

# Structure of the ergothioneine-biosynthetic amidohydrolase EgtC

Allegra Vit,<sup>[b,c]</sup> Gabriel T. Mashabela,<sup>[a]</sup> Wulf Blankenfeldt,<sup>[b,c]</sup> and Florian P. Seebeck<sup>[a]\*</sup>

**Abstract:** The ubiquitous sulfur metabolite ergothioneine is biosynthesized by oxidative attachment of a sulfur atom to the imidazole ring of N $\alpha$ -trimethylhistidine. Most actinobacteria, including *Mycobacterium tuberculosis*, use  $\gamma$ -glutamyl cysteine as a sulfur donor. In subsequent steps the carbon scaffold of  $\gamma$ -glutamyl cysteine is removed by the glutamine amidohydrolase EgtC and the  $\beta$ -lyase EgtE. We solved the crystal structure of EgtC from *M. smegmatis* in complex with its physiological substrate. The active site of EgtC is surprisingly conserved even in homologs that are clearly not involved in ergothioneine production. The phylogenetic distribution of EgtC-like enzymes indicate that their last common ancestor may have emerged for a different purpose than ergothioneine production.

## Introduction

Ergothioneine (**1**, Figure 1) is a sulfur containing histidine derivative that occurs in plants, fungi, animal including humans and a broad range of bacteria. Because of its 2-mercaptoimidazole side chain, ergothioneine has long been suspected to play a role in cellular redox homeostasis.<sup>[1]</sup> Several recent studies on the cytoprotective effects of ergothioneine on human cells,<sup>[2]</sup> in zebrafish<sup>[3]</sup> in lower eukaryotes<sup>[4]</sup> and in bacteria<sup>[5]</sup> provide growing support for this idea.<sup>[6]</sup> Furthermore, the latest discovery that ergothioneine also serves as a cofactor in bacterial secondary metabolite biosynthesis<sup>[7]</sup> indicates that this molecule may be even more versatile.

As a prerequisite to discover such functions, we started to characterize the ergothioneine biosynthetic machinery in Mycobacteria.<sup>[8]</sup> Almost all sequenced actinobacteria, including the pathogen *Mycobacterium tuberculosis* or the saprophyte *Mycobacterium smegmatis*, contain a five-gene cluster (egtABCDE) which enables the cells to produce ergothioneine from the primary metabolites cysteine,

glutamate, histidine and S-adenosylmethionine (SAM, Figure 1).<sup>[8a]</sup> The  $\gamma$ -glutamyl cysteine ( $\gamma$ GC) synthase EgtA<sup>[9]</sup> and the SAM-dependent histidine methyltransferase EgtD<sup>[8b]</sup> provide the substrates for EgtB-catalyzed production of the central intermediate  $\gamma$ -glutamyl cysteine sulfoxide trimethyl histidinyl conjugate (sulfoxide **3**, Figure 1). This compound is substrate to the amidohydrolase EgtC, which cleaves the  $\gamma$ -glutamyl tail to produce sulfoxide **4** (Figure 1), which in turn is converted to ergothioneine by the pyridoxal 5-phosphate-dependent  $\beta$ -lyase EgtE.<sup>[8a]</sup> This pathway is not universal. For example, *Neurospora crassa* and *Schizosaccharomyces pombe* use cysteine directly as sulfur donor to make sulfoxide **4** and consequently bypass the EgtC catalyzed step (gray, Figure 1).<sup>[10]</sup> Given that the complete *egtABCDE* gene cluster<sup>[11]</sup> is an exclusive trait of actinobacterial genomes it may be even possible that the five step biosynthetic route is largely limited to this phylum. On the other hand, EgtC-like enzymes frequently occur in proteobacteria, cyanobacteria and fungi, including *N. crassa*. It is impossible to infer from primary sequence data alone whether these homologs are bona fide EgtCs or whether they serve unrelated functions.

Sequence homology places EgtC in a large superfamily of Ntn-hydrolases (CL0052) named for their common use of an N-terminal serine, threonine or cysteine residue as catalytic nucleophile.<sup>[12]</sup> This superfamily comprises the proteasome,<sup>[13]</sup> penicillin amidase,<sup>[14]</sup>  $\gamma$ -glutamyl-transpeptidase<sup>[15]</sup> and many amidohydrolases that remove glutamyl- or asparaginyl-residues from ammonia or more complex amines.<sup>[16]</sup> Despite the common Ntn-hydrolase fold and similar catalytic activities, members of this superfamily do not necessarily share recognizable sequence homology.<sup>[16]</sup> In keeping this family tradition, EgtC shares almost no recognizable sequence similarity with any other protein for which both structure and function are known.

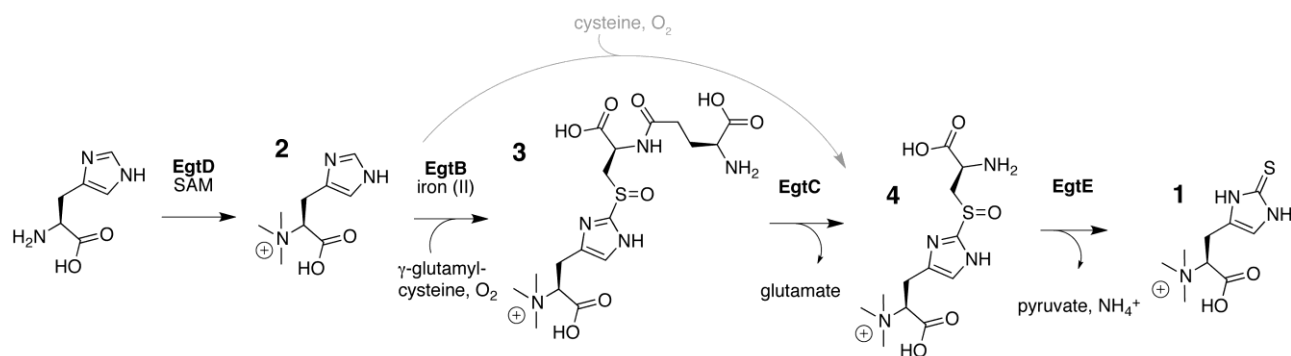
In this paper we report the crystal structure of mycobacterial EgtC in complex with either the substrate sulfoxide **3**, or the product glutamate. These structures reveal a set of active site residues that determine the substrate specificity of EgtC. Sequence comparison with homologs from organisms that do not use EgtC in ergothioneine biosynthesis find these specificity determinants more conserved than the physiological function. Based on phylogenetic distribution and active site conservation we conclude that the last common ancestor of EgtC-like enzymes may have emerged for a different purpose than ergothioneine biosynthesis. In addition, the presented structures define the sulfoxide stereochemistry of sulfoxide **3**. This information sheds new light onto the catalytic mechanism of the iron (II) dependent C-S bond forming sulfoxide synthase EgtB.

[a] Dr. G. T. Mashabela, Prof. Dr. F.P. Seebeck  
Department for Chemistry  
University of Basel  
St. Johanns-Ring 19, 4056, Basel, Switzerland  
E-mail: florian.seebeck@unibas.ch

[b] A. Vit, Prof. Dr. W. Blankenfeldt  
<sup>3</sup>Structure and Function of Proteins  
Helmholtz Centre for Infection Research  
Inhoffenstr. 7, 38124 Braunschweig, Germany

[c] previous address:  
Department of Biochemistry  
University of Bayreuth  
Universitätsstr. 30 95447 Bayreuth

[d] These authors contributed equally to this work.



**Figure 1.** Mycobacterial ergothioneine biosynthesis is catalyzed by enzymes EgtA-E.<sup>[8a]</sup> The glutamine amidohydrolase EgtC cleaves the  $\gamma$ -glutamyl moiety of the central intermediate sulfoxide **3**. In gray: Cysteine serves as sulfur donor in the abbreviated ergothioneine biosynthetic pathway described for *N. crassa*.<sup>[10a, 10b]</sup>

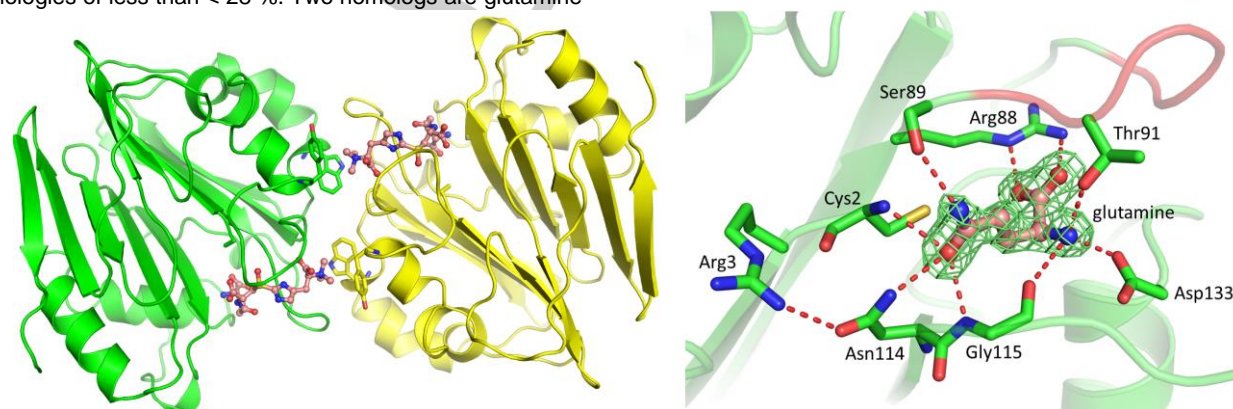
## Results and Discussion

**Overall structure.** We crystallized EgtC<sub>wt</sub> in the apo form and in complex with glutamine/glutamate. The first structure was determined using selenomethionine labeled protein. To obtain a stable complex with the substrate sulfoxide **3** (termed 4NK in the PDB file) we constructed an EgtC variant in which the N-terminal cysteine nucleophile was mutated to alanine (EgtC<sub>C2A</sub>). The asymmetric unit of these crystals contained either twelve protein chains (EgtC<sub>wt</sub> apo, EgtC<sub>C2A</sub>\_4NK) in space group P2<sub>1</sub> or four protein chains in space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> (EgtC<sub>C2A</sub>\_Gln). The crystal contacts between neighboring chains are conserved in both crystal forms. In space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, the asymmetric unit is formed by an EgtC tetramer. Oligomer prediction by PISA,<sup>[17]</sup> and analytical size exclusion chromatography of purified EgtC indicated that this quaternary structure is also relevant for the protein in solution.

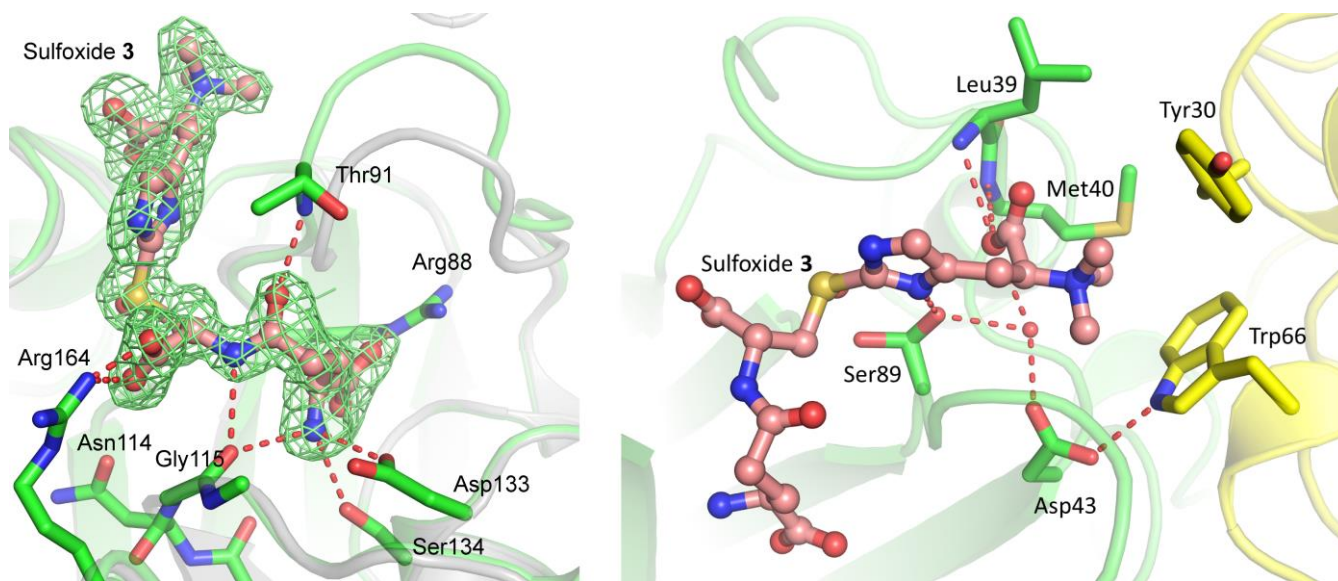
The tertiary structure of EgtC consists of a typical Ntn-hydrolase fold characterized by a four-layer  $\alpha + \beta$  structure with two antiparallel 5- and 6-stranded  $\beta$ -sheets.<sup>[12]</sup> A DALI five closest structural homologs with remarkable sequence homology (Z-score > 20, r.m.s.d. < 2.5 Å) despite sequence homologies of less than < 28 %. Two homologs are glutamine

phosphoribosyl-pyrophosphate amidotransferase (PDB: 1ECB and 1A00; EC 2.4.2.14), one is a glutamine-fructose-6-phosphate transaminase (PDB: 300J; EC 2.6.1.16) and two are hypothetical proteins (PDB: 3MDN and 1TE5). The characteristic N-terminal nucleophile in EgtC is provided by the 1,2-aminothiol function of the N-terminal cysteine.<sup>[12]</sup> To liberate their N-terminal Ser-, Thr- or Cys-nucleophile, Ntn-amidohydrolases usually undergo autocatalytic proteolysis. The observed electron density is consistent with an N-terminal cysteine in EgtC. Because the *egtC* gene encodes this residue directly after Met1, we believe that activation of recombinant EgtC is afforded by endogenous methionine aminopeptidase activity in *E. coli*.<sup>[18]</sup>

The apo form and the complexes with glutamine/glutamate or with sulfoxide **3** adopt very similar conformations (r.m.s.d. = 0.176 Å) (Figure S1). The most mobile element is a loop consisting of residues Ala90 – Pro95 (Figures 2 and 3). In two structures (apo form and the glutamine/glutamate complex) this loop closes over the glutamyl binding site and positions Thr91 for hydrogen bonding with the ligand. Yet, in the sulfoxide **3** complex this loop adopts a more open conformation that stabilizes an unproductive substrate-binding mode (see below).



**Figure 2.** **Left:** Dimer of EgtC<sub>C2A</sub> in complex with sulfoxide **3**. The substrate binding site maps to the dimer interface. **Right:** EgtC in complex with glutamine/glutamate (salmon, with  $m|F_o|-D|F_c|$  electron density;  $\sigma$ -level = 2.5). The mobile loop consisting of residues Ala90 – Pro95 is shown in red.



**Figure 3.** **Left:** EgtC<sub>C2A</sub> in complex with sulfoxide **3** (salmon, with  $m|F_o| - D|F_c|$  electron density;  $\sigma$ -level = 2.5). Comparison with EgtC in complex with glutamine/glutamate (gray) shows an open conformation of the loop between Ala90 and Pro95. **Right:** Recognition of the betaine-moiety of sulfoxide **3**. Met40 and Asp43 from chain A (green) combine with Trp66 and Tyr30 from chain B (yellow) to form a hydrophobic pocket with excess negative charge to recognize the trimethyl ammonium moiety of sulfoxide **3**. Ser89 (resolved in two conformations) makes the only specific contact to the imidazole ring of the substrate.

**EgtC co-crystallized with glutamine.** We crystallized EgtC<sub>wt</sub> in the presence of 2.5 mM glutamine as a simple substrate analog. The resulting crystals contained a ligand in the EgtC active site that may be glutamine or glutamate. The  $\alpha$ -amino acid moiety of the ligand is recognized by the side chains of Arg88 (2.7 Å and 3.1 Å), Asp133 (2.7 Å), and the backbone Gly115 (3.2 Å) (Figure 2). In this structure the Ala90 – Pro95 loop adopts the closed conformation allowing Thr91 to form an additional hydrogen bond to the  $\alpha$ -amino group of the ligand (2.9 Å). The side chain of the ligand reaches into the catalytic site represented by the catalytic nucleophile Cys2 and an oxy-anion hole formed by the side chain of Asn114 and the backbone at residue Gly115 which both coordinate the carbonyl oxygen of the ligand (3.0 Å and 2.8 Å).<sup>[19]</sup> To form this oxy anion hole Gly115 adopts a conformation which would be difficult to attain by an L-amino acid ( $\varphi = 148^\circ/\psi = -147^\circ$ ). Consistently, the Asn114-Gly115 motif is highly conserved among all members of the GATase6/GATase4 enzyme family.<sup>[20]</sup> The side chain of Asn114 is further immobilized by a hydrogen bond to Arg3 (2.9 Å). This interaction is common among EgtC-like proteins but it is absent in sequences outside this group of enzymes.

The side chain of the ligand also interacts with the N-terminal amino group (3.1 Å) and the side chain of Ser89 (2.9 Å). The observation that both heteroatoms on the ligand side chain interact with hydrogen bond donors in addition to the cationic N-terminus indicates that the ligand may not be glutamine, but rather its hydrolysis product glutamate.

**EgtC in complex with sulfoxide 3.** In the co-crystal of EgtC<sub>C2A</sub> with sulfoxide **3** we found a different binding mode for the glutamyl moiety together with an “open” conformation of

the Ala90 – Pro95 loop. The  $\alpha$ -amino acid function of the glutamyl moiety is still anchored by the same two salt bridges to Arg88 (2.7 Å and 3.5 Å) and Asp133 (3.0 Å) (Figure 3), but the C $\alpha$ -C $\beta$  bond points away from the active site, placing the scissile amid bond almost 8 Å away from the oxy anion-hole and the catalytic nucleophile. Instead, this amide bond is immobilized by two hydrogen bonds from the backbone at Thr91 (3.3 Å) and Gly115 (3.2 Å) (Figure 3). In this position, the substrate is clearly safe from enzyme catalyzed hydrolysis. However, cursory computer modeling indicates that sulfoxide **3** can easily adopt a hydrolysis-competent binding mode providing that the Ala90 - Pro95 loop assumes a closed conformation. The trimethylhistidine moiety of sulfoxide **3** binds to a pocket at the interface between two EgtC chains (Figures 2 and 3). Residues Met40 and Asp43 from chain A and Tyr30 and Trp66 from chain B combine to a large hydrophobic pocket with excess negative charge providing a typical binding site for the quaternary ammonium cation (Figure 3).<sup>[21]</sup> The carboxylate of the TMH moiety is coordinated by the backbone amides of residues Leu39 (3.3 Å) and Met40 (3.2 Å). Finally, Ser89 (2.8 Å) makes the only specific interaction to the imidazole ring on sulfoxide **3**.

**Conservation of substrate-binding residues.** The three EgtC structures in complex with either substrate or product identify six residues contacting the  $\gamma$ -glutamyl moiety (Arg88, Thr91, Asn114, Gly115, Asp133 and Ser134), five residues recognizing the TMH fragment (Tyr30, Met40, Asp43, Trp66 and Ser89) and Arg164 as a ligand to the cysteinyl sulfoxide fragment. Most actinobacterial EgtC homologs provide the same pattern of ligand interactions, which is consistent with their involvement in ergothioneine biosynthesis.<sup>[8a, 11]</sup> More

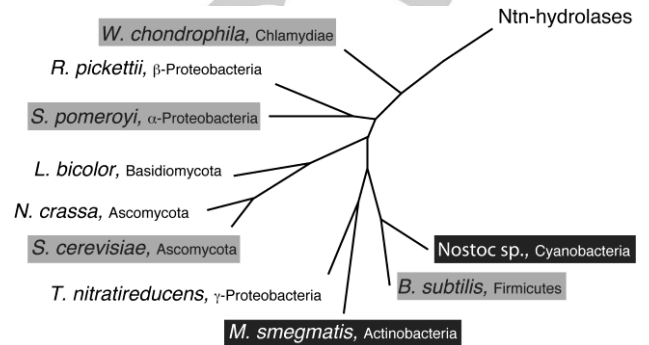


surprisingly, we found most of these active site residues also in EgtC homologs that must have different functions. To detail this observation we compared representative EgtC sequences from four classes of organisms: a) bacteria that encode EgtC in a *egtB/C/D* ergothioneine biosynthesis cluster (*M. smegmatis* and *Nostoc. sp.*); b) bacteria that encode EgtC in a different locus than EgtB and EgtD (*Ralstonia pickettii* and *Thioalkalivibrio nitratireducens*); c) bacteria that lack *egtB/D* homologs (*Bacillus subtilis*, *Waddlia chondrophila* and *Silicibacter pomeroyi*); and finally d) fungi which do (*Laccaria bicolor*, *N. crassa*) or do not (*Saccharomyces cerevisiae*) contain *egtB/D* genes.

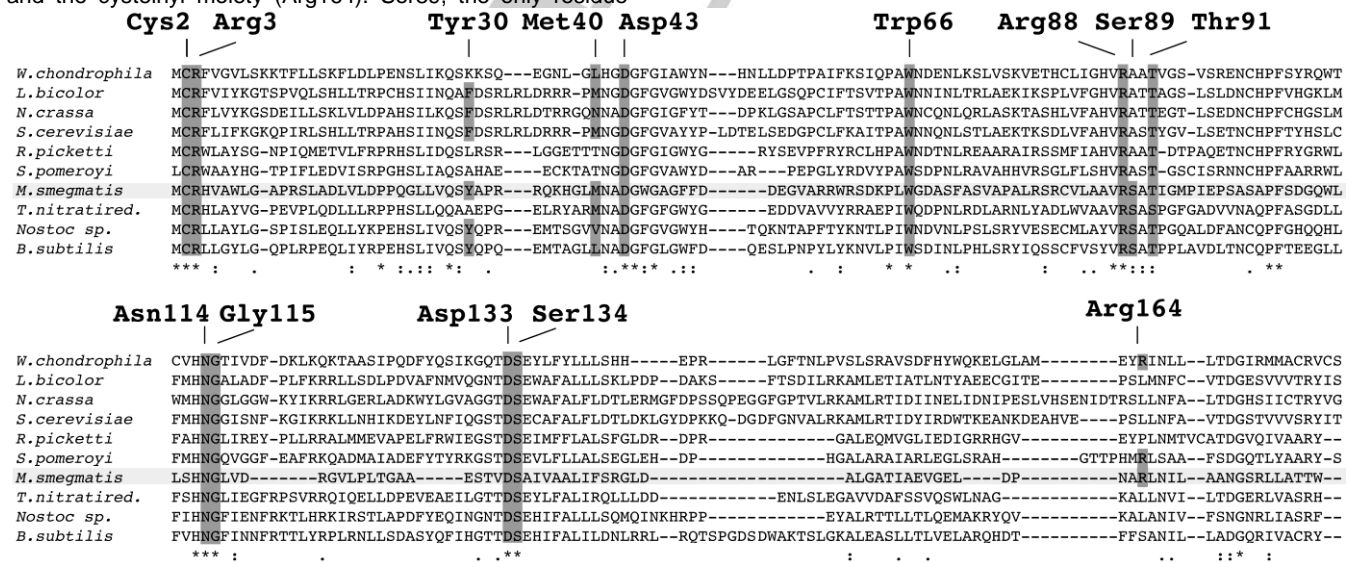
A phylogenetic tree<sup>[22]</sup> based on these homologs reveals the family of EgtC-like enzymes as a homogeneous group which does not distinguish between EgtC homologs from organisms which do produce ergothioneine (blue and green, Figure 4) and homologs from organisms which do not (gray, Figure 4). At the same time the family of EgtC-like enzymes is well separated from its closest neighbors, which is a group of uncharacterized glutamine amidohydrolase termed YafJ-like enzymes.<sup>[23]</sup> One apparent distinction between these two enzyme classes is the identity of residue 3, following the N-terminal nucleophile Cys2. Whereas EgtC-like enzymes contain arginine as third residue (Arg3), YafJ-like proteins contain glutamate or glutamine at this position.

In an alignment<sup>[24]</sup> of the same set of EgtC sequences we find the N-terminal nucleophile motif (Cys2-Arg3) and all glutamyl-binding residues conserved (Arg88, Thr91, Asn114, Gly115, Asp133 and Ser134) (Figure 5). Less conserved are active site residues which recognize TMH (Tyr30 and Ser89) and the cysteinyl moiety (Arg164). Ser89, the only residue

which contacts the imidazole ring of the substrate, occurs exclusively in close relatives of mycobacterial EgtC (*T. nitratireducens*, *Nostoc. sp* and *B. subtilis*, Figure 4 and Figure 5). More distant sequences contain a conserved alanine at this position.



**Figure 4.** Phylogenetic tree based on EgtC-like protein sequences (EgtC gene names or locus tags in brackets). **gray:** Species which do not contain *egtB* or *egtD* genes (*Bacillus subtilis*: BEST7613\_5058; *Saccharomyces cerevisiae*: NP\_014208.1; *Waddlia chondrophila*: YP\_003709474.1; *Silicibacter pomeroyi*: Spo2029, PDB: 3MDN); **white:** species which contain *egtB/C/D* genes which are not organized in a cluster (*Ralstonia pickettii* 12J: Rpic\_4440; *Thioalkalivibrio nitratireducens*: YP\_007216429.1; *Neurospora crassa*: CAD70842.1; *Laccaria bicolor*: XP\_001873785.1); **black:** species which contain *egtB/C/D* genes assembled in one cluster (*M. smegmatis*: MSMEG\_6248; *Nostoc sp.* PCC 7524: Nos7524\_2665).<sup>[22]</sup>

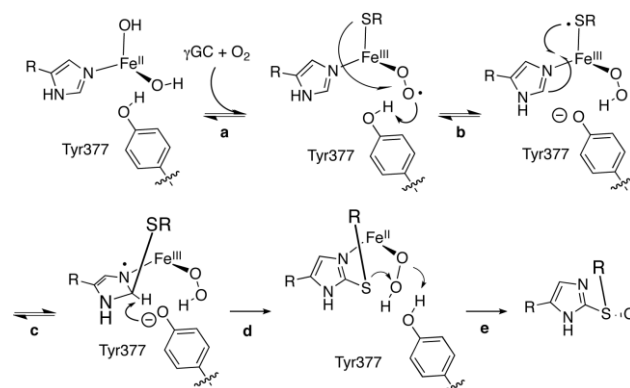


**Figure 5.** Alignment of EgtC sequence from *M. smegmatis* (light gray, residues 1 – 180) with homologs from bacterial and eukaryotic organisms (see Figure 4).<sup>[22]</sup>  
<sup>24)</sup> Conserved residues which make specific contacts to the sulfoxide substrate are highlighted (gray, Ser/Thr and Tyr/Phe are treated as equal).

**Origin of EgtC.** A recent phylogenetic study identified many bacterial species which encode the ergothioneine biosynthesis proteins EgtD and EgtB but lack recognizable EgtC-homologs.<sup>[11]</sup> Conversely, our structure-based surveil identified EgtC-like enzyme in organisms which do not contain EgtD/B homologs, or do not involve EgtC in ergothioneine production (Figure 1).<sup>[10a, 10b]</sup> Both perspectives agree that the phylogenetic distribution of EgtC and EgtD/B are poorly correlated. Therefore, we think that the last common ancestor of EgtC-like enzymes evolved for a different purpose than ergothioneine production. In addition, from the active site conservation across EgtC-like enzymes we can infer that this ancestral function involved a substrate with significant similarities to sulfoxide **3**. This similarity made the ancestral EgtC-like enzyme an easy recruit for the assembly of the five-step ergothioneine biosynthesis in actinobacteria. One adaption in the active site seems to have been the change of Ala to Ser at position 89. As discussed above, Ser89 may contribute to specific binding of sulfoxide **3** through a hydrogen bond to its TMH moiety (Figure 3).

**Avoiding the cysteine problem.** The reason why actinobacteria chose  $\gamma$ GC as sulfur donor rather than cysteine, is not quite clear (Figure 1). One possible motivation for this adaption may have been the "cysteine problem":<sup>[25]</sup> The 1,2-aminothiol function of cysteine efficiently binds to redox active transition metals. These complexes catalyze production of reactive oxygen species.<sup>[25]</sup> Cells which grow under high partial O<sub>2</sub> pressures may need to limit the concentration of intracellular cysteine.  $\gamma$ GC provides a much safer cysteine equivalent because its  $\alpha$ -amino function is  $\gamma$ -glutamylated. In addition, because  $\gamma$ GC biosynthesis is ATP-driven, this building block could accumulate despite a very low steady state cysteine concentration. Hence, an ancestral actinobacterium may have adopted EgtC and EgtA by horizontal gene transfer to optimize ergothioneine biosynthesis under O<sub>2</sub> rich and/or cysteine poor conditions.

**Sulfoxide stereochemistry.** Finally, we would like to highlight the stereochemistry of sulfoxide **3**. The resolution of the EgtC<sub>C2A</sub> complex is below 1.8 Å which unambiguously establishes an S-configuration at the sulfur atom of the ligand (Figure S2). We can not make much of this information in terms of EgtC catalysis, because we do not yet know whether the sulfoxide function is in any way involved in EgtC catalysis. However, the stereochemistry of sulfoxide **3** illuminates an interesting aspect of the catalytic mechanism of EgtB, the sulfoxide synthase that makes this intermediate (Figure 1). We have recently described the structure of EgtB from *Mycobacterium thermoresistibile* in complex with  $\gamma$ GC, TMH and manganese (II).<sup>[8c]</sup> Based on this structure we proposed a catalytic mechanism as depicted in Figure 6. This mechanism predicts that the sulfoxide product



**Figure 6.** Proposed reaction mechanism of EgtB. Binding of TMH,  $\gamma$ GC and O<sub>2</sub> as ligands to the iron center forms the first reaction intermediate consisting of an iron (III)-superoxo species (a). This species is protonated to generate an iron (III)-hydroperoxo species and a  $\gamma$ GC based thiyl radical (b). This thiyl radical attacks TMH to form the thioimidazole function (c and d). Finally, sulfoxidation of the thioether function completes the catalytic cycle as the enzyme is released in the reduced iron (II) state. The relative position of the TMH ligand and the O<sub>2</sub>-binding site observed in the crystal structure<sup>[8c]</sup> predict that sulfoxide **3** must have an S-configuration at the sulfur center.

should have the same S-configuration at the sulfur atom as we have now confirmed by crystallography.

## Conclusions

In this report we described the crystal structure of the ergothioneine biosynthetic amidohydrolase EgtC from *M. smegmatis*. EgtC removes the  $\gamma$ -glutamyl moiety from sulfoxide **3**. Even though this activity may be specific to actinobacterial ergothioneine biosynthesis, EgtC-like enzymes occur in a broad range of bacterial and fungal organisms, where they presumably serve other functions. Conservation of most active site residues across all EgtC-like enzymes suggests that these homologs transform substrates with significant similarity to sulfoxide **3**. This similarity may have facilitated the recruitment of EgtC to  $\gamma$ GC-dependent ergothioneine biosynthesis by actinobacteria.

## Experimental Section

**Protein production.** The open reading frame of *egtC* (locus name: MSMEG\_6248; UniProtKB entry: A0R5M9) was amplified from *Mycobacterium smegmatis* strain ATCC 700084 / mc(2)155 genomic DNA.<sup>[8a]</sup> The resulting PCR product was digested with PciI and XhoI while the vector pET28a(+) (Novagen) was cut with NcoI and XhoI prior to ligation with the insert. The resulting plasmid pET28a(+)\_EgtC\_His<sub>6</sub> was sequenced and encodes a protein with the sequence

MCRHVAWLGA PRSLADLVLDPPQGLLVQSYAPRRQKHGL  
MNADGWGAGFFDDDG VARRWRS DKPLWGDASFASVAPA  
LRSRCVVA AVRSATIGMSIEPSASAPFSDGQWLLSHNGLVA

RGVLP LTGAAESTVDSAILAALIFSRGLDALGATIAEVGELDP  
 NARLNILAANGSRL LATTWGD TSVLRRPDGVV LASEPYDD  
 DPGWSDIPDRHLVDVRDAHV VVTPLSSHHHHHHH,

where underlined letters indicate mutations with respect to the EgtC sequence deposited in the UniProtKB database and italic letters indicate the linker and the affinity tag. This construct was used to produce native and seleno-L-methionine labeled EgtC for SAD phasing of the apo structure.

The pET28a(+)\_EgtC\_His<sub>6</sub> construct was regenerated to remove several mutations using the previous cloning strategy. The protein encoded by pET28a(+)\_EgtC\_His<sub>6</sub>\_new,

MCRHVAWL G APRSLADLVLDPPQGLLVQSYAPRRQKHGL  
 MNADGWGAGFFDDDGVARRWRS DKPLWGDASFASVAPA  
 LRSRCVLA AVRSATIGMPIEPSASAPFSDGQWLLSHNGLVD  
 RGVLP LTGAAESTVDSAILAALIFSRGLDALGATIAEVGELDP  
 NARLNILAANGSRL LATTWGD TSVLRRPDGVV LASEPYDD  
 DPGWSDIPDRHLVDVRDAHV VVTPLLEHHHHHHH,

was used for structure determination of the EgtC<sub>wt</sub>\_Gln complex.

An inactive variant EgtC<sub>C2A</sub> was generated by PCR amplification using primers EgtC\_NcoI\_s (5'-ACT GTC CCA TGG CCC GGC ATG TGG CGT-3') and EgtC\_Xho\_as (5'-ACT ATC CTC GAG CAG GGG TGT CAC GAC GAC-3') from plasmid pET28a(+)\_EgtC\_His<sub>6</sub>\_new. The PCR product as well as pET28a(+) were digested with NcoI and XhoI prior to ligation of both components. The protein encoded on plasmid pET28a(+)\_EgtC<sub>C2A</sub> is

MARHVAWL G APRSLADLVLDPPQGLLVQSYAPRRQKHGL  
 MNADGWGAGFFDDDGVARRWRS DKPLWGDASFASVAPA  
 LRSRCVLA AVRSATIGMPIEPSASAPFSDGQWLLSHNGLVD  
 RGVLP LTGAAESTVDSAILAALIFSRGLDALGATIAEVGELDP  
 NARLNILAANGSRL LATTWGD TSVLRRPDGVV LASEPYDD  
 DPGWSDIPDRHLVDVRDAHV VVTPLLEHHHHHHH.

*E. coli* BL21(DE3) cells were employed in recombinant production of EgtC<sub>wt</sub> and EgtC<sub>C2A</sub> in LB (lysogeny broth) medium supplemented with 50 µg ml<sup>-1</sup> kanamycin. For production of seleno-L-methionine labeled protein, *E. coli* BL21-Codon-Plus(DE3)-RIL cells were grown in artificial LeMaster medium suppressing L-methionine biosynthesis.<sup>[26]</sup> Expression was induced with 0.5 mM IPTG at an OD<sub>600</sub> of 0.8. After vigorous shaking for 16 h at 25 °C, cells were harvested by centrifugation, resuspended in buffer A (50 mM phosphate pH 7.5, 300 mM NaCl, 5 mM β-mercaptoethanol) and lysed by sonication or in a microfluidizer. The lysate was clarified by centrifugation at 48000g for 1 h at 4 °C. All EgtC variants were purified on immobilized Ni<sup>2+</sup> using 1 ml NTA Agarose (Qiagen) or a HiTrap IMAC sepharose FF column (GE Healthcare). After elution with buffer B (buffer A containing 500 mM imidazole), the protein was loaded onto a HiLoad 16/60 Superdex™75 pregrade size exclusion

chromatography column (GE Healthcare) primed with 20 mM Tris/HCl pH 7.5, 150 mM NaCl, 5 mM β-mercaptoethanol and connected to an ÄKTApriime FPLC system (GE Healthcare). Fractions containing pure protein were pooled and concentrated to 25 mg ml<sup>-1</sup>. The identity of the proteins and the incorporation of seleno-L-methionine were verified using ESI mass spectrometry. Purified proteins were flash frozen in liquid nitrogen and stored at -80 °C if not used immediately. No differences in crystallization behavior were observed between fresh and frozen protein.

**Sulfoxide production.** Sulfoxide **3** was produced in an 50 ml reaction containing 1.5 µM EgtB from *Mycobacterium thermoresistibile*,<sup>[8c]</sup> 12 mM γGC, 10 mM TMH, 2 mM ascorbate, 2 mM TCEP, 10 µM FeSO<sub>4</sub> and 20 mM potassium phosphate buffer at pH 7.5. After incubation for > 12 h at room temperature the filtered solution was acidified to pH 6.0 by addition of ammonium acetate. This solution was incubated with 4 ml Dowex resin for two hours. The resin was washed with 8 ml nano-pure water. The flow-through and washing fractions were combined, and incubated with 3 ml Dowex resin at pH 3. After washing with 12 ml nano-pure water, sulfoxide **3** was eluted with 250 mM ammonium hydroxide.

**Crystallization.** Crystallization conditions for the EgtC variants were determined by the sitting drop vapor diffusion method using The JCSG Core Suites I-IV (and The PACT and The PEGS for EgtC<sub>wt</sub>) and then optimized using a hanging drop setup. Optimized crystals of native EgtC<sub>wt</sub> and seleno-L-methionine labeled EgtC<sub>wt</sub> were obtained using 1 µl of 15 mg ml<sup>-1</sup> EgtC mixed with 1 µl reservoir equilibrated against 500 µl reservoir (0.1 M MES pH 6.4, 10-12% (w/v) PEG 20.000) at 20 °C. Rod-shaped crystals grew after 2-4 days. They were cryoprotected in 0.1 M MES pH 6.4, 10-12% (w/v) PEG 20.000, 35% (w/v) PEG 3350 and flash cooled in liquid nitrogen. Co-crystallization of EgtC<sub>wt</sub> (protein produced from pET28a(+)\_EgtC\_His<sub>6</sub>\_new) with L-glutamine was achieved by mixing of 1 µl EgtC (20 mg ml<sup>-1</sup>) pre-incubated with 2.5 mM L-glutamine with 1 µl mother liquor equilibrated against 500 µl of reservoir (0.1 M sodium acetate pH 4.7, 50%(v/v) ethylene glycol). Since no further cryoprotection was required, the crystals were directly flash cooled in liquid nitrogen for data collection.

Initial crystal hits of EgtC<sub>C2A</sub> with sulfoxide **3** (termed 4NK in PDB entry 4ZFL) were optimized by mixing 1 µL of 16 mg ml<sup>-1</sup> EgtC<sub>C2A</sub> pre-incubated with 10 mM sulfoxide **3** and 1 µL of reservoir equilibrated against 500 µL of reservoir (0.1 M sodium citrate pH 5.2-5.4, 20% (w/v) PEG 6000) at 20 °C. Crystals were cryoprotected in reservoir solution supplemented with 20% (v/v) glycerol and 10 mM sulfoxide **3** before plunging them in liquid nitrogen.

**Data collection, data processing, structure solution and refinement.**



Diffraction data sets of a native as well as a seleno-L-methionine labeled EgtC<sub>wt</sub> crystal were collected on beamline PXII of the SLS (Swiss Light Source, Paul Scherrer Institute, Villingen, Switzerland) at cryogenic temperatures. Native EgtC<sub>wt</sub> data were collected at a wavelength of  $\lambda = 0.9792 \text{ \AA}$ , while single anomalous diffraction (SAD) data were collected at the Se K edge ( $\lambda = 0.9786 \text{ \AA}$ ). According to Xtriage, anomalous differences were significant to 4.6 Å resolution.<sup>[27]</sup> Diffraction data of an EgtC<sub>wt</sub>-Gln complex were collected on beamline BL14.1 of BESSY II Synchrotron (Helmholtz Centre Berlin, Berlin, Germany) at a wavelength of  $\lambda = 0.9184 \text{ \AA}$ , while diffraction data of an EgtC<sub>CC2A</sub>-4NK crystals were collected on beamline P11 at PETRA III (DESY, EMBL Hamburg, Germany). Data indexing employed XDS and data scaling was performed with AIMLESS from the CCP4 suite.<sup>[28]</sup>

Initial phasing and model building were achieved with Autosol from the Phenix software package, combining single anomalous diffraction data of a Se-Met labeled crystal with a native data set of EgtC<sub>wt</sub> in the same crystal form. The second crystal form was phased by molecular replacement, using one EgtC monomer in MOLREP.<sup>[29]</sup> Structure determination of EgtC<sub>CC2A</sub>-4NK was achieved using a monomer of the EgtC<sub>wt</sub> apo structure in PHASER from Phenix.<sup>[30]</sup>

Ligand restraints were prepared with PRODRG<sup>[31]</sup> and eLBOW.<sup>[32]</sup> Model building was done in COOT, followed by refinement in phenix.refine.<sup>[30c, 33]</sup> The structures were validated with MolProbity and deposited as PDB entries 4ZFJ (EgtC<sub>wt</sub>), 4ZFK (EgtC<sub>wt</sub>-Gln), and 4ZFL (EgtC<sub>CC2A</sub>-4NK).<sup>[34]</sup> Details of data collection and refinement statistics are provided in Table S1 and S2. Figures were prepared with PyMOL.<sup>[35]</sup>

## Acknowledgements

The authors are indebted to R. S. Goody who enabled the initiation of this project at the Max Planck Institute for Molecular Physiology in Dortmund. The authors would like to thank the X-ray communities at the Max Planck Institute in Heidelberg and Dortmund, at the University of Bayreuth and at the Helmholtz Centre for Infection Research for help with data collection. The Swiss Light Source (Paul Scherrer Institute, Villingen, Switzerland), BESSY II Synchrotron (Helmholtz Centre Berlin, Berlin, Germany) and PETRA III (DESY, EMBL Hamburg, Germany) are acknowledged for providing beamline access. This project was supported by the Swiss National Foundation. F.P.S. is supported by the "Professur für Molekulare Bionik". A.V. is supported by the HZI Graduate School for Infection Research. G.T.M. is a recipient of a Swiss government fellowship for excellence.

**Keywords:** Ergothioneine • mycobacteria • Ntn-hydrolase • biosynthesis • sulfoxide stereochemistry

- [1] P. E. Hartman, *Methods Enzymol.* **1990**, *186*, 310-318.
- [2] aL. Saing, Y. C. Wei, C. J. Tseng, *Clin. Exp. Pharmacol. Physiol.* **2015**, *Epub ahead of print*; bB. D. Paul, S. H. Snyder, *Cell. Death Differ.* **2009**, *47*, S149-S149; cR. W. S. Li, C. Yang, A. Sit, Y. W. Kwan, S. M. Y. Lee, M. P. M. Hoi, S. W. Chan, M. Hausman, P. M. Vanhoutte, G. P. H. Leung, *J. Pharmacol. Exp. Ther.* **2014**, *350*, 691 - 700; dD. Gründemann, *Prev. Med.* **2012**.
- [3] C. Pfeiffer, M. Bach, T. Bauer, J. C. da Ponte, E. Schörming, D. Gründemann, *Free Radic. Biol. Med.* **2015**, *Epub ahead of print*.
- [4] al. K. Cheah, R. L. Ong, J. Gruber, T. S. Yew, L. F. Ng, C. B. Chen, B. Halliwell, *Free Radic. Res.* **2013**, *47*, 1036 - 1045; bM. H. Bello, V. Barrera-Perez, D. Morin, L. Epstein, *Fungal Genet. Biol.* **2012**, *49*, 160-172; cM. H. Bello, J. C. Mogannam, D. Morin, L. Epstein, *Fungal Genetics and Biology* **2014**, *73*, 120 - 127.
- [5] aP. Ta, N. Buchmeier, G. L. Newton, M. Rawat, R. C. Fahey, *J. Bacteriol.* **2011**, *193*, 1981-1990; bS. Nakajima, Y. Satoh, K. Yanashima, T. Matsui, T. Dairi, *J. Biosci. Bioeng.* **2015**, *Epub ahead of print*; cS. C. Emani, M. J. Williams, I. J. Wiid, N. F. Hiten, A. J. Viljoen, R. D. Pietersen, P. D. van Helden, B. Baker, *Antimicrob. Agents Chemother.* **2013**, *57*, 3202 - 3207.
- [6] I. K. Cheah, B. Halliwell, *Biochim. Biophys. Acta.* **2012**, *1822*, 784-793.
- [7] Q. Zhao, M. Wang, D. Xu, Q. Zhang, W. Liu, *Nature* **2015**, *Epub ahead of print*.
- [8] aF. P. Seebeck, *J. Am. Chem. Soc.* **2010**, *132*, 6632-6633; bA. Vit, L. E. Misson, W. Blankenfeldt, F. P. Seebeck, *Chembiochem* **2015**, *16*, 119125; cK. V. Goncharenko, A. Vit, W. Blankenfeldt, F. P. Seebeck, *Angew Chem Int Ed Engl.* **2015**, *Epub ahead of print*.
- [9] G. Harth, S. Maslesa-Galic, M. V. Tullius, M. A. Horwitz, *Mol. Microbiol.* **2005**, *58*, 1157-1172.
- [10] aW. Hu, H. Song, A. Sae Her, D. W. Bak, N. Naowarajna, S. J. Elliot, L. Qin, X. Chen, P. Liu, *Org. Lett.* **2014**, *16*, 5382 - 5385; bD. B. Melville, S. Eich, M. L. Ludwig, *J. Biol. Chem.* **1957**, *224*, 871-877; cT. Pluskal, M. Ueno, M. Yanagida, *PLoS One* **2014**, *9*, e97774.
- [11] G. W. Jones, S. Doyle, D. A. Fitzpatrick, *Gene* **2014**, *549*, 161-170.
- [12] J. A. Brannigan, G. Dodson, H. J. Duggleby, P. C. Moody, J. L. Smith, D. R. Tomchick, A. G. Murzin, *Nature* **1995**, *378*, 416 - 419.
- [13] J. Lowe, D. Stock, B. Jap, P. Zwicker, W. Baumeister, R. Huber, *Science* **1995**, *268*, 533 - 539.
- [14] H. J. Duggleby, S. P. Tolley, C. P. Hill, E. J. Dodson, G. Dodson, P. C. Moody, *Nature* **1995**, *373*, 264 - 268.
- [15] I. Castellano, A. Merlino, *Cell. Mol. Life Sci.* **2012**, *69*, 3381 - 3394.
- [16] aK. Michalska, K. Brzezinski, M. K. Jaskolski, *J. Biol. Chem.* **2005**, *280*, 28484 - 28491; bJ. L. Smith, E. J. Zaluzec, J. P. Wery, L. Niu, R. L. Switzer, H. Zalkin, Y. Satow, *Science* **1994**, *264*, 1427 - 1433; cR. H. van den Heuvel, D. I. Svergun, M. V. Petoukhov, A. Coda, B. Curti, S. Ravasio, M. A. Vanoni, A. Mattevi, *J. Mol. Biol.* **2003**, *330*, 113 - 128; dM. B. West, Y. Chen, S. Wickham, A. Heroux, K. Cahill, M. H. Hanigan, B. H. Mooers, *J. Biol. Chem.* **2013**, *289*, 31902 - 31913; eG. Orsmando, R. D. de la Garza, B. J. Green, M. Peng, P. A. Rea, T. J. Ryan, J. F. r. Gregory, A. D. Hanson, *J. Biol. Chem.* **2005**, *280*, 28877 - 28884; fC. Oinonen, J. Rouvinnen, *Protein Sci.* **2000**, *9*, 2329 - 2337.
- [17] E. B. Krissinel, K. Henrick, *J. Mol. Biol.* **2007**, *372*, 774 - 797.
- [18] W. T. Lowther, B. W. Matthews, *Biochim. Biophys. Acta* **2000**, *1477*, 157 - 167.
- [19] aH. G. Schnitzer, S. K. Boehlein, J. D. Stewart, N. G. J. Richards, S. M. Schuster, *Biochemistry* **1999**, *38*, 3677-3682;

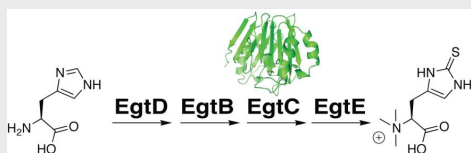
- bR. Wu, S. Richer, R. G. Zhang, V. J. Anderson, D. M. Missiakas, A. Joachimiak, *J. Biol. Chem.* **2009**, *284*, 24406 - 24414.
- [20] R. D. Finn, A. MBateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heder, K. Hetherington, J. Mistry, E. L. Sonnhammer, J. Tate, M. Punta, *Nucleic Acids Res.* **2014**, *42*, D222 - D230.
- [21] aD. A. Dougherty, D. A. Stauffer, *Science* **1990**, *250*, 1558 - 1560; bA. S. Mahadevi, G. N. Sastry, *Chem. Rev.* **2013**, *113*, 2100 - 2138.
- [22] A. Dereeper, V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie, O. Gascuel, *Nucleic Acids Res.* **2008**, *36*, W465-469.
- [23] aR. Barten, T. F. Meyer, *Mol. Gen. Genet.* **1998**, *258*, 34-44; bD. Blaha, H. Sanguin, P. Robe, R. Nalin, R. Bally, Y. Moenne-Loccoz, *FEMS Microbi. Lett.* **2005**, *244*, 157-163.
- [24] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopet, J. D. Thompson, T. J. Gibson, D. G. Higgins, *Bioinformatics* **2007**, *23*, 2947-2948.
- [25] R. C. Fahey, *Biochim. Biophys. Acta.* **2013**, *1830*, 3182 - 3198.
- [26] D. M. LeMaster, F. M. Richards, *Biochemistry* **1985**, *24*, 7263 - 7268.
- [27] P. H. Zwart, R. W. Grosse-Kunstleve, P. D. Adams, *CCP4 Newsl.* **2005**, *Winter*.
- [28] aW. Kabsch, *Acta Crystallog. D Biol. Crystallogr.* **2010**, *66*, 125 - 132; bP. R. Evans, *Acta Crystallog. D Biol. Crystallogr.* **2011**, *67*, 282 - 292; cM. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, Evans, P. R., R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, et al., *Acta Crystallog. D Biol. Crystallogr.* **2011**, *67*.
- [29] A. Vagin, A. Teplyakov, *J. Appl. Cryst.* **1997**, *30*, 1022 - 1025.
- [30] aG. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallog. D Biol. Crystallogr.* **2011**, *67*, 355 - 367; bP. D. Adams, P. V. Afonie, G. Bunkoczi, V. B. Chen, I. W. Davies, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwillinger, P. H. Zwart, *Acta Crystallog. D Biol. Crystallogr.* **2010**, *66*, 213 - 221; cP. V. Afonie, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, *Acta Crystallog. D Biol. Crystallogr.* **2012**, *68*, 352-367.
- [31] A. W. Schüttelkopf, D. M. F. van Aalten, *Acta Crystallog. D Biol. Crystallogr.* **2004**, *60*, 1355 - 1363.
- [32] N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, *Acta Crystallog. D Biol. Crystallogr.* **2009**, *65*, 1074 - 1080.
- [33] P. Emsley, K. D. Cowtan, *Acta Crystallog. D Biol. Crystallogr.* **2004**, *60*, 2126 - 2132.
- [34] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallog. D Biol. Crystallogr.* **2009**, *66*, 12-21.



Entry for the Table of Contents (Please choose one layout)

Layout 2:

FULL PAPER



Allegra Vit, Gabriel T. Mashabela, Wulf Blankenfeldt, and Florian P. Seebeck\*

**Structure of the ergothioneine-biosynthetic amidohydrolase EgtC**

Text for Table of Contents