⁷ In the taurine dioxygenase TauD, oxidation of active site Tyr73 also occurs but is part of a deactivating side reaction.¹³⁸ In mammalian CDO the active site Tyr157 is cross-linked with a nearby Cys residue (Cys93) as a result of an oxygen-dependent side reaction.¹³⁹ Since mutation to Phe conserves most of the CDO activity, Tyr157 is unlikely to participate as an essential redox partner during cysteine deoxygenation. On the other hand, there are also enzymes that use tyrosine side chains to activate and orient iron-bound oxygen species with no apparent redox participation. A computational study on the algal prolyl-4-hydroxylase concluded that the active site tyrosine (Tyr140) controls the reactivity of the oxo-ferryl species by hydrogen bonding to the iron bound oxygen atom.¹⁴⁰ How Tyr140 evades oxidation remains an open question. The same is true for the sulfoxide synthases *MthEgtB* and *CthEgtB*. Even though Tyr377 and Tyr93/Tyr94 appear intimately involved with oxygen binding and activation, we have no evidence that these residues participate in any redox activity, and it is not yet clear why they do not.

In summary, the structural and kinetic evidence discussed above is consistent with the following interpretation: Catalysis by *CthEgtB* requires two hydrogen bond donors in the oxygen binding site. Tyr94 hydrogen bonds with the proximal oxygen of the iron (III) superoxide complex to render this species more oxidative. Tyr93 transfers a proton to the distal oxygen of the iron (III) superoxide. This transfer is coupled to electron transfer from the substrate cysteine to form an iron (III) hydroperoxide and a thiyl radical (Figure 5). The relative positions of Tyr93 and Tyr94 compounded by their specific functions provide evidence that proton transfer occurs to the distal and not to the proximal oxygen. Analogous conclusions based on *MthEgtB* would be more ambiguous because in the available crystal structures Tyr377 adopts a position that allows interaction with either atom of the iron-coordinated superoxide (Figure 4). From this point forth, the main tenets of any mechanistic proposal will have to be consistent with the structures of both sulfoxide synthase types (*MthEgtB* and *CthEgtB*). We believe that this test will be of significant help in the elimination or validation of mechanistic proposals for the sulfoxide synthase catalyzed reaction.

The emergence of sulfoxide synthase diversity. The differences in the active sites of *CthEgtB* and *MthEgtB* are also interesting from an evolutionary perspective. Apparently, the family of sulfoxide synthases does not comply with the general expectation that essential catalytic residues and active site geometries are conserved among enzymes with similar functions.¹¹² Residues Tyr377 and Tyr93 map to completely different locations within the protein scaffold (Figure 4B), despite their identical roles in catalysis. Relocation of catalytic residues in enzymes with identical function has been documented both in natural and laboratory protein evolution.^{112, 141-146} The importance of spatial conservation – rather than conservation in the primary

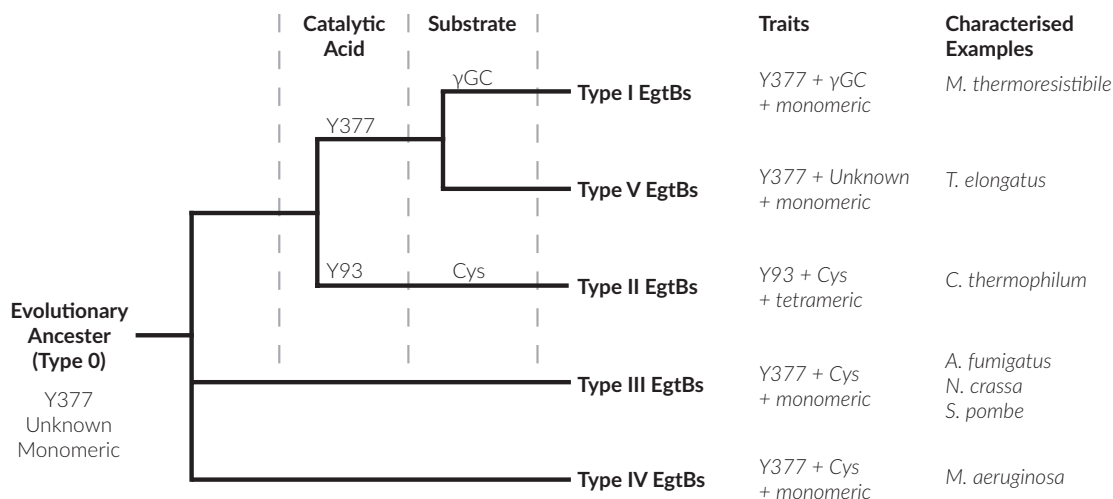
sequence – for catalysis has been noted. *Functional residue hopping* is typically facilitated by conservation of the rest of the active site. By contrast, the migration of a catalytic tyrosine in *CthEgtB* (Tyr93) with respect to *MthEgtB* (Tyr377) is accompanied by the introduction of a second catalytic tyrosine (Tyr94), change of substrate specificity, and conversion of a stiff active site (as in *MthEgtB*) to an open active site covered by two flexibles. This complete remodeling raises the question as to what evolutionary path could lie between the two catalyst types, and what evolutionary pressure may have caused the conversion from one to the other type.

To address this question and to probe the evolutionary relationships of sulfoxide synthases, we constructed a phylogenetic tree based on seven EgtB homologs with characterized activities (Figure 6A).^{40, 47, 49-50, 118-37} Also, we inspected the alignment of the respective sequences for the absence or presence of motifs that are important for function in *MthEgtB* or *CthEgtB* (Figure 6B). Based on this analysis we distinguished five types of ergothioneine biosynthetic sulfoxide synthase (types I –V). Type I sulfoxide synthases utilize γ -GC, and type II – IV use cysteine as sulfur donor.^{40, 47, 115-116, 118, 125-126} Most type IV sulfoxide synthases are OvoA-like enzymes that use histidine and cysteine to produce a sulfoxide intermediate in ovothiol biosynthesis.⁶³ However, in the presence of TMH OvoA from *E. tasmaniensis* was shown to produce sulfoxide **3**, highlighting the functional similarity of between EgtBs and OvoAs.⁷⁰ Indeed, the OvoA homologs from *Microcystis aeruginosa* and other cyanobacteria have evolved to make sulfoxide **3** as their main physiological product. Hence, we included this cyanobacterial enzyme as representative of type VI ergothioneine biosynthetic sulfoxide synthase.⁴⁷ Finally, we also included the cyanobacterial EgtB from *Thermosynechococcus elongatus* as a representative of type V sulfoxide synthases. This type has not been fully characterized yet. Despite significant similarity to type I enzymes, the type V EgtBs accept neither cysteine nor γ -GC as substrate (Stampfli & Seebeck, unpublished).

Comparing these five enzyme types we arrived at the following conclusions (Figure 6): i) the motifs for iron and TMH binding are conserved in all types, ii) tyrosines equivalent to Tyr377 in *MthEgtB* are conserved except in type II; iii) the γ -GC-recognition motif DXXR motif is exclusive to type I; iv) the RXXR motif that is responsible for γ -GC-binding in *MthEgtB*, and for cysteine-binding in *CthEgtB* occur in type I, II, and with slight variation (KXXR) in type V. In contrast, type III and type IV sulfoxide synthases lack this motif, suggesting that these enzymes bind cysteine in a completely different mode. Since most types use a Tyr377-like catalytic acid and are monomeric enzymes, we conclude that an ancestral sulfoxide synthase most likely shared these properties. Type II-IV use cysteine as sulfur donor, arguing that the ancestral enzyme too used this very common sulfur metabolite. But why is the cysteine-binding pocket not conserved among the three types? Why should an ancestral enzyme undergo dramatic active site remodeling to arrive at a functionally equivalent solution? A more likely explanation for the emergence of these enzymes with different substrate binding pockets would be that the ancestral sulfoxide synthase used a different sulfur donor than cysteine. Possible candidates could be hydrogen sulfite, thiosulfate or thiophosphate. This scenario suggests that this ancestral enzyme type (type 0) first entered different bacterial and fungal lineages and then adapted

independently to utilize cysteine or γ -GC for ergothioneine biosynthesis. Consequently, there may not be an evolutionary path directly connecting extant (type I - V) sulfoxide synthases.

A



B

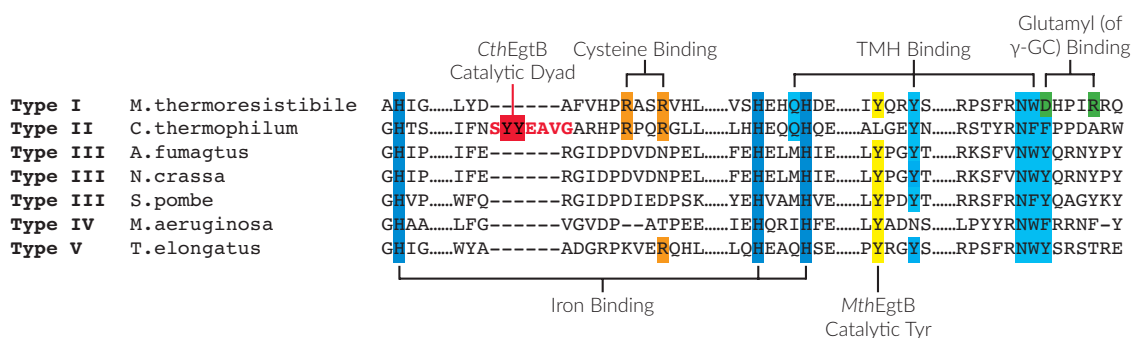


Figure 6. A: A qualitative phylogenetic tree containing characterised EgtB homologs *C. thermophilum*, *M. thermoresistibile*¹⁴⁷⁻¹⁴⁸, *M. aeruginosa*¹⁴⁹, *A. fumigatus*¹¹⁸, *S. pombe*⁴⁰, *N. crassa*¹¹⁶ and uncharacterised *T. Elongatus*. (... = omitted residues; --- = gaps) B: Sequence alignment of characterised EgtB homologues *C. thermophilum*, *M. thermoresistibile*¹⁴⁷⁻¹⁴⁸, *M. aeruginosa*¹⁴⁹, *A. fumigatus*¹¹⁸, *S. pombe*⁴⁰, *N. crassa*¹¹⁶ and uncharacterised *T. Elongatus*. Residues highlighted in red indicate active site loop 1 of CthEgtB. Key residues for binding and catalysis are highlighted and labelled.

Conclusion

Based on the crystal structure of *CthEgtB* in complex with iron (II) and TMH we identified Tyr93 and Tyr94 as essential catalytic residues. Mutation of either residue to phenylalanine significantly reduced sulfoxide synthase activity, whereas only the Tyr94Phe mutation affected the enzymes ability to consume cysteine. These different phenotypes, compounded by their different positions in the oxygen binding pocket provide evidence that Tyr93 serves as a proton donor and Tyr94 serves as a hydrogen bond donor, both facilitating the reduction of the initial iron (III) superoxide species. This reduction is the first catalytic step towards sulfoxide production. Comparison with the structure of EgtB from *M. thermoresistibile* (*MthEgtB*) and with primary sequences of other bacterial and fungal homologs revealed that the class of ergothioneine biosynthetic sulfoxide synthases is characterized by remarkable active site diversity. Detailed characterization of type I - V sulfoxide synthases will provide a powerful approach understand the catalytic mechanism of oxidative C-S bond formation.^{123, 147-148} Finally, the observed diversity indicates that these sulfoxide synthase types may have emerged from an ancestral enzyme with different substrate specificity than any known extant homolog.

Addendum on CthEgtB Substrate Selectivity

Does EgtB from *Chloracidobacterium thermophilum* accept γ -glutamyl cysteine as a substrate?

Following the publication of the preceding story in *J. Am. Chem. Soc.*, a second paper on CthEgtB was published by the Liu group.¹¹³ This paper states that CthEgtB utilizes γ -GC as a substrate with a k_{cat} of $17.5 \pm 0.4 \text{ min}^{-1}$, less than two times slower than the k_{cat} for the native substrate cysteine ($26.6 \pm 0.7 \text{ min}^{-1}$) and with K_M values of $5.9 \pm 0.9 \text{ mM}$ and $205 \pm 18 \text{ }\mu\text{M}$ respectively (Table 1). We found this result surprising, as we had not observed such activity in our studies. Additionally, the *in vitro* characterization of another type II EgtB, EgtB from *Burkholderia pseudomallei* did not produce any detectable amount of the γ -GC-sulfoxide product in the presence of 1 mM γ -GC.¹⁵⁰

We decided to investigate this claim with our HPLC assay. In contrast, the Liu group quantifies activity with an oxygen electrode and therefore measures the sum of all oxygen-consuming reactions rather than the consumption or production of a single substrate or product, as in our method.¹⁵⁰ CthEgtB.WT was incubated with γ -GC at three different concentrations, 1, 3 and 10 mM, along with a control reaction containing 1 mM cysteine. The reaction was followed over 10 minutes (Figure 1A). All reactions showed formation of a product. However, within the measured time frame, the product of the γ -GC reactions had the same retention time as the cysteine-sulfoxide in the control reaction (Figure 1B). This indicates that γ -GC contains a significant amount of cysteine as an impurity. The reaction course with 1 mM γ -GC also supports this, as the reaction rate is not linear over 10 minutes as the cysteine impurity becomes limiting. This allowed estimation of the cysteine impurity to be around 2.5-3%, which is consistent with the analytical data sheet for the γ -GC purchased from BACHEM ((Des-Gly)-Glutathione (reduced) ammonium salt, product number: 4025354.0250) which gives a purity by HPLC of 98%.

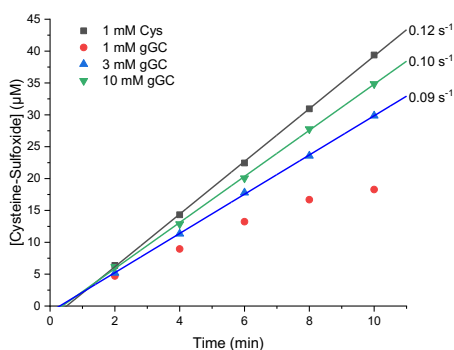


Figure 1A. CthEgtB.WT + Cys or γ -GC. Assay conditions included: 100 mM phosphate, pH 8, 100 mM NaCl, 12 mM TCEP, 2 mM ascorbic acid, 6 μM FeSO_4 and 1 mM TMH. The sulfur donor was varied with 1 mM Cys or γ -GC at 1, 3 and 10 mM. The reaction was initiated with 0.6 μM of CthEgtB.WT and aliquots were taken at 2, 4, 6, 8 and 10 minutes.

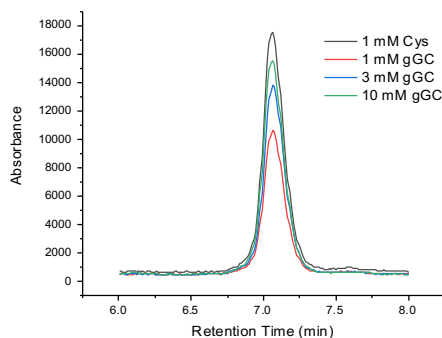


Figure 1B. HPLC trace of the reaction of CthEgtB with γ -GC & Cys, quenched at 6 minutes. The products of the γ -GC reactions have the same retention time as the cysteine-sulfoxide product of the cysteine control reaction.

The rates for the reaction of *CthEgtB* with 3 and 10 mM γ -GC were linear within the measured time frame, giving rates of cysteine-sulfoxide production of 0.09 s^{-1} and 0.10 s^{-1} respectively. These rates are very similar to that measured for 1 mM Cys as a substrate (0.12 s^{-1}). These results indicate that the rates of oxygen consumption measured by the Liu group for “ γ -GC-sulfoxide production” could instead correspond to turnover of the cysteine impurity to give the cysteine-sulfoxide.

The reactions of *CthEgtB* with Cys and γ -GC were also analyzed after one hour. In the reaction with γ -GC the HPLC traces showed the presence of a slight shoulder on the cysteine-sulfoxide peak. This shoulder had a slightly shorter retention time than the Cys-sulfoxide, and was not present in the cysteine control, which is consistent with the γ -GC sulfoxide product, indicating that perhaps γ -GC could be slowly consumed by *CthEgtB*. (Figure 2).

To test this observation the following experiments were set up: (1) *CthEgtB* + γ -GC at 1, 3 and 10 mM as previously, yet with a higher enzyme concentration and longer assay times to determine the rates produced; (2) Control reactions with *MthEgtB* an excess of γ -GC and defined amounts of TMH to quantify amounts of γ -GC-sulfoxide produced. (3) A reaction with γ -GC with both *CthEgtB* and *MthEgtB* to produce a sample in which we know both sulfoxides are present.

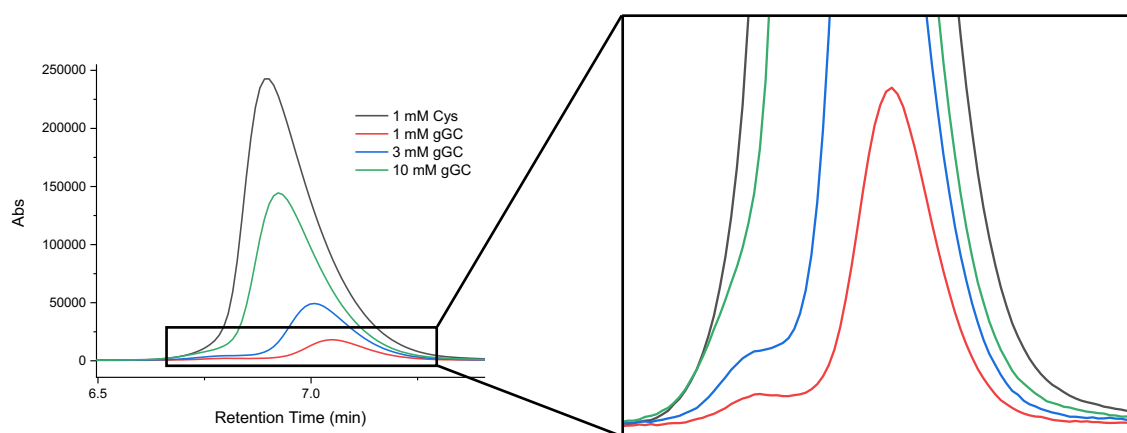


Figure 2. HPLC trace of the reaction of *CthEgtB*.WT with γ -GC & Cys, quenched at 1 hour. The γ -GC reactions appear to give a second product that appears as a shoulder on the cysteine-sulfoxide. This shoulder on the cysteine-sulfoxide peak is more apparent in the zoom in (right).

The optimized conditions for the reaction of *CthEgtB* with γ -GC show that with higher enzyme concentrations and longer time points (0.5, 1, 1.5, 2, 2.45 hours and overnight), the observed shoulder becomes much more apparent and grows with time, while that of the cysteine-sulfoxide also slightly increases (Figure 3).

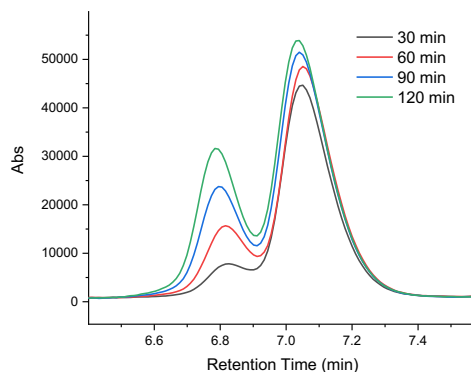


Figure 3. HPLC traces for the reaction of *CthEgtB*.WT with 3 mM γ -GC. Assay conditions included: 100 mM phosphate, pH 8, 100 mM NaCl, 15 mM TCEP, 2 mM ascorbic acid, 12.5 μ M FeSO₄, 1 mM TMH and 3 mM γ -GC. The reaction was initiated with 2.5 μ M of *CthEgtB*.WT.

The HPLC traces of the control reactions of *MthEgtB* with γ -GC (Figure 4A, black) and *CthEgtB* with cysteine (Figure 4A, blue) were superimposed with a HPLC trace from the reaction of *CthEgtB* with 1 mM γ -GC quenched at 165 minutes (Figure 4A, red). This overlay (Figure 4B) confirms that the peak with shorter retention time is likely the γ -GC-sulfoxide and again shows that the higher retention time peak is the cysteine sulfoxide. This is further supported by comparison to experiment (3) in which γ -GC was incubated with both *CthEgtB* and *MthEgtB* showing both sulfoxide products (Figure 4B).

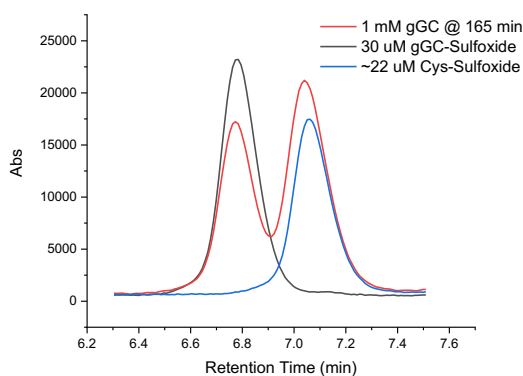


Figure 4A. Overlay of the reaction of *CthEgtB*.WT with 1 mM γ -GC at 165 minutes with a sample of 30 μ M γ -GC-sulfoxide and approximately 22 μ M of the Cys-sulfoxide.

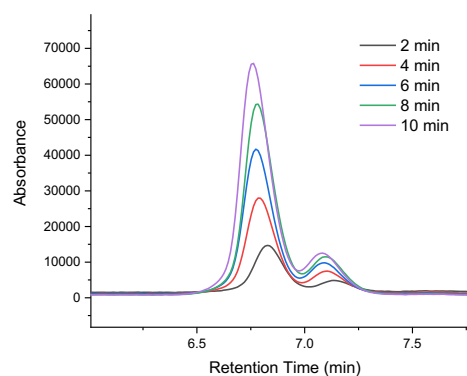


Figure 4B. *CthEgtB*. WT & *MthEgtB*.WT + γ -GC. Assay conditions included: 100 mM phosphate, pH 8, 100 mM NaCl, 2 mM TCEP, 2 mM ascorbic acid, 24 μ M FeSO₄, 1 mM TMH and 1 mM γ -GC. The reaction was initiated with 0.6 μ M of *CthEgtB* and 0.6 μ M of *MthEgtB*.WT. Aliquots were taken at 2, 4, 6, 8 and 10 minutes.

To approximate the rate of γ -GC production by *CthEgtB*, the γ -GC-sulfoxide peaks in the 1, 3 and 10 mM γ -GC reactions were quantified to give rates of 0.0007 s⁻¹, 0.002 s⁻¹ and 0.003 s⁻¹ respectively (Figure 5A) The rate for *CthEgtB* γ -GC -sulfoxide production in the presence of 10 mM γ -GC is 40 times slower than the rate of Cys-sulfoxide production by *CthEgtB* under saturating conditions. These rates were plotted against the concentrations of γ -GC used to approximate a Michaelis–Menten fit (Figure 5B). Fitting of the Michaelis–Menten equation gave an approximate K_M of 3.8 mM and a k_{cat} of 0.004 s⁻¹ (Table 1). While 10 mM γ -GC may

not saturate CthEgtB, the approximate k_{cat} with γ -GC as a substrate is 30 times slower than when cysteine is used as substrate.

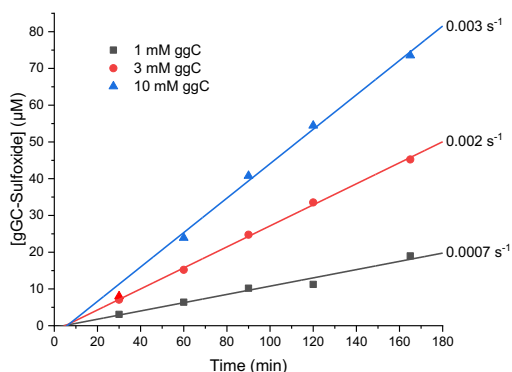


Figure 5A. Rates of γ -GC-sulfoxide production by CthEgtB. The rates of γ -GC-sulfoxide production by CthEgtB.WT in the presence of three different γ -GC concentrations (1, 3 and 10 mM). Assay conditions included: 100 mM phosphate, pH 8, 100 mM NaCl, 15 mM TCEP, 2 mM ascorbic acid, 12.5 μ M FeSO₄ and 1 mM TMH. γ -GC was varied at 1, 3 and 10 mM. The reaction was initiated with 2.5 μ M of CthEgtB and aliquots were taken at 0.5, 1, 1.5, 2 and 2.75 hours.

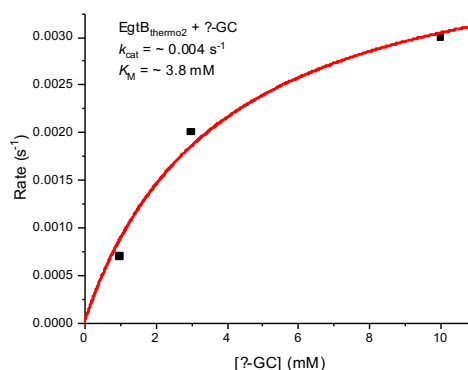


Figure 5B. Approximate Michaelis-Menten plot for CthEgtB with variable [γ -GC].

Our results show that γ -GC can be a substrate for CthEgtB, however it is a very poor substrate. Comparison of our results to those of the Liu group shows large differences (Table 1). While the K_M for γ -GC are around the same order of magnitude, the k_{cat} values differ greatly (70 fold). This leads to a large discrepancy in the ratio of the catalytic efficiencies for each substrate, 43 for the Liu group, 5000 in this study. This clearly shows that the Liu group considerably overestimates the ability of CthEgtB to consume γ -GC. In addition, it is likely that the rates measured for oxygen consumption by the Liu group do not correspond to γ -GC consumption, and instead consumption of the Cys present as an impurity of γ -GC.

This finding greatly undermines the significance of their “enzyme engineering” and consequently the whole paper.

Table 1. Kinetic parameters reported and measured for the reaction of CthEgtB with either Cys or γ -GC.

	Donor	k_{cat} (μ M min ⁻¹)	K_M (μ M)	k_{cat} / K_M (min ⁻¹)	$(k_{cat} / K_M(\text{Cys})) / (k_{cat} / K_M(\gamma\text{GC}))$
Liu ¹⁵¹	Cys	26.6 \pm 0.7	205 \pm 18	0.13 \pm 0.01	43
	γ -GC	17.5 \pm 0.4	(5.9 \pm 0.9)E ³	(3.0 \pm 0.5)E ⁻³	
This study	Cys	8.4 \pm 3	27 \pm 3	0.3	5000
	γ -GC	0.24*	3.8*	(6)E ^{-5*}	

*estimation

Chapter 2:

An Alternative Active Site Architecture for O₂ Activation in the Ergothioneine Biosynthetic EgtB from *Chloracidobacterium thermophilum*

Supplementary Information

Supplementary Figures

Table S1: Bacterial Phylogenies in which type II EgtBs have been identified. EgtBs have been identified by conservation of the N-terminal YY catalytic diad.

Acidobacteria, Actinobacteria, Bacteroidetes, Chloriflexi, Chlorobi, Cyanobacteria, Firmicutes, Gemmatimonadetes, Ignavibacteriariae, Planctomycetes, Proteobacteria, Rhodothermaeota, Spirochaetes, Verrucomicrobia

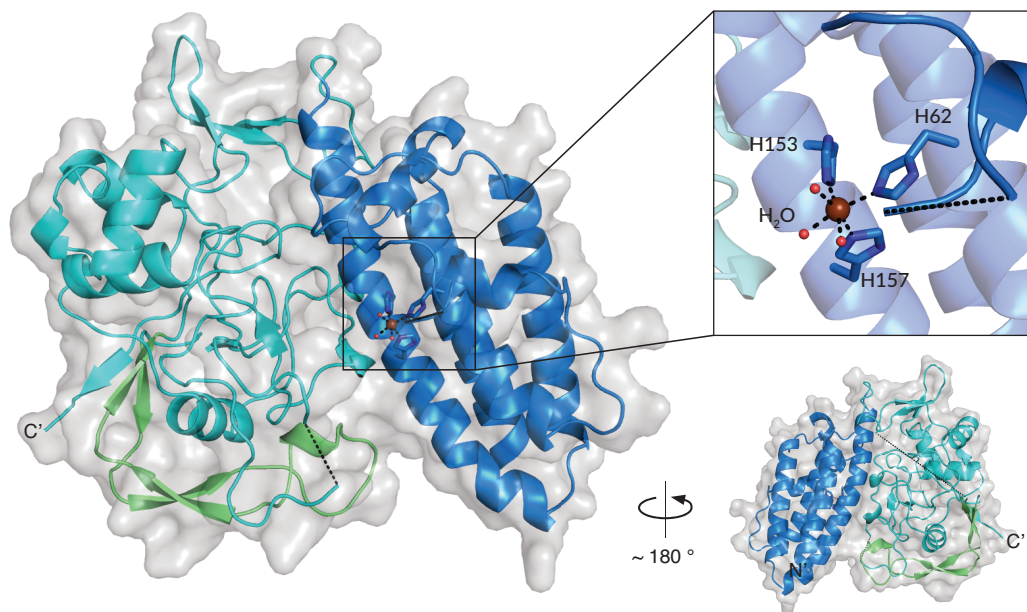


Figure S1: Cartoon and surface representation of monomer A of the *CthEgtB* native structure, PDB: 6QKI, 2.0 Å. The N-terminal domain, with a DinB_2 like fold (17-176) is shown in dark blue. The C-terminal domain, an FGE sulfatase like fold, consists of an extended beta sheet (198-234), green, and a C-type lectin fold, light blue (235-434). The N & C termini are denoted by the letters N' and C'. The catalytic iron is shown as a brown sphere, its co-ordination sphere to three histidines (blue), and three water molecules (red) are shown. Three regions lack electron density, including a loop in vicinity of the active site, these regions are represented by a dashed line.

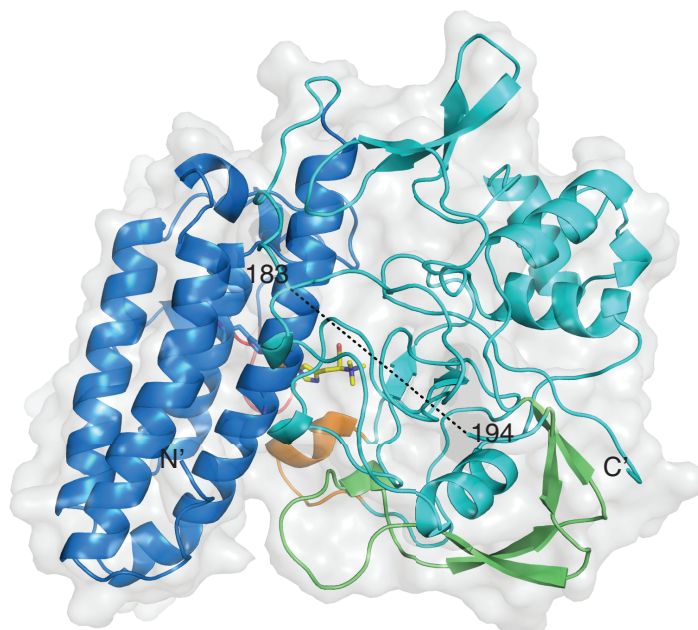


Figure S2: Cartoon and surface representation *CthEgtB* monomer A in complex with TMH. This shows the back view, the front view is shown in the main text, Figure 2A. PDB: 6QKJ, 2.2 Å. The N-terminal domain, with a DinB_2 like fold (17-176) is shown in dark blue. The C-terminal FGE-like domain is shown in green (198-234) and light blue (235-434). Active Site loop 1 (93-99) in red, and active site loop 2 (378-386) in orange, close over the active site. In the native structure these two loops are not resolved. The substrate TMH (yellow), iron (orange), chloride ligands (green) and the metal coordinating histidines (blue) are shown to indicate the location of the active site. Residues 184-193, the linker region connecting the N and C terminal domain were not modelled, due to a lack of electron density (black dotted line).

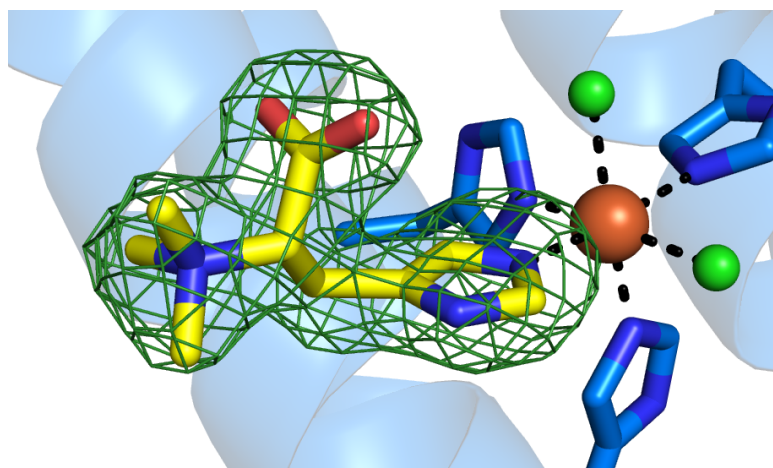


Figure S3: Omit map of TMH in the *CthEgtB*/Fe/TMH structure ($m | F_o | - D | F_c |$, electron density; $\sigma = 3.0$). TMH is bound to the iron (brown) which is also co-ordinated to 3 histidines (sticks in blue) and two chloride ions (green).

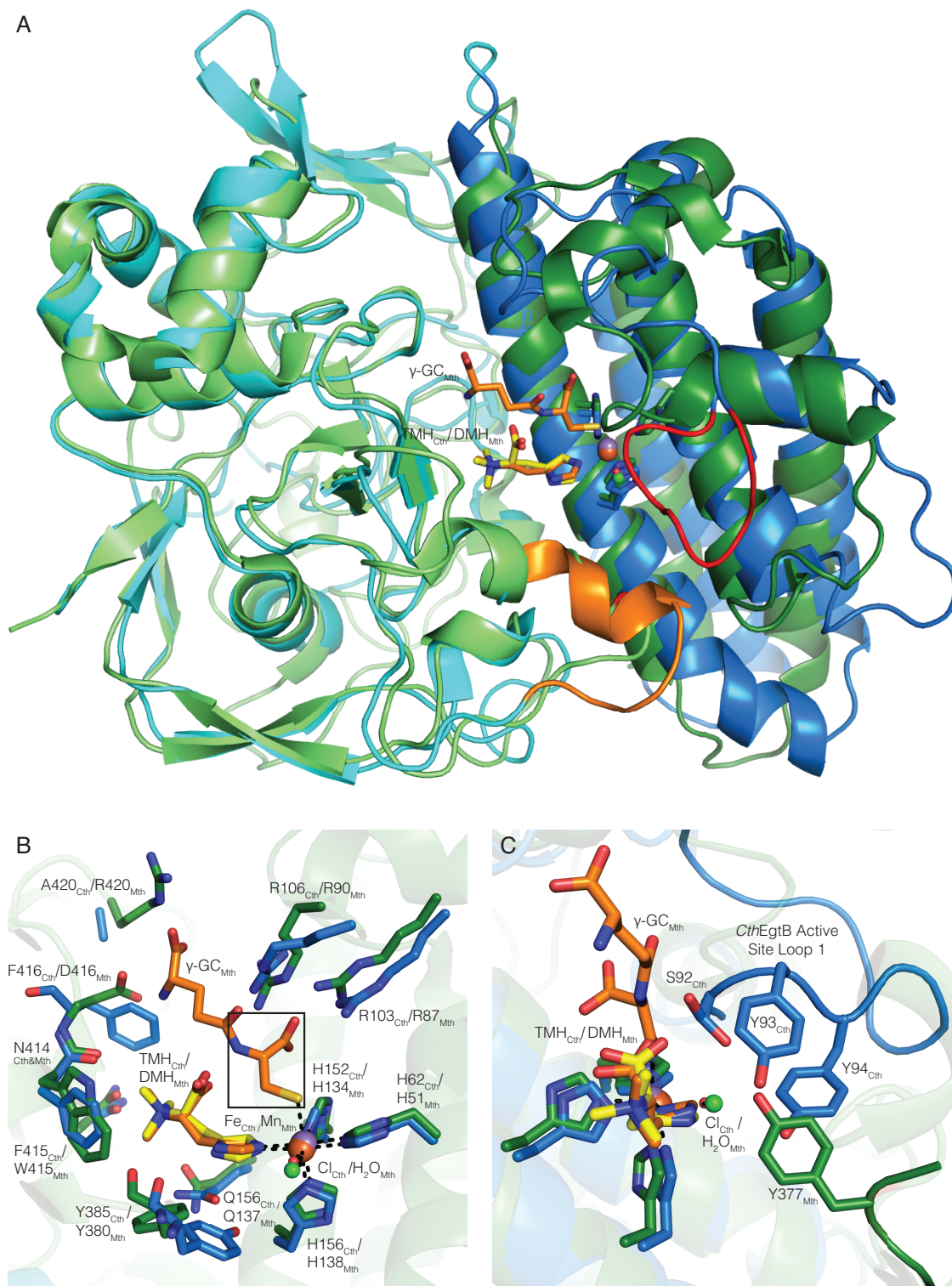


Figure S4: Structural comparison of the overall structures and active site of *CthEgtB* with *MthEgtB*. Superposition of *CthEgtB* TMH structure (blue, orange & red) and *MthEgtB* structure (green) (PDB: 4X8D). iron is shown in brown, Mn in purple, TMH in the *CthEgtB* structure in yellow, and *MthEgtB* ligands, DMH and gGC, in orange, Cl ions and water as spheres in green and red respectively. A: Superimposition shows a high structural similarity between the two structures in overall structure. B: A front on view of the active site, focusing on the metal co-ordination sphere and TMH and γ -GC/Cys binding sites. Active site loop 1 of *CthEgtB* has been omitted for clarity. Box: Cysteine is modelled into the active site of *CthEgtB* by mimicking the position of the cysteinyl moiety of γ -GC in *MthEgtB*. C: This view focuses on the oxygen

An Alternative Active Site Architecture for O₂ Activation

binding site and is rotated by approximately 90 °to the left from the front view. Active site loop 2 has been omitted for clarity in the *CthEgtB* structure. The two oxygen binding sites are dramatically different, with S92, Y93 & Y94 of *CthEgtB* siting on active site loop 1 pointing in towards the oxygen binding site (blue). Y377 of *MthEgtB* points in towards the proposed from a different location of the protein scaffold (green).

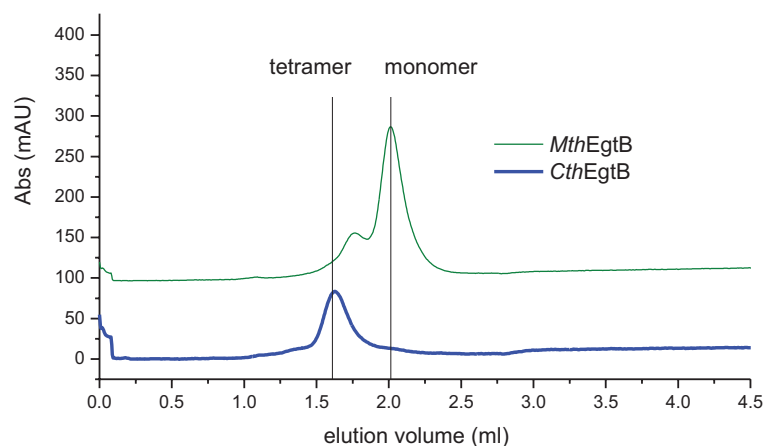


Figure S5. Analytical Gel Filtration of *CthEgtB* and *MthEgtB* under native conditions. Superdex200 size exclusion chromatography, at room temperature using 20 mM Tris-HCl pH 7.4, 150 mM NaCl, as running buffer (flow rate: 0.5 mL/min)

Experimental

Materials

All standard reagents were purchased from Aldrich/Sigma if not otherwise stated. Synthetic oligonucleotides were purchased from Microsynth, Switzerland.

Recombinant EgtB constructs

The gene for EgtB from *Chloracidobacterium thermophilum* (*CthEgtB*, WP_014099805.1) was codon-optimized for protein production in *E. coli* and purchased from Genscript. The gene was ligated as NdeI-XhoI fragments into a pET28a cloning vector. The corresponding constructs induced *E. coli* cells to produce the following proteins:

→ sequence of pET28 *CthEgtB*:

```
GSSHHHHHSSGLVPRGSHMEARSHPEPIQSGEVSDRKAWQRHYRAVRVSEAICQPLETEDYVVQPMPDVSP
KWLHGHTSWFFETFILKSLADYRPFHPRYDYIFNSYYEAVGARHPRPQRGLLTRPTVSEVYAYRAHVDAVERF
IAHSDTRTWAALQPILELGLHHEQQHQELLLLTDIKAILATNPLDPVYRPQPQLPSPVEQLSPTGDWHIVEGGRY
AIGHAGRGFAFDNEGPRHDVLLRPCRIAARPVTNGEFLAFMADGGYRRPELWLSDGWAAVTARGWEAPLYWRQAA
DGTWETLTLHGVPVAPYEPVCHISFYEADAYARWAGKRLPTEAEWEVVAARLPVTGNFYESGVLHPRPVSVA
FYGDVWWTASPYVGYPGFRPVSGALGEYNGKFMCNQMVLRRGGSCATSLTHIRSTYRNFFPPDARWQFTGVR
LAE
DMS
```

$m/z(\text{CthEgtB})$: calc.: 51131 Da, meas.: 51130 & 51307 Da (+ 177 Da, corresponds to α -N-gluconoylation of his tag¹⁵²)

$\epsilon_{280}(\text{CthEgtB})$: 115280 M⁻¹ cm⁻¹

EgtB variants

CthEgtBY93F, *CthEgtBY94F* and *CthEgtBY93F_Y94F* were constructed by primer extension using the primers shown below. The resulting fragments were cloned into pET28a vectors. For protein crystallization we cloned the *CthEgtB* gene into a modified vector, pET19m, to encode an EgtB fusion construct with an N-terminal His₆-tag followed by a TEV (tobacco etch virus) protease cleavage site.

Y93Fs: 5' - ACATCTTCAACTCTTTTTTACGAAGCGGTTGGT -3'

Y93Fa: 5' - ACCAACCGCTTCGTAAAAAGAGTTGAAGATGT -3'

Y94Fs: 5' - ACATCTTCAACTCTTATTTTTGAAGCGGTTGGT -3'

Y94Fa: 5' - ACCAACCGCTTCAAATAAGAGTTGAAGATGT -3'

CthEgtBs: 5' - ACATCTTCAACTCTTTTTTTGAAGCGGTTGGT -3'

CthEgtBa: 5' - ACCAACCGCTTCAAAAAAAGAGTTGAAGATGT -3'

$m/z(CthEgtBY93F)$: calc.: 51115 Da meas.: 51291 Da (+ 176 Da, corresponds to α -N-gluconoylation of his tag¹⁵²); $\epsilon_{280}(CthEgtBY93F)$: 113790 M⁻¹ cm⁻¹

$m/z(CthEgtBY94F)$: calc.: 51115 Da meas.: 51291 Da (+ 176 Da, corresponds to α -N-gluconoylation of his tag¹⁵²); $\epsilon_{280}(CthEgtBY94F)$: 113790 M⁻¹ cm⁻¹

$m/z(CthEgtBY93F_Y94F)$: calc.: 51099 Da, meas.: 51100 & 51275 (+ 176 Da, corresponds to α -N-gluconoylation of his tag¹⁵²); $\epsilon_{280}(CthEgtBY93F_Y94F)$: 112300 M⁻¹ cm⁻¹

$m/z(CthEgtB_{WT_nohis})$: calc.: 49294 Da meas.: 49294 Da; $\epsilon_{280}(CthEgtB_{WT_nohis})$: 115280 M⁻¹ cm⁻¹

Recombinant protein production

CaCl₂ competent BL21.pLysS cells were transformed with a pET28*CthEgtB* plasmid for kinetics or a pET19m*CthEgtB* plasmid for crystallization, following standard heat shocking procedures. An overnight preculture (Lysogeny Broth (LB) medium (5 mL), containing the appropriate antibiotics (kanamycin (50 mg/L) and chloramphenicol (34 mg/L) for pET28, and ampicillin (100 mg/L) and chloramphenicol (34 mg/L) for pET19) was inoculated with the transformed cells and incubated overnight at 37 °C whilst shaking at 180 rpm. 1 mL of pre-culture was used to inoculate growth cultures of Luria Broth containing the appropriate antibiotics at concentrations previously stated. Cells were grown at 37 °C with shaking (180 rpm) until the optical density at 600 nm (OD₆₀₀) reached 0.6. Expression of the plasmid encoded gene was induced by the addition of IPTG to a final concentration of 0.1 mM. Protein expression was allowed to continue for 3 hours at 37 °C. Cells were harvested by centrifugation at 9000 g for 25 minutes at 4 °C, the supernatant was discarded and the cell pellet was frozen at - 20 °C until required for purification. The cell pellet was thawed

on ice and resuspended in Lysis buffer (200 mM NaCl, 50 mM Tris, pH 8.0) and was lysed by sonication (2 minutes, 50% duty cycle, output of 5) three times with two-minute intervals in between. Cellular debris was then pelleted by centrifugation at 4,100 g for 40 minutes, at 4 °C. The supernatant was incubated with Ni-NTA Agarose slurry at 4°C for 20 minutes while rotating. The agarose beads were washed with washing buffer 1 (200 mM NaCl, 50 mM Tris, pH 8.0, 10 mM imidazole) and washing buffer 2 (200 mM NaCl, 50 mM Tris, pH 8.0, 20 mM imidazole). *CthEgtB* was eluted and collected in fractions by washing with elution buffer (200 mM NaCl, 50 mM Tris, pH 8.0, 250 mM imidazole). Protein concentration was determined for each collected fraction with a nano-drop 2000/2000c spectrophotometer. Fractions identified as containing *CthEgtB* by concentration and SDS PAGE Gel were pooled and dialyzed overnight into dialysis buffer twice (50 mM Tris, pH 8, 50 mM NaCl). For kinetics, the pET28 construct was then concentrated and frozen in aliquots for kinetic analysis. The pET19 construct for crystallization was dialyzed against 20 mM Tris pH 8, 200 mM NaCl and 1 mM DTT with TEV protease at final protein ratio of 10:1 for 16 h at 4°C. After dialysis the sample was ran over a Ni-NTA column to remove the TEV protease and His tag. Afterwards *CthEgtB* WT was concentrated with an Amicon centrifugal filter device (cut-off 10 kDa) and loaded onto a Superdex 200 pg 26/600 pre-equilibrated with 20 mM Tris pH 8, 200 mM NaCl. The *CthEgtB* WT peak fractions were collected and concentrated in an Amicon centrifugal filter device (cut-off 10 kDa) to a final concentration of 6.5 mg/mL and immediately used for crystallization trials.

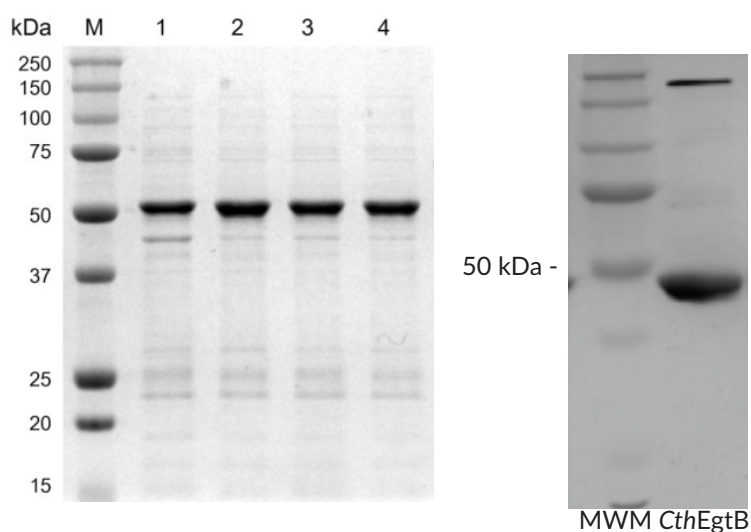


Figure S6: Left: SDS page gel with *CthEgtB* variants used for kinetic analysis: M – Molecular weight marker; 1 – *CthEgtB*WT; 2 – *CthEgtBY93F*; 3 – *CthEgtBY94F*; 4 – *CthEgtBY93F_Y94F*. Right: SDS Page gel of *CthEgtB* used for crystallization, 10 µg are loaded. A thick band is observed just under the 50 kDa marker.

Enzymatic Assay

Sulfoxide synthase activities of *CthEgtB* variants were measured in reactions containing 100 mM phosphate buffer, pH 8, 100 mM NaCl, 2 mM TCEP, 2 mM ascorbate, FeSO₄ (4 equiv. to protein concentration), TMH and cysteine. Reactions were initiated by addition of enzyme and were incubated at 26°C. Aliquots of the reactions were quenched by addition of phosphoric acid. Reaction products were quantified by cation exchange HPLC using 20 mM phosphoric acid at pH 2 with a NaCl gradient as a mobile phase.¹⁵³ Chromatograms were recorded at 265 nm. Cysteine dioxygenase activity was quantified by monitoring the consumption of cysteine. To quantify cysteine, 40 μL reaction aliquots were quenched by the addition of 40 μL acetonitrile and 10 μL of 20 mM 4-bromomethyl-7-methoxycoumarin (BMC) in DMSO. After incubation for 30 min the mixture was diluted with one volume equivalent of aqueous 0.1% TFA solution. Coumarin-adducts (Figure S7) were identified by comparison to authentic samples using RP-HPLC. Chromatograms were recorded at 330 nm. Averages of at least three independently determined rates of sulfoxide production or cysteine consumption were fitted to the function $v = V_{\max}[s]/(K_M + [s])$ (Figures 8a – 8g). The Michaelis-Menten parameters k_{cat} and k_{cat}/K_M were determined in the presence of co-substrate at a concentration at least 3-fold higher than the corresponding K_M and in air saturated buffers.

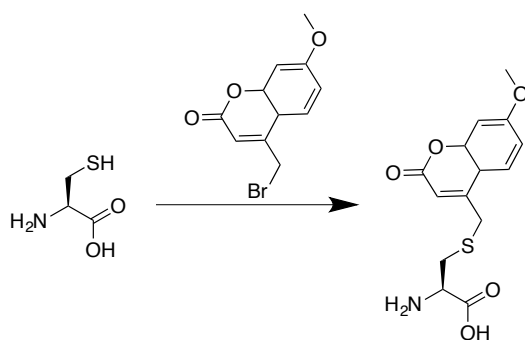
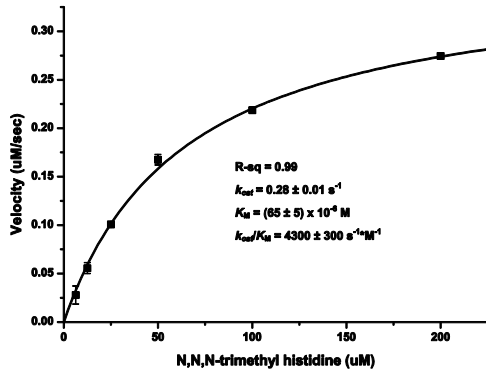
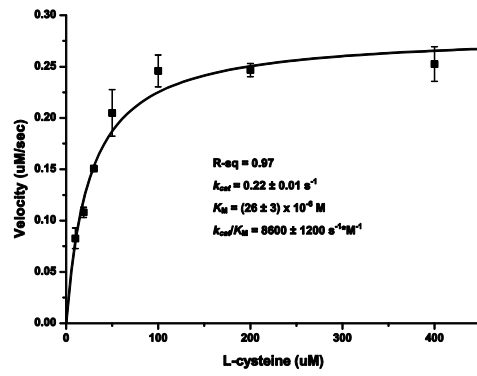


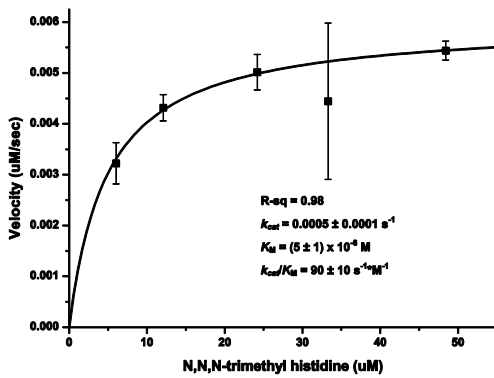
Figure S7: Formation of the methyl-7-methoxycoumarin adduct of Cys.



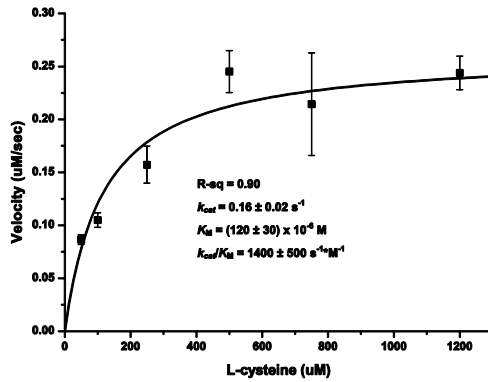
S8a: Sulfoxide Formation - *CthEgtB_{wT}*, [CYS] = 500 μM



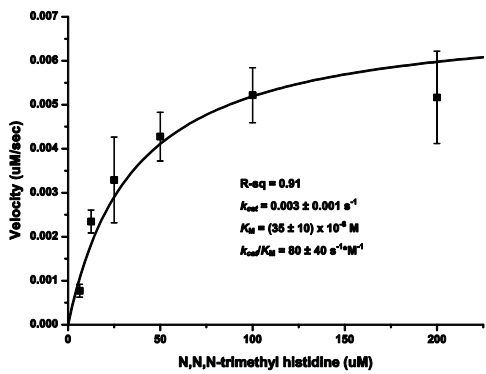
S8b: Sulfoxide Formation - *CthEgtB_{wT}*, [TMH] = 200 μM



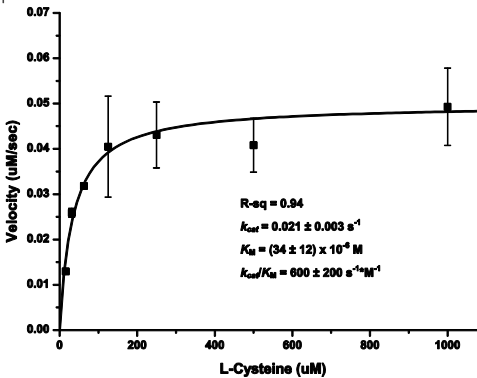
S8c: Sulfoxide Formation - *CthEgtB_{y93F}*, [CYS] = 2000 μM



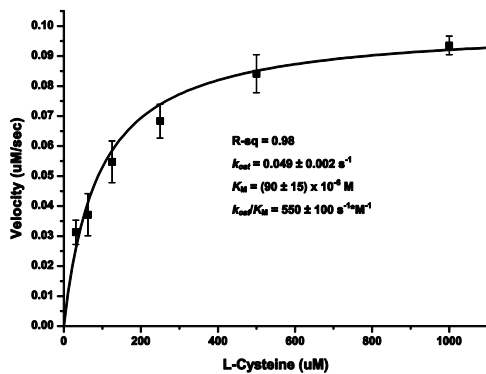
S8d: Cysteine Consumption - *CthEgtB_{wT}*, [TMH] = 500 μM



S8e: Sulfoxide Formation - *CthEgtB_{y94F}*, [CYS] = 500 μM



S8f: Cysteine Consumption - *CthEgtB_{y94F}*, [TMH] = 500 μM



S8g: Cysteine Consumption - *CthEgtB_{y93F_y94F}*, [TMH] = 500 μM

Characterization of reaction products by ¹H NMR

CthEgtB containing reaction mixtures were analyzed by ¹H NMR to identify the formed products. The reactions contained 100 mM phosphate buffer - pH 8.0, 100 mM NaCl, 2 mM ascorbate, 2 mM TCEP, 4 μM FeSO₄, 1 mM TMH, 0.5 mM *L*-cysteine and 1 μM of *CthEgtB* variant to a final volume of 2 ml. These solutions were incubated overnight at room temperature. After lyophilization the residue was dissolved in D₂O. ¹H NMR (500 MHz, D₂O) was measured with 128 scans and analyzed by MestReNova software. Substrates and products were identified based on α- and β-protons of the cysteine moiety and the aromatic protons of TMH.

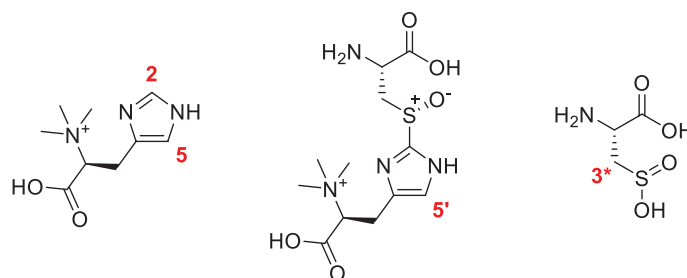


Figure S9. Sulfoxide synthase substrates and products. C-H functions used for identification by ¹H NMR (see Figures S10 & S11) are labeled in red.

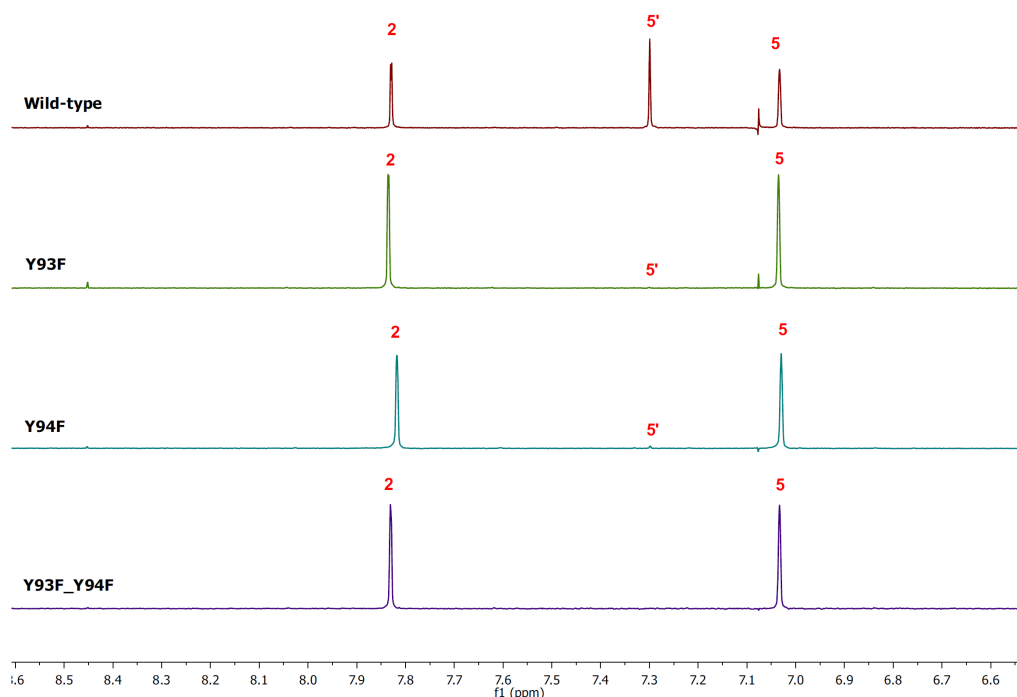


Figure S10. Aromatic region of the *CthEgtB* variants catalyzed reactions in excess of TMH.

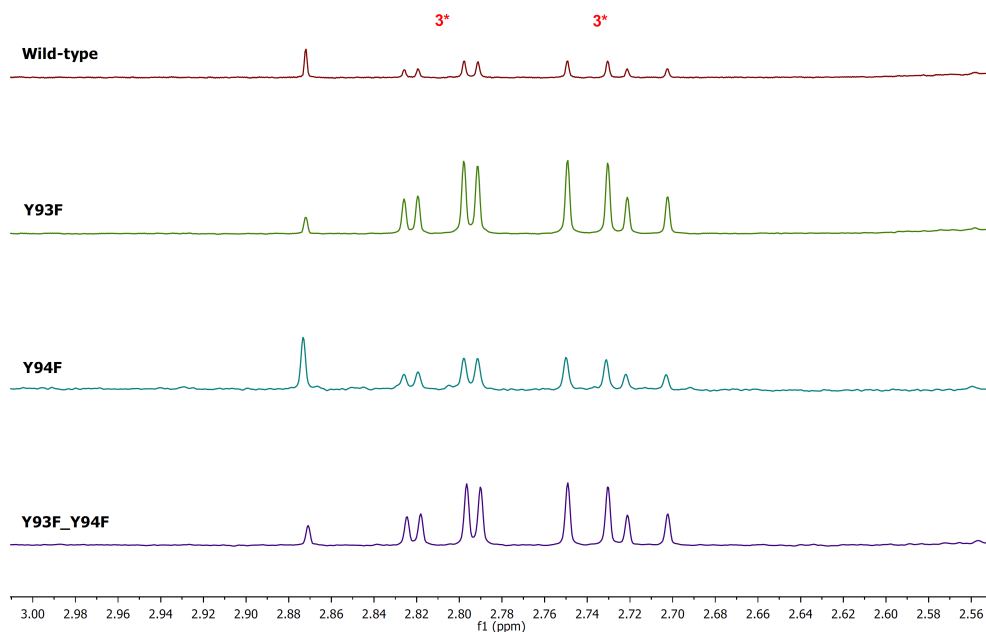


Figure S11. β -protons of the sulfinic acid in *CthEgtB* variants catalyzed reactions in excess of TMH.

Solvent KIE

Solvent KIE were determined in reactions containing 100 mM HEPES, pH 8.0 or 7.6 (final pD = 8.0), 100 mM NaCl, 2 mM TCEP, 4 equiv. (to enzyme) FeSO_4 , and 2 mM ascorbate.⁵ Premixtures were lyophilized and then dissolved in H_2O or D_2O . For *CthEgtB*_{wt} the solvent KIE was determined by full Michaelis-Menten analysis. The concentration of TMH was 500 μM , the cysteine concentration was varied between 12-400 μM . The reactions were initiated by addition of 1.3 μM *CthEgtB*_{wt} and incubated at 26°C. The rate of sulfoxide production was determined as described above. The solvent KIE of *CthEgtB*_{Y93F} and *CthEgtB*_{Y94F} was determined in reactions containing 1 mM TMH and 1.2 mM cysteine. Reactions were initiated by addition of 14 μM *CthEgtB*_{Y93F} or 10 μM *CthEgtB*_{Y94F} and incubated at 26°C. The rate of sulfoxide production was determined as described above.

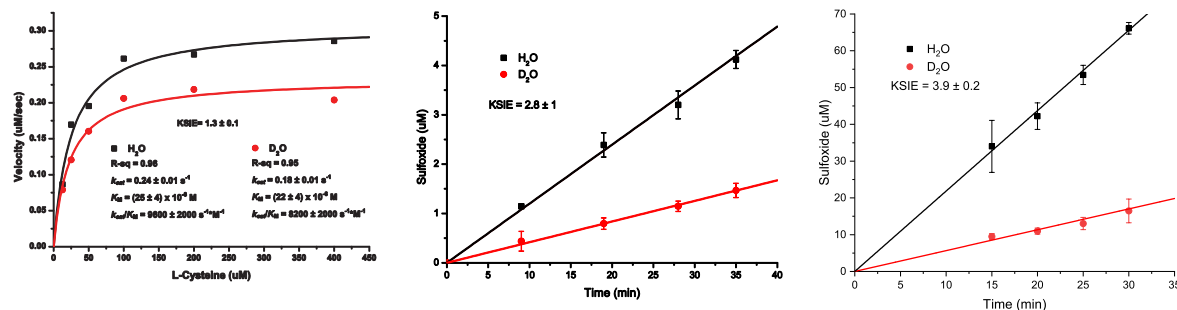


Figure S12: Solvent kinetic isotope effects (KSIE) of *CthEgtB*.WT (left), *CthEgtB*Y93F (middle), and *CthEgtB*Y94 (right)

Sequence Alignments & Phylogenetic Trees

All sequences alignments were done using MUSCLE.¹⁵⁴ Phylogenetic trees were built using phylogeny.fr¹⁵⁵ and redrawn in indesign. Sequences of other characterised EgtBs and that of *T. Elongatus*. Grey text indicates the methyl transfer domains of the fungal EgtBs.

→ sequence of EgtB from *Mycolicibacterium thermoresistibile* (G7CFI3)

MTGVAVPHRAELARQLIDARNRTLRLVDFDDAELRRQYDPLMSPLVWDLAHIQQEELWLLRGGDPRRPGLLEPA
VEQLYDAFVHPRASRVHLLPLSPAQARRFCATVRSVAVLDALDRLPEDADTFAFGMVVSHEHQHDETMLQALNLR
GEPPLGSGTALPPGRPGVAGTSVLVPGGPFVLGVDLADEPYALDNERPAHVVDVPAFRIGRVPVTNAEWRAFIDD
GGYRQRRWWS DAGWAYRCEAGLTAPQFWNPDGTRTRFRGHVEDI PPDEPVQHVTYFEAEAYA AWAGARLPTEIEWE
KACAWDPATGRRRRYPWGDAAPTAALANLGGDALRPAPVGAYPAGASACGAEQMLGDVWEWTS SPLRPWPFGFTPM
IYQRYSPFFEGAGSGDYRVL RGGSWAVAADILRPSFRNWDHPIRRQIFAGVRLAWDVDRQTARPGVGGC

→ sequence of EgtB from *Thermosynechococcus elongatus* (WP_011057240)

MNQKFTPSETSTSWKPF DREALWQAFQQQRQFTLQLVAPLSEVILCAQPHPLYSPVGWHLGHIGYTEAFWLLPED
AAFREDRYWYAADGRPKVERQHLP RRHQLLEYLAEIRQRTGDRHLHGLSDRQWQQEARLWWI LQHEAQHSETMQ
MVLAMQGI FTTLPPSLSL PQDHQRI PAGRYAIGSEDL LALDNEQPSQCVDTLPFRIDAAPVTWREFLAFVEAGGY
QRREWWSSSGWEWREAE EITSPFYPI PENLDLPMWGLSFYEAEAYGHSQ GKRLP SEREWEIAAQQGLLN RGYVWE
WTQSPFAPYPGFQSY PYRGYSAPYFDGEHFV LKGGSHWTRP L LKRPSFRNWYSRSTREVFAGARYVHQENLISQ

→ sequence of EgtB from *Microcystis aeruginosa* (WP_008197519)

MEKIQAQAPISLNNCSRENILDYFDNAWQLEELLKLSIIKEETFYCNPD DLRNPLIFYLGHAAAFYLNKLQLVNL
LKKSPNDPYELLFGVGVDPATPEELNSAIAQIQWPGVAKVWEYRQVYEVVEI IKNIPLNLP IHPQHS LWALMM
GIEHQRVHFETSSMLLRQLPLDCLQRPPGWHYAPAFGQAYPNQMVEVSGGMVEIGKPNYPIYGDWNEYGYRQVQ
VNNFLVSKYMITNGEFKEFVHDGGYENPCYWDEEAWQWKNHYQVKYKFWLVDEGGNYQYRAMFDVFDLPLDWPV
EVNYYEAIAYCRFQGGIRLMTAEWNLVSYGSQKNRCYTLNDFDDYNLNLKFCSP TPVGM LKNAGNNSEIYD
LRGNVWEWLEDDFNPLTDFQPHYLYADNSTPFFNSQH KMMLGGAWVTNGTEILPYYRNWFRNFYQHAGFRIAQS
L

→ sequence of EgtB from *Neurospora crassa* (Q7RX33.3)

MPSAESMTPSSALGQLKATGQHVLSK LQQQTSNADIIDIRRAVEINLKTEITSMFRPKDGPRQLPTLLLLYNERG
LQLFERITYLEEYYLTNDEIKILTKHATEMASFI PSGAMI IELGSGNLRKVNLLLEALDNAGKAIDYYALDLSRE
ELERTLAQVPSYKHVKCHGLLGTYYDDGRDWLKAPENINKQKCIHLHGSSIGNFNRS DAATFLKGF TDV LGPNDKM
LIGVDACNDPARVYHAYNDKVGITHEFILNGLRNANEIIGETA FIEGDWRVIGEYVYDEEGGRHQAFYAPTRDTM
VMGELIRSHDRIQIEQSLKYSKEESERLWSTAGLEQVSEWYTGNEYGLHLLAKSRMSFSLIPSVYARSALPTLDD
WEALWATWDVVT RQMLPQEELLEKPIKLRNACIFYLGHIP TFLDIQLTKTKQAPSEPAHFCKIFERGI DPVDN
PELCHAHSEIPDEWPPVEI LTYQETVRSRLRGLYAHGIANIPRNVGRAI WVGFEHELMHIETLLYMM LQSDKTL
IPTHIPRPDFDKLARKAESERVPNQWFKI PAQEITIGLDDPEDGSDINKHYGWDNEKPPRRVQVA AFQAQGRPIT
NEEYAQYLLEKNI DKL PASWARLDNENI SNGTTNSVSGHHSNR TSKQLPSSFLEKTAVRTVYGLVPLKHALDWP
VFASYDELAGCAAYMGRIPTFEETR SIYAYADALKKKKEAERQLGRTVPAVNAHL TNNGVEITPPSSPSSETPA
ESSSPSDSNTTLITTEDLFSDLGDANVGFHNWHPMPITSKGNTLVGQGELGGVWEWTS SVLRKWE GFEPMELYPG
YTADFFDEKHNI VLGGSWATHPRIAGRKS FVNWYQRNYPYAWVGARVVRDL

→ sequence of EgtB from *Schizosaccharomyces pombe* (O94632.4)

MTEIENIGALEVLFSPESIEQSLKRCQLPSTLLYDEKGLRLFDEITNLKEYYLYESEL DILKKFSDSIANQLLSP
DLPNTVIELGCGNMRKTKLLLD AFEKKGCDVHFYALDLNEAELQKGLQELRQT TNYQHVKVSGICGCFERLLQCL
DRFRSEPNRSRISMLYL GASIGNFDRKSAASFLRSFASRLNIHDNLLISFDHRNKAELVQLAYDDPYRITEKFEKN
ILASVNAVFGENLFDENDWEYKSVYDEDLGVHRAYLQAKNEVTVIKGP MFFQFKPSHLILIEESWKNSDQECRQI
IEKGFDFKLVSKYESTIADYSTYVITKQFPAMLQPLQPCPSLA EWDALRKVWLFITNKLLNKDNMYTAWIPLRHP
PIFYIGHVPVFNDIY LTKIVKNKATANKKHFWEWFQRGIDPDIEDPSKCHWHSEVPESWSPDQLREYEKESWEY
HIVKLCAMDELSTSEKRILWLCYEHVAMHVETTLYIYVQSFQANQTVSICGSLPEPAEKLTKAPLWVNPETE
IAVGMLT TQYTSVGSNLQSSDLSAHENTDEL F YFAWDNEKPMRKKLVSSFSIANRPI SNGEYLD FINKKSKTER

VYPKQWAEIDGTLTYIRTMYGLLPLDDYLGWPVMTSYDDLNNYASSQGCRLPTEDELNCFYDRVLERTDEPYVSTE
GKATGFQQLHPLALSDNSSNQIFTGAWEWTSVLEKHEDFEPEELYPDYTRDFFDGKHNVLGGSFATATRISNR
RSFRNFYQAGYKYAWIGARLVKN

→ sequence of EgtB from *Aspergillus fumigatus* (OXN26701)

MSPLPCPSKKVEIVDIHRNDVKFSLVNEIRKGLNPPEGTPRSLPTMLLYDAQGLKLFEEITYVDEYYLTNAEIEV
LQNHSSKIVERVPENAQLLELGSGLNRKIKILLQEFERTGKHVDYYALDLSLSELQRTFVEVSSDEYSHVDLHGL
HGTYYDALAWLSNPQNLQRPTVVMMSGSSIGNFSREGAAEFQAQFARLLKPSDLMIIGLDACTDPDKVYKAYNDS
KGITQRFYENGLLHANAVLGYEAFQLSEWEVVTDYDVAGGRHRAFYSQKQNVITDGVLLQKGEKLVFEEATKYSP
QQREQLWRDANLVLCEELGNSSEEYHIHLLSPPTLSLPSQPSEYAANPVPSFKFQSLWTAWDTVTKAMVPREEL
LAKPIKLRNALIFYFGHIPFTDIHLTRALGGSPTFPRNYRQIFERGI DPVDNPEHCHSHSEIPDEWPPLAEIL
DYQDRVRSRVDSVLQRDDITQNRCLGEALWIGFEHEAMHLETFLYMLLQSDKTLPPPLADRPDFEKLHFHQRANA
KPNEWFAIPEQTLSIGFDDTDEQSLPDVSFGWDNEKPQRTITVRAFEAQAHAITNGEYAKYLQATRQRRRPESWV
LTHSDENYPI SKGVTLESSQATKDFMDNFAVRTVFGPVPLEFAQDWPVMASYDELALYA EWVGCRLPTYEEVKSI
YNNSAQLKETRQHEPSDHESNGVKGINRDMVTNGHSHVHVDKPRTPERQPIQPPSQSTMPVFVDLHGCNVGFKHW
HPTPVIQNGDRLAGHGELGGVWEWTSTPLTPHDGFKAMDIYPGYTADFFDGKHNIVLGGSWATHPRIAGRITTFVN
WYQHNYPYTWAGARLVRSQ

Native Crystallization

Initial crystallization conditions of *CthEgtB* were determined with the vapor diffusion method in a sitting drop 96-well format. Drops were set up using a dispensing robot (Crystal Gryphon, Art Robbins) mixing 0.2 μL of *EgtB* solution (6.5 mg mL⁻¹) with 0.2 μL of reservoir solution equilibrated against 30 μL reservoir solution. The screens were stored at 30 °C. Several initial hits were identified and optimized in 24-well plates using sitting drop format. Crystallization conditions that led to the native structure contained 13% PEG 3350, 0.25 M MgAc, pH 6.5, mixed in a 1 μL : 1 μL ratio *CthEgtB* at 6.5 mg mL⁻¹, equilibrated against 50 μL . Plates were incubated at 30°C with crystals appearing within a few days.

TMH crystallization

The TMH containing structure was obtained through co-crystallisation. Crystallization conditions of *CthEgtB* with TMH were determined with the vapor diffusion method in a sitting drop 96-well format using commercially available screens, PEG-ION HT and Morpheus HT. *CthEgtB* was mixed and incubated with TMH (650 μM) for 30 minutes preceding crystallization. Drops were set up using a dispensing robot (Crystal Gryphon, Art Robbins) mixing different ratios of *EgtB*-TMH solution (6.5mg mL⁻¹) with reservoir solution (0.2 μL :0.1 μL , 0.1 μL :0.2 μL and 0.2 μL :0.2 μL) and were equilibrated against 30 μL reservoir solution. The screens were stored at 30 °C and crystals appeared within several days. Several initial hits were identified, flash cooled in liquid nitrogen and taken a synchrotron for data collection. A crystal from C2 of Morpheus (.09 M NPS (0.3M Sodium nitrate, 0.3 Sodium phosphate dibasic, 0.3M Ammonium sulfate), 0.1 Buffer system 1 (0.1 M Buffer system 1 Imidazole; MES monohydrate (acid)), pH 6.5, 50 % precipitant mix 2 (40% v/v Ethylene glycol; 20 % w/v PEG 8000) mixed in a 2:1 ratio of protein to reservoir solution diffracted well.

Data collection, data processing, structure solution and refinement

Data of native crystals were collected at the Xo6SA (PXI) beamline using a Pilatus 6M detector, while data for the TMH co-crystalised crystal was collected at beamline Xo6DA (PXIII) using a Pilatus 2M-F detector, both at the Swiss Light Source (SLS), Villigen, Switzerland. The collected diffraction data were indexed and integrated using MOSFLM¹⁵⁶ or XDS¹⁵⁷ and were scaled using aimless¹⁵⁸. The native structure was solved by molecular replacement using a homology model built from the published native structure of *MthEgtB* (PDB: 4X8E) as a search model using PHENIX-Phaser.¹⁵⁹ The TMH structure was solved by molecular replacement using the native structure. The initial model of native *MthEgtB* was built using AUTOBUILD of the PHENIX package.¹⁶⁰ Several rounds of iterative model building and refinement were performed using Coot¹⁶¹ and Refmac¹⁶² or PHENIX¹⁶³. 5% of the data were excluded from refinement and used for cross-validation. Data collections and refinement statistics are summarized in Table 2 and Table 3, respectively. Stereochemical validation of the final models were performed using MolProbity.¹⁶⁴ Interfaces of proteins were analyzed by PISA¹⁶⁵ and Figures were prepared PyMOL Molecular Graphics System, Version 1.7 (Schrödinger).

Table S2: Data Collection Statistics

	<i>CthEgtB</i> Native Structure PDB: 6QKI	<i>CthEgtB</i> TMH Structure PDB: 6QKJ
X-ray Source	SLS X06SA (PXI)	SLS X06DA (PXIII)
X-Ray detector	Pilatus 6M	Pilatus 2M
Wavelength (Å)	1.0332	1.0332
Space group	P 1 21 1	C 2 2 21
Cell dimensions a, b, c (Å)	85.0 127.5 89.5	108.1 201.3 108.3
Cell Angles α , β , γ (°)	90 113 90	90 90 90
Solvent content (%)	46	59
Molecules in asymmetric unit	4	2
Resolution limits (Å)	29.98 - 2.04 (2.12 - 2.04)	29.68 - 2.2 (2.29 - 2.20)
R _{merge} [†]	0.128 (0.574)	0.268 (2.006)
R _{meas} [‡]	0.069 (0.408)	0.279 (2.099)
CC ½	0.993 (0.521)	0.995 (0.569)
<I/ σ (I)>	7.7 (2.3)	11.2 (1.5)
Total reflections	714035 (90207)	783779 (68532)
Unique reflections	109407 (15540)	60188 (5960)
Completeness	98.8 (96.4)	99.92 (99.87)
Multiplicity	6.5 (5.8)	13.0 (11.5)
Mosaicity	0.63	0.77

[†] $R_{\text{merge}} = \frac{\sum hkl \sum i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum hkl \sum i I_i(hkl)}$, where $I_i(hkl)$ is the observed intensity for a reflection and $\langle I(hkl) \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.

[‡] $R_{\text{meas}} = \frac{\sum hkl [N/(N-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum hkl \sum i I_i(hkl)}$, where $I_i(hkl)$ is the observed intensity for a reflection, $\langle I(hkl) \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections and N is the number of observations of intensity $I(hkl)$.

Table S3: Refinement Statistics

	<i>CthEgtB</i> Native Structure PDB: 6QKI	<i>CthEgtB</i> TMH Structure PDB: 6QKJ
Resolution limits (Å)	29.98 - 2.044 (2.117 - 2.044)	29.68 - 2.2 (2.279 - 2.2)
Rwork * (%)	20.6 (35.5)	19.4 (31.5)
Rfree ** (%)	24.9 (35.9)	23.8 (32.6)
Number of non-H atoms	12935	6992
macromolecules	12710	6671
ligands	4	46
solvent	221	275
Protein residues	1572	824
Clashscore ***	2.34	1.76
R.m.s.d from ideal		
Bond lengths (Å°)	0.019	0.015
Bond angles (u)	2.17	1.86
Ramachandran favored *** (%)	96.36	97.18
Ramachandran outliers *** (%)	0.71	0.49
Average B values (Å ²)	37.27	35.74
macromolecules	37.42	35.95
ligands	26.92	32.96
solvent	29.14	30.93

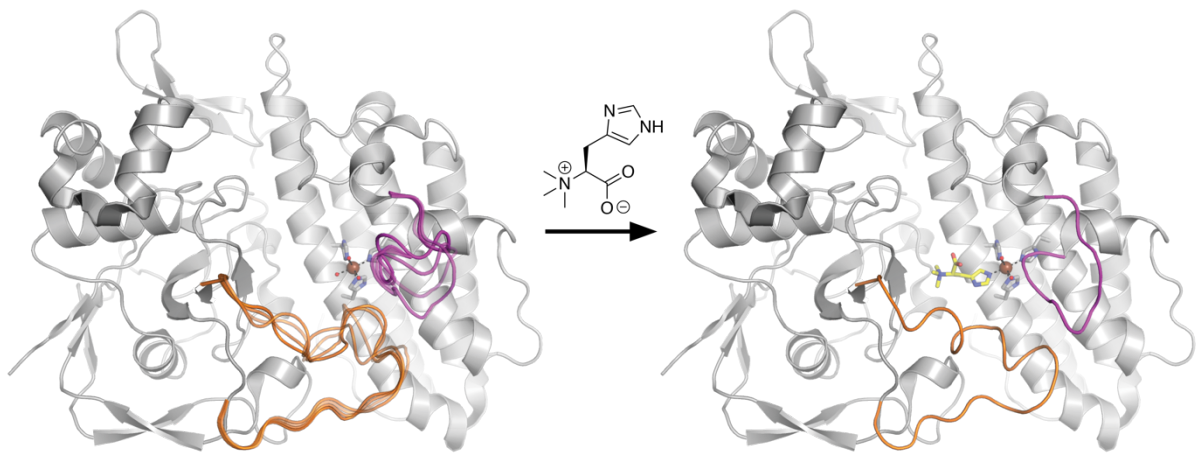
Numbers in parentheses refer to the outer shell.

* $R_{work} = \frac{\sum |hkl| |F_{obs} - F_{calc}|}{\sum |hkl| F_{obs}}$

** Rfree is the R value calculated for 5% of the data set that was not included in the refinement.

*** Molprobit.

3. The Role of Oligomerisation and Loop Folding in a type II EgtB



Homo-oligomerization is well established to confer global stability or to install functional complexity in proteins. We provide an example of a type II EgtB in which tetramerization not only provides global stability but also stabilises a specific conformation of two flexible active site loops. In contrast, other EgtB types are monomeric and adopt a rigid scaffold lacking any flexible loops. Through mutational analysis we dissect the roles of important loop residues, which reveals that tetramerization was an essential strategy to stabilize active enzyme form and to optimize enzymatic performance. This may have been a means of enabling rapid evolution and diversification, and may provide valuable lessons for enzyme engineering.

Introduction

Protein oligomerization is widespread in nature.¹⁶⁶⁻¹⁶⁹ Over 60% of structures in the protein data bank (PDB) are shown to self-associate to dimers or higher order oligomers. Homo-oligomerization is generally accepted as a fundamental strategy to improve stability and introduce functional complexity.¹⁶⁶⁻¹⁶⁷ In simple thermodynamic terms, stabilization by oligomerization can be understood as an increase in enthalpic stabilization that progressively outcompetes the effect of the entropic gains associated with unfolding.¹⁶⁹ Kinetic factors may also contribute to stabilization, in that reduction of the surface/volume ratio reduces the number of labile sites at which a given structure can start to unzip.^{167, 170-171} Thus, oligomerization appears as a fool-proof strategy to make robust proteins.

The classical example of protein function that emerges from oligomerization has been provided by the description of the co-operative oxygen (O₂) binding behavior of tetrameric hemoglobin. Binding of O₂ to a hemoglobin monomer was shown to induce subtle conformational changes that in turn increase the affinity for O₂ in the three monomers. The resulting non-linear O₂ binding behavior is crucial for efficient delivery of O₂ from the lungs to peripheral tissues.¹⁷²

In our research on the biosynthesis of ergothioneine, we have encountered a type of sulfoxide synthase that forms tetramers. The data discussed in this report suggests that oligomerization of these proteins does not necessarily serve global stabilization and does not coordinate the activity of the symmetry-related active sites. Rather, we argue that oligomerization serves specific stabilization of active site loops in their active conformation. The comparison to monomeric homologs with similar function suggests that oligomerization may have facilitated rapid – or, indeed sloppy – restructuring of the active site architecture. Oligomerization as a means to active site loop stabilization may provide novel incentive for enzyme engineers to also consider oligomeric structures for enzyme development.

The protein of interest in the present study is EgtB from *Chloracidobacterium thermophilum* (*CthEgtB*), a type II sulfoxide synthase. This iron-dependent enzyme catalyzes oxidative coupling of *N,N,N*- α -trimethylhistidine (**2**, TMH, Figure 1) with cysteine to produce sulfoxide **3**, as the central step in the biosynthesis of ergothioneine (**1**, EGT) (Fig. 1).^{31, 49} The crystal structure of *CthEgtB*, compounded by size-exclusion chromatography data suggest that this protein forms a stable homo-tetramer. Conservation of specific residues that are involved in inter-chain interactions, suggest that most type II sulfoxide synthase adopt the same quaternary structure.¹⁷³ By contrast, other types of bacterial sulfoxide synthases (type I EgtB from *Mycobacterium thermoresistibile*, *MthEgtB*; type IV EgtB from *M. aeruginosa*) (Goncharenko2015, Liao2017) are predominantly monomeric, suggesting that oligomerization is not essential for sulfoxide synthase activity.
49, 173

Type II sulfoxide synthases also differ from type I sulfoxide synthases because of a highly flexible active site structure. The binding site of the metal and the three substrates is covered by two active site loops. The flexibility of these loops was first evidenced by the lack of electron density in most analyzed crystals,

presumably due to structural disorder in this region. In the closed conformation one of the loops delivers two vital catalytic residues (Tyr93 and Tyr94) to the active site. In contrast, crystal structures of *MthEgtB* showed that type I sulfoxide synthases contain a rigid active site that remains invariant upon substrate binding.

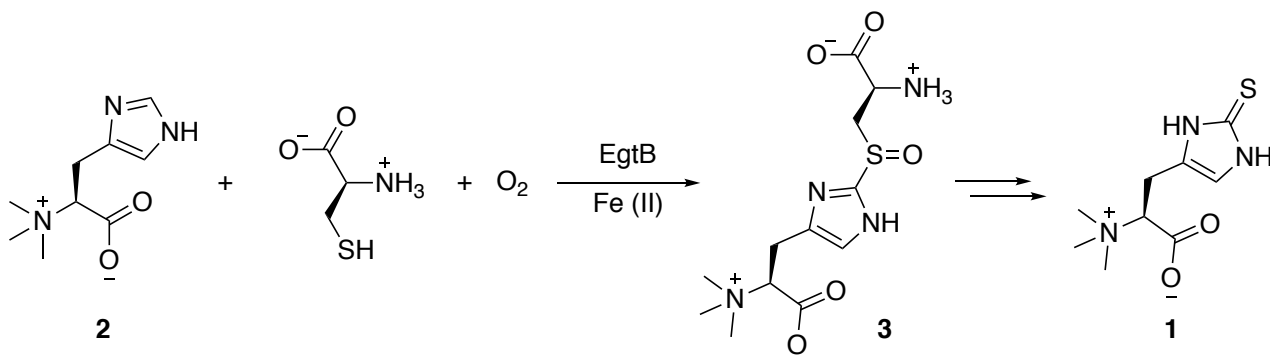


Figure 1. Reaction catalyzed by *CthEgtB*.

These observations raise the question as to how loop folding is integrated into the sequence of steps that lead from the native enzyme to the reactive complex in which all substrates bound and all catalytic residues in place. Furthermore, we discuss the question of how the quaternary structure modulates the conformational preferences of these loops.

Results

Alternative loop conformations in CthEgtB Crystal structures. In the published CthEgtB structures the active site loops 1 (residues 92 – 98) and 2 (374 – 392) either adopt an “active” conformation (structure 2, Fig. 2 B&C, PDB: 6QKJ, 606M, 605L) or remain unresolved due to structural disorder (structure 1, Fig 2. C, PDB: 6QKI).^{113, 173}

In the active conformation (**structure 2**), loop 1 adopts an extended conformation and points the side chain of residues Ser92, Tyr93 and Tyr94 towards the metal binding site. As shown in previous work, Tyr93 and 94 are important for catalytic activity. Therefore we termed this loop conformation as “active”. The extended conformation is likely stabilized by three hydrogen bonds between the backbone of loop 1 and the side chains of Glu384 and K388 on loop 2 (structure 2, Fig. 2B&D) These contacts suggest that the active conformation of loop 1 is dependent on the proper positioning of loop 2. Most residues of loop 2 (residues 380 - 309) are rolled up into three turns of a π -helix (3_{10} -helix). The side chain of Asn386 and the backbone carbonyl of Tyr385 hydrogen bond to a crystallographic water (H₂O_690, 6qKJ) which in turn hydrogen bonds with the imidazole ring of the substrate TMH. The aromatic side chain of Tyr385 packs against the bottom face of the imidazole ring of TMH. Furthermore, the side chain of Asn392 and the backbone of Lys388 form hydrogen bonds to a loop segment from a neighboring monomer (residues 255 – 261). These numerous polar contacts may be stabilizing the helical structure of loop 2.

Among the many crystals of CthEgtB grown under slightly different conditions, we identified two crystals in which the active site loops of CthEgtB adopt unique alternative conformations. Data collection and refinement statistics for these new data sets are summarized in Supplementary Tables 1 and 2.

Structure 3 was obtained by soaking of native crystals with TMH. The crystal belongs to space group P2₁ with four monomers in the asymmetric unit. This CthEgtB/Fe/TMH complex structure was solved to 2.8 Å by molecular replacement with the native CthEgtB structure (PDB: 6QKI) as a search model. The electron density revealed a continuous polypeptide chain from residues 17 to 434 with the exception of segments 93–98 (loop 1), 182–193, and 379–381 (part of loop 2). The visible part of loop 2 shows that the π -helix has given way to an extended conformation that pushes residues 381-384 closer into the active site (Fig. 2E) (Supplementary Figure 2). Glu384 is possibly directly coordinated to the metal center. Superposition of structures 2 and 3 visualizes that loop 1 in structure 2 occupies the same space as loop 2 in structure 3. Hence, the shapes of the two loops seem interdependent because they limit each others conformational freedom (Supplementary Figure 3). The TMH ligand could be unambiguously modeled into the (Fo – Fc) difference density map as a ligand of the metal center. (Supplementary Figure 1) The side chain of Tyr385 still stacks with the imidazole ring of TMH, but the hydrogen bond interactions to Asn386 and the backbone carbonyl of Tyr385 are lost.

Structure 4 was observed in a crystal of the CthEgtB/Fe/TMH complex that was obtained through co-crystallization. The crystal diffracted to a resolution of 2.4 Å and belonged to space group C2 with four

monomers in the asymmetric unit. The electron density revealed a continuous polypeptide chain with the exception of the interdomain segment 182–192. The (2Fo – Fc) electron density map revealed weak density for the C- α backbone of active site loop 1, for which side chain conformations could not be resolved (Supplementary Figure 4). The side chains were initially included for modelling, to ensure that the modelled conformation could fit each side chain. However, these were removed in the final stages of refinement. TMH could be modeled into the (Fo – Fc) difference density map (Supplementary Figure 1). Loop 2 occupies the same conformation and makes the same interactions with TMH and the neighboring monomer as described for structure 2 (Figure 2F). The C- α backbone conformation of loop 1 is restricted by the available electron density, while side chain conformations remain uncertain. However, the position of the backbone suggests that catalytic residues Tyr93 and Tyr94 are pulled out of the active site.

The structural behavior of loop 1 and 2 in the four structures provides important information about the mechanism of substrate binding. Structure 1 (Native *CthEgtB*) features an active site that appears wide open, because the loops are not resolved. This raised the question as to whether loop unfolding opens a channel through which the substrates access the active site. Inspection of structures 3 and 4 show that the alternative conformations of loop 1 and 2 do not open any new channel. Alternatively, a large tunnel that accesses the active site of *CthEgtB* from the water filled core of the tetramer is a more likely entrance for TMH and cysteine. This tunnel remains open and invariant in all four structures. The same tunnel occurs in type I sulfoxide synthases which do not contain any flexible loops near the active site.⁴⁹

CthEgtB forms a stable homotetramer with the shape of a donut, with D₂ symmetry. The first interface is formed between ends of the N-terminal α -helix bundle (DinB-2-like domain). The second interface is located between the C-terminal domain (FGE-like domain Figure 3). Because the two flexible loops are located close to the the second interface, their conformation could impact the shape to the tetramer. However, comparison of all four structures showed that the relative orientations of monomeric units do not change. Hence, we have no structural evidence that binding of TMH and/or folding of the loops would induce conformational changes across the tetramer.

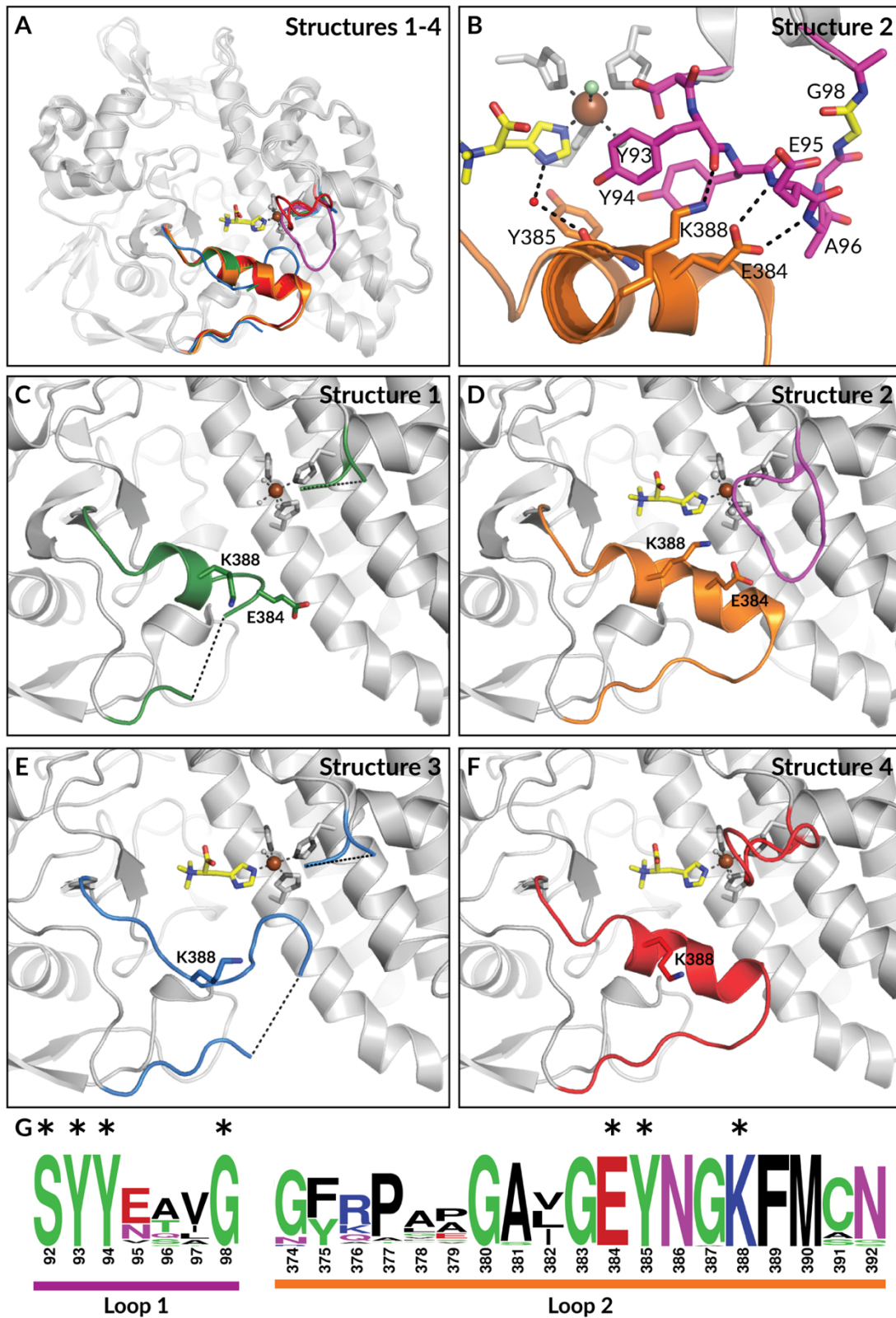


Figure 2. A. Superposition of the four *CthEgtB* structures, each with a distinct loop conformation, shown as cartoons. Loop Regions are colored by structure B. Conformation of active site loops in the *CthEgtB*.WT/Fe/TMH closed loop structure. Important residues and interactions are shown as sticks and dashes respectively. C. Loop conformation in the native structure with disordered loops (PDB: 6QKI, green). D. Loop conformations in the TMH containing, closed structure with ordered loops closing over the active site, loop 1 in purple, loop 2 in orange (PDB: 6QKJ). E. Loop conformations in the partially open loop 2 structure in blue. F. Loop conformations in the open loop 1 structure in red. G. Sequence logo for active site loops 1 and 2 (N = 100). An asterisk indicates an important and conserved residue.

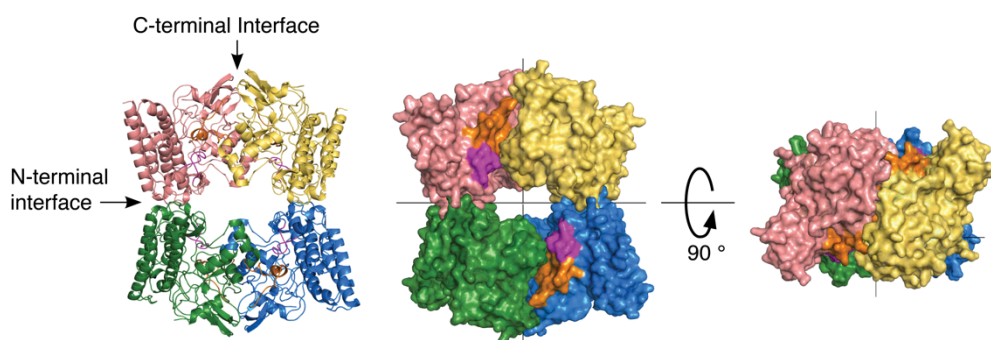


Figure 3. Cartoon and surface representation of the *CthEgtB* closed loop tetrameric (PDB: 6QKJ). Each monomer is in a different color, while active site loop 1 is in magenta and loop 2 in orange, making up part of the C-terminal interface.

Identification of Gly98 as a hinge region. A sequence alignment of sulfoxide synthases type I – V show that active site loop 1 in type II corresponds to a six-residue insertion that is not present in the other types.¹⁷³ Four of these inserted residues (Ser92, Tyr93, Tyr94 and Gly98) are strictly conserved. Gly98, is the last residue in the disordered loop 1 in structures 1 and 3, indicating this residue may function as a hinge connecting the flexible loop with the rigid body of the protein.¹⁷⁴⁻¹⁷⁷ Indeed, Gly98 adopts different conformations in structure 2 ($\phi = 155^\circ$, $\psi = -164^\circ$) and structure 4 ($\phi = 65^\circ$, $\psi = 23^\circ$). Both of these conformations are unfavorable for other residues due to steric side chain-to-backbone interactions.¹⁷³ To test the idea that Gly98 is conserved because this residue stabilizes the active conformation of loop 1 we examined the effect on activity of a Gly to Ala mutation at this position. Structural modelling suggest that the additional methyl group does not cause steric clashes with other parts of the protein so that the observed effects can be attributed to conformational effects.

Impact of the Gly98 mutation on catalytic activity. To this end we produced recombinant *CthEgtB*_{G98A} and examined its activity using an established set of assays.^{49, 173} Briefly, time-dependent sulfoxide production was monitored by HPLC. Maximal rates were measured (k_{cat}) in the presence of a constant and saturating concentration of TMH and cysteine. The affinity of TMH ($K_{D, TMH}$) for the enzyme was examined by isothermal calorimetry titration (ITC). In addition to the sulfoxide, *CthEgtB* is known to also produce cysteine dioxide as a minor side product.¹⁷³ The ratio between sulfoxide and cysteine dioxide was determined by NMR analysis of exhausted reactions that started with an excess of TMH.

As the only significant effect of the Gly98 to Ala mutation we observed that *CthEgtB*_{G98A} - unlike wild type - is not fully active at the start of the reaction. After a lag-phase of about 30 minutes the productivity of the variant is indistinguishable from that of the wild type (Table 1). The lag-phase indicates that an inactive form of *CthEgtB*_{G98A} converts to an active form in a unimolecular process with the rate of $4.2 \times 10^{-8} \pm 6.0 \times 10^{-5} \text{ s}^{-1}$. Similar rates were measured at 1 or 3 mM TMH, suggesting that the rate limiting step in activation does not involve substrate binding.

Binding of TMH was examined by ITC using Mn^{II} reconstituted *CthEgtB* variants. The enzymes were reconstituted with Mn^{II} instead of Fe^{II} to avoid unwanted heat producing redox reactions during the titration experiment. However, comparable ITC measurements show that Mn^{II} or Fe^{II} containing *CthEgtB* bind TMH with a K_D of 40 μ M or 10 μ M respectively, suggesting that Mn^{II} is a reliable substitute for Fe^{II}. More importantly, we found that Mn^{II} containing *CthEgtB*_{G98A} also binds TMH with a K_D of 40 μ M. The same Mn^{II} - containing proteins were used to assess their thermal stability by Differential Scanning Fluorimetry (DSF) (Supplementary Figure 5). Again, both proteins behaved indistinguishably (Table 1). The melting point of enzymes increased by 10°C by the addition of 10 mM TMH, indicating that substrate binding does contribute significantly to the stability of the entire protein.

In conclusion, the G98A mutant proved remarkably inconspicuous. The only observed effect of this mutation is the introduction of a lag-phase preceding steady-state sulfoxide production. Apparently, the variant enzyme is resting in an inactive form. Conversion to the active form occurs as a unimolecular process that is slow with respect to catalytic turnover.

Table 1. Measured kinetic and thermodynamic parameters for *CthEgtB*.WT and various mutants.

	k_{cat} (μ M s ⁻¹) * [TMH] = 1 mM	k_{cat} (μ M s ⁻¹) * [TMH] = 3 mM	% Sulfoxide : Sulfinic Acid	K_D - TMH (μ M) EgtB•Fe	K_D - TMH (μ M) EgtB•Mn	Melting Temp. (°C) Native	Melting Temp. (°C) + 10 mM TMH	% tetramer @ 20 μ M
<i>CthEgtB</i> _{WT}	0.05 ± 0.01	-	80:20 ± 10	9.8 ± 0.2	37 ± 1	71.3 ± 0.1	79.8 ± 0.1	100
<i>CthEgtB</i> _{G98A}	0.08 ± 0.01	0.06 ± 0.01	80:20 ± 10	-	38 ± 3	68.8 ± 0.1	79.1 ± 0.1	99
<i>CthEgtB</i> _{CF}	0.09 ± 0.01	0.10 ± 0.01	70:30 ± 10	-	67 ± 4	51.9 ± 0.1	57.8 ± 0.1	74
<i>CthEgtB</i> _{CF/G98A}	0.07 ± 0.01	0.06 ± 0.01	80:20 ± 10	-	> 500	50.7 ± 0.1	56.6 ± 0.1	42
<i>CthEgtB</i> _{CF/G98V}	0.04 ± 0.01	0.05 ± 0.01	70:30 ± 10	-	> 500	50.2 ± 0.1	56.4 ± 0.1	29

* 0.5 μ M enzyme was used.

Global elimination of cysteine destabilizes the tetramer. Since the two active site loops are close to one of the monomer-monomer interfaces (Figure 3), we explored the possibility that the behavior of these loops may be connected to the quaternary structure. Indeed, the Gly98 to Ala mutation proved much more informative when inserted into a *CthEgtB* variant with reduced stability of the quaternary structure. In the context of an unrelated experiment, we constructed a variant of *CthEgtB* in which all cysteine residues were mutated to residues (Ala, Phe or Ser) that occupy these positions in closely related homologs (*CthEgtB*_{C37A/C231F/C303S/C391S/C401A} = *CthEgtB*_{CF}). None of these cysteine residues are conserved or participate in the active site of *CthEgtB*_{CF} (Supplementary Figure 6). Hence, we were not surprised to find this variant as active as wild type and to produce sulfoxide and cysteine dioxide in a similar ratio (Table 1). Also, the affinity for TMH was reduced only two-fold.

On the other hand, the activity of *CthEgtB*_{CF} is characterized by the same lag-phase as observed with *CthEgtB*_{G98A} (Figure 4), and the melting point of *CthEgtB*_{CF} is lowered by 22 °C (Table 1). Examination of the quaternary structure of the *CthEgtB* variants by analytical gel filtration indicated that this loss of thermal stability may, to a large part, arise from destabilization of the quaternary structure (Supplementary Figure 7). *CthEgtB* and *CthEgtB*_{G98A} elute from the size exclusion column in a single peak with a retention time consistent with a tetrameric structure. The oligomerization state remained invariant at concentrations

ranging from 2 μM to 100 μM , or in the presence of 20 % MeOH (Supplementary Figure 8). *CthEgtB*.CF on the other hand, behaved predominantly as a tetramer (90 %) when injected at high concentration (100 μM), but 50 % of the protein eluted as a monomer at low concentration (2 μM). Notably, the tetrameric and the monomeric peaks are well resolved and their elution time was not dependent on concentration, indicating that interconversion of the two species is slow compared to the time scale of the experiment.

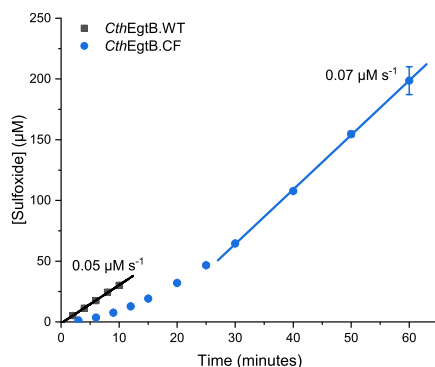


Figure 4. Sulfoxide production as a function of time, catalyzed by *CthEgtB*.WT and *CthEgtB*.CF. A lag phase is observed in the reaction with *CthEgtB*.CF.

In short, *CthEgtB*_{CF} provides a fully active enzyme with weakened quaternary structure and an inactive resting state that is slow to activate. Introduction of the Gly98 to Ala mutation into this protein resulted in a variant (*CthEgtB*_{CF/G98A}) that is fully active at high substrate concentration and after a lag-phase, and produces the same product ratio as wild type (Table 1). However, the mutation affected substrate binding and the quaternary structure. We were unable to detect binding by ITC, indicating that the affinity for TMH is decreased by at least ten-fold. The Gly98 to Ala mutation did not reduce thermal stability significantly as inferred by DSF, but analytical gel filtration revealed that *CthEgtB*_{CF/G98A} elutes predominantly as a monomer at low concentrations (Figure SX). These effects are even more pronounced when Gly98 was replaced with Val (*CthEgtB*_{CF/G98V}). At low concentrations this protein elutes almost completely as a monomer (Supplementary Figure 14). Upon closer inspection of the chromatograms for *CthEgtB*_{CF/G98A} and *CthEgtB*_{CF/G98V} one can spot an additional peak consistent with a dimeric species. Possibly, the mutation on loop 1 is specifically destabilizing on interface 2 leading to disassembly into dimers instead into monomers (Figure 5 D&E). It is remarkable that *CthEgtB*_{CF/G98A} and *CthEgtB*_{CF/G98V} are still almost as active as wild type in the presence of saturating concentrations of substrates (Table 1). These results raise the question as to whether the quaternary structure is at all necessary for catalysis.

To address this question we examined the oligomeric state of *CthEgtB*_{CF/G98A} and *CthEgtB*_{CF} before and after preincubation with 10 mM *N,N*-Dimethyl-L-histidine (DMH), and with 10 mM DMH in the elution buffer. DMH is synthetically more accessible than TMH but is recognized by EgtB with similar efficiencies.¹⁷⁸ The presence of DMH shifted the equilibrium towards the tetrameric form for *CthEgtB*_{CF/G98A} (25 % to 80 % at 20 μM) and for *CthEgtB*_{CF} (65 % to 90 % at 20 μM) (Supplementary Figure 7). Similarly, the presence of 10 mM TMH increased the melting temperatures of both variants by 6 °C (Table 1). These results suggest that

under the conditions used to assess catalytic activity, all *CthEgtB* is tetrameric. We have no evidence that the activity of the *CthEgtB* variants is dominated by the monomeric form.

Discussion

Our key objective is to understand the catalytic mechanism of O₂ activating enzymes. Sulfoxide synthases are a unique class of enzymes that are distinguished from other non-heme enzyme by their structure and reactivity. Furthermore the availability of different EgtB types enables us to compare different solutions to the same problem. *CthEgtB* is distinct from other EgtB types due to its homo-tetramerization and containment of two dynamic active site loops, which appear to fold in a substrate dependent manner. Herein we disentangle the role of Gly98 in loop folding and quaternary structure and discuss a possible evolutionary trajectory to the type II EgtBs.

Approach to mutant characterization. Through a combination of kinetic and thermodynamic measurements we are able to dissect a mutant phenotype to determine if a mutation manifests itself in (1) the formation of the reactive complex, influencing the quaternary structure and or substrate K_D . Alternatively, a mutation could (2) interfere with the reactivity of the reactive complex, which would be demonstrated by perturbation in catalytic activity and/or product distribution. We assess quaternary structure through size-exclusion chromatography (SEC), while measurement of the K_D for TMH by Isothermal titration calorimetry (ITC) provides a measure of substrate affinity that is independent of catalysis, thereby representing the substrate binding events that precede formation of the reactive complex. In contrast, if catalysis or the geometry of the reactive complex is perturbed, V_{max} and product distribution will be altered respectively. As the substrates bind with an ordered binding mechanism and we hypothesized that the G98A mutation may affect substrate binding, we assessed activity through measurement of V_{max} by following the production of the sulfoxide product in a reaction mixture containing FeSO₄, ascorbate and TCEP under aerobic conditions, following published procedures.^{31, 41, 49-50, 120} By ¹H NMR we can determine the ratio of sulfoxide to sulfinic acid production. Sulfinic acid is a common side reaction for many of the sulfoxide synthases, and becomes the dominant activity upon perturbation of the active site geometry.^{50, 71-72, 111, 173} Even minimal perturbations favor the less complex dioxygen pathway.^{111, 179} As the two competing pathways share a common intermediate that follows formation of the reactive complex, product distribution is a valuable measure of the perturbation of active site geometry of the reactive complex.⁵⁰ Finally, thermal stability is assessed by differential scanning fluorimetry (DSF), a thermal shift assay which measures the intrinsic tryptophan fluorescence that changes upon protein denaturation.¹⁸⁰⁻¹⁸¹

Catalysis is unperturbed. In contrast, the mutation of Gly98 presented no phenotype. However, in the context of the destabilized cysteine-free system, the G98A & V mutants destabilized the closed loop conformation, as indicated by impaired TMH binding. Quaternary structure is also further destabilized compared to the CF variant, how primarily through the disruption of just one interface, most likely the C-terminal interface. This mutant phenotype indicates that steps in the formation of the reactive complex are perturbed. Despite this destabilization, catalysis proceeds as normal in the presence of saturating substrates, with the exception of the lag phase. This lag phase is also consistent with changes in how the reactive species is formed. The reactive species that does form, has the same geometry and features of the WT active site, as indicated by the identical production distribution.

TMH binding induces concerted loop folding. In the destabilized system, CF G98A & V mutants, formation of the reactive complex becomes limited by loop folding. The effect of this mutation is curious, as TMH is bound and packed by active site loop 2, while Gly98 sits on active site loop 1, and is at its closest 16 Å away from TMH. This provides evidence for concerted loop closing, which is coupled to TMH binding. The closed, catalytically relevant conformation of active site loop 2 is stabilized by the closing of active site loop 1 through enthalpic contributions, via hydrogen bonding interactions of Lys388 and Glu384, and entropic contributions, as the closed loop 1 conformation reduces the degrees of freedom available for loop 2. Thereby, through mutation of Gly98, the closed loop conformation of loop 1 is destabilized, which consequently destabilizes active site loop 2, which forms part of the TMH pocket. The partially open loop 2 structure has TMH bound, yet active site loop 2 is not in a closed conformation, likely while active site loop 1 is not closed and therefore does not stabilize the catalytically relevant conformation. It is likely that the WT tetramer is so stable that it militates the effect of the single mutant in *CthEgtB*.G98A, providing evidence that the quaternary structure stabilizes the enzyme:substrate complex.

TMH binding stabilizes the *CthEgtB* tetramer. In addition to stabilizing a specific active site loop conformation, tetramerization also appears to provide global stability, as the *CthEgtB* variant has a much lower thermal stability ($\Delta = 22$ °C). Despite this high stability, DSF measurements indicated that the tetramer is further stabilized by the binding of TMH alone (9 °C increase in melting temperature). Incubation of all the mutants with 10 mM TMH provided stabilization by approximately 6 °C. Incubation of *CthEgtB*.CF and *CthEgtB*.CF.G98A with TMH analogue, DMH, was found to dramatically shift the oligomeric equilibrium observed by SEC to populate primarily the tetrameric form. These results further indicate a link between TMH binding, the active site loops and oligomeric structure. However consequently, we cannot determine if the monomeric species is active or not. The enhanced stability upon TMH binding provides further evidence that either one or both close upon TMH binding. This enhancement is also observed for the WT and CF *CthEgtB* constructs, indicating that the loops are dynamically moving between open and closed conformations, advocating for a breathing motion of the tetramer. The two new crystal structures provide two snap shots of both loops in open forms, showing that a dynamic range of conformations can be taken up by these two loops. This provides an example in which evolution of quaternary structures optimises dynamic properties.¹⁸² Under substrate-saturating conditions, tetramer formation is aided by TMH binding to form the Michaelis Menten complex, which, once formed, allows for reactivity as active as the WT. This observation should provide valuable lessons for computer-assisted enzyme design, which for the most part focusses on monomeric designs. This insight could be leveraged for rational design of enzymes to overcome issues with substrate binding or loops that have inherent flexibility.

Loop folding may provide metabolic regulation. These findings show that the *CthEgtB* active site and subsequently the oxygen-binding site is only formed upon binding of TMH. This presents an alternative way to facilitate O₂ binding to its active site: a dynamic response to substrate binding rather than a rigid, permanently composed O₂ channel as in *MthEgtB*. It is possible that the dynamic *CthEgtB* active site could provide a form of metabolic regulation. TMH is therefore the first committed intermediate towards EGT

biosynthesis. EgtD, the methyl transferase responsible for the production of TMH from histidine, is the gateway to EGT biosynthesis and is subject to stringent feedback product regulation.¹⁸³ The catalytically relevant *CthEgtB* active site, and oxygen binding pocket is only stabilized and formed when TMH is present and bound. This may also provide an alternative strategy to preventing undesirable iron-Cysteine or iron-O₂ chemistry before all substrates are present.

Oligomerisation enables evolutionary divergence. The evolutionary emergence of the EgtB family and various types is an intriguing topic which we recently addressed. We proposed a common EgtB ancestor was monomeric and used an alternative thiol donor could account for most of the variety observed in the present EgtB family.¹⁷³ However, this does not provide any insight into the evolutionary path or selection pressures that lead to complete active site remodeling and introduction of flexible loops to form the type II EgtBs. This study did not provide an obvious advantage for the loop architecture. What is apparent is that active site loops 1 and 2 correspond to six and two residue insertions respectively and comprise part of the C-terminal interface. Interestingly, insertions have a significant tendency to be located on interfaces and it has been suggested that small insertions or deletions represent an important evolutionary mechanism for oligomerization, having a profound impact on complex stability.^{170, 184} Furthermore, in many oligomeric proteins, amino acid residues that form the intermonomer interfaces are located on extended loop regions.¹⁶⁸ It is therefore tempting to suggest that the insertions in type II EgtBs are not only catalytic loops but also complex enabling regions that may have been an essential step in tetramer evolution. This is supported by (1) the observation that all type II EgtBs have these catalytic loops and are believed to be tetrameric. (2) loops correspond to small residue insertions that make up part of the C-terminal interface and (3) our finding that the loops and quaternary structure are connected, with the tetramer stabilizing the active enzyme form. Perhaps the stability gained upon tetramerization enabled the radical active site redesign observed for the type II EgtBs as an ancestral EgtB underwent rapid diversification into types I – V. This provides an example in which oligomerization enables evolutionary divergence, providing an explanation for how the dramatic active site remodeling and catalytic residue hopping of the type II EgtBs could have developed.

Conclusion

Homo-oligomerization is well established to confer global stability or to install a means of functional regulation such as allostery or co-operativity. The interplay between oligomerization, dynamic movement and catalysis is less understood. We provide an example in which the homotetrameric structure of a type II EgtB reduces dynamic fluctuations to stabilize the active conformation of catalytic loops and increase specificity. We show that i) loop folding is coupled to TMH binding, ii) the quaternary structure stabilizes the enzyme:substrate complex and iii) loop folding is a limiting factor. The connection between loop folding and quaternary-structure suggests that tetramerization was an essential strategy to stabilize active enzyme form and to optimize enzymatic performance. We also propose that oligomerization enabled divergent evolution of the EgtB family, providing an explanation for the extreme active site plasticity and divergence of type II EgtBs.

Addendum on *CthEgtB* Ser92

What is the role of Ser92 on the *CthEgtB* active site loop 1?

Active site loop 1 contains four conserved residues Ser93, Tyr93, Tyr94 and G98. The first three are delivered to the active site upon loop closing so that Ser92 points into the active site (Figure 3C) In the closed crystal structure, the Ser92 β -hydroxyl group is observed in two conformations of equal occupancy that can hydrogen bond to either the axial or with the equatorial chloride ligands (O-Cl_{ax}, 2.9 Å; and O-Cl_{eq}, 3.1 Å) (Figure 1). We surmised that S92 could either be involved in catalysis as Y93 or Y94 or in stabilizing the activation conformation of loop 1, as Gly98 is. To investigate the role of S92, two point mutants were generated: a Ser to Ala mutation (*CthEgtB.S92A*), in which the hydrogen capabilities of the amino acid side group was completely removed and a Ser to Cys mutation (*CthEgtB.S92C*) in which serine was replaced with a more acidic hydrogen bond donor ($pK_a(\text{Ser}) = 16$ vs $pK_a(\text{Cys}) = 9$). Both mutants were characterized using Michaelis-Menten kinetics, varying both TMH and cysteine (Table 1, Figure 2) The serine to cysteine mutation catalyzes multiple turnovers without any sign of inactivation, and exclusively affected k_{cat} , with a 30-fold decrease in reactivity. The serine to alanine mutation affected k_{cat} (7-fold decrease), product distribution, and cysteine K_M , which increased to 500 μM , 14 times higher than that for the WT. This indicates that hydrogen bonding is important for substrate binding, while the introduction of a stronger hydrogen bond impedes catalysis.

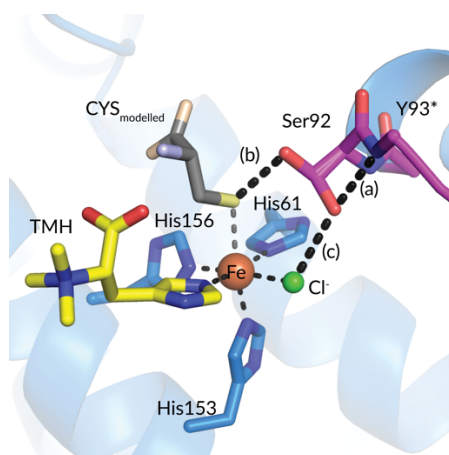


Figure 1. Conformations and possible hydrogen bonding interactions (labelled (a) – (c).) of Ser92 in the *CthEgtB.WT/Fe/TMH* structure with closed loops (PDB: 6QKJ). * The sidechain of Tyr93 has been omitted for clarity.

Table1. 1 Michaelis-Menten Parameters for *CthEgtB.WT* and mutants, *CthEgtB.S92A* & *CthEgtB.S92C*.

	TMH			Cysteine			% Sulfoxide : Sulfinic Acid
	k_{cat} (s ⁻¹)	K_M (μM)	k_{cat}/K_M (M ⁻¹ s ⁻¹)	k_{cat} (s ⁻¹)	K_M (μM)	k_{cat}/K_M (M ⁻¹ s ⁻¹)	
<i>CthEgtB.WT</i>	0.050 ± 0.002	24 ± 4	2080	0.06 ± 0.003	35 ± 6	1700	80:20 ± 10
<i>CthEgtB.S92A</i>	0.008 ± 0.001	19 ± 3	430	0.008 ± 0.001	490 ± 60	20	50:50 ± 10
<i>CthEgtB.S92C</i>	0.0022 ± 0.0001	69 ± 6	32	0.0019 ± 0.0001	40 ± 8	50	80:20 ± 10

Active site loop 1 contains four residues which are conserved among all type II EgtBs: Ser92, Tyr93, Tyr94 and Gly98. The roles of Tyr93 and Tyr94 have been established as a catalytic acid and lewis acid respectively, involved in stabilizing the iron bound superoxide. This study explores the role of two other conserved residues, Ser92 and Gly98, for a complete functional assignment of conserved loop 1 residues. Mutation of Ser92 to an alanine affected k_{cat} , K_M and product distribution, while replacement with cysteine only affected k_{cat} , indicating that the function of the serine side chain as a hydrogen bond donor is essential to stabilize the reactive species complex.

Structural Role of S92. Kinetic analysis of Ser92 mutants shows that a hydrogen-bonding interaction is important for the binding of substrate cysteine and that the strength of the hydrogen bond plays a role in modulating reactivity. Three possible hydrogen bond partners are viable for Ser92: i) the backbone amide of Tyr93 as a hydrogen bond donor (O-N: 2.9 Å, C β -O- N: 73.3° C β -O-S). ii) the substrate cysteine thiolate as an acceptor (O-Cl_{ax}: 3.1 Å, C β -O- Cl_{ax}: 117.8°) and iii) the distal oxygen, also an acceptor (O-Cl_{eq}: 2.9 Å, C β -O- Cl_{eq}: 97.8°). (Figure 6 (a)-(c)). The loss of an interaction to the amide backbone upon mutation of Ser to Ala could destabilize the closed loop conformation, which would account for the change in K_M and product distribution. However loop 1 is held in place by several other interactions. While S92C replaces a good H acceptor (R-OH) with a poor acceptor (R-SH), this hydrogen-bonding partner alone does not account for the decrease in reactivity upon substitution of Ser for Cys. This indicates that Ser92 must act as an H bonding donor. To predict the identity of the hydrogen bond acceptor, we theorized how substitution of serine with a more acidic residue would influence catalysis for both hydrogen bond acceptors. Hydrogen-bonding acceptor/donor pairs with matching pK_a s generally form stronger H bonds than those with unmatched pK_a s. Therefore a hydrogen bond to substrate thiol would be stronger for cysteine than serine, which would increase the redox potential of metal thiolate compared to the WT, disfavoring formation of thiyl radical, slowing down proton-coupled electron transfer in the proposed mechanism.⁶⁰⁻⁶¹ This could rationalize the 30-fold decrease in reactivity observed for CthEgtB.S92C. A hydrogen -bonding interaction between Ser92 and cysteine substrate seems, however, counterintuitive by virtue of the findings from a Ala82Ser mutation in MthEgtB, in which engineering of a serine to substrate cysteine hydrogen bond was found to be counterproductive for reactivity. However, the CthEgtB and MthEgtB active sites are completely different, exemplified by the requirement of CthEgtB for a second catalytic tyrosine while MthEgtB requires only one. Therefore a hydrogen bond to the substrate thiol in CthEgtB may not be counterproductive. Alternatively, S92 could hydrogen bond to the distal oxygen of the molecular oxygen/superoxo species. The greater acidity of cysteine in place of Ser92 would increase the electron affinity of the iron(III) superoxide species for electrons. We believe this would aid reduction of oxygen in the first catalytic step. Following this logic, it is unlikely this substitution could account for the CthEgtB.S92C phenotype or the increase in cysteine K_M for the CthEgtB.S92A mutant. From this analysis, the thiolate of substrate cysteine seems the most probable hydrogen-bonding partner for Ser92, playing an important role in substrate binding and orientation.

CthEgtB.WT, S92A & S92C Michaelis-Menten Curves. All measurements are in triplicate.

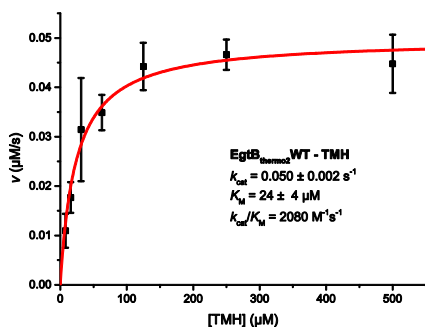


Figure 2A: *CthEgtB.WT*, variable TMH. Assay conditions included: 100 mM phosphate, pH 8, 100 mM NaCl, 3 mM TCEP, 2 mM ascorbic acid, 4 μM FeSO_4 , 1.5 mM cysteine, TMH was varied at 500 μM , 250 μM , 125 μM , 62.5 μM , 31.25 μM , 15.625 μM and 7.81 μM . The reaction was initiated with 1 μM of *CthEgtB.WT*.

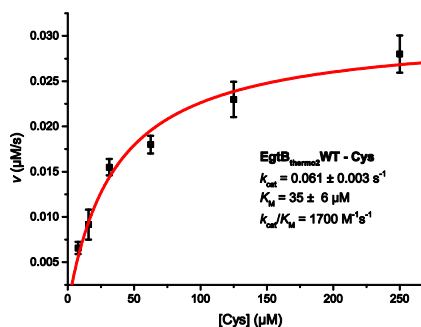


Figure 2B: *CthEgtB.WT*, variable Cysteine. Assay conditions included: 100 mM phosphate, pH 8, 100 mM NaCl, 2 mM TCEP, 2 mM ascorbic acid, 4 μM FeSO_4 , 0.5 mM TMH, Cysteine was varied at 500 μM , 250 μM , 125 μM , 62.5 μM , 31.25 μM , 15.625 μM and 7.81 μM . The reaction was initiated with 0.5 μM of *CthEgtB.WT*.

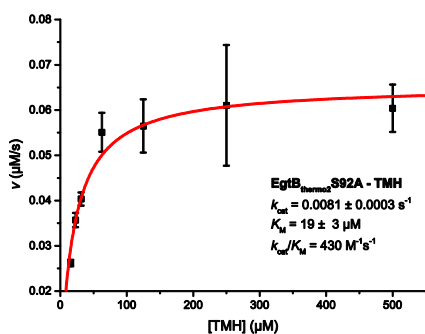


Figure 2C: *CthEgtB.S92A*, variable TMH. Assay conditions included: 100 mM phosphate, pH 8, 100 mM NaCl, 3 mM TCEP, 2 mM ascorbic acid, 32 μM FeSO_4 , 1.5 mM cysteine, TMH was varied at 500 μM , 250 μM , 125 μM , 62.5 μM , 31.25 μM , 15.625 μM and 7.81 μM . The reaction was initiated with 8 μM of *CthEgtB.S92A*.

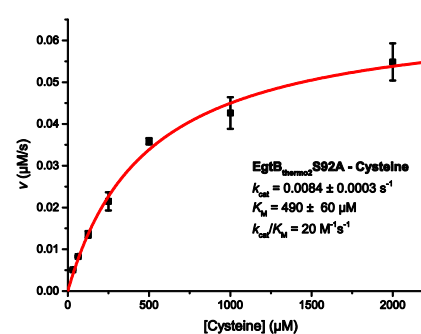


Figure 2D: *CthEgtB.S92A*, variable Cysteine. Assay conditions included: 100 mM phosphate, pH 8, 100 mM NaCl, 2 mM TCEP, 2 mM ascorbic acid, 32 μM FeSO_4 , 0.5 mM TMH, Cysteine was varied at 2000 μM , 1000 μM , 500 μM , 250 μM , 125 μM , 62.5 μM and 31.25 μM . The reaction was initiated with 8 μM of *CthEgtB.S92A*.

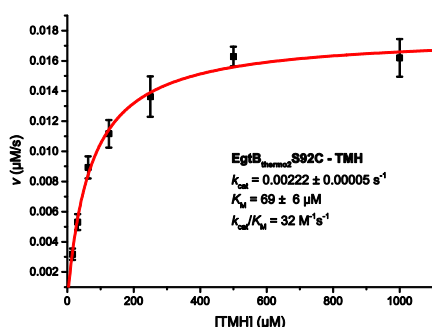


Figure 2E: *CthEgtB.S92C*, variable TMH. Assay conditions included: 100 mM phosphate, pH 8, 100 mM NaCl, 3 mM TCEP, 2 mM ascorbic acid, 32 μM FeSO_4 , 1.5 mM cysteine, TMH was varied at 1000 μM , 500 μM , 250 μM , 125 μM , 62.5 μM , 31.25 μM and 15.625 μM . The reaction was initiated with 8 μM of *CthEgtB.S92C*.

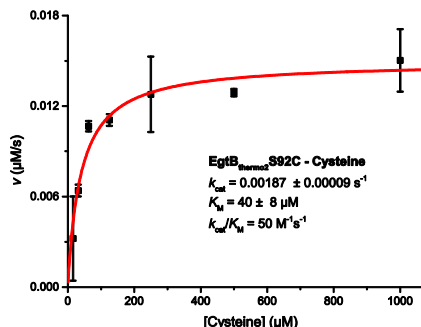


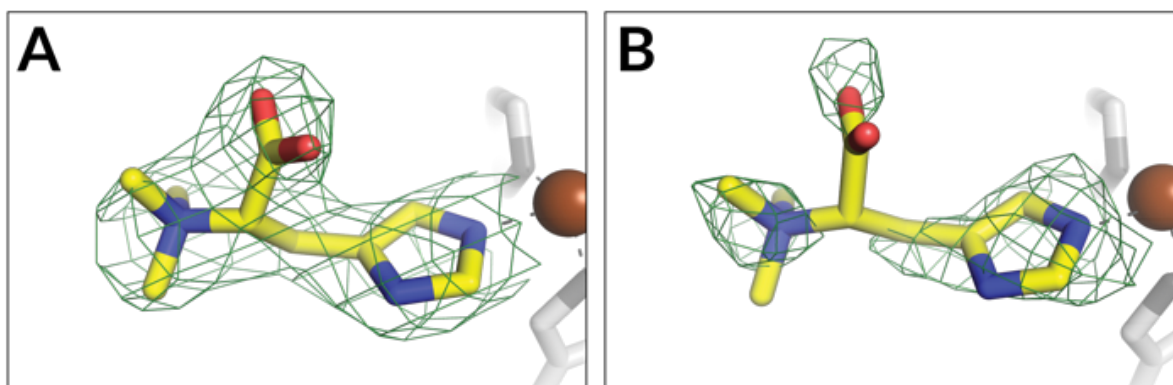
Figure 2F: *CthEgtB.S92C*, variable Cysteine. Assay conditions included: 100 mM phosphate, pH 8, 100 mM NaCl, 2 mM TCEP, 2 mM ascorbic acid, 32 μM FeSO_4 , 0.5 mM TMH, Cysteine was varied at 1000 μM , 500 μM , 250 μM , 125 μM , 62.5 μM , 31.25 μM and 15.63 μM . The reaction was initiated with 8 μM of *CthEgtB.S92C*.

Chapter 3:

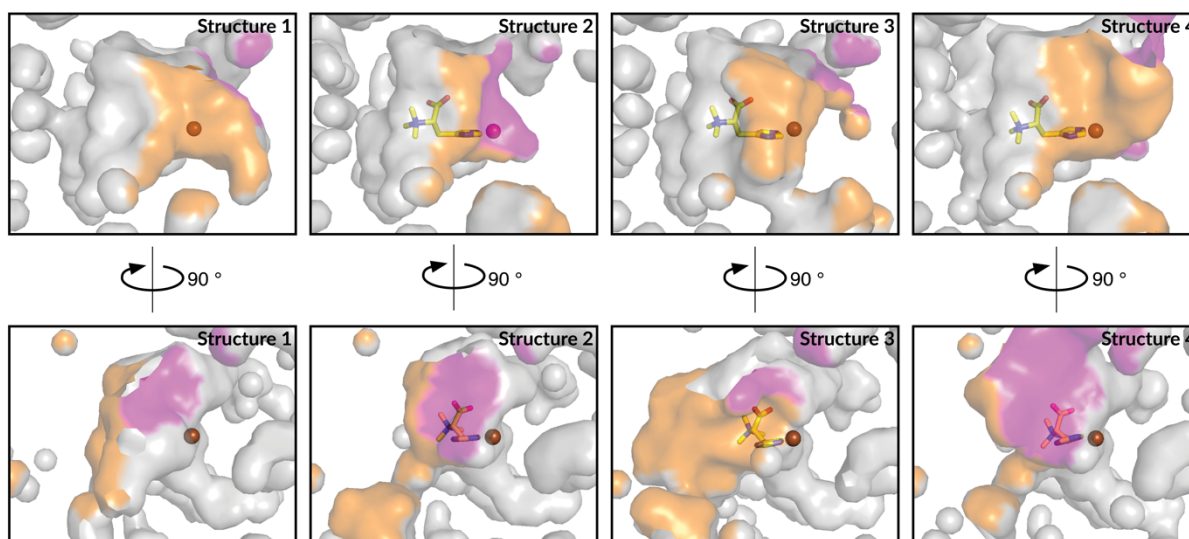
The Role of Oligomerisation and Loop Folding in a type II EgtB

Supplementary Information

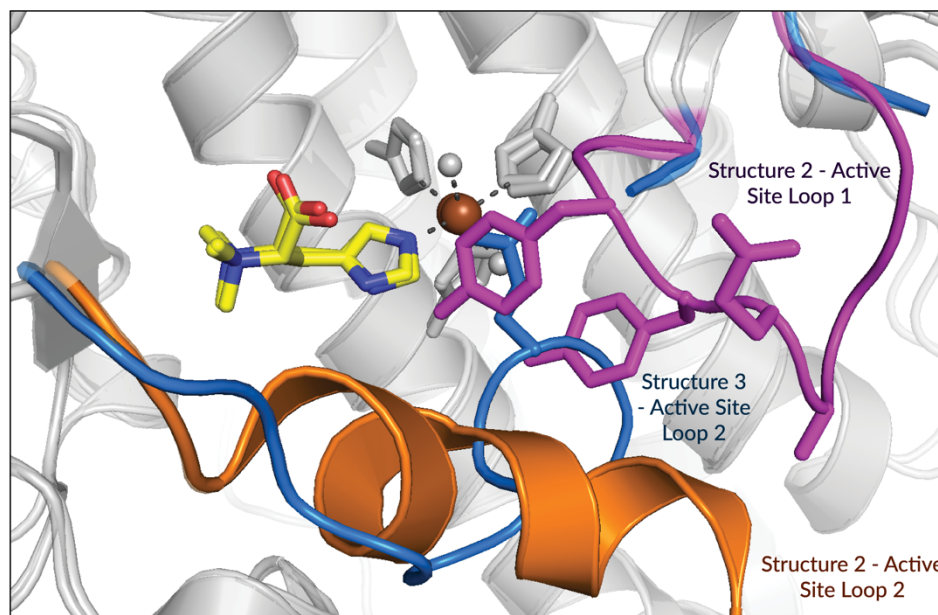
Supplementary Figures



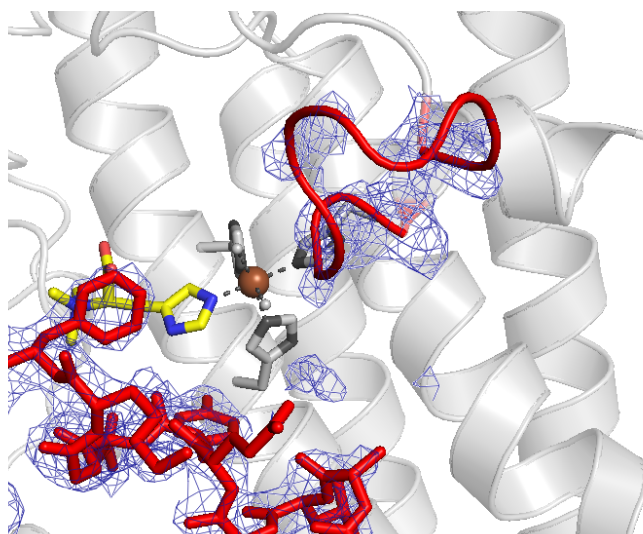
Supplementary Figure 1. Omit map of TMH in A: structure 3 and B: structure 4 ($m|F_o| - D|F_c|$, electron density; $\sigma = 3.0$).



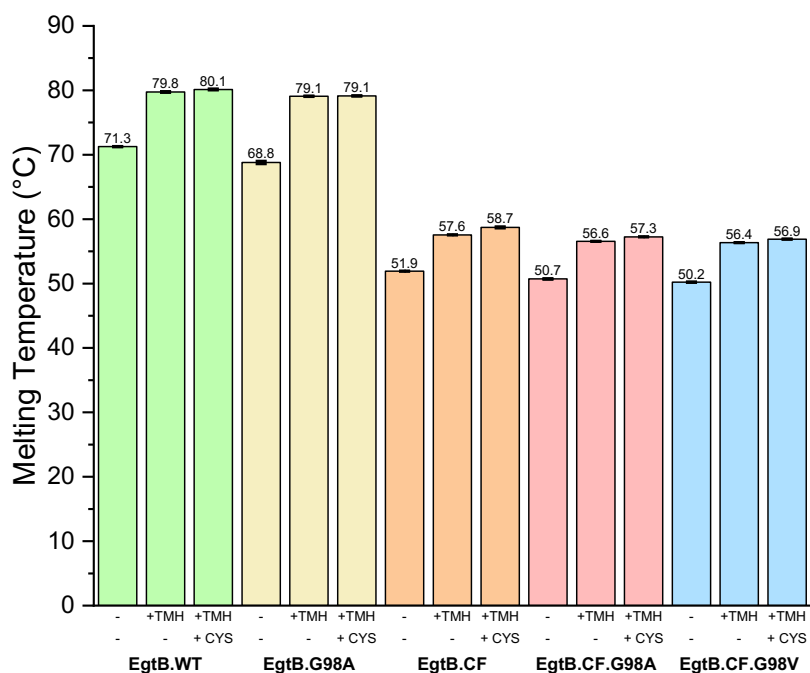
Supplementary Figure 2. TMH binding pocket for each of the four *CthEgtB* structures with various loop conformation. Surface areas are colored according to the segment which forms the pocket surface: Active loop 1 (magenta), active site loop 2 (orange) and remaining structure (gray).



Supplementary Figure 3. Superimposition of the *CthEgtB.WT/Fe/TMH* closed and partially closed structures. The conformation of loop 1 in the closed structure clashes with the conformation of active site loop 2 in the partially closed structure, indicating that closing of loop 1 decreases the degrees of freedom available for loop 2.

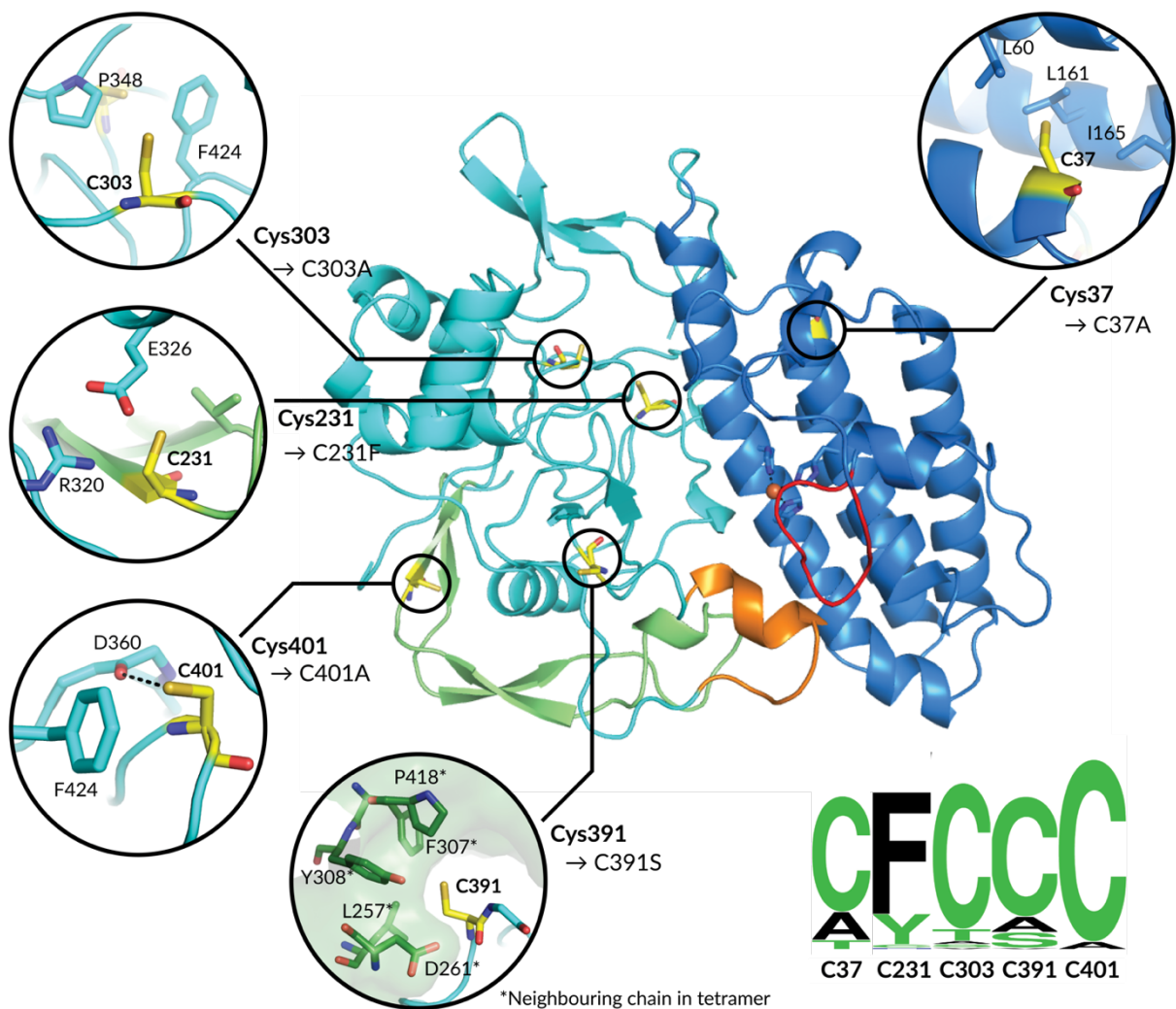


Supplementary Figure 4. Electron density ($2m|F_o| - D|F_c|$; $\sigma = 1.0$) for active site loop 1 of structure 4. This enables modelling of the C- α backbone, while side chain conformations could not be resolved.

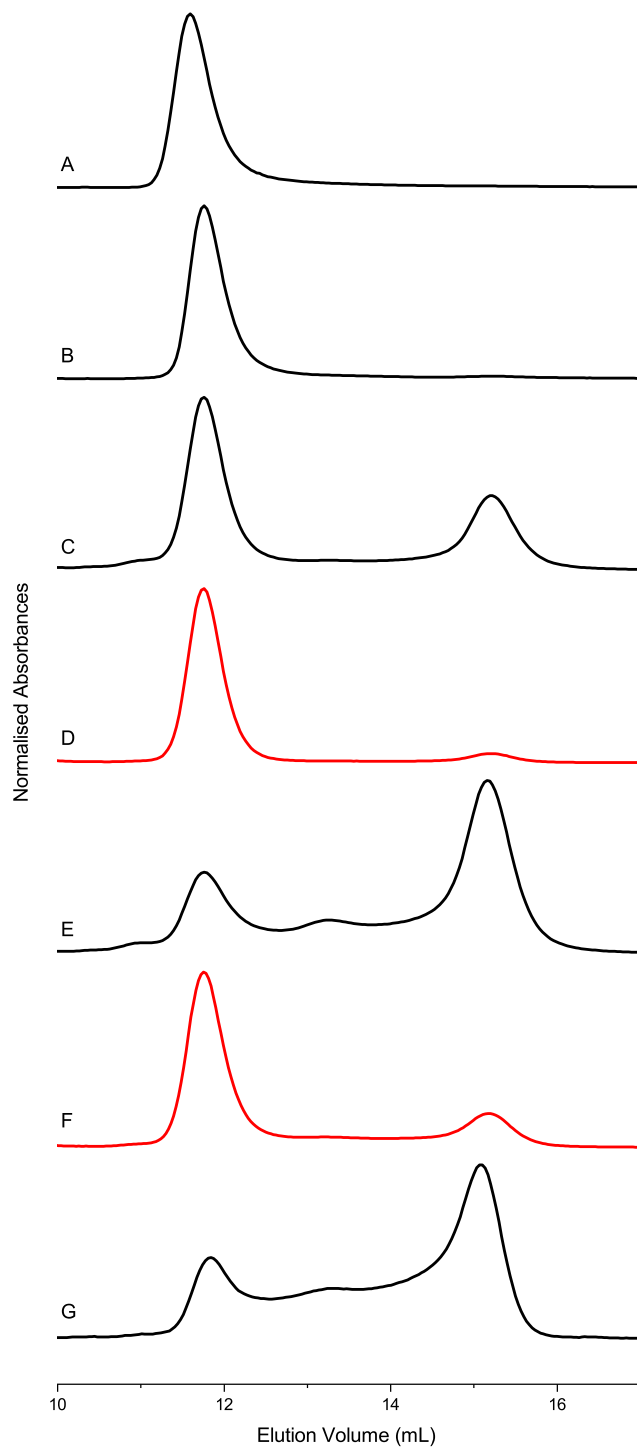


Supplementary Figure 5. Differential Scanning Fluorimetry (DSF) of *CthEgtB* variants with TMH and Cysteine: 4.9 μ M of *CthEgtB*.WT, *CthEgtB*.CF, *CthEgtB*.CF.G98A and *CthEgtB*.CF.G98V were incubated with 24.4 μ M $MnCl_2$ in 10 mM TCEP, 100 mM NaCl and 100 mM phosphate, pH 8. The TMH samples contained 10 mM TMH, the TMH + CYS samples contained 10 mM TMH and 10 mM CYS, the cysteine samples contained 10 mM CYS, while the control reactions lacked TMH.

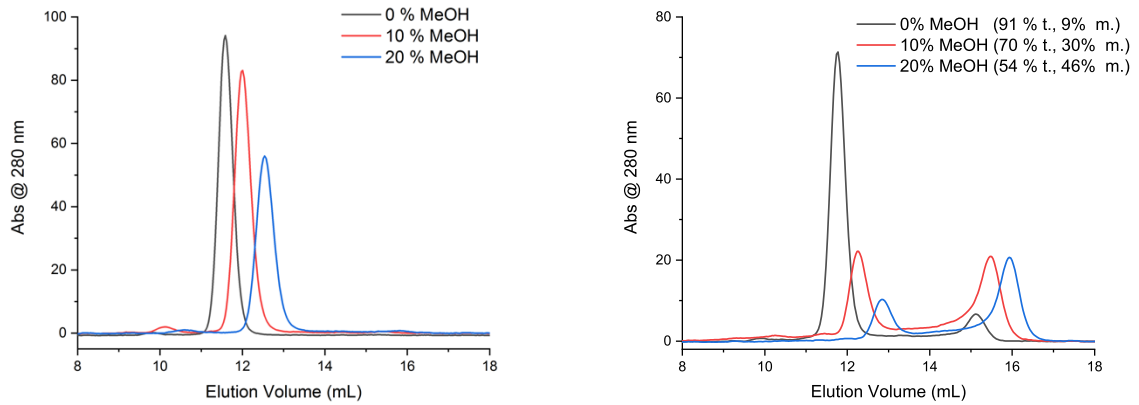
Cysteine – free construct. The cysteine-free (CF) variant is a *CthEgtB* variant in which all cysteine residues were mutated to the next most populated residue at this position ($CthEgtB_{C37A/C231F/C303S/C391S/C401A} = CthEgtB_{CF}$). All five cysteines are distant from the active site, are not conserved in all type II EgtBs and make limited side chain interactions. Cys37, Cys303 and Cys401 pack into a hydrophobic environment within the monomer, the former hydrogen bonding to the carbonyl backbone of Asp360. Cys391 is buried upon tetramerization, packing into the hydrophobic interface of neighboring monomer at the C-terminal Interface. Mutation to serine may reduce complementarity at the interface. Cys231 also packs into a hydrophobic environment, and its mutation to Phe could also introduce steric stress. The mutation of the latter two residues is most likely to cause slight destabilization of the *CthEgtB*.CF.



Supplementary Figure 6. Location and local environment of the five cysteines in *CthEgtB* WT, and the residues each cysteine is mutated to in the *CthEgtB*.CF construct. A sequence logo (N=100) is shown for each of the cysteine residues.



Supplementary Figure 7. Analytical Gel Filtration of *CthEgtB* variants. A. *CthEgtB*.WT B. *CthEgtB*.G98A C. *CthEgtB*.CF. D. *CthEgtB*.CF + 10 mM DMH E. *CthEgtB*.CF.G98A. F *CthEgtB*.CF.G98A + 10 mM DMH and G *CthEgtB*.CF.G98V. A-F contain 20 μ M enzyme, while G contains 40 μ M. Samples incubated with 10 mM DMH were run with a running buffer containing 10 mM DMH.



Supplementary Figure 8. Analytical Gel Filtration of *CthEgtB_{WT}* (left) *CthEgtB_{CF}* (right) in running buffer containing 0, 10 and 20 % Methanol: 100 μ L of 1 mg/mL (100 μ g total) *CthEgtB_{WT}* and *CthEgtB_{CF}* was injected onto a Superdex 200 Increase 10/300 column at room temperature with 200 mM NaCl, 50 mM Tris-HCl pH 8 containing either 0, 10 or 20 % MeOH, as running buffer (flow rate: 0.5 mL/min). Samples were pre-incubated with 5 eq. of FeSO₄, 10 eq. of Ascorbic Acid and 2 mM DTT.

Experimental

Recombinant *CthEgtB* constructs

The gene for EgtB from *Chloracidobacterium thermophilum* (*CthEgtB*, WP_014099806.1) was codon-optimized for protein production in *E. coli* and purchased from Genscript. The gene was ligated as NdeI-XhoI fragments into a pET28a cloning vector.

Sequence of pET28.*CthEgtB*.WT:

```
GSSHHHHHHSSGLVPRGSHMEAAARSHPEPIQSGEVSDRKAWQRHYRAVRAVSEAICQPLETEDYVVQMPDVSPP
KWHLGHTSWFFETFILKSLADYRPFHPRYDYIFNSYYEAVGARHPRPQRGLLTRPTVSEVYAYRAHVDAVERF
IAHSDTRTWAAALQPILELGLHHEQQHQELLLTDIKAILATNPLDPVYRPQPQPLPSPVEQLSPTGDWHIVEGGRY
AIGHAGRGFAFDNEGPRHDVLLRPCRIAAARPVNTNGEFLAFMADGGYRRPELWLSDGWAAVTARGWEAPLYWRQAA
DGTWETLTLHGVPVAPYEPVCHISFYEADAYARWAGKRLPTEAEWEVVAARLPVTGNFYESGVLHPRPVSVSAA
FYGDVWVWTASPYVGYPGFRPVSGALGEYNGKFMCNQMVLRGGSCATSLTHIRSTYRNFFPPDARWQFTGVRLAE
DMS
```

m/z (*CthEgtB*.WT): calc.: 51131.48 Da, meas.: 51130.22 & 51307.36 Da (+177 corresponds to α -N-6-Phosphogluconoylation of his tag).¹⁵²

ϵ_{280} (*CthEgtB*.WT): 115280 M⁻¹ cm⁻¹

pET19.*CthEgtB*.WT for crystallization

For crystallization of the partially closed structure we cloned the *CthEgtB* gene into a modified vector, pET19m, to encode an EgtB fusion construct with an N-terminal His₆-tag followed by a TEV (tobacco etch virus) protease cleavage site.

```
CthEgtBs: 5' - acatcttcaactcttttttgaagcggttggt -3'
CthEgtBa: 5' - accaaccgcttcaaaaaagagttgaagatgt -3'
```

m/z (pET19.*CthEgtB*.WT_{cleaved,his}): calc.: 49293.54 Da meas.: 49293.95 Da

ϵ_{280} (pET19.*CthEgtB*.WT_{cleaved,his}): 115280 M⁻¹ cm⁻¹

CthEgtB Cysteine-Free (CF) and Loop Mutants

The cysteine free and loop mutants were purchased from General Biosystems in a pET28a plasmid. The sequences are as follows:

Sequence of pET28.*CthEgtB*.G98A:

```
MGSSHHHHHHSSGLVPRGSHMEAAARSHPEPIQSGEVSDRKAWQRHYRAVRAVSEAICQPLETEDYVVQMPDVSPP
KWHLGHTSWFFETFILKSLADYRPFHPRYDYIFNSYYEAVGARHPRPQRGLLTRPTVSEVYAYRAHVDAVERF
IAHSDTRTWAAALQPILELGLHHEQQHQELLLTDIKAILATNPLDPVYRPQPQPLPSPVEQLSPTGDWHIVEGGRY
AIGHAGRGFAFDNEGPRHDVLLRPCRIAAARPVNTNGEFLAFMADGGYRRPELWLSDGWAAVTARGWEAPLYWRQAA
```

DGTWETLTLHGVPVAPYEPVCHISFYEADAYARWAGKRLPTEAEWEVVAARLPVTGNFYESGVLHPRPVSVA
 FYGDVWVWTASPYVGYPGFRPVS GALGEYNGKFMNCQMVLRGGSCATSLTHIRSTYRNFFPPDARWQFTGVRLAE
 DMS

m/z (CthEgtB.G98A): calc.: 51034.36 Da, meas.: 51035.24 Da

ϵ_{280} (CthEgtB.G98A): 116770 M⁻¹ cm⁻¹

Sequence of pET28.CthEgtB.CF:

MGHHHHHAENLYFQGHMEAAARSHPEPIQSGEVSDRKAWQRHYRAVRVAVSEIAIQPLETEDYVVQPMPDVSPPKW
 HLGHTSWFFETFILKSGLADYRPFHPRYDYIFNSYYEAVGARHPRPQRGLLTRPTVSEVYAYRAHVDAEVERFIA
 HSDTRTWAALQPILELGLHHEQQHQEELLTLDIKAILATNPLDPVYRPQPQLPSPVEQLSPTGDWHIVEGGRYAI
 GHAGRGFAFDNEGPRHDVLLRPFRIAARPVNNGEFLAFMADGGYRRPELWLSDGWAAVTARGWEAPLYWRQAADG
 TWETLTLHGVPVAPYEPVSHISFYEADAYARWAGKRLPTEAEWEVVAARLPVTGNFYESGVLHPRPVSVA
 AAFYGDVWVWTASPYVGYPGFRPVS GALGEYNGKFMNSQMVLRGSSAATSLTHIRSTYRNFFPPDARWQFTGVRLAEDM
 S

m/z (CthEgtB.CF): calc.: 50987.18 Da, meas.: 50988.51 Da

ϵ_{280} (CthEgtB.CF): 116770 M⁻¹ cm⁻¹

Sequence of pET28.CthEgtB.CF.G98A:

MGHHHHHAENLYFQGHMEAAARSHPEPIQSGEVSDRKAWQRHYRAVRVAVSEIAIQPLETEDYVVQPMPDVSPPKW
 HLGHTSWFFETFILKSGLADYRPFHPRYDYIFNSYYEAVAAARHPRPQRGLLTRPTVSEVYAYRAHVDAEVERFIA
 HSDTRTWAALQPILELGLHHEQQHQEELLTLDIKAILATNPLDPVYRPQPQLPSPVEQLSPTGDWHIVEGGRYAI
 GHAGRGFAFDNEGPRHDVLLRPFRIAARPVNNGEFLAFMADGGYRRPELWLSDGWAAVTARGWEAPLYWRQAADG
 TWETLTLHGVPVAPYEPVSHISFYEADAYARWAGKRLPTEAEWEVVAARLPVTGNFYESGVLHPRPVSVA
 AAFYGDVWVWTASPYVGYPGFRPVS GALGEYNGKFMNSQMVLRGSSAATSLTHIRSTYRNFFPPDARWQFTGVRLAEDM
 S

m/z (CthEgtB.CF.G98A): calc.: 51001.20Da, meas.: 50999.6361 Da

ϵ_{280} (CthEgtB.CF.G98A): 116770 M⁻¹ cm⁻¹

Sequence of pET28.CthEgtB.CF.G98V:

MGHHHHHAENLYFQGHMEAAARSHPEPIQSGEVSDRKAWQRHYRAVRVAVSEIAIQPLETEDYVVQPMPDVSPPKW
 HLGHTSWFFETFILKSGLADYRPFHPRYDYIFNSYYEAVVARHPRPQRGLLTRPTVSEVYAYRAHVDAEVERFIA
 HSDTRTWAALQPILELGLHHEQQHQEELLTLDIKAILATNPLDPVYRPQPQLPSPVEQLSPTGDWHIVEGGRYAI
 GHAGRGFAFDNEGPRHDVLLRPFRIAARPVNNGEFLAFMADGGYRRPELWLSDGWAAVTARGWEAPLYWRQAADG
 TWETLTLHGVPVAPYEPVSHISFYEADAYARWAGKRLPTEAEWEVVAARLPVTGNFYESGVLHPRPVSVA
 AAFYGDVWVWTASPYVGYPGFRPVS GALGEYNGKFMNSQMVLRGSSAATSLTHIRSTYRNFFPPDARWQFTGVRLAEDM
 S

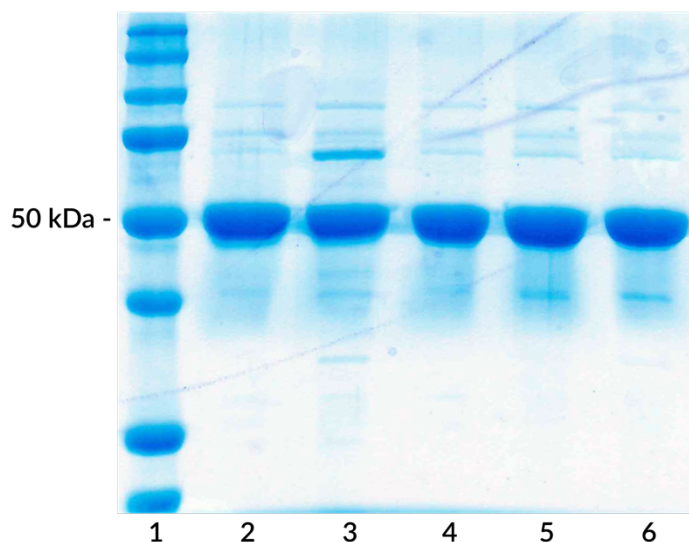
m/z (CthEgtB.CF.G98V): calc.: 51029.26 Da, meas.: 51028.18 Da

ϵ_{280} (CthEgtB.CF.G98V): 116770 M⁻¹ cm⁻¹

Recombinant Protein Purification

CaCl₂ competent BL21.pLysS (DE3) cells were transformed with either a pET28CthEgtB or pET19mCthEgtB plasmid following standard heat-shocking procedures. An overnight preculture of rich Luria Broth (LB) media (32 g tryptone, 20 g yeast extract and 5 g NaCl per L of media) containing the appropriate antibiotics (kanamycin (50 mg/L) and chloramphenicol (34 mg/L) for the pET28 construct, and ampicillin (100 mg/mL)

and chloramphenicol (34 mg/L for the pET19m construct) was inoculated with the transformed cells and incubated overnight at 37 °C whilst shaking at 180 rpm. 1 mL of pre-culture was used to inoculate growth cultures of rich LB media containing the appropriate antibiotics at concentrations previously stated. Cells were grown at 37 °C with shaking (180 rpm) until the optical density at 600 nm (OD₆₀₀) reached 0.8. Expression of the plasmid encoded gene was induced by the addition of IPTG to a final concentration of 0.1 mM. Protein expression was allowed to continue for 18 hours at 18 °C. Cells were harvested by centrifugation at 9000 g for 20 minutes at 4 °C, the supernatant was discarded and the cell pellet was frozen at -20 °C until required for purification. The cell pellet was thawed on ice and resuspended in Lysis buffer (10 % (v/v) glycerol, 200 mM NaCl, 20 mM Tris, pH 8.0 and 10 mM imidazole) and was lysed in an Emulsiflex-C3 (Avestin). Cellular debris was removed by centrifugation at 25000 g for 40 minutes, at 4 °C. The supernatant was incubated with Ni-NTA Agarose slurry at 4°C for 15 minutes. The agarose beads were washed with washing buffer 1 (10 % (v/v) glycerol, 200 mM NaCl, 20 mM Tris, pH 8.0 and 20 mM imidazole). *CthEgtB* was eluted and collected in fractions by washing with elution buffer (200 mM NaCl, 50 mM Tris, pH 8.0, 250 mM imidazole). Protein concentration was determined for each collected fraction with a nano-drop 2000/2000c spectrophotometer. Fractions identified as containing *CthEgtB* by concentration and SDS PAGE Gel were pooled and incubated with 10 mM EDTA and 2 mM DTT to remove any metals. For kinetics and ITC, the protein was dialyzed overnight into dialysis buffer at least twice (50 mM Tris, pH 8, 50 mM NaCl), concentrated in an Amicon centrifugal filter device (cut-off 10 kDa) and then frozen in aliquots. For analysis by gel filtration, protein was further purified by Size Exclusion Chromatography (SEC) using a Superdex 200 pg 26/600 column with 200 mM NaCl, 50 mM Tris, pH 8 as the running buffer before freezing. For crystallography, protein was dialyzed against 50 mM NaCl, 50 mM Tris, pH 8, and 1 mM DTT with TEV protease at a final protein ratio of 12:1 for 16 hours at 4°C. After dialysis, the sample was run over a Ni-NTA column to remove the TEV protease and His tag. The flow through was purified by SEC using a Superdex 200 pg 26/600 column with 200 mM NaCl, 50 mM Tris, pH 8 as the running buffer. The *CthEgtB* peak fractions were collected and concentrated for immediate use in crystallization trials.

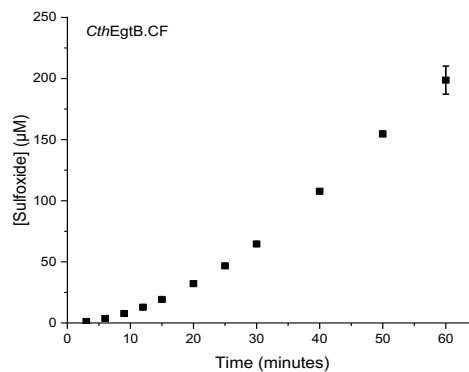
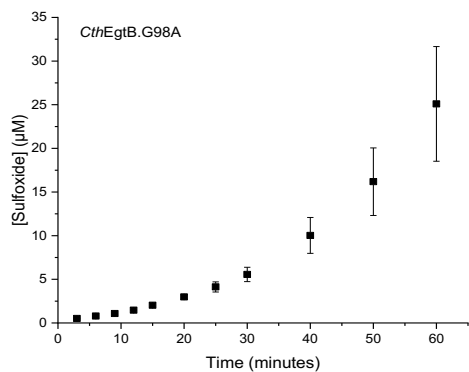


Supplementary Figure 9. SDS PAGE Gel of *CthEgtB* variants, 5 μ g is loaded in each lane. 1. Molecular Weight Marker 2. WT 3. G98A 4. CF 5. CF/G98A 6. CF/G98V

Enzyme Kinetics

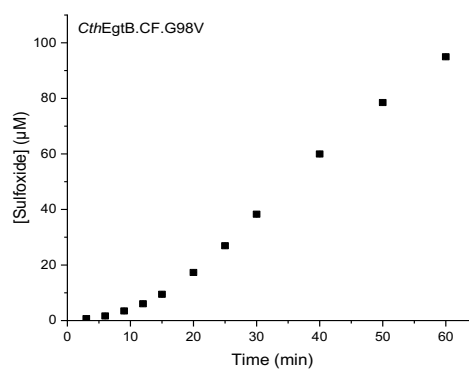
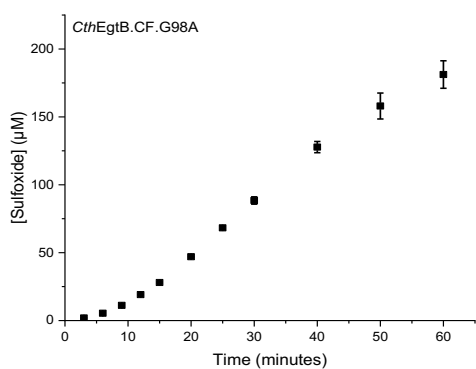
Sulfoxide synthase activities of *CthEgtB* variants were measured in reactions containing 100 mM phosphate buffer, pH 8, 100 mM NaCl, 2 mM TCEP, 2 mM ascorbate and FeSO_4 (4 equiv. to protein concentration), TMH and cysteine. Reactions were initiated by addition of enzyme and were incubated at 26°C. Aliquots of the reactions were quenched by addition of phosphoric acid. Reaction products were quantified by cation exchange HPLC using 20 mM phosphoric acid at pH 2 with a NaCl gradient as a mobile phase.² Chromatograms were recorded at 265 nm. Kinetic parameters for the WT and S92 mutants represent the averages of at least three independently-determined rates of sulfoxide production, which were fitted to the function $v = V_{\text{max}}[s]/(K_M + [s])$. The Michaelis-Menten parameters k_{cat} and k_{cat}/K_M were determined in the presence of co-substrate at a concentration at least 3 times higher than the corresponding K_M and in air-saturated buffers. For k_{cat} measurements of the loop and tetramer mutants, rates of sulfoxide production were determined in the presence of 1 mM and 3 mM of both substrates. All values are the averages of three independent measurements. The longer time scale plots used to visualize the lag phase were measured with 0.5 μ M enzyme, in the presence of 1 mM TMH and Cysteine and represents the average of two individual measurements.

The Role of Oligomerization and Loop Folding in a type II EgtB



7A. *CthEgtB*_{G98A}

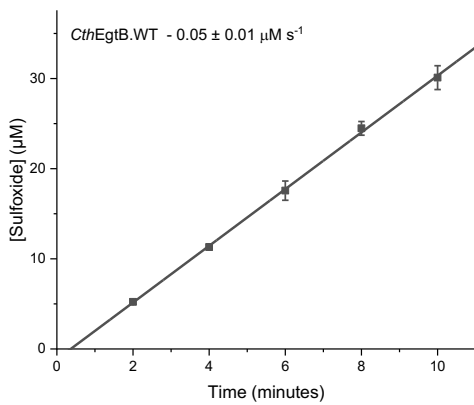
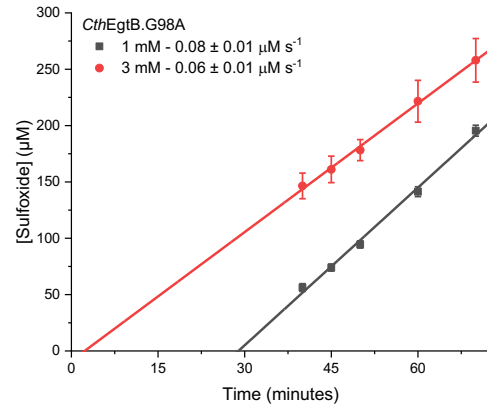
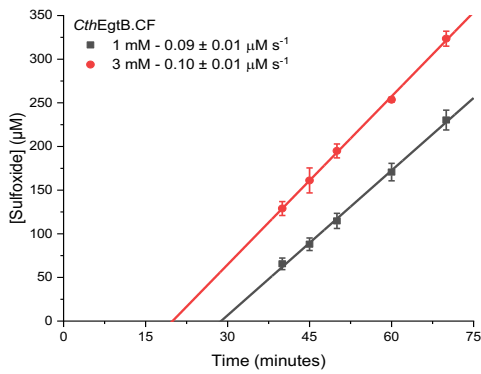
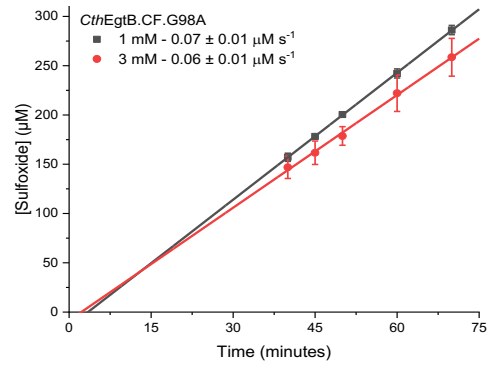
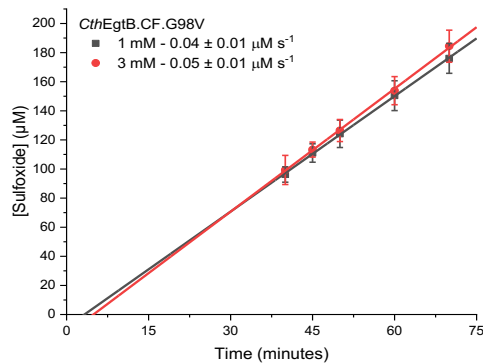
7B. *CthEgtB*_{CF}



7C. *CthEgtB*_{CF/G98A}

7D. *CthEgtB*_{CF/G98V}

Supplementary Figure 10. Sulfoxide Production as a function of time of four *CthEgtB* loop mutants. Reactions contained 1 mM TMH and Cysteine and 0.5 µM enzyme. Data shown is the average of two individual measurements.

k_{cat} measurements for *CthEgtB* WT and various loop mutantsFigure 10A. V_{max} of *CthEgtB*.WT with 1 mM of both substrates.Figure 10B. V_{max} of *CthEgtB*.G98AFigure 10C. V_{max} of *CthEgtB*.CFFigure 10D. V_{max} of *CthEgtB*.CF.G98AFigure 10E. V_{max} of *CthEgtB*.CF.G98V

Supplementary Figure 11. V_{max} Measurements of *CthEgtB*.WT and mutants. All measurements are triplicate. 0.5 μM of each enzyme was used in the presence of 2 μM FeSO_4 .

Characterization of reaction products by ^1H NMR

Reaction mixtures containing *CthEgtB* were analyzed by ^1H NMR to identify the formed products. The reactions contained 100 mM phosphate buffer - pH 8.0, 100 mM NaCl, 2 mM ascorbate, 2 mM TCEP, 4 μM FeSO_4 , 1 mM TMH, 0.5 mM L-cysteine and between 1 and 5 μM of *CthEgtB* variant to a final volume of 2 ml. These solutions were incubated overnight at room temperature. After lyophilization, the residue was dissolved in D_2O . ^1H NMR (500 MHz, D_2O) was measured with at least 128 scans and analyzed by MestReNova software. Substrates and products were identified based on α - and β -protons of the cysteine moiety and the aromatic protons of TMH. All measures were done in triplicate, from which the standard deviation was calculated.

DSF

DSF measurements were performed in a Prometheus NT.48 (Nanotemper) with a constant temperature gradient (20 to 95°C at 1°C/min). 4.9 μM of *CthEgtB.WT*, *CthEgtB.G98A*, *CthEgtB.CF*, *CthEgtB.CF.G98A* and *CthEgtB.CF.G98V* were incubated with 24.4 μM MnCl_2 in 10 mM TCEP, 100 mM NaCl and 100 mM Phosphate, pH 8. The TMH samples contained 10 mM TMH, the TMH + CYS samples contained 10 mM TMH and 10 mM CYS, while the control reactions lacked both substrates. All values represent the average of three measurements.

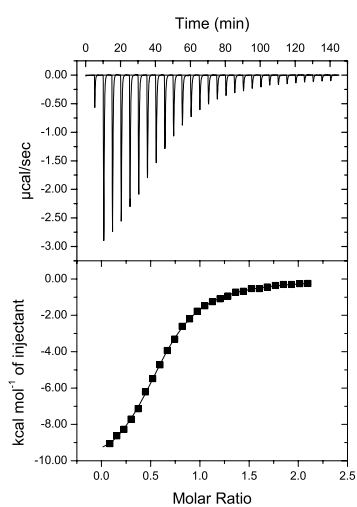
Isothermal titration calorimetry (ITC)

100 μ M of *CthEgtB* WT or variant was reconstituted with five equivalents of either FeSO_4 (and 10 equivalents of ascorbate) or MnCl_2 in a buffering solution of 50 mM NaCl and 50 mM Tris, pH 8, and was transferred into the sample cell (cell volume 1.4204 mL) of an isothermal titration calorimeter (VP200-ITC system, MicroCal LLC). A 1 mM solution of TMH was added to the protein solution by syringe in 1 x 2 μ L injection followed by 10 μ L injections every five minutes. All ITC measurements were conducted at 26°C. Data were plotted as the power needed to maintain the reference and sample cell at the same temperature against time and as kcal/mol of injectant against the molar ratio of ligand and protein. The program Origin7 (OriginLab Corporation) was used to analyze the data. The data from titration of TMH and corresponding calculated dissociation constants are shown on the following page.

The Role of Oligomerization and Loop Folding in a type II EgtB

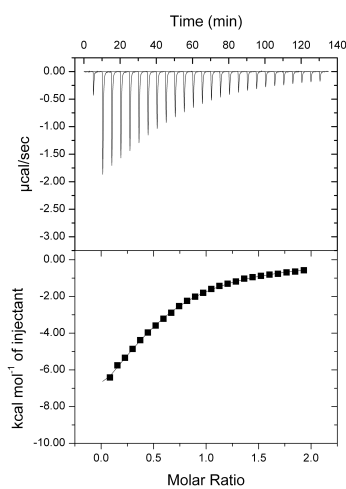
CthEgtB_{WT}/Fe + TMH

$K_D = 9.8 \pm 0.2 \mu\text{M}$



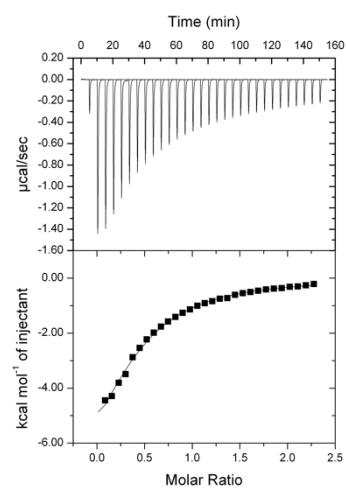
CthEgtB_{WT}/Mn + TMH

$K_D = 37 \pm 1 \mu\text{M}$



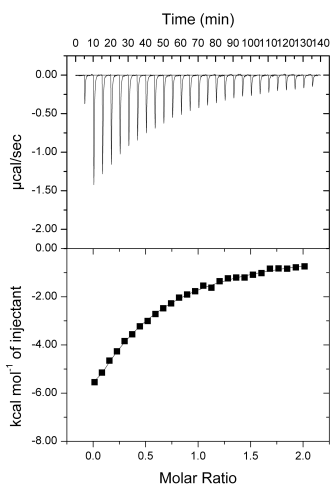
CthEgtB_{G98A}/Mn + TMH

$K_D = 38 \pm 3 \mu\text{M}$



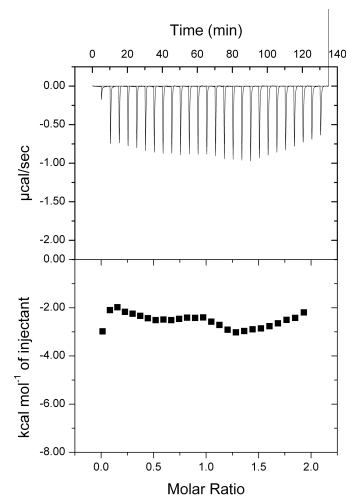
CthEgtB_{CF}/Mn + TMH

$K_D = 67 \pm 4 \mu\text{M}$



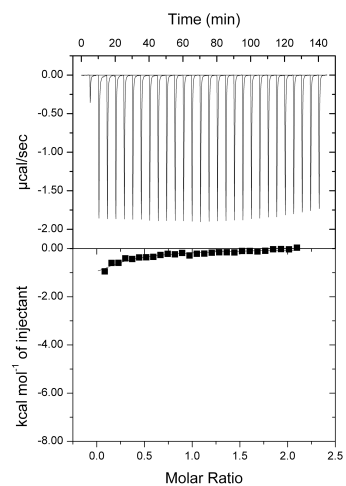
CthEgtB_{CF/G98A}/Mn + TMH

No significant binding



CthEgtB_{CF/G98V}/Mn + TMH

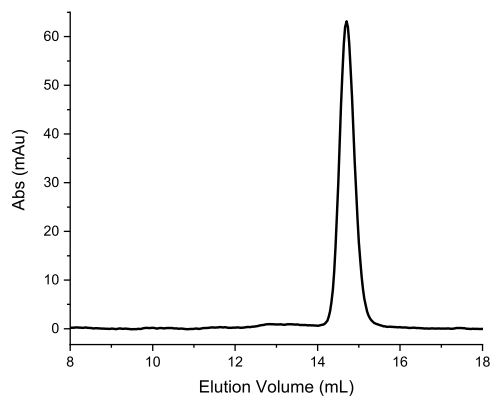
No significant binding



Supplementary Figure 12. ITC measurements for the titration of TMH into *CthEgtB_{WT}* or a variant.

Analytical Gel Filtration

All gel filtration was run on a Superdex 200 Increase 10/300 column at room temperature, using 200 mM NaCl, 50 mM Tris-HCl pH 8 as running buffer (flow rate: 0.5 mL/min). Samples were incubated overnight at the given concentrations to ensure concentration-dependent equilibrium was reached. Tested enzymes were reconstituted with either Mn or Fe and ascorbate.



Supplementary Figure 13. Analytical Gel Filtration of *MthEgtB* 100 μ L of 1 mg/mL (100 μ g total) of *MthEgtB*.WT was injected onto a Superdex 200 Increase 10/300 column at room temperature with 200 mM NaCl, 50 mM Tris-HCl pH 8 as running buffer (flow rate: 0.5 mL/min). Samples were pre-incubated with 5 eq. of FeSO_4 , 10 eq. of ascorbic acid and 2 mM DTT.

Size Exclusion Chromatography (SEC)

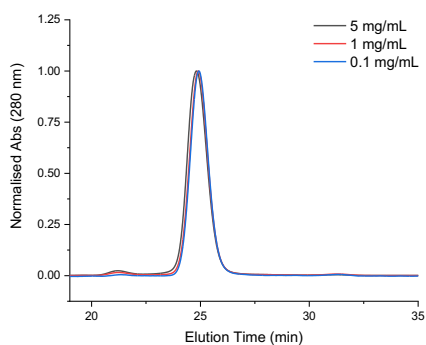


Figure 14A. SEC of *CthEgtB_{WT}* at 5, 1 and 0.1 mg/mL.

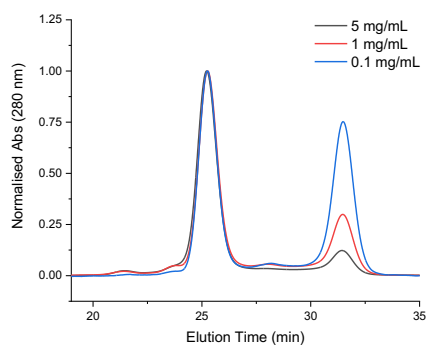


Figure 14B. SEC of *CthEgtB_{CF}* at 5, 1 and 0.1 mg/mL.

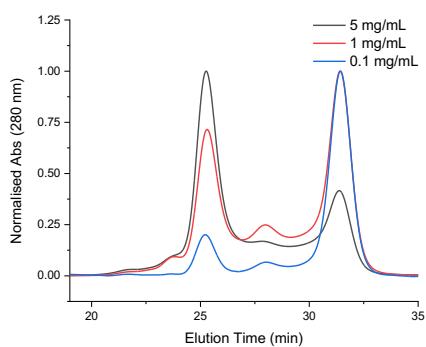


Figure 14C. SEC of *CthEgtB_{CF/G98A}* at 5, 1 and 0.1 mg/mL.

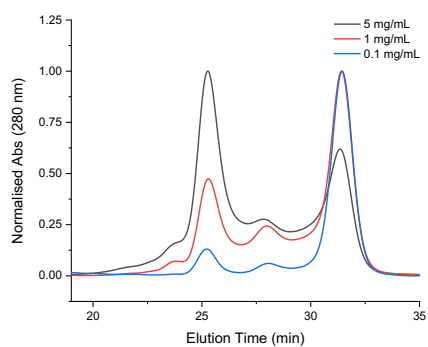


Figure 14D. SEC of *CthEgtB_{CF/G98V}* at 5, 1 and 0.1 mg/mL.

Supplementary Figure 14. SEC plotted by enzyme (globally normalized).

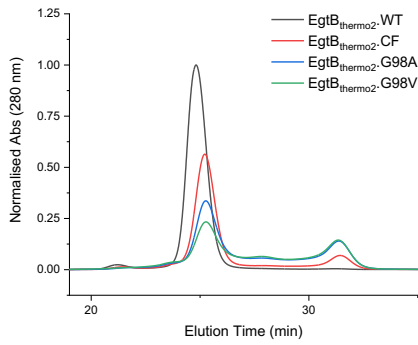


Figure 15A. SEC of *CthEgtB* WT, CF, CF/G98A & CF/G98V at 100 μ M (5 mg/mL).

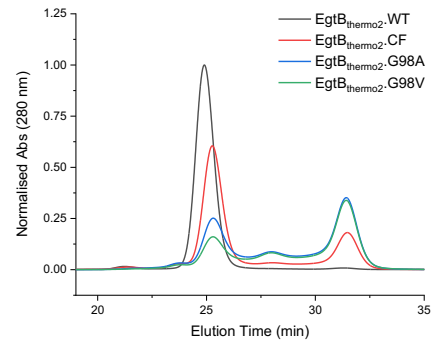


Figure 15B. SEC of *CthEgtB* WT, CF, CF/G98A & CF/G98V at 20 μ M (1 mg/mL).

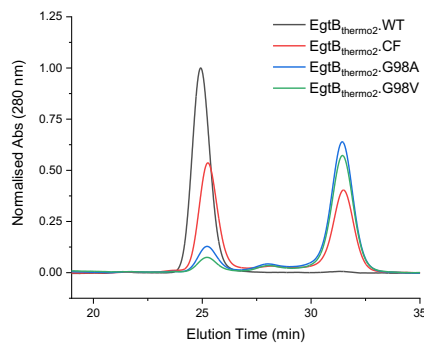


Figure 15C. SEC of *CthEgtB* WT, CF, CF/G98A & CF/G98V at 2 μ M (0.1 mg/mL).

Supplementary Figure 15. SEC plotted by concentration (Individually normalized)

Crystallization

Crystallization of the *CthEgtB* Structure 3

Structure 2 was obtained through the soaking of native crystal structures. Initial crystallization conditions for the crystallization of *CthEgtB* were determined with the vapor diffusion method in a sitting drop 96-well format. Drops were set up using a dispensing robot (Crystal Gryphon, Art Robbins) mixing 0.2 μL of EgtB solution (6.5 mg mL^{-1}) with 0.2 μL of reservoir solution, equilibrated against 30 μL reservoir solution. The screens were stored at 30 °C. Several initial hits were identified and optimized in 24-well plates using sitting drop format. Crystallization conditions that led to the native crystals used for soaking contained 23 % PEG 3350, 0.25 M MgAc mixed in a 0.5 : 0.5 μL ratio with *CthEgtB* at (6.5 mg mL^{-1}), equilibrated against 50 μL . Plates were incubated at 30°C with crystals appearing within a few days. The soaking solution was prepared by mixing reservoir mother liqueur with 10 mM TMH and 15 % (v/v) glycerol. Solution was added on top of grown crystals and they were left for 24 hours before flash freezing and storage in liquid nitrogen.

Crystallization of the *CthEgtB* Structure 4

Structure 4 was obtained through co-crystallization of His₆-tagged *CthEgtB*. crystallization conditions of *CthEgtB* with TMH were determined with the vapor diffusion method in a sitting drop 96-well format using the commercially available screen, Morpheus HT. *CthEgtB* was mixed and incubated with TMH (10 mM) for 1 hour preceding crystallization. Drops were set up using a dispensing robot (Crystal Gryphon, Art Robbins) mixing different ratios of EgtB-TMH solution (6.5 mg mL^{-1}) with reservoir solution (0.2 μL :0.1 μL , 0.1 μL :0.2 μL and 0.2 μL :0.2 μL) and were equilibrated against 30 μL reservoir solution. The screens were stored at 30 °C and crystals appeared within several days. Several initial hits were identified, flash cooled in liquid nitrogen and a synchrotron was taken for data collection. A crystal from E12 of Morpheus 0.12 M Ethylene Glycols (0.3M Diethylene glycol; 0.3M Triethylene glycol; 0.3M Tetraethylene glycol; 0.3M Pentaethylene glycol), 0.1 Buffer system 3 (Tris (base); BICINE), pH 8.5, 37.4 % precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350) mixed in a 1:2 ratio of protein to reservoir solution diffracted well.

Data collection, data processing, structure solution and refinement

Data of structure 3 crystal was collected at the Xo6SA (PXI) beamline using a Pilatus 6M-F detector, while data for the structure 4 crystal were collected at beamline Xo6DA (PXIII) using a Pilatus 2M-F detector, both at the Swiss Light Source (SLS), Villigen, Switzerland. The collected diffraction data were indexed and integrated using XDS¹⁵⁷ and were scaled using aimless¹⁵⁸. Both structures were solved by molecular replacement, structure 3 used the native structure (PDB: 6QKI)¹⁷³ as search model, while structure 4 used the closed TMH structure (PDB: 6QKJ)¹⁷³ with the loop regions removed as search model.¹⁵⁹ Several rounds of iterative model building and refinement were performed using Coot¹⁶¹ and Refmac¹⁶² or PHENIX¹⁶³. 5% of the data were excluded from refinement and used for cross-validation. Data collections and refinement statistics are summarized in Supplementary Tables 1 and 2 respectively. Stereochemical validation of the final models was performed using MolProbity.¹⁶⁴ Interfaces of proteins were analysed by PISA¹⁶⁵ and Figures were prepared PyMOL Molecular Graphics System, Version 1.7 (Schrödinger).

Supplementary Table 1. Data Collection Statistics

	Structure 3	Structure 4
X-ray Source	XO6SA (PXI)	Xo6DA (PXIII)
X-Ray detector	PILATUS 6MF	PILATUS 2MF
Wavelength (Å)	0.99	1.00
Space group	P 1 2 1 1	C 1 2 1
Cell dimensions a, b, c (Å)	85.1, 127.1, 88.9	84.6, 117.0, 199.3
Cell Angles α , β , γ (°)	90, 113.2, 90	90, 93.8, 90
Solvent content (%)	45	49
Molecules in asymmetric unit	4	4
Resolution limits (Å)	49.34-2.82 (2.921-2.82)	48.96-2.4 (2.486-2.4)
$R_{\text{merge}}^{\dagger}$	0.6372 (0.9168)	0.1541 (1.116)
$R_{\text{meas}}^{\ddagger}$	0.6564 (0.9666)	0.1669 (1.206)
CC $\frac{1}{2}$	0.875 (0.663)	0.996 (0.615)
$\langle I/\sigma(I) \rangle$	23.27 (3.43)	12.54 (1.81)
Total reflections	698050 (29240)	518160 (53270)
Unique reflections	40032 (2618)	75693 (7580)
Completeness	17.4 (9.6)	6.8 (7.0)
Multiplicity	93.16 (62.90)	99.95 (99.99)
Mosaicity	0.33	0.23

$\dagger R_{\text{merge}} = \sum hkl \sum i |I_i(hkl) - \overline{I(hkl)}| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection and $\overline{I(hkl)}$ is the average intensity obtained from multiple observations of symmetry-related reflections.

$\ddagger R_{\text{meas}} = \sum hkl [N/(N-1)]^{1/2} \sum i |I_i(hkl) - \overline{I(hkl)}| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection, $\overline{I(hkl)}$ is the average intensity obtained from multiple observations of symmetry-related reflections and N is the number of observations of intensity $I(hkl)$.

Supplementary Table 1. Refinement Statistics

	<i>CthEgtB</i> Open Loop 2 Structure	<i>CthEgtB</i> Open Loop 1 Structure
Resolution limits (Å)	49.34-2.82 (2.921-2.82)	48.96-2.4 (2.486-2.4)
Rwork *	0.1971 (0.2631)	0.1861 (0.2468)
Rfree **	0.2439 (0.3039)	0.2258 (0.3006)
Number of non-H atoms	12936	13596
macromolecules	12856	13068
ligands	60	60
solvent	20	468
Protein residues	1592	1632
Clashscore ***	11.65	10.44
R.m.s.d from ideal		
Bond lengths (Å)	0.009	0.008
Bond angles (u)	1.54	1.21
Ramachandran favored *** (%)	93.65	95.42
Ramachandran outliers *** (%)	1.35	1.49
Average B values (Å ²)	48.74	41.15
macromolecules	48.82	41.26
ligands	37.51	38.53
solvent	32.71	38.52

Numbers in parentheses refer to the outer shell.

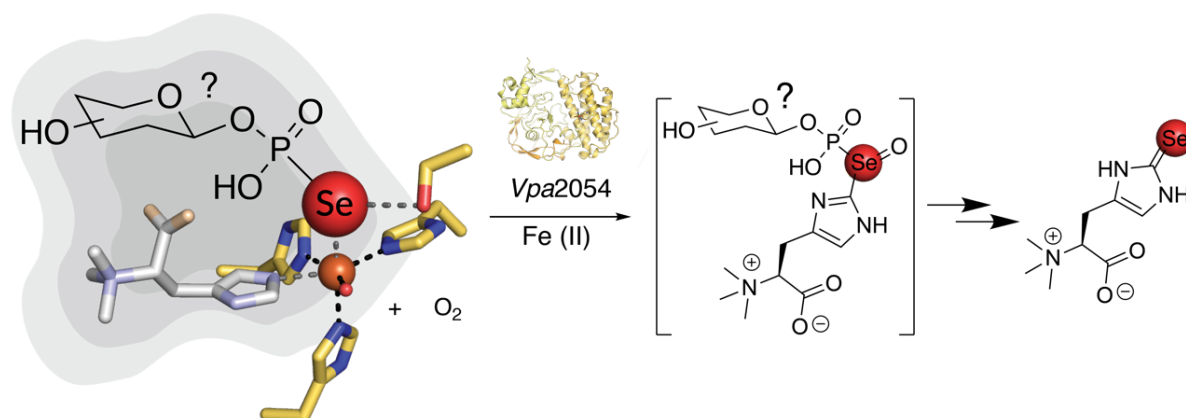
* $R_{work} = \frac{\sum |hkl| |F_{obs}| - |F_{calc}|}{\sum |hkl| |F_{obs}|}$

** Rfree is the R value calculated for 5% of the data set that was not included in the refinement.

*** Molprobability.

4. The Structure of a Sulfoxide Synthase Homologue implicated in Carbon-Selenium Bond Formation reveals insights into Selenium Activation

Anja R. Stampfli, Sebastian Flückiger and Florian P. Seebeck.



Selenium-containing secondary metabolites and their biosynthetic enzymes are incredibly rare in nature. Consequently, the discovery of selenoneine, a selenium isolog of ergothioneine, was met with great curiosity, and in particular its biosynthetic origin. It has been suggested that ergothioneine could be produced via the ergothioneine biosynthetic pathway. However, for reasons discussed within, this proposition seems implausible. To provide an alternate approach for selenoneine production, we solved the structure of Vpa2054, a sulfoxide synthase from *Variovorax paradoxus*, which is implicated in oxidative C-Se bond formation. Conservation of Vpa2054 in a three-gene cluster is consistent with a pathway that utilises TMH and a selenophosphorylated sugar to form the C-Se bond of selenoneine. This hypothesis is supported by structural and biochemical characterization of two members of the three-operon gene cluster. We discuss features that may enable selenium-based activation of molecular oxygen at an iron center and highlights the amenability of the EgtB scaffold for evolutionary variance.

Introduction

Ergothioneine (EGT) is a sulfur-containing metabolite produced by many bacterial and fungal species. It is assimilated through diet into mammals, including humans, where cellular concentrations of up to 1 mM have been found.^{14,22} While the antioxidant properties have clearly been demonstrated *in vitro*, the *in vivo* function of EGT remains elusive.¹⁴ Following the first isolation in tuna fish, selenoneine, the selenium isolog of EGT, was found to accumulate in humans with seafood-rich diets, raising considerable interest in its physiological function and implications for human health.¹⁸⁵⁻¹⁸⁷ The *in vivo* function of selenoneine is also unclear. While sulfur and selenium atoms have a similar valency, electronegativity and size, the two elements display distinctly different reactivities.^{53,58} In brief, selenols are more acidic and reductive compared to thiols and will readily react with oxygen to auto-oxidise without the necessity of a transition metal catalyst.¹⁸⁸⁻¹⁸⁹ The recent total synthesis and characterization of selenoneine highlighted a number of functional differences between EGT and selenoneine, suggesting that the two isologs can fill distinct functional niches.¹⁹⁰ Examples of natural selenium-containing secondary metabolites are sparse. Practically all selenium metabolites are isologs of corresponding sulfur compounds and lack their own dedicated biosynthetic machinery.⁵³

The biosynthetic origin of selenoneine is unknown. It has been suggested that selenoneine could be produced *via* the ergothioneine biosynthetic pathway. In most bacteria, ergothioneine is produced in an oxygen-dependent pathway that utilises an iron-dependent sulfoxide synthase, EgtB, to form the key C-S bond of ergothioneine. EgtB utilises *N,N,N*- α -trimethylhistidine (TMH), *L*-cysteine (cys) (or a derivative) and oxygen as substrates in an oxidative C-S bond-forming reaction, forming a sulfoxide, which is subsequently converted to ergothioneine (Figure 1A).^{31, 173} Utilisation of selenocysteine as a building block by EgtB rather than cysteine would, in theory, produce selenoneine (Figure 1B). The detection of small amounts of selenoneine production by *Schizosaccharomyces pombe* grown in selenocysteine-enriched media supports this idea.^{40, 191} However, SeCys seems an unlikely biosynthetic building block, as cellular SeCys concentrations are typically exceedingly low in any organism, with no free pool existing unlike other amino acids.¹⁹² We are not aware of any biosynthetic pathway that utilizes free SeCys as a genuine substrate. Furthermore, the bacterial sulfoxide synthases are highly inefficient catalysts for oxidative C-Se bond formation. In contrast, the fungal sulfoxide synthases (type IV) do not distinguish between Cys and SeCys.⁴¹ These findings indicate that either 1. selenoneine production is not catalysed by the enzymes involved in ergothioneine biosynthesis or; 2. that selenoneine is produced utilising the fungal EgtB enzymes. However it is only a fungal, and not a bacterial metabolite.

Subtle active site adaptations can dramatically improve the ability of enzymes to make C-Se bonds. As the structure of the type III sulfoxide synthases are not known, the structural basis that facilitates oxidative C-Se bond formation in the fungal enzymes is unclear. An alternative approach to understanding the structural features which enable selenium derivatization is to study an enzyme which naturally accepts a selenium-containing substrate. We identified a distinct class of EgtB homologues that are encoded in conjugation with a selenophosphate synthetase. Selenophosphate synthetases catalyse the formation of selenophosphate from selenide, and are believed to be a signature of selenium utilization in biology, indicating these enzymes

may produce selenium-containing metabolites.¹⁹³⁻¹⁹⁶ Therefore, we characterised *Vpa2054*, from *Variovorax paradoxus*, to gain insight into the likely substrates of this enzyme, and to determine active site features that might facilitate C-Se bond formation at an iron center.

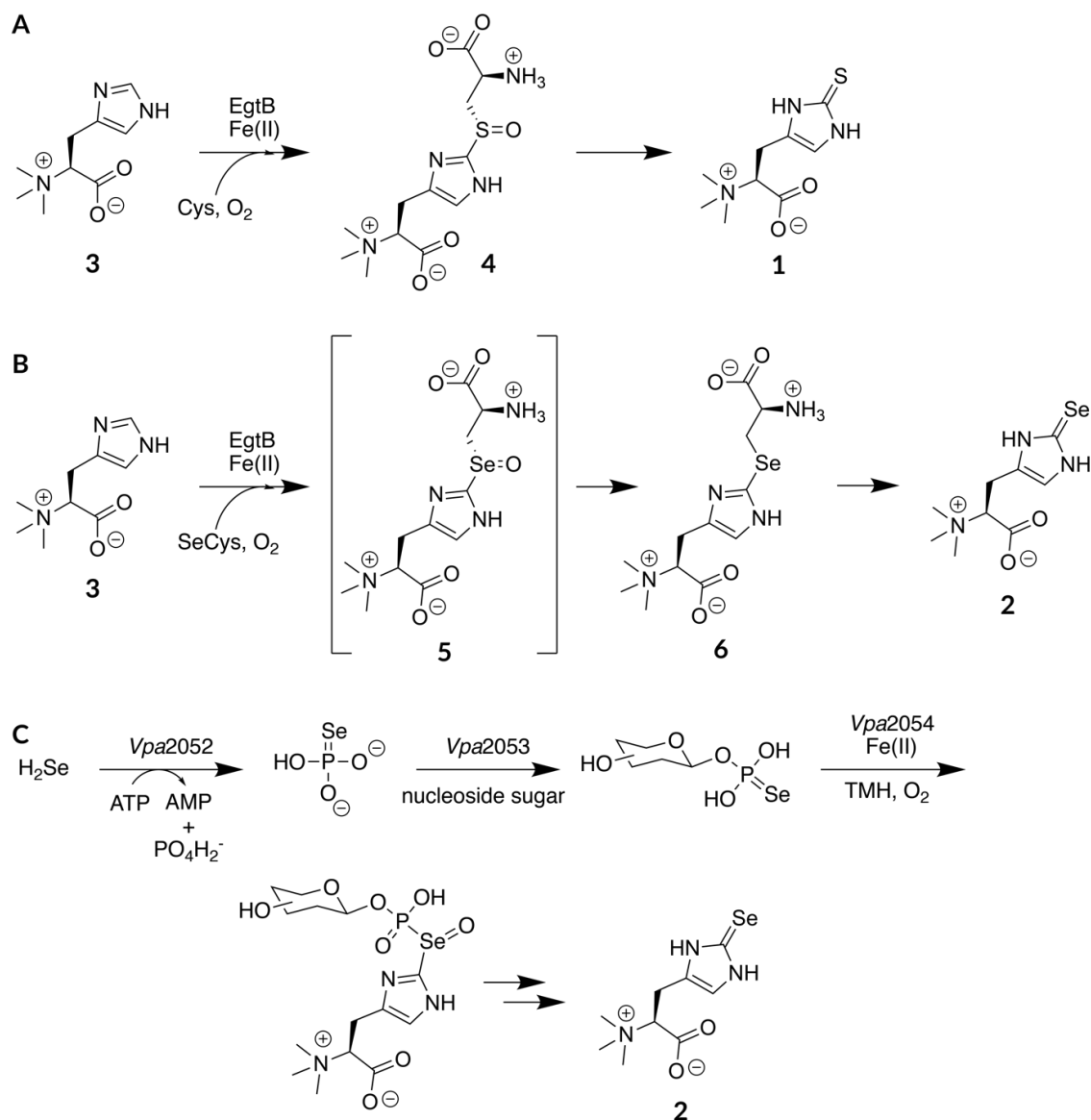


Figure 1 A. Many bacteria and fungi utilise an iron dependent sulfoxide synthase, EgtB, to form the key C-S bond of Ergothioneine (**1**). TMH (**3**), cysteine and molecular oxygen are converted to sulfoxide (**4**) which is converted to ergothioneine by the PLP-dependent-lyase, EgtE. B. Use of a SeCys building block could produce selenoxide (**5**), which is rapidly reduced to (**6**), and decomposes to selenoneine (**2**). C. Hypothesis for an alternative route to selenoneine biosynthesis in *Variovorax paradoxus* by a three-gene operon encoding for a selenophosphate synthetase, a glycosyl transferase, and an EgtB homologue to form the C-Se bond of selenoneine.

Results

A conserved 3-gene operon. Exploration of the EgtB bioinformatic space identified a sub-cluster of type-1 enzymes, primarily from β -proteobacteria, with a unique genomic environment to characterized EgtB proteins (Figure 2 A&B) Enzymes of this cluster are consistently co-encoded with a selenophosphate synthetase and a glycosyl transferase (GT) (Figure 2 D&E) and lack residues associated with cysteine and γ -GC binding. We speculated that this subclass might catalyse C-Se bond formation providing a basis for selenoneine biosynthesis in bacteria.

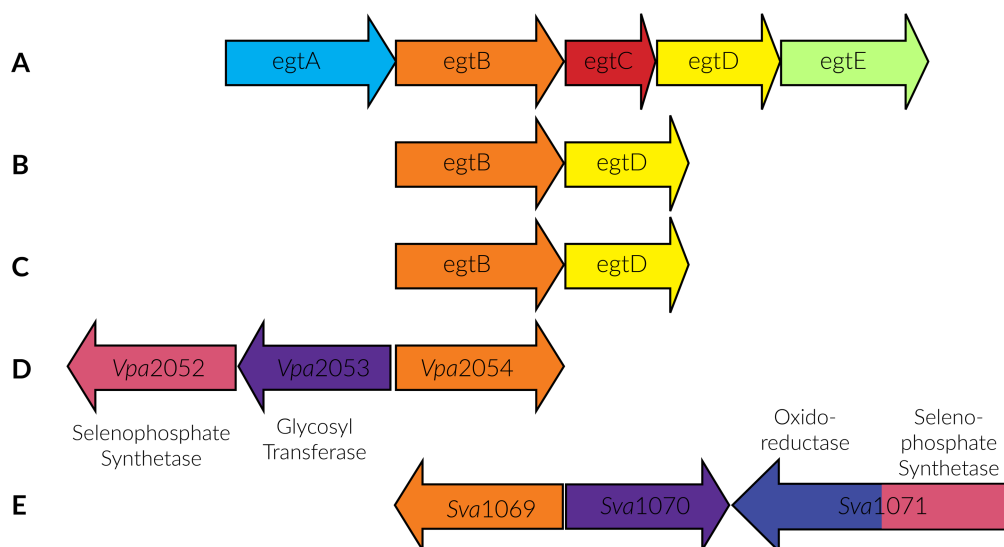


Figure 2 Genomic environment for the sulfoxide synthases from A. *Mycolicibacterium thermoresistibile*, B. *Chloracidobacterium thermophilum*, and C & D. *Variovorax paradoxus* and E. *Sulfurifustis variabilis*.

Crystal Structure Determination of Vpa2054. The crystal structure of Vpa2054 was determined by X-ray crystallography. Vpa2054 was crystallized as the native protein in complex with iron(II). The crystal diffracted to a resolution of 2.7 Å and belonged to space group P2₁ with two monomers in the asymmetric unit. For data collection and refinement statistics, see supplementary Table 1. The crystal structure was solved by molecular replacement using the native structure of *MthEgtB* (PDB: 4X8E) as a search model. *MthEgtB* and Vpa2054 share 31% sequence identity. The electron density revealed a continuous polypeptide chain from residues 6 to 412 with the exception of a small interdomain segment (187-196) (Figure 3A).

Vpa2054 Overall Structure. The asymmetric unit is comprised of two monomers and 45 % solvent content. The two monomers pack into long fibril units, indicating that the packing observed is an artefact of crystal packing and has no biological relevance. The overall Vpa2054 structure closely resembles that of characterized EgtBs, with a two domain fusion (Figure 3A). The N-terminal domain is comprised of a four helix bundle, most closely related to zinc-dependent thiol-S transferases (DinB_2 domain).¹²² The α -helices are arranged in an up, down, down, up conformation. An extensive span of random coil with several short helices (45 residues, α 1- α 2 loop) connects the second and third helices, folding close to the front of the active site.

An interdomain section connects the N-terminal DinB_2 domain to the C-terminal FGE-like domain. No electron density is visible for this interdomain section, likely owing to high flexibility in this region. This section is also disordered in the homologue, *CthEgtB*.¹⁷³ In the FGE-like domain, named after the copper-dependent formylglycine generating enzyme (FGE), an extended β -sheet wraps around a C-lectin type fold which is void of much secondary structure.¹⁹⁷ Instead, numerous ionic interactions appear to stabilize this fold. The active site is located within a 10 Å deep and 10 Å wide tunnel at the interface of the two domains. A single iron is bound by three histidines from the α -helical domain; His71 (Fe-N, 2.3 Å), His167 (Fe-N, 2.2 Å) and His171 (Fe-N, 2.2 Å) at the bottom of this tunnel in a facial coordination manner. The binding of two water molecules (both Fe-O, 2.5 Å) gives iron a square pyramidal geometry, leaving an open site at the axial position. While this site is typically filled by either a water/chloride ligand or thiol substrate in structures of homologues, a third water molecule is not observed in the axial position. This is however likely due to the low resolution of the structure. Tyr362 points into the active site from the C-terminal domain, with the phenolic group hydrogen bonding to one of the iron-coordinating water molecules (O-O, 2.5 Å). The hydroxyl group of Ser112 is also in close proximity to this water molecule, (O-O, 4.2 Å) and is within hydrogen bonding distance to the axial coordination site on iron (Figure 3B). Ser112 resides on the $\alpha 1$ - $\alpha 2$ loop close to the iron center and narrows the tunnel to the binding pocket at this point, largely closing off the active site to the solvated exterior (Supplementary Figure 1).

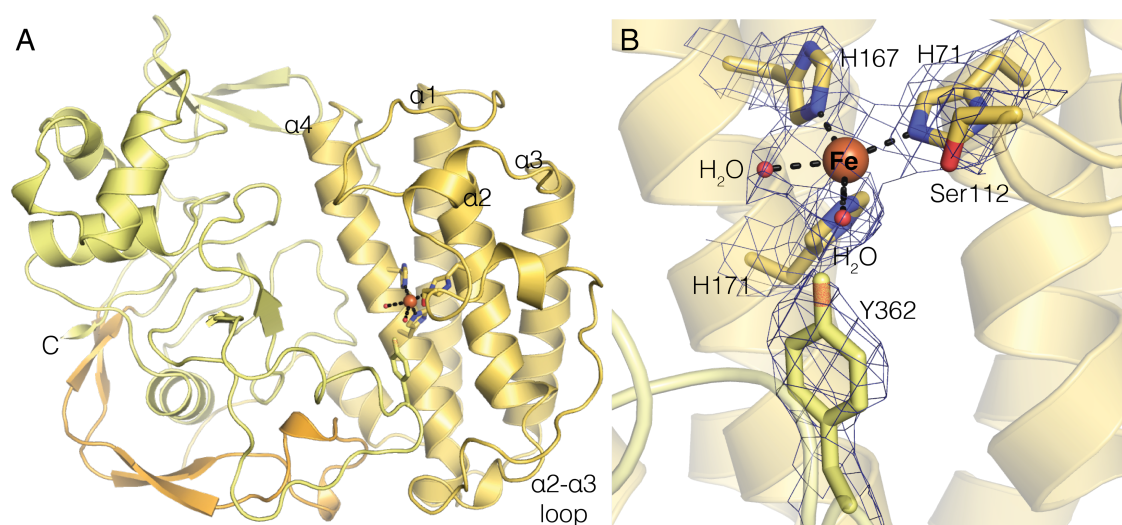


Figure 3. Structural Analysis of the *Vpa2054* native structure. A) Cartoon of a *Vpa2054* monomer. The N-terminal DinB_2-like domain (residues 6–183) is shown in yellow while the C-terminal FGE-like domain is shown in orange (residues 184–237) and light yellow (residues 237–412). Iron (brown), the metal coordinating histidine (yellow) and water molecules (red spheres) are shown to indicate the location of the active site B) Active site of *Vpa2054*, showing iron, histidine coordination sphere, two water molecules and residues in second coordination sphere, Tyr362 and Ser112. The $2m|F_o| - D|F_c|$ map is shown for iron, the three histidine ligands, two coordinating water molecules and Tyr362 and Ser112, contoured at σ -level = 1

TMH Binding Site. The *Vpa2054* crystal structure reveals that the binding pocket for the first substrate closely resembles the binding pockets of TMH in *CthEgtB* and *MthEgtB* (Figure 4) In the *Vpa2054* structure,

the side chains of Phe396, Phe397, Asn395 and backbone carbonyl of Phe396 create a negatively-polarized and aromatic box which is typical of quaternary ammonium binding proteins.¹⁹⁸ With this pocket as an anchoring point, we modelled TMH into the *Vpa2054* native structure (Figure 4A). The modelled TMH fits well into the active site, occupying an identical conformation as TMH in the *MthEgtB* and *CthEgtB* structures.^{49, 173} The described aromatic box packs around the *N*- α -trimethylamine moiety of TMH and the three N-methyl groups are in close proximity to Asn395 and the backbone carbonyl of Phe396 (Figure 4C). The Met170 side chain sits in a conformation in which it cannot directly interact with TMH, instead forming the base of the binding pocket underneath the substrate imidazole ring. N π of this ring could directly coordinate to iron. In both *MthEgtB* and *CthEgtB*, the N π hydrogen bonds to the backbone carbonyl of Tyr380_{Mth}/Tyr385_{Cth} via a bridging water molecule. In *Vpa2054*, Tyr366 occupies an identical conformation to Tyr380 in *MthEgtB*, and points into the binding pocket, indicating that a water-mediated interaction between the backbone carbonyl of Tyr366 and N π of TMH is also feasible in *Vpa2054*. The carboxylate of TMH points into the center of the active site, towards where the second substrate presumably binds. The observations from our model indicate that no conformational changes are required of the native structure for TMH binding. Overall, the active site appears rigid and closed, resembling the active site of the type I EgtBs.⁴⁹ In contrast, the native structure of a type II EgtB shows an open and exposed active site until substrate binding induces conformational changes in two loop regions to fold over and sequester the active site (Chapters two and three).

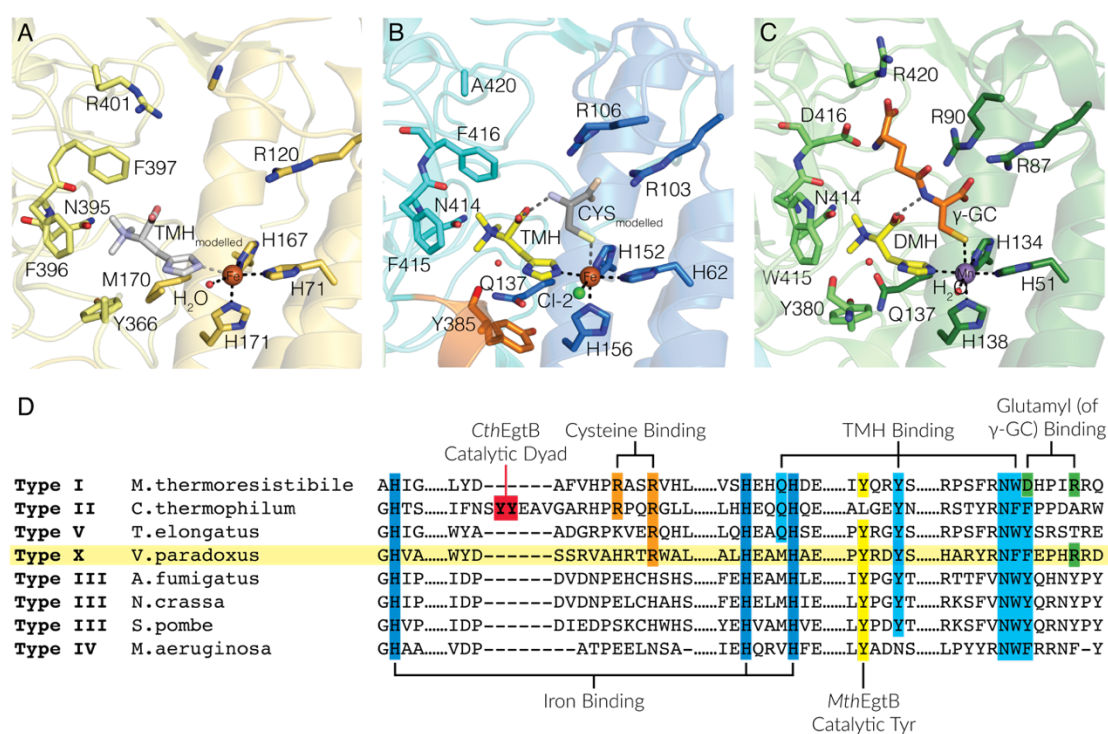


Figure 4. Substrate binding sites of A. *Vpa2054* native structure. B. *CthEgtB* TMH bound structure and C. *MthEgtB*. Mn.DMH, γ -GC complex. Key residues for binding and catalysis are highlighted and labeled. D) Sequence alignment of characterized EgtB homologues *M. thermoresistibile*⁴⁹⁻⁵⁰, *C. thermophilum*¹⁷³, *V. paradoxus* (highlighted), *M. aeruginosa*⁴⁷, *A. fumigatus*³⁸, *N. crassa*,³³ *S. pombe*¹⁹⁹ and uncharacterized *T. elongatus*¹⁷³. Residues highlighted in red indicate active site loop 1 of *CthEgtB*. Key residues for binding and catalysis are highlighted and labeled.

A polar substrate binding pocket. The *Vpa2045* structure itself in comparison to the *CthEgtB* and *MthEgtB* structures does not advocate for an obvious second substrate. On visual inspection, the *Vpa2045* binding pocket appears larger than that observed for cysteine binding in *CthEgtB*, yet smaller than that for γ -GC binding in *MthEgtB* (Supplementary Figure 2). The surface of the binding pocket appears to be very polar, with numerous hydrogen bonding donors and acceptors at the surface of the binding pocket. (Figure 5A&B). Arg401 and Arg374 (the latter is not shown) form a positively charged region (Figure 5B) above the TMH binding site. The oxygen atoms of the Tyr177 phenolic group and backbone carbonyls of two consecutive backbone carbonyls, Pro60 and Glu61, point into the active site forming the back and top of the binding pocket, while the side chains of Asn63, Glu68, Arg120 and Ser112 form the right side of the pocket and point inwards to the binding pocket. To gain insight into which residues lining the binding pocket might facilitate important interactions to the second substrate, we investigated the conservation of each of these residues. To do so, the 50 closest *Vpa2054* homologues were cumulated from a pBLAST search. Homologues that were not co-encoded with a selenophosphate transferase and a glycosyl transferase were removed to ensure that all sequences were *Vpa2054*-like. Analysis of the remaining 36 sequences for residue conservation revealed that only five of the nine residues that line the binding pocket are strictly conserved in all sequences analyzed: Asn63, Ser112, Arg120, Glu68 and Arg401 (Supplementary Figure 3). This observation indicates that these residues are likely important for substrate binding and should be of prominence in developing criteria for deciphering the second substrate.

Substrate Binding Assays. To test our hypothesis that TMH is the first substrate, we investigated the binding of potential substrates with a fluorescence-based thermal shift assay. This technique assesses protein stability through measurement of a melting curve, and can be used to determine if there are any changes in protein thermal stability upon the binding of a ligand.²⁰⁰ The stabilizing effect by a given ligand is proportional to its concentration and affinity to the protein and can therefore be used to screen for potential substrates.²⁰¹ In an initial assessment, we evaluated the stabilizing properties of various histidine derivatives. Incubation of *Vpa2054* with 10 mM TMH increased the melting point of *Vpa2054* by 9 °C in comparison to a control reaction containing only buffer. Incubation of *Vpa2054* with histidine or DMH did not result in an increase in melting temperature.⁵⁹ An identical degree of stabilization (+ 8.5 °C) was observed for *CthEgtB* with 10 mM TMH (Chapter 3). TMH binding was further quantified by the measurement of a K_D of $2.5 \pm \mu\text{M}$ using Isothermal titration calorimetry (ITC). This value is in the same range as values measured for *MthEgtB* (2 μM) and *CthEgtB* (10 μM) (Chapter 3). In a further screening, we assessed the binding of glucose- β -1-thiophosphate (G1-SP), as the three-gene operon is consistent with a pathway that utilizes a selenophosphate sugar. Glucose was chosen due to its abundance in nature, and thiophosphate as an analogue of selenophosphate, as the latter is very reactive and air sensitive.²⁰² Incubation of *Vpa2054* with TMH and G1-SP did not further increase the melting temperature, suggesting that G1-SP does not bind to *Vpa2054* and is therefore an unlikely substrate. However it is possible that following the binding of TMH, *Vpa2054* is energetically in the most stable conformation and that binding of the second ligand does not elicit a further response. Such an effect has been seen in *CthEgtB*, where TMH binding stabilizes *CthEgtB*, yet the

addition of cysteine or cysteine alone does not increase the melting temperature, despite being a native substrate (Chapter 3).

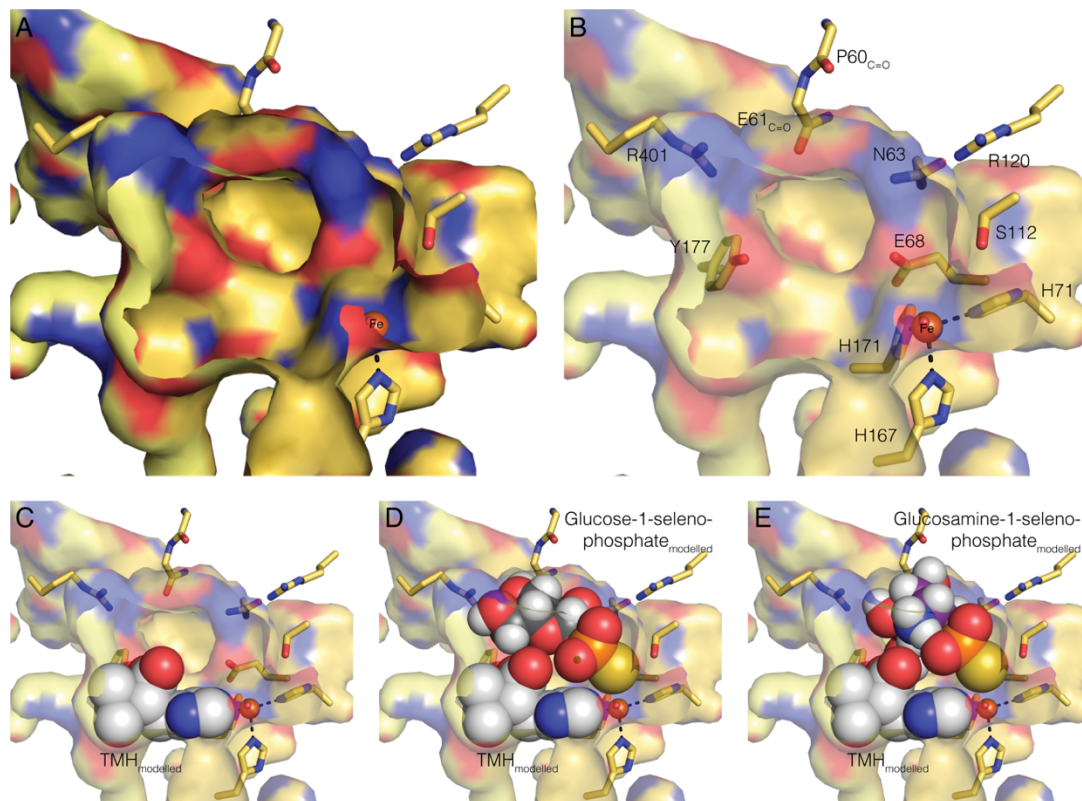


Figure 5. The *Vpa2054* substrate binding pocket shown in inverse surface mode and colored by atom. A. Binding pocket. B. Binding Pocket showing residues that comprise the binding pocket, colored by atom. C. Binding pocket with TMH modelled and shown as spheres. D. Binding pocket with TMH and Glucose-1-seleno-phosphate modelled and shown as spheres. E. Binding pocket with TMH (sphere) and Glucosamine-1-selenophosphate (spheres) modelled.

Vpa2053, a glycosyl transferase. To gain insight into the second substrate of *Vpa2054*, we examined the gene encoded beside *Vpa2054*, *Vpa2053*, annotated as a GT. GTs catalyse the transfer of a saccharide moiety from a glycosyl donor to an acceptor molecule.²⁰³⁻²⁰⁴ The activated sugar donor is often a nucleotide sugar, however lipid phosphor-sugar donors and sugar-1-phosphates are also widely used, while the possible acceptor substrates are incredibly diverse, including mono-, di-, or oligo- carbohydrates, proteins, lipids, DNA, and numerous other small molecules.²⁰³⁻²⁰⁴ This diversity of each of the three reaction components involved in glycosyl transfer reactions renders functional assignment incredibly difficult. Prediction of substrate specificity is notoriously challenging as prokaryotic GTs share low sequence identities, and no clear sequence motifs for substrate specificity have been established.²⁰⁵⁻²⁰⁶ A pBLAST search of *Vpa2053* did not return any characterized homologues with significant similarity. From sequential analysis, *Vpa2053* is classified as a clan II, inverting GT, yet this does not provide any aid in substrate identification. Further difficulties ensued upon attempts to produce *Vpa2053*, as the used construct was insoluble. We therefore focused our attention on the glycosyl transferase from *Sulfurifustis variabilis* (*SvaGT*), which shares 40 %

sequence identity with *Vpa2053* and is co-encoded with a *Vpa2054*-like EgtB and selenophosphate synthetase (Figure 2E). *SvaGT* was producible and UV-vis absorbance of the purified protein contained a strong absorption band at 260 nm, which is indicative of co-purification with a bound nucleoside. This result supports the genetic annotation of *Vpa2053* as a GT and indicates that the donor is likely to be a nucleotide sugar. The fluorescence-based thermal shift assay method was utilized to screen for the preferred nucleoside. In the presence of uridine diphosphate (UDP), *SvaGT* was stabilized by 9 °C, while adenosine diphosphate (ADP) and guanine diphosphate (GDP) increased melting temperatures by less than 4 °C, providing evidence that the substrate is comprised of a UDP component. Two commercially available UDP-sugars, UDP-glucose and UDP-galactose, were tested for stabilization of *Vpa2053*, increasing the melting temperatures of *SvaGT* by 8.4 °C and 6.4 °C, respectively. These results indicate that a monosaccharide-UDP is a likely substrate for *SvaGT*, with UDP-glucose being a viable candidate. However, no activity was detected upon use of *SvaGT* and UDP-glucose combined with thiophosphate as an acceptor molecule.⁵⁹

Discussion

Biosynthetic Hypothesis. Inspection of the *Variovorax paradoxus* genome revealed that two genes are annotated as sulfoxide synthases, indicating a divergent function for one of these genes. The genomic environment of the first gene is similar to that of type II EgtBs, co-encoded with EgtD, a methyl transferase (Figure 2B&C). The second annotated sulfoxide synthase gene (*Vpa2054*) sits within a conserved gene cluster consisting of a selenophosphate synthetase, and a glycosyl transferase (Figure 2D). The presence of selenophosphate synthetase indicates that the gene cluster may produce a selenium-containing metabolite. Several comparative genomics studies have suggested that the selenophosphate synthetase gene is an indicator of Se utilization in biology.¹⁹³⁻¹⁹⁵ The product, selenophosphate, is an essential selenium donor for the formation of SeCys and methylaminomethyl-2-selenouridine.¹⁹⁵ We hypothesise that selenophosphate is the glycosyl acceptor in the glycosyl transfer reaction catalysed by *Vpa2053*. This would produce a selenophosphate-sugar conjugate which could be utilised by *Vpa2054* to form a C-Se between the selenophosphate sugar and TMH, and oxidation of the conjugate to a selenoxide. Under reducing cellular conditions, the selenoxide would be reduced and would decompose to form selenoneine. The possibility that selenoneine might be biosynthesized *via* an alternative route is intriguing, as examples of natural selenium-containing secondary metabolites are sparse.

Vpa2054 Substrates. All EgtB homologues characterized thus far utilize TMH as the first substrate.^{33, 47, 49-50, 173, 199, 207} The modelling of TMH into the *Vpa2054* native structure and conservation of crucial residues in *MthEgtB* and *CthEgtB* demonstrate that TMH is a feasible substrate. In addition, the similarity of the measured TMH binding properties to those in reported EgtBs provides strong evidence that TMH is the first *Vpa2054* substrate. In contrast, the second substrate remains elusive. Variance in the donor substrate and modes of binding the donor substrate has been observed among the EgtB family.^{33, 47, 49-50, 173, 199, 207} Type I EgtBs utilize γ -GC, the substrate of EgtB from *T. elongatus* is unknown, and, while types II-IV utilize cysteine, the binding pocket is not conserved among the three types. (Stampfli & Seebeck, unpublished). The *Vpa2054* structure reveals that the native second substrate is polar, likely to be larger than cysteine, and contains numerous hydrogen bond donors and acceptors. Structural and bioinformatic analysis revealed that five residues, Asn63, Ser112, Arg120, Glu68 and Arg401, are likely important for substrate binding. Of these residues Arg401 and Arg120 are conserved in other homologues. Arg401 is conserved in the *MthEgtB* structure (Arg420_{Mth}), forming a salt bridge to the glutamyl moiety of γ -GC (Figure 4C&D). However, in contrast to Arg420_{Mth}, Arg374_{Vpa} folds down into the pocket in a different conformation from Arg420_{Mth}, creating a smaller binding pocket. Arg401 forms a salt bridge to Asp403, and is therefore unlikely to compensate for a charged group in the substrate. The other key functionality involved in γ -GC binding, D416_{Mth}, is not conserved, indicating that γ -GC or its selenium isolog are unlikely substrates. (Figure 4C&D). Arg120_{Vpa} is also conserved in both *MthEgtB* and *CthEgtB*, which in these homologues, is believed to bind the carboxylate group of cysteine and the cysteine moiety of γ -GC along with a second arginine in an RXXR motif. The second arginine in the RXXR motif is not conserved in *Vpa2054*, and is instead replaced by a histidine. While Arg120 occupies a similar conformation in *Vpa2054* as Arg103_{Mth} and Arg87_{Cth} in *MthEgtB* and

CthEgtB, respectively, the C α backbone occupies a different conformation, such that His123 does not structurally align with Arg106_{Mth} and Arg90_{Cth}. The proximity of Arg120 to the iron site and its conservation suggests that it may provide an important interaction, potentially to the selenophosphate group, which would coordinate to the axial iron site. The structure modelled with TMH indicates that this substrate primarily forms the base of the second substrate binding pocket, suggestive of an ordered binding mechanism, as observed in *CthEgtB*.⁴¹ The carboxylate of TMH points into the second substrate binding pocket (Figure 5B). In *MthEgtB*, the carboxylate of DMH forms a hydrogen bond to the amide functionality of γ -GC (Figure 4C).⁴⁹ Studies with substrate analogues in *CthEgtB* show that a salt bridge interaction occurs between the TMH carboxylate and cysteine amine that is important for binding of the second substrate and catalysis.¹¹¹ As the TMH binding mode appears to be conserved in *Vpa2054*, it is likely the carboxylate requires either a hydrogen bond, or more likely, a salt bridge partner in the second substrate, to compensate for this charge.

Vpa2054 Modelling. The finding that the *Vpa2053* homologue, *SvaGT*, binds and stabilizes a UDP nucleoside, supports the annotation of this enzyme as a GT and identifies UDP-glucose as a potential substrate. UDP-glucose is one of the most abundant nucleoside sugars. The reaction of UDP-glucose with selenophosphate would produce 1- β -selenophosphoglucose, due to the classification of *Vpa2053* as an inverting GT. We also hypothesized that 1-selenophosphorylated glucosamine could also be a viable substrate, as the C2-amino group could compensate for and interact with the carboxylate of TMH. We wanted to determine if these propositions are feasible. Therefore, TMH was modelled into the *Vpa2054* active site with 1-selenophosphoglucose and 1-selenophosphoglucosamine. From this analysis we wanted to establish if 1) the binding pocket is large enough to accommodate a monosaccharide-selenophosphate conjugate and 2) if any favorable interactions are possible, in particular to the carboxylate of the modelled TMH and Arg120. Ligand restraints were generated for both isomers of 1-selenophosphoglucose and 1-selenophosphoglucosamine, and the molecules were approximately modelled into the active site, so that the selenium atom coordinates to the iron center (Figure 5 D&E). The approximations indicate that 1) the binding pocket is large enough to fit TMH and either of the selenophosphate sugars, both of which pack well into the active site and do not present any obvious clashes. A larger ligand would very likely result in clashes, and is therefore unlikely to be a substrate 2) The β conformation of 1-selenophosphorylated glucosamine appears more compatible with the binding pocket. And 3) As modelled, Arg120 is in close proximity to the selenophosphate group, as are the C2-amine of glucosamine and carboxylate of TMH, which could potentially provide favorable interactions. These impressions indicate that a selenophosphorylated sugar is a viable substrate and that the *Vpa2054* structure does not contradict the bioinformatic-based hypothesis. However, we were unable to observe any reactivity between G1-ST and TMH, possibly due to not having identified the correct sugar substrate, or the thiophosphate group being an inadequate selenophosphate analogue.

Oxygen Binding Site. Formation of a C-Se bond would also require molecular oxygen as an oxidant. In *CthEgtB* and *MthEgtB* oxygen is proposed to directly to iron, providing an octahedral binding geometry. Conservation of the oxygen pocket, indicates this is also likely for *Vpa2054*. Tyr362 hydrogen bonds to the water molecule at the proposed oxygen-binding site (O-O, 2.5 Å). In *EgtB* types I and III a tyrosine conserved in sequence (Tyr377 in *MthEgtB*) is a vital catalytic residue to facilitate the reduction of molecular oxygen to

initiate C-S bond formation.^{50, 72} In the *Vpa2054* structure, Tyr362 occupies an identical conformation to Tyr377 in *MthEgtB*. This conservation suggests a similar function and conservation of the oxygen binding site and mechanism of activation in *Vpa2054* (Figure 4D).

Comparison to other EgtB structures. The overall structure of *Vpa2054* is very similar to the *CthEgtB* and *MthEgtB* structures, with rmsd values of 1.4 Å (241 C α atoms) and 1.2 Å (248 C α atoms), respectively. (Supplementary Figure 4). The C-terminal domains are almost identical, as are the helical fragments of the N-terminal domain. The interhelical segments are however less conserved. In particular, the α 1- α 2 loop varies greatly between the three structures available, in the number of residues and conformation that this section takes up. In *CthEgtB*, this α 1- α 2 loop forms a flexible loop that contains catalytic residues that closes over the active site to form the oxygen binding site. This active site plasticity with respect to the oxygen binding is a hallmark of the EgtB family, largely stemming from the diversity in structure and sequence of this interhelical loop region. The deviances in this loop region have implications on reactivity, as the differences in active sites manifests themselves in the ability of different homologues to utilize SeCys as a substrate.⁴¹ However, due to the lack of secondary structural elements in the α 1- α 2 loop, predictions on the structural determinants that drive the observed reactivity differences are unsubstantiated. While the active sites of *CthEgtB* and *Vpa2054* look completely different, the α 1- α 2 loop also contributes to the *Vpa2054* active site, with Ser112 pointing in close proximity to iron. How this residue may facilitate selenium reactivity will be discussed.

Hydrogen Bonding supports oxidative C-Se bond formation. While enzymes involved in sulfur metabolism do not typically discriminate against selenium, another cysteine-utilising iron oxygenase, cysteine dioxygenase (CDO), does not tolerate sulfur-to-selenium substitutions and is unable to oxidise selenocysteine.⁵³⁻⁵⁶ This suggests that iron oxygenases are perhaps more sensitive to S-Se substitution due to their requirement for finely tuned electronics to mediate oxygen activation at an iron centre. In CDO, binding of the cysteine thiolate to iron is a pre-requisite for O₂ activation, and the exact positioning is important for determining the reaction outcome.²⁰⁸⁻²⁰⁹ While SeCys binds to CDO in an identical fashion as cysteine, CDO is unable to oxidize SeCys. Calculations show this lack of reactivity stems from the stronger electron density donor ability of selenium *versus* sulfur to iron, consequently lowering the reduction potential. The Brunold group stipulates this increased donor ability impairs oxygen binding⁵⁶, while calculations by the Che group indicate that oxygen can bind and be activated to form a selenium-based radical. Yet, the increased donor ability of selenium hinders the ability of iron to access the iron(III) oxidation state, halting catalysis.²⁰⁹ A qualitative analysis of *CthEgtB* indicates that increased donor ability of using SeCys as a substrate manifests itself in all three key catalytic steps. One approach heavily used in biology to modulate the reduction potential and electronic properties of a metal-thiolate system is the use of hydrogen bonding to a metal-bound thiolate.²¹⁰⁻²¹³ Removal of hydrogen-bonding interactions in a [2Fe-2S] cluster decreased the redox potential by up to 132 mV,²¹⁴ the reason being that hydrogen bonding to a metal thiolate will alter the covalency of the metal-ligand bond, decreasing the charge donation, leading to a more positive redox potential.⁶⁰⁻⁶¹ In *MthEgtB*, engineering of a hydrogen bond (Ala82Ser mutation) to metal thiolate disfavoured thiyl radical formation and provided evidence for proton-coupled electron transfer (PCET) in an

early step in the mechanistic proposal.⁶² Therefore hydrogen bonding of Ser112 to a selenium-containing substrate would reduce the electron density donated by selenium, counteracting the reduced reduction potential arising from the stronger electron density donor ability of selenium. This would increase the reduction potential of the metal-thiolate, perhaps enabling the active site to support selenium-mediated oxygen activation at an iron centre. If so, the *Vpa2054* structure provides another example from the EgtB family of how variance in the $\alpha 1$ - $\alpha 2$ loop leads to dramatic active site plasticity and, in this case, divergent reactivity.

Conclusion

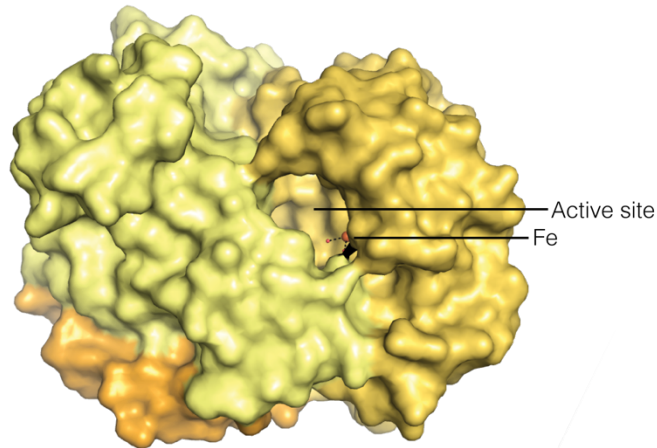
The *Vpa2054* structure presents another unique EgtB active site, highlighting the plasticity of the active site in this enzyme family and the amenability of the $\alpha 1$ - $\alpha 2$ loop for evolutionary variance. The crystal structure, together with the preliminary biochemical characterization, provides strong evidence that TMH is the first substrate. A selenophosphorylated glucoside as the second substrate is consistent with the conserved gene cluster, and preliminary characterization of a GT homologous to *Vpa2053*. The formation of a selenophosphorylated sugar would provide a selenium delivery system to form the C-Se bond of selenoneine. The structure additionally provides a platform to determine if substrate predictions are plausible and afforded residues that are likely to be important in substrate recognition and binding. This study narrows down the possibilities for proposals for the pathway encoded for by the three-operon gene cluster. If the second substrate does contain selenium, the finding that a dedicated enzyme exists for selenoneine would cement the importance of selenoneine as a biologically relevant and important selenium metabolite. In contrast, other selenium-containing metabolites, aside from SeCys, do not have dedicated pathways, with their production dependent on cellular selenium concentrations and utilizing the same catalytic machinery as their sulfur isologs. This finding would also represent the first iron-dependent enzyme with a native selenium-based substrate for the activation of molecule oxygen. We hypothesize that Ser112 might play an important role in hydrogen bonding to the selenium substrate, modulating the redox potential to enable selenium-based activation of molecular oxygen at an iron center in the EgtB scaffold. The electronic requirements of this would be an intriguing study that would provide an interesting parallel to the fungal EgtBs which do not distinguish SeCys from Cys.

Chapter 4:

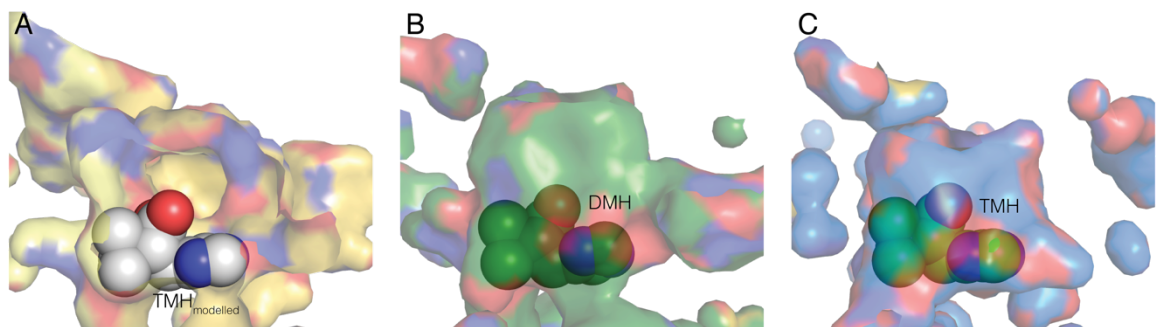
The Structure of a Sulfoxide Synthase Homologue implicated in Carbon-Selenium Bond Formation reveals insights into Selenium Activation

Supplementary Information

Supplementary Figures



Supplementary Figure 1. Surface View of *Vpa2054* native structure monomer. A) Cartoon of a *Vpa2054* monomer. The N-terminal DinB_2-like domain (residues 6–183) is shown in yellow while the C-terminal FGE-like domain is shown in orange (residues 184–237) and light yellow (residues 237–412). The active site is located between the two domains in a deep tunnel, iron (brown) and coordinating water molecules (red spheres) are shown.

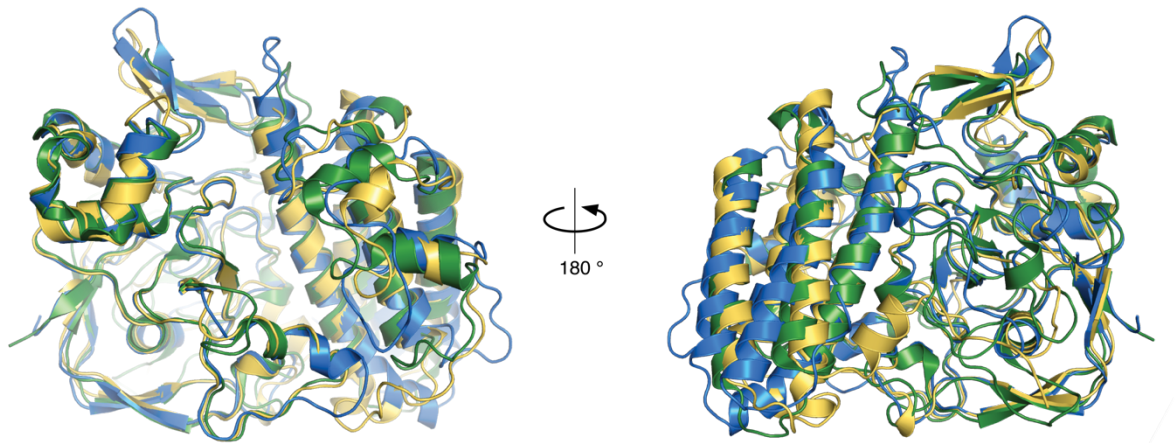


Supplementary Figure 2. Substrate binding pockets colored by atom. A: *Vpa2054*, TMH is modelled in and shown as spheres. B: *MthEgtB* (PDB: 4X8D) with DMH shown as spheres and C: *CthEgtB* (PDB: 6QKJ) with TMH shown as spheres.



Supplementary Figure 3. Sequence logo of 36 *Vpa2054*-like enzymes, all of which contain a selenophosphate transferase and a glycosyl transferase in their close genomic environment (± 10 genes).

The Structure of a Sulfoxide Synthase implicated in C-Se Bond Formation



Supplementary Figure 4. Structural comparison of the overall structures of *Vpa2054* (yellow), *CthEgtB* (blue) and *MthEgtB* (green). Superimposition shows a high structural similarity between the three structures in overall structure.

Experimental

Recombinant *Vpa2054* construct

The gene for EgtB from *Variovorax paradoxus* (*Vpa2054*, WP_012747184.1) was codon-optimized for protein production in *E. coli* and was purchased from Genscript in a pUC57 plasmid. The gene was ligated into the pET19 and pET28 expression vectors using the restriction enzymes NdeI/XhoI.

Sequence of pET28.*Vpa2054*:

(M) GSSHHHHHSSGLVPRGSHMDSTLPVYSVAGAPEALALRAGPPASVRAALLAARRRTLDDLADDFRAALGDAY
 PGIGYAPELNPPWLWELGHVAVWFQEWIWIGRNRQRARGVACEPDHAREPSLLPQADAWYDSSRVAHRTRWALPLPDA
 EATRGYLERTFAQTLALLDELPPDAHDDALYFFRLVALHEAMHAEAAAAYMAEGLGISLREGGPGPQLAEDAELEL
 PARRFRIGSEAGAGFAFDNELLPHDVAIGPLRIDAQAVSWARFLPFVEAGGYENPAWWS DAGRWLARQPLRQPA
 CLRAAGTGWQQQRGGRWLPLGPAQAAVHLNAHEAEAWCRWAGRRLPTEAEWECAALTLPGFAWGRVWEWTSSPFE
 PYPGFAPHPYRDYSAPWFGTRRVLRGACHATSAALAHARYRNFFEPHRRDIFAGFRSCRAAGG

m/z (*CthEgtB.WT*): calc.: 48096.0 Da, meas.: 48094.2 Da

$\epsilon_{280}(\text{CthEgtB.WT})$: 122380 M⁻¹ cm⁻¹

Sequence of pET19.*Vpa2054*:

MGHHHHHAENLYFQ/GHMDSTLPVYSVAGAPEALALRAGPPASVRAALLAARRRTLDDLADDFRAALGDAYPGIG
 YAPELNPPWLWELGHVAVWFQEWIWIGRNRQRARGVACEPDHAREPSLLPQADAWYDSSRVAHRTRWALPLPDAEATR
 GYLERTFAQTLALLDELPPDAHDDALYFFRLVALHEAMHAEAAAAYMAEGLGISLREGGPGPQLAEDAELEL
 PARRFRIGSEAGAGFAFDNELLPHDVAIGPLRIDAQAVSWARFLPFVEAGGYENPAWWS DAGRWLARQPLRQPAC
 LRAAGTGWQQQRGGRWLPLGPAQAAVHLNAHEAEAWCRWAGRRLPTEAEWECAALTLPGFAWGRVWEWTSSPFE
 PYPGFAPHPYRDYSAPWFGTRRVLRGACHATSAALAHARYRNFFEPHRRDIFAGFRSCRAAGG

m/z (*CthEgtB.WT*): calc.: 46258.1 Da, meas.: 46257.0 Da

$\epsilon_{280}(\text{CthEgtB.WT})$: 122380 M⁻¹ cm⁻¹

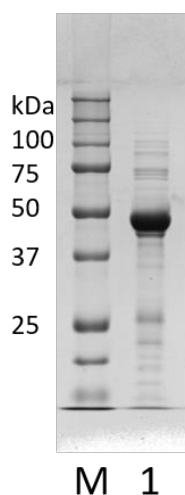
Recombinant Protein Purification

CaCl₂ competent BL21.pLysS (DE3) cells were transformed with a pET28*Vpa2054* or pET19*Vpa2054* plasmid following standard heat shocking procedures. An overnight preculture of Luria Broth (LB) media (10 g tryptone, 5 g yeast extract and 10 g NaCl per L of media) containing the appropriate antibiotics: kanamycin (50 mg/L) and chloramphenicol (34 mg/L) for the pET28 vector, ampicillin (100 mg/L) and chloramphenicol (34 mg/L) for pET19, were inoculated with the transformed cells and incubated overnight at 37 °C whilst shaking at 180 rpm. 1 mL of pre-culture was used to inoculate growth cultures of rich LB media containing the appropriate antibiotics at concentrations previously stated. Cells were grown at 37 °C with shaking (180 rpm) until the optical density at 600 nm (OD₆₀₀) reached 0.8. Cells were cooled down to 18 °C and expression of the plasmid encoded gene was induced by the addition of IPTG to a final concentration of 0.1 μM. Protein

The Structure of a Sulfoxide Synthase implicated in C-Se Bond Formation

expression was allowed to continue for 16 hours at 18 °C. Cells were harvested by centrifugation at 7000 rpm for 20 minutes at 4 °C, the supernatant was discarded and the cell pellet was frozen at - 20 °C until required for purification. The cell pellet was thawed on ice and resuspended in Lysis buffer (300 mM NaCl, 50 mM Na₂HPO₄, pH 8) and was lysed in an Emulsiflex-C3 (Avestin). Cellular debris was removed by centrifugation at 4000 g for one hour, at 4 °C. The supernatant was incubated with Ni- NTA Agarose slurry at 4°C for 20 minutes. The agarose beads were washed with washing buffer 1 (300 mM NaCl, 50 mM Na₂HPO₄, pH 8 and 20 mM imidazole). *Vpa2054* was eluted and collected in fractions by washing with elution buffer (300 mM NaCl, 50 mM Na₂HPO₄, pH 8, 250 mM imidazole). Protein concentration was determined for each collected fraction with a nano-drop 2000/2000c spectrophotometer. Fractions identified as containing *Vpa2054* by concentration and SDS PAGE Gel were pooled and incubated with 10 mM EDTA at 0°C to remove any metals. The pET28 construct used for binding studies and activity assays was dialyzed into dialysis buffer (50 mM Tris, pH 8, 50 mM NaCl), at least twice for a minimum of four hours, before flash freezing and storage for further use.

The pET19 construct used for crystallography was dialyzed against 50 mM NaCl, 50 mM Tris, pH 8, and 1 mM DTT with TEV protease at a final protein ratio of 12:1 for 16 hours at 4°C. After dialysis, the sample was run over a Ni-NTA column to remove the TEV protease and His tag. The flow through was purified by Size Exclusion Chromatography (SEC) using a Superdex 200 pg 26/600 column with 200 mM NaCl, 50 mM Tris, pH 8 as the running buffer. The *Vpa2054* peak fractions were collected and concentrated for immediate use in crystallization trials.



Supplementary Figure 5. SDS PAGE Gel containing molecular weight marker (Lane M) and recombinantly produced and purified *Vpa2054* (Lane 1). Adapted from ⁵⁹.

Biophysical Characterization

Please see the PhD thesis of Sebastian Flückiger 'Mechanistic Studies of Sulfur-Carbon Bond Formation by Metal-Dependent Enzymes' (2018)⁵⁹ for experimental details on the fluorescence-based thermal Shift Assay and Isothermal titration calorimetry (ITC) and activity assays.

Crystallization

Crystallization conditions for the crystallization of *Vpa2054* were determined with the vapor diffusion method in a sitting drop 96-well format. Drops were set up using a dispensing robot (Crystal Gryphon, Art Robbins) mixing different ratios of *Vpa2054* solution (31 mg mL⁻¹) with reservoir solution (0.2 μL:0.1 μL, 0.1 μL:0.2 μL and 0.2 μL:0.2 μL) and were equilibrated against 35 μL reservoir solution. The screens were stored at 20 °C. A crystal from E12 of Morpheus 0.12 M Ethylene Glycols (0.3M Diethylene glycol; 0.3M Triethylene glycol; 0.3M Tetraethylene glycol; 0.3M Pentaethylene glycol), 0.1 Buffer system 3 (Tris (base); BICINE), pH 8.5, 37.4 % precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350) mixed in a 1:2 ratio of protein to reservoir solution diffracted well.

Data collection, data processing, structure solution and refinement

Data of the *Vpa2054* crystals were collected at the Xo6SA (PXI) beamline using a EIGER 16M X Detector at the Swiss Light Source (SLS), Villigen, Switzerland. The collected diffraction data were indexed and integrated using XDS¹⁵⁷ and were scaled using aimless¹⁵⁸. The structure was solved by molecular replacement using the native structure of *MthEgtB* (PDB: 4X8B)⁴⁹ as the search model. The initial model of native was built using AUTOBUILD of the PHENIX package.¹⁶⁰ Several rounds of iterative model building and refinement were performed using Coot¹⁶¹ and Refmac¹⁶² or PHENIX¹⁶³. 5% of the data were excluded from refinement and used for cross-validation. Data collections and refinement statistics are summarized respectively. Stereochemical validation of the final models was performed using MolProbity.¹⁶⁴ Interfaces of proteins were analysed by PISA¹⁶⁵ and Figures were prepared in PyMOL Molecular Graphics System, Version 1.7 (Schrödinger).

Modelling

TMH (PDB ligand: AVJ) is modelled into the active site of *Vpa2054* through superimposition of the TMH-bound structures of *MthEgtB* (PDB:4X8E) onto *Vpa2054*. The ligand restraints for 1-β-selenophosphoglucose and both isomers of 1-selenophosphoglucosamine were generated using grade.²¹⁵ Coot was used to place the compounds into the *Vpa2054* structure to approximate if binding was plausible.¹⁶¹ Ligands were placed in such a way that (1) The selenium group co-ordinates directly to iron, in an analogous binding mode as γ-GC in the quaternary *MthEgtB* structure (PDB:4X8D) (2) Arg120 could co-ordinate to the phosphate group (3) the C2 amide of glucosamine can interact with the TMH carboxylate and (4) the whole ligand does not clash with any protein residues.

The structure of a sulfoxide synthase homologue implicated in C-Se bond formation

Supplementary Table 1. Data Collection and Refinement Statistics

Data Collection Statistics	Vpa2054	Refinement Statistics	Vpa2054
X-ray Source	X06SA (PXI)	Resolution limits (Å)	40.5-2.7 (2.8-2.7)
X-Ray detector	EIGER 16M (Dectris)	Rwork *	0.1979 (0.2740)
Wavelength (Å)	0.99	Rfree **	0.2532 (0.3568)
Space group	P 1 21 1	Number of non-H atoms	6298
Cell dimensions a, b, c (Å)	51.3, 120.9 63.7	macromolecules	6281
Cell Angles α , β , γ (°)	90, 91.6, 90	ligands	2
Solvent content (%)	38	solvent	15
Molecules in asymmetric unit	2	Protein residues	796
Resolution limits (Å)	40.5-2.7 (2.80-2.7)	Clashscore ***	5.64
$R_{\text{merge}}^{\dagger}$	0.06551 (0.2539)	R.m.s.d from ideal	
$R_{\text{meas}}^{\ddagger}$	0.09264 (0.2539)	Bond lengths (Å)	0.003
CC $\frac{1}{2}$	0.99 (0.861)	Bond angles (u)	0.68
$\langle I/\sigma(I) \rangle$	6.92 (2.47)	Ramachandran favoured *** (%)	97.08
Total reflections	39361 (3865)	Ramachandran outliers *** (%)	1.55
Unique reflections	20954 (2102)	Average B values (Å ²)	35.4
Completeness	1.9 (1.8)	macromolecules	35.43
Multiplicity	96.96 (98.96)	ligands	27.41
Mosaicity	0.17	solvent	24.57

$\dagger R_{\text{merge}} = \sum hkl \sum i |I_i(hkl) - \langle I(hkl) \rangle| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection and $\langle I(hkl) \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections.

$\ddagger R_{\text{meas}} = \sum hkl [N/(N-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl) \rangle| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection, $\langle I(hkl) \rangle$ is the average intensity obtained from multiple observations of symmetry-related reflections and N is the number of observations of intensity $I(hkl)$.

Numbers in parentheses refer to the outer shell.

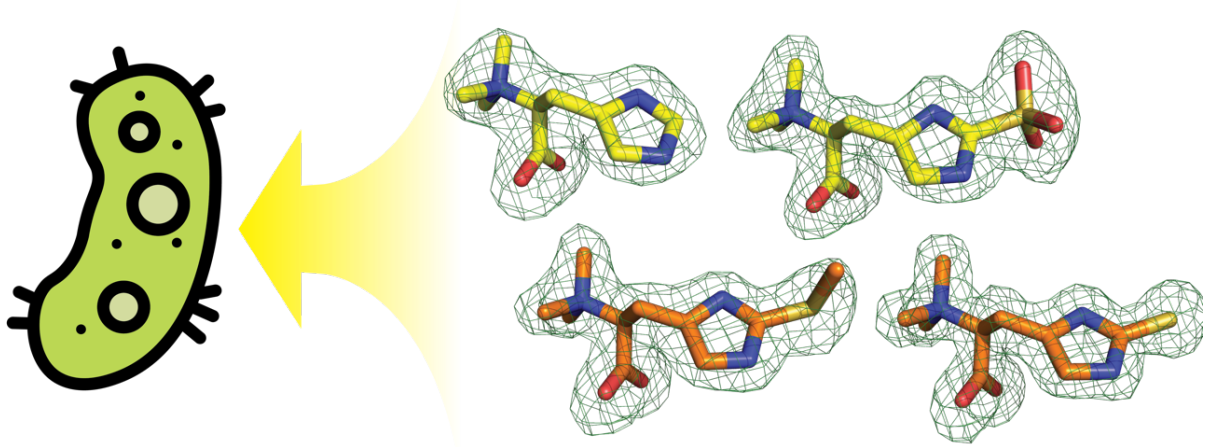
* $R_{\text{work}} = \sum hkl ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum hkl |F_{\text{obs}}|$

** Rfree is the R value calculated for 5% of the data set that was not included in the refinement.

*** Molprobit.

5. The Structural Characterization of Ergothioneine Solute Binding Proteins

This research will be form part of a publication, along with the work of Mariia Believa and Alice Maurer.



The molecular basis for ergothioneine biosynthesis is well established, as is the chemistry of a key EGT-degrading enzyme, ergothionase. A discrepancy was observed between the organisms that can produce ergothioneine and degrade ergothioneine. As most ergothionases are intracellular, a transporter system for EGT must exist. Herein we report the structural characterisation of two types of solute-binding proteins that bind EGT or ergothioneine sulfonic acid (EGTSO_3^-) with high affinities. Solute-binding proteins are a key component of the ABC transporter system which facilitates solute uptake into bacterial cells. The finding that a specific transport system exists for EGTSO_3^- , provides clear evidence that EGTSO_3^- is a relevant EGT degradation product in nature. Comparison of structures of the two SBP types identified residues that determine substrate selectivity. The application of these motifs to explore the sequence space of betaine solute-binding proteins revealed several key insights into EGT utilization and evolutionary history of the EGT related SBPs. This study presents a molecular handle and genomic precedent to identify organisms that utilize ergothioneine without biosynthesizing or degrading it, Until now, no approach was available to identify such organisms. This greatly increases the prevalence of known ergothioneine utilization, highlighting the ubiquity and importance of ergothioneine to many life forms.

Introduction

Ergothioneine and its derivatives. The biosynthesis of ergothioneine (EGT) and the enzymes involved are well-established, as is the transporter for its uptake into human cells.^{20, 31-32, 216} In contrast, very little is currently known about the degradative pathways of EGT and the biologically relevant forms and oxidation states that EGT can exist in. Recent work reported the characterization of “ergothionase”, an ammonia lyase that catalyzes the 1,2-elimination of trimethylamine from EGT to form thiouraconic acid (Figure 1). The oxidized analogue, ergothioneine-sulfonic acid (EGTSO₃⁻) was also found to be accepted as a substrate with a similar catalytic efficiency to EGT.²¹⁷ It is unknown if this oxidized metabolite is a biologically-relevant substrate or if the characterized ergothionase is merely promiscuous towards EGTSO₃⁻. Oxidation of EGT to the sulfonic acid can be mediated by reactive oxygen species. Stimulated human neutrophils were found to produce EGTSO₃⁻. However, the levels of hypochlorite produced through induced oxidative burst are artificially high, and it is therefore unknown if EGTSO₃⁻ is a relevant metabolite under normal physiological conditions.²¹⁸⁻²¹⁹ Reactive oxygen species can also lead to oxidative desulfurization, resulting in *N,N,N*- α -trimethyl histidine (TMH).⁷ Another proposed degradation pathway is methylation, to form *S*-methyl ergothioneine (MeEGT) (Figure 1). While this metabolite has been isolated, the enzyme responsible for methylation has not been identified.²²⁰

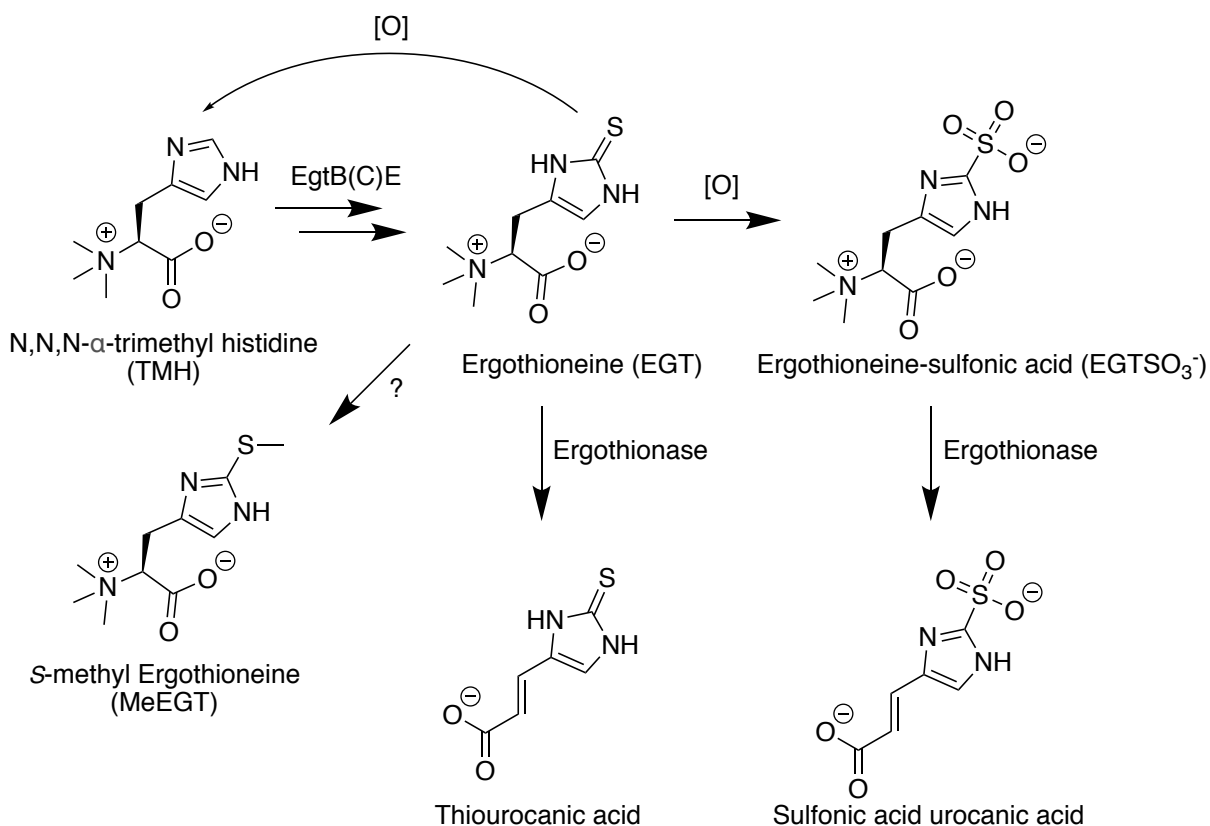


Figure 1. EGT metabolites. Oxidative degradation of EGT could produce *N*- α -trimethyl histidine and ergothioneine-sulfonic acid.²¹⁹ Methylation of EGT by an unknown methyl donor leads to *S*-methyl Ergothioneine. Elimination of the trimethylamine moiety of EGT and EGTSO₃⁻ by ergothionase leads to their corresponding thiouracanic acids.²¹⁷

Ergothionase Genomic Environment. Ergothionases are typically intracellular, yet many of the organisms with an ergothionase, do not biosynthesize EGT. Furthermore, EGT is not membrane permeable, which, in conjunction with the disparity between organisms that degrade and produce EGT, indicates that a system for EGT acquisition must exist.²⁰ To identify proteins that might be involved in EGT uptake, the genomic environment of genes encoding for ergothionase were analysed. This bioinformatic analysis revealed that genes annotated as components of the ATP-binding cassettes (ABC) transport system, including an ABC solute binding protein, were sometimes clustered with ergothionase genes, as exemplified in the *Citrobacter koseri* genome (Figure 2). ABC transporters transport compounds across the cellular membrane, being driven by ATP hydrolysis. These importers require a solute binding protein to bind and deliver the substrate to the periplasmic gate of the transporter. The close genomic environment of the SBP to ergothionases suggested that an ABC transportation system may exist for EGT and/or its degradation products.



Figure 2. Genomic environment of ergothionase in *Citrobacter koseri*. In numerous *Citrobacter* species the ergothionase gene is flanked by genes annotated as components of the ABC transport system, including both the transmembrane and ATP binding domains and a solute binding protein.

ATP-binding Cassette (ABC) Transport Systems. ATP-binding cassettes (ABC) transport systems are a large and important class of active transport system for the uptake of nutrients from the periplasm and delivery to the cytoplasm in gram-negative bacteria. The ABC transport system consists of five proteins organized into three main components (Figure 3): (i) a pair of integral membrane protein domains, the transmembrane domain (TMD). These components span the cytoplasmic membrane a minimum of six times each and form the channel through which the solute passes through. (ii) A pair of ATPase domains, which are associated with the cytoplasmic surface. This is the site of ATP hydrolysis, which is coupled to movement of the substrate through the channel. And finally, (iii) the solute binding protein (SBP) (also called periplasmic binding protein) provides the specificity for solute transfer by recognizing and binding the substrate and delivering it to membrane protein domain for translocation.²²¹⁻²²²

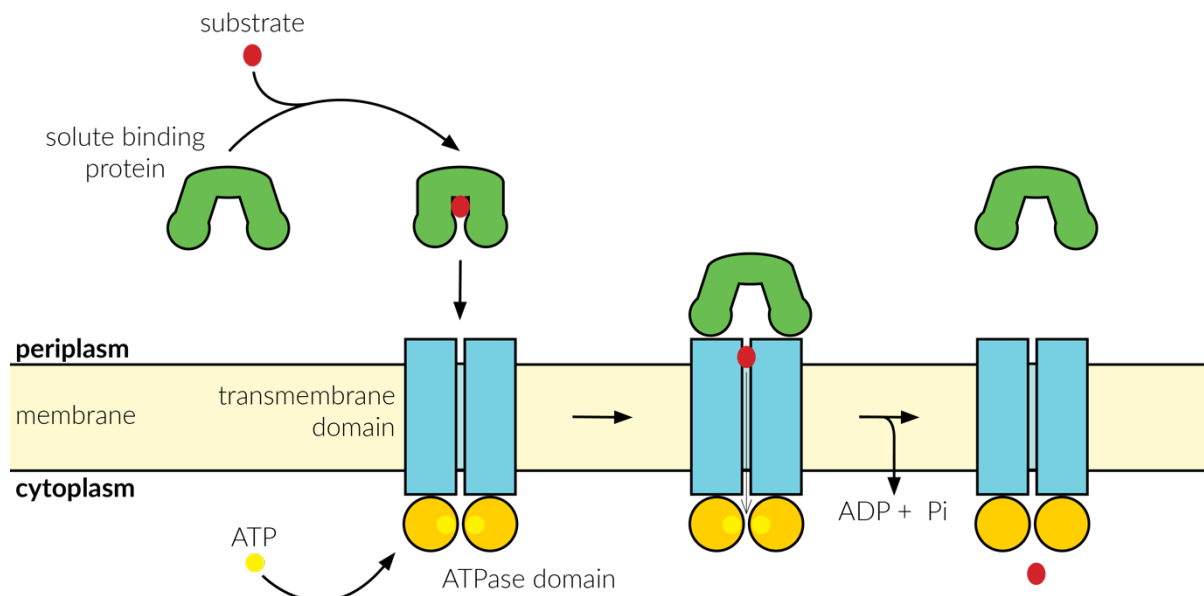


Figure 3. Simplified schematic for the mechanism and components of ABC importers. Three domains comprise the ABC transport system to transfer a solute from the periplasm to the cytoplasm using ATP.

Solute Binding Proteins. SBPs are absolutely essential for the function of the ABC transporters. They are abundant receptors, ensnaring their ligands with high binding affinities (K_D of 10-10000 nM) and large association rate constants ($1-10 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$). In the periplasm, SBP are typically found in concentrations of 0.1 - 1 mM, greatly surpassing the concentration of the membrane-associated components of the transporter system, and often the binding solute. These features enable rapid response to the presence of a ligand.²²¹

SBPs are monomers, which can range in size from 25000 to 60000 kD. They facilitate the transport of an extensive range of substrates of diverse sizes including amino acids, sugars, peptides, oxyanions, polyamines, vitamins, metal ions and their chelates. Their solubility and relatively high periplasmic concentrations lends itself to excellent overexpression and purification. Furthermore, SBP are very amenable to crystallization, in most cases diffracting to high resolutions. This amenability provides a plethora of structural data on the SBPs.²²¹

Solute Binding Protein Structure. Despite the huge diversity among the substrates and the often low-sequence homology between different SBPs, there is a common structural architecture to the solute binding proteins. This architecture consists of two similar globular domains. Each domain consists of a central beta-pleated sheet, which is flanked by alpha helices. The two domains are connected by either two or three (in rare cases, one) connecting segments, referred to as switch segments. This results in the two domains being comprised of non-contiguous regions of the polypeptide (Figure 4).²²¹

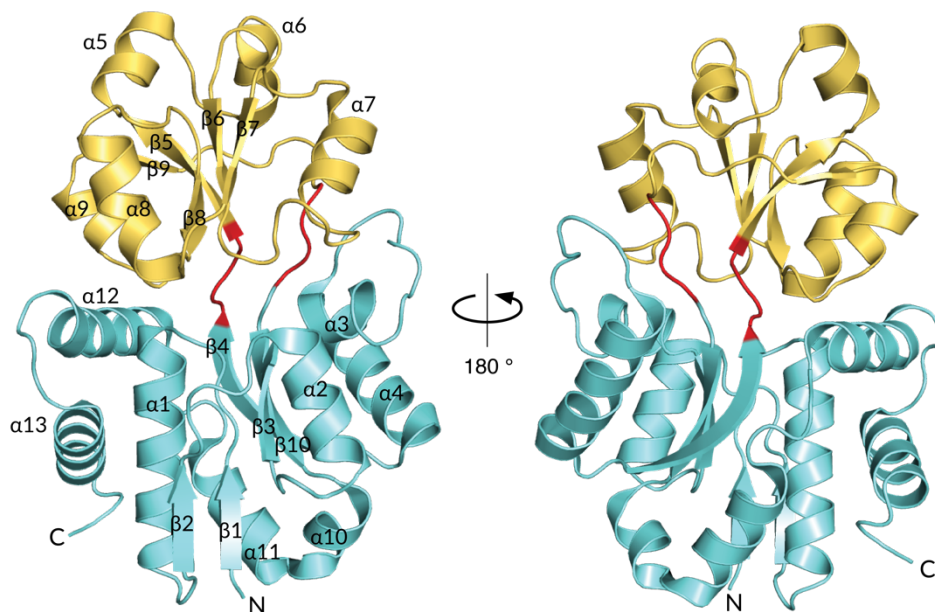


Figure 4. Structure of a type II Solute Binding Protein, ProX from *Archeoglobus fulgidus* (PDB: 1SW1, 1.9 Å). The two domains are in blue and yellow while the switch segment with two segments is colored in red. The N & C terminals are labelled and secondary structures labelled.

The protein substrate-binding site is located in the cleft between the two domains. Binding is associated with a large movement of the domains towards one another, closing in around the substrate so that the substrate is typically completely sequestered from solvent. This is referred to as the closed conformation. The structures of the domains themselves do not significantly change when moving from an open to closed conformation. Instead, their relative angles to one another differ, with the interdomain switch segments acting as a hinge (Figure 5). In solution, the unliganded open form is in equilibrium with unliganded closed form. The ligand binds to an open form of protein, interacting with just one of the domains. It is common that one domain will provide a larger substrate-binding surface area than the other domain. When the two domains do come together, the ligand is buried within the protein, interacting with both domains and shifting the equilibrium towards the closed conformation. This complete enclosure of the substrate gives rise to a multitude of protein-ligand interactions, accounting for the remarkable selectivity and high affinities.

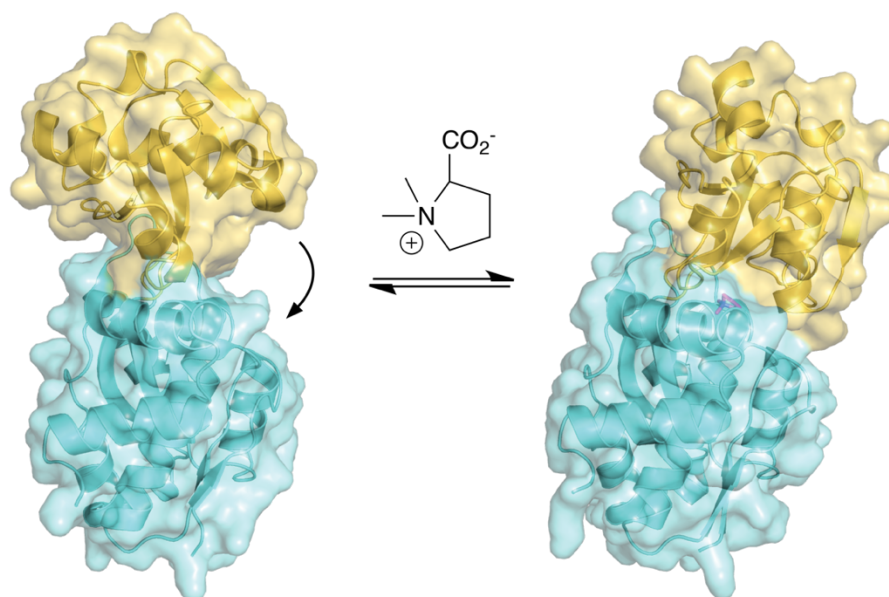


Figure 5. Structures of ProX from *Archeoglobus fulgidus* in an open, unliganded form (PDB: 1SW5, 1.8 Å) and in a closed form in complex with glycine betaine (PDB: 1SW5, 1.9 Å). Both the cartoon and surface representations are shown for the two colored domains, while proline betaine is shown in a stick representation in pink.

Betaine Binding Solute Binding Proteins. Several betaine solute binding proteins have been functionally and structurally characterized. One family of SBPs, called ProX, binds glycine betaine (GB) and proline betaine (PB) with high affinity. Two variants: a bacterial and archaeal system, have been well characterized. The bacterial species, *Bacillus subtilis*, has six transport systems, which are used in response to stress, three of which, OpuA-C, are of the ABC superfamily. The solute binding domains of each are labelled with a second C. OpuAC and OpuCC both bind GB and PB, while OpuBC binds choline.²²³⁻²²⁷ Phylogenetic analysis suggests these five solute-binding proteins cluster into three clades with distinct active sites. The bacterial ProX and OpuAC form one clade, the archaeal ProX forms its own, while OpuBC and OpuCC form a third (Figure 6). The betaine-binding sites of an enzyme from each clade will be discussed.

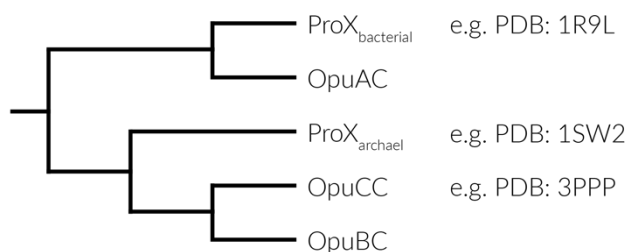


Figure 6. A qualitative phylogenetic tree containing characterized solute binding proteins specific for betaines: OpuAC, OpuBC and OpuCC from *Bacillus subtilis* and ProX from *Escherichia coli* and archaeon *Archeoglobus fulgidus*.

Bacterial ProX. In 2004, crystal structures of a bacterial ProX, a betaine binding SBP were reported. ProX_{bacterial} is reported to bind both GB and PB selectively and with high affinity, with K_D values of approximately 1 μ M and 5 μ M respectively.^{198, 228-230} Liganded structures revealed that both ligands are

entirely engulfed in the closed conformation and that cation- π interactions between the positive charge of the quaternary amine of the ligand and three tryptophan residues were key determinants for substrate selectivity (Figure 7A). Trp65 and Trp144 sit almost parallel to one another, on either side of the trimethylated amine of GB or PB. These two tryptophans, along with Trp188, form three faces of a rectangular aromatic box, forming the basis of binding quaternary amine binding. Site-directed mutagenesis was used to validate the individual contributions of each of the Trp residues in substrate binding. The carboxylic group of GB points out of this aromatic box and makes hydrogen-bonding contacts to the backbone amides of Gly141 and Cys142. These residues sit on a loop held in a particular conformation by a disulfide bond. The carboxylic group also makes another hydrogen bond to the imidazole of His69.³⁷

Archaeal ProX. The ProX from an archaeal species, ProX_{archaeal}, also binds GB and PB, with K_D values of 50 and 60 nM respectively.²³¹ Structures solved of ProX_{archaeal} in complex with its natural substrates revealed a different substrate-binding site to that of ProX_{bacterial} (Figure 7B&A respectively). Despite a low-sequence similarity between ProX and ProX_{archaeal} (29%), both SBPs provide a similar solution for the binding of cationic quaternary amines by utilizing cation- π interactions. Instead of using tryptophan, four tyrosines: Tyr63, Tyr111, Tyr190 and Tyr214, form four faces of a box, forming an “aromatic girdle of tyrosines” that encompass the trimethylated amine of GB. In contrast to the ProX_{bacterial} binding site, ProX_{archaeal} also makes use of non-classical hydrogen bonds between the methyl groups of the quaternary ammonium and phenolic groups of the four tyrosines and a mainchain carbonyl. The mainchain carbonyl of Asp109 points in towards the quaternary ammonium headgroup, forming the bottom of the π -box. This aromatic girdle provides a negative surface potential to complement the cationic charge on the trimethylated amine. The carboxylic group of GB points out of the aromatic cage and forms a salt bridge to Lys13 and Arg149 and a hydrogen bond to Thr66. A second contrast to ProX_{bacterial} is that the solutes are not completely closed off from solvent. One water molecule sits within the active site, separating the ligand from the rest of the bulk solvent. This water molecule makes no direct contact to GB, yet makes two hydrogen bonds to residues in close proximity.²³¹

OpuCC. OpuCC also utilizes an aromatic girdle to bind GB, as revealed through numerous crystal structures in complex with various betaines; GB, choline, ectoine and carnitine. The quaternary amines of these molecules are all bound in an identical fashion to the archaeal ProX with an aromatic girdle comprised of four tyrosines, Tyr91, Tyr137, Tyr217 and Tyr241 (Figure 7C). As in ProX_{archaeal}, a mainchain carbonyl, Asn135 points in towards the cationic ammonium group, forming the base of the aromatic cage. Residues close to the box contribute differently to bind the different substrates. The carboxylic group of GB is fixed in place by three hydrogen bonds, to Thr94, Gln39 and a water molecule, which also makes hydrogen-binding contact to Gln39. This also contrasts with the archaeal and bacterial ProX, as neither makes direct contact to a water molecule. However this may be a result of the substrate promiscuity, as the longer chained betaine, carnitine substrates occupy more space and do not make contact to any water molecules.

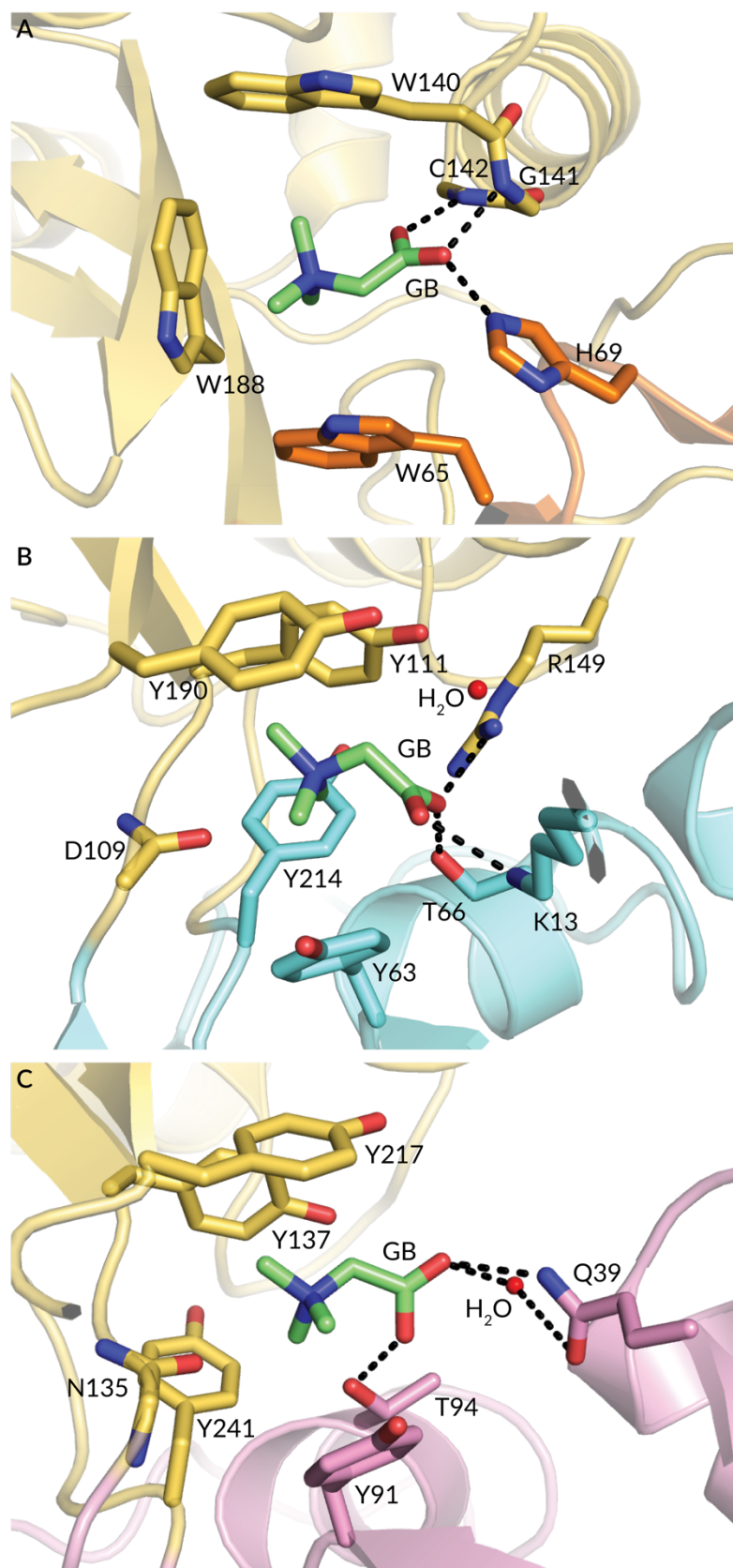


Figure 7. Active sites of three SBP that bind glycine betaine (GB). GB is shown in green, while important substrate-binding residues are labelled and shown as sticks and coloured according to their domain. A. ProX from *E. coli* in complex with GB (PDB: 1R9L, 1.6 Å). B. ProX from *Archeoglobus fulgidus* in complex with GB (PDB: 1SW2, 2.1 Å) and C. OpuCC from *Bacillus subtilis* in complex with GB (PDB: 3PPP, 2.4 Å).

Solute Binding Protein Phylogenetic Analysis. To explore the idea that EGT and its derivatives may have a specific solute-binding protein as part of an ABC transport system, a qualitative phylogenetic tree was created from characterized betaine-binding proteins. The sequences of three potential EGT solute-binding proteins from the close genomic environment of EGT were included (Figure 8). A preliminary sequence alignment and analysis indicates that the tyrosine girdle that is responsible for quaternary amine binding in ProX_{archaeal} and OpuCC/BC is conserved among the hypothetical EGT binding proteins. This indicates that the solute bound by the SBPs of interest likely contains a quaternary ammonium. In the phylogenetic tree, the SBP homologues that are co-encoded with ergothionases cluster away from those that bind GB/PB and choline-binding SBP, suggesting these are unlikely to be the native solutes. This is consistent with our hypothesis that these proteins may be involved in EGT transportation (Figure 8). The sequences associated with EGT also split into a further two clades, suggesting further diversity among the sequences.

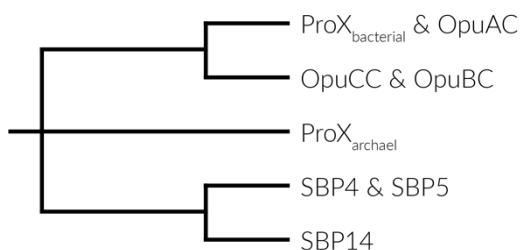


Figure 8. A qualitative phylogenetic tree containing characterized solute-binding proteins specific for betaines and three SBPs implicated for EGT binding: OpuAC, OpuBC and OpuCC from *Bacillus subtilis*, ProX from *Escherichia coli* and archaeon *Archaeoglobus fulgidus*, SBP4 from *Martellella endophytica*, SBP5 from *Trabulsiella odontotermitis*, SBP14 from *Bacillus sp. FJAT-27264*.

Functional Characterization, performed by Mariia Beliaeva. Several potential EGT binding SBPs were selected, produced and their affinities tested for a range of substrates including betaine, histidine, EGT, EGTSO₃⁻, MeEGT and TMH. This was performed by Mariia Beliaeva. Affinities were determined by Isothermal titration calorimetry (ITC). Of the various SBP investigated, three representative proteins have been selected, SBP4, 5 and 14, the K_Ds for which are listed in Table 1. SBP4 and SBP5 both bind EGTSO₃⁻ with high affinities of 9 and 11 nM respectively. EGT also binds, but with much weaker affinities of 35 and 32 μM respectively. Both proteins have no measurable affinity for MeEGT. SBP5 also has no measurable affinity for TMH, while SBP4 binds TMH with a K_D of 23 μM. SB14 binds all four EGT variants, with the thiol variants in the low nM range. EGT binds with the highest affinity (K_D = 20 nM), while TMH is bound with a K_D of 5 μM. These results implicate SBP4 and 5 as EGTSO₃⁻ binding SBPs and SBP14 as a EGT SBPs.

Table 1. K_D values of four solute binding proteins for EGT and three derivatives, the solute with the highest affinity of each protein is highlighted as bold. A dash (-) indicates that no significant binding was measured ($K_D > 0.5$ mM).

	Egt	EGTSO ₃ ⁻	MeEGT	TMH
SBP4	35 ± 5 μM	9 ± 2 nM	-	23 ± 3 μM
SBP5	34 ± 5 μM	26 ± 1.3 nM	-	-
SBP14	20 ± 2.3 nM	170 ± 11 nM	26 ± 1.3 nM	4.90 ± 0.61 μM

Aim of this Chapter. The goal of this work is to structurally characterise several different EGT-related solute-binding proteins with different ligands. A structural basis will enable the determinants of substrate selectivity to be identified through structural comparison, compounded with a bioinformatic approach. This chapter then aims to test these hypotheses in the framework of a sequence similarity network, which allows for a complete coverage of sequence space. This will finally be utilised to analyse the distribution of the SBPs and their genomic environment to gain further insight into the organisms that do not produce EGT but utilise it, and evolutionary insights into the genes.

This chapter encompasses the structural characterization of three SBPs with various ligands, listed below. Each protein and its structure(s) will be discussed consecutively, followed by a structural and bioinformatic comparison and discussion.

EGTSO₃⁻ specific binders:

- SBP4, a solute binding protein from *Martelella endophytica* (proteobacteria)
- SBP5, a solute binding protein from *Trabulsiella odontotermis* (enterobacteria)

EGT specific binders:

- SBP14, a solute binding protein from *Bacillus* sp. FJAT-27264 (firmicute)

Results and Discussion

Structural Characterization of SBP5

Crystal structure determination of SBP5. Functional characterization identified SBP5 as an EGTSO_3^- -specific SBP. To gain insight into the binding mode and selectivity determinants, structures of SBP5 in complex with EGTSO_3^- and EGT were determined by X-ray crystallography. Structures in complex with EGTSO_3^- and EGT were obtained by co-crystallization, with crystals diffracting to resolutions of 2.1 and 1.2 Å respectively. In the presence of EGTSO_3^- , SBP5 crystallized in space group $P2_12_12_1$ with cell constants $a, b, c = 43.8, 107.6$ and 227.9 Å respectively. While in the presence of EGT, SBP5 crystallized in space group $P12_11$ with cell constants $a, b, c = 46.0, 45.5$ and 61.9 Å respectively and $\beta = 103.6^\circ$. For data collection and refinement statistics, see Supplementary tables 4 and 5 respectively. The first crystal structure (with EGT as a ligand) was solved by molecular replacement using a search model derived from four published structures of SBP5 analogues (PDB: 3PPN, 4Z7E, 1SW4, 3R6U). The EGTSO_3^- liganded structure was solved by molecular replacement with the SBP5-EGT model. The electron density for both structures revealed a continuous polypeptide chain from residues 25 to 295 (Figure 9). The first two residues of the construct used for crystallization lacked electron density and have therefore not been modelled. Both ligands could be unambiguously modeled into the $(F_o - F_c)$ difference density map (Figure 10).

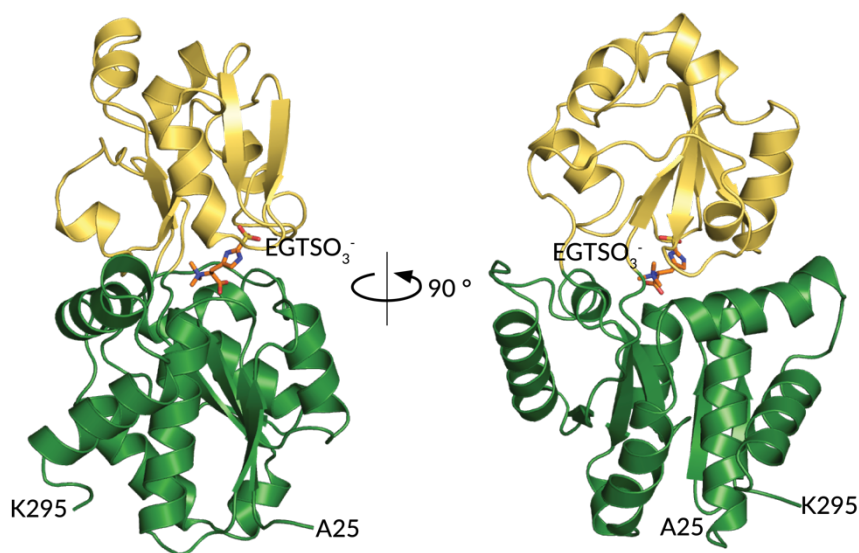


Figure 9. Structural Analysis of SBP5 in complex with EGTSO_3^- . Cartoon of the SBP5- EGTSO_3^- structure. Domain 1 (residues 25 – 127 & 232 – 295) is shown in green while the domain 2 (residues 128 - 231) is shown in yellow. EGTSO_3^- (orange) is bound as a ligand and indicates the location of the substrate-binding site.

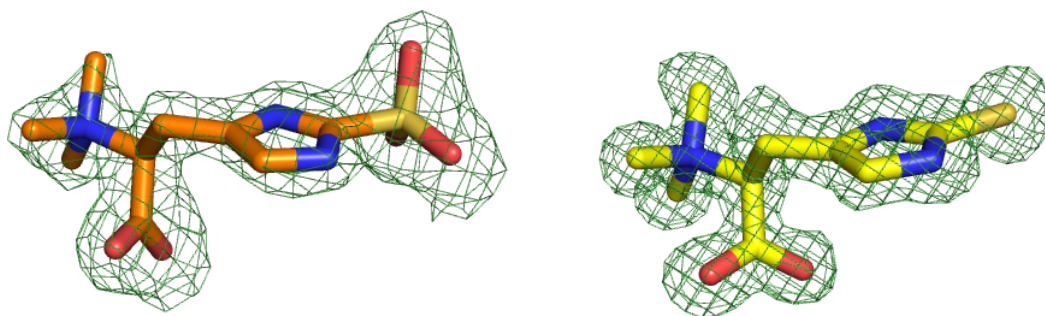


Figure 10. OMIT maps for SBP5 bound ligands. EGTSO₃⁻ (2.1 Å) and Egt (1.2 Å). ($m|F_o| - D|F_c|$, electron density; $\sigma = 3.0$).

The SBP5 structure revealed a fold typical of the type II SBPs, and like all SBPs can be divided into two domains. Domain 1 (residues 25 - 127 & 232 - 295) comprises two β sheets, one with two parallel strands, and the second with three antiparallel sheets, of which one strand is much shorter than the other two. Domain 1 also contains eight α -helices. Domain 2 (residues 128 - 231) contains two beta sheets, a two-stranded parallel sheet and a two-stranded antiparallel sheet, and six short α -helical coils.

SBP5 Ergothioneine-Sulfonic Acid Binding. EGTSO₃⁻ is bound in a deep groove between the two domains, in which solvent access is minimal. The trimethyl amine moiety resides closest to the center of the protein, while the sulfur group points outwards towards the protein surface. The trimethyl moiety is bound by an aromatic girdle of four aromatic residues (Tyr83, Tyr129, Phe209 and Tyr233). However, in contrast to the canonical four tyrosines, one tyrosine is replaced by a phenylalanine. (Figure 11A). The backbone carbonyl group of Asn127 points towards the trimethyl moiety making non-typical hydrogen bonds to each of the ammonium methyl groups (3.1, 3.1 and 3.4 Å). Such an interaction is also observed in the archaeal ProX and in EgtB structures with TMH bound.^{49, 173, 231} The side chain of Thr86 also makes such a contact to a methyl group (3.2 Å). The EGT carboxyl group forms salt bridges with Lys33 (2.8 & 3.4 Å) and Arg166 (3.0 & 3.3 Å), and a hydrogen-bonding contact to Thr86 (2.6 Å). In the substrate imidazole ring, N τ hydrogen bonds to Tyr129 (2.9 Å), while N ϵ does not make any contacts. The oxygens of the sulfonic acid moiety make numerous hydrogen-bonding contacts, forming interactions to the side chains of Gln207 (2.9 Å) and Tyr129 (3.1 Å) and to the backbone amides of four consecutive residues, which make up a loop region between β -sheet 6 and α -helix-7: Gly161 (2.8 Å), Ala162 (3.2 Å), Glu163 (2.9 Å) and Phe164 (3.1 and 3.5 Å). The side chain of Glu163 points in towards the active site, making contact to and positioning Arg167 (3.1 Å).

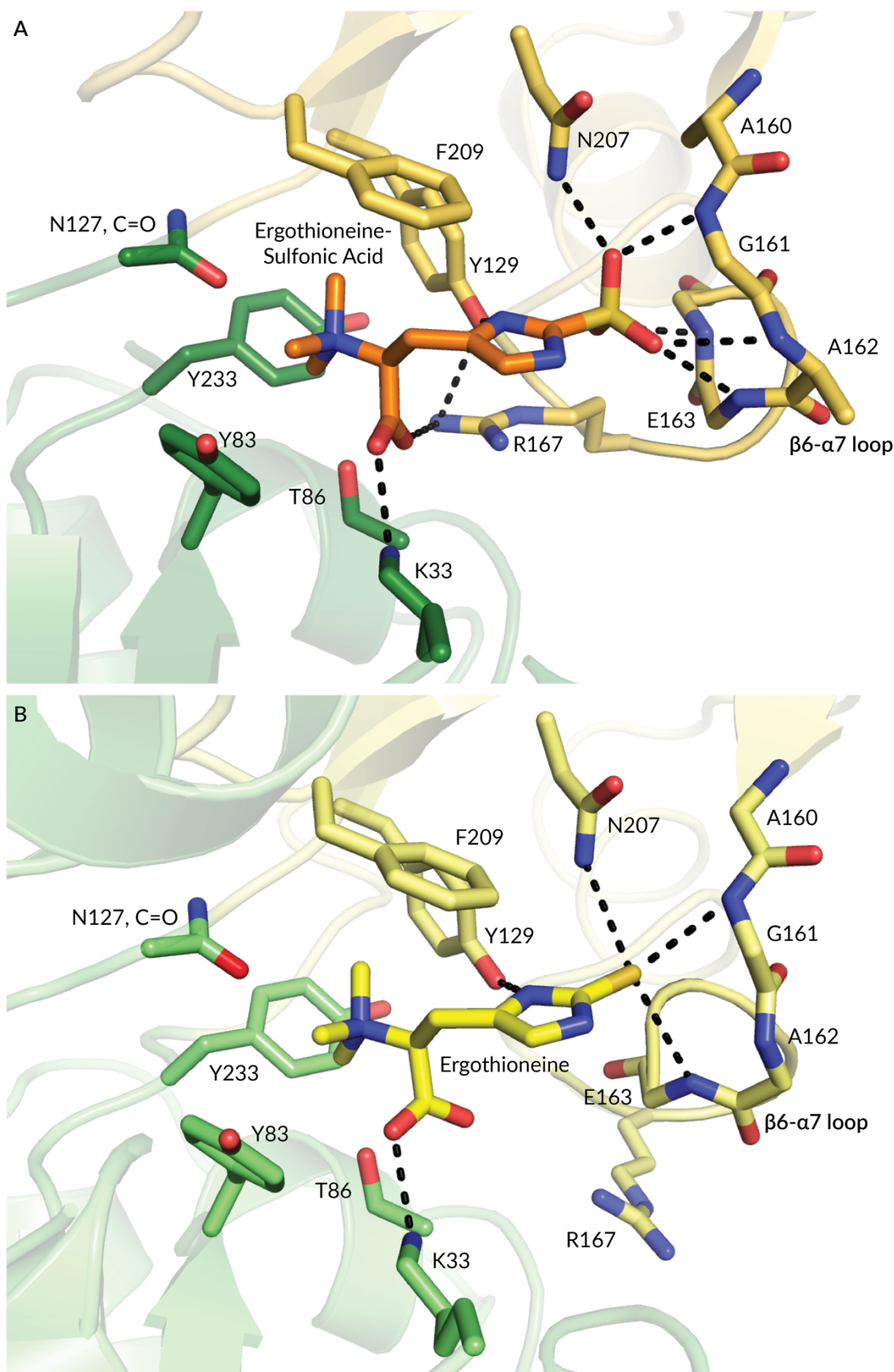


Figure 11. Active Site of SBP5 with A. EGTSO_3^- bound and B. EGT bound. Substrate binding residues are shown as sticks and are colored according to their domain location. The $\beta 6$ $\alpha 7$ loop rearranges on the presence of the sulfonic acid moiety.

SBP5 Ergothioneine Binding. EGT binds in a similar fashion to EGTSO_3^- , with a slightly different tilt with respect to the rest of the protein (Figure 12). However, several key interactions are lost in comparison to EGTSO_3^- structure. The EGT carboxyl group forms salt bridges with Lys33 (2.8 & 3.4 Å) and a hydrogen bond to Thr86 (2.8 & 3.2 Å). In the substrate imidazole ring, $\text{N}\pi$ hydrogen bonds to Tyr129 (2.9 Å), however this residue moves further out of the structure in comparison to its location in the EGTSO_3^- structure, giving a greater hydrogen-bonding distance, and presumably a weaker interaction. $\text{N}\pi$ of the imidazole ring now forms a water-mediated hydrogen bond to the backbone amine of Ala162. The sulfur hydrogen bonds to the backbone of Gly161 (3.3 Å), Glu163 (3.4 Å) and the side chain of Gln207 (3.4 Å). Only two hydrogen-bonding contacts are made to the $\beta 6$ - $\alpha 7$ loop, in contrast to the numerous interactions (7 hydrogen bonds) that the sulfonic head group of EGTSO_3^- makes to this loop. Through the loss of these interactions the $\beta 6$ - $\alpha 7$ loop in the EGT structure folds into another conformation, which moves Arg167 out of the active site. In the EGTSO_3^- structure, Arg167 forms a key salt bridge to the substrate carboxylate. An interaction is also lost between the sulfonic moiety and Tyr129, which sits further out of the active site. The movement of Tyr129 and rearrangement of the $\beta 6$ - $\alpha 7$ loop and consequently movement of Arg167 further impacts EGT binding, as these structural rearrangements result in a less well packing substrate (Figure 13). A larger binding pocket results, with space around the carboxylic group that is filled by several water molecules. In contrast EGTSO_3^- is tightly encased, providing a greater surface area for substrate binding.

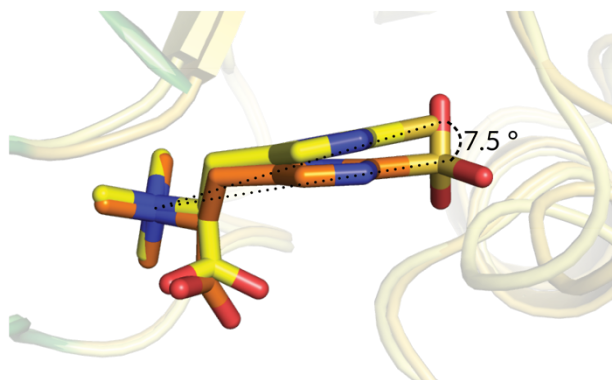


Figure 12. Conformations of EGT and EGTSO_3^- in SBP4. The quaternary ammonium of both ligands occupy the same space, taking this atom as a fixed point, EGT is tilted up by 7.5° in the active site in comparison to EGTSO_3^- .

The loss of interactions; the salt bridge to the carboxylate and numerous hydrogen bonding interactions to the sulfonic head group, combined with the less-encased binding pocket account for the almost 3000-fold decrease in binding affinity of SBP5 for EGT (K_D of 34 μM) versus EGTSO_3^- (K_D of 26 nM).

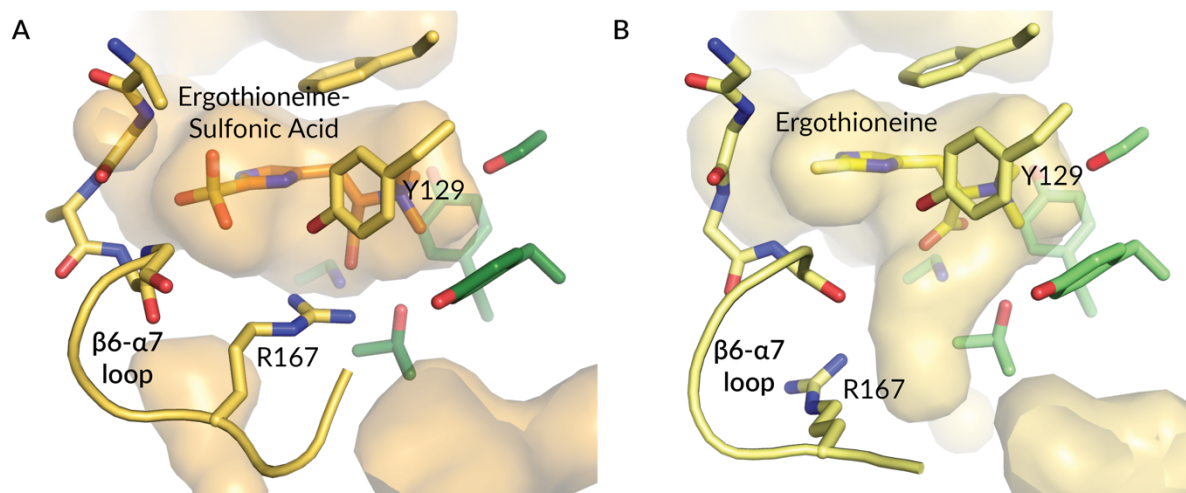


Figure 13. Comparison of the substrate binding pockets in the EGT and EGTSO₃⁻ bound structures in SBP5. A much larger, and less well-packed cavity results with EGT as a ligand.

Structural Characterization of SBP4

Crystal structure determination of SBP4 with Ergothioneine. In parallel to SBP5 crystallization, the same experiments were carried out for SBP4. SBP4 & SBP5 are closely related, and are both EGTSO_3^- binders. However, only structures of SBP4 in the presence of EGT were obtained. SBP4 crystallized in space group $P2_12_12_1$ with cell constants $a = 39.7 \text{ \AA}$, $b = 53.7 \text{ \AA}$ and $c = 127.4 \text{ \AA}$ with a crystal diffracting to a resolution of 1.4 \AA . For data collection and refinement statistics, see Supplementary tables 4 and 5 respectively. The structure was solved by molecular replacement using the structure of SBP5 + EGT as search model. SBP4 and SBP5 share 60 % sequence identity. The electron density revealed a continuous polypeptide chain from residues 25 to 295 (Figure 14). The first three residues of the construct used for crystallization lacked electron density and have therefore not been modelled. EGT could be unambiguously modeled into the $(F_o - F_c)$ difference density map (Figure 15).

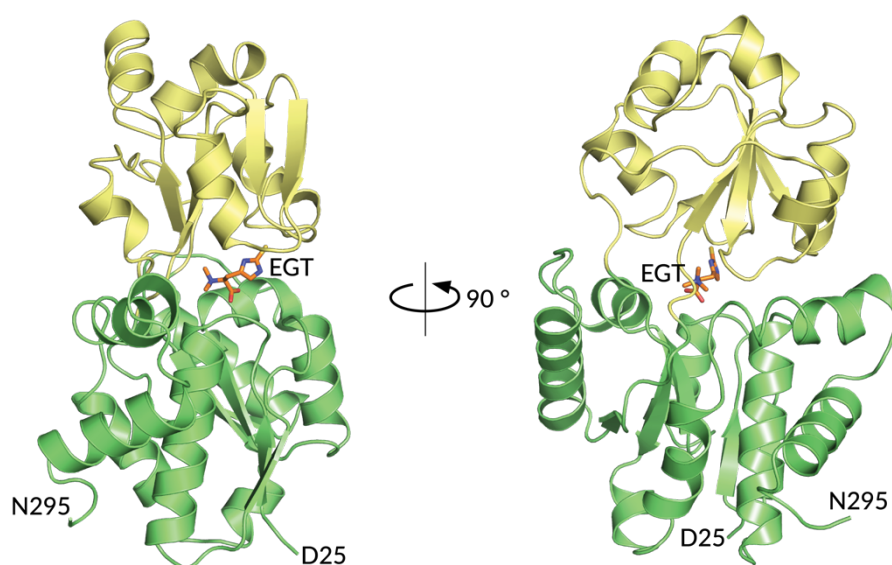


Figure 14. Structural Analysis of SBP4 in complex with EGT. Cartoon of the SBP5-EGT Structure. Domain 1 (residues 25 – 153 & 260 – 295) is shown in green while the domain 2 (residues 154 - 259) is shown in yellow. EGT (orange) is bound as a ligand and indicates the location of the substrate binding site.

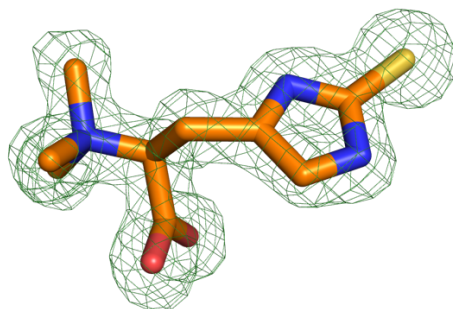


Figure 15. Omit map for SBP4 bound EGT. $(m|F_o| - D|F_c|)$ electron density; $\sigma = 3.0$.

SBP4 Ergothioneine Binding. EGT binds in a deep groove between the two domains, binding in a similar mode to EGT in SBP5. The aromatic girdle is comprised of four tyrosines, Tyr82, Tyr128, Tyr208 and Tyr232, in contrast to the 3Tyr-1Phe motif of SBP5 (Figure 16). The backbone carbonyl group of Gln126 points towards the trimethyl moiety, making non-typical hydrogen bonds to each of the ammonium methyl groups (3.2, 3.2 and 3.3 Å). The EGT carboxyl group forms salt bridges with Lys32 (2.8 & 3.4 Å) and Arg166 (2.8 & 3.2 Å) as in the SBP5 EGTSO₃⁻ structure. In the substrate imidazole ring, N τ hydrogen bonds to Tyr128 (2.9 Å) while N π forms a water-mediated hydrogen bond to Glu162 (2.8 & 3.4 Å). Glu162 has no equivalent in SBP5, yet appears to hydrogen bond to the EGT sulfur. Also in close proximity to the sulfur are three atoms, the electron density of which allows for their unambiguous placement (Figure 17). The central atom is 2.0 Å from sulfur, and 2.0 and 1.7 Å from the peripheral atoms. These short distances rule out hydrogen bonding, and are best explained by a central metal that coordinates to EGT and two crystallographic water molecules. These water molecules make further contacts to the amine backbone of Glu162, Gly160, Phe103 and the side chains of Asn206 and Glu162. The most likely central metals are Na⁺ and Zn²⁺ due to significant containment in the crystallization conditions. A trigonal planar geometry is however unusual for both metals.²³²⁻²³³ This unusual metal binding geometry likely arises because the protein is electrostatically optimized to bind a sulfonate.

ITC showed that EGTSO₃⁻ binds with an almost 4000-fold greater affinity to SBP4 than EGT does, indicating that this is the preferred substrate (Table 1). The structure itself indicates this, with the ordered water molecules occupying the space most likely taken up by the oxygen atoms of the EGTSO₃⁻ head group. It is likely this head group would make similar interactions to those that water molecules make in the EGT-bound structure. While crystals of SBP4 containing EGTSO₃⁻ were grown and diffracted well, the data proved difficult to solve.

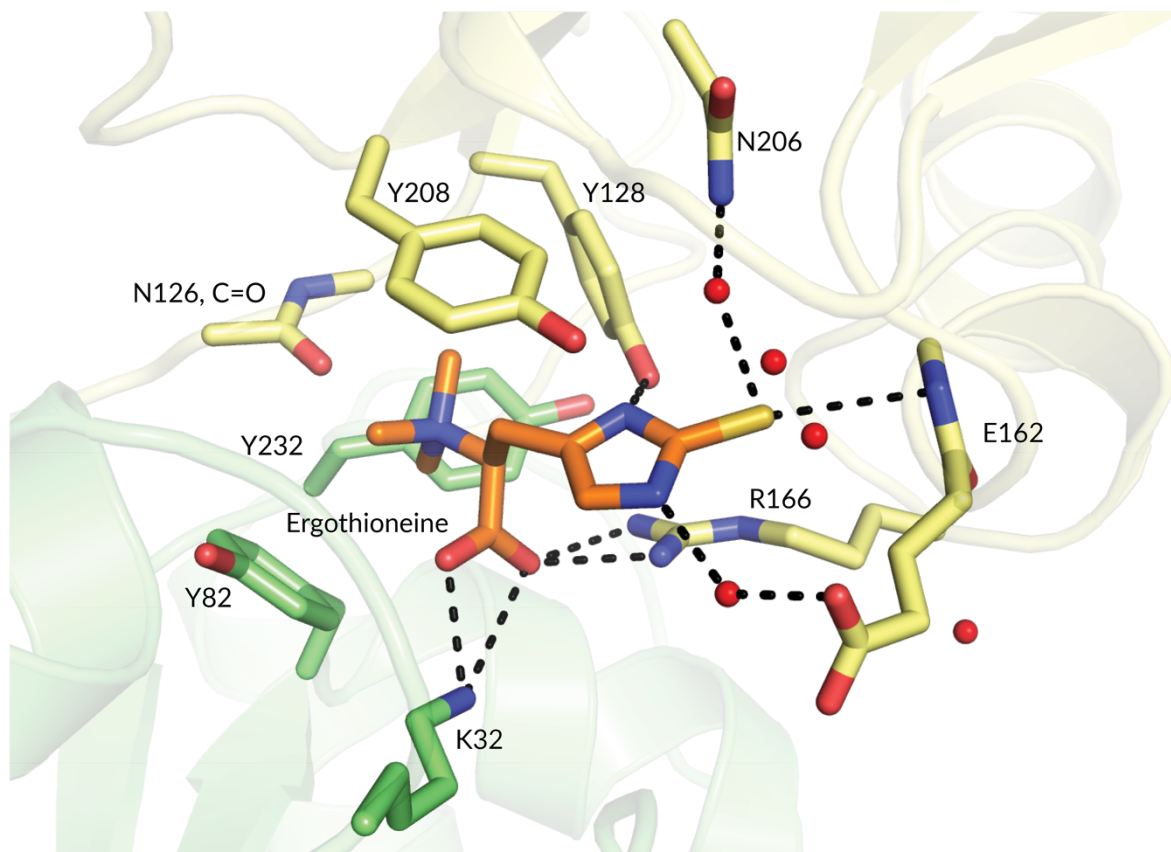


Figure 16. EGT binding site. Substrate-binding residues are shown as sticks and are colored according to their domain location.

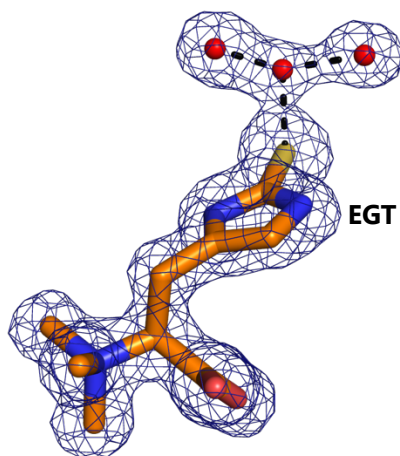


Figure 17. Electron density for EGT and coordinating atoms. $(2m|F_o| - D|F_c|)$ electron density; $\sigma = 1.0$.

Structural Characterization of SBP14

Crystal structure determination of SBP14. SBP14 represents a SBP from a different region of sequence space than SBP4 and 5. Any differences between the two groups appear to manifest themselves in the substrate selectivities, as SBP14 was found to bind EGT with high affinity. To characterize these differences, SBP14 was crystallized in the presence of several ligands. In the presence of Egtor related derivative, SBP14 crystallized in space group $P2_13$ with cell constants $a,b,c = 98.2 \text{ \AA}$. Liganded structures with EGT, MeEGT, EGTSO_3^- or TMH were obtained by co-crystallization, with crystals diffracting to a resolution of 1.6, 1.9, 1.8 and 2.0 \AA respectively. For data collection and refinement statistics, see Supplementary tables 4 and 5 respectively. The first crystal structure, with MeEGT as a ligand, was solved by molecular replacement using the native structure of SBP5 + EGT as search model. SBP5 and SBP14 share 41 % sequence identity. The other liganded structures were solved by molecular replacement with the SBP14-MeEGT model. The electron density for all structures revealed a continuous polypeptide chain from residues 54 to 322 (Figure 18). The first 13 residues of the construct used for crystallization lacked electron density and have therefore not been modelled. All ligands could be unambiguously modeled into the $(F_o - F_c)$ difference density map (Figure 19).

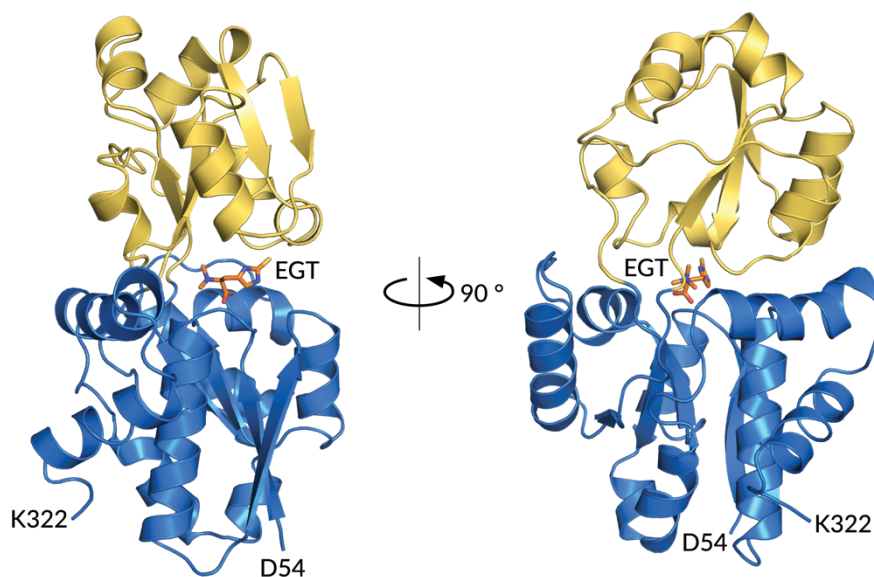


Figure 18. Structural Analysis of SBP5 in complex with EGT. Cartoon of the SBP5-EGT structure. Domain 1 (residues 54 – 153 & 260 – 322) is shown in dark blue while the domain 2 (residues 154 – 259) is shown in yellow. EGT (orange) is bound as a ligand and indicates the location of the substrate-binding site.

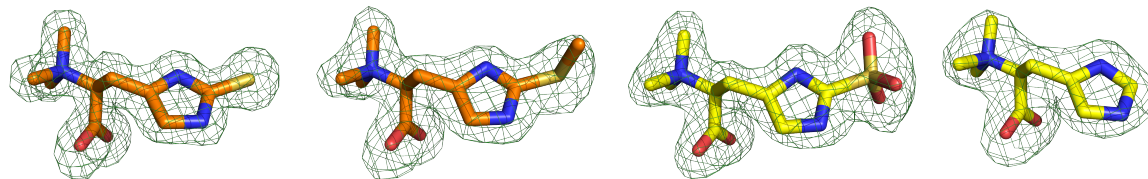


Figure 19. Omit maps for SBP14 bound ligands. Respectively: EGT, MeEGT, EGTSO₃⁻ and TMH ($m|F_o| - D|F_c|$, electron density; $\sigma = 3.0$).

SBP14 Ergothioneine Binding. The ligand with highest affinity, EGT, is bound in a deep groove between the two domains, in which solvent access is minimal. The trimethyl amine moiety resides closest to the center of the protein, while the sulfur group points outwards towards the protein surface. As with SBP4 and 5, the trimethyl moiety is bound by the archetypical aromatic girdle. The backbone carbonyl group of Gln154 points towards the trimethyl moiety and appears to make a dipolar contact (3.3 Å) to a methyl group. The EGT carboxyl group forms salt bridges with Lys61 (2.8 Å) and Arg194 (2.9 & 3.1 Å). In the substrate imidazole ring, N τ hydrogen bonds to Tyr156 (2.9 Å), while N ϵ makes contact to Asp91 (2.8 & 3.4 Å). Asp91 makes water-mediated contacts to Lys61 and Arg194, suggesting its charge is saturated. The head group is rather isolated, with the sulfur making hydrogen bonding to only Gln92 (3.4 Å) (Figure 20A).

In contrast to SBP5, all four liganded structures of SBP15 adopt superimposable positions. Structural alignment of the four liganded structures showed that all four substrates (EGT, MeEGT, EGTSO₃⁻ & TMH) bind in an identical fashion and do not lead to any rearrangement or change of active site residue conformation (Figure 20AB). For MeEGT and EGTSO₃⁻ no additional interactions are made from varying sulfur moiety in comparison to the EGT structure. While in the EGTSO₃⁻ bound structure, Gln92 can make contact to the sulfonic oxygen (3.0 Å) instead of the sulfur as in EGT and MeEGT. The lack of interactions around the head group may explain the lack of selectivity between the EGT metabolites.

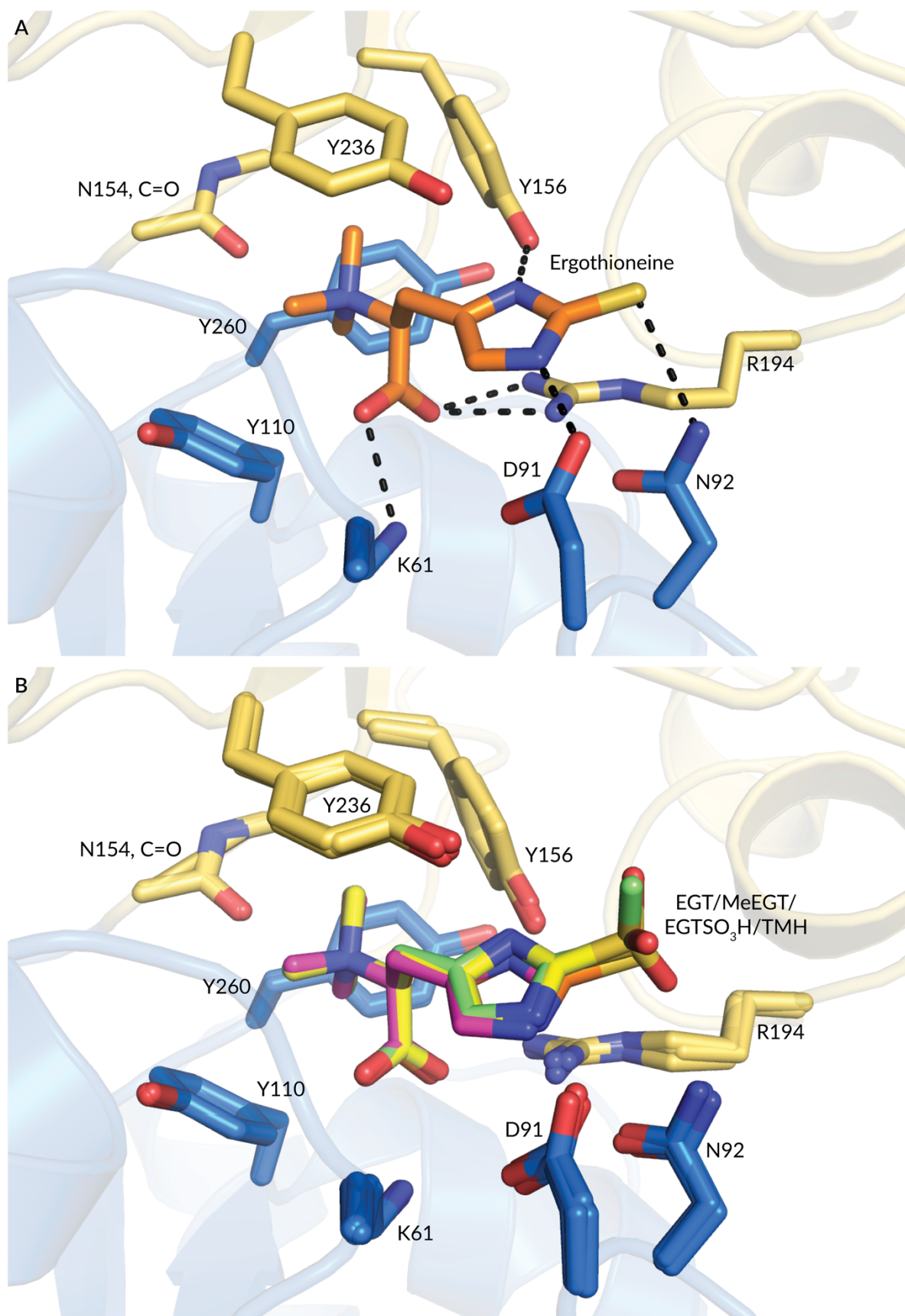


Figure 20. Substrate binding site of SBP14 with A. EGT bound Figure B: Alignment of the four liganded SBP14 with EGT in orange, TMH in pink, MeEGT in green and EGTSO₃H/TMH in yellow. Substrate-binding residues are shown as sticks and are colored according to their domain locations.

Structural Comparison

The overall structures of SBP4, 5 and 14 are very similar, with rmsd values of 0.71 Å (232 to 232 C- α atoms) for SBP4 & SBP5, 1.28 Å (241 to 241 C- α atoms) for SBP4 and SBP14, and 1.54 Å (244 to 244 C- α atoms) for SBP5 and SBP14. A large proportion of the active site is also well-conserved among the three structures. However, some key differences exist around the binding of the imidazole ring and sulfur head group (Figure 21). The active site residues closer to the center of the protein, which are responsible for binding of the quaternary ammonium group and carboxylic group, are well conserved. All three proteins make use of an aromatic girdle, with a carbonyl backbone to create a negatively-charged aromatic cage for binding of the trimethyl-amine group. A lysine and arginine are conserved in all three homologues, forming a salt bridge to the substrate carboxylate. It should be noted that in SBP5, the interaction to the latter arginine is dependent on folding of a loop in a specific conformation around the sulfonic head moiety. One of the tyrosines in the aromatic girdle hydrogen bonds to N τ of the substrate imidazole ring. The interactions around the rest of the substrate imidazole ring and the sulfur head group differ among the three proteins. In SBP14, N τ has a direct hydrogen-bonding partner, Asp91, while in SBP4, such an interaction to Glu162 is mediated by a water molecule. In contrast, N τ has no hydrogen-bonding partner in SBP5. For binding of the sulfur head group, an asparagine is conserved in the EGTSO₃⁻ binders. In addition to this asparagine, a glycine-rich loop that folds over the sulfonic acid moiety makes numerous consecutive interactions to the amine backbone of this loop. In contrast, the SBP14 sulfur moiety makes only one interaction to a Asn92, and is otherwise surrounded by an open unoccupied space.

From the comparison and analysis of these three structures, there are four key differences in the substrate binding sites. In SBP14, Asp91 and Asn92 are involved in EGT binding, while no equivalent residues exist in the EGTSO₃⁻ specific SBPs. Instead Asn206/207 and glycine-rich β 6- α 7 loop appear to be unique to the EGTSO₃⁻ SBP type. These residues will be explored in a bioinformatic analysis in the following section to determine if they are conserved among close and distant homologues.

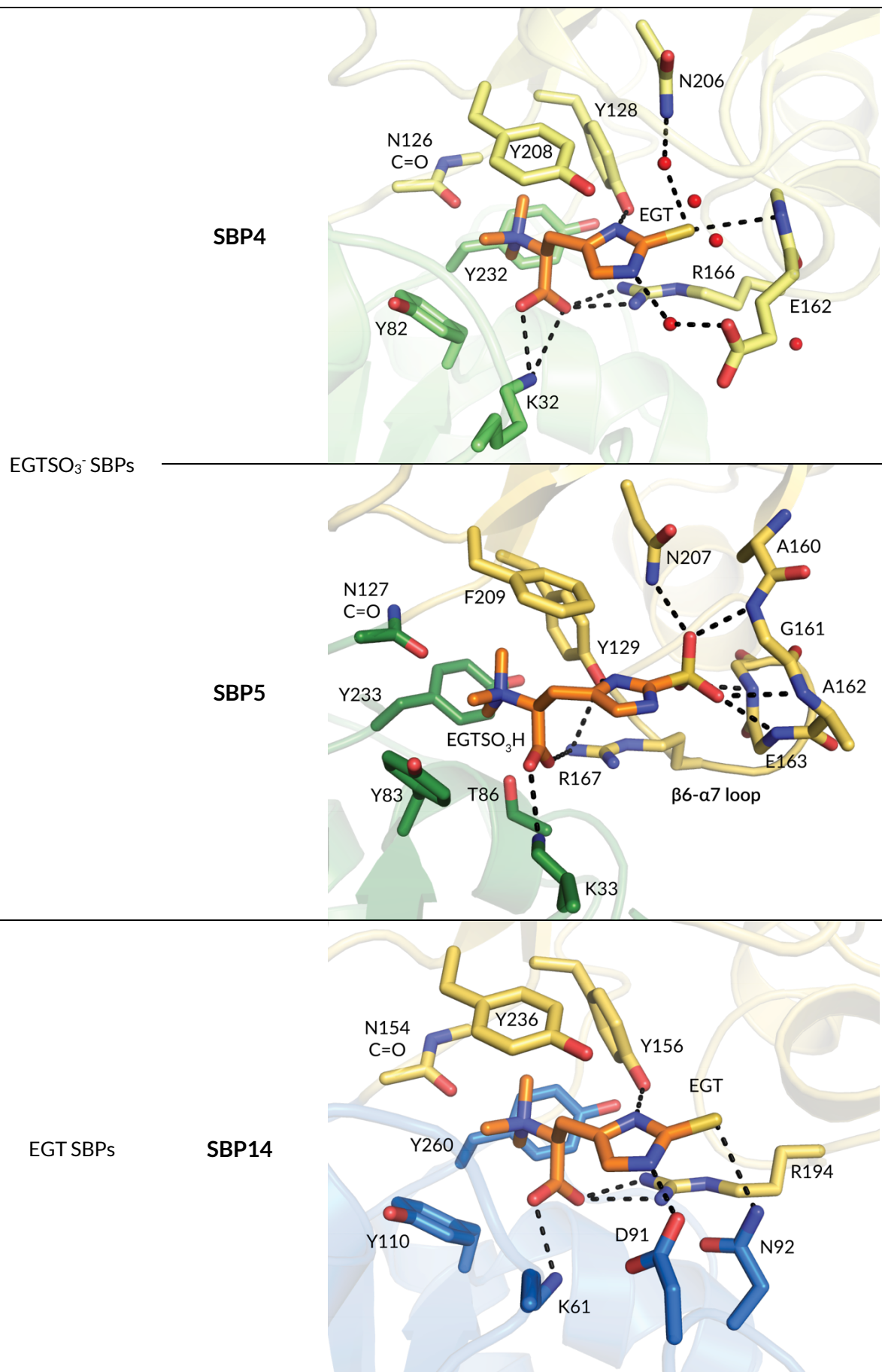


Figure 21. Comparison of the substrate binding sites of SBP4, 5 and 14.

Selectivity of the solute binding proteins. SBP5 binds EGT, and EGTSO_3^- , yet not MeEGT and TMH. Lack of TMH binding could be rationalized by the lack of a sulfur moiety, however MeEGT has such a group, yet has no measurable binding affinity. The sterics of the methyl group cannot be accountable, as the much larger sulfonic head group binds with an even greater affinity. Perhaps electronics could rationalize the substrate selectivity of SBP5. A similar selectivity is observed in the ergothionase from *Treponema denticola* (*TdeErgothionase*). In solution and at a physiological pH, EGT exists predominantly in its thione form.¹⁴ In the EGT activating enzyme, ergothionase, EGT appears to be bound in a zwitterionic form, as a thiolate and imidazolium ring. Ergothionase accepted EGT and EGTSO_3^- as substrates, yet not MeEGT or TMH. The ergothionase active site permits itself to solvate the zwitterionic side chain, with an anionic head group and a positive charge on the imidazolium ring. It has been proposed that this positive charge activates the substrate for catalysis by increasing acidity of methylene group. TMH and MeEGT, neither of which are accepted as substrates, cannot attain a zwitterionic state and bind to *TdeErgothionase* with a neutral side chain (Figure 22). This was corroborated by mutations in the side chains to disrupt charge complementarity in the enzyme:ligand complexes.

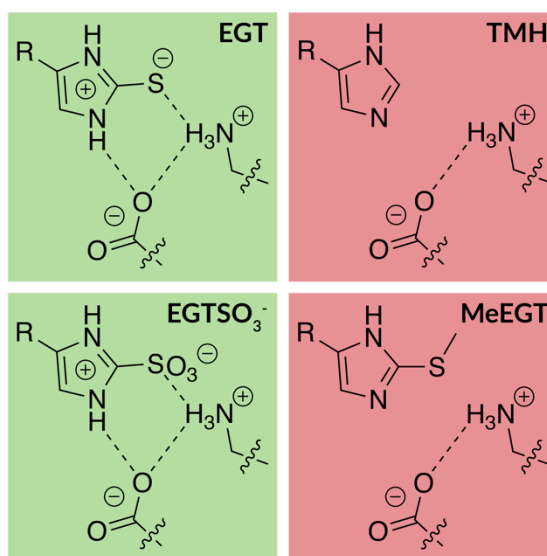


Figure 22: Proposed substrate-binding modes of EGT and derivatives in *TdeErgothionase*. Ligands that can form a positive charge on the imidazole ring and a negative charge on the sulfur head group complement the charges in the active site (green boxes) and are more active than substrates with a neutral imidazole ring (red boxes).²¹⁷

We were interested in determining if the binding of EGT as a zwitterion is a common theme among EGT-binding proteins. Formation of a zwitterionic species would be accompanied by polarization of the C-S bond. Ligand restraints were generated with grade for both the thione and thiol forms of EGT.²¹⁵ The thione C=S bond distance was set to 1.677 Å with estimated standard deviation (esd) of 0.016 Å, while the thiol equivalent had a C-S bond distance of 1.722 Å with an esd of 0.030. Despite having atomic resolution structures, the differences in bond length between C=S and C-S (0.045 Å) are too little to determine if polarization of the C-S bond is occurring. Analysis of the residues forming the binding pockets of SBP4 & 5 indicates a zwitterionic binding mode is not likely, as no charged residues interact directly with the imidazole ring or the sulfur head group. In SPB14, Asp91 makes a direct contact to $\text{N}\pi$ of the substrate imidazole ring,

yet Asp91 also makes water-mediated contact to two likely positively-charged groups. Additionally, the guanidinium side chain of Arg194 packs under the imidazole ring. Therefore a charged imidazole is unlikely. Additionally, SBP14 binds all substrates and MeEGT and EGT with similar affinities, providing further evidence that EGT and its derivatives are bound as a neutral thione without polarization, which is the most stable form of EGT. This conceptually makes sense, as SBPs merely want to bind, not activate their substrates for catalysis.

Bioinformatic Analysis

Our ultimate goal is to predict which organisms utilize EGT and EGTSO₃⁻. The identification of two SBP types that are specific for EGT and EGTSO₃⁻ and are part of an ABC transporter system provides the molecular basis towards our goal. The structural characterization of three SBPs, provides a platform for this approach. The comparison of these structures enables identification of the structural determinants that distinguish EGT or EGTSO₃⁻ binding in the SBPs. The motifs will be tested through their correlation to clustering in a sequence similarity network, and analysis of the SBP genomic environment. If corroborated these structural motifs will then be leveraged to explore the distribution of the EGT related SBPs.

To determine the substrate-binding residues that are exclusive to a SBP type sequence, sequence LOGOs of EGT SBPs (N = 10) and EGTSO₃⁻ SBPs (N=11) were generated and aligned using MUSCLE (Figure 23). The sequences are numbered corresponding to SBP5 and SBP14 respectively. Interactions from the crystal structures were noted and were annotated as follows: Sidechain interactions: ◆ NMe₃ ▶ CO₂⁻ ◆ SO₃⁻ ● Thione ★ imidazole and backbone interactions: ✧ NMe₃ ▷ CO₂⁻ ◇ SO₃⁻ ○ Thione ☆ imidazole (Figure 23). These interactions were then analyzed, looking for high conservation within the specific type, yet no conservation between the two groups. This analysis afforded several residues which we propose are indicative of binding EGT or EGTSO₃⁻ binding, highlighted in yellow.

Structural Characterization of Ergothioneine Solute Binding Proteins

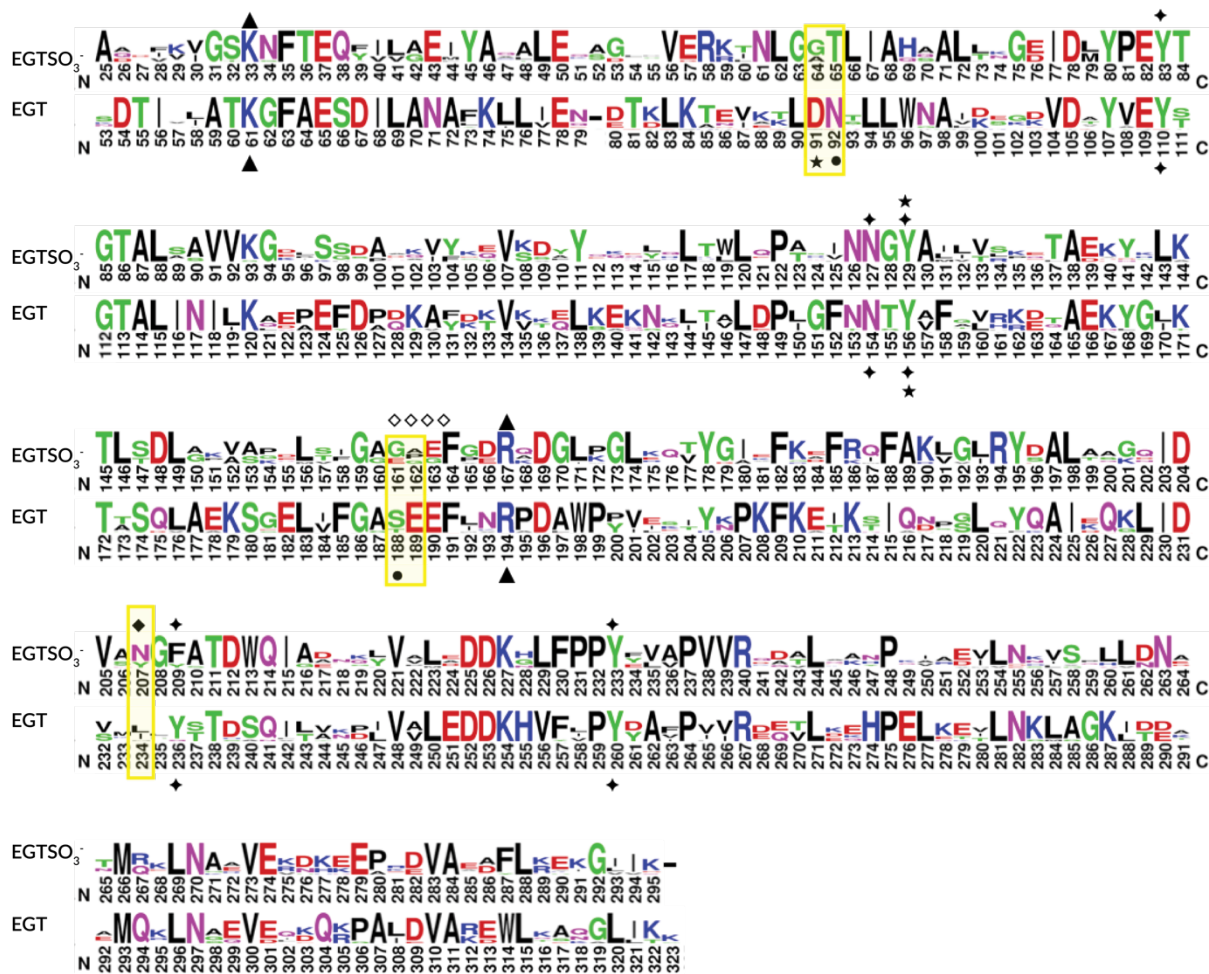


Figure 23. Alignment of sequence logos for EGTSO₃⁻ and EGT binding SBPs. Sequence numbers correspond to gene numbering of SBP5 and SBP14 for the EGTSO₃⁻ and EGT binders respective. Sidechain interactions: ◆ NMe₃ ▲ CO₂ ◆ SO₃⁻ ● Thione ★ imidazole and backbone interactions: ◇ NMe₃ △ CO₂ ◇ SO₃⁻ ○ Thione ☆ imidazole. Yellow box indicates residues proposed to be an indicative motif.

In the structural analysis, Asp91 and Asn92 were identified as potential indicators of EGT binding. Both residues are conserved in the EGT-binding logo, yet have no basis in the EGTSO₃⁻ acid binders. In place of Asp91 and Asn92, the EGTSO₃⁻ binders have a conserved Gly/Ala-Thr motif, highlighting these residues as indicators of preferred solute. In the EGTSO₃⁻ SBPs, it is proposed that Asn207 is indicative of sulfonic acid binding. This residue is largely conserved among the sulfonic acid binders. However, some sequences contain a Tyr at this position, yet Tyr could also facilitate a hydrogen bond. At the equivalent position, the EGT binders contain a non-polar residue, which is unable to facilitate any hydrogen-bonding interaction and therefore contributes to the substrate selectivity. The β6-α7 loop of SBP4 and 5 is unique to the sulfonic acid binders, wrapping around EGTSO₃⁻ presumably in response to the sulfonate head group. This loop is not comprised of a strictly-conserved motif. However, this region always contains at least one glycine residue, if not several. There also appears to be covariance between positions 161 and 163 of this loop, with a glycine and glutamate at each of the positions in the EGTSO₃⁻ SBPs. At this position EGT SBPs have a conserved Ser/Thr188 & Glu189 motif. In summary, the combination of these residues from this structural and bioinformatic analysis highlights the following residues as indicators of EGT or EGTSO₃⁻. These residues are represented

schematically in a sequence LOGO containing the residues identified as being a substrate selectivity determinant (Figure 24).

- EGTSO₃⁻ binding: G/A64, T65, N/Y207 and E/G161-G/E163.
- EGT binding: D91, N92, S/T188 and E189.

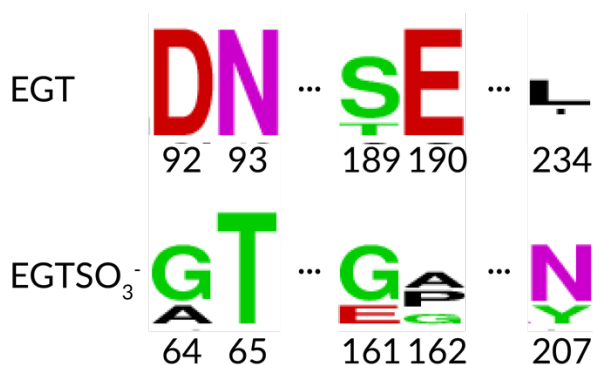


Figure 24. Sequence logo of residues important for substrate selectivity.

Sequence Similarity Network (SSN)

With these sequence motifs in hand, we examined the clustering of SBPs within a sequence similarity network (SSN). SSNs allow for the visualization of relationships among protein sequences. In such a network the most related proteins are grouped together in a cluster. Combined with the use of UniProt Reference Clusters (UniRef) as sequence inputs, we can obtain complete coverage of sequence space to introduce a quantitative aspect to a bioinformatic assessment.²³⁴ Using this approach, we wanted to determine if our hypotheses on the SBP structural determinants are consistent with a much larger and unbiased set of sequences, rather than hand-selected examples. Use of SSN can also leverage our selectivity determinants to answer the following questions: (i) What type of lyases are in the close genomic environment of the EGTSO₃⁻ SBPs? (ii) What can we learn about the evolutionary origins of EGT-related SBPs and diversity between the two types? (iii) Are there lyases in the close genomic environment of the EGT SBPs?

(i) What type of lyases are in the close genomic environment of the Ergothioneine-sulfonic acid SBPs?

This chapter has shown that at least two EGT-related SBPs exist, which can be grouped by their preferred substrate (EGT or EGTSO₃⁻). Our group recently discovered the same typology applies to the ergothionases, with the discovery of EGTSO₃⁻-specific lyases (type II ergothionases). The sequence motifs that distinguish the two types (type I -EGT, type ii - EGTSO₃⁻) have also been established. This work was carried out by Mariia Beliaeva and Alice Mauer and will be published in conjunction with the work in this chapter. Based off this finding, we were intrigued to determine if the EGTSO₃⁻ SBPs are co-encoded with type II lyases. To explore this, a SSN was generated using pBlast of the 2000 closest SBP5 homologues. These 2000 sequences also included the EGT binding SBP13 & 14 as out groups to ensure that we have a complete coverage of all sulfonic acid binding sequences. SBP13 is another EGT-binding SBP, yet no structure exists, which has been

included in this analysis as an additional reference point for an EGT SBP. Sequences with lengths greater than 380 were removed from the dataset leaving 910 sequences. An alignment score of 90 (corresponding to ~41% pairwise identity) produced fractionation in which SBP4 & 5 are in the same cluster, yet not in the same cluster as the EGT SBPs SBP13 & 14 (Figure 25). The class containing SBP4 and SBP5 was denoted as the EGTSO₃⁻ binding cluster (orange Figure 25) and contains 82 sequences (A full list and description is given in Supplementary Table 1. To confirm that these are EGTSO₃⁻ binders, a multiple sequence alignment of all 82 sequences revealed conservation of all of the sequence motifs indicative of sulfonic acid binding. This validates our hypothesis on the structural determinants of EGTSO₃⁻ binding. The sequences in this cluster all come from Proteobacteria and are primarily of the orders Rhizobiales and Enterobacterales. Two species are from the order Rhodobacterales. To explore the genomic environment, this list of 82 sequences was reduced to 32 representative sequences by removing sequences with greater than 90% sequence similarity. The presence and position of a lyase relative to the SBP gene was noted (± 10 genes), and classified as a type I (EGT-specific) or type II (EGTSO₃⁻-specific) lyase through the presence of sequence motifs. The genomic environment for 31 of the 32 sequences could be analysed. Of these 31 sequences, three SBP genes (10%) did not contain a lyase with ± 10 genes. The remaining 28 (90%) contained a type II lyase in close proximity. Of these 28, three species (10%) contained a second SBP and type I Lyase specific for EGT. The full results of this analysis are listed in Supplementary Table 2.

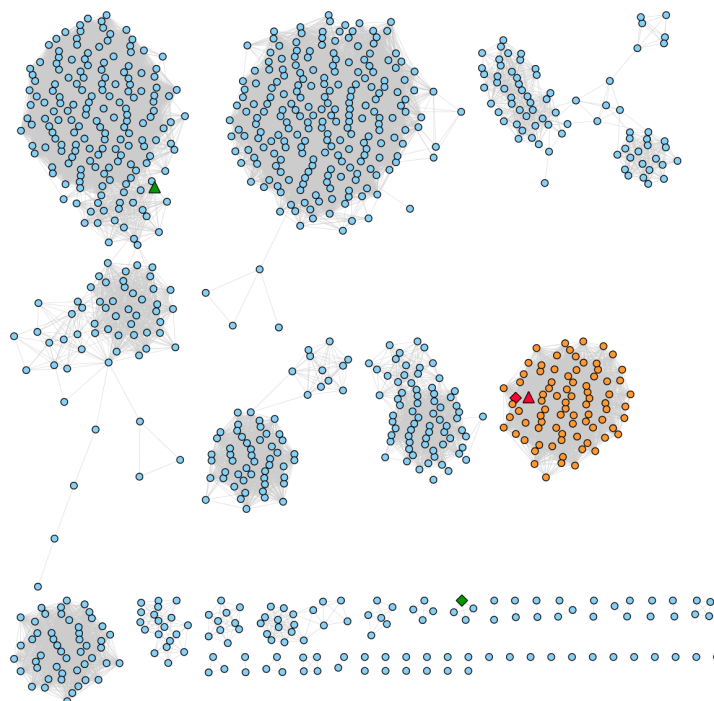


Figure 25. SSN generated of 910 sequences from a pblast of SBP5 homologues, separated with an alignment score of 90. The sulfonic acid binding cluster (orange) contains 82 sequences of EGTSO₃⁻ binders. Characterized homologues are labelled as follows: EGT SBPs in green, SBP14: diamond and SBP13: triangle. EGTSO₃⁻ SBPs in red, SBP5: diamond and SBP4: triangle SBP4.

This analysis very clearly indicates that the majority (90 %) of the EGTSO₃⁻ specific SBPs contain a type II lyase in close genomic environment, indicating that organisms that take up EGTSO₃ typically degrade it. It also confirms the predictive power of the sequence motifs proposed in the previous section.

(ii) What can we learn about the evolutionary origins of Ergothioneine-related SBPs and diversity among the two types?

The application of the previous strategy to the EGT SBPs binders is much more complex because (1) many bacterial species may uptake EGT yet not degrade it and (2) there appears to be much greater sequence diversity amongst the EGT SBPs than amongst the sulfonic-acid binding SBPs. Difficulties immediately transpire from the latter in attempts to generate a SSN with a sequence set that will encompass all EGT SBPs and some sulfonic acid binders as outgroups. A pBlast of 10000 sequences (the SSN maximum) using SBP14 as an input does not include any sulfonic acid binders. As a result, it is difficult to ensure a complete set of EGT SBPs. Therefore, a different approach was taken to acquire sequences for SSN generation: sequences were derived from members of an InterPro family.

IPR007210. All the SBPs characterized by Mariia with a high affinity ($K_D < 50$ nM) for EGT or a derivative (SBP-4,5,13 and 14) belong to the interpro family IPR007210. The IPR007210 family is the substrate-binding domain of ABC-type GB transport system and encompasses 46,647 sequences (accessed 29.10.2019). The use of an interpro family ensures complete coverage of the SBP sequence space. Of these sequences, 15 are reviewed by Swiss-Prot and have manual annotation. Crystal structures of eight family members also exist in the PDB. The literature was reviewed for each candidate (see Supplementary Table 3) and a list of 11 well-characterized IPR007210 family members with high affinity ligands (K_D or $K_M < 50$ μ M) was curated. This list includes the OpuAC,²²⁴ OpuBC¹²⁻¹³ and OpuCC^{227, 235-238} SBPs of *Bacillus subtilis*, the OpuAC of *Lactococcus lactis*²³⁹ and the OpuCC of *Listeria monocytogenes*²⁴⁰⁻²⁴² that bind and transport GB, choline, PB and carnitine. ProX from *E. coli*³⁴⁻³⁶ and archaea *A. fulgidus*²⁴³⁻²⁴⁴ bind GB, as do OusX from *Erwinia chrysanthemi*²⁴⁵ and GbuC from *Listeria monocytogenes*²⁴⁶⁻²⁴⁸. ChoX from *Rhizobium meliloti* is a choline specific SBP²⁴⁹, while TauA from *E. coli* binds taurine with very high affinity.²⁵⁰ In addition to these published members, we have verified function for a further four SBPs from the IPR007210 family; two sulfonic acid binders (SBP4 & 5) and two EGT binders (SBP13 & 14). All the characterized SBPs act as reference points in the SSN. The use of an interpro family ensures complete coverage of the sequence space.

IPR007210 SSN Generation. A SSN network was generated from the RefID90 database for the IPR007210 family. In UniRef90, the sequences that share over 90% sequence identity over 80% of the sequence length are grouped together and represented by a single sequence.²⁵¹ For the IPR007210 family, the use of UniRef90 reduces the number of sequences in the dataset from 47140 to 15,758. Of these sequences, only those with sequence lengths between 230 and 400 amino acids were used (Supplementary Figure 1). This length filtering excluded one of the characterized examples, OpuAC of *Lactococcus lactis*, and reduced that data set to 12,095 sequences. An alignment score of 78 corresponds to sequence identity of approximately

47 %, results in fractionation in which the EGT-related SBPs (SBP4,5,13&14) are clustered together yet are in distinct clusters from literature characterized homologues (Supplementary Figure 2).

Alignment score of 78. The SSN with an alignment score threshold of 78 places the characterized IPR007210 members into seven distinct clusters (Figure 26B). The EGT SBPs (SB13 and 14) and EGTSO₃⁻ binders (SBP4 & 5) sit within the same cluster, which is loosely grouped in contrast to the other clusters. As EGT and EGTSO₃⁻ SBPs are in the same cluster, it suggests a close evolutionary relationship. The relationships of the different clusters were explored by increasing and decreasing the alignment threshold (Figure 26A&C). This primarily affected the EGT-related cluster, while the other six clusters were largely unaffected.

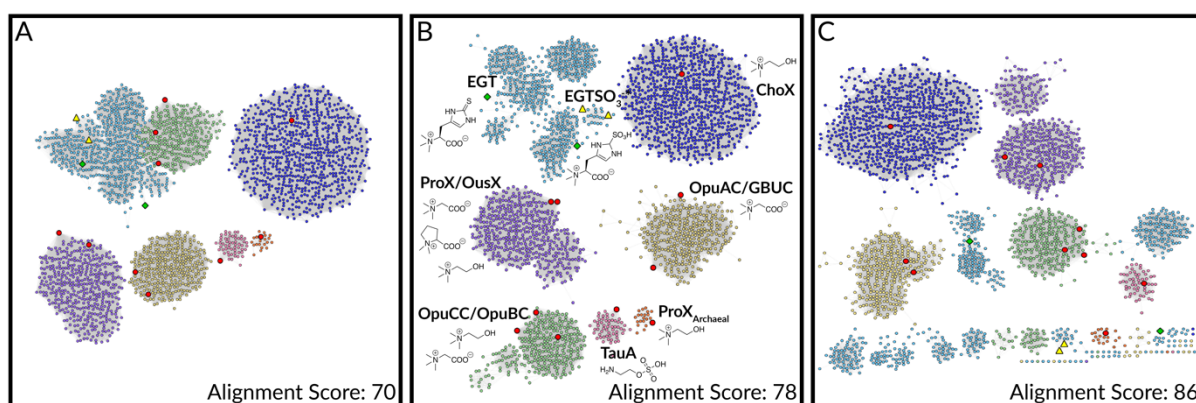


Figure 26. SSN of IPR007210 RefID90 family members (75% RefID shown). Only clusters with characterized homologues are shown (see Figures 36-38 for full SSN). Members from the literature are shown as red points, in B these are labelled, and their preferred ligand(s) shown. EGT and EGTSO₃⁻ binders characterized in this work are shown as green diamonds and yellow triangles respectively.

The most dramatic effect of increasing the alignment score is the splitting of the EGT SBP cluster into at least nine significant clusters (Figure 26C). In contrast, the clusters of other characterized SBPs are largely unaffected. At this alignment score, the two characterized EGT SBPs (SBP13 and 14) sit within different clusters, while the EGTSO₃⁻ SBPs form a distinct cluster. These results indicate there is much more diversity among the EGT binders than within the SBPs that bind EGTSO₃⁻ and other betaine binding SBPs. This greater diversity suggests that EGT SBPs have a long evolutionary history, having accumulated substantial sequence diversity through neutral drift over a long time. This is consistent with the suggestion that EGT is an ancient molecule.³² In contrast, the EGTSO₃⁻ SBPs sit within a distinct and well-defined cluster with less sequence diversity. It is likely all the EGTSO₃⁻ binders evolved from a single EGT binding ancestor. This cluster maps well to the EGTSO₃⁻ cluster in the previous bioinformatic study (Q(i), page 177). The agreement between SSNs generated using different sequence inputs validates our approach.

Decreasing the alignment score results in 6 major clusters; the EGT-related SBPs merge with the OpuCC/OpuCC to form one large cluster (Figure 26C). Another discrete cluster contains ProX_{archaeal}, which is described in the literature as a GB SBP. However, ProX_{archaeal} also binds EGT with a high affinity ($K_D(\text{EGT})$: 56

nM, $K_D(\text{GB})$: 110 nM (Beliaeva & Seebeck, unpublished). As $\text{ProX}_{\text{archaeal}}$ does not cluster with other characterized EGT-binding proteins at an alignment score of 70, yet unrelated betaine binders do, it is likely that EGT binding SBPs evolved twice from different evolutionary origins. These findings suggest that the general theme of diversity and multiple evolutionary emergences in EGT biosynthesis may also apply to proteins involved in the utilization and degradation of EGT.

.(iii) Are there Lyases in the close genomic environment of the Ergothioneine SBPs?

The EGTSO_3^- SBPs are commonly co-encoded with a type II lyase. We were interested to see if the abundance of the EGT SBPs are co-encoded with a lyase. The EGT SBPs appear much more prevalent and diverse. To explore the genomic environment of EGT SBPs, the sequences identified as being EGT related SBPs from the IPR007210 SSN at an alignment score of 78 (colored blue) were taken and separated into nine major clusters at the alignment score of 86 (Figure 27). The genomic environment for each of these clusters was analyzed through input into Genome Neighborhood Tool (GNT). For each cluster, the co-occurrence of an aromatic lyase ± 10 genes either side of the SBP gene is given and annotated (Figure 27). Of the nine major clusters, only three have significant occurrence of a lyase in the close genomic environment. In the EGTSO_3^- binding cluster (purple), containing SBP4&5, 86 of 97 sequences (89 %) contain a lyase in the close genomic environment 89%. This is very consistent with the analysis afforded for Q(i) (90% co-occurrence was found). In the largest cluster (red), containing SBP13, 33 of 328 sequences contain a lyase in the close genomic environment (10 %), while the cluster containing SBP14 also has a higher percentage of lyases in genomic environment with a co-occurrence of 13 out of 17 sequences (76 %). The lyases corresponding to these latter groups have not been classified. This shows that a large proportion of organisms that utilize EGT do not degrade it. This provides evidence that EGT is an important molecule, even for organisms that (1) do not produce EGT and (2) cannot degrade EGT. This identifies numerous organisms in which EGT may play a physiological role, indicating that EGT chemistry is much more prevalent in nature than originally hypothesized.

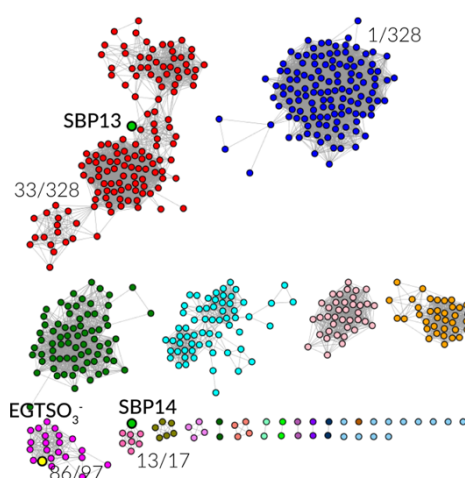


Figure 27. Co-occurrence of an aromatic lyase and SBP for each of the EGT related SBPs at an alignment score threshold of 86. Labels indicate co-occurrence of an aromatic lyase with the SBP.

Conclusion

Herein we describe the characterisation of the SBP components of an ABC transporter system for EGT and EGTSO₃⁻. High-resolution crystal structures of EGT and EGTSO₃⁻ specific SBP complexed with ligands have been solved. These atomic resolution structures allowed for recognition of the residues involved in substrate binding, and in combination with a bioinformatic analysis provide a fingerprint for EGT and EGTSO₃⁻ binding in the SBPs. These studies provide evidence that the oxidized analogue, EGTSO₃⁻ is indeed a relevant degradation product in nature. Further leverage of these structural determinants revealed that EGT SBPs are diverse and are not commonly encoded with an ergothionase. This identifies organisms which (1) do not product EGT, (2) cannot degrade EGT, yet (3) have an ABC system for its transport, thereby dramatically increasing the organisms to which we know EGT is relevant, highlighting the ubiquity and prevalence of EGT in nature.

Chapter 5:
Structural Characterization of Ergothioneine Solute Binding Proteins
Supplementary Information

Supplementary Information

Supplementary Table 1 Organism name, Uniprot ID, phylum and order of the 82 EGTSO₃⁻ binders from SSN analysis (Figure 25).

Organism	UniProt ID	Phylum	Order
Rhizobium sp. NFR03.	A0A1H9WZ83	Proteobacteria	Rhizobiales
Rhizobium leguminosarum bv. viciae (strain 3841)	Q1M367	Proteobacteria	Rhizobiales
Rhizobium sp. NFR07.	A0A1I0VTX2	Proteobacteria	Rhizobiales
Rhizobium pisi.	A0A427N3G9	Proteobacteria	Rhizobiales
Brenneria sp. CFCC 11842.	A0A2U1UCT5	Proteobacteria	Enterobacterales
Chelativorans sp. (strain BNC1)	Q11E26	Proteobacteria	Rhizobiales
Citrobacter koseri	A0A427PID5	Proteobacteria	Enterobacterales
Mesorhizobium sp. YR577.	A0A1I7EFP7	Proteobacteria	Rhizobiales
Rhizobium leguminosarum.	A0A444IPI3	Proteobacteria	Rhizobiales
Rhizobium leguminosarum.	A0A444HS03	Proteobacteria	Rhizobiales
Rhizobium leguminosarum bv. viciae.	A0A222U5L6	Proteobacteria	Rhizobiales
Rhizobium leguminosarum.	A0A444ND25	Proteobacteria	Rhizobiales
Roseovarius sp. TM1035.	A6E628	Proteobacteria	Rhodobacterales
Rhizobium etli 8C-3.	A0A1L5PI31	Proteobacteria	Rhizobiales
Bosea sp. 32-68-6.	A0A258D508	Proteobacteria	Rhizobiales
Bosea sp. RCAM04685.	A0A370L1I2	Proteobacteria	Rhizobiales
Enterobacter sp. 10-1.	A0A267RJN0	Proteobacteria	Enterobacterales
Paracoccus aminophilus JCM 7686.	S5XNL6	Proteobacteria	Rhodobacterales
Citrobacter koseri (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696)	A8AKL8	Proteobacteria	Enterobacterales
Phyllobacterium bourgognense.	A0A368YFG5	Proteobacteria	Rhizobiales
Bosea sp. 12-68-7.	A0A257ZK53	Proteobacteria	Rhizobiales
Agrobacterium radiobacter (strain K84 / ATCC BAA-868)	B9JID4	Proteobacteria	Rhizobiales
Agrobacterium rhizogenes NBRC 13257.	A0A061MY32	Proteobacteria	Rhizobiales
Rhizobium leguminosarum bv. trifolii WSM1689.	W0IR34	Proteobacteria	Rhizobiales
Citrobacter koseri	A0A381H7D7	Proteobacteria	Enterobacterales
Rhizobium sp. L9.	A0A2A6KSN5	Proteobacteria	Rhizobiales
Rhizobium sp. L43.	A0A2A6HRU7	Proteobacteria	Rhizobiales
Klebsiella pneumoniae.	A0A377VMY4	Proteobacteria	Enterobacterales
Rhizobium sp. H4.	A0A2A6R4N7	Proteobacteria	Rhizobiales
Citrobacter koseri	A0A078LCQ8	Proteobacteria	Enterobacterales
Rhizobium sp. WYCCWR10014.	A0A198YHL2	Proteobacteria	Rhizobiales
Bosea sp. LC85.	A0A085EW29	Proteobacteria	Rhizobiales
Rhizobium jaguaris.	A0A387FMD6	Proteobacteria	Rhizobiales
Mesorhizobium sp. DCY119.	A0A3A3HP34	Proteobacteria	Rhizobiales
Agrobacterium rhizogenes.	A0A071I930	Proteobacteria	Rhizobiales
Aureimonas flava.	A0A3A1WN43	Proteobacteria	Rhizobiales
Rhizobium leguminosarum bv. viciae.	A0A1Q8HLL0	Proteobacteria	Rhizobiales
Rhizobium leguminosarum bv. trifolii.	A0A1B8R9I8	Proteobacteria	Rhizobiales
Rhizobium leguminosarum.	A0A1B1CIA7	Proteobacteria	Rhizobiales
Bosea sp. RAC05.	A0A1B3NKL2	Proteobacteria	Rhizobiales
Rhizobium lusitanum.	A0A1C3XES6	Proteobacteria	Rhizobiales

Structural Characterization of Ergothioneine Solute Binding Proteins

<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> .	A0A2L1CP29	Proteobacteria	Rhizobiales
<i>Rhizobium leguminosarum</i> .	A0A2K9ZE43	Proteobacteria	Rhizobiales
<i>Marteella endophytica</i> .	A0A0D5LL55	Proteobacteria	Rhizobiales
<i>Rhizobium laguerreae</i> .	A0A1S9GZC8	Proteobacteria	Rhizobiales
<i>Marteella mediterranea</i> DSM 17316.	A0A1U9Z8G6	Proteobacteria	Rhizobiales
<i>Salmonella</i> sp. HMSC13B08.	A0A1F2K6J4	Proteobacteria	Enterobacterales
<i>Roseovarius mucosus</i> .	A0A1V0RMQ5	Proteobacteria	Rhodobacterales
<i>Bosea</i> sp. BIWAKO-01.	A0A1E1V194	Proteobacteria	Rhizobiales
<i>Rhizobium loessense</i> .	A0A1G4QW84	Proteobacteria	Rhizobiales
<i>Mesorhizobium ephedrae</i> .	A0A2P7RPG1	Proteobacteria	Rhizobiales
<i>Rhizobiales</i> bacterium.	A0A2W4SYJ6	Proteobacteria	Rhizobiales
<i>Mesorhizobium amorphae</i> .	A0A2W5BL08	Proteobacteria	Rhizobiales
<i>Phyllobacterium endophyticum</i> .	A0A2P7ANQ4	Proteobacteria	Rhizobiales
<i>Phyllobacterium zundukense</i> .	A0A2N9VTN0	Proteobacteria	Rhizobiales
<i>Phyllobacterium brassicacearum</i> .	A0A2P7BER5	Proteobacteria	Rhizobiales
<i>Phyllobacterium sophorae</i> .	A0A2P7B9Y0	Proteobacteria	Rhizobiales
<i>Rhizobium</i> sp. CF142.	J2L8I1	Proteobacteria	Rhizobiales
<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM2297.	J0KYX8	Proteobacteria	Rhizobiales
<i>Rhizobium</i> sp. AP16.	J2WFE3	Proteobacteria	Rhizobiales
<i>Phyllobacterium</i> sp. YR531.	J2VG25	Proteobacteria	Rhizobiales
<i>Trabulsiella odontotermitis</i> .	A0A0L0GWC6	Proteobacteria	Enterobacterales
<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM597.	I9NEV0	Proteobacteria	Rhizobiales
<i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM2012.	J0C0E3	Proteobacteria	Rhizobiales
<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> (strain WSM1455)	J0B3W4	Proteobacteria	Rhizobiales
<i>Phyllobacterium myrsinacearum</i> .	A0A2S9JQ57	Proteobacteria	Rhizobiales
<i>Citrobacter koseri</i>	A0A3S4IH65	Proteobacteria	Enterobacterales
<i>Rhizobium leguminosarum</i> .	A0A2Z4YNM0	Proteobacteria	Rhizobiales
<i>Salmonella enterica</i> I.	A0A402QTJ3	Proteobacteria	Enterobacterales
<i>Phyllobacterium</i> sp. OV277.	A0A1H0UIN0	Proteobacteria	Rhizobiales
<i>Rhizobium anhuiense</i> .	A0A3S0SSK1	Proteobacteria	Rhizobiales
<i>Rhizobium</i> sp. NXC14.	A0A1W6Q350	Proteobacteria	Rhizobiales
<i>Rhizobium leguminosarum</i> .	A0A3S4AEK8	Proteobacteria	Rhizobiales
<i>Rhizobium</i> sp. 24NR.	A0A3S3SBG1	Proteobacteria	Rhizobiales
<i>Bosea lupini</i> .	A0A1H7TC34	Proteobacteria	Rhizobiales
<i>Citrobacter koseri</i>	A0A336Q393	Proteobacteria	Enterobacterales
<i>Bosea</i> sp. AAP35.	A0A0N1LDZ1	Proteobacteria	Rhizobiales
<i>Hoeflea marina</i> .	A0A317PSX8	Proteobacteria	Rhizobiales
<i>Bosea vaviloviae</i> .	A0A0N1F3W2	Proteobacteria	Rhizobiales
<i>Rhizobium acidisoli</i> .	A0A0N1MCN8	Proteobacteria	Rhizobiales
<i>Bosea lathyri</i> .	A0A1H6C5Q0	Proteobacteria	Rhizobiales

Supplementary Table 2. Representative Sequences of EGTSO₃⁻ cluster (Organism name & Uniprot ID) and notation of lyase in genomic environment. Relative location in genome is noted.

Organism	UniProt ID	First Lyase ± 10 genes			Second Lyase ± 10 genes			Represented Sequences
		Location	UniProtID	Type	Location	UniProtID	Type	
Agrobacterium radiobacter (strain K84 / ATCC BAA-868)	B9JID4	-1	B9JID5	II	-10	B9JIE4	I	
Rhizobium lusitanum.	AOA1C3XE56	-1	AOA1C3XE19	II				
Rhizobium jaguaris.	AOA387FMD6	None found +- 10 genes						
Rhizobium etli 8C-3.	AOA1L5PI31	-1	AOA1L5PHN5	II				Rhizobium sp. NXC14.
Aureimonas flava.	AOA3A1WN43	+1	AOA3A1WNB2	II				
Paracoccus aminophilus JCM 7686.	S5XNL6	-2	S5XUM9	II				
Bosea sp. RCAM04685.	AOA370L1I2	+1	AOA370L0D9	II				Bosea sp. AAP35., Bosea sp. RAC05., Bosea lathyri., Bosea sp. 12-68-7. Bosea sp. 32-68-6., Rhizobiales bacterium.
Bosea vaviloviae.	AOA0N1F3W2	+1	AOA0N0MC47	II				
Rhizobium sp. 24NR.	AOA3S3SBG1	+6	AOA3S3VTV3	II				
Phyllobacterium sp. OV277.	AOA1H0UIN0	+7	AOA1H0UID8	II				Phyllobacterium myrsinacearum.
Phyllobacterium zundukense.	AOA2N9VTN0	+6	AOA2P7RPG6	II				Phyllobacterium endophyticum., Phyllobacterium sophorae., Phyllobacterium brassicacearum., Phyllobacterium bourgognense., Phyllobacterium sp. YR531.
Mesorhizobium sp. YR577.	AOA1I7EFP7	+6	AOA1I7EFU9	II				Mesorhizobium sp. DCY119.
Marteella mediterranea DSM 17316.	AOA1U9Z8G6	-1	AOA1U9Z848	II				
Hoeflea marina.	AOA317PSX8	-1	AOA317PQ12	II				
Rhizobium acidisoli.	AOA0N1MCN8	+6	AOA0N1DQ59	II				Rhizobium leguminosarum., Rhizobium leguminosarum bv. trifolii WSM597., Rhizobium leguminosarum bv. trifolii WSM2297., Rhizobium sp. CF142. Salmonella sp. HMSC13B08., Enterobacter sp. 10-1., Klebsiella pneumoniae., Citrobacter koseri (strain ATCC BAA-895 / CDC 4225-83 / SGSC4696)
Citrobacter koseri	AOA078LCQ8	+1	AOA078LHJ6	II				
Trabulsiella odontotermitis.	AOA0LOGWC6	+1	AOA0LOGXG1	II				Rhizobium leguminosarum. Rhizobium leguminosarum bv. trifolii., Rhizobium loessense., Rhizobium leguminosarum bv. viciae., Rhizobium laguerreae., Rhizobium sp. L43., Rhizobium sp. L9., Rhizobium sp. H4., Rhizobium anhuiense, Rhizobium pisi., Rhizobium leguminosarum
Rhizobium sp. WYCCWR10014.	AOA198YHL2	-1	AOA198YF95	II				

Structural Characterization of Ergothioneine Solute Binding Proteins

Agrobacterium rhizogenes NBRC 13257.	A0A061MY32	-1	A0A061MWB1	II	-10	A0A061MY23	I	bv. viciae (strain WSM1455), Rhizobium leguminosarum bv. trifolii WSM2012., Rhizobium leguminosarum bv. trifolii WSM1689.
Chelativorans sp. (strain BNC1)	Q11E26	+1	Q11E25	II				Agrobacterium rhizogenes, Rhizobium sp. AP16.
Marteella endophytica.	A0A0D5LL55	-1	A0A0D5LLV5	II	8	Q11E18	I	
Roseovarius mucosus.	A0A1V0RMQ5	+1	A0A1V0RMQ0	II				Roseovarius sp. TM1035.
Bosea sp. LC85.	A0A085EW29	+1	A0A085EW28	II				Bosea sp. BIWAKO-01.
Bosea lupini.	A0A1H7TC34	+1	A0A1H7TC09	II				
Mesorhizobium ephedrae.	A0A2P7RPG1	None found +- 10 genes						
Salmonella enterica I.	A0A402QTJ3	None found (+- 10 genes unusual genome)						
Rhizobium sp. NFR03.	A0A1H9WZ83	+6	A0A1H9WZC8	II				Rhizobium sp. NFR07.
Citrobacter koseri	A0A3S4IH65	+1	A0A3S5DP65	II				
Rhizobium leguminosarum bv. viciae.	A0A2L1CP29	-1	A0A2L1CPB6	II				Rhizobium leguminosarum bv. viciae (strain 3841)
Brenneria sp. CFCC 11842.	A0A2U1UCT5	+1	A0A2U1UCU2	II				
Mesorhizobium amorphae.	A0A2W5BL08	-2	A0A2W4ZPB6	II				

Supplementary Table 3. Characterized Members of the IPR007210 family and their binding affinities for various ligands with references. Boxes highlighted in green indicates a ligand with significant binding (K_D or $K_M < 50 \mu\text{M}$). Those in grey either substrates are unknown or have low affinities and have not been included in our bioinformatic analysis.

Entry Name	Name	Organism	Ligand 1		Ligand 2		Ligand 3		Struct.	Notes	Ref
			Ligand	K_D (μM)	Ligand	K_D (μM)	Ligand	K_D (μM)			
Reviewed (Swiss-Prot) - Manually annotated UniProtKB:											
PROX_ECOLI	P0AFM2	<i>Escherichia coli</i> (strain K12)	Glycine Betaine	0.7 ²³⁰ 1.4 ²²⁸	Proline Betaine	5.2 ²²⁹	NA	NA	1R9Q 1R9L ¹⁹⁸		198, 228, 230, 252
GBUC_LISM4	Q9RR44	<i>Listeria monocytogenes</i> serotype 1/2a (strain 10403S)	Glycine Betaine	K_M : 4.4 ²⁴⁶ K_M : 2-8 ²⁴⁷ K_M : 6 ²⁴⁸	Carnitine	4000 ²⁴⁸	NA	NA	NA		246- 248, 253- 255
OPUBC_BACSU	Q45462	<i>Bacillus subtilis</i> (strain 168)	Choline	30 ²²⁶ 180 ²⁵⁶ K_M : 1 ²²⁷	Arsenocholine	2100 ²⁵⁶	NA	NA	3R6U ²²⁶ 5NXY ²⁵⁶ 6EYG 6EYH 6EYL 6EYQ		226- 227, 256
OPUCC_BACSU	O32243	<i>Bacillus subtilis</i> (strain 168)	Glycine Betaine	K_M : 2.8 ²³⁸ K_M : 5.1 ²³⁷⁻²³⁸	Carnitine	K_M : 5 ²³⁶	Choline	38 μM ²²⁷	3PPN -R ²²⁵	Broad substrate range. Other Ligands: - Choline-O-sulfate- K_M - 4 μM ²³⁸ - other substrates in mM	225, 227, 236- 238
YEZH_ECOLI	P33362	<i>Escherichia coli</i> (strain K12)	Glycine Betaine	2700 ²⁵⁷					4WE p ²⁵⁷	Unknown ligand	257
TAUA_ECOLI	Q47537	<i>Escherichia coli</i> (strain K12)	Taurine	0.011 ²⁵⁰					NA	Gene cluster is necessary for grow with taurine, ²⁵⁸	250, 258
OUSX_DICD3	E0SCY3	(<i>Erwinia chrysanthemi</i> (strain 3937))	Glycine Betaine	1.6 ²⁴⁵	Choline	2 ²⁴⁵			NA		245
OSMX_SALTY	Q8ZPK2	<i>Salmonella typhimurium</i>	Choline-O-sulfate	"Higher than GB" ²⁵⁹	Glycine Betaine	"low" ²⁵⁹			NA	Growth of WT cells vs an OsuM transporter mutant ²⁵⁹	259
OPUAC_BACSU	P46922	<i>Bacillus subtilis</i> (strain 168)	Glycine Betaine	17 ²⁶⁰	Proline Betaine	295 ²⁶⁰			2B4 M ²⁶⁰ 2B4L ²⁶⁰		260- 261
PROX_SALTY	Q8ZML1	<i>Salmonella typhimurium</i>	Annotated by similarity to ProX from <i>E. coli</i> , no affinities found.								
OPUCC_LISM4	G2JZ42	<i>Listeria monocytogenes</i> serotype 1/2a	Same as Q9KHT7								
PROX_ECO57	P0AFM3	<i>Escherichia coli</i> O157:H7	Annotated by similarity to ProX from <i>E. coli</i> , no affinities found.								
OPUCC_LISMN	Q9KHT7	<i>Listeria monocytogenes</i>	Glycine Betaine	K_M : 10 ²⁴⁰⁻²⁴¹	Carnitine	K_M : 10 ²⁴⁰⁻²⁴²					240- 242
COSB_HALS3	B0R6A8	<i>Halobacterium salinarum</i>	Vague.								
ORF1_CHRSD	Q9L4K3	<i>Chromohalobacter salexigens</i>	Function inferred from homology.								
IPR007210 Members with a PDB Entry:											
Q92N37_RHI ME	Q92N37	<i>Rhizobium meliloti</i>	Choline	2.7 ²⁴⁹	Acetyl Choline	145 ²⁴⁹			2REG ²⁶² 2REJ ²⁶² 2RIN ²⁶²		249, 262

Structural Characterization of Ergothioneine Solute Binding Proteins

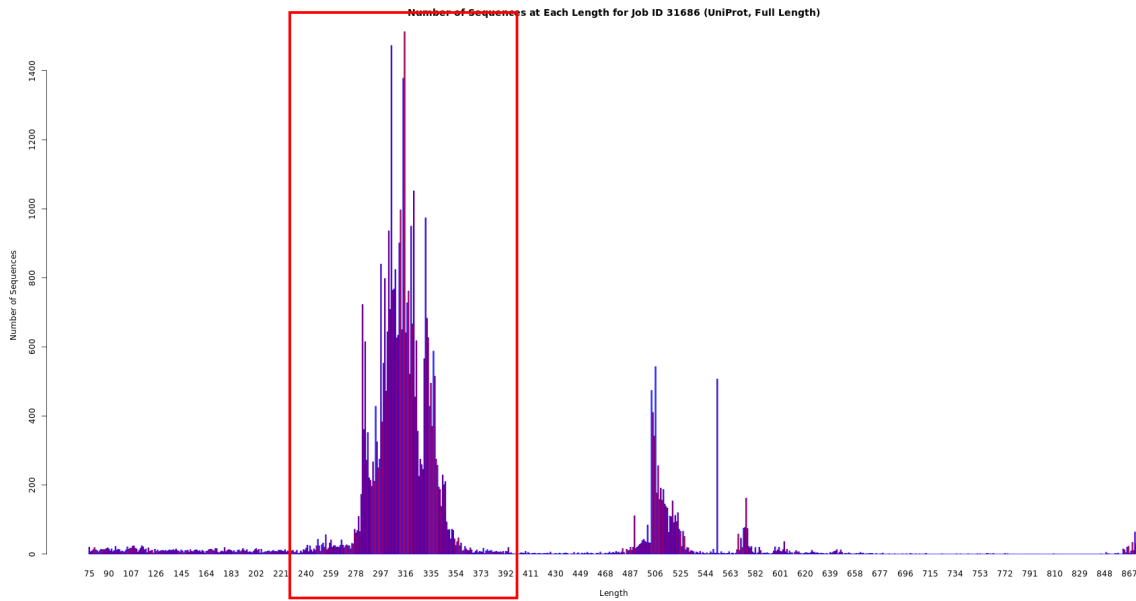
Q7DAU8_LAC LA	Q7DAU8	<i>Lactococcus lactis</i>	Glycine Betaine	4-6 ²³⁹	Proline Betaine	**				WRF 1 ²⁶²⁻²⁶³ 3HC Q ²⁷ 3L6G ²³⁹ 3L6H ²³⁹	** inhibition studies K _i of 41 uM ²³⁹	239
A0A0H3K0Z1	A0A0H3K0Z1	<i>Staphylococcus aureus</i>								3O66		-
O29280_ARCF U	O29280	<i>Archaeoglobus fulgidus</i>	Glycine Betaine	0.06	Proline Betaine	0.05	EGT	0.016*		1SW1 ⁻² ²⁴³ 1SW4 ⁻⁵ ²⁴³ 3MA M ²⁴⁴	EGT ₃ : 0.033 uM	226, 243
Unpublished IPR007210 members characterised by Maria:												
SBP4	A0A0D5LL55	<i>Marteella endophytica</i>	EGT ₃	0.008	EGT	35	TMH	23				
SBP5	A0A0LOGW C6	<i>Trabulsiella odontotermitis</i>	EGT ₃	0.011	EGT	32						
SBP13	A0A0C2R5J 3	<i>Cohnella kolymensis</i>	EGT	0.020	MeEGT	0.46 & 2						
SBP14	A0A1B8WB Y4	<i>Bacillus sp. FJAT-27264</i>	EGT	0.02	MeEGT	0.026	EGT ₃	0.172			TMH - 5 uM	

Genes used for EGT₃ binders sequence Logo (based off SBP5)

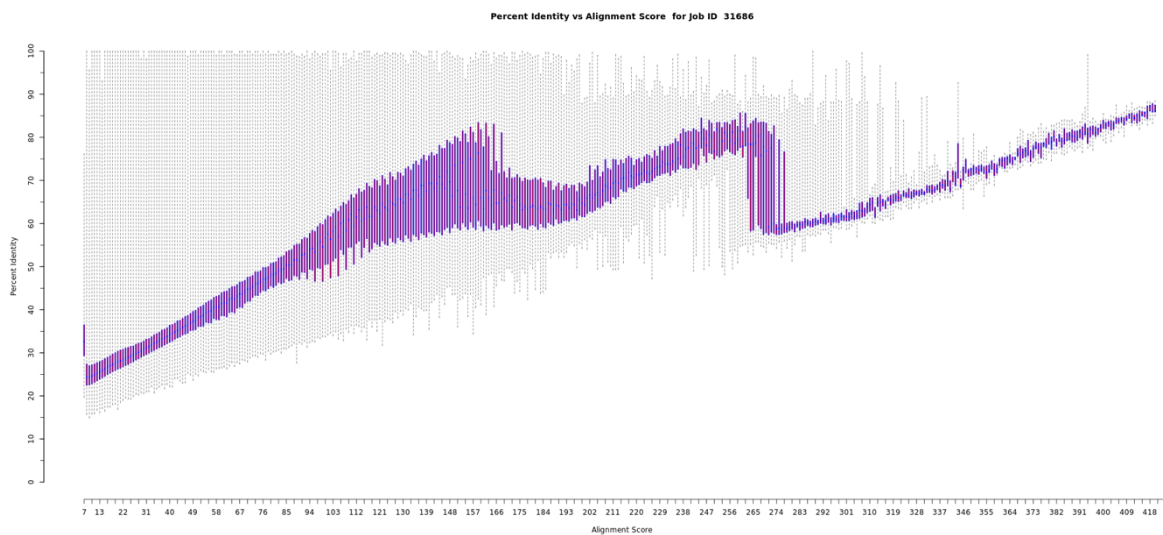
WP_049857012.1, WP_045679504.1, WP_049857012.1, WP_047458689.1, WP_046852135.1,
 WP_017800827.1, WP_081506961.1, WP_119539141.1, WP_105733597.1, WP_112524158.1,
 WP_115055625.1, WP_110030243.1

Genes used for EGT binders sequence Logo (based off SBP14)

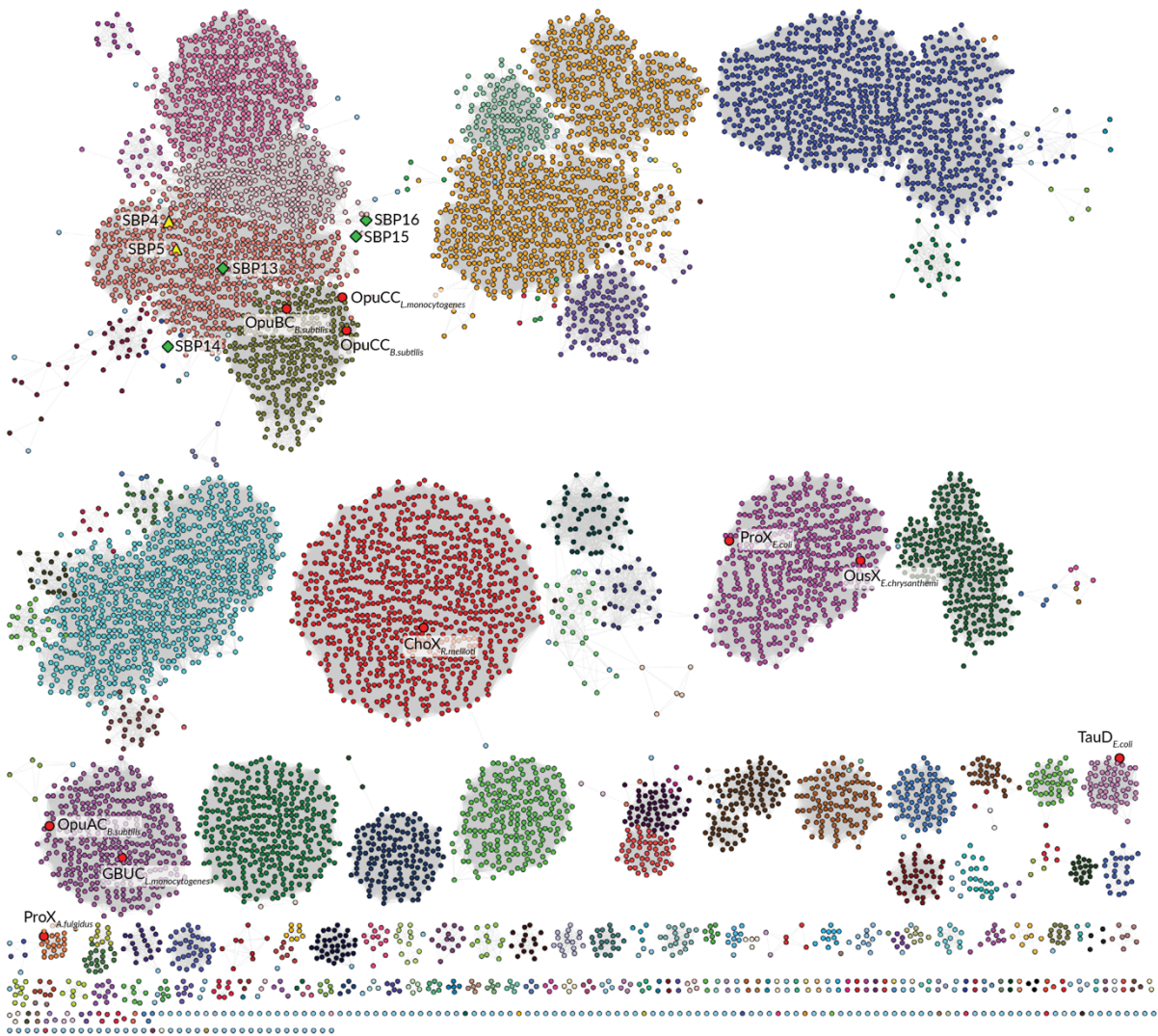
WP_066368978.1, WP_053782311.1, WP_090949738.1, WP_099854310.1, WP_020432100.1,
 OPH57908.1, WP_079413395.1, WP_110771752.1, WP_053372609.1, WP_051237002.1



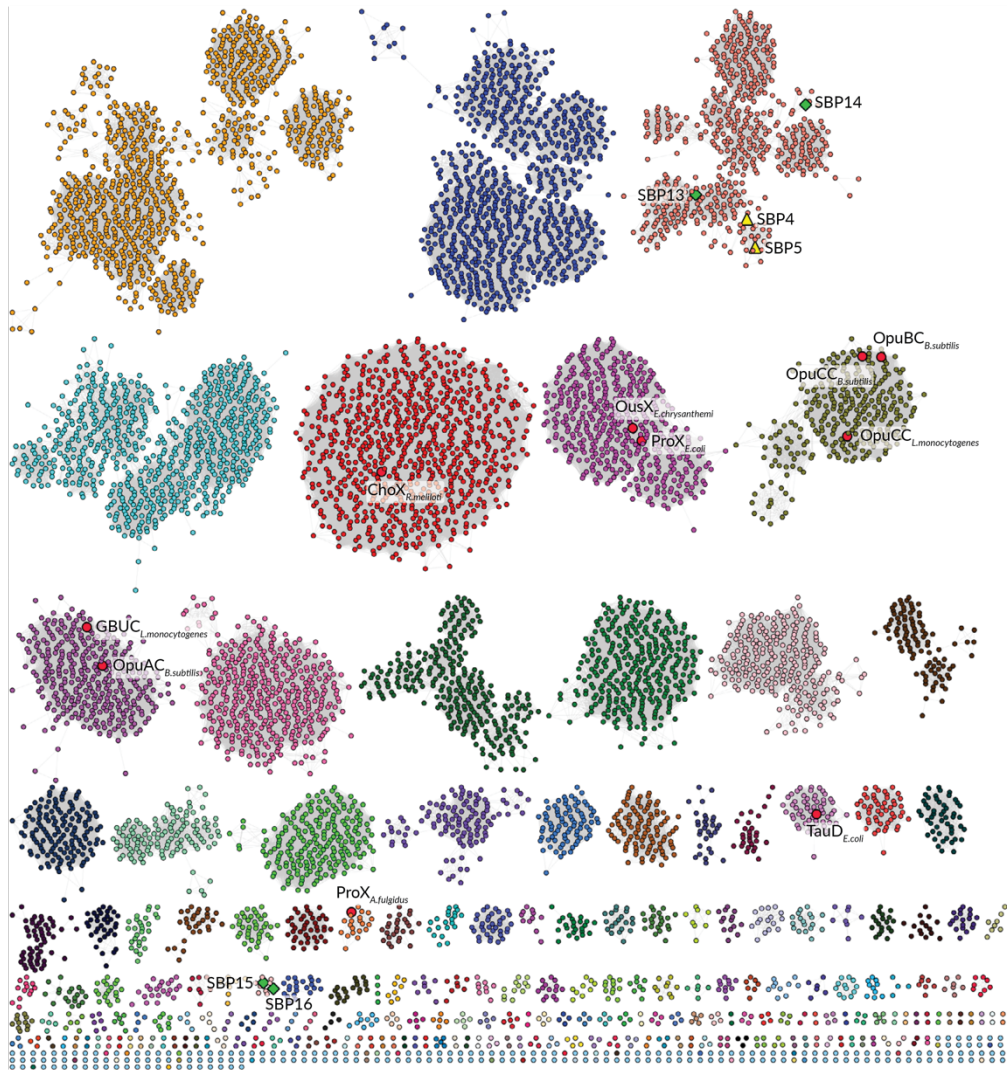
Supplementary Figure 1. Histogram for the length distribution for all sequences used in the generation of a SSN using the Ref90 of IPR007210 family. The red box indicates the selection criteria for SSN generation, sequences less than 230 aa or greater than 400 amino acids were excluded from network generation.



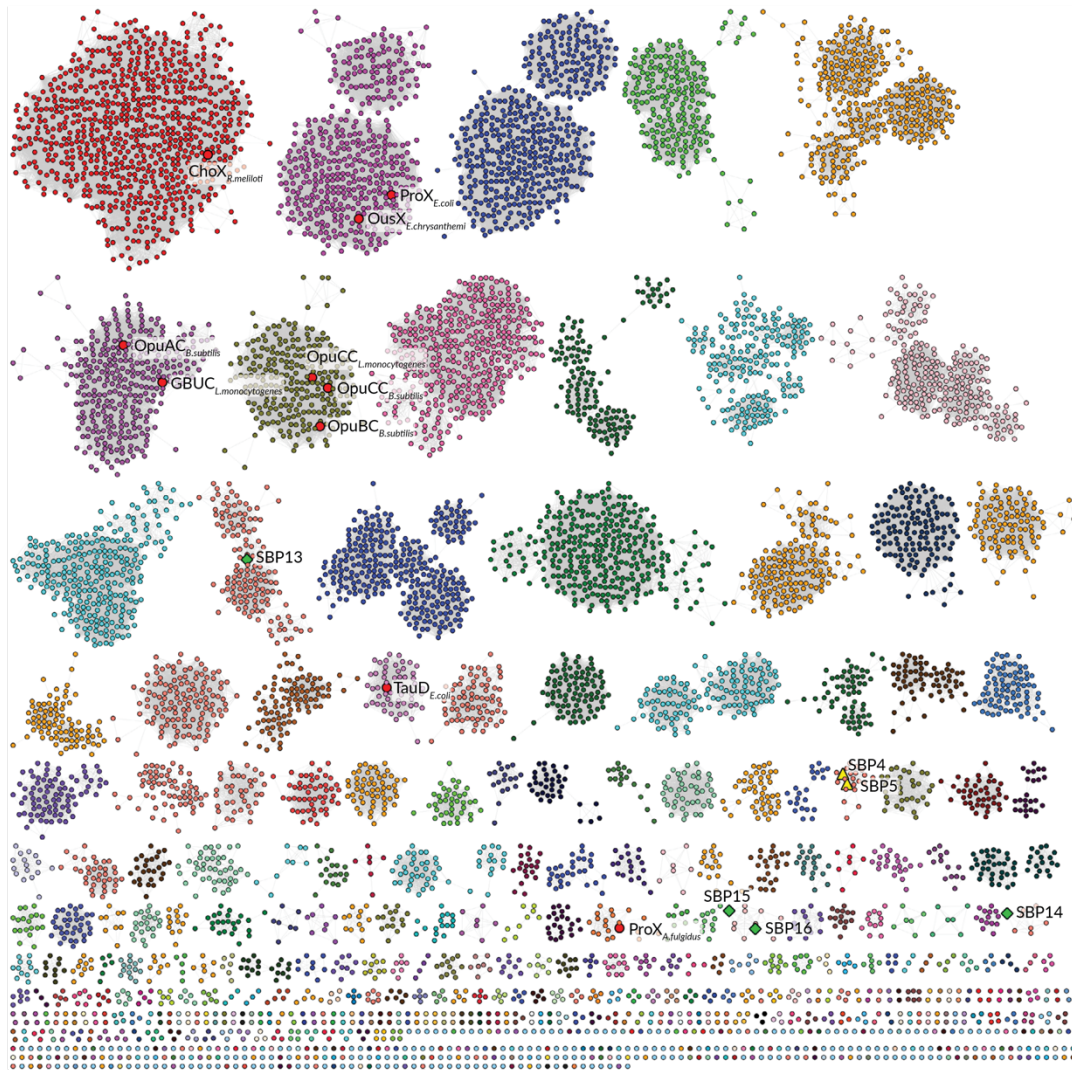
Supplementary Figure 2. Pairwise percent sequence identity as a function of alignment score for the SSN generated from the 90RefID generated from family IPR007210.



Supplementary Figure 3. SSN of RefID90 of family IPR007210. Alignment score of 70, coloured by clustering at an alignment score of 78. Shown as a 75 % Ref ID.



Supplementary Figure 4. SSN of RefID90 of family IPR007210. Alignment score of 78. Shown as a 75 % Ref ID.



Supplementary Figure 5. SSN of RefID90 of family IPR007210. Alignment score of 86, coloured by clustering at an alignment score of 78. Shown as a 75 % Ref ID.

Experimental

Proteins

SBP4 – Solute Binding Protein from *Marteella endophytica* (WP_045679504.1)

MGHHHHHAENLYFQGSADPIKVGSKNFTEQFILGEMYAILLEDAGYDVDRKINLGGTLIAHQALTTG
EIDLYPEYTG TALNSVVKGEVSSDADAVYKEVSDYYLNEGLVWLEPTGINNGYAIIVRQDTADEYGL
KTLSDLGKVS GDLVFGGGPEFPDRADGLPGLKAVYDAEFKEFKQFAKLGLRYDALMQGDIDVANGYAT
DWQIGA EGLVPLDDDKGLFPPYYVAPVVRKDVLDANPGMADVLNSLAPHVDNAV MRKLN AEVEKDHEE
PEDVARAYLEEIGLIKN

$m/z(\text{SBP4}_{\text{tev_cleaved}})$: calc.: 29802.37 Da, meas.: 29801.89 Da

$\epsilon_{280}(\text{SBP4}_{\text{tev_cleaved}})$: 33350 $\text{M}^{-1} \text{cm}^{-1}$

SBP5 – Solute Binding Protein from *Trabulsiella odontotermitis* (WP_049857012.1)

MGHHHHHAENLYFQSAADIVVGSKNFTEQYILAEIYARTLEQSGLSVERKTNLGGTLIAHAALLKG
EIDMYPEYTG TALS AVVKGK PSSGAAQVFKQVSDAYQQYHLLTWLQPARVNNGYALLVSKQTAEKYHL
KTLSDLARAPELTI GAGAEFGDRQDGLKGLEQTYGITFKAFRQFAKVGLRYDALAAGQIDVANGFAT
DWQIAENQYVALEDDKHLFPPYE VAPVVRDATLQKYPQIREIVNNVSQLLDNATMQKLN AEVEKNKEE
PRDVAEDFLREKGI IK

$m/z(\text{SBP5}_{\text{tev_cleaved}})$: calc.: 30115.00 Da, meas.: 30114.73 Da

$\epsilon_{280}(\text{SBP5}_{\text{tev_cleaved}})$: 30370 $\text{M}^{-1} \text{cm}^{-1}$

SBP14 – Solute Binding Protein from *Bacillus* sp. FJAT-27264 (WP_066368978)

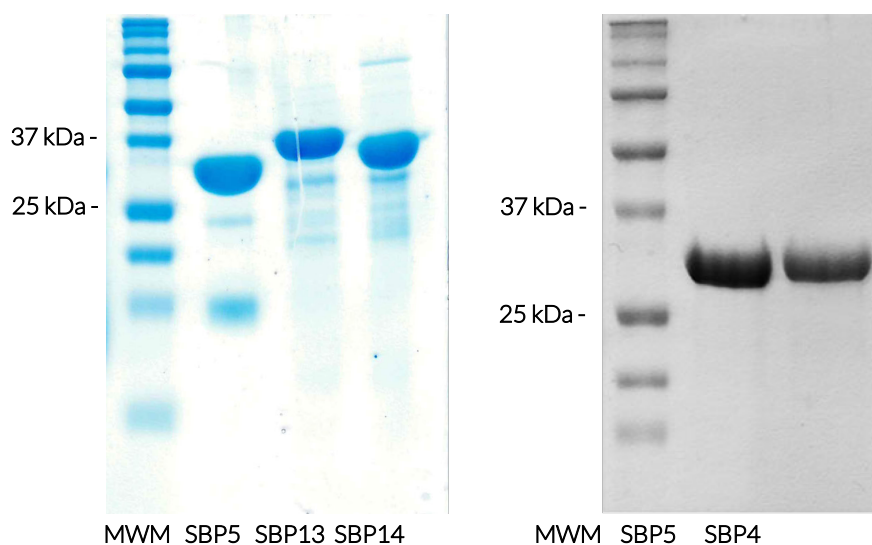
MGHHHHHAENLYFQGSGCSNSKASEGKDTIILATKGF AESDILANAFKLLIENDTKLKTEVKTL DNT
LLWNAIDSGD VDAYVEYS GTALINILKQQPEFDPDKAFKTVVVTQLKEKNKLI ALDPLGFNNTYVFSVR
KDTAEKYGLKTTSQLAEKSGELVFGASEEFLKRPDAWPYVESIYKPKFKETKSIQNSSLQYQAIKQKL
IDVMLAYSTDSQILANDLVALEDDKHVFPYDAFPV VREQTLDEHPELKETLNKLAGKLDESAMQKLN
AEVEQDQKPAIDVAREWLKSQGLIK

$m/z(\text{SBP14}_{\text{tev_cleaved}})$: calc.: 31547.79 Da, meas.: 31545.78 Da

$\epsilon_{280}(\text{SBP14}_{\text{tev_cleaved}})$: 29910 $\text{M}^{-1} \text{cm}^{-1}$

Protein Purification

The SBPs were produced, purified by Ni-NTA beads, the his₆ tag was cleaved off, and a second Ni-NTA purification was carried out by Mariia. These proteins were then further purified by size exclusion chromatography (S300) using 200 mM NaCl, 50 mM Tris, pH 8. Fractions containing pure SBP were pooled and concentrated to at least 35 mg/mL before immediate use in crystallization experiments. SDS PAGE Gel of samples used for protein crystallization show a strong band corresponding to the correct molecular weight (Supplementary Figure 6)).



Supplementary Figure 6. SDS PAGE Gel of SBP4, SBP5, SBP13 & SBP14 used for crystallization. MWM: molecular weight marker Left Gel: 10 µg are loaded of each protein.

Crystallization

SBP4

The SBP4 EGT-containing structure was obtained through co-crystallization. SBP4 (20 mg/mL) was incubated with EGT (4 mM) for at least 30 minutes before crystallization. Initial crystallization conditions were determined with the vapor diffusion method in a sitting drop 96-well format. Drops were set up using a dispensing robot (Crystal Gryphon, Art Robbins), mixing different ratios of SBP4-EGT solution with reservoir solution (0.2 μ L:0.1 μ L, 0.1 μ L:0.2 μ L and 0.2 μ L:0.2 μ L) and were equilibrated against 30 μ L reservoir solution. The screens were stored at 20 °C in an automated imaging system (Formulatrix).

In drop 3 (2 μ L protein : 2 μ L reservoir solution) with condition A12 of PACT *premier*[™] HT-96 (Molecular Dimensions) a tiny crystal appeared within a few hours after plate set up and continued to grow in size for another 3 days until dimensions of approximately 400 μ M x 90 μ M were reached (Supplementary Figure 7) A12 conditions comprising 0.01 M Zinc chloride, 0.1 M Sodium acetate, pH 5 and 20 % w/v PEG 6000. 30 % glycerol were used as cryoprotectant for the single crystal, which was fished and flash-cooled, stored in liquid nitrogen and taken to the synchrotron for data collection.



Supplementary Figure 7. SBP5 crystal from A12 of PACT *premier*[™] HT-96 (Hampton research) crystallization screen. EGT is co-crystallized. The image was taken 3 days following plate set. The crystal is approximately 400 μ M x 90 μ M in size.

SBP5

The SBP5 structures in complex with EGT and EGTSO_3^- were obtained through co-crystallization. Initial crystallization conditions were determined with the vapor diffusion method in a sitting drop 96-well format. Drops were set up using a dispensing robot (Crystal Gryphon, Art Robbins), mixing different ratios of SBP5-ligand solution with reservoir solution (0.2 μL :0.1 μL , 0.1 μL :0.2 μL and 0.2 μL :0.2 μL) and were equilibrated against 30 μL reservoir solution. The screens were stored at 20 °C in an automated imaging system (Formulatrix).

SBP5 + EGT

SBP5 (25 mg/mL) was incubated with EGT (4 mM) for at least 30 minutes before crystallization.

In drop 2 (1 μL protein : 2 μL reservoir solution) with condition A9 of PEG ION HT (Hampton Research) a tiny crystal appeared 12 hours after plate set up and continued to grow in size for another 4.5 days until dimensions of approximately 170 μM x 90 μM were reached (Supplementary Figure 8). A9 conditions comprising 0.2 M Ammonium chloride, 20% (w/v) Polyethylene glycol 3,350, pH 6.3. 30% glycerol were used as cryo protectant for the single crystal, which was fished and flash cooled, stored in liquid nitrogen and then taken to the synchrotron for data collection.



Supplementary Figure 8. SBP5 crystal from A9 of PEG ION HAT (Hampton research) crystallization screen. EGT is co-crystallized. The image was taken 5 days following plate set up, once the crystal stopped growing. The crystal is approximately 170 μM x 90 μM in size.

SBP5 + EGTSO_3^-

Crystals of SBP5 containing EGTSO_3^- grew in G4 of the Morpheus® HT-96 screen (Molecular Dimensions) containing 0.1 M carboxylic acids (0.2M Sodium formate; 0.2M Ammonium acetate; 0.2M Sodium citrate tribasic dihydrate; 0.2M Sodium potassium tartrate tetrahydrate; 0.2M Sodium oxamate), 0.1 M Buffer system 1, pH 6.5 (Imidazole; MES monohydrate (acid)) & 50 % Precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350). Upon mixing of 0.1 μL of SBP5-Ergthioneine-sulfonic acid mixture (28 mg/mL thawed SBP5 + 1.05 mM EGTSO_3^-) with 0.2 μL of the crystallization condition, crystals appeared two weeks after screen set up. No photo was taken. The crystals were then harvested and stored in $\text{N}_2(\text{l})$ until data collection.

SBP14 Crystallization

The SBP14 liganded structures were obtained through co-crystallization. SBP14 (28 mg/mL) was incubated with 1.2 equivalents of ligand (1.05 mM) for at least 30 minutes before crystallization. Crystallization conditions were identified using the vapor diffusion method in a sitting drop 96-well format. Drops were set up using a dispensing robot (Crystal Gryphon, Art Robbins), mixing different ratios of the SBP14-ligand solution with reservoir solution at three different ratios (0.2 μ L:0.1 μ L, 0.1 μ L:0.2 μ L and 0.2 μ L:0.2 μ L) and were equilibrated against 30 μ L reservoir solution. The screens were stored at 20 °C in an automated imaging system (Formulatrix).

SBP14 + EGT

Crystals of SBP14 containing EGT grew in E12 of the Morpheus® HT-96 screen (Molecular Dimensions) containing 0.12 M ethylene glycols (0.3M Diethylene glycol; 0.3M Triethylene-glycol; 0.3M Tetraethylene glycol; 0.3M Pentaethylene glycol), 0.1 M Buffer system 3, pH 8.5 (Tris (base); BICINE) & 37.5 % Precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350) (Supplementary Figure 9). Upon mixing of 0.2 μ L of SBP14-EGT mixture (28 mg/mL fresh SBP14 + 1.05 mM EGT) with 0.2 μ L of the crystallization condition, crystals appeared within 12 hours of screen set up. These cubic crystals continued to grow for another 12 hours with dimensions of approximately 60 μ m before the crystals were harvested using 25 % glycerol as a cryoprotectant and were stored in N₂(l) until data collection.



Supplementary Figure 9. Crystal of SBP14 in complex with EGT from E12 of Morpheus® HT-96 screen (Molecular Dimensions) crystallization screen. The image was taken 1 day after plate set up. The crystal is approximately 60 μ m x 60 μ m in size.

SBP14 + EGTSO₃⁻

Crystals of SBP14 containing EGTSO₃⁻ grew in F8 of the Morpheus® HT-96 screen (Molecular Dimensions) containing 0.12 M Monosaccharides (0.2M D-Glucose; 0.2M D-Mannose; 0.2M D-Galactose; 0.2M L-Fucose; 0.2M D-Xylose; 0.2M N-Acetyl-D-Glucosamine), 0.1 M Buffer system 2, pH 7.5 (Sodium HEPES; MOPS (acid)) & 37.5 % Precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350) (Supplementary Figure 10). Upon mixing of 0.2 μ L of SBP14-EGTSO₃⁻ mixture (28 mg/mL thawed SBP14 + 1.05 mM EGTSO₃⁻) with 0.1 μ L of the crystallization condition, crystals appeared within 12 hours of screen set up. These crystals continued to grow for 48 hours before the crystals were harvested and stored in N₂(l) until data collection.

Structural Characterization of Ergothioneine Solute Binding Proteins



Supplementary Figure 10. Crystal of SBP14 in complex with EGTSO_3^- from F8 of Morpheus® HT-96 screen (Molecular Dimensions) crystallization screen. The image was taken 1 day following plate set up. The crystal is approximately $50\ \mu\text{M}$ x $50\ \mu\text{M}$ in size.

SBP14 + MeEGT

Crystals of SBP14 containing MeEGT grew in G4 of the Morpheus® HT-96 screen (Molecular Dimensions) containing 0.1 M carboxylic acids (0.2M Sodium formate; 0.2M Ammonium acetate; 0.2M Sodium citrate tribasic dihydrate; 0.2M Sodium potassium tartrate tetrahydrate; 0.2M Sodium oxamate), 0.1 M Buffer system 1, pH 6.5 (Imidazole; MES monohydrate (acid)) & 50 % Precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350) (Supplementary Figure 11). Upon mixing of $0.2\ \mu\text{L}$ of SBP14-MeEGT mixture (28 mg/mL fresh SBP14 + 1.05 mM MeEGT) with $0.2\ \mu\text{L}$ of the crystallization condition, crystals appeared within 12 hours of screen set up. These crystals continued to grow for another 12 hours before the crystals were harvested using 25 % glycerol as a cryoprotectant and were stored in $\text{N}_2(\text{l})$ until data collection.



Supplementary Figure 11. Crystal of SBP14 in complex with MeEGT from G4 of Morpheus® HT-96 screen (Molecular Dimensions) crystallization screen. The image was taken 3 days following plate set up. The crystal is approximately $40\ \mu\text{M}$ x $50\ \mu\text{M}$ in size.

SBP14 + TMH (sls20181031/1242/8)

Crystals of SBP14 containing TMH grew in B4 of the Morpheus® HT-96 screen (Molecular Dimensions) containing 0.09 M halogens (0.3M Sodium fluoride; 0.3M Sodium bromide; 0.3M Sodium iodide), 0.1 M Buffer system 1, pH 6.5 (Imidazole; MES monohydrate (acid)) & 37 % Precipitant mix 4 (25% v/v MPD; 25% PEG 1000; 25% w/v PEG 3350) (Supplementary Figure 12). Upon mixing of $0.2\ \mu\text{L}$ of SBP14-TMH mixture (28 mg/mL thawed SBP14 + 1.05 mM TMH) with $0.2\ \mu\text{L}$ of the crystallization condition, crystals appeared within 24 hours of screen set up. These crystals continued to grow for 2.5 days before the crystals were harvested and stored in $\text{N}_2(\text{l})$ until data collection.



Supplementary Figure 12. Crystal of SBP14 in complex with TMH from B4 of Morpheus® HT-96 screen (Molecular Dimensions) crystallization screen. The image was taken 3 days following plate set up. The crystal is approximately 50 μM x 50 μM in size.

Data collection, data processing, structure solution and refinement

Data of the various solute-binding proteins were collected at the Swiss Light Source (SLS), Villigen, Switzerland on either the Xo6DA (PXIII) beamline using a Pilatus 2M-F detector, or the X06SA (PXI) beamline using an EIGER 16M (Dectris) detector. The collected diffraction data were indexed and integrated using XDS¹⁵⁷ and were scaled using aimless¹⁵⁸.

The SBP5 EGT-containing structure was solved by molecular replacement using a search model derived from four published structures of SBP5 analogues (PDB: 3PPN, 4Z7E, 1SW4, 3R6U) using PHENIX-Phaser.^{159, 225-226, 243, 264} The initial model of SBP5 was built using AUTOBUILD of the PHENIX package.¹⁶⁰ The SBP5 EGTSO₃⁻ structure was solved by molecular replacement using the SBP5-EGT structure. The SBP4 in complex with EGT containing structure was solved by molecular replacement using the SBP5 EGT-containing structure as a search model using PHENIX-Phaser.^{159, 225-226, 243, 264} The initial model of SBP4 was built using AUTOBUILD of the PHENIX package.¹⁶⁰ The SBP14 MeEGT-containing structure was solved by molecular replacement using the SBP5 EGT containing structure as a search model using PHENIX-Phaser and an initial model was built in AUTOBUILD.^{37, 159, 226, 243, 264, 160} All subsequent SBP14 structures used the SBP14 MeEGT as a search model using PHENIX-Phaser.^{159, 225-226, 243, 264}

Several rounds of iterative model building and refinement were performed using Coot¹⁶¹ and Refmac¹⁶² or PHENIX¹⁶³. Ligand restraints were prepared on the Grade Web server, from Global Phasing Ltd. and could be unambiguously modelled into the ($F_o - F_c$) difference density map.²¹⁵ 5% of the data was excluded from refinement and used for cross-validation. Data collection and refinement statistics are summarized in Table 1 & 2 respectively. Figures were prepared in PyMOL.

Structural Characterization of Ergothioneine Solute Binding Proteins

Supplementary Table 4a. Data collection for structures of the various Solute Binding Proteins.

	SBP4 + EGT	SBP5 + EGT	SBP5 + EGTSO ₃ ⁻
X-ray Source	Xo6DA (PXIII)	XO6DA (PXIII)	XO6SA (PXI)
X-Ray detector	PILATUS 2MF	PILATUS 2MF	EIGER 16M (Dectris)
Wavelength (Å)	0.98	0.98	1.00
Space group	P 21 21 21	P 1 21 1	P 21 21 21
Cell dimensions a, b, c (Å)	39.7, 53.7, 127.4	46.0, 45.45, 61.9	43.8, 107.6, 228.0
Cell Angles α, β, γ (°)	90 90 90	90 103.6 90	90, 90, 90
Solvent content (%)	46	41	43
Molecules in asymmetric unit	1	1	4
Resolution limits (Å)	37.9-1.4 (1.45-1.4)	44.7-1.2 (1.24-1.2)	42.0-1.94 (2.00-1.94)
R _{merge} [†]	0.09914 (0.8078)	0.03414 (0.1509)	0.1383 (1.805)
R _{meas} [‡]	0.1083 (0.8759)	0.04105 (0.1977)	0.1439 (1.876)
CC ½	0.998 (0.797)	0.999 (0.957)	0.999 (0.765)
<I/σ(I)>	12.37 (2.10)	18.03 (4.15)	13.19 (1.60)
Total reflections	342416 (35257)	228055 (12924)	1075789 (107393)
Unique reflections	53924 (5292)	75189 (7385)	80953 (7969)
Multiplicity	6.3 (6.7)	3.0 (2.0)	13.3 (13.5)
Completeness	98.75 (97.66)	98.79 (95.82)	99.88 (99.81)
Mosaicity	0.17	0.27	0.15

[†] $R_{\text{merge}} = \sum hkl \sum i |I_i(hkl) - \overline{I(hkl)}| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection and $\overline{I(hkl)}$ is the average intensity obtained from multiple observations of symmetry-related reflections.

[‡] $R_{\text{meas}} = \sum hkl [N/(N-1)]^{1/2} \sum i |I_i(hkl) - \overline{I(hkl)}| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection, $\overline{I(hkl)}$ is the average intensity obtained from multiple observations of symmetry-related reflections and N is the number of observations of intensity $I(hkl)$.

Supplementary Table 4b. Data collection for structures of the various Solute Binding Proteins.

	SBP14 + EGT	SBP14 + MeEGT	SBP14 + EGTSO ₃ ⁻	SBP14 + TMH
X-ray Source	X06DA (PXIII)	X06SA (PXI)	X06DA (PXIII)	X06SA (PXI)
X-Ray detector	PILATUS 2MF	EIGER 16M (Dectris)	PILATUS 2MF	EIGER 16M (Dectris)
Wavelength (Å)	1.00	1.00	1.00	1.00
Space group	P 21 3	P 21 3	P 21 3	P 21 3
Cell dimensions a, b, c (Å)	98.2, 98.2, 98.2	98.4, 98.4, 98.4	98.6, 98.6, 98.6	98.7, 98.7, 98.7
Cell Angles α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Solvent content (%)	51	51	51	51
Molecules in asymmetric unit	1	1	1	1
Resolution limits (Å)	43.9 - 1.6 (1.66 - 1.6)	44.0 - 1.84 (1.91 - 1.84)	49.3 - 1.9 (1.97 - 1.9)	44.1 - 2.0 (2.07 - 2.0)
R _{merge} [†] (%)	0.09101 (2.217)	0.1171 (4.577)	0.1121 (2.623)	0.0983 (3.96)
R _{meas} [‡]	0.09218 (2.246)	0.1187 (4.634)	0.1135 (2.658)	0.09958 (4.008)
CC ½	1 (0.736)	0.999 (0.537)	1 (0.765)	1 (0.531)
<I/ σ (I)>	35.51 (2.04)	19.30 (0.94)	33.32 (1.82)	24.04 (1.20)
Total reflections	1671237 (162598)	1085797 (112905)	1016011 (97492)	881799 (90252)
Unique reflections	41820 (4134)	27881 (2778)	25484 (2529)	21925 (2166)
Multiplicity	40.0 (39.3)	38.9 (40.6)	39.9 (38.5)	40.2 (41.6)
Completeness	99.98 (99.98)	99.99 (100.00)	99.53 (96.64)	99.95 (99.95)
Mosaicity	0.10	0.21	.20	0.11

[†] $R_{\text{merge}} = \sum hkl \sum i |I_i(hkl) - \overline{I(hkl)}| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection and $\overline{I(hkl)}$ is the average intensity obtained from multiple observations of symmetry-related reflections.

[‡] $R_{\text{meas}} = \sum hkl [N/(N-1)]^{1/2} \sum i |I_i(hkl) - \overline{I(hkl)}| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity for a reflection, $\overline{I(hkl)}$ is the average intensity obtained from multiple observations of symmetry-related reflections and N is the number of observations of intensity $I(hkl)$.

Structural Characterization of Ergothioneine Solute Binding Proteins

Supplementary Table 5a. Refinement Statistics for structures of the various Solute Binding Proteins

	SBP4 + EGT	SBP5 + EGT	SBP5 + EGTSO ₃ ⁻
Resolution limits (Å)	37.9-1.4 (1.45-1.4)	44.7-1.2 (1.24-1.2)	42.0-1.94 (2.00-1.94)
Rwork *	0.1694 (0.2364)	0.1529 (0.2175)	0.2018 (0.2607)
Rfree **	0.1987 (0.2733)	0.1788 (0.2417)	0.2402 (0.3058)
Number of non-H atoms	2563	2597	8847
macromolecules	2120	2164	8497
ligands	30	15	72
solvent	413	418	278
Protein residues	271	271	1086
Clashscore ***	4.97	3.43	9.56
R.m.s.d from ideal			
Bond lengths (Å°)	0.008	0.01	0.006
Bond angles (u)	1.27	1.37	1.04
Ramachandran favored *** (%)	97.4	97.77	96.75
Ramachandran outliers *** (%)	0	0	0.37
Average B values (Å ²)	18.87	16.11	36.17
macromolecules	16.68	13.48	36.08
ligands	18.9	8.63	29.86
solvent	30.1	29.96	40.61

Numbers in parentheses refer to the outer shell.

* $R_{work} = \sum |hkl| |F_{obs} - F_{calc}| / \sum |hkl| F_{obs}$

** Rfree is the R value calculated for 5% of the data set that was not included in the refinement.

*** Molprobit.

Supplementary Table 5b. Data collection for structures of the various Solute Binding Proteins.

	SBP14 + EGT	SBP14 + MeEGT	SBP14 + EGTSO ₃ ⁻	SBP14 + TMH
Resolution limits (Å)	43.9 - 1.6 (1.66 - 1.6)	44.0-1.84 (1.91-1.84)	49.3-1.9 (1.97-1.9)	44.1-2.0 (2.07-2.0)
Rwork * (%)	0.1868 (0.2279)	0.2034 (0.2660)	0.1968 (0.3841)	0.2047 (0.3237)
Rfree ** (%)	0.2037 (0.2468)	0.2429 (0.2745)	0.2349 (0.4389)	0.2513 (0.3635)
Number of non-H atoms	2376	2229	2263	2193
macromolecules	2153	2153	2153	2151
ligands	15	18	16	14
solvent	208	58	94	28
Protein residues	269	269	269	269
Clashscore ***	1.84	1.61	2.3	2.07
R.m.s.d from ideal				
Bond lengths (Å°)	0.007	0.008	0.008	0.004
Bond angles (u)	1.21	1.18	1.13	0.98
Ramachandran favored ***				
(%)	98.88	98.13	98.5	98.13
Ramachandran outliers ***				
(%)	0	0	0	0
Average B values (Å ²)	26.77	42.47	38.69	57.43
macromolecules	26.25	42.61	38.81	57.59
ligands	18.4	33.48	26.43	46.38
solvent	32.76	40.06	38.03	50.94

Numbers in parentheses refer to the outer shell.

* Rwork = $\sum |hkl| |F_{obs}| - |F_{calc}| / \sum |hkl| |F_{obs}|$

** Rfree is the R value calculated for 5% of the data set that was not included in the refinement.

*** Molprobit.

Conclusion and Outlook

The biological role of ergothioneine and its evolutionary history remains a complex question. Understanding of any biochemical process must start with a detailed understanding of the chemistry of the enzymes involved, their catalytic mechanisms and their structures. This thesis contributes to the development of a molecular blue print of ergothioneine biochemistry. The approach used to do so involves the structural and functional characterisation of homologues from different areas of sequence space. This has revealed divergent enzymes and proteins, which through comparative enzymology have afforded several mechanistic and evolutionary insights into ergothioneine biochemistry.

Chapter two characterizes a divergent EgtB. The discovery of EgtB as part of the aerobic biosynthetic pathway represents an entirely new catalyst type that is distinct in both reactivity and structure from other iron oxygenases. While EgtB, and other sulfoxide synthases have been the focus of numerous studies, its catalytic mechanism is still disputed and its evolutionary history unknown. However, all mechanistic models are based on the crystal structure of *MthEgtB*. To provide an alternative model system, we solved the crystal structure of *CthEgtB*, a divergent homologue. This structure reveals a completely different configuration of active site residues that are involved in oxygen binding and activation. This structure informed mutagenesis studies, the outcomes of which provide evidence for a catalytic tyrosine dyad. One tyrosine plays an essential role as a catalytic acid, while the other acts as a Lewis acid, both facilitating the reduction of the initial iron (III) superoxide species, the first catalytic step towards sulfoxide production. The drastic active site differences of *CthEgtB* to *MthEgtB* provide us with a powerful and unique approach to tease out mechanistic features and to isolate the core catalytic principles of these catalysts via comparative enzymology. Re-evaluation of the contentious mechanistic proposals in view of the structure of *CthEgtB* provides a new test that may help to distinguish between different models. We believe that these additional constraints will be of significant help in the elimination or validation of mechanistic proposals for the sulfoxide synthase-catalyzed reaction.

Furthermore, comparison of the two EgtB structures and to the primary sequences of other bacterial and fungal homologs identified several sequence motifs predictive of active site features and function. The conservation of these features, or lack thereof, revealed that the class of ergothioneine biosynthetic sulfoxide synthases is characterized by remarkable active site diversity. It also led to the identification of another divergent EgtB homologue from *T. elongatus*, which remains uncharacterized. Finally, the observed diversity indicates that these sulfoxide synthase types may have emerged from an ancestral enzyme with different substrate specificity to any known extant homolog, gaining insight into evolutionary divergence explanation for the incredible diversity observed.

The evolutionary model proposed in chapter 2 explains the diversity of substrates and substrate-binding modes in the EgtB family. However, it does not account for the dramatic active site remodeling and catalytic residue hopping observed in *CthEgtB*. The introduction of two active site loops in *CthEgtB* is also

accompanied by tetramerization. In **chapter three**, two novel *CthEgtB* crystal structures are presented that each show an alternative loop conformation, providing evidence of the dynamic conformations that can be taken by the two active site loops. Structural and bioinformatic analysis identified two residues important for loop folding and substrate binding, Ser92 and Gly98. Mutation of Gly98 in a destabilized tetramer perturbed one of the tetrameric interfaces and hindered TMH binding through destabilizing loop folding. This showed that i) Gly98 acts as a hinge for loop folding ii) loop folding is coupled to TMH binding, iii) the quaternary structure stabilizes the enzyme:substrate complex and iv) loop folding is a limiting factor. Mutational analysis of Ser92 revealed that it is an important hydrogen bond donor for substrate-binding and formation of a catalytically relevant loop conformation. The assignment of roles for Ser92 and Gly98, in conjunction with the catalytic tyrosine dyad in chapter 2, provides a complete functional assignment of the four conserved active site loop 1 residues. We also propose that oligomerization enabled divergent evolution of the EgtB family, providing an explanation for the extreme active site plasticity and divergence of type II EgtBs.

Further exploration of the EgtB sequence space identified another divergent EgtB, *Vpa2054*. The genomic environment and conserved three-gene cluster implicates *Vpa2054* in C-Se bond formation. In **chapter four** the structural characterization of *Vpa2054* is reported. The structure shows a similar overall architecture compared to other EgtBs, yet the active site and substrate-binding sites differ. This again highlights the plasticity of the active sites in the EgtB family. The crystal structure, together with the preliminary biochemical characterization, provides strong evidence that TMH is the first substrate. TMH is the key component of all ergothioneine biosynthetic pathways characterized to date, suggesting that this enzyme could make ergothioneine or its seleno- isolog, selenoneine. While the second substrate is unknown, a selenophosphorylated glucoside is consistent with the conserved gene cluster and the *Vpa2054* substrate-binding pocket. This study narrows down the possibilities for proposals for the pathway encoded for by the three-operon gene cluster, and highlights the amenability of the EgtB scaffold for evolutionary divergence to accept alternative substrates and perhaps, alter reactivity. This work provides a platform for further studies to identify the second *Vpa2054* substrate and product of the three-operon gene cluster.

The identification of an enzyme for ergothioneine degradation (ergothionase) revealed a discrepancy in the organisms that can produce ergothioneine and degrade ergothioneine. As ergothionases are typically intracellular, a transport system for ergothioneine must therefore exist. **Chapter five** describes the structural characterization of two types of solute-binding proteins (a key component of the ABC transporter systems) that are specific for ergothioneine or ergothioneine sulfonic acid. The comparison of several high-resolution crystal structures of EGT and EGTSO₃⁻ specific SBP complexed with ligands revealed sequence motifs indicative of substrate binding. These studies provide evidence for a transport system for EGT and its oxidized analogue, EGTSO₃⁻, revealed EGTSO₃⁻ is indeed a relevant degradation product in nature. This molecular basis was applied to explore the sequence space of the solute binding proteins and revealed several key insights into EGT utilization and evolutionary history. (i) EGT SBPs are incredibly diverse, more diverse than any other betaine SBP (ii). This great diversity suggests EGT is an old molecule. (iii) EGT SBPs have evolved at least twice from independent origins. These findings are consistent with the hypotheses from the anaerobic biosynthetic pathway. (iv) The EGTSO₃⁻ SBPs are less diverse, are typically co-encoded with

EGT SO_3^- specific lyase, and likely evolved from a EGT-specific SBP. (v) EGT SBPs are not commonly encoded with a lyase. The latter is important as it means we can identify organisms which (1) do not produce EGT, (2) cannot degrade EGT and (3) have an ABC system for its transport. From the crystal structures, we have a molecular handle and genomic precedent to identify organisms in which ergothioneine may play a physiological role. Until now, no approach was available for identifying such organisms. This dramatically increases the number of organisms to which we know EGT may be relevant. This highlights the ubiquity and prevalence of EGT in nature.

The findings of this thesis provide numerous new examples of diversity in proteins involved in ergothioneine biosynthesis and utilization. The EgtB family in particular shows remarkable active site diversity, to the extent that key catalytic residues and substrates are not conserved among subtypes (Chapter 2). This highlights the amenability of the EgtB scaffold for evolutionary diversification, to lead to different active sites, substrates and perhaps even reactivity (Chapter 4). The oligomerization of the EgtB scaffold may even have facilitated the complete active site remodeling of *CthEgtB* (Chapter 3). Chapter 2 provided several sequence motifs that are indicative of a structural feature or behavior. Further exploration of the EgtB sequence space with these motifs is likely to reveal further divergent EgtBs. The characterization of these, along with detailed characterization of EgtB types iii-iv will provide further approaches to disentangle the catalytic mechanism, explore the evolutionary history and perhaps discover new reactivity and pathways. The structural characterization of a type III fungal enzyme would be particularly fruitful due to its ability to accept selenocysteine as a substrate, and may provide an intriguing parallel to *Vpa2054* (chapter 4) once the substrates are delineated.

As for EgtB, this thesis also develops a molecular blue print for the utilization of EGT and its derivatives, enabling the identification of new organisms to which ergothioneine is relevant. As we further explore the biochemistry of ergothioneine, its production, degradation and utilization, the organisms which make use of ergothioneine become more abundant, indicating the physiological importance of ergothioneine to numerous life forms. These results provide further evidence that ergothioneine is an old molecule, and that its biochemical components have emerged numerous times from independent origins, again advocating for a fundamental role in life. Incorporation of the detailed molecular basis of ergothioneine described in this thesis and elsewhere into a larger framework, which includes interconnected metabolic pathways and whole organisms, can provide a crucial mechanistic understanding of the role of ergothioneine to cells, organisms, their regulation and at a very basic level, even life.

References

1. Guo, A. C.; Jewison, T.; Wilson, M.; Liu, Y.; Knox, C.; Djoumbou, Y.; Lo, P.; Mandal, R.; Krishnamurthy, R.; Wishart, D. S., ECMDDB: the E. coli Metabolome Database. *Nucleic Acids Res* 2013, 41 (Database issue), D625-D630.
2. Jewison, T.; Knox, C.; Neveu, V.; Djoumbou, Y.; Guo, A. C.; Lee, J.; Liu, P.; Mandal, R.; Krishnamurthy, R.; Sinelnikov, I.; Wilson, M.; Wishart, D. S., YMDB: the Yeast Metabolome Database. *Nucleic Acids Res* 2011, 40 (D1), D815-D820.
3. Nelson, D. L.; Cox, M. M., *Lehninger Principles of Biochemistry*, 4th Edition. W. H. Freeman: 2004; p 1200 pp.
4. Lewis, C. A.; Wolfenden, R., Uroporphyrinogen decarboxylation as a benchmark for the catalytic proficiency of enzymes. *Proceedings of the National Academy of Sciences* 2008, 105 (45), 17328-17333.
5. de Verneuil, H.; Sassa, S.; Kappas, A., Purification and properties of uroporphyrinogen decarboxylase from human erythrocytes. A single enzyme catalyzing the four sequential decarboxylations of uroporphyrinogens I and III. *Journal of Biological Chemistry* 1983, 258 (4), 2454-60.
6. Loiseau, D., Fermentation of raffinose by beer yeast. *Compt. rend.* 109, 614-5.
7. Buchner, E., Alcoholic fermentation without yeast cells. *Ber.* 1897, 30, 117-24.
8. Sumner, J. B., The isolation and crystallisation of the enzyme urease: preliminary paper. *Journal of Biological Chemistry* 1926, 69 (2), 435-441.
9. Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R., Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* 1982, 31 (1), 147-157.
10. Michaelis, L.; Menten, M. L.; Johnson, K. A.; Goody, R. S., The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry* 2011, 50 (39), 8264-9.
11. Pauling, L., *Molecular Architecture and Biological Reactions*. *Chemical & Engineering News Archive* 1946, 24 (10), 1375-1377.
12. Silverman, R. B., *Organic Chemistry of Enzyme-Catalyzed Reactions*. 2 ed.; Academic Press: 2002.
13. Copley, S. D., Toward a Systems Biology Perspective on Enzyme Evolution. *Journal of Biological Chemistry* 2012, 287 (1), 3-10.
14. Cheah, I. K.; Halliwell, B., Ergothioneine; antioxidant potential, physiological function and role in disease. *Biochim Biophys Acta* 2012, 1822 (5), 784-93.
15. Tanret, C., New Base Obtained from Ergot of Rye. Ergothioneine. *Compt. rend.* 1909, 149, 222-4.
16. Halliwell, B.; Cheah, I. K.; Tang, R. M. Y., Ergothioneine – a diet-derived antioxidant with therapeutic potential. *FEBS Letters* 2018, 592 (20), 3357-3366.

17. Carlsson, J.; Kierstan, M. P. J.; Brocklehurst, K., Reactions of L-ergothioneine and some other aminothiones with 2,2'- and 4,4'-dipyridyl disulphides and of L-ergothioneine with iodoacetamide. 2-Mercaptoimidazoles, 2- and 4-thiopyridones, thiourea and thioacetamide as highly reactive neutral sulphur nucleophiles. *Biochemical Journal* 1974, 139 (1), 221-235.
18. Jocelyn, P. C., *Biochemistry of the SH group; the occurrence, chemical properties, metabolism and biological function of thiols and disulphides*. London ; New York : Academic Press, 1972.: 1972.
19. Melville, D. B., Ergothioneine. In *Vitamins and Hormones*, 1959; Vol. 17, pp 155-204.
20. Gründemann, D.; Harlfinger, S.; Golz, S.; Geerts, A.; Lazar, A.; Berkels, R.; Jung, N.; Rubbert, A.; Schömig, E., Discovery of the ergothioneine transporter. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102 (14), 5256-5261.
21. Melville, D. B.; Horner, W. H.; Lubschez, R., Tissue ergothioneine. *The Journal of biological chemistry* 1954, 206 (1), 221-228.
22. Shires, T. K.; Brummel, M. C.; Pulido, J. S.; Stegink, L. D., Ergothioneine Distribution in Bovine and Porcine Ocular Tissues. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology* 1997, 117 (1), 117-120.
23. Salt, H. B., The ergothioneine content of the blood in health and disease. *Biochemical Journal* 1931, 25 (5), 1712-1719.
24. Leone, E.; Mann, T., Ergothioneine in the Seminal Vesicle Secretion. *Nature* 1951, 168 (4266), 205-206.
25. Franzoni, F.; Colognato, R.; Galetta, F.; Laurenza, I.; Barsotti, M.; Di Stefano, R.; Bocchetti, R.; Regoli, F.; Carpi, A.; Balbarini, A.; Migliore, L.; Santoro, G., An in vitro study on the free radical scavenging capacity of ergothioneine: comparison with reduced glutathione, uric acid and trolox. *Biomedicine & Pharmacotherapy* 2006, 60 (8), 453-457.
26. Whiteman, M.; Halliwell, B., Thiols and disulphides can aggravate peroxynitrite-dependent inactivation of α 1-antiproteinase. *FEBS Letters* 1997, 414 (3), 497-500.
27. Motohashi, N.; Mori, I., Thiol-induced hydroxyl radical formation and scavenger effect of thiocarbamides on hydroxyl radicals. *Journal of Inorganic Biochemistry* 1986, 26 (3), 205-212.
28. Akanmu, D.; Cecchini, R.; Aruoma, O. I.; Halliwell, B., The antioxidant action of ergothioneine. *Archives of Biochemistry and Biophysics* 1991, 288 (1), 10-16.
29. Paul, B. D.; Snyder, S. H., The unusual amino acid L-ergothioneine is a physiologic cytoprotectant. *Cell Death And Differentiation* 2009, 17, 1134.
30. Deiana, M.; Rosa, A.; Casu, V.; Piga, R.; Assunta Dessì, M.; Aruoma, O. I., L-Ergothioneine modulates oxidative damage in the kidney and liver of rats in vivo: studies upon the profile of polyunsaturated fatty acids. *Clinical Nutrition* 2004, 23 (2), 183-193.
31. Seebeck, F. P., In Vitro Reconstitution of Mycobacterial Ergothioneine Biosynthesis. *Journal of the American Chemical Society* 2010, 132 (19), 6632-6633.
32. Burn, R.; Misson, L.; Meury, M.; Seebeck, F. P., Anaerobic Origin of Ergothioneine. *Angew Chem Int Ed Engl* 2017, 56 (41), 12508-12511.

33. Hu, W.; Song, H.; Sae, H. A.; 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 (20), 5382-5.
34. Melville, D. B.; Eich, S.; Ludwig, M. L., THE BIOSYNTHESIS OF ERGOTHIONEINE. *Journal of Biological Chemistry* 1957, 224 (2), 871-877.
35. Ishikawa, Y.; Melville, D. B., The Enzymatic α -N-Methylation of Histidine. *Journal of Biological Chemistry* 1970, 245 (22), 5967-5973.
36. Ishikawa, Y.; Israel, S. E.; Melville, D. B., Participation of an Intermediate Sulfoxide in the Enzymatic Thiolation of the Imidazole Ring of Hercynine to Form Ergothioneine. *Journal of Biological Chemistry* 1974, 249 (14), 4420-4427.
37. Bello, M. H.; Barrera-Perez, V.; Morin, D.; Epstein, L., The *Neurospora crassa* mutant Nc Δ Egt-1 identifies an ergothioneine biosynthetic gene and demonstrates that ergothioneine enhances conidial survival and protects against peroxide toxicity during conidial germination. *Fungal Genet. Biol.* 2012, 49 (2), 160-172.
38. Sheridan, K. J.; Lechner, B. E.; O'Keeffe, G.; Keller, M. A.; Werner, E. R.; Lindner, H.; Jones, G. W.; Haas, H.; Doyle, S., Ergothioneine Biosynthesis and Functionality in the Opportunistic Fungal Pathogen, *Aspergillus fumigatus*. *Sci. Rep.* 2016, 6, 35306.
39. Vit, A.; Misson, L.; Blankenfeldt, W.; Seebeck, F. P., Ergothioneine biosynthetic methyltransferase EgtD reveals the structural basis of aromatic amino acid betaine biosynthesis. *Chembiochem* 2015, 16 (1), 119-25.
40. Pluskal, T.; Ueno, M.; Yanagida, M., Genetic and metabolomic dissection of the ergothioneine and selenoneine biosynthetic pathway in the fission yeast, *S. pombe*, and construction of an overproduction system. *PLoS One* 2014, 9 (5), e97774.
41. Goncharenko, K. V.; Flückiger, S.; Liao, C.; Lim, D.; Stampfli, A. R.; Seebeck, F. P., Selenocysteine as a substrate, an inhibitor and a mechanistic probe for bacterial and fungal iron-dependent sulfoxide synthases. *Chemistry – A European Journal* 2019, 0 (ja).
42. Jones, G. W.; Doyle, S.; Fitzpatrick, D. A., The evolutionary history of the genes involved in the biosynthesis of the antioxidant ergothioneine. *Gene* 2014, 549 (1), 161-70.
43. Paul, B. D.; Snyder, S. H., The unusual amino acid L-ergothioneine is a physiologic cytoprotectant. *Cell Death & Differentiation* 2010, 17 (7), 1134-1140.
44. Gamage, A. M.; Liao, C.; Cheah, I. K.; Chen, Y.; Lim, D. R. X.; Ku, J. W. K.; Chee, R. S. L.; Gengenbacher, M.; Seebeck, F. P.; Halliwell, B.; Gan, Y.-H., The proteobacterial species *Burkholderia pseudomallei* produces ergothioneine, which enhances virulence in mammalian infection. *FASEB J.* 2018, 32 (12), 6395-6409.
45. Leisinger, F.; Burn, R.; Meury, M.; Lukat, P.; Seebeck, F. P., Structural and Mechanistic Basis for Anaerobic Ergothioneine Biosynthesis. *J. Am. Chem. Soc.* 2019, 141 (17), 6906-6914.
46. Fischer, W. W.; Hemp, J.; Valentine, J. S., How did life survive Earth's great oxygenation? *Current Opinion in Chemical Biology* 2016, 31, 166-178.

47. Liao, C.; Seebeck, F. P., Convergent Evolution of Ergothioneine Biosynthesis in Cyanobacteria. *ChemBioChem* 2017, 18 (21), 2115-2118.
48. Vit, A.; Mashabela, G. T.; Blankenfeldt, W.; Seebeck, F. P., Structure of the Ergothioneine-Biosynthesis Amidohydrolase EgtC. *ChemBioChem* 2015, 16 (10), 1490-1496.
49. Goncharenko, K. V.; Vit, A.; Blankenfeldt, W.; Seebeck, F. P., Structure of the sulfoxide synthase EgtB from the ergothioneine biosynthetic pathway. *Angew Chem Int Ed Engl* 2015, 54 (9), 2821-4.
50. Goncharenko, K. V.; Seebeck, F. P., Conversion of a non-heme iron-dependent sulfoxide synthase into a thiol dioxygenase by a single point mutation. *Chem Commun (Camb)* 2016, 52 (9), 1945-8.
51. Pierce, B. S.; Gardner, J. D.; Bailey, L. J.; Brunold, T. C.; Fox, B. G., Characterization of the Nitrosyl Adduct of Substrate-Bound Mouse Cysteine Dioxygenase by Electron Paramagnetic Resonance: Electronic Structure of the Active Site and Mechanistic Implications. *Biochemistry* 2007, 46 (29), 8569-8578.
52. Li, W.; Pierce, B. S., Steady-state substrate specificity and O₂-coupling efficiency of mouse cysteine dioxygenase. *Archives of Biochemistry and Biophysics* 2015, 565, 49-56.
53. Birringer, M.; Pilawa, S.; Flohe, L., Trends in selenium biochemistry. *Nat. Prod. Rep.* 2002, 19 (6), 693-718.
54. Arnaudguilhem, C.; Bierla, K.; Ouerdane, L.; Preud'homme, H.; Yiannikouris, A.; Lobinski, R., Selenium metabolomics in yeast using complementary reversed-phase/hydrophilic ion interaction (HILIC) liquid chromatography–electrospray hybrid quadrupole trap/Orbitrap mass spectrometry. *Analytica Chimica Acta* 2012, 757, 26-38.
55. Stadtman, T. C., SELENIUM BIOCHEMISTRY. *Annual Review of Biochemistry* 1990, 59 (1), 111-127.
56. Blaesi, E. J.; Gardner, J. D.; Fox, B. G.; Brunold, T. C., Spectroscopic and Computational Characterization of the NO Adduct of Substrate-Bound Fe(II) Cysteine Dioxygenase: Insights into the Mechanism of O₂ Activation. *Biochemistry* 2013, 52 (35), 6040-6051.
57. De Silva, V.; Woznichak, M. M.; Burns, K. L.; Grant, K. B.; May, S. W., Selenium Redox Cycling in the Protective Effects of Organoselenides against Oxidant-Induced DNA Damage. *Journal of the American Chemical Society* 2004, 126 (8), 2409-2413.
58. Reich, H. J.; Hondal, R. J., Why Nature Chose Selenium. *ACS Chemical Biology* 2016, 11 (4), 821-841.
59. Flückiger, S. Mechanistic Studies of Sulfur-Carbon Bond Formation by Metal-Dependent Enzymes. 2018.
60. Solomon, E. I.; Gorelsky, S. I.; Dey, A., Metal–thiolate bonds in bioinorganic chemistry. *Journal of Computational Chemistry* 2006, 27 (12), 1415-1428.
61. Zheng, P.; Takayama, S.-i. J.; Mauk, A. G.; Li, H., Hydrogen Bond Strength Modulates the Mechanical Strength of Ferric-Thiolate Bonds in Rubredoxin. *Journal of the American Chemical Society* 2012, 134 (9), 4124-4131.
62. Goncharenko, K. V. The Catalytic Mechanism of the Iron-Dependent Sulfoxide Synthase EgtB. University of Basel, 2017.

63. Braunschauen, A.; Seebeck, F. P., Identification and characterization of the first ovoidiol biosynthetic enzyme. *J. Am. Chem. Soc.* 2011, 133 (6), 1757.
64. Turner, E.; Hager, L.; Shapiro, B., Ovoidiol replaces glutathione peroxidase as a hydrogen peroxide scavenger in sea urchin eggs. *Science* 1988, 242 (4880), 939-941.
65. Weaver, K. H.; Rabenstein, D. L., Thiol/Disulfide Exchange Reactions of Ovoidiol A with Glutathione. *The Journal of Organic Chemistry* 1995, 60 (6), 1904-1907.
66. Holler, T. P.; Hopkins, P. B., Ovoidiols as biological antioxidants. The thiol groups of ovoidiol and glutathione are chemically distinct. *Journal of the American Chemical Society* 1988, 110 (14), 4837-4838.
67. Naowarajna, N.; Huang, P.; Cai, Y.; Song, H.; Wu, L.; Cheng, R.; Li, Y.; Wang, S.; Lyu, H.; Zhang, L.; Zhou, J.; Liu, P., In Vitro Reconstitution of the Remaining Steps in Ovoidiol A Biosynthesis: C-S Lyase and Methyltransferase Reactions. *Organic Letters* 2018, 20 (17), 5427-5430.
68. Misson, L. L., C.; Seebeck, F. P., Characterisation of OvoA. Unpublished.
69. Mashabela, G. T. M.; Seebeck, F. P., Substrate specificity of an oxygen dependent sulfoxide synthase in ovoidiol biosynthesis. *Chem Commun (Camb)* 2013, 49 (70), 7714-6.
70. Song, H.; Leninger, M.; Lee, N.; Liu, P., Regioselectivity of the Oxidative C-S Bond Formation in Ergothioneine and Ovoidiol Biosyntheses. *Organic Letters* 2013, 15 (18), 4854-4857.
71. 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 Ovoidiol Biosynthesis. *Org. Lett.* 2014, 16 (8), 2122-2125.
72. Chen, L.; Deng, Z.; Zhao, C.; Chen, L.; Naowarajna, N.; Song, H.; Wang, S.; Zhao, C.; Liu, P.; Song, H.; Wang, J., Use of a Tyrosine Analogue To Modulate the Two Activities of a Nonheme Iron Enzyme OvoA in Ovoidiol Biosynthesis, Cysteine Oxidation versus Oxidative C-S Bond Formation. *J Am Chem Soc* 2018, 140 (13), 4604-4612.
73. Chen, L.; Naowarajna, N.; Chen, B.; Xu, M.; Quill, M.; Wang, J.; Deng, Z.; Zhao, C.; Liu, P., Mechanistic Studies of a Nonheme Iron Enzyme OvoA in Ovoidiol Biosynthesis Using a Tyrosine Analogue, 2-Amino-3-(4-hydroxy-3-(methoxyl) phenyl) Propanoic Acid (MeOTyr). *ACS Catal.* 2019, 9 (1), 253-258.
74. Stipanuk, M. H.; Ueki, I.; Dominy, J. E., Jr.; Simmons, C. R.; Hirschberger, L. L., Cysteine dioxygenase: a robust system for regulation of cellular cysteine levels. *Amino Acids* 2009, 37 (1), 55-63.
75. Joseph, C. A.; Maroney, M. J., Cysteine dioxygenase: structure and mechanism. *Chemical Communications* 2007, (32), 3338-3349.
76. Driggers, C. M.; Cooley, R. B.; Sankaran, B.; Hirschberger, L. L.; Stipanuk, M. H.; Karplus, P. A., Cysteine Dioxygenase Structures from pH4 to 9: Consistent Cys-Persulfenate Formation at Intermediate pH and a Cys-Bound Enzyme at Higher pH. *Journal of Molecular Biology* 2013, 425 (17), 3121-3136.
77. Ye, S.; Wu, X. a.; Wei, L.; Tang, D.; Sun, P.; Bartlam, M.; Rao, Z., An Insight into the Mechanism of Human Cysteine Dioxygenase: KEY ROLES OF THE THIOETHER-BONDED TYROSINE-CYSTEINE COFACTOR. *Journal of Biological Chemistry* 2007, 282 (5), 3391-3402.

78. Tchesnokov, E. P.; Faponle, A. S.; Davies, C. G.; Quesne, M. G.; Turner, R.; Fellner, M.; Souness, R. J.; Wilbanks, S. M.; de Visser, S. P.; Jameson, G. N. L., An iron–oxygen intermediate formed during the catalytic cycle of cysteine dioxygenase. *Chemical Communications* 2016, 52 (57), 8814-8817.
79. Blaesi, E. J.; Fox, B. G.; Brunold, T. C., Spectroscopic and Computational Investigation of Iron(III) Cysteine Dioxygenase: Implications for the Nature of the Putative Superoxo-Fe(III) Intermediate. *Biochemistry* 2014, 53 (36), 5759-5770.
80. Bushnell, E. A.; Fortowsky, G. B.; Gauld, J. W., Model iron-oxo species and the oxidation of imidazole: insights into the mechanism of OvoA and EgtB? *Inorg Chem* 2012, 51 (24), 13351-6.
81. Faponle, A. S.; Seebeck, F. P.; de Visser, S. P., Sulfoxide Synthase versus Cysteine Dioxygenase Reactivity in a Nonheme Iron Enzyme. *J. Am. Chem. Soc.* 2017, 139 (27), 9259-9270.
82. Wei, W. J.; Siegbahn, P. E.; Liao, R. Z., Theoretical Study of the Mechanism of the Nonheme Iron Enzyme EgtB. *Inorg Chem* 2017, 56 (6), 3589-3599.
83. Tian, G.; Su, H.; Liu, Y., Mechanism of Sulfoxidation and C–S Bond Formation Involved in the Biosynthesis of Ergothioneine Catalyzed by Ergothioneine Synthase (EgtB). *ACS Catal.* 2018, 8, 5875.
84. Price, J. C.; Barr, E. W.; Tirupati, B.; Bollinger, J. M.; Krebs, C., The First Direct Characterization of a High-Valent Iron Intermediate in the Reaction of an α -Ketoglutarate-Dependent Dioxygenase: A High-Spin Fe(IV) Complex in Taurine/ α -Ketoglutarate Dioxygenase (TauD) from *Escherichia coli*. *Biochemistry* 2003, 42 (24), 7497-7508.
85. Tamanaha, E.; Zhang, B.; Guo, Y.; Chang, W.-c.; Barr, E. W.; Xing, G.; St. Clair, J.; Ye, S.; Neese, F.; Bollinger, J. M.; Krebs, C., Spectroscopic Evidence for the Two C–H-Cleaving Intermediates of *Aspergillus nidulans* Isopenicillin N Synthase. *Journal of the American Chemical Society* 2016, 138 (28), 8862-8874.
86. Bollinger Jr, J. M.; Price, J. C.; Hoffart, L. M.; Barr, E. W.; Krebs, C., Mechanism of Taurine: α -Ketoglutarate Dioxygenase (TauD) from *Escherichia coli*. *European Journal of Inorganic Chemistry* 2005, 2005 (21), 4245-4254.
87. Naowarajna, N.; Cheng, R.; Chen, L.; Quill, M.; Xu, M.; Zhao, C.; Liu, P., Mini-Review: Ergothioneine and Othiol Biosyntheses, an Unprecedented Trans-Sulfur Strategy in Natural Product Biosynthesis. *Biochemistry* 2018, 57 (24), 3309-3325.
88. Hoyle, C. E.; Bowman, C. N., Thiol–Ene Click Chemistry. *Angewandte Chemie International Edition* 2010, 49 (9), 1540-1573.
89. Solomon, E. I.; Goudarzi, S.; Sutherlin, K. D., O₂ Activation by Non-Heme Iron Enzymes. *Biochemistry* 2016, 55 (46), 6363-6374.
90. Mbughuni, M. M.; Chakrabarti, M.; Hayden, J. A.; Bominaar, E. L.; Hendrich, M. P.; xfc; nck, E.; Lipscomb, J. D., Trapping and spectroscopic characterization of an Fe^{III}-superoxo intermediate from a nonheme mononuclear iron-containing enzyme. *Proceedings of the National Academy of Sciences of the United States of America* 2010, 107 (39), 16788-16793.

91. Zhu, H.; Peck, S. C.; Bonnot, F.; van der Donk, W. A.; Klinman, J. P., Oxygen-18 Kinetic Isotope Effects of Nonheme Iron Enzymes HEPD and MPnS Support Iron(III) Superoxide as the Hydrogen Abstraction Species. *Journal of the American Chemical Society* 2015, 137 (33), 10448-10451.
92. Xing, G.; Diao, Y.; Hoffart, L. M.; Barr, E. W.; Prabhu, K. S.; Arner, R. J.; Reddy, C. C.; Krebs, C.; Bollinger, J. M., Evidence for C-H Cleavage by an Iron-Superoxide Complex in the Glycol Cleavage Reaction Catalyzed by Myo-Inositol Oxygenase. *Proceedings of the National Academy of Sciences of the United States of America* 2006, 103 (16), 6130-6135.
93. van der Donk, W. A.; Krebs, C.; Bollinger, J. M., Substrate activation by iron superoxo intermediates. *Current Opinion in Structural Biology* 2010, 20 (6), 673-683.
94. Hong, S.; Sutherland, K. D.; Park, J.; Kwon, E.; Siegler, M. A.; Solomon, E. I.; Nam, W., Crystallographic and spectroscopic characterization and reactivities of a mononuclear non-haem iron(III)-superoxo complex. *Nature Communications* 2014, 5 (1), 5440.
95. Mashabela, G. T.; Seebeck, F. P., Substrate specificity of an oxygen dependent sulfoxide synthase in ovolithol biosynthesis. *Chem Commun (Camb)* 2013, 49 (70), 7714-6.
96. Matthews, W. S.; Bares, J. E.; Bartmess, J. E.; Bordwell, F. G.; Cornforth, F. J.; Drucker, G. E.; Margolin, Z.; McCallum, R. J.; McCollum, G. J.; Vanier, N. R., Equilibrium acidities of carbon acids. VI. Establishment of an absolute scale of acidities in dimethyl sulfoxide solution. *Journal of the American Chemical Society* 1975, 97 (24), 7006-7014.
97. Bordwell, F. G., Equilibrium acidities in dimethyl sulfoxide solution. *Accounts of Chemical Research* 1988, 21 (12), 456-463.
98. Richard-Greenblatt, M.; Bach, H.; Adamson, J.; Pena-Diaz, S.; Li, W.; Steyn, A. J. C.; Av-Gay, Y., Regulation of ergothioneine biosynthesis and its effect on *Mycobacterium tuberculosis* growth and infectivity. *J. Biol. Chem.* 2015, 290 (38), 23064-23076.
99. Nakajima, S.; Satoh, Y.; Yanashima, K.; Matsui, T.; Dairi, T., Ergothioneine protects *Streptomyces coelicolor* A3(2) from oxidative stresses. *J. Biosci. Bioeng.* 2015, 120 (3), 294-298.
100. Osawa, R.; Kamide, T.; Satoh, Y.; Kawano, Y.; Ohtsu, I.; Dairi, T., Heterologous and High Production of Ergothioneine in *Escherichia coli*. *J. Agric. Food Chem.* 2018, 66 (5), 1191-1196.
101. Bello, M. H.; Mogannam, J. C.; Morin, D.; Epstein, L., Endogenous ergothioneine is required for wild-type levels of conidiogenesis and conidial survival but does not protect against 254 nm UV-induced mutagenesis or kill. *Fungal Genet. Biol.* 2014, 73, 120-127.
102. Sao, E. C.; Williams, M. J.; Van, H. P. D.; Wiid, I. J.; Baker, B.; Sao, E. C.; Carolis, C.; Taylor, M. J. C., Generation and characterization of thiol-deficient *Mycobacterium tuberculosis* mutants. *Sci Data* 2018, 5, 180184.
103. Saini, V.; Cumming, B. M.; Guidry, L.; Lamprecht, D. A.; Adamson, J. H.; Reddy, V. P.; Chinta, K. C.; Mazorodze, J. H.; Glasgow, J. N.; Richard-Greenblatt, M.; Gomez-Velasco, A.; Bach, H.; Av-Gay, Y.; Eoh, H.; Rhee, K.; Steyn, A. J. C., Ergothioneine Maintains Redox and Bioenergetic Homeostasis Essential for Drug Susceptibility and Virulence of *Mycobacterium tuberculosis*. *Cell Rep.* 2016, 14 (3), 572-585.

104. Pfeiffer, C.; Bach, M.; Bauer, T.; Campos da Ponte, J.; Schoemig, E.; Gruendemann, D., Knockout of the ergothioneine transporter ETT in zebrafish results in increased 8-oxoguanine levels. *Free Radical Biol. Med.* 2015, 83, 178-185.
105. Kato, Y.; Kubo, Y.; Iwata, D.; Kato, S.; Sudo, T.; Sugiura, T.; Kagaya, T.; Wakayama, T.; Hirayama, A.; Sugimoto, M.; Sugihara, K.; Kaneko, S.; Soga, T.; Asano, M.; Tomita, M.; Matsui, T.; Wada, M.; Tsuji, A., Gene Knockout and Metabolome Analysis of Carnitine/Organic Cation Transporter OCTN1. *Pharm. Res.* 2010, 27 (5), 832-840.
106. Nakamura, T.; Sugiura, S.; Kobayashi, D.; Yoshida, K.; Yabuuchi, H.; Aizawa, S.; Maeda, T.; Tamai, I., Decreased Proliferation and Erythroid Differentiation of K562 Cells by siRNA-induced Depression of OCTN1 (SLC22A4) Transporter Gene. *Pharm. Res.* 2007, 24 (9), 1628-1635.
107. Cheah, I. K.; Ong, R. L. S.; Gruber, J.; Yew, T. S. K.; Ng, L. F.; Chen, C. B.; Halliwell, B., Knockout of a putative ergothioneine transporter in *Caenorhabditis elegans* decreases lifespan and increases susceptibility to oxidative damage. *Free Radical Res.* 2013, 47 (12), 1036-1045.
108. Tanaka, N.; Kawano, Y.; Satoh, Y.; Dairi, T.; Ohtsu, I., Gram-scale fermentative production of ergothioneine driven by overproduction of cysteine in *Escherichia coli*. *Sci. Rep.* 2019, 9 (1), 1-10.
109. Liu, P.; Song, H.; Hu, W. Ergothioneine production through metabolic engineering. WO2014100752A1, 2014.
110. Jiang, W.; Liu, Q.; Zhou, T.; Yang, P. Manufacture of ergothioneine with large fungus. CN103184246A, 2013.
111. Burn, R. Reto Burn PhD thesis. 2019.
112. Todd, A. E.; Orengo, C. A.; Thornton, J. M., Plasticity of enzyme active sites. *Trends Biochem. Sci.* 2002, 27 (8), 419-426.
113. Naowarajna, N.; Irani, S.; Hu, W.; Cheng, R.; Zhang, L.; Li, X.; Chen, J.; Zhang, Y. J.; Liu, P., Crystal Structure of the Ergothioneine Sulfoxide Synthase from *Candidatus Chloracidobacterium thermophilum* and Structure-Guided Engineering To Modulate Its Substrate Selectivity. *ACS Catal.* 2019, Ahead of Print.
114. Gao, S. S.; Naowarajna, N.; Cheng, R.; Liu, X.; Liu, P., Recent examples of α -ketoglutarate-dependent mononuclear non-haem iron enzymes in natural product biosyntheses. *Nat. Prod. Rep.* 2018, 35 (8), 792 - 837.
115. Askari, A.; Melville, D. B., The reaction sequence in ergothioneine biosynthesis: hercynine as an intermediate. *J. Biol. Chem.* 1962, 237 (5), 1615-&.
116. Bello, M. H.; Barrera-Perez, V.; Morin, D.; Epstein, L., The *Neurospora crassa* mutant NcDEgt-1 identifies an ergothioneine biosynthetic gene and demonstrates that ergothioneine enhances conidial survival and protects against peroxide toxicity during conidial germination. *Fungal Genet. Biol.* 2012, 49 (2), 160-172.
117. Cumming, B. M.; Chinta, K. C.; Reddy, V. P.; Steyn, A. J. C., Role of Ergothioneine in Microbial Physiology and Pathogenesis. *Antioxid. Redox Signal.* 2018, 28 (6), 431 - 444.

118. Sheridan, K. J.; Lechner, B. E.; Keeffe, G. O.; Keller, M. A.; Werner, E. R.; Lindner, H.; Jones, G. W.; Haas, H.; Doyle, S., Ergothioneine Biosynthesis and Functionality in the Opportunistic Fungal Pathogen, *Aspergillus fumigatus*. *Sci Rep.* 2016, 6, 35306.
119. Aik, W. S.; McDonough, M. A.; Thalhammer, A.; Chowdhury, R.; Schofield, C. J., Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* 2012, 22 (6), 691 - 700.
120. Meury, M.; Knop, M.; Seebeck, F. P., Structural Basis for Copper-Oxygen Mediated C-H Bond Activation by the Formylglycine-Generating Enzyme. *Angew Chem Int Ed Engl.* 2017, 56 (27), 8115 - 8119.
121. Dierks, T.; Dickmanns, A.; Preusser-Kunze, A.; Schmidt, B.; Mariappan, M.; von Figura, K.; Ficner, R.; Rudolph, M. G., Molecular Basis for Multiple Sulfatase Deficiency and Mechanism for Formylglycine Generation of the Human Formylglycine-Generating Enzyme. *Cell* 2005, 121 (4), 541.
122. Newton, G. L.; Leung, S. S.; Wakabayashi, J. L.; Rawat, M.; Fahey, R. C., The DinB Superfamily Includes Novel Mycothiol, Bacillithiol, and Glutathione S-Transferases. *Biochemistry* 2011, 50 (49), 10751-10760.
123. Chen, L.; Naowarojna, N.; Song, H.; Wang, S.; Wang, J.; Deng, Z.; Zhao, C.; P., L., Use of a tyrosine analog to modulate the two activities of a non-heme iron enzyme OvoA in ovothiol biosynthesis, cysteine oxidation versus oxidative C-S bond formation. *J. Am. Chem. Soc.* 2018, 140 (13), 4604 - 4612.
124. Tian, G.; Su, H.; Liu, Y., Mechanism of Sulfoxidation and C-S Bond Formation Involved in the Biosynthesis of Ergothioneine Catalyzed by Ergothioneine Synthase (EgtB) . *ACS Catalysis* 2018, 8, 5875 - 5889.
125. Gamage, A. M.; Liao, C.; Cheah, I. K.; Chen, Y.; Lim, D. R. X.; Ku, J. W. K.; Chee, R. S. L.; Gengenbacher, M.; Seebeck, F. P.; Halliwell, B.; Gan, Y. H., The proteobacterial species *Burkholderia pseudomallei* produces ergothioneine which enhances virulence in mammalian infection. *FASEB, J* 2018, 11:fj201800716.
126. Hu, W.; Song, H.; Sae Her, A.; Bak, D. W.; Naowarojna, N.; Elliot, 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 (20), 5382 - 5385.
127. Kulik, H. J.; Blasiak, L. C.; Marzari, N.; Drennan, C. L., First-Principles Study of Non-heme Fe(II) Halogenase SyrB2 Reactivity. *Journal of the American Chemical Society* 2009, 131 (40), 14426-14433.
128. Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L., Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* 2006, 440 (7082), 368-371.
129. Song, H.; Her, A. S.; Raso, F.; Zheng, Z.; Huo, Y.; Liu, P., Cysteine Oxidation Reactions Catalyzed by a Mononuclear Nonheme Iron Enzyme (OvoA) in Ovothiol Biosynthesis. *Org. Lett.* 2014, 16 (8), 2122-2125.

130. Mbughuni, M. M.; Chakrabarti, M.; Hayden, J. A.; Bominaar, E. L.; Hendrich, M. P.; Münck, E.; Lipscomb, J. D., Trapping and spectroscopic characterization of an FeIII-superoxo intermediate from a nonheme mononuclear iron-containing enzyme. *Proc. Natl. Acad. Sci. U. S. A.* 2010, 107 (39), 16788 - 16793.
131. Hong, S.; Sutherlin, K. D.; Park, J.; Kwon, E.; Siegler, M. A.; Solomon, E. I.; Nam, W., Crystallographic and spectroscopic characterization and reactivities of a mononuclear non-haem iron(III)-superoxo complex. *Nat. Commun.* 2014, 5, 5440.
132. Vit, A.; Mashabela, G. T.; Blankenfeldt, W.; Seebeck, F. P., Structure of the Ergothioneine-Biosynthesis Amidohydrolase EgtC. *ChemBioChem* 2015, 16 (10), 1490 - 1496.
133. Gardner, J. D.; Pierce, B. S.; Fox, B. G.; Brunold, T. C., Spectroscopic and Computational Characterization of Substrate-Bound Mouse Cysteine Dioxygenase: Nature of the Ferrous and Ferric Cysteine Adducts and Mechanistic Implications. *Biochemistry* 2010, 49, 6033-6041.
134. Kumar, D.; Thiel, W.; de Visser, S. P., Theoretical study on the mechanism of the oxygen activation process in cysteine dioxygenase enzymes. *J. Am. Chem. Soc.* 2011, 133, 3869 - 3882.
135. Ye, S.; Wu, X.; Wei, L.; Tang, D.; Sun, P.; Bartlam, M.; Rao, Z., An insight into the mechanism of human cysteine dioxygenase. Key roles of the thioether-bonded tyrosine-cysteine cofactor. *J. Biol. Chem.* 2007, 282 (5), 3391 - 3402.
136. Mantri, M.; Zhang, Z.; McDonough, M. A.; Schofield, C. J., Autocatalysed oxidative modifications to 2-oxoglutarate dependent oxygenases. *FEBS J.* 2012, 279 (9), 1563 - 1575.
137. Yan, W.; Song, H.; Song, F.; Guo, Y.; Wu, C. H.; Her, A. S.; Pu, Y.; Wang, S.; Naowarajna, N.; Weitz, A.; Hendrich, M. P.; Costello, C. E.; Zhang, L.; Liu, P.; Zhang, Y. J., Endoperoxide formation by an α -ketoglutarate-dependent mononuclear non-haem iron enzyme. *Nature* 2015, 527 (7579), 539 - 543.
138. Ryle, M. J.; Liu, A.; Muthukumar, R. B.; Ho, R. Y.; Koehntop, K. D.; McCracken, J.; Que, L. J.; Hausinger, R. P., O₂- and alpha-ketoglutarate-dependent tyrosyl radical formation in TauD, an alpha-keto acid-dependent non-heme iron dioxygenase. *Biochemistry* 2003, 42 (7), 1854 - 1862.
139. Dominy, J. E. J.; Hwang, J.; Guo, S.; Hirschberger, L. L.; Zhang, S.; Stipanuk, M. H., Synthesis of amino acid cofactor in cysteine dioxygenase is regulated by substrate and represents a novel post-translational regulation of activity. *J. Biol. Chem.* 2008, 283 (18), 12188 - 12201.
140. Timmins, A.; Saint-André, M.; de Visser, S. P., Understanding How Prolyl-4-hydroxylase Structure Steers a Ferryl Oxidant toward Scission of a Strong C-H Bond. *J. Am. Chem. Soc.* 2017, 139 (29), 9855 - 9866.
141. Todd, A. E.; Orengo, C. A.; Thornton, J. M., Evolution of function in protein superfamilies, from a structural perspective. *Journal of Molecular Biology* 2001, 307 (4), 1113-1143.
142. Schrag, J. D.; Winkler, F. K.; Cygler, M., Pancreatic lipases: Evolutionary intermediates in a positional change of catalytic carboxylates? *Journal of Biological Chemistry* 1992, 267 (7), 4300-4303.
143. Hasson, M. S.; Schlichting, I.; Moulai, J.; Taylor, K.; Barrett, W.; Kenyon, G. L.; Babbitt, P. C.; Gerlt, J. A.; Petsko, G. A.; Ringe, D., Evolution of an enzyme active site: The structure of a new crystal

- form of muconate lactonizing enzyme compared with mandelate racemase and enolase. *Proc Natl Acad Sci U S A*. 1998, 95 (18), 10396-10401.
144. Wilce, M. C. J.; Board, P. G.; Feil, S. C.; Parker, M. W., Crystal structure of a theta-class glutathione transferase. *EMBO Journal* 1995, 14 (10), 2133-2143.
145. Skirgaila, R.; Grazulis, S.; Bozic, D.; Huber, R.; Siksnys, V., Structure-based redesign of the Catalytic/Metal binding site of Cfr 10I restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues1. *J. Mol. Biol.* 1998, 279 (2), 473-481.
146. Giger, L.; Caner, S.; Obexer, R.; Kast, P.; Baker, D.; Ban, N.; Hilvert, D., Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* 2013, 9 (8), 494-498.
147. Goncharenko, K. V.; Seebeck, F. P., Conversion of a non-heme iron-dependent sulfoxide synthase into a thiol dioxygenase by a single point mutation. *Chem. Commun.* 2016, 52 (9), 1945 - 1948.
148. Goncharenko, K. V.; Vit, A.; Blankenfeldt, W.; Seebeck, F. P., Structure of the Sulfoxide Synthase EgtB from the Ergothioneine Biosynthetic Pathway. *Angew. Chem. Int. Ed. Engl.* 2015, 54 (9), 2821 - 2824.
149. Liao, C.; Seebeck, F. P., Convergent Evolution of Ergothioneine Biosynthesis in Cyanobacteria. *Chembiochem* 2017, 18 (21), 2115 - 2118.
150. Gamage, A. M.; Cheah, I. K.; Chen, Y.; Lim, D. R. X.; Ku, J. W. K.; Chee, R. S. L.; Halliwell, B.; Gan, Y.-H.; Liao, C.; Seebeck, F. P.; Gengenbacher, M., The proteobacterial species *Burkholderia pseudomallei* produces ergothioneine, which enhances virulence in mammalian infection. *FASEB J* 2018, fj201800716.
151. Naowarojna, N.; Irani, S.; Hu, W.; Cheng, R.; Zhang, L.; Li, X.; Chen, J.; Zhang, Y. J.; Liu, P., Crystal Structure of the Ergothioneine Sulfoxide Synthase from *Candidatus Chloracidobacterium thermophilum* and Structure-Guided Engineering To Modulate Its Substrate Selectivity. *ACS Catalysis* 2019, 9 (8), 6955-6961.
152. Geoghegan, K. F.; Dixon, H. B. F.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; McColl, A. S.; Kamath, A. V.; Stroh, J. G., Spontaneous α -N-6-Phosphogluconoylation of a "His Tag" in *Escherichia coli*: The Cause of Extra Mass of 258 or 178 Da in Fusion Proteins. *Analytical Biochemistry* 1999, 267 (1), 169-184.
153. Braunshausen, A.; Seebeck, F. P., Identification and characterization of the first ovolthiol biosynthetic enzyme. *J Am Chem Soc* 2011, 133 (6), 1757-9.
154. Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004, 32 (5), 1792-1797.
155. Dereeper, A.; Guignon, V.; Blanc, G.; Audic, S.; Buffet, S.; Chevenet, F.; Dufayard, J. F.; Guindon, S.; Lefort, V.; Lescot, M.; Claverie, J. M.; Gascuel, O., Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 2008, 36 (suppl_2), W465-W469.
156. Battye, T. G. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. W., iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallographica Section D* 2011, 67 (4), 271-281.

157. Kabsch, W., XDS. *Acta Crystallographica Section D* 2010, 66 (2), 125-132.
158. Evans, P. R.; Murshudov, G. N., How good are my data and what is the resolution? *Acta Crystallographica Section D* 2013, 69 (7), 1204-1214.
159. McCoy, A.; Grosse-Kunstleve, R.; Adams, P.; Winn, M.; Storoni, L.; Read, R., PHASER crystallographic software. 2007; Vol. 40, p 658-674.
160. Zwart, P. H.; Afonine, P. V.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; McKee, E.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Storoni, L. C.; Terwilliger, T. C.; Adams, P. D., Automated Structure Solution with the PHENIX Suite. In *Structural Proteomics: High-Throughput Methods*, Kobe, B.; Guss, M.; Huber, T., Eds. Humana Press: Totowa, NJ, 2008; pp 419-435.
161. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallographica Section D* 2010, 66 (4), 486-501.
162. Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D* 2011, 67 (4), 355-367.
163. Adams, P. D.; Afonine, P. V.; Bunkoczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* 2010, 66 (2), 213-221.
164. Chen, V. B.; Arendall, W. B., III; 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 Crystallographica Section D* 2010, 66 (1), 12-21.
165. Krissinel, E.; Henrick, K., Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology* 2007, 372 (3), 774-797.
166. Marianayagam, N. J.; Sunde, M.; Matthews, J. M., The power of two: protein dimerization in biology. *Trends in Biochemical Sciences* 2004, 29 (11), 618-625.
167. Matthews, J. M.; Sunde, M., Dimers, Oligomers, Everywhere. In *Protein Dimerization and Oligomerization in Biology*, Matthews, J. M., Ed. Springer New York: New York, NY, 2012; pp 1-18.
168. Griffin, M. D. W.; Gerrard, J. A., The Relationship between Oligomeric State and Protein Function. In *Protein Dimerization and Oligomerization in Biology*, Matthews, J. M., Ed. Springer New York: New York, NY, 2012; pp 74-90.
169. Goodsell, D. S.; Olson, A. J., Structural Symmetry and Protein Function. *Annual Review of Biophysics and Biomolecular Structure* 2000, 29 (1), 105-153.
170. Hashimoto, K.; Panchenko, A. R., Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proceedings of the National Academy of Sciences* 2010, 107 (47), 20352-20357.
171. Mei, G.; Di Venere, A.; Rosato, N.; Finazzi-Agrò, A., The importance of being dimeric. *Febs j* 2005, 272 (1), 16-27.

172. Monod, J.; Wyman, J.; Changeux, J.-P., On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology* 1965, 12 (1), 88-118.
173. Stampfli, A. R.; Goncharenko, K. V.; Meury, M.; Dubey, B. N.; Schirmer, T.; Seebeck, F. P., An Alternative Active Site Architecture for O₂ Activation in the Ergothioneine Biosynthetic EgtB from *Chloracidobacterium thermophilum*. *Journal of the American Chemical Society* 2019, 141 (13), 5275-5285.
174. Okoniewska, M.; Tanaka, T.; Yada, R. Y., The pepsin residue glycine-76 contributes to active-site loop flexibility and participates in catalysis. *Biochem J* 2000, 349 (Pt 1), 169-177.
175. Larson, E. M.; Larimer, F. W.; Hartman, F. C., Mechanistic insights provided by deletion of a flexible loop at the active site of ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochemistry* 1995, 34 (14), 4531-4537.
176. Schneider, T. R.; Gerhardt, E.; Lee, M.; Liang, P.-H.; Anderson, K. S.; Schlichting, I., Loop Closure and Intersubunit Communication in Tryptophan Synthase. *Biochemistry* 1998, 37 (16), 5394-5406.
177. Tanaka, T.; Yamaguchi, H.; Kato, H.; Nishioka, T.; Katsube, Y.; Oda, J., Flexibility impaired by mutations revealed the multifunctional roles of the loop in glutathione synthetase. *Biochemistry* 1993, 32 (46), 12398-12404.
178. Seebeck, F. P., In vitro reconstitution of Mycobacterial ergothioneine biosynthesis. *J Am Chem Soc* 2010, 132 (19), 6632-3.
179. Chen, L.; Naowarajna, N.; Song, H.; Wang, S.; Wang, J.; Deng, Z.; Zhao, C.; Liu, P., 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.* 2018, 140 (13), 4604-4612.
180. He, F.; Hogan, S.; Latypov, R. F.; Narhi, L. O.; Razinkov, V. I., High throughput thermostability screening of monoclonal antibody formulations. *Journal of Pharmaceutical Sciences* 2010, 99 (4), 1707-1720.
181. Wanner, R.; Breitsprecher, D.; Duhr, S.; Baaske, P.; Winter, G., Thermo-Optical Protein Characterization for Straightforward Preformulation Development. *Journal of Pharmaceutical Sciences* 2017, 106 (10), 2955-2958.
182. Griffin, M. D. W.; Dobson, R. C. J.; Pearce, F. G.; Antonio, L.; Whitten, A. E.; Liew, C. K.; Mackay, J. P.; Trewthella, J.; Jameson, G. B.; Perugini, M. A.; Gerrard, J. A., Evolution of Quaternary Structure in a Homotetrameric Enzyme. *Journal of Molecular Biology* 2008, 380 (4), 691-703.
183. Misson, L.; Burn, R.; Vit, A.; Hildesheim, J.; Beliaeva, M. A.; Blankenfeldt, W.; Seebeck, F. P., Inhibition and Regulation of the Ergothioneine Biosynthetic Methyltransferase EgtD. *ACS Chemical Biology* 2018, 13 (5), 1333-1342.
184. Hashimoto, K.; Madej, T.; Bryant, S. H.; Panchenko, A. R., Functional States of Homooligomers: Insights from the Evolution of Glycosyltransferases. *Journal of Molecular Biology* 2010, 399 (1), 196-206.

185. Yamashita, M.; Yamashita, Y.; Ando, T.; Wakamiya, J.; Akiba, S., Identification and Determination of Selenoneine, 2-Selenyl-N α , N α , N α -Trimethyl-L-Histidine, as the Major Organic Selenium in Blood Cells in a Fish-Eating Population on Remote Japanese Islands. *Biological Trace Element Research* 2013, 156 (1), 36-44.
186. Yamashita, Y.; Yabu, T.; Yamashita, M., Discovery of the strong antioxidant selenoneine in tuna and selenium redox metabolism. *World J Biol Chem* 2010, 1 (5), 144-150.
187. Yamashita, M.; Yamashita, Y.; Suzuki, T.; Kani, Y.; Mizusawa, N.; Imamura, S.; Takemoto, K.; Hara, T.; Hossain, M. A.; Yabu, T.; Touhata, K., Selenoneine, a Novel Selenium-Containing Compound, Mediates Detoxification Mechanisms against Methylmercury Accumulation and Toxicity in Zebrafish Embryo. *Marine Biotechnology* 2013, 15 (5), 559-570.
188. Nauser, T.; Steinmann, D.; Koppenol, W. H., Why do proteins use selenocysteine instead of cysteine? *Amino Acids* 2012, 42 (1), 39-44.
189. Luther, G. W.; Findlay, A.; MacDonald, D.; Owings, S.; Hanson, T.; Beinart, R.; Girguis, P., Thermodynamics and Kinetics of Sulfide Oxidation by Oxygen: A Look at Inorganically Controlled Reactions and Biologically Mediated Processes in the Environment. *Frontiers in Microbiology* 2011, 2 (62).
190. Lim, D.; Gründemann, D.; Seebeck, F. P., Total Synthesis and Functional Characterization of Selenoneine. *Angewandte Chemie International Edition* 2019, 58 (42), 15026-15030.
191. Turrini, N. G.; Kroepfl, N.; Jensen, K. B.; Reiter, T. C.; Francesconi, K. A.; Schwerdtle, T.; Kroutil, W.; Kuehnelt, D., Biosynthesis and isolation of selenoneine from genetically modified fission yeast. *Metallomics* 2018, 10 (10), 1532-1538.
192. Lu, J.; Holmgren, A., Selenoproteins. *J. Biol. Chem.* 2009, 284 (2), 723 - 727.
193. Zhang, Y.; Turanov, A. A.; Hatfield, D. L.; Gladyshev, V. N., In silico identification of genes involved in selenium metabolism: evidence for a third selenium utilization trait. *BMC Genomics* 2008, 9 (1), 251.
194. Romero, H.; Zhang, Y.; Gladyshev, V. N.; Salinas, G., Evolution of selenium utilization traits. *Genome Biology* 2005, 6 (8), R66.
195. Veres, Z.; Kim, I. Y.; Scholz, T. D.; Stadtman, T. C., Selenophosphate synthetase. Enzyme properties and catalytic reaction. *Journal of Biological Chemistry* 1994, 269 (14), 10597-603.
196. Veres, Z.; Tsai, L.; Scholz, T. D.; Politino, M.; Balaban, R. S.; Stadtman, T. C., Synthesis of 5-methylaminomethyl-2-selenouridine in tRNAs: 31P NMR studies show the labile selenium donor synthesized by the selD gene product contains selenium bonded to phosphorus. *Proceedings of the National Academy of Sciences of the United States of America* 1992, 89 (7), 2975-2979.
197. Meury, M.; Knop, M.; Seebeck, F. P., Structural Basis for Copper-Oxygen Mediated C-H Bond Activation by the Formylglycine-Generating Enzyme. *Angew. Chem., Int. Ed.* 2017, 56 (28), 8115-8119.
198. Schiefner, A.; Breed, J.; Boesser, L.; Kneip, S.; Gade, J.; Holtmann, G.; Diederichs, K.; Welte, W.; Bremer, E., Cation- π interactions as determinants for binding of the compatible solutes glycine

- betaine and proline betaine by the periplasmic ligand-binding protein ProX from *Escherichia coli*. *J. Biol. Chem.* 2004, 279 (7), 5588-5596.
199. Pluskal, T.; Ueno, M.; Yanagida, M., Genetic and metabolomic dissection of the ergothioneine and selenoneine biosynthetic pathway in the fission yeast, *S. pombe*, and construction of an overproduction system. *PLoS One* 2014, 9 (5), e97774/1-e97774/12, 12 pp.
200. Niesen, F. H.; Berglund, H.; Vedadi, M., The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* 2007, 2 (9), 2212-2221.
201. Vedadi, M.; Niesen, F. H.; Allali-Hassani, A.; Fedorov, O. Y.; Finerty, P. J.; Wasney, G. A.; Yeung, R.; Arrowsmith, C.; Ball, L. J.; Berglund, H.; Hui, R.; Marsden, B. D.; Nordlund, P.; Sundstrom, M.; Weigelt, J.; Edwards, A. M., Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proceedings of the National Academy of Sciences* 2006, 103 (43), 15835-15840.
202. Glass, R. S.; Singh, W. P.; Jung, W.; Veres, Z.; Scholz, T. D.; Stadtman, T., Monoselenophosphate: Synthesis, characterization, and identity with the prokaryotic biological selenium donor, compound SePX. *Biochemistry* 1993, 32 (47), 12555-12559.
203. Liang, D.-M.; Liu, J.-H.; Wu, H.; Wang, B.-B.; Zhu, H.-J.; Qiao, J.-J., Glycosyltransferases: mechanisms and applications in natural product development. *Chemical Society Reviews* 2015, 44 (22), 8350-8374.
204. Schmid, J.; Heider, D.; Wendel, N. J.; Sperl, N.; Sieber, V., Bacterial Glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products. *Frontiers in microbiology* 2016, 7, 182-182.
205. Gloster, T. M., Advances in understanding glycosyltransferases from a structural perspective. *Current Opinion in Structural Biology* 2014, 28, 131-141.
206. Sánchez-Rodríguez, A.; Tytgat, H. L. P.; Winderickx, J.; Vanderleyden, J.; Lebeer, S.; Marchal, K., A network-based approach to identify substrate classes of bacterial glycosyltransferases. *BMC genomics* 2014, 15 (1), 349-349.
207. Sheridan, K. J.; Lechner, B. E.; Keeffe, G. O.; Keller, M. A.; Werner, E. R.; Lindner, H.; Jones, G. W.; Haas, H.; Doyle, S., Ergothioneine Biosynthesis and Functionality in the Opportunistic Fungal Pathogen, *Aspergillus fumigatus*. *Sci. Rep.* 2016, 6, 35306.
208. Badiei, Y. M.; Siegler, M. A.; Goldberg, D. P., O₂ Activation by Bis(imino)pyridine Iron(II)-Thiolate Complexes. *Journal of the American Chemical Society* 2011, 133 (5), 1274-1277.
209. Che, X.; Gao, J.; Liu, Y.; Liu, C., Metal vs. chalcogen competition in the catalytic mechanism of cysteine dioxygenase. *J. Inorg. Biochem.* 2013, 122, 1-7.
210. Lin, I. J.; Gebel, E. B.; Machonkin, T. E.; Westler, W. M.; Markley, J. L., Changes in hydrogen-bond strengths explain reduction potentials in 10 rubredoxin variants. *Proceedings of the National Academy of Sciences of the United States of America* 2005, 102 (41), 14581.
211. Dey, A.; Okamura, T.-a.; Ueyama, N.; Hedman, B.; Hodgson, K. O.; Solomon, E. I., Sulfur K-Edge XAS and DFT Calculations on P450 Model Complexes: Effects of Hydrogen Bonding on Electronic Structure and Redox Potentials. *Journal of the American Chemical Society* 2005, 127 (34), 12046-12053.

212. Li, H.; Webb, S. P.; Ivanic, J.; Jensen, J. H., Determinants of the Relative Reduction Potentials of Type-1 Copper Sites in Proteins. *Journal of the American Chemical Society* 2004, 126 (25), 8010-8019.
213. Dong, S.; Ybe, J. A.; Hecht, M. H.; Spiro, T. G., H-Bonding Maintains the Active Site of Type 1 Copper Proteins: Site-Directed Mutagenesis of Asn38 in Poplar Plastocyanin. *Biochemistry* 1999, 38 (11), 3379-3385.
214. Birrell, J. A.; Laurich, C.; Reijerse, E. J.; Ogata, H.; Lubitz, W., Importance of Hydrogen Bonding in Fine Tuning the [2Fe-2S] Cluster Redox Potential of HydC from *Thermotoga maritima*. *Biochemistry* 2016, 55 (31), 4344-4355.
215. Limited, G. P. Grade Web Server. <http://grade.globalphasing.org/cgi-bin/grade/server.cgi>.
216. Tschirka, J.; Kreisor, M.; Betz, J.; Gründemann, D., Substrate Selectivity Check of the Ergothioneine Transporter. *Drug Metabolism and Disposition* 2018, 46 (6), 779.
217. Maurer, A.; Leisinger, F.; Lim, D.; Seebeck, F. P., Structure and Mechanism of Ergothionease from *Treponema denticola*. *Chemistry* 2019.
218. Servillo, L.; D'Onofrio, N.; Casale, R.; Giovane, A.; Cautela, D.; Castaldo, D.; Balestrieri, M. L., Ergothioneine products derived by superoxide oxidation in endothelial cells exposed to high-glucose. *Free Radic Biol Med* 2017, 108, 8-18.
219. Servillo, L.; Castaldo, D.; Cautela, D.; Casale, R.; D'Onofrio, N.; Giovane, A.; Balestrieri, M. L., An uncommon redox behavior sheds light on the cellular antioxidant properties of ergothioneine. *Free Radic Biol Med* 2015, 79, 228-36.
220. Cheah, I. K.; Tang, R. M. Y.; Yew, T. S. Z.; Lim, K. H. C.; Halliwell, B., Administration of Pure Ergothioneine to Healthy Human Subjects: Uptake, Metabolism, and Effects on Biomarkers of Oxidative Damage and Inflammation. *Antioxid. Redox Signaling* 2017, 26 (5), 193-206.
221. Wilkinson, A. J.; Verschueren, K. H. G. In *Crystal structures of periplasmic solute-binding proteins in ABC transport complexes illuminate their function*, Elsevier Science Ltd.: 2003; pp 187-207.
222. Tam, R.; Saier, M. H., Jr., Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol Rev* 1993, 57 (2), 320-46.
223. Smits, S. H. J.; Hoeing, M.; Lecher, J.; Jebbar, M.; Schmitt, L.; Bremer, E., The compatible-solute-binding protein OpuAC from *Bacillus subtilis*: ligand binding, site-directed mutagenesis, and crystallographic studies. *J. Bacteriol.* 2008, 190 (16), 5663-5671.
224. Horn, C.; Sohn-Boesser, L.; Breed, J.; Welte, W.; Schmitt, L.; Bremer, E., Molecular Determinants for Substrate Specificity of the Ligand-binding Protein OpuAC from *Bacillus subtilis* for the Compatible Solutes Glycine Betaine and Proline Betaine. *J. Mol. Biol.* 2006, 357 (2), 592-606.
225. Du, Y.; Shi, W.-W.; He, Y.-X.; Yang, Y.-H.; Zhou, C.-Z.; Chen, Y., Structures of the substrate-binding protein provide insights into the multiple compatible solute binding specificities of the *Bacillus subtilis* ABC transporter OpuC. *Biochem. J.* 2011, 436 (2), 283-289.
226. Pittelkow, M.; Tschapek, B.; Smits, S. H. J.; Schmitt, L.; Bremer, E., The Crystal Structure of the Substrate-Binding Protein OpuBC from *Bacillus subtilis* in Complex with Choline. *Journal of Molecular Biology* 2011, 411 (1), 53-67.

227. Kappes, R. M.; Kempf, B.; Kneip, S.; Boch, J.; Gade, J.; Meier-Wagner, J.; Bremer, E., Two evolutionarily closely related ABC transporters mediate the uptake of choline for synthesis of the osmoprotectant glycine betaine in *Bacillus subtilis*. *Molecular Microbiology* 1999, 32 (1), 203-216.
228. May, G.; Faatz, E.; Villarejo, M.; Bremer, E., Binding protein dependent transport of glycine betaine and its osmotic regulation in *Escherichia coli* K12. *Mol Gen Genet* 1986, 205 (2), 225-33.
229. Haardt, M.; Kempf, B.; Faatz, E.; Bremer, E., The osmoprotectant proline betaine is a major substrate for the binding-protein-dependent transport system ProU of *Escherichia coli* K-12. *Mol Gen Genet* 1995, 246 (6), 783-6.
230. Barron, A.; Jung, J. U.; Villarejo, M., Purification and characterization of a glycine betaine binding protein from *Escherichia coli*. *J. Biol. Chem.* 1987, 262 (24), 11841-6.
231. Schiefner, A.; Holtmann, G.; Diederichs, K.; Welte, W.; Bremer, E., Structural Basis for the Binding of Compatible Solutes by ProX from the Hyperthermophilic Archaeon *Archaeoglobus fulgidus*. *J. Biol. Chem.* 2004, 279 (46), 48270-48281.
232. Harding, M., Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallographica Section D* 2002, 58 (5), 872-874.
233. Harding, M., Geometry of metal-ligand interactions in proteins. *Acta Crystallographica Section D* 2001, 57 (3), 401-411.
234. Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L., Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 2015, 1854 (8), 1019-1037.
235. Du, Y.; Shi, W.-W.; He, Y.-X.; Yang, Y.-H.; Zhou, C.-Z.; Chen, Y., Structures of the substrate-binding protein provide insights into the multiple compatible solute binding specificities of the *Bacillus subtilis* ABC transporter OpuC. *Biochemical Journal* 2011, 436 (2), 283.
236. Kappes, R. M.; Bremer, E., Response of *Bacillus subtilis* to high osmolarity: uptake of carnitine, crotonobetaine and γ -butyrobetaine via the ABC transport system OpuC. *Microbiology* 1998, 144 (1), 83-90.
237. Kappes, R. M.; Kempf, B.; Bremer, E., Three transport systems for the osmoprotectant glycine betaine operate in *Bacillus subtilis*: characterization of OpuD. *Journal of Bacteriology* 1996, 178 (17), 5071.
238. Nau-Wagner, G.; Boch, J.; Le Good, J. A.; Bremer, E., High-Affinity Transport of Choline-O-Sulfate and Its Use as a Compatible Solute in *Bacillus subtilis*. *Applied and Environmental Microbiology* 1999, 65 (2), 560.
239. Wolters, J. C.; Berntsson, R. P. A.; Gul, N.; Karasawa, A.; Thunnissen, A.-M. W. H.; Slotboom, D.-J.; Poolman, B., Ligand Binding and Crystal Structures of the Substrate-Binding Domain of the ABC Transporter OpuA. *PLOS ONE* 2010, 5 (4), e10361.

240. Fraser, K. R.; Harvie, D.; Coote, P. J.; Byrne, C. P., Identification and characterization of an ATP binding cassette L-carnitine transporter in *Listeria monocytogenes*. *Applied and Environmental Microbiology* 2000, 66 (11), 4696.
241. Verheul, A.; Glaasker, E.; Poolman, B.; Abee, T., Betaine and L-carnitine transport by *Listeria monocytogenes* Scott A in response to osmotic signals. *Journal of bacteriology* 1997, 179 (22), 6979-6985.
242. Verheul, A.; Rombouts, F. M.; Beumer, R. R.; Abee, T., An ATP-dependent L-carnitine transporter in *Listeria monocytogenes* Scott A is involved in osmoprotection. *Journal of Bacteriology* 1995, 177 (11), 3205.
243. Schiefner, A.; Holtmann, G.; Diederichs, K.; Welte, W.; Bremer, E., Structural Basis for the Binding of Compatible Solutes by ProX from the Hyperthermophilic Archaeon *Archaeoglobus fulgidus*. *Journal of Biological Chemistry* 2004, 279 (46), 48270-48281.
244. Tschapek, B.; Pittelkow, M.; Sohn-Boesser, L.; Holtmann, G.; Smits, S. H. J.; Gohlke, H.; Bremer, E.; Schmitt, L., Arg149 Is Involved in Switching the Low Affinity, Open State of the Binding Protein AfProX into Its High Affinity, Closed State. *J. Mol. Biol.* 2011, 411 (1), 36-52.
245. Choquet, G.; Jehan, N.; Pissavin, C.; Blanco, C.; Jebbar, M., OusB, a Broad-Specificity ABC-Type Transporter from *Erwinia chrysanthemi*, Mediates Uptake of Glycine Betaine and Choline with a High Affinity. *Applied and Environmental Microbiology* 2005, 71 (7), 3389.
246. Gerhardt, P. N.; Smith, L. T.; Smith, G. M., Sodium-driven, osmotically activated glycine betaine transport in *Listeria monocytogenes* membrane vesicles. *Journal of Bacteriology* 1996, 178 (21), 6105.
247. Ko, R.; Smith, L. T.; Smith, G. M., Glycine betaine confers enhanced osmotolerance and cryotolerance on *Listeria monocytogenes*. *Journal of Bacteriology* 1994, 176 (2), 426.
248. Mendum, M. L.; Smith, L. T., Gbu glycine betaine porter and carnitine uptake in osmotically stressed *Listeria monocytogenes* cells. *Applied and Environmental Microbiology* 2002, 68 (11), 5647.
249. Dupont, L.; Garcia, I.; Poggi, M.-C.; Alloing, G.; Mandon, K.; Le Rudulier, D., The *Sinorhizobium meliloti* ABC transporter Cho is highly specific for choline and expressed in bacteroids from *Medicago sativa* nodules. *Journal of bacteriology* 2004, 186 (18), 5988-5996.
250. Javaux, C.; Joris, B.; De Witte, P., Functional Characteristics of TauA Binding Protein from TauABC *Escherichia coli* System. *The Protein Journal* 2007, 26 (4), 231-238.
251. Suzek, B. E.; Wang, Y.; Huang, H.; McGarvey, P. B.; Wu, C. H.; Consortium, t. U., UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 2014, 31 (6), 926-932.
252. Haardt, M.; Kempf, B.; Faatz, E.; Bremer, E., The osmoprotectant proline betaine is a major substrate for the binding-protein-dependent transport system ProU of *Escherichia coli* K-12. *Molecular and General Genetics MGG* 1995, 246 (6), 783-796.
253. Ko, R.; Smith, L. T., Identification of an ATP-driven, osmoregulated glycine betaine transport system in *Listeria monocytogenes*. *Applied and Environmental Microbiology* 1999, 65 (9), 4040.

254. Gerhardt, P. N. M.; Tombras Smith, L.; Smith, G. M., Osmotic and chill activation of glycine betaine porter II in *Listeria monocytogenes* membrane vesicles. *Journal of Bacteriology* 2000, 182 (9), 2544.
255. Angelidis, A. S.; Smith, G. M., Three transporters mediate uptake of glycine betaine and carnitine by *Listeria monocytogenes* in response to hyperosmotic stress. *Applied and Environmental Microbiology* 2003, 69 (2), 1013.
256. Hoffmann, T.; Warmbold, B.; Smits, S. H. J.; Tschapek, B.; Ronzheimer, S.; Bashir, A.; Chen, C.; Rolbetzki, A.; Pittelkow, M.; Jebbar, M.; Seubert, A.; Schmitt, L.; Bremer, E., Arsenobetaine: an ecophysiologicaly important organoarsenical confers cytoprotection against osmotic stress and growth temperature extremes. *Environmental Microbiology* 2018, 20 (1), 305-323.
257. Lang, S.; Cressatti, M.; Mendoza, K. E.; Coumoundouros, C. N.; Plater, S. M.; Culham, D. E.; Kimber, M. S.; Wood, J. M., YehZYXW of *Escherichia coli* Is a Low-Affinity, Non-Osmoregulatory Betaine-Specific ABC Transporter. *Biochemistry* 2015, 54 (37), 5735-5747.
258. van der Ploeg, J. R.; Weiss, M. A.; Saller, E.; Nashimoto, H.; Saito, N.; Kertesz, M. A.; Leisinger, T., Identification of sulfate starvation-regulated genes in *Escherichia coli*: a gene cluster involved in the utilization of taurine as a sulfur source. *Journal of Bacteriology* 1996, 178 (18), 5438.
259. Frossard, S. M.; Khan, A. A.; Warrick, E. C.; Gately, J. M.; Hanson, A. D.; Oldham, M. L.; Sanders, D. A.; Csonka, L. N., Identification of a third osmoprotectant transport system, the *osmU* system, in *Salmonella enterica*. *Journal of Bacteriology* 2012, 194 (15), 3861.
260. Horn, C.; Sohn-Bösser, L.; Breed, J.; Welte, W.; Schmitt, L.; Bremer, E., Molecular Determinants for Substrate Specificity of the Ligand-binding Protein OpuAC from *Bacillus subtilis* for the Compatible Solutes Glycine Betaine and Proline Betaine. *Journal of Molecular Biology* 2006, 357 (2), 592-606.
261. Kempf, B.; Bremer, E., OpuA, an Osmotically Regulated Binding Protein-dependent Transport System for the Osmoprotectant Glycine Betaine in *Bacillus subtilis*. *Journal of Biological Chemistry* 1995, 270 (28), 16701-16713.
262. Oswald, C.; Smits, S. H. J.; Höing, M.; Sohn-Bösser, L.; Dupont, L.; Le Rudulier, D.; Schmitt, L.; Bremer, E., Crystal Structures of the Choline/Acetylcholine Substrate-binding Protein ChoX from *Sinorhizobium meliloti* in the Liganded and Unliganded-Closed States. *Journal of Biological Chemistry* 2008, 283 (47), 32848-32859.
263. Oswald, C.; Smits Sander, H. J.; Höing, M.; Bremer, E.; Schmitt, L., Structural analysis of the choline-binding protein ChoX in a semi-closed and ligand-free conformation. *bchm* 2009, 390 (11), 1163.
264. Ruiz, J. S.; Schuurman-Wolters, K. G.; Poolman, B., Crystal Structure of the Substrate-Binding Domain from *Listeria monocytogenes* Bile-Resistance Determinant Bile. *Crystals* 2016, 6 (12).

Acknowledgements

First and foremost and I would like to thank Prof. Florian Seebeck for giving me a position in his group and providing me with a challenging project to work on. I have really valued your creativity, and openness to new ideas and suggestions. I particularly want to thank you for your support in figuring out my next step. I'm looking forward to see the exciting things that will come out of the group in the future.

I would like to thank Prof. Wulf Blankenfeldt for accepting to be my co-examiner, taking the journey to Basel and reading my thesis.

A big thank you to Prof. Jonathan De Roo for accepting to be chair.

I want to thank Prof. Tilman Schirmer for introducing me to crystallography and for valuable comments in the crafting of the first EgtB Story.

A huge thank you to all the past and current members of the Seebeck group: Leti, Roxana, Pascal, Matthias, Marcel, Kristina, Sebastian, Reto, Tom, Niko, Liao, Davey, Alice, Dima, Mariia, Florian L., Jiaming, Xiaojin, Egor and Camille. I feel fortunate to have been able to work with such great colleagues and friends. It takes a special group of people to discuss bean/bin bags for over an hour. Special thanks go to Mariia for our collaboration on the SBP project, I think working together on the project worked really smoothly and efficiently. Further thanks to Mariia for all her help with ITC, running DSF measurements, and willingness to help out at a moments notice. To my fellow kiwi Davey, I'm so grateful to have had you in the group – finally someone who makes my accent seem ok by comparison. Thank you for all the awful puns, clothes-lining, kiwi-banter, discussions and English-editing. To Liao (猪猪叔叔) thank you for helping me out with the EgtB kinetics, and being the go to person in the lab for any chemistry problem. I've enjoyed all of our hiking and skiing trips together (particularly the first ski trip - the snowboarding not so much). To Reto, I'm glad to have started our PhD's together, thank you for the discussions about science and life. To Sebastian or Siibi I'm glad we could work on a couple projects together, you and your ants were fun desk neighbors for the first few years. To Flo or little Florian: Best masters student ever! and an amazing cryptographer. I'm grateful that we were desk neighbors (maybe you were less so). Thanks for all the jokes, crystallographic discussions and being the person who heard the most of my frustrations. Further thanks to my other friends and colleagues across the chemistry department from teaching and the PCC.

I would also like to thank the members of the Schirmer group for many good times and scientific discussions. With this comes the Maier group, in particular Roman, for looking after all things crystallography, and incredible knowledge on the topic. I thank both groups for advice on all things biophysics, some weird late night synchrotron trips and fun aperos.

Acknowledgements

A huge thank you to my third group, the Bollinger-Krebs group at Penn State, especially Marty and Carsten. I really enjoyed my time in Penn State - it was an invaluable experience learning about iron-oxygenases from the best and gaining experience in SF! Also thank you for bringing me back to help out and take part in the 2018 Penn State Workshop. Special thanks to Carsten, Eva, Bo, Chris, Beth and Juan for all the fond memories both inside and outside of the lab.

I am very thankful for my two proof readers, Davey and Mum, without you this thesis would not be what it is. I'm grateful for your input and the short notice at which you were able to edit everything.

A big thank you to the secretaries, werkstatt and analytical team who keep the department running and working well.

Special thanks goes to the Biozentrum Ladies: Gohanna, Anna, Francesca, Freddy and Leonie. I'm incredibly grateful for you all. Our ladies evenings and coffee tea/breaks have been a real crutch throughout my PhD. I've loved all the scientific discussions, advice, venting, hot seats, crémant and sushi evenings. I'm lucky to be surrounded by such inspiring, strong women. I can't wait to see where we all end up.

Thank you to my Kiwi friends, especially Krissy, who have supported me from afar and have had to put up with my sporadic contact and sparse visits over the past five years. A special thank you to those who visited me here in Switzerland. I can't wait to spend a proper few months with you all again.

To my Swiss family (mami und papi 2-6), thank you for all of the support, particularly helping me get established when I first moved to Switzerland. I felt immediately taken up in the family and was always grateful for the mini breaks in Solothurn to get from the stress of Basel. Thank you to Lisbeth for all of our wonderful trips together. To Rani and Thomas, thank you for being so welcoming and supportive. Thank you to Mum, Dad and Elisa for their support from afar. I'm very lucky to have grown up in such a supportive family with endless patience.

To Raj, thank you for all the love and support over the past years. Especially over the past few months while we have both been finishing our PhD's - I've been pleasantly surprised at how well we have handled the stress of finishing within a month of one another. I'm incredibly grateful for you and your endless support. Du bist das Beste!

Curriculum vitae

ANJA STAMPFLI

✉ anja.stampfli@hotmail.com

Nationality: New Zealand & Swiss

EDUCATION

- Jan. 2020 **PhD in Chemistry, *summa cum laude***
Supervised by Prof. Florian Seebeck, Department of Chemistry, Universität Basel (CH).
- Jan. 2014 **BSc. (hons), First Class** (equivalent to European Masters)
Supervised by Prof. Emily Parker, Department of Chemistry, University of Canterbury (NZ).
- Apr. 2014 **BSc. - Double major in Chemistry and Biochemistry**
(GPA: 8.6/9) University of Canterbury (NZ).

RESEARCH EXPERIENCE (in addition to Education listed above)

- Jul. 2017 – Oct. 2017 **Research Exchange to Bollinger-Krebs Group, Penn State University (USA)**
A four month exchange to learn and utilize stopped-flow, freeze-quench methods & Mössbauer spectroscopy in the lab of Prof. Carsten Krebs & Prof. Martin Bollinger at Penn State University, PA (USA).
- Summer 2013 - 2014 **UC Summer Research Scholarship: 'Discovering markers for diabetic complications.'**
Supervised by Assoc. Prof. Renwick Dobson, University of Canterbury & Biomolecular Interaction Centre, University of Canterbury (NZ).
- 2013 **Bachelor Thesis:** Structural characterisation and bioinformatics analysis of N-acetylmannosamine kinase (NanK) from the sialic acid catabolic pathway in MRSA. Supervised by Assoc. Prof. Renwick Dobson, Biomolecular Interaction Centre, University of Canterbury (NZ).
- Summer 2012 – 2013 **Summer Studentship** examining a novel Tyrosine Aminotransferase. Supervised by Assoc. Prof. Renwick Dobson, Biological Sciences, University of Canterbury (NZ).

PUBLICATIONS

- BBA*, 2021 M. Lassé; **A. Stampfli**; T. Orban; R. K. Bothara; J. Gerrard; A. Fairbanks; N. Pattinson & R. Dobson. Reaction dynamics and residue identification of haemoglobin modification by acrolein, a lipid-peroxidation by-product, *BBA.*, 12.
- Curr Opin Chem Biol.*, 2020 **A. Stampfli**, F. Seebeck. The catalytic mechanism of sulfoxide synthases (2020) *Curr Opin Chem Biol.*, 59.
- Curr. Opin. Struct. Biol.*, 2020 **A. Stampfli**, W. Blankenfeldt, F. Seebeck. Structural basis of ergothioneine biosynthesis (2020) *Curr. Opin. Struct. Biol.*, 65.
- Chem. Eur. J.*, 2020 K. Goncharenko, S. Flüchiger, C. Liao, D. Lim, **A. Stampfli** & F. Seebeck. Selenocysteine as a substrate, an inhibitor and a mechanistic probe for bacterial and fungal iron-dependent sulfoxide synthases. (2020) *Chem. Eur. J.*, 26.
- J. Am. Chem. Soc.*, 2019 **A. Stampfli**, K. Goncharenko, M. Meury, B. Dubey, T. Schirmer & F. Seebeck. An alternative active site architecture for O₂ activation in the ergothioneine biosynthetic EgtB from *Chloracidobacterium thermophilum* (2019) *J. Am. Chem. Soc.*, 141.

- Acta Crystal*, 2014 R. A. North, S. Seizova, A. Stampfli, S. A. Kessans, H. Suzuki, M. D. W. Griffin, M. Kvensakul & R. C. J. Dobson. Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of N-acetylmannosamine kinase from methicillin-resistant *Staphylococcus aureus*. (2014) *Acta Crystal*. 70.

TEACHING & SUPERVISION

- 2018 Teacher for stopped-flow at the 2018 Penn State Bioinorganic Workshop (US)
2016 – 2018 Supervision of Protein Production Courses, Universität Basel (CH) (4 weeks)
2016 – 2018 Lab practical teacher of Organic Chemistry, Universität Basel (CH) (500 + hrs)
2017 Supervision of a Masters Student, Seebeck Group, Universität Basel (CH)
2014 Laboratory demonstrator in General Chemistry, University of Canterbury (NZ)
< 2014 Three years experience in tutoring high school chemistry and biology (NZ)

AWARDS

- 2021 Unibas Chemistry Departmental Doctorate Award 2021 (CH).
2019 SNF Early. Post doc Mobility Fellowship (CH).
2017 antelope@university. A competitive career program for female doctoral students, including three program elements: expert exchange, career camp and coaching (CH).
2016 Chemistry Department 2016 Christmas Symposium– Poster Prize
2014 UC Prize for top graded third year Chemistry Student - best overall grades University of Canterbury Senior Scholarship
Dean of Science Congratulatory Letter, University of Canterbury
2013 University of Canterbury Summer Research Scholarship
Canterbury Scientific Ltd. Emerging Leader Scholarship

ORAL PRESENTATIONS

- 2018 Chemical Biology Conference, Geneva (CH), Satellite event
2017 PhD Chemistry Community (PCC) Research Seminar
2016 Biozentrum PhD Students – PhD Club Presentation

POSTER PRESENTATIONS

- 2018 Penn State Frontiers in Metallobiochemistry Symposium (USA)
2018 Chemical Biology Conference, Geneva (CH)
2016 Swiss Chemical Society Fall Meeting, Zurich (CH)
Christmas Symposium, Basel (CH)

TRAININGS & WORKSHOPS

- 2019 Project Management - Toolbox for Scientists, Manage Science - 2 day workshop (CH)
2019 Successful Team Leadership Workshop, RTTA - 2 days, Basel (CH).
2019 Successful Team Leadership Workshop, RTTA - 2 day workshop (CH).
2018 Penn State Bioinorganic Workshop – 6 day workshop (USA).
2018 Negotiating with Success, Negoservices – 2 day workshop (CH).
2018 Publishing Research Articles: Strategies and Steps – 1 day workshop (CH).
2017 antelope@university
A successful career program for female doctoral students, including three program elements: expert exchange, career camp and coaching (CH).

- 2016 **Macromolecular Crystallography School - MCS2016**
School/Workshop on Advanced Crystallographic techniques, held at the CBE (Department of Crystallography and Structural Biology) of the Institute of Physical-Chemistry "Rocasolano", CSIC (Spanish National Research Council), in Madrid (ES).

RESPONSIBILITIES

- 2017 - 2018 Postgraduate Chemistry Committee, Universität Basel - Committee member
2016 - present Gryphon (Art Robbins) Crystallisation Robot maintenance and trainer
2014 University of Canterbury Chemsoc (Chem. society) committee member (NZ)

