

# **Novel lipoate-binding proteins and their role in sulfur oxidation**

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

**Tomohisa Sebastian Tanabe**

aus

Neuss

Bonn, 2024



Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät  
der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachterin: Priv.-Doz. Dr. Christiane Dahl
2. Gutachter: Prof. Dr. Uwe Deppenmeier
3. Gutachter: Prof. Dr. Matthias Boll

Tag der Promotion: 14.03.2024

Erscheinungsjahr: 2024













# Contents

<b>I Summary</b>	<b>1</b>
<b>II Introduction</b>	<b>3</b>
1 A new pathway of sulfur oxidation involving a heterodisulfide reductase-like complex	3
1.1 The proposed sHdr system and its lipoate-binding protein . . . . .	3
1.2 Heterodisulfide reductases and related complexes . . . . .	6
2 The annotation of sulfur metabolism related proteins . . . . .	7
2.1 A historical view on the development of annotation . . . . .	8
2.2 Proteins in sulfur metabolism . . . . .	9
2.3 Sulfur metabolism prediction with HMS-S-S and HMSS2 . . . . .	10
3 The lipoate-binding protein of the sHdr system . . . . .	12
3.1 Lipoate assembly pathways . . . . .	12
3.2 The lipoate-binding protein LbpA . . . . .	14
4 Sulfur trafficking in the sHdr-system . . . . .	15
4.1 Sulfur transferases and sulfur transfer systems . . . . .	16
4.2 Sulfur import and the regulating of dissimilatory sulfur oxidation . . . . .	18
<b>III Aims of the thesis</b>	<b>21</b>
<b>IV Publicatins and manuscripts included in this thesis</b>	<b>23</b>
Chapter 1 HMS-S-S: A tool for the identification of Sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies . . . . .	25
Chapter 2 HMSS2: An advanced tool for the analysis of sulphur metabolism, including organosulphur compound transformation, in genome and metagenome assemblies .	61
Chapter 3 HAMSTER: Automated Hidden Markov model generation from collinear syntenic blocks . . . . .	93
Chapter 4 Identification of a novel lipoic acid biosynthesis pathway reveals the complex evolution of lipoate assembly in prokaryotes . . . . .	101
Chapter 5 A cascade of sulfur transferases delivers sulfur to the sulfur-oxidizing heterodisulfide reductase-like complex . . . . .	141
Chapter 6 Fe/S proteins in microbial sulfur oxidation . . . . .	197
Chapter 7 A metabolic puzzle: Consumption of C1 compounds and thiosulfate in <i>Hyphomicrobium denitrificans</i> X <sup>T</sup> . . . . .	229
Chapter 8 In the Alphaproteobacterium <i>Hyphomicrobium denitrificans</i> SoxR Serves a Sulfane Sulfur-Responsive Repressor of Sulfur Oxidation . . . . .	261
Chapter 9 The type I sHdr system is co-occurring with thiosulfate oxidation via a truncated Sox system . . . . .	291
<b>V Discussion</b>	<b>297</b>
1 A novel pathway for lipoate assembly in bacteria . . . . .	297
2 The sulfur oxidation in the type I sHdr pathway . . . . .	300

3	The type II sHdr system . . . . .	302
4	The evolution of the lipoate biosynthesis and sHdr pathway . . . . .	302
5	Perspectives and outlook . . . . .	304
<b>VI</b>	<b>References</b>	<b>305</b>
<b>VII</b>	<b>Acknowledgments</b>	<b>319</b>

## I Summary

The global sulfur cycle is a highly complex network of interconnected processes that involve the conversion of organic and inorganic sulfur compounds across multiple redox states. Driven by prokaryotes through the oxidation, reduction, or even disproportionation of organic and inorganic sulfur compounds it plays a crucial role in maintaining a balanced biosphere. Exploring these processes and the involved prokaryotes is highly relevant for the understanding of current and ancient ecosystems. In recent years, the development of next-generation sequencing technologies and genome reconstruction from metagenomes has led to an increasing number of accessible prokaryotic genomes from a previously unknown diversity of species. Assigning metabolic pathways to this genetic information based on experimentally validated functions has become one of the most important tasks to make use of the increasing amount of data. This work combines bioinformatic and biochemical techniques to explore the dissimilatory sulfur oxidation.

For the prediction genes and gene clusters with a function in dissimilatory sulfur oxidation, or reduction in (meta-)genomes a software named HMS-S-S was developed. With the advanced version HMSS2 proteins involved in the conversion, degradation and transport of organic sulfur compounds were added. Both tools were shown to predict genes and gene clusters with a reasonable confidence using profiled hidden Markov models. The generation of profiled hidden Markov models was also automated as an independent software to keep up with the discovery of new metabolic pathways.

These tools were used to elucidate the mechanisms of a dissimilatory sulfur oxidation through a sulfur-oxidizing heterodisulfide reductase (sHdr) complex that interacts with a lipoate-binding protein (LbpA). Before it is functionally active, the lipoate cofactor must be covalently attached to its cognate enzyme. Here, a novel LbpA-specific lipoate biosynthesis pathway was demonstrated to be active in bacteria using a genetic approach. This pathway operated autonomously and concurrently with the canonical lipoate biosynthesis pathways. Examining the occurrence of the novel and canonical lipoate biosynthesis across prokaryotic biodiversity revealed a much broader distribution of the lipoate assembly systems than previously recognized and located the evolutionary origin of the novel biosynthesis pathway inside the archaeal domain. Genetically the sHdr system is tightly connected to several sulfur transferases, which are proposed to relay protein-bound sulfur to the LbpA-sHdr complex. Indeed it was possible to detect sulfane sulfur transfer activity and to determine the catalytically active cysteines. With sulfur transfer assays it was possible to reconstruct the sulfur relay system of four bacteria. Assessing the co-occurrence of these sulfur transferases in the *shdr* gene cluster revealed a general importance of these sulfur relay systems. Two regulatory sulfur-sensing proteins were also characterized, which likely receive sulfur from these sulfur transferases. Additionally, the co-occurrence of the sHdr system and other sulfur-oxidizing systems was examined. In total this combination of genetic, biochemical and bioinformatic methods contributed to the understanding of the complexity of microbial sulfur-oxidation.



## II Introduction

### 1 A new pathway of sulfur oxidation involving a heterodisulfide reductase-like complex

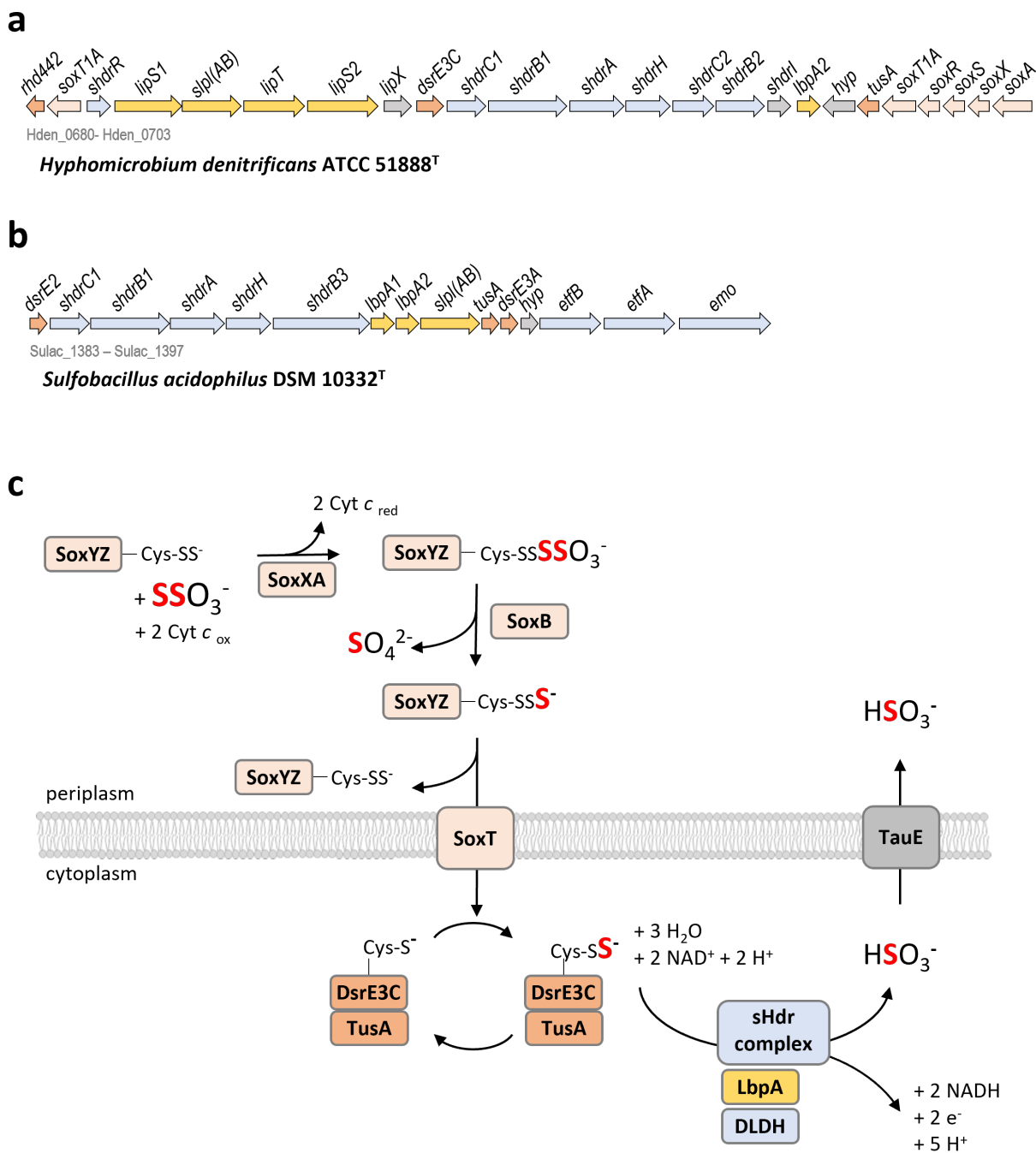
The turnover of sulfur compounds is one of the earliest microbial activities in Earth's history (Canfield & Raiswell 1999, Canfield *et al.* 2006). Global biogeochemical cycling of sulfur compounds directly affects Earth's oxidation state, surface conditions and climate regulation (Fike *et al.* 2015). Prokaryotes contribute to the sulfur cycle through the oxidation, reduction and even disproportionation. The sulfur cycle is not simply the switching between sulfide (the most reduced state, -II) and sulfate (the most oxidized state, +VI). Instead, it involves sulfur compounds of all intermediate oxidation states. There are enzymatic machines for the conversion between each of these oxidation states, which can function alone or in modular combinations, resulting in complex and sometimes even redundant systems (Kümpel *et al.* 2024). Sulfate reduction is mainly driven by bacteria and archaea that operate the Dsr system, which is named after the key enzyme dissimilatory sulfite reductase (DsrAB). Conversely, many photo- and chemotrophic sulfur oxidizers operate a reverse Dsr system, oxidizing reduced sulfur compounds. However, several of the established sulfur-oxidizing archaea and bacteria do not encode for the DsrAB or any other essential components of the system. How these prokaryotes utilize sulfur compounds as an electron source has been enigmatic until several transcriptomic and proteomic studies attributed this activity to a heterodisulfide reductase-like complex (sHdr) (Quatrini *et al.* 2009, Ehrenfeld *et al.* 2013, Christel *et al.* 2016, Koch & Dahl 2018). The relevance of the *shdr* genes for the sulfur oxidation capacity was finally proven in *Hyphomicrobium denitrificans* by reverse genetics (Koch & Dahl 2018). This approach also revealed a new function for the lipoate cofactor by identifying a lipoate-binding protein (LbpA) as an essential component of the system (Cao *et al.* 2018).

#### 1.1 The proposed sHdr system and its lipoate-binding protein

The sHdr system combines a heterodisulfide reductase-like complex with a lipoate-binding protein for the dissimilatory oxidation of reduced sulfur compounds. This system is present in a vast number of well-established sulfur-oxidizing bacteria and archaea that lack DsrAB (Venceslau *et al.* 2014). Two different types of this system have been recognized so far. The type I system is encoded by a *shdrC1B1AHC2B2* gene cluster and has been described in sulfur-oxidizing bacteria such as *Hyphomicrobium denitrificans* (Fig. 1a) (Koch & Dahl 2018, Cao *et al.* 2018), *Aquifex aeolicus* (Boughanemi *et al.* 2016), *Acidithiobacillus caldus* (Koch & Dahl 2018), *Thioalkalivibrio* sp. (Koch & Dahl 2018, Berben *et al.* 2019), or the archaeon *Metallosphaera cuprina* (Liu *et al.* 2014). The relevance of this cluster for the sulfur oxidizing capacity has been demonstrated in numerous studies (Quatrini *et al.* 2009, Ehrenfeld *et al.* 2013, Christel *et al.* 2016, Koch & Dahl 2018). It was also possible to show the existence of a sHdr complex formed by at least sHdrC1B1AC2B2 (Boughanemi *et al.* 2016). The type II sHdr system has been described to be encoded in a *shdrC1B1AH* gene cluster together with two genes for putative iron-sulfur cluster containing proteins and *etfAB* encoding for two electron-transferring flavoproteins (Fig. 1b) (Justice *et al.* 2014, Cao *et al.* 2018).

Both types are usually accompanied by genes for diverse sulfur transferases and up to three genes for lipoate-binding protein (LbpA) (Justice *et al.* 2014, Cao *et al.* 2018). Lipoate is an organosulfur cofactor with highly important function in the central carbon metabolism. Until the discovery of LbpA-sHdr system the three  $\alpha$ -ketoacid dehydrogenases, acetoin dehydrogenase, and the glycine cleavage complex were the only known lipoate-depending multienzyme complexes (Cronan 2016). LbpA was revealed to be essential for the functionality of the sHdr pathway in *H. denitrificans* (Cao *et al.* 2018). In the proposed model for the sHdr pathway, LbpA is a key component for the energy conversion step, as electrons could be transferred directly from the cofactor to  $\text{NAD}^+$  due to the low standard biological redox potential of lipoamide/dihydrolipoamide. This facultative aerobic bacterium *Hyphomicrobium denitrificans* utilizes thiosulfate as an supplementary electron source via two different processes. Thiosulfate can be oxidized to tetrathionate catalyzed by the diheme cytochrome c thiosulfate dehydrogenase TsdA in the periplasm (Brito *et al.* 2015). Tetrathionate is not further utilized by the bacterium and is only formed when high concentrations of thiosulfate are available in the environment (Koch & Dahl 2018). The second mechanism is the oxidation of thiosulfate to sulfate via a truncated Sox system in the periplasm and the sHdr pathway in the cytoplasm. A model for thiosulfate oxidation through the sHdr pathway in *Hyphomicrobium denitrificans* has been postulated based on the proteins encoded by the type I *shdr* gene cluster (Koch & Dahl 2018). In the following the model as proposed at that time is described (Fig. 1c) (Koch & Dahl 2018, Cao *et al.* 2018).

Initially, thiosulfate is decomposed to sulfite and protein-bound sulfane sulfur via a truncated Sox system. SoxXA catalyzes the covalent attachment of the thiosulfate sulfane sulfur to a cysteine persulfide adduct of SoxY from SoxYZ (Sauvé *et al.* 2007, Grabarczyk & Berks 2017). Sulfohydrolyase SoxB then hydrolytically cleaves the sulfone group as sulfate from the SoxY-S-thiosulfonate (Sauvé *et al.* 2009). The SoxY-bound sulfane sulfur cannot be further oxidized to the sulfonate by the truncated Sox system because it lacks the required sulfur dehydrogenase SoxCD. To complete the oxidation, the sulfane sulfur is imported into the cytoplasm by an unknown mechanism, where it is oxidized by the sHdr pathway. A protein of the YeeE/YedE transporter protein family, encoded downstream of *soxRS*, was proposed to facilitate this import due to the presence of three conserved cysteine residues (Gristwood *et al.* 2011). It was later discovered that this protein has significant similarity to a transporter associated with the Sox system. Accordingly, it was then renamed to SoxT (Li *et al.* 2023b). After the import the sulfur is proposed to be transferred in a protein-bound form to the sHdr complex. The sulfur transferase TusA was considered as a good candidate for channeling the sulfur to LbpA possibly in combination with a DsrE-type sulfur transferase (Cao *et al.* 2018). The LbpA then binds the sulfur via a disulfide bond at one of the sulfur atoms of the lipoate cofactor. The resulting dihydrolipoamide bound sulfane sulfur could then be presented to the catalytic sites of the sHdr complex. Finally, the sulfane sulfur is oxidized to sulfite at the sHdr complex by an unknown mechanism releasing four electrons. Two electrons stemming from the sulfane sulfur oxidation at the sHdr complex are directed to the lipoate cofactor for the reduction to dihydrolipoamide (Cao *et al.* 2018). A dihydrolipoamide dehydrogenase may then re-oxidize dihydrolipoamide coupled to the reduction of  $\text{NAD}^+$  to  $\text{NADH}$  (Koch & Dahl 2018, Cao *et al.* 2018). The other two electrons are eventually transferred to an unknown acceptor.



**Figure 1: The *shdr* gene clusters and the sHDr system.** (a) The type I *shdr* gene cluster of *H. denitrificans*. (b) The type II *shdr* gene cluster of *Sulfolobus acidophilus*. (c) Model for the thiosulfate oxidation by *H. denitrificans* via a truncated Sox system and the sHDr system as proposed by Koch & Dahl (2018). The sHDr complex is formed by the components sHDrC1, sHDrB1, sHDrA, sHDrH, sHDrC2 and sHDrB2. LbpA, lipoate-binding proteins; DLDH, dihydrolipoamide dehydrogenase.

The generated sulfite is finally exported into the periplasm by the sulfite exporter TauE, where it is chemically oxidized with molecular oxygen to sulfate (Weinitschke *et al.* 2007, Koch & Dahl 2018).

This work aimed to elucidate several of the proposed functions and to characterize the involved enzymes. These included lipoate cofactor biosynthesis, initial thiosulfate oxidation via the Sox system, and sulfur transfer to the sHdr complex. In parallel to this biochemical characterization, tools were developed to determine the occurrence of the sHdr pathway across the diversity of prokaryotes.

### 1.2 Heterodisulfide reductases and related complexes

Heterodisulfide reductases (Hdr) are high molecular weight protein complexes that coordinate multiple co-factors such as non-cubane iron-sulfur clusters, cubane iron-sulfur clusters or flavin-adenine dinucleotides (FAD). In addition to sulfur oxidizing bacteria operating the sHdr system, heterodisulfide reductases and related complexes are found in methanogenic archaea (Kaster *et al.* 2011b, Wagner *et al.* 2017), acetogenic bacteria (Mock *et al.* 2014), sulfate reducing archaea (Mander *et al.* 2004) and bacteria (Ramos *et al.* 2015) and aromatic compound degraders (Huwiler *et al.* 2019). The heterodisulfide reductases of most of these pathways form hexameric (HdrABC)<sup>2</sup> complex, whereby larger complexes can also be formed with additional interaction partners (Wagner *et al.* 2017). Many of the known heterodisulfide reductases are proposed to catalyze a flavin-based electron bifurcation (Appel *et al.* 2021). Transfer of electrons to a low-potential acceptor in an endergonic reaction. Electron bifurcation is a concept in which each of two electron from a medium-potential donor is transferred to a high-potential acceptor in an exergonic reduction coupled to the endergonic reduction of a low-potential acceptor. Thus, the endergonic reduction is enabled by the exergonic reduction with electrons from the same medium-potential donor (Herrmann *et al.* 2008, Li *et al.* 2008, Buckel & Thauer 2013). So far, flavin-based electron bifurcation has been demonstrated *in vitro* for the heterodisulfide reductase from the methanogenic archaeon *Methanothermobacter marburgensis* (Kaster *et al.* 2011b). It is also considered a plausible reaction mechanism for most known Hdr complexes (Appel *et al.* 2021). The sHdr complex is not included into this consideration as spectroscopic analysis of the sHdrA component contradicted an electron-bifurcating reaction mechanism (Ernst *et al.* 2021). The evolutionary process shaping the heterodisulfide reductase from a methanogenic mHdr complex to the heterodisulfide reductase present in the sHdr system would yield insight into the ancient sulfur cycling itself and the mechanism that changed the reaction mechanism of the heterodisulfide reductases. All currently known forms of methanogenesis require a heterodisulfide reductase complex (Garcia *et al.* 2022). Methanogenesis is one of the oldest metabolic processes in Earth history, predating even the sulfur cycle. The oldest geological evidence of microbial methanogenesis has been dated back to more than 3,46 Ga ago (Ueno *et al.* 2006). Current methanogenic archaea are thought to have diversified not more than 3,51 Ga ago, with methanogenesis itself likely to have evolved earlier in a common ancestor of the ancient methanogens (Wolfe & Fournier 2018). It is possible, that the last common ancestor of all archaea already possessed methanogenic capacity (Wang *et al.* 2021, Mei *et al.* 2023). Since the presence of the last common archaeal ancestor has been dated between 3,95 Ga and 3,37 Ga (Mahendrarajah *et al.* 2023), prokaryotic methanogenesis may have occurred in the late Hadean eon about 3,8 Ga



ago (Battistuzzi *et al.* 2004, Wolfe & Fournier 2018). Thus, the methanogenic Hdr complex could be of similar age. The time span for the appearance of methanogenesis is considerably earlier than the appearance of the sulfur cycle, which has been dated to the Archaen eon. Geological evidence for microbial sulfate reduction has been found in several 3,5 Ga old environments (Shen *et al.* 2001, Roerdink *et al.* 2012, McLoughlin *et al.* 2012, Zhelezinskaia *et al.* 2014). Microbial sulfur oxidation was evident not later than 2,5 Ga (Czaja *et al.* 2016). First geological indications of an active sulfur cycle have been found in 3,2 Ga old samples (Nabhan *et al.* 2020). Calculation of the first occurrence of the key components of the Dsr system also revealed the first evidence not more than 3,5 Ga ago (Mateos *et al.* 2023). The first Dsr systems probably reduced sulfate as the reductive type Dsr system is phylogenetically older (Müller *et al.* 2015). The sulfur-oxidizing reverse Dsr system evolved at least twice from the reductive type (Neukirchen *et al.* 2023). Until now, most systems for oxidative sulfur dissimilation were considered to be significantly younger than the reductive Dsr system (Mateos *et al.* 2023). Predation of sulfate reduction before microbial sulfur oxidation is a scenario consistent with the geological evidence. (Czaja *et al.* 2016). The evolution of the sHdr system as a whole and its individual components like the lipoate-binding protein has not yet been elucidated. For this task tool were developed to discriminate between the various heterodisulfide reductases and related complexes (Tanabe & Dahl 2022, 2023).

## 2 The annotation of sulfur metabolism related proteins

The growing use of multi-omics technologies and improvements in sequencing techniques have led to an increasing amount of sequencing data from different environments and uncultivated strains. This recent development has increased the knowledge of microbial interactions and metabolic activities in various environments. However, this development can only be pursued further if individual metabolic pathways can be predicted with a reasonable degree of certainty based on experimentally validated knowledge of the biochemical processes and with adequate computational resources. This assignment of functions to gene products from genomes of cultured and non-cultured organisms can be approached from several directions, including sequence similarity and genomic context (Hawkins & Kihara 2007). Each of these approaches in itself requires computational capacity due to the enormous amount of data to be processed and difficulty of the problem to be solved. As a consequence to the required resources the assignment of names to sequences based on similarity is the principle used by annotation pipelines including the most popular prokaryotic annotation pipeline PGAP from NCBI (Xiao *et al.* 2015). The genomic context, which can provide valuable information about the function of the encoded genes (Hawkins & Kihara 2007), is not a criterion for functional annotation in the PGAP pipeline (Li *et al.* 2021). In the field of sulfur metabolism, these methods often result in unpredictable functional assignments. This challenge is compounded by the striking resemblance between proteins engaged in sulfur metabolism and those participating in the transformation of other elements. To address this issue effectively, the HMS-S-S and HMSS2 tools were developed, significantly enhancing the accuracy of sulfur metabolism predictions. These tools use profiled hidden Markov models (HMM) in conjunction with genomic context analysis for functional annotation (Tanabe & Dahl 2022, 2023).

## 2.1 A historical view on the development of annotation

Functional annotation is essential for understanding microbial physiology and interaction. The question of a nomenclature for the meaningful categorization of enzymes has been discussed ever since.

One significant issue has been the lack of a consistent vocabulary. Each annotator approached the naming of curated genomes differently. Some opted for abbreviations in their designations, while others provided a complete description. Genes with the same predicted function may be described by interchangeable synonyms. For electronic data processing and even for human readers, this inconsistency was and is a difficult hurdle to overcome when comparing two proteins (Hawkins & Kihara 2007). One case of this inconsistency is the sulfur transferase TusA, which is also named YhhP (Ishii *et al.* 2000, Katoh *et al.* 2000), Yedf (Ballesté-Delpierre *et al.* 2017), YeeD (Tanaka *et al.* 2020), TsuB (Tanaka *et al.* 2020) and SirA (Yamashino *et al.* 1998). More complexity is added since Sir is also used for sporulation inhibitor of replication protein SirA or the LuxR/UhpA family response regulator (Rahn-Lee *et al.* 2009). Indeed, the naming of *tusA*, *yedF*, and *yeeD* is useful for *Escherichia coli*, as it allows the three co-occurring homologous to be distinguished. However, this principle cannot be extended indefinitely in a meaningful way and requires knowledge about the concrete function.

Several strategies have been used to solve this problem, some of which are still in use today (Hawkins & Kihara 2007, Cruz *et al.* 2017). Gene ontology terms (GO term) and enzyme commission numbers (E.C. numbers) are two of the earliest and most widely utilized classification systems for proteins that provided a uniform vocabulary to protein naming (Hawkins & Kihara 2007). These approaches are complemented by clustering proteins into orthologous groups. In this concept two homologous genes are orthologous if they have evolved from a common ancestor after a speciation event (Tatusov *et al.* 1997). In contrast, homologs that are not the result of a speciation event, i.e., a duplication event, are called paralogs. Extending this concept on multiple homologs gave rise to the similar concept of orthologous groups. In these groups each pair of homologs derived from different lineages are orthologs (Tatusov *et al.* 1997). By definition these groups form monophyletic clades in accurate phylogenetic trees (Gabaldón & Koonin 2013). Assignment of orthologous groups is essentially depending on a all-against-all protein sequence comparison via BLAST or an BLAST alternative (Tatusov *et al.* 1997). Widely used the orthology databases are The Clusters of Orthologous Groups (COG) (Galperin *et al.* 2021), Protein ANalysis THrough Evolutionary Relationships (PANTHER) (Thomas *et al.* 2022), and Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups (EggNOG) (Hernández-Plaza *et al.* 2023), which currently contribute to the UniProt database (Consortium 2023). All these concepts are unified by the ability to cluster proteins in meaningful named groups. Originally the concept of orthology only referred to genes, but was later extended to protein sequences (Gabaldón & Koonin 2013). While this approach enables the clustering of sequences into meaningful groups and facilitates predictions of functions for newly discovered proteins, there are properties that cannot be derived from orthology alone. Two orthologs are not necessarily functionally equivalent, nor are two functionally equivalent genes or proteins necessarily also orthologous (Gabaldón & Koonin 2013). Also paralogs are not necessarily functionally different (Gabaldón & Koonin 2013). Thus, orthology can

support the functional annotation, but is not be used without other criteria for function prediction.

A more functional approach was used by the complementary Pfam and the discontinued TIGRFAM database that provide Hidden Markov Models for protein annotation (Bateman & Haft 2002). Profiled HMMs are probabilistic models that captures the information about the conservation of each amino acid at each position of a sequence. This results in a position-specific probability of occurrence of amino acids, insertions and deletions (Eddy 1998). To create a profiled HMM the probabilities are deduced from a multiple sequence alignments. Unlike BLAST, the HMMs can discriminate between conserved and variable amino acids and sequence motifs when calculating sequence similarity. The selection of sequences for the alignment is handled differently by Pfam and TIGRFAM. Pfam focus is the clustering of similar sequences that rather correspond to protein domains than whole proteins. Focus of TIGRFAM was the clustering of homologous proteins with equivalent specific function, that has been equivalent since their last common ancestor (Haft *et al.* 2001, Bateman & Haft 2002). These were used to create profiled Hidden Markov Models for functional annotation. This database was moved to NCBI in 2018 and incorporated in NCBI's Prokaryotic Genome Annotation Pipeline (PGAP), terminating the creation of new TIGRFAM protein clusters. Until then TIGRFAM covered over 4000 different enzymes and proteins (Haft *et al.* 2018). In PGAP the TIGRFAM HMMs, Pfam HMMs are used together with HMMs from NCBI protein clusters. The latter provide protein clusters that are formed by highly similar sequences that are merged by annotators into larger clusters based on E.C. numbers, previous annotation and literature (Klimke *et al.* 2009). Furthermore, there are several thousands of BLASTRules. Annotation by BLASTRule assigns the name of a reference protein to a unknown protein if a certain threshold of similarity is met. PGAP then hierarchically orders matches from BLASTRules and HMMs to assign names during annotation (Li *et al.* 2021). Genomic context is not a criterion in any of the hierarchies (Li *et al.* 2021).

## 2.2 Proteins in sulfur metabolism

The profiled HMMs are broadly used to assess the sulfur metabolism capacity of prokaryotes. The efforts of most sulfur metabolism related ecological studies focus on the reduction of sulfur from sulfate (+VI) to sulfide (-II). In contrast, sulfide oxidation, the conversion of inorganic sulfur compounds of intermediate oxidation states and the turnover of organic sulfur compounds are only studied to a limited extent. As a consequence the publicly available HMMs and prediction tool are mostly limited to sulfate reduction and oxidation via dissimilatory sulfite reductase (Anantharaman *et al.* 2018, Neukirchen & Sousa 2021). Although a lack of knowledge regarding the oxidation of sulphur and the conversion of organic sulphur compounds has been identified, the microbial diversity responsible for this conversion has only been studied to a limited extent (Wolf *et al.* 2022). Correcting this deficiency is complicated by the complexity of the pathways involved and the existence of parallel pathways in some cases even within the same organism. In addition, several redox proteins involved in the conversion of sulfur compounds are orthologous to proteins that are involved in the turnover of other elements increasing the potential of false interpretations (Tanabe & Dahl 2022).

Heterodisulfide reductases and related complexes are only one example of a multienzyme com-

plexes that may serve in completely divergent metabolisms. The proteins of the Dsr pathway may serve as another example for this complexity. This pathway is named after its key enzyme, the dissimilatory sulfite reductase DsrAB, and consists of up to 16 currently named Dsr proteins (Rabus *et al.* 2015). A wide range of bacteria and archaea implement this pathway for the reduction of sulfite to sulfide (Müller *et al.* 2015, Anantharaman *et al.* 2018, Chernyh *et al.* 2020). The *dsrAB* genes are often encoded together with its allosteric activator *dsrD* (Ferreira *et al.* 2022). The latter gene is a widely used marker gene for sulfate reduction, although *dsrD* is not necessarily present in the genome of sulfate reducers (Anantharaman *et al.* 2018). The remaining genes encoding for the Dsr system are usually encoded in a second syntenic gene cluster elsewhere in the genome. In an array of sulfur oxidizers the same enzymes are operated in the reverse direction, generating sulfate from reduced sulfur compounds (Pott & Dahl 1998). In these bacteria the *dsr* genes are usually organized in a single syntenic gene cluster in the absence of *dsrD*. Phylogenetically, DsrAB and most of the other Dsr protein forms two clades, one of which can be assigned to the oxidative type and the other to the reductive type (Neukirchen *et al.* 2023). However, the presence of either type does not strictly correlate with the catalyzed direction, as some bacteria like *Desulfurivibrio alkaliphilus* oxidize sulfur with a reductive type Dsr-system (Thorup *et al.* 2017). In addition, the type of is not necessarily consistent across all Dsr proteins and mixed type Dsr systems may occur (Löffler *et al.* 2020b). To to make statements about metabolic capacity despite these difficulties, rules based on the simultaneous occurrence of different marker genes and the presence of the reductive or oxidative Dsr system type have been postulated to circumvent the challenge of inconclusive markers (Anantharaman *et al.* 2018). An approach that is also applicable for many other reactions of the sulfur metabolism.

### 2.3 Sulfur metabolism prediction with HMS-S-S and HMSS2

HMS-S-S was developed to systematically assess the ability of microbes to utilize inorganic sulfur compounds (Tanabe & Dahl 2022). This tool uses profiled Hidden Markov Models (HMM) to annotate 164 proteins involved in the conversion and transport of sulfate, sulfite, thiosulfate, tetrathionate, zero-valent sulfur, polysulfide and sulfide. A special feature of HMS-S-S is the recognition of the genomic context for each detected protein. A prerequisite for function predictions is an initial detection by homology with a decent confidence. The genomic context can be utilized as an additional information to support function predictions. The HMMs of HMS-S-S were created using manually curated protein sequences derived from prokaryotes with described sulfur metabolism. Initial sequence clustering was performed by sequence similarity, the orthology and genomic context to ensure the functional equivalence of all sequences in one cluster (Tanabe & Dahl 2022). Sequences were then aligned using M-coffee, an algorithm that combines multiple alignment algorithms to increase alignment quality (Wallace *et al.* 2006). Similar to the BLAST algorithm, searching genomes with HMMs returns a list with detected proteins and hit-specific scores and an E-values. HMMs are highly sensitive to distant homologs. Typically HMM-specific similarity thresholds are introduced to exclude too divergent protein sequences and to reduce the rate of false positive hits (Hong *et al.* 2020, Garber *et al.* 2020, Neukirchen & Sousa 2021, Zhou *et al.* 2022). The cutoffs for HMS-S-S HMMs were calculated from the protein sequences in the align-

ment using a standardized nested cross-validation procedure. These cutoffs were then validated in a cross-validation procedure to ensure their reliability. Additionally, the HMMs were validated using sequences from proteins that were not used to generate them, resulting in a similar or better performance to tools where the cutoff values were assigned through manually inspection to exclude false hits (Garber *et al.* 2020, Neukirchen & Sousa 2021, Zhou *et al.* 2022). Although the latter is a much simpler concept, it is highly dependent on individual choices and can lead to biases (Chicco 2017).

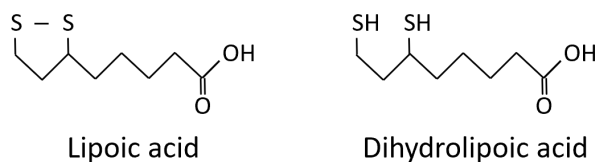
With the development of HMSS2 the capabilities of HMS-S-S have been extended to include convenient functions for generating data sets for phylogenetic analysis, display by iTol, synteny-based functional annotation and 134 HMMs for organic sulfur metabolism (Tanabe & Dahl 2023). Additional HMMs include the degradation, transport and synthesis of sulfoquinovose and DMSP, which are the most abundant terrestrial and aquatic organic sulfur compounds, respectively (Kiene *et al.* 2000, Goddard-Borger & Williams 2017). Sulfur from these compounds or their degradation products can either be assimilated or is excreted as sulfate (Ruff *et al.* 2003), sulfite (Koch & Dahl 2018, Li *et al.* 2023), thiosulfate (de Zwart *et al.* 1997), tetrathionate (Boden *et al.* 2010), or sulfide (Peck *et al.* 2019). Linking organic and inorganic sulfur metabolism, these compounds may serve as sources of sulfur for assimilation or may be involved in the metabolism of sulfur-reducing, -oxidizing, or -disproportionating prokaryotes (Jørgensen *et al.* 2019, Vigneron *et al.* 2021). The underlying metabolic pathways are often constructed in a modular structure. The genes corresponding to a single module or even entire pathways are often found in a distinguishable gene arrangement. In general closely related prokaryotes tend to maintain the order of genes in their genomes. However, the conservation of gene order decreases rapidly with increasing evolutionary distance between prokaryotic species (Huynen & Bork 1998). The preservation of gene order between distantly related organisms is considered to be the result of evolutionary pressure that maintains similarity. This conservation may be due to a crucial role that the genes in question play in a metabolic pathway, where the specific arrangement of the genes is critical for the efficient functioning and regulation of these biological processes (Tamames 2001). As the general gene cluster arrangement of genes involved in sulfur metabolism is regularly maintained between currently known species this can be used to support the annotation quality of currently unknown species (Friedrich *et al.* 2008, Anantharaman *et al.* 2018, Tanabe & Dahl 2022, 2023).

HMS-S-S provides an algorithm to detect neighboring genes and detect specified patterns. With the upgrade from HMS-S-S to HMSS2, this feature has been extended to be used for annotation as well (Tanabe & Dahl 2023). After the initial detection of homologs by HMMs, HMS-S-S receives from the user a list of grouped genes, each forming a named gene cluster. The algorithm then checks whether there are gene clusters in the analyzed genomes that consist of the specified groups of genes. If a gene cluster is found that matches one of the predefined gene clusters, the number of genes from these two clusters is calculated simultaneously. In the more advanced version HMSS2, incomplete gene clusters in the genome are further examined to determine whether there are genes that are physically located in the gene cluster, but whose gene products did not have sufficient sequence similarity to meet the cutoff when searched using the Hidden Markov Model. These candidate proteins are then tested to see if they have remote similarity to one of the missing genes

and if lowering the cutoff would lead to the completion of the gene cluster. If this is the case, the gene is annotated despite the sequence divergence (Tanabe & Dahl 2023).

### 3 The lipoate-binding protein of the sHdr system

The lipoate-binding protein LbpA is an essential component of the sHdr system (Cao *et al.* 2018). Lipoate is an organosulfur cofactor formed by insertion of two sulfur atoms at the C6 and C8 positions of an octanoic acid (Fig. 2). Unlike other cofactors, lipoate must be covalently bound to its cognate enzyme before it is functionally active. Until the discovery of LbpA, only the multienzyme complexes  $\alpha$ -ketoacid dehydrogenases, acetoin dehydrogenase, and the glycine cleavage complex had been described as lipoate-dependent (Cronan *et al.* 2005, Cronan 2016). In these five complexes the lipoate cofactor binds the substrate via a thioether or thioester bond and presents it to the active sites (Spalding & Prigge 2010). Lipoate is reduced from oxidized lipoamide to reduced dihydrolipoamide during catalysis. Reoxidation can reduce  $\text{NAD}^+$  to NADH due to the low biological standard redox potentials of the cofactor ( $E^0' = -0.29$  V). Analogously LbpA might bind sulfane sulfur through the lipoate cofactor, which is reduced during the sulfur oxidation (Ernst *et al.* 2021). The sHdr system could therefore also generate NADH by transferring electrons from LbpA to  $\text{NAD}^+$ . LbpA differs from other lipoic acid-dependent enzymes. The lipoate biosynthetic machinery that provides lipoate for the other complexes cannot lipoylate LbpA (Cao *et al.* 2018). Without this cofactor, apo-LbpA is catalytically inactive. Therefore, the lipoate assembly pathway for LbpA is of interest for the understanding of the sHdr system, its evolution and for the investigation of the reaction mechanism of LbpA.



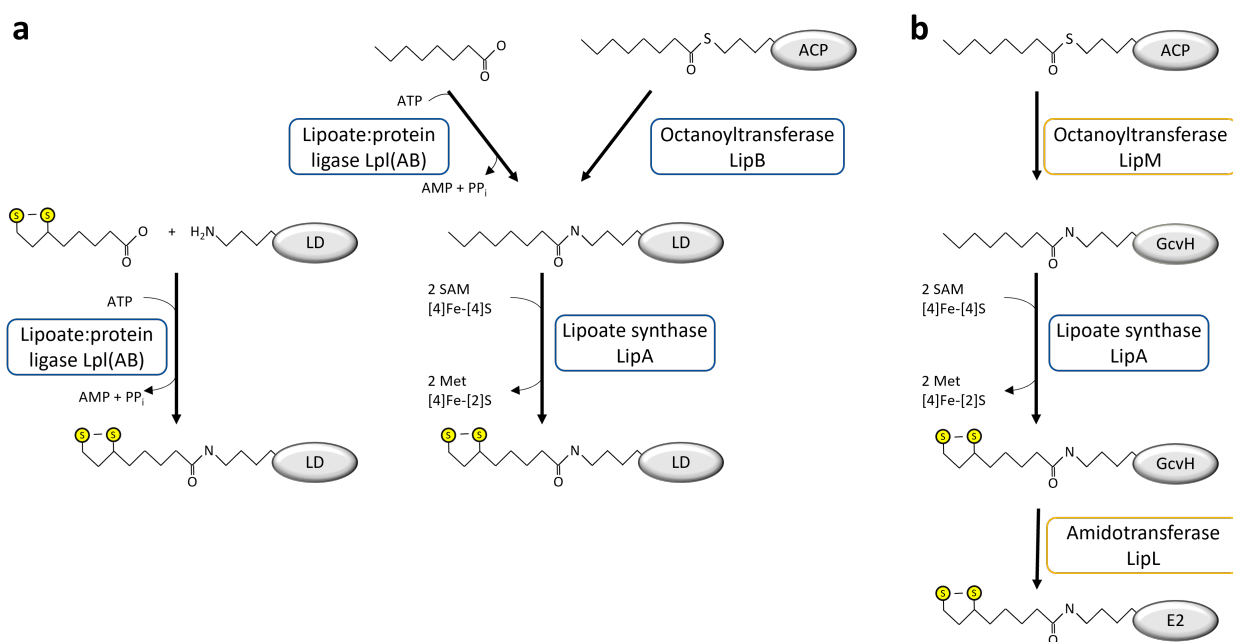
**Figure 2: Structures of oxidized lipoic acid and its reduced derivative dihydrolipoic acid.**

#### 3.1 Lipoate assembly pathways

Lipoate is an organosulfur cofactor consisting of octanoic acid with sulfur atoms inserted at the C6 and C8 positions. To be functionally active, lipoate must be covalently bound to the  $\epsilon$ -amino group of a conserved lysine. The cognate enzymes are the H protein of the glycine cleavage complex and the E2 subunits of the pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase, branched-chain 2-oxoacid dehydrogenase, and acetoin dehydrogenase enzyme complexes (Cronan 2016). All these proteins are supplied by the same lipoate biosynthesis machinery. Up to this point, the lipoate biosynthesis pathway was considered a general-purpose pathway. This view was challenged when LbpA was not lipoylated by this machinery. Instead, a *bona fide* lipoate:protein ligase was shown to lipoylate LbpA in vitro with free lipoate (Cao *et al.* 2018). How the lipoate of LbpA is assembled de novo in *H. denitrificans* remained to be discovered (Cao *et al.* 2018). In fact, this work has shown

that the assembly of lipocate on LbpA takes place via a completely new pathway (Tanabe *et al.* 2023b).

Lipoate can be covalently attached to its cognate enzyme by a lipoate:protein ligase or is assembled on the protein by a de novo assembly system (Fig. 3). The canonical mechanism for de novo assembly has been described in Pseudomonadota (formerly Proteobacteria) and Bacillota (Fig. 3). Lipoylation in both phyla share the same general mechanisms, but the enzymes involved are different, and therefore both pathways are described in detail below. In Pseudomonadota the de novo lipoate assembly pathway is formed by a lipoate synthase LipA and an octanoyltransferase LipB (Fig. 3a). The de novo synthesis starts with the transfer of octanoic acid from the octanoyl acyl carrier protein to the lipoate-dependent enzyme. Octanoyl is an intermediate product of fatty acid biosynthesis. The octanoyltransferase LipB catalysis this transfer without the consumption of ATP. Acceptor proteins recognized by LipB in Pseudomonadota are the E2 subunits of  $\alpha$ -ketoacid dehydrogenases, acetoin dehydrogenase, or the H-protein GcvH of the glycine cleavage complex. The two sulfur atoms at the C6 and C8 positions are added by the lipoate synthase LipA (Cronan 2016). Lipoate synthases are members of the radical S-adenosyl-l-methionine enzyme superfamily that generate 5'-deoxyadenosyl radicals from SAM for the radical-based insertion. The donor for both sulfur atoms is a single [4Fe-4S] cluster, which is consumed along with two SAM molecules. Both atoms are inserted sequentially without releasing a mercaptoctanoylated intermediate (Cicchillo & Booker 2005, Douglas *et al.* 2006, Lanz *et al.* 2014, Cronan 2016).



**Figure 3: Canonical lipoate assembly** (a) Described lipoate assembly in *E. coli* and other Pseudomonadota. Lipoate or octanoate is bound to the  $\epsilon$ -amino group of a conserved lysine of the lipoyl domains (LD) of the  $\alpha$ -ketoacid dehydrogenases, acetoin dehydrogenase, or the H-protein of the glycine cleavage complex (GcvH). (b) Described lipoate assembly in *Bacillus subtilis* and other Bacillota. Lipoate is exclusively assembled on GcvH and is then distributed to the lipoate-binding E2 subunits of the  $\alpha$ -ketoacid dehydrogenases or acetoin dehydrogenase. Lipoylation by lipoate:protein ligase also occurs in Bacillota, but is not shown here.

The de novo synthesis described in *Bacillus subtilis* differs from this LipB/LipA pathway (Fig. 3b). In *B. subtilis*, octanoyltransferase LipB is replaced by octanoyltransferase LipM, which transfers octanoyl residues exclusively to GcvH (Christensen & Cronan 2010, Martin *et al.* 2011). Lipolate synthase LipA then adds the sulfur atoms on the GcvH-bound octanoyl residue (Cronan 2016). The assembled lipolate cofactors are then redistributed to the E2 subunits by an amidotransferase LipL (Christensen *et al.* 2011,b, Teoh *et al.* 2019). LipM and LipB share the same general catalytic mechanism and a similar folding, but have low primary sequence similarity (Cronan 2016). Within the protein family of small cofactor transferase proteins, LipB and LipM are each more closely related to the lipolate:protein ligases than to each other. The octanoyltransferases LipB and LipM have therefore been proposed to have evolved from a common ancestral protein by two routes in a functionally and mechanistically convergent evolutionary process (Cronan 2016).

Another enzyme that catalyzes the attachment of free lipolate is the lipolate:protein ligase (Lpl). The first lipolate:protein ligase was purified from *E. coli* and characterized as a single domain enzyme (Morris *et al.* 1994, Green *et al.* 1995). It was later recognized that the lipolate:protein ligases require a second domain named LplB for the ligase activity (Christensen & Cronan 2009, Posner *et al.* 2009). Most characterized canonical lipolate:protein ligases are fusions proteins of the catalytic domain LplA and the small accessory domain LplB (Fujiwara *et al.* 2005, Christensen & Cronan 2009, Martin *et al.* 2011, Cao & Cronan 2015, Jin, Chen, Wang, Zhu, Liu, Shi, Xin & Liu 2020). Lipolate:protein ligases have in general a rather broad substrate spectrum and bind free lipolate, octanoate and similar carbon compounds to modify of apo-proteins in an ATP-dependent reaction (Cronan 2016). Due to this octanoylation activity the lipolate:protein ligase allows the growth of *E. coli* strains lacking the LipB octanoyltransferase (Hermes & Cronan 2009). Other proteins with lipoylation or octanoylation are not present in *E. coli* (Morris *et al.* 1995).

In this work, a novel lipolate biosynthesis pathway is shown to be active in *H. denitrificans* (Tanabe *et al.* 2023b). This pathway required at least a *bone-fide* lipolate:protein ligase (sLpl(AB)) and two lipolate synthases LipS1 and LipS2 and specifically assembled the lipolate cofactor on LbpA. This novel pathway is simultaneously active and not redundant to the canonical lipolate assembly machinery (Tanabe *et al.* 2023b). HMSS2 in combination with TIGRFAM HMMs were used to assess the presence of lipolate assembly systems across the currently known diversity of prokaryotes, revealing a much broader distribution than previously described (Tanabe *et al.* 2023b). The novel lipolate assembly pathway was also not restricted to sulfur-oxidizing prokaryotes using the sHdr system, but was much more widespread, especially among archaea. Phylogenetic analysis also revealed the evolutionary origin of the machinery and suggested an archaeal origin of the novel lipolate assembly pathway (Tanabe *et al.* 2023b).

### 3.2 The lipolate-binding protein LbpA

The *shdr* genes are usually tightly linked to at least one gene encoding for the lipolate-binding protein LbpA. often the *shdr* genes are accompanied by two and in a few cases even up to three paralogous *lbpA* genes (Cao *et al.* 2018). Phylogenetically, LbpAs have been shown to cluster in two distinct clades, with the included lipolate-binding proteins named LbpA1 and LbpA2 accordingly. Both protein types are commonly present in *shdr* clusters with more than a single *lbpA*. The



reason for the presence of two different LbpA proteins sulfur oxidizers is unexplained. It is possible that more than a single LbpA monomer functionally active in the sHdr-LbpA complex. This would be analogously to the lipoate-binding E2 subunits of the pyruvate dehydrogenases which contain up to three lipoyl domains. LbpA are phylogenetically most closely related lipoate-binding glycine cleavage system H (GcvH) proteins. GcvH and LbpA proteins share both small globular proteins with a similar structure. Despite considerable sequence similarity, the LbpA proteins from the sulfur oxidizers *Thiorhodospira sibirica*, *Thioalkalivibrio sp.* K90mix and *Aquifex aeolicus* cannot functionally replace the GcvH protein in *Bacillus subtilis* (Cao *et al.* 2018,b). The most conspicuous common features of LbpA that is not present in GcvH are two strictly conserved cysteine residues, one near the N-terminus and the other located close to the C-terminus. The cysteine at the C-terminus is also part of a pentapeptide that is similar to the sulfur-binding and transmitting swing arm of SoxY (Sauvé *et al.* 2007). Thus, these cysteines, the lipoate cofactor or both might have a sulfur mediating function.

#### 4 Sulfur trafficking in the sHdr-system

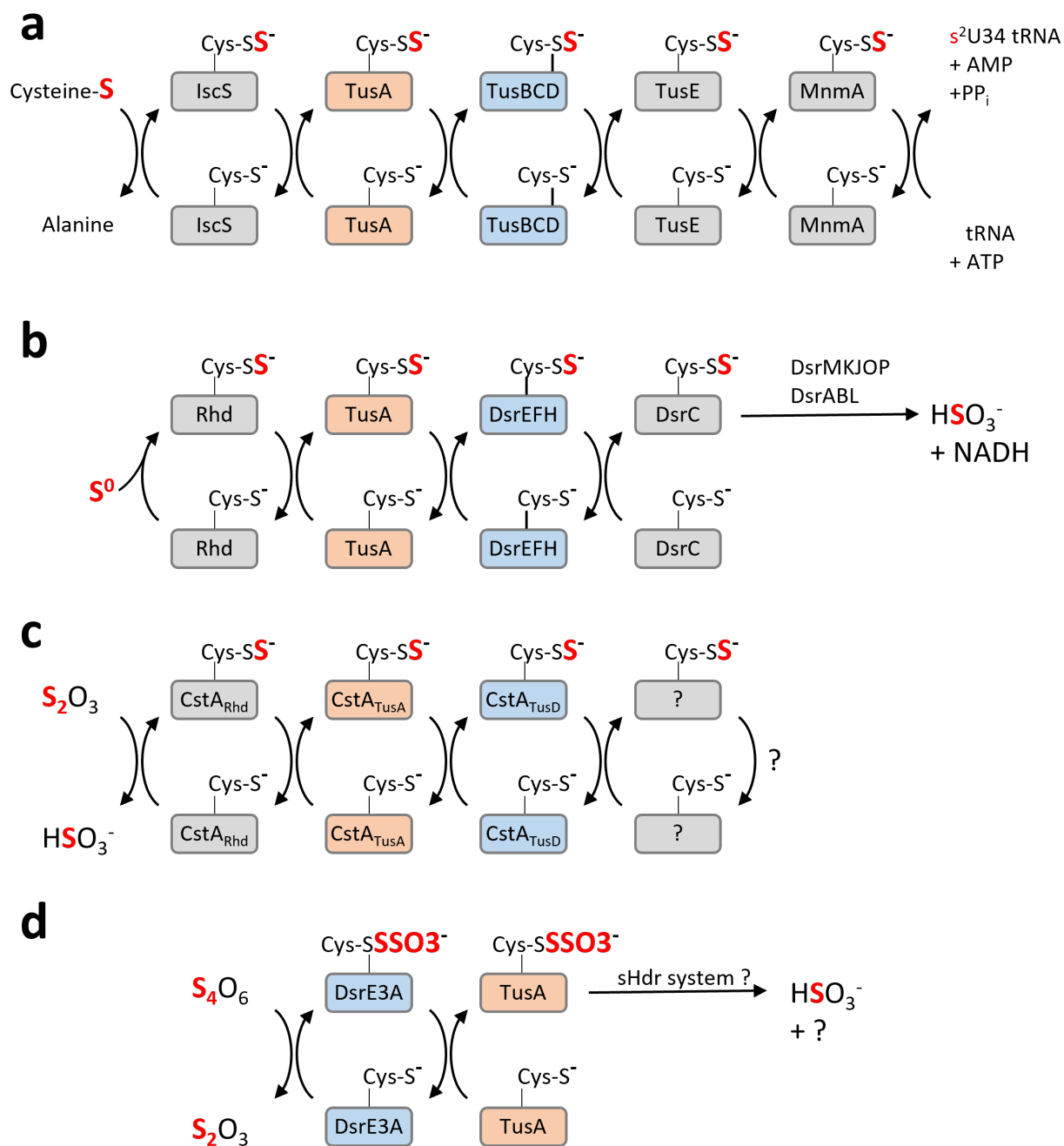
Cellular sulfur transfer is of general importance for prokaryotes, as the concentration of protein-bound and free sulfur species as well as the redox state of sulfur must be strictly regulated. Due to the versatile effects of reduced sulphur compounds, the sulphur does not diffuse freely through the cytoplasm, but is directed to the various metabolic pathways in a controlled sequential, protein-bound transfer (Dahl 2015, Tanabe *et al.* 2019). This transport is catalyzed by sulfur transferases, which often form interconnected sulfur relay systems and may supply sulfur to multiple anabolic and catabolic pathways (Dahl 2015, Tanabe *et al.* 2019). These pathways can be the biosynthesis of important cofactors or cell components like iron sulfur clusters, biotin, thiamin, molybdopterin, cysteine, methionine and thiolated tRNA (Kessler 2006, Shigi 2014, Kümpel *et al.* 2024). Sulfur oxidizing prokaryotes utilizes reduced sulfur compounds as a source of electrons for anaerobic or aerobic chemotrophic and anaerobic phototrophic growth (Dahl 2015). Sulfide, polysulfide and sulfite are may serve as electron source in prokaryotes during dissimilatory sulfur oxidation (Kappler & Dahl 2001). At low concentrations, these sulfur species have a function in cell signaling, redox homeostasis, and metabolic regulation (Mishanina *et al.* 2015, Filipovic *et al.* 2018). Higher concentrations of free intracellular sulfide are harmful and inhibit cytochrome c oxidase (Nicholls 1975). Polysulfide and sulfite can damage proteins, DNA and lipids through spontaneous chemical modification and contribute to oxidative stress through radical formation (Mishanina *et al.* 2015). Sulfur oxidizing prokaryotes transport these compounds in protein bound form and regulate their cytoplasmic concentration through transporters. The existance of a sulfur transferring system to the sHdr complex is well possible. However, the involved transporters and sulfur transferases channeled sulfur to the sHdr complex and the processes influences by the sulfur transporting system remain to be revealed.

#### 4.1 Sulfur transferases and sulfur transfer systems

The *shdr* genes are often genetically linked to the genes for several sulfur transferases of the TusA and DsrE family. These sulfur transferases are proposed to successively transport protein-bound sulfane sulfur to the sHdr complex (Koch & Dahl 2018, Cao *et al.* 2018). The TusA proteins are very versatile sulfur transferases involved in various anabolic and catabolic metabolisms (Tanabe *et al.* 2019). Confirmed functions of TusA proteins include the tRNA thiolation (Ikeuchi *et al.* 2006), molybdenum cofactor biosynthesis (Dahl *et al.* 2013b), thiosulfate assimilation (Tanaka *et al.* 2020), sulfide detoxification (Shen *et al.* 2015) and dissimilatory sulfur oxidation (Stockdreher *et al.* 2014). Regulation of sulfur oxidation (Li *et al.* 2023b) and motility (Ballesté-Delpierre *et al.* 2017) are proposed functions. Furthermore, TusA and DsrE fusion proteins are possibly a component of selenocysteine biosynthesis (Shaw *et al.* 2012, Peng *et al.* 2016). Each individual TusA acts specifically on its own function(s), but does not substitute for paralogs involved in other pathways (Tanabe *et al.* 2019). Thus, the three *E. coli* paralogs named *tusA*, *yedF* and *yeeD* respectively act only in tRNA thiolation and molybdenum cofactor biosynthesis (Dahl *et al.* 2013b), cell motility (Ballesté-Delpierre *et al.* 2017) and thiosulfate assimilation (Tanaka *et al.* 2020). These homologs are not redundant and do not compensate for each other's activity. Deletions of individual homologs therefore lead to different phenotypes (Tanabe *et al.* 2019). Prokaryotes with a single TusA may also require it for multiple essential and non essential metabolic functions (Stockdreher *et al.* 2014). Deletion attempts usually lead to pleiotropic effects or non-viable cells (Dahl *et al.* 2013b, Filiatrault *et al.* 2013, Tomblin *et al.* 2013, Stockdreher *et al.* 2014, Tanabe *et al.* 2019). The diverse metabolic pathways in which TusA and DsrE interact are summarized below.

The TusA family was originally named for the 2-thiouridine modification of uracil 34 on tRNA for lysine, glutamine, or glutamate (Ikeuchi *et al.* 2006). The sulfur for this modification is channeled to the tRNA via seven proteins (IscS, TusA, TusBCD, TusE, and MnmA) that form a persulfide shuttling sulfur relay system (Ikeuchi *et al.* 2006). First, the sulfur atom of a free cysteine is used by cysteine desulfurase IscS to generate a persulfide at the active site cysteine of IscS. This persulfide is then successively transported via the active site cysteines of the sulfur transferases TusA, TusD of the heterotrimeric TusBCD complex and TusE to MnmA. Finally, the sulfane sulfur is used by MnmA to generate the 2-thiouridine modification at uracil 34 (Ikeuchi *et al.* 2006, Shigi 2014). The same TusA is also involved in molybdenum cofactor biosynthesis of *Escherichia coli* (Dahl *et al.* 2013b).

The sulfur transfer in the reverse Dsr system of *Allochromatium vinosum* is analogous to this process. The involved proteins TusA, DsrEFH and DsrC are homologous to the TusA, TusBCD and TusE for thiouridinylation. During this process, sulfur is activated from periplasmic sulfur globules and transferred to TusA by an unknown mechanism, possibly involving a rhodanese type sulfur transferase. The sulfane sulfur is then relayed via the active site cysteines of TusA, DsrE from the the DsrEFH complex to DsrC (Stockdreher *et al.* 2012, 2014). Persulfurized DsrC serves as substrate to the dissimilatory sulfite reductase DsrAB releasing sulfite and reduction equivalents (Stockdreher *et al.* 2012). Deletion of the single *tusA* gene in *Allochromatium vinosum* resulted in a barely viable strain incapable of sulfur oxidation. It was concluded that this *tusA* is required for further functions beyond sulfur transport to the Dsr system (Stockdreher *et al.* 2014).



**Figure 4: Canonical lipote assembly** (a) Sulfur relay system from the cysteine desulfurase IscS via the sulfur transferases TusA, TusD (in the TusBCD complex) and TusE to the ATP pyrophosphatase MnmA for the biosynthesis of s<sup>2</sup> U34 tRNA thionucleosides in *E. coli*. (b) Sulfur transfer to the reverse Dsr system in *Allochromatium vinosum* via rhodanese Rhd, TusA, DsrE (in the DsrEFH) complex, to sulfur transferase DsrC. The DsrC-bound sulfur is oxidized by the reverse Dsr system to sulfite. (c) Reaction mechanism of the sulfur transferase CstA in *Staphylococcus aureus*. Sulfur is transferred between the Rhd, TusA and TusD domain of the multidomain sulfur transferase CstA. The terminal acceptor of the sulfane sulfur is unknown. Thiosulfate is generated during sulfide detoxification by sulfide:quinone oxidoreductase and persulfidedioxygenase CstB. (d) In vitro activity of the sulfur transferases DsrE3A and TusA *Metallosphaera cuprina*. For better clarity DsrE3A is depicted as a monomer with a single thiosulfonate, and not as the isolated homotrimer with two thiosulfonate groups per monomeric domain.

Sulfide detoxification in *Staphylococcus aureus* and related species is another process in which sulfur is transferred by TusA and a DsrE-type protein. In this process, sulfane sulfur is shuttled between the rhodanese, TusA, TusD/DsrE domain of the fusion protein CstA. Three enzymes are required for the removal of excess sulfide in *S. aureus*. These are the sulfide:quinone oxidoreductase (SQR), the persulfide dioxygenase-sulfur transferase CstB, and multidomain sulfur transferase CstA. The SQR catalyzes the two-electron oxidation of sulfide into sulfane sulfur coupled to the reduction of a quinone (Shen *et al.* 2016b). CstB first oxidizes this persulfide with molecular oxygen to sulfite and subsequently consumes it again with another persulfide to release thiosulfate (Shen *et al.* 2015). The sulfane sulfur from thiosulfate is activated by the CstA rhodanese domain and transferred successively via the active site cysteines of the rhodanese, TusA and TusD/DsrE domains to a yet undefined acceptor molecule (Higgins *et al.* 2015).

While a common feature of these three processes is the transfer of TusA to DsrE/TusD, sulfur can be transferred unidirectionally in the opposite direction. In the sulfur-oxidizing and sHdr system operating archaeon *Metallosphaera cuprina*, genes for a DsrE-like protein, DsrE3A, and TusA are encoded in the vicinity of the *shdr* genes. Recombinant DsrE3A mobilized thiosulfonate from tetrathionate and transferred the thiosulfonate to TusA. Transfer in the opposite direction was not possible and resulted in the cleavage of thiosulfonate from TusA (Liu *et al.* 2014). Both proteins mobilized thiosulfonate from tetrathionate but not persulfides from sulfide, glutathione persulfide, polysulfide, thiosulfate, or sulfite (Liu *et al.* 2014). The thiosulfonated *M. cuprina* TusA has been proposed as a thiosulfonate donor for the sHdr system (Liu *et al.* 2014).

Several sulfur transferases have been proposed to channel transport sulfur to the sHdr complex based on their genetic association with the *shdr* genes (Koch & Dahl 2018). In this thesis, the putative sulfur transferases TusA, DsrE3 and Rhd442 of *H. denitrificans*, *Aquifex aeolicus*, *Thioalkalivibrio sp.* K90mix and *Thiorhodospira sibirica* were biochemically characterized for their ability to mobilize sulfur from inorganic compounds and for their transferase activity. Analysis of the distribution of these sulfur transferases and transporter systems among the *shdr* encoding prokaryotes with HMSS2 also revealed differences in the co-occurrence of specific sulfur transferases with type I and type II sHdr systems. Furthermore, the regulation of these transferases, transporters and the whole sHdr system was investigated.

## 4.2 Sulfur import and the regulating of dissimilatory sulfur oxidation

The transporters catalyzing the importing of protein-bound sulfur into cytoplasm for the oxidation by the sHdr system is currently unknown. The protein family of the YeeE/YedE transporters has been proposed to be sulfur transporters due to its structural prediction and the presence of three conserved cysteines (Gristwood *et al.* 2011). Two transporter proteins SoxT1A and SoxT1B of the YeeE/YedE transporter class can possibly facilitate this transfer connecting the periplasmic Sox system and the cytoplasmic sHdr system of *H. denitrificans* (Koch & Dahl 2018).

A YeeE/YedE transporter from *E. coli* was structurally resolved and biochemically characterized as thiosulfate importer (Tanaka *et al.* 2020). In *E. coli* thiosulfate and sulfate are imported for sulfur assimilation by the ATP-dependent CysUWA transporter for sulfur assimilation (Sirko *et al.* 1990). In the absence of this transporter, the YeeE/YedE transporter was recognized to be an

alternative route for thiosulfate assimilation (Tanaka *et al.* 2020). The structure of the transporter showed 13 helices, including 9 transmembrane helices and three cysteines facing the internal channel of the transporter. The cysteines were assumed to form Cysteine-S-H-SSO<sub>3</sub> hydrogen bonds with thiosulfate during transport from the periplasm to the cytoplasm (Tanaka *et al.* 2020). This transporter is encoded immediately adjacent to *yeeD*, one of the three TusA paralogs of *E. coli* (Tanabe *et al.* 2019, Tanaka *et al.* 2020). YeeD shares only low sequence similarity with other TusA proteins and accordingly forms a distinct clades in the TusA protein family (Tanabe *et al.* 2019, 2023c). Crystallization of the YeeE/YedE-YeeD complex has revealed a physically interact of these two proteins (Ikei *et al.* 2023). Incubation of YeeD with thiosulfate indicated that YeeD can mobilize sulfur from the inorganic compound. However, this activity should be interpreted with some caution, as the thiosulfate concentration used was much higher than the physiologically occurring (Ikei *et al.* 2023). If true this catalytic activity would distinguish YeeD from TusA associated with sHdr, which does not react with thiosulfate (Tanabe *et al.* 2023c).

The YeeE/YedE class transporters that are located in the *shdr* gene cluster of *H. denitrificans* are phylogenetically closely related to the SoxT1 transporter of the Sox system. In *Pseudoaminobacter salicylatoxidans* *soxT1* was located in a *soxSRT1* operon adjacent to the arsenic response regulator ArsR family protein SoxR (Lahiri *et al.* 2006) and the thioredoxin SoxS (Friedrich *et al.* 2008). The transcription of *soxSRT* differed from the transcription of the remaining *sox* genes (*soxV*, *soxW*, *soxXA*, *soxB*, *soxCD*, *soxYZ*) that were located downstream of *soxSRT* (Lahiri *et al.* 2006). Deletion of the *soxT1* resulted in a complete impairment of sulfur oxidation in *P. salicylatoxidans*. Thus, the transporter SoxT1 and the regulator SoxR were considered to form the thiosulfate sensing system regulating the transcription of the *sox* genes (Lahiri *et al.* 2006). A similar operon structure was found in the *sox* gene clusters of other bacteria (Friedrich *et al.* 2008). Comparison of the genomes also revealed the presence of a second SoxT like transporter in the *sox* gene cluster of *Paracoccus denitrificans* which was named SoxT2 (Friedrich *et al.* 2008, Tanabe & Dahl 2022). The SoxT transporters associated with the sHdr system might therefore have a regulating, a sulfur importing or both functions. Here, the co-occurrence of *soxT1* and *soxT2* transporters as well as the presence of the regulatory proteins *soxR* and the homologous *shdrR* in association with sHdr systems was analyzed using HMSS2 (Li *et al.* 2023b). SoxR and sHdrR are both members of the arsenic response regulator ArsR family and where shown to regulate the transcription of *sox* and *shdr* genes in *H. denitrificans* (Li *et al.* 2023,b).



### III Aims of the thesis

The sHdr system and its lipoate-binding protein are essential components for the dissimilatory sulfur oxidation. This is a completely new function for the lipoate cofactor that remains to be elucidated. How the lipoate cofactor is de novo assembled gets attached to lipoate-binding protein is yet to be shown, since the canonical lipoate biosynthesis machinery does not recognize LbpA as a target protein. A model for the functionality of the sHdr system has been proposed for *H. denitrificans*. This model has some open questions regarding the route of the thiosulfate to the cytoplasm. It is yet to uncover which sulfur transferases carry out the sulfur transfer towards the sHdr complex and LbpA. Thiosulfate decomposition by a truncated Sox system is known from other sulfur-oxidizing bacteria, but remains to be confirmed in *H. denitrificans*. Furthermore, the prokaryotic diversity that might operate the sHdr system, as well as variations of the sHdr system itself, are currently unknown. This thesis aimed to address these questions by:

- The development of software solutions for the annotation of genes related to organic and inorganic sulfur metabolism in (meta)genomes.
- The investigation of the mechanism and evolution of the system that binds lipoate to the lipoate-binding protein of the sHdr system.
- characterization of the sulfur transferases present in the sHdr systems of sulfur-oxidizing bacteria
- confirmation of an active truncated Sox system
- assessing the distribution and composition of sHdr systems across the prokaryotic domains.





## IV Publications and manuscripts included in this thesis

---

Kümpel, C., Grosser, M., **Tanabe, T.S.**, Dahl, C. (2024) Fe/S proteins in microbial sulfur oxidation. *Biochim. Biophys. Acta - Mol. Cell Res.* (under review)

**Tanabe, T. S.**, Bach, E., D'Ermo, G., Mohr, M. G., Hager, N., Pfeiffer, N., Guiral, M., Dahl, C. (2023) A cascade of sulfur transferases delivers sulfur to the sulfur-oxidizing heterodisulfide reductase-like complex. *bioRxiv*, 2023.12.18.572138; doi.org/10.1101/2023.12.18.572138

**Tanabe, T. S.**, & Dahl, C. (2023) HMSS2: An advanced tool for the analysis of sulphur metabolism, including organosulphur compound transformation, in genome and metagenome assemblies. *Mol. Ecol. Resour.*, 23(8), 1930–1945; doi.org/10.1111/1755-0998.13848

Li, J., Törkel, K., Koch, J., **Tanabe, T. S.**, Hsu, H. Y. & Dahl, C. (2023) In the Alphaproteobacterium *Hyphomicrobium denitrificans* SoxR Serves a Sulfane Sulfur-Responsive Repressor of Sulfur Oxidation. *Antioxidants*, 12(8), 1620; doi.org/10.3390/antiox12081620

**Tanabe, T. S.**, Grosser, M., Hahn, L., Kümpel, C., Hartenfels, H., Vtulkin, E., Flegler, W., & Dahl, C. (2023). Identification of a novel lipoic acid biosynthesis pathway reveals the complex evolution of lipoate assembly in prokaryotes. *PLoS Biol.*, 21(6), e3002177; doi.org/10.1371/journal.pbio.3002177

Li, J., Koch, J., Flegler, W., Garcia Ruiz, L., Hager, N., Ballas, A., **Tanabe, T. S.**, & Dahl, C. (2022). A metabolic puzzle: Consumption of C1 compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>. *Biochim. Biophys. Acta - Bioenerg.*, 1864(1), 148932; doi.org/10.1016/j.bbabi.2022.148932

**Tanabe, T. S.**, & Dahl, C. (2022). HMS-S-S: A tool for the identification of Sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies. *Mol. Ecol. Resour.*, 22(7), 2758–2774; doi.org/10.1111/1755-0998.13642



---

# HMS-S-S: A tool for the identification of Sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies

Tanabe, T. S., & Dahl, C.

---

Studying microbial physiology and their impact on ecosystems is currently mainly driven by the rapidly increasing number of genomic and metagenomic assemblies. With the advent of next generation sequencing technologies, the constantly development of accelerated algorithms for the assembling of genomes from metagenomes the availability of genomes exponentially increased over the past few years (Koonin *et al.* 2021). This did not only lead to an enormous increase of the accessible genomes in public databases but also to the discovery of new phyla and superphyla (Koonin *et al.* 2021, Parks *et al.* 2022). Assigning function to this biodiversity is especially challenging for the sulfur metabolism due to the highly complex nature of the associated pathways. Sulfur oxidizing or reducing systems typically consist of multiple modules that can operate independently in the absence of other modules. Some organisms even encode the genetic potential for redundant systems that may be active depending on environmental conditions or available substrate. Predicting metabolic capacity is further complicated by the similarity of sulfur compound conversion systems to enzymes for the turnover of other elements.

HMS-S-S is designed to detect and analyze the sulfur metabolism in genomic assemblies. The tool uses profiled Hidden Markov models (HMM) and the publicly available HMMER3 algorithm to identify proteins and the corresponding operon structures related to inorganic sulfur metabolism related (Tanabe & Dahl 2022). Because HMMs require alignments of homologous sequences to detect protein sequences of the same type, the quality of the alignment is critical to the reliability of the HMM in detecting similar sequences. A novel approach was developed to ensure that the HMMs were built from alignments containing only functionally equivalent sequences. The aligned sequences were collected exclusively from genomes of validly described prokaryotes with established sulfur metabolism. All sequences in the alignment had to share the same orthologous group as well as a similar conserved genomic background or form a monophyletic clade with the other training sequences in a phylogenetic tree (Emms & Kelly 2019, Tanabe & Dahl 2022). The alignment was then computed by combining multiple alignment algorithms and then trimmed using a transition score cutoff to increase the overall alignment quality (Wallace *et al.* 2006).

The reliability of the generated HMMs was tested in several procedures to proof a correct prediction of sulfur metabolism related enzymes. The validation process included a cross-validation procedure, a prediction test with sequences from uncultured divergent species and in several case

---

Tanabe, T. S., & Dahl, C. (2022). HMS-S-S: A tool for the identification of Sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies. *Mol. Ecol. Resour.*, 22(7), 2758–2774; doi.org/10.1111/1755-0998.13642

studies. For the latter, the analysis of genomes from active sulfur cycling environments was replicated using HMS-S-S and compared to the original results. These included prokaryotic genomes assembled from environmental samples derived from sulfur glaciers (Trivedi *et al.* 2020), marine sediments (Pfeffer *et al.* 2012, Kjeldsen *et al.* 2019, Müller *et al.* 2020, Flieder *et al.* 2021), hot springs (Watanabe *et al.* 2019) and acidic peatlands (Hausmann *et al.* 2018). The performance of the HMMs was also compared with other tool of similar purpose. This comparison included the BLASTp and PsiBLAST algorithms, due to their functional similarity to the HMMER algorithm. In addition, the HMMs created here were compared to those of DiSCo, a specialized tool for the detection of the Dsr system (Neukirchen & Sousa 2021). This comparison showed a similar or better performance of the corresponding HMS-S-S HMMs. HMS-S-S finally included a total of 164 HMMs covering the proteins involved in the sulfate and sulfite conversion, the reductive type Dsr system, the oxidative Dsr system, the reductases for polysulfide, thiosulfate, sulfur, and sulfite, the conversion of thiosulfate/tetrathionate, the Sox system for periplasmic thiosulfate oxidation, the sulfide and elemental sulfur oxidation, the sHdr system and the cytoplasmic sulfur transfer (Tanabe & Dahl 2022). The reductive and oxidative Dsr protein types were reliably distinguished by HMS-S-S. Furthermore, HMS-S-S also has an operon structure recognition algorithm and can recognize and report known sulfur metabolism operon structures to the user. Since these structures tend to be highly conserved in distantly related species that share the same pathway, this feature adds another criterion to the reliability of the prediction.

T.S.T. contributed to this study by conceptualization, investigation, data curation, formal analysis, validation, visualization and writing of the original manuscript: T.S.T. conceptualized the program and the HMM development procedure. T.S.T. carried out the literature research, the programming, the data curation for the production and validation of the HMMs, the development and implementation of the operon structure detection algorithm and the visualization of the results. T.S.T. also contributed to writing the manuscript and processing the revisions.



Received: 2 February 2022 | Revised: 25 April 2022 | Accepted: 11 May 2022

DOI: 10.1111/1755-0998.13642

## RESOURCE ARTICLE

MOLECULAR ECOLOGY  
RESOURCES WILEY

# HMS-S-S: A tool for the identification of Sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies

Tomohisa Sebastian Tanabe | Christiane Dahl

Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

**Correspondence**

Tomohisa Sebastian Tanabe and Christiane Dahl, Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich Wilhelms-Universität Bonn, Meckenheimer Allee 168, Bonn 53227, Germany.  
Emails: [sótotana@uni-bonn.de](mailto:sótotana@uni-bonn.de) (T.S.T.), [chdahl@uni-bonn.de](mailto:chdahl@uni-bonn.de) (C.D.)

**Funding information**

Studienstiftung des Deutschen Volkes; Deutsche Forschungsgemeinschaft, Grant/Award Number: Da 351/13-1

**Handling Editor:** Alex C. Buerkle

**Abstract**

Sulphur compounds are used in a variety of biological processes including respiration and photosynthesis. Sulphide and sulphur compounds of intermediary oxidation state can serve as electron donors for lithotrophic growth while sulphate, thiosulphate and sulphur are used as electron acceptors in anaerobic respiration. The biochemistry underlying the manifold transformations of inorganic sulphur compounds occurring in sulphur metabolizing prokaryotes is astonishingly complex and knowledge about it has immensely increased over the last years. The advent of next-generation sequencing approaches as well as the significant increase of data availability in public databases has driven focus of environmental microbiology to probing the metabolic capacity of microbial communities by analysis of this sequence data. To facilitate these analyses, we created HMS-S-S, a comprehensive equivalent hidden Markov model (HMM)-supported tool. Protein sequences related to sulphur compound oxidation, reduction, transport and intracellular transfer are efficiently detected and related enzymes involved in dissimilatory sulphur oxidation as opposed to sulphur compound reduction can be confidently distinguished. HMM search results are coupled to corresponding genes, which allows analysis of co-occurrence, synteny and genomic neighbourhood. The HMMs were validated on an annotated test data set and by cross-validation. We also proved its performance by exploring meta-assembled genomes isolated from samples from environments with active sulphur cycling, including members of the cable bacteria, novel Acidobacteria and assemblies from a sulphur-rich glacier, and were able to replicate and extend previous reports.

**KEYWORDS**

hidden Markov model (HMM) database, sulphur metabolism, sulphur oxidation, sulphur reduction

## 1 | INTRODUCTION

Dissimilatory sulphur metabolism is probably one of the earliest biological strategies of energy conservation (Canfield et al., 2006;

Canfield & Raiswell, 1999; Grein et al., 2013). Organisms participate in the sulphur cycle in two fundamentally different ways: (1) The assimilation of sulphur compounds serves for the biosynthesis of sulphur-containing biomolecules for example, cysteine, methionine

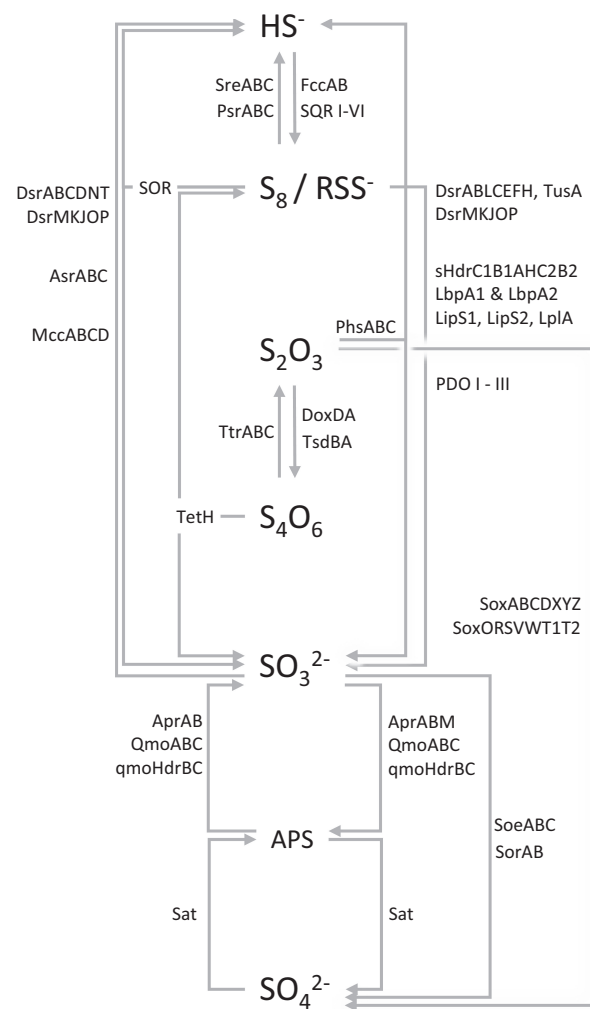
This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

or enzyme cofactors such as biotin or sulpholipids. (2) During dissimilation, sulphur compounds serve as electron donors or acceptors for energy-conserving processes. Understanding the principles underlying dissimilatory sulphur metabolism is especially important, because sulphur-based energy conservation goes along with mass transformation and the relevant organisms are thus the main drivers of biogeochemical cycling of the element.

Sulphur cycling is by no means as easy as switching back and forth between the most oxidized and reduced states, sulphate (+6) and sulphide (-2) (Figure 1). While there is indeed a huge number of sulphate respirers that form sulphide as the end-product, there is also a variety of prokaryotes that reduce compounds of intermediate oxidation states, such as sulphite, thiosulphate, organic sulphoxides, inorganic polysulphides and/or organic disulphanes. On the other hand, chemo- and photolithotrophic sulphur oxidizers apply a complex network of pathways not only enabling the use of sulphide but also an array of other reduced sulphur compounds as electron donors. This may lead to formation of intermediate sulphur species or sulphate as the end product. The intermediates are again all substrates for further microbial oxidation, reduction or disproportionation (Figure 1).

Many new proteins and even completely novel pathways involved in sulphur metabolism have been discovered over the past years. Our understanding of their combination and concerted action has improved not only by biochemical and genetic studies but also by approaches such as strain-resolution genome reconstruction from metagenomes, single-cell genomics, and other molecular “omics” technologies (Anantharaman et al., 2018; Flieder et al., 2021; Hausmann et al., 2018; Wasmund et al., 2017; Zecchin et al., 2018). In fact, new sequencing techniques have yielded a vast amount of data that, once assigned to individual genomes and thoroughly analysed, can provide invaluable assistance in elucidating the metabolic capacity and, moreover, microbial networks at the sites studied. However, this can only succeed if, on the basis of a profound knowledge of the fundamental biochemical processes, individual metabolic pathways can be predicted with a reasonably high degree of certainty. With regard to dissimilatory sulphur metabolism, this point in particular gives rise to quite considerable difficulties because the pathways employed are very complex and parallel metabolic routes have been developed in different prokaryotes or are even applied in the same organism. Even more complexity is added by the fact that a number of redox proteins closely related to those applied in sulphur dissimilation serve functions in the turnover of other elements. This holds true for molybdoenzyme families (Leimkühler & Lobbi-Nivol, 2016), quinone oxidoreductase complexes (Duarte et al., 2021) and also for heterodisulphide reductase-like enzymes originally characterized in methanogenic archaea (Grein et al., 2013). Thus, sulphur dissimilation, metal reduction, hydrogen oxidation, nitrogen metabolism and methanogenesis are all driven at least in part by enzymatic machineries sharing primary structure and protein complex composition, regularly leading to uninformative annotations or even misannotations



**FIGURE 1** The most important conversions of dissimilatory sulphur metabolism. For clarity, assimilatory reactions as well as transformations involving organic sulphur compounds are not shown. Note that the figure does not take into account the intracellular localization of the enzymes involved. Apr, APS reductase; Asr, anaerobic sulphite reductase, Dox, Thiosulphate:quinone oxidoreductase; Dsr, dissimilatory sulphite reductase; Fcc, flavocytochrome c sulphide dehydrogenase; sHdr, sulphur-oxidizing heterodisulphide reductase-like system; LbpA, lipopate-binding protein; LipS, lipoyl synthase; LplA, lipopate: protein ligase; MCC, octaheme c-copper sulphite reductase; PDO, persulphide dioxygenase; Phs, thiosulphate reductase; Psr, polysulphide reductase; Qmo, quinone-interacting membrane-bound oxidoreductase; Sat, sulphate adenyllyltransferase; Soe, sulphite-oxidizing enzyme; Sor, sulphite:acceptor oxidoreductase; Sox, sulphur-oxidizing multienzyme complex; Sqr, sulphide:quinone oxidoreductase; Sre, sulphur reductase; TetH, tetrathionate hydrolase; Tsd, thiosulphate dehydrogenase; Ttr, tetrathionate reductase; S<sub>8</sub>, elemental sulphur; R-SS<sup>-</sup>, cysteine-bound persulphide; <sup>-</sup>S-S<sub>n</sub>-S<sup>-</sup>, polysulphide

of sulphur-related genes by established gene annotation pipelines (Duarte et al., 2021; Rückert, 2016).

Functional annotation is usually based on comparing the similarity of sequences. A probable function is inferred by a combination of clusters of orthologous groups assignment, profile hidden Markov models (HMM) prediction and reciprocal best Blast hit with a reference protein (Tatusova et al., 2016). Indeed, several public databases provide orthologous groups to determine relatedness. However, in the context of sulphur metabolism, some proteins do not fall within any existing orthologous group. In other cases, closely related proteins encoded in the same genetic context and thus probably performing the same function are found in different orthologous groups (Yu et al., 2020). Even more confusion is caused by the same orthologous group accommodating proteins with clearly distinct functions (e.g., QmoA and QmoB and mHdrA share COG1048, although they are involved in dissimilatory sulphate reduction and methanogenesis, respectively). As an addition to orthologous groups, HMMs from PFAM (El-Gebali et al., 2019) and TIGRFAMs (Haft et al., 2013) are broadly used in annotation pipelines. In this probabilistic approach, similarities between a target sequence and a set of reference sequences are determined. Since PFAM HMMs are created for protein domains rather than for the complete length of a protein, proteins with completely different functions can be found within the same PFAM protein family. On the other hand, most TIGRFAMs HMMs are meant to be equivalent, that is, homologues of conserved function (Haft et al., 2013). In terms of function, the TIGRFAMs HMMs are consequently the most suitable for annotation. However, proteins involved in sulphur metabolism are barely covered by existing TIGRFAMs HMMs, and the development of new TIGRFAMs models has ended because of the enormous computational effort needed.

The unsatisfactory coverage of sulphur metabolism-related proteins by orthologous groups and HMMs is the major cause for the almost exclusive focus of comparative genomic studies on only few well-established metabolic routes, for example, the Dsr (dissimilatory sulphite reductase) pathway (Anantharaman et al., 2018). Recently, a number of additional HMMs became available through a specialized tool for Dsr-dependent sulphur metabolism that now allows for differentiation between oxidative and reductive enzyme types (Neukirchen & Sousa, 2021). However, the presence of reductive or oxidative type enzymes is not strictly correlated with the direction in which the pathway is operated in living cells (Thorup et al., 2017). To overcome this problem, rules have been proposed based on the simultaneous presence of various *dsr* genes (Anantharaman et al., 2018). This approach does not fully take into account that the arrangement of the individual genes may also provide crucial aid in predicting function. The presence and genomic arrangement of genes encoding heterodisulphide reductases (Hdr) and related complexes makes up a similar case. The complexes can be involved in a whole range of metabolic functions including methanogenesis and sulphur oxidation, albeit their individual subunits closely resemble each other (Appel

et al., 2021). However, the genomic context provides a valuable marker for functional assignment for example, sulphur related *shdr* genes are usually collocated with genes for lipoate-binding proteins (*lbpA*), while the homologous *hdr* genes from sulphate reducers often occur together with *flx* genes encoding a FAD-containing NADH dehydrogenase (Appel et al., 2021; Cao et al., 2018; Grein et al., 2013; Koch & Dahl, 2018; Ramos et al., 2015). Current synteny prediction algorithms use the orthologous relationship between proteins based on the reciprocal best Blast hit criterion to either support “nonhomology”-based functional assignment by expert annotators (Vallenet et al., 2006) or to compare synteny between genomes on a whole-genome level (Proost et al., 2012; Svetitsky et al., 2019, 2020; Wang et al., 2012). These are neither suitable for scanning sulphur-related gene clusters in large sets of genomes within a reasonable period of time nor do they yield an output that is informative without even further elaborate analyses. To fill this gap, we developed HMS-S-S (Hidden Markov model Supported Sulphur Search system), a tool optionally supported by a graphical user interface. It uses 164 newly prepared equivalent and type-specific HMMs. These are validated by cross-validation, by application on a test data set, that is independent of that used for training, and by application to publicly available genomes and metagenomes from a range of terrestrial and marine habitats with active sulphur cycling. The HMMs are used to find sulphur metabolism-associated sequences in user-submitted genomes. On their own systems, users can locally extend the application by adding their own HMMs as well as HMMs from PFAM, TIGRFAMs or other compatible sources. In the new tool, the possibility is implemented to examine the results of HMM-supported searches for their synteny. Thereby, gene clusters and probable operons become immediately apparent. Gene cluster names are automatically assigned by the program and can be manually curated. The application is designed to easily and quickly filter the results stored in the relational database for the presence, absence or co-occurrence of genes/gene clusters. The application can be downloaded from <https://github.com/TSTanabe/HMSSS>.

## 2 | MATERIALS AND METHODS

### 2.1 | Data set generation

Genomic assemblies, general genomic features, assembly statistics and protein FASTA files were downloaded from NCBI RefSeq (Haft et al., 2018) or GenBank (Sayers et al., 2019) (Figure 2a). The HMM training data set (Figure 2b, Table S1) contained 561 assemblies from the NCBI RefSeq database and were manually selected based on described physiology of the organism/genus, published sulphur-related enzymatic activity, and/or published presence of sulphur-related genes or close homologues thereof. The training set covered three archaeal phyla with seven classes. Among the 14 bacterial phyla, the Proteobacteria represented the phylum with the highest number of classes (7). A test data set (Figure 2c, Table S2) was generated

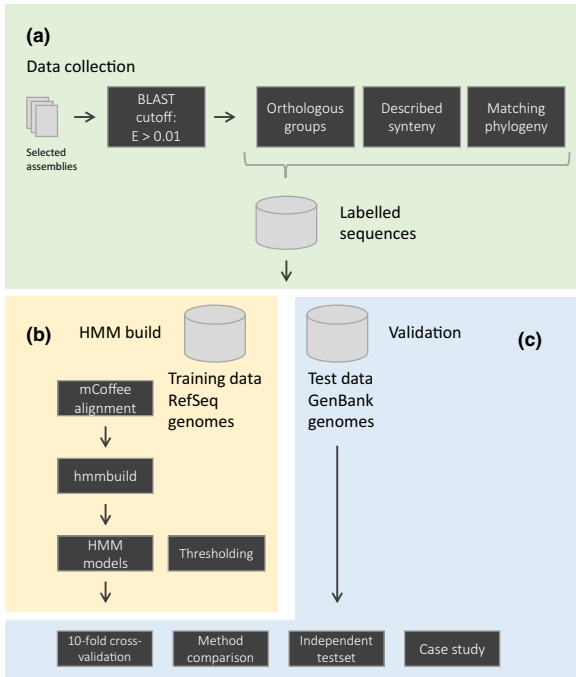


FIGURE 2 HMM design and validation process

that contained 846 further assemblies from GenBank. The test data set included sequences originating from metagenome assembled genomes (MAGs) acquired from environments with active sulphur-cycling (Anantharaman et al., 2016, 2018; Hausmann et al., 2018) and reflected a greater biological diversity than the training set. The test set covered eight archaeal and 28 bacterial phyla. The Proteobacteria still represented the phylum with most classes (nine). The comparison data set included all 1407 assemblies.

## 2.2 | Sequence annotation

The workflow for sequence annotation is depicted in Figure 2a. First, a reference list of proteins related to dissimilatory sulphur metabolism was generated (Table S3, Figure 1). Sequences corresponding to proteins/protein subunits whose functions have been characterized in the literature were downloaded from UniProt (Consortium, 2017). Those sequences were then used as queries in BLASTp v.6.0 (Camacho et al., 2009) searches with an E-value cutoff of 0.01 against the 1407 assemblies of the comparison data set. As a first criterion for annotation of the results, they were grouped into orthologues by Orthofinder (Emms & Kelly, 2019). Then, the groups were named according to the reference protein within the group. As a second criterion, we tested in each case whether the gene environment matched structures previously described in the literature. Sequences that failed one of the criteria were annotated according to phylogenetic tree construction with IQ-TREE (Minh et al., 2020) or remained unlabelled.

## 2.3 | Hidden-Markov-model generation

For HMM generation (Figure 2b), sets of sequences originating from the 561 training assemblies were aligned with M-coffee8 (Wallace et al., 2006). Columns with 95% gaps or positions with an alignment score  $< 4$  were trimmed. These curated alignments were then used as seeds for the generation of HMMs using the hmmbuild command from HMMER. The minimum number of sequences used as a basis for HMM generation was set to 10. Only five seed alignments did not meet this criterion, that is, thiocyanate dehydrogenase (Tikhonova et al., 2020; Tsallagov et al., 2019), the three different subunits of thiocyanate hydrolase (Berben et al., 2019; Katayama et al., 2006), and carbonyl sulphide hydrolase (Ogawa et al., 2013).

## 2.4 | Performance metric calculation

For performance assessment of HMMs, a confusion matrix was created by comparing the assignments made by the HMM with the previously assigned annotations. Hits were counted as true positives (TP) when they matched the sequence annotation or as false positives (FP) when the annotation did not match. All sequences whose annotation matched the respective HMM but were not assigned to the HMM counted as false negative (FN). All other sequences were counted as true negatives (TN). For assessment, balanced accuracy (Brodersen et al., 2010), F1-score (Forman & Scholz, 2010), and Matthew-correlation-coefficient (MCC) (Chicco & Jurman, 2020) were calculated. Values corrected for the skewness of the data set (Jeni et al., 2013) were also calculated (Table S4).

## 2.5 | Cross-validation and thresholding

A nested cross-validation procedure (Varma & Simon, 2006) with a 10-fold outer and 5-fold inner loop was applied to the HMMs to validate performance and assignment of the trusted, noise and optimal thresholds (Figure 2c). Between fold generation for each loop, the order of the sequences in the alignment for HMM generation was randomized. All aligned sequences were split into 10 folds and served as potential TP, all sequences annotated differently from the tested HMM were added to the fold and served as potential TN. For inner folds the TPs of each of the 10 folds were divided equally into 5 folds. The inner folds of all HMMs were calculated first to determine the cutoff scores. The noise cutoff corresponded to the score of the lowest scoring TP hit. The trusted cutoff corresponded to the score of the highest scoring FP hit. The optimized cutoff corresponded to the median of the threshold scores with the highest F1-scores over all inner folds. Each outer fold was combined with all other HMMs generated here in a library before search to test for interferences between the models. The confusion matrix was then created using the optimized thresholds as cutoff. For sequences with multiple hits, the HMM was taken for annotation that scored highest over cutoff. Balanced accuracy was calculated as the mean from all accuracies



from each fold.  $F_1$  score and MCC were calculated as the sum of the confusion matrices from all folds (Forman & Scholz, 2010). The same procedure without fold generation was done for the independent test data set (Chicco, 2017).

## 2.6 | Method comparison

Using the comparison data set, the performance of the HMMs generated in our work was assessed by comparison with BLASTp, Psi-BLAST, ProteinOrtho6 (Lechner et al., 2014) and DiSCo (Neukirchen & Sousa, 2021) (Figure 2c). These methods also served for comparative purposes in previous studies (Neukirchen & Sousa, 2021; Skewes-Cox et al., 2014). Protein assignments by the respective tool were compared with the originally assigned annotations (cf. Section 2.3) to generate a confusion matrix and thereby to assess the performance. The proteins from the initial reference list (cf. Section 2.2) served as query for BLASTp and Psi-BLAST. Sequences with multiple query hits were named by the highest scoring query.

ProteinOrtho6 is a reciprocal-best-blast-hit based tool for orthologous groups. For input into this program, all comparison data set sequences originating from the same assembly were saved in individual FASTA formatted files. ProteinOrtho6 was then run with diamond in default mode and resulting orthologous groups were analysed. When sequences with different annotations appeared in the same group, protein type assignment for this group followed the most frequently encountered annotation. Annotations matching the groups type assignment were counted as true positives. Sequences with annotations not fitting to the type assignment of the group were considered as false positive for the type assignment of the group and as false negative for the protein type annotation. All other sequences were considered as true negatives. Orthologous groups with less than three sequences were not considered.

DiSCo is a HMM type-specific protein sequence predictor for dissimilatory sulphur metabolism. DiSCo was run with default setting for the assignments. If a single protein type was covered by multiple HMMs, the confusion matrix was created from the sum of all hits for this protein, respectively.

## 2.7 | Acquisition of case study metagenomes

For our case studies, nucleotide FASTA data for 76 recently published MAGs was downloaded from the respective repository. ORFs were predicted and translated via Prodigal (Hyatt et al., 2010) with default settings. Protein FASTA files were then searched for sulphur metabolism via HMS-S-S.

Neutrophilic sulphur-oxidizing bacteria (Watanabe et al., 2019): 10 genomes isolated from the water of Lake Harutori and Jozankei hot spring. Assemblies were downloaded from NCBI Bioproject PRJDB7001.

Svalbard Fiords (Flieder et al., 2021): 12 MAGs originating from marine sediment samples from Smeerenburgfjorden, Kongsfjorden

and Van Keulenfjorden, Norway. MAGs were downloaded from NCBI Bioproject PRJNA623111.

Acidic peatland (Hausmann et al., 2018): 13 MAGs originated from samples taken from 10–20 cm depth acidic peatland soil Schlöppnerbrunnen II, Germany. Assemblies were downloaded from NCBI Bioproject PRJEB24926.

Cable bacteria: 10 MAGs originated from a sulphidic sediment sampled in Aarhus Bay, Denmark (Kjeldsen et al., 2019; Pfeffer et al., 2012) and from an enrichment culture from a former coal gasification site in Gliwice, Poland (Müller et al., 2019). Assemblies were downloaded from NCBI Bioprojects PRJNA389779, PRJNA278504 and PRJNA475330.

Borup Fiord Pass (Trivedi et al., 2020): 31 MAGs originated from samples of a sulphur-rich glacier from the Canadian High Arctic. Samples were collected from multiple locations on the glacier over a period of three years. Assemblies were downloaded from Figshare referenced by (Trivedi et al., 2020).

## 2.8 | HMS-S-S workflow

HMS-S-S is an optionally graphical user interface-supported program to study dissimilatory sulphur metabolism and synteny. Requirements are HMMER3, MySQL and Perl which can be easily installed on Linux systems. The overall workflow of HMS-S-S is outlined in Figure 3. The user specifies a directory containing assemblies with contigs or scaffolds in nucleotide FASTA format or alternatively GFF3 files with protein FASTA files. All files in the directory will then be processed in consecutive order. File names and sequence identifiers must be unique. For assemblies from NCBI, assembly statistical files and NCBI taxon database can be used to integrate phylogenetic information for each assembly. For nucleotide FASTA, the open-reading frames (ORFs) are predicted and translated by Prodigal (Hyatt et al., 2010). Using hmmersearch from the HMMER3 package, a custom library of profile HMMs with custom bit score cutoffs is then queried against these ORFs. Hits are matched to their genomic features and stored in a local relational database.

In the second step the data can be processed. Hits from each set can be analysed for colocalization in the respective genome. Hits are considered colocalized when their genes have a user-defined maximum nucleotide distance to the next hit. All hits that meet this condition are assigned to a syntenic block. Each of these blocks can then be matched against user-defined patterns of gene names. If a block matches one or more of these patterns, it is assigned to corresponding keywords. Furthermore, there is an option for adding taxonomic information to the results. All results from data processing are stored in the local database.

The output of the stored results is supported by the user interface. The interface supports filtering of the results according to various criteria, for example, narrowing down to defined phylogenetic groups or to genes specified with co-occurrence and/or syntenic blocks in the same genome. Genes with a specific genomic environment can also be output. As a final output, the results can be displayed in a table as

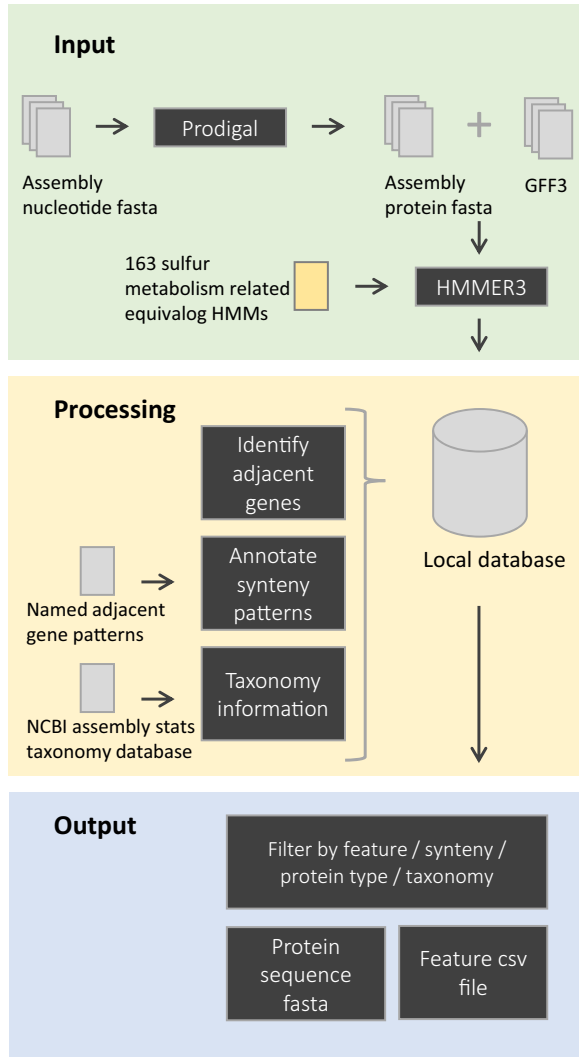


FIGURE 3 HMS-S-S algorithm overview. External programs are prodigal and HMMER3

a CSV file. All identified proteins can be output in the CSV file along with the associated species and all higher taxonomic information summarized, as well as protein identifiers, locus tags, contig, start, end, strand, HMM, bit score and the keyword. Optionally, the output can also be a protein FASTA together with the sequence. The output can be customized via the interface. It is also possible to pass complex queries directly to the local database using MySQL via the interface.

### 3 | RESULTS

We intended to create a comprehensive database of reliable Hidden Markov models (HMMs) based on archaeal and bacterial proteins related to dissimilatory sulphur metabolism. HMM generation was not solely based on sequence similarity but additionally integrated synteny for allocation of a protein to a specific functional group. In a few cases, phylogenies had to be constructed to

guarantee unambiguous sequence annotation. The second important point of our approach was a strict separation between training data sets and test data sets for the generation and later validation, respectively, of the HMMs developed. The HMM training data sets originated from 561 assemblies from the NCBI RefSeq database and were manually selected based on published information. The test set for validation consisted of 846 independent assemblies reflecting a much broader phylogenetic range than the training set.

#### 3.1 | HMM development: Sulphate/sulphite conversions

In dissimilatory sulphate reducers, sulphate has first to be activated to adenosine 5'-phosphosulphate (APS) at the expense of ATP catalysed by sulphate adenylyltransferase (Sat). APS is then reduced to sulphite and AMP is released. This reaction is catalysed by APS reductase (AprAB). In many sulphur oxidizers, the reaction sequence occurs in reverse enabling generation of ATP via substrate level phosphorylation. Two HMMs were set up for Sat, one each for sulphate reducers and sulphur oxidizers while three HMMs cover each subunit of the APS reductases from sulphate reducers and the two different lineages from sulphur oxidizers. The recently discovered AprAB subunits with a probable assimilatory function (Chernyh et al., 2020) fall in separate HMMs. Electron transport to/from APS reductase is provided by the membrane-bound QmoABC complex. Each of its subunits is represented in one HMM. In some sulphur oxidizers, the complete QmoABC electron-accepting unit is replaced by AprM (Meyer & Kuever, 2007a, 2007b). In addition, some other dissimilatory sulphur oxidizers and Gram-positive sulphate reducers do not encode the QmoC subunit but instead carry genes for HdrBC. The latter usually occur as subunits of heterodisulphide reductase from methanogens (Kaster, Moll, et al., 2011; Wagner et al., 2017). For clear distinction from the methanogen proteins (mHdrB and mHdrC), the two HMMs representing the Qmo-associated polypeptides were named qHdrB and qHdrC.

In many dissimilatory sulphur oxidizers, the oxidation of sulphite to sulphate is catalysed by the membrane-bound, cytoplasmically oriented iron-sulphur molybdoenzyme SoeABC, which may be present in addition to the APS reductase-Sat pathway (Dahl et al., 2013). Each Soe subunit is represented by one HMM. Sulphite can also be oxidized directly to sulphate via SorAB, a periplasmic haemomolybdoprotein (Kappler et al., 2000). In some organisms, periplasmic sulphite dehydrogenase is a homodimer of the molybdoprotein SorT, that is closely related to SorA (Wilson & Kappler, 2009). Accordingly, SorT and SorA are represented by the same HMM.

#### 3.2 | HMM development: Dsr-based sulphur reduction/oxidation

In dissimilatory sulphate reducers, reduction of sulphite is achieved by an interplay of an array of Dsr proteins, named for the key enzyme dissimilatory sulphite reductase, DsrAB (Rabus et al., 2015). This enzyme is also active in the cytoplasm of many dissimilatory sulphur oxidizers,

where it is crucial for the generation of sulphite (Pott & Dahl, 1998). The reductive- and oxidative-type enzymes are covered by two HMMs for each subunit. The subunits of the electron-transporting DsrMK(JOP) complex as well as the sulphur-binding cosubstrate protein DsrC and the probable siroheme-amidating enzyme DsrN also each make up one HMM specific for sulphite reduction versus sulphur oxidation. The sulphurtransferase DsrEFH and the iron-sulphur flavoprotein DsrL are present in the vast majority of sulphur oxidizer genomes and have documented essential function during sulphur oxidation in the purple sulphur bacterium *Allochromatium vinosum* (Dahl et al., 2008; Löffler, Feldhues, et al., 2020). DsrEFH and reductive- as well as oxidative-type DsrC are well distinguished from their paralogues TusBCD and TusE (Ikeuchi et al., 2006; Numata et al., 2006), that are involved in 2-thiouridine biosynthesis and fall in separate HMMs. Detailed phylogenetic analyses proved that DsrL proteins fall into three groups, one of which occurs exclusively in sulphur oxidizers (Löffler, Wallerang, et al., 2020). We could not separate these three types, but only create one HMM for all DsrL types because the training data set did not include enough sequences of the NADP(H)-reactive type DsrL-2. Currently there are very few genomes with the respective gene among RefSeq genomes. HMMs were also implemented for DsrD, an allosteric activator of DsrAB (Ferreira et al., 2022), and DsrT, a protein of unknown function (Holkenbrink et al., 2011), as well as for DsrR and DsrS, that are involved in posttranslational modification of Dsr proteins in a number of sulphur oxidizers (Grimm et al., 2010, 2011).

### 3.3 | HMM development: Sulphite, polysulphide, thiosulphate and Sulphur reduction

For enzymes catalysing sulphite reduction, 10 HMMs were made. These include three HMMs for anaerobic sulphite reductase AsrABC (Huang & Barrett, 1991) catalysing hydrogen sulphide production from sulphite. Octaheme sulphite reductase MccABCD (Kern et al., 2011) is covered by four HMMs. The A, B and C subunits of polysulphide, thiosulphate and sulphur reductases (Psr/Phs/Sre) were combined in one HMM, respectively, therefore a total of three HMMs cover these pathways.

### 3.4 | HMM development: Thiosulphate/tetrathionate conversion and tetrathionate oxidation

Thiosulphate to tetrathionate conversion can either be catalysed via DoxDA (Müller et al., 2004) or via thiosulphate dehydrogenase TsdBA (Denkmann et al., 2012). Tetrathionate reductase TtrABC catalyses the reverse reaction, regulated by response regulator TtrRS (Hensel et al., 1999). Each subunit is covered by an HMM, making a total of seven HMMs for thiosulphate/tetrathionate conversion and two HMMs for the regulator. Tetrathionate hydrolase (TetH) converts tetrathionate to sulphite, sulphur and thiosulphate and is represented by one HMM (Kanao et al., 2007).

### 3.5 | HMM development: Thiosulphate oxidation in periplasm

Thiosulphate oxidation to sulphate in the periplasm is catalysed by the complete Sox system involving proteins SoxYZ, SoxXA, SoxB and SoxCD (Friedrich et al., 2005). SoxV and SoxW (Appia-Ayme & Berks, 2002), SoxO (Pyne et al., 2017), SoxF and SoxE (Bardischewsky et al., 2006), SoxG and SoxH (Rother et al., 2001), SoxT1 and SoxT2 (Friedrich et al., 2008; Lahiri et al., 2006), SoxS (Rother et al., 2008), and SoxR (Mandal et al., 2007; Rother et al., 2005) may be present depending on the organism. A total of 16 HMMs were constructed representing all Sox proteins except SoxR, which is covered by the same HMM as the transcriptional repressor sHdR (cf. Section 3.8). An additional HMM for the sulphur transporter YeeE was made in order to distinguish between SoxT1, SoxT2 and YeeE, dedicated to thiosulphate uptake, and related transporters (Tanaka et al., 2020).

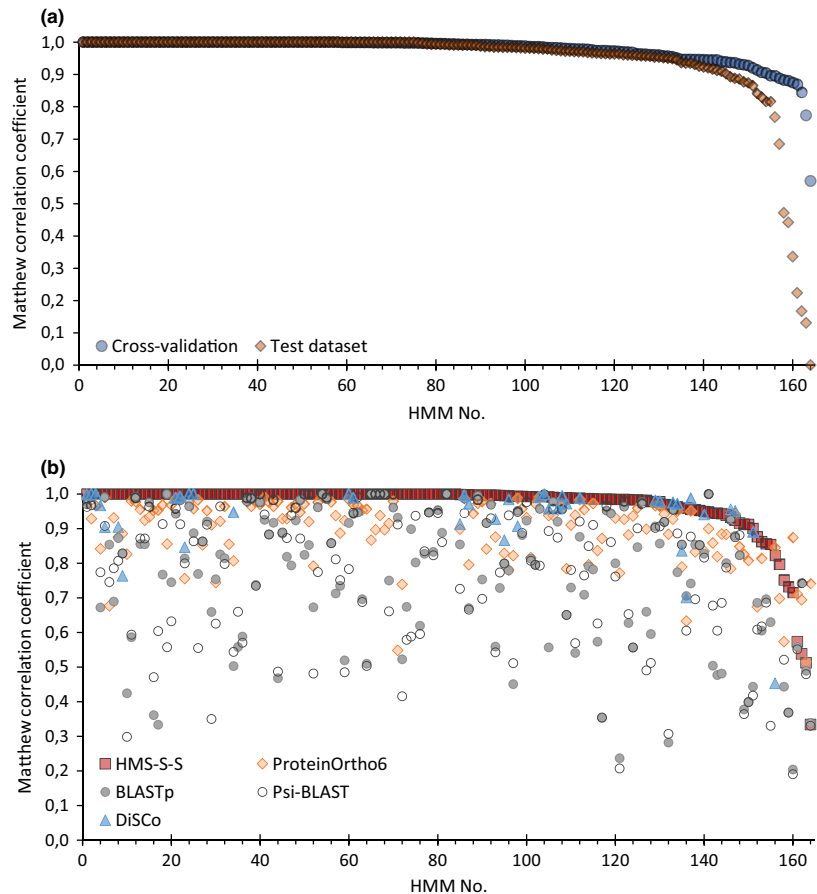
### 3.6 | HMM development: Sulphide and elemental sulphur oxidation

Many organisms contain genes for sulphide oxidation via sulphide:quinone oxidoreductase (SQR) and sulphur (or persulphide) dioxygenase (SDO). Based on the structure, six SQR classes can be distinguished (Marcia et al., 2010). For SDOs, which are all part of the beta-lactamase superfamily, three classes have been described (Liu, Xin, et al., 2014). Accordingly, nine HMMs for these classes were constructed. Additionally, two HMMs for periplasmic sulphide oxidation via flavocytochrome c sulphide dehydrogenase FccAB are present (Dolata et al., 1993). One HMM was trained for sulphur oxygenase/reductase (SOR), which converts elemental sulphur to sulphide and sulphite (Kletzin, 1992).

### 3.7 | HMM development: sHdR system

As an alternative to the reverse Dsr system, sulphur can also be oxidized via the sHdR system. This system consists of the sHdR complex sHdRC1B1AHC2B2, lipolate binding proteins LbpA1 and LbpA2 and a proposed novel pathway for LbpA maturation consisting of radical SAM domain proteins 1 and 2 (LipS1 and LipS2), a geranylgeranyl reductase-like FAD-containing NAD(P)-binding protein and a LplA-like lipolate: protein ligase (Cao et al., 2018; Ernst et al., 2021; Koch & Dahl, 2018). Genes for proteins sHdRC2B2 can be replaced by genes for electron-transferring flavoprotein EtfAB and two proteins with CCG signature motifs of noncubane [4Fe-4S] cluster-binding heterodisulphide reductase active sites (Cao et al., 2018). HMMs for several proteins rarely encoded in vicinity of the main sHdR complex proteins like sHdrl, sHdrt or the transcriptional regulator sHdR were also constructed. A total of 22 HMMs were included for the sHdR system.

**FIGURE 4** Validation of the 164 HMMs generated in this work. (a) Cross-validation of the HMMs in HMS-S-S. performance of the HMMs created in this work was assessed by cross-validation (blue dots) and on an independent test set (red diamonds). For each of the 164 HMMs in the cross-validation experiment the Matthew correlation coefficient is plotted. The HMMs are ranked by their MCC (x-axis). (b) Method comparison. Performance of the HMMs created in this study in classification of 164 different Sulphur metabolism-related proteins compared to the performance of other methods. All methods were tested on the same comparison data set. The HMMs from HMS-S-S are ranked by their MCC (x-axis). The MCC values (y-axis) resulting from the other procedures are plotted at the corresponding HMM



### 3.8 | HMM development: Sulphur transfer

Another component common to many different sulphur mobilization routes is a small highly conserved sulphur-binding protein referred to as TusA due to its firmly established function in the sulphur relay system involved in the formation of (c)mmn5s2U-modified nucleosides (2-thiouridine) (Ikeuchi et al., 2006). TusA-like proteins are implicated in different biosynthetic pathways, sulphur-based energy metabolism, detoxification as well as in regulatory circuits (Tanabe et al., 2019). There is one HMM for TusA. In addition, HMMs for TusB, -C, -D, and TusE, which resemble but are functionally distinct from DsrE, -F, -H, and DsrC, were set up to distinguish between these homologues. Sulphur transferase DsrE subclasses (Boughanemi et al., 2016; Liu, Stockdreher, et al., 2014) were covered by 7 HMMs.

### 3.9 | HMM development: Others

Models for QrcABCD, TmcABCD and HmcABCDEF complexes from sulphate reducers, which are homologous to and share composition with DsrMKJOP and Hdr complexes (Grein et al., 2013), were trained in order to guarantee clear separation. The same consideration led to the inclusion of models for mHdrABC from methanogenic

and sulphate-reducing organisms, as well as for FlxABCD (Ramos et al., 2015), MvhADG (Kaster, Goenrich, et al., 2011), McrABCDG (Kaster, Goenrich, et al., 2011), HdrDE and HdrF.

### 3.10 | Validation

In a first validation approach, the HMMs created for HMS-S-S were validated by cross-validation (Chicco, 2017; Refaeilzadeh et al., 2009). As the HMMs used in HMS-S-S were designed to specifically find equivalogues, the ability to correctly classify proteins was tested and measured. In addition to the left-out sequences, which served as true positives, true negative sequences from close homologues as well as unrelated sequences were added. The first goal of the cross-validation was to show robustness by demonstrating generalizability as well as equivalence purity of the training data. The second goal was to show the usability of the cutoffs calculated here. Performance was measured by the Matthews correlation coefficient (MCC) where 1 means perfect, 0 means random performance on assignment and -1 means total disagreement. Values of 1 or close to 1 are reached when both, TP and TN, are recognized with high probability. While the occurrence of either FP or FN reduces the score, the simultaneous occurrence of both errors

decreases the score to a disproportionately greater extent (Chicco & Jurman, 2020). Of the 164 HMMs covering proteins of dissimilatory sulphur metabolism and homologues thereof, 162 stayed above an MCC of 0.80 (Figure 4, Table S4). Only the HMMs for DsrE5 and qHdrC performed below 0.8 (Table S4). In the case of qHdrC, this was due to low precision even though no false negatives occurred. For DsrE5, FP and FN hits occurred (Table S4).

In a second validation approach, the reliability of HMS-S-S's classification and recognition was evaluated using an independent test data set consisting of (meta-)genomes. Homology between test and training sequences was taken into account by selecting the test sequences from a much broader and different phylogenetic range than the training set. Thus, the criterion was fulfilled that training and test data sets did not overlap (Refaeilzadeh et al., 2009). This excluded bias resulting from sequence similarity during the validation of generalizability. When HMS-S-S was validated on this data set, 155 of 164 HMMs scored >0.8 (Table S4). HMMs with a MCC <0.8 were SoxO, SoxF, PsrB/PhsB, SDOIII and the response regulator proteins TtrRS and sHdrr. These generated more false positive hits than true positives but no false negative hits. Therefore, they were considered to be sensitive but unprecise. An exception was oxidative-type Sat, where there were more true positives than false positives, with a MCC of 0.76 which resulted from interference with the reductive-type Sat HMM (Table S4).

### 3.11 | Method comparison

HMS-S-S was compared with BLASTp, Psi-BLAST, ProteinOrtho 6 and DiSCo in terms of its ability to recognize sequences and classify them into equivalogues. BLASTp is the most commonly used algorithm for protein sequence similarity searches, while Psi-BLAST is the HMMER equivalent of the BLAST algorithm. Since BLAST requires a query, the proteins in the reference list used for sequence annotation (Table S3) were taken as queries. ProteinOrtho6 is a tool to cluster protein sequences into co-orthologous groups by bidirectional best blast hit strategy and synteny. This tool has the advantage of having been created for the classification into orthologue groups of very large numbers of genomes (Lechner et al., 2014). DiSCo is a novel tool that also created HMMs for the detection and classification of sulphur metabolism proteins with a specific focus on the Dsr and Qmo-Apr-Sat pathways (Neukirchen & Sousa, 2021). Sequences for method comparison were taken from the comparison data set. All methods were tested on default settings with the same data set to ensure comparability. The results are summarized in Figure 5 and Table S5.

On the comparison dataset, HMS-S-S achieved an MCC higher than 0.8 for 156 protein types. The best hit approach with BLAST and Psi-BLAST showed an overall good recall but low precision. Thus, classification by BLASTp resulted in a MCC >0.8 for only 87 cases. Similar results were obtained with Psi-BLAST. Here, a MCC >0.8 was achieved in 83 cases. Performance of both methods suffered from false positive and false negative hits in the classification.

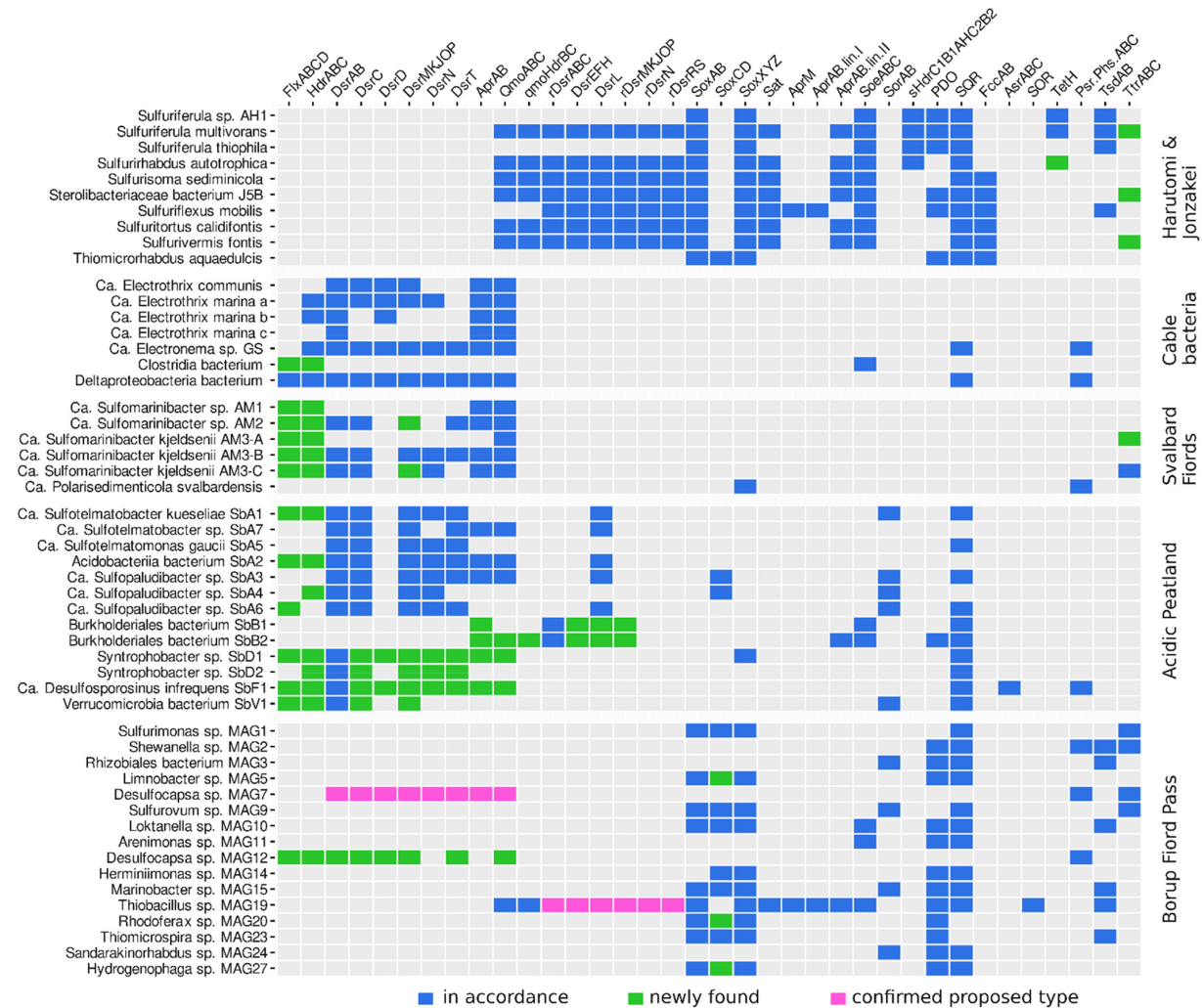
Only for TtrR and McrC classification by BLASTp or Psi-BLAST resulted in a higher MCC, and therefore better performance, than for HMS-S-S. Clustering into orthologous groups by ProteinOrtho6 resulted in 120 cases in an MCC >0.8. ProteinOrtho6 outperformed classification by (Psi)-BLAST(p) but performance suffered from nonequivalogous sequence assignments to the orthologous groups. Furthermore, the computation time was much longer than for all other methods compared. Most of the HMMs from HMS-S-S had a similar performance as the ones from DiSCo, with HMS-S-S performing slightly better. The HMMs for oxidative- and reductive-type DsrMOP and AprM were an exception, as DiSCo achieved up to 0.03 higher MCCs. Compared to DiSCo, HMS-S-S achieved significantly better results in the classification of QmoABC and HdrABC from methanogens and sulphate reducers. Here, performance values were higher by up to 0.1. In accordance with our results a balanced accuracy between 0.70 and 0.78 QmoABC HMMs from DiSCo was also measured during the original validation of DiSCo (Neukirchen & Sousa, 2021).

### 3.12 | Case studies

In the next step, functionality of the HMMs was tested against complete genomes derived from meta-assembled genomes. All ORFs from all genomes were predicted by Prodigal (Hyatt et al., 2010) and then searched for sulphur-related proteins by HMS-S-S. All results are combined in Figures 5 and S1.

#### 3.12.1 | Neutrophilic sulphur-oxidizing bacteria

A study on Beta- and Gammaproteobacteria isolated from Lake Harutori and Jozankei hot springs provided ten different genomes for characterized sulphur oxidizers (Watanabe et al., 2019) and thus provided an ideal basis for challenging HMS-S-S performance. In fact, the tool yielded predictions in full accordance with the original analysis of the genomes. SQR type I, an incomplete Sox system, lacking SoxCD, and SoeABC were found in all genomes. *Thiomicrothabodus aquaedulcis* was the only organism with additional genes for SQR III, V, VI and SoxCD completing the Sox system. FccB (catalytic flavoprotein subunit) was found in six assemblies, while the electron-transferring hemoprotein FccA was not encoded in three of the six assemblies. In the MAGs from the genera *Sulfuriferula* and *Sulfurirhabdus* neither FccA nor FccB were detected. Thiosulphate dehydrogenase TsdAB was present in four genomes (*Sulfuriferula multivorans*, *Sulfuriferula thiophila*, *Sulfuriferula* sp. AH1, *Sulfuriflexus mobilis*). Utilization of tetrathionate as electron donor during lithoautotrophic growth on nitrate has been observed for *Sulfuriferula multivorans* and *Sulfurivermis fontis* (Kojima et al., 2017; Watanabe et al., 2015). While tetrathionate degradation is initiated by tetrathionate hydrolase TetH in *Sulfuriferula multivorans*, a gene encoding this enzyme is not present in *Sulfurivermis fontis* according to our analysis as well as to



**FIGURE 5** Application of HMS-S-S to published metagenomes. Sulphur-metabolism-related proteins detected by HMS-S-S in recently described assemblies of neutrophilic sulphur oxidizers (Watanabe et al., 2019), cable bacteria (Kjeldsen et al., 2019; Pfeffer et al., 2012), novel acidobacteria (Flieder et al., 2021; Hausmann et al., 2018; Müller et al., 2019), and sulphur glacier meta-assembled genomes (Trivedi et al., 2020) are plotted in blue when they had also been detected in previously published work. Boxes are coloured pink when HMS-S-S confirmed a reductive/oxidative type for a specific protein that had previously been proposed but could by then not be definitely established. Green boxes indicate proteins newly found by HMS-S-S. With exception of *Candidatus Polarisedimenticola svalbardensis*, the figure shows only those MAGs that have a number of hits in the “in accordance” category greater than or equal to four. A full version is available as Figure S1

Watanabe et al. (2019). With the exception of *Sulfuriferula* sp. AH1, *Sulfuriferula thiophila* and *Thiomicrothabodus aquaedulcis*, genes for the oxidative-type rDsr system, oxidative-type AprAB and Sat, reductive-type QmoAB and the Qmo-associated qHdrBC complex were found in all MAGs by HMS-S-S, consistent with the previous report. In full agreement with earlier results, *Sulfuriflexus mobilis* encodes AprAB from lineage I, AprM and QmoABqHdrBC, while all AprAB-APS reductases from the other organisms belonged to lineage II. Complete *shdr* operons were found in the three genomes from *Sulfuriferula* species and in *Sulfurirhabdus autotrophica*. In addition to the original analysis, tetrathionate reductase *ttrBCA*

operons were found in *Sulfuriferula multivorans*, *Sulfurivermis fontis* and *Sterolibacteriaceae bacterium* J5B. However, growth on tetrathionate as electron acceptor has not been tested for these species (Kojima et al., 2017; Watanabe et al., 2015).

### 3.12.2 | Acidobacteria from sulphidic fjord sediments

Bacteria of the phylum Acidobacteriota are widespread and abundant in many marine sediments. A recent study on Norwegian

fjord sediments allowed first insights into the metabolic potential of uncultured Acidobacteria lineages and provided evidence that they play important roles in sedimentary biogeochemical element cycles, especially in the sulphur cycle (Flieder et al., 2021). A total of six MAGs from new Acidobacteria were analysed in this study: Two belonged to *Candidatus* Sulfuromarinibacter sp. (AM1 & AM2), three were assigned to *Candidatus* Sulfuromarinibacter kjeldsenii (AM3A-C) and a single MAG was assigned to *Polarisedimenticola svaldbardensis* (AM4). In accordance with the original results (Flieder et al., 2021), HMS-S-S detected the reductive-type Dsr system in AM2 and AM3, while Psr/PhsABC was found in AM4. In accordance with the proposed metabolic potential, HMS-S-S identified previously undetected reductive-type DsrMK and DsrMKJP in the AM2 and AM3 assemblies, respectively. A complete set of reductive-type AprAB and QmoABC is encoded in a single operon in all five studied assemblies. Beyond the original work, HMS-S-S detected *hdrCBA-flxCBA* genes encoding NADH dehydrogenase/heterodisulphide reductase involved in ethanol metabolism (Ramos et al., 2015) in all analysed acidobacterial MAGs. With exception of AM2, HdrABC complexes were encoded adjacent to FlxABC. An *hdrCBA-flxCBA* operon structure has not only been reported for a number of sulphate-reducing Deltaproteobacteria but also for an Acidobacterium (*Holophaga foetida*) incapable of sulfate reduction (Ramos et al., 2015). In *Candidatus* Sulfuromarinibacter kjeldsenii MAG AM3-A, HMS-S-S found genes for tetrathionate reductase subunits TtrA and TtrC, which were previously only reported for assembly AM3-B of the same species (Flieder et al., 2021). In summary, HMS-S-S detected all previously reported genes and additional hits coincided with the proposed metabolic potential of these Acidobacteria (Flieder et al., 2021; Hausmann et al., 2018).

### 3.12.3 | Peatland metagenome

Sulphur-cycling microorganisms have a strong impact on organic matter decomposition in wetlands. Hausmann et al. (2018) applied a functional metagenomics approach to an acidic peatland and recovered draft genomes of seven novel Acidobacteria species with the potential for dissimilatory sulphur metabolism. Corroborating this work, HMS-S-S identified sulphite reduction via a complete reductive-type Dsr system and DsrL in peatland Acidobacteria assemblies SbA1–SbA7. Sulphate activation and reduction via Sat, AprAB and QmoABC was found in assemblies SbA2, 3 and 7, while this pathway was not detected in the other four acidobacterial assemblies. In SbA1 and SbA2, genes for the HdrABC complex and *flxABC* genes were present. In accordance with Hausmann et al. (2018), HMS-S-S detected SorAB in SbA1, SbA3, SbA4 and SbA6. In addition, SbA3 and SbA4 gave positive results with SoxCD. The function of these enzymes in the probable sulphate/sulphite reducers remains unclear. HMS-S-S identified the potential for sulphite reduction via the reductive-type Dsr

system in four additional assemblies, SbD1, SbD2, SbV1 and SbF1. These assemblies are from *Syntrophobacter*, Verrucomicrobia, and *Desulfosporosinus* species, all of which were predicted to perform sulphate reduction (Anantharaman et al., 2018). A sulphur-oxidizing metabolism is likely for two Burkholderiales bacteria assemblies (SbB1 and SbB2) due to the presence of an oxidative-type rDsr system and sulphite oxidation via SoeABC as well as via AprAB lineage II, QmoABqHdrBC and Sat.

### 3.12.4 | Cable bacteria

Another set of MAGs originated from projects with a special focus on deltaproteobacterial cable bacteria. A total of nine MAGs originated from a sulphidic sediment sampled in Aarhus Bay, Denmark (Kjeldsen et al., 2019; Pfeffer et al., 2012) and from an enrichment culture from a former coal gasification site in Gliwice, Poland (Müller et al., 2019). Cable bacteria form long filaments that couple sulphur oxidation at one end of the filament with oxygen reduction at the other end, transporting the electrons over several centimetres (Kjeldsen et al., 2019). The energy metabolism of the filament in the anoxic zones might also be constituted by sulphur disproportionation (Müller et al., 2019).

The first nearly complete genome of a groundwater cable bacterium was reconstructed together with three MAGs from other species from Gliwice (Müller et al., 2019). In this assembly all genes for sulphite reduction via the reductive-type Dsr system including *dsrD* and *dsrT* were found by HMS-S-S. Genes encoding the enzymes for sulphate activation and reduction (AprAB, QmoABC and Sat) were also present. Furthermore, genes for HdrA and HdrC, as well as for polysulphide/thiosulphate reductase PsrA/PhsA and SQR were detected. In the other three noncable bacterial genomes, dissimilatory sulphur metabolism-related genes were not detected by HMS-S-S but a *Clostridium* assembly encoded a HdrABC-FlxABC gene cluster. In all four MAGs from Aarhus Bay assigned to *Candidatus* Electrothrix and in the one MAG assigned to *Candidatus* Electronema, reductive-type DsrAB was found. Due to the highest completeness achieved for *Candidatus* Electronema aureum GS, this MAG provided further valuable insights into the performance of HMS-S-S. Reductive-type DsrCMKDNT were detected while DsrJOP were missing. Genes for the complete reductive-type Apr, Qmo, Sat pathway were also found. In addition, HdrAC, a polysulphide/thiosulphate reductase PsrA/PhsA candidate and sulphide-oxidizing SQR were correctly predicted for *Candidatus* Electronema sp. GS by HMS-S-S (Kjeldsen et al., 2019). Searches in the *Candidatus* Electrothrix assemblies provided similar results, but these often lacked single genes due to low completeness (21 to 64%). Other systems like oxidative-type Dsr, sHdr or Sox were neither predicted by HMS-S-S nor by the original description (Kjeldsen et al., 2019). In summary, these results are in accordance with the proposal that cable bacterial metabolism includes sulphur compound oxidation via SQR and a reductive-type Dsr system operated in reverse (Kjeldsen et al., 2019).

## 3.12.5 | Sulphur glaciers

In 2020, Trivedi et al. provided valuable information on the so far understudied biological sulphur cycling in polar, low-temperature ecosystems. The authors investigated 31 MAGs originating from a sulphur-rich glacial environment in the Canadian High Arctic (Trivedi et al., 2020). Using HMS-S-S, a complete Sox system for thiosulphate oxidation in the periplasm, including SoxCD, was found in seven of these MAGs (MAGs 1, 5, 9, 10, 15, 20, 27), while Trivedi et al. (2020) found soxXYZBC in only four MAGs. In MAGs 14, 19, 23 and 29, various sox genes are missing, possibly due to incomplete MAGs. In MAG19, which was assigned to *Thiobacillus* sp., three operons related to sulphite oxidation (*aprMBA* [lineage I], *qmoAB-qhdrBC*, *sat-aprB* lineage II) were identified by HMS-S-S. The same three operons were reported for the well-studied sulphur-oxidizing bacterium *Thiobacillus denitrificans* (Meyer & Kuever, 2007a). Furthermore, sulphur oxygenase reductase, and a complete set of oxidative-type Dsr genes, were found in this MAG. Thus, we confirmed the hypothesis that MAG 19 operates a oxidative-type Dsr system (Trivedi et al., 2020). Thiosulphate oxidation to tetrathionate via TsdBA was detected in MAG 2, 3, 10, 15, 19 and 23. Except for MAG 2, *tsdA* and *tsdB* were encoded next to each other. MAGs 3, 9 and 14 showed potential for sulphite oxidation in the periplasm via SorAB, while MAGs 10, 11 and 19 encode genes for sulphite oxidation in the cytoplasm via SoeABC. SQR was present in 24 of 31 MAGs, with SDO in 17 of those MAGs. All these findings were in full accordance with previous results (Trivedi et al., 2020). Two MAGs were most probably derived from sulphate-reducing organisms. In MAG 7 and MAG 12, which are both assigned to *Desulfocapsa*, all *dsr* genes for dissimilatory sulphite reduction were identified including *dsrABD*, *dsrC*, *dsrMKJOP*, *dsrN* and *dsrT*. AprAB, QmoABC and Sat were all present in MAG7, while MAG12 lacked AprAB and QmoA, albeit it is not completely excluded that they are missing in the assembly because the respective contig ends with *qmoBC*. MAG12 was previously not mentioned to have sulphate reduction potential (Trivedi et al., 2020).

## 4 | DISCUSSION

Here, we developed HMS-S-S, an equivalogous HMM-based searching tool for sulphur-related proteins. A special feature of our tool is the provision of information about operon structures to support equivalence prediction and possibilities for further analysis of the synteny offered. Functionality was demonstrated by validation, comparison with other tools and by analysis of recently published meta-assembled genomes.

The quality of 164 novel HMMs was ensured by literature-based selection of genomes derived from the RefSeq and GenBank. Training and test sequences were annotated according to strict criteria and were used separately from each other. For the resulting HMMs, cutoffs were calculated and validated by cross-validation and on an independent test set using MCC as performance metric (Chicco, 2017). The functional capability of the HMMs in searching,

recognizing and correctly annotating could thus be confirmed. Even though the cutoff values are validated, it is likely that they will need to be adjusted for newly discovered phyla, as is already being done in other studies (Anantharaman et al., 2018; Jaffe et al., 2020).

HMS-S-S does not only cover marker proteins, but also less commonly known sulphur-related proteins missed by common annotation pipelines like DsrL (Löffler, Feldhues, et al., 2020; Löffler, Wallerang, et al., 2020), sHdrH (Koch & Dahl, 2018), SoxEFGH (Friedrich et al., 2001) or LbpA (Cao et al., 2018). Here, 122 HMMs for proteins were created which are not covered by DiSCo (Neukirchen & Sousa, 2021). Furthermore, only 14 of our 164 HMMs have an equivalent in PFAM (El-Gebali et al., 2019) or TIGRFAMs. Even though curated PFAM can be used in functional annotation, this is only true for roughly half of all PFAM families (Li et al., 2021). A common problem arising from this in functional annotation is a lack of specificity for the proposed reaction. PFAM family Pf07992 may serve as a prominent example as it includes not only the FAD-containing nonelectron bifurcating sulphur-oxidizing sHdrA (Ernst et al., 2021) and electron-bifurcating mHdrA from methanogens (Wagner et al., 2017) but also the flavoprotein sulphide quinone oxidoreductase. In contrast, HMS-S-S clearly distinguishes not only between the flavoprotein subunits of the sHdr, mHdr, and Qmo complexes but also between all their other subunits (Appel et al., 2021). The tool differentiates between the six types of SQRs and FccA (Sousa et al., 2018) as well as between oxidative and reductive types of the Dsr, Apr and Sat. Polysulphide reductase, thiosulphate reductase and sulphur reductase are an exception since an equivalogous HMM could not be constructed. The subunits of these proteins are neither distinguishable by primary sequence or relative gene order nor by phylogenetic tree construction (Boughanemi et al., 2020; Duval et al., 2008). Validation proved that related complexes like Ttr, Soe, antimon reductase or arsenite oxidase do not interfere with the Psr/Phs/Sre HMM in accordance with phylogenetic analyses (Boughanemi et al., 2020), Ttr and Soe can therefore be distinguished from Psr/Phs/Sre. In summary, we created a set of reliable HMM and substantially extended current resources in protein annotation.

HMS-S-S was compared with Psi-BLAST, BLASTp and ProteinOrtho6 in terms of precision and recall in functional annotation (Figure 5). Comparison methods were chosen as they had also been used in the validation of DiSCo (Neukirchen & Sousa, 2021). As reported by previous studies the profiled HMM-based approach in HMS-S-S was more precise than the unidirectional best hit strategy of BLAST based methods (Kaushik et al., 2016; Neukirchen & Sousa, 2021; Skewes-Cox et al., 2014). In contrast to BLAST, ProteinOrtho6 follows a bidirectional best hit strategy paired with synteny analysis to cluster sequences into orthologous groups. MCC scores from this clustering were lower than those of HMS-S-S, as the orthologous groups were not completely conserved in function. Orthologues are the result of a speciation event and form monophyletic clades in a phylogenetic tree. They are therefore considered to catalyse the same reactions which forms the basis for functional annotation (Gabaldón & Koonin, 2013). Recently this view was challenged by a comparison of experimentally shown function and function predicted based on orthology (Stambouliau



et al., 2020). Furthermore, orthologous groups include orthologues, co-orthologues and paralogues and do not necessarily imply functional equality (Gabaldón & Koonin, 2013), which is in accordance with the results observed here. DiSCo, the only specialized tool for Dsr-dependent sulphur metabolism, operates on a similar HMM based approach as HMS-S-S, but does not include synteny analysis (Neukirchen & Sousa, 2021). The performance of HMS-S-S and DiSCo for oxidative-type and reductive-type Dsr, AprAB and Sat was shown to be comparably good, but HMS-S-S performs better for QmoABC and HdrABC. Since DiSCo itself is a validated tool, this comparison also shows the reliability of our HMMs.

In a final step, the ability of HMS-S-S for annotation of sulphur metabolism proteins and corresponding genes was validated on recently described assemblies and proved the consistency of the HMS-S-S results with the described features of the assemblies (Figure 5).

## 5 | CONCLUSIONS

In summary, HMS-S-S is a curated comprehensive HMM based tool for annotation and synteny analysis of the sulphur metabolism. Using equivalogs for HMM training resulted in 164 HMMs with high precision and recall. This also closes a gap in the coverage of sulphur metabolism prediction by HMMs. Demonstrated validation showed a similar or higher performance than that of other currently available tools. HMS-S-S also encompasses the possibility of studying synteny, also in the combination with other HMMs from public databases or custom models and can therefore be extended to the needs of the user.

## AUTHOR CONTRIBUTIONS

Tomohisa Sebastian Tanabe and Christiane Dahl conceived the study. Tomohisa Sebastian Tanabe developed and implemented the method and performed the analyses. Tomohisa Sebastian Tanabe analysed and interpreted the data. Both authors wrote and approved the final version of the manuscript.

## ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (grant no. 351/13-1 to CD). TST received a scholarship from the Studienstiftung des Deutschen Volkes.

## CONFLICT OF INTEREST

The authors declare that they have no competing interests.

## DATA AVAILABILITY STATEMENT

HMS-S-S database program files are available at <https://github.com/TSTanabe/HMSSS>

## ORCID

Tomohisa Sebastian Tanabe  <https://orcid.org/0000-0003-2154-7980>

[org/0000-0003-2154-7980](https://orcid.org/0000-0003-2154-7980)

Christiane Dahl  <https://orcid.org/0000-0001-8288-7546>

## REFERENCES

- Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., Thomas, B. C., Singh, A., Wilkins, M. J., Karaoz, U., Brodie, E. L., Williams, K. H., Hubbard, S. S., & Banfield, J. F. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications*, 7, 13219. <https://doi.org/10.1038/ncomms13219>
- Anantharaman, K., Hausmann, B., Jungbluth, S. P., Kantor, R. S., Lavy, A., Warren, L. A., Rappé, M. S., Pester, M., Loy, A., Thomas, B. C., & Banfield, J. F. (2018). Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME Journal*, 12, 1715–1728. <https://doi.org/10.1038/s41396-018-0078-0>
- Appel, L., Willistein, M., Dahl, C., Ermler, U., & Boll, M. (2021). Functional diversity of prokaryotic HdrA(BC) modules: Role in flavin-based electron bifurcation processes and beyond. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1862, 148379. <https://doi.org/10.1016/j.bbabi.2021.148379>
- Appia-Ayme, C., & Berks, B. C. (2002). SoxV, an orthologue of the CcdA disulfide transporter, is involved in thiosulfate oxidation in *Rhodovulum sulfidophilum* and reduces the periplasmic thioredoxin SoxW. *Biochemical and Biophysical Research Communications*, 296(3), 737–741. [https://doi.org/10.1016/S0006-291x\(02\)00936-1](https://doi.org/10.1016/S0006-291x(02)00936-1)
- Bardischewsky, F., Quentmeier, A., & Friedrich, C. G. (2006). The flavoprotein SoxF functions in chemotrophic thiosulfate oxidation of *Paracoccus pantotrophus* in vivo and in vitro. *FEMS Microbiology Letters*, 258(1), 121–126. <https://doi.org/10.1111/j.1574-6968.2006.00210.x>
- Berben, T., Overmars, L., Sorokin, D. Y., & Muyzer, G. (2019). Diversity and distribution of sulfur oxidation-related genes in *Thioalkalivibrio*, a genus of chemolithoautotrophic and haloalkaliphilic sulfur-oxidizing bacteria. *Frontiers in Microbiology*, 10, 160. <https://doi.org/10.3389/fmicb.2019.00160>
- Boughanemi, S., Infossi, P., Giudici-Ortoni, M. T., Schoepp-Cothenet, B., & Guiral, M. (2020). Sulfite oxidation by the quinone-reducing molybdenum sulfite dehydrogenase SoeABC from the bacterium *Aquifex aeolicus*. *Biochimica et Biophysica Acta-Bioenergetics*, 1861(11), 148279. <https://doi.org/10.1016/j.bbabi.2020.148279>
- Boughanemi, S., Lyonnet, J., Infossi, P., Bauzan, M., Kosta, A., Lignon, S., Giudici-Ortoni, M. T., & Guiral, M. (2016). Microbial oxidative sulfur metabolism: Biochemical evidence of the membrane-bound heterodisulfide reductase-like complex of the bacterium *Aquifex aeolicus*. *FEMS Microbiology Letters*, 363(15), fnw156. <https://doi.org/10.1093/femsle/fnw156>
- Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. In *2010 International Conference on Pattern Recognition* (pp. 3121–3124). IEEE. <https://doi.org/10.1109/icpr.2010.764>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 421. <https://doi.org/10.1186/1471-2105-10-421>
- Canfield, D. E., & Raiswell, R. (1999). The evolution of the sulfur cycle. *American Journal of Science*, 299(7–9), 697–723. <https://doi.org/10.2475/ajs.299.7-9.697>
- Canfield, D. E., Rosing, M. T., & Bjerrum, C. (2006). Early anaerobic metabolisms. *Philosophical Transactions of the Royal Society London B Biological Sciences*, 361(1474), 1819–1834; discussion 1835–1816. <https://doi.org/10.1098/rstb.2006.1906>
- Cao, X., Koch, T., Steffens, L., Finkensieper, J., Zigann, R., Cronan, J. E., & Dahl, C. (2018). Lipoate-binding proteins and specific lipoate-protein ligases in microbial sulfur oxidation reveal an atypical role for an old cofactor. *eLife*, 7, e37439. <https://doi.org/10.7554/eLife.37439>
- Chernyh, N. A., Neukirchen, S., Frolov, E. N., Sousa, F. L., Miroshnichenko, M. L., Merkel, A. Y., Pimenov, N. V., Sorokin, D.

- Y., Ciordia, S., Mena, M. C., Ferrer, M., Golyshin, P. N., Lebedinsky, A. V., Pereira, I. A. C., & Bonch-Osmolovskaya, E. A. (2020). Dissimilatory sulfate reduction in the archaeon '*Candidatus Vulcanisaeta moutnovskia*' sheds light on the evolution of sulfur metabolism. *Nature Microbiology*, 5, 1428–1438. <https://doi.org/10.1038/s41564-020-0776-z>
- Chico, D. (2017). Ten quick tips for machine learning in computational biology. *BioData Mining*, 10, 35. <https://doi.org/10.1186/s13040-017-0155-3>
- Chico, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6. <https://doi.org/10.1186/s12864-019-6413-7>
- Consortium, T. U. (2017). UniProt: The universal protein knowledge-base. *Nucleic Acids Research*, 45(D1), D158–D169. <https://doi.org/10.1093/nar/gkw1099>
- Dahl, C., Franz, B., Hensen, D., Kesselheim, A., & Zigann, R. (2013). Sulfite oxidation in the purple sulfur bacterium *Allochromatium vinosum*: Identification of SoeABC as a major player and relevance of SoxYZ in the process. *Microbiology*, 159, 2626–2638. <https://doi.org/10.1099/mic.0.071019-0>
- Dahl, C., Schulte, A., Stockdreher, Y., Hong, C., Grimm, F., Sander, J., Kim, R., Kim, S. H., & Shin, D. H. (2008). Structural and molecular genetic insight into a wide-spread bacterial sulfur oxidation pathway. *Journal of Molecular Biology*, 384, 1287–1300. <https://doi.org/10.1016/j.jmb.2008.10.016>
- Denkman, K., Grein, F., Zigann, R., Siemen, A., Bergmann, J., van Helmont, S., Nicolai, A., Pereira, I. A. C., & Dahl, C. (2012). Thiosulfate dehydrogenase: A wide-spread unusual acidophilic c-type cytochrome. *Environmental Microbiology*, 14, 2673–2688. <https://doi.org/10.1111/j.1462-2920.2012.02820.x>
- Dolata, M. M., van Beumen, J. J., Ambler, R. P., Meyer, T. E., & Cusanovich, M. A. (1993). Nucleotide sequence of the heme subunit of flavocytochrome c from the purple phototrophic bacterium, *Chromatium vinosum*. A 2.6-kilobase pair DNA fragment contains two multiheme cytochromes, a flavoprotein, and a homolog of human ankyrin. *Journal of Biological Chemistry*, 268(19), 14426–14431.
- Duarte, A. G., Barbosa, A. C. C., Ferreira, D., Manteigas, G., Domingos, R. M., & Pereira, I. A. C. (2021). Redox loops in anaerobic respiration—The role of the widespread NrfD protein family and associated dimeric redox module. *Biochimica et Biophysica Acta Bioenergetics*, 1862(7), 148416. <https://doi.org/10.1016/j.bbabi.2021.148416>
- Duval, S., Ducluzeau, A. L., Nitschke, W., & Schoepp-Cothenet, B. (2008). Enzyme phylogenies as markers for the oxidation state of the environment: The case of respiratory arsenate reductase and related enzymes. *BMC Evolutionary Biology*, 8, 206. <https://doi.org/10.1186/1471-2148-8-206>
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1), D427–D432. <https://doi.org/10.1093/nar/gky995>
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1), 238. <https://doi.org/10.1186/s13059-019-1832-y>
- Ernst, C., Kayashta, K., Koch, T., Venceslau, S. S., Pereira, I. A. C., Demmer, U., Ermler, U., & Dahl, C. (2021). Structural and spectroscopic characterization of a HdrA-like subunit from *Hyphomicrobium denitrificans*. *FEBS Journal*, 288, 1664–1678. <https://doi.org/10.1111/febs.15505>
- Ferreira, D., Barbosa, A. C. C., Oliveira, G. P., Catarino, T., Venceslau, S. S., & Pereira, I. A. C. (2022). The DsrD functional marker protein is an allosteric activator of the DsrAB dissimilatory sulfite reductase. *Proceedings of the National Academy of Sciences of the United States of America*, 119(4), e2118880119. <https://doi.org/10.1073/pnas.2118880119>
- Flieder, M., Buongiorno, J., Herbold, C. W., Hausmann, B., Rattei, T., Lloyd, K. G., Loy, A., & Wasmund, K. (2021). Novel taxa of Acidobacteriota implicated in seafloor sulfur cycling. *ISME Journal*, 15, 3159–3180. <https://doi.org/10.1038/s41396-021-00992-0>
- Forman, G., & Scholz, M. (2010). Apples-to-apples in cross-validation studies. *ACM SIGKDD Explorations Newsletter*, 12(1), 49–57. <https://doi.org/10.1145/1882471.1882479>
- Friedrich, C. G., Bardischewsky, F., Rother, D., Quentmeier, A., & Fischer, J. (2005). Prokaryotic sulfur oxidation. *Current Opinion in Microbiology*, 8(3), 253–259. <https://doi.org/10.1016/j.mib.2005.04.005>
- Friedrich, C. G., Quentmeier, A., Bardischewsky, F., Rother, D., Orawski, G., Hellwig, P., & Fischer, J. (2008). Redox control of chemotrophic sulfur oxidation of *Paracoccus pantotrophus*. In C. Dahl & C. G. Friedrich (Eds.), *Microbial sulfur metabolism* (pp. 139–150). Springer.
- Friedrich, C. G., Rother, D., Bardischewsky, F., Quentmeier, A., & Fischer, J. (2001). Oxidation of reduced inorganic sulfur compounds by bacteria: Emergence of a common mechanism? *Applied and Environmental Microbiology*, 67(7), 2873–2882. <https://doi.org/10.1128/AEM.67.7.2873-2882.2001>
- Gabaldón, T., & Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*, 14(5), 360–366. <https://doi.org/10.1038/nrg3456>
- Grein, F., Ramos, A. R., Venceslau, S. S., & Pereira, I. A. C. (2013). Unifying concepts in anaerobic respiration: Insights from dissimilatory sulfur metabolism. *Biochimica et Biophysica Acta*, 1827(2), 145–160. <https://doi.org/10.1016/j.bbabi.2012.09.001>
- Grimm, F., Cort, J. R., & Dahl, C. (2010). DsrR, a novel IscA-like protein lacking iron- and FeS-binding function involved in the regulation of sulfur oxidation in *Allochromatium vinosum*. *Journal of Bacteriology*, 192(6), 1652–1661. <https://doi.org/10.1128/JB.101269-09>
- Grimm, F., Franz, B., & Dahl, C. (2011). Regulation of dissimilatory sulfur oxidation in the purple sulfur bacterium *Allochromatium vinosum*. *Frontiers in Microbiology*, 2, 51. <https://doi.org/10.3389/fmicb.2011.00051>
- Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwatz, M., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Zheng, C., Thibaud-Nissen, F., Geer, L. Y., ... Pruitt, K. D. (2018). RefSeq: An update on prokaryotic genome annotation and curation. *Nucleic Acids Research*, 46(D1), D851–D860. <https://doi.org/10.1093/nar/gkx1068>
- Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., & Beck, E. (2013). TIGRFAMs and genome properties in 2013. *Nucleic Acids Research*, 41(Database issue), D387–D395. <https://doi.org/10.1093/nar/gks1234>
- Hausmann, B., Pelikan, C., Herbold, C. W., Kostlbacher, S., Albertsen, M., Eichorst, S. A., Glavina Del Rio, T., Huemer, M., Nielsen, P. H., Rattei, T., Stingl, U., Tringe, S. G., Trojan, D., Wentrup, C., Woebken, D., Pester, M., & Loy, A. (2018). Peatland Acidobacteria with a dissimilatory sulfur metabolism. *ISME Journal*, 12, 1729–1742. <https://doi.org/10.1038/s41396-018-0077-1>
- Hensel, M., Hinsley, A. P., Nikolaus, T., Sawers, G., & Berks, B. C. (1999). The genetic basis of tetrathionate respiration in *Salmonella typhimurium*. *Molecular Microbiology*, 32(2), 275–287. <https://doi.org/10.1046/j.1365-2958.1999.01345.x>
- Holkenbrink, C., Ocon Barbas, S., Mellerup, A., & Frigaard, N. U. (2011). Sulfur globule oxidation in green sulfur bacteria is dependent on the dissimilatory sulfite reductase system. *Microbiology*, 157, 1229–1239. <https://doi.org/10.1099/mic.0.044669-0>
- Huang, C. J., & Barrett, E. L. (1991). Sequence analysis and expression of the *Salmonella typhimurium asr* operon encoding production of hydrogen sulfide from sulfite. *Journal of Bacteriology*, 173(4), 1544–1553. <https://doi.org/10.1128/jb.173.4.1544-1553.1991>

- Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11, 119. <https://doi.org/10.1186/1471-2105-11-119>
- Ikeuchi, Y., Shigi, N., Kato, J., Nishimura, A., & Suzuki, T. (2006). Mechanistic insights into sulfur relay by multiple sulfur mediators involved in thiouridine biosynthesis at tRNA wobble positions. *Molecular Cell*, 21(1), 97–108. <https://doi.org/10.1016/j.molcel.2005.11.001>
- Jaffe, A. L., Castelle, C. J., Matheus Carnevali, P. B., Gribaldo, S., & Banfield, J. F. (2020). The rise of diversity in metabolic platforms across the Candidate Phyla Radiation. *BMC Biology*, 18(1), 69. <https://doi.org/10.1186/s12915-020-00804-5>
- Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing imbalanced data recommendations for the use of performance metrics. In *International Conference on Affective Computing and Intelligent Interaction and Workshops* (Vol. 2013, pp. 245–251). IEEE. <https://doi.org/10.1109/ACII.2013.47>
- Kanao, T., Kamimura, K., & Sugio, T. (2007). Identification of a gene encoding a tetrathionate hydrolase in *Acidithiobacillus ferrooxidans*. *Journal of Biotechnology*, 132(1), 16–22. <https://doi.org/10.1016/j.jbiotec.2007.08.030>
- Kappler, U., Bennett, B., Rethmeier, J., Schwarz, G., Deutzmann, R., McEwan, A. G., & Dahl, C. (2000). Sulfite:cytochrome c oxidoreductase from *Thiobacillus novellus* - purification, characterization, and molecular biology of a heterodimeric member of the sulfite oxidase family. *Journal of Biological Chemistry*, 275(18), 13202–13212. <https://doi.org/10.1074/jbc.275.18.13202>
- Kaster, A. K., Goenrich, M., Seedorf, H., Liesegang, H., Wollherr, A., Gottschalk, G., & Thauer, R. K. (2011). More than 200 genes required for methane formation from H<sub>2</sub> and CO<sub>2</sub> and energy conservation are present in *Methanothermobacter marburgensis* and *Methanothermobacter thermoautotrophicus*. *Archaea*, 2011, 973848. <https://doi.org/10.1155/2011/973848>
- Kaster, A. K., Moll, J., Parey, K., & Thauer, R. K. (2011). Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), 2981–2986. <https://doi.org/10.1073/pnas.1016761108>
- Katayama, Y., Hashimoto, K., Nakayama, H., Mino, H., Nojiri, M., Ono, T. A., Nyunoya, H., Yohda, M., Takio, K., & Odaka, M. (2006). Thiocyanate hydrolase is a cobalt-containing metalloenzyme with a cysteine-sulfinic acid ligand. *Journal of the American Chemical Society*, 128(3), 728–729. <https://doi.org/10.1021/ja057010q>
- Kaushik, S., Nair, A. G., Mutt, E., Subramanian, H. P., & Sowdhamini, R. (2016). Rapid and enhanced remote homology detection by cascading hidden Markov model searches in sequence space. *Bioinformatics*, 32(3), 338–344. <https://doi.org/10.1093/bioinformatics/btv538>
- Kern, M., Klotz, M. G., & Simon, J. (2011). The *Wolinella succinogenes* mcc gene cluster encodes an unconventional respiratory sulphite reduction system. *Molecular Microbiology*, 82(6), 1515–1530. <https://doi.org/10.1111/j.1365-2958.2011.07906.x>
- Kjeldsen, K. U., Schreiber, L., Thorup, C. A., Boesen, T., Bjerg, J. T., Yang, T., Dueholm, M. S., Larsen, S., Risgaard-Petersen, N., Nierychlo, M., Schmid, M., Boggild, A., van de Vossenbergh, J., Geelhoed, J. S., Meysman, F. J. R., Wagner, M., Nielsen, P. H., Nielsen, L. P., & Schramm, A. (2019). On the evolution and physiology of cable bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 116(38), 19116–19125. <https://doi.org/10.1073/pnas.1903514116>
- Kletzin, A. (1992). Molecular characterization of the *sor* gene, which encodes the sulfur oxygenase/reductase of the thermoacidophilic archaeum *Desulfurolobus ambivalens*. *Journal of Bacteriology*, 174(18), 5854–5859. <https://doi.org/10.1128/jb.174.18.5854-5859.1992>
- Koch, T., & Dahl, C. (2018). A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds. *ISME Journal*, 12(10), 2479–2491. <https://doi.org/10.1038/s41396-018-0209-7>
- Kojima, H., Watanabe, M., & Fukui, M. (2017). *Sulfurivermis fontis* gen. nov., sp. nov., a sulfur-oxidizing autotroph, and proposal of *Thioprofundaceae* fam. *International Journal of Systematic and Evolutionary Microbiology*, 67(9), 3458–3461. <https://doi.org/10.1099/ijsem.0.002137>
- Lahiri, C., Mandal, S., Ghosh, W., Dam, B., & Roy, P. (2006). A novel gene cluster *soxSRT* is essential for the chemolithotrophic oxidation of thiosulfate and tetrathionate by *Pseudaminobacter salicylatoxidans* KCT001. *Current Microbiology*, 52(4), 267–273. <https://doi.org/10.1007/s00284-005-0176-x>
- Lechner, M., Hernandez-Rosales, M., Doerr, D., Wieseke, N., Thevenin, A., Stoye, J., Hartmann, R. K., Prohaska, S. J., & Stadler, P. F. (2014). Orthology detection combining clustering and synteny for very large datasets. *PLoS One*, 9(8), e105015. <https://doi.org/10.1371/journal.pone.0105015>
- Leimkühler, S., & Iobbi-Nivol, C. (2016). Bacterial molybdoenzymes: Old enzymes for new purposes. *FEMS Microbiology Reviews*, 40(1), 1–18. <https://doi.org/10.1093/femsre/fuv043>
- Li, W., O'Neill, K. R., Haft, D. H., DiCuccio, M., Chetvernin, V., Badretin, A., Coulouris, G., Chitsaz, F., Derbyshire, M. K., Durkin, A. S., Gonzales, N. R., Gwartz, M., Lanczycki, C. J., Song, J. S., Thanki, N., Wang, J., Yamashita, R. A., Yang, M., Zheng, C., ... Thibaud-Nissen, F. (2021). RefSeq: Expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Research*, 49(D1), D1020–D1028. <https://doi.org/10.1093/nar/gkaa1105>
- Liu, H., Xin, Y., & Xun, L. (2014). Distribution, diversity, and activities of sulfur dioxygenases in heterotrophic bacteria. *Applied and Environmental Microbiology*, 80(5), 1799–1806. <https://doi.org/10.1128/AEM.03281-13>
- Liu, L. J., Stockdreher, Y., Koch, T., Sun, S. T., Fan, Z., Josten, M., Sahl, H. G., Wang, Q., Luo, Y. M., Liu, S. J., Dahl, C., & Jiang, C. Y. (2014). Thiosulfate reduction mediated by DsrE/TusA homologs from acidothermophilic sulfur-oxidizing archaeon *Metallosphaera cuprina*. *Journal of Biological Chemistry*, 289(39), 26949–26959. <https://doi.org/10.1074/jbc.M114.591669>
- Löffler, M., Feldhues, J., Venceslau, S. S., Kammler, L., Grein, F., Pereira, I. A. C., & Dahl, C. (2020). DsrL mediates electron transfer between NADH and rDsrAB in *Allochrochromatium vinosum*. *Environmental Microbiology*, 22(2), 783–795. <https://doi.org/10.1111/1462-2920.14899>
- Löffler, M., Wallerang, K., Venceslau, S. S., Pereira, I. A. C., & Dahl, C. (2020). The iron-sulfur flavoprotein DsrL as NAD(P)H: acceptor oxidoreductase in oxidative and reductive dissimilatory sulfur metabolism. *Frontiers in Microbiology*, 11, 578209. <https://doi.org/10.3389/fmicb.2020.578209>
- Mandal, S., Chatterjee, S., Dam, B., Roy, P., & Das Gupta, S. K. (2007). The dimeric repressor SoxR binds cooperatively to the promoter(s) regulating expression of the sulfur oxidation (*sox*) operon of *Pseudaminobacter salicylatoxidans* KCT001. *Microbiology*, 153, 80–91.
- Marcia, M., Ermler, U., Peng, G. H., & Michel, H. (2010). A new structure-based classification of sulfide: quinone oxidoreductases. *Proteins: Structure Function and Bioinformatics*, 78(5), 1073–1083. <https://doi.org/10.1002/prot.22665>
- Meyer, B., & Kuever, J. (2007a). Molecular analysis of the distribution and phylogeny of dissimilatory adenosine-5'-phosphosulfate reductase-encoding genes (*aprBA*) among sulfur-oxidizing prokaryotes. *Microbiology*, 153(Pt 10), 3478–3498. <https://doi.org/10.1099/mic.0.2007/008250-0>
- Meyer, B., & Kuever, J. (2007b). Phylogeny of the alpha and beta subunits of the dissimilatory adenosine-5'-phosphosulfate (APS) reductase

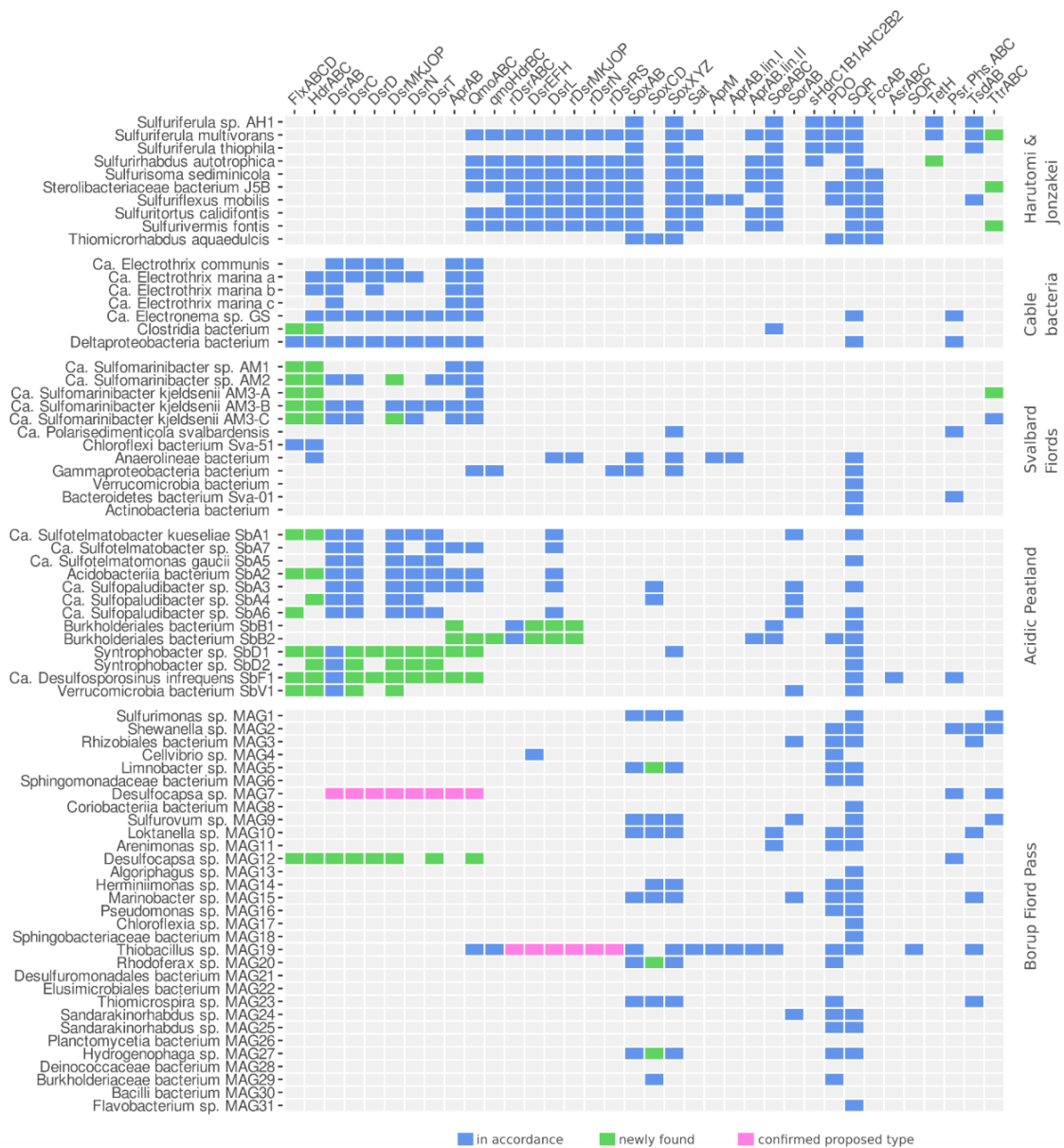
- from sulfate-reducing prokaryotes - origin and evolution of the dissimilatory sulfate-reduction pathway. *Microbiology*, 153(Pt 7), 2026–2044. <https://doi.org/10.1099/mic.0.2006/003152-0>
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Müller, F. H., Bandejas, T. M., Urich, T., Teixeira, M., & Kletzin, A. (2004). Coupling of the pathway of sulphur oxidation to dioxygen reduction: Characterization of a novel membrane-bound thiosulphate:quinone oxidoreductase. *Molecular Microbiology*, 53(4), 1147–1160. <https://doi.org/10.1111/j.1365-2958.2004.04193.x>
- Müller, H., Marozava, S., Probst, A. J., & Meckenstock, R. U. (2019). Groundwater cable bacteria conserve energy by sulfur disproportionation. *ISME Journal*, 14(2), 623–634. <https://doi.org/10.1038/s41396-019-0554-1>
- Neukirchen, S., & Sousa, F. L. (2021). DiSCo: A sequence-based type-specific predictor of Dsr-dependent dissimilatory Sulphur metabolism in microbial data. *Microbial Genomics*, 7(7), 000603. <https://doi.org/10.1099/mgen.0.000603>
- Numata, T., Fukai, S., Ikeuchi, Y., Suzuki, T., & Nureki, O. (2006). Structural basis for sulfur relay to RNA mediated by heterohexameric TusBCD complex. *Structure*, 14(2), 357–366. <https://doi.org/10.1016/j.str.2005.11.009>
- Ogawa, T., Noguchi, K., Saito, M., Nagahata, Y., Kato, H., Ohtaki, A., Nakayama, H., Dohmae, N., Matsushita, Y., Odaka, M., Yohda, M., Nyunoya, H., & Katayama, Y. (2013). Carbonyl sulfide hydrolase from *Thiobacillus thioparus* strain TH115 is one of the beta-carbonic anhydrase family enzymes. *Journal of the American Chemical Society*, 135(10), 3818–3825. <https://doi.org/10.1021/ja307735e>
- Pfeffer, C., Larsen, S., Song, J., Dong, M., Besenbacher, F., Meyer, R. L., Kjeldsen, K. U., Schreiber, L., Gorby, Y. A., El-Naggar, M. Y., Leung, K. M., Schramm, A., Risgaard-Petersen, N., & Nielsen, L. P. (2012). Filamentous bacteria transport electrons over centimetre distances. *Nature*, 491(7423), 218–221. <https://doi.org/10.1038/nature11586>
- Pott, A. S., & Dahl, C. (1998). Sirohaem-sulfite reductase and other proteins encoded in the *dsr* locus of *Chromatium vinosum* are involved in the oxidation of intracellular sulfur. *Microbiology*, 144, 1881–1894. <https://doi.org/10.1099/00221287-144-7-1881>
- Proost, S., Fostier, J., De Witte, D., Dhoedt, B., Demeester, P., Van de Peer, Y., & Vandepoele, K. (2012). I-ADHoRe 3.0--fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Research*, 40(2), e11. <https://doi.org/10.1093/nar/gkr955>
- Pyne, P., Alam, M., & Ghosh, W. (2017). A novel *soxO* gene, encoding a glutathione disulfide reductase, is essential for tetrathionate oxidation in *Advenella kashmirensis*. *Microbiological Research*, 205, 1–7. <https://doi.org/10.1016/j.micres.2017.08.002>
- Rabus, R., Venceslau, S. S., Wohlbrand, L., Voordouw, G., Wall, J. D., & Pereira, I. A. (2015). A post-genomic view of the ecophysiology, catabolism and biotechnological relevance of sulphate-reducing prokaryotes. *Advances in Microbial Physiology*, 66, 55–321. <https://doi.org/10.1016/bs.ampbs.2015.05.002>
- Ramos, A. R., Grein, F., Oliveira, G. P., Venceslau, S. S., Keller, K. L., Wall, J. D., & Pereira, I. A. (2015). The FlxABC-HdrABC proteins correspond to a novel NADH dehydrogenase/heterodisulfide reductase widespread in anaerobic bacteria and involved in ethanol metabolism in *Desulfovibrio vulgaris* Hildenborough. *Environmental Microbiology*, 17(7), 2288–2305. <https://doi.org/10.1111/1462-2920.12689>
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In L. Liu & T. Özsu (Eds.), *Encyclopedia of database systems* (pp. 532–538). Springer.
- Rother, D., Heinrich, H. J., Quentmeier, A., Bardischewsky, F., & Friedrich, C. G. (2001). Novel genes of the *sox* gene cluster, mutagenesis of the flavoprotein SoxF, and evidence for a general sulfur-oxidizing system in *Paracoccus pantotrophus* GB17. *Journal of Bacteriology*, 183(15), 4499–4508. <https://doi.org/10.1128/JB.183.15.4499-4508.2001>
- Rother, D., Orawski, G., Bardischewsky, F., & Friedrich, C. G. (2005). SoxRS-mediated regulation of chemotrophic sulfur oxidation in *Paracoccus pantotrophus*. *Microbiology*, 151(5), 1707–1716. <https://doi.org/10.1099/mic.0.27724-0>
- Rother, D., Ringk, J., & Friedrich, C. G. (2008). Sulfur oxidation of *Paracoccus pantotrophus*: The sulfur-binding protein SoxYZ is the target of the periplasmic thiol-disulfide oxidoreductase SoxS. *Microbiology*, 154(7), 1980–1988. <https://doi.org/10.1099/mic.0.2008/018655-0>
- Rückert, C. (2016). Sulfate reduction in microorganisms-recent advances and biotechnological applications. *Current Opinion in Microbiology*, 33, 140–146. <https://doi.org/10.1016/j.mib.2016.07.007>
- Sayers, E. W.,avanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Research*, 47(D1), D94–D99. <https://doi.org/10.1093/nar/gky989>
- Skewes-Cox, P., Sharpton, T. J., Pollard, K. S., & DeRisi, J. L. (2014). Profile hidden Markov models for the detection of viruses within metagenomic sequence data. *PLoS One*, 9(8), e105067. <https://doi.org/10.1371/journal.pone.0105067>
- Sousa, F. M., Pereira, J. G., Marreiros, B. C., & Pereira, M. M. (2018). Taxonomic distribution, structure/function relationship and metabolic context of the two families of sulfide dehydrogenases: SQR and FCS. *Biochimica et Biophysica Acta Bioenergetics*, 1859(9), 742–753. <https://doi.org/10.1016/j.bbabi.2018.04.004>
- Stambouliau, M., Guerrero, R. F., Hahn, M. W., & Radivojac, P. (2020). The ortholog conjecture revisited: The value of orthologs and paralogs in function prediction. *Bioinformatics*, 36(Suppl\_1), i219–i226. <https://doi.org/10.1093/bioinformatics/btaa468>
- Svetlitsky, D., Dagan, T., Chalifa-Caspi, V., & Ziv-Ukelson, M. (2019). CSBFinder: Discovery of colinear syntenic blocks across thousands of prokaryotic genomes. *Bioinformatics*, 35(10), 1634–1643. <https://doi.org/10.1093/bioinformatics/bty861>
- Svetlitsky, D., Dagan, T., & Ziv-Ukelson, M. (2020). Discovery of multi-operon colinear syntenic blocks in microbial genomes. *Bioinformatics*, 36(Suppl\_1), i21–i29. <https://doi.org/10.1093/bioinformatics/btaa503>
- Tanabe, T. S., Leimkühler, S., & Dahl, C. (2019). The functional diversity of the prokaryotic sulfur carrier protein TusA. *Advances in Microbial Physiology*, 75, 233–277. <https://doi.org/10.1016/bs.ampbs.2019.07.004>
- Tanaka, Y., Yoshikae, K., Takeuchi, A., Ichikawa, M., Mori, T., Uchino, S., Sugano, Y., Hakoshima, T., Takagi, H., Nonaka, G., & Tsukazaki, T. (2020). Crystal structure of a YeeE/YedE family protein engaged in thiosulfate uptake. *Science advances*, 6(35), eaba7637. <https://doi.org/10.1126/sciadv.aba7637>
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetverin, V., Nawrocki, E. P., Zaslavsky, L., Lomsadze, A., Pruitt, K. D., Borodovsky, M., & Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, 44(14), 6614–6624. <https://doi.org/10.1093/nar/gkw569>
- Thorup, C., Schramm, A., Findlay, A. J., Finster, K. W., & Schreiber, L. (2017). Disguised as a sulfate reducer: Growth of the deltaproteobacterium *Desulfovibrio alkaliphilus* by sulfide oxidation with nitrate. *MBio*, 8(4), e00671–e006717. <https://doi.org/10.1128/mBio.00671-17>
- Tikhonova, T. V., Sorokin, D. Y., Hagen, W. R., Khrenova, M. G., Muiyzer, G., Rikitina, T. V., Shabalin, I. G., Trofimov, A. A., Tsallagov, S. I., & Popov, V. O. (2020). Trinuclear copper biocatalytic center forms an active site of thiocyanate dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America*, 117(5290), 201922133. <https://doi.org/10.1073/pnas.1922133117>
- Trivedi, C. B., Stamps, B. W., Lau, G. E., Grasby, S. E., Templeton, A. S., & Spear, J. R. (2020). Microbial metabolic redundancy is a

- key mechanism in a sulfur-rich glacial ecosystem. *mSystems*, 5(4), e00504-20. <https://doi.org/10.1128/mSystems.00504-20>
- Tsallagov, S. I., Sorokin, D. Y., Tikhonova, T. V., Popov, V. O., & Muyzer, G. (2019). Comparative genomics of *Thiohalobacter thiocyanaticus* HRh1T and *Guyparkeria* sp. SCN-R1, halophilic chemolithoautotrophic sulfur-oxidizing Gammaproteobacteria capable of using thiocyanate as energy source. *Frontiers in Microbiology*, 10, 898. <https://doi.org/10.3389/fmicb.2019.00898>
- Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., Lajus, A., Pascal, G., Scarpelli, C., & Medigue, C. (2006). MaGe: A microbial genome annotation system supported by synteny results. *Nucleic Acids Research*, 34(1), 53–65. <https://doi.org/10.1093/nar/gkj406>
- Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7, 91. <https://doi.org/10.1186/1471-2105-7-91>
- Wagner, T., Koch, J., Ermler, U., & Shima, S. (2017). Methanogenic heterodisulfide reductase (HdrABC-MvhAGD) uses two noncubane [4Fe-4S] clusters for reduction. *Science*, 357(6352), 699–703. <https://doi.org/10.1126/science.aan0425>
- Wallace, I. M., O'Sullivan, O., Higgins, D. G., & Notredame, C. (2006). M-coffee: Combining multiple sequence alignment methods with T-coffee. *Nucleic Acids Research*, 34(6), 1692–1699. <https://doi.org/10.1093/nar/gkl091>
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40(7), e49. <https://doi.org/10.1093/nar/gkr1293>
- Wasmund, K., Mussmann, M., & Loy, A. (2017). The life sulfuric: Microbial ecology of sulfur cycling in marine sediments. *Environmental Microbiology*, 9(4), 323–344. <https://doi.org/10.1111/1758-2229.12538>
- Watanabe, T., Kojima, H., & Fukui, M. (2015). *Sulfuriferula multivorans* gen. nov., sp. nov., isolated from a freshwater lake, reclassification of *'Thiobacillus plumbophilus'* as *Sulfuriferula plumbophilus* sp. nov., and description of *Sulfuricellaceae* fam. nov. and *Sulfuricellales* ord. nov. *International Journal of Systematic and Evolutionary Microbiology*, 65(Pt 5), 1504–1508. <https://doi.org/10.1099/ijs.0.000129>
- Watanabe, T., Kojima, H., Umezawa, K., Hori, C., Takasuka, T. E., Kato, Y., & Fukui, M. (2019). Genomes of neutrophilic sulfur-oxidizing chemolithoautotrophs representing 9 proteobacterial species from 8 genera. *Frontiers in Microbiology*, 10, 316. <https://doi.org/10.3389/fmicb.2019.00316>
- Wilson, J. J., & Kappler, U. (2009). Sulfite oxidation in *Sinorhizobium meliloti*. *Biochimica et Biophysica Acta*, 1787(12), 1516–1525. <https://doi.org/10.1016/j.bbabi.2009.07.005>
- Yu, X., Zhou, J., Song, W., Xu, M., He, Q., Peng, Y., Tian, Y., Wang, C., Shu, L., Wang, S., Yan, Q., Liu, J., Tu, Q., & He, Z. (2020). SCycDB: A curated functional gene database for metagenomic profiling of Sulphur cycling pathways. *Molecular Ecology Resources*, 21(3), 924–940. <https://doi.org/10.1111/1755-0998.13306>
- Zecchin, S., Mueller, R. C., Seifert, J., Stingl, U., Anantharaman, K., von Bergen, M., Cavalca, L., & Pester, M. (2018). Rice paddy *Nitrospirae* carry and express genes related to sulfate respiration: Proposal of the new genus "Candidatus Sulfobium". *Applied and Environmental Microbiology*, 84(5), e02224-17. <https://doi.org/10.1128/AEM.02224-17>

#### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Tanabe, T. S., & Dahl, C. (2022). HMS-S-S: A tool for the identification of Sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies. *Molecular Ecology Resources*, 22, 2758–2774. <https://doi.org/10.1111/1755-0998.13642>



**Fig. S1.** Application of HMS-S-S to published metagenomes. Sulfur-metabolism-related proteins detected by HMS-S-S in recently described assemblies of neutrophilic sulfur oxidizers (Watanabe et al., 2019), cable bacteria (Kjeldsen et al., 2019; Pfeffer et al., 2012), novel acidobacteria (Flieder et al., 2021; Hausmann et al., 2018; Müller et al., 2019), and sulfur glacier meta-assembled genomes (Trivedi et al., 2020) are plotted in blue when they had also been detected in previously published work. Boxes are colored pink when HMS-S-S confirmed a reductive/oxidative type for a specific protein that had previously been proposed but could by then not be definitely established. Green boxes indicate proteins newly found by HMS-S-S.

<sup>0</sup>The supplementary tables S1,S2 and S5 are not printable and available through the online publication.  
<https://doi.org/10.1111/1755-0998.13642>

## References

- Flieder, M., Buongiorno, J., Herbold, C. W., Hausmann, B., Rattei, T., Lloyd, K. G., Loy, A., & Wasmund, K. (2021). Novel taxa of Acidobacteriota implicated in seafloor sulfur cycling. *ISME Journal*. <https://doi.org/10.1038/s41396-021-00992-0>
- Hausmann, B., Pelikan, C., Herbold, C. W., Kostlbacher, S., Albertsen, M., Eichorst, S. A., Glavina Del Rio, T., Huemer, M., Nielsen, P. H., Rattei, T., Stingl, U., Tringe, S. G., Trojan, D., Wentrup, C., Woebken, D., Pester, M., & Loy, A. (2018). Peatland Acidobacteria with a dissimilatory sulfur metabolism. *ISME Journal*, 12, 1729-1742. <https://doi.org/10.1038/s41396-018-0077-1>
- Kjeldsen, K. U., Schreiber, L., Thorup, C. A., Boesen, T., Bjerg, J. T., Yang, T., Dueholm, M. S., Larsen, S., Risgaard-Petersen, N., Nierychlo, M., Schmid, M., Boggild, A., van de Vossenberg, J., Geelhoed, J. S., Meysman, F. J. R., Wagner, M., Nielsen, P. H., Nielsen, L. P., & Schramm, A. (2019). On the evolution and physiology of cable bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 116(38), 19116-19125. <https://doi.org/10.1073/pnas.1903514116>
- Müller, H., Marozava, S., Probst, A. J., & Meckenstock, R. U. (2019). Groundwater cable bacteria conserve energy by sulfur disproportionation. *ISME Journal*, 14(2), 623-634. <https://doi.org/10.1038/s41396-019-0554-1>
- Pfeffer, C., Larsen, S., Song, J., Dong, M., Besenbacher, F., Meyer, R. L., Kjeldsen, K. U., Schreiber, L., Gorby, Y. A., El-Naggar, M. Y., Leung, K. M., Schramm, A., Risgaard-Petersen, N., & Nielsen, L. P. (2012). Filamentous bacteria transport electrons over centimetre distances. *Nature*, 491(7423), 218-221. <https://doi.org/10.1038/nature11586>
- Trivedi, C. B., Stamps, B. W., Lau, G. E., Grasby, S. E., Templeton, A. S., & Spear, J. R. (2020). Microbial metabolic redundancy is a key mechanism in a sulfur-rich glacial ecosystem. *mSystems*, 5(4). <https://doi.org/10.1128/mSystems.00504-20>
- Watanabe, T., Kojima, H., Umezawa, K., Hori, C., Takasuka, T. E., Kato, Y., & Fukui, M. (2019). Genomes of neutrophilic sulfur-oxidizing chemolithoautotrophs representing 9 proteobacterial species from 8 genera. *Frontiers in Microbiology*, 10, 316. <https://doi.org/10.3389/fmicb.2019.00316>

Table S3. Reference proteins for dataset annotation

Pathway	HMM name	RefSeqId	HMS-S-S annotation/function	COG	Pfam family	TIGRfam family	Reference
Dissimilatory sulfate reduction	redDsrA	WP_010937709	reductive-type dissimilatory siroheme-sulfite reductase	COG2221	pf01077	TIGR02064	Wolfe and Cowan 1994
	redDsrB	WP_010937710.1	reductive-type dissimilatory siroheme-sulfite reductase	COG2221	pf01077	TIGR02066	Wolfe and Cowan 1994
	redDsrC	WP_010940042.1	sulfur-binding protein substrate of DsrAB, reductive type	COG2920	pf04358	TIGR03342	Santos et al. 2015
	redDsrM	WP_010938586.1	electron-transferring DsrMK(JOP) complex, reductive type	COG2181	pf02665	TIGR00351	Pires et al. 2006
	redDsrK	WP_010938585.1	electron-transferring DsrMK(JOP) complex, reductive type	COG0247	pf13183	TIGR00273	Pires et al. 2006
	redDsrJ	WP_010938584.1	electron-transferring DsrMK(JOP) complex, reductive type				Pires et al. 2006
	redDsrO	WP_010938583.1	electron-transferring DsrMK(JOP) complex, reductive type	COG0437	pf13247	TIGR03149	Pires et al. 2006
	redDsrP	WP_010938582.1	complex, reductive type	COG5557	pf03916	TIGR03148	Pires et al. 2006
	DsrD	WP_010937711.1	unknown function		pf08679		Hittel and Voordouw 2000
	redDsrN	WP_010937712.1	amidation of siroheme, reductive type	COG1797	pf07685	TIGR07685	Mathews et al. 1995, Lübke et al. 2006
	DsrT	WP_012611240.1	unknown function		pf14361		Holkenbrink et al. 2011, Hausmann et al. 2018
	redAprA	WP_010938147.1	reductive-type adenylylsulfate reductase flavoprotein	COG1053	pf00890		Meyer and Kuever 2007a
	redAprB	WP_010938146.1	reductive-type adenylylsulfate reductase iron-sulfur protein	COG1146	pf12139		Meyer and Kuever 2007a
	QmoA	WP_010938148.1	Quinone reactive oxidoreductase, electron transfer to/from APS reductase	COG1148			Ramos et al. 2012, Duarte et al. 2016
	QmoB	WP_010938149.1	Quinone reactive oxidoreductase, electron transfer to/from APS reductase	COG1148	pf07992		Ramos et al. 2012, Duarte et al. 2016
	QmoC	WP_011312158.1	Quinone reactive oxidoreductase, electron transfer to/from APS reductase	COG1150	pf02665		Ramos et al. 2012, Duarte et al. 2016
	qHdrB	WP_011312157.1	qHdrBC replace QmoC in some organisms	COG1150	pf02754		Meyer and Kuever 2007a, Peireira et al. 2011
	qHdrC	WP_010938590.1	qHdrBC replace QmoC in some organisms	COG2048	pf13187		Meyer and Kuever 2007a, Peireira et al. 2011
	redSat	WP_010938590.1	reductive-type sulfate adenylyltransferase	COG2046	pf01747		Parey et al. 2013, Hanna et al. 2002, Gavel et al. 1998
	Assimilatory reduction of APS	assAprA	WP_006007809.1	assimilatory-type AprAB adenylylsulfate reductase		pf00890	
assAprB		WP_006007810.1	assimilatory-type AprAB adenylylsulfate reductase		pf12139		Chernyh et al. 2020, Neukirchen et al. 2021
Sulfite reduction	AsrA	WP_000985204.1	Anaerobic sulfite reductase		pf17179	TIGR02910	Huang and Barrett 1991
	AsrB	WP_000017587.1	Anaerobic sulfite reductase		pf10418	TIGR02911	Huang and Barrett 1991
	AsrC	WP_000020685.1	Anaerobic sulfite reductase		pf13187	TIGR02912	Huang and Barrett 1991
	MccA	WP_011138326.1	Octaheme sulfite reductase	COG0484			Kern et al. 2011, Hermann et al. 2015
	MccB	WP_011138327.1	Octaheme sulfite reductase	COG0545			Kern et al. 2011, Hermann et al. 2015
	MccC	WP_011138328.1	Octaheme sulfite reductase	COG0437			Kern et al. 2011, Hermann et al. 2015
	MccD	WP_011138329.1	Octaheme sulfite reductase	COG3301			Kern et al. 2011, Hermann et al. 2015
	Reduction of polysulfides/thiosulfate/sulfur	PsrAPhsASreA	WP_129545342.1	Polysulfide/Thiosulfate/Sulfur reductase	COG0243	pf04879	
PsrBPhsBSreB		WP_011138082.1	Polysulfide/Thiosulfate/Sulfur reductase	COG0437	pf13247		Stoffels et al. 2012, Krafft et al. 1995, Laska et al. 2003
PsrCPhsCSreC		WP_011138083.1	Polysulfide/Thiosulfate/Sulfur reductase	COG3301	pf03916		Stoffels et al. 2012, Krafft et al. 1995, Laska et al. 2003
Reduction of tetrathionate	TtrA	WP_000002375.1	Tetrathionate reductase		pf04879		Hensel et al. 1999
	TtrB	WP_000269830.1	Tetrathionate reductase		pf13247		Hensel et al. 1999
	TtrC	WP_000149760.1	Tetrathionate reductase		pf03916		Hensel et al. 1999
	TtrR	WP_000190927.1	Transcriptional regulator component				Hensel et al. 1999
	TtrS	WP_001214413.1	Transcriptional regulator component				Hensel et al. 1999
Thiosulfate/tetrathionate conversion	TsdA	WP_012969337.1	Thiosulfate dehydrogenase	COG3258	pf13442		Denkmann et al. 2012, Kurth et al. 2016
	TsdB	WP_013124333.1	Thiosulfate dehydrogenase	COG2863	pf01494		Denkmann et al. 2012, Kurth et al. 2016
	DoxA	WP_152941534.1	Thiosulfate:quinone oxidoreductase		pf07680		Müller et al. 2004
DoxD	WP_013774923.1	Thiosulfate:quinone oxidoreductase		pf04173		Müller et al. 2004	
Tetrathionate oxidation	Tth	WP_012535754.1	Tetrathionate hydrolase	COG1520	pf01011		Kanao et al. 2007
	SoxO	WP_014750231.1	Periplasmic glutathione disulfide reductase				Pyne et al. 2017
Elemental sulfur oxidation	SOR	WP_010880260.1	Cytoplasmic sulfur oxygenase/reductase		pf07682		Kletzin et al. 1992, Pelletier et al. 2008
Thiosulfate oxidation in periplasm	SoxA	WP_011750382.1	Thiosulfate-oxidizing multienzyme system, c-type cytochrome	COG3258			Friedrich et al. 2005



SoxB	WP_011750383.1	Thiosulfate-oxidizing multienzyme system, thiosulfohydrolase	COG0737	pf02872	TIGR04486	Friedrich et al. 2005
SoxC	WP_011750384.1	Thiosulfate-oxidizing multienzyme system, sulfane dehydrogenase	COG2041	pf03404	TIGR04555	Friedrich et al. 2005
SoxD	WP_011750385.1	Thiosulfate-oxidizing multienzyme system, sulfane dehydrogenase	COG3258	pf00034		Friedrich et al. 2005
SoxX	WP_011750379.1	Thiosulfate-oxidizing multienzyme system, c-type cytochrome	COG2010	pf00034		Friedrich et al. 2005
SoxY	WP_011750380.1	Thiosulfate-oxidizing multienzyme system, substrate binding	COG5501	pf13501		Friedrich et al. 2005
SoxZ	WP_011750381.1	Thiosulfate-oxidizing multienzyme system, substrate binding	COG5501	pf08770	TIGR04490	Friedrich et al. 2005
SoxE	WP_011750386.1	diheme cytochrome c	COG2863	pf00034		Friedrich et al. 2005
SoxF	WP_011750387.1	flavoprotein	COG0446	pf07992		Bardischewsky et al 2006a
SoxG	WP_011750388.1	periplasmic, MBL fold metallo-hydrolase	COG0491	pf00753		Rother et al. 2001, Friedrich et al. 2005
SoxH	WP_011750389.1	periplasmic, MBL fold metallo-hydrolase	COG0491	pf00753		Rother et al. 2001, Friedrich et al. 2005
SoxS	WP_011750376.1	periplasmic thioredoxin, thiol-disulfide oxidoreductase	COG0526			Rother et al. 2008
SoxV	WP_011750377.1	membrane protein, disulfide transporter, reduces SoxW	COG0785	pf02683		Appia-Ayme and Berks 2002
SoxW	WP_011750378.1	periplasmic thioredoxin	COG2143	pf13098		Appia-Ayme and Berks 2002, Bardischewsky et al 2006b
SoxT1	WP_011750374.1	YeeE-like transporter	COG2391	pf04143		Lahiri et al 2006, Friedrich et al 2005
SoxT2	WP_041530453.1	YeeE-like transporter	COG2391	pf04143		Friedrich et al. 2008
<b>Sulfide oxidation</b>						
FccA	WP_012970308.1	Sulfide dehydrogenase flavoprotein	COG2863			Dolata et al. 1993
FccB	WP_012970307.1	Sulfide dehydrogenase hemoprotein	COG0446	pf07992		Dolata et al. 1993
SQRI	WP_010881436.1	Sulfide:quinone oxidoreductase	COG0446	pf07992		Marcia et al. 2010, Arieli et al. 1994
SQRII	WP_010932704.1	Sulfide:quinone oxidoreductase	COG0446	pf07992		Marcia et al. 2010, Hildebrandt and Grieshaber 2008, vande Weghe and Ow 1999
SQRIII	WP_010931811.1	Sulfide:quinone oxidoreductase	COG0447	pf07992		Marcia et al. 2010, Chan et al. 2009, Reinartz et al. 1998
SQRIV	WP_010932556.1	Sulfide:quinone oxidoreductase	COG0448	pf07992		Marcia et al. 2010, Chan et al. 2009, Brito et al. 2009
SQRV	WP_010932765.1	Sulfide:quinone oxidoreductase	COG0448	pf07992		Marcia et al. 2010, Chan et al. 2009
SQRVI	WP_010932765.1	Sulfide:quinone oxidoreductase	COG0448	pf07992		Marcia et al. 2010, Chan et al. 2009
<b>Sulfide detoxification: sulfane sulfur oxidation</b>						
SDOI	WP_014003505.1	Sulfur dioxygenase		pf00753		Liu and Xun 2014
SDOII	WP_003101427.1	Sulfur dioxygenase		pf00037		Liu and Xun 2014
SDOIII	WP_000465472.1	Sulfur dioxygenase		pf00753		Liu and Xun 2014
<b>Sulfane sulfur oxidation in cytoplasm</b>						
oxDsrA	WP_012970464.1	oxidative-type dissimilatory sulfite reductase	COG2221	pf01077	TIGR02064	Pott and Dahl 1998
oxDsrB	WP_012970465.1	oxidative-type dissimilatory sulfite reductase	COG2221	pf01077	TIGR02066	Pott and Dahl 1998
DsrL	WP_012970472.1	NAD(P)H-acceptor oxidoreductase	COG0493	pf14691	TIGR01316	Dahl et al. 2005, Löffler et al 2020a,b
oxDsrM	WP_012970470.1	electron-transferring DsrMK(JOP) complex, oxidative type	COG2181	pf02665	TIGR00351	Pott and Dahl 1998, Grein et al 2010a
oxDsrK	WP_012970471.1	electron-transferring DsrMK(JOP) complex, oxidative type	COG2047	pf13183	TIGR00273	Pott and Dahl, 1998, Grein et al 2010a
oxDsrJ	WP_012970473.1	electron-transferring DsrMK(JOP) complex, oxidative type				Dahl et al. 2005, Grein et al 2010a,b
oxDsrI	WP_012970474.1	electron-transferring DsrMK(JOP) complex, oxidative type	COG0437	pf13247	TIGR03149	Dahl et al. 2005, Grein et al 2010a
oxDsrO	WP_012970475.1	electron-transferring DsrMK(JOP) complex, oxidative type	COG5557	pf03916	TIGR03148	Dahl et al. 2005, Grein et al 2010a
oxDsrP	WP_012970476.1	amidation of siroheme, oxidative type	COG1797	pf07685	TIGR00379	Dahl et al. 2005, Lübbe et al. 2006
oxDsrN	WP_012970477.1	IscA-like protein involved in <i>dsr</i> posttranslational control	COG0316	pf01521	TIGR00049	Dahl et al. 2005, Grimm et al. 2010
DsrR	WP_012970478.1	Participation in <i>dsr</i> posttranslational control				Dahl et al. 2005, Grimm et al. 2011
DsrS	WP_012970478.1	sulfur-binding protein substrate of DsrAB, oxidative type				Pott and Dahl 1998, Cort et al. 2008, Santos et al. 2015
oxDsrC	WP_012970469.1	sulfur transferase DsrEFH	COG2920	pf04358	TIGR03342	Pott and Dahl 1998, Dahl et al. 2008
DsrE	WP_012970466.1	sulfur transferase DsrEFH	COG1553	pf02635	TIGR03012	Pott and Dahl 1998, Dahl et al. 2008
DsrF	WP_012970467.1	sulfur transferase DsrEFH	COG2923	pf02635	TIGR03010	Pott and Dahl 1998, Dahl et al. 2008
DsrH	WP_012970468.1	sulfur transferase DsrEFH	COG2168	pf04358	TIGR03011	Pott and Dahl 1998, Dahl et al. 2008
DsrE2	WP_012971779.1	sulfur transferase	COG2210	pf13686		Liu et al. 2014, Stockdreher et al. 2014
DsrE3A	WP_013737284.1	sulfur transferase	arCOG02066	pf02635		Liu et al. 2014
DsrE3B	WP_004868635.1	sulfur transferase		pf02635		Liu et al. 2014
DsrE3C	WP_004868827.1	sulfur transferase		pf02635		Liu et al. 2014
DsrE4	WP_013738325.1	sulfur transferase		pf02635		Liu et al. 2014
DsrE5	WP_013738307.1	sulfur transferase		pf02635		Liu et al. 2014
sHdrA	WP_014003386.1	Sulfur oxidizing heterodisulfide reductase-like system		pf07992		Koch and Dahl 2018
sHdrB1	WP_014003387.1	Sulfur oxidizing heterodisulfide reductase-like system		pf02754		Koch and Dahl 2018
sHdrB2	WP_004868633.1	Sulfur oxidizing heterodisulfide reductase-like system		pf02754		Koch and Dahl 2018

	sHdrC1	WP_004868548.1	Sulfur oxidizing heterodisulfide reductase-like system	pf13183		Koch and Dahl 2018
	sHdrC2	WP_004868632.1	Sulfur oxidizing heterodisulfide reductase-like system			Koch and Dahl 2018
	sHdrH	WP_004868631.1	sHdr system component of unknown function			Koch and Dahl 2018
	sHdrCCG1	WP_028963020.1	non-cubane [4Fe-4S] cluster-binding sHdr system component	pf02754		Cao et al. 2018
	sHdrCCG2	WP_028963017.2	non-cubane [4Fe-4S] cluster-binding sHdr system component	pf02754		Cao et al. 2018
	sEtfA	WP_028963019.2	electron-transferring flavoprotein associated with sHdr	pf01012		Cao et al. 2018
	sEtfB	WP_028963018.2	electron-transferring flavoprotein associated with sHdr	pf02754		Cao et al. 2018
	sHdrT	WP_014003382.1	TK90_0645-like membrane protein of unknown function	pf03349		Cao et al. 2018
	LbpA1	WP_004868634.1	Lipoate-binding protein	pf01597		Cao et al. 2018
	LbpA2	WP_004868636.1	Lipoate-binding protein	pf01597		Cao et al. 2018
	sHdrI	WP_013214732.1	Hden_0695-like protein of unknown function associated with sHdr/LbpA			Koch and Dahl 2018
	sHdrRSoxR	WP_013214719.1	ArsR-type transcriptional regulator			Rother et al. 2005, Koch and Dahl 2018
<b>Sulfite oxidation in cytoplasm</b>	AprM	WP_012970333.1	oxidative-type adenylylsulfate reductase, membrane subunit			Meyer and Kuever, 2007b
	oxAprAI	WP_012970335.1 WP_010932546.1	oxidative-type adenylylsulfate reductase	pf00890		Meyer and Kuever, 2007b
	oxAprAll	WP_012970334.1	oxidative-type adenylylsulfate reductase	pf00890		Meyer and Kuever, 2007b
	oxAprBI	WP_010932545.1	oxidative-type adenylylsulfate reductase	pf12139		Meyer and Kuever, 2007b
	oxAprBII	WP_010932545.1 WP_012970332.1	oxidative-type adenylylsulfate reductase, lineage II	pf12139		Meyer and Kuever, 2007b
	oxSat	WP_012971673.1	oxidative-type sulfate adenylyltransferase	COG2046 pf01747		Laue and Nelson 1994
	SoeA	WP_012971672.1	sulfite-oxidizing enzyme	COG0243 pf04879, pf00384, pf01568		Dahl et al. 2013
	SoeB	WP_012971671.1	sulfite-oxidizing enzyme	COG0437 pf13247, pf12800		Dahl et al. 2013
	SoeC	WP_012971671.1	sulfite-oxidizing enzyme	COG3302 pf04976		Dahl et al. 2013
<b>Sulfite oxidation in periplasm</b>	SorA/T	WP_013168043.1	sulfite-oxidizing enzyme	COG2041 pf03404		Kappler et al. 2000, Kappler and Enemark 2014
	SorB	WP_013168044.1	sulfite-oxidizing enzyme	COG2010		Kappler et al. 2000, Kappler and Enemark 2014
<b>Cysteine biosynthesis</b>	YeeE	NP_416517	Thiosulfate uptake	COG2391 pf04143		Tanaka et al. 2020
<b>Growth on CS<sub>2</sub></b>	Cs2H	WP_038471897.1	CS <sub>2</sub> lyase	pf00484		Smeulders et al. 2013
<b>Selenocysteine synthesis</b>	TusADsrE	WP_012799650.1	Selenium transfer/homeostasis	COG0425 pf01206	TIGR03527	Lin et al. 2015, Zhang et al. 2008
<b>2-thiouridine synthesis</b>	TusA	WP_012971778.1	Sulfur transferase	COG0425 pf01206		Ikeuchi et al. 2006, Tanabe et al. 2019
	TusB	WP_000903373.1	Sulfur transferase TusBCD	COG2168 pf02635		Ikeuchi et al. 2006, Numata et al. 2006
	TusC	WP_000820714.1	Sulfur transferase TusBCD	pf02635		Ikeuchi et al. 2006, Numata et al. 2006
	TusD	WP_001209680.1	Sulfur transferase TusBCD	COG1553 pf04358		Ikeuchi et al. 2006, Numata et al. 2006
	TusE	WP_000904442.1	Sulfur transferase	COG2920 pf04358	TIGR03342	Ikeuchi et al. 2006
<b>Electron transport</b>	QrcA	WP_010940677.1	Menaquinone-reducing membrane complex involved in sulfate reduction	pf02085		Venceslau et al. 2010
	QrcB	WP_010937997.1	Menaquinone-reducing membrane complex involved in sulfate reduction	COG0243 pf00384		Venceslau et al. 2010
	QrcC	WP_010937996.1	Menaquinone-reducing membrane complex involved in sulfate reduction	COG0437 pf00037		Venceslau et al. 2010
	QrcD	WP_010937995.1	Menaquinone-reducing membrane complex involved in sulfate reduction	COG5557 pf03916		Venceslau et al. 2010
	HmcA	WP_014524248.1	Transmembrane complex mediating electron transfer from hydrogen to sulfate	COG0484 pf02085		Rossi et al. 1993, Dolla et al. 2000
	HmcB	WP_010937841.1	Transmembrane complex mediating electron transfer from hydrogen to sulfate	COG0437 pf13247		Rossi et al. 1993, Dolla et al. 2000
	HmcC	WP_010937840.1	Transmembrane complex mediating electron transfer from hydrogen to sulfate	COG5557 pf03916		Rossi et al. 1993, Dolla et al. 2000
	HmcD	WP_010937839.1	Transmembrane complex mediating electron transfer from hydrogen to sulfate	COG2181		Rossi et al. 1993, Dolla et al. 2000
	HmcE	WP_010937838.1	Transmembrane complex mediating electron transfer from hydrogen to sulfate	COG2181		Rossi et al. 1993, Dolla et al. 2000
	HmcF	WP_010937837.1	Transmembrane complex mediating electron transfer from hydrogen to sulfate	COG0247 pf02754		Rossi et al. 1993, Dolla et al. 2000
	TmcA	WP_010937572.1	Transmembrane transfer of electron from periplasmic hydrogen oxidation	pf02085		Pereira et al. 2006

	TmcB	WP_010937573.1	Transmembrane transfer of electron from periplasmic hydrogen oxidation	COG0247	pf02754	Pereira et al. 2006	
	TmcC	WP_010937574.1	Transmembrane transfer of electron from periplasmic hydrogen oxidation	COG2181		Pereira et al. 2006	
	TmcD	WP_010937575.1	Transmembrane transfer of electron from periplasmic hydrogen oxidation	COG0823		Pereira et al. 2006	
<b>LbpA biosynthesis</b>	GerGerRed	WP_014003384.1	Flavoprotein similar to geranylgeranyl reductase		pf01494 TIGR02032	Cao et al. 2018	
	LipS1	WP_004868637.1	Radical S-adenosyl domain containing protein		pf04055	Cao et al. 2018, Jin et al. 2020	
	LipS2	WP_014003383.1	Radical S-adenosyl domain containing protein		pf04055	Cao et al. 2018, Jin et al. 2020	
	LplA	WP_014003385.1	Lipoate:protein ligase			Cao et al. 2018	
	Hden0687	WP_013214724.1	Hden_0687-like protein of unknown function	COG0518		Cao et al. 2018	
<b>Glycine cleavage system</b>	GcvH	WP_417380	Lipoate-binding glycine cleavage system protein	COG0509	pf01597	TIGR00527 Okamudura-Ikeda et al. 1993	
	DHDL	NP_414658	Dihydroipoamide dehydrogenase	COG1249	pf07992	TIGR01350 Carothers et al. 1989	
<b>Methanogenesis</b>	McrA	WP_013296337.1	Methyl-coenzyme M reductase		pf02249	TIGR03256 Kaster et al. 2011	
	McrB	WP_013296341.1	Methyl-coenzyme M reductase		pf02241	TIGR03257 Kaster et al. 2011	
	McrC	WP_013296339.1	Methyl-coenzyme M reductase			TIGR03264 Kaster et al. 2011	
	McrD	WP_013296340.1	Methyl-coenzyme M reductase			Kaster et al. 2011	
	McrG	WP_013296338.1	Methyl-coenzyme M reductase		pf02240	TIGR03259 Kaster et al. 2011	
	mHdRA	WP_010939674.1	methanogen-type heterodisulfide reductase	COG1148	pf00037	Ramos et al. 2015	
	mHdB	WP_010939675.1	methanogen-type heterodisulfide reductase	COG2048	pf02754	Ramos et al. 2015	
	mHdRC	WP_010939676.1	methanogen-type heterodisulfide reductase	COG1150	pf13187	Ramos et al. 2015	
	mHdRD	WP_048045886.1	membrane-bound heterodisulfide reductase	arCOG00333	pf02754	Mander et al. 2004	
	mHdRE	WP_011033778.1	membrane-bound heterodisulfide reductase	arCOG05014	pf02665	Mander et al. 2004	
	MvhA	WP_013296307.1	[NiFe] hydrogenase associated with mHdRABC		pf00374	Kaster et al. 2011	
	MvhD	WP_013296309.1	[NiFe] hydrogenase associated with mHdRABC		pf02662	Kaster et al. 2011	
	MvhG	WP_013296308.1	[NiFe] hydrogenase associated with mHdRABC		pf01058	Kaster et al. 2011	
	<b>NADH dehydrogenase/heterodisulfide reductase</b>	FlxA	WP_010939671.1	Part of NADH dehydrogenase/heterodisulfide reductase	COG0543	pf10418	Ramos et al. 2015
		FlxB	YP_011613.1	Part of NADH dehydrogenase/heterodisulfide reductase	COG1453	pf17179	Ramos et al. 2015
FlxC		WP_010940765.1	Part of NADH dehydrogenase/heterodisulfide reductase	COG0479	pf13237	Ramos et al. 2015	
FlxD		WP_010940766.1	Part of NADH dehydrogenase/heterodisulfide reductase	COG1908	pf02662	Ramos et al. 2015	
HdrF		WP_011878115.1	Multidomain heterodisulfide reductase-like protein	COG0247	pf10418	Strittmatter et al. 2009	

## References

- Appia-Ayme, C., & Berks, B. C. (2002). SoxV, an orthologue of the CcdA disulfide transporter, is involved in thiosulfate oxidation in *Rhodovulum sulfidophilum* and reduces the periplasmic thioredoxin SoxW. *Biochemical and Biophysical Research Communications*, 296(3), 737-741. [https://doi.org/10.1016/S0006-291X\(02\)00936-1](https://doi.org/10.1016/S0006-291X(02)00936-1)
- Arieli, B., Shahak, Y., Taglicht, D., Hauska, G., & Padan, E. (1994). Purification and characterization of sulfide-quinone reductase, a novel enzyme driving anoxygenic photosynthesis in *Oscillatoria limnetica*. *Journal of Biological Chemistry*, 269(8), 5705-5711.
- Bardischewsky, F., Fischer, J., Holler, B., & Friedrich, C. G. (2006). SoxV transfers electrons to the periplasm of *Paracoccus pantotrophus* - an essential reaction for chemotrophic sulfur oxidation. *Microbiology*, 152(Pt 2), 465-472. <https://doi.org/10.1099/mic.0.28523-0>
- Bardischewsky, F., Quentmeier, A., & Friedrich, C. G. (2006). The flavoprotein SoxF functions in chemotrophic thiosulfate oxidation of *Paracoccus pantotrophus* in vivo and in vitro. *FEMS Microbiology Letters*, 258(1), 121-126. <https://doi.org/10.1111/j.1574-6968.2006.00210.x>
- Brito, J. A., Sousa, F. L., Stelter, M., Bandejas, T. M., Vonrhein, C., Teixeira, M., Pereira, M. M., & Archer, M. (2009). Structural and functional insights into sulfide:quinone oxidoreductase. *Biochemistry*, 48(24), 5613-5622. <https://doi.org/10.1021/bi9003827>
- Cao, X., Koch, T., Steffens, L., Finkensieper, J., Zigann, R., Cronan, J. E., & Dahl, C. (2018). Lipoate-binding proteins and specific lipoate-protein ligases in microbial sulfur oxidation reveal an atypical role for an old cofactor. *Elife*, 7, e37439. <https://doi.org/10.7554/eLife.37439>
- Chan, L. K., Morgan-Kiss, R., & Hanson, T. E. (2009). Functional analysis of three sulfide:quinone oxidoreductase homologs in *Chlorobaculum tepidum*. *Journal of Bacteriology*, 191(3), 1026-1034. <https://doi.org/10.1128/JB.01154-08>
- Cort, J. R., Selan, U. M., Schulte, A., Grimm, F., Kennedy, M. A., & Dahl, C. (2008). *Allochroatrium vinosum* mDsrC: solution-state NMR structure, redox properties and interaction with DsrEFH, a protein essential for purple sulfur bacterial sulfur oxidation. *Journal of Molecular Biology*, 382(3), 692-707. <https://doi.org/10.1016/j.jmb.2008.07.022>
- Dahl, C., Engels, S., Pott-Sperling, A. S., Schulte, A., Sander, J., Lübke, Y., Deuster, O., & Brune, D. C. (2005). Novel genes of the dsr gene cluster and evidence for close interaction of Dsr proteins during sulfur oxidation in the phototrophic sulfur bacterium *Allochroatrium vinosum*. *Journal of Bacteriology*, 187(4), 1392-1404. <https://doi.org/10.1128/JB.187.4.1392-1404.2005>
- Dahl, C., Schulte, A., Stockdreher, Y., Hong, C., Grimm, F., Sander, J., Kim, R., Kim, S. H., & Shin, D. H. (2008). Structural and molecular genetic insight into a wide-spread bacterial sulfur oxidation pathway. *Journal of Molecular Biology*, 384, 1287-1300. <https://doi.org/10.1016/j.jmb.2008.10.016>
- Denkmann, K., Grein, F., Zigann, R., Siemen, A., Bergmann, J., van Helmont, S., Nicolai, A., Pereira, I. A. C., & Dahl, C. (2012). Thiosulfate dehydrogenase: a wide-spread unusual acidophilic cytochrome. *Environmental Microbiology*, 14, 2673-2688. <https://doi.org/10.1111/j.1462-2920.2012.02820.x>
- Dolata, M. M., van Beumen, J. J., Ambler, R. P., Meyer, T. E., & Cusanovich, M. A. (1993). Nucleotide sequence of the heme subunit of flavocytochrome c from the purple phototrophic bacterium, *Chromatium vinosum*. A 2.6-kilobase pair DNA fragment contains two multiheme cytochromes, a flavoprotein, and a homolog of human ankyrin. *Journal of Biological Chemistry*, 268(1), 109-114.
- Dolla, A., Pohorelic, B. K. J., Voordouw, J. K., & Voordouw, G. (2000). Deletion of the *hmc* operon of *Desulfotribium vulgaris* subsp. *vulgaris* Hildenborough hampers hydrogen metabolism and low-redox-potential niche establishment. *Archives of Microbiology*, 174(3), 143-151. <https://doi.org/10.1007/s002030000183>
- Duarte, A. G., Santos, A. A., & Pereira, I. A. (2016). Electron transfer between the QmoABC membrane complex and adenosine 5'-phosphosulfate reductase. *Biochimica et Biophysica Acta*, 1857(4), 380-386. <https://doi.org/10.1016/j.bbabi.2016.01.001>
- Ferreira, D., Barbosa, A. C. C., Oliveira, G. P., Catarino, T., Venceslau, S. S., & Pereira, I. A. C. (2022). The DsrD functional marker protein is an allosteric activator of the DsrAB dissimilatory sulfite reductase. *Proceedings of the National Academy of Sciences of the United States of America*, 119(4), e2118880119. <https://doi.org/10.1073/pnas.2118880119>
- Friedrich, C. G., Bardischewsky, F., Rother, D., Quentmeier, A., & Fischer, J. (2005). Prokaryotic sulfur oxidation. *Current Opinion in Microbiology*, 8(3), 253-259. <https://doi.org/10.1016/j.cmi.2005.04.005>

- Friedrich, C. G., Quentmeier, A., Bardischewsky, F., Rother, D., Orawski, G., Hellwig, P., & Fischer, J. (2008). Redox control of chemotrophic sulfur oxidation of *Paracoccus pantotrophus*. In C. Dahl & C. G. Friedrich (Eds.), *Microbial sulfur metabolism* (pp. 139–150). Berlin Heidelberg: Springer.
- Gavel, O. Y., Bursakov, S. A., Calvete, J. J., George, G. N., Moura, J. J. G., & Moura, I. (1998). ATP sulfurylases from sulfate-reducing bacteria of the genus *Desulfovibrio*. A novel metalloprotein containing cobalt and zinc. *Biochemistry*, 37(46), 16225–16232. <https://doi.org/10.1021/bi9816709>
- Grein, F., Pereira, I. A., & Dahl, C. (2010). Biochemical characterization of individual components of the *Allochrochromatium vinosum* DsrMKJOP transmembrane complex aids understanding of complex function in vivo. *Journal of Bacteriology*, 192(24), 6369–6377. <https://doi.org/10.1128/JB.00849-10>
- Grein, F., Venceslau, S. S., Schneider, L., Hildebrandt, P., Todorovic, S., Pereira, I. A. C., & Dahl, C. (2010). DsrJ, an essential part of the DsrMKJOP complex in the purple sulfur bacterium *Allochrochromatium vinosum*, is an unusual trimeric cytochrome c. *Biochemistry*, 49(38), 8290–8299. <https://doi.org/10.1021/bi1007673>
- Grimm, F., Cort, J. R., & Dahl, C. (2010). DsrR, a novel IscA-like protein lacking iron- and FeS-binding function involved in the regulation of sulfur oxidation in *Allochrochromatium vinosum*. *Journal of Bacteriology*, 192(6), 1652–1661. <https://doi.org/10.1128/JB.01269-09>
- Grimm, F., Franz, B., & Dahl, C. (2011). Regulation of dissimilatory sulfur oxidation in the purple sulfur bacterium *Allochrochromatium vinosum*. *Frontiers in Microbiology*, 2, 51. <https://doi.org/10.3389/fmicb.2011.00051>
- Hanna, E., MacRae, I. J., Medina, D. C., Fisher, A. J., & Segel, I. H. (2002). ATP sulfurylase from the hyperthermophilic chemolithotroph *Aquifex aeolicus*. *Archives of Biochemistry and Biophysics*, 406(2), 275–288. [https://doi.org/10.1016/S0003-9861\(02\)00428-9](https://doi.org/10.1016/S0003-9861(02)00428-9)
- Hausmann, B., Pelikan, C., Herbold, C. W., Kostlbacher, S., Alberts, M., Eichorst, S. A., Glavina Del Rio, T., Huemer, M., Nielsen, P. H., Ratte, T., Stingl, U., Tringe, S. G., Trojan, D., Wentrup, C., Wobcken, D., Pester, M., & Loy, A. (2018). Peatland Acidobacteria with a dissimilatory sulfur metabolism. *ISME Journal*, 12, 1729–1742. <https://doi.org/10.1038/s41396-018-0077-1>
- Hensel, M., Hinsley, A. P., Nikolaus, T., Sawers, G., & Berks, B. C. (1999). The genetic basis of tetrathionate respiration in *Salmonella typhimurium*. *Molecular Microbiology*, 32(2), 275–287. <https://doi.org/10.1046/j.1365-2958.1999.01345.x>
- Hermann, B., Kern, M., La Pietra, L., Simon, J., & Einsle, O. (2015). The octahaem MccA is a haem c-copper sulfite reductase. *Nature*, 520(7549), 706–709. <https://doi.org/10.1038/nature14109>
- Hildebrandt, T. M., & Grieshaber, M. K. (2008). Three enzymatic activities catalyze the oxidation of sulfide to thiosulfate in mammalian and invertebrate mitochondria. *FEBS Journal*, 275(13), 3352–3361. <https://doi.org/10.1111/j.1742-4658.2008.06482.x>
- Holkenbrink, C., Ocon Barbas, S., Møllerup, A., & Frigaard, N. U. (2011). Sulfur globule oxidation in green sulfur bacteria is dependent on the dissimilatory sulfite reductase system. *Microbiology*, 157, 1229–1239. <https://doi.org/10.1099/mic.0.044669-0>
- Huang, C. J., & Barrett, E. L. (1991). Sequence analysis and expression of the *Salmonella typhimurium* *asr* operon encoding production of hydrogen sulfide from sulfite. *Journal of Bacteriology*, 173(4), 1544–1553. <https://doi.org/10.1128/jb.173.4.1544-1553.1991>
- Ikeuchi, Y., Shigi, N., Kato, J., Nishimura, A., & Suzuki, T. (2006). Mechanistic insights into sulfur relay by multiple sulfur mediators involved in thiouridine biosynthesis at tRNA wobble positions. *Molecular Cell*, 21(1), 97–108. <https://doi.org/10.1016/j.molcel.2005.11.001>
- Jin, J. Q., Hachisuka, S. I., Sato, T., Fujiwara, T., & Atomi, H. (2020). A structurally novel lipoyl synthase in the hyperthermophilic archaeon *Thermococcus kodakarensis*. *Applied and Environmental Microbiology*, 86(23). <https://doi.org/10.1128/AEM.01359-20>
- Kanao, T., Kamimura, K., & Sugio, T. (2007). Identification of a gene encoding a tetrathionate hydrolase in *Acidithiobacillus ferrooxidans*. *Journal of Biotechnology*, 132(1), 16–22. <https://doi.org/10.1016/j.jbiotec.2007.08.030>
- Kappler, U., Bennett, B., Rethmeier, J., Schwarz, G., Deutzmann, R., McEwan, A. G., & Dahl, C. (2000). Sulfite:cytochrome c oxidoreductase from *Thiobacillus novellus* - Purification, characterization, and molecular biology of a heterodimeric member of the sulfite oxidase family. *Journal of Biological Chemistry*, 275(18), 13202–13212.
- Kappler, U., & Enemark, J. H. (2014). Sulfite-oxidizing enzymes. *Journal of Biological Inorganic Chemistry*, 20(2), 253–264. <https://doi.org/10.1007/s00775-014-1197-3>
- Kaster, A. K., Goenrich, M., Seedorf, H., Liesegang, H., Wollherr, A., Gottschalk, G., & Thauer, R. K. (2011). More than 200 genes required for methane formation from H<sub>2</sub> and CO<sub>2</sub> and energy conservation are present in *Methanothermobacter marburgensis* and *Methanothermobacter thermautotrophicus*. *Archaea*, 2011, 973848. <https://doi.org/10.1155/2011/973848>
- Kletzin, A. (1992). Molecular characterization of the *dsr* gene, which encodes the sulfur oxygenase/reductase of the thermoacidophilic archaeum *Desulfurolobus ambivalens*. *Journal of Bacteriology*, 174(18), 5854–5859. <https://doi.org/10.1128/jb.174.18.5854-5859.1992>
- Kern, M., Klotz, M. G., & Simon, J. (2011). The *Wolinella succinogenes* *mcc* gene cluster encodes an unconventional respiratory sulphite reduction system. *Molecular Microbiology*, 82(6), 1515–1530. <https://doi.org/10.1111/j.1365-2958.2011.07906.x>
- Koch, T., & Dahl, C. (2018). A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds. *ISME Journal*, 12(10), 2479–2491. <https://doi.org/10.1038/s41396-018-0209-7>
- Krafft, T., Gross, R., & Kröger, A. (1995). The function of *Wolinella succinogenes* *psr* genes in electron transport with polysulphide as the terminal electron acceptor. *European Journal of Biochemistry*, 230(2), 601–606.
- Kurth, J. M., Brito, J. A., Reuter, J., Flegler, A., Koch, T., Franke, T., Klein, E. M., Rowe, S. F., Butt, J. N., Denkmann, K., Pereira, I. A. C., Archer, M., & Dahl, C. (2016). Electron accepting units of the diheme cytochrome c TsdA, a bifunctional thiosulfate dehydrogenase/tetrathionate reductase. *Journal of Biological Chemistry*, 291(48), 24804–24818.
- Lahiri, C., Mandal, S., Ghosh, W., Dam, B., & Roy, P. (2006). A novel gene clusters *soxSRT* is essential for the chemolithotrophic oxidation of thiosulfate and tetrathionate by *Pseudomonas salicylatoloxidans* KCT001. *Current Microbiology*, 52(4), 267–273. <https://doi.org/10.1007/s00284-005-0176-x>
- Laska, S., Lottspeich, F., & Kletzin, A. (2003). Membrane-bound hydrogenase and sulfur reductase of the hyperthermophilic and acidophilic archaeon *Acidianus ambivalens*. *Microbiology*, 149(Pt 9), 2357–2371. <https://doi.org/10.1099/mic.0.26455-0>
- Laue, B., & Nelson, D. C. (1994). Characterization of the gene encoding the autotrophic ATP sulfurylase from the bacterial endosymbiont of the hydrothermal vent tubeworm *Riftia pachyptila*. *Journal of Bacteriology*, 176(12), 3723–3729. <https://doi.org/10.1128/jb.176.12.3723-3729.1994>
- Lin, J., Peng, T., Jiang, L., Ni, J. Z., Liu, Q., Chen, L., & Zhang, Y. (2015). Comparative genomics reveals new candidate genes involved in selenium metabolism in prokaryotes. *Genome Biology and Evolution*, 7(3), 664–676. <https://doi.org/10.1093/gbe/evw022>
- Liu, H., Xin, Y., & Xun, L. (2014). Distribution, diversity, and activities of sulfur dioxygenases in heterotrophic bacteria. *Applied and Environmental Microbiology*, 80(5), 1799–1806. <https://doi.org/10.1128/AEM.03281-13>
- Liu, L. J., Stockdreher, Y., Koch, T., Sun, S. T., Fan, Z., Josten, M., Sahl, H. G., Wang, Q., Luo, Y. M., Liu, S. J., Dahl, C., & Jiang, C. Y. (2014). Thiosulfate transfer mediated by DsrE/TusA homologs from acidothermophilic sulfur-oxidizing archaeon *Metallosphaera cuprina*. *Journal of Biological Chemistry*, 289(39), 26949–26959. <https://doi.org/10.1074/jbc.M114.591669>
- Löffler, M., Feldhues, J., Venceslau, S. S., Kammler, L., Grein, F., Pereira, I. A. C., & Dahl, C. (2020). DsrL mediates electron transfer between NADH and rDsrAB in *Allochrochromatium vinosum*. *Environmental Microbiology*, 22(2), 783–795. <https://doi.org/10.1111/1462-2920.14899>
- Löffler, M., Wallerang, K., Venceslau, S. S., Pereira, I. A. C., & Dahl, C. (2020). The iron-sulfur flavoprotein DsrL as NAD(P)H:acceptor oxidoreductase in oxidative and reductive dissimilatory sulfur metabolism. *Frontiers in Microbiology*, 11, 578209. <https://doi.org/10.3389/fmicb.2020.578209>
- Lübbe, Y. J., Youn, H. S., Timkovich, R., & Dahl, C. (2006). Siro(haem)amide in *Allochrochromatium vinosum* and relevance of DsrL and DsrN, a homolog of cobyrinic acid *c* diamide synthase for sulphur oxidation. *FEMS Microbiology Letters*, 261(2), 194–202. <https://doi.org/10.1111/j.1574-6968.2006.00343.x>
- Mander, G. J., Pierik, A. J., Huber, H., & Hedderich, R. (2004). Two distinct heterodisulfide reductase-like enzymes in the sulfate-reducing archaeon *Archaeoglobus profundus*. *European Journal of Biochemistry*, 271(6), 1106–1116. <https://doi.org/10.1111/j.1432-1033.2004.04013.x>
- Marcia, M., Ermler, U., Peng, G. H., & Michel, H. (2010). A new structure-based classification of sulfide:quinone oxidoreductases. *Proteins: Structure Function and Bioinformatics*, 78(5), 1073–1083. <https://doi.org/10.1002/prot.22665>
- Matthews, J. C., Timkovich, R., Liu, M. Y., & LeGall, J. (1995). Siroamide: A prosthetic group isolated from sulfite reductases in the genus *Desulfovibrio*. *Biochemistry*, 34(15), 5248–5251. <https://doi.org/10.1021/bi00015a039>
- Meyer, B., & Kuever, J. (2007a). Molecular analysis of the distribution and phylogeny of dissimilatory adenosine-5'-phosphosulfate reductase-encoding genes (*apsrA*) among sulfur-oxidizing prokaryotes. *Microbiology*, 153(Pt 10), 3478–3498. <https://doi.org/10.1099/mic.0.2007/008250-0>
- Meyer, B., & Kuever, J. (2007b). Phylogeny of the alpha and beta subunits of the dissimilatory adenosine-5'-phosphosulfate (APS) reductase from sulfate-reducing prokaryotes - origin and evolution of the dissimilatory sulfate-reduction pathway. *Microbiology*, 153(Pt 7), 2026–2044. <https://doi.org/10.1099/mic.0.2006/003152-0>
- Müller, F. H., Bandejas, T. M., Ulrich, T., Teixeira, M., & Kletzin, A. (2004). Coupling of the pathway of sulphur oxidation to dioxygen reduction: characterization of a novel membrane-bound thiosulphate:quinone oxidoreductase. *Molecular Microbiology*, 53(4), 1147–1160. <https://doi.org/10.1111/j.1365-2958.2004.04193.x>
- Numata, T., Fukai, S., Ikeuchi, Y., Suzuki, T., & Nureki, O. (2006). Structural basis for sulfur relay to RNA mediated by heterohexameric TusBCD complex. *Structure*, 14(2), 357–366. <https://doi.org/10.1016/j.str.2005.11.009>
- Okamura-Ikeda, K., Ohmura, Y., Fujiwara, K., & Motokawa, Y. (1993). Cloning and nucleotide sequence of the *gcv* operon encoding the *Escherichia coli* glycine-cleavage system. *European Journal of Biochemistry*, 216(2), 539–548. <https://doi.org/10.1111/j.1432-1033.1993.tb18172.x>
- Parey, K., Demmer, U., Warkentin, E., Wynen, A., Ermler, U., & Dahl, C. (2013). Structural, biochemical and genetic characterization of ATP sulfurylase from *Allochrochromatium vinosum*. *PLoS One*, 8(9), e74707. <https://doi.org/10.1371/journal.pone.0074707>
- Pereira, P. M., Teixeira, M., Xavier, A. V., Louro, R. O., & Pereira, I. A. C. (2006). The Tmc complex from *Desulfovibrio vulgaris* Hildenborough is involved in transmembrane electron transfer from periplasmic hydrogen oxidation. *Biochemistry*, 45(34), 10359–10367. <https://doi.org/10.1021/bi0610294>
- Pires, R. H., Venceslau, S. S., Morais, F., Teixeira, M., Xavier, A. V., & Pereira, I. A. C. (2006). Characterization of the *Desulfovibrio desulfuricans* ATCC 27774 DsrMKJOP complex - a membrane-bound redox complex involved in the sulfate respiratory pathway. *Biochemistry*, 45(1), 249–262. <https://doi.org/10.1021/bi0515265>
- Pott, A. S., & Dahl, C. (1998). Sirohaem-sulfite reductase and other proteins encoded in the *dsr* locus of *Chromatium vinosum* are involved in the oxidation of intracellular sulfur. *Microbiology*, 144, 1881–1894. <https://doi.org/10.1099/00221287-144-7-1881>

- Pyne, P., Alam, M., & Ghosh, W. (2017). A novel soxO gene, encoding a glutathione disulfide reductase, is essential for tetrathionate oxidation in *Advenella kashmirensis*. *Microbiological Research*, 205, 1-7. <https://doi.org/10.1016/j.micres.2017.08.002>
- Ramos, A. R., Grein, F., Oliveira, G. P., Venceslau, S. S., Keller, K. L., Wall, J. D., & Pereira, I. A. (2015). The FlxABC-D-HdrABC proteins correspond to a novel NADH dehydrogenase/heterodisulfide reductase widespread in anaerobic bacteria and involved in ethanol metabolism in *Desulfovibrio vulgaris* Hildenborough. *Environmental Microbiology*, 17(11), 3233-3244. <https://doi.org/10.1111/1365-3113.12403>
- Ramos, A. R., Keller, K. L., Wall, J. D., & Pereira, I. A. C. (2012). The membrane QmoABC complex interacts directly with the dissimilatory adenosine 5'-phosphosulfate reductase in sulfate reducing bacteria. *Frontiers in Microbiology*, 3, 137. <https://doi.org/10.3389/fmicb.2012.00137>
- Reinartz, M., Tschäpe, J., Brüser, T., Trüper, H. G., & Dahl, C. (1998). Sulfide oxidation in the phototrophic sulfur bacterium *Chromatium vinosum*. *Archives of Microbiology*, 170(1), 59-68. <https://doi.org/10.1007/s002030050615>
- Rossi, M., Pollock, W. B. R., Reiji, M. W., Keon, R. G., Fu, R., & Voordouw, G. (1993). The *hmc* operon of *Desulfovibrio vulgaris* subsp. *vulgaris* Hildenborough encodes a potential transmembrane redox protein complex. *Journal of Bacteriology*, 175(15), 4699-4711. <https://doi.org/10.1128/jb.175.15.4699-4711.1993>
- Rother, D., Heinrich, H. J., Quentmeier, A., Bardischewsky, F., & Friedrich, C. G. (2001). Novel genes of the *sox* gene cluster, mutagenesis of the flavoprotein SoxF, and evidence for a general sulfur-oxidizing system in *Paracoccus pantotrophus* GB17. *Journal of Bacteriology*, 183(15), 4499-4508. <https://doi.org/10.1128/JB.183.15.4499-4508.2001>
- Rother, D., Orawski, G., Bardischewsky, F., & Friedrich, C. G. (2005). SoxRS-mediated regulation of chemotrophic sulfur oxidation in *Paracoccus pantotrophus*. *Microbiology*, 151(5), 1707-1716. <https://doi.org/10.1099/mic.0.27724-0>
- Rother, D., Ringk, J., & Friedrich, C. G. (2008). Sulfur oxidation of *Paracoccus pantotrophus*: The sulfur-binding protein SoxYZ is the target of the periplasmic thiol-disulfide oxidoreductase SoxS. *Microbiology*, 154(7), 1980-1988. <https://doi.org/10.1099/mic.0.2008/018655-0>
- Santos, A. A., Venceslau, S. S., Grein, F., Leavitt, W. D., Dahl, C., Johnston, D. T., & Pereira, I. A. (2015). A protein trisulfide couples dissimilatory sulfate reduction to energy conservation. *Science*, 350(6267), 1541-1545. <https://doi.org/10.1126/science.1263558>
- Smeulders, M. J., Pol, A., Venselaar, H., Barends, T. R., Hermans, J., Jetten, M. S., & Op den Camp, H. J. (2013). Bacterial C<sub>5</sub> hydrolases from *Acidithiobacillus thiooxidans* strains are homologous to the archaeal catenane C<sub>5</sub> hydrolase. *Journal of Bacteriology*, 195(18), 4046-4056. <https://doi.org/10.1128/JB.00627-13>
- Stockdreher, Y., Sturm, M., Josten, M., Sahl, H. G., Dobler, N., Zigann, R., & Dahl, C. (2014). New proteins involved in sulfur trafficking in the cytoplasm of *Allochromatium vinosum*. *Journal of Biological Chemistry*, 289(18), 12390-12403. <https://doi.org/10.1074/jbc.M113.536425>
- Stoffels, L., Krehenbrink, M., Berks, B. C., & Udden, G. (2012). Thiosulfate reduction in *Salmonella enterica* is driven by the proton motive force. *Journal of Bacteriology*, 194(2), 475-485. <https://doi.org/10.1128/JB.06014-11>
- Strittmatter, A. W., Liesegang, H., Rabus, R., Decker, I., Amann, J., Andres, S., Henne, A., Fricke, W. F., Martinez-Arias, R., Bartels, D., Goesmann, A., Krause, L., Puhler, A., Klenk, H. P., Richter, M., Schuler, M., Glockner, F. O., Meyerdieck, A., Gottschalk, G., & Amann, R. (2009). Genome sequence of *Desulfobacterium autotrophicum* HRM2, a marine sulfate reducer. *Genome Announcements*, 1(1), 1-5. <https://doi.org/10.1128/genomea.00001-09>
- Tanabe, T. S., Leimkühler, S., & Dahl, C. (2019). The functional diversity of the prokaryotic sulfur carrier protein TusA. *Advances in Microbial Physiology*, 75, 233-277. <https://doi.org/10.1016/bs.ampbs.2019.07.004>
- Tanaka, Y., Yoshikawa, K., Takeuchi, A., Ichikawa, M., Mori, T., Uchino, S., Sugano, Y., Hakoshima, T., Takagi, H., Nonaka, G., & Tsukazaki, T. (2020). Crystal structure of a YeeE/YedE family protein engaged in thiosulfate uptake. *Science Advances*, 6(35), eaba7637. <https://doi.org/10.1126/sciadv.aba7637>
- Vande Weghe, J. G., & Ow, D. W. (1999). A fission yeast gene for mitochondrial sulfide oxidation. *Journal of Biological Chemistry*, 274(19), 13250-13257. <https://doi.org/10.1074/jbc.274.19.13250>
- Venceslau, S. S., Lino, R. R., & Pereira, I. A. C. (2010). The Qrc membrane complex, related to the alternative complex III, is a menaquinone reductase involved in sulfate respiration. *Journal of Biological Chemistry*, 285(30), 22772-22781. <https://doi.org/10.1074/jbc.M110.124305>
- Wolfe, B. M., Lui, S. M., & Cowan, J. A. (1994). Desulfoviridin, a multimeric-dissimilatory sulfite reductase from *Desulfovibrio vulgaris* (Hildenborough). Purification, characterization, kinetics and EPR studies. *European Journal of Biochemistry*, 223(1), 79-89. <https://doi.org/10.1111/j.1432-1033.1994.tb18968.x>
- Zhang, Y., Turanov, A. A., Hatfield, D. L., & Gladyshev, V. N. (2008). In silico identification of genes involved in selenium metabolism: evidence for a third selenium utilization trait. *BMC Genomics*, 9, 251. <https://doi.org/10.1186/1471-2164-9-251>

Table S4. HMM evaluation by (cross-)validation

Model	10-fold cross validation on training dataset						Confusion matrix					
	Metrics						True		False		True	
	MCC	F1-score	Balanced accuracy	Cohens kappa	Precision	Recall	positives	False positives	False negatives	positives	negatives	
redDsrA	1	1	1	1	1	1	1520	0	0	305750		
redDsrB	1	1	1	1	1	1	1520	0	0	305690		
redDsrC	0,96046387	0,96026861	0,96627488	0,96003639	0,98961938	0,9326087	1716	18	124	305412		
DsrD	1	1	1	1	1	1	940	0	0	306350		
redDsrJ	1	1	1	1	1	1	1030	0	0	306180		
redDsrK	0,99655305	0,99656357	0,99657534	0,99654711	1	0,99315068	1450	0	10	303740		
redDsrM	0,95662617	0,95588235	0,99980375	0,95568645	0,91549296	1	1300	120	0	305620		
redDsrN	0,99239858	0,99240266	0,99292289	0,99233765	0,99904398	0,98584906	1045	1	15	305829		
redDsrO	0,94664603	0,94547093	0,99932791	0,94527393	0,89734816	0,99904762	1049	120	1	306160		
redDsrP	0,93920509	0,9377519	0,99834918	0,93752552	0,88503804	0,99714286	1047	136	3	305834		
DsrT	0,99457536	0,99457861	0,99460784	0,99456065	1	0,98921569	1009	0	11	306200		
redAprA	0,95102012	0,95046591	0,99424086	0,95030129	0,91501416	0,98877551	969	90	11	306250		
redAprB	0,94056592	0,94080997	0,97174413	0,94056132	0,9378882	0,94375	1208	80	72	305570		
redSat	0,96409328	0,96376812	0,96830168	0,96368637	0,99253731	0,93661972	665	5	45	305545		
QmoA	1	1	1	1	1	1	1530	0	0	305780		
QmoB	0,99673165	0,99674267	0,99675325	0,99672631	1	0,99350649	1530	0	10	305610		
QmoC	1	1	1	1	1	1	1010	0	0	305930		
qHdrB	0,99880719	0,9988081	0,99880952	0,99880648	1	0,99761905	419	0	1	306920		
qHdrC	0,56983714	0,49056604	0,99956056	0,49024057	0,325	1	130	270	0	306940		
mHdrA	0,99226149	0,99224806	0,99998348	0,99223155	0,98461538	1	640	10	0	302700		
mHdrB	0,97374673	0,97345133	0,99995087	0,97340223	0,94827586	1	550	30	0	305310		
mHdrC	0,97281373	0,97254207	0,999042	0,97249156	0,94818653	0,99818182	549	30	1	306660		
HdrF	0,96304998	0,96275072	0,99409954	0,96272933	0,93854749	0,98823529	168	11	2	303769		
FlxA	0,99085117	0,99082569	0,99998363	0,99080932	0,98181818	1	540	10	0	305460		
FlxB	0,9573489	0,95652174	0,99991831	0,9564402	0,91666667	1	550	50	0	306000		
FlxC	0,99886136	0,99886234	0,99886364	0,99886071	1	0,99772727	439	0	1	306490		
FlxD	0,89633704	0,89324619	0,98795027	0,89308589	0,82329317	0,97619048	410	88	10	303432		
AsrA	0,99761458	0,99761337	0,99761905	0,99761173	1	0,9952381	209	0	1	305830		
AsrB	0,99761457	0,99761337	0,99761905	0,99761172	1	0,9952381	209	0	1	304220		
AsrC	0,98318058	0,98305085	0,98333333	0,98303915	1	0,96666667	203	0	7	299230		
MccA	1	1	1	1	1	1	210	0	0	306870		
MccB	0,99761457	0,99761337	0,99761905	0,99761172	1	0,9952381	209	0	1	304350		
MccC	1	1	1	1	1	1	210	0	0	307110		
MccD	0,99060367	0,99056604	0,99999349	0,99055953	0,98130841	1	210	4	0	307126		
PsrAPhsASreA	0,93034913	0,93039648	0,97135187	0,93026757	0,91826087	0,94285714	528	47	32	306343		
PsrBPhsBSreB	0,86979145	0,86997194	0,93045116	0,86974524	0,87901701	0,86111111	465	64	75	306476		
PsrCPhsCSreC	0,93151503	0,92994924	0,93875824	0,92983711	0,98920086	0,87739464	458	5	64	306535		
FccA	0,99267019	0,99265477	0,99270833	0,99264333	1	0,98541667	473	0	7	305800		
FccB	0,93689181	0,93685625	0,97965801	0,93662595	0,91513561	0,95963303	1046	97	44	305883		
oxDsrA	1	1	1	1	1	1	690	0	0	306480		
oxDsrB	1	1	1	1	1	1	720	0	0	306540		
oxDsrC	0,97360942	0,97364771	0,9919903	0,97355398	0,96340348	0,98411215	1053	40	17	304010		
DsrE	1	1	1	1	1	1	700	0	0	306520		
DsrF	1	1	1	1	1	1	690	0	0	306470		
DsrH	1	1	1	1	1	1	700	0	0	306580		
oxDsrJ	0,98264908	0,98253607	0,98283582	0,98249858	1	0,96567164	647	0	23	306640		
oxDsrK	0,99524765	0,99524779	0,99527027	0,99523636	1	0,99054054	733	0	7	306130		
oxDsrM	0,91394165	0,91044776	0,91780822	0,91025358	1	0,83561644	610	0	120	306440		
oxDsrO	0,90652043	0,9037037	0,91779189	0,90349288	0,98387097	0,83561644	610	10	120	306200		
oxDsrP	0,9206726	0,91771683	0,9239726	0,9175369	1	0,84794521	619	0	111	306310		
DsrL	0,99329448	0,99328859	0,99333333	0,993272	1	0,98666667	740	0	10	301330		
DsrN	1	1	1	1	1	1	630	0	0	304560		
DsrR	1	1	1	1	1	1	420	0	0	304410		
DsrS	1	1	1	1	1	1	110	0	0	306950		
DsrE2	0,97263288	0,9725777	0,99257455	0,97255309	0,96028881	0,98518519	266	11	4	304769		
DsrE3A	0,9987912	0,99879373	0,99879518	0,99879046	1	0,99759036	828	0	2	306490		
DsrE3B	1	1	1	1	1	1	420	0	0	306920		
DsrE3C	0,97842824	0,97826087	0,9787234	0,97819563	1	0,95744681	900	0	40	306390		
DsrE4	0,97135067	0,97097625	0,97179487	0,97094045	1	0,94358974	368	0	22	306950		
DsrE5	0,77348159	0,77253219	0,90892966	0,77227401	0,73170732	0,81818182	270	99	60	306881		
AprM	1	1	1	1	1	1	320	0	0	307020		
oxAprAI	1	1	1	1	1	1	320	0	0	307020		
oxAprAll	0,99843465	0,99843505	0,9984375	0,99843343	1	0,996875	319	0	1	306940		
assAprA	1	1	1	1	1	1	160	0	0	307010		
oxAprBI	1	1	1	1	1	1	320	0	0	307020		
oxAprBII	0,97979983	0,97972973	0,98333008	0,97971019	0,99315068	0,96666667	290	2	10	306968		
assAprB	0,87520513	0,86757991	0,99990536	0,86748693	0,76612903	1	190	58	0	306382		
oxSat	0,93668798	0,93543759	0,9938724	0,93536426	0,88828338	0,98787879	326	41	4	305959		
TsdA	0,99330449	0,99328859	0,99333333	0,99328207	1	0,98666667	296	0	4	306780		
TsdB	0,97568581	0,97540984	0,976	0,97539031	1	0,952	238	0	12	307090		
SoeA	1	1	1	1	1	1	1100	0	0	305510		
SoeB	1	1	1	1	1	1	1100	0	0	305820		
SoeC	1	1	1	1	1	1	1120	0	0	304170		
SorA	0,94480475	0,9435337	0,99802804	0,94348333	0,89619377	0,99615385	259	30	1	306840		
SorB	0,96005396	0,95930233	0,99995438	0,95925678	0,92178771	1	330	28	0	306862		
SOR	1	1	1	1	1	1	240	0	0	307080		
SQRI	1	1	1	1	1	1	1200	0	0	306090		
SQRII	0,97880641	0,97859327	0,9999886	0,97858187	0,95808383	1	160	7	0	306883		
SQRIII	0,9631188	0,96262887	0,96397516	0,96243917	1	0,92795031	1494	0	116	305260		
SQRIV	0,99214866	0,99212598	0,9921875	0,99211784	1	0,984375	315	0	5	306990		
SQRV	1	1	1	1	1	1	840	0	0	306500		
SQRFVI	0,98333018	0,98324022	0,98351648	0,98319125	1	0,96703297	880	0	30	306210		
SDOI	0,99802294	0,99802761	0,99933884	0,99802272	0,99737188	0,99868421	759	2	1	306578		
SDOII	0,97941456	0,97926635	0,9796875	0,97920273	1	0,959375	921	0	39	306360		
SDOIII	0,92764368	0,92766794	0,98468945	0,92669656	0,888079	0,97095116	3777	476	113	302274		
SoxA	1	1	1	1	1	1	1480	0	0	305170		

SoxB	1	1	1	1	1	1	1120	0	0	304230
SoxC	1	1	1	1	1	1	500	0	0	306240
SoxD	0,9980417	0,99804305	0,99999673	0,99803979	0,99609375	1	510	2	0	306048
SoxE	1	1	1	1	1	1	100	0	0	307070
SoxF	0,88131972	0,88145049	0,93882208	0,88131198	0,88515406	0,87777778	316	41	44	306799
SoxG	0,9916284	0,99159664	0,99166667	0,99159336	1	0,98333333	118	0	2	304660
SoxH	0,99102307	0,99099099	0,99107143	0,99098278	1	0,98214286	275	0	5	304380
SoxO	0,99705284	0,99705015	0,99705882	0,9970485	1	0,99411765	169	0	1	302580
SoxS	0,88332255	0,87671233	0,8902439	0,87656778	1	0,7804878	320	0	90	306540
SoxT1	0,96996419	0,96969697	0,99726305	0,96965948	0,94601542	0,99459459	368	21	2	306539
SoxT2	0,89648005	0,8957346	0,92952427	0,89566304	0,93564356	0,85909091	189	13	31	306867
SoxV	0,96859205	0,96813354	0,96911765	0,96809907	1	0,93823529	319	0	21	304370
SoxW	0,98909435	0,98909091	0,99754949	0,98907614	0,98313253	0,99512195	408	7	2	304643
SoxX	0,99843137	0,99843505	0,9984375	0,99843014	1	0,996875	957	0	3	305180
SoxY	1	1	1	1	1	1	1580	0	0	304700
SoxZ	1	1	1	1	1	1	1590	0	0	304980
DoxA	1	1	1	1	1	1	90	0	0	307230
DoxD	0,98838936	0,98832685	0,98846154	0,98832196	1	0,97692308	127	0	3	306940
Tth	0,9857789	0,98569571	0,98589744	0,98567779	1	0,97179487	379	0	11	306940
TtrA	0,94598288	0,94473684	0,99854422	0,9446685	0,8975	0,99722222	359	41	1	306419
TtrB	0,88875846	0,88888889	0,94437923	0,88875846	0,88888889	0,88888889	320	40	40	306650
TtrC	0,94672305	0,94565217	0,99708094	0,94558715	0,9015544	0,99428571	348	38	2	306802
TtrR	0,94470972	0,944	0,99162577	0,94395236	0,90769231	0,98333333	236	24	4	293396
TtrS	0,90516163	0,90196078	0,99278415	0,90188313	0,8313253	0,98571429	207	42	3	287638
GerGerRed	1	1	1	1	1	1	300	0	0	304180
LpIA	0,9982711	0,99827288	0,99827586	0,99826961	1	0,99655172	578	0	2	305260
sHdRj	0,99544254	0,99543379	0,99545455	0,99543215	1	0,99090909	109	0	1	305700
sHdrl	1	1	1	1	1	1	180	0	0	307020
LipS1	1	1	1	1	1	1	490	0	0	306490
LipS2	0,99736169	0,99736148	0,99736842	0,99735821	1	0,99473684	378	0	2	306450
GcvH	0,99432947	0,99438422	0,99441558	0,99431339	1	0,98883117	3807	0	43	303490
LbpA1	0,98843548	0,98840885	0,99891985	0,9883909	0,97912317	0,99787234	469	10	1	306280
LbpA2	0,99912079	0,99912204	0,99912281	0,99912041	1	0,99824561	569	0	1	306560
DHDL	0,99582297	0,9958159	0,99583333	0,99581425	1	0,99166667	119	0	1	302600
sEtFA	0,99372872	0,99371069	0,99375	0,99370905	1	0,9875	79	0	1	305280
sEtFB	1	1	1	1	1	1	70	0	0	305470
sHdRA	0,99852507	0,99852725	0,99852941	0,99852398	1	0,99705882	678	0	2	306580
sHdRB1	1	1	1	1	1	1	670	0	0	306570
sHdRB2	1	1	1	1	1	1	600	0	0	306640
sHdRC1	1	1	1	1	1	1	680	0	0	306620
sHdRC2	1	1	1	1	1	1	590	0	0	306680
sHdRCG1	0,98772638	0,98765432	0,99999675	0,98765107	0,97560976	1	80	2	0	307258
sHdRCG2	0,94599201	0,94520548	0,99284574	0,94519247	0,90789474	0,98571429	69	7	1	306843
sHdRH	1	1	1	1	1	1	660	0	0	306610
sHdRT	1	1	1	1	1	1	250	0	0	304680
sHdRSoxR	1	1	1	1	1	1	520	0	0	301120
Cs2H	1	1	1	1	1	1	130	0	0	307210
McrA	1	1	1	1	1	1	280	0	0	307050
McrB	1	1	1	1	1	1	270	0	0	307070
McrC	1	1	1	1	1	1	170	0	0	307130
McrD	1	1	1	1	1	1	250	0	0	307070
McrG	1	1	1	1	1	1	260	0	0	307080
TusA	0,98486585	0,98508791	0,99959213	0,98475365	0,97075596	0,9998494	6639	200	1	300490
TusADsrE	0,9447861	0,94348894	0,94651163	0,94326426	1	0,89302326	1152	0	138	306050
TusB	0,98056471	0,98039216	0,99998372	0,98037588	0,96153846	1	250	10	0	307080
TusC	0,98184036	0,98175182	0,99813349	0,98173555	0,9676259	0,9962963	269	9	1	307061
TusD	0,97370679	0,97338403	0,97407407	0,97336125	1	0,94814815	256	0	14	307070
TusE	0,94670075	0,94551845	0,9980993	0,94546811	0,89966555	0,9962963	269	30	1	307040
YeeE	0,95499133	0,95460123	0,96308542	0,95454073	0,98481013	0,92619048	389	6	31	305494
HmcA	0,98559505	0,98550725	0,99998406	0,98549131	0,97142857	1	340	10	0	313610
HmcB	0,99864979	0,99865047	0,99999841	0,99864888	0,99730458	1	370	1	0	313589
HmcC	0,99864614	0,99864682	0,99864865	0,99864523	1	0,9972973	369	0	1	313590
HmcD	1	1	1	1	1	1	250	0	0	313710
HmcE	0,99856881	0,99856938	0,99857143	0,99856779	1	0,99714286	349	0	1	313610
HmcF	0,94577549	0,94451962	0,99850766	0,94445443	0,89717224	0,99714286	349	40	1	313570
MvhA	1	1	1	1	1	1	660	0	0	312990
MvhD	0,84476449	0,84485981	0,91845636	0,84472753	0,85283019	0,83703704	226	39	44	313651
MvhG	1	1	1	1	1	1	600	0	0	313140
QrcA	1	1	1	1	1	1	570	0	0	313390
QrcB	0,96920012	0,96925566	0,98303739	0,96919503	0,9724026	0,96612903	599	17	21	313323
QrcC	0,99916472	0,99916597	0,99916667	0,99916438	1	0,99833333	599	0	1	313360
QrcD	1	1	1	1	1	1	580	0	0	313380
TmcA	0,97784761	0,97763578	0,978125	0,97760231	1	0,95625	459	0	21	313480
TmcB	0,99068167	0,99065421	0,99998405	0,99063825	0,98148148	1	530	10	0	313420
TmcC	1	1	1	1	1	1	530	0	0	313430
TmcD	1	1	1	1	1	1	540	0	0	313420
mHdR	0,98626466	0,98617512	0,98636364	0,98617034	1	0,97272727	107	0	3	313850
mHdRE	0,96142771	0,96069869	0,99998566	0,96068437	0,92436975	1	110	9	0	313841

Validation on independent test dataset

Model	Metrics						Confusion matrix			
	MCC	F1-score	Balanced	Cohens	Precision	Recall	True positives	False	True	
			accuracy	kappa				positives		negatives
redDsrA	0,99287039	0,99285714	0,99998783	0,99284497	0,9858156	1	139	2	0	82165
redDsrB	0,97845638	0,97826087	0,99996349	0,97822438	0,95744681	1	135	6	0	82165
redDsrC	0,91659255	0,91324201	0,92016807	0,9131273		0,84033613	100	0	19	82187
DsrD	1	1	1	1	1	1	67	0	0	82239
redDsrJ	0,98358151	0,98347107	0,99997567	0,98344675	0,96747967	1	119	4	0	82183
redDsrK	0,99301919	0,99300699	0,99998783	0,99299482	0,98611111	1	142	2	0	82162
redDsrM	0,95703079	0,9561753	0,99993308	0,9561085	0,91603053	1	120	11	0	82175
redDsrN	0,98245473	0,98245614	0,98836601	0,9824379	0,98823529	0,97674419	84	1	2	82219
redDsrO	0,92039235	0,91735537	0,99987834	0,91723453	0,84732824	1	111	20	0	82175
redDsrP	0,93085873	0,92857143	0,99990268	0,9284746	0,86666667	1	104	16	0	82186
DsrT	1	1	1	1	1	1	103	0	0	82203
redAprA	0,9857295	0,98564593	0,99998175	0,98562769	0,97169811	1	103	3	0	82200
redAprB	0,97489152	0,97478992	0,99569608	0,97475342	0,95867769	0,99145299	116	5	1	82184
redSat	0,83052862	0,82242991	0,86064358	0,82231668	0,95652174	0,72131148	44	2	17	82243
QmoA	1	1	1	1	1	1	84	0	0	82222
QmoB	0,99356302	0,99354839	0,99999392	0,99354231	0,98717949	1	77	1	0	82228
QmoC	0,98951647	0,98947368	0,99998784	0,98946152	0,97916667	1	94	2	0	82210
qHdrB	1	1	1	1	1	1	11	0	0	82295
qHdrC	1	1	1	1	1	1	15	0	0	82291
mHdrA	1	1	1	1	1	1	98	0	0	82208
mHdrB	0,99375731	0,99375	0,99998783	0,99373783	0,98757764	1	159	2	0	82145
mHdrC	0,99725795	0,99726027	0,99999391	0,99725419	0,99453552	1	182	1	0	82123
HdrF	0,98771758	0,98765432	0,99998784	0,98764216	0,97560976	1	80	2	0	82224
FlxA	1	1	1	1	1	1	23	0	0	82283
FlxB	1	1	1	1	1	1	23	0	0	82283
FlxC	1	1	1	1	1	1	33	0	0	82273
FlxD	1	1	1	1	1	1	46	0	0	82260
AsrA	1	1	1	1	1	1	25	0	0	82281
AsrB	1	1	1	1	1	1	25	0	0	82281
AsrC	1	1	1	1	1	1	25	0	0	82281
MccA	1	1	1	1	1	1	12	0	0	82294
MccB	0,92580885	0,92307692	0,99998785	0,92306484	0,85714286	1	12	2	0	82292
MccC	1	1	1	1	1	1	12	0	0	82294
MccD	1	1	1	1	1	1	12	0	0	82294
PsrAPhsASreA	0,8164817	0,8	0,99998177	0,7999825	0,66666667	1	6	3	0	82297
PsrBPhsBSreB	0,47134437	0,36363636	0,99987242	0,36356044	0,22222222	1	6	21	0	82279
PsrCPhsCSreC	0,8164817	0,8	0,99998177	0,7999825	0,66666667	1	6	3	0	82297
FccA	0,93674277	0,93478261	0,99996353	0,93474629	0,87755102	1	43	6	0	82257
FccB	0,95894252	0,95867769	0,96773586	0,95864731	0,98305085	0,93548387	58	1	4	82243
oxDsrA	0,99069145	0,99065421	0,99074074	0,99064813	1	0,98148148	53	0	1	82252
oxDsrB	0,9913358	0,99130435	0,99137931	0,99129827	1	0,98275862	57	0	1	82248
oxDsrC	0,92321475	0,92041522	0,99986005	0,92027614	0,8525641	1	133	23	0	82150
DsrE	1	1	1	1	1	1	69	0	0	82237
DsrF	1	1	1	1	1	1	63	0	0	82243
DsrH	1	1	1	1	1	1	71	0	0	82235
oxDsrJ	0,96074556	0,96	0,96153846	0,95997572	1	0,92307692	48	0	4	82254
oxDsrK	0,98471995	0,98461538	0,98484848	0,98460323	1	0,96969697	64	0	2	82240
oxDsrM	0,90995288	0,90598291	0,9140625	0,90591663	1	0,828125	53	0	11	82242
oxDsrO	0,8865569	0,88235294	0,90177964	0,8822807	0,97826087	0,80357143	45	1	11	82249
oxDsrP	0,89317293	0,89108911	0,91665451	0,89102257	0,95744681	0,83333333	45	2	9	82250
DsrL	0,99162562	0,99159664	0,99166667	0,99159056	1	0,98333333	59	0	1	82246
DsrN	0,98056875	0,98039216	0,99998784	0,98038	0,96153846	1	50	2	0	82254
DsrR	1	1	1	1	1	1	38	0	0	82268
DsrS	1	1	1	1	1	1	3	0	0	82303
DsrE2	0,95168583	0,95061728	0,99990262	0,95052013	0,90588235	1	154	16	0	82136
DsrE3A	0,97558235	0,97530864	0,99997568	0,97528433	0,95180723	1	79	4	0	82223
DsrE3B	1	1	1	1	1	1	14	0	0	82292
DsrE3C	1	1	1	1	1	1	8	0	0	82298
DsrE4	1	1	1	1	1	1	19	0	0	82287
DsrE5	1	1	1	1	1	1	25	0	0	82281
AprM	1	1	1	1	1	1	34	0	0	82272
oxAprAl	1	1	1	1	1	1	33	0	0	82273
oxAprAll	0,9813008	0,98113208	0,98148148	0,981126	1	0,96296296	26	0	1	82279
assAprA	0,93538592	0,93333333	0,99996961	0,93330308	0,875	1	35	5	0	82266
oxAprBl	0,93538592	0,93333333	0,9375	0,93330308	1	0,875	35	0	5	82266
oxAprBll	0,9813008	0,98113208	0,99999392	0,981126	0,96296296	1	26	1	0	82279
assAprB	1	1	1	1	1	1	115	0	0	82191
oxSat	0,76785973	0,75324675	0,96763862	0,75313566	0,63043478	0,93548387	29	17	2	82258
TsdA	0,98345798	0,98333333	0,99998784	0,98332118	0,96721311	1	59	2	0	82245
TsdB	1	1	1	1	1	1	47	0	0	82259
SoeA	0,99118323	0,99115044	0,99122807	0,99114436	1	0,98245614	56	0	1	82249
SoeB	0,9913358	0,99130435	0,99137931	0,99129827	1	0,98275862	57	0	1	82248
SoeC	0,99118323	0,99115044	0,99122807	0,99114436	1	0,98245614	56	0	1	82249
SorA	0,91282654	0,90909091	0,99995138	0,90904269	0,83333333	1	40	8	0	82258
SorB	0,87699405	0,86956522	0,99992707	0,86949352	0,76923077	1	40	12	0	82254
SOR	1	1	1	1	1	1	25	0	0	82281
SQRl	1	1	1	1	1	1	125	0	0	82181
SQRll	1	1	1	1	1	1	40	0	0	82266
SQRlll	1	1	1	1	1	1	243	0	0	82063
SQRlV	1	1	1	1	1	1	33	0	0	82273
SQRV	1	1	1	1	1	1	484	0	0	81822
SQRVI	1	1	1	1	1	1	67	0	0	82239
SDOI	0,96571885	0,96517413	0,99995743	0,9651316	0,93269231	1	97	7	0	82202
SDOII	1	1	1	1	1	1	117	0	0	82189
SDOIII	0,33615457	0,20503597	0,99462235	0,20305464	0,11422846	1	114	884	0	81308
SoxA	0,9940177	0,99401198	0,99998783	0,9939998	0,98809524	1	166	2	0	82138



SoxB	0,97855448	0,97835498	0,99996958	0,97832458	0,95762712	1	113	5	0	82188
SoxC	0,97055817	0,97014925	0,99997568	0,97012496	0,94202899	1	65	4	0	82237
SoxD	0,98471995	0,98461538	0,99998784	0,98460323	0,96969697	1	64	2	0	82240
SoxE	1	1	1	1	1	1	12	0	0	82294
SoxF	0,68470231	0,64788732	0,97902083	0,64775134	0,4893617	0,95833333	23	24	1	82258
SoxG	0,94589381	0,94444444	0,99998785	0,94443233	0,89473684	1	17	2	0	82287
SoxH	0,97013661	0,96969697	0,99999392	0,9696909	0,94117647	1	16	1	0	82289
SoxO	0,44212694	0,32727273	0,99955036	0,32702654	0,19565217	1	18	74	0	82214
SoxS	0,96822818	0,96774194	0,99998177	0,96772372	0,9375	1	45	3	0	82258
SoxT1	0,95969737	0,95890411	0,99998177	0,95888591	0,92105263	1	35	3	0	82268
SoxT2	1	1	1	1	1	1	21	0	0	82285
SoxV	0,95996032	0,95918367	0,99997569	0,9591594	0,92156863	1	47	4	0	82255
SoxW	0,95647746	0,95555556	0,99997569	0,95553129	0,91489362	1	43	4	0	82259
SoxX	1	1	1	1	1	1	116	0	0	82190
SoxY	1	1	1	1	1	1	157	0	0	82149
SoxZ	1	1	1	1	1	1	158	0	0	82148
DoxA	1	1	1	1	1	1	10	0	0	82296
DoxD	1	1	1	1	1	1	8	0	0	82298
Th	0,95342781	0,95238095	0,99996352	0,95234456	0,90909091	1	60	6	0	82240
TtrA	1	1	1	1	1	1	74	0	0	82232
TtrB	0,96358895	0,96296296	0,99996352	0,96292653	0,92857143	1	78	6	0	82222
TtrC	0,96643436	0,96644295	0,98646825	0,96641255	0,96	0,97297297	72	3	2	82229
TtrR	0,2232652	0,09817672	0,98220932	0,09742703	0,05169867	0,97222222	35	642	1	81628
TtrS	0,13049773	0,03429691	0,9880213	0,03348901	0,01744766	1	35	1971	0	80300
GerGerRed	1	1	1	1	1	1	24	0	0	82282
LpIA	0,95740383	0,95652174	0,99997569	0,95649747	0,91666667	1	44	4	0	82258
sHdRj	1	1	1	1	1	1	7	0	0	82299
sHdRl	1	1	1	1	1	1	8	0	0	82298
LipS1	1	1	1	1	1	1	23	0	0	82283
LipS2	1	1	1	1	1	1	24	0	0	82282
GcvH	1	1	1	1	1	1	322	0	0	81984
LbpA1	0,9860073	0,98591549	0,98611111	0,98590942	1	0,97222222	35	0	1	82270
LbpA2	0,97296082	0,97297297	0,98648041	0,97296082	0,97297297	0,97297297	36	1	1	82268
DHDL	0,86600962	0,85714286	0,99998177	0,857125	0,75	1	9	3	0	82294
sEtFA	0,97181941	0,97142857	0,99999392	0,9714225	0,94444444	1	17	1	0	82288
sEtFB	1	1	1	1	1	1	17	0	0	82289
sHdRA	1	1	1	1	1	1	47	0	0	82259
sHdRB1	0,97978399	0,97959184	0,99998784	0,97957968	0,96	1	48	2	0	82256
sHdRB2	0,98373278	0,98360656	0,99999392	0,98360048	0,96774194	1	30	1	0	82275
sHdRC1	0,96487782	0,96428571	0,99997568	0,96426143	0,93103448	1	54	4	0	82248
sHdRC2	1	1	1	1	1	1	31	0	0	82275
sHdRCCG1	0,84158961	0,82926829	0,99995747	0,829227	0,70833333	1	17	7	0	82282
sHdRCCG2	0,88850172	0,88235294	0,9999757	0,88232897	0,78947368	1	15	4	0	82287
sHdRH	1	1	1	1	1	1	47	0	0	82259
sHdRT	0,94867753	0,94736842	0,99999392	0,94736236	0,9	1	9	1	0	82296
sHdRSoxR	0,16674039	0,05521811	0,98959954	0,05410059	0,02839296	1	50	1711	0	80545
Cs2H	1	1	1	1	1	1	4	0	0	82302
McrA	1	1	1	1	1	1	81	0	0	82225
McrB	1	1	1	1	1	1	76	0	0	82230
McrC	0,90313377	0,89855072	0,99991489	0,89846648	0,81578947	1	62	14	0	82230
McrD	1	1	1	1	1	1	49	0	0	82257
McrG	1	1	1	1	1	1	73	0	0	82233
TusA	0,99729291	0,9973262	0,99996305	0,99728925	0,99466667	1	1119	6	0	81181
TusADsrE	1	1	1	1	1	1	98	0	0	82208
TusB	0,96411667	0,96350365	0,9999696	0,96347329	0,92957746	1	66	5	0	82235
TusC	1	1	1	1	1	1	66	0	0	82240
TusD	1	1	1	1	1	1	60	0	0	82246
TusE	0,97013661	0,96969697	0,99999392	0,9696909	0,94117647	1	16	1	0	82289
YeeE	1	1	1	1	1	1	177	0	0	82129
HmcA	1	1	1	1	1	1	5	0	0	82460
HmcB	0,96361827	0,96296296	0,99999394	0,96295691	0,92857143	1	13	1	0	82451
HmcC	0,98373279	0,98360656	0,99999393	0,98360049	0,96774194	1	30	1	0	82434
HmcD	1	1	1	1	1	1	1	0	0	82464
HmcE	1	1	1	1	1	1	3	0	0	82462
HmcF	1	1	1	1	1	1	3	0	0	82462
MvhA	1	1	1	1	1	1	16	0	0	82449
MvhD	0	0	0	0	0	0	0	0	0	82465
MvhG	0,97181942	0,97142857	0,99999394	0,97142251	0,94444444	1	17	1	0	82447
QrcA	1	1	1	1	1	1	13	0	0	82452
QrcB	0,95345681	0,95238095	0,99999394	0,9523749	0,90909091	1	10	1	0	82454
QrcC	0,87445342	0,86666667	0,99997574	0,86664284	0,76470588	1	13	4	0	82448
QrcD	1	1	1	1	1	1	29	0	0	82436
TmcA	1	1	1	1	1	1	7	0	0	82458
TmcB	1	1	1	1	1	1	7	0	0	82458
TmcC	0,96361827	0,96296296	0,99999394	0,96295691	0,92857143	1	13	1	0	82451
TmcD	1	1	1	1	1	1	6	0	0	82459
mHdRd	1	1	1	1	1	1	22	0	0	82443
mHdRE	0,95036463	0,94915254	0,9999818	0,94913439	0,90322581	1	28	3	0	82434

## Skew corrected 10-fold cross-validation

Model	Metrics				Confusion matrix					
	MCC	F1-score	Balanced accuracy	Cohens kappa	Precision	Recall	True positives	False positives	False negatives	True negatives
redDsrA	1	1	1	1	1	1	1520	0	0	305750
redDsrB	1	1	1	1	1	1	1520	0	0	305690
redDsrC	0,98954494	0,99456726	0,99494023	0,98949873	0,98961938	0,99956487	1716	18	124	305412
DsrD	1	1	1	1	1	1	940	0	0	306350
redDsrJ	1	1	1	1	1	1	1030	0	0	306180
redDsrK	0,99996696	0,99998343	0,99998343	0,99996696	1	0,99996685	1450	0	10	303740
redDsrM	0,91547777	0,95588235	0,95773129	0,91191965	0,91549296	1	1300	120	0	305620
redDsrN	0,99900134	0,99949691	0,99950765	0,99900093	0,99904398	0,9995025	1045	1	15	305829
redDsrO	0,89737072	0,94589574	0,94869826	0,89213573	0,89734816	0,99999673	1049	120	1	306160
redDsrP	0,88515104	0,93900912	0,94264445	0,8786069	0,88503804	0,99999017	1047	136	3	305834
DsrT	0,99996388	0,99998184	0,99998183	0,99996388	1	0,99996369	1009	0	11	306200
redAprA	0,91540347	0,95560472	0,95791443	0,91184519	0,91501416	0,99996369	969	90	11	306250
redAprB	0,93928112	0,96783318	0,97045619	0,93747059	0,9378882	0,99975046	1208	80	72	305570
redSat	0,99261983	0,99617665	0,99644893	0,9925948	0,99253731	0,99984278	665	5	45	305545
QmoA	1	1	1	1	1	1	1530	0	0	305780
QmoB	0,99996717	0,99998353	0,99998353	0,99996717	1	0,99996707	1530	0	10	305610
QmoC	1	1	1	1	1	1	1010	0	0	305930
qHdrB	0,99999674	0,99999837	0,99999833	0,99999674	1	0,99999673	419	0	1	306920
qHdrC	0,32490356	0,49056604	0,66240357	0,19096584	0,325	1	130	270	0	306940
mHdrA	0,98461513	0,99224806	0,99230744	0,9844968	0,98461538	1	640	10	0	302700
mHdrB	0,94827345	0,97345133	0,97413552	0,94693752	0,94827586	1	550	30	0	305310
mHdrC	0,94822553	0,97340271	0,97413387	0,94688746	0,94818653	0,99999673	549	30	1	306660
HdrF	0,93887929	0,96829659	0,97027296	0,93701593	0,93854749	0,99999334	168	11	2	303769
FlxA	0,98181789	0,99082569	0,9909088	0,98165262	0,98181818	1	540	10	0	305600
FlxB	0,91666043	0,95652174	0,95832709	0,91320079	0,91666667	1	550	50	0	306000
FlxC	0,99999673	0,99999836	0,99999833	0,99999673	1	0,99999673	439	0	1	306490
FlxD	0,8249765	0,90306994	0,91337411	0,80995043	0,82329317	0,99996625	410	88	10	303432
AsrA	0,99999672	0,99999836	0,99999828	0,99999672	1	0,99999671	209	0	1	305830
AsrB	0,99999671	0,99999835	0,99999827	0,99999671	1	0,9999967	209	0	1	304220
AsrC	0,9999762	0,9999879	0,99998772	0,9999762	1	0,9999758	203	0	7	299230
MccA	1	1	1	1	1	1	210	0	0	306870
MccB	0,99999671	0,99999835	0,99999828	0,99999671	1	0,9999967	209	0	1	304350
MccC	1	1	1	1	1	1	210	0	0	307110
MccD	0,98130829	0,99056604	0,99147667	0,98113363	0,98130841	1	210	4	0	307126
PsrAPhsASreA	0,92029986	0,95733818	0,96129981	0,91715176	0,91826087	0,99988924	528	47	32	306343
PsrBPhsBSreB	0,88622255	0,93548934	0,94689683	0,87985617	0,87901701	0,99971595	465	64	75	306476
PsrCPhsCSreC	0,98963299	0,99445344	0,99518136	0,98958379	0,98920086	0,99976211	458	5	64	306535
FccA	0,99997694	0,99998839	0,99998794	0,99997694	1	0,99997677	473	0	7	305800
FccB	0,91654845	0,9556191	0,959067	0,91310387	0,91513561	0,99985017	1046	97	44	305883
oxDsrA	1	1	1	1	1	1	690	0	0	306480
oxDsrB	1	1	1	1	1	1	720	0	0	306540
oxDsrC	0,96362503	0,98133331	0,98195123	0,96296801	0,96340348	0,99994319	1053	40	17	304010
DsrE	1	1	1	1	1	1	700	0	0	306520
DsrF	1	1	1	1	1	1	690	0	0	306470
DsrH	1	1	1	1	1	1	700	0	0	306580
oxDsrJ	0,99992367	0,99996117	0,99996095	0,99992366	1	0,99992233	647	0	23	306640
oxDsrK	0,99997703	0,99998846	0,99998793	0,99997702	1	0,99997692	733	0	7	306130
oxDsrM	0,99957007	0,99976574	0,9997658	0,99956998	1	0,99953159	610	0	120	306440
oxDsrO	0,98474744	0,99163928	0,99300865	0,98464388	0,98387097	0,99953124	610	10	120	306200
oxDsrP	0,99960529	0,99978637	0,99978638	0,99960521	1	0,99957282	619	0	111	306310
DsrL	0,99996659	0,99998318	0,99998318	0,99996659	1	0,99996637	740	0	10	301330
DsrN	1	1	1	1	1	1	630	0	0	304560
DsrR	1	1	1	1	1	1	420	0	0	304410
DsrS	1	1	1	1	1	1	110	0	0	306950
DsrE2	0,96055751	0,97973578	0,98047263	0,95978132	0,96028881	0,99998668	266	11	4	304769
DsrE3A	0,99999347	0,99999673	0,9999967	0,99999347	1	0,99999346	828	0	2	306490
DsrE3B	1	1	1	1	1	1	420	0	0	306920
DsrE3C	0,99986656	0,99993183	0,99993151	0,99986656	1	0,99986366	900	0	40	306390
DsrE4	0,99992619	0,99996202	0,99996184	0,99992619	1	0,99992405	368	0	22	306950
DsrE5	0,74998988	0,84498513	0,88474112	0,72008951	0,73170732	0,99976117	270	99	60	306881
AprM	1	1	1	1	1	1	320	0	0	307020
oxAprA	1	1	1	1	1	1	320	0	0	307020
oxAprAll	0,99999674	0,99999837	0,99999832	0,99999674	1	0,99999673	319	0	1	306940
assAprA	1	1	1	1	1	1	160	0	0	307010
oxAprBl	1	1	1	1	1	1	320	0	0	307020
oxAprBlI	0,99323092	0,99654684	0,99675724	0,99320846	0,99315068	0,99996663	290	2	10	306968
assAprB	0,76611207	0,86757991	0,88328734	0,73970315	0,76612903	1	190	58	0	306382
oxSat	0,88886569	0,94083108	0,94475836	0,88273577	0,88828338	0,99998677	326	41	4	305959
TsdA	0,99998687	0,99999339	0,99999314	0,99998687	1	0,99998679	296	0	4	306780
TsdB	0,99995994	0,99997948	0,99997935	0,99995994	1	0,99995895	238	0	12	307090
SoeA	1	1	1	1	1	1	1100	0	0	305510
SoeB	1	1	1	1	1	1	1100	0	0	305820
SoeC	1	1	1	1	1	1	1120	0	0	304170
SorA	0,89636493	0,94525401	0,94826963	0,89102747	0,89619377	0,99999673	259	30	1	306840
SorB	0,92178442	0,95930233	0,96499341	0,9187357	0,92178771	1	330	28	0	306862
SOR	1	1	1	1	1	1	240	0	0	307080
SQR1	1	1	1	1	1	1	1200	0	0	306090
SQR1I	0,95808337	0,97859327	0,97941113	0,95720568	0,95808383	1	160	7	0	306883
SQR1II	0,9996054	0,99979529	0,99979522	0,99960532	1	0,99959066	1494	0	116	305260
SQR1IV	0,99998358	0,99999173	0,99999159	0,99998358	1	0,99998345	315	0	5	306990
SQRV	1	1	1	1	1	1	840	0	0	306500
SQRVI	0,99990037	0,99994935	0,99994935	0,99990036	1	0,99998987	880	0	30	306210
SDOI	0,99737033	0,99868258	0,9986996	0,99736689	0,99737188	0,99999673	759	2	1	306578
SDOII	0,99987002	0,99993366	0,99993365	0,99987001	1	0,99986733	921	0	39	306360
SDOIII	0,88907086	0,94055223	0,94522577	0,88304345	0,888079	0,99961574	3777	476	113	302274
SoxA	1	1	1	1	1	1	1480	0	0	305170

SoxB	1	1	1	1	1	1	1120	0	0	304230
SoxC	1	1	1	1	1	1	500	0	0	306240
SoxD	0,99609374	0,99804305	0,99807686	0,99608611	0,99609375	1	510	2	0	306048
SoxE	1	1	1	1	1	1	100	0	0	307070
SoxF	0,89127318	0,93900672	0,95013195	0,88543433	0,88515406	0,99983666	316	41	44	306799
SoxG	0,99999338	0,99999666	0,99999642	0,99999338	1	0,99999332	118	0	2	304660
SoxH	0,99998342	0,99999164	0,99999135	0,99998342	1	0,99998327	275	0	5	304380
SoxO	0,99999669	0,99999834	0,99999824	0,99999669	1	0,99999668	169	0	1	302580
SoxS	0,99966523	0,99981195	0,99981198	0,99966517	1	0,99962397	320	0	90	306540
SoxT1	0,9461452	0,97225582	0,97316796	0,94469795	0,94601542	0,99999344	368	21	2	306539
SoxT2	0,93980467	0,96669697	0,97258458	0,93800951	0,93564356	0,99988243	189	13	31	306867
SoxV	0,99992874	0,99996323	0,99996318	0,99992874	1	0,99992647	319	0	21	304370
SoxW	0,98316621	0,99149129	0,99166304	0,98302476	0,98313253	0,9999934	408	7	2	304643
SoxX	0,99999015	0,99999507	0,99999493	0,99999015	1	0,99999014	957	0	3	305180
SoxY	1	1	1	1	1	1	1580	0	0	304700
SoxZ	1	1	1	1	1	1	1590	0	0	304980
DoxA	1	1	1	1	1	1	90	0	0	307230
DoxD	0,99999011	0,999995	0,99999471	0,99999011	1	0,99999	127	0	3	306940
Th	0,99996364	0,99998156	0,99998152	0,99996364	1	0,99996312	379	0	11	306940
TtrA	0,89761839	0,94598009	0,9488946	0,89240848	0,8975	0,99999673	359	41	1	306419
TtrB	0,89428284	0,94111149	0,94992078	0,88875846	0,88888889	0,99985329	320	40	40	306650
TtrC	0,90179609	0,94822594	0,95107076	0,897001	0,9015544	0,99999344	348	38	2	306802
TtrR	0,90837421	0,95160663	0,95482677	0,90419838	0,90769231	0,99998614	236	24	4	293396
TtrS	0,83230808	0,90789038	0,91702271	0,8184848	0,8313253	0,99998942	207	42	3	287638
GerGerRed	1	1	1	1	1	1	300	0	0	304180
LpIA	0,99999344	0,99999671	0,99999667	0,99999344	1	0,99999343	578	0	2	305260
sHdrJ	0,99999671	0,99999835	0,9999982	0,99999671	1	0,9999967	109	0	1	305700
sHdrI	1	1	1	1	1	1	180	0	0	307020
LipS1	1	1	1	1	1	1	490	0	0	306490
LipS2	0,99999346	0,99999672	0,99999665	0,99999346	1	0,99999344	378	0	2	306450
GcvH	0,99985754	0,99992836	0,99992835	0,99985752	1	0,99985674	3807	0	43	303490
LbpA1	0,97914132	0,98944988	0,98958133	0,97892396	0,97912317	0,99999673	469	10	1	306280
LbpA2	0,99999674	0,99999837	0,99999834	0,99999674	1	0,99999673	569	0	1	306560
DHDL	0,99999668	0,99999833	0,9999982	0,99999668	1	0,99999667	119	0	1	302600
sEtFA	0,9999967	0,99999834	0,99999813	0,9999967	1	0,99999668	79	0	1	305280
sEtFB	1	1	1	1	1	1	70	0	0	305470
sHdrA	0,99999347	0,99999673	0,99999669	0,99999347	1	0,99999346	678	0	2	306580
sHdrB1	1	1	1	1	1	1	670	0	0	306570
sHdrB2	1	1	1	1	1	1	600	0	0	306640
sHdrC1	1	1	1	1	1	1	680	0	0	306620
sHdrC2	1	1	1	1	1	1	590	0	0	306680
sHdrCCG1	0,97560968	0,98765432	0,98888857	0,97531233	0,97560976	1	80	2	0	307258
sHdrCCG2	0,9084884	0,95172264	0,95624685	0,90432102	0,90789474	0,99999669	69	7	1	306843
sHdrH	1	1	1	1	1	1	660	0	0	306610
sHdrT	1	1	1	1	1	1	250	0	0	304680
sHdrRSoxR	1	1	1	1	1	1	520	0	0	301120
Cs2H	1	1	1	1	1	1	130	0	0	307210
McrA	1	1	1	1	1	1	280	0	0	307050
McrB	1	1	1	1	1	1	270	0	0	307070
McrC	1	1	1	1	1	1	170	0	0	307130
McrD	1	1	1	1	1	1	250	0	0	307070
McrG	1	1	1	1	1	1	260	0	0	307080
TusA	0,97074532	0,98515939	0,98536901	0,97031779	0,97075596	0,99999667	6639	200	1	300490
TusADsrE	0,99952231	0,9997476	0,99974753	0,9995222	1	0,99949533	1152	0	138	306050
TusB	0,96153786	0,98039216	0,98076863	0,96079876	0,96153846	1	250	0	0	307080
TusC	0,9676802	0,98354504	0,98392638	0,9671584	0,9676259	0,99999673	269	9	1	307061
TusD	0,99995316	0,99997596	0,99997579	0,99995316	1	0,99995192	256	0	14	307070
TusE	0,8998251	0,94718163	0,94999391	0,89483607	0,89966555	0,99999673	269	30	1	307040
YeeE	0,98525701	0,992293	0,99317585	0,98515144	0,98481013	0,99989045	389	6	31	305494
HmcA	0,97142813	0,98550725	0,98603099	0,97102012	0,97142857	1	340	10	0	313610
HmcB	0,99730458	0,99865047	0,99868417	0,99730095	0,99730458	1	370	1	0	313589
HmcC	0,99999681	0,9999984	0,99999836	0,99999681	1	0,9999968	369	0	1	313590
HmcD	1	1	1	1	1	1	250	0	0	313710
HmcE	0,99999681	0,9999984	0,99999836	0,99999681	1	0,9999968	349	0	1	313610
HmcF	0,89729499	0,94579803	0,94871044	0,8920523	0,89717224	0,9999968	349	40	1	313570
MvhA	1	1	1	1	1	1	660	0	0	312990
MvhD	0,86308418	0,92049926	0,93861529	0,85385331	0,85283019	0,99983245	226	39	44	313651
MvhG	1	1	1	1	1	1	600	0	0	313140
QrcA	1	1	1	1	1	1	570	0	0	313390
QrcB	0,9727885	0,98597451	0,98664609	0,97242211	0,9724026	0,99993064	599	17	21	313323
QrcC	0,99999681	0,9999984	0,99999838	0,99999681	1	0,9999968	599	0	1	313360
QrcD	1	1	1	1	1	1	580	0	0	313380
TmcA	0,99993148	0,99996497	0,99996494	0,99993148	1	0,99992995	459	0	21	313480
TmcB	0,98148119	0,99065421	0,99074045	0,98130975	0,98148148	1	530	10	0	313420
TmcC	1	1	1	1	1	1	530	0	0	313430
TmcD	1	1	1	1	1	1	540	0	0	313420
mHdrD	0,99999031	0,99999509	0,99999474	0,99999031	1	0,99999017	107	0	3	313850
mHdrE	0,92436875	0,96069869	0,9624989	0,92151752	0,92436975	1	110	9	0	313841

## Skew corrected validation on independent test dataset

Model	Metrics						Confusion matrix			
	MCC	F1-score	Balanced accuracy	Cohens kappa	Precision	Recall	True positives	False positives	False negatives	True negatives
redDsrA	0,98581543	0,99285714	0,99290763	0,98571484	0,9858156	1	139	2	0	82165
redDsrB	0,95744532	0,97826087	0,97872192	0,95654073	0,95744681	1	135	6	0	82165
redDsrC	0,99974692	0,99986247	0,99986249	0,99974689	1	0,99972497	100	0	19	82187
DsrD	1	1	1	1	1	1	67	0	0	82239
redDsrJ	0,96747891	0,98347107	0,98373907	0,96695039	0,96747967	1	119	4	0	82183
redDsrK	0,98611094	0,99300699	0,99305539	0,9860145	0,98611111	1	142	2	0	82162
redDsrM	0,91602539	0,9561753	0,95801012	0,91251304	0,91603053	1	120	11	0	82175
redDsrN	0,98834583	0,99407054	0,99424035	0,9882785	0,98823529	0,9999751	84	1	2	82219
redDsrO	0,8473125	0,91735537	0,92364838	0,83581396	0,84732824	1	111	20	0	82175
redDsrP	0,86665542	0,92857143	0,93332209	0,8578553	0,86666667	1	104	16	0	82186
DsrT	1	1	1	1	1	1	103	0	0	82203
redAprA	0,97169761	0,98564593	0,98584855	0,97129726	0,97169811	1	103	3	0	82200
redAprB	0,95883361	0,97889707	0,97950087	0,95798803	0,95867769	0,99998773	116	5	1	82184
redSat	0,96212327	0,97764081	0,98398337	0,96142501	0,95652174	0,99971352	44	2	17	82243
QmoA	1	1	1	1	1	1	84	0	0	82222
QmoB	0,98717941	0,99354839	0,99358967	0,98709723	0,98717949	1	77	1	0	82228
QmoC	0,97916642	0,98947368	0,98958309	0,97894945	0,97916667	1	94	2	0	82210
qHdrB	1	1	1	1	1	1	11	0	0	82295
qHdrC	1	1	1	1	1	1	15	0	0	82291
mHdrA	1	1	1	1	1	1	98	0	0	82208
mHdrB	0,98757749	0,99375	0,99378867	0,98750034	0,98757764	1	159	2	0	82145
mHdrC	0,99453549	0,99726027	0,99726773	0,99452056	0,99453552	1	182	1	0	82123
HdrF	0,97560947	0,98765432	0,98780459	0,97531211	0,97560976	1	80	2	0	82224
FlxA	1	1	1	1	1	1	23	0	0	82283
FlxB	1	1	1	1	1	1	23	0	0	82283
FlxC	1	1	1	1	1	1	33	0	0	82273
FlxD	1	1	1	1	1	1	46	0	0	82260
AsrA	1	1	1	1	1	1	25	0	0	82281
AsrB	1	1	1	1	1	1	25	0	0	82281
AsrC	1	1	1	1	1	1	25	0	0	82281
MccA	1	1	1	1	1	1	12	0	0	82294
MccB	0,85714137	0,92307692	0,92856994	0,84705713	0,85714286	1	12	2	0	82292
MccC	1	1	1	1	1	1	12	0	0	82294
MccD	1	1	1	1	1	1	12	0	0	82294
PsrAPhsASreA	0,66666262	0,8	0,83332928	0,61537944	0,66666667	1	6	3	0	82297
PsrBPhsBSreB	0,22220017	0,36363636	0,61108906	0,09409985	0,22222222	1	6	21	0	82279
PsrCPhsCSreC	0,66666262	0,8	0,83332928	0,61537944	0,66666667	1	6	3	0	82297
FccA	0,8775471	0,93478261	0,93877159	0,87011326	0,87755102	1	43	6	0	82257
FccB	0,98353836	0,99142744	0,9920374	0,98340454	0,98305085	0,99994801	58	1	4	82243
oxDsrA	0,99998773	0,99999381	0,99999381	0,99998773	1	0,99998761	53	0	1	82252
oxDsrB	0,99998774	0,99999381	0,99999381	0,99998774	1	0,99998763	57	0	1	82248
oxDsrC	0,85254651	0,92041522	0,92626446	0,84181212	0,8525641	1	133	23	0	82150
DsrE	1	1	1	1	1	1	69	0	0	82237
DsrF	1	1	1	1	1	1	63	0	0	82243
DsrH	1	1	1	1	1	1	71	0	0	82235
oxDsrJ	0,99994935	0,99997366	0,99997366	0,99994935	1	0,99994732	48	0	4	82254
oxDsrK	0,9999753	0,99998746	0,99998746	0,9999753	1	0,99997492	64	0	2	82240
oxDsrM	0,99985239	0,99991925	0,99991926	0,99985238	1	0,99983851	53	0	11	82242
oxDsrO	0,98020613	0,9889296	0,99114476	0,98001616	0,97826087	0,9998336	45	1	11	82249
oxDsrP	0,96073945	0,97819805	0,9820768	0,95997878	0,95744681	0,99986871	45	2	9	82250
DsrL	0,99998774	0,99999382	0,99999382	0,99998774	1	0,99998764	59	0	1	82246
DsrN	0,96153801	0,98039216	0,98076878	0,96079892	0,96153846	1	50	2	0	82254
DsrR	1	1	1	1	1	1	38	0	0	82268
DsrS	1	1	1	1	1	1	3	0	0	82303
DsrE2	0,90587405	0,95061728	0,95293287	0,90146576	0,90588235	1	154	16	0	82136
DsrE3A	0,95180611	0,97530864	0,9759025	0,9506462	0,95180723	1	79	4	0	82223
DsrE3B	1	1	1	1	1	1	14	0	0	82292
DsrE3C	1	1	1	1	1	1	8	0	0	82298
DsrE4	1	1	1	1	1	1	19	0	0	82287
DsrE5	1	1	1	1	1	1	25	0	0	82281
AprM	1	1	1	1	1	1	34	0	0	82272
oxAprA1	1	1	1	1	1	1	33	0	0	82273
oxAprAll	0,99998761	0,99999369	0,99999369	0,99998761	1	0,99998738	26	0	1	82279
assAprA	0,87499668	0,93333333	0,93749668	0,86725291	0,875	1	35	5	0	82266
oxAprBl	0,99993488	0,99996527	0,99996527	0,99993488	1	0,99993054	35	0	5	82266
oxAprBlI	0,96296275	0,98113208	0,98148126	0,96227736	0,96296296	1	26	1	0	82279
assAprB	1	1	1	1	1	1	115	0	0	82191
oxSat	0,63803912	0,77332556	0,82288004	0,5786465	0,63043478	0,99997402	29	17	2	82258
TsdA	0,96721273	0,98333333	0,98360617	0,96667553	0,96721311	1	59	2	0	82245
TsdB	1	1	1	1	1	1	47	0	0	82259
SoeA	0,99998773	0,99999381	0,99999381	0,99998773	1	0,99998762	56	0	1	82249
SoeB	0,99998774	0,99999381	0,99999381	0,99998774	1	0,99998763	57	0	1	82248
SoeC	0,99998773	0,99999381	0,99999381	0,99998773	1	0,99998762	56	0	1	82249
SorA	0,83332658	0,90909091	0,91665991	0,81966429	0,83333333	1	40	8	0	82258
SorB	0,76921782	0,86956522	0,88460244	0,7434787	0,76923077	1	40	12	0	82254
SOR	1	1	1	1	1	1	25	0	0	82281
SQR1	1	1	1	1	1	1	125	0	0	82181
SQR1I	1	1	1	1	1	1	40	0	0	82266
SQR1II	1	1	1	1	1	1	243	0	0	82063
SQR1IV	1	1	1	1	1	1	33	0	0	82273
SQRV	1	1	1	1	1	1	484	0	0	81822
SQRVI	1	1	1	1	1	1	67	0	0	82239
SDOI	0,93268963	0,96517413	0,96634348	0,93042978	0,93269231	1	97	7	0	82202
SDOII	1	1	1	1	1	1	117	0	0	82189
SDOIII	0,11368237	0,20503597	0,55656945	0,02551758	0,11422846	1	114	884	0	81308
SoxA	0,98809509	0,99401198	0,99404748	0,98802424	0,98809524	1	166	2	0	82138

SoxB	0,95762588	0,97835498	0,97881233	0,95672894	0,95762712	1	113	5	0	82188
SoxC	0,94202766	0,97014925	0,97101316	0,94035025	0,94202899	1	65	4	0	82237
SoxD	0,96969661	0,98461538	0,98484813	0,96923768	0,96969697	1	64	2	0	82240
SoxE	1	1	1	1	1	1	12	0	0	82294
SoxF	0,49460376	0,65714012	0,74995719	0,39310941	0,4893617	0,99998732	23	24	1	82258
SoxG	0,8947357	0,94444444	0,94736728	0,88922951	0,89473684	1	17	2	0	82287
SoxH	0,94117613	0,96969697	0,9705879	0,93944919	0,94117647	1	16	1	0	82289
SoxO	0,19558139	0,32727273	0,59775531	0,07368553	0,19565217	1	18	74	0	82214
SoxS	0,93749893	0,96774194	0,96874893	0,9355498	0,9375	1	45	3	0	82258
SoxT1	0,92105131	0,95890411	0,96052499	0,91794537	0,92105263	1	35	3	0	82268
SoxT2	1	1	1	1	1	1	21	0	0	82285
SoxV	0,92156687	0,95918367	0,96078256	0,91850122	0,92156863	1	47	4	0	82255
SoxW	0,91489172	0,95555556	0,95744492	0,9112843	0,91489362	1	43	4	0	82259
SoxX	1	1	1	1	1	1	116	0	0	82190
SoxY	1	1	1	1	1	1	157	0	0	82149
SoxZ	1	1	1	1	1	1	158	0	0	82148
DoxA	1	1	1	1	1	1	10	0	0	82296
DoxD	1	1	1	1	1	1	8	0	0	82298
Th	0,90908789	0,95238095	0,95454244	0,90497409	0,90909091	1	60	6	0	82240
TtrA	1	1	1	1	1	1	74	0	0	82232
TtrB	0,92856901	0,96296296	0,96428329	0,92602481	0,92857143	1	78	6	0	82222
TtrC	0,960494	0,97957984	0,9805063	0,95971621	0,96	0,999975	72	3	2	82229
TtrR	0,05218695	0,09831455	0,52634617	0,00543351	0,05169867	0,9999875	35	642	1	81628
TtrS	0,01724099	0,03429691	0,50851839	0,00059433	0,01744766	1	35	1971	0	80300
GerGerRed	1	1	1	1	1	1	24	0	0	82282
LplA	0,91666481	0,95652174	0,95833148	0,91320554	0,91666667	1	44	4	0	82258
sHdRj	1	1	1	1	1	1	7	0	0	82299
sHdRl	1	1	1	1	1	1	8	0	0	82298
LipS1	1	1	1	1	1	1	23	0	0	82283
LipS2	1	1	1	1	1	1	24	0	0	82282
GcvH	1	1	1	1	1	1	322	0	0	81984
LbpA1	0,99998767	0,99999375	0,99999375	0,99998767	1	0,9999875	35	0	1	82270
LbpA2	0,97331605	0,98629529	0,9868357	0,97296082	0,97297297	0,99998751	36	1	1	82268
DHDL	0,74999658	0,85714286	0,87499658	0,7199958	0,75	1	9	3	0	82294
sEtFA	0,94444413	0,97142857	0,9722219	0,94290342	0,94444444	1	17	1	0	82288
sEtFB	1	1	1	1	1	1	17	0	0	82289
sHdRA	1	1	1	1	1	1	47	0	0	82259
sHdRB1	0,95999953	0,97959184	0,97999953	0,95920018	0,96	1	48	2	0	82256
sHdRB2	0,96774175	0,98360656	0,98387078	0,96722173	0,96774194	1	30	1	0	82275
sHdRC1	0,93103292	0,96428571	0,96551568	0,92866075	0,93103448	1	54	4	0	82248
sHdRC2	1	1	1	1	1	1	31	0	0	82275
sHdRCG1	0,70832455	0,82926829	0,85415788	0,66819705	0,70833333	1	17	7	0	82282
sHdRCG2	0,78946964	0,88235294	0,8947328	0,76791325	0,78947368	1	15	4	0	82287
sHdRH	1	1	1	1	1	1	47	0	0	82259
sHdRT	0,89999945	0,94736842	0,94999945	0,89502702	0,9	1	9	1	0	82296
sHdRSoxR	0,02810441	0,05521811	0,51390939	0,00157847	0,02839296	1	50	1711	0	80545
Cs2H	1	1	1	1	1	1	4	0	0	82302
McrA	1	1	1	1	1	1	81	0	0	82225
McrB	1	1	1	1	1	1	76	0	0	82230
McrC	0,81577668	0,89855072	0,90788194	0,79915335	0,81578947	1	62	14	0	82230
McrD	1	1	1	1	1	1	49	0	0	82257
McrG	1	1	1	1	1	1	73	0	0	82233
TusA	0,99466647	0,9973262	0,99733314	0,99465225	0,99466667	1	1119	6	0	81181
TusADsrE	1	1	1	1	1	1	98	0	0	82208
TusB	0,92957547	0,96350365	0,96478674	0,92710227	0,92957746	1	66	5	0	82235
TusC	1	1	1	1	1	1	66	0	0	82240
TusD	1	1	1	1	1	1	60	0	0	82246
TusE	0,94117613	0,96969697	0,9705879	0,93944919	0,94117647	1	16	1	0	82289
YeeE	1	1	1	1	1	1	177	0	0	82129
HmcA	1	1	1	1	1	1	5	0	0	82460
HmcB	0,92857103	0,96296296	0,96428531	0,92602697	0,92857143	1	13	1	0	82451
HmcC	0,96774175	0,98360656	0,98387078	0,96722173	0,96774194	1	30	1	0	82434
HmcD	1	1	1	1	1	1	1	0	0	82464
HmcE	1	1	1	1	1	1	3	0	0	82462
HmcF	1	1	1	1	1	1	3	0	0	82462
MvhA	1	1	1	1	1	1	16	0	0	82449
MvhD	0	0	0	0	0	0	0	0	0	82465
MvhG	0,94444413	0,97142857	0,9722219	0,94290342	0,94444444	1	17	1	0	82447
QrcA	1	1	1	1	1	1	13	0	0	82452
QrcB	0,90909041	0,95238095	0,95454495	0,90497683	0,90909091	1	10	1	0	82454
QrcC	0,76470152	0,86666667	0,88234858	0,73798595	0,76470588	1	13	4	0	82448
QrcD	1	1	1	1	1	1	29	0	0	82436
TmcA	1	1	1	1	1	1	7	0	0	82458
TmcB	1	1	1	1	1	1	7	0	0	82458
TmcC	0,92857103	0,96296296	0,96428531	0,92602697	0,92857143	1	13	1	0	82451
TmcD	1	1	1	1	1	1	6	0	0	82459
mHdrD	1	1	1	1	1	1	22	0	0	82443
mHdrE	0,90322422	0,94915254	0,95161131	0,89856559	0,90322581	1	28	3	0	82434



---

## HMSS2: An advanced tool for the analysis of sulphur metabolism, including organosulphur compound transformation, in genome and metagenome assemblies

Tanabe, T. S., & Dahl, C.

---

Organic sulfur compounds are biologically relevant in archaea, bacteria, and eukaryotes and can serve simultaneously as a sulfur, carbon, or energy source. In aquatic environments dimethylsulfoniopropionate (DMSP) is currently the most abundant compound produced by phytoplankton in pelagic waters (Kiene *et al.* 2000) and by prokaryotes in the dark deep waters and sediments (Zheng *et al.* 2020). Bacterial DMSP degradation is also the major source for dimethylsulfide (DMS), which affects ocean chemistry and modulates climate change and bridges the aquatic and terrestrial sulfur pools. In the coming decades taurine and sulfolactate may become the predominant sulfonates as the phytoplankton composition is expected to change, due to the global warming, with unforeseen effects on the availability of organic and inorganic sulfur compounds in the oceanic habitat (Moran & Durham 2019). In terrestrial environments, sulfonated lipids are the most abundant group of sulfonated compounds (Goddard-Borger & Williams 2017). They are found as part of the thylakoid membranes of chloroplasts, but can also be synthesized by algae, bacteria and archaea (Goddard-Borger & Williams 2017, Hou *et al.* 2022). The degradation of organic sulfur compounds in terrestrial and aquatic habitats and their effects on the ecosystems and human health are not well understood (Wolf *et al.* 2022, Folz *et al.* 2023).

The HMS-S-S was further developed to HMSS2 to include proteins required for organic sulfur metabolism (Tanabe & Dahl 2023). The profiled Hidden Markov models (HMM) were generated according to similar standards as the HMMs for inorganic sulfur metabolism in HMS-S-S. However, due to the higher prevalence of organic sulfur metabolism, more genomes were included. Sequence selection and alignment methods were the same as described for HMS-S-S. For validation, a cross-validation procedure and an independent test set consisting of sequences not used for the alignment were used to assess the performance of each HMM. In addition, 24 selected genomes of prokaryotes with described capacity for organic sulfur compound conversion were searched with the HMMs created for HMSS2. In this test it was possible to replicate the previously identified and described genomic capacity for each genome tested. A total of 134 HMMs were added to HMSS2, covering transport, transformation, degradation, and assimilation processes in the organic sulfur cycle. These included the proteins for the DMSP degradation, the DMS and methanethiol assimilation, the dissimilation of dimethylsulfone and methanesulfonate, the alkanesulfonate oxidation and transport, sulfoquinovose biosynthesis, degradation and transport, the

---

Tanabe, T. S., & Dahl, C. (2023) HMSS2: An advanced tool for the analysis of sulphur metabolism, including organosulphur compound transformation, in genome and metagenome assemblies. *Molecular ecology resources*, 23(8), 1930 - 1945; doi.org/10.1111/1755-0998.13848

2,3-dihydroxypropane sulfonate transport and degradation, the isethionate and taurine degradation, the sulfoacetaldehyde formation and the cysteine biosynthesis. The corresponding pathways and involved enzymes are reviewed in the article to provide an overview of the metabolism included in HMSS2. In addition to the new HMMs, several technical improvements and new features have been implemented in HMSS2. The computational calculation time per genome required by HMSS2 has been increased by a factor of 4 compared to HMS-S-S. HMSS2 also implements new algorithms to increase the detection performance by taking into account the genomic vicinity of each detected gene sensitivity and reliability, a feature that is not present in the HMMER3, BLAST or similar tools. Detection of multidomain proteins has also been added to HMSS2. In HMS-S-S, multidomain proteins were only annotated by highest scoring single domain, which was usually the longest polypeptide. Any other protein domain was ignored. HMSS2 now correctly annotated multidomain proteins with each non-overlapping domain that is detected. In addition, a feature was added to create files compatible with the iTol visualization platform (Letunic & Bork 2021). HMSS2 is now able to create dataset and range files compatible with phylogenetic trees visualized with iTol. The overall usability has also been significantly improved by simplifying the required input data and the installation process, by providing a pre-compiled version that can be directly executed (Tanabe & Dahl 2023).

T.S.T. contributed to this study by conceptualization, investigation, data curation, formal analysis, validation, visualization and writing of the original manuscript: T.S.T. conceptualized the entire program structure and improvements. The modified source code of HMS-S-S was rewritten by T.S.T. in python3. For the HMMs T.S.T. carried out the literature research to find suitable reference proteins and organisms, collected the sequences and created and validated of the HMMs. The algorithmic improvements were developed and implemented by T.S.T. Improvements, modifications and overall functionality of HMSS2 was tested by T.S.T. Benchmark tests were set up, performed and analyzed by T.S.T. The article figures and text were drafted by T.S.T who also contributed to writing the final manuscript and processing the revisions.





Received: 11 May 2023 | Revised: 8 July 2023 | Accepted: 17 July 2023

DOI: 10.1111/1755-0998.13848

## RESOURCE ARTICLE

MOLECULAR ECOLOGY  
RESOURCES WILEY

# HMSS2: An advanced tool for the analysis of sulphur metabolism, including organosulphur compound transformation, in genome and metagenome assemblies

Tomohisa Sebastian Tanabe | Christiane Dahl

Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

**Correspondence**

Tomohisa Sebastian Tanabe and Christiane Dahl, Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich Wilhelms-Universität Bonn, Meckenheimer Allee 168, 53115 Bonn, Germany.  
Email: [s6totana@uni-bonn.de](mailto:s6totana@uni-bonn.de) and [chdahl@uni-bonn.de](mailto:chdahl@uni-bonn.de)

**Funding information**

Deutsche Forschungsgemeinschaft, Grant/Award Number: Da 351/13-1; Studienstiftung des Deutschen Volkes

**Handling Editor:** Frederic Austerlitz

**Abstract**

The global sulphur cycle has implications for human health, climate change, biogeochemistry and bioremediation. The organosulphur compounds that participate in this cycle not only represent a vast reservoir of sulphur but are also used by prokaryotes as sources of energy and/or carbon. Closely linked to the inorganic sulphur cycle, it involves the interaction of prokaryotes, eukaryotes and chemical processes. However, ecological and evolutionary studies of the conversion of organic sulphur compounds are hampered by the poor conservation of the relevant pathways and their variation even within strains of the same species. In addition, several proteins involved in the conversion of sulphonated compounds are related to proteins involved in sulphur dissimilation or turnover of other compounds. Therefore, the enzymes involved in the metabolism of organic sulphur compounds are usually not correctly annotated in public databases. To address this challenge, we have developed HMSS2, a profiled Hidden Markov Model-based tool for rapid annotation and synteny analysis of organic and inorganic sulphur cycle proteins in prokaryotic genomes. Compared to its previous version (HMS-S-S), HMSS2 includes several new features. HMM-based annotation is now supported by nonhomology criteria and covers the metabolic pathways of important organosulphur compounds, including dimethylsulphoniopropionate, taurine, isethionate, and sulphoquinovose. In addition, the calculation speed has been increased by a factor of four and the available output formats have been extended to include iTol compatible data sets, and customized sequence FASTA files.

**KEYWORDS**

dimethylsulphoniopropionate, hidden Markov model database, organosulphur compounds, sulphoquinovose, sulphur metabolism

## 1 | INTRODUCTION

The global organic sulphur cycle occurs in both terrestrial and aquatic environments and involves the interplay of prokaryotes, eukaryotes and chemical processes. Millions of megatonnes of sulfonated

compounds are produced annually by biological and industrial processes. These compounds not only represent a vast reservoir of sulphur but can also be used by prokaryotes as a sources of energy and carbon (Moran & Durham, 2019). Understanding the mechanisms and ecological interactions of prokaryotes in the organic sulphur

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

cycle is of great importance because the decomposition of organic sulphur compounds affects human health, bacterial virulence in infection (Dhouib et al., 2021), global warming, bioremediation processes such as wastewater treatment (Schäfer et al., 2010), and is linked to the biogeochemical cycling of sulphur between habitats (Koch & Dahl, 2018).

Sulphonated compounds can range from small size with only a C<sub>1</sub> carbon skeleton up to sulphonated lipids with long-chain alkanes, amino acids such as cysteine, or sulphur-containing cofactors with complex structures such as lipoate (Boden & Hutt, 2019; Goddard-Borger & Williams, 2017; Moran & Durham, 2019). Although new sulphonated compounds are constantly being discovered, the metabolic function, synthesis or degradation pathways are often not yet clear (Thume et al., 2018). Only the most abundant sulphonated compounds, such as sulphanoinovose, dimethylsulphoniopropionate (DMSP), taurine, isethionate, cysteine and methionine, have been studied biochemically in terms of synthesis and degradation pathways.

In aquatic environments, the antistress molecule DMSP is the most well-known organosulphur compound (Kiene et al., 2000). Mainly produced by macroalgae and phytoplankton, it is emitted by around 600 million tonnes per year. Bacterial DMSP degradation in the oceans, salt marshes and coastal regions is the major source of dimethylsulphide (DMS), which is released at a rate of about 300 million tonnes per year (Moran & Durham, 2019). As a volatile compound, DMS affects atmospheric chemistry and global warming by forming cloud condensation nuclei that increase the reflection of solar radiation (Schäfer et al., 2010). In the context of the global sulphur cycle, DMS acts as a link between the terrestrial, atmospheric and aquatic environments (Lovell et al., 1972). DMS-derived carbon and sulphur are used as electron acceptors or donors during dissimilation, or are assimilated via the intermediates dimethylsulphone and methanesulphonate (Figure 1).

Sulphonated lipids are estimated to be the largest reservoir of sulphur in terrestrial ecosystems (Goddard-Borger & Williams, 2017).

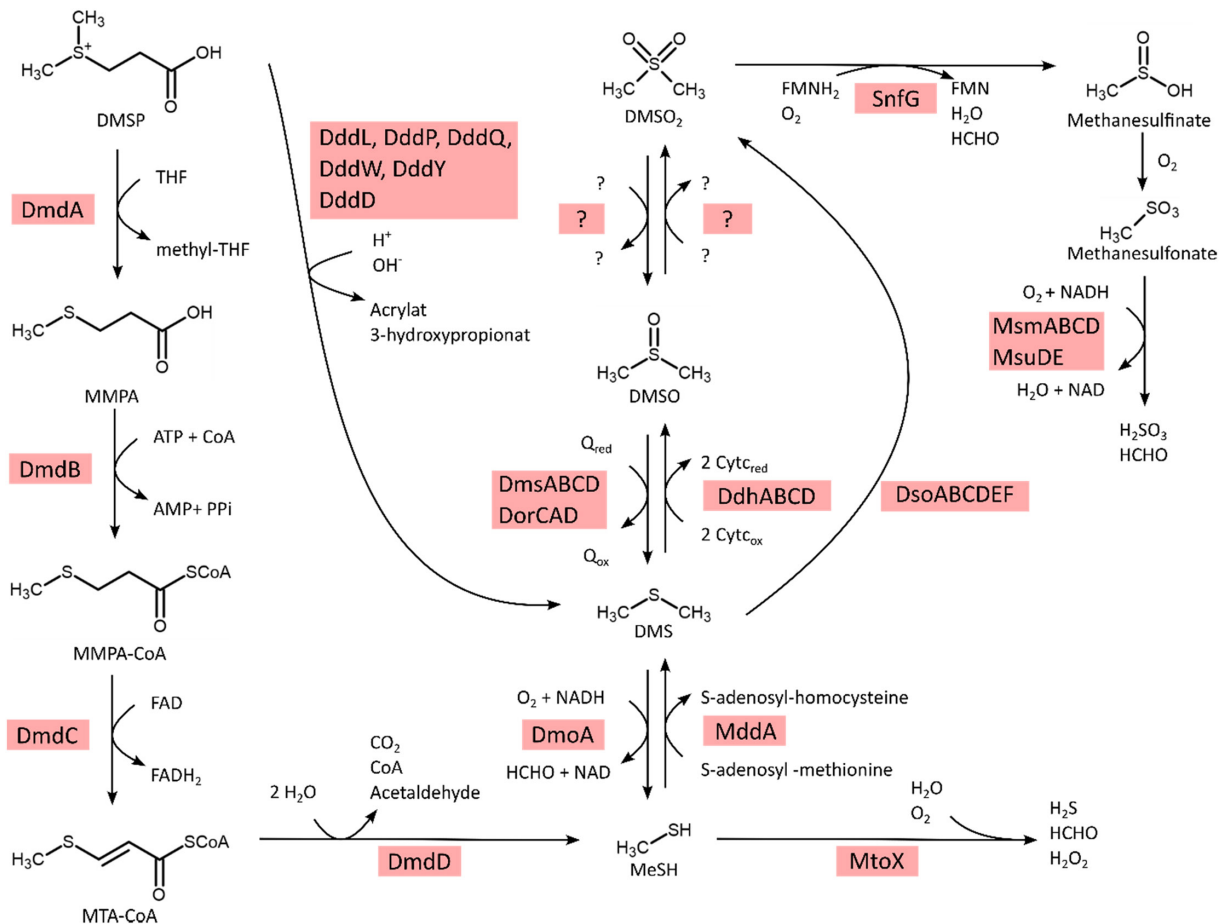


FIGURE 1 Prokaryotic metabolism of C<sub>1</sub> organosulphur compounds. All proteins shown have a corresponding HMM in HMSS2. Cyt<sub>c</sub>, Cytochrome c; DMSP, dimethylsulphoniopropionate; DHPs, 2,3-dihydroxypropane-1-sulphonate; DMS, dimethylsulphide; DMSO, dimethylsulphone, DMSO<sub>2</sub> dimethylsulphoxide; FMN, flavin mononucleotide; FMNH<sub>2</sub>, reduced flavin mononucleotide; MeSH, methanethiol; MMPA, methylmercaptopyropionate; MMPA-CoA, 3-methylmercaptopyropionyl-CoA; MTA-CoA, methylthioacryloyl-CoA; THF, tetrahydrofolate.

Sulphoquinovose is a sulphonated glucose derivate and the most common part of the head group of sulpholipids which are integral part of thylakoid membranes of chloroplasts and photosynthetic systems. Mainly produced by plants, algae and cyanobacteria, its turnover rate has been estimated at around 10 billion tonnes per year (Goddard-Borger & Williams, 2017). The bacterial decomposition of sulphoquinovose involves several different pathways similar to the degradation of glucose (Figure 2a), with the exception that smaller sulphonated compounds are often released, since complete utilization with release of free sulphur by a single organism is often not possible (Wei et al., 2022). Release and scavenging of sulphonated intermediates is achieved by various transport systems (Figure 2b). Sulphoquinovose decomposition and release of inorganic sulphur is then completed by pathways linked to taurine, isethionate and/or sulphoacetate (Figure 2c). In summary, prokaryotic utilization of these organic compounds as sources of sulphur, carbon and energy is far from being a uniform process and new metabolic pathways for the degradation of sulphonated compound are constantly being discovered (Boden et al., 2010; Koch & Dahl, 2018; Sharma et al., 2022; Wolf et al., 2022).

These processes are also closely linked to the availability of inorganic sulphur as the released sulphur is either assimilated or excreted as sulphate (Ruff et al., 2003), sulphite (Koch & Dahl, 2018; Li et al., 2023; Sharma et al., 2022), thiosulphate (De Zwart et al., 1997), tetrathionate (Boden et al., 2010) or sulphide (Peck et al., 2019). Indeed, the complete consumption of the volatile sulphonated C<sub>1</sub>-compound DMS coupled with the oxidation of the thiosulphate formed as an intermediate, has been reported for a single organism, providing a new link between the organic and inorganic sulphur cycles (Koch & Dahl, 2018). However, the fate of the sulphur released from sulphonated compounds is often not known or assumed to be the same as in dissimilatory sulphur oxidation or reduction. The physiology and interactions of bacterial communities that release sulphur from sulphonated carbon compounds have been sparsely explored and the few existing studies are based on, or assume, sulphur cycling via dissimilatory sulphite reductases (Burrichter et al., 2021; Hanson et al., 2021; Wolf et al., 2022).

Ecological studies of organic sulphur compounds are difficult because their metabolism is poorly conserved across bacterial phylogeny and can even vary between strains of the same species. Thus, even within a species, predictions based on taxonomic assignment are not possible (Schäfer et al., 2010). As the functional annotation pipelines of public databases mainly focus on the synthesis of methionine and cysteine, the enzymes involved in the metabolism of organic sulphur compounds are usually not correctly annotated. Inaccurate annotation in public databases is exacerbated by the fact that several proteins involved in the conversion of sulphonated compounds are related to proteins involved in sulphur dissimilation or the turnover of other compounds, for example, the DMSO reductase family (Leimkühler & Iobbi-Nivol, 2016) or quinone oxidoreductase complexes (Duarte et al., 2021). For these reasons, the abundance of microbes utilizing organic sulphur compounds is likely to be underestimated (Carrion et al., 2019) and the role of sulphonated compounds

is understudied (Wolf et al., 2022). Thus, there is a knowledge gap of the link between inorganic and organic sulphur cycling in ecological systems.

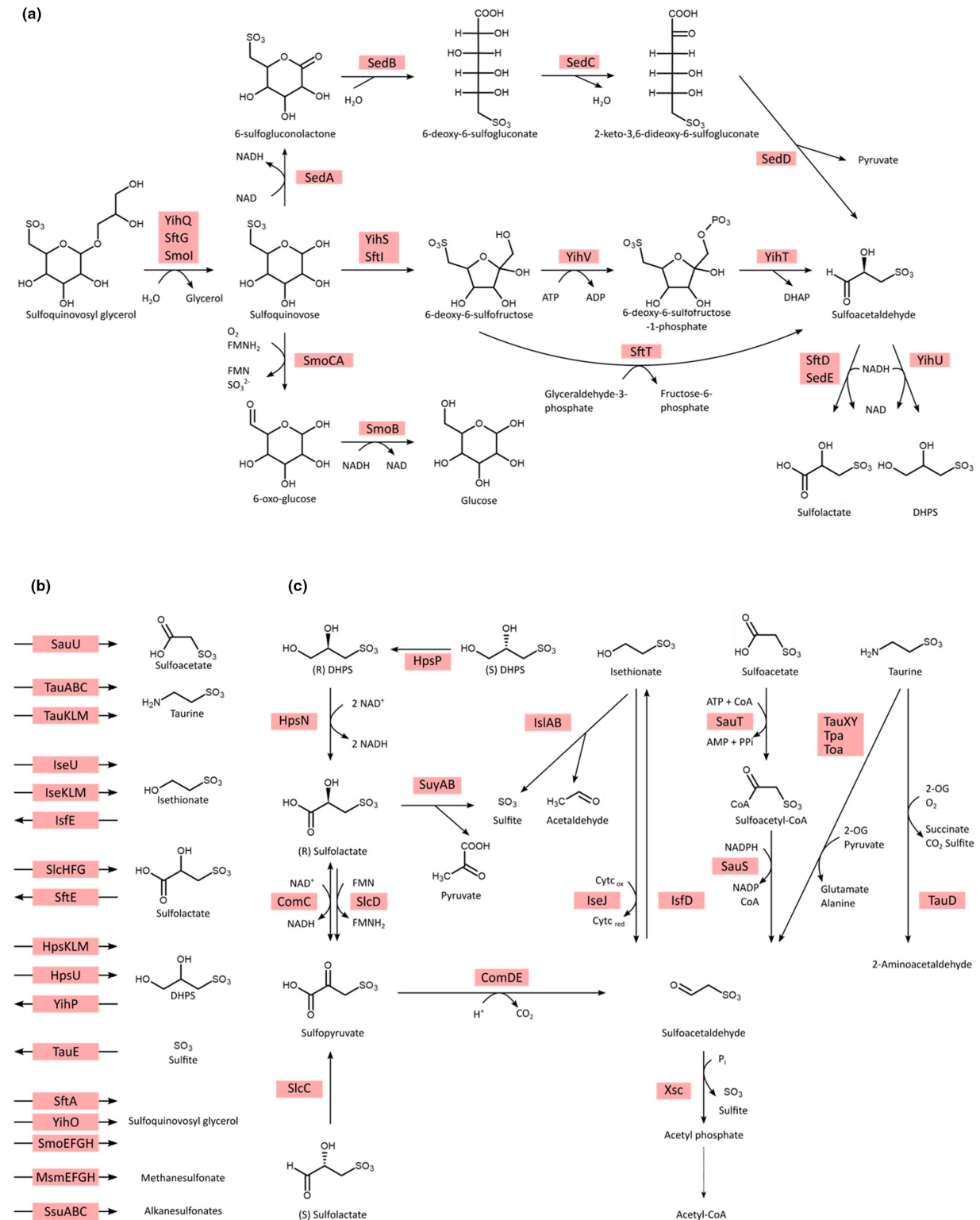
To fill this gap, we have extended HMS-S-S (Tanabe & Dahl, 2022). This tool was originally developed for rapid detection and annotation of inorganic sulphur dissimilation in prokaryotic genomes. With the substantial extension presented here, it now includes not only inorganic sulphur metabolism enzymes but also enzymes with characterized or at least strongly indicated function in the metabolism of sulphonated sulphur compounds. These include sulphoquinovose synthesis and degradation pathways, DMSP metabolism, taurine and isethionate conversion, and transport systems for various sulphonated compounds. For all these pathways, we developed individual profiled hidden Markov Models (HMM) and validated score thresholds by cross-validation and with an independent test data set. HMS-S-S itself has been completely redesigned, improving usability and output formats, and extending the file manipulation tool. By optimizing the underlying algorithms, the overall computing speed has been increased by a factor of four. Due to the complete overhaul, we have renamed the tool "HMSS2". HMSS2 now covers the known metabolism of inorganic and organic sulphur compounds, facilitating the exploration of the microbe-driven natural sulphur cycle.

## 2 | METHODS

### 2.1 | HMSS2 improvements and workflow

Algorithmic improvements were made on the speed and user-friendliness by process optimization and the implementation of additional features. HMSS2 algorithms are now completely written in Python and precompiled versions are available. In this way, the number of dependencies required to be installed by the user has been greatly reduced to just two external programs. HMMER and Prodigal are still required but installing and configuring of MySQL is no longer necessary. The installation was further simplified by preparation of a precompiled executable that will run directly on a Unix system.

HMSS2 includes the basic design of HMS-S-S with further automation. User-supplied input requires a directory containing files in FASTA nucleotide format, consisting of scaffolds or contigs. Alternatively, it is possible to provide amino acid sequences in FASTA files and the corresponding features in GFF3 formatted files. All files in the directory will then be processed in consecutive order. Nucleotide input files are first searched for open-reading frames and translated into protein sequences by Prodigal. This step is omitted if protein sequences are provided. Profile HMM are then queried against the protein sequences of the current file with validated bit score cut-offs via *hmmsearch*. Hits are saved in a local database together with corresponding genomic features and protein amino acid sequences. The local database now uses the SQLite database engine and an improved database table structure



**FIGURE 2** Prokaryotic metabolism of organosulphur compounds with two or more carbon atoms and relevant transporters. (a) Pathways of sulfoquinovosyl glycerol degradation. (b) Transport systems for import and export of organic sulphur compounds. (c) Degradation pathways of C<sub>2</sub> and C<sub>3</sub> organosulphur compounds. Usually, the same cell does not contain all the pathways. All proteins show have a corresponding HMM in HMSS2. Cyt<sub>c</sub>, Cytochrome c; DHPS, 2,3-dihydroxypropane-1-sulphonate; FMN, flavin mononucleotide; FMNH<sub>2</sub>, reduced flavin mononucleotide; 2-OG, 2-oxoglutarate.

that allows to save multidomain proteins with all domains. In the next step, the detected proteins are searched for genetic colocalization. This is done via the genomic features and a maximum nucleotide distance between two genes to be syntenic. Syntenic gene clusters are then compared with a set of predefined and named gene patterns. A new feature of HMSS2 is the detection of co-linear gene clusters. This is a special type of synteny where the genes occur in exactly the same order as the gene pattern. Gene clusters that are similar to the pattern(s) provided are then named by characteristic keywords. NCBI, GTDB taxonomy files or custom files with a similar format can be used to assign taxonomic information. As the taxonomy may change over time, it is recommended that the user updates this information locally as required. Results can be retrieved from the local database filtered by protein domains and/or keywords via HMSS2. The standard output now includes FASTA formatted files and iTol data sets.

## 2.2 | Training data set generation, annotation and HMM development

Data sets were generated from genomic data downloaded from NCBI RefSeq (Haft et al., 2018) or GenBank (Sayers et al., 2019) as of September 2022. The HMM training data set contained all assemblies from the NCBI RefSeq database with an assembly level of a complete chromosome. The independent test data consisted of assemblies originating from GenBank, again with an assembly level of the complete chromosome. GenBank covers a greater number of phyla and a wider range of quality and is, therefore, not entirely similar to the training data from RefSeq. Sequence annotation for Hidden-Markov-model generation was performed using the training data set and list of reference proteins for organic sulphur metabolism (Table S1). Methods for annotating the training and independent test data sets and for HMM generation were used as described previously (Tanabe & Dahl, 2022).

## 2.3 | Performance metric calculation

Performance was determined using balanced accuracy (Brodersen et al., 2010), F1-score (Forman & Scholz, 2010), and the Matthew correlation-coefficient (MCC; Chicco & Jurman, 2020). The metric values were additionally corrected for the data set's skewness (Jeni et al., 2013; Table S2). Values for each Hidden Markov Model were calculated from a confusion matrix obtained by comparing the annotation of the training/test data set and annotation assigned by the HMMs. Matching assignments were considered as true positives (TP), while mismatching assignments were considered as false positives (FP), if the HMM recognized a sequence unrelated to the HMM training sequences. All sequences that were not recognized by the HMM but matched the annotation were counted as false negative (FN), and all other sequences were recorded as true negatives.

## 2.4 | Thresholding and cross-validation

Thresholding and cross-validation were executed as previously described (Tanabe & Dahl, 2022). For each HMM, bit scores for noise cutoff, trusted cutoff, and an optimized threshold were determined prior to cross-validation. The noise cutoff corresponded to the score of the lowest scoring TP hit. The trusted cutoff corresponded to the score of the highest scoring FP hit. The optimized cutoff was computed during a nested cross-validation procedure with a 10-fold outer loop and a five-fold inner loop (Varma & Simon, 2006). The optimized cutoff corresponded to the median of the thresholds with the highest F1 scores across all inner folds. Outer folds were analysed after all thresholds were set.

Each cross-validation fold was generated from the HMM training data. Sequences were randomly sorted into the 10 outer folds of equal size, followed by the equal deviation of each outer fold into five inner folds. A cross-validation procedure was then performed on all folds. The inner folds were used to determine the optimized thresholds. The overall performance of each HMM was then done with a confusion matrix created for the outer folds using the optimized thresholds as a cutoff. Balanced accuracy was calculated as the average of all accuracies from each fold. F1 score and MCC were calculated as the sum of the confusion matrices from all folds (Forman & Scholz, 2010). The same procedure without fold generation was performed for the independent test data set (Chicco, 2017).

## 2.5 | Performance testing

The performance of HMSS2 was compared with that of HMS-S version 1 (Tanabe & Dahl, 2022). The HMM library included all 164 HMMs of the original library, detecting dissimilatory sulphur metabolism. A quadratic increasing number of randomly selected genomes ranging from 2 to 64 were chosen from the training data set described for version 1 and used as input for the performance comparison. The input data were in FASTA nucleotide format. Each run was repeated three times with newly randomized input data to reduce performance bias caused by the input data. Both program versions were benchmarked for the execution time required for the workflow from data entry to the final annotated hits with appropriately named gene clusters, but without taxonomy assignment. Time was measured as the required wall-clock runtime when running HMS-S or HMSS2 with four parallel threads on an Intel Core i7-6700 CPU.

## 3 | RESULTS

Here, we created a comprehensive database of reliable HMMs based on archaeal and bacterial proteins associated with organic sulphur metabolism. The same approach has already been used for the enzymes of dissimilatory metabolism of inorganic sulphur compounds (Tanabe & Dahl, 2022). Not only sequence similarity but also integrated synteny was considered to assign a protein

to a specific functional group. The HMMs created here focus on the most abundant organic sulphur compounds in terrestrial and aquatic environments. The compounds covered here include DMSP, dimethyl sulphide (DMS), dimethyl sulphoxide (DMSO), dimethyl sulphone (DMSO<sub>2</sub>; Figure 1), 2,3-dihydroxypropane-1-sulphonate (DHPS), isethionate, taurine and membrane sulpholipids (Figure 2). The HMMs for the enzymes of the metabolic pathways for degradation of individual compounds are described in full below. Normally, prokaryotes do not code for the entire degradation pathways, but only for parts of them (Boden & Hutt, 2019; Liu et al., 2021).

### 3.1 | HMM development: DMSP degradation

DMSP is primarily produced by single-celled phytoplankton and algal seaweeds, where it acts as an osmolyte and anti-stress molecule (Kiene et al., 2000). Degradation of DMSP either requires a demethylation pathway or a DMSP lyase (Figure 1). The demethylation pathway is encoded by the *dmdABCD* gene cluster and starts with the demethylation of DMSP via DmdA to form methylmercaptopyruvate. This intermediate is further catabolized by DmdB, DmdC and finally DmdD with the release of acetaldehyde and methanethiol (Bullock et al., 2014; Reisch et al., 2011). For each of the enzymes, one HMM was generated, making four in total. Several nonorthologous DMSP lyases, DddL, DddP, DddQ, DddW and DddY, have been characterized which convert DMSP to acrylate with the release of DMS and acrylate. The latter is then converted to 3-hydroxypropionate by AcuNK (Curson et al., 2011) or to propionyl-CoA by AcuI (Todd et al., 2012). DMSP lyase DddD catalyses formation of propionyl-CoA and DMS from DMSP in a single reaction without the formation of an acrylate intermediate. 3-hydroxypropionate can be further converted to acetyl-CoA via DddA and DddC (Curson et al., 2011). HMMs were generated for AcuI, AcuN, AcuK, DddA and all DMSP lyases. As there were less than ten sequences identified for DddQ, DddW and DddC, HMMs could not be constructed for these three enzymes.

### 3.2 | HMM development: assimilation of methanethiol and DMS

DMS and methanethiol are C<sub>1</sub>-organosulphur compounds derived mainly from the degradation of DMSP. Both can be assimilated by bacteria as a source of sulphur and carbon, where methanethiol is first converted to DMS, followed by oxidation and assimilation (Figure 1). The conversion of methanethiol to DMS is catalysed by methanethiol S-methyltransferase, MddA. This membrane-bound enzyme transfers a single sulphur atom from S-adenosylmethionine to methanethiol (Carrion et al., 2015). The resulting DMS can be further oxidized by either DMS cytochrome c reductase, DdhABCD, also known as DMS dehydrogenase (McDevitt, Hanson, et al., 2002), or by multicomponent DMS monooxygenase DsoABCDEF (Horinouchi

et al., 1999). The periplasmic DdhABC DMS dehydrogenase couples the oxidation of DMS to the reduction of two c-type cytochromes, producing DMSO as the final product. DdhD is a cytoplasmic protein that is not part of the DMS dehydrogenase but has a proposed function in the assembly of the DdhAB complex and its secretion via the Tat pathway (McDevitt, Hugenholtz, et al., 2002). For DdhA and DdhB, it was possible to generate individual HMMs, while this was not the case for DdhC and DdhD which had less than ten validly annotated sequences in the training data set. The multicomponent DMS monooxygenase DsoABCDEF oxidizes DMS in a two-step reaction to DMSO<sub>2</sub> with DMSO as intermediate. As the sulphur moiety is specifically oxidized, this enzyme is also referred to in the literature as assimilatory DMS S-monooxygenase (Boden & Hutt, 2019). A total of six HMMs were generated for this complex. After the oxidation of DMS to DMSO<sub>2</sub>, the next step in sulphur assimilation is the oxygen-dependent conversion of DMSO<sub>2</sub> to methanesulphinate, catalysed by FMN-dependent DMSO<sub>2</sub> monooxygenase SnfG (Wicht, 2016). SnfG was represented by a single HMM. Methanesulphinate is chemically oxidized to methanesulphonate, which is further oxidized to sulphite and formaldehyde by the assimilatory methanesulphonate monooxygenase MsuDE in a NADH- and oxygen-dependent reaction. For MsuDE, a HMM was trained for each subunit.

### 3.3 | HMM development: dissimilation of DMSO<sub>2</sub>

Dimethylsulphone is mainly derived from oxidation of DMS. The degradation of dimethyl sulphone (DMSO<sub>2</sub>) begins with its reduction to DMSO by a DMSO<sub>2</sub> reductase in an NADH-dependent reaction (Figure 1). Although the activity has been measured in crude extracts of some methylotrophic Actinobacteria and Alphaproteobacteria (Borodina et al., 2000, 2002), the enzyme has not been characterized. DMSO is then further reduced to DMS. Two types of DMSO reductases have so far been characterized (Boden & Hutt, 2019). The first, membrane-bound enzyme is composed of the three subunits, DmsABC, and uses electrons from the quinol pool for DMSO reduction (Bilous & Weiner, 1985). For this enzyme one HMM for each subunit was trained. The second DMSO reductase uses NADH for this purpose and probably consists of only one subunit with high similarity to DmsA, indicated by its cross-reaction with DmsA antibodies. A separate HMM could not be trained for this enzyme, because it is only known by its activity in crude extracts (Borodina et al., 2002). In addition to the Dms-type DMSO reductases, a soluble periplasmic DMSO reductase, DorCAD, has been characterized (McEwan et al., 1998). The corresponding genes are regulated by DorS and DorR (Kappler & Schäfer, 2014). For each of these five proteins/subunits, we constructed one HMM. The DMS, which is released by DMSO reductase of both types, is oxidized to methanethiol (CH<sub>3</sub>SH) and formaldehyde by a DMS monooxygenase, DmoAB, in another NADH-consuming reaction (Boden et al., 2011). As only *dmoA* has been validly identified so far, we trained a HMM specifically for DmoA, but not for DmoB. Further oxidation of methanethiol by a methanethiol oxidase MtoX leads to the final release of

sulphide and another molecule of formaldehyde (Eyice et al., 2017). A single HMM was trained for MtoX.

### 3.4 | HMM development: dissimilation of methanesulphonate

Methanesulphonate is formed by spontaneous chemical oxidation of DMS in the atmosphere (Figure 1). It is used by diverse aerobic bacteria as a sulphur source and by some specialized methylotrophic prokaryotes as a source of carbon and energy (Kelly & Murrell, 1999). The dissimilatory methanesulphonate monooxygenase catalyses the conversion of methanesulphonate to formaldehyde and sulphite (Henriques & De Marco, 2015). This enzyme is encoded by the *msmABCD* operon, which is often located adjacent to the *msmEFGH* operon, usually in the opposite direction. The latter encodes a putative ABC-type transporter (Figure 2b) proposed to facilitate the import of methanesulphonate into the cytoplasm (Henriques & De Marco, 2015). Six HMMs were developed to represent each of these proteins. MsmC and MsmD had to be excluded due to the small number of sequences in the training data sets.

### 3.5 | HMM development: alkanesulphonate oxidation and transporters

The *ssuEADCB* gene cluster encodes the two-component alkanesulphonate monooxygenase SsuDE and the alkanesulphonate ABC-transporter SsuABC (Figure 2b). Alkanesulphonate monooxygenase catalyses the oxidation of various sulphonated alkanes as substrates with variable affinity, including phenylated organic compounds like *N*-phenyltaurine. After transport into the cell via SsuABC, the sulphonate is cleaved by SsuDE in a reaction dependent on NADH and molecular oxygen (Eichhorn et al., 1999). Electrons are provided by SsuE via an FMN cofactor. SsuD then cleaves the sulphonate group and oxidizes the terminal carbon atom. For this pathway five HMMs, one for each encoded protein, were created.

### 3.6 | HMM development: sulphoquinovose synthesis

Sulphoquinovose (SQ) is a sulphonated derivate of glucose where the 6-hydroxyl group is substituted by a sulphonate group. SQ is a constituent of the unique head group of the membrane-bound glycolipid sulphoquinovosyl diacylglycerol (SQDG) present in thylakoid membranes and photosynthetic prokaryotes. On a genetic level, five genes *sqdA*, *sqdB*, *sqdC*, *sqdD* and *sqdX* have been described to be involved in SQDG synthesis in bacteria so far (Benning & Somerville, 1992a, 1992b; Guler et al., 2000; Rossak et al., 1995). The functions of *SqdA* and *SqdC* have not been completely resolved (Benning & Somerville, 1992b; Rossak et al., 1997). The synthesis begins with the exchange of the 6-hydroxyl group of

uridine-diphosphate (UDP)-glucose for a sulphonate group by UDP-sulphoquinovose synthase, *SqdB*. The formation of SQDG is then catalysed by SQDG synthase, *SqdD* or *SqdX* (Rossak et al., 1995). A total of five HMMs was trained to detect the enzymes of this pathway.

### 3.7 | HMM development: sulphoquinovose degradation and transport

As sulphoquinovose is a sulphonated derivate of glucose, it is catabolized in a similar manner and can serve as a carbon and energy source (Hanson et al., 2021). Several pathways resembling glucose degradation have been characterized, including the sulpho-Embden-Meyerhof-Parnas pathway (Denger et al., 2014), the sulpho-Entner-Doudoroff pathway (Felux et al., 2015), the transaldolase-based pathway related to the pentose phosphate pathway (Frommeyer et al., 2020) and a complete degradation pathway based on a sulphoquinovose monooxygenase (Sharma et al., 2022; Figure 2a).

The sulpho-Embden-Meyerhof-Parnas pathway (Figure 2a) begins with import of sulphoquinovose by the transporter *YihO*. A sulpholipid  $\alpha$ -glucosidase *YihQ* may also be involved and other SQ derivatives may also be imported. Analogous to the EMP pathway, SQ is then cleaved to dihydroxyacetonephosphate (DHAP) and 3-sulpholactaldehyde (SLA) via the isomerase *YihS*, kinase *YihV* and aldolase *YihT*. In an NADH-dependent reaction, the reductase *YihU* then reduces SLA to the final product 2,3-dihydroxypropane sulphonate (DHPS), which is transported out of the cell again via *YihP*. A separate HMM was created for each of the *Yih* proteins.

The Sulpho-Entner-Doudoroff is analogous to the ED pathway (Figure 2a). As there was no specific abbreviated name assigned to these enzymes by the original publication (Felux et al., 2015), we assigned names to enhance HMSS2 output readability. SQ is cleaved by a dehydrogenase *SedA*, a lactonase *SedB*, a dehydratase *SedC*, and an aldolase *SedD* to pyruvate and SLA. Another dehydrogenase, *SedE*, then oxidizes SLA in an NAD-dependent reaction to 3-sulpholactate (SL), which is then exported. A separate HMM was generated for each of the proteins mentioned, for a total of five HMMs.

The third SQ degradation pathway contains a transaldolase as the key enzyme (Figure 2a; Frommeyer et al., 2020). SQ is imported into this pathway via the transporter *SftA* and converted to sulphofructose by the isomerase *SftI*. This product, together with glycerinaldehyde-3-phosphate, is then converted by the transaldolase *SftT* to SLA and fructose-6-phosphate. SLA, in turn, is converted to SL in an NAD-dependent reaction by the dehydrogenase *SftD* and exported via the transporter *SftE* or reduced to DHPS in an NADH-dependent reaction by the reductase *SftR*. A separate HMM was generated for each of the *Sft* proteins, for a total of six HMMs.

The fourth known degradation pathway for SQ (Figure 2a) differs from the others described so far, because it involves oxidation of the entire molecule, including cleavage of sulphur (Sharma et al., 2022). The pathway described begins with the import of sulphoquinovosyl glycerol by an ABC transporter called *SmoEFGH*. In the cytoplasm,

sulphoquinovosyl glycerol is cleaved by the sulphoquinovosidase SmoI to SQ. In contrast to the other pathways, SQ is now transformed to 6-oxo-glucose and sulphite by an alkanesulphonate monooxygenase, SmoC. The electrons for this reaction come from NADPH via the flavin reductase SmoA. 6-oxo-glucose is converted in another NADPH-dependent reaction by SmoB into glucose, which is then available for glycolysis. Eight HMMs were generated for this pathway, one for each protein. An additional HMM was trained for SmoD, a putative regulator encoded in the *smo* operon.

### 3.8 | HMM development: 2,3-dihydroxypropane sulphonate transporters and degradation

According to the postulated pathway for degradation of 2,3-dihydroxypropane sulphonate (DHPS; Figure 2c), the compound is either taken up by the TRAP transporter HpsKLM or by HpsU (Figure 2b). The DHPS-3-dehydrogenase HpsN then converts (R)-DHPS to sulpholactate with concomitant formation of two equivalents of NADH. For (S)-DHPS, it was postulated that this compound is first converted to the (R)-DHPS enantiomer via (R)-DHPS-2-dehydrogenase HpsP and (S)-DHPS-2-dehydrogenase HpsO (Mayer et al., 2010). The resulting (R)-sulpholactate can be further converted in several ways: The (R)-sulpholactate sulpholyase SuyAB catalyses a desulphonation reaction, releasing sulphite and pyruvate. The (S)-enantiomer of sulpholactate is first converted to sulphopyruvate by SlcC and then to (R)-sulpholactate by ComC (Mayer et al., 2010). Both enantiomers were postulated to be transported by the exporter SlcHFG (Mayer et al., 2010; Figure 2b). One HMM was created for each protein/subunit of the DHPS degradation pathway.

### 3.9 | HMM development: isethionate and taurine degradation

Isethionate and taurine are C<sub>2</sub>-sulphonates which are produced by eukaryotes from cysteine or methionine (Moran & Durham, 2019). Bacterial degradation of these compounds includes sulphoacetaldehyde as an intermediate which is a point of convergence with sulphoacetate degradation (Weinitschke, Hollemeyer, et al., 2010; Figure 2c). Two different transporters are proposed for the import of isethionate (Figure 2b). These are the TRAP transporters IseKLM and IseU from the major facilitator superfamily. After import into the cytoplasm, isethionate is oxidized to sulphoacetaldehyde by the isethionate dehydrogenase IseJ (Weinitschke, Sharma, et al., 2010). In some organisms, isethionate is not converted, but the sulphonate group is cleaved off by isethionate sulphite lyase IslAB, releasing sulphite and acetaldehyde (Peck et al., 2019).

Taurine import is postulated to be facilitated by the ABC transporter TauAB1B2C or the TRAP transporter TauKLM (Figure 2b). There are several possibilities for the further pathway. Taurine can either be oxygenated by TauD to form 1-hydroxy-2-aminoethane sulphonic acid, which decomposes to aminoacetaldehyde and

sulphite (Eichhorn et al., 1999), or it is oxidized in NADH-dependent reaction by the taurine dehydrogenase TauXY, which produces sulphoacetaldehyde. The same product is also produced by the transfer of the amino group to pyruvate by taurine: pyruvate aminotransferase Tpa (Bruggemann et al., 2004) or to 2-oxoglutarate by taurine:2-oxoglutarate aminotransferase Toa (Krejci et al., 2010).

Sulphoacetaldehyde can be converted by the NADPH-dependent sulphoacetaldehyde reductase IsfD to isethionate which is then exported by the IsfE transporter (Krejci et al., 2010). Another possible fate of sulphoacetaldehyde is desulphonation coupled to a phosphorylation by sulphoacetaldehyde acetyltransferase Xsc to acetyl phosphate which is further converted to acetyl-CoA by phosphate acetyltransferase Pta (Weinitschke, Sharma, et al., 2010). Sulphite released in the each of these processes is exported via TauE (Weinitschke et al., 2007). An individual HMM was developed for each individual protein/subunit mentioned here. An exception was made for TauB1 and TauB2, which were combined into a single HMM due to their similarity. Additionally, we trained an HMM for TauZ, a protein of unknown function, and the regulator TauR. Both are commonly found genetically associated with other *tau* genes.

### 3.10 | HMM development: sulphoacetaldehyde formation

Sulphoacetaldehyde is not only produced by taurine and isethionate degradation but also by the dissimilation of sulphoacetate (Weinitschke, Hollemeyer, et al., 2010). The transporter SauU is hypothesized to facilitate the entry of sulphoacetate into the cell (Figure 2b). Subsequently, sulphoacetate is activated by sulphoacetate-CoA ligase, SauT and finally reduced to sulphoacetaldehyde via sulphoacetaldehyde dehydrogenase, SauS, consuming NADPH. SauS, SauT and SauU (Weinitschke, Hollemeyer, et al., 2010) were each represented by a HMM respectively. Sulphoacetaldehyde can also be produced by decarboxylation of sulphopyruvate (Figure 2c) catalysed by ComDE (Denger et al., 2009). These two subunits are each represented by a HMM.

### 3.11 | HMM development: cysteine synthesis

Cysteine is an essential amino acid with a thiol side chain. Here, we started to cover the relevant proteins with HMMs primarily based on knowledge collected with enterobacterial model organisms. Biosynthesis begins with the import of sulphate or thiosulphate into the bacterial cell via CysUWA (Aguilar-Barajas et al., 2011) or YeeE/YedE-like (Tanaka et al., 2020) transporters. Sulphate is reduced to sulphide which is then incorporated into O-acetylserine to synthesize cysteine (Kredich, 1996). In *E. coli*, sulphate is activated by ATP sulphurylase CysDN (Leyh et al., 1988) to adenosine 5'-phosphosulphate (APS), which can be further activated by APS kinase CysC to 3'-phosphoadenosine-5'-phosphosulphate (PAPS). PAPS reductase CysH then reduces the activated compound to



sulphite. In some bacteria, including most cyanobacteria, APS can be reduced to sulphite directly, without phosphorylation to PAPS (Bick et al., 2000). The assimilatory APS reductases catalysing this reaction exhibit similarity to the assimilatory PAPS reductases (Abola et al., 1999; Bick et al., 2000) and are covered by the same HMM (CysH) in this work. In Enterobacteria, sulphite is reduced to sulphide via CysIJ. Finally, cysteine is synthesized from sulphide and O-acetyl-L-serine by the cysteine synthase CysK. A total of 10 new HMMs was generated for the mentioned proteins/subunits. An HMM for YeeE/YedE-like transporters was already available through HMS-S-S (Tanabe & Dahl, 2022).

### 3.12 | HMM validation: cross validation and independent test data set

The HMMs developed were validated by cross-validation and with an independent test data set. In cross-validation, sequences unrelated to the tested HMM training data were added as true negative examples in addition to the omitted training sequences (Chicco, 2017; Refaeilzadeh et al., 2009). The omitted sequences from each fold served as true positive examples. Cross-validation was performed using the optimized thresholds calculated prior to cross-validation. Thus, the threshold values should also be checked for their suitability. Performance was measured using the MCC. This metric ranges from -1 to 1, with 0 corresponding to random assignment, 1 corresponding to perfect assignment with no misclassification, and -1 corresponding to complete misclassification. Here, the individual occurrence of FP or FN lowers the score on the MCC, while the combination of both misclassifications lowers the score more dramatically than the single occurrence of either type of error (Chicco & Jurman, 2020).

The majority of the HMMs developed showed high precision and recall in the cross-validation and on the test data set (Figure 3). Of the 134 HMMs covering proteins of organic sulphur compound metabolism, 127 stayed above an MCC of 0.80 during the cross-validation

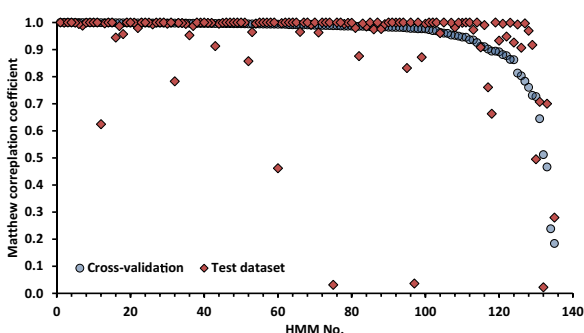


FIGURE 3 Validation of the 134 HMMs generated in this work. Performance was assessed by cross-validation (blue dots) and on an independent test data set (red diamonds). For each HMM Matthew correlation coefficient was calculated. HMMs were ranked by their performance in cross-validation.

(Figure 3; Table S2). The evaluation of the 134 HMMs against the independent test data set resulted in 120 HMMs with an MCC of 0.80 or higher. HMMs for the alkanesulphonate transporter subunits SsuB and SsuC failed the cross-validation threshold of 0.8 slightly by 0.02 points but performed better on the independent test data set. These were the only cases where the cross-validation performance was insufficient but the performance on the test data set was above the threshold. From the HMMs with an MCC >0.8 during cross-validation, seven scored below 0.8 in the test data set. These were MsmG with an MCC of 0.78, Smol (0.76), MsmB (0.66), DddA (0.62), DorA (0.46) and SftD (0.03). For SftD, MsmB, MsmG and DddA this was due to a high number of sequences which were falsely classified as negative, probably due to a low training sequence diversity. Thus, these HMMs had a high precision and did not generate high numbers of false positive hits, but they performed low in recognition resulting in a high number of unrecognized sequences. The opposite was the case for the DorA HMM, which generated too many false positive hits but no FN ones. Sulphoquinovosidase Smol interfered in the detection with sulphoquinovosidase named YihQ. The same holds true for transporters HpsU and IseU. All sequences that were falsely classified by one of these two HMMs belonged to the other HMM. Together these two HMMs performed well in detecting of isethionate and DHPS transporters of the major facilitator superfamily. The situation was similar for YihO and SftA which are both postulated sulphoquinovose importers that catalyse the same function in the context of sulphoquinovose degradation. In summary, 112 of 134 HMMs were successfully tested via cross-validation and with an independent data set. Two other pairs of HMMs can be used together, for the safe detection of sulphoquinovosidase and the transporters YihO and SftA.

### 3.13 | HMM validation: case study

HMS2 was also validated with 24 complete genomes from bacteria with organic sulphur compound metabolism (Table S3), which were screened for the presence of enzymes for the utilization of taurine, isethionate, DHPS, sulphoquinovose and DMS (Figure 4).

Proteins for taurine utilization were found mainly in the known taurine-utilizing genera *Octadecabacter*, *Roseobacter*, *Roseovarius* and *Ruegeria* of the Roseobacterales, including the taurine degraders *Roseovarius nubinhibens* (Denger et al., 2009) and *Ruegeria pomeroyi* (Gorzynska et al., 2006). These strains encoded for the TauABC taurine importer, Tpa and Xsc constituting the complete degradation pathway from free taurine via sulphaacetaldehyde to acetyl phosphate with the release of sulphite. *Roseobacter denitrificans* additionally possessed genes for the taurine dehydrogenase TauXY and the taurine:2-oxoglutarate aminotransferase Toa, which can also convert taurine to sulphaacetaldehyde. The sulphaacetaldehyde acetyltransferase Xsc was present in all genomes examined. This is probably due to the fact that sulphaacetaldehyde is not only exclusively an intermediate of taurine degradation but also of isethionate, sulphaacetate and DHPS degradation, and possibly of other

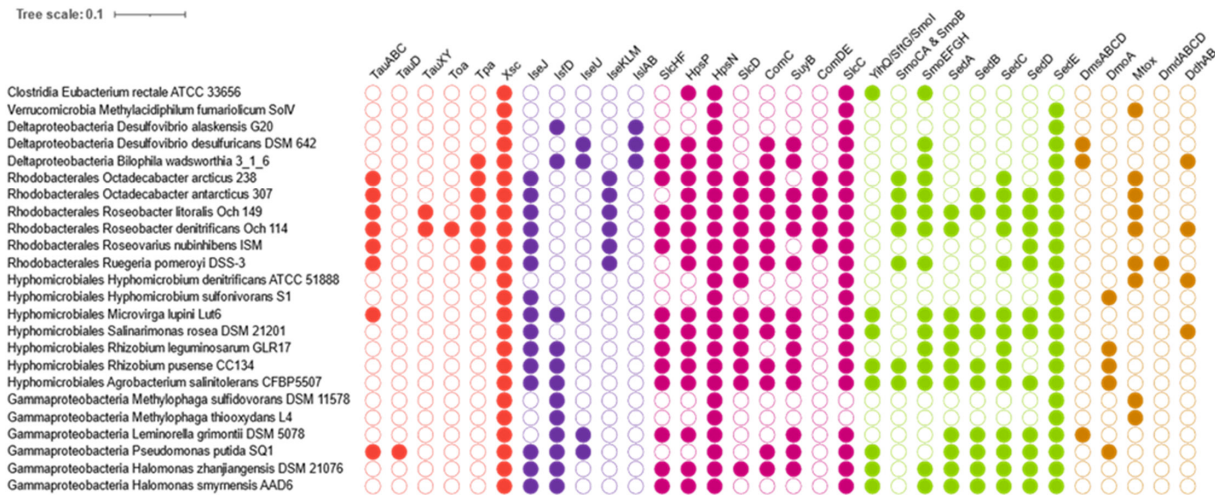


FIGURE 4 Presence/absence of proteins involved in the metabolism of organic sulphur compounds. Occurrence of genes for proteins involved in taurine degradation, isethionate degradation, 2,3-dihydroxypropane-1-sulphonate, sulphoquinovose and DMS metabolism, is indicated by filled orange, violet, purple, green and light brown circles respectively. The function of the individual proteins can be deduced from Figures 1 and 2.

as yet unknown pathways (Weinitschke, Hollemeyer, et al., 2010). In line with this possibility, genes encoding isethionate dehydrogenase IseJ, which converts isethionate to sulphoacetaldehyde, were found in almost all analysed Rhodobacterales, Hyphomicrobiales and Gammaproteobacteria genomes, consistent with earlier reports (Weinitschke, Sharma, et al., 2010). *Leminorella grimontii*, *Hyphomicrobium denitrificans* and all *Methylophaga* species were exceptions, consistent with the inability of *H. denitrificans* and *Methylophaga* to consume organosulphur compounds with more than one carbon atom.

Isethionate desulphonation via isethionate sulphite-lyase IseA has been found in microcompartments of *Bilophila wadsworthia* (Burrichter et al., 2021). In accordance, HMSS2 detected the importer IseU and IseA in this organism. A similar desulphonation pathway without microcompartments was postulated for *Desulfovibrio alaskensis* and *D. desulfuricans* (Burrichter et al., 2021). In *D. desulfuricans*, HMSS2 also found IseU and IseA, suggesting that this organisms, like *B. wadsworthia*, may scavenge free isethionate via IseU. In contrast, *D. alaskensis* encodes IseA but not IseU. Instead, it contains sulphoacetaldehyde reductase IseD (or SarD), which is also present in *Bilophila wadsworthia*. In both cases, this enzyme may provide an endogenous source of isethionate (Burrichter et al., 2021).

Most analysed genomes possessed the potential for sulphopyruvate and (R)-sulpholactate generation from DHPS and (L)-sulpholactate. The potential of (R)-DHPS oxidation via HpsN generating 2 NADH equivalents was found in all analysed strains and most Ise encoded for isomerization of (S)-DHPS to (R)-DHPS via HpsP (17/24 genomes). The predicted presence of genes for desulphonation of sulphopyruvate by ComDE and sulpholactate by SuyAB as found here is also in accordance with previous reports for the Roseobacterales clade (Chen et al., 2021; Denger et al., 2009),

the Hyphomicrobiales (Chen et al., 2021), *Desulfovibrio desulfuricans* and *B. wadsworthia* (Hanson et al., 2021). Even without the ability to desulphonate sulphopyruvate or sulpholactate, the conversion of DHPS to sulphopyruvate or sulpholactate and export of these as end products provides 2–3 NADH equivalents and thus a growth advantage for the organism.

Sulphoquinovose degradation via the sulpho-Entner–Doudoroff pathway is present in eight bacteria, including *Pseudomonas putida* and other bacteria for which this pathway has been described or postulated (Felux et al., 2015). The complete sulphoquinovose degradation pathway based on a sulphoquinovose monooxygenase was found in seven proteobacteria in accordance with previous reports (Sharma et al., 2022). The other known sulphoquinovose degradation pathways were not detected, which is likely due to the presence of the Sulpho-Emden–Meyerhof–Parnas pathway (Denger et al., 2014) primarily in Enterobacterales and the transaldolase-dependent sulphoquinovose degradation in Firmicutes (Frommeyer et al., 2020). Bacteria from these taxonomic groups were not included in the case study.

DMS degradation has been described for *Methylophaga thiooxydans*, *Methylophaga sulfidovorans* (Kröber & Schäfer, 2019), *Hyphomicrobium denitrificans* (Koch & Dahl, 2018), and *Hyphomicrobium sulfonivorans* (Boden et al., 2011). According to our HMSS2 analysis, *H. sulfonivorans* encoded for DmoA, while all other three encoded only for methanethiol oxidase MtoX. DmoA was missing and the organisms must contain a so far unknown DMS monooxygenase. In accordance with previous reports, MtoX was also found in *Methylacidiphilum fumariolicum* (Schmitz et al., 2022), and several Roseobacterales, including *Ruegeria pomeroyi* (Eyice et al., 2017). The latter is a known degrader of DMSP to methanethiol via DmdA, B, C and DmdD (Reisch et al., 2011) which were all detected by the HMMs created here.

In summary, our case study on characterized organosulphur compound degraders has shown that in all cases the detection by HMSS2 agrees with the published analyses of other authors.

### 3.14 | HMSS2 improvements

HMSS2 has a redesigned engine and additional features for protein annotation and output format customization (Figure 5). Proteins with multiple domains are now stored with all domains and not just the domain with the highest score. This was accomplished by improving the local relational database structure. This requires that the recognized domain regions in the primary sequence do not overlap, so that domains with high scores are not overwritten by lower scores. On the other hand, high-scoring domains may still overwrite one or more lower-scoring domains during annotation.

Gene arrangement can now be used by HMSS2 for annotation as a nonhomologous criterion. Hits below the threshold are also considered and annotated if they lie within a gene cluster and the

potentially assigned annotation would complete a known gene cluster arrangement. Thus, a gene that highly likely occurs within a gene cluster must reach a lower cutoff than normal to be detected if it is encoded within such a cluster.

The output formats have been greatly expanded, and new features were added to improve usability and readability. It is still possible to retrieve sequences filtered by protein type, the genomic proximity and the presence of proteins or gene clusters in the same genome. HMSS2 automatically recovers a list of all hits with genomic features and a separate protein sequence file in FASTA format. Additionally, two subsets of the latter file are created. One subset includes all hits that are unique to their genome, respectively, while another subset includes all hits that occur at least twice in the same genome. Multidomain proteins, retrieved by the requested protein type, are listed separately if at least one other domain has been annotated.

An output module for iTol compatible data sets was also included. This module integrates the generation of iTol data sets for the presence/absence of the keywords/domains for each

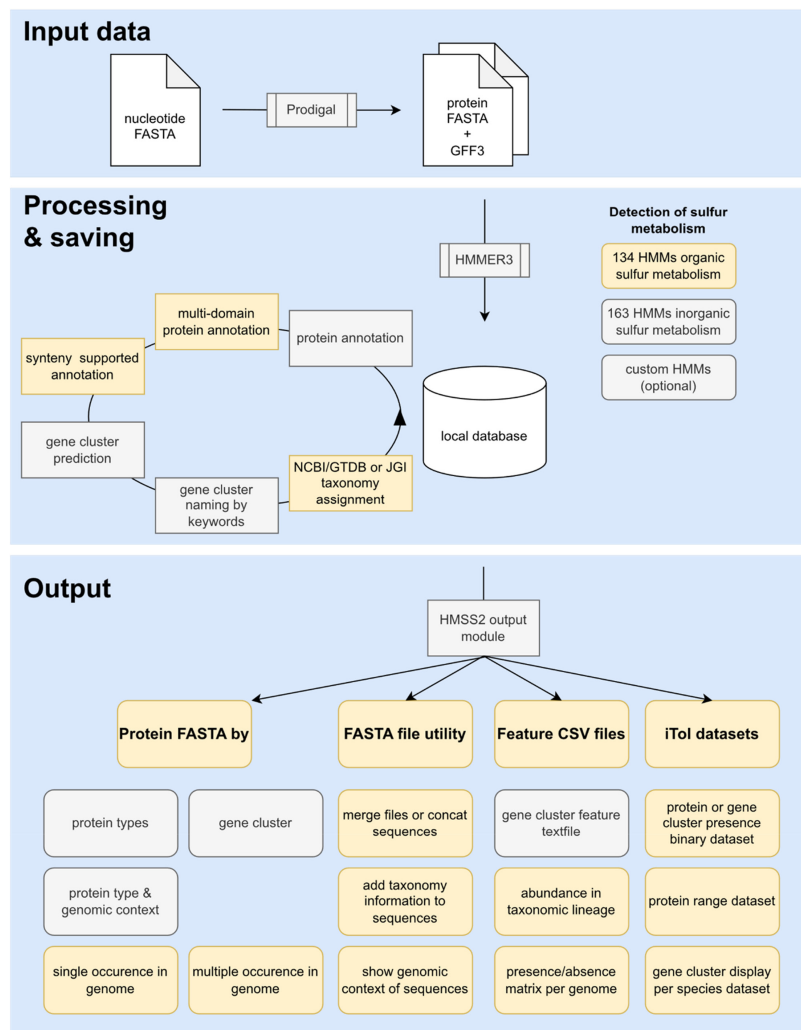
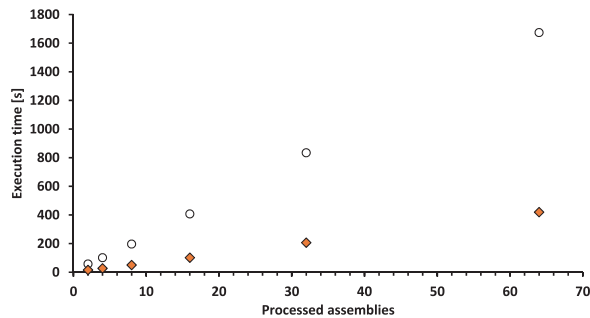


FIGURE 5 Algorithm overview of HMSS2. New features added in HMSS2 are highlighted in yellow. The only external programs required are HMMER3 and Prodigal.



**FIGURE 6** Computing time required by HMS-S-S compared to HMSS2. Test were performed in triplicate with defined numbers of randomly selected sulphur-oxidizing or sulphur-reducing prokaryotes and 164 HMMs. White circles: HMS-S-S, orange diamonds: HMSS2.

genome. Range data sets, which mark specific proteins in a phylogenetic tree, can now also be generated by HMSS2, as well as iTol-compatible data sets for displaying gene clusters. HMSS2 also comes with several utilities to modify the output protein FASTA files. It is now possible to assign the taxonomic name of the source organism to each sequence. Files can now be filtered by length, merged without duplicating sequence identifiers and sequences from multiple FASTA files originating from the same organism can be concatenated into a single sequence. With a FASTA-formatted file as input, a list of neighbouring genes is now accessible to support searches for conserved but previously undiscovered gene constellations.

The execution time of the HMSS2 was compared to that of HMS-S-S to demonstrate the scalability and efficiency of HMSS2. For this test, increasing numbers of genomes were randomly selected from the assemblies of the training data set and gene clusters were annotated and determined with the 164 HMMs of the original library. Time measurements were performed in triplicate with random selection of input assemblies for each replicate. The execution time was then averaged over all replicates. Comparison between the two versions showed a large difference in the required execution time (Figure 6; Table S4). The observed increase in execution speed for HMSS2 became more significant as the number of genomes processed increased and scaled linearly with the number of input assemblies. While HMS-S-S required around 26 min to process 64 assemblies, HMSS2 needed only 7 min for this task. Thus, the introduced improvements led to a fourfold accelerated computation speed for HMSS2.

## 4 | DISCUSSION

Here, we present a substantial update that provides an HMM-based search tool for proteins involved in the metabolism of inorganic and organic sulphur compounds. The high accuracy of the advanced tool presented here provides a reliable basis for genome analysis and is

further supported by the genomic context detection. The HMSS2 algorithm now uses homologous and nonhomologous criteria already in the protein annotation step, not just for the later identification of gene clusters. In addition, the overall execution time was accelerated by fourfold compared to the previous version, further speeding up the detection of sulphur metabolism pathways in genomes and metagenomes. With the increasing number of available genomes, faster protein annotation is required to handle the immense amount of available data.

We also significantly broadened the applicability of HMSS2 by adding the conversion of sulphonated carbon compounds. HMSS2 now covers pathways from the entire sulphur cycle, enabling studies on the link between the cycles of inorganic and organic sulphur compounds. In addition to providing operon structure information to support equivalence prediction, the accessibility and display of the annotated proteins has been greatly enhanced. Not only can sequences now be filtered by annotation but also the presence of genes and genomic context can be displayed using other specialized applications, further extending the capabilities of synteny analysis. Such analyses are not only limited to studies of the ecological role of prokaryotes but also include the evolution of metabolic pathways (Garcia et al., 2022), distribution of new pathways (Sharma et al., 2022) and genomic context visualization (Garcia et al., 2019; Letunic & Bork, 2021).

The expansion to the metabolism of organic sulphur compounds resulted in the generation of 134 additional HMMs in addition to the 164 HMMs previously included in HMS-S-S, almost doubling the total number of proteins included. The accuracy of the newly generated HMMs and the respective thresholds were demonstrated by cross-validation and a test data set. Observed deviations between both testing methods are likely due to an uneven distribution and abundance of protein sequences influencing the number and diversity of testable sequences. The quality of the 134 novel HMMs was ensured by selection of high-quality genomes derived from the RefSeq and GenBank databases. The overall development process had already been successfully applied for the proteins of inorganic sulphur metabolism (Tanabe & Dahl, 2022). The test data set was obtained from the full diversity of phyla accessible from GenBank and should therefore reflect the widest possible range of sequence variation. However, although the cutoff values have been validated, they are likely to need adjustment for newly discovered phyla (Anantharaman et al., 2018; Jaffe et al., 2020).

The diversity of proteins involved in the metabolism of organic sulphur compounds covered by HMSS2 also includes less prominent pathways for degradation and conversion of compounds such as sulphoquinovose or DMS. Although a considerable proportion of sulphur in the biosphere is bound in substrates or intermediates of these pathways, they are not commonly included in annotation pipelines and often unrecognized or incorrectly annotated. This is illustrated by fact that only 16 of the 124 proteins included here for the conversion of sulphoquinovose, taurine, isethionate or DMSP have an exact counterpart in PFAM (El-Gebali et al., 2019)

or TIGRFAMs. In contrast, eight of ten HMMs covering sulphate assimilation for cysteine biosynthesis have a TIGRFAM equivalent. A common problem in the functional annotation of enzymes involved in metabolism of organic sulphur compounds are enzymes, such as DmsA or DorA, that belong to the DMSO reductase superfamily. This family includes tetrathionate reductase, polysulphide reductase and thiosulphate reductase, as well as several other proteins unrelated to sulphur metabolism. Tertiary structure and complex composition is conserved throughout all members of this family (McEwan et al., 2010) and substrate specificity may only arise through a small number of conserved amino acids at the active site (Struwe et al., 2021). The validation performed here showed that related complexes in the DMSO reductase family did not negatively affect the HMMs for DmsA and DorA. Furthermore, the reliability of prediction is raised when genomic context is paired with the prediction made by the HMM detection as already discussed above.

## 5 | CONCLUSIONS

In summary, HMSS2 is an advanced comprehensive HMM-based tool for annotation and synteny analysis of prokaryotic sulphur metabolism. It has a higher speed and a much wider coverage than its predecessor HMS-S-S and now includes proteins involved in the metabolism of inorganic and organic sulphur compounds. The use of curated functionally equivalent sequences for HMM training resulted in HMMs with high precision and recall. This also fills a gap in the coverage of sulphur metabolism prediction by HMMs. The application possibilities also include the combination with other HMMs from public databases or user-defined models and can therefore be extended according to the user's needs. The improved output formats are also applicable to ecology and evolutionary research.

## AUTHOR CONTRIBUTIONS

TST and CD conceived the study. TST developed and implemented the method, performed the analyses and analysed and interpreted the data. Both authors wrote and approved the final version of the manuscript.

## ACKNOWLEDGEMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (grant Da 351/13-1 to CD). TST received a scholarship from the Studienstiftung des Deutschen Volkes. Open Access funding enabled and organized by Projekt DEAL.

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no competing interests.

## DATA AVAILABILITY STATEMENT

HMSS2 program files are available at <https://github.com/TSTanabe/HMSS2>.

## ORCID

Tomohisa Sebastian Tanabe  <https://orcid.org/0000-0003-2154-7980>

[org/0000-0003-2154-7980](https://orcid.org/0000-0003-2154-7980)

Christiane Dahl  <https://orcid.org/0000-0001-8288-7546>

## REFERENCES

- Abola, A. P., Willits, M. G., Wang, R. C., & Long, S. R. (1999). Reduction of adenosine-5'-phosphosulfate instead of 3'-phosphoadenosine-5'-phosphosulfate in cysteine biosynthesis by *Rhizobium meliloti* and other members of the family Rhizobiaceae. *Journal of Bacteriology*, 181, 5280–5287. <https://doi.org/10.1128/JB.181.17.5280-5287.1999>
- Aguilar-Barajas, E., Diaz-Perez, C., Ramirez-Diaz, M. I., Riveros-Rosas, H., & Cervantes, C. (2011). Bacterial transport of sulfate, molybdate, and related oxyanions. *Biomaterials*, 24(4), 687–707. <https://doi.org/10.1007/s10534-011-9421-x>
- Anantharaman, K., Hausmann, B., Jungbluth, S. P., Kantor, R. S., Lavy, A., Warren, L. A., Rappé, M. S., Pester, M., Loy, A., Thomas, B. C., & Banfield, J. F. (2018). Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME Journal*, 12, 1715–1728. <https://doi.org/10.1038/s41396-018-0078-0>
- Benning, C., & Somerville, C. R. (1992a). Identification of an operon involved in sulfolipid biosynthesis in *Rhodobacter sphaeroides*. *Journal of Bacteriology*, 174(20), 6479–6487. <https://doi.org/10.1128/jb.174.20.6479-6487.1992>
- Benning, C., & Somerville, C. R. (1992b). Isolation and genetic complementation of a sulfolipid-deficient mutant of *Rhodobacter sphaeroides*. *Journal of Bacteriology*, 174(7), 2352–2360. <https://doi.org/10.1128/jb.174.7.2352-2360.1992>
- Bick, J. A., Dennis, J. J., Zylstra, G. J., Nowack, J., & Leustek, T. (2000). Identification of a new class of 5'-adenylylsulfate (APS) reductases from sulfate-assimilating bacteria. *Journal of Bacteriology*, 182, 135–142.
- Bilous, P. T., & Weiner, J. H. (1985). Dimethyl sulfoxide reductase activity by anaerobically grown *Escherichia coli* HB101. *Journal of Bacteriology*, 162(3), 1151–1155. <https://doi.org/10.1128/jb.162.3.1151-1155.1985>
- Boden, R., Borodina, E., Wood, A. P., Kelly, D. P., Murrell, J. C., & Schäfer, H. (2011). Purification and characterization of dimethylsulfide monooxygenase from *Hyphomicrobium sulfonivorans*. *Journal of Bacteriology*, 193(5), 1250–1258.
- Boden, R., & Hutt, L. P. (2019). Bacterial metabolism of C<sub>1</sub> sulfur compounds. In F. Rojo (Ed.), *Aerobic utilization of hydrocarbons, oils and lipids. Handbook of hydrocarbon and lipid microbiology* (pp. 1–43). Springer Nature Switzerland AG.
- Boden, R., Kelly, D. P., Murrell, J. C., & Schäfer, H. (2010). Oxidation of dimethylsulfide to tetrathionate by *Methylophaga thiooxidans* sp. nov.: A new link in the sulfur cycle. *Environmental Microbiology*, 12(10), 2688–2699. <https://doi.org/10.1111/j.1462-2920.2010.02238.x>
- Borodina, E., Kelly, D. P., Rainey, F. A., Ward-Rainey, N. L., & Wood, A. P. (2000). Dimethylsulfone as a growth substrate for novel methylophilic species of *Hyphomicrobium* and *Arthrobacter*. *Archives of Microbiology*, 173(5–6), 425–437.
- Borodina, E., Kelly, D. P., Schumann, P., Rainey, F. A., Ward-Rainey, N. L., & Wood, A. P. (2002). Enzymes of dimethylsulfone metabolism and the phylogenetic characterization of the facultative methylophilic *Arthrobacter sulfonivorans* sp. nov., *Arthrobacter methylophilus* sp. nov., and *Hyphomicrobium sulfonivorans* sp. nov. *Archives of Microbiology*, 177(2), 173–183. <https://doi.org/10.1007/s00203-001-0373-3>
- Brodersen, K. H., Ong, C. S., Stephan, K. E., & Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. 2010 20th International Conference on Pattern Recognition, 3121–3124. <https://doi.org/10.1109/icpr.2010.764>

- Bruggemann, C., Denger, K., Cook, A. M., & Ruff, J. (2004). Enzymes and genes of taurine and isethionate dissimilation in *Paracoccus denitrificans*. *Microbiology*, 150(Pt 4), 805–816. <https://doi.org/10.1099/mic.0.26795-0>
- Bullock, H. A., Reisch, C. R., Burns, A. S., Moran, M. A., & Whitman, W. B. (2014). Regulatory and functional diversity of methylmercaptopyruvate coenzyme A ligases from the dimethylsulfoniopropionate demethylation pathway in *Ruegeria pomeroyi* DSS-3 and other proteobacteria. *Journal of Bacteriology*, 196(6), 1275–1285. <https://doi.org/10.1128/JB.00026-14>
- Burricher, A. G., Dorr, S., Bergmann, P., Haiss, S., Keller, A., Fournier, C., Franchini, P., Isono, E., & Schleheck, D. (2021). Bacterial microcompartments for isethionate desulfonation in the taurine-degrading human-gut bacterium *Bilophila wadsworthia*. *BMC Microbiology*, 21(1), 340. <https://doi.org/10.1186/s12866-021-02386-w>
- Carrion, O., Curson, A. R. J., Kumaresan, D., Fu, Y., Lang, A. S., Mercade, E., & Todd, J. D. (2015). A novel pathway producing dimethylsulfide in bacteria is widespread in soil environments. *Nature Communications*, 6, 6579. <https://doi.org/10.1038/ncomms7579>
- Carrion, O., Pratscher, J., Richa, K., Rostant, W. G., Farhan Ul Haque, M., Murrell, J. C., & Todd, J. D. (2019). Methanethiol and dimethylsulfide cycling in Stiffkey saltmarsh. *Frontiers in Microbiology*, 10, 1040. <https://doi.org/10.3389/fmicb.2019.01040>
- Chen, X., Liu, L., Gao, X., Dai, X., Han, Y., Chen, Q., & Tang, K. (2021). Metabolism of chiral sulfonate compound 2,3-dihydroxypropane-1-sulfonate (DHPS) by *Roseobacter* bacteria in marine environment. *Environment International*, 157, 106829. <https://doi.org/10.1016/j.envint.2021.106829>
- Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData Mining*, 10, 35. <https://doi.org/10.1186/s1304-0-017-0155-3>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6. <https://doi.org/10.1186/s12864-019-6413-7>
- Curson, A. R., Todd, J. D., Sullivan, M. J., & Johnston, A. W. (2011). Catabolism of dimethylsulphoniopropionate: Microorganisms, enzymes and genes. *Nature Reviews Microbiology*, 9(12), 849–859. <https://doi.org/10.1038/nrmicro2653>
- De Zwart, J., Sluis, J., & Kuennen, J. G. (1997). Competition for dimethyl sulfide and hydrogen sulfide by *Methylophaga sulfidovorans* and *Thiobacillus thioparus* T5 in continuous cultures. *Applied and Environmental Microbiology*, 63(8), 3318–3322. <https://doi.org/10.1128/aem.63.8.3318-3322.1997>
- Denger, K., Mayer, J., Buhmann, M., Weintschke, S., Smits, T. H., & Cook, A. M. (2009). Bifurcated degradative pathway of 3-sulfolactate in *Roseovarius nubinhibens* ISM via sulfoacetaldehyde acetyltransferase and (S)-cysteate sulfolylase. *Journal of Bacteriology*, 191(18), 5648–5656. <https://doi.org/10.1128/JB.00569-09>
- Denger, K., Weiss, M., Felux, A. K., Schneider, A., Mayer, C., Spitteller, D., Huhn, T., Cook, A. M., & Schleheck, D. (2014). Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature*, 507(7490), 114–117. <https://doi.org/10.1038/nature12947>
- Dhouib, R., Nasreen, M., Othman, D., Ellis, D., Lee, S., Essilfie, A. T., Hansbro, P. M., McEwan, A. G., & Kappler, U. (2021). The DmsABC sulfoxide reductase supports virulence in non-typeable *Haemophilus influenzae*. *Frontiers in Microbiology*, 12, 686833. <https://doi.org/10.3389/fmicb.2021.686833>
- Duarte, A. G., Barbosa, A. C. C., Ferreira, D., Manteigas, G., Domingos, R. M., & Pereira, I. A. C. (2021). Redox loops in anaerobic respiration—The role of the widespread NrfD protein family and associated dimeric redox module. *Biochimica et Biophysica Acta Bioenergetics*, 1862(7), 148416. <https://doi.org/10.1016/j.bbabi.2021.148416>
- Eichhorn, E., van der Ploeg, J. R., & Leisinger, T. (1999). Characterization of a two-component alkanesulfonate monooxygenase from *Escherichia coli*. *Journal of Biological Chemistry*, 274(38), 26639–26646. <https://doi.org/10.1074/jbc.274.38.26639>
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1), D427–D432. <https://doi.org/10.1093/nar/gky995>
- Eyice, O., Myronova, N., Pol, A., Carrión, O., Todd, J. D., Smith, T. J., Gurman, S. J., Cuthbertson, A., Mazard, S., Mennink-Kersten, M. A., Bugg, T. D., Andersson, K. K., Johnston, A. W., Op den Camp, H. J., & Schäfer, H. (2017). Bacterial SBP56 identified as a Cu-dependent methanethiol oxidase widely distributed in the biosphere. *ISME Journal*, 12(1), 145–160. <https://doi.org/10.1038/ismej.2017.148>
- Felux, A. K., Spitteller, D., Klebensberger, J., & Schleheck, D. (2015). Entner-Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proceedings of the National Academy of Sciences of the United States of America*, 112(31), E4298–E4305. <https://doi.org/10.1073/pnas.1507049112>
- Forman, G., & Scholz, M. (2010). Apples-to-apples in cross-validation studies. *ACM SIGDD Explorations Newsletter*, 12(1), 49–57. <https://doi.org/10.1145/1882471.1882479>
- Frommeyer, B., Fiedler, A. W., Oehler, S. R., Hanson, B. T., Loy, A., Franchini, P., Spitteller, D., & Schleheck, D. (2020). Environmental and intestinal phylum Firmicutes bacteria metabolize the plant sugar sulfoquinovose via a 6-deoxy-6-sulfofructose transaldolase pathway. *iScience*, 23(9), 101510. <https://doi.org/10.1016/j.isci.2020.101510>
- Garcia, P. S., D'Angelo, F., Ollagnier de Choudens, S., Dussouchaud, M., Bouveret, E., Gribaldo, S., & Barras, F. (2022). An early origin of iron-sulfur cluster biosynthesis machineries before earth oxygenation. *Nature Ecology & Evolution*, 6(10), 1564–1572. <https://doi.org/10.1038/s41559-022-01857-1>
- Garcia, P. S., Jauffrit, F., Grangeasse, C., & Brochier-Armanet, C. (2019). GeneSpy, a user-friendly and flexible genomic context visualizer. *Bioinformatics*, 35(2), 329–331. <https://doi.org/10.1093/bioinformatics/bty459>
- Goddard-Borger, E. D., & Williams, S. J. (2017). Sulfoquinovose in the biosphere: Occurrence, metabolism and functions. *Biochemical Journal*, 474(5), 827–849. <https://doi.org/10.1042/BCJ20160508>
- Gorzynska, A. K., Denger, K., Cook, A. M., & Smits, T. H. M. (2006). Inducible transcription of genes involved in taurine uptake and dissimilation by *Silicibacter pomeroyi* DSS-3<sup>T</sup>. *Archives of Microbiology*, 185(5), 402–406. <https://doi.org/10.1007/s00203-006-0106-8>
- Guler, S., Essigmann, B., & Benning, C. (2000). A cyanobacterial gene, *sqdX*, required for biosynthesis of the sulfolipid sulfoquinovosyldiacylglycerol. *Journal of Bacteriology*, 182(2), 543–545. <https://doi.org/10.1128/JB.182.2.543-545.2000>
- Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetverin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwatz, M., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Zheng, C., Thibaud-Nissen, F., Geer, L. Y., ... Pruitt, K. D. (2018). RefSeq: An update on prokaryotic genome annotation and curation. *Nucleic Acids Research*, 46(D1), D851–D860. <https://doi.org/10.1093/nar/gkx1068>
- Hanson, B. T., Dimitri Kits, K., Löffler, J., Burricher, A. G., Fiedler, A., Denger, K., Frommeyer, B., Herbold, C. W., Rattei, T., Karcher, N., Segata, N., Schleheck, D., & Loy, A. (2021). Sulfoquinovose is a select nutrient of prominent bacteria and a source of hydrogen sulfide in the human gut. *The ISME Journal*, 15, 2779–2791. <https://doi.org/10.1038/s41396-021-00968-0>
- Henriques, A. C., & De Marco, P. (2015). Methanesulfonate (MSA) catabolic genes from marine and estuarine bacteria. *PLoS One*, 10(5), e0125735. <https://doi.org/10.1371/journal.pone.0125735>

- Horinouchi, M., Yoshida, T., Nojiri, H., Yamane, H., & Omori, T. (1999). Polypeptide requirement of multicomponent monooxygenase DsoABCDEF for dimethyl sulfide oxidizing activity. *Bioscience, Biotechnology and Biochemistry*, 63(10), 1765–1771. <https://doi.org/10.1271/bbb.63.1765>
- Jaffe, A. L., Castelle, C. J., Matheus Carnevali, P. B., Gribaldo, S., & Banfield, J. F. (2020). The rise of diversity in metabolic platforms across the candidate phyla radiation. *BMC Biology*, 18(1), 69. <https://doi.org/10.1186/s12915-020-00804-5>
- Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing imbalanced data recommendations for the use of performance metrics. *Int Conf Affect Comput Intell Interact Workshops*, 2013, 245–251. <https://doi.org/10.1109/ACII.2013.47>
- Kappler, U., & Schäfer, H. (2014). Transformations of dimethylsulfide. *Metal Ions in Life Sciences*, 14, 279–313.
- Kelly, D. P., & Murrell, J. C. (1999). Microbial metabolism of methane-sulfonic acid. *Archives of Microbiology*, 172(6), 341–348. <https://doi.org/10.1007/s002030050770>
- Kiene, R. P., Linn, L. J., & Bruton, J. A. (2000). New and important roles for DMSP in marine microbial communities. *Journal of Sea Research*, 43(3–4), 209–224. [https://doi.org/10.1016/s1385-1101\(00\)00023-x](https://doi.org/10.1016/s1385-1101(00)00023-x)
- Koch, T., & Dahl, C. (2018). A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds. *ISME Journal*, 12(10), 2479–2491. <https://doi.org/10.1038/s41396-018-0209-7>
- Kredich, N. M. (1996). Biosynthesis of cysteine. In F. C. Neidhardt (Ed.), *Escherichia coli and Salmonella typhimurium. Cellular and molecular biology* (pp. 514–527). American Society for Microbiology.
- Krejci, Z., Hollemeyer, K., Smits, T. H. M., & Cook, A. M. (2010). Isethionate formation from taurine in *Chromohalobacter salexigens*: Purification of sulfoacetaldehyde reductase. *Microbiology*, 156(Pt 5), 1547–1555. <https://doi.org/10.1099/mic.0.036699-0>
- Kröber, E., & Schäfer, H. (2019). Identification of proteins and genes expressed by *Methylophaga thiooxydans* during growth on dimethylsulfide and their presence in other members of the genus. *Frontiers in Microbiology*, 10, 1132. <https://doi.org/10.3389/fmicb.2019.01132>
- Leimkühler, S., & Iobbi-Nivol, C. (2016). Bacterial molybdoenzymes: Old enzymes for new purposes. *FEMS Microbiology Reviews*, 40(1), 1–18. <https://doi.org/10.1093/femsre/fuv043>
- Letunic, I., & Bork, P. (2021). Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296. <https://doi.org/10.1093/nar/gkab301>
- Leyh, T. S., Taylor, J. C., & Markham, G. D. (1988). The sulfate activation locus of *Escherichia coli* K12: Cloning, genetic, and enzymatic characterization. *Journal of Molecular Chemistry*, 263, 2409–2416.
- Li, J., Koch, J., Flegler, W., Garcia Ruiz, L., Hager, N., Ballas, A., Tanabe, T. S., & Dahl, C. (2023). A metabolic puzzle: Consumption of C<sub>1</sub> compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1864, 148932. <https://doi.org/10.1016/j.bbabi.2022.148932>
- Liu, J., Wei, Y., Ma, K., An, J., Liu, X., Liu, Y., Ang, E. L., Zhao, H., & Zhang, Y. (2021). Mechanistically diverse pathways for sulfoquinovose degradation in bacteria. *ACS Catalysis*, 11(24), 14740–14750. <https://doi.org/10.1021/acscatal.1c04321>
- Lovelock, J. E., Maggs, R. J., & Rasmussen, R. A. (1972). Atmospheric dimethyl sulphide and the natural Sulphur cycle. *Nature*, 237(5356), 452–453. <https://doi.org/10.1038/237452a0>
- Mayer, J., Huhn, T., Habeck, M., Denger, K., Hollemeyer, K., & Cook, A. M. (2010). 2,3-Dihydroxypropane-1-sulfonate degraded by *Cupriavidus pinatubonensis* JMP134: Purification of dihydroxypropanesulfonate 3-dehydrogenase. *Microbiology*, 156(Pt 5), 1556–1564. <https://doi.org/10.1099/mic.0.037580-0>
- McDevitt, C. A., Hanson, G. R., Noble, C. J., Cheesman, M. R., & McEwan, A. G. (2002). Characterization of the redox centers in dimethyl sulfide dehydrogenase from *Rhodovulum sulfidophilum*. *Biochemistry*, 41(51), 15234–15244.
- McDevitt, C. A., Hugenholtz, P., Hanson, G. R., & McEwan, A. G. (2002). Molecular analysis of dimethyl sulphide dehydrogenase from *Rhodovulum sulfidophilum*: Its place in the dimethyl sulphoxide reductase family of microbial molybdopterine-containing enzymes. *Molecular Microbiology*, 44(6), 1575–1587.
- McEwan, A. G., Hanson, G. R., & Bailey, S. (1998). Dimethylsulphoxide reductase from purple phototrophic bacteria: Structures and mechanism(s). *Biochemical Society Transactions*, 26, 390–396.
- McEwan, A. G., Ridge, J. P., McDevitt, C. A., & Hugenholtz, P. (2010). The DMSO reductase family of microbial molybdenum enzymes; molecular properties and role in the dissimilatory reduction of toxic elements. *Geomicrobiology Journal*, 19(1), 3–21. <https://doi.org/10.1080/014904502317246138>
- Moran, M. A., & Durham, B. P. (2019). Sulfur metabolites in the pelagic ocean. *Nature Reviews Microbiology*, 17, 665–678. <https://doi.org/10.1038/s41579-019-0250-1>
- Peck, S. C., Denger, K., Burrichter, A., Irwin, S. M., Balskus, E. P., & Schleheck, D. (2019). A glycol radical enzyme enables hydrogen sulfide production by the human intestinal bacterium *Bilophila wadsworthia*. *Proceedings of the National Academy of Sciences of the United States of America*, 116(8), 3171–3176. <https://doi.org/10.1073/pnas.1815661116>
- Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In L. Liu & T. Özsu (Eds.), *Encyclopedia of database systems* (pp. 532–538). Springer.
- Reisch, C. R., Stoudemayer, M. J., Varaljay, V. A., Amster, I. J., Moran, M. A., & Whitman, W. B. (2011). Novel pathway for assimilation of dimethylsulphoniopropionate widespread in marine bacteria. *Nature*, 473(7346), 208–211. <https://doi.org/10.1038/nature10078>
- Rossak, M., Schafer, A., Xu, N., Gage, D. A., & Benning, C. (1997). Accumulation of sulfoquinovosyl-1-O-dihydroxyacetone in a sulfolipid-deficient mutant of *Rhodobacter sphaeroides* inactivated in *sqdC*. *Archives of Biochemistry and Biophysics*, 340(2), 219–230. <https://doi.org/10.1006/abbi.1997.9931>
- Rossak, M., Tietje, C., Heinz, E., & Benning, C. (1995). Accumulation of UDP-sulfoquinovose in a sulfolipid-deficient mutant of *Rhodobacter sphaeroides*. *Journal of Biological Chemistry*, 270(43), 25792–25797. <https://doi.org/10.1074/jbc.270.43.25792>
- Ruff, J., Denger, K., & Cook, A. M. (2003). Sulphoacetaldehyde acetyltransferase yields acetyl phosphate: Purification from *Alcaligenes defragrans* and gene clusters in taurine degradation. *Biochemical Journal*, 369(Pt 2), 275–285. <https://doi.org/10.1042/BJ20021455>
- Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Research*, 47(D1), D94–D99. <https://doi.org/10.1093/nar/gky989>
- Schäfer, H., Myronova, N., & Boden, R. (2010). Microbial degradation of dimethylsulphide and related C<sub>1</sub>-Sulphur compounds: Organisms and pathways controlling fluxes of Sulphur in the biosphere. *Journal of Experimental Botany*, 61(2), 315–334. <https://doi.org/10.1093/jxb/erp355>
- Schmitz, R. A., Mohammadi, S. S., van Erven, T., Berben, T., Jetten, M. S. M., Pol, A., & Op den Camp, H. J. M. (2022). Methanethiol consumption and hydrogen sulfide production by the thermoacidophilic methanotroph *Methylacidiphilum fumarolicum* SolV. *Frontiers in Microbiology*, 13, 857442. <https://doi.org/10.3389/fmicb.2022.857442>
- Sharma, M., Lingford, J. P., Petricevic, M., Snow, A. J. D., Zhang, Y., Jarva, M. A., Mui, J. W., Scott, N. E., Saunders, E. C., Mao, R., Epa, R., da Silva, B. M., Pires, D. E. V., Ascher, D. B., McConville, M. J., Davies, G. J., Williams, S. J., & Goddard-Borger, E. D. (2022). Oxidative desulfurization pathway for complete catabolism of sulfoquinovose by bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 119(4), e2116022119. <https://doi.org/10.1073/pnas.2116022119>

- Struwe, M. A., Kalimuthu, P., Luo, Z., Zhong, Q., Ellis, D., Yang, J., Khadanand, K. C., Harmer, J. R., Kirk, M. L., McEwan, A. G., Clement, B., Bernhardt, P. V., Kobe, B., & Kappler, U. (2021). Active site architecture reveals coordination sphere flexibility and specificity determinants in a group of closely related molybdoenzymes. *Journal of Biological Chemistry*, 296, 100672. <https://doi.org/10.1016/j.jbc.2021.100672>
- Tanabe, T. S., & Dahl, C. (2022). HMS-S-S: A tool for the identification of Sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies. *Molecular Ecology Resources*, 22(7), 2758–2774. <https://doi.org/10.1111/1755-0998.13642>
- Tanaka, Y., Yoshikaie, K., Takeuchi, A., Ichikawa, M., Mori, T., Uchino, S., Sugano, Y., Hakoshima, T., Takagi, H., Nonaka, G., & Tsukazaki, T. (2020). Crystal structure of a YeeE/YedE family protein engaged in thiosulfate uptake. *Science Advances*, 6(35), eaba7637. <https://doi.org/10.1126/sciadv.aba7637>
- Thume, K., Gebser, B., Chen, L., Meyer, N., Kieber, D. J., & Pohnert, G. (2018). The metabolite dimethylsulfoxonium propionate extends the marine organosulfur cycle. *Nature*, 563(7731), 412–415. <https://doi.org/10.1038/s41586-018-0675-0>
- Todd, J. D., Curson, A. R., Sullivan, M. J., Kirkwood, M., & Johnston, A. W. (2012). The *Ruegeria pomeroyi acul* gene has a role in DMSP catabolism and resembles *yhdH* of *E. coli* and other bacteria in conferring resistance to acrylate. *PLoS One*, 7(4), e35947. <https://doi.org/10.1371/journal.pone.0035947>
- Varma, S., & Simon, R. (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7, 91. <https://doi.org/10.1186/1471-2105-7-91>
- Wei, Y., Tong, Y., & Zhang, Y. (2022). New mechanisms for bacterial degradation of sulfoquinovose. *Bioscience Reports*, 42(10), BSR20220314. <https://doi.org/10.1042/BSR20220314>
- Weinitschke, S., Denger, K., Cook, A. M., & Smits, T. H. M. (2007). The DUF81 protein TauE in *Cupriavidus necator* H16, a sulfite exporter in the metabolism of C<sub>2</sub> sulfonates. *Microbiology*, 153, 3055–3060. <https://doi.org/10.1099/mic.0.2007/009845-0>
- Weinitschke, S., Hollemeyer, K., Kusian, B., Bowien, B., Smits, T. H., & Cook, A. M. (2010). Sulfoacetate is degraded via a novel pathway involving sulfoacetyl-CoA and sulfoacetaldehyde in *Cupriavidus necator* H16. *Journal of Biological Chemistry*, 285(46), 35249–35254. <https://doi.org/10.1074/jbc.M110.127043>
- Weinitschke, S., Sharma, P. I., Stingl, U., Cook, A. M., & Smits, T. H. (2010). Gene clusters involved in isethionate degradation by terrestrial and marine bacteria. *Applied and Environmental Microbiology*, 76(2), 618–621. <https://doi.org/10.1128/AEM.01818-09>
- Wicht, D. K. (2016). The reduced flavin-dependent monooxygenase SfnG converts dimethylsulfone to methanesulfinate. *Archives of Biochemistry and Biophysics*, 604, 159–166. <https://doi.org/10.1016/j.abb.2016.07.001>
- Wolf, P. G., Cowley, E. S., Breister, A., Matatov, S., Lucio, L., Polak, P., Ridlon, J. M., Gaskins, H. R., & Anantharaman, K. (2022). Diversity and distribution of sulfur metabolic genes in the human gut microbiome and their association with colorectal cancer. *Microbiome*, 10(1), 64. <https://doi.org/10.1186/s40168-022-01242-x>

#### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Tanabe, T. S., & Dahl, C. (2023). HMSS2: An advanced tool for the analysis of sulphur metabolism, including organosulphur compound transformation, in genome and metagenome assemblies. *Molecular Ecology Resources*, 23, 1930–1945. <https://doi.org/10.1111/1755-0998.13848>



Table S1: Reference proteins for dataset annotation

Pathway	HMM name	Uniprot Accession	Annotation/function	COG	Pfam family	TIGRfam family	Reference
Methylated sulfur compounds							
DMS degradation	DddA	C8YX89	alcohol dehydrogenase		pf05199, pf00732		Curson, Todd, et al., 2011
	DddC	C8YX90	aldehyde dehydrogenase		pf00171	TIGR01722	Curson, Todd, et al., 2011
	DddD	C8YX85	DMSP lyase		pf02515		Curson, Todd, et al., 2011
	DddK	Q4FNM4	DMSP lyase	COG0662	pf07883		Peng et al., 2019
	DddL	Q3J6L0	DMSP lyase	COG0662	pf16867		Sullivan et al., 2011
	DddP	A35K19	DMSP lyase	COG0006	pf00557		Kirkwood et al., 2010
	DddQ	Q5LT18	DMSP lyase	COG1917	pf16867		Todd et al., 2011
	DddT	C8YX86	BCCT-type transporter		pf02028	TIGR00842	Curson, Todd, et al., 2011
	DddW	Q5LW89	DMSP lyase	COG0662	pf07883		Todd et al., 2012
	DddY	E7DDH2	DMSP lyase		pf16867		Curson, Sullivan, et al., 2011
	AcuN	C8YX88	Acrylate hydrolase		pf02515		Curson, Todd, et al., 2011
	AcuK	C8YX87	Acrylate hydrolase		pf00378		Curson, Todd, et al., 2011
	AcuI	Q3J6K9	Acrylate hydrolase		pf08240, pf00107	TIGR02823	Curson, Sullivan, et al., 2011
DMSP demethylase	DmdA	Q5LS57	Dimethylsulfoniopropionate demethylase	COG0404	pf01571, pf08669		Reisch et al., 2008
	DmdB	Q5LRT0	3-methylmercaptopyruvate-CoA ligase	COG0318	pf00501, pf13193		Bullock et al., 2014
	DmdC	Q5LLW7	3-methylmercaptopyruvate-CoA dehydrogenase	COG1960	pf00441, pf12806, pf02770, pf02771		Reisch et al., 2011
	DmdD	Q5LLW6	Methylthioacryloyl-CoA hydratase	COG1024	pf00378		Reisch et al., 2011
DMS	MddA	A0A0F6P9C0	Methanethiol S-methyltransferase		pf04191		Carrion et al., 2015
	DmoA	E9JFX9	DMS monooxygenase		pf00296	TIGR03860	Boden et al., 2011
	DmsA	P18775	DMSO reductase	COG0243	pf04879, pf00384, pf01568		Bilous et al., 1988
	DmsB	P18776	DMSO reductase	COG0437	pf13247, pf12800	TIGR02951	Bilous et al., 1988
	DmsC	P18777	DMSO reductase	COG3302	pf04976		Bilous et al., 1988
	DmsD	P69853					
	DorS	D5APK9	DMSO reductase regulator	COG2205	pf02518, pf00512, pf01627, pf00072		Kappler et al., 2002, Shaw et al., 1999
	DorR	D5APK8	DMSO reductase regulator		pf00072		Kappler et al., 2002, Shaw et al., 1999
	DorC	D5APK7	DMSO reductase	COG3005	pf03264		McEwan et al., 1998
	DorD	D5AP70	DMSO reductase	COG3381	pf02613		McEwan et al., 1998
	DorA	D5APK5	DMSO reductase	COG0243	pf00384, pf01568		McEwan et al., 1998
Assimilatory DMS monooxygenase	DsoA	Q32428	Assimilatory DMS monooxygenase		pf06099		Horinouchi et al., 1999
	DsoB	Q32429	Assimilatory DMS monooxygenase		pf02332		Horinouchi et al., 1999
	DsoC	Q32430	Assimilatory DMS monooxygenase		pf02406		Horinouchi et al., 1999
	DsoD	Q32431	Assimilatory DMS monooxygenase		pf02332		Horinouchi et al., 1999
	DsoE	Q32432	Assimilatory DMS monooxygenase		pf04663		Horinouchi et al., 1999
	DsoF	Q32433	Assimilatory DMS monooxygenase		pf00970		Horinouchi et al., 1999
Assimilatory DMSO2 monooxygenase	SnfG	Q3KC85	DMSO <sub>2</sub> monooxygenase		pf00296		Wicht, 2016
Assimilatory methanesulfonate monooxygenase	MsuE	Q3K9A2	methanesulfonate sulfonotase	COG0431	pf03358		Kertesz et al., 1999
	MsuD	Q911C2	methanesulfonate sulfonotase		pf00296	TIGR03565	Kertesz et al., 1999
	MsuC	Q911C3	methanesulfonate sulfonotase		pf00296		Kertesz et al., 1999
Dissimilatory Methanesulfonate monooxygenase	MsmA	Q9X404	Methanesulfonate monooxygenase, hydroxylase		pf13577		de Marco et al., 1999
	MsmB	Q9X405	Methanesulfonate monooxygenase, hydroxylase		pf13577		de Marco et al., 1999
	MsmC	P70752	Methanesulfonate monooxygenase, ferredoxin component		pf00355		de Marco et al., 1999
	MsmD	Q9X406	Methanesulfonate monooxygenase, reductase		pf00970, pf00111, pf00175		de Marco et al., 1999
	MsmE	Q9X403	Methanesulfonate ABC Transporter				de Marco et al., 1999
	MsmF	Q9X402	Methanesulfonate ABC Transporter		pf00528		de Marco et al., 1999
	MsmG	Q5F4T8	Methanesulfonate ABC Transporter		pf00005		de Marco et al., 1999
	MsmH	Q5F4T9	Methanesulfonate ABC Transporter		pf00528		de Marco et al., 1999
Isethionate & Taurine	IseI	Q5LQX3	Isethionate dehydrogenase		pf05199		Weintschke et al., 2010
	IseU	D1FK56	Isethionate transporter		pf00083		Weintschke et al., 2010
	IseK	Q5LQX4	Isethionate TRAP transporter		pf06808		Weintschke et al., 2010
	IseL	Q5LQX5	Isethionate TRAP transporter		pf04290		Weintschke et al., 2010
	IseM	Q5LQX6	Isethionate TRAP transporter		pf03480		Weintschke et al., 2010
	IseR	Q5LQX7	Isethionate dissimilation regulator		pf09339, pf01614		Weintschke et al., 2010
	TauA	Q47537	Taurine ABC transporter	COG4521	pf04069		van der Ploeg et al., 1996
	TauB	Q6RH47	Taurine ABC transporter		pf00005		Bruggemann et al., 2004
	TauC	Q47539	Taurine ABC transporter	COG0600	pf00528		van der Ploeg et al., 1996
	TauK	F7ZLE6	Taurine TRAP transporter		pf03480		Bruggemann et al., 2004
	TauL	F7ZLE7	Taurine TRAP transporter		pf04290		Bruggemann et al., 2004
	TauM	F7ZLE8	Taurine TRAP transporter		pf06808		Bruggemann et al., 2004
	TauD	P37610	Taurine dioxygenase	COG2175	pf02668		Eichhorn et al., 1997
	TauE	Q0K020	Sulfite exporter	COG0730	pf01925		Weintschke et al., 2007
	TauR	D5AKX9	Taurine dissimilation regulator		pf00155		Wiethaus et al., 2008
	TauX	F7ZLF0	Taurine dehydrogenase		pf12680		Bruggemann et al., 2004
	TauY	F7ZLF1	Taurine dehydrogenase		pf01266		Bruggemann et al., 2004
	TauZ	Q6RH51	Hypothetical protein				Bruggemann et al., 2004
	Tpa	Q9APM5	Taurine : pyruvate aminotransferase		pf00202		Laue & Cook, 2000
	Toa	Q1R186	Taurine : 2-oxoglutarate aminotransferase		pf00202		Krejci et al., 2010
	Xsc	Q84H44	Sulfoacetaldehyde acetyltransferase		pf00205	TIGR03457	Ruff et al., 2003
	IsfD	Q1R183	Sulfoacetaldehyde reductase	COG4221	pf00106		Krejci et al., 2010
	IsfE	Q1R184	Isethionate exporter	COG0730	pf01925		Krejci et al., 2010
	IsIA	E5Y378	Isethionate sulfite-lyase				Peck et al., 2019
	IsIB	E5Y377	Isethionate sulfite-lyase				Peck et al., 2019

DHPS, Sulfolactate, Sulfoacetate	HpsU	Q46N55	DHPS uptake	COG0477	pf07690	Mayer et al., 2010	
	HpsO	Q46N54	DHPS 2-dehydrogenases	COG1028		Mayer et al., 2010	
	HpsP	Q46N56	DHPS 2-dehydrogenases	COG1063	pf08240, pf16912	Mayer et al., 2010	
	HpsN	Q46N53	DHPS 3-dehydrogenase	COG0141	pf00815	Mayer et al., 2010	
	HpsK	A3SJ79	DHPS TRAP transporter	COG1638	pf03480	Denger et al., 2009	
	HpsL	A3SJ73	DHPS TRAP transporter	COG3090	pf04290	Denger et al., 2009	
	HpsM	A3SJ74	DHPS TRAP transporter	COG1593	pf06808	Denger et al., 2009	
	SuyA	Q1QWP1	(R)-Sulfolactate sulfo-lyase	COG2721	pf08666	Denger & Cook, 2010	
	SuyB	Q1QWP0	(R)-Sulfolactate sulfo-lyase	COG2721	pf08666	Denger & Cook, 2010	
	Pta	A3SR26	Phosphate acetyltransferase	COG0280	pf01515	TIGR00651	
	ComC	Q1QWN5	Sulfo-pyruvate decarboxylase	COG2055	pf02615	Denger & Cook, 2010	
	ComD	A3SN11	Sulfo-pyruvate decarboxylase	COG4032		TIGR03845	
	ComE	A3SN10	Sulfo-pyruvate decarboxylase	COG0028	pf02775	TIGR03846	
	SicH	A3SN09	Sulfolactate TRAP Transporter	COG3181	pf03401	Denger et al., 2009	
	SicF	A3SN08	Sulfolactate TRAP Transporter	COG3333	pf01970	Denger et al., 2009	
	SicG	A3SN07	Sulfolactate TRAP Transporter		pf07331	Denger et al., 2009	
	SicD	A3SN06	sulfolactate dehydrogenase		pf01070	Denger et al., 2009	
	CuyA	A3SQG3	L-cysteate sulfo-lyase	COG2515	pf00291	TIGR01275	
	CuyZ	A3SQG2	Hypothetical membrane protein	COG2855	pf03601	Denger et al., 2009	
	SicC	Q1QWNG	(S)-sulfolactate dehydrogenase	COG0111	pf00389, pf02826	Denger & Cook, 2010	
	SauS	Q0K845	Sulfoacetaldehyde dehydrogenase	COG1012	pf00171	Weinitschke et al., 2010	
	SauT	Q0K844	Sulfoacetate-CoA ligase	COG0318	pf00501, pf13193	Weinitschke et al., 2010	
	SauU	Q0K843	Sulfoacetate transporter	COG2271	pf07690	Weinitschke et al., 2010	
	Alkanesulfonate	SsuD	P80645	Alkanesulfonate monooxygenase	COG2141	pf00296	TIGR03565
				Alkanesulfonate monooxygenase, FMN reductase (NADPH)	COG0431	pf03358	TIGR03567
		SsuE	P80646	Alkanesulfonate monooxygenase, FMN reductase (NADPH)	COG0431	pf03358	TIGR03567
		SsuA	P75853	Alkanesulfonate ABC transporter	COG0715	pf09084	TIGR01728
SsuB		P0AA11	Alkanesulfonate ABC transporter	COG1116	pf00005	Eichhorn et al., 1999	
SsuC		P75851	Alkanesulfonate ABC transporter	COG0600	pf00528	Eichhorn et al., 1999	
Sulfoquinovose synthesis	SqdA	Q3J501	Phospholipid/glycerol acyltransferase	COG0204	pf01553	Benning & Somerville, 1992	
	SqdB	Q3J3A8	UDP-sulfoquinovose synthase	COG0451	pf01370	Benning & Somerville, 1992	
	SqdC	Q3J3B0	Putative sulfolipid biosynthesis protein	COG3825	pf05762	Benning & Somerville, 1992	
			Glycosyl transferase (Sulfolipid biosynthesis) protein	COG5597	pf01501	Benning & Somerville, 1992	
	SqdD	Q3J3A9	Glycosyl transferase (Sulfolipid biosynthesis) protein	COG5597	pf01501	Benning & Somerville, 1992	
Sulfoquinovose degradation	SqdX	Q9R6U1	SqdX	COG0438	pf13439, pf00534	Guler et al., 2000	
Sulfo-EMP	YihO	P32136	Putative sulfoquinovose importer	COG2211		TIGR00792	
			Putative 2,3-dihydroxypropane-1-sulfonate exporter	COG2211		TIGR00792	
	YihP	P32137	Putative 2,3-dihydroxypropane-1-sulfonate exporter	COG2211		Denger et al., 2014	
	YihQ	P32138	Sulfoquinovosidase	COG1501	pf01055	Okuyama et al., 2004	
	YihR	P32139	Mutarotase	COG2017	pf01263	Denger et al., 2014	
	YihS	P32140	Sulfoquinovose isomerase	COG2942		Denger et al., 2014	
	YihT	P32141	Sulfofructosephosphate aldolase	COG3684	pf01791	Denger et al., 2014	
	YihU	P0A9V8	3-sulfolactaldehyde reductase	COG2084	pf14833, pf03446	Denger et al., 2014	
	YihV	P32143	Sulfofructose kinase	COG0524	pf00294	Denger et al., 2014	
Sulfoquinovose degradation Sulfo-ED	SedE	P0DOV9	3-sulfolactaldehyde dehydrogenase		pf00171	Felix et al., 2015	
	SedC	P0DOV7	6-deoxy-6-sulfo-D-gluconate dehydratase		pf00920	Felix et al., 2015	
	SedA	P0DOV5	Sulfoquinovose 1-dehydrogenase			Felix et al., 2015	
	SedB	P0DOV6	6-deoxy-6-sulfo-gluconolactonase		pf08450	Felix et al., 2015	
			2-dehydro-3,6-dideoxy-6-sulfo-gluconate aldolase		pf03328	Felix et al., 2015	
	SedD	P0DOV8	aldolase		pf03328	Felix et al., 2015	
Sulfoquinovose degradation Sulfo-monooxygenase	SmoA	Q7CS24	flavin reductase			Sharma et al., 2022	
	SmoB	A9CEY6	Oxidoreductase			Sharma et al., 2022	
	SmoC	A9CEY7	alkanesulfonate monooxygenase			Sharma et al., 2022	
	SmoD	A9CEY8	GntR-like regulator			Sharma et al., 2022	
	SmoE	Q7CS28	ABC Transporter ATP binding domain			Sharma et al., 2022	
	SmoF	A9CEY9	ABC transporter substrate binding domain			Sharma et al., 2022	
	SmoG	Q7CS30	ABC transporter membrane spanning domain			Sharma et al., 2022	
	SmoH	Q7CS31	ABC transporter membrane spanning domain			Sharma et al., 2022	
SmoI	A9CEZ0	sulfoquinovosidase			Sharma et al., 2022		
Sulfoquinovose degradation transaldolase	SftA		sulfoquinovose importer			Frommeyer et al., 2020	
	SftT		6-deoxy-6-sulfofructose transaldolase			Frommeyer et al., 2020	
	SftX		unknown function			Frommeyer et al., 2020	
	SftG		SQG hydrolase			Frommeyer et al., 2020	
	SftI		SQ isomerase			Frommeyer et al., 2020	
	SftF		transcriptional regulator			Frommeyer et al., 2020	
	SftD		SLA dehydrogenase			Frommeyer et al., 2020	
Sulfate assimilation	CysM	P16703	Cystein synthase B		pf00291	TIGR01138	
			Sulfate/thiosulfate import ATP-binding protein				
	CysA	P16676	CysA		pf00005	TIGR00968	
			Sulfate transport system permease protein				
	CysW	P0AE80	CysW		pf00528		
	CysU	P16701	Sulfate transport system permease protein			Sirko et al., 1990	
	CysC	P0A611	Adenylyl-sulfate kinase	COG0529		TIGR00455	
	CysN	P23845	Sulfate adenylyltransferase subunit 1	COG2895	pf00009	TIGR02034	
	CysD	P21156	Sulfate adenylyltransferase subunit 2	COG0175	pf01507	TIGR02039	
	CysH	P17854	Phosphoadenosine 5'-phosphosulfate reductase	COG0175	pf1507	TIGR00434	
			Sulfitte reductase [NADPH] hemoprotein beta-component				
	CysI	P17846	Sulfitte reductase [NADPH] hemoprotein beta-component	COG0155	pf01077	TIGR02041	
		Sulfitte reductase [NADPH] flavoprotein alpha-component					
CysJ	P38038	Sulfitte reductase [NADPH] flavoprotein alpha-component	COG0369	pf00667, pf00258, pf00175	TIGR01931		
CysK	P0ABK5	Cystein synthase A	COG0031	pf00291	TIGR01139		
					Byrne et al., 1988		

## References

- Benning, C., & Somerville, C. R. (1992). Identification of an operon involved in sulfolipid biosynthesis in *Rhodobacter sphaeroides*. *Journal of Bacteriology*, 174(20), 6479-6487. <https://doi.org/10.1128/jb.174.20.6479-6487.1992>
- Bilus, P. T., Cole, S. T., Anderson, W. T., & Weiner, J. H. (1988). Nucleotide sequence of the *dmsABC* operon encoding the anaerobic dimethylsulphoxide reductase of *Escherichia coli*. *Molecular Microbiology*, 2, 785-795.
- Boden, R., Borodina, E., Wood, A. P., Kelly, D. P., Murrell, J. C., & Schäfer, H. (2011). Purification and characterization of dimethylsulfide monooxygenase from *Hyphomicrobium sulfonivorans*. *Journal of Bacteriology*, 193(5), 1250-1258.
- Bruggemann, C., Denger, K., Cook, A. M., & Ruff, J. (2004). Enzymes and genes of taurine and isethionate dissimilation in *Paracoccus denitrificans*. *Microbiology (Reading)*, 150(Pt 4), 805-816. <https://doi.org/10.1099/mic.0.26795-0>
- Bullock, H. A., Reisch, C. R., Burns, A. S., Moran, M. A., & Whitman, W. B. (2014). Regulatory and functional diversity of methylmercaptopyruvate coenzyme A ligases from the dimethylsulfoniopyruvate demethylation pathway in *Ruegeria pomeroyi* DSS-3 and other proteobacteria. *Journal of Bacteriology*, 196(6), 1275-1285. <https://doi.org/10.1128/JB.00026-14>
- Byrne, C. R., Monroe, R. S., Ward, K. A., & Kredich, N. M. (1988). DNA sequences of the *cysK* regions of *Salmonella typhimurium* and *Escherichia coli* and linkage of the *cysK* regions to *ptsH*. *Journal of Bacteriology*, 170(7), 3150-3157. <https://doi.org/10.1128/jb.170.7.3150-3157.1988>
- Carrion, O., Curson, A. R. J., Kumaresan, D., Fu, Y., Lang, A. S., Mercade, E., & Todd, J. D. (2015). A novel pathway producing dimethylsulphide in bacteria is widespread in soil environments. *Nature Communications*, 6, 6579. <https://doi.org/10.1038/ncomms7579>
- Curson, A. R., Sullivan, M. J., Todd, J. D., & Johnston, A. W. (2011). DddY, a periplasmic dimethylsulfoniopyruvate lyase found in taxonomically diverse species of Proteobacteria. *ISME Journal*, 5(7), 1191-1200. <https://doi.org/10.1038/ismej.2010.203>
- Curson, A. R., Todd, J. D., Sullivan, M. J., & Johnston, A. W. (2011). Catabolism of dimethylsulfoniopyruvate: microorganisms, enzymes and genes. *Nature Reviews Microbiology*, 9(12), 849-859. <https://doi.org/10.1038/nrmicro2653>
- de Marco, P., Moradas-Ferreira, P., Higgins, T. P., McDonald, I., Kenna, E. M., & Murrell, J. C. (1999). Molecular analysis of a novel methanesulfonic acid monooxygenase from the methylotroph *Methylosulfonomonas methylolora*. *Journal of Bacteriology*, 181(7), 2244-2251. <https://doi.org/10.1128/JB.181.7.2244-2251.1999>
- Denger, K., & Cook, A. M. (2010). Racemase activity effected by two dehydrogenases in sulfolactate degradation by *Chromohalobacter salexigens*: purification of (S)-sulfolactate dehydrogenase. *Microbiology (Reading)*, 156(Pt 3), 967-974. <https://doi.org/10.1099/mic.0.034736-0>
- Denger, K., Mayer, J., Buhmann, M., Weinitzschke, S., Smits, T. H., & Cook, A. M. (2009). Bifurcated degradative pathway of 3-sulfolactate in *Roseovarius nubinhibens* ISM via sulfoacetaldehyde acetyltransferase and (S)-cysteate sulfolase. *Journal of Bacteriology*, 191(18), 5648-5656. <https://doi.org/10.1128/JB.00569-09>
- Denger, K., Weiss, M., Felix, A. K., Schneider, A., Mayer, C., Spitteller, D., Huhn, T., Cook, A. M., & Schleheck, D. (2014). Sulphoglycolysis in *Escherichia coli* K-12 closes a gap in the biogeochemical sulphur cycle. *Nature*, 507(7490), 114-117. <https://doi.org/10.1038/nature12947>
- Eichhorn, E., van der Ploeg, J. R., Kertesz, M. A., & Lelsinger, T. (1997). Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from *Escherichia coli*. *Journal of Biological Chemistry*, 272(37), 23031-23036. <https://doi.org/10.1074/jbc.272.37.23031>
- Eichhorn, E., van der Ploeg, J. R., & Lelsinger, T. (1999). Characterization of a two-component alkanesulfonate monooxygenase from *Escherichia coli*. *Journal of Biological Chemistry*, 274(38), 26639-26646. <https://doi.org/10.1074/jbc.274.38.26639>
- Felix, A. K., Spitteller, D., Klebensberger, J., & Schleheck, D. (2015). Entner-Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proceedings of the National Academy of Sciences of the United States of America*, 112(31), E4298-4305. <https://doi.org/10.1073/pnas.1507049112>
- Frommeyer, B., Fiedler, A. W., Oehler, S. R., Hanson, B. T., Loy, A., Franchini, P., Spitteller, D., & Schleheck, D. (2020). Environmental and intestinal phylum Firmicutes bacteria metabolize the plant sugar sulfoquinovose via a 6-deoxy-6-sulfofructose transaldolase pathway. *iScience*, 23(9), 101510. <https://doi.org/10.1016/j.isci.2020.101510>
- Güler, S., Essigmann, B., & Benning, C. (2000). A cyanobacterial gene, *sqdX*, required for biosynthesis of the sulfolipid sulfoquinovosyl diacylglycerol. *Journal of Bacteriology*, 182(2), 543-545. <https://doi.org/10.1128/JB.182.2.543-545.2000>
- Horinouchi, M., Yoshida, T., Nojiri, H., Yamane, H., & Omori, T. (1999). Polypeptide requirement of multicomponent monooxygenase DsoABCDEF for dimethyl sulfide oxidizing activity. *Bioscience, biotechnology and biochemistry*, 63(10), 1765-1771. <https://doi.org/10.1271/bbb.63.1765>
- Kappler, U., Huston, W. M., & McEwan, A. G. (2002). Control of dimethylsulfide reductase expression in *Rhodobacter capsulatus*: the role of carbon metabolites and the response regulators DorR and RegA. *Microbiology (Reading)*, 148(Pt 2), 605-614. <https://doi.org/10.1099/00221287-148-2-605>
- Kertesz, M. A., Schmidt-Larbig, K., & Wuest, T. (1999). A novel reduced flavin mononucleotide-dependent methanesulfonate sulfonase encoded by the sulfur-regulated *msu* operon of *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 181(5), 1464-1473. <https://doi.org/10.1128/JB.181.5.1464-1473.1999>
- Kirkwood, M., Le Brun, N. E., Todd, J. D., & Johnston, A. W. B. (2010). The *dddP* gene of *Roseovarius nubinhibens* encodes a novel lyase that cleaves dimethylsulfoniopyruvate into acrylate plus dimethyl sulfide. *Microbiology (Reading)*, 156(Pt 6), 1900-1906. <https://doi.org/10.1099/mic.0.038927-0>
- Krejciak, Z., Hollemeyer, K., Smits, T. H. M., & Cook, A. M. (2010). Isethionate formation from taurine in *Chromohalobacter salexigens*: purification of sulfoacetaldehyde reductase. *Microbiology (Reading)*, 156(Pt 5), 1547-1555. <https://doi.org/10.1099/mic.0.036699-0>
- Laue, H., & Cook, A. M. (2000). Biochemical and molecular characterization of taurine:pyruvate aminotransferase from the anaerobe *Bilophila wadsworthii*. *European Journal of Biochemistry*, 267(23), 6841-6848. <https://doi.org/10.1046/j.1432-1033.2000.01782.x>
- Leyh, T. S., Vogt, T. F., & Suo, Y. (1992). The DNA sequence of the sulfate activation locus from *Escherichia coli* K-12. *Journal of Biological Chemistry*, 267, 10405-10410.
- Mayer, J., Huhn, T., Habek, M., Denger, K., Hollemeyer, K., & Cook, A. M. (2010). 2,3-Dihydroxypropane-1-sulfonate degraded by *Cupriavidus pinatubonensis* JMP134: purification of dihydroxypropanesulfonate 3-dehydrogenase. *Microbiology (Reading)*, 156(Pt 5), 1556-1564. <https://doi.org/10.1099/mic.0.037580-0>
- McEwan, A. G., Hanson, G. R., & Bailey, S. (1998). Dimethylsulphoxide reductase from purple phototrophic bacteria: structures and mechanism(s). *Biochemical Society Transactions*, 26, 390-396.

**Table S2: HMM performance evaluation**  
**10-fold cross validation on training dataset**

Model	Metrics						Confusion matrix				
	MCC	F1-score	Balanced accuracy	Cohens kappa	Precision	Recall	True positives	False positives	False negatives	True negatives	
AcuI	0,99789225	0,9978903	0,99789474	0,99789003		1	0,99578947	946	0	4	7406310
AcuK	0,86364754	0,85487977	0,9993158	0,85453287	0,74691053	0,99933862	15110	5120	10	7236490	
AcuN	0,9536098	0,95269704	0,99916719	0,95260087	0,91089172	0,99852424	14209	1390	21	7319620	
ComC	0,99398169	0,99396378	994	0,99396358		1	988	247	0	3	7403780
ComD	0,96362346	0,96296296	0,96428571	0,96296229		1	0,92857143	130	0	10	7422220
ComE	0,98922749	0,98916968	0,98928571	0,98916947		1	0,97857143	137	0	3	7420710
CuyA	0,96076827	0,96	0,96153846	0,95999933		1	0,92307692	120	0	10	7412320
CuyZ	0,95991	0,95910781	0,96071429	0,95910707		1	0,92142857	129	0	11	7412460
CysA	0,89166289	0,89078097	0,92611464	0,89075048	0,93296433	0,85224719	1517	109	263	6087121	
CysC	0,94612185	0,94468873	0,94758772	0,94467264		1	0,89517544	2041	0	239	7405500
CysD	0,99987973	0,99987979	0,99987981	0,99987973		1	0,99975962	4159	0	1	7395740
CysH	0,99504876	0,99505139	0,99758748	0,99504876	0,99492515	0,99517766	3921	20	19	7402750	
CysI	0,99589229	0,99588888	0,99962297	0,99588665	0,99255029	0,99925	3997	30	3	7402220	
CysJ	0,9971292	0,99712719	0,99973754	0,99712644	0,99478895	0,99947644	1909	10	1	7402660	
CysM	0,98822785	0,98815931	0,98829787	0,98815856		1	0,97659574	459	0	11	7339290
CysN	0,99639087	0,99638719	0,9998731	0,99638523	0,99304693	0,99975	3999	28	1	7372632	
CysU	0,99907344	0,99907322	0,99907407	0,99907301		1	0,99814815	1617	0	3	7381730
CysW	0,99657726	0,99657214	0,99658385	0,9965714		1	0,9931677	1599	0	11	7406260
DddA	0,99908817	0,99908842	0,99908925	0,99908775		1	0,99817851	5480	0	10	7457840
DddD	0,99628928	0,99628253	0,9962963	0,99628239		1	0,99259259	268	0	2	7464200
DddL	0,991137	0,99109792	0,99117647	0,99109772		1	0,98235294	167	0	3	7465510
DddP	0,99658488	0,99657925	0,99659091	0,99657905		1	0,99318182	437	0	3	7463740
DddT	0,98742068	0,98734177	0,9875	0,98734157		1	975	117	0	3	7423010
DddY	0,98601297	0,98591549	0,98611111	0,98591516		1	0,97222222	175	0	5	7465500
DdhA	0,9527004	0,95196507	0,99545388	0,95196433	0,91596639	0,99090909	109	10	1	7465210	
DdhB	0,99974622	0,99974626	0,99999993	0,99974619	0,99949264		1	1970	1	0	7455389
DmdA	1	1	1	1		1	1	2190	0	0	7442890
DmdB	0,91245953	0,90986636	0,99226827	0,90982381	0,84566596	0,98461538	3200	584	50	7406046	
DmdC	1	1	1	1		1	1	2400	0	0	7322970
DmdD	0,99350688	0,99348727	0,99352941	0,9934858		1	0,98705882	1678	0	22	7457530
DmoA	0,97833658	0,97810219	0,97857143	0,97810199		1	0,95714286	67	0	3	7430070
DmsA	0,99813885	0,9981378	0,99814126	0,99813712		1	0,99628253	2680	0	10	7342500
DmsB	0,99813191	0,99813084	0,99813433	0,99813016		1	0,99626866	2670	0	10	7381650
DmsC	0,99951898	0,999519	0,99951923	0,99951886		1	0,99903846	2078	0	2	7420650
DmsD	0,99923809	0,999238	0,99923858	0,9992378		1	0,99847716	1967	0	3	7420400
DorA	0,99373028	0,99371069	0,99375	0,99371062		1	0,9875	79	0	1	7394540
DorC	0,98260671	0,98245614	0,98275862	0,98245547		1	0,96551724	280	0	10	7414430
DorR	0,9893366	0,98928025	0,98939394	0,98927975		1	0,97878788	323	0	7	7075320
DorS	0,97647069	0,97626113	0,99848379	0,97626	0,95639535	0,9969697	329	15	1	7068985	
DsoA	0,99791442	0,99791232	0,99791667	0,99791225		1	0,99583333	239	0	1	7422540
DsoB	0,99582395	0,9958159	0,99583333	0,99581523		1	0,99166667	1190	0	10	7421560
DsoC	0,99946788	0,9994678	0,99946809	0,99946773		1	0,99893617	939	0	1	7421830
DsoD	0,99879908	0,99879856	0,9988	0,99879836		1	0,9976	1247	0	3	7420530
DsoE	0,98907004	0,98901099	0,98913043	0,98901032		1	0,97826087	450	0	10	7422320
DsoF	0,99152131	0,991522	0,99516068	0,99152058	0,99272433	0,99032258	1228	9	12	7391041	
HpsK	0,94556006	0,94555874	0,97142736	0,94555618	0,94827586	0,94285714	330	18	20	7422022	
HpsL	0,988165	0,98809524	0,98823529	0,98809497		1	0,97647059	166	0	4	7411500
HpsM	0,9701412	0,96969697	0,99999866	0,96969563	0,94117647		1	320	20	0	7465190
HpsN	0,99030313	0,99025686	0,99035088	0,99025612		1	0,98070175	559	0	11	7381890
HpsO	0,64489211	0,64492754	0,82445481	0,64487966	0,6409465	0,64895833	623	349	337	7164021	
HpsP	0,99838559	0,99838449	0,9983871	0,99838429		1	0,99677419	927	0	3	7359040
HpsU	0,51224668	0,50909091	0,78635353	0,50907463	0,45818182	0,57272727	126	149	94	7372471	
IseI	0,73106738	0,69664984	0,76725352	0,69660946		1	0,53450704	759	0	661	7432050
IseK	0,97681383	0,97654584	0,97708333	0,9765451		1	0,95416667	229	0	11	7422470
IseL	0,99564255	0,99563319	0,99565217	0,99563305		1	0,99130435	228	0	2	7422550
IseM	0,97979524	0,97959184	0,98	0,97959116		1	0,96	240	0	10	7417310
IseR	0,98885968	0,9888	0,98892405	0,98879763		1	0,9778481	1545	0	35	7390700
IseU	0,18428822	0,16238318	0,65429188	0,16223218	0,11014263	0,30888889	278	2246	622	7358804	
IsfD	0,94963094	0,94901316	0,9579359	0,94900895	0,98464164	0,91587302	577	9	53	7365891	
IsfE	0,99163145	0,99159664	0,99166667	0,99159644		1	0,98333333	177	0	3	7422540
IsIA	0,99819523	0,99819495	0,9981982	0,9981936		1	0,9963964	5530	0	20	7443950
IsIB	0,99990541	0,99990547	0,99990548	0,99990541		1	0,99981096	5289	0	1	7450010
MddA	1	1	1	1		1	1	100	0	0	7463510
MsmA	0,89442659	0,88888889	0,9	0,88888822		1	0,8	40	0	10	7407390
MsmB	0,89442659	0,88888889	0,9	0,88888822		1	0,8	40	0	10	7422710
MsmE	0,98198018	0,98181818	0,98214286	0,98181785		1	0,96428571	135	0	5	7422580
MsmF	0,9364407	0,93442623	0,93846154	0,93442516		1	0,87692308	114	0	16	7421060
MsmG	0,99791442	0,99791232	0,99791667	0,99791225		1	0,99583333	239	0	1	7392290
MsmH	0,97665008	0,97637795	0,97692308	0,97637755		1	0,95384615	124	0	6	7420390
MsuC	0,99765329	0,99765074	0,99765625	0,99765054		1	0,9953125	637	0	3	7405130
MsuD	0,72670918	0,72021661	0,91560955	0,72019606	0,63535032	0,83125	399	229	81	7410881	
MsuE	0,99673359	0,99672846	0,99673913	0,99672826		1	0,99347826	457	0	3	7415850
Mtox	0,99409958	0,99408284	0,99411765	0,99408217		1	0,98823529	840	0	10	7464490
SedA	0,99544396	0,99543379	0,99545455	0,99543358		1	0,99090909	327	0	3	7249810
SedB	0,99442872	0,99441341	0,99444444	0,99441321		1	0,98888889	267	0	3	7406010
SedC	0,99498723	0,99497487	995	0,99497467		1	0,99	297	0	3	7352230
SedD	0,99398169	0,99396378	994	0,99396358		1	988	247	0	3	7403320
SedE	0,80295031	0,8	0,93749648	0,79999511	0,73684211		875	140	50	20	7107370

Pta	0,92581947	0,92307692	0,92857143	0,92307625	1	0,85714286	60	0	10	7377390
SauS	0,9949873	0,99497487	995	0,99497474	1	0,99	198	0	2	7405330
SauT	0,98994921	0,98989899	0,99	0,98989871	1	0,98	196	0	4	7046530
SauU	0,96953518	0,96907216	0,97	0,96907135	1	0,94	188	0	12	7367790
SftA	0,23835699	0,16763006	0,78996423	0,16761098	0,09797297	0,58	58	534	42	7464176
SftD	0,97833639	0,97810219	0,97857143	0,97810179	1	0,95714286	134	0	6	7446000
SftG	0,46655019	0,37220844	0,96871678	0,37218687	0,23219814	0,9375	150	496	10	7464784
SftI	0,99544403	0,99543379	0,99545455	0,99543366	1	0,99090909	218	0	2	7465390
SftT	0,991137	0,99109792	0,99117647	0,99109772	1	0,98235294	167	0	3	7465460
SlcC	0,99900592	0,99900563	0,99900662	0,99900543	1	0,99801325	1507	0	3	7278150
SlcD	0,97356281	0,97321429	0,97391304	0,97321347	1	0,94782609	218	0	12	7392530
SlcF	0,99593749	0,99592944	0,99594595	0,99592924	1	0,99189189	367	0	3	7392270
SlcH	0,99854954	0,99854862	0,99855072	0,99854849	1	0,99710145	688	0	2	7373310
SmoA	0,98219363	0,98203593	0,98235294	0,98203512	1	0,96470588	328	0	12	7404750
SmoB	0,9953377	0,9953271	0,99534884	0,99532683	1	0,99069767	426	0	4	7403600
SmoC	0,98473126	0,98461538	0,98484848	0,98461471	1	0,96969697	320	0	10	7421940
SmoD	0,98473126	0,98461538	0,98484848	0,98461471	1	0,96969697	320	0	10	7422260
SmoE	0,98391657	0,98380297	0,98406114	0,98378725	1	0,96812227	6651	0	219	6962020
SmoF	0,99917049	0,999171	0,99992391	0,99917026	0,99849375	0,99984917	6629	10	1	7408960
SmoG	0,9990484	0,99904908	0,99919577	0,99904819	0,99970722	0,99839181	6829	2	11	7248468
SmoH	0,99831551	0,99831613	0,99985193	0,99831454	0,99692938	0,99970674	6818	21	2	7257079
Smol	0,90192663	0,89977728	0,92083286	0,89977426	0,96650718	0,84166667	202	7	38	7417043
SnfG	0,99961112	0,99961125	0,9996114	0,99961105	1	0,9992228	3857	0	3	7456000
SqdA	0,99163132	0,99159664	0,99166667	0,9915963	1	0,98333333	295	0	5	7462030
SqdB	0,99811129	0,99810964	0,99811321	0,99810951	1	0,99622642	528	0	2	7419130
SqdC	0,99903793	0,99903754	0,99903846	0,99903747	1	0,99807692	519	0	1	7422130
SqdD	0,99807494	0,99807322	0,99807692	0,99807308	1	0,99615385	518	0	2	7422120
SsuA	0,9983835	0,99838449	0,9983871	0,99838219	1	0,99677419	10506	0	34	7402660
SsuB	0,78318199	0,78313004	0,90324511	0,78284354	0,76079534	0,80681576	7576	2382	1814	7314438
SsuC	0,98202783	0,98195059	0,98404884	0,9819271	0,99620053	0,96810256	9439	36	311	7384894
SsuD	0,76079213	0,75714495	0,84436697	0,75703435	0,8405467	0,6888	2583	490	1167	7416220
SsuE	0,99828553	0,99828473	0,99828767	0,99828406	1	0,99657534	2910	0	10	7417210
SuyA	0,99207365	0,99204244	0,99210526	0,99204224	1	0,98421053	187	0	3	7422080
SuyB	0,99777517	0,99777283	0,99777778	0,99777269	1	0,99555556	448	0	2	7403070
TauA	0,99672172	0,99672445	0,99824153	0,99672169	0,9969634	0,99648562	6238	19	22	7416041
TauB	0,81336136	0,80358269	0,97589368	0,80337603	0,69511131	0,95216741	6370	2794	320	7348696
TauC	0,98423387	0,98414701	0,98514707	0,98413341	0,99839974	0,97029549	6239	10	191	7383790
TauD	0,99802888	0,99802761	0,9980315	0,99802694	1	0,99606299	2530	0	10	7403260
TauE	0,98107005	0,98089172	0,98125	0,98089091	1	0,9625	308	0	12	7421700
TauK	0,93544621	0,93355155	0,99858792	0,93353786	0,87753846	0,9972028	1426	199	4	7383041
TauL	0,99433835	0,99432279	0,99435484	0,99432232	1	0,98870968	613	0	7	7422160
TauM	0,88141188	0,87823344	0,97997787	0,87820759	0,80930233	0,96	1392	328	58	7410962
TauR	0,99464511	0,99463153	0,99466019	0,99463078	1	0,98932039	1019	0	11	7331990
TauX	0,99874908	0,99874844	0,99875	0,9987483	1	0,9975	798	0	2	7421570
TauY	0,99835008	0,99834893	0,99835165	0,99834872	1	0,9967033	907	0	3	7390280
TauZ	0,86238596	0,8544	0,99441497	0,85437008	0,75211268	0,98888889	1335	440	15	7462700
Toa	0,87735742	0,87701613	0,92646738	0,87700781	0,90248963	0,85294118	435	47	75	7328573
Tpa	0,91064605	0,91049383	0,94696813	0,91048987	0,92767296	0,89393939	295	23	35	7331417
Xsc	0,99373014	0,99371069	0,99375	0,99371049	1	0,9875	237	0	3	7298040
YihO	0,99721822	0,99721448	0,99722222	0,99721435	1	0,99444444	358	0	2	7413070
YihP	0,98368329	0,98365123	0,98783764	0,98365042	0,99175824	0,97567568	361	3	9	7412507
YihQ	0,9856101	0,98550725	0,98571429	0,98550657	1	0,97142857	340	0	10	7406520
YihR	0,98601296	0,98591549	0,98611111	0,98591516	1	0,97222222	175	0	5	7417870
YihS	0,98639326	0,98630137	0,98648649	0,9863007	1	0,97297297	360	0	10	7416790
YihT	0,98675372	0,98666667	0,98684211	0,98666599	1	0,97368421	370	0	10	7421830
YihU	0,98675371	0,98666667	0,98684211	0,98666599	1	0,97368421	370	0	10	7364040
YihV	0,98675372	0,98666667	0,98684211	0,98666599	1	0,97368421	370	0	10	7408050

## Validation on independent test dataset

Model	Metrics						Confusion matrix				
	MCC	F1-score	Balanced	Cohens	Precision	Recall	True positives	False positives	False negatives	True negatives	
			accuracy	kappa							
AcuI	0,99871352	0,9987372	0,99873879	0,99871269		1	0,99747758	3559	0	9	183651
AcuK	0,9992585	0,9992609	0,99999732	0,99925822	0,9985229		1	676	1	0	186542
AcuN	0,99933832	0,99934896	0,99998914	0,9993381	0,99869876		1	3070	4	0	184145
ComC	1	1	1	1		1	1	45	0	0	187174
ComD	0,96076636	0,96	0,96153846	0,95999733		1	0,92307692	12	0	1	187206
ComE	1	1	1	1		1	1	12	0	0	187207
CuyA	1	1	1	1		1	1	36	0	0	187183
CysA	0,93296123	0,93190963	0,93625036	0,93071954		1	0,87250072	3011	0	440	183768
CysC	0,99887822	0,99889111	0,99889233	0,99887759		1	0,99778467	2252	0	5	184962
CysD	0,9991807	0,99919394	0,99919459	0,99918036		1	0,99838918	3099	0	5	184115
CysH	0,99978073	0,99978341	0,99978346	0,99978071		1	0,99956691	2308	0	1	184910
CysI	0,99879589	0,99881141	0,99881282	0,99879517		1	0,99762564	2521	0	6	184692
CysJ	0,98726426	0,98738116	0,98753841	0,98718317		1	0,97507682	2856	0	73	184290
CysM	0,0312999	0,01033414	0,50249645	0,00960132	0,21126761	0,00529661		15	56	2817	184331
CysN	0,99870197	0,99872286	0,99872449	0,99870113		1	0,99744898	3128	0	8	184083
CysU	0,99658879	0,99663174	0,99664304	0,99658297		1	0,99328609	2663	0	18	184538
CysW	0,99870983	0,99872797	0,99872958	0,99870899		1	0,99745917	2748	0	7	184464
DddA	0,62474379	0,56333333	0,69628071	0,56219575	0,99705015	0,39256678		338	1	523	186357
DddD	0,91286605	0,90909091	0,91666667	0,90908561		1	0,83333333	10	0	2	187207
DddL	1	1	1	1		1	1	10	0	0	187209
DddP	1	1	1	1		1	1	47	0	0	187172
DddT	1	1	1	1		1	1	4011	0	0	183208
DddY	0,87579115	0,86831276	0,88363636	0,86814458		1	0,76727273	211	0	64	186944
DdhA	0,98164457	0,98148148	0,98181818	0,98147614		1	0,96363636	53	0	2	187164
DdhB	0,9944208	0,9944483	0,99447895	0,99440524		1	0,9889579	1433	0	16	185770
DmdA	1	1	1	1		1	1	28	0	0	187191
DmdB	0,90824723	0,90466191	0,91296012	0,90405737		1	0,82592025	1077	0	227	185915
DmdC	0,99929528	0,99930314	0,99930362	0,99929503		1	0,99860724	2151	0	3	185065
DmdD	1	1	1	1		1	1	11	0	0	187208
DmoA	1	1	1	1		1	1	1350	0	0	185869
DmsA	0,99322411	0,99343955	0,9993131	0,99320728	0,98788444	0,99905749		6360	78	6	180775
DmsB	0,99989112	0,99989385	0,99989387	0,99989111		1	0,99978773	4710	0	1	182508
DmsC	0,9996324	0,99964059	0,99975769	0,99963239	0,99976036	0,99952084		4172	1	2	183044
DmsD	1	1	1	1		1	1	2054	0	0	185165
DorA	0,4620604	352	0,99978365	0,35187449	0,21359223		1	22	81	0	187116
DorC	1	1	1	1		1	1	1501	0	0	185718
DorR	0,96293972	0,96226415	0,96363636	0,96225348		1	0,92727273	51	0	4	187164
DorS	0,99543618	0,99543379	0,99545455	0,99542576		1	0,99090909	327	0	3	186889
DsoA	1	1	1	1		1	1	24	0	0	187195
DsoB	1	1	1	1		1	1	40	0	0	187179
DsoC	1	1	1	1		1	1	94	0	0	187125
DsoD	0,9870936	0,98701299	0,98717949	0,98701032		1	0,97435897	38	0	1	187180
DsoE	1	1	1	1		1	1	16	0	0	187203
DsoF	0,96549649	0,96491228	0,96610169	0,96490161		1	0,93220339	55	0	4	187160
HpsK	1	1	1	1		1	1	10	0	0	187209
HpsL	1	1	1	1		1	1	6	0	0	187213
HpsM	1	1	1	1		1	1	3	0	0	187216
HpsN	0,99176676	0,99173554	0,99180328	0,99173287		1	0,98360656	60	0	1	187158
HpsO	0,70709734	0,66666667	0,75	0,6666548		1	0,5	5	0	5	187209
HpsP	1	1	1	1		1	1	13	0	0	187206
HpsU	0,02250096	0,00175953	0,50043509	0,00170629	0,6	0,00088106		3	2	3402	183812
IseI	0,91738964	0,91543271	0,99825278	0,91404732	0,84456588	0,99928161		2782	512	2	183923
IseK	1	1	1	1		1	1	13	0	0	187206
IseL	1	1	1	1		1	1	34	0	0	187185
IseM	0,83204141	0,81818182	0,84615385	0,81817149		1	0,69230769	9	0	4	187206
IseR	1	1	1	1		1	1	6	0	0	187213
IseU	0,27987863	0,21065675	0,56210362	0,2050713	0,64968153	0,12570865		510	275	3547	182887
IsfD	0,99938363	0,99938613	0,9993865	0,99938344		1	0,99877301	814	0	1	186404
IsfE	1	1	1	1		1	1	18	0	0	187201
IsIA	0,99986981	0,99987253	0,99987255	0,9998698		1	0,99974509	3922	0	1	183296
IsIB	1	1	1	1		1	1	29	0	0	187190
MddA	1	1	1	1		1	1	3	0	0	187216
MsmA	1	1	1	1		1	1	2	0	0	187217
MsmB	0,66322572	0,61111111	0,72	0,61098413		1	0,44	44	0	56	187119
MsmE	1	1	1	1		1	1	1	0	0	187218
MsmF	1	1	1	1		1	1	1	0	0	187218
MsmG	0,78285627	0,76	0,80645161	0,75996979		1	0,61290323	19	0	12	187188
MsmH	0,87172389	0,86363636	0,88	0,86357343		1	0,76	76	0	24	187119
MsuC	0,99479984	0,99479167	0,99481865	0,99478632		1	0,98963731	191	0	2	187026
MsuD	0,49511426	0,39497908	0,62304484	0,39375251		1	0,24608968	236	0	723	186260
MsuE	1	1	1	1		1	1	123	0	0	187096
Mtox	1	1	1	1		1	1	10	0	0	187209
SedA	1	1	1	1		1	1	11	0	0	187208
SedB	1	1	1	1		1	1	1214	0	0	186005
SedC	0,96449211	0,96394231	0,96519722	0,96386211		1	0,93039443	401	0	30	186788
SedD	1	1	1	1		1	1	46	0	0	187173
SedE	0,90677106	0,91739092	0,98443189	0,90268437	0,84904127	0,99770861		26125	4645	60	156389

Pta	1	1	1	1	1	1	159	0	0	187060
SauS	0,85713752	0,85714286	0,92856876	0,85713752	0,85714286	0,85714286	6	1	1	187211
SauT	1	1	1	1	1	1	3	0	0	187216
SauU	1	1	1	1	1	1	3	0	0	187216
SftD	0,03634413	0,00267023	0,50066845	0,00263831	1	0,0013369	3	0	2241	184975
SftG	0,70010788	0,68181818	0,94114442	0,68178272	0,55555556	0,88235294	15	12	2	187190
SftI	1	1	1	1	1	1	16	0	0	187203
SftT	1	1	1	1	1	1	11	0	0	187208
SlcC	0,94439019	0,94285714	0,94594595	0,94284649	1	0,89189189	33	0	4	187182
SlcD	1	1	1	1	1	1	44	0	0	187175
SlcF	0,99557697	0,99560761	0,99562682	0,99556719	1	0,99125364	1700	0	15	185504
SlcH	1	1	1	1	1	1	46	0	0	187173
SmoA	1	1	1	1	1	1	11	0	0	187208
SmoB	1	1	1	1	1	1	40	0	0	187179
SmoC	0,97467683	0,97435897	975	0,97435631	1	0,95	19	0	1	187199
SmoD	1	1	1	1	1	1	15	0	0	187204
SmoE	0,97615015	0,97742998	0,97792815	0,97586583	1	0,95585629	11866	0	548	174805
SmoF	1	1	1	1	1	1	22	0	0	187197
SmoG	1	1	1	1	1	1	22	0	0	187197
SmoH	1	1	1	1	1	1	22	0	0	187197
Smol	0,76086965	0,73333333	0,78947368	0,73331349	1	0,57894737	11	0	8	187200
SnfG	0,98846744	0,98841699	0,98854962	0,98840094	1	0,97709924	256	0	6	186957
SqdA	1	1	1	1	1	1	6	0	0	187213
SqdB	1	1	1	1	1	1	50	0	0	187169
SqdC	1	1	1	1	1	1	3	0	0	187216
SqdD	1	1	1	1	1	1	22	0	0	187197
SsuA	0,99945741	0,99946265	0,99946294	0,99945726	1	0,99892589	1860	0	2	185357
SsuB	0,99699462	0,99701978	0,99702863	0,99699011	1	0,99405727	1840	0	11	185368
SsuC	0,99974112	0,99974379	0,99974385	0,99974109	1	0,9994877	1951	0	1	185267
SsuD	0,96938972	0,96961476	0,97874271	0,96931272	0,98186813	0,95766345	1787	33	79	185320
SsuE	1	1	1	1	1	1	1601	0	0	185618
SuyA	1	1	1	1	1	1	19	0	0	187200
SuyB	0,99814108	0,99814471	0,99814815	0,99813936	1	0,9962963	538	0	2	186679
TauA	0,99972787	0,99973053	0,9997306	0,99972783	1	0,99946121	1855	0	1	185363
TauB	0,99274753	0,99278318	0,99283489	0,99272123	1	0,98566978	1582	0	23	185614
TauC	1	1	1	1	1	1	1865	0	0	185354
TauD	0,99538628	0,99542149	0,99544236	0,99537564	1	0,99088472	1848	0	17	185354
TauE	1	1	1	1	1	1	3	0	0	187216
TauK	0,97332593	0,97297297	0,99999733	0,9729703	0,94736842	1	18	1	0	187200
TauL	1	1	1	1	1	1	1	0	0	187218
TauM	0,99530089	0,99531616	0,9959133	0,9952947	0,99882491	0,99183197	850	1	7	186361
TauR	0,99616839	0,99617445	0,99618902	0,99616105	1	0,99237805	651	0	5	186563
TauX	0,95713066	0,95625	0,95808383	0,95621265	1	0,91616766	153	0	14	187052
TauY	0,97979328	0,97959184	0,99999733	0,97958917	0,96	1	24	1	0	187194
TauZ	0,92599246	0,9245283	0,93749733	0,924507	0,98	875	49	1	7	187162
Toa	0,94813711	0,94706724	0,95033745	0,94686694	0,9984917	0,90068027	662	1	73	186483
Tpa	0,99038001	0,99039781	0,99451252	0,99037907	0,99175824	0,9890411	361	3	4	186851
Xsc	0,99775266	0,99775281	0,99775785	0,99775014	1	0,9955157	222	0	1	186996
YihO	0,9528618	0,9520362	0,95423143	0,95175209	1	0,90846287	1052	0	106	186061
YihP	0,99389021	0,99391833	0,99480431	0,99388071	0,9982548	0,98961938	1144	2	12	186061
YihQ	0,98519136	0,98517577	0,98539232	0,98508173	1	0,97078464	1163	0	35	186021
YihR	0,99637497	0,99638989	0,99640288	0,9963684	1	0,99280576	1104	0	8	186107
YihS	0,97986482	0,97979381	0,98019402	0,97966215	1	0,96038804	1188	0	49	185982
YihT	0,9947954	0,99482223	0,9948489	0,99478186	1	0,9896978	1441	0	15	185763
YihU	0,9989633	0,99897084	0,9989719	0,99896277	1	0,9979438	1456	0	3	185760
YihV	0,99860708	0,99861687	0,99861878	0,99860611	1	0,99723757	1444	0	4	185771

## Skew corrected 10-fold cross validation on training dataset

Model	Metrics				Confusion matrix						
	MCC	F1-score	Balanced accuracy	Cohens kappa	Precision	Recall	True positives	False positives	False negatives	True negatives	
AcuI	0,99999946	0,99999973	0,99999973	0,99999946		1	0,99999946	946	0	4	7406310
AcuK	0,74690483	0,85512117	0,87345026	0,71619387	0,74691053	0,99999862	15110	5120	10	7236490	
AcuN	0,91094105	0,95336692	0,95549663	0,906993	0,91089172	0,99999713	14209	1390	21	7319620	
ComC	0,99999959	0,99999979	0,99999979	0,99999959		1	0,99999959	247	0	3	7403780
ComD	0,99999986	0,99999927	0,99999927	0,99999986		1	0,99999855	130	0	10	7422220
ComE	0,99999959	0,99999979	0,99999974	0,99999959		1	0,99999959	137	0	3	7420710
CuyA	0,99999859	0,99999927	0,99999927	0,99999859		1	0,99999854	120	0	10	7412320
CuyZ	0,99999845	0,99999919	0,99999919	0,99999845		1	0,99999839	129	0	11	7412460
CysA	0,93757189	0,96529614	0,97139422	0,93563314	0,93296433	0,99994931	1517	109	263	6087121	
CysC	0,99996584	0,99998197	0,99998197	0,99996584		1	0,99996395	2041	0	239	7405500
CysD	0,99999986	0,99999993	0,99999993	0,99999986		1	0,99999986	4159	0	1	7395740
CysH	0,99493474	0,99745484	0,99747345	0,99492194	0,99492515	0,99999742	3921	20	19	7402750	
CysI	0,99255264	0,99626102	0,99627777	0,99252491	0,99255029	0,99999959	3997	30	3	7402220	
CysJ	0,99479017	0,9973876	0,99739576	0,9947766	0,99478895	0,99999986	1909	10	1	7402660	
CysM	0,99999848	0,99999923	0,99999923	0,99999848		1	0,99999847	459	0	11	7339290
CysN	0,99304765	0,99651127	0,99652474	0,99302348	0,99304693	0,99999986	3999	28	1	7372632	
CysU	0,99999959	0,99999998	0,99999979	0,99999959		1	0,99999959	1617	0	3	7381730
CysW	0,99999851	0,99999925	0,99999925	0,99999851		1	0,9999985	1599	0	11	7406260
DddA	0,99999866	0,99999933	0,99999933	0,99999866		1	0,99999866	5480	0	10	7457840
DddD	0,99999973	0,99999987	0,99999986	0,99999973		1	0,99999973	268	0	2	7464200
DddL	0,99999959	0,99999998	0,99999978	0,99999959		1	0,99999959	167	0	3	7465510
DddP	0,9999996	0,99999998	0,99999979	0,9999996		1	0,9999996	437	0	3	7463740
DddT	0,99999959	0,99999979	0,99999978	0,99999959		1	0,99999959	117	0	3	7423010
DddY	0,99999932	0,99999966	0,99999963	0,99999932		1	0,99999931	175	0	5	7465500
DdhA	0,91631627	0,95614029	0,95833321	0,91282814	0,91596639	0,99999986	109	10	1	7465210	
DdhB	0,99949264	0,99974626	0,99974747	0,99949251	0,99949264		1	1970	1	0	7455389
DmdA	1	1	1	1		1	1	2190	0	0	7442890
DmdB	0,84665983	0,91637742	0,92384954	0,83506628	0,84665696	0,99999314	3200	584	50	7406046	
DmdC	1	1	1	1		1	1	2400	0	0	7322970
DmdD	0,99999703	0,99999851	0,9999985	0,99999703		1	0,99999701	1678	0	22	7457530
DmoA	0,99999959	0,99999979	0,99999976	0,99999959		1	0,99999958	67	0	3	7430070
DmsA	0,99999864	0,99999932	0,99999932	0,99999864		1	0,99999863	2680	0	10	7342500
DmsB	0,99999864	0,99999932	0,99999932	0,99999864		1	0,99999864	2670	0	10	7381650
DmsC	0,99999973	0,99999987	0,99999986	0,99999973		1	0,99999973	2078	0	2	7420650
DmsD	0,9999996	0,99999998	0,99999998	0,9999996		1	0,9999996	1967	0	3	7420400
DorA	0,99999986	0,99999993	0,99999992	0,99999986		1	0,99999986	79	0	1	7394540
DorC	0,99999863	0,9999993	0,9999993	0,99999863		1	0,9999986	280	0	10	7414430
DorR	0,999999	0,99999949	0,99999947	0,999999		1	0,99999899	323	0	7	7075320
DorS	0,95645836	0,97771167	0,97836122	0,95551137	0,95639535	0,99999986	329	15	1	7068985	
DsoA	0,99999986	0,99999993	0,99999993	0,99999986		1	0,99999986	239	0	1	7422540
DsoB	0,99999865	0,99999932	0,99999932	0,99999865		1	0,99999864	1190	0	10	7421560
DsoC	0,99999987	0,99999993	0,99999993	0,99999987		1	0,99999987	939	0	1	7421830
DsoD	0,9999996	0,99999998	0,99999998	0,9999996		1	0,99999959	1247	0	3	7420530
DsoE	0,99999864	0,99999931	0,99999931	0,99999864		1	0,99999862	450	0	10	7422320
DsoF	0,99275765	0,99634807	0,99639917	0,99273145	0,99272433	0,99999836	1228	9	12	7391041	
HpsK	0,94967754	0,97344997	0,97567404	0,94841333	0,94827586	0,99999714	330	18	20	7422022	
HpsL	0,99999945	0,99999972	0,99999971	0,99999945		1	0,99999945	166	0	4	7411500
HpsM	0,9411764	0,96969697	0,97058816	0,93944946	0,94117647		1	320	20	0	7465190
HpsN	0,9999985	0,99999924	0,99999924	0,9999985		1	0,99999848	559	0	11	7381890
HpsO	0,68554579	0,78116911	0,86699333	0,63946444	0,6409465	0,99992752	623	349	337	7164021	
HpsP	0,99999959	0,99999998	0,99999979	0,99999959		1	0,99999959	927	0	3	7359040
HpsU	0,52263821	0,62842453	0,7988262	0,42910633	0,45818182	0,99997774	126	149	94	7372471	
IseI	0,99987235	0,99991681	0,99991681	0,99987234		1	0,99983363	759	0	661	7432050
IseK	0,99999848	0,99999922	0,99999922	0,99999848		1	0,99999845	229	0	11	7422470
IseL	0,99999973	0,99999986	0,99999985	0,99999973		1	0,99999973	228	0	2	7422550
IseM	0,99999862	0,99999993	0,99999993	0,99999862		1	0,9999986	240	0	10	7417310
IseR	0,99999521	0,99999758	0,99999758	0,99999521		1	0,99999516	1545	0	35	7390700
IseU	0,1773364	0,19842431	0,64319585	0,06103837	0,11014263	0,99972652	278	2246	622	7358804	
IseD	0,98527083	0,99225753	0,99313088	0,98516258	0,98464164	0,99999214	577	9	53	7365891	
IseE	0,99999959	0,99999979	0,99999979	0,99999959		1	0,99999959	177	0	3	7422540
IseA	0,99999731	0,99999865	0,99999865	0,99999731		1	0,9999973	5530	0	20	7443950
IseB	0,99999987	0,99999993	0,99999993	0,99999987		1	0,99999987	5289	0	1	7450010
MddA	1	1	1	1		1	1	100	0	0	7463510
MsmA	0,99999848	0,99999916	0,99999916	0,99999848		1	0,99999831	40	0	10	7407390
MsmB	0,99999848	0,99999916	0,99999916	0,99999848		1	0,99999832	40	0	10	7422710
MsmE	0,99999931	0,99999965	0,99999963	0,99999931		1	0,9999993	135	0	5	7422580
MsmF	0,99999769	0,99999877	0,99999873	0,99999769		1	0,99999754	114	0	16	7421060
MsmG	0,99999986	0,99999993	0,99999993	0,99999986		1	0,99999986	239	0	1	7392290
MsmH	0,99999917	0,99999958	0,99999954	0,99999917		1	0,99999915	124	0	6	7420390
MsuC	0,99999959	0,99999998	0,99999979	0,99999959		1	0,99999959	637	0	3	7405130
MsuD	0,65583412	0,77701648	0,89753768	0,60152166	0,63535032	0,99998685	399	229	81	7410881	
MsuE	0,99999959	0,99999998	0,99999979	0,99999959		1	0,99999959	457	0	3	7415850
Mtox	0,99999865	0,99999932	0,99999932	0,99999865		1	0,99999864	840	0	10	7464490
SedA	0,99999958	0,99999979	0,99999979	0,99999958		1	0,99999958	327	0	3	7249810
SedB	0,99999959	0,99999998	0,99999979	0,99999959		1	0,99999959	267	0	3	7406010
SedC	0,99999959	0,99999979	0,99999978	0,99999959		1	0,99999959	297	0	3	7352230
SedD	0,99999959	0,99999979	0,99999978	0,99999959		1	0,99999959	247	0	3	7403320
SedE	0,74926501	0,84848369	0,90086907	0,71909807	0,73684211	0,99999678	140	50	20	7107370	



Pta	0,99999853	0,99999921	0,99999921	0,99999853	1	0,99999842	60	0	10	7377390
SauS	0,99999973	0,99999986	0,99999985	0,99999973	1	0,99999973	198	0	2	7405330
SauT	0,99999943	0,99999971	0,99999969	0,99999943	1	0,99999942	196	0	4	7046530
SauU	0,99999832	0,99999913	0,99999913	0,99999832	1	0,99999827	188	0	12	7367790
SftA	0,12429956	0,17846138	0,54110548	0,03043258	0,09797297	0,99999903	58	534	42	7464176
SftD	0,99999918	0,99999958	0,99999953	0,99999918	1	0,99999916	134	0	6	7446000
SftG	0,23797098	0,37688432	0,62196281	0,10719087	0,23219814	0,99999857	150	496	10	7464784
SftI	0,99999973	0,99999986	0,99999986	0,99999973	1	0,99999973	218	0	2	7465390
SftT	0,99999959	0,99999998	0,99999979	0,99999959	1	0,99999959	167	0	3	7465460
SlcC	0,99999959	0,99999979	0,99999979	0,99999959	1	0,99999959	1507	0	3	7278150
SlcD	0,99999833	0,99999914	0,99999914	0,99999833	1	0,99999829	218	0	12	7392530
SlcF	0,99999959	0,99999998	0,99999979	0,99999959	1	0,99999959	367	0	3	7392270
SlcH	0,99999973	0,99999986	0,99999986	0,99999973	1	0,99999973	688	0	2	7373310
SmoA	0,99999835	0,99999916	0,99999916	0,99999835	1	0,99999832	328	0	12	7404750
SmoB	0,99999946	0,99999973	0,99999972	0,99999946	1	0,99999945	426	0	4	7403600
SmoC	0,99999863	0,99999931	0,99999927	0,99999863	1	0,99999861	320	0	10	7421940
SmoD	0,99999863	0,99999931	0,99999931	0,99999863	1	0,99999861	320	0	10	7422260
SmoE	0,99996803	0,99998375	0,99998375	0,99996803	1	0,99996751	6651	0	219	6962020
SmoF	0,99849373	0,99924624	0,99924622	0,99849259	0,99849375	0,99999987	6629	10	1	7408960
SmoG	0,99970593	0,99985283	0,99985325	0,99970589	0,99970722	0,99999848	6829	2	11	7248468
SmoH	0,99692954	0,99846219	0,99846509	0,99692483	0,99692938	0,99999972	6818	21	2	7257079
SmoI	0,96907451	0,98296543	0,98599689	0,96859699	0,96650718	0,99999391	202	7	38	7417043
SnfG	0,99999966	0,99999998	0,99999998	0,99999966	1	0,99999966	3857	0	3	7456000
SqdA	0,99999932	0,99999966	0,99999965	0,99999932	1	0,99999932	295	0	5	7462030
SqdB	0,99999973	0,99999986	0,99999986	0,99999973	1	0,99999973	528	0	2	7419130
SqdC	0,99999987	0,99999993	0,99999993	0,99999987	1	0,99999987	519	0	1	7422130
SqdD	0,99999973	0,99999986	0,99999986	0,99999973	1	0,99999973	518	0	2	7422120
SsuA	0,99999954	0,99999977	0,99999977	0,99999954	1	0,999999539	10506	0	34	7402660
SsuB	0,77870481	0,86403493	0,89956499	0,75508249	0,76079534	0,99969281	7576	2382	1814	7314438
SsuC	0,99621809	0,99807498	0,998139	0,99621125	0,99620053	0,9999565	9439	36	311	7384894
SsuD	0,86201583	0,91327106	0,94240922	0,85265216	0,8405467	0,99977162	2583	490	1167	7416220
SsuE	0,99999865	0,99999932	0,99999932	0,99999865	1	0,99999865	2910	0	10	7417210
SuyA	0,99999959	0,99999979	0,99999978	0,99999959	1	0,99999959	187	0	3	7422080
SuyB	0,99999973	0,99999986	0,99999986	0,99999973	1	0,99999973	448	0	2	7403070
TauA	0,99696575	0,99847791	0,99848564	0,99696116	0,9969634	0,99999702	6238	19	22	7416041
TauB	0,70015221	0,82012111	0,8526726	0,65793447	0,69511131	0,99995429	6370	2794	320	7348696
TauC	0,99839721	0,99918592	0,99921027	0,99839601	0,99839974	0,99997334	6239	10	191	7383790
TauD	0,99999865	0,99999932	0,99999932	0,99999865	1	0,99999864	2530	0	10	7403260
TauE	0,99999835	0,99999916	0,99999916	0,99999835	1	0,99999832	308	0	12	7421700
TauK	0,87768681	0,93477525	0,93891936	0,8702699	0,87753846	0,99999946	1426	199	4	7383041
TauL	0,99999905	0,99999952	0,99999951	0,99999905	1	0,99999905	613	0	7	7422160
TauM	0,81239543	0,89459828	0,90779777	0,79517371	0,80930233	0,99999185	1392	328	58	7410962
TauR	0,99999849	0,99999924	0,99999924	0,99999849	1	0,99999848	1019	0	11	7331990
TauX	0,99999973	0,99999986	0,99999986	0,99999973	1	0,99999973	798	0	2	7421570
TauY	0,99999959	0,99999998	0,99999979	0,99999959	1	0,99999959	907	0	3	7390280
TauZ	0,75314312	0,85852015	0,87710736	0,72385963	0,75211268	0,99999797	1335	440	15	7462700
Toa	0,90901945	0,94874051	0,9596309	0,90490148	0,90248963	0,9999988	435	47	75	7328573
Tpa	0,93124652	0,96247713	0,96788936	0,92888968	0,92767296	0,99999466	295	23	35	7331417
Xsc	0,99999959	0,99999979	0,99999978	0,99999959	1	0,99999958	237	0	3	7298040
YihO	0,99999973	0,99999986	0,99999986	0,99999973	1	0,99999973	358	0	2	7413070
YihP	0,99185644	0,99586145	0,99624936	0,9918233	0,99175824	0,99999876	361	3	9	7412507
YihQ	0,99999863	0,99999931	0,99999931	0,99999863	1	0,99999861	340	0	10	7406520
YihR	0,99999932	0,99999965	0,99999961	0,99999932	1	0,99999931	175	0	5	7417870
YihS	0,99999863	0,99999931	0,99999931	0,99999863	1	0,99999861	360	0	10	7416790
YihT	0,99999863	0,99999931	0,99999931	0,99999863	1	0,99999862	370	0	10	7421830
YihU	0,99999862	0,99999931	0,99999931	0,99999862	1	0,99999861	370	0	10	7364040
YihV	0,99999863	0,99999931	0,99999931	0,99999863	1	0,99999861	370	0	10	7408050

## Skew corrected validation on independent test dataset

Model	Metrics				Confusion matrix						
	MCC	F1-score	Balanced accuracy	Cohens kappa	Precision	Recall	True positives	False positives	False negatives	True negatives	
AcuI	0,99995093	0,99997544	0,99997544	0,99995093		1	0,99995087	3559	0	9	183651
AcuK	0,99852289	0,9992609	0,99926144	0,9985218	0,9985229	1		676	1	0	186542
AcuN	0,99869875	0,99934896	0,99934937	0,9986979	0,99869876	1		3070	4	0	184145
ComC	1	1	1	1	1	1		45	0	0	187174
ComD	0,99999444	0,99999711	0,99999711	0,99999444		1	0,99999421	12	0	1	187206
ComE	1	1	1	1	1	1		12	0	0	187207
CuyA	1	1	1	1	1	1		36	0	0	187183
CysA	0,99743733	0,99862978	0,99863165	0,99743405		1	0,9972633	3011	0	440	183768
CysC	0,99997294	0,99998645	0,99998645	0,99997294		1	0,99997291	2252	0	5	184962
CysD	0,99997282	0,9999864	0,9999864	0,99997282		1	0,9999728	3099	0	5	184115
CysH	0,99999459	0,99999729	0,99999729	0,99999459		1	0,99999459	2308	0	1	184910
CysI	0,99996748	0,99998372	0,99998372	0,99996748		1	0,99996744	2521	0	6	184692
CysJ	0,99959898	0,99979692	0,99979696	0,9995989		1	0,99959393	2856	0	73	184290
CysM	0,21611931	0,23207904	0,61902116	0,21502083	0,21126761	0,25743867		15	56	2817	184331
CysN	0,99995649	0,99997822	0,99997822	0,99995649		1	0,99995643	3128	0	8	184083
CysU	0,99990214	0,9999509	0,9999509	0,99990213		1	0,99990181	2663	0	18	184538
CysW	0,99996201	0,99998098	0,99998098	0,999962		1	0,99996196	2748	0	7	184464
DddA	0,99299585	0,99497166	0,99587087	0,99299166	0,99705015	0,99290183		338	1	523	186357
DddD	0,99998825	0,99999359	0,99999359	0,99998825		1	0,99998718	10	0	2	187207
DddL	1	1	1	1	1	1		10	0	0	187209
DddP	1	1	1	1	1	1		47	0	0	187172
DddT	1	1	1	1	1	1		4011	0	0	183208
DddY	0,99960589	0,99977696	0,99977701	0,99960581		1	0,99955401	211	0	64	186944
DdhA	0,99998911	0,99999446	0,99999446	0,99998911		1	0,99998891	53	0	2	187164
DdhB	0,9999134	0,99995646	0,99995646	0,99991339		1	0,99991292	1433	0	16	185770
DmdA	1	1	1	1	1	1		28	0	0	187191
DmdB	0,99865217	0,99926138	0,99926192	0,99865126		1	0,99852385	1077	0	227	185915
DmdC	0,99998378	0,99999188	0,99999188	0,99998378		1	0,99998377	2151	0	3	185065
DmdD	1	1	1	1	1	1		11	0	0	187208
DmoA	1	1	1	1	1	1		1350	0	0	185869
DmsA	0,9878543	0,9938889	0,99392868	0,98778135	0,98788444	0,99996679		6360	78	6	180775
DmsB	0,99999452	0,99999726	0,99999726	0,99999452		1	0,99999452	4710	0	1	182508
DmsC	0,99974949	0,99987477	0,99987477	0,99974947	0,99976036	0,99998907		4172	1	2	183044
DmsD	1	1	1	1	1	1		2054	0	0	185165
DorA	0,21355589	352	0,60675977	0,08723384	0,21359223	1		22	81	0	187116
DorC	1	1	1	1	1	1		1501	0	0	185718
DorR	0,99997779	0,99998848	0,99998848	0,99997779		1	0,99997695	51	0	4	187164
DorS	0,99998387	0,9999919	0,9999919	0,99998387		1	0,99998387	327	0	3	186889
DsoA	1	1	1	1	1	1		24	0	0	187195
DsoB	1	1	1	1	1	1		40	0	0	187179
DsoC	1	1	1	1	1	1		94	0	0	187125
DsoD	0,99999459	0,99999726	0,99999726	0,99999459		1	0,99999452	38	0	1	187180
DsoE	1	1	1	1	1	1		16	0	0	187203
DsoF	0,99997785	0,99998854	0,99998854	0,99997785		1	0,99997707	55	0	4	187160
HpsK	1	1	1	1	1	1		10	0	0	187209
HpsL	1	1	1	1	1	1		6	0	0	187213
HpsM	1	1	1	1	1	1		3	0	0	187216
HpsN	0,99999461	0,99999728	0,99999728	0,99999461		1	0,99999457	60	0	1	187158
HpsO	0,99995994	0,99997329	0,99997329	0,99995994		1	0,99994659	5	0	5	187209
HpsP	1	1	1	1	1	1		13	0	0	187206
HpsU	0,1615472	0,0844842	0,52242715	0,08202711	0,6	0,04544133		3	2	3402	183812
IseI	0,84441961	0,91572949	0,92214212	0,8324913	0,84456588	0,99998915		2782	512	2	183923
IseK	1	1	1	1	1	1		13	0	0	187206
IseL	1	1	1	1	1	1		34	0	0	187185
IseM	0,99997389	0,99998457	0,99998457	0,99997389		1	0,99996914	9	0	4	187206
IseR	1	1	1	1	1	1		6	0	0	187213
IseU	0,7116046	0,74259301	0,9014717	0,70176845	0,64968153	0,86651386		510	275	3547	182887
IseD	0,99999463	0,99999731	0,99999731	0,99999463		1	0,99999463	814	0	1	186404
IseE	1	1	1	1	1	1		18	0	0	187201
IseA	0,99999454	0,99999727	0,99999727	0,99999454		1	0,99999454	3922	0	1	183296
IseB	1	1	1	1	1	1		29	0	0	187190
IseD	1	1	1	1	1	1		3	0	0	187216
MsmA	1	1	1	1	1	1		2	0	0	187217
MsmB	0,99951054	0,99966003	0,99966015	0,99951042		1	0,99932029	44	0	56	187119
MsmE	1	1	1	1	1	1		1	0	0	187218
MsmF	1	1	1	1	1	1		1	0	0	187218
MsmG	0,99991566	0,99994771	0,99994771	0,99991565		1	0,99989542	19	0	12	187188
MsmH	0,99985151	0,99991563	0,99991563	0,9998515		1	0,99983126	76	0	24	187119
MsuC	0,99998925	0,9999946	0,9999946	0,99998925		1	0,9998919	191	0	2	187026
MsuD	0,99028527	0,99217501	0,99223577	0,99023809		1	0,98447154	236	0	723	186260
MsuE	1	1	1	1	1	1		123	0	0	187096
Mtox	1	1	1	1	1	1		10	0	0	187209
SedA	1	1	1	1	1	1		11	0	0	187208
SedB	1	1	1	1	1	1		1214	0	0	186005
SedC	0,99983341	0,99991369	0,9999137	0,9998334		1	0,9998274	401	0	30	186788
SedD	1	1	1	1	1	1		46	0	0	187173
SedE	0,84691606	0,91820093	0,92258896	0,83547183	0,84904127	0,99962669		26125	4645	60	156389

Pta	1	1	1	1	1	1	159	0	0	187060
SauS	0,86601933	0,92307427	0,93749659	0,85713752	0,85714286	0,99999377	6	1	1	187211
SauT	1	1	1	1	1	1	3	0	0	187216
SauU	1	1	1	1	1	1	3	0	0	187216
SftD	0,31335744	0,18079696	0,54969125	0,1788263	1	0,09938251	3	0	2241	184975
SftG	0,57065653	0,71428263	0,79308962	0,4913127	0,55555556	0,99998789	15	12	2	187190
SftI	1	1	1	1	1	1	16	0	0	187203
SftT	1	1	1	1	1	1	11	0	0	187208
SlcC	0,99997734	0,99998802	0,99998802	0,99997734	1	0,99997604	33	0	4	187182
SlcD	1	1	1	1	1	1	44	0	0	187175
SlcF	0,99991879	0,99995921	0,99995922	0,99991879	1	0,99991843	1700	0	15	185504
SlcH	1	1	1	1	1	1	46	0	0	187173
SmoA	1	1	1	1	1	1	11	0	0	187208
SmoB	1	1	1	1	1	1	40	0	0	187179
SmoC	0,99999452	0,99999719	0,99999719	0,99999452	1	0,99999438	19	0	1	187199
SmoD	1	1	1	1	1	1	15	0	0	187204
SmoE	0,99680295	0,99836283	0,99836551	0,99679784	1	0,99673102	11866	0	548	174805
SmoF	1	1	1	1	1	1	22	0	0	187197
SmoG	1	1	1	1	1	1	22	0	0	187197
SmoH	1	1	1	1	1	1	22	0	0	187197
Smol	0,99994173	0,99996309	0,9999631	0,99994173	1	0,99992619	11	0	8	187200
SnfG	0,99996753	0,99998358	0,99998358	0,99996753	1	0,99996716	256	0	6	186957
SqdA	1	1	1	1	1	1	6	0	0	187213
SqdB	1	1	1	1	1	1	50	0	0	187169
SqdC	1	1	1	1	1	1	3	0	0	187216
SqdD	1	1	1	1	1	1	22	0	0	187197
SsuA	0,9999892	0,9999946	0,9999946	0,9999892	1	0,9999892	1860	0	2	185357
SsuB	0,99994048	0,99997015	0,99997015	0,99994048	1	0,99994031	1840	0	11	185368
SsuC	0,9999946	0,9999973	0,9999973	0,9999946	1	0,9999946	1951	0	1	185267
SsuD	0,98180821	0,9906327	0,99108727	0,98165823	0,98186813	0,99955514	1787	33	79	185320
SsuE	1	1	1	1	1	1	1601	0	0	185618
SuyA	1	1	1	1	1	1	19	0	0	187200
SuyB	0,99998927	0,99999462	0,99999462	0,99998927	1	0,99998925	538	0	2	186679
TauA	0,9999946	0,9999973	0,9999973	0,9999946	1	0,9999946	1855	0	1	185363
TauB	0,9998752	0,99993715	0,99993715	0,99987519	1	0,9998743	1582	0	23	185614
TauC	1	1	1	1	1	1	1865	0	0	185354
TauD	0,99990787	0,99995372	0,99995372	0,99990787	1	0,99990745	1848	0	17	185354
TauE	1	1	1	1	1	1	3	0	0	187216
TauK	0,94736829	0,97297297	0,97368408	0,94598526	0,94736842	1	18	1	0	187200
TauL	1	1	1	1	1	1	1	0	0	187218
TauM	0,99879199	0,9993932	0,99939831	0,99879135	0,99882491	0,99996213	850	1	7	186361
TauR	0,9999731	0,9999865	0,9999865	0,9999731	1	0,99997299	651	0	5	186563
TauX	0,99992174	0,99995915	0,99995916	0,99992173	1	0,99991831	153	0	14	187052
TauY	0,9599999	0,97959184	0,9799999	0,95920056	0,96	1	24	1	0	187194
TauZ	0,98118719	0,98987805	0,99120665	0,98101176	0,98	0,99995726	49	1	7	187162
Toa	0,99815363	0,99902835	0,99910343	0,99815311	0,9984917	0,99956557	662	1	73	186483
Tpa	0,99178144	0,99585134	0,99591303	0,99174802	0,99175824	0,99997836	361	3	4	186851
Xsc	0,99999464	0,99999731	0,99999731	0,99999464	1	0,99999463	222	0	1	186996
YihO	0,99940195	0,99968654	0,99968664	0,99940177	1	0,99937328	1052	0	106	186061
YihP	0,998199	0,99909411	0,99910385	0,99819761	0,9982548	0,99993483	1144	2	12	186061
YihQ	0,99980905	0,9999031	0,99990311	0,99980904	1	0,99980622	1163	0	35	186021
YihR	0,99995686	0,99997835	0,99997835	0,99995686	1	0,9999567	1104	0	8	186107
YihS	0,99973117	0,99986285	0,99986287	0,99973114	1	0,99972574	1188	0	49	185982
YihT	0,99991884	0,99995921	0,99995921	0,99991883	1	0,99991842	1441	0	15	185763
YihU	0,99998383	0,99999191	0,99999191	0,99998383	1	0,99998382	1456	0	3	185760
YihV	0,99997844	0,9999892	0,9999892	0,99997844	1	0,99997841	1444	0	4	185771

Table S3: Organisms for case study

Phylum	Class	Order	Species	Reference regarding sulfur metabolism
Firmicutes	Clostridia	Eubacteriales	<i>Eubacterium rectale</i> ATCC 33656	Sharma et al., 2022
Verrucomicrobia	Methylocidiphilae	Mehtyladidiphilales	<i>Methylocidiphilum fumarolicum</i> SoV	Schmitz et al., 2022
Proteobacteria	Deltaproteobacteria	Desulfovibrionales	<i>Desulfovibrio alaskensis</i> G20	Burrichter et al., 2021
	Deltaproteobacteria	Desulfovibrionales	<i>Desulfovibrio desulfuricans</i> DSM 642	Burrichter et al., 2021 Burrichter et al., 2021, Hanson et al., 2021
	Deltaproteobacteria	Desulfovibrionales	<i>Bilophila wadsworthia</i> 3_1_6	Denger et al., 2009
	Alphaproteobacteria	Rhodobacterales	<i>Octadecabacter arcticus</i> 238	Denger et al., 2009
	Alphaproteobacteria	Rhodobacterales	<i>Octadecabacter antarcticus</i> 307	Denger et al., 2009
	Alphaproteobacteria	Rhodobacterales	<i>Roseobacter litoralis</i> Och 149	Denger et al., 2009
	Alphaproteobacteria	Rhodobacterales	<i>Roseobacter denitrificans</i> Och 114	Denger et al., 2009
	Alphaproteobacteria	Rhodobacterales	<i>Roseovarius nubinhibens</i> ISM	Denger et al., 2009 Denger et al., 2009, Gorzyska et al., 2006
	Alphaproteobacteria	Rhodobacterales	<i>Ruegeria pomeroyi</i> DSS-3	Koch & Dahl, 2018
	Alphaproteobacteria	Hyphomicrobiales	<i>Hyphomicrobium denitrificans</i> ATCC 51888 <sup>T</sup>	Boden et al., 2011
	Alphaproteobacteria	Hyphomicrobiales	<i>Hyphomicrobium sulfonivorans</i> S1	Felux et al., 2015
	Alphaproteobacteria	Hyphomicrobiales	<i>Microvirga lupini</i> Lut6	Felux et al., 2015
	Alphaproteobacteria	Hyphomicrobiales	<i>Salinarimonas rosea</i> DSM 21201	Felux et al., 2015
	Alphaproteobacteria	Hyphomicrobiales	<i>Rhizobium leguminosarum</i> GLR17	Felux et al., 2015
	Alphaproteobacteria	Hyphomicrobiales	<i>Rhizobium pusense</i> CC134	Sharma et al., 2022
	Alphaproteobacteria	Hyphomicrobiales	<i>Agrobacterium salinitolerans</i> CFBP5507	Sharma et al., 2022
	Gammaproteobacteria	Thiotrichales	<i>Methylophaga sulfidovorans</i> DSM 11578	Kröber & Schäfer, 2019
	Gammaproteobacteria	Thiotrichales	<i>Methylophaga thiooxydans</i> L4	Kröber & Schäfer, 2019
	Gammaproteobacteria	Enterobacterales	<i>Leminorella grimantii</i> DSM 5078	Felux et al., 2015
	Gammaproteobacteria	Pseudomonadales	<i>Pseudomonas putida</i> SQ1	Felux et al., 2015
	Gammaproteobacteria	Oceanospirillales	<i>Halomonas zhanjiangensis</i> DSM 21076	Felux et al., 2015
	Gammaproteobacteria	Oceanospirillales	<i>Halomonas smynensis</i> AAD6	Felux et al., 2015

## References

- Boden, R., Borodina, E., Wood, A. P., Kelly, D. P., Murrell, J. C., & Schäfer, H. (2011). Purification and characterization of dimethylsulfide monooxygenase from *Hyphomicrobium sulfonivorans*. *Journal of Bacteriology*, 193(5), 1250-1258.
- Burrichter, A. G., Dorr, S., Bergmann, P., Haiss, S., Keller, A., Fournier, C., Franchini, P., Isono, E., & Schleheck, D. (2021). Bacterial microcompartments for isethionate desulfonation in the taurine-degrading human-gut bacterium *Bilophila wadsworthia*. *BMC Microbiology*, 21(1), 340. <https://doi.org/10.1186/s12866-021-02386-w>
- Felux, A. K., Spittler, D., Klebensberger, J., & Schleheck, D. (2015). Entner-Doudoroff pathway for sulfoquinovose degradation in *Pseudomonas putida* SQ1. *Proceedings of the National Academy of Sciences of the United States of America*, 112(31), E4298-4305. <https://doi.org/10.1073/pnas.1507049112>
- Gorzyska, A. K., Denger, K., Cook, A. M., & Smits, T. H. M. (2006). Inducible transcription of genes involved in taurine uptake and dissimilation by *Silicibacter pomeroyi* DSS-3<sup>T</sup>. *Archives of Microbiology*, 185(5), 402-406. <https://doi.org/10.1007/s00203-006-0106-8>
- Hanson, B. T., Dimitri Kits, K., Löffler, J., Burrichter, A. G., Fiedler, A., Denger, K., Frommeyer, B., Herbold, C. W., Rattei, T., Karcher, N., Segata, N., Schleheck, D., & Loy, A. (2021). Sulfoquinovose is a select nutrient of prominent bacteria and a source of hydrogen sulfide in the human gut. *ISME J*. <https://doi.org/10.1038/s41396-021-00968-0>
- Koch, T., & Dahl, C. (2018). A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds. *ISME Journal*, 12(10), 2479-2491. <https://doi.org/10.1038/s41396-018-0209-7>
- Kröber, E., & Schäfer, H. (2019). Identification of proteins and genes expressed by *Methylophaga thiooxydans* during growth on dimethylsulfide and their presence in other members of the genus. *Frontiers in Microbiology*, 10, 1132. <https://doi.org/10.3389/fmicb.2019.01132>
- Schmitz, R. A., Mohammadi, S. S., van Erven, T., Berben, T., Jetten, M. S. M., Pol, A., & Op den Camp, H. J. M. (2022). Methanethiol consumption and hydrogen sulfide production by the thermoacidophilic methanotroph *Methylocidiphilum fumarolicum* SoV. *Frontiers in Microbiology*, 13, 857442. <https://doi.org/10.3389/fmicb.2022.857442>
- Sharma, M., Lingford, J. P., Petricevic, M., Snow, A. J. D., Zhang, Y., Jarva, M. A., Mui, J. W., Scott, N. E., Saunders, E. C., Mao, R., da Silva, B. M., Pires, D. E. V., Ascher, D. B., McConville, M. J., Davies, G. J., Williams, S. J., & Goddard-Borger, E. D. (2022). Oxidative desulfurization pathway for complete catabolism of sulfoquinovose by bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 119(4). <https://doi.org/10.1073/pnas.2116022119>

Table S4: HMS-S-S vs. HMSS2 Benchmark

HMS-S-S				HMSS2			
Processed assemblies	Time [s]	Average time [s]	Standard deviation [s]	Processed assemblies	Time [s]	Average time [s]	Standard deviation [s]
2	49,6420059			2	21,9698892		
2	76,8069198			2	10,6381671		
2	48,363564	58,2708299	16,06544661	2	15,1214478	15,9098347	5,706850832
4	99,7836242			4	28,0088732		
4	98,4819531			4	29,5106945		
4	107,466704	101,91076	4,855406186	4	24,6718907	27,3971528	2,476723009
8	213,351081			8	42,3750913		
8	194,844924			8	57,6438847		
8	182,181549	196,792518	15,67577017	8	52,5083995	50,8424585	7,769525861
16	458,234438			16	108,003326		
16	392,999264			16	94,1815164		
16	371,884647	407,706116	45,01433357	16	103,61809	101,934311	7,063068834
32	827,781968			32	219,433296		
32	819,331038			32	199,738819		
32	856,199976	834,43766	19,31458874	32	201,004987	206,7257	11,02329501
64	1682,1316			64	425,39541		
64	1677,98398			64	419,53285		
64	1661,64178	1673,91912	10,83284484	64	415,001567	419,976609	5,211111619



---

## HAMSTER: Automated Hidden Markov model generation from collinear syntenic blocks

Tanabe, T. S., & Dahl, C.

---

The degradation of organic sulfur compounds is by no means a linear process, but usually requires microbial consortia to achieve complete biomineralization, as there are multiple degradation strategies even for the same compound (Goddard-Borger & Williams 2017, Burrichter *et al.* 2018, Wolf *et al.* 2022). Knowledge of the microbial interspecies interaction that drive these transformations is advancing (Peck *et al.* 2019, Ye *et al.* 2023). This progress is accompanied by the discovery of new enzymes and even completely new metabolic pathways acting on organic sulfur compounds (Boden *et al.* 2010, Miller *et al.* 2018, Koch & Dahl 2018, Sharma *et al.* 2022, Wolf *et al.* 2022, Tanabe *et al.* 2023b). At the same time novel organic sulfur compounds are constantly discovered, giving rise to numerous metabolic pathways to be discovered (Walker *et al.* 2017, Thume *et al.* 2018, Hou *et al.* 2022). Simultaneously, the number of available genomes is increasing rapidly and steadily (Koonin *et al.* 2021). With the expectation that the number of newly characterized enzymes will increase and that multiple pathways will need to be added to HMSS2, there is also a need to accelerate the production of the HMMs. The HAMSTER pipeline optimizes and automates the process of HMM production, reducing data entry to a minimum.

The construction of HMMs, as done for HMS-S-S is a time-consuming process that requires some manual intervention between and knowledge of the naturally occurring gene arrangements (Tanabe & Dahl 2022). Of particular importance is the initial correct annotation of proteins with a conserved function, so-called equilog proteins (Haft *et al.* 2003), for the production of HMMs. With HAMSTER, the data input was reduced to a minimum of a single protein sequence and a set of genomes. A HMM is then automatically generated for the input sequence. The HMM is then validated as described previously (Tanabe & Dahl 2022). This automatization made some adjustment necessary. Adaptations were especially made to the assignment of genes to a genomic context and the expected significantly larger amount of data to be processed. The original annotation process for obtaining functionally equivalent sequences involved an initial BLAST search for possible homologs. Detected sequences were then sorted into orthologous groups and examined for genomic context. Only sequences that had the same context as the reference protein were retained. If this was not possible, e.g. because the genes lacked a conserved genomic context, a phylogeny was calculated to allow classification into clades. HAMSTER replaces BLAST with MMseqs2, which is 400 times faster than BLAST with similar sensitivity (Steinegger & Söding 2017). Subsequent clustering into clades of similar sequences was performed by linclust (Steinegger & Söding 2018). This allows HAMSTER to perform the initial grouping much faster, and linear scaling of the runtime. Since protein sequences are initially provided by MMseqs2 without knowledge of a specific genomic context, finding conserved gene clusters requires comparing all possible combinations

of clusters. This is a computationally complex task due to the biological possibility of gene loss, duplication, or insertion. Defining the truly corresponding region between two gene clusters with repeating genes is therefore challenging. To accomplish this task even with a large number of gene clusters, HAMSTER implements a modified version of the matchpoint based algorithm for fast detection of collinear syntenic gene clusters (Svetlitsky *et al.* 2020). Proteins that fall into the same cluster of similarity and have the same or similar conserved gene arrangement are then used to generate and validate HMMs. The parameters for sequence similarity, coverage and identity, as well as the number of rearrangements allowed in gene clusters to be considered similar, can be freely defined by the user.

T.S.T. contributed to this study by conceptualization, investigation, validation and writing of the original manuscript: T.S.T. conceptualized the program and the HMM development procedure. T.S.T. programmed the pipeline for the creation and validation of the HMMs and implemented the CSB-finder-S algorithm based on the written formal description. T.S.T. contributed to writing the manuscript.



1 **Title: HAMSTER: Automated Hidden Markov model**  
2 **generation from collinear syntenic blocks**

3 **Authors: Tomohisa Sebastian Tanabe\* & Christiane Dahl\***

4 Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich-Wilhelms-Universität Bonn,  
5 Bonn, Germany

6 **Running title: Hidden Markov model generation from syntenic blocks**

7  
8 **Article type: Application note**

9  
10 \*Corresponding author: Institut für Mikrobiologie & Biotechnologie  
11 Rheinische Friedrich Wilhelms-Universität Bonn  
12 Meckenheimer Allee 168  
13 53227 Bonn, Germany  
14 Tel. +49-228-735591  
15 Fax +49-228-737476  
16 E-mail: s6totana@uni-bonn.de  
17

18

19

20 **Keywords:** Hidden Markov model (HMM) database, collinear syntenic blocks, cross-  
21 validation

22

23

24

25

26 **Abbreviations:**

27 **Abstract**

28 Motivation

29 Hidden Markov models are frequently utilized in ecological research to examine the metabolic  
30 potential of prokaryotic communities. Similarly, evolutionary studies use HMMs because of  
31 their high sensitivity in detecting divergent members of large protein families to study the  
32 development of specific pathways. Generating HMM is a laborious task that is commonly done  
33 by collection of functionally equivalent sequences and the manual selection and inspection of  
34 the sequence data. Automating the process of generating HMMs can accelerate the overall  
35 process and enable the use of HMMs for detecting the increasing number of genes and pro-  
36 teins described.

37 Results

38 Here, we present HAMSTER, an automated Hidden Markov model generation from collinear  
39 syntenic blocks. HAMSTER only requires a set of genome assemblies and a single reference  
40 protein sequence for the generation of an HMM. The sequences for the generation of the  
41 HMM are automatically collected and sorted based on their mutual similarity and conserved  
42 genomic vicinity. HAMSTER takes care of all downstream processes as processing the align-  
43 ment, selecting score cutoff values and validation of the HMM performance. Parameters for  
44 these tasks can be modified by the user.

45

## 46 **1 Introduction**

47 Profiled Hidden Markov models (HMM) are one of the most popular and powerful techniques  
48 for the functional annotation of genes and proteins. In the past years several HMM based  
49 specialized tools for the elucidation of specific metabolisms have been developed (Garber et  
50 al., 2020; Neukirchen & Sousa, 2021; Tanabe & Dahl, 2023; Zhou et al., 2022). These tools  
51 were applied to explore the metabolic capacity of prokaryotes from various habitats or to  
52 reconstruct the evolution of whole pathways (Garcia et al., 2022; Tanabe et al., 2023). To  
53 annotate the increasing number of characterized metabolic pathways, it is necessary to  
54 generate new profiled HMMs that can detect and annotate proteins or genes with a high  
55 degree of reliability. Profiled HMMs are probabilistic models that describe the occurrence of  
56 amino acids at each position of a set of aligned homologous sequences. The alignment-based  
57 modeling of the primary sequence also enables the capture of gaps and insertions at each  
58 position, as well as the degree of conservation for each residue. This sensitivity makes HMMs  
59 useful for detecting even distantly related homologous sequences. (Eddy, 1998). The creation  
60 of profiled HMM alignment typically involves using a set of functionally equivalent sequences  
61 derived from a diverse range of prokaryotes. To ensure high sensitivity towards distantly  
62 related sequences, an HMM-specific cutoff score is necessary to prevent the detection of  
63 functionally divergent proteins. Collecting these sequences can be computationally intensive  
64 and requires synteny and gene similarity analysis. (Garber et al., 2020; Neukirchen & Sousa,  
65 2021). The process of selecting HMM-specific score cutoffs usually involves manual inspection  
66 to determine the optimal score threshold for distinguishing true and false hits. (Garber et al.,  
67 2020; Neukirchen & Sousa, 2021). However, automated selection of a cutoff score to balance  
68 sensitivity and specificity can also improve the reliability of the HMM (Srivastava et al., 2007).  
69 This work presents a new fully automated pipeline for generating, calculating score cutoffs,  
70 and validating HMMs. The pipeline collects sequences for alignment based on their genomic  
71 vicinity and sequence similarity. It only requires a single reference sequence per HMM and a  
72 set of genome assemblies.

## 73 2 Implementation

74 Hamster is a python-based command-line tool. It requires a single representative  
75 protein sequence for each protein and a set of genome assemblies of any level (e.g. complete  
76 genomes, chromosomes, scaffolds, contigs). Assemblies can either be provided GFF3 files with  
77 protein FASTA files or as nucleotide FASTA. In the latter case the open-reading frames are  
78 automatically predicted and translated via Prodigal before further processing (Hyatt et al.,  
79 2010).

80 First homologues sequences to the input sequences are retrieved from all genome  
81 assemblies via mmseqs2 and stored in a local sqlite database (Steinegger & Söding, 2017). The  
82 genes that are genetically adjacent to the retrieved sequences are also saved. All these  
83 sequences are then clustered through linclust from the mmseqs2 package into clusters of  
84 similar sequences (Steinegger & Söding, 2017). Default parameters for these sequence search  
85 and clustering steps are provided but can be modified by the user as needed. The detected  
86 proteins and genes are then annotated by the cluster of similarity and the most similar  
87 reference protein. Using gene localization of the sequences detected with mmseqs2 and the  
88 preliminary annotation, conserved genomic vicinities are determined. For this task a modified  
89 implementation of the CSBfinder-S algorithm was used to discover conserved collinear  
90 syntenic gene clusters (Svetlitsky et al., 2020). This algorithm compares all gene clusters and  
91 automatically sorts them into groups of conserved collinear syntenic blocks using Jaccard  
92 similarity. The user can adjust the minimal size of collinear syntenic blocks, the minimal  
93 number of appearances, and the Jaccard similarity, as in CSBfinder (Svetlitsky et al., 2020).  
94 HMMs are then generated from sequences corresponding to the same sequence similarity  
95 cluster and gene cluster group. Collected protein sequences are aligned with M-coffee  
96 combining the alignment algorithms Clustal, MUSCLE and MAFFT (Wallace et al., 2006). The  
97 aligned positions in each alignment are filtered using a transitive consistency score to remove  
98 any unreliable positions. (Wallace et al., 2006). To establish HMM-specific score cutoffs, the  
99 HMMs undergo validation through a cross-validation method. The alignment is divided into  
100 five equal folds, with four folds used to create a temporary HMM and the sequences of the  
101 fifth fold serving as positive example sequences to be detected by the temporary HMM. This  
102 process is repeated five times until each fold has been tested for detection using the remaining

103 four folds. This approach is complemented by including all sequences that are not within the  
104 same sequence similarity cluster, which serve as true negative examples. Additionally, cutoff  
105 values are assigned during the procedure to optimize the results. The HMM, along with the  
106 cutoff values and a validation report, are provided to the user.

### 107 **3 Conclusion**

108 HAMSTER generates profiled Hidden Markov models from genomic data with only a single  
109 sequence input. The internal sequence collection uses information on sequence similarity and  
110 genomic vicinity to select functionally equivalent protein sequences. This allows for automa-  
111 tized pipeline execution of HMM generation for recognizing proteins encoded in syntenic gene  
112 clusters. The HMMER3 algorithm utilizes HMMs and cutoff values that are highly flexible for  
113 use in various applications.

### 114 **ACKNOWLEDGEMENTS**

115 This work was supported by the Deutsche Forschungsgemeinschaft (grant Da 351/13-1 to  
116 Christiane Dahl). TST received a scholarship from the Studienstiftung des Deutschen Volkes.

### 117 **AUTHOR CONTRIBUTIONS**

118 TST conceived the study. TST developed and implemented the method. TST wrote the  
119 manuscript.

### 120 **DATA AVAILABILITY STATEMENT**

121 Hamster is available from <https://github.com/TSTanabe/HAMSTER>

122

123 **References**

- 124 Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics (Oxford, England)*, *14*(9), 755–763.
- 125 Garber, A. I., Nealson, K. H., Okamoto, A., McAllister, S. M., Chan, C. S., Barco, R. A., & Merino, N.  
126 (2020). Fegenie: A Comprehensive Tool for the Identification of Iron Genes and Iron Gene  
127 Neighborhoods in Genome and Metagenome Assemblies. *Frontiers in Microbiology*, *11*, 37.
- 128 Garcia, P. S., D'Angelo, F., Ollagnier de Choudens, S., Dussouchaud, M., Bouveret, E., Gribaldo, S., &  
129 Barras, F. (2022). An early origin of iron-sulfur cluster biosynthesis machineries before Earth  
130 oxygenation. *Nature Ecology & Evolution*, *6*(10), 1564–1572.
- 131 Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal:  
132 Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*,  
133 *11*, 119.
- 134 Neukirchen, S., & Sousa, F. L. (2021). DiSCo: A sequence-based type-specific predictor of Dsr-  
135 dependent dissimilatory sulphur metabolism in microbial data. *Microbial Genomics*, *7*(7),  
136 000603.
- 137 Srivastava, P. K., Desai, D. K., Nandi, S., & Lynn, A. M. (2007). HMM-ModE--improved classification  
138 using profile hidden Markov models by optimising the discrimination threshold and modifying  
139 emission probabilities with negative training sequences. *BMC Bioinformatics*, *8*, 104.
- 140 Steinegger, M., & Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the  
141 analysis of massive data sets. *Nature Biotechnology*, *35*(11), 1026–1028.
- 142 Svetlitsky, D., Dagan, T., & Ziv-Ukelson, M. (2020). Discovery of multi-operon colinear syntenic blocks  
143 in microbial genomes. *Bioinformatics (Oxford, England)*, *36*(Suppl\_1), i21-i29.
- 144 Tanabe, T. S., & Dahl, C. (2023). HMSS2: An advanced tool for the analysis of sulphur metabolism,  
145 including organosulphur compound transformation, in genome and metagenome assemblies.  
146 *Molecular Ecology Resources*, *23*(8), 1930–1945.
- 147 Tanabe, T. S., Grosser, M., Hahn, L., Kümpel, C., Hartenfels, H., Vtulkin, E., Flegler, W., & Dahl, C. (2023).  
148 Identification of a novel lipoic acid biosynthesis pathway reveals the complex evolution of  
149 lipoate assembly in prokaryotes. *PLoS Biology*, *21*(6), e3002177.
- 150 Wallace, I. M., O'Sullivan, O., Higgins, D. G., & Notredame, C. (2006). M-Coffee: Combining multiple  
151 sequence alignment methods with T-Coffee. *Nucleic Acids Research*, *34*(6), 1692–1699.
- 152 Zhou, Z., Tran, P. Q., Breister, A. M., Liu, Y., Kieft, K., Cowley, E. S., Karaoz, U., & Anantharaman, K.  
153 (2022). METABOLIC: High-throughput profiling of microbial genomes for functional traits,  
154 metabolism, biogeochemistry, and community-scale functional networks. *Microbiome*, *10*(1),  
155 33.

---

## Identification of a novel lipoic acid biosynthesis pathway reveals the complex evolution of lipoate assembly in prokaryotes

Tanabe, T. S., Grosser, M., Hahn, L., Kämpel, C. Hartenfels, H., Vtulkin, E., Flegler, W., & Dahl, C.

---

Lipoate is an organosulfur cofactor that has key functions in central carbon metabolism and dissimilatory sulfur oxidation. The lipoate-dependent multienzyme complexes characterized so far are three  $\alpha$ -ketoacid dehydrogenases, acetoin dehydrogenase and the glycine cleavage complex and the lipoate-binding protein of the sHdr system (Cronan 2016). Unlike many other cofactors lipoate is only active when it is covalently attached to its cognate enzyme. Even the biosynthesis of the cofactor takes place on an already protein-bound octanoyl precursor and not on free octanoic acid (Zhao *et al.* 2003, Cronan 2016). It was assumed that all lipoate-requiring enzymes are supplied with lipoate by a biosynthetic machinery, which generally consists of lipoate:protein ligase, lipoate synthase LipA and octanoyltransferase LipB or LipM (Cronan 2016). This view was challenged when this canonical biosynthesis machinery did not assemble lipoate on LbpA from *H. denitrificans* in *E. coli* (Cao *et al.* 2018). Instead *in-vitro* lipoylation via the *bona-fide* lipoate:protein ligase sLpl(AB) was required for the production of the holo-LbpA. The existence of such a specific lipoate biosynthesis pathway for LbpA was even more confusing as *H. denitrificans* genetically encodes for canonical LipA and LipB lipoate biosynthesis (Cao *et al.* 2018).

Here, a biochemical and genetic proof for a novel assembly pathway in bacteria was presented. This machinery assembles lipoate on LbpA by octanoylation of apo-LbpA via the *bona-fide* lipoate:protein ligase sLpl(AB) and free octanoic acid. Sulfur insertion at C6 and C8 are then individually catalyzed by two distinct lipoate synthases LipS1 and LipS2 (Tanabe *et al.* 2023b). The essential function of sLpl(AB) in the lipoate assembly was demonstrated by the generation of the genetically modified *H. denitrificans* strains. The thiosulfate oxidation rates of the strains *H. denitrificans*  $\Delta$  *tsdA* *lbpA*-His and *H. denitrificans*  $\Delta$  *tsdA*  $\Delta$  *slpl*(AB) *lbpA*-His, were compared with the strains *H. denitrificans*  $\Delta$  *tsdA* and *H. denitrificans*  $\Delta$  *tsdA*  $\Delta$  *lbpA*. Purified LbpA-His from these strains was applied to a gel shift assay to determine the presence of the holo- or apo-LbpA. These experiments also demonstrated the non-redundant functionality of the canonical and novel lipoate assembly machinery. These experiments were complemented by the production of recombinant LbpA in the presence of the lipoate synthases LipS1/LipS2 and sLpl(AB) in *E. coli*. Mass spectrometry of the purified LbpA revealed the presence of mercaptanoylated species, consistent with sequential sulfur insertion by LipS1/LipS2 (Neti *et al.* 2022). This modification is unlikely to be a result of the

---

Tanabe, T. S., Grosser, M., Hahn, L., Kämpel, C. Hartenfels, H., Vtulkin, E., Flegler, W., & Dahl, C. (2023). Identification of a novel lipoic acid biosynthesis pathway reveals the complex evolution of lipoate assembly in prokaryotes *PLoS Biol.*, 21(6), e3002177; doi.org/10.1371/journal.pbio.3002177

LipA activity, since these lipoate synthases insert both sulfurs without the release of a mercaptoacyl-acylated species (Cicchillo & Booker 2005, Douglas *et al.* 2006, Lanz *et al.* 2014, Cronan 2016).

The second question was the reason for two non-redundant, parallel active lipoate assembly systems in the same organisms. To reconstruct the evolution of the lipoate synthesis, HMSS2 was used to search the genome taxonomy database (GTDB) for the presence of lipoate assembly systems, lipoate-requiring enzymes and the sHdr system (Tanabe & Dahl 2023, Parks *et al.* 2022). This analysis revealed a much broader distribution of the canonical and novel pathways than previously described. Canonical assembly was found in nearly all bacterial and archaeal phyla and was by no means restricted to the phyla Pseudomonadota (formerly Proteobacteria) and Bacillota (formerly Firmicutes), the only phyla for which this system was previously known. Canonical and novel pathway also co-occurred in most of the genomes analyzed. Although there was a high correlation between the presence of the sHdr system and the novel pathway, the latter was not limited to sulfur-oxidizing bacteria. In fact, it was highly abundant in archaeal phyla which were described as having highly diverse physiology. Phylogeny of the lipoate:protein ligases that could be confidently assigned to either the novel or the canonical pathway resulted in a bipartite split between these two ligases. One clade consisted of ligases exclusively associated with the canonical pathway and were all of bacterial origin. In contrast, ligases associated with the novel pathway were derived from archaeal and bacterial genomes and clustered in a second clade, with archaeal ligases being the most ancient ones. Thus, the novel pathway was considered to be of archaeal origin. The evolutionary history of the lipoate synthases LipS1 and LipS2 did also not contradict this evolutionary scenario. With these phylogenies it was possible to formulate an evolutionary scenario for the substrate specificity of the novel pathway towards LbpA. In this scenario sLpl(AB) co-evolved with an LbpA-like protein as its substrate in the archaeal domain. Both were then horizontally transferred into a bacterium that already possessed the canonical pathway. In the bacterium, the specific protein-protein interaction of the sLpl(AB) LbpA-like protein was retained and sLpl(AB) did not recognize any other lipoate-requiring domains as substrates. In this scenario, the two pathway would operate independently due to their different evolutionary histories (Tanabe *et al.* 2023b).

T.S.T. contributed to this study by conceptualization, Data curation, formal analysis, investigation, validation, visualization and writing of the original manuscript: T.S.T. conceptualized and constructed the genetically modified strain *H. denitrificans*  $\Delta$  *tsdA lbpA-His*. T.S.T. produced LbpA-His from *H. denitrificans*  $\Delta$  *tsdA lbpA-His* and *H. denitrificans*  $\Delta$  *tsdA*  $\Delta$  *slpl(AB) lbpA-His*. All experiments regarding the evolution and distribution of the were conceptualized, performed, analyzed, validated and visualized by T.S.T., who also contributed to writing the original draft and all revisions. The Supplementary Tables S2, S3, S4, S5 and Supplementary Data S2 and S3 are not printable due to their format and/or size, but are accessible through the online version of the publication.



## RESEARCH ARTICLE

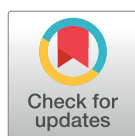
# Identification of a novel lipoic acid biosynthesis pathway reveals the complex evolution of lipoate assembly in prokaryotes

Tomohisa Sebastian Tanabe, Martina Grosser, Lea Hahn<sup>‡</sup>, Carolin Kümpel, Hanna Hartenfels, Evelyn Vtulkin, Wanda Flegler, Christiane Dahl<sup>‡\*</sup>

Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

<sup>‡</sup> Current address: Institut für Integrierte Naturwissenschaften, Arbeitsgruppe Mikrobiologie, Universität Koblenz, Koblenz, Germany

\* [chdahl@uni-bonn.de](mailto:chdahl@uni-bonn.de)



## OPEN ACCESS

**Citation:** Tanabe TS, Grosser M, Hahn L, Kümpel C, Hartenfels H, Vtulkin E, et al. (2023) Identification of a novel lipoic acid biosynthesis pathway reveals the complex evolution of lipoate assembly in prokaryotes. *PLoS Biol* 21(6): e3002177. <https://doi.org/10.1371/journal.pbio.3002177>

**Academic Editor:** Thomas A. Richards, University of Oxford, UNITED KINGDOM

**Received:** March 8, 2023

**Accepted:** May 31, 2023

**Published:** June 27, 2023

**Copyright:** © 2023 Tanabe et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting Information](#) files. Structural data needed for [S1 Fig](#) is available in the RCSB Protein Data Bank (accession 3AB9).

**Funding:** This work was funded by the Deutsche Forschungsgemeinschaft (<https://www.dfg.de/>, grants Da 351/13-1 and Da 351/14-1 to CD). TST received a scholarship from the Studienstiftung des Deutschen Volkes (<https://www.studienstiftung.de/>). The funders had no role in study design, data

## Abstract

Lipoic acid is an essential biomolecule found in all domains of life and is involved in central carbon metabolism and dissimilatory sulfur oxidation. The machineries for lipoate assembly in mitochondria and chloroplasts of higher eukaryotes, as well as in the apicoplasts of some protozoa, are all of prokaryotic origin. Here, we provide experimental evidence for a novel lipoate assembly pathway in bacteria based on a sLpI(AB) lipoate:protein ligase, which attaches octanoate or lipoate to apo-proteins, and 2 radical SAM proteins, LipS1 and LipS2, which work together as lipoyl synthase and insert 2 sulfur atoms. Extensive homology searches combined with genomic context analyses allowed us to precisely distinguish between the new and established pathways and map them on the tree of life. This not only revealed a much wider distribution of lipoate biogenesis systems than expected, in particular, the novel sLpI(AB)–LipS1/S2 pathway, and indicated a highly modular nature of the enzymes involved, with unforeseen combinations, but also provided a new framework for the evolution of lipoate assembly. Our results show that dedicated machineries for both de novo lipoate biogenesis and scavenging from the environment were implemented early in evolution and that their distribution in the 2 prokaryotic domains was shaped by a complex network of horizontal gene transfers, acquisition of additional genes, fusions, and losses. Our large-scale phylogenetic analyses identify the bipartite archaeal LpIAB ligase as the ancestor of the bacterial sLpI(AB) proteins, which were obtained by horizontal gene transfer. LipS1/S2 have a more complex evolutionary history with multiple of such events but probably also originated in the domain archaea.

## Introduction

$\alpha$ -Lipoic acid is a cofactor found in all domains of life and is involved in key reactions of central carbon metabolism and dissimilatory sulfur oxidation [1–4]. In this eight-carbon saturated fatty acid, sulfur atoms replace the hydrogen atoms of carbons 6 and 8 of the acyl chain [5].

collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

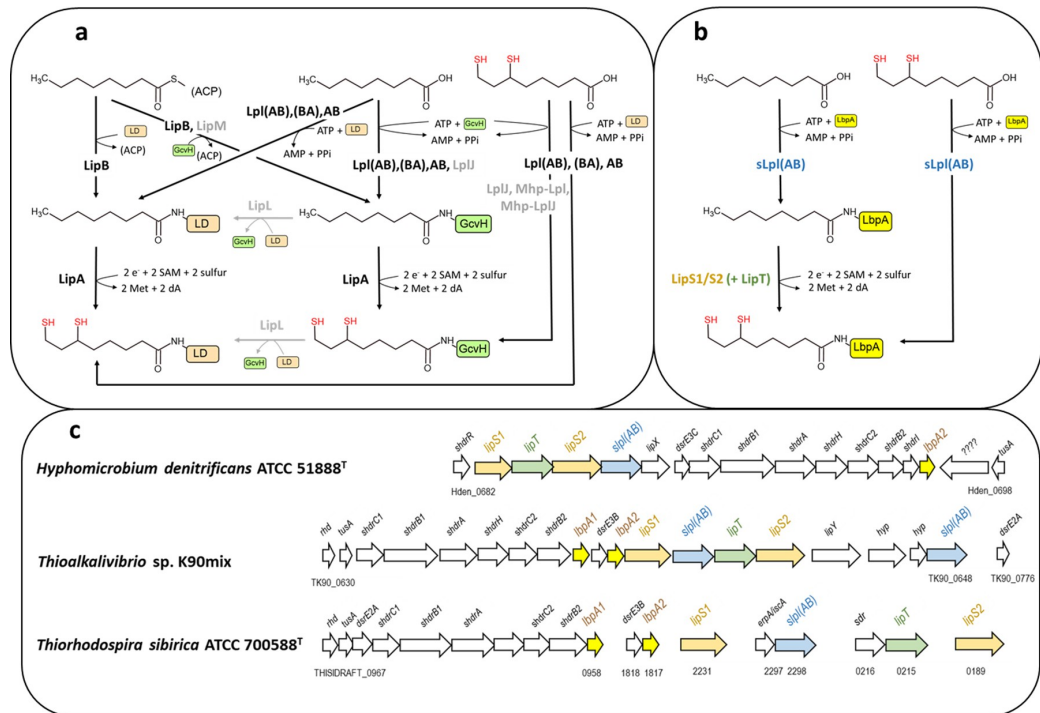
**Abbreviations:** ACP, acyl carrier protein; GTDB, Genome Taxonomy Database; LB, lysogeny broth; SAM, S-adenosylmethionine.

Only a few, but particularly important, lipic acid-dependent enzyme systems have been described [2,3,6]. These include 3  $\alpha$ -ketoacid dehydrogenases, such as pyruvate dehydrogenase, whose E2 subunits bind lipic acid. In the glycine cleavage complex, lipocate is bound to the glycine cleavage H protein (GcvH) [3]. Lipoylated proteins also play an important role in combatting reactive oxygen species [7,8]. In 2018, we discovered another lipocate-binding protein (LbpA) homologous to GcvH (S1 Fig) and demonstrated that it is an essential component of the sulfur-oxidizing heterodisulfide reductase-like (sHdr) system present in a wide range of bacterial and archaeal dissimilatory sulfur oxidizers [2,9].

Two posttranslational machineries are known to construct lipoyl moieties [1,3]: The first requires an acyl carrier protein (ACP)-bound octanoyl residue from endogenous fatty acid biosynthesis to be transferred to the  $\epsilon$ -amino groups of conserved lysine residues in the accepting apo-proteins. In the second, free lipocate or octanoate are hooked up to the target lysine. Irrespective of the initial step, a sulfur atom must be added to each of the octanoyl C<sub>6</sub> and C<sub>8</sub> atoms to complete lipocate biosynthesis (Fig 1A). Using free precursors involves the enzyme lipocate:protein ligase that activates the precursors to lipoyl- or octanoyl-AMP at the expense of ATP before transfer to the target protein. In many bacteria, including *Escherichia coli*, the ligase consists of 2 fused domains, the catalytic domain LplA and the accessory domain LplB [10,11]. We denote these enzymes Lpl(AB) or in the circularly permuted case [12], Lpl(BA). Ligases with tight substrate specificity have been described that transfer free precursors only to GcvH and, in one case, also to the E2 subunit of 2-oxoglutarate dehydrogenase [13–15]. Additional amidotransferases or ligases are then necessary for modification of other lipoyl domains [13,14,16]. Bipartite lipocate:protein ligases forming a functional LplA-LplB heterodimer (denoted LplAB here) have so far been found primarily in archaea [10]. In the absence of free precursors, an octanoyl residue derived from ACP is attached to the target protein by an octanoyltransferase, LipB or LipM [1,16–18]. LipB has been found mainly in Proteobacteria, serves as an all-purpose transferase and provides octanoate or lipocate to all known lipocate-requiring pathways except the sHdr-system [2,19]. LipM has been proposed to occur predominantly in Firmicutes and to transfer octanoyl residues exclusively to GcvH. An amidotransferase, LipL, is required for the transfer of octanoyl or lipoyl moieties from GcvH to the E2-subunits of pyruvate and branched-chain  $\alpha$ -ketoacid dehydrogenases [13,14,20]. Despite poor sequence conservation, an evolutionary relationship has been detected between lipocate:protein ligases and octanoyltransferases as well as biotin:protein ligases (BirA) [6]. Once the octanoyl residues arrive at their target proteins, they become substrates for lipocate synthase LipA, a member of the radical S-adenosylmethionine (SAM) superfamily [21], which sequentially adds 2 sulfur atoms in a single reaction, first at position C<sub>6</sub> and then at C<sub>8</sub> [1,22,23].

Initial evidence for an alternative pathway for lipocate assembly in prokaryotes emerged, when we showed that the GcvH-like LbpA proteins involved in sHdr-based sulfur oxidation are not modified by the canonical *E. coli* and *Bacillus subtilis* lipoyl attachment machineries [2]. Instead, the bacterial and archaeal *shdr-lbpA* clusters are accompanied by a set of genes encoding a specific lipoylation pathway (Fig 1B and 1C) that includes lipocate:protein ligases (sLpl(AB)) and 2 proteins of the radical SAM superfamily, originally termed RadSAM1 and RadSAM2. sLpl(AB) lipocate:protein ligases from sulfur oxidizers not only lipoylate LbpA acceptor proteins from the same organism *in vitro* but also show cross-species functionality among sulfur oxidizers while failing to recognize lipoyl domains/proteins from organisms lacking components of a sHdr-LbpA sulfur-oxidizing system [2].

Recently, proteins from the thermophilic archaeon *Thermococcus kodakarensis* similar to sLpl(AB) ligase (*Tk*-Lpl-N and *Tk*-Lpl-C) and RadSAM1 and RadSAM2 from sulfur oxidizers (now termed LipS1 and LipS2) were shown to exert octanoate/lipocate:protein ligase and LipA-like lipocate synthase functions, respectively, on chemically synthesized peptide substrates in



**Fig 1. Processes and components of lipolate assembly pathways.** (a) Main known steps of established lipolate assembly pathways. Enzymes and steps not occurring in *E. coli* but described for other organisms are printed in gray. LipM and LipL have been demonstrated in Firmicutes, *B. subtilis* [24], *Staphylococcus aureus* [25], and *Listeria monocytogenes* [26], as well as in Tenericutes, *Mycoplasma hyopneumoniae* [15,16]. (b) Predicted novel lipolate assembly pathway. The pathway is substantiated by experiments reported here as well as by published work on proteins from the 3 model organisms depicted in c [2,9] and by genetic and biochemical work on LipS1 and LipS2 from the archaeon *Thermococcus kodakarensis* [27]. Lipolate:protein ligases from sulfur oxidizers were originally reported not to contain a carboxy-terminal LplB domain based on superposition of the structure modeled for *Thioalkalivibrio* sp. K90mix by using the automated mode of SWISS\_Model on *E. coli* Lpl(AB). We challenged this view and indeed, modeling by AlphaFold [28] as well as sequence alignments yielded clear proof of the presence of the LplB domain (S2 Fig). (c) Genetic arrangement of 3 novel systems for lipolate assembly in Proteobacteria. Colors correspond to the biochemical roles as depicted in b. For *Ts. sibirica* locus tags are given according to JGI-IMG. LipT is an FAD-binding NAD(P)H-dependent oxidoreductase possibly delivering electrons for the LipS1/S2-catalyzed sulfur insertion step. The genes *lipY* and *lipX* encode a putative fatty acid transporter and a putative glutamine amidotransferase, respectively. ACP, acyl carrier protein; GcvH, glycine cleavage system protein H; LbpA, lipolate-binding protein; LD, lipoyl domains of the 2-oxoacid dehydrogenases.

<https://doi.org/10.1371/journal.pbio.3002177.g001>

vitro. Genetic analysis provided further evidence that these proteins are involved in archaeal lipolate biosynthesis [16,27,29] (Fig 1B). We have previously shown that the *sLpl(AB)*-encoded protein:lipolate ligase from the gammaproteobacterial sulfur oxidizer *Thioalkalivibrio* sp. K90mix accepts only free precursors, i.e., octanoate or lipolate, and lacks octanoyltransferase activity [2]. The archaeal ligase is also thought to be restricted to free substrates, mainly because ACPs do not occur in archaea [30].

Here, we provide conclusive experimental evidence for the existence of a novel sLpl(AB)-LipS1/S2-based lipolate assembly pathway not only in archaea but also in bacteria and raise the question of whether it is restricted to thermophilic archaea and sulfur-oxidizing bacteria or is of more general importance. To date, no studies have been published on the origin and evolution of lipolate assembly machineries despite the importance of lipolate for almost all living organisms. This prompted us to carry out an exhaustive large-scale analysis including a large fraction of prokaryotic diversity, and in particular, the archaea for which knowledge about

lipolate assembly is scarce. We mapped the novel lipolate synthesis pathway on the tree of life revealing an enormously wide distribution. Finally, our analyses show that the novel lipolate synthesis pathway evolved in the archaeal domain.

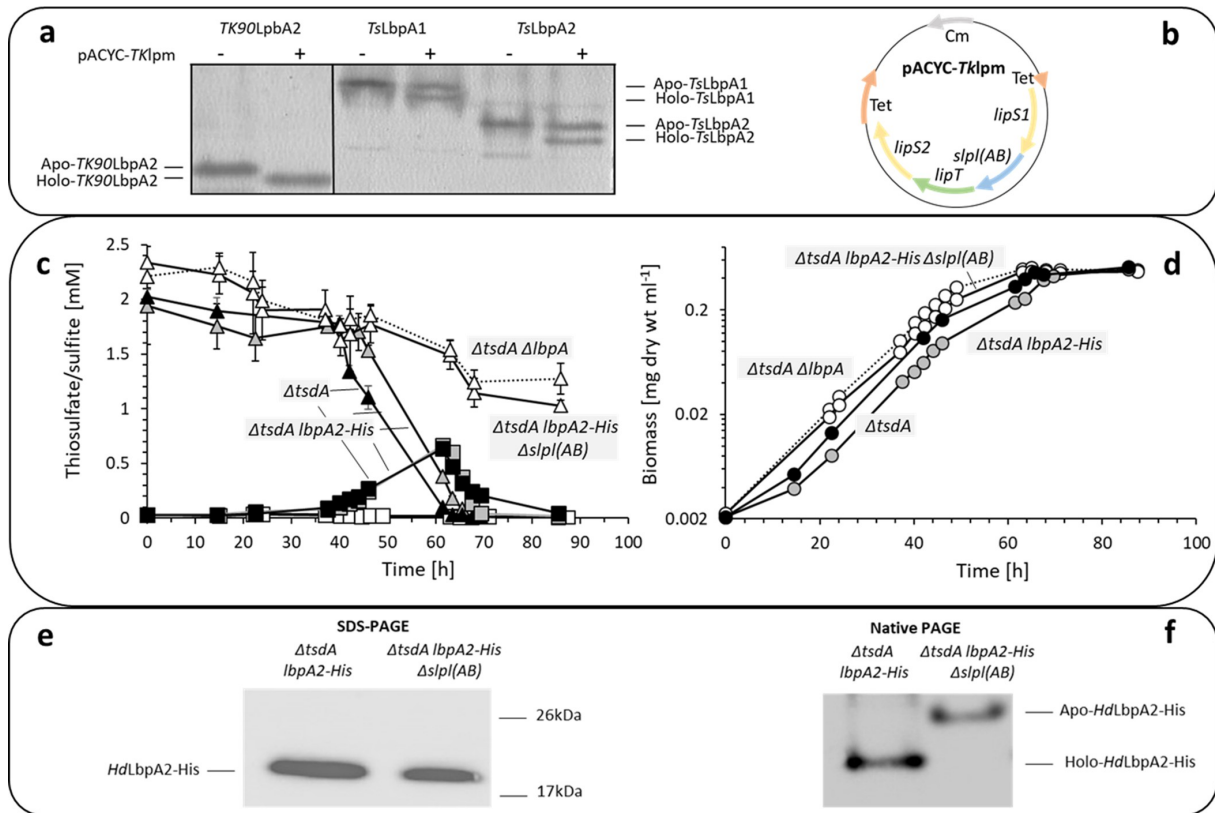
## Results

### Biochemical and genetic proof for an sLpl(AB)-LipS1/S2-based lipolate assembly pathway in bacteria

The initial proposal of a novel route for maturation of lipolate-binding proteins in bacteria relied on the detection of conspicuous *lipS1-sLpl(AB)-lipT-lipS2* gene clusters, in vitro assays with sLpl(AB) lipolate:protein ligase from a model sulfur oxidizer and genetic complementation studies in *E. coli* and *B. subtilis* [2]. Here, we set out to collect conclusive experimental evidence for the functionality of the pathway in bacteria. A focus was kept on the biochemically characterized sLpl(AB) ligases from the sulfur oxidizers *Thiorhodospira sibirica* and *Thioalkalivibrio* sp. K90mix [2].

First, 3 Strep-tagged LbpA lipolate acceptor proteins from these 2 bacteria were recombinantly produced in *E. coli*, with or without a helper plasmid from which the *Thioalkalivibrio* assembly genes *lipS1-sLpl(AB)-lipT-lipS2* were expressed under control of the pACYC184 *tet* promoter (Fig 2A and 2B). Native gel electrophoresis showed the faster mobility expected for holo-LbpAs only when produced in the presence of the helper plasmid. This is due to the lack of the positive charge when the lipolate-binding lysine is modified by covalent attachment of lipolate or octanoate (Fig 2A). Mass spectrometric analyses confirmed posttranslational modification of all 3 LbpA acceptor proteins by a 157-Da mercaptooctanoyl moiety in the presence of the helper plasmid (S3 Fig), fully consistent with in vitro results for the archaeal system where LipS2 first catalyzes sulfur attachment at C<sub>8</sub> of an artificial octanoyllysyl peptide substrate and LipS1 then inserts the second sulfur at C<sub>6</sub> [29]. Although this last step was not efficiently catalyzed in the heterologous environment, our experiments clearly confirm specific modification of sulfur oxidizer LbpA acceptor proteins by lipolate assembly proteins from a sulfur oxidizer.

In a second approach, 4 strains of the Alphaproteobacterium *Hyphomicrobium denitrificans* were studied. The organism is accessible for manipulative genetics, the necessity of its LbpA2 protein for the oxidation of thiosulfate is documented [2,9] and the respective genes are located in immediate vicinity of a *lipS1-lipT-lipS2-sLpl(AB)* cluster (Fig 1C). *H. denitrificans*  $\Delta$ *tsdA* served as the reference strain. It lacks thiosulfate dehydrogenase (TsdA) that catalyzes the formation of the dead-end product tetrathionate. Thus, the strain can oxidize thiosulfate exclusively via the sHdr-LbpA pathway [9,31] that substantially simplifies its elucidation by reverse genetics. When grown in the presence of methanol as a carbon source and thiosulfate as an additional electron source, it excretes toxic sulfite, which causes growth retardation [31]. Functionality of the sHdr-LbpA pathway is thus easily detectable by sulfite formation and diminished growth rate (Fig 2C and 2D). The second strain studied carries a  $\Delta$ *lbpA2* deletion in a  $\Delta$ *tsdA* background, is unable to oxidize thiosulfate, and served as control. In the third strain, *lbpA2-His*, encoding carboxy-terminally His-tagged LbpA2, replaces the original *lbpA* gene in the chromosome of *H. denitrificans*  $\Delta$ *tsdA*, so that LbpA2 can be purified from this strain. The fourth strain carries an *in frame* deletion of *sLpl(AB)* in a  $\Delta$ *tsdA lbpA2-His* background. Thus, the modification of the lipolate acceptor LbpA2 can be compared in presence or absence of the sLpl(AB) ligase. *H. denitrificans*  $\Delta$ *tsdA* and  $\Delta$ *tsdA lbpA2-His* oxidized thiosulfate completely and excreted up to 0.6 mM sulfite (Fig 2C and 2D). This demonstrates that the addition of the carboxy-terminal His-tag does not prevent proper function of the LbpA2 protein. In contrast, in both, the  $\Delta$ *tsdA*  $\Delta$ *lbpA2* and the  $\Delta$ *tsdA lbpA2-His*  $\Delta$ *sLpl(AB)* strains, thiosulfate degradation was very slow, sulfite formation was not observed (Fig 2C), and the growth



**Fig 2. Biochemical and genetic evidence for a novel lipocate assembly pathway in bacteria.** (a) LbpAs from *Thioalkalivibrio* sp. K90mix and *Ts. sibirica* were produced in *E. coli* BL21(DE3)  $\Delta$ iscR, a strain designed for improved synthesis of iron-sulfur proteins [32], either with or without a helper plasmid (pACYC-Tklpm) carrying genes *lipS1-sLpl(AB)-lipT-lipS2* from *Thioalkalivibrio* sp. K90mix (shown in b) under control of the constitutive pACYC184 *tet* promoter. Holo-LbpAs migrate faster in native PAGE due to loss of the positive lysine charge upon modification. In the heterologous host, *TsLbpA* proteins are—albeit not fully—modified by the assembly proteins stemming from a different species, i.e., *Thioalkalivibrio*. (c) Thiosulfate (triangles) and sulfite (boxes) concentrations for 4 different *H. denitrificans* strains during growth on methanol (24.4 mM) as a carbon source in the presence of 2 mM thiosulfate. (d) Growth of *H. denitrificans* strains. Symbols and lines in c and d correspond to *H. denitrificans* strains as follows: filled black symbols, solid lines: *H. denitrificans*  $\Delta$ tsdA; symbols filled gray, solid lines: *H. denitrificans*  $\Delta$ tsdA *lbpA2-His*; open symbols, dotted lines: *H. denitrificans*  $\Delta$ tsdA  $\Delta$ lbpA2; open symbols, solid lines: *H. denitrificans*  $\Delta$ tsdA *lbpA2-His*  $\Delta$ sLpl(AB). For all measurements, standard deviations based on 3 technical replicates are indicated, but too small to be visible for determination of biomass and sulfite. (e) SDS-PAGE of *HdlbpA2-His* enriched from *H. denitrificans*  $\Delta$ tsdA *lbpA2-His* (left lane, 2.5  $\mu$ g protein) and  $\Delta$ tsdA  $\Delta$ sLpl(AB) *lbpA2-His* (right lane, 1.5  $\mu$ g protein). (f) Native gel mobility shift assay for *HdlbpA2-His* enriched from *H. denitrificans*  $\Delta$ tsdA *lbpA2-His* (left lane, 2.5  $\mu$ g protein) and  $\Delta$ tsdA  $\Delta$ sLpl(AB) *lbpA2-His* (right lane, 1.3  $\mu$ g protein). *HdlbpA2-His* proteins were visualized after Western blotting using an Anti-His peroxidase conjugate. The data underlying panels c and d is provided as S1\_data.xlsx.

<https://doi.org/10.1371/journal.pbio.3002177.g002>

rates were higher than those for strains  $\Delta$ tsdA and  $\Delta$ tsdA *lbpA2-His* (Fig 2D). This confirmed the crucial function of LbpA2 in the cytoplasmic sHdr-LbpA sulfur oxidation pathway and more importantly, showed that the absence of sLpl(AB) lipocate:protein ligase had the same effect as the complete absence of the LbpA2 protein, strongly suggesting that sLpl(AB) ligase is essential for the modification and thus the proper function of LbpA2. We found evidence in support of this hypothesis by enriching the LbpA2 acceptor proteins from the *H. denitrificans* strains  $\Delta$ tsdA *lbpA2-His* and  $\Delta$ tsdA *lbpA2-His*  $\Delta$ sLpl(AB) producing them with a His-tag and comparing their behavior in SDS and native PAGE. While the *H. denitrificans* strain lacking sLpl(AB) ligase produced only apo-LbpA2, the holo-protein was produced in the strain containing the complete assembly pathway as evident from the native gel mobility shift (Fig 2E and 2F).

### Lipocate assembly systems are unevenly distributed in bacteria and archaea

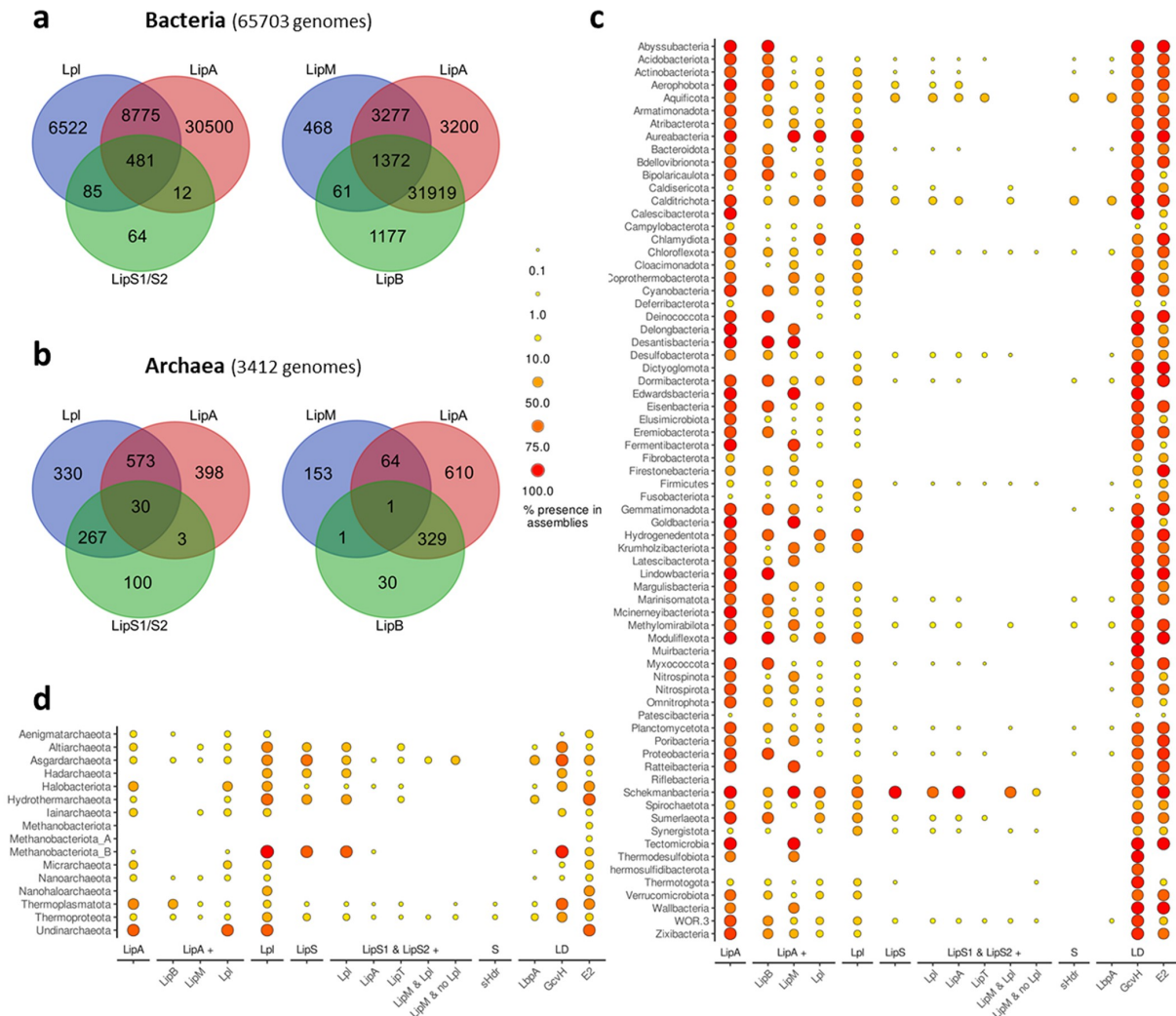
The experiments described above clearly established the functionality of a LipS1/S2-based lipocate assembly pathway in bacteria. We now asked how lipocate:protein ligases and octanoyltransferases (LipB, LipM) and the different lipocate synthases (LipA, LipS1/S2) are distributed among the prokaryotes and analyzed the genomes available in the Genome Taxonomy Database (GTDB, release R207). In GTDB, all genomes are sorted according to validly published taxonomies. In order to accurately identify and discriminate the enzymes involved in lipocate assembly pathways, a task severely hampered by the fact that central components are part of very large multi-protein families, we used HMS-S-S, a tool that specifically finds sulfur metabolism-related proteins [33] and extended it with publicly available HMMs for canonical lipocate synthesis enzymes as well as for well-studied lipocate-binding proteins, such as GcvH (S1 Table).

In both prokaryotic domains, scavenging of free lipocate or octanoate by lipocate:protein ligase is widespread (Fig 3). Lpl/LipA combinations were found in only 14% and 8% of the bacterial and archaeal genomes, respectively, while Lpl/LipS1/S2 were found in 0.9% and 17%, respectively, leaving the majority of prokaryotes unable to use free octanoate for lipocate assembly (Fig 3A and 3B). In bacteria, de novo lipocate synthesis by LipB/LipA or LipM/LipA is more common than lipocate scavenging (Fig 3). The co-occurrence of both octanoyltransferases, LipB and LipM, is very rare and there is usually no strict correlation between phylum and enzyme type. Contrary to previous assumptions [19], LipB is found in the Firmicutes and LipM is not restricted to a specific phylum. The Proteobacteria, together with a few other phyla, are the exception rather than the rule, containing only LipB and never LipM (Fig 3C). As mentioned in the Introduction, *B. subtilis* octanoyltransferase LipM is specific for GcvH and an amidotransferase (LipL) is required for modification of other lipoyl-binding domains/proteins. Surprisingly, not all bacteria, including Firmicutes, which carry out LipM/LipA-based lipocate biosynthesis, also possess LipL, although lipocate-binding domains/proteins are encoded in these genomes (S2 and S3 Tables). This suggests a much broader substrate specificity of the octanoyltransferase LipM than previously described or the existence of (an) unknown amidotransferase(s) that functionally replace(s) the amidotransferase LipL. Among the archaea, octanoyltransferase/lipocate synthase combinations LipB/LipA or LipM/LipA were also present (Fig 3), suggesting de novo lipocate synthesis from ACP-bound octanoate. This was unexpected, because ACP is generally absent in archaea [34–36] and cannot serve as direct octanoate donor. Coenzymes A or M are possible alternatives [35–37].

In both prokaryotic domains, the LipS1/S2-type of lipoyl synthase is less common than LipA lipocate synthase but remarkably widespread (Fig 3). Co-occurrence of lipocate synthase LipA is a common feature for *lipS1/S2*-containing bacteria but rare in LipS1/S2-encoding archaea. Our analyses do not confirm the proposal that LipS1/S2 are specific to thermophilic archaea [27]. The Asgardarchaeota, in which LipS1/S2 is more abundant than LipA, are a counter-example, as all sequences, including that of the only cultured representative *Candidatus Prometheoarchaeum syntrophicum*, are from temperate to cold environments [38,39]. Halobacteriota and Altiaarchaeota genomes with *lipS1/S2* also stem from temperate/cold habitat samples [40,41]. In addition, LipS1/S2 is present in many mesophilic bacteria, i.e., *H. denitrificans*, *Thioalkalivibrio* sp., or *Sporomusa* sp.

### LipS1/S2-type lipocate synthases and their cooperation partners

Regarding the novel lipocate synthesis pathway, our analyses confirm that *lipS1/S2* genes are often associated with genes for lipocate:protein ligases (Figs 1C and 3) [2], usually with a Lpl (AB) domain structure (S5 Table). A *lipS2-lplA-lplB-lipS1* arrangement seems typical for



**Fig 3. Taxonomic distribution of the lipopeate synthesis systems, lipopeate scavenging, and lipopeate requiring proteins.** Venn diagrams show the abundance and overlap of lipopeate:protein ligases (Lpl), octanoyl transferase (LipB, LipM), and lipopeate synthases (LipA, LipS1/S2) in the bacteria (a) and the archaea (b). Panels c and d visualize the taxonomic distribution of these enzymes, the sulfur-oxidizing sHdr system (S) and lipopeate-binding domains (LD). For each bacterial (c) and archaeal phylum (d), the percentage of genomes possessing these proteins is indicated by dots of different sizes and colors. Note that the proportion was normalized to the size of the phylum and does not show absolute counts or overall phylum size. The data underlying parts a and b are provided in S2 and S3 Tables, respectively. S4 Table supplies the data underlying parts c and d.

<https://doi.org/10.1371/journal.pbio.3002177.g003>

Archaea but is also found with some rearrangement in the bacterial phyla Chloroflexota, Aerophobota, and Synergistota. Occasionally, direct linkage of genes for canonical lipopeate synthesis with *lipS1/S2* is observed, e.g., in several *Sporomusa* species (phylum Firmicutes). In some *lipS1/S2*-containing bacteria (e.g., members of the Schekmanbacteria, Synergistota, and Thermotogota) and archaea (members of the Asgardarchaeota, Thermoproteota, and Thermoplasmatota), a gene for lipopeate:protein ligase is absent (Fig 3 and S2 Table). Instead, *lipS1/S2* co-occur with a gene for octanoyltransferase LipM. In the bacterial cases, this implies that *lipS1/S2* insert sulfur into target proteins that have been octanoylated by a transferase reaction. For

the archaea, as discussed above, the question of the substrate for the LipM homologs is unresolved.

The *lipT* encoded FAD-containing oxidoreductase is a likely candidate to provide electrons, probably derived from NAD(P)H, for the reductive sulfur insertion catalyzed by LipS1/S2 (Fig 1). Indeed, *lipT* occurs almost exclusively in bacteria containing *lipS1/S2* (91% of the cases), often in a *lipS1-lpl(AB)-lipS2-lipT* arrangement (Fig 1C and S5 Table). The picture is different for archaea, where only 22.3% of the *lipT*-containing genomes also contain *lipS1* and *lipS2* (S4 Table). Approximately 53% and 17% of the LipS1/S2-containing bacteria and archaea, respectively, also encode LipT (S2, S3 and S4 Tables).

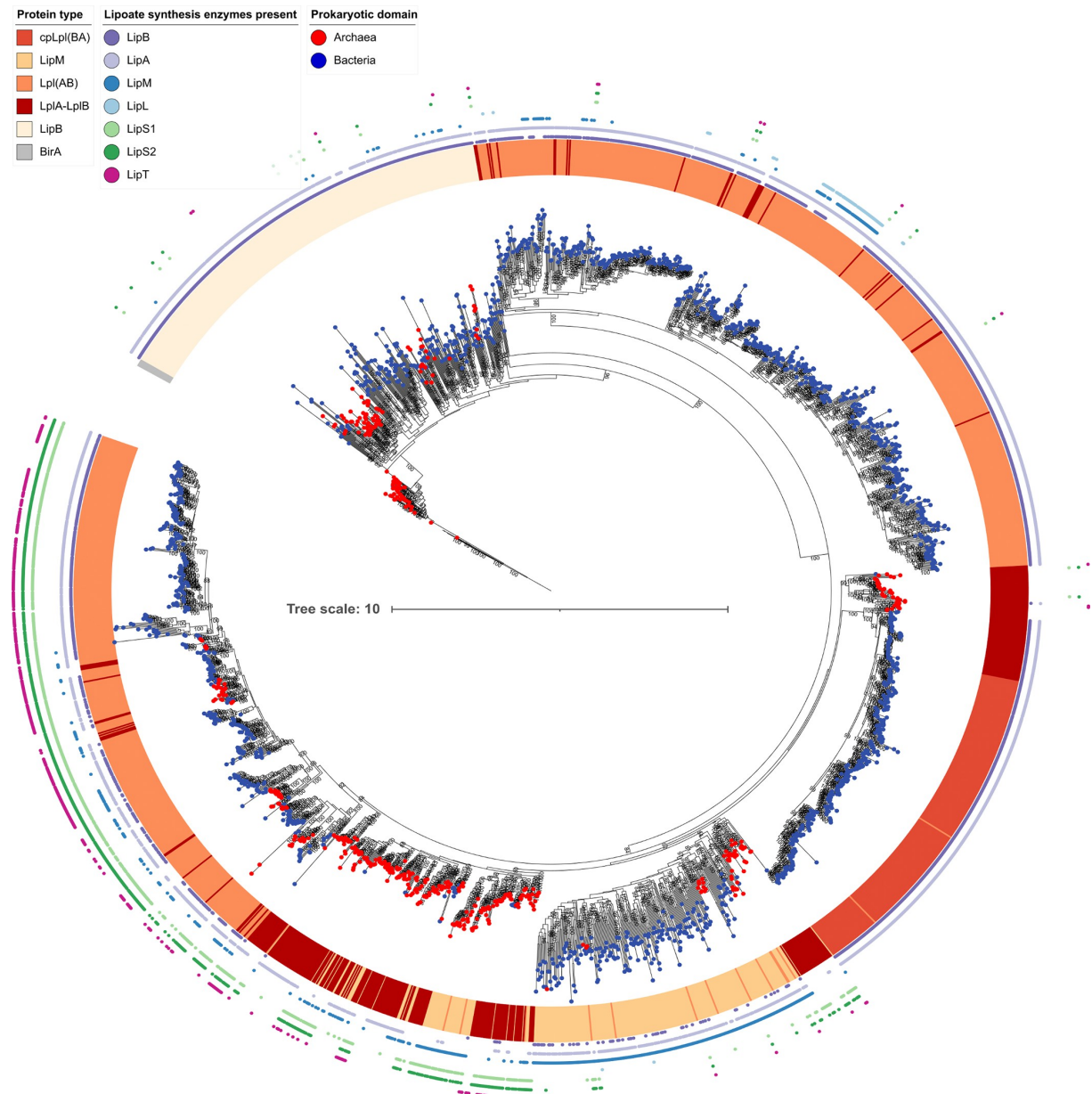
The *lipS1/S2* genes were first detected in bacterial and archaeal sulfur oxidizers that use the sHdr pathway for sulfur oxidation [2]. LbpA proteins are essential components of this pathway [2] and are encoded in *shdr*-containing genomes with very few exceptions, probably due to incompleteness of the respective assemblies (Fig 3 and S2 and S3 Tables), raising the question of whether the assembly of LbpA proteins is strictly dependent on LipS1/S2. While the majority of sHdr-containing prokaryotes are indeed equipped with LipS1/S2 (74.3% and 88.6% for bacteria and archaea, respectively; Fig 1C and S2, S3 and S4 Tables), the reverse is not true, i.e., LipS1/S2 are not restricted to sulfur oxidizers (42% and 8% of LipS1/S2-containing bacteria and archaea, respectively, have sHdr; S2, S3 and S4 Tables).

### Evolution of lipocate:protein ligases and octanoyltransferases

All lipocate:protein ligases and octanoyltransferases belong to the cofactor transferase family PF03099. Calculating rooted phylogenetic trees for these proteins was expected to bring new insights into the origin and evolution of these enzymes. If they initially originated in archaea, the tree should be a priori rooted in the archaeal domain and similarly for bacteria. The structurally related biotin ligase BirA, which is also a member of the cofactor transferase family, was chosen as a suitable outgroup to root the tree. The tree for the complete lipocate:protein ligase/octanoyltransferase dataset contains 3 clearly delineated clades with high bootstrap support (Fig 4). The first clade contains the bacterial and archaeal LipB octanoyltransferases and resides between the BirA root and all other analyzed proteins. The second clade harbors a group of lipocate:protein ligases derived exclusively from bacteria with LipA but usually without LipS1/S2 lipocate synthase. Broad substrate range *E. coli* Lpl(AB) as well as the narrow substrate range ligases LplJ from *B. subtilis*, and Mhp-LplJ and Mhp-Lpl from *M. hypopneumoniae* reside in this clade.

For the proteins of clade 3, an archaeal origin is inferred, since bipartite LplABs ligases from the archaeal phyla Thermoproteota and Thermoplasmatota including the characterized *Thermoplasma acidophilum* enzyme [10] are the deepest branching sequences. Three well-supported subgroups (bootstrap  $\geq 92$ ) branch off from these, each again with deep-branching archaeal proteins. In the first subgroup, bipartite LplABs from Nano- and Haloarchaeota form the deepest branches, which are immediately adjacent to LplABs from Burkholderiales (class Gammaproteobacteria according to GTDB). The remaining sequences in the subgroup are circularly permuted Lpl(BA) proteins, nearly exclusively stemming from Actinomycetota and including the characterized *S. coelicolor* Lpl(BA) [12]. This topology suggests an evolutionary history with lateral transfer of LplAB from Archaea to Gammaproteobacteria followed or accompanied by rearrangement of the gene order and final fusion of the genes upon transfer to the phylum Actinomycetota, where the gene was then vertically transmitted. The second subgroup consists of archaeal and bacterial bipartite LplAB ligases and LipM-type octanoyltransferases and provides insights into the origin of LipM: (1) Archaeal LplAB ligases, mostly encoded near *lipS1/lipS2* and including the characterized *Thermococcus kodakarensis* protein





**Fig 4. Rooted phylogenetic tree for the complete lipocate:protein ligase/octanoyltransferase dataset.** The tree was rooted with the structurally related biotin ligase BirA as an outgroup. Red or blue dots placed on each leaf identify the source organisms as archaea or bacteria, respectively. The ligase/transferase type is color-coded in the next circle. In the outermost rings, the presence of other lipocate synthesis enzymes occurring in the same genome is labeled. The data underlying this figure is provided in Supplementary S3 Data.

<https://doi.org/10.1371/journal.pbio.3002177.g004>

[30], are the most deeply branching sequences and appear to be the ancestors of a large number of bacterial LipMs, which thus arose from a single interdomain horizontal gene transfer event. A scenario is supported in which archaeal LplAB lost its accessory peptide LplB, developed into LipM and simultaneously or later moved into a bacterial host. (2) In the remaining

part of the subgroup, many archaeal and some bacterial LplABs ligases are mixed with many archaeal and some bacterial LipM octanoyltransferases, indicating that the described horizontal gene transfer, loss of LplB and transformation of the remaining catalytic domain LplA into an octanoyltransferase was not a singular event but happened multiple times. The third subgroup of clade 3 is made up of even further archaeal and bacterial LplABs and Lpl(AB) lipocate protein ligases, the vast majority of which originate from organisms containing LipS1/S2. The genetically and biochemically characterized sLpl(AB) ligases from proteobacterial sulfur oxidizers fall into this group.

We challenged the idea that clade 3 has an origin within the archaea by calculating 3 separate trees for this group. The trees were rooted by the most closely related, similarly sized and biochemically characterized bacterial lipocate:protein ligases from clade 2 and confidence levels were increased by not including LipM octanoyltransferases and/or circularly permuted cpLplBA lipocate:protein ligases (Figs 5, S4 and S5). The trees have a high to very high bootstrap support especially for the higher order splits and all 3 indeed show a root in the archaeal domain with Thermoproteota and Thermoplasmatota proteins as the deepest branching sequences. Moreover, it is clear that bacterial Lpl(AB)s, which co-occur in the same organism with LipS1/S2, originate from an archaeal ancestor. Several horizontal gene transfer events are also evident. The 2 earliest were transfer of LplAB ligases from Hadarchaeota to Synergistota and from Altiarchaeota to Chloroflexota. On the other hand, several transfers from bacteria back to archaea can be delineated, e.g., into members of the Thorarchaeota, Baldrarchaeota, Jordarchaeota, Sifarchaeota, Thermoplasmatota, and Thermoproteota.

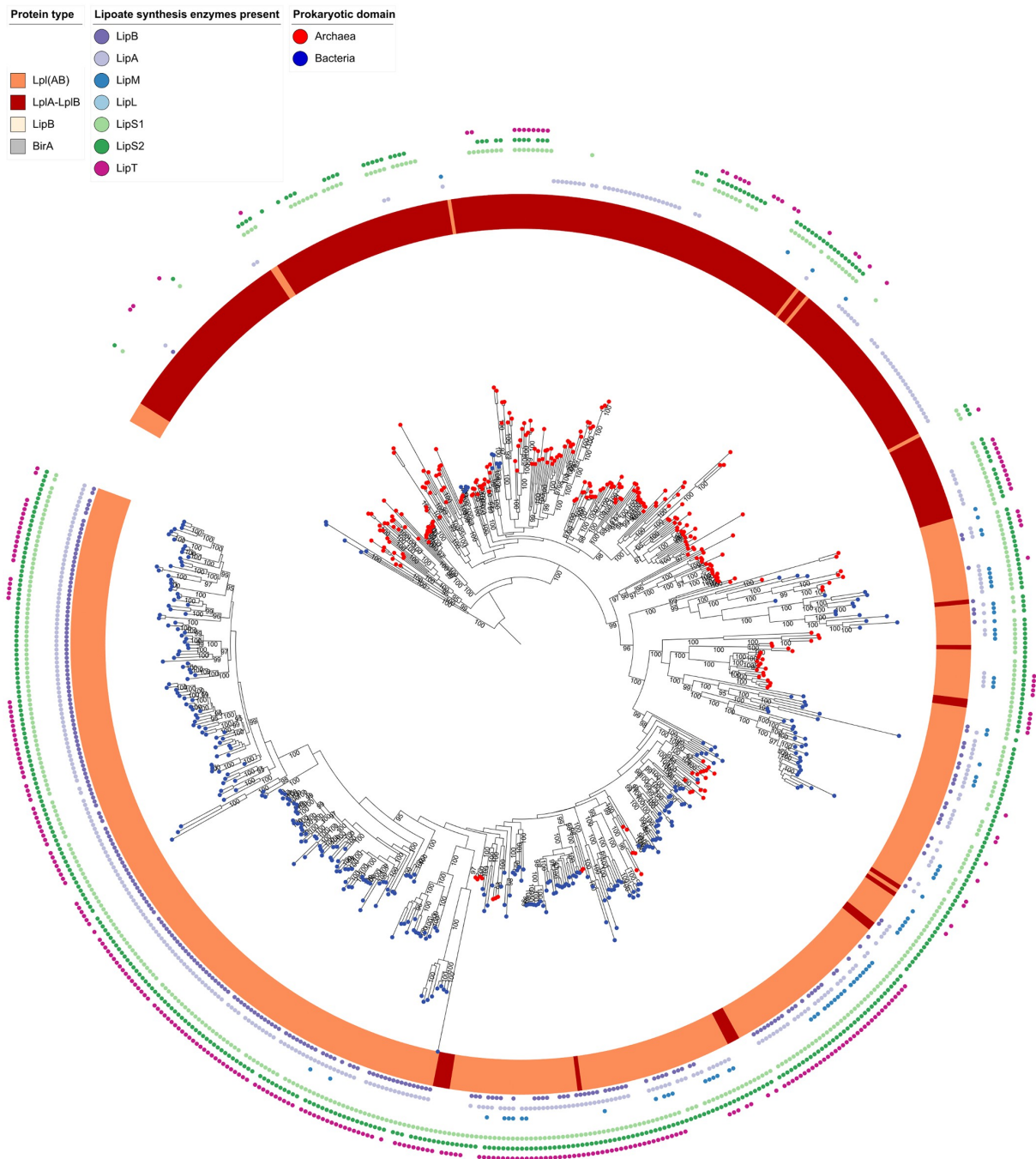
### Phylogenetic analysis of the radical SAM proteins LipS1 and LipS2

The tree obtained for the concatenated LipS1/S2 proteins neither shows a 2 domain split nor a long branching separation, as would be expected for an ancient protein that already existed in a universal common ancestor [42]. This indicates an origin of LipS1/S2 either in the bacterial or in the archaeal domain, followed by horizontal gene transfer into the other domain [43,44]. To obtain further insights, separate trees were constructed for the bacterial and the archaeal proteins (Fig 6B and 6C). If LipS1/S2 lipocate synthase originated in either the archaea or the bacteria and was inherited vertically, the domain-specific tree should essentially follow the taxonomy of that domain. However, neither the tree for archaeal LipS1/S2 nor that for the bacterial equivalents fit this concept. With the exception of Proteobacteria and Aquificota, monophyletic clusters were not recovered within the bacteria (Fig 6B) and among the archaea only the LipS1/S2 lipoyl synthases from the Hadarchaeota, Methanobacteriota, and Halobacteriota appear monophyletic (Fig 6C). The trees indicate that LipS1/S2 has undergone multiple gene transfers between and within the 2 prokaryotic domains.

Finally, some insight into the origin of LipS1/S2 lipocate synthase was gained by rooting separate LipS1 and LipS2 trees by the radical SAM enzyme biotin synthase BioB, which belongs to the same megacluster of similarity and, just like each of the LipS proteins, catalyzes the insertion of a single sulfur atom [29,45]. For LipS2, the resulting tree topology was rather well supported and placement of BioB in the split between 2 archaeal clades was stable (S6 Fig). This points at an origin of LipS2 in the archaea. The final topology of the rooted LipS1 tree is consistent with an archaeal origin but bootstrap support is not high enough to draw firm conclusions (S6 Fig).

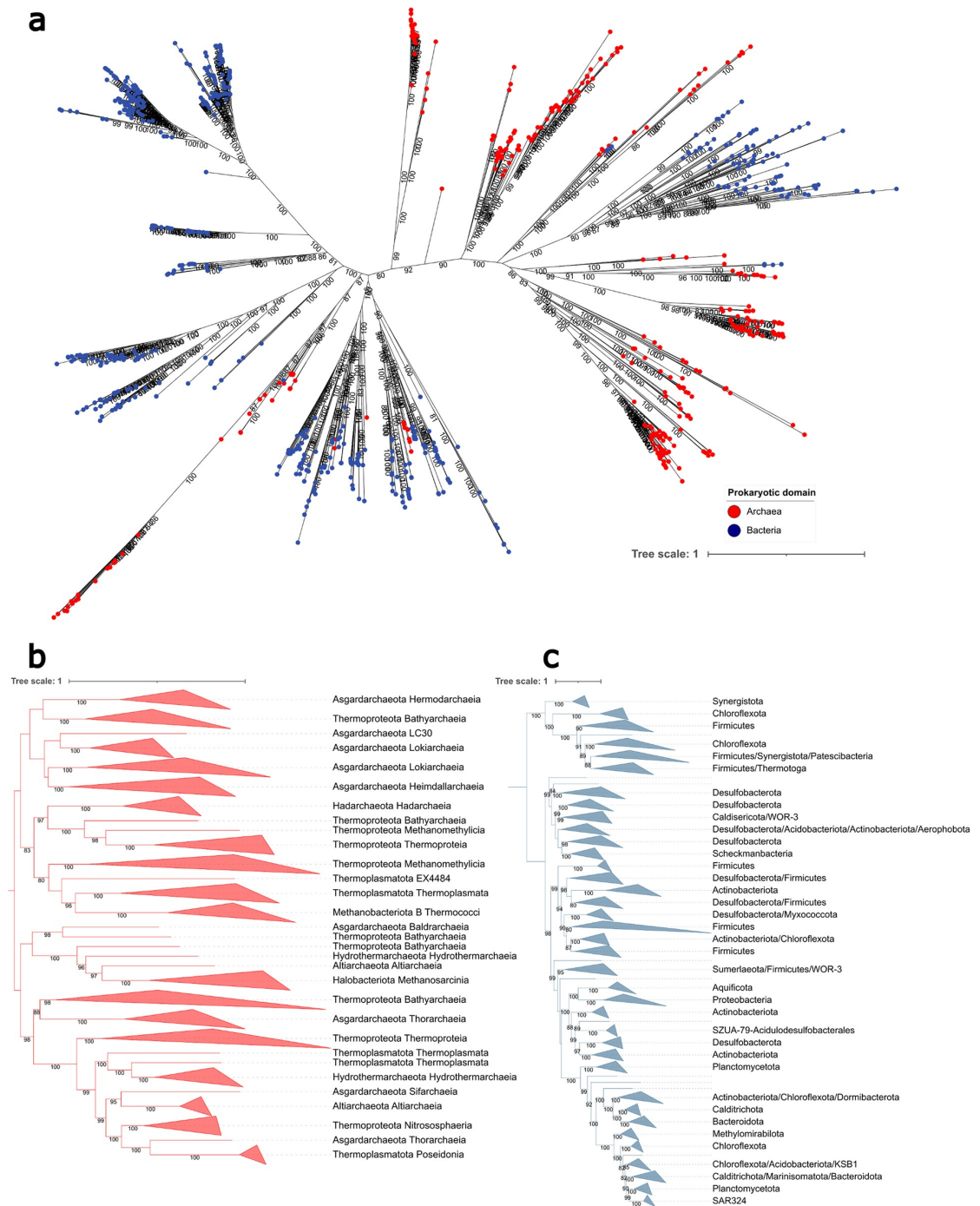
### Discussion

Our work proves the existence of a LipS1/S2-based lipoylation pathway not only in archaea but also in bacteria. In both prokaryotic domains, the pathway starts with the incorporation of



**Fig 5. Phylogeny for clade 3 lipote:protein ligases without LipM and cpLpl(BA).** LipMs do not have LplB domains and their sequences are consequently shorter. If a sequence is incomplete, parts of the information used to calculate the phylogenetic tree are missing. This can lead to erroneous estimates of the relationships between sequences and can bias the result and weakens statistical significance of the calculation. In addition, Lpl(BA) clearly shows an individual evolution and may also cause weakening of statistical support. The data underlying this figure is provided in [S3 Data](#).

<https://doi.org/10.1371/journal.pbio.3002177.g005>



**Fig 6. Phylogenetic trees for LipS1/S2.** (a) To investigate the evolution of LipS1 and LipS2, their sequences were concatenated, as both units are usually found in synteny, are catalytically active together and should therefore be under the same evolutionary pressure [29]. Incomplete sequences and concatenated sequences from genomes lacking either LipS1 or LipS2 were removed from the analysis. The lower panels show schematic representation of phylogenetic trees generated using only archaeal sequences (red, b) or bacterial sequences (blue, c). Bacterial clades represented by single sequences were left out to increase readability. The data underlying this figure is provided in [S3 Data](#).

<https://doi.org/10.1371/journal.pbio.3002177.g006>

a free octanoic acid and then LipS1/S2 insert 2 sulfur atoms. The lipoate:protein ligase that catalyzes the first step originates from archaea and thus has a completely different evolutionary history than the previously known Lpl(AB) enzymes in clade 2, which are exclusively found in bacteria. For LipS1/S2, which catalyze the second step, an archaeal origin is also likely. The majority of bacteria that use reduced sulfur compounds as electron donors by the sulfur-oxidizing heterodisulfide reductase-like (sHdr) complex in conjunction with LbpA proteins, are equipped with the originally archaeal lipoate synthesis machinery. Since heterodisulfide reductases are enzymes typical of methanogenic archaea, co-transfer of the genes seems a likely possibility.

Our large-scale integrated phylogenomic analyses open new perspectives on the general evolution and the complexity of lipoate synthesis. Not only do they reveal a much wider distribution of lipoate assembly systems than expected, in particular, the novel sLpl(AB)–LipS1/S2 pathway, and indicate a highly modular nature of the enzymes involved, with unforeseen combinations, but they also provide a new framework for the origin and subsequent evolution of lipoate assembly systems in the 2 prokaryotic domains. Our results are relevant to the entire diversity of life, since the machineries for lipoate synthesis in the mitochondria and plastids of eukaryotes are all of prokaryotic origin and may bear more complexity than currently assumed [1,46,47]. Contrary to the intuitive assumption, mitochondrial LipAs are more closely related to the LipA lipoate synthases of archaea than to those of  $\alpha$ -proteobacteria [47]. In addition, plastidial LipAs are not most closely related to those of cyanobacteria, as would be expected if they were derived from the primary endosymbiont that led to the first plastid. Instead, they form a sister group to the mitochondrial LipAs and probably arose by a gene duplication [47]. Furthermore, protozoa with mitochondria and plastids require 2 different ligases for lipoate metabolism in the respective organelles [1].

The tripartite topology of the phylogenetic tree for the octanoyltransferases and lipoate:protein ligases has provided particularly important insights. LipB octanoyltransferases (clade 1) are close to the root of the tree and thus represent the ancestors of all the other enzymes. Clade 1 gave rise to 2 other groups, one of which evolved in bacteria (clade 2) and the other is probably of archaeal origin (clade 3). In clade 2, the acquisition of an additional gene encoding LplB, and in many cases its fusion with the original LipB octanoyltransferase unit, allowed the conversion of an octanoyltransferase into an ATP-dependent lipoate:protein ligase. The most deeply branching proteins in clade 3 also have additional LplB, suggesting that the acquisition of the corresponding gene and functional transformation into a ligase was again the starting point for further evolution. Our conclusions are consistent with previous suggestions, based on structural data, that the lipoate:protein ligase LplA initially lacked an LplB unit, which was first acquired as a separate unit before fusion occurred [48].

We infer that bipartite LplAB enzymes represent the most original lipoate:protein ligase in archaea. Free lipoate may therefore have been the primary source of protein lipoylation in ancient archaea. There was no lipoyl synthase in these organisms, as we still find today, for example, in *Thermoplasma acidophilum*. Over time, lipoyl synthases appeared in archaea, either LipA or LipS1/S2 and very rarely both, enabling the use of octanoate. Gene clusters then formed, as we find in almost all recent archaea. In bacteria, clade 2 Lpl(AB) ligases appear to co-operate with LipA and in those bacteria that encode sLpl(AB) ligase, the genes for LipS1/S2 are almost always also present. Genetic linkage of the genes in operons is the usual case. These observations, together with the highly probable archaeal origin for sLpl(AB) ligase and LipS1/S2, suggest that the genes were transferred together from archaea to bacteria, possibly even together with genes for their specific lipoylation substrates. Such a scenario also fully explains why the canonical Lpl(AB)/LipA combination and the sLpl(AB)/LipS1/S2 system can coexist in the same organism and specifically lipoylate only their cognate substrates. Further

experimental work on bacterial and also archaeal lipolate:protein ligases and their potential substrates will allow testing these functional hypotheses.

Lipolate-requiring reductive glycine synthesis was placed at the basis of the tree of life, as part of the phenotype of the last common ancestor [19]. Accordingly, the last common ancestor has been proposed to contain radical SAM-dependent enzymes and LipA lipoyl synthase is considered among the most ancient enzymes in this class [21]. In contrast, genes encoding LipS1/S2 are absent in most phyla close to the root of the archaeal domain [49] (Iainarchaeota, Nanoarchaeota, Nanoarchaeota, Aenigmataarchaeota, and Micrarchaeota) and likely emerged later than LipA.

In conclusion, our results pave the way for further work on lipolate biochemistry as a key process for ancient and modern life.

## Methods

### Bacterial strains, plasmids, primers, and growth conditions

S6 Table lists the bacterial strains and plasmids that were used for this study. *E. coli* strains were grown on complex lysogeny broth (LB) medium [50] or on glycerol-containing M63 minimal medium [51] under aerobic conditions at 37°C unless otherwise indicated. *E. coli* 10B was used for molecular cloning. *E. coli* BL21 (DE3) was used for recombinant protein production. *H. denitrificans* strains were cultivated in minimal media kept at pH 7.2 with 100 mM 3-(*N*-Morpholino)propanesulfonic acid (MOPS) buffer as described before [9,31]. Media contained 24.4 mM methanol. Thiosulfate was added as needed. Antibiotics for *E. coli* and *H. denitrificans* were used at the following concentrations (in  $\mu\text{g ml}^{-1}$ ): ampicillin, 100; kanamycin, 50; streptomycin, 200; chloramphenicol, 25.

### Recombinant DNA techniques

Standard techniques for DNA manipulation and cloning were used unless otherwise indicated [52]. Restriction enzymes, T4 ligase and Q5 polymerase were obtained from New England Biolabs (Ipswich, United Kingdom) and used according to the manufacturer's instructions. Oligonucleotides for cloning were obtained from Eurofins MWG (Ebersberg, Germany). Plasmid DNA from *E. coli* was purified using the GenJet Plasmid Miniprep kit (Thermo Scientific, Waltham, United States of America). Chromosomal DNA from *H. denitrificans* strains was prepared using the Simplex easy DNA kit (GEN-IAL GmbH, Troisdorf, Germany).

### Construction of helper plasmids with lipolate assembly genes from *Thioalkalivibrio* sp. K90mix

For the construction of pACYC-*Tklpm* (lipolate protein maturation), the potential lipolate assembly genes *lipS1-slp(AB)-lipT-S2-Y* (TK90\_0641–0644, Fig 1C) were amplified with primers TK90\_0641–0644 fw and TK90\_0641–0644 rev using genomic *Thioalkalivibrio* sp. K90mix DNA as the template. The amplicon was digested with *DraI* and *BspHI* and cloned into the *EcoRV/BspHI* sites of pACYC184. This plasmid contained the *lipS1-slp(AB)-lipT-lipS2* sequence in the same orientation as the *tet* gene of the vector, thus allowing outreading transcription from the *tet* promoter.

### Production of LbpA proteins in *E. coli* BL21 (DE3) $\Delta$ *iscR* with and without helper plasmids

Recombinant LbpA proteins were produced in the *E. coli* BL21 (DE3)  $\Delta$ *iscR*, and 500-ml batches of LB medium containing 100 mM MOPS buffer, pH 7.4, 25 mM glucose, and 2 mM

iron ammonium citrate as well as ampicillin and kanamycin (and chloramphenicol for strains containing plasmid pACYC-*Tklpm*) were inoculated with 1% (v/v) *E. coli* precultures and cultivated in 1-L flasks at 37°C and 200 rpm until an OD<sub>600</sub> of 0.8 was reached. At this point, cysteine (0.5 mM), sodium fumarate (25 mM), and IPTG (0.1 mM) were added. Incubation continued for 14 to 16 h at 30°C and 180 rpm. Cells were harvested by centrifugation (11,000 × g, 12 min, 4°C) and lysed by sonication. After removal of insoluble cell material by centrifugation (16,100 × g, 30 min, 4°C), the LbpA proteins were purified by Strep-Tactin affinity according to the manufacturer's instructions (IBA Lifesciences, Göttingen, Germany). The proteins were then transferred to salt-free 100 mM Tris-HCl (pH 8.0) buffer and underwent further analysis.

### Construction of *H. denitrificans* mutant strains

The *tsdA* gene was deleted from *H. denitrificans*  $\Delta$ *lbpA* [2] by using plasmid pK18*mobsacB*  $\Delta$ *tsdATc*, transferring it by electroporation, and selecting double crossover recombinants as described previously [9]. For chromosomal integration of the gene encoding HdLbpA2 with a carboxy-terminal His-tag, the modified gene and upstream as well as downstream sequences were amplified by SOE PCR using primers Fwd5'  $\Delta$ *lbpA*, KI\_HdLbpA2-His-Up-rev, KI\_HdLbpA2-His-Down-fw, and Fwd3'  $\Delta$ *lbpA* (S6 Table). The final plasmid pK18*mobsacB-lbpA2-his-Tc* was transferred into *H. denitrificans*  $\Delta$ *tsdA*  $\Delta$ *lbpA*. For markerless *in frame* deletion of the *H. denitrificans* *slp(AB)* gene by splicing overlap extension [53], PCR fragments were constructed using the primers listed in S6 Table. It should be noted that the GTG start codon of the *slp(AB)* gene overlaps the stop codon of the preceding gene *lipS2* in the sequence GTGA. Similarly, the TGA stop codon of *slp(AB)* overlaps the start of gene *lipX* in the sequence ATGA. The ribosomal binding site for *lipX* translation must thus be embedded in the *slp(AB)* gene. To avoid affecting signals for *lipX* translation, the *in frame* deletion of *slp(AB)* was designed to leave the last 35 bp of the gene untouched. The 2,070 bp fragment, which implements deletion of a 1,029 bp fragment encoding amino acids 8 to 349 of sLp(AB), was digested with XbaI and cloned into the XbaI site of pK18*mobsacB-Tc* [31]. The final plasmid pK18*mobsacB* $\Delta$ *slp(AB)*-Tc was transferred into *H. denitrificans*  $\Delta$ *tsdA* *lbpA2-His*. The genotypes of the *H. denitrificans* mutant strains generated in this study were confirmed by PCR.

### Purification of His-tagged HdLbpA2

A total of 50 ml *H. denitrificans* precultures were grown in 100-ml Erlenmeyer flasks in methanol-containing medium with 100 mM MOPS (pH 7.4), chloramphenicol and streptomycin up to an OD<sub>600</sub> of 0.7. The main culture had a volume of 1 l in a 2-L Erlenmeyer flask and was inoculated to an OD<sub>600</sub> of 0.006 with the preculture. The main culture medium contained 2 mM thiosulfate. Cultures were incubated at 30°C and 150 rpm. Cells were harvested by centrifugation at 10,100 × g for 20 min at 4°C, when the cultures were actively oxidizing thiosulfate. Cells were stored at -20°C. Cells were resuspended in 10 ml phosphate buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl (pH 7.4)) per g cell material. A spatula tip of DNase RNAse, lysozyme, and protease inhibitor were added. Cells were broken by ultrasonication at 4°C for 10 min per g cells (Branson sonifier, 55% power) followed by centrifugation (16,100 × g, 4°C, 30 min) and ultracentrifugation (1 h, 4°C, 145,000 × g). Affinity chromatography on 1 ml HisTrap TALON crude columns (Cytiva, Marlborough, Massachusetts, USA) was performed with an Äkta-FPLC system according to the manufacturer's instructions. Enriched fractions were concentrated with Amicon Ultra-0.5, Ultracel 3k membrane, 3 kDa centrifugal filter units and subjected to electrophoreses on 12.5% Tricine-SDS or on 15% native gels stained with RotiBlue Quick (Carl Roth GmbH, Karlsruhe, Germany) for 1 h. Western blot analysis

was performed using nitrocellulose membranes (Amersham Protran 0.2  $\mu\text{m}$  NC, GE Healthcare, Solingen, Germany) and a Trans-Blot Turbo Transfer system (Bio-Rad Laboratories, Munich, Germany). Before western blotting, native gels were incubated for 20 min in Bjerrum buffer (48 mM Tris, 39 mM glycine, 0.04% SDS (pH 9.2)). Whatman filter paper needed during the blotting process was also soaked in Bjerrum buffer. For western blotting of SDS gels, gels and membrane were incubated for 20 min in Towbin buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol (pH 8.3)) before loading the blotting chamber. After overnight blocking at 4°C with 4 mM  $\text{KH}_2\text{PO}_4$ , 16 mM  $\text{Na}_2\text{HPO}_4$ , 115 mM NaCl, pH 7.4 + 0.05% (v/v) Tween 20, 4% milk powder, the membrane was washed 3 times for 5 min in PBS-Tween-buffer (4 mM  $\text{KH}_2\text{O}_4$ , 16 mM  $\text{Na}_2\text{HPO}_4$ , 115 mM NaCl + 0.1% Tween 20). Proteins were detected with Anti-His-HRP-conjugate (1:5,000) using the SignalFire™ ECL reagent system (Cell Signaling Technology, Cambridge, UK) and visualized with a ChemiDoc Imaging System (BioRad Laboratories, Munich, Germany).

### Characterization of phenotypes, quantification of sulfur compounds, and protein content

Growth experiments with *H. denitrificans* were run in Erlenmeyer flasks with media containing 24.4 mM methanol and varying concentrations of thiosulfate as necessary [31]. Thio-sulfate and sulfite concentrations and biomass content were determined by previously described methods [31,54]. All growth experiments were repeated 3 to 5 times. Representative experiments with 2 biological replicates for each strain are shown. All quantifications are based on at least 3 technical replicates.

### Mass spectrometry and Edman degradation

MALDI-TOF measurements and Edman degradation were performed at the Core Facility Protein Synthesis & BioAnalytics, Pharmaceutical Institute, University of Bonn.

### Dataset generation

Archaeal and bacterial genomes were downloaded from GTDB (release R207). In GTDB, all genomes are sorted according to validly published taxonomies, they are pre-validated and have high quality (completeness minus 5\*contamination must be higher than 50%). One representative of each of the current 65,703 species clusters was analyzed. It should be noted that GTDB is built on recently standardized bacterial and archaeal taxonomies derived by normalization of the evolutionary distance between taxonomic levels [49,55]. Among the bacteria, 148 phyla are currently distinguished. For the archaea, GTDB lists 16 phyla. Recent reclassification of the archaea was accompanied by merging parts of the Euryarchaeota with the TACK super-phylum into a single phylum [49]. Due to these massive but necessary taxonomic rearrangements, conclusions drawn on the distribution and occurrence of genes or groups of genes in higher taxonomic ranks need special care when comparing with previous work. Open reading frames were determined using Prodigal [56] and subsequently annotated for sulfur-related proteins via HMS-S-S [33]. Lipoic acid synthesis proteins and known lipoic acid-dependent enzymes were searched and annotated with HMMs from the TIGRFAMs and Pfams databases with trusted cutoffs. The HMM with the highest above-threshold bitscore was selected for each protein. Lipoate:protein ligases were checked for accessory domain LplB using the pfam PF10437 and annotated accordingly as LplAB. The publicly available HMMs for lipoate synthases, octanoyltransferases, and lipoate:protein ligases are generally sufficiently sensitive to detect all relevant sequences and also precise enough to make a good distinction between related proteins with different functions. However, there are limitations to these HMMs, as



they are based on the state of knowledge at the time they were generated and are not updated on a regular basis. Within the lipocate:protein ligases, it is for example not possible to distinguish between LplAB, Lpl(AB), sLpl(AB) from sulfur oxidizers, Lpl(BA), LplJ or Mhp-LplJ. All existing HMMs have been trained by Lpl(AB) and detect catalytic LplA domains of any type even in the absence of the accessory domain LplB. LplJ and LplA sequences are too similar to be distinguished. Thus, we annotated all detected lipocate:protein ligases as Lpl. The main distinguishing feature between the previously known ligases and sLpl(AB) and sLplAB enzymes is the genomic context and/or the type of lipocate synthase present in the organism. LipA and LipS1/S2 can be distinguished very reliably by HMMs.

### Phylogenetic inference

In order to investigate the evolutionary history of the lipocate:protein ligases and octanoyltransferases of the cofactor transferase family PF03099, sequences were separately retrieved from archaeal and bacterial assemblies. A dataset representing classical Lpl(AB) sequences was assembled from LipA-containing genomes not containing *lipS1* or *lipS2*. A second dataset for lipocate:protein ligases originated from gene clusters containing *lipS1/lipS2* and a single copy of the ligase gene. Thus, we ensured the highest possible probability that this dataset specifically included LipS1/S2-coupled ligases. LplA and LplB encoded in syntenic gene clusters were concatenated to match the Lpl(AB) domain structure. *Tc. kodakarensis* LplAB presented the only exception to this rule and was included in the analysis because the protein is biochemically characterized although the relevant genes do not form a gene cluster. For the circularly permuted LplBA ligases, the sequences of the 2 domains were separated and rejoined in the LplAB order and noted as cpLpl(AB). Octanoyltransferases LipB and LipM from bacterial genomes encoding LipA were clustered by similarity to limit the analysis to a reasonable number of sequences but at the same time maintain the diversity of the underlying dataset, while all corresponding archaeal sequences could be taken into account. A further dataset for LipM octanoyltransferase sequences originated from gene clusters containing *lipS1/lipS2* and a single copy of the octanoyltransferase gene, again in order to guarantee the highest possible probability that this dataset specifically includes LipS1/S2-coupled octanoyltransferases.

In order to exclude paralogous sequences, only proteins from assemblies encoding for 1 copy of each, LipS1 and LipS2, were considered and concatenated.

Sequences were aligned using MAFFT [57]. Ambiguously aligned regions and other sources of bias, such as highly variable characters, were removed using BMGE [58] (entropy threshold = 0.95, minimum length = 1, matrix = BLOSUM30), thereby trimming the alignments to regions suitable for phylogenetic inference. Phylogenetic inference by maximum likelihood is widely used in molecular systematics and involves substitution model parameters, branch lengths, and tree topology. In this work, maximum likelihood phylogenies were inferred using IQ-TREE v1.6.12 [59] implemented on the “bonna” high performance clusters of the University of Bonn. As a first step, the best-fitting model of sequence evolution that led to the available data was selected using ModelFinder, a method that combines amino acid substitution model used in other popular model-selection methods [60]. Branch support was then calculated by 3 different tree topology tests, SH-aLRT (2,000 replicates) [61], aBayes (2,000 replicates) [62], and ultrafast bootstrap (2,000 replicates) [63]. Finally, trees were displayed using iTol [64].

Where phylogeny of concatenated sequences was inferred, alignments were made individually and concatenated before trimming by BMGE. As an exception, LplA and LplB encoded in the same gene cluster were concatenated before the alignments as these domains are commonly found to be fused. Octanoyltransferases LipB and LipM lacking a C-terminal accessory domain were added to the LplAB alignments using MAFFT [57]—add-fragment function with—keeplength.

All trees are available in Newick format with associated iTol records in a compressed/zip file archive as Supporting information ([S3 Data](#)).

## Supporting information

**S1 Fig. Structural superposition of *E. coli* GcvH with *H. denitrificans* LbpA2.** Structural data for the *E. coli* protein (colored beige) are available in the RCSB Protein Data Bank (accession 3AB9). The structure of the *H. denitrificans* LbpA2 (colored light blue) was generated with AlphaFold2.

(PDF)

**S2 Fig. Comparison of bacterial and archaeal lipoate:protein ligases.** (a) Sequence alignment of Lpl(AB) from *E. coli* (Ecoli), sLpl(AB) from *H. denitrificans* (Hden), sLpl(AB) from *Thioalkalivibrio* sp. K90mix (TK90), concatenated LplBA from *Thermoplasma acidophilum* acidophilum (Thaci), and concatenated LplBA from *Tc. kodakarensis* (Tkoda). The LplB domain/polypeptide is highlighted by a box. The alignment was produced with T-Coffee. Alignment scores are graded from green (bad) to red (good). The right panel shows structural superposition of the *E. coli* Lpl(AB) with *H. denitrificans* sLplAB (b), *Thioalkalivibrio* sp. K90mix sLpl(AB) (c), *Thermoplasma acidophilum* LplAB (d), and *Thermococcus kodakarensis* LplAB (e). All structures were generated with AlphaFold2. The *E. coli* structure is shown in green in each case. The LplB domain is boxed in all panels.

(PDF)

**S3 Fig. Mass spectrometric analyses of 3 different recombinant LbpA proteins.** Tk90LbpA2 from *Thioalkalivibrio* sp. K90mix (a), TsLbpA2 (b), and TsLbpA1 (c) from *Thiorhodospira sibirica* produced in *E. coli* BL21 (DE3)  $\Delta$ iscR in the absence (green spectra) or presence (blue spectra) of helper plasmid pACYC-Tklpm carrying genes *lipS1-spl(AB)-lipT-lipS2* from *Thioalkalivibrio* sp. K90mix. The mass spectrometric data is provided as “[S2 Data](#).”

(PDF)

**S4 Fig. Phylogeny for clade 3 lipoate:protein ligases including LipM, but excluding cpLpl (BA).** Introducing LipM did not change the support of an archaeal origin of clade 3. The data underlying this figure can be found in [S3 Data](#).

(PDF)

**S5 Fig. Phylogenetic tree of for clade 3 lipoate:protein ligases including cpLpl(BA) but excluding LipM.** Introducing cpLpl(BA) did not change the support of an archaeal origin of the clade 3. The data underlying this figure can be found in [S3 Data](#).

(PDF)

**S6 Fig. Rooted individual phylogenetic trees for LipS2 (a) and LipS1 (b).** Sequences were derived from bacterial and archaeal genomes encoding exactly for 1 LipS1 and 1 LipS2. The trees were rooted by biotin synthase BioB as an outgroup. Archaeal sequences are marked in red, bacterial ones in blue. The data underlying this figure can be found in [S3 Data](#).

(PDF)

**S1 Table. Hidden Markov models for extension of HMS-S-S.**

(PDF)

**S2 Table. Presence/absence in bacteria of lipoate synthesis systems, lipoate scavenging, and lipoate requiring proteins.**

(XLSX)

**S3 Table. Presence/absence in archaea of lipote synthesis systems, lipote scavenging, and lipote requiring proteins.**

(XLSX)

**S4 Table. Numerical data underlying Fig 3C and 3D.**

(XLSX)

**S5 Table. List of genomic arrangements.**

(XLSX)

**S6 Table. Strains, plasmids, and primers.**

(PDF)

**S1 Data. Numerical values underlying Fig 2C and 2D.**

(XLSX)

**S2 Data. Mass spectra LbpA proteins.**

(RAR)

**S3 Data. Newick treefiles and iTol datasets.**

(RAR)

**S1 Raw Images. Raw gel and blot images underlying Fig 2A, 2E and 2F.**

(PDF)

## Acknowledgments

We gratefully acknowledge the support of the Core Facility “Protein Synthesis and Bioanalytics” of the University of Bonn for performing mass spectrometry and Edman degradation. We gratefully acknowledge the access to the Bonna HPC cluster hosted by the University of Bonn along with the support provided by its High Performance Computing & Analytics Lab.

## Author Contributions

**Conceptualization:** Tomohisa Sebastian Tanabe, Martina Grosser, Carolin Kumpel, Christiane Dahl.

**Data curation:** Tomohisa Sebastian Tanabe, Martina Grosser, Lea Hahn, Carolin Kumpel.

**Formal analysis:** Tomohisa Sebastian Tanabe, Martina Grosser, Carolin Kumpel, Christiane Dahl.

**Funding acquisition:** Christiane Dahl.

**Investigation:** Tomohisa Sebastian Tanabe, Martina Grosser, Lea Hahn, Carolin Kumpel, Hanna Hartenfels, Evelyn Vtulkin, Wanda Flegler.

**Project administration:** Christiane Dahl.

**Supervision:** Christiane Dahl.

**Validation:** Tomohisa Sebastian Tanabe, Martina Grosser, Carolin Kumpel, Christiane Dahl.

**Visualization:** Tomohisa Sebastian Tanabe.

**Writing – original draft:** Tomohisa Sebastian Tanabe, Christiane Dahl.

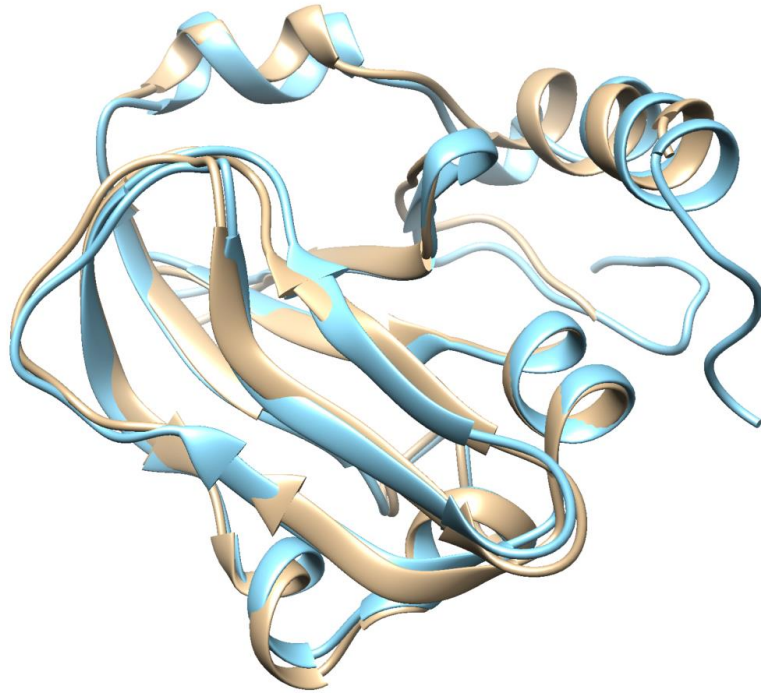
## References

1. Cronan JE. Assembly of Lipoic Acid on Its Cognate Enzymes: An Extraordinary and Essential Biosynthetic Pathway. *Microbiol Mol Biol Rev.* 2016; 80(2):429–450. <https://doi.org/10.1128/MMBR.00073-15> PMID: 27074917
2. Cao X, Koch T, Steffens L, Finkensieper J, Ziggann R, Cronan JE, et al. Lipocate-Binding Proteins and Specific Lipocate-Protein Ligases in Microbial Sulfur Oxidation Reveal an Atypical Role for an Old Cofactor. *elife.* 2018; 7:e37439. <https://doi.org/10.7554/eLife.37439> PMID: 30004385
3. Spalding MD, Prigge ST. Lipoic Acid Metabolism in Microbial Pathogens. *Microbiol Mol Biol Rev.* 2010; 74(2):200–228. <https://doi.org/10.1128/MMBR.00008-10> PMID: 20508247
4. Dunn MF. Vitamin Formation from Fatty Acid Precursors. In: Geiger O, editor. *Biogenesis of Fatty Acids, Lipids and Membranes. Handbook of Hydrocarbon and Lipid Microbiology.* Cham: Springer Nature Switzerland AG; 2019. p. 259–271.
5. Reed LJ, DeBusk BG, Gunsalus IC, Hornberger CS Jr. Crystalline A-Lipoic Acid: A Catalytic Agent Associated with Pyruvate Dehydrogenase. *Science.* 1951; 114(2952):93–94.
6. Perham RN. Swinging Arms and Swinging Domains in Multifunctional Enzymes: Catalytic Machines for Multistep Reactions. *Annu Rev Biochem.* 2000; 69:961–1004. <https://doi.org/10.1146/annurev.biochem.69.1.961> PMID: 10966480
7. Cussiol JR, Alegria TG, Szweida LI, Netto LE, Ohr (Organic Hydroperoxide Resistance Protein) Possesses a Previously Undescribed Activity, Lipoyl-Dependent Peroxidase. *J Biol Chem.* 2010; 285(29):21943–21950. <https://doi.org/10.1074/jbc.M110.117283> PMID: 20463026
8. Nyvltova E, Smutna T, Tachezy J, Hrdy I, OsmC and Incomplete Glycine Decarboxylase Complex Mediate Reductive Detoxification of Peroxides in Hydrogenosomes of *Trichomonas vaginalis*. *Mol Biochem Parasitol.* 2016; 206(1–2):29–38. <https://doi.org/10.1016/j.molbiopara.2016.01.006> PMID: 26794804
9. Koch T, Dahl C. A Novel Bacterial Sulfur Oxidation Pathway Provides a New Link between the Cycles of Organic and Inorganic Sulfur Compounds. *ISME J.* 2018; 12(10):2479–2491. <https://doi.org/10.1038/s41396-018-0209-7> PMID: 29930335
10. Christensen QH, Cronan JE. The *Thermoplasma acidophilum* LplA-LplB Complex Defines a New Class of Bipartite Lipocate-Protein Ligases. *J Biol Chem.* 2009; 284(32):21317–21326. <https://doi.org/10.1074/jbc.M109.015016> PMID: 19520844
11. Green DE, Morris TW, Green J, Cronan JE Jr., Guest JR. Purification and Properties of the Lipocate Protein Ligase of *Escherichia coli*. *Biochem J.* 1995;309(Pt 3):853–862. <https://doi.org/10.1042/bj3090853> PMID: 7639702
12. Cao X, Cronan JE. The *Streptomyces coelicolor* Lipocate-Protein Ligase Is a Circularly Permuted Version of the *Escherichia coli* Enzyme Composed of Discrete Interacting Domains. *J Biol Chem.* 2015; 290(11):7280–7290. <https://doi.org/10.1074/jbc.M114.626879> PMID: 25631049
13. Rasetto NB, Lavatelli A, Martin N, Mansilla MC. Unravelling the Lipoyl-Relay of Exogenous Lipocate Utilization in *Bacillus subtilis*. *Mol Microbiol.* 2019; 112(1):302–316. <https://doi.org/10.1111/mmi.14271> PMID: 31066113
14. Martin N, Christensen QH, Mansilla MC, Cronan JE, de Mendoza D. A Novel Two-Gene Requirement for the Octanoyltransfer Reaction of *Bacillus subtilis* Lipoic Acid Biosynthesis. *Mol Microbiol.* 2011; 80(2):335–349. <https://doi.org/10.1111/j.1365-2958.2011.07597.x> PMID: 21338420
15. Zhu K, Chen H, Jin J, Wang N, Ma G, Huang J, et al. Functional Identification and Structural Analysis of a New Lipocate Protein Ligase in *Mycoplasma hyopneumoniae*. *Front Cell Infect Microbiol.* 2020; 10:156. <https://doi.org/10.3389/fcimb.2020.00156> PMID: 32373550
16. Jin J, Chen H, Wang N, Zhu K, Liu H, Shi D, et al. A Novel Lipocate-Protein Ligase, Mhp-LplJ, Is Required for Lipoic Acid Metabolism in *Mycoplasma hyopneumoniae*. *Front Microbiol.* 2021; 11:631433. <https://doi.org/10.3389/fmicb.2020.631433> PMID: 33584596
17. Zhao X, Miller JR, Cronan JE. The Reaction of LipB, the Octanoyl- [Acyl Carrier Protein]:Protein N-Octanoyltransferase of Lipoic Acid Synthesis, Proceeds through an Acyl-Enzyme Intermediate. *Biochemistry.* 2005; 44(50):16737–46. <https://doi.org/10.1021/bi051865y> PMID: 16342964
18. Christensen QH, Cronan JE. Lipoic Acid Synthesis: A New Family of Octanoyltransferases Generally Annotated as Lipocate Protein Ligases. *Biochemistry.* 2010; 49(46):10024–10036. <https://doi.org/10.1021/bi101215f> PMID: 20882995
19. Braakman R, Smith E. Metabolic Evolution of a Deep-Branching Hyperthermophilic Chemoautotrophic Bacterium. *PLoS ONE.* 2014; 9(2):e87950. <https://doi.org/10.1371/journal.pone.0087950> PMID: 24516572

20. Christensen QH, Martin N, Mansilla MC, de Mendoza D, Cronan JE. A Novel Amidotransferase Required for Lipoic Acid Cofactor Assembly in *Bacillus subtilis*. *Mol Microbiol.* 2011; 80(2):350–363. <https://doi.org/10.1111/j.1365-2958.2011.07598.x> PMID: 21338421
21. Broderick JB, Duffus BR, Duschene KS, Shepard EM. Radical S-Adenosylmethionine Enzymes. *Chem Rev.* 2014; 114(8):4229–4317. <https://doi.org/10.1021/cr4004709> PMID: 24476342
22. Douglas P, Kriek M, Bryant P, Roach PL. Lipoyl Synthase Inserts Sulfur Atoms into an Octanoyl Substrate in a Stepwise Manner. *Angew Chem Int Ed Eng.* 2006; 45(31):5197–5199. <https://doi.org/10.1002/anie.200601910> PMID: 16835858
23. Lanz ND, Pandelia ME, Kakar ES, Lee KH, Krebs C, Booker SJ. Evidence for a Catalytically and Kinetically Competent Enzyme-Substrate Cross-Linked Intermediate in Catalysis by Lipoyl Synthase. *Biochemistry.* 2014; 53(28):4557–72. <https://doi.org/10.1021/bi500432r> PMID: 24901788
24. Cronan JE. Advances in Synthesis of Biotin and Assembly of Lipoic Acid. *Curr Opin Chem Biol.* 2018; 47:60–66. <https://doi.org/10.1016/j.cbpa.2018.08.004> PMID: 30236800
25. Zorzoli A, Grayczyk JP, Alonzo F 3rd. *Staphylococcus aureus* Tissue Infection During Sepsis Is Supported by Differential Use of Bacterial or Host-Derived Lipoic Acid. *PLoS Pathog.* 2016; 12(10):e1005933. <https://doi.org/10.1371/journal.ppat.1005933> PMID: 27701474
26. Christensen QH, Hagar JA, O'Riordan MX, Cronan JE. A Complex Lipoate Utilization Pathway in *Listeria monocytogenes* J Biol Chem. 2011; 286(36):31447–31456. <https://doi.org/10.1074/jbc.M111.273607> PMID: 21768091
27. Jin JQ, Hachisuka SI, Sato T, Fujiwara T, Atomi H. A Structurally Novel Lipoyl Synthase in the Hyperthermophilic Archaeon *Thermococcus kodakarensis*. *Appl Environ Microbiol.* 2020; 86(23). <https://doi.org/10.1128/AEM.01359-20> PMID: 32978128
28. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature.* 2021; 596(7873):583–589. <https://doi.org/10.1038/s41586-021-03819-2> PMID: 34265844
29. Neti SS, Sil D, Warui DM, Esakova OA, Solinski AE, Serrano DA, et al. Characterization of LipS1 and LipS2 from *Thermococcus kodakarensis*: Proteins Annotated as Biotin Synthases, Which Together Catalyze Formation of the Lipoyl Cofactor. *ACS Bio Med Chem Au.* 2022; 2(5):509–520. <https://doi.org/10.1021/acsbioimedchemau.2c00018> PMID: 36281299
30. Jin JQ, Sato T, Hachisuka SI, Atomi H. A Lipoate-Protein Ligase Is Required for *De Novo* Lipoyl-Protein Biosynthesis in the Hyperthermophilic Archaeon *Thermococcus kodakarensis*. *Appl Environ Microbiol.* 2022; 88(13):e0064422. <https://doi.org/10.1128/aem.00644-22> PMID: 35736229
31. Li J, Koch J, Flegler W, Garcia Ruiz L, Hager N, Ballas A, et al. A Metabolic Puzzle: Consumption of C<sub>1</sub> Compounds and Thiosulfate in *Hyphomicrobium denitrificans* X<sup>1</sup>. *Biochim Biophys Acta Bioenerg.* 2022; 1864:148932. <https://doi.org/10.1016/j.bbabi.2022.148932> PMID: 36367491
32. Akhtar MK, Jones PR. Deletion of *IscR* Stimulates Recombinant Clostridial Fe-Fe Hydrogenase Activity and H<sub>2</sub>-Accumulation in *Escherichia coli* BL21 (DE3). *Appl Microbiol Biotechnol.* 2008; 78(5):853–862. <https://doi.org/10.1007/s00253-008-1377-6> PMID: 18320190
33. Tanabe TS, Dahl C. HMS-S-S: A Tool for the Identification of Sulphur Metabolism-Related Genes and Analysis of Operon Structures in Genome and Metagenome Assemblies. *Mol Ecol Resour.* 2022; 22(7):2758–2774. <https://doi.org/10.1111/1755-0998.13642> PMID: 35579058
34. Lombard J, Lopez-Garcia P, Moreira D. An ACP-Independent Fatty Acid Synthesis Pathway in Archaea: Implications for the Origin of Phospholipids. *Mol Biol Evol.* 2012; 29(11):3261–3265. <https://doi.org/10.1093/molbev/mss160> PMID: 22718911
35. Dibrova DV, Galperin MY, Mulikidjanian AY. Phylogenomic Reconstruction of Archaeal Fatty Acid Metabolism. *Environ Microbiol.* 2014; 16(4):907–918. <https://doi.org/10.1111/1462-2920.12359> PMID: 24818264
36. Wang K, Sybers D, Maklad HR, Lemmens L, Lewylye C, Zhou X, et al. A TetR-Family Transcription Factor Regulates Fatty Acid Metabolism in the Archaeal Model Organism *Sulfolobus acidocaldarius*. *Nat Commun.* 2019; 10(1):1542. <https://doi.org/10.1038/s41467-019-09479-1> PMID: 30948713
37. Seitz KW, Dombrowski N, Eme L, Spang A, Lombard J, Sieber JR, et al. Asgard Archaea Capable of Anaerobic Hydrocarbon Cycling. *Nat Commun.* 2019; 10(1):1822. <https://doi.org/10.1038/s41467-019-09364-x> PMID: 31015394
38. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Backstrom D, Juzokaite L, Vancaester E, et al. Asgard Archaea Illuminate the Origin of Eukaryotic Cellular Complexity. *Nature.* 2017; 541(7637):353–358. <https://doi.org/10.1038/nature21031> PMID: 28077874
39. Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, et al. Isolation of an Archaeon at the Prokaryote-Eukaryote Interface. *Nature.* 2020; 577(7791):519–525. <https://doi.org/10.1038/s41586-019-1916-6> PMID: 31942073

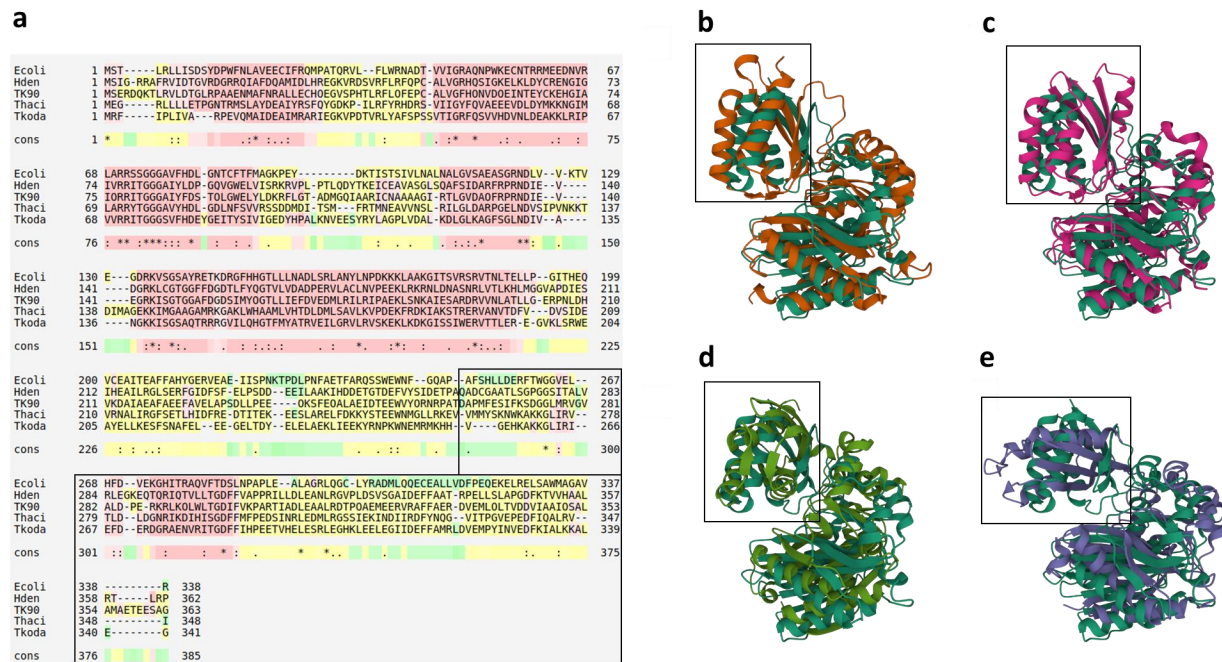
40. Probst AJ, Weinmaier T, Raymann K, Perras A, Emerson JB, Rattei T, et al. Biology of a Widespread Uncultivated Archaeon That Contributes to Carbon Fixation in the Subsurface. *Nat Commun.* 2014; 5:5497. <https://doi.org/10.1038/ncomms6497> PMID: 25425419
41. Kotelnikova S, Macario AJ, Pedersen K. *Methanobacterium subterraneum* sp. nov., a New Alkaliphilic, Eurythermic and Halotolerant Methanogen Isolated from Deep Granitic Groundwater. *Int J Syst Evol Microbiol.* 1998; 48(Pt 2):357–367. <https://doi.org/10.1099/00207713-48-2-357> PMID: 9731274
42. Catchpole RJ, Forterre P. The Evolution of Reverse Gyrase Suggests a Nonhyperthermophilic Last Universal Common Ancestor. *Mol Biol Evol.* 2019; 36(12):2737–2747. <https://doi.org/10.1093/molbev/msz180> PMID: 31504731
43. Berkemer SJ, McGlynn SE. A New Analysis of Archaea-Bacteria Domain Separation: Variable Phylogenetic Distance and the Tempo of Early Evolution. *Mol Biol Evol.* 2020; 37(8):2332–2340. <https://doi.org/10.1093/molbev/msaa089> PMID: 32316034
44. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, et al. The Physiology and Habitat of the Last Universal Common Ancestor. *Nat Microbiol.* 2016; 1(9):16116. <https://doi.org/10.1038/nmicrobiol.2016.116> PMID: 27562259
45. Oberg N, Precord TW, Mitchell DA, Gerlt JA. Radicalsam.Org: A Resource to Interpret Sequence-Function Space and Discover New Radical Sam Enzyme Chemistry. *ACS Bio Med Chem Au.* 2022; 2(1):22–35. <https://doi.org/10.1021/acsbiochemau.1c00048> PMID: 36119373
46. Gueguen V, Macherel D, Jaquinod M, Douce R, Bourguignon J. Fatty Acid and Lipoic Acid Biosynthesis in Higher Plant Mitochondria. *J Biol Chem.* 2000; 275(7):5016–5025. <https://doi.org/10.1074/jbc.275.7.5016> PMID: 10671542
47. Crawford MJ, Thomsen-Zieger N, Ray M, Schachtner J, Roos DS, Seeber F. *Toxoplasma gondii* Scavenges Host-Derived Lipoic Acid Despite Its *De Novo* Synthesis in the Apicoplast. *EMBO J.* 2006; 25(13):3214–3222. <https://doi.org/10.1038/sj.emboj.7601189> PMID: 16778769
48. Wood ZA, Weaver LH, Brown PH, Beckett D, Matthews BW. Co-Repressor Induced Order and Biotin Repressor Dimerization: A Case for Divergent Followed by Convergent Evolution. *J Mol Biol.* 2006; 357(2):509–523. <https://doi.org/10.1016/j.jmb.2005.12.066> PMID: 16438984
49. Rinke C, Chuvochina M, Mussig AJ, Chaumeil PA, Davin AA, Waite DW, et al. A Standardized Archaeal Taxonomy for the Genome Taxonomy Database. *Nat Microbiol.* 2021; 6(7):946–959. <https://doi.org/10.1038/s41564-021-00918-8> PMID: 34155373
50. Bertani G. Lysogeny at Mid-Twentieth Century: P1, P2, and Other Experimental Systems. *J Bacteriol.* 2004; 186(3):595–600. <https://doi.org/10.1128/JB.186.3.595-600.2004> PMID: 14729683
51. Elbing K, Brent R. Media Preparation and Bacteriological Tools. *Curr Protoc Mol Biol.* 2002; Chapter 1: Unit 1.1. <https://doi.org/10.1002/0471142727.mb0101s59> PMID: 18265292
52. Ausubel FA, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al. *Current Protocols in Molecular Biology.* New York: John Wiley & Sons; 1997.
53. Horton RM. PCR Mediated Recombination and Mutagenesis: Soeing Together Tailor-Made Genes. *Mol Biotechnol.* 1995; 3(2):93–99. <https://doi.org/10.1007/BF02789105> PMID: 7620981
54. Dahl C. Insertional Gene Inactivation in a Phototrophic Sulphur Bacterium: APS-Reductase-Deficient Mutants of *Chromatium vinosum*. *Microbiology.* 1996; 142:3363–3372. <https://doi.org/10.1099/13500872-142-12-3363> PMID: 9004500
55. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A Standardized Bacterial Taxonomy Based on Genome Phylogeny Substantially Revises the Tree of Life. *Nat Biotechnol.* 2018; 36(10):996–1004. <https://doi.org/10.1038/nbt.4229> PMID: 30148503
56. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinformatics.* 2010; 11:119. <https://doi.org/10.1186/1471-2105-11-119> PMID: 20211023
57. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol.* 2013; 30(4):772–780. <https://doi.org/10.1093/molbev/mst010> PMID: 23329690
58. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): A New Software for Selection of Phylogenetic Informative Regions from Multiple Sequence Alignments. *BMC Evol Biol.* 2010; 10:210. <https://doi.org/10.1186/1471-2148-10-210> PMID: 20626897
59. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol.* 2015; 32(1):268–274. <https://doi.org/10.1093/molbev/msu300> PMID: 25371430
60. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat Methods.* 2017; 14(6):587–589. <https://doi.org/10.1038/nmeth.4285> PMID: 28481363

61. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol.* 2010; 59(3):307–321. <https://doi.org/10.1093/sysbio/syq010> PMID: 20525638
62. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. Survey of Branch Support Methods Demonstrates Accuracy, Power, and Robustness of Fast Likelihood-Based Approximation Schemes. *Syst Biol.* 2011; 60(5):685–699. <https://doi.org/10.1093/sysbio/syr041> PMID: 21540409
63. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol.* 2018; 35(2):518–522. <https://doi.org/10.1093/molbev/msx281> PMID: 29077904
64. Letunic I, Bork P. Interactive Tree of Life (iTOL): An Online Tool for Phylogenetic Tree Display and Annotation. *Bioinformatics.* 2007; 23(1):127–128. <https://doi.org/10.1093/bioinformatics/btl529> PMID: 17050570

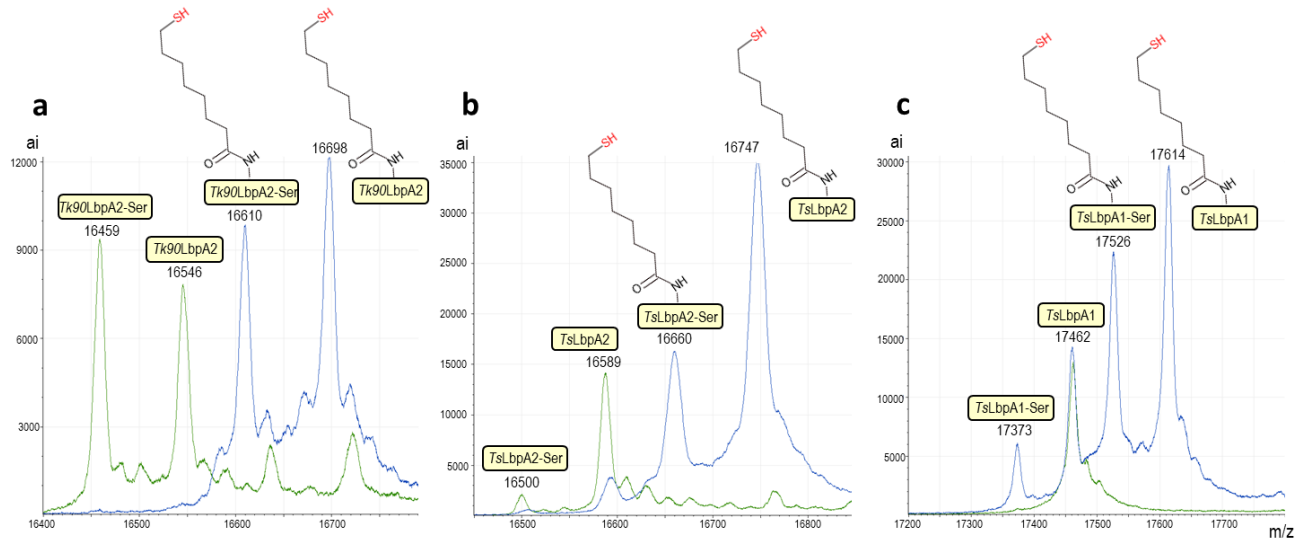


**Fig. S1. Structural superposition of *E. coli* GcvH with *H. denitrificans* LbpA2.** Structural data for the *E. coli* protein (colored beige) are available in the RCSB Protein Data Bank (accession 3AB9). The structure of the *H. denitrificans* LbpA2 (colored light blue) was generated with AlphaFold2.





**Fig S2. Comparison of bacterial and archaeal lipote:protein ligases.** (a) Sequence alignment of Lpl(AB) from *E. coli* (Ecoli), sLpl(AB) from *H. denitrificans* (Hden), sLpl(AB) from *Thioalkalivibrio* sp. K90mix (TK90), concatenated LplBA from *Tp. acidophilum* (Thac1), and concatenated LplBA from *Tc. kodakarensis* (Tkoda). The LplB domain/polypeptide is highlighted by a box. The alignment was produced with T-Coffee. Alignment scores are graded from green (bad) to red (good). The right panel shows structural superposition of the *E. coli* Lpl(AB) with *H. denitrificans* sLp(AB) (b), *Thioalkalivibrio* sp. K90mix sLp(AB) (c) *Thermoplasma acidophilum* Lp(AB) (d) and *Thermococcus kodakarensis* Lp(AB) (e). All structures were generated with AlphaFold2. The *E. coli* structure is shown in green in each case. The LplB domain is boxed in all panels.

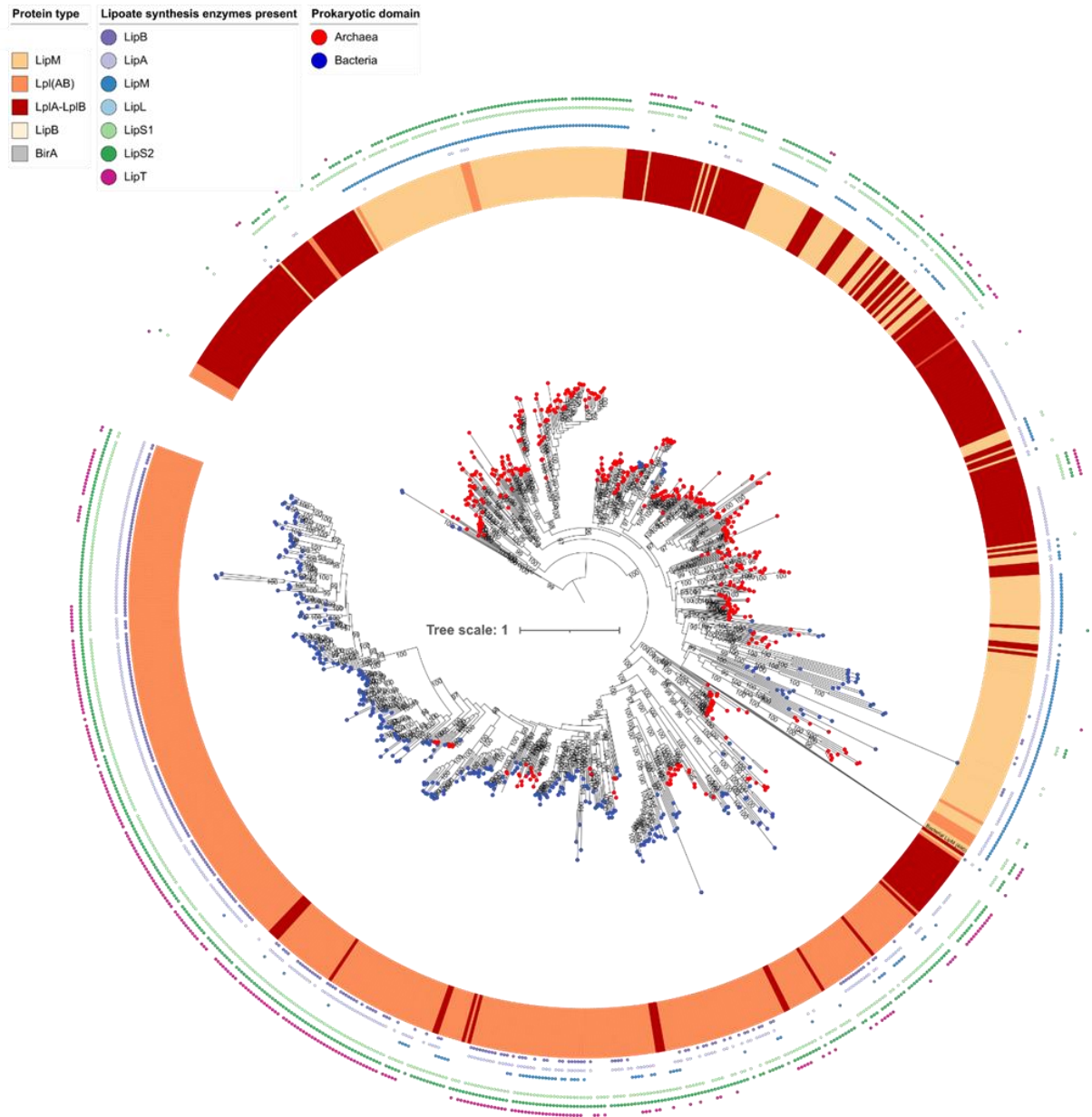


**Fig. S3. Mass spectrometric analyses of three different recombinant LbpA proteins** (*Tk90LbpA2* from *Thioalkalivibrio* sp. K90mix (a), *TsLbpA2* (b) and *TsLbpA1* (c) from *Thiorhodospira sibirica*) produced in *E. coli* BL21 (DE3)  $\Delta$ *iscR* [1] in the absence (green spectra) or presence (blue spectra) of helper plasmid pACYC-*Tk1pm* carrying genes *lipS1-slp(AB)-lipT-lipS2-lipY* from *Thioalkalivibrio* sp. K90mix. The mass spectrometric data is provided as “S2 data.rar”. Calculated and obtained masses for the three proteins, their expected modifications and the conditions for detection are given in the table below. The  $\Delta$ *iscR* strain is engineered for improved synthesis of iron-sulfur proteins like LipS1 and LipS2 by the removal of the gene for IscR, a transcriptional negative regulator of the *isc* (iron-sulfur cluster biosynthesis) operon. Predicted masses for the recombinant apo-proteins were 16547.47 Da (*Tk90LbpA2*), 16589.64 Da (*TsLbpA2*) and 17465.71 Da (*TsLbpA1*). In addition to the full length apo-proteins, polypeptides with masses 87-89 Da lower were recovered. This mass difference is consistent with the absence of the amino-terminal amino acid serine, derived from the Strep-tag in part of the protein population. Our hypothesis was confirmed by Edman degradation of *Tk90LbpA2*. In the presence of the helper plasmid, all LbpA protein species analyzed exhibited masses 151-160 Da higher, which is fully compatible with the addition of a monothiolated octanoyl moiety. Based on the results obtained for the archaeal LipS1/S2 proteins [2], we suggest that this is the 8-mercaptooctanyl intermediate.

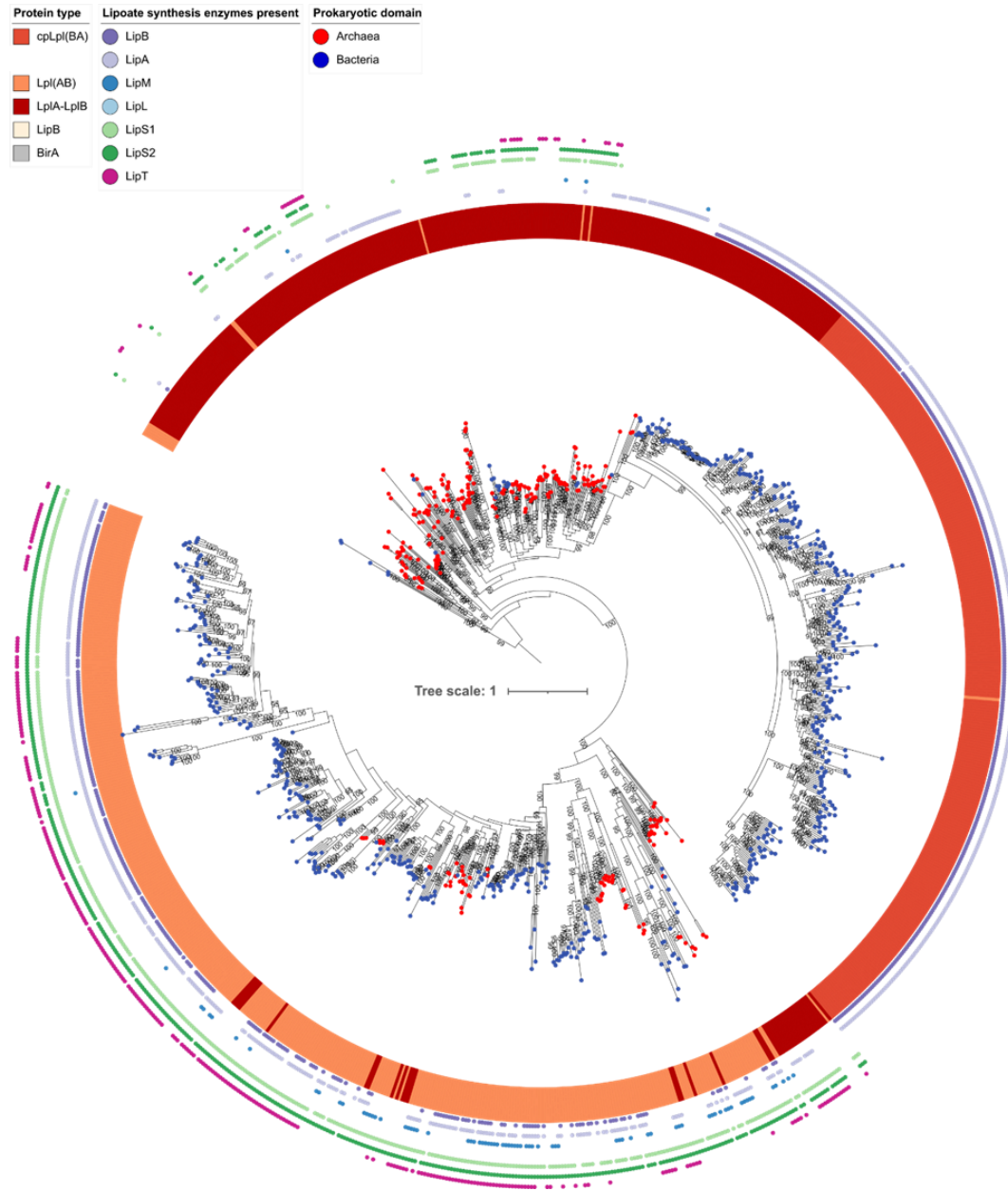
Protein	Modification	Calculated/obtained mass [Da]	Condition for obtained mass
<b><i>TK90LbpA2</i></b>	Apo-protein	16547/16546 ( $\Delta$ : 1)	Without helper plasmid
	Octanoylated	16672/nd	Not detected
	Monothiolated octanoyl	16704/16698 ( $\Delta$ : 6)	With helper plasmid
	Holo-protein	16736/nd	Not detected
	Apo-protein-Ser	16459/16459 ( $\Delta$ : 0)	Without helper plasmid
	Octanoylated-Ser	16584/nd	Not detected
	Monothiolated octanoyl-Ser	16616/16610 ( $\Delta$ : 6)	With helper plasmid
	Holo-protein-Ser	16648/nd	Not detected
<b><i>TsLbpA2</i></b>	Apo-protein	16589/16589 ( $\Delta$ : 0)	Without and with helper plasmid
	Octanoylated	16714/nd	Not detected
	monothiolated octanoyl	16746/16747 ( $\Delta$ : 1)	With helper plasmid
	Holo-protein	16778/nd	Not detected
	Apo-protein-Ser	16501/16500 ( $\Delta$ : 1)	Without helper plasmid
	Octanoylated-Ser	16626/nd	Not detected
	Monothiolated octanoyl -Ser	16658/16660 ( $\Delta$ : 2)	With helper plasmid
	Holo-protein-Ser	16690/nd	Not detected
<b><i>TsLbpA1</i></b>	Apo-protein	17465/17462 ( $\Delta$ : 3)	Without and with helper plasmid
	Octanoylated	17590/nd	Not detected
	Monothiolated octanoyl	17622/17614 ( $\Delta$ : 8)	With helper plasmid
	Holo-protein	17654/nd	Not detected
	Apo-protein -Ser	17377/17373 ( $\Delta$ : 4)	With helper plasmid
	Octanoylated-Ser	17502/nd	Not detected
	Monothiolated octanoyl -Ser	17534/17526 ( $\Delta$ : 8)	With helper plasmid
	Holo-protein-Ser	17566/nd	Not detected

**References**

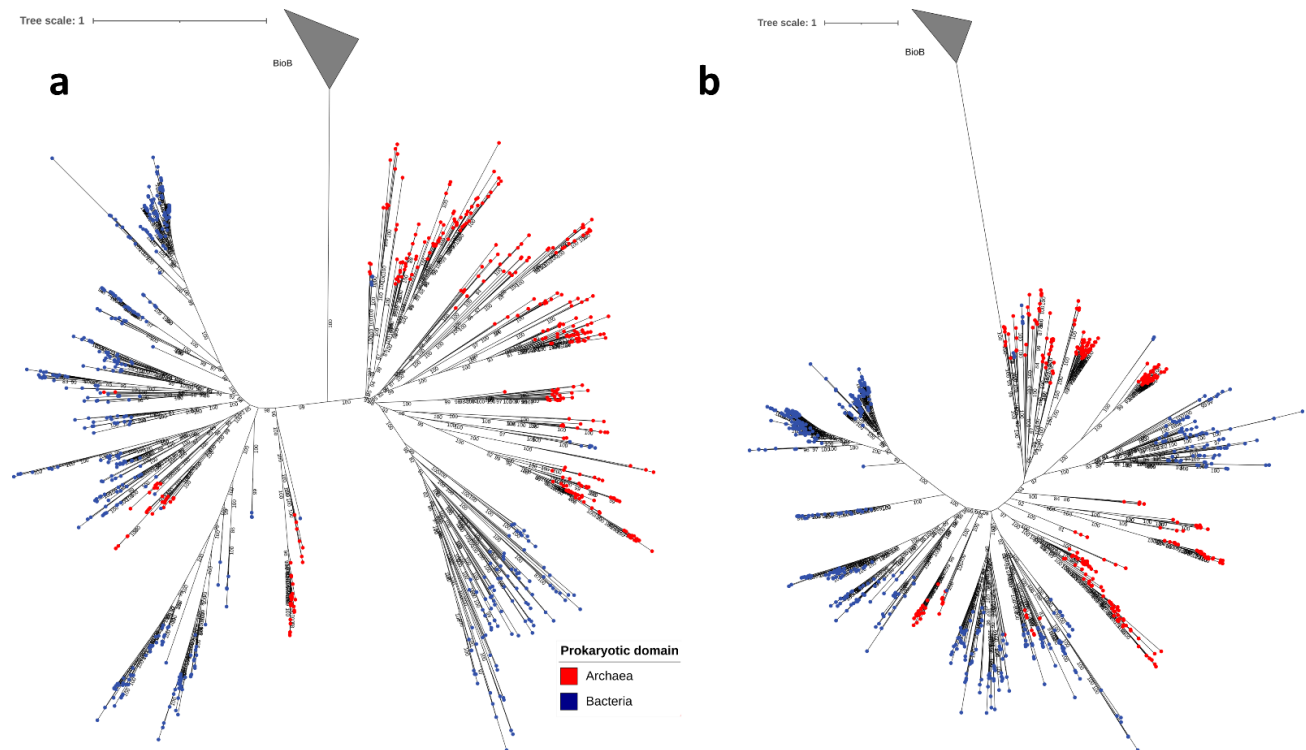
1. Akhtar MK, Jones PR. Deletion of *iscR* Stimulates Recombinant Clostridial Fe-Fe Hydrogenase Activity and H<sub>2</sub>-Accumulation in *Escherichia coli* BL21(DE3). *Appl. Microbiol. Biotechnol.* 2008; 78(5):853-862. <https://doi.org/10.1007/s00253-008-1377-6>
2. Neti SS, Sil D, Warui DM, Esakova OA, Solinski AE, Serrano DA, et al. Characterization of LipS1 and LipS2 from *Thermococcus kodakarensis*: Proteins Annotated as Biotin Synthases, Which Together Catalyze Formation of the Lipoyl Cofactor. *ACS Bio Med Chem Au.* 2022; 2(5):509-520. <https://doi.org/10.1021/acsbiochemau.2c00018>



**Fig. S4. Phylogeny for clade 3 lipoteichoic acid synthase:protein ligases including LipM, but excluding cpLpl(BA).** Introducing LipM did not change the support of an archaeal origin of clade 3. The data underlying this Figure can be found in Supplementary data S3.



**Fig. S5. Phylogenetic tree of for clade 3 lipote:protein ligases including cpLpl(BA) but excluding LipM.** Introducing cpLpl(BA) did not change the support of an archaeal origin of the clade 3. The data underlying this Figure can be found in Supplementary data S3.



**Fig. S6. Rooted individual phylogenetic trees for LipS2 (a) and LipS1 (b).** Sequences were derived from bacterial and archaeal genomes encoding exactly for one LipS1 and one LipS2. The trees were rooted by biotin synthase BioB as an outgroup. Archaeal sequences are marked in red, bacterial ones in blue. The data underlying this Figure can be found in Supplementary data S3.

**Table S1. Hidden Markov Models for extension of HMS-S-S [1]**

Hidden Markov Model	Threshold score	Source	Enzyme
TIGR00121_birA	83.1	Tigrfam	Bifunctional ligase/repressor BirA
TIGR00214_LipB	83.55	Tigrfam	octanoyltransferase
TIGR00510_LipA	310.25	Tigrfam	lipoate synthase
TIGR00545_LplA	201.85	Tigrfam	lipoate:protein ligase
PRK05481_LipA	203	InterPro	lipoate synthase
PF10437_LplB	24	pfam	accessory domain lipoate:protein ligase
PF03099_Lpl-Lip-bpl	21	pfam	small cofactor transferase family
IPR024897_LipL	100	InterPro	amidotransferase
cd16444_LipB	191.17	InterPro	octanoyltransferase
MF02118_LipM	130	HAMAP	octanoyltransferase
MF02119_LipL	100.1	HAMAP	amidotransferase
cd16443_LplA	100.1	InterPro	lipoate:protein ligase
cd16442_BPL	115	InterPro	biotin:protein ligase
TIGR00461_GcvP	859.2	Tigrfam	glycine cleavage P protein
TIGR00527_GcvH	115.85	Tigrfam	glycine cleavage H protein
TIGR00528_GcvT	235.85	Tigrfam	glycine cleavage T protein
TIGR00759_PdhE1	859.9	Tigrfam	pyruvat dehydrogenase E1
TIGR01348_PdhE2	552.1	Tigrfam	pyruvat dehydrogenase E2
TIGR01350_DHDL	408.8	Tigrfam	Dihydrolipoamide dehydrogenase
TIGR03181_PdhE1a	411.35	Tigrfam	pyruvat dehydrogenase E1
TIGR03182_PdhE1a	393.85	Tigrfam	pyruvat dehydrogenase E1
PRK09405_PdhE1	1326	InterPro	pyruvat dehydrogenase E1
cd02017_TPP_PdhE1	479	InterPro	pyruvat dehydrogenase E1
PRK09212_PdhE1b	321	InterPro	pyruvat dehydrogenase E1
PRK11855_PdhE2	576	InterPro	pyruvat dehydrogenase E2
PRK09404_KdhE1	764	InterPro	a-ketoglutarate dehydrogenase E1
PRK05704_KdhE2	449	InterPro	a-ketoglutarate dehydrogenase E2
TIGR00239_KdhE1	770	Tigrfam	a-ketoglutarate dehydrogenase E1
TIGR03186_KdhE2	1408	Tigrfam	a-ketoglutarate dehydrogenase E2
PF12573_BcdhE1a_N	26	pfam	branched-chain dehydrogenase E1
PRK11856_BcdhE2	250	InterPro	branched-chain dehydrogenase E1
PRK05976_DHDL	469	InterPro	Dihydrolipoamide dehydrogenase
PRK06327_DHDL	639	InterPro	Dihydrolipoamide dehydrogenase
PRK06467_DHDL	752	InterPro	Dihydrolipoamide dehydrogenase
PRK06912_DHDL	374	InterPro	Dihydrolipoamide dehydrogenase
PRK07818_DHDL	389	InterPro	Dihydrolipoamide dehydrogenase
PRK07846_DHDL	501	InterPro	Dihydrolipoamide dehydrogenase
COG1071_AodhE1	283	eggNOG	acetoin dehydrogenase
COG0022_AodhE1b	406	eggNOG	acetoin dehydrogenase
PRK14875_AodhE2	346.9	InterPro	acetoin dehydrogenase

## References

1. Tanabe TS, Dahl C. HMS-S-S: A Tool for the Identification of Sulphur Metabolism-Related Genes and Analysis of Operon Structures in Genome and Metagenome Assemblies. *Mol. Ecol. Resour.* 2022; 22(7):2758-2774. <https://doi.org/10.1111/1755-0998.13642>

**Table S6. Strains, plasmids and primers.** The locus tags given for *Ts. sibirica* refer to IMG JGI nomenclature.

Strains primers or plasmids	Relevant genotype, description or sequence	Reference or source
<b>Strains</b>		
<i>Thiorhodospira sibirica</i> ATCC 700588 <sup>T</sup>	Wild type	[1]
<i>Thioalkalivibrio</i> sp. K90 mix	Wild type	[2]
<i>Hyphomicrobium denitrificans</i> ATCC 51888 <sup>T</sup>	Wild type	[3]
<i>E. coli</i> 10β	Δ( <i>ara-leu</i> ) 7697 <i>araD139 fhuA ΔlacX74 galK16 galE15 e14- φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL</i> (Str <sup>R</sup> ) <i>rph spoT1 Δ(mrr-hsdRMS-mcrBC)</i>	New England Biolabs
<i>E. coli</i> BL21 (DE3)	<i>F- ompT hsdS<sub>B</sub>(r<sub>B</sub>m<sub>B</sub>) gal dcm</i> (DE3)	Novagen
<i>E. coli</i> BL21 (DE3) Δ <i>iscR</i>	<i>F- ompT hsdS<sub>B</sub>(r<sub>B</sub>m<sub>B</sub>) gal dcm <i>iscR::kan</i></i> (DE3)	[4]
<i>H. denitrificans</i> Δ <i>tsdA</i>	In-frame deletion of <i>tsdA</i> in <i>H. denitrificans</i> ATCC 51888	[5]
<i>H. denitrificans</i> Δ <i>tsdA</i> Δ <i>lbpA</i>	In-frame deletion of <i>lbpA</i> (Hden_0696) in <i>H. denitrificans</i> Δ <i>tsdA</i>	This work
<i>H. denitrificans</i> Δ <i>tsdA</i> <i>lbpA</i> -His	Complementation of <i>lbpA</i> -His into strain <i>Hyphomicrobium denitrificans</i> Δ <i>tsdA</i> Δ <i>lbpA</i>	This work
<i>H. denitrificans</i> Δ <i>tsdA</i> Δ <i>spl</i> (AB) <i>lbpA</i> -His	In-frame deletion of <i>splA</i> (Hden_686) in in <i>H. denitrificans</i> Δ <i>tsdA</i> <i>lbpA</i> -His	This work
<b>Primers</b>		
THISIDRAFT_1817 ( <i>lbpA2</i> ) fw	TTTT <b>CATATG</b> AGCGCTTGGAGCCACCCGAGTTCGAA AAAGGCGCCGGCACTGCAAAAGGC (NdeI)	This work
THISIDRAFT_1817 ( <i>lbpA2</i> ) rev	TTTT <b>CTCGAG</b> TTAAGCCCGCACCC (XhoI)	This work
TK90_0638 ( <i>lbpA1</i> ) strep_fw	TTTT <b>CATATG</b> AGCGCTTGGAGCCACCCGAGTTCGAA AAAGGCGCCGACTGCAACGGTTGC (NdeI)	This work
TK90_0638 ( <i>lbpA1</i> ) strep_rev	TTTT <b>CTCGAG</b> CTATCGCAGCTTGCCTG (XhoI)	This work
TK90_0640 ( <i>lbpA2</i> ) strep_fw	TTTT <b>CATATG</b> AGCGCTTGGAGCCACCCGAGTTCGAA AAAGGCGCCGGCGCAGTACGGGGT (NdeI)	This work
TK90_0640 ( <i>lbpA2</i> ) strep_rev	TTTT <b>CTCGAG</b> TTAGCAGCCGCCGAA (XhoI)	This work
TK90_0641-0644 fw	GCAGTT <b>AAAC</b> CGGAGCAAGCTGATGAC (DraI)	This work
TK90_0641-0644 rev	GCAT <b>CATGA</b> TTATGCGGACTTCTCGTCC (BspHI)	This work
Hden_0686_fw-up	ATATT <b>CTAGA</b> ATCTGCTGCTGAC ATATCCTGAAGG (XbaI)	This work
Hden_0686_fw-down	GTGAGTATAGGGCGGCGCAACCGTGGTTCACGCG GCTTT	This work
Hden_0686_rev-up	AAAGCCGCGTGAACCACGGTTGCGCGCCGCCCTATAC TCAC	This work
Hden_0686_rev-down	ATATT <b>CTAGA</b> GTGCAAATGACCTCGACGTCGTAGC (XbaI)	This work
Fwd5'_Δ <i>lbpA</i>	AAA <b>AGCATGCC</b> ACCAAGGGACGGCTCGCT (SphI)	[6]
Fwd3'_Δ <i>lbpA</i>	AAA <b>ATCTAG</b> AGCGATCGCTCGATGGAAAA (XbaI)	[6]
KI_HdLbpA2-His-Up-rev	TCAGTGGTGGTGGTGGTGGTGGCCGCCGCCGCCGCC GCAACCTGCGAAAC	Tanabe
KI_HdLbpA2-His-Down-fw	GGCGGCGGCGGCCACCACCACCACCACCTGAGCT CTACGGCCGCTCT	Tanabe
<b>Plasmids</b>		
pET22b	Ap <sup>r</sup>	Novagen
pACYC184	Cm <sup>r</sup> , Tc <sup>r</sup>	[7]
pACYC- <i>Tkpm</i>	Cm <sup>r</sup> , DraI/BspHI fragment covering <i>lipS1-spl(AB)-lipT-lipS2</i> from <i>Thioalkalivibrio</i> sp. K90mix in EcoRV/BspHI of pACYC184	This work
pET15b	Ap <sup>r</sup>	Novagen
pET22b	Ap <sup>r</sup>	Novagen
pET28b	Km <sup>r</sup>	Novagen



pHP45Ω-Tc	Ap <sup>r</sup> , Tc <sup>r</sup>	[8]
pk18 <i>mobsacB</i> -Tc	pHP45ΩTc tetracycline cassette inserted into pk18 <i>mobsacB</i> using SmaI	[9]
pk18 <i>mobsacB</i> Δ <i>tsdA</i> Tc	Km <sup>r</sup> , Tc <sup>r</sup> , 2.01 kb fragment implementing deletion of a 996 bp <i>tsdA</i> fragment in pk18 <i>mobsacB</i> with additional tetracycline resistance	[5]
pk18 <i>mobsacB</i> Δ <i>sLpA</i> -Tc	Km <sup>r</sup> , Tc <sup>r</sup> , 2.07 kb SOE PCR fragment implementing deletion of a 1029 bp fragment encoding amino acids 8 to 349 of sLpA cloned into pk18 <i>mobsacB</i> -Tc using XbaI restriction sites	This work
pk18 <i>mobsacB</i> - <i>lbpA</i> -His	Km <sup>r</sup> , SOE PCR fragment implementing chromosomal integration of <i>lbpA</i> joined with a His-tag encoding sequence cloned into pk18 <i>mobsacB</i> using XbaI and SphI restriction sites	This work
pk18 <i>mobsacB</i> - <i>lbpA</i> -His-Tc	Km <sup>r</sup> , Tc <sup>r</sup> , pk18 <i>mobsacB</i> - <i>lbpA</i> -his with tetracycline resistance gene from pHP45Ω cloned into SmaI site	This work
pET-TsLbpA1-N-Strep	Ap <sup>r</sup> , N-terminal Strep-tag, NdeI-EcoRI fragment of PCR amplified <i>lbpA1</i> from <i>Thiorhodospira sibirica</i> (THISIDRAFT_ RS04590, former ThisiDRAFT_1533) in pET22b	[6]
pET-TsLbpA2-N-Strep	Ap <sup>r</sup> , N-terminal Strep-tag, NdeI-XhoI fragment of PCR amplified <i>lbpA2</i> from <i>Thiorhodospira sibirica</i> (THISIDRAFT_ RS08775) in pET22b	This work
pET-TK90LbpA2-N-Strep	Ap <sup>r</sup> , N-terminal Strep-tag, NdeI-XhoI fragment of PCR amplified <i>lbpA2</i> from <i>Thioalkalivibrio</i> sp. K90mix (TK90_0640) in pET22b	This work

## References

- Bryantseva IA, Gorlenko VM, Kompantseva EI, Imhoff JF, Söling J, Mityushina L. *Thiorhodospira sibirica* gen. nov., sp. nov., a New Alkaliphilic Purple Sulfur Bacterium from a Siberian Soda Lake. *Int. J. Syst. Bacteriol.* 1999; 49:697-703. <https://doi.org/10.1099/00207713-49-2-697>
- Muyzer G, Sorokin DY, Mavromatis K, Lapidus A, Foster B, Sun H, et al. Complete Genome Sequence of *Thioalkalivibrio* sp. K90mix. *Stand. Genomic Sci.* 2011; 5(3):doi:10.4056/sigs.2315092. <https://doi.org/10.4056/sigs.2315092>
- Urakami T, Sasaki J, Suzuki KI, Komagata K. Characterization and Description of *Hyphomicrobium denitrificans* sp. nov. *Int. J. Syst. Bacteriol.* 1995; 45(3):528-532. <https://doi.org/10.1099/00207713-45-3-528>
- Akhtar MK, Jones PR. Deletion of *iscR* Stimulates Recombinant Clostridial Fe-Fe Hydrogenase Activity and H<sub>2</sub>-Accumulation in *Escherichia coli* BL21(DE3). *Appl. Microbiol. Biotechnol.* 2008; 78(5):853-862. <https://doi.org/10.1007/s00253-008-1377-6>
- Koch T, Dahl C. A Novel Bacterial Sulfur Oxidation Pathway Provides a New Link between the Cycles of Organic and Inorganic Sulfur Compounds. *ISME J.* 2018; 12(10):2479-2491. <https://doi.org/10.1038/s41396-018-0209-7>
- Cao X, Koch T, Steffens L, Finkensieper J, Zigann R, Cronan JE, et al. Lipoate-Binding Proteins and Specific Lipoate-Protein Ligases in Microbial Sulfur Oxidation Reveal an Atypical Role for an Old Cofactor. *eLife.* 2018; 7:e37439. <https://doi.org/10.7554/eLife.37439>
- Chang AC, Cohen SN. Construction and Characterization of Amplifiable Multicopy DNA Cloning Vehicles Derived from the p15a Cryptic Miniplasmid. *J. Bacteriol.* 1978; 134(3):1141-1156.
- Fellay R, Frey J, Krisch HM. Interposon Mutagenesis of Soil and Water Bacteria: A Family of DNA Fragments Designed for in Vitro Insertional Mutagenesis of Gram-Negative Bacteria. *Gene.* 1987; 52:147-154. [https://doi.org/10.1016/0378-1119\(87\)90041-2](https://doi.org/10.1016/0378-1119(87)90041-2)
- Li J, Koch J, Flegler W, Garcia Ruiz L, Hager N, Ballas A, et al. A Metabolic Puzzle: Consumption of C<sub>1</sub> Compounds and Thiosulfate in *Hyphomicrobium denitrificans* X<sup>1</sup>. *Biochim. Biophys. Acta - Bioenergetics.* 2023; 1864:148932. <https://doi.org/10.1016/j.bbabi.2022.148932>

Figure 2c

*Hyphomicrobium denitrificans* strain

<i>ΔtsdA</i>					
Time [h]	Thiosulfate [mM]	Standard deviation [mM]	Time [h]	Sulfite [mM]	Standard deviation [mM]
0	2,022	0,067	0	0,026	0,008
14,5	1,893	0,124	14,5	0,028	0,013
40	1,779	0,067	22,5	0,042	0,007
42	1,340	0,053	37,5	0,093	0,019
46	1,103	0,108	40	0,139	0,008
61,5	0,095	0,002	42	0,174	0,005
63,5	0,031	0,001	44	0,197	0,029
65,5	0,029	0,001	46	0,269	0,010
67,5	0,003	0,001	61,5	0,637	0,069
69,5	0,010	0,001	63,5	0,470	0,023
85,5	0,000	0,000	65,5	0,317	0,014
			67,5	0,242	0,007
			69,5	0,206	0,007
			85,5	0,044	0,009

<i>tsdA lbpA2-His</i>			<i>slp(AB)</i>		
Time [h]	Thiosulfate [mM]	Standard deviation [mM]	Time [h]	Sulfite [mM]	Standard deviation [mM]
0	2,207	0,102	0	0,030	0,003
15	2,291	0,131	15	0,020	0,007
22	2,162	0,091	22	0,030	0,007
24	1,987	0,124	24	0,034	0,001
37	1,811	0,073	37	0,029	0,006
40,25	1,758	0,065	40,25	0,018	0,011
42,25	1,818	0,094	42,25	0,023	0,007
46,5	1,855	0,101	44,5	0,015	0,016
63	1,537	0,090	46,5	0,025	0,004
68	1,242	0,113	49	0,021	0,008
86	1,274	0,142	63	0,021	0,005
			65	0,009	0,005
			68	0,014	0,007
			71	0,015	0,004
			87,5	0,009	0,001

<i>tsdA lbpA2</i>					
Time [h]	Thiosulfate [mM]	Standard deviation [mM]	Time [h]	Sulfite [mM]	Standard deviation [mM]
0	2,337	0,143	0	0,030	0,006
15	2,223	0,173	15	0,023	0,008
22	2,056	0,373	22	0,034	0,002
24	1,900	0,273	24	0,037	0,002
37	1,905	0,176	37	0,029	0,001
40,25	1,623	0,137	40,25	0,025	0,008
42,25	1,685	0,343	42,25	0,023	0,006
46,5	1,768	0,167	44,5	0,017	0,016
63	1,490	0,130	46,5	0,020	0,007
68	1,144	0,130	49	0,022	0,003
86	1,028	0,046	63	0,018	0,003
			65	0,008	0,010
			68	0,015	0,001
			71	0,016	0,002
			87,5	0,008	0,002

<i>tsdA lbpA2-His</i>					
Time [h]	Thiosulfate [mM]	Standard deviation [mM]	Time [h]	Sulfite [mM]	Standard deviation [mM]
0	1,940	0,350	0	0,031	0,009
14,5	1,754	0,206	14,5	0,030	0,012
22,5	1,644	0,209	22,5	0,059	0,007
37,5	1,750	0,062	37,5	0,079	0,006
44	1,702	0,170	40	0,098	0,005
46	1,529	0,049	42	0,130	0,006
61,5	0,377	0,012	44	0,163	0,033
63,5	0,180	0,020	46	0,244	0,003
65,5	0,084	0,008	61,5	0,665	0,064
67,5	0,003	0,002	63,5	0,598	0,039
69,5	0,008	0,002	65,5	0,375	0,015
85,5	0,006	0,005	67,5	0,136	0,016
			69,5	0,043	0,007
			85,5	0,016	0,001

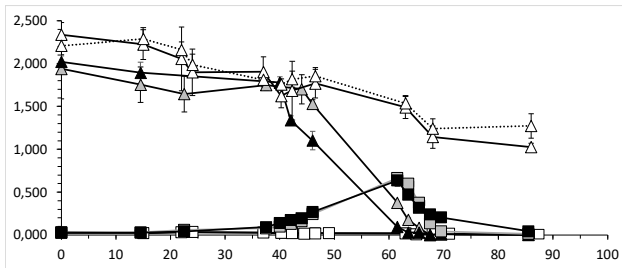


Figure 2d

*Hyphomicrobium denitrificans* strain

Delta *tsdA*

Time [h]	OD430 A	OD430 B	OD430 Mean	Standard deviation	Mean biomass	Standard deviation biomass
0	0,007	0,007	0,0070	0,0000	0,0021	0
14,5	0,015	0,019	0,0170	0,0020	0,0051	0,0006
22,5	0,045	0,044	0,0445	0,0005	0,0134	0,00015
42	0,361	0,364	0,3625	0,0015	0,1088	0,00045
46	0,539	0,540	0,5395	0,0005	0,1619	0,00015
61,5	1,110	1,122	1,1160	0,0060	0,3348	0,0018
63,5	1,315	1,337	1,3260	0,0110	0,3978	0,0033
65,5	1,550	1,530	1,5400	0,0100	0,4620	0,003
67,5	1,510	1,400	1,4550	0,0550	0,4365	0,0165
85,5	1,730	1,710	1,7200	0,0100	0,5160	0,003

*tsdA lbpA2-His sfp(AB)*

Time [h]	OD430 A	OD430 B	OD430 Mean	Standard deviation	Mean biomass	Standard deviation biomass
0	0,008	0,007	0,0075	0,0007	0,0023	0,0002
22	0,076	0,073	0,0745	0,0021	0,0224	0,0006
24	0,100	0,097	0,0985	0,0021	0,0296	0,0006
37	0,349	0,332	0,3405	0,0120	0,1022	0,0036
40,25	0,510	0,495	0,5025	0,0106	0,1508	0,0032
42,25	0,635	0,620	0,6275	0,0106	0,1883	0,0032
44,5	0,785	0,720	0,7525	0,0460	0,2258	0,0138
46,5	0,955	0,885	0,9200	0,0495	0,2760	0,0148
49	1,125	1,070	1,0975	0,0389	0,3293	0,0117
63	1,680	1,590	1,6350	0,0636	0,4905	0,0191
65	1,760	1,630	1,6950	0,0919	0,5085	0,0276
68	1,670	1,570	1,6200	0,0707	0,4860	0,0212
71	1,620	1,630	1,6250	0,0071	0,4875	0,0021
87,5	1,575	1,645	1,6100	0,0495	0,4830	0,0148

*tsdA lbpA2*

Time [h]	OD430 A	OD430 B	OD430 Mean	Standard deviation	Mean biomass	Standard deviation biomass
0,0	0,008	0,007	0,0075	0,0007	0,0023	0,0002
22,0	0,068	0,059	0,0635	0,0064	0,0191	0,0019
24,0	0,087	0,078	0,0825	0,0064	0,0248	0,0019
37,0	0,281	0,249	0,2650	0,0226	0,0795	0,0068
40,3	0,425	0,375	0,4000	0,0354	0,1200	0,0106
42,3	0,505	0,430	0,4675	0,0530	0,1403	0,0159
44,5	0,590	0,570	0,5800	0,0141	0,1740	0,0042
46,5	0,730	0,650	0,6900	0,0566	0,2070	0,0170
49,0	0,885	0,795	0,8400	0,0636	0,2520	0,0191
63,0	1,530	1,540	1,5350	0,0071	0,4605	0,0021
65,0	1,520	1,590	1,5550	0,0495	0,4665	0,0148
68,0	1,560	1,580	1,5700	0,0141	0,4710	0,0042
71,0	1,510	1,520	1,5150	0,0071	0,4545	0,0021
87,5	1,515	1,580	1,5475	0,0460	0,4643	0,0138

*tsdA lbpA2-His*

Time [h]	OD430 A	OD430 B	OD430 Mean	Standard deviation	Mean biomass	Standard deviation biomass
0,0	0,007	0,007	0,0070	0,0000	0,0021	0,0000
14,5	0,0140	0,0120	0,0130	0,0010	0,0039	0,0003
22,5	0,0260	0,0280	0,0270	0,0010	0,0081	0,0003
37,5	0,1390	0,1350	0,1370	0,0020	0,0411	0,0006
40,0	0,1650	0,1750	0,1700	0,0050	0,0510	0,0015
42,0	0,2150	0,2050	0,2100	0,0050	0,0630	0,0015
44,0	0,2700	0,2780	0,2740	0,0040	0,0822	0,0012
46,0	0,3340	0,3100	0,3220	0,0120	0,0966	0,0036
61,5	0,8050	0,7750	0,7900	0,0150	0,2370	0,0045
63,5	0,8850	0,8350	0,8600	0,0250	0,2580	0,0075
67,5	1,2100	1,3700	1,2900	0,0800	0,3870	0,0240
69,5	1,4900	1,3500	1,4200	0,0700	0,4260	0,0210
85,5	1,5500	1,7100	1,6300	0,0800	0,4890	0,0240

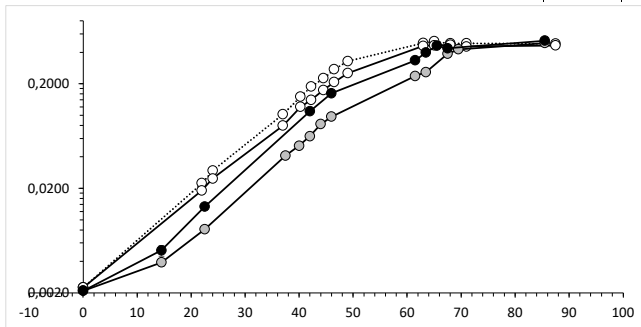


Figure 2a, lanes on the left

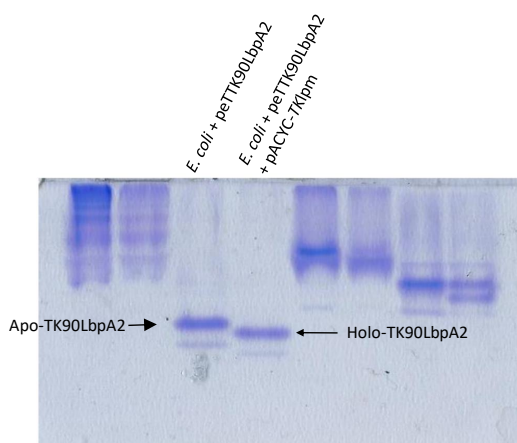
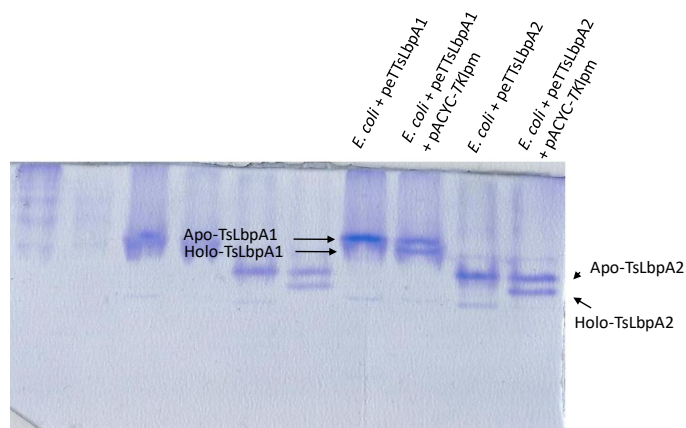
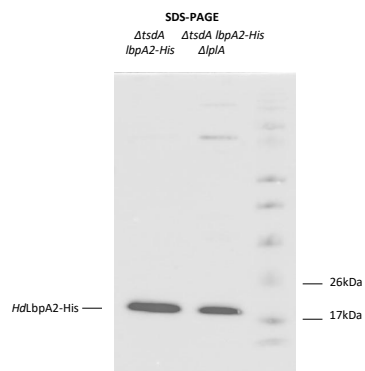


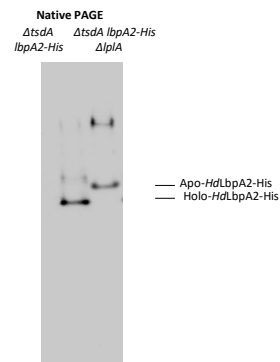
Figure 2a, lanes on right



**Figure 2e**



**Figure 2f**





---

## A cascade of sulfur transferases delivers sulfur to the sulfur-oxidizing heterodisulfide reductase-like complex

Tanabe, T. S., Bach, E., D'Ermo, G., Mohr, M. G., Hager, N., Pfeiffer, N., Guiral, M., Dahl, C.

---

In the cytoplasm, sulfur is generally transported in protein-bound form, as reduced sulfur compounds are highly reactive. The process is catalyzed by sulfur transferases that successively relay sulfur compounds between the transferases and finally to a specific target protein. These sulfur transfer networks can supply sulfur to multiple processes simultaneously, and therefore play an important role in the distribution and routing of sulfur within the cell (Dahl 2015, Tanabe *et al.* 2019). Genes for several putative sulfur transferases are associated with the sHdr genes, including the versatile sulfur transferase TusA and several DsrE-type sulfur transferases. Here, the sulfur transferases that channel the sulfur to the sHdr system from the sHdr system operating sulfur-oxidizing bacteria *Hyphomicrobium denitrificans*, *Aquifex aeolicus*, *Thioalkalivibrio sp.* K90mix and *Thiorhodospira sibirica* have been characterized.

The *shdr* and *lbpA* genes of the four selected model organisms have been described to be genetically linked to the sulfur transferases *tusA*, *dsrE2*, *dsrE3B*, *dsrE3C* and rhodanese *rhd442*. Furthermore *dsrE3A* has proposed to transfer sulfur in the sulfur-oxidizing archaeon *Metallosphaera cuprina*. The presence of these specific sulfur transferases and the sHdr system was mapped across the diversity of prokaryotes in the the genomce taxonomy database GTDB R207 to elucidate the general co-occurrence patterns. TusA was found to be always present in the analyzed sulfur-oxidizing archaea and bacteria underlining the importance of TusA for diverse metabolism (Tanabe *et al.* 2019). The sHdr associated TusA proteins were also shown to be phylogenetically distinguishable from the characterized TusA proteins associated with other anabolic processes but not from the TusA proteins involved in sulfur-oxidation via the Dsr system. This analysis also revealed the distribution of sHdr type I and type II systems. DsrE3B and DsrE3C were particularly correlated with the type I systems, while type II sHdr system were nearly exclusively linked to DsrE3A. In addition, it was possible to reconstruct the phylogeny of DsrE protein family including a large number of sequences. Similar phylogenies had been postulated in previous studies on the basis of a small number of sequences.

The recombinant TusA from *H. denitrificans*, *A. aeolicus*, *Thioalkalivibrio sp.* K90mix and *T. sibirica*, recombinant DsrE3C and Rhd442 from *H. denitrificans* and recombinant DsrE3B from *Thioalkalivibrio sp.* K90mix and *Ts. sibirica* were further analyzed.

The recombinant TusA proteins were incubated *in vitro* with various inorganic sulfur compounds to test their ability to mobilize sulfur. In these experiments it was possible to produce both sulfane

---

Tanabe, T. S., Bach, E., D'Ermo, G., Mohr, M. G., Hager, N., Pfeiffer, N., Guiral, M., Dahl, C. (2023) A cascade of sulfur transferases delivers sulfur to the sulfur-oxidizing heterodisulfide reductase-like complex. *bioRxiv*, 2023.12.18.572138; doi.org/10.1101/2023.12.18.572138. *Protein Science* under review

sulfur and thiosulfonate modifications. Based on the physiology of the original organisms, it was assumed that the sulfane-sulfur modification also occurs under physiological conditions. TusA from *H. denitrificans* and *A. aeolicus* were also produced with a cysteine to serine exchange. With these proteins, it was possible to determine the catalytically active cysteine that mediates sulfur binding and transfer.

Similarly, the sulfur mobilization capacity of DsrE3C and DsrE3B was determined. The catalytically active cysteine of *H. denitrificans* DsrE3C was also elucidated by the preparation of cysteine exchange mutants. *H. denitrificans* DsrE3C and *Thioalkalivibrio sp.* K90mix were observed to form homotrimeric complexes as indicated by gel permeation chromatography and on native polyacrylamide gel electrophoresis. To investigate the function of DsrE3C *in vivo* the strains *H. denitrificans*  $\Delta$ *tsdA*  $\Delta$ *dsrE3C*, *H. denitrificans*  $\Delta$ *tsdA dsrE3C* Cys<sup>83</sup>Ser and *H. denitrificans*  $\Delta$ *tsdA dsrE3C* Cys<sup>84</sup>Ser were constructed. The specific thiosulfate oxidation capacity of all three strains was severely reduced compared to *H. denitrificans*  $\Delta$ *tsdA*. Conversely, the growth rate of the mutant strains was higher and the growth inhibition caused by the production of toxic sulfite was not observed (Li *et al.* 2023). The essential function of DsrE3C in *H. denitrificans* was therefore confirmed.

To determine the direction of the sulfur transfer a transfer activity from persulfurated TusA proteins to the respective DsrE3 proteins was tested. Also a possible transfer from persulfurated DsrE3 to the respective TusA proteins were tested. The sulfur transfer was found to be efficient from the respective DsrE3B or DsrE3C to TusA, but the opposite direction was not efficient.

In an additional experiment the possible interaction partners of recombinant *A. aeolicus* TusA were determined. DsrE3B, DsrE3C, lipote-binding protein LbpA and several sHdr subunits were co-purified with recombinant TusA from *A. aeolicus* cell extract revealing a physical interaction of these proteins.

T.S.T. contributed to this study by conceptualization, data curation, formal analysis, investigation, visualization and writing of the original manuscript: T.S.T. conceptualized, performed, analyzed, validated and visualized all experiments regarding the distribution, phylogeny and structure simulation of the sulfur transferases and sHdr system. T.S.T. conceptualized and supervised the construction of the strains *H. denitrificans*  $\Delta$ *tsdA dsrE3C* Cys<sup>83</sup>Ser and *H. denitrificans*  $\Delta$ *tsdA dsrE3C* Cys<sup>84</sup>Ser. Furthermore T.S.T. analyzed, curated and visualized the data of the growth experiments and sulfur transfer assays. T.S.T. conceptualized and performed the sulfur transfer assays with recombinant *H. denitrificans* TusA and the DsrE3C Cys<sup>83</sup>Ser and DsrE3C Cys<sup>84</sup>Ser. The sulfur mobilization experiments with *H. denitrificans*, *Thioalkalivibrio sp.* K90mix and *T. sibirica* TusA and *H. denitrificans* DsrE3C, sulfur transfer experiment for *Thioalkalivibrio sp.* K90mix and *T. sibirica* TusA with the corresponding DsrE3B proteins, as well as the sulfur transfer between *H. denitrificans* TusA and DsrE3C proteins, were also performed by T.S.T. as part of a master's thesis.



1 **Title:** A cascade of sulfur transferases delivers sulfur to the sulfur-  
2 oxidizing heterodisulfide reductase-like complex

3

4

5 **Running title:** Shuttling sulfur to the sHdr-system

6 **Authors:** Tomohisa Sebastian Tanabe,<sup>1</sup> Elena Bach,<sup>1</sup> Giulia D'Ermo,<sup>2</sup> Marc Gregor

7 Mohr,<sup>1</sup> Natalie Hager,<sup>1</sup> Niklas Pfeiffer,<sup>1</sup> Marianne Guiral,<sup>2</sup> Christiane Dahl<sup>1\*</sup>

8 <sup>1</sup>Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich-Wilhelms-Universität Bonn,

9 Bonn, Germany

10 <sup>2</sup>CNRS, Bioénergétique et Ingénierie des Protéines, Aix Marseille Université, IMM, 13009

11 Marseille, France

12 **\*Correspondence:** Christiane Dahl, Institut für Mikrobiologie & Biotechnologie, Rheinische

13 Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 168, D-53115

14 Bonn, Germany, Phone +49 228 732119, Fax: +49 228 737576,

15 E-Mail: [chdahl@uni-bonn.de](mailto:chdahl@uni-bonn.de)

16

17 Tomohisa Sebastian Tanabe: [s6totana@uni-bonn.de](mailto:s6totana@uni-bonn.de)

18 Elena Bach: [elena.bach97@web.de](mailto:elena.bach97@web.de)

19 Giulia D'Ermo: [giuliadermo@gmail.com](mailto:giuliadermo@gmail.com)

20 Natalie Hager: [s6nahage@uni-bonn.de](mailto:s6nahage@uni-bonn.de)

21 Marc Gregor Mohr: [s6mcmohr@uni-bonn.de](mailto:s6mcmohr@uni-bonn.de)

22 Niklas Pfeiffer: [niklas.pfeiffer98@gmx.de](mailto:niklas.pfeiffer98@gmx.de), Current affiliation: Labor Dr. Wisplinghoff, Horbeller Str.

23 18-20, 50858 Köln, Germany

24 Marianne Guiral: [guiral@imm.cnrs.fr](mailto:guiral@imm.cnrs.fr)

25

26 Number of manuscript pages: 35

27 Figures: 9

28 Tables: 1

29 **Supplementary Material:** Figs S1-S4, Tables S1-S7

30

**31 ABSTRACT**

32 A heterodisulfide reductase-like complex (sHdr) and novel lipoate-binding proteins (LbpAs) are  
33 central players of a wide-spread pathway of dissimilatory sulfur oxidation. Bioinformatic  
34 analysis demonstrate that the cytoplasmic sHdr-LbpA systems are always accompanied by sets  
35 of sulfur transferases (DsrE proteins, TusA, rhodanases). The exact composition of these sets  
36 may vary depending on the organism and sHdr system type. To enable generalizations, we  
37 studied model sulfur oxidizers from distant bacterial phyla, i.e. Aquificota and  
38 Pseudomonadota. DsrE3C of the chemoorganotrophic Alphaproteobacterium  
39 *Hyphomicrobium denitrificans* and DsrE3B from the Gammaproteobacteria *Thioalkalivibrio* sp.  
40 K90mix, an obligate chemolithotroph, and *Thiorhodospira sibirica*, an obligate  
41 photolithotroph, are homotrimers that donate sulfur to TusA. Additionally, the hyphomicrobial  
42 rhodanese-like protein Rhd442 exchanges sulfur with both TusA and DsrE3C. The latter is  
43 essential for sulfur oxidation in *Hm. denitrificans*. TusA from *Aquifex aeolicus* (AqTusA)  
44 interacts physiologically with AqDsrE, AqLbpA and AqsHdr proteins. This is particularly  
45 significant as it establishes a direct link between sulfur transferases and the sHdr-LbpA  
46 complex that oxidizes sulfane sulfur to sulfite. *In vivo*, it is unlikely that there is a strict  
47 unidirectional transfer between the sulfur-binding enzymes studied. Rather, the sulfur  
48 transferases form a network, each with a pool of bound sulfur. Sulfur flux can then be shifted  
49 in one direction or the other depending on metabolic requirements. A single pair of sulfur-  
50 binding proteins with a preferred transfer direction, such as a DsrE3-type protein towards  
51 TusA, may be sufficient to push sulfur into the sink where it is further metabolized or needed.

**52 SIGNIFICANCE STATEMENT**

53 A network of bacterial sulfur transferases is uncovered and characterized that ultimately  
54 delivers sulfur to a complex cytoplasmic sulfur-oxidizing metalloenzyme, sHdr, that resembles  
55 heterodisulfide reductase from methanogenic archaea and interacts with lipoate-binding  
56 proteins. Similar sets of sulfur transferases occur in phylogenetically distant bacteria,  
57 underscoring the fundamental importance of the work.

58 **KEYWORDS:** dissimilatory sulfur oxidation, sulfur transferases, *Aquifex aeolicus*, sHdr  
59 pathway, DsrE

## 60 1. INTRODUCTION

61 In the cytoplasm of prokaryotes, as in most (if not all) other biological contexts, reduced  
62 sulfur is normally handled in a protein-bound state due to its reactivity.<sup>(1-5)</sup> Often, multiple  
63 sulfur transferases form a network to direct the sulfur to the correct metabolic pipeline or its  
64 target molecule. In these networks, individual sulfur-trafficking proteins may provide sulfur to  
65 multiple target reactions/proteins. Such networks are not only important for the biosynthesis  
66 of sulfur-containing cellular components but also essential in many Bacteria and Archaea,  
67 which carry out dissimilatory sulfur oxidation with energy conservation via respiratory or  
68 photosynthetic electron transport.<sup>(6,7)</sup> Here, the enzymatic production of persulfide sulfur, the  
69 successive transfer of sulfur as a persulfide between multiple proteins, and the oxidation of  
70 sulfane sulfur in protein-bound form are all crucial steps. Rhodanases, TusA and DsrE-like sulfur  
71 transferases are central and common elements in these processes. They have an established  
72 role in the rDsr pathway of sulfur oxidation where they work together to transfer sulfur as a  
73 cysteine persulfide to DsrC, which ultimately presents the sulfur to the oxidizing enzymatic  
74 unit, dissimilatory sulfite reductase, DsrAB.<sup>(6,8,9)</sup> In the model organism *Allochromatium*  
75 *vinosum*, sulfur atoms are successively transferred from the rhodanase Rhd\_2599 to TusA,<sup>(8)</sup>  
76 then to a conserved cysteine residue of DsrE, the active site subunit of the heterohexameric  
77 DsrE<sub>2</sub>F<sub>2</sub>H<sub>2</sub> complex,<sup>(10)</sup> and from there to DsrC.<sup>(9)</sup> The function of the membrane-bound  
78 AvDsrE2A protein, another sulfur transferase involved in this enzymatic relay, remains  
79 unclear.<sup>(8)</sup>

80 Recently, the sulfur-oxidizing heterodisulfide reductase-like sHdr complex and novel  
81 lipolate-binding proteins (LbpAs) have been identified as central players of an additional  
82 widespread sulfur oxidation pathway.<sup>(11-13)</sup> It exists in a significant group of organisms that  
83 comprise the volatile organic sulfur compound degrader *Hyphomicrobium denitrificans*, along  
84 with many chemo- and photolithoautotrophic bacteria and archaea, that include several  
85 environmentally relevant sulfur oxidizers such as *Acidithiobacillus sp.*, *Thioalkalivibrio* or  
86 *Sulfobacillus* species and the hyperthermophile *Aquifex aeolicus*. Two types of gene *shdr*  
87 clusters can be differentiated (Figure 1): Type I with an *shdrC1B1AHC2B2* arrangement and  
88 type II with an *shdrC1B1AHB3-etfAB-emo* arrangement.<sup>(11,14)</sup> While the importance of the type  
89 I cluster for sulfur dissimilation has been proven,<sup>(11-13)</sup> less is known for type II, although  
90 prokaryotes encoding the type II system are known sulfur oxidizers.<sup>(14)</sup> Irrespective of the type  
91 of gene cluster, the *shdr-lbpA* genes are conspicuously often associated with genes for

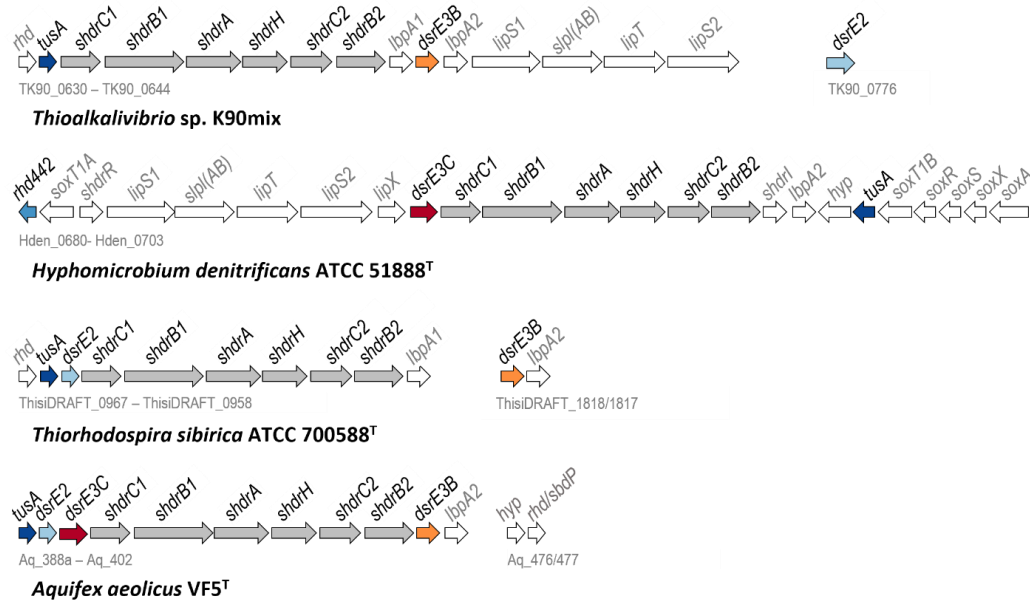
92 accessory sulfurtransferases similar to but not identical with those that fuel the rDsr pathway  
93 (Figure 1).<sup>(11,12)</sup>

94 Biochemical information regarding sulfur transferases associated with sHdr is extremely  
95 limited. The only sHdr-associated sulfur transferases studied thus far come from the archaeon  
96 *Metallosphaera cuprina*. Its DsrE3A protein has been biochemically characterized as a  
97 thiosulfonate transferase.<sup>(15)</sup> Bound thiosulfate is transferred from DsrE3A to TusA but not vice  
98 versa<sup>(15)</sup> implying that DsrE3A functions as a thiosulfate donor to TusA *in vivo*. The DsrE2A-type  
99 sulfur transferases, which are encoded in certain *shdr* gene clusters (Figure 1), have been  
100 proposed as potential membrane anchors for the sHdr-like complex.<sup>(16)</sup>

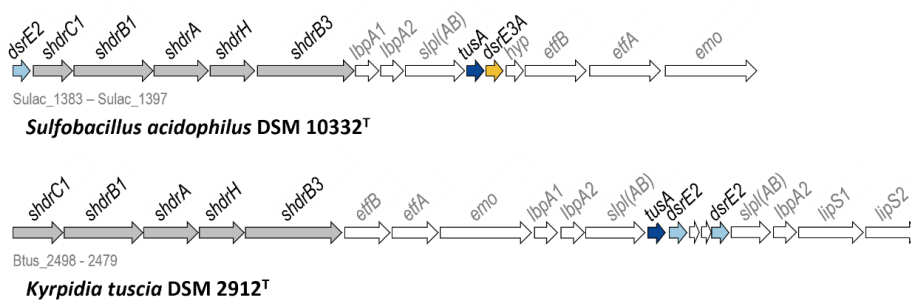
101 Here we set out to shed more light on sulfur transfer to the sHdr system. Initially, we  
102 analyse the potential correlation between the prevalence of *shdr* and sulfur transferases  
103 genes. To enable generalizations, we concentrate on the type I sHdr system in bacteria and our  
104 model organisms stem from two distant phyla: the Aquificota (*Aquifex aeolicus*) and the  
105 Pseudomonadota, which are represented by sulfur oxidizers from two different classes, the  
106 Alphaproteobacteria (*Hyphomicrobium denitrificans*) and the Gammaproteobacteria  
107 (*Thiorhodospira sibirica* and *Thioalkalivibrio* sp. K90mix). From these organisms, we investigate  
108 the proteins Rhd442, TusA, DsrE3B and DsrE3C to determine whether and how they mobilize  
109 and transfer sulfur. In the genetically accessible *Hm. denitrificans*, we gather information on  
110 the importance of DsrE3C *in vivo*.

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Type I sHdr system



## Type II sHdr system



111

112 **FIGURE 1.** Representative *shdr* gene clusters in sulfur oxidizers. The KEGG/NCBI locus tag identifiers for the first and last genes are shown below each cluster. Genes for Tusa, DsrE2, DsrE3A, DsrE3B and DsrE3C are indicated in dark blue, light blue, yellow, orange and dark red, respectively. Genes for probable components of the sulfur-oxidizing heterodisulfide reductase-like (sHdr) complex are shaded in gray. EMO, ETF:(methyl)menaquinone oxidoreductase; Etf, electron transfer protein. EtfAB and EMO have been proposed to direct electrons stemming from sulfane sulfur oxidation to menaquinone.<sup>(17)</sup> LbpA, lipoyl-binding protein; sLpI(AB), lipoyl:protein ligase; LipS1/S2, lipoyl synthase; LipT, FAD-binding NAD(P)H-dependent oxidoreductase possibly delivering electrons for the LipS1/S2-catalyzed sulfur insertion step<sup>(13,17)</sup>; Rhd, rhodanese; SbdP, sulfur-binding-donating protein.

121

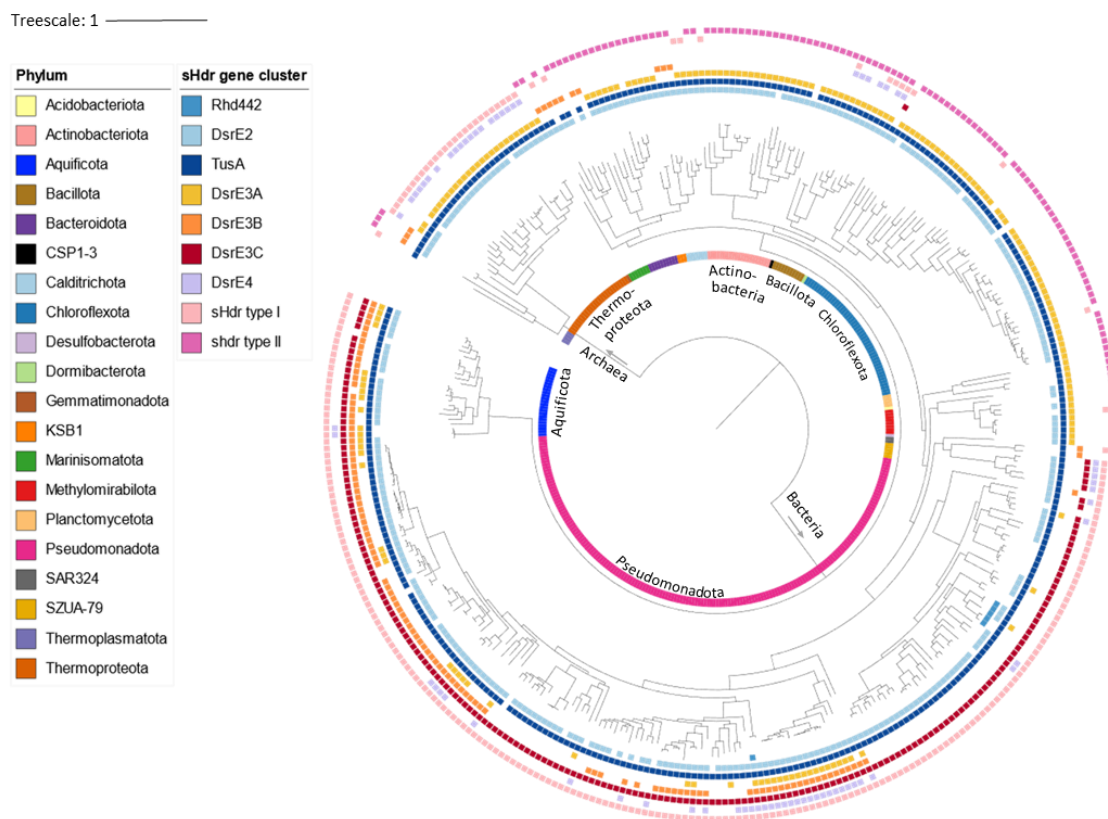
## 122 **2 RESULTS**

### 123 **2.1 Distribution of the sHdr system**

124 Clusters of genes encoding the sHdr pathway for sulfane sulfur oxidation in the  
 125 cytoplasm fall into two distinct categories<sup>(11,14,17)</sup> (Figs. 1 and 2). The type I and type II sHdr  
 126 systems share several core proteins, namely the Fe/S-flavoprotein sHdrA, the electron carrier  
 127 protein sHdrC1, that binds two cubane [4Fe-4S] clusters and the proposed catalytic subunit  
 128 sHdrB1 that probably coordinates two noncubane Fe/S clusters.<sup>(18)</sup> sHdrC2 is another  
 129 ferredoxin-like electron carrier. sHdrB2 has the potential to bind two classical noncubane Fe/S  
 130 clusters and probably acts as a disulfide reductase.<sup>(18)</sup> Organisms with type II sHdr systems  
 131 encode a protein that we term sHdrB3. This is a fusion of sHdrC2 and sHdrB2, albeit it can bind  
 132 only one noncubane Fe/S cluster.<sup>(17)</sup> Electron transfer protein EtfAB and  
 133 ETF:(methyl)menaquinone oxidoreductase EMO are encoded within type II *shdr* clusters and  
 134 have been proposed to direct electrons stemming from sulfane sulfur oxidation to  
 135 menaquinone.<sup>(17)</sup>

136 Of the entire GTDB representative genome collection (release R207), 397 assemblies from  
 137 20 phyla contain core *shdr* genes. For 353 of these assemblies, the concatenated sequences  
 138 for 16 ribosomal proteins could be used as phylogenetic markers to compute a species  
 139 tree,<sup>(19,20)</sup> that served as the basis for mapping the distribution of type I and II sHdr systems  
 140 using HMSS2<sup>(21)</sup> (Figure 2, Table S1). Among the Bacteria, the Pseudomonadota, the Aquificota,  
 141 and the phylum SZUA-79 have exclusively the type I system. Among the Archaea, the type I  
 142 *shdr* gene set occurs in the Thermoproteota, a phylum harboring well-established sulfur  
 143 oxidizers like *Sulfolobus sp.* and *Acidianus sp.* The type II sHdr system is found in 130 assemblies  
 144 from one archaeal and 12 bacterial phyla. The majority of the respective bacterial assemblies  
 145 belong to the Chloroflexota, Actinobacteriota and Bacillota. In some bacterial phyla, there are  
 146 species that have either type I or type II sHdr systems (e.g. Marinisomatota or Chloroflexota).  
 147 All sHdr-containing members of the Bacillota contain the type II system. In addition, four  
 148 *Sulfobacillus* species, which are well-established sulfur oxidizers,<sup>(14,22-24)</sup> and one further  
 149 member of the order Sulfobacillales bear the genetic capacity for both sHdr systems (Figure 2,  
 150 Table S1).

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



151

152 **FIGURE 2.** Taxonomic distribution of type I or type II sHdr systems in sulfur-oxidizing prokaryotes. The  
 153 distribution of TusA, Rhd442 and DsrE-type sulfur transferases is also visualized. The data underlying  
 154 the figure is provided in Table S1.

## 155 2.2 Distribution of sHdr-associated sulfurtransferases

156 Here, we intended to further illuminate the general association of *shdr* gene clusters  
 157 with genes for different sulfurtransferases<sup>(2,11,15,17)</sup> (TusA, DsrE-type sulfur transferases,  
 158 rhodanese Rhd442). In order to do so, we first needed to clearly describe and validate the  
 159 various classes of proteins using sequence similarity networks (SSN). Clusters in SSNs reflect  
 160 phylogenetic clades.<sup>(25)</sup>

161 TusA family proteins are a central hub in sulfur transfer during various anabolic and  
 162 catabolic processes.<sup>(1)</sup> In *Escherichia coli*, the three homologous TusA-family proteins, TusA,  
 163 YedF and YeeD have distinct functions and cannot substitute for each other.<sup>(26)</sup> Besides  
 164 contributing to tRNA thiolation,<sup>(27)</sup> TusA mediates sulfur transfer for molybdenum cofactor  
 165 biosynthesis and affects iron-sulfur cluster assembly as well as the activity of major regulatory  
 166 proteins.<sup>(26,28,29)</sup> YeeD is a component of thiosulfate uptake for sulfur assimilation.<sup>(30)</sup> YedF in

167 some way affects flagella formation and motility.<sup>(31)</sup> We retrieved all sequences clustering with  
168 these three proteins from Uniprot and subjected them to a SSN analysis together with TusA-  
169 like proteins that are genetically linked with sulfur-oxidation systems. Indeed, TusA, YedF, and  
170 YeeD each form distinct clusters of similar sequences (Figure 3a). In addition, three further  
171 clusters are clearly discernible. Two of them consist of TusA-like proteins linked with rDsr  
172 systems and a third comprises TusAs linked with sHdr and rDsr-systems. In summary, TusAs  
173 genetically associated with sulfur oxidation systems can be confidently distinguished from  
174 classical TusA, YedF and YeeD and form distinct phylogenetic clades (Figure 3a).

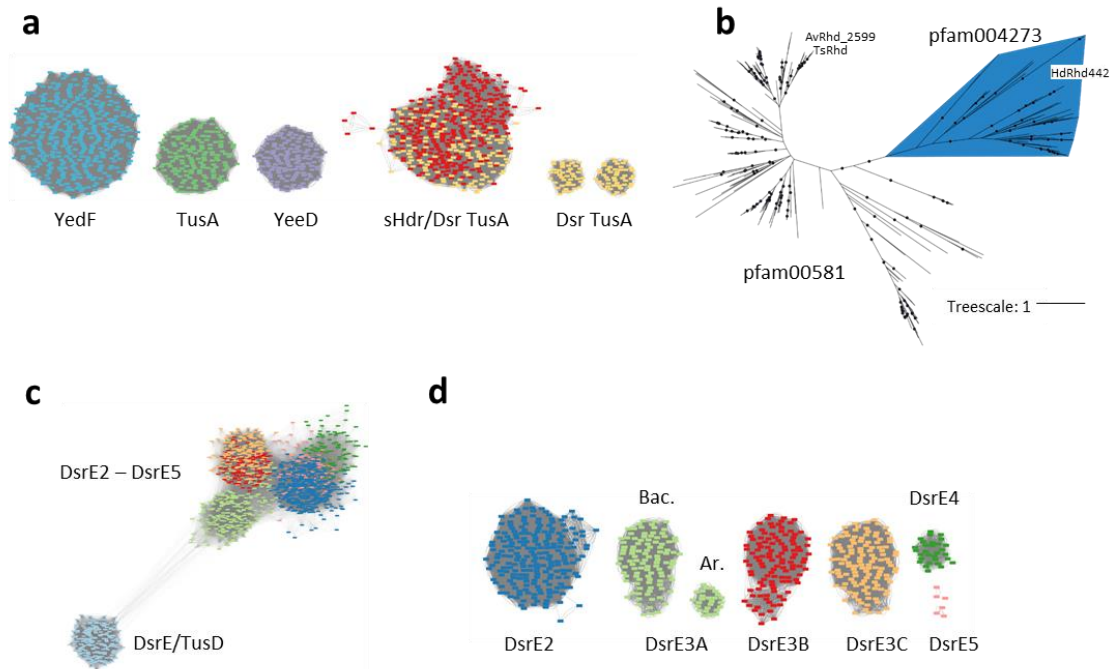
175 Rhodanases were originally identified and named on the basis of their ability to catalyze  
176 the transfer of a sulfane sulfur atom from thiosulfate to cyanide yielding SCN<sup>-</sup> (rhodanide,  
177 thiocyanate) as the product.<sup>(32)</sup> Rhodanases and rhodanese-like proteins are very  
178 widespread<sup>(33)</sup> and therefore we did not consider it useful to create an SSN analysis spanning  
179 all prokaryotes, but instead limited ourselves to sulfur oxidizers with rDsr and/or sHdr systems.  
180 Within a first group of these, the corresponding gene is present in a syntenic cluster with genes  
181 for DsrE2 and TusA (e.g. in *Ts. sibirica*, Figure 1). A second group contains proteins related to  
182 rhodanese Rhd442 encoded in the vicinity of the *shdr* gene cluster in *Hm. denitrificans* X<sup>T</sup>  
183 (Figure 1). By establishing a phylogenetic tree (Figure 3b), it became apparent that the DsrE2-  
184 TusA associated rhodanases belong to a different clade than the Rhd442-related proteins  
185 (Figure 3b). The former belong to a protein family (pfam00581) that also encompasses  
186 Rhd\_2599 from *Ac. vinosum*<sup>(8)</sup> and SbdP from *Aq. aeolicus*.<sup>(34)</sup> While the first is part of the relay  
187 delivering sulfur to the rDsr pathway of sulfur oxidation,<sup>(8)</sup> SbdP can load long sulfur chains and  
188 interacts with sulfur reductase and sulfur oxygenase reductase, i.e. key enzymes of sulfur  
189 energy metabolism.<sup>(34)</sup> Rhd442 belongs to a separate protein family, pfam004273 (DUF442)  
190 that forms a monophyletic group in the rhodanese tree presented Figure 3b.

191 Originally, the DsrE family has been categorized into five well distinguishable  
192 phylogenetic groups, DsrE, DsrE2, DsrE3 (with subgroups DsrE3A, DsrE3B and DsrE3C), DsrE4  
193 and DsrE5.<sup>(15,16,35)</sup> Here, we clustered all DsrE-type sequences from the representative  
194 genomes of GTDB (release R207) by sequence similarity network (SSN) analysis. In a first  
195 approach, the proteins were not filtered for association with dissimilatory sulfur oxidation.  
196 Classical TusD/DsrE were clearly distinguishable and only distantly related to the  
197 sulfurtransferases from the other subclasses (Figure 3c). In a second approach, an SSN was  
198 calculated with all DsrE homologs associated with rDsr and/or sHdr systems, excluding the



bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

199 divergent TusD/DsrE proteins (Figure 3d). For DsrE2, DsrE3 and DsrE4 robust clades mirroring  
 200 the original DsrE groups and subgroups were retrieved.<sup>(15)</sup> DsrE3A formed two robust  
 201 phylogenetic clades, containing archaeal and bacterial sequences, respectively (Figure 3d).  
 202 DsrE5 proteins present an exception as they do not cluster as a coherent group due to high  
 203 sequence dissimilarity. While functional predictions for this group are not available, members  
 204 of the DsrE4 group are proposed to play a role in detoxification of reactive sulfur species.<sup>(15)</sup>  
 205 As already pointed out, archaeal DsrE3A is an established thiosulfonate carrier whereas DsrE3B  
 206 and DsrE3C have not been studied on a biochemical level yet. Although a genetic association  
 207 indicates a function in the sHdr system, the mechanism and specific roles of DsrE3B and DsrE3C  
 208 have not been described so far.



209

210 **FIGURE 3.** Sequence similarity network (SSN) and phylogenetic analyses for sulfur transferases relevant  
 211 for dissimilatory sulfur oxidation. (a) SSN for the TusA family (pfam01206) members clustering at 50 %  
 212 identity with TusA, YedF and YeeD from *E. coli* and TusA-like proteins that are genetically linked with  
 213 sHdr and/or Dsr sulfur-oxidation systems. Connections below a threshold score of 100 were removed.  
 214 (b) Maximum likelihood tree for ~100-aa single domain rhodanases that are encoded in synteny with  
 215 DsrE2 and TusA or that are related to Rhd442 from *Hm. denitrificans* X<sup>T</sup>. Bootstrap values exceeding  
 216 95% are indicated by dots. (c) SSN for all DsrE-type sequences from the representative genomes of  
 217 GTDB (release R207). (d) SSN for all DsrE homologs associated with rDsr and/or sHdr systems, excluding  
 218 the divergent TusD/DsrE proteins. Connections below a threshold score of 135 were removed. Bac.,  
 219 Bacteria; Ar, Archaea.

220 Once the sulfurtransferases could be confidently categorized, their co-occurrence in  
 221 sHdr-containing organisms was studied using HMSS2.<sup>(21)</sup> In addition, they were mapped onto  
 222 the phylogenetic tree shown in Figure 2. Of all 397 studied assemblies, 387 encode at least  
 223 one TusA and 302 of the *shdr* clusters are directly linked to *tusA* genes. The *dsrE2* gene is  
 224 present in 345 (87%) assemblies and in 199 (50.1%) of the *shdr* gene clusters, making it the  
 225 second most frequently occurring gene. The *dsrE3A* (200), *dsrE3B* (132), and *dsrE3C* (223)  
 226 genes are less prevalent. However, when these genes are present, 70%, 90% and 52%,  
 227 respectively, are located in the immediate vicinity of *shdr* core genes. The rhodanese-encoding  
 228 gene *rhd442* is present in only three *shdr* gene clusters from the family *Hyphomicrobiaceae*  
 229 (Alphaproteobacteria).

230 The distribution of DsrE-type sulfurtransferases appears to relate to the type of sHdr  
 231 system (Figure 2). Archaea with type I sHdr system particularly often contain *shdr* genes linked  
 232 with *tusA*, *dsrE3A* and *dsrE4*. This is different in Bacteria with type I sHdr system, where a  
 233 *tusA/dsrE3C* combination occurs in the same genome at a notably high frequency. Other DsrE-  
 234 type sulfur transferases may be present but are less abundant. Some members of the  
 235 Pseudomonadota and Aquificota simultaneously encode DsrE3A and DsrE3B. The highest  
 236 number of different sulfurtransferases is found in genomes from the gammaproteobacterial  
 237 orders Acidithiobacillales, Acidiferrobacterales and Ectothiorhodospirillales (here in the genus  
 238 *Thioalkalivibrio* and in the family Acidihalobacteraceae). These organisms have the genetic  
 239 potential for DsrE2, DsrE3A, DsrE3B, DsrE3C and DsrE4. Type II sHdr systems are predominantly  
 240 associated with TusA and DsrE3A, while DsrE3B and DsrE3C are only very rarely present (Figure  
 241 2). Among the type II-containing archaea of the phylum Thermoplasmatota, the sulfur  
 242 transferase DsrE3B is the only transferase present and is genetically associated with the *shdr*  
 243 cluster.

### 244 2.3 Properties of sHdr associated TusA

245 The TusA proteins from *Hm. denitrificans* (HdTusA), *Thioalkalivibrio* sp. K90mix (TkTusA),  
 246 *Ts. sibirica* (TsTusA) and *Aq. aeolicus* (AqTusA), all of which are encoded in type I *shdr* gene  
 247 clusters, were selected as model proteins for further analysis. All four proteins contain a highly  
 248 conserved cysteine within an N-terminal CPXP motif, which is characteristic for the active site  
 249 of TusA proteins (Figure 4a). Leucine and isoleucine have been described at the X position for  
 250 TusA proteins involved in sulfur oxidation.<sup>(1)</sup> A second cysteine is present at equivalent



272 To elucidate the capacity of TusA from our four sHdr-containing model sulfur oxidizers  
273 for sulfur mobilization from inorganic and organic sulfur compounds, the purified proteins  
274 were incubated with 5 mM polysulfide ( $-SS_nS-$ ), thiosulfate ( $S_2O_3^{2-}$ ), tetrathionate ( $S_4O_6^{2-}$ ), and  
275 oxidized glutathione (GSSG) and analyzed by MALDI-ToF mass spectrometry. All reacted with  
276 polysulfide, resulting in mass increases of 32 Da, 64 Da or 96 Da, corresponding to the addition  
277 of one to three sulfur atoms (Table 1, Table S2). After incubation with tetrathionate, mass  
278 increases corresponding to addition of one or two sulfur atoms, a thiosulfonate group ( $-SSO_3^-$ ),  
279 or  $-SSSO_3^-$  were detected. With oxidized glutathione, mass increases of 305 Da were observed  
280 for TkTusA, AqTusA and HdTusA, which corresponds to covalently bound glutathione. The  
281 mobilization abilities for the bacterial TusAs AqTusA, TkTusA, TsTusA and HdTusA differed from  
282 those reported for archaeal TusA, which only mobilized thiosulfonate from tetrathionate but  
283 not sulfane sulfur from polysulfide. Like McTusA,<sup>(15)</sup> none of the studied bacterial TusA proteins  
284 were modified by thiosulfate.

285 To identify the sulfur-binding cysteine with certainty, the cysteine of the CPXP motif was  
286 replaced by serine in both, HdTusA and AqTusA, as was the C-terminal partially conserved  
287 cysteine of AqTusA. No additional peaks were observed when AqTusA Cys<sup>17</sup>Ser or HdTusA-  
288 Cys<sup>13</sup>Ser were incubated with sulfur compounds. AqTusA Cys<sup>54</sup>Ser reacted with polysulfide and  
289 tetrathionate just as wild type AqTusA. The cysteine of the CPXP motif was thus confirmed as  
290 the sulfur-binding cysteine.

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

291 **TABLE 1.** Sulfur loading of native and variant TusA and DsrE3 proteins. Numbers in parentheses  
292 represent mass increases. –, no modification. Further information is available in Table S2.

Protein	Expected masses [Da]	Detected masses [Da]	Modification
HdTusA + Polysulfide	9149	9149	-
		9180 ( $\Delta 31$ )	-S
		9208 ( $\Delta 60$ )	-S <sub>2</sub>
		9238 ( $\Delta 90$ )	-S <sub>3</sub>
HdTusA Cys <sup>13</sup> Ser + Polysulfide	9135	9135	-
HdTusA + S <sub>4</sub> O <sub>6</sub> <sup>2-</sup>	9149	9150	-
		9148 ( $\Delta 34$ )	-S
		9212 ( $\Delta 62$ )	-S <sub>2</sub>
		9262 ( $\Delta 112$ )	-S-SO <sub>3</sub>
AqTusA + Polysulfide	9610	9296 ( $\Delta 146$ )	-S <sub>2</sub> -SO <sub>3</sub>
		9612	-
		9643 ( $\Delta 31$ )	-S
		9595	-
AqTusA Cys <sup>17</sup> Ser + Polysulfide	9595	9595	-
AqTusA Cys <sup>54</sup> Ser + Polysulfide	9595	9596	-
		9627 ( $\Delta 31$ )	-S
AqTusA + S <sub>4</sub> O <sub>6</sub> <sup>2-</sup>	9610	9610	-
		9642 ( $\Delta 32$ )	-S
		9724 ( $\Delta 114$ )	-S-SO <sub>3</sub>
		9755 ( $\Delta 145$ )	-S <sub>2</sub> -SO <sub>3</sub>
AqTusA Cys <sup>17</sup> Ser + S <sub>4</sub> O <sub>6</sub> <sup>2-</sup>	9595	9595	-
AqTusA Cys <sup>54</sup> Ser + S <sub>4</sub> O <sub>6</sub> <sup>2-</sup>	9595	9597	-
		9630 ( $\Delta 33$ )	-S
		9710 ( $\Delta 113$ )	-S-SO <sub>3</sub>
HdDsrE3C + Polysulfide	15520	15524	-
		15561 ( $\Delta 38$ )	-S
		15588 ( $\Delta 64$ )	-S <sub>2</sub>
		15621 ( $\Delta 98$ )	-S <sub>3</sub>
HdDsrE3C Cys <sup>83</sup> Ser + Polysulfide	15504	15503	-
		15535 ( $\Delta 32$ )	-S
HdDsrE3C Cys <sup>84</sup> Ser + Polysulfide	15504	15503	-
HdDsrE3C + S <sub>4</sub> O <sub>6</sub> <sup>2-</sup>	15520	15517	-
		15610 ( $\Delta 93$ )	-S <sub>3</sub>
		15629 ( $\Delta 112$ )	-S-SO <sub>3</sub> <sup>-</sup>
		15658 ( $\Delta 141$ )	-S <sub>2</sub> -SO <sub>3</sub> <sup>-</sup>
HdDsrE3C + GSSG	15520	15518	-
		15824 ( $\Delta 305$ )	-SG
TkDsrE3B + Polysulfide	17201	17205	-
		17236 ( $\Delta 31$ )	-S
		17320 ( $\Delta 115$ )	-S-SO <sub>3</sub> <sup>-</sup>
TsDsrE3B + Polysulfide	16619	16619	-
		16653 ( $\Delta 34$ )	-S
		16683 ( $\Delta 64$ )	-S <sub>2</sub>
		16736 ( $\Delta 117$ )	-S <sub>2</sub> -SO <sub>3</sub> <sup>-</sup>

## 293 **2.4 Properties of sHdr-associated Rhd442**

294 The ability of rDsr associated rhodanese Rhd\_2599 to transfer sulfur to the TusA protein  
295 encoded next to its gene has already been demonstrated for *Ac. vinosum*<sup>(8)</sup> and there is no  
296 reason to doubt that closely related enzymes from other sulfur oxidizers (Figure 3b) exert the  
297 same function. All these proteins are single domain rhodaneses featuring a classical  
298 CRXGC[R/T] motif.<sup>(33)</sup> On the other hand, information about the putative single domain  
299 rhodanese Rhd442 encoded in the vicinity of hypomicrobial type I *shdr* clusters (Figure 1) is  
300 limited. Rhd442 has been described as a domain fused to sulfide:quinone oxidoreductase  
301 (SQR), e.g. in *Cupriavidus pinatubonensis* and that domain (CpRhd442) was shown to have  
302 rhodanese activity.<sup>(38)</sup> The activity depended on the cysteines residue in the CRXGXR motif that  
303 the domain shares with classical rhodanese.<sup>(33,39)</sup> A sequence reminiscent of that motif is also  
304 present in sHdr-associated Rhd442 (Figure 5a). The AlphaFold models of HdRhd442 and  
305 CpRhd442 are similar (Figure 5b) and there is also high consistency with the crystal structure  
306 of a non-classical phosphatase from *Neisseria meningitidis* (PDB 2F46<sup>(40)</sup>) (Figure 5c).  
307 Recombinant HdRhd442 catalyzed sulfur transfer from thiosulfate to cyanide with a maximum  
308 specific activity of 360 mU/mg in the assay described by Ray et al. <sup>(41)</sup> Furthermore, the protein  
309 proved able to mobilize sulfur from polysulfide as shown by characteristic mass increases  
310 (Figure 5 d,e).



## 323 2.5 Function and properties of sHdr-associated DsrE proteins

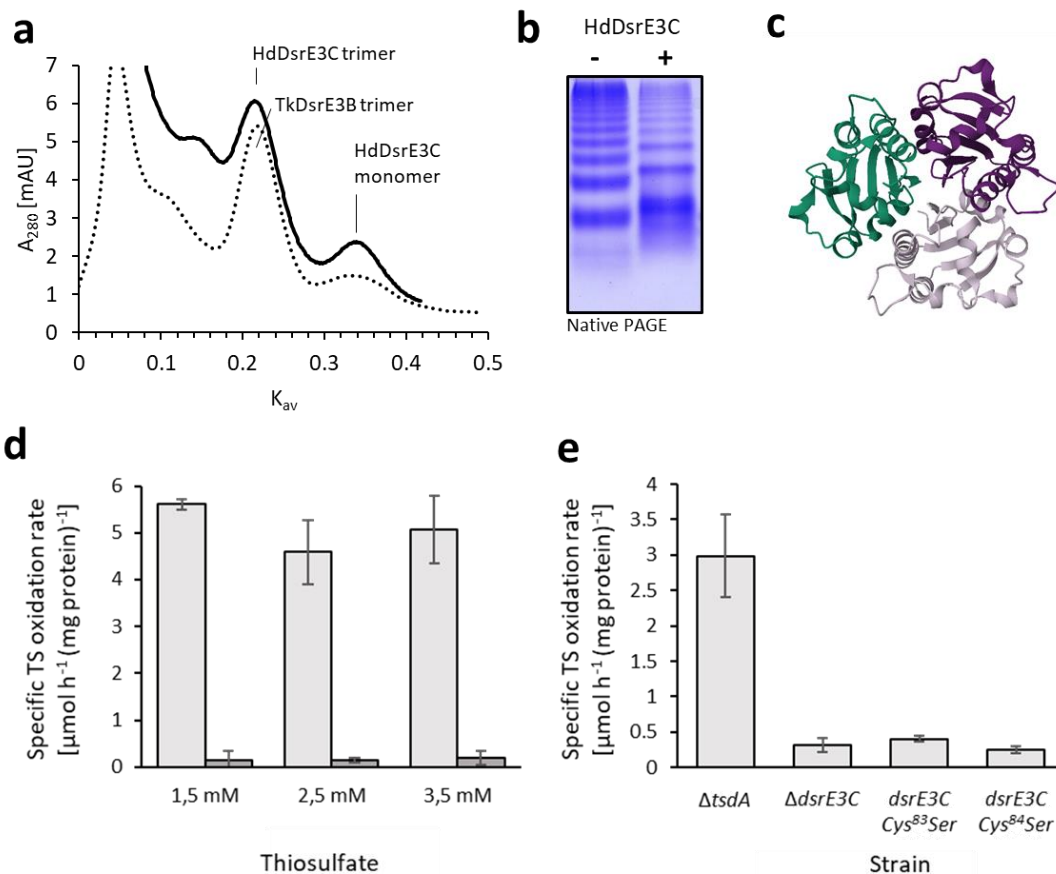
324 Analysis of recombinant DsrE3B from *Thioalkalivibrio sp.* K90mix and DsrE3C from *Hm.*  
325 *denitrificans* by gel permeation chromatography showed two peaks, one corresponding to the  
326 monomeric size and the other corresponding to a trimer (Figure 6a). Applying HdDsrE3C to  
327 native PAGE revealed a ladder of bands corresponding to higher oligomers (Figure 6b).  
328 Reduction of DsrE3C with 5 mM DTT resulted in a shift of the band pattern towards lower  
329 oligomers in relation to the non-reduced protein (Figure 6b). The structure of the trimeric  
330 complex was modelled by AlphaFold (Figure 6c). Notably, the attempt to predict a hexamer for  
331 HdDsrE3C resulted in a complex which oligomerized by protein-protein interaction at the  
332 surface of only two subunits of each trimer, leaving two subunits for further docking with other  
333 units.

334 Recombinant DsrE3C and DsrE3B had masses corresponding to the polypeptides without  
335 the N-terminal starting methionine. In all cases of DsrE3B, modified species were detected  
336 that correspond to glycosylated derivatives (+178 Da) (Table 1, Figure S3, Figure 7b). This is  
337 due to the well documented addition of glucose to the His-tag of the recombinant proteins  
338 produced in *E. coli*.<sup>(42)</sup> Sulfur mobilization assays showed reaction of HdDsrE3C with  
339 polysulfide, tetrathionate, and GSSG but not with thiosulfate. The DsrE3B proteins were  
340 persulfurated upon incubation with polysulfide. Oxidized species also occurred, probably due  
341 to the presence of oxygen. DsrE3C from *Hm. denitrificans* has an additional cysteine (Cys<sup>83</sup>)  
342 residing right next to the conserved cysteine (Cys<sup>84</sup>). To unambiguously identify the active site  
343 sulfur binding cysteine, both cysteines of HdDsrE3C were replaced with serine. While the  
344 Cys<sup>83</sup>Ser mutation did not significantly affect the sulfur binding properties of HdDsrE3C, the  
345 HdDsrE3C Cys<sup>84</sup>Ser variant was no longer persulfurated by polysulfide. In conclusion, Cys<sup>84</sup> of  
346 HdDsrE3C was identified as the sulfur-binding cysteine.

347



bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



348

349 **FIGURE 6.** Function and properties of DsrE3B and DsrE3C proteins. (a) Gel permeation  
350 chromatography of HdDsrE3C and TkDsrE3B on Hiloal 16/60 Superdex 75. (b) Native-PAGE of  
351 3.5  $\mu$ g HdDsrE3C as purified (-) and reduced with 5 mM DTT (+). (c) AlphaFold model of the  
352 HdDsrE3C trimer. (d) Specific thiosulfate (TS) oxidation rates for *Hm. denitrificans*  $\Delta$ *tsdA* (light  
353 gray columns) and *Hm. denitrificans*  $\Delta$ *tsdA*  $\Delta$ *dsrE3C* (dark gray columns) at the indicated initial  
354 thiosulfate concentrations. The corresponding growth curves are provided in Figure S1.  
355 Precultures contained 2 mM thiosulfate. (e) Specific thiosulfate (TS) oxidation rates of *Hm.*  
356 *denitrificans*  $\Delta$ *tsdA* compared to a strain lacking the complete *dsrE3C* gene and two strains  
357 carrying *dsrE3C* genes encoding the indicating cysteine to serine exchanges grown with  
358 2.5 mM thiosulfate. Note that specific thiosulfate oxidation rates are not fully comparable to  
359 the experiments shown in (d) because the growth experiments shown here were performed  
360 in a plate reader.

361 *Hm. denitrificans* is accessible to genetic manipulation and indeed, the analysis of a  
362 mutant strain lacking *dsrE3C* provided important information. The deletion was established in  
363 the *Hm. denitrificans*  $\Delta$ *tsdA* reference strain<sup>(12,43)</sup> that completely oxidizes thiosulfate via the  
364 sHdr-pathway (Figure 6d, Figure S1). Compared with the reference strain, *Hm. denitrificans*

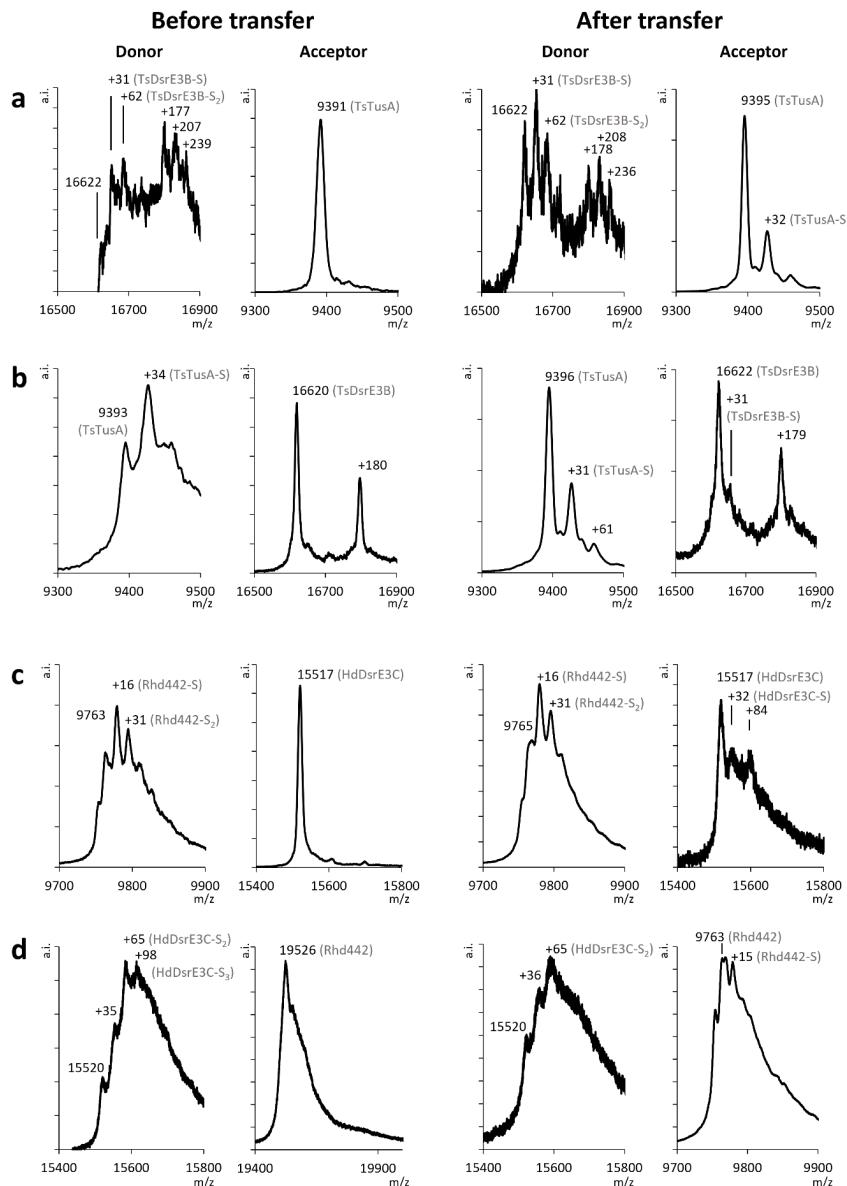
365 *ΔtsdA ΔdsrE3C* oxidized thiosulfate with a significantly decreased specific oxidation rate (Figure  
366 6d). Thus, DsrE3C is crucial for the functionality of the sHdr system. In addition, the importance  
367 of the conserved Cys<sup>84</sup> and the non-conserved Cys<sup>83</sup> of HdDsrE3C was studied *in vivo* by  
368 replacing them with serine. Both exchanges resulted in significantly decreased specific  
369 thiosulfate oxidation rates compared to the reference strain (Figure 6e). We conclude that,  
370 although it is not involved in sulfur binding, Cys<sup>83</sup> is important for DsrE3C function *in vivo*.

## 371 2.6 Interactions of sHdr-associated sulfurtransferases

372 The persulfurated DsrE3 proteins from our proteobacterial model organisms *Ts. sibirica*,  
373 *Thioalkalivibrio* sp. K90 and *Hm. denitrificans* were used as donors for the TusA proteins from  
374 the same organism, resulting in efficient sulfur transfer as shown by mass spectrometry (Figure  
375 7a, Figure S2a, Figure S3a, Table S3). In contrast, in the opposite direction, when TusA acted as  
376 donor for the DsrE3 proteins, only a small amount of sulfur was added, if any was added at all  
377 (Figure 7b, Figure S2b, Figure S3b, Table S3). These *in vitro* results point at DsrE3 proteins  
378 transferring sulfur to TusA *in vivo*, whereas the opposite direction is unfavorable. Neither  
379 recombinant HdTusA-Cys<sup>13</sup>Ser nor HdDsrE3C-Cys<sup>84</sup>Ser accepted sulfur from the native  
380 persulfurated donor proteins, whereas HdDsrE3C-Cys<sup>83</sup>Ser accepted sulfur from persulfurated  
381 HdTusA (Table S3). The Cys<sup>83</sup>Ser mutation did not significantly affect the sulfur transfer  
382 properties of HdDsrE3C, while the HdDsrE3C Cys<sup>84</sup>Ser variant was neither persulfurated by  
383 polysulfide (see above) nor by HdTusA.

384 Persulfurized HdRhd442 was tested as sulfur donor for HdTusA and HdDsrE3C. Sulfane  
385 sulfur from HdDsrE3C was efficiently transferred to HdRhd442 (Figure 7c). Transfer from  
386 HdDsrE3C to HdRhd442 was also efficient (Figure 7d). With HdTusA as the acceptor molecule  
387 for HdRhd442, the transfer was inefficient as the intensity of the signal corresponding to the  
388 sulfated species was low compared to the signal for the unmodified protein (Figure S4a).  
389 Transfer in the opposite direction was much more efficient (Figure S4b).

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



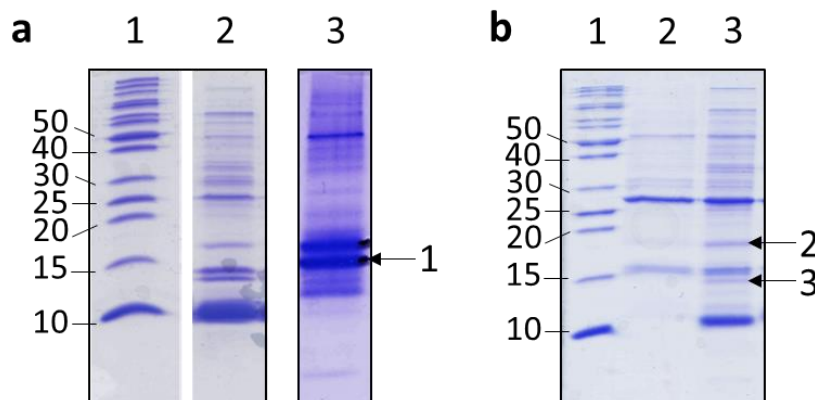
390

391 **FIGURE 7.** Sulfur transfer between TusA, DsrE3 proteins and Rhd442. Sulfur transfer reaction between  
 392 TusA and DsrE3B from *Ts. sibirica* and between Rhd442 and DsrE3C from *Hm. denitrificans* are shown  
 393 as examples. The full set of experiments is available as Figures S2, S3 and S4. (a) Left panels: TsDsrE3B  
 394 as persulfurated donor after treatment with polysulfide and unmodified reduced TsTusA as acceptor;  
 395 right panels: DsrE3B (donor) and TusA (acceptor) after the transfer reaction. The TsDsrE3B species  
 396 exhibiting additional 178 Da are due to glucosylation of the His-Tagged protein.<sup>(42)</sup> (b) Left panels:  
 397 TsTusA as persulfurated donor after treatment with polysulfide and unmodified reduced TsDsrE3B as  
 398 acceptor; right panels: TsTusA and TsDsrE3B after the transfer reaction. (c) Left panels: HdRhd442 as  
 399 persulfurated donor after treatment with polysulfide and unmodified reduced HdDsrE3C as acceptor.  
 400 Right panels: HdRhd442 (donor) and HdDsrE3C (acceptor) after the transfer reaction. (d) Left panels:  
 401 HdDsrE3C as persulfurated donor after treatment with polysulfide and unmodified reduced HdRhd442  
 402 as acceptor. Right panels: HdDsrE3C (donor) and HdRhd442 (acceptor) after the transfer reaction.

## 403 2.7 TusA and DsrE3C, DsrE3B and DsrE2 from *Aq. aeolicus* interact

404 *Aq. aeolicus* is an excellent model organism for identifying interactions between sHdr-  
405 associated sulfur transferases and other proteins. Membrane fractions, cell extracts and  
406 partially purified proteins from this hyperthermophile have repeatedly served as the basis for  
407 cross-linking, co-purification and co-migration approaches, which have enabled the  
408 identification of physiological protein partners.<sup>(16,34,44)</sup> Here, we incubated pure recombinant  
409 His-tagged AqTusA with *Aq. aeolicus* soluble extracts, prepared from cells that had been grown  
410 in the presence of various ratios of hydrogen and thiosulfate,<sup>(16)</sup> at room temperature as  
411 specified in the Material and methods section. In each case, after re-purification, AqTusA  
412 showed a mass gain of 32 Da, indicating a bound sulfane sulfur. This may be attributed to a  
413 transfer of sulfur originating from the cell extract since the heterologously produced AqTusA  
414 exhibited no such masses after purification (Table S1). In the first experiment, proteins  
415 interacting with TusA after 10 min incubation were assessed through mass spectrometry after  
416 re-purification and native PAGE (Figure 8a). This approach confirmed co-purification of TusA  
417 with AqDsrE3C (aq\_390) and AqLbpA (aq\_402) alongside 12 additional proteins (Table S4).  
418 Other sulfur transferases were not identified. In the second trial, AqTusA and *Aq. aeolicus* cell  
419 extract were incubated at room temperature for thirty minutes, followed by the identification  
420 by mass spectrometry of all proteins recovered after affinity chromatography. Cell extract  
421 without the addition of His-tagged TusA served as negative control. 253 proteins were found  
422 in the sample but not in the negative control and were therefore considered to be co-purified  
423 with AqTusA. Among them, AqDsrE3C (aq\_390) was identified with one of the highest scores  
424 (Table S5). AqDsrE2A (aq\_389) also eluted specifically in the presence of TusA but was  
425 identified with lower score and peptide number. Note that some subunits of the sHdr complex  
426 were also captured in this experiment (and not in the control, Table S5), some of which were  
427 identified with very high scores (the HdrB1 and HdrB2 subunits). All the other sHdr subunits  
428 were identified but also present (albeit with lower scores and peptide numbers) in the control  
429 sample, together with LbpA2 (aq\_402) (Table S5). To confirm the interaction of TusA with DsrE-  
430 type and LbpA proteins, samples were applied to a SDS PAGE after re-purification (Figure 8b).  
431 AqDsrE3C and AqLbpA2 were clearly detected by mass spectrometry in one band that contains  
432 also AqDsrE3B (Aq\_401) and 25 additional proteins (band 3, lane 3, Figure 8 and Table S6). In  
433 another band, AqDsrE2A (aq\_389) was found together with LbpA3 (aq\_1657), a putative thiol  
434 peroxidase (aq\_488), and 14 additional proteins (band 2, lane 3, Figure 8 and Table S6).

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

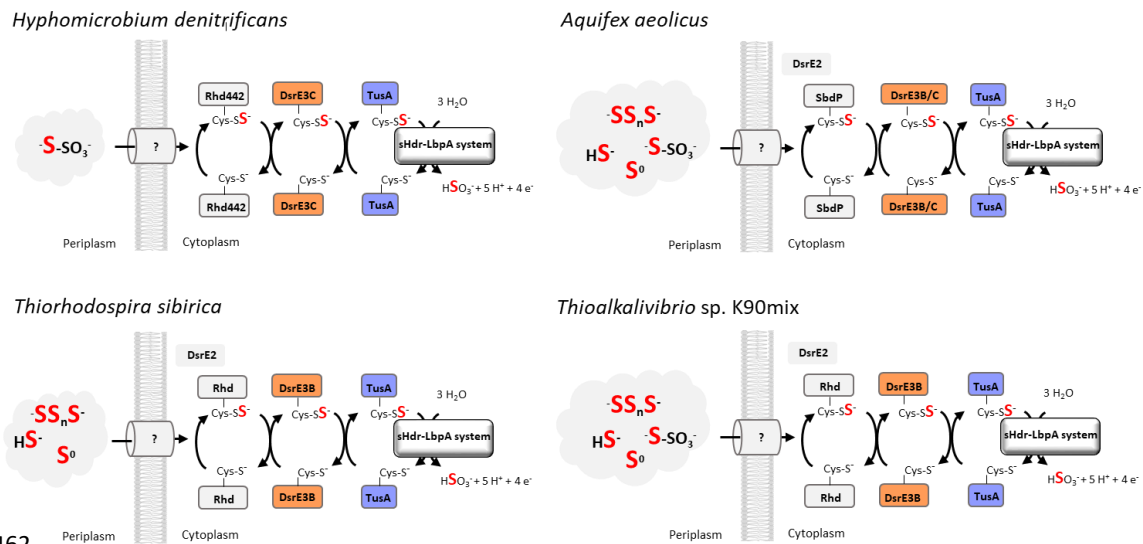


435

436 **FIGURE 8.** Re-purification experiments of AqTusA after incubated with *Aq. aeolicus* soluble  
 437 extract in different conditions. (a) Proteins, eluted from the first re-purification column, were  
 438 separated on a 18% SDS PAGE (lane 2) and a 15% native PAGE (lane 3). Six  $\mu\text{g}$  of total proteins  
 439 were loaded in lanes 2 and 3. Proteins present in the band 1, indicated by an arrow, were  
 440 identified by mass spectrometry (Table S4x). (b) Proteins, eluted from the second re-  
 441 purification column, were separated on a 18% SDS PAGE. Lane 2 corresponds to the proteins  
 442 retained in the control experiment (extract only) and lane 3 to the proteins potentially  
 443 interacting with AqTusA. Same volumes of elution fractions were loaded in lanes 2 and 3  
 444 corresponding to 2 and 10  $\mu\text{g}$  of total proteins, respectively. Proteins occurring in bands 2 and  
 445 3 were identified by mass spectrometry (Table S6). Molecular weight markers (kDa) for the SDS  
 446 gels are shown in lanes 1.

### 447 3 DISCUSSION

448 When bound to a large molecule like a protein, persulfidic sulfane sulfur can be handled  
 449 with high specificity.<sup>(4)</sup> The sulfur transferases examined in this work perfectly illustrate this  
 450 concept. Several independent lines of evidence have been combined to strongly suggest an  
 451 important function for TusA- and DsrE3-type sulfur transferases in sulfur-oxidizing prokaryotes  
 452 utilizing the sHdr pathway. First, the respective genes not only frequently occur in this  
 453 physiological group (Figure 2), but in the majority of cases they co-localize with *shdr* genes  
 454 (Figures 1 and 2), similar to what has been observed for the second well-established  
 455 cytoplasmic oxidation pathway in sulfur-dependent lithotrophs, rDsr.<sup>(8)</sup> Rhodanese Rhd442 is  
 456 present in only a small number of sulfur oxidizers and therefore appears to be of minor  
 457 importance. Second, deletion of the gene for DsrE3C from *Hm. denitrificans* resulted in a  
 458 thiosulfate oxidation-negative phenotype, emphasizing the importance of the sulfur  
 459 transferase. Third, the studied proteins all demonstrated the ability to bind sulfur upon  
 460 incubation with inorganic polysulfide and transfer it to interaction partners, therefore acting  
 461 as components of a sulfur trafficking cascade or network.



462

463 **FIGURE 9.** Sulfur relay systems in the four bacterial model organisms studied in this work. *Hm.*  
 464 *denitrificans* oxidizes thiosulfate and dimethyl sulfide. Thiosulfate is an intermediate of DMS  
 465 oxidation.<sup>(11,12)</sup> *Aq. aeolicus*<sup>(37)</sup> and *Thioalkalivibrio* sp. K90mix<sup>(45)</sup> metabolize a wide range of inorganic  
 466 sulfur substrates including thiosulfate, whereas *Ts. sibirica* is unable to oxidize thiosulfate.<sup>(46)</sup> It was  
 467 originally predicted that DsrE2A from *Aq. aeolicus* (Aq\_389) is bound to the membrane via two  
 468 transmembrane segments. However, recent deep-learning based programs like DeepTMHMM<sup>(47)</sup>  
 469 challenge this view not only for the *Aquifex* protein but also for the corresponding enzymes found in  
 470 *Ts. sibirica* and *Thioalkalivibrio* sp. K90mix.

471 The investigated DsrE3B and DsrE3C proteins from bacteria were characterized as stable  
 472 homotrimers in solution that have a tendency to assemble into higher oligomers, particularly  
 473 hexamers. This is consistent with the available data on DsrE3A proteins from archaea. The  
 474 homotrimeric form of the protein is observed in solution for the *Ms. cuprina* protein, and the  
 475 *Saccharolobus solfataricus* (SSO1125) protein crystallizes as a trimer (PDB ID 3MC3). The  
 476 DsrE2B sulfur transferases from the same archaea also form homotrimers (PDB ID 2QS7).<sup>(15)</sup> In  
 477 sulfur oxidizers utilizing the rDsr pathway, the heterohexameric DsrE<sub>2</sub>F<sub>2</sub>H<sub>2</sub> complex plays a vital  
 478 role in the sulfur relay system, which supplies the oxidizing enzyme rDsrAB. The subunits are  
 479 arranged in two stacked DsrEFH rings, each resembling the trimeric rings observed in DsrE3A  
 480 crystals (PDB ID 3MC3) and DsrE3C AlphaFold models (Figure 6c). DsrE, DsrF and DsrH are  
 481 related to each other.<sup>(10)</sup> In summary, the formation of trimers is a general and common  
 482 property of DsrE and DsrE3-type proteins, that has been conserved throughout evolution. It is  
 483 probable, that the three distinct subunits in DsrEFH have evolved due to gene duplications and  
 484 specialization of ancestral DsrE-type proteins. In fact, from an evolutionary perspective, the  
 485 DsrE3 proteins are older than the DsrE in DsrEFH.<sup>(48)</sup> Filamentation has not been reported for

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

486 the DsrEFH complex nor for any other DsrE-type; it is a novel characteristic of DsrE3C from *Hm.*  
487 *denitrificans*. As previously stated, all DsrE-like proteins possess a single conserved cysteine  
488 residue (Cys<sup>84</sup> in HdDsrE3C). The importance of the equivalent cysteine for sulfur binding and  
489 transfer has so far only been shown *in vitro* for the *Ms. cuprina* protein DsrE3A and both *in*  
490 *vivo* and *in vitro* for the distantly related DsrE from the *A. vinosum* DsrEFH complex. Here, we  
491 prove the crucial role of this residue through *in vivo* and *in vitro* studies on DsrE3C from *Hm.*  
492 *denitrificans*. Furthermore, we show that the adjacent Cys<sup>83</sup> is important for DsrE3C function  
493 *in vivo*, even though it does not participate in sulfur binding.

494 Just like archaeal McDsrE3A,<sup>(15)</sup> all DsrE3B, DsrE3C and TusA proteins from bacteria that  
495 were examined in this study displayed reaction with tetrathionate *in vitro*. However, it is  
496 impossible that thiosulfonates derived from tetrathionate are the physiologically relevant form  
497 of sulfur processed in the cytoplasm of these bacteria. Neither *Hm. denitrificans*<sup>(12)</sup> nor  
498 *Thioalkalivibrio sp.* K90mix<sup>(45)</sup>, *Aq. aeolicus*<sup>(49)</sup> or *Ts. sibirica*<sup>(46)</sup> can metabolize tetrathionate. If  
499 we seek a type of sulfur that every organism studied can use as a cytoplasmic sulfur currency,  
500 then only polysulfide or sulfane sulfur is left, as *Ts. sibirica* cannot oxidize thiosulfate either  
501 (Figure 9). In addition, HdTusA and HdDsrE3C are unable to mobilize sulfane sulfur from  
502 thiosulfate *in vitro*.

503 In dissimilatory sulfur oxidizers, the oxidation of sulfide and thiosulfate is always initiated  
504 outside of the cytoplasm.<sup>(2,6)</sup> In many cases, sulfane sulfur is then imported into the cytoplasm  
505 where it is further oxidized (Figure 9). It is not yet clear how this import is accomplished. *Hm.*  
506 *denitrificans* possesses two candidate transporters (SoxT1A and SoxT1B, Figure 1) encoded in  
507 close proximity to the genes encoding Sox proteins that are involved in the initial steps of  
508 thiosulfate oxidation and the genes for the cytoplasmic sHdr system.<sup>(43,50)</sup> However, evidence  
509 supporting the proposed sulfur transport has yet to be presented. Related transporters are not  
510 encoded in *Aq. aeolicus* and *Ts. sibirica* such that they cannot be of general importance.  
511 Rhodanese-like sulfurtransferases including Rhd442, with its comparatively narrow  
512 distribution, are potential primary sulfur acceptors and distributors in the cytoplasm. They  
513 occur in all examined organisms (Figures 1, 2, 9) and have well established sulfur transfer  
514 activity. HdRhd442 was shown here to interact with HdDsrE3C. This protein, in turn, was  
515 established as an indispensable component of the sulfur-handling cascade that feeds the type  
516 I sHdr system of *Hm. denitrificans*. *In vitro*, both HdDsrE3C and the DsrE3B proteins from  
517 gammaproteobacterial model organisms shuttle sulfane sulfur to TusA (Figures 7, 9). Efficient

518 sulfur transfer to TusA from the same organism was observed with HdDsrE3C, TkDsrE3B and  
 519 TsDsrE3B, while transfer in the opposite direction was either undetectable or comparatively  
 520 inefficient. Unidirectional transfer has also been observed for DsrE3A and TusA from *Ms.*  
 521 *cuprina*, where a thiosulfonate group is moved from DsrE3A to TusA but not vice versa.<sup>(15)</sup> In  
 522 the absence of a DsrE3C homolog, as in *Thioalkalivibrio* sp. K90mix, the sulfurtransferase  
 523 DsrE3B may functionally substitute for DsrE3C. It is very important to note that each sulfur  
 524 transferase may interact with multiple partners as illustrated by TusA from *Aq. aeolicus* which  
 525 was purified along with three DsrE, two AqLbpA and several sHdr proteins. This finding is  
 526 particularly significant as it establishes a direct link between sulfur transferases and the sHdr-  
 527 LbpA complex where sulfane sulfur is oxidized to sulfite.<sup>(11,12,18)</sup>

## 528 4 CONCLUSIONS

529 We conclude that cytoplasmic sHdr systems for sulfane sulfur oxidation are always  
 530 accompanied by sets of sulfur transferases. The exact composition of these sets may vary  
 531 (Figure 9). *In vivo*, a strict unidirectional transfer of sulfur between the components is unlikely.  
 532 Rather, it can be assumed that a network of sulfur-binding proteins exists, each with a pool of  
 533 bound sulfur. Sulfur flux can be shifted in one direction or the other depending on the  
 534 metabolic requirements. A single pair of sulfur-binding proteins with a preferred transfer  
 535 direction, such as a DsrE3-type protein toward TusA, may well be sufficient to push sulfur into  
 536 the sink where it is further metabolized or needed. Multiple possible interactions are most  
 537 easily exemplified for the TusA protein. In organisms such as *Hm. denitrificans*, which differ  
 538 from *E. coli* by containing just one *tusA* gene, the same protein must likely fulfill a variety of  
 539 functions, including providing substrate for sulfur oxidation, and supplying sulfur for tRNA  
 540 thiolation and biosynthesis of cofactors and Fe/S clusters.

## 541 5 MATERIALS AND METHODS

### 542 5.1 Bacterial strains, plasmids, primers, and growth conditions

543 Table S7 lists the bacterial strains, and plasmids that were used for this study. *Escherichia*  
 544 *coli* strains were grown on complex lysogeny broth (LB) medium<sup>(51)</sup> under aerobic conditions  
 545 at 37°C unless otherwise indicated. *E. coli* 10β was used for molecular cloning. *E. coli* BL21  
 546 (DE3) was used for recombinant protein production. *Hm. denitrificans* strains were cultivated  
 547 in minimal media kept at pH 7.2 with 24.4 mM methanol and 100 mM 3-(*N*-Morpholino)  
 548 propanesulfonic acid (MOPS) buffer as described before.<sup>(12,43)</sup> Thiosulfate was added as



549 needed. Antibiotics for *E. coli* and *Hm. denitrificans* were used at the following concentrations  
550 (in  $\mu\text{g ml}^{-1}$ ): ampicillin, 100; kanamycin, 50; streptomycin, 200; chloramphenicol, 25.

## 551 **5.2 Recombinant DNA techniques**

552 Standard techniques for DNA manipulation and cloning were used unless otherwise  
553 indicated.<sup>(52)</sup> Restriction enzymes, T4 ligase and Q5 polymerase were obtained from New  
554 England Biolabs (Ipswich, UK) and used according to the manufacturer's instructions.  
555 Oligonucleotides for cloning were obtained from Eurofins MWG (Ebersberg, Germany).  
556 Plasmid DNA from *E. coli* was purified using the GenJET Plasmid Miniprep kit (Thermo  
557 Scientific, Waltham, USA). Chromosomal DNA from *Hm. denitrificans*, *Ts. sibirica* and  
558 *Thioalkalivibrio* strains was prepared using the Simplex easy DNA kit (GEN-IAL GmbH, Troisdorf,  
559 Germany). Total DNA from 20 mg of *Aq. aeolicus* cells was extracted using the phenol-  
560 chloroform extraction method. Cells were resuspended in 150  $\mu\text{l}$  of TEN buffer (10 mM Tris  
561 pH8, 150 mM NaCl, 10 mM EDTA) and 300  $\mu\text{l}$  of SDS-EB (100 mM Tris pH8, 400 mM NaCl, 40  
562 mM EDTA, 2% SDS). 2  $\mu\text{l}$  of RNase A solution (4 mg  $\text{ml}^{-1}$ ) (Promega) were added before  
563 incubation at 37°C for 15 min. 350  $\mu\text{l}$  phenol/chloroform-isoamylalcohol (25:24:1 mixture,  
564 Biosolve) were added and the mixture was vortexed for 30 s and spun for 3 min at 14,000 $\times$ g.  
565 The upper phase was transferred to a new tube with 300  $\mu\text{l}$  of chloroform-isoamylalcohol,  
566 vortexed and spun again. The upper phase was then incubated with 600  $\mu\text{l}$  isopropanol for 30  
567 min at -25°C and spun at 4°C for 30 min. The DNA pellet was washed with 70% ethanol, left to  
568 air dry and dissolved in 100  $\mu\text{l}$  H<sub>2</sub>O for 30 min at 65°C. 1  $\mu\text{l}$  was used for 25  $\mu\text{l}$  PCR reaction.

## 569 **5.3 Construction of *Hm. denitrificans* mutant strains**

570 For markerless deletion of the *Hm. denitrificans dsrE3C* (Hden\_0688) gene by splicing  
571 overlap extension (SOE) (Horton, 1995), PCR fragments were constructed using the primers  
572 Hden0688\_Up\_Fw, Hden0688\_Up\_Rev, Hden0688\_Down\_Fw, Hden0688\_Down\_Rev (Table  
573 S1). The resulting 2.08 kb SOE PCR fragment was cloned into the XbaI and PstI sites of  
574 pK18mobscaB-Tc. The final construct pK18mobsacB\_Tc\_Δ*dsrE3C* was electroporated into *H.*  
575 *denitrificans* Δ*tsdA* and transformants were selected using previously published procedures  
576 <sup>(11,12)</sup>. Single crossover recombinants were Cm<sup>r</sup> and Tc<sup>r</sup>. Double crossover recombinants were  
577 Tc<sup>s</sup> and survived in the presence of sucrose due to loss of both, the vector-encoded  
578 levansucrase (SacB) and the tetracycline resistance gene. For chromosomal integration of the  
579 genes encoding DsrE3C Cys<sup>83</sup>Ser and DsrE3C Cys<sup>84</sup>Ser, the modified genes and upstream as  
580 well as downstream sequences were amplified by SOE PCR using primers Hden0688\_Up\_Fw,

581 Hden0688\_Down\_Rev, Hden0688\_C83S\_Fw, Hden0688\_C83S\_Rev, and Hden0688\_Up\_Fw,  
582 Hden0688\_Down\_Rev, Hden0688\_C84S\_Fw, Hden0688\_C84S\_Rev (Table S1), respectively.  
583 The final plasmids *pk18mobsacB-dsrE3C-C83S-Tc* and *pk18mobsacB-dsrE3C-C84S-Tc* were  
584 transferred into *Hm. denitrificans*  $\Delta$ *tsdA*  $\Delta$ *dsrE3C* and double crossover recombinants were  
585 selected as described previously.<sup>(12)</sup> The genotypes of the *Hm. denitrificans* mutant strains  
586 generated in this study were confirmed by PCR.

#### 587 **5.4 Characterization of phenotypes and quantification of sulfur compounds**

588 Growth experiments with *Hm. denitrificans* were run in Erlenmeyer flasks with media  
589 containing 24.4 mM methanol and varying concentrations of thiosulfate as necessary.<sup>(43)</sup>  
590 Thiosulfate and sulfite concentrations and biomass content were determined by previously  
591 described methods.<sup>(43,53)</sup> All growth experiments were repeated three to five times.  
592 Representative experiments with two biological replicates for each strain are shown. All  
593 quantifications are based on at least three technical replicates. Alternatively, growth  
594 experiments were run in 48-well microtiter plates. Plates were continuously shaken at 200 rpm  
595 and growth was followed by measuring optical density at 600 nm every 5 min using an Infinite  
596 200Pro (Tecan, Crailsheim, Germany) plate reader. Samples for thiosulfate determination were  
597 taken as previously described.<sup>(43)</sup>

#### 598 **5.5 Cloning, site-directed mutagenesis, overproduction, purification and size exclusion 599 chromatography of recombinant proteins**

600 The 378-bp *dsrE3C* gene was amplified from *Hm. denitrificans* genomic DNA with the  
601 primers Hden0688 (*dsrE3C*)\_NdeI\_fw and Hden0688 (*dsrE3C*)\_BamHI\_rev (Table S1) and  
602 cloned between the NdeI and BamHI sites of pET15b(+), resulting in pET15b-Hd-DsrE3C.  
603 Analogous procedures were followed for *dsrE3B* from *Thioalkalivibrio* sp. K90mix which was  
604 cloned between the NdeI and BamHI sites and *dsrE3B* from *Ts. sibirica* which was cloned  
605 between the XhoI and NdeI sites generating the plasmids pET15b-TkDsrE3B and pET15b-  
606 TsDsrE3B. The *tusA* genes were amplified from *Hm. denitrificans*, *Thioalkalivibrio* sp. K90mix  
607 or *Ts. sibirica* genomic DNA with primers adding a sequence for a C-terminal Strep-tag and  
608 cloned between the NdeI and EcoRI sites of pET-22b(+). The *tusA* gene from *Aq. aeolicus*  
609 (aq\_388a, coding for the protein WP\_024015099.1) was amplified from genomic DNA with  
610 primers Aq388a\_NdeI\_fw and Aq388a\_XhoI\_rev (Table S1) introducing at the C-terminal  
611 position of the protein the two amino acids Leu and Glu directly followed by a 6His-tag, and  
612 cloned into the pET24a expression plasmid to generate the plasmid pET24a-AqTusA. Cysteine

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

613 to serine exchanges were implemented to HdDsrE3C by SOE PCR using primers sets  
614 Hden0688\_NdeI\_fw and Hden0688\_BamHI\_rev, Hden0688\_C83S\_Fw, Hden0688\_C83S\_Rev  
615 and Hden0688\_NdeI\_fw and Hden0688\_BamHI\_rev, Hden0688\_C84S\_Fw,  
616 Hden0688\_C84S\_Rev, respectively, resulting in plasmids pET15b-DsrE3C-C83S and pET15b-  
617 DsrE3C-C84S. A cysteine to serine exchange was introduced into *Hm. denitrificans* TusA using  
618 the same method and the primers listed in Table S7. The QuikChange site-directed  
619 mutagenesis Kit (Stratagene) was used to generate the *Aq. aeolicus* *tusA* mutated genes  
620 (coding for AqTusA Cys<sup>17</sup>Ser and AqTusA Cys<sup>54</sup>Ser) with the primers Aq388a\_C17S\_fw,  
621 Aq388a\_C17S\_rev and Aq388a\_C54S\_fw, Aq388a\_C54S\_rev using the pET24a-AqTusA  
622 plasmid.

623 Recombinant DsrE3B, DsrE3C and TusA proteins were produced in *E. coli* BL21(DE3).  
624 Overnight precultures were used to inoculate fresh LB medium with a ratio of 1:50 (v/v).  
625 Synthesis of recombinant proteins was induced by the addition of 0.1 or 1mM (for AqTusA)  
626 IPTG when cultures had reached an OD<sub>600</sub> of 0.6–0.8, followed by incubation for 2.5 h at 37  
627 °C. Cells were harvested by centrifugation (11,000 × g, 20 min, 4°C). Strep-tagged proteins  
628 were resuspended in 50 mM Tris-HCl buffer (pH 7.5) containing 150 mM NaCl. His-tagged  
629 proteins were resuspended in buffer containing 20 mM sodium-phosphate (20 mM Tris-HCl  
630 for AqTusA), 500 mM NaCl and 50 mM imidazole (pH 7.4). Cells were lysed by sonication (or  
631 with a cell disruptor for AqTusA). Insoluble cell material was subsequently removed by  
632 centrifugation (16,100 × g, 30 min, 4°C). His-tagged and Strep-tagged proteins were purified  
633 with Ni-NTA Agarose (Jena Bioscience) (except for AqTuA that was purified with a HiScreen Ni  
634 FF column (Cytiva)) and Strep-Tactin Superflow (IBA Lifesciences, Göttingen, Germany),  
635 respectively, according to the manufacturer's instructions. The proteins were then transferred  
636 to salt-free 50 mM Tris-HCl buffer (pH 7.5) or to 20 mM Tris-HCl buffer pH 7.4 for AqTusA and  
637 stored at -70°C. Size exclusion chromatography on HiLoad 16/60 Superdex™ 75 (Cytiva,  
638 Freiburg, Germany) was performed as described in Li et al.<sup>(50)</sup>

### 639 5.6 Protein-protein interaction in cell extracts

640 To detect interaction between AqTusA and DsrE-like proteins, TusA from *Aq. aeolicus*  
641 (200 µg) was incubated with crude soluble extract of *Aq. aeolicus* at room temperature and  
642 re-purified via His tag affinity-chromatography (HisTrap FF 1 ml column, Cytiva) according to  
643 the manufacturer's instructions. Soluble extracts were prepared as previously described<sup>(49)</sup>  
644 except that cells were resuspended in 50 mM Tris-HCl pH 7.3. Two different experiments were

645 run: in the first one, purified AqTusA was incubated for 10 min with 500  $\mu$ l of an extract  
646 obtained from cells grown with excess of hydrogen in the presence of thiosulfate (called 100%  
647 H<sub>2</sub>) as described previously;<sup>(16)</sup> in the second trial, the protein was incubated for 30 min with  
648 2 ml of extract prepared from cells grown with a lower amount of H<sub>2</sub> in the presence of  
649 thiosulfate (referred as 30% H<sub>2</sub>, condition in which the Hdr amount in cells is higher<sup>(16)</sup>). In both  
650 cases, the AqTusA-extract mixtures were immediately frozen in liquid nitrogen after incubation  
651 and thawed just prior to purification. Proteins were eluted with a buffer containing 50 mM  
652 Tris-HCl pH 7.3, 150 mM NaCl and 250 mM imidazole. A control experiment was run with the  
653 same extract but without TusA. After dialysis on a Vivaspin concentrator (molecular mass  
654 cutoff 3000 Da) with 50 mM Tris HCl pH 7.3, eluted proteins from both columns underwent  
655 18% SDS PAGE or 15% Tris-Glycine native PAGE (without any reducing agent).<sup>(54)</sup> Resulting  
656 bands were cut out of the gel and analyzed by LC-MSMS as described previously.<sup>(16)</sup> For the  
657 second experiment, total proteins in the elution fraction of the column were identified after  
658 the proteins were introduced in a 5% acrylamide stacking gel, as described in Prioretti *et al.*<sup>(44)</sup>  
659 under the name “stacking method”. Protein concentrations were determined with the BCA  
660 protein assay kit from Sigma-Aldrich.

### 661 5.7 Sulfur binding and transfer experiments

662 For sulfur binding experiments with the recombinant sulfur transferases from *Hm.*  
663 *denitrificans*, *Thioalkalivibrio* sp. K90mix and *Ts. sibirica*, 1.5 nmol of the proteins were  
664 incubated with 5 mM polysulfide, thiosulfate, tetrathionate or GSSG in 20  $\mu$ l 50 mM Tris-HCl  
665 pH 7.5. The polysulfide stock solution needed for these experiments was prepared, diluted and  
666 used as previously described.<sup>(50,55)</sup> For sulfur transfer experiments, 1.5 nmol of the putative  
667 sulfur-donating protein were incubated with 0.5 mM polysulfide for 30 minutes at room  
668 temperature in 20  $\mu$ l 50 mM Tris-HCl pH 7.5. Acceptor proteins were reduced with 1 mM DTT  
669 under the same conditions. Excess of polysulfide or DTT was removed with micro Bio-Spin 6  
670 columns (GE Healthcare, Munich, Germany) equilibrated with 50 mM Tris-HCl pH 7.5. Efficient  
671 removal of polysulfide was confirmed by adding a filtered protein-free polysulfide solution to  
672 an acceptor protein, which was then tested for persulfuration via mass spectrometry. Donor  
673 and acceptor proteins were mixed in a 1:1 ratio to a final volume of 40  $\mu$ l and incubated for 30  
674 min at room temperature. The mixtures were then stored at -70°C. For mass spectrometry,  
675 samples of 20  $\mu$ l were desalted by ZipTipC4 Pipette tips (Merck Millipore, Darmstadt,  
676 Germany), crystallized in a 2',6'-dihydroxyacetophenon matrix and measured by MALDI-ToF

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

677 (matrix-assisted laser desorption ionization-time-of-flight) mass spectrometry at the Core  
678 Facility Protein synthesis & BioAnalytics, Pharmaceutical Institute, University of Bonn as  
679 previously described.<sup>(54)</sup>

680 For sulfur loading of AqTusA, 2 µg of protein was incubated for 30 min at 65°C, with  
681 5 mM (final concentration) of the tested sulfur compound, in a final volume of 5 µl in 50 mM  
682 Tris-HCl pH 7.3. If reduction was desired, samples were incubated, with 10 mM DTT (final  
683 concentration) for 45 min at room temperature after the sulfur loading. Samples were  
684 immediately frozen in liquid nitrogen and stored at -80°C until further use. Before MALDI-ToF  
685 mass spectrometry analysis, the samples were desalted and concentrated with ZipTip C18  
686 (Merck Millipore) using 0.1% trifluoroacetic acid (TFA) as desalting solution and 70%  
687 acetonitrile/0.1% TFA as elution solution. 1 µl of samples (~41 pmol) mixed with 1 µl of matrix  
688 α-cyano-4-hydroxycinnamic acid were analyzed using the mass spectrometer Microflex II  
689 (Bruker). Three µg of AqTusA, incubated with cell extracts and re-purified on a HisTrap column  
690 (see paragraph 5.6), was diluted in 50 mM Tris-HCl pH 7.3 in a volume of 5 µl, desalted with  
691 ZipTipC18 and analyzed by MALDI-TOF.

## 692 **5.8 Generation of datasets for phylogenetic and similarity network analyses**

693 Archaeal and bacterial genomes were downloaded from Genome Taxonomy Database  
694 (GTDB, release R207). In GTDB, all genomes are sorted according to validly published  
695 taxonomies, they are pre-validated and have high quality (completeness minus  
696 5\*contamination must be higher than 50%). One representative of each of the current 65,703  
697 species clusters was analyzed. Open reading frames were determined using Prodigal<sup>(56)</sup> and  
698 subsequently annotated for sulfur related proteins via HMSS2.<sup>(21)</sup> Annotation was extended by  
699 HMMs from TIGRFAMs<sup>(57)</sup> and Pfam<sup>(58)</sup> databases representing the 16 syntenic ribosomal  
700 proteins Rpl2, 3, 4, 5, 6, 14, 15, 16, 18, 22, and 24, and Rps3, 8, 10, 17, and 19. A type I sHdr  
701 system was considered to be present if the core genes *shdrC1B1AHC2B2* were present in a  
702 syntenic gene cluster. For a type II sHdr system gene cluster *shdrC1B1AHB3* and *etfAB* had to  
703 be present in a single syntenic gene cluster.<sup>(11,14)</sup>

## 704 **5.9 Phylogenetic tree inference and structural modelling**

705 For phylogenetic tree inference, proteins were aligned using MAFFT<sup>(59)</sup> and trimmed with  
706 BMGE<sup>(60)</sup> (entropy threshold = 0.95, minimum length = 1, matrix = BLOSUM30). Alignments  
707 were then used for maximum likelihood phylogeny inference using IQ-TREE v1.6.12<sup>(61)</sup>  
708 implemented on the “bonna” high performance clusters of the University of Bonn. The best-

709 fitting model of sequence evolution was selected using ModelFinder.<sup>(62)</sup> Branch support was  
710 then calculated by SH-aLRT (2000 replicates),<sup>(63)</sup> aBayes (2000 replicates)<sup>(64)</sup> and ultrafast  
711 bootstrap (2000 replicates).<sup>(65)</sup> Finally, trees were displayed using iTol.<sup>(66)</sup> For species tree  
712 inference, results for each ribosomal protein were individually aligned, trimmed and  
713 subsequently concatenated before they were used for phylogenetic tree construction.  
714 Structural models of proteins and protein complexes were generated using AlphaFold2.<sup>(67)</sup>

#### 715 **5.10 Sequence similarity network analysis**

716 Amino acid sequences for the sequence similarity network were derived from the search  
717 of the GTDB dataset with HMSS2.<sup>(21)</sup> Groups were chosen based on their sequence similarity  
718 and genomic context, which depended on the specific question being investigated. From all  
719 selected sequences a meaningful and diverse set was derived via dereplication with mmseqs2  
720 linclust<sup>(68)</sup> with default settings. Sequence similarity analysis was performed by an all versus all  
721 comparison with mmseqs2 search.<sup>(68,69)</sup> The similarity matrix was modified in cytoscape and  
722 edged were filtered stepwise until optimal clustering was observed. This status was  
723 characterized as the minimum number of clusters while maintaining the highest number of  
724 connectivity within a group of annotated proteins, visualizing the grouping of proteins on a  
725 deep branching level.

#### 726 **ACKNOWLEDGEMENTS**

727 This work was funded by the Deutsche Forschungsgemeinschaft (Grant Da 351/13-1).  
728 Tomohisa Sebastian Tanabe received a scholarship from the Studienstiftung des Deutschen  
729 Volkes. We gratefully acknowledge the access to the Bonna HPC cluster hosted by the  
730 University of Bonn along with the support provided by its High Performance Computing &  
731 Analytics Lab. We gratefully thank for the support of France-Germany Hubert Curien Procope  
732 program for its exchange funding (n° 40444VM). We gratefully acknowledge the support of the  
733 Core Facility “Protein Synthesis and Bioanalytics” of the University of Bonn. The authors are  
734 also grateful to Régine Lebrun from the Core Facility “Protein Synthesis Mass spectrometry  
735 and Bioanalytics” proteomic platform of the University Mediterranean Institute of  
736 Microbiology (IMM) in Marseille (France), which is a part of the network “Marseille  
737 Protéomique”, IBISA, and the Proteomic Platform IBISA of the IMM, CNRS, Aix-Marseille  
738 Université for performing mass spectrometry. We thank Souh la Boughanemi and Robert van

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

739 Lis (BIP-CNRS, Marseille) for *Aq. aeolicus tusA* cloning and site directed mutagenesis. We thank  
740 Alina Ballas for help with *Hm. denitrificans* growth experiments.

#### 741 **CONFLICT OF INTEREST**

742 The authors declare no competing interests.

#### 743 **ORCID**

744 *Christiane Dahl*  0000-0001-8288-7546

745 *Tomohisa Sebastian Tanabe*  0000-0003-2154-7980

#### 746 **SUPPLEMENTARY MATERIAL**

747 Additional supporting information may be found online in the supporting Information section  
748 at the end of this article

#### 749 **REFERENCES**

- 750 1. Tanabe TS, Leimkühler S, Dahl C (2019) The functional diversity of the prokaryotic sulfur carrier  
751 protein TusA. *Adv Microb Physiol* 75:233-277.
- 752 2. Dahl C (2015) Cytoplasmic sulfur trafficking in sulfur-oxidizing prokaryotes. *IUBMB Life* 67:268-274.
- 753 3. Mueller EG (2006) Trafficking in persulfides: delivering sulfur in biosynthetic pathways. *Nat Chem*  
754 *Biol* 2:185-194.
- 755 4. Kessler D (2006) Enzymatic activation of sulfur for incorporation into biomolecules in prokaryotes.  
756 *FEMS Microbiol Rev* 30:825-840.
- 757 5. Gojon G (2020) On H<sub>2</sub>S prodrugs. *Antioxid Redox Signal* 33:999-1002.
- 758 6. Dahl C. A biochemical view on the biological sulfur cycle. In: Lens P, Ed. (2020) *Environmental*  
759 *technologies to treat sulfur pollution: principles and engineering*. IWA Publishing, London, pp. 55-  
760 96.
- 761 7. -. Sulfur metabolism in phototrophic bacteria. In: Hallenbeck PC, Ed. (2017) *Modern topics in the*  
762 *phototrophic prokaryotes: Metabolism, bioenergetics and omics*. Springer International Publishing,  
763 Cham, pp. 27-66.
- 764 8. Stockdreher Y, Sturm M, Josten M, Sahl HG, Dobler N, Zigann R, Dahl C (2014) New proteins involved  
765 in sulfur trafficking in the cytoplasm of *Allochrochromatium vinosum*. *J Biol Chem* 289:12390-12403.
- 766 9. Stockdreher Y, Venceslau SS, Josten M, Sahl HG, Pereira IAC, Dahl C (2012) Cytoplasmic  
767 sulfurtransferases in the purple sulfur bacterium *Allochrochromatium vinosum*: evidence for sulfur  
768 transfer from DsrEFH to DsrC. *PLoS One* 7:e40785.
- 769 10. Dahl C, Schulte A, Stockdreher Y, Hong C, Grimm F, Sander J, Kim R, Kim SH, Shin DH (2008) Structural  
770 and molecular genetic insight into a wide-spread bacterial sulfur oxidation pathway. *J Mol Biol*  
771 384:1287-1300.
- 772 11. Cao X, Koch T, Steffens L, Finkensieper J, Zigann R, Cronan JE, Dahl C (2018) Lipoate-binding proteins  
773 and specific lipoate-protein ligases in microbial sulfur oxidation reveal an atypical role for an old  
774 cofactor. *eLife* 7:e37439.
- 775 12. Koch T, Dahl C (2018) A novel bacterial sulfur oxidation pathway provides a new link between the  
776 cycles of organic and inorganic sulfur compounds. *ISME J* 12:2479-2491.
- 777 13. Tanabe TS, Grosser M, Hahn L, Kümpel C, Hartenfels H, Vtulkin E, Flegler W, Dahl C (2023)  
778 Identification of a novel lipoic acid biosynthesis pathway reveals the complex evolution of lipoate  
779 assembly in prokaryotes. *PLoS Biol* 21:e3002177.

- 780 14. Justice NB, Norman A, Brown CT, Singh A, Thomas BC, Banfield JF (2014) Comparison of  
781 environmental and isolate *Sulfobacillus* genomes reveals diverse carbon, sulfur, nitrogen, and  
782 hydrogen metabolisms. *BMC Genomics* 15:1107.
- 783 15. Liu LJ, Stockdreher Y, Koch T, Sun ST, Fan Z, Josten M, Sahl HG, Wang Q, Luo YM, Liu SJ, Dahl C, Jiang  
784 CY (2014) Thiosulfate transfer mediated by DsrE/TusA homologs from acidothermophilic sulfur-  
785 oxidizing archaeon *Metallosphaera cuprina*. *J Biol Chem* 289:26949-26959.
- 786 16. Boughanemi S, Lyonnet J, Infossi P, Bauzan M, Kosta A, Lignon S, Giudici-Ortoni MT, Guiral M  
787 (2016) Microbial oxidative sulfur metabolism: biochemical evidence of the membrane-bound  
788 heterodisulfide reductase-like complex of the bacterium *Aquifex aeolicus*. *FEMS Microbiol Lett*  
789 363:fnw156.
- 790 17. Kumpel C, Grosser M, Tanabe TS, Dahl C (2023) Fe/S proteins in microbial sulfur oxidation.  
791 *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* submitted.
- 792 18. Ernst C, Kayashta K, Koch T, Venceslau SS, Pereira IAC, Demmer U, Ermler U, Dahl C (2021) Structural  
793 and spectroscopic characterization of a HdrA-like subunit from *Hyphomicrobium denitrificans*. *FEBS*  
794 *J* 288:1664-1678.
- 795 19. Jaffe AL, Castelle CJ, Matheus Carnevali PB, Gribaldo S, Banfield JF (2020) The rise of diversity in  
796 metabolic platforms across the Candidate Phyla Radiation. *BMC Biology* 18:69.
- 797 20. Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, Williams KH, Tringe SG,  
798 Banfield JF (2013) Community genomic analyses constrain the distribution of metabolic traits across  
799 the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* 1:22.
- 800 21. Tanabe TS, Dahl C (2023) HMSS2: an advanced tool for the analysis of sulfur metabolism, including  
801 organosulfur compound transformation, in genome and metagenome assemblies. *Mol Ecol Resour*  
802 23:1930-1945.
- 803 22. Norris PR, Clark DA, Owen JP, Waterhouse S (1996) Characteristics of *Sulfobacillus acidophilus* sp.  
804 nov. and other moderately thermophilic mineral-sulphide-oxidizing bacteria. *Microbiology* 142:775-  
805 783.
- 806 23. Zhang X, Liu X, Liang Y, Guo X, Xiao Y, Ma L, Miao B, Liu H, Peng D, Huang W, Zhang Y, Yin H (2017)  
807 Adaptive evolution of extreme acidophile *Sulfobacillus thermosulfidooxidans* potentially driven by  
808 horizontal gene transfer and gene loss. *Appl Environ Microbiol* 83:e03098-03016.
- 809 24. Watanabe T, Kojima H, Umezawa K, Hori C, Takasuka TE, Kato Y, Fukui M (2019) Genomes of  
810 neutrophilic sulfur-oxidizing chemolithoautotrophs representing 9 proteobacterial species from 8  
811 genera. *Front Microbiol* 10:316.
- 812 25. Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC (2009) Using sequence similarity networks for  
813 visualization of relationships across diverse protein superfamilies. *PLoS One* 4:e4345.
- 814 26. Dahl JU, Radon C, Buhning M, Nimtze M, Leichert LI, Denis Y, Jourlin-Castelli C, Iobbi-Nivol C, Mejean  
815 V, Leimkühler S (2013) The sulfur carrier protein TusA has a pleiotropic role in *Escherichia coli* that  
816 also affects molybdenum cofactor biosynthesis. *J Biol Chem* 288:5426-5442.
- 817 27. Ikeuchi Y, Shigi N, Kato J, Nishimura A, Suzuki T (2006) Mechanistic insights into sulfur relay by  
818 multiple sulfur mediators involved in thioridine biosynthesis at tRNA wobble positions. *Molecular*  
819 *Cell* 21:97-108.
- 820 28. Ishii Y, Yamada H, Yamashino T, Ohashi K, Katoh E, Shindo H, Yamazaki T, Mizuno T (2000) Deletion  
821 of the *yhhP* gene results in filamentous cell morphology in *Escherichia coli*. *Bioscience*  
822 *Biotechnology and Biochemistry* 64:799-807.
- 823 29. Yildiz T, Leimkühler S (2021) TusA is a versatile protein that links translation efficiency to cell division  
824 in *Escherichia coli*. *J Bacteriol* 203.
- 825 30. Tanaka Y, Yoshikaie K, Takeuchi A, Ichikawa M, Mori T, Uchino S, Sugano Y, Hakoshima T, Takagi H,  
826 Nonaka G, Tsukazaki T (2020) Crystal structure of a YeeE/YedE family protein engaged in thiosulfate  
827 uptake. *Science Advances* 6:eaba7637.
- 828 31. Balleste-Delpierre C, Fernandez-Orth D, Ferrer-Navarro M, Diaz-Pena R, Odena-Caballol A, Oliveira  
829 E, Fabrega A, Vila J (2017) First insights into the pleiotropic role of *vrf* (*yedF*), a newly characterized  
830 gene of *Salmonella Typhimurium*. *Sci Rep* 7:15291.



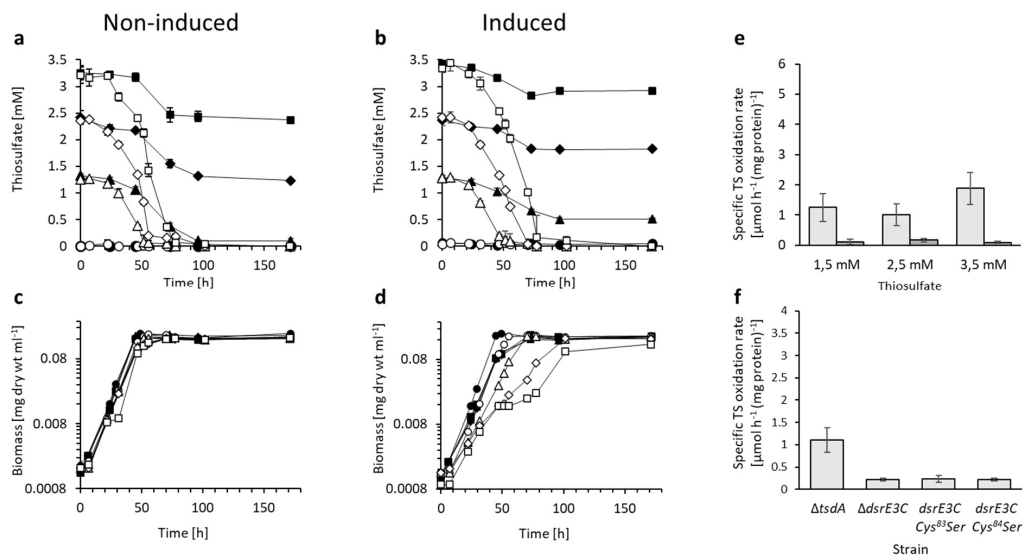
bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

- 831 32. Westley J (1973) Rhodanese. *Advances in Enzymology and Related Areas of Molecular Biology*  
832 39:327-368.
- 833 33. Bordo D, Bork P (2002) The rhodanese/Cdc25 phosphatase superfamily - Sequence-structure-  
834 function relations. *EMBO Rep* 3:741-746.
- 835 34. Aussignargues C, Giuliani MC, Infossi P, Lojou E, Guiral M, Giudici-Orticoni MT, Ilbert M (2012)  
836 Rhodanese functions as sulfur supplier for key enzymes in sulfur energy metabolism. *J Biol Chem*  
837 287:19936-19948.
- 838 35. Tanabe TS, Dahl C (2022) HMS-S-S: a tool for the identification of sulphur metabolism-related genes  
839 and analysis of operon structures in genome and metagenome assemblies. *Mol Ecol Resour*  
840 22:2758-2774.
- 841 36. Shi R, Proteau A, Villarroya M, Moukadiri I, Zhang LH, Trempe JF, Matte A, Armengod ME, Cygler M  
842 (2010) Structural basis for Fe-S cluster assembly and tRNA thiolation mediated by IscS protein-  
843 protein interactions. *PLoS Biol* 8:e1000354.
- 844 37. Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA,  
845 Keller M, Aujay M, Huber R, Feldman RA, Short JM, Olsen GJ, Swanson RV (1998) The complete  
846 genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392:353-358.
- 847 38. Xin Y, Liu H, Cui F, Liu H, Xun L (2016) Recombinant *Escherichia coli* with sulfide:quinone  
848 oxidoreductase and persulfide dioxygenase rapidly oxidises sulfide to sulfite and thiosulfate via a  
849 new pathway. *Environ Microbiol* 18:5123-5136.
- 850 39. Ran M, Li Q, Xin Y, Ma S, Zhao R, Wang M, Xun L, Xia Y (2022) Rhodaneses minimize the  
851 accumulation of cellular sulfane sulfur to avoid disulfide stress during sulfide oxidation in bacteria.  
852 *Redox Biology* 53:102345.
- 853 40. Krishna SS, Tautz L, Xu Q, McMullan D, Miller MD, Abdubek P, Ambing E, Astakhova T, Axelrod HL,  
854 Carlton D, Chiu HJ, Clayton T, DiDonato M, Duan L, Elsliger MA, Grzechnik SK, Hale J, Hampton E,  
855 Han GW, Haugen J, Jaroszewski L, Jin KK, Klock HE, Knuth MW, Koesema E, Morse AT, Mustelin T,  
856 Nigoghossian E, Oommachen S, Reyes R, Rife CL, van den Bedem H, Weekes D, White A, Hodgson  
857 KO, Wooley J, Deacon AM, Godzik A, Lesley SA, Wilson IA (2007) Crystal structure of NMA1982 from  
858 *Neisseria meningitidis* at 1.5 angstroms resolution provides a structural scaffold for nonclassical,  
859 eukaryotic-like phosphatases. *Proteins* 69:415-421.
- 860 41. Ray WK, Zeng G, Potters MB, Mansuri AM, Larson TJ (2000) Characterization of a 12-kilodalton  
861 rhodanese encoded by *glpE* of *Escherichia coli* and its interaction with thioredoxin. *J Bacteriol*  
862 182:2277-2284.
- 863 42. Geoghegan KF, Dixon HBF, Rosner PJ, Hoth LR, Lanzetti AJ, Borzilleri KA, Marr ES, Pezzullo LH, Martin  
864 LB, LeMotte PK, McColl AS, Kamath AV, Stroh J (1999) Spontaneous  $\alpha$ -N-6-phosphogluconoylation  
865 of a "His Tag" in *Escherichia coli*: The cause of extra mass of 258 or 178 Da in fusion proteins. *Anal*  
866 *Biochem* 267:169-184.
- 867 43. Li J, Koch J, Flegler W, Garcia Ruiz L, Hager N, Ballas A, Tanabe TS, Dahl C (2023) A metabolic puzzle:  
868 consumption of C<sub>1</sub> compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>. *Biochim Biophys*  
869 *Acta Bioenerget* 1864:148932.
- 870 44. Prioretti L, D'Ermo G, Infossi P, Kpebe A, Lebrun R, Bauzan M, Lojou E, Guigliarelli B, Giudici-Orticoni  
871 M-T, Guiral M (2023) Carbon fixation in the chemolithoautotrophic bacterium *Aquifex aeolicus*  
872 involves two low-potential ferredoxins as partners of the PFOR and OGOR enzymes. *Life* 13:627.
- 873 45. Muyzer G, Sorokin DY, Mavromatis K, Lapidus A, Foster B, Sun H, Ivanova N, Pati A, D'Haeseleer P,  
874 Woyke T, Kyrpides NC (2011) Complete genome sequence of *Thioalkalivibrio* sp. K90mix. *Stand*  
875 *Genomic Sci* 5:doi:10.4056/sigs.2315092.
- 876 46. Bryantseva IA, Gorlenko VM, Kompantseva EI, Imhoff JF, Sölling J, Mityushina L (1999)  
877 *Thiorhodospira sibirica* gen. nov., sp. nov., a new alkaliphilic purple sulfur bacterium from a Siberian  
878 soda lake. *Int J Syst Bacteriol* 49:697-703.
- 879 47. Hallgren J, Tsirigos KD, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, Krogh A,  
880 Winther O (2022) DeepTMHMM predicts alpha and beta transmembrane proteins using deep  
881 neural networks. *bioRxiv*.

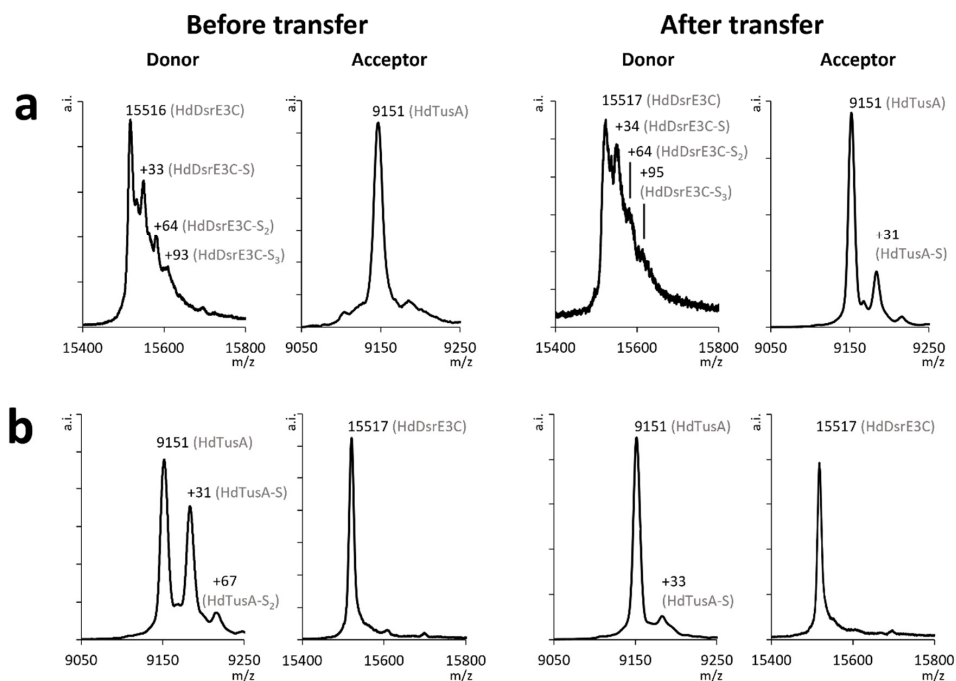
- 882 48. Neukirchen S, Pereira IAC, Sousa FL (2023) Stepwise pathway for early evolutionary assembly of  
883 dissimilatory sulfite and sulfate reduction. *ISME J* 17:1680-1692.
- 884 49. Boughanemi S, Infossi P, Giudici-Ortoniconi MT, Schoepp-Cothenet B, Guiral M (2020) Sulfite oxidation  
885 by the quinone-reducing molybdenum sulfite dehydrogenase SoeABC from the bacterium *Aquifex*  
886 *aeolicus*. *Biochim Biophys Acta Bioenerg* 1861:148279.
- 887 50. Li J, Törkel K, Tanabe TS, Hsu HY, Dahl C (2023) In the Alphaproteobacterium *Hyphomicrobium*  
888 *denitrificans* SoxR serves as a sulfane sulfur-responsive repressor of sulfur oxidation. *Antioxidants*  
889 12:1620.
- 890 51. Bertani G (2004) Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J*  
891 *Bacteriol* 186:595-600.
- 892 52. Ausubel FA, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. 1997. *Current protocols*  
893 *in molecular biology*, John Wiley & Sons, New York.
- 894 53. Dahl C (1996) Insertional gene inactivation in a phototrophic sulphur bacterium: APS-reductase-  
895 deficient mutants of *Chromatium vinosum*. *Microbiology* 142:3363-3372.
- 896 54. Guiral M, Prunetti L, Lignon S, Lebrun R, Moinier D, Giudici-Ortoniconi MT (2009) New insights into  
897 the respiratory chains of the chemolithoautotrophic and hyperthermophilic bacterium *Aquifex*  
898 *aeolicus*. *J Proteome Res* 8:1717-1730.
- 899 55. Ikeda S, Satake H, Hisano T, Terazawa T (1972) Potentiometric argentimetric method for the  
900 successive titration of sulphide and dissolved sulphur in polysulphide solutions. *Talanta* 19:1650-  
901 1654.
- 902 56. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene  
903 recognition and translation initiation site identification. *BMC Bioinformatics* 11:119.
- 904 57. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire  
905 MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang  
906 M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F (2021) RefSeq: expanding the prokaryotic genome  
907 annotation pipeline reach with protein family model curation. *Nucleic Acids Res* 49:D1020-D1028.
- 908 58. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin  
909 L, Raj S, Richardson LJ, Finn RD, Bateman A (2021) Pfam: The protein families database in 2021.  
910 *Nucleic Acids Res* 49:D412-D419.
- 911 59. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7:  
912 improvements in performance and usability. *Mol Biol Evol* 30:772-780.
- 913 60. Crisuolo A, Gribaldo S (2010) BMGE (Block Mapping and Gathering with Entropy): a new software  
914 for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*  
915 10:210.
- 916 61. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic  
917 algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268-274.
- 918 62. Kalyanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS (2017) ModelFinder: fast model  
919 selection for accurate phylogenetic estimates. *Nat Methods* 14:587-589.
- 920 63. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and  
921 methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0.  
922 *Systematic Biology* 59:307-321.
- 923 64. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O (2011) Survey of branch support methods  
924 demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes.  
925 *Systematic Biology* 60:685-699.
- 926 65. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: Improving the ultrafast  
927 bootstrap approximation. *Mol Biol Evol* 35:518-522.
- 928 66. Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display  
929 and annotation. *Nucleic Acids Res* 49:W293-W296.
- 930 67. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek  
931 A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S,  
932 Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M,

bioRxiv preprint doi: <https://doi.org/10.1101/2023.12.18.572138>; this version posted December 18, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](#).

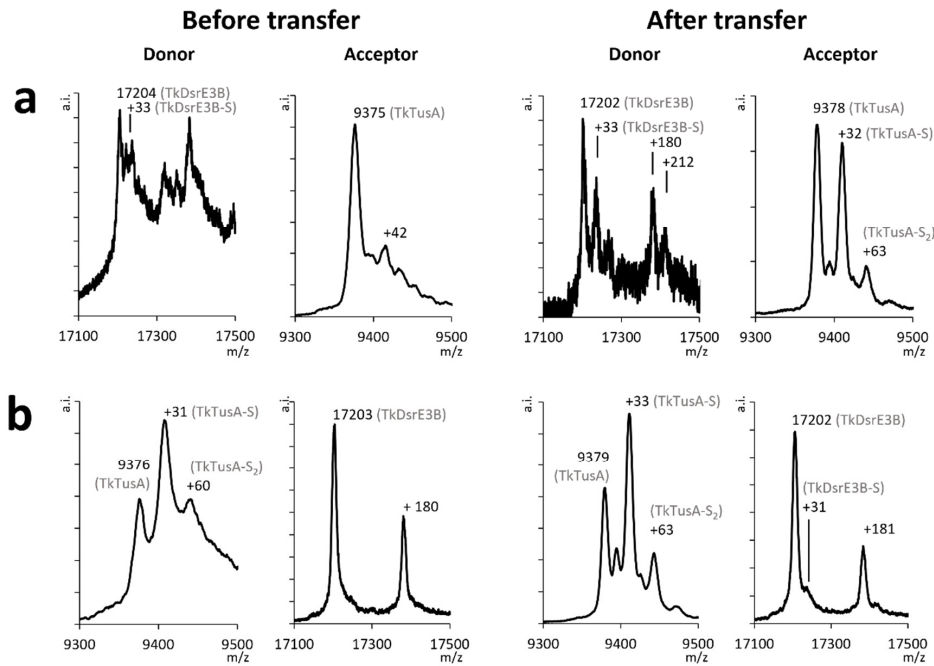
- 933 Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D  
934 (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583-589.
- 935 68. Steinegger M, Söding J (2017) MMseqs2 enables sensitive protein sequence searching for the  
936 analysis of massive data sets. *Nat Biotechnol* 35:1026-1028.
- 937 69. Mirdita M, Steinegger M, Söding J (2019) MMseqs2 desktop and local web server app for fast,  
938 interactive sequence searches. *Bioinformatics* 35:2856-2858.
- 939



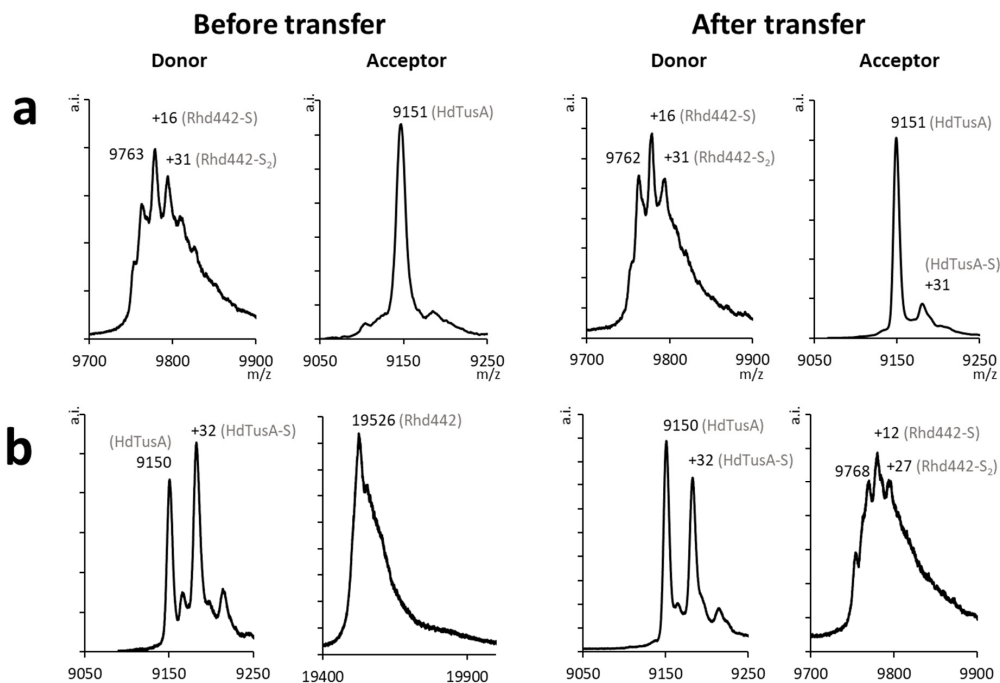
**FIGURE S1.** Thiosulfate consumption (a,b) and growth (c,d) of *H. denitrificans* Δ*tsdA* (open symbols) and *H. denitrificans* Δ*tsdA* Δ*dsrE3C* (filled symbols) in a medium with 24.4 mM methanol and different thiosulfate concentrations (3.5 mM, boxes, 2.5 mM, diamonds, 1.5 mM, triangles, no thiosulfate, circles). Precultures contained either no thiosulfate (non-induced, panels a and c) or 2 mM thiosulfate (induced, panels b and d). In general, induced strains oxidized thiosulfate at a higher specific oxidation rate than non-induced strains. When the Δ*tsdA* reference strain is grown with thiosulfate as an additional electron source, it excretes toxic sulfite, which causes growth retardation (J. Li, Koch, et al., 2023). Functionality of the sHdr-LbpA pathway is thus easily detectable. Growth inhibition of the reference strain was proportional to the initial thiosulfate concentration. Growth retardation was not observed for strain *H. denitrificans* Δ*tsdA* Δ*dsrE3C*. (e) Specific thiosulfate oxidation rate for the non-induced strains *H. denitrificans* Δ*tsdA* (light gray) and *H. denitrificans* Δ*tsdA* Δ*dsrE3C* (dark gray). (f) Specific thiosulfate oxidation rates of *H. denitrificans* Δ*tsdA* compared to the mutant lacking the *dsrE3C* gene and two strains carrying *dsrE3C* genes encoding the indicating cysteine to serine exchanges. Cultures were grown with 2.5 mM thiosulfate and without prior induction. Note that specific thiosulfate oxidation rates are not fully comparable to the experiments shown in (e) because the growth experiments shown here were performed in a plate reader.



**FIGURE S2.** (a) The efficient transfer of sulfane sulfur from DsrE3C to TusA. (b) The transfer of sulfane sulfur from HdTusA to HdDsrE3C led to loss of sulfur from persulfurated HdTusA. The persulfuration of TusA was confirmed before mixing with HdDsrE3C (b donor before transfer), but the signal for this species was almost completely absent in the sample after incubation with HdDsrE3C (b donor after transfer). Neither a mass increase of 32 Da nor a change in ionization properties was observed for the HdDsrE3C after this reaction (b acceptor after transfer). Therefore, we conclude, that a sulfur transfer was not possible from HdTusA to HdDsrE3C.



**FIGURE S3.** (a) The efficient transfer of sulfane sulfur from TkdSrE3B to TkTusA. Left panels: TkdSrE3B as persulfurated donor after treatment with polysulfide and unmodified reduced TkTusA as acceptor; right panels: TkdSrE3B (donor) and TkTusA (acceptor) after the transfer reaction. (b) Transfer of sulfane sulfur from TkTusA to TkdSrE3B is inefficient. Left panels: TkTusA as persulfurated donor after treatment with polysulfide and unmodified reduced TkdSrE3B as acceptor; right panels: TkTusA and TkdSrE3B after the transfer reaction.



**FIGURE S4.** (a) Transfer of sulfane sulfur between HdRh442 to HdTusA. Left panels: HdRh442 as persulfurated donor after treatment with polysulfide and unmodified reduced HdTusA as acceptor. Right panels: HdRh442 (donor) and HdTusA (acceptor) after the transfer reaction. (b) Left panels: HdTusA as persulfurated donor after treatment with polysulfide and unmodified reduced HdRh442 as acceptor. Right panels: HdTusA (donor) and HdRh442 (acceptor) after the transfer reaction.















**TABLE S2.** Sulfur loading of TusA proteins from *H. denitrificans*, *Ts. sibirica*, *Thioalkalivibrio* sp. K90mix and *Aquifex aeolicus* with various inorganic sulfur compounds and oxidized glutathione (GSSG) detected by mass spectrometry. Numbers in parentheses represent mass increases. -, no modification; G, glutathione.

Protein	Expected masses [Da]	Detected masses [Da]	Modification
HdTusA + S <sub>2</sub> O <sub>3</sub> <sup>2-</sup>	9149	9149	-
HdTusA + S <sub>4</sub> O <sub>6</sub> <sup>2-</sup>	9149	9150	-
		9148 (Δ34)	-S
		9212 (Δ62)	-S <sub>2</sub>
		9262 (Δ112)	-S <sub>2</sub> O <sub>3</sub>
		9296 (Δ146)	-S <sub>3</sub> O <sub>3</sub>
HdTusA + GSSG	9149	9148	-
		9453 (Δ304)	-SG
HdTusA + Polysulfide	9149	9149	-
		9180 (Δ31)	-S
		9208 (Δ60)	-S <sub>2</sub>
		9238 (Δ90)	-S <sub>3</sub>
Aq TusA + S <sub>2</sub> O <sub>3</sub> <sup>2-</sup>	9610	9611	-
Aq TusA + GSH	9610	9612	-
Aq TusA + GSSH	9610	9612	-
Aq TusA + S <sub>4</sub> O <sub>6</sub> <sup>2-</sup>	9610	9610	-
		9642 (Δ32)	-S
		9724 (Δ114)	-S <sub>2</sub> O <sub>3</sub>
		9755 (Δ145)	-S <sub>3</sub> O <sub>3</sub>
Aq TusA + Polysulfide	9610	9612	-
		9643 (Δ31)	-S
Aq TusA + Na <sub>2</sub> S	9610	9611	-
		9643 (Δ32)	-S
TkTusA + S <sub>2</sub> O <sub>3</sub> <sup>2-</sup>	9376	9375	-
TkTusA + S <sub>4</sub> O <sub>6</sub> <sup>2-</sup>	9376	9376	-
		9410 (Δ32)	-S
		9490 (Δ111)	-S <sub>2</sub> O <sub>3</sub>
		9525 (Δ146)	-S <sub>3</sub> O <sub>3</sub>
		9549 (Δ171)	-S <sub>4</sub> O <sub>3</sub>
TkTusA + GSSG	9376	9375	-
		9680 (Δ306)	-SG
TkTusA + Polysulfide	9376	9380	None
		9412 (Δ32)	-S
		9445 (Δ64)	-S <sub>2</sub>
TsTusA + S <sub>2</sub> O <sub>3</sub> <sup>2-</sup>	9393	9398	-
TsTusA + S <sub>4</sub> O <sub>6</sub> <sup>2-</sup>	9393	9393	-
		9426 (Δ32)	-S
		9505 (Δ112)	-SO <sub>3</sub>
TsTusA + GSSG	9393	9392	-
TsTusA + Polysulfide	9393	9395	-
		9427 (Δ32)	-S
		9459 (Δ64)	-S <sub>2</sub>

**Table S3. Detection of sulfur transfer between DsrE proteins, TusA and Rhd442.** Numbers in parentheses represent mass increases. -, no modification.

Donor					Acceptor				
Protein	Expected masses [Da]	Detected masses [Da]	Modifications	Protein	Expected masses [Da]	Detected masses [Da]	Modification		
HdTusA-S	9149 + 32x	9151 9184 (Δ32)	- -S	↑	HdDsrE3C	15520 + 32x?	15517	-	
HdDsrE3C-S	15520 + 32x	15517 15551 (Δ34) 15581 (Δ64) 15612 (Δ95)	- -S -S <sub>2</sub> -S <sub>3</sub>	↑	HdTusA	9149 + 32x?	9151 9182 (Δ32)	- -S	
HdDsrE3C-S	15520 + 32x?	15519 15541 (Δ 32)	- -S	↑	HdTusA Cys <sup>83</sup> Ser	9133 + 32x?	9135	-	
HdDsrE3C Cys <sup>83</sup> Ser -S	15503 + 32x?	15503 15535 (Δ32) 15567 (Δ64)	- -S -S <sub>2</sub>	↑	HdTusA	9149 + 32x?	9149 9181 (Δ32)	- -S	
HdDsrE3C Cys <sup>84</sup> Ser -S	15503 + 32x?	15503	-	↑	HdTusA	9149 + 32x?	9148	-	
TkTusA-S	9376 + 32x	9379 9412 (Δ33) 9442 (Δ63)	- -S -S <sub>2</sub>	↑	TKDsrE3B	17202 + 32x?	17202 17234 (Δ31) 17383 (Δ181)	- -S -glucose	
TkDsrE3B-S	17202 + 32x	17202 17235 (Δ33) 17382 (Δ180) 17414 (Δ212)	- -S -glucose -S + -glucose	↑	TkTusA	9376 + 32x?	9378 9410 (Δ32) 9441 (Δ63)	- -S -S <sub>2</sub>	
TsTusA-S	9393 + 32x	9396 9427 (Δ31) 9457 (Δ61)	- -S -S <sub>2</sub>	↑	TsDsrE3B	16620 + 32x?	16622 16653 (Δ31) 16801 (Δ179)	- -S -glucose	
TsDsrE3B-S	16620 + 32x	16622 16653 (Δ31) 16684 (Δ62) 16720 (Δ98)	NM -S -S <sub>2</sub> -S <sub>3</sub>	↑	TsTusA	9393 + 32x?	9395 9427 (Δ32)	- -S	
HdRhd442-S	19526 + 32x	19554 9762 (z = 2) 9778 (Δ16) (z = 2) 9793 (Δ31) (z = 2)	-S - -S -S <sub>2</sub>	↑	HdTusA	9149 + 32x?	9149 9181 (Δ32)	-S	
HdRhd442-S	19526 + 32x	19554 9763 (z = 2) 9779 (Δ16) (z = 2) 9795 (Δ32) (z = 2) 9810 (Δ47) (z = 2)	-S - -S -S <sub>2</sub> -S <sub>3</sub>	↑	HdDsrE3C	15520 + 32x?	15519 15550 (Δ32)	-S	

Table S4: Proteins identified by mass spectrometry after re-purification of His-tagged AqTusA incubated with an *Aq. aeolicus* soluble extract (Band 1).

Accession	Description	Locus tag	MW	Coverage	Score	# PSMs	# Unique Peptides
	<b>TusA_tag</b>	aq_388a	9,611	<b>90</b>	850	<b>237</b>	<b>10</b>
O66901	hydrogenase expression/formation protein B	aq_671	28,997	38	37	13	10
O67481	OsmC/Ohr family protein	aq_1515	15,448	46	16	6	5
O67877	Acetoin utilization protein	aq_2110	34,807	21	10	4	4
O67476	O-methyltransferase	aq_1507	24,296	17	9	4	4
O66829	Uncharacterized RNA pseudouridine synthase	aq_554	27,845	13	8	3	3
<b>O66711</b>	<b>DsrE3C</b>	<b>aq_390</b>	15,584	<b>31</b>	8	<b>3</b>	<b>3</b>
<b>O66720</b>	<b>LbpA</b>	<b>aq_402</b>	15,813	<b>27</b>	7	<b>3</b>	<b>3</b>
O66599	Rieske-I iron sulfur protein	aq_234	26,858	14	6	3	3
O66859	DUF2795 domain-containing protein	aq_600	18,31	17	6	3	3
O67802	Putative carboxymethylenebutenolidase	aq_1997	26,34	10	4	2	2
O67045	iron-sulfur cluster assembly scaffold protein IscU	aq_896	17,519	13	4	2	2
O66523	30S ribosomal protein S16	aq_123	13,03	17	4	2	2
O67717	CoA-binding domain-containing protein	aq_1869	15,326	15	4	2	2
O66857	DUF302 domain-containing protein	aq_598	14,51	19	3	2	2

Proteins from band 1 of the native gel (Figure 8 a) were analyzed.

Accession: accession number in UniProt database. MW: theoretical molecular weight of protein in kDa. Coverage: percent protein sequence coverage by the matching peptides. # Peptides: number of distinct peptides matching to protein sequence and unique to this protein. # PSMs: peptide spectrum match number (given by the algorithm corresponding to the total number of identified peptide sequences for the protein, including those redundantly identified). The *Aq. aeolicus* TusA protein (WP\_024015099) is not referenced in Uniprot and the corresponding sequence was manually added in the database.



**Table S5:** Proteins, encoded by the *Aq. aeolicus* *hdr*-like cluster, identified by mass spectrometry after re-purification of His-tagged TusA incubated with an *Aq. aeolicus* soluble extract (second experiment).

Accession	Description	Locus tag	MW	Control				TusA			
				Coverage	Score	# Peptides	# PSMs	Coverage	Score	# Peptides	# PSMs
	TusA_tag	aq_388a	9,611	-	-	-	-	88	861	13	211
O66711	DsrE3C	aq_390	15,584	-	-	-	-	46	91	6	28
O66718	heterodisulfide reductase subunit B HdrB2	aq_400	37,221	-	-	-	-	42	68	14	20
A0A193BL08	heterodisulfide reductase subunit B HdrB1	aq_392	51,938	-	-	-	-	32	46	12	15
O66712	heterodisulfide reductase subunit C HdrC1	aq_391	29,438	-	-	-	-	35	19	7	7
O66710	DsrE2A	aq_389	19,88	-	-	-	-	13	7	2	2
O66715	heterodisulfide reductase subunit A HdrA	aq_395	38,517	27	25	8	9	67	83	18	25
O66717	heterodisulfide reductase subunit C HdrC2	aq_398	29,823	7	7	3	3	20	16	5	6
O66720	LbpA2	aq_402	15,813	13	7	2	2	32	20	3	5

“TusA” corresponds to the incubation of AqTusA with the soluble extract (second trial) and “Control” to the incubation of the extract without the AqTusA (negative control). Accession: accession number in UniProt database. MW: theoretical molecular weight of protein in kDa. Coverage: percent protein sequence coverage by the matching peptides. # Peptides: number of distinct peptides matching to protein sequence and unique to this protein. # PSMs: peptide spectrum match number (given by the algorithm corresponding to the total number of identified peptide sequences for the protein, including those redundantly identified). The *Aq. aeolicus* TusA protein (WP\_024015099) is not referenced in Uniprot and the corresponding sequence was manually added in the database. HdrB1 was originally encoded by a pseudogene and corresponded to locus tags aq\_392 and aq\_394.<sup>(1)</sup> “-” means that the protein was not identified in the sample. The proteins AqsHdrH (aq\_397) and AqDsrE3B (aq\_401) were not identified in any sample in this experiment.

## REFERENCES

1. Boughanemi S, Lyonnet J, Infossi P, Bauzan M, Kosta A, Lignon S, Giudici-Orticoni MT, Guiral M (2016) Microbial oxidative sulfur metabolism: biochemical evidence of the membrane-bound heterodisulfide reductase-like complex of the bacterium *Aquifex aeolicus*. *FEMS Microbiol Lett* 363:fnw156.

Table S6: Proteins identified by mass spectrometry after re-purification of His-tagged AqTusA incubated with an *Aq. aeolicus* soluble extract (bands 2 and 3).

Accession	Description	Locus tag	MW	Coverage	Score	# PSMs	# Unique Peptides
<b>Band 2</b>							
O66869	Protease I	aq_618	18.961	71	321	104	13
O67566	Small ribosomal subunit protein uS8	aq_1651	19.445	39	16	6	6
O66780	Thiol peroxidase	aq_488	18.561	31	11	5	5
O67587	ATP-dependent protease subunit HsIV	aq_1671	19.276	38	10	5	5
O66541	Citrate synthase	aq_150	29.057	5	8	3	2
<b>O66710</b>	<b>DsrE2A</b>	<b>aq_389</b>	<b>19.88</b>	<b>13</b>	<b>8</b>	<b>4</b>	<b>2</b>
O67527	ATP synthase subunit delta	aq_1588	20.723	19	7	3	3
O66981	Cytochrome c552 CycB1	aq_792	18.474	9	7	3	2
O67903	Peptidoglycan associated lipoprotein	aq_2147	22.842	16	7	3	3
O67613	dTTP/UTP pyrophosphatase	aq_1718	21.037	16	5	2	2
O66616	Uncharacterized protein	aq_255	18.849	8	5	2	2
O66764	Archaemetzincin	aq_459	20.324	11	5	2	2
O67100	Arginine biosynthesis bifunctional protein ArgJ	aq_970	41.717	6	4	2	2
<b>O67573</b>	<b>LbpA3</b>	<b>aq_1657</b>	<b>18.215</b>	<b>19</b>	<b>4</b>	<b>2</b>	<b>2</b>
O67742	Orotate phosphoribosyltransferase	aq_1907	20.297	14	4	2	2
O67722	Large ribosomal subunit protein uL13	aq_1877	17.057	11	4	2	2
<b>Band 3</b>							
O67294	Copper chaperone PCu(A)C	aq_1253	17.191	55	158	54	11
<b>O66711</b>	<b>DsrE3C</b>	<b>aq_390</b>	<b>15.584</b>	<b>56</b>	<b>46</b>	<b>15</b>	<b>7</b>
O66523	Small ribosomal subunit protein bS16	Aq_123	13.03	39	44	17	8
O67481	OsmC family peroxiredoxin	aq_1515	15.448	49	40	14	6
O66486	Small ribosomal subunit protein uS13	aq_074	14.331	28	26	9	6
<b>O66720</b>	<b>LpbA2</b>	<b>aq_402</b>	<b>15.813</b>	<b>15</b>	<b>21</b>	<b>6</b>	<b>4</b>
O66537	Uncharacterized protein	aq_142	16.311	38	16	6	5
O67065	Ferredoxin-1	aq_919.1	10.755	41	15	5	3
O66485	Small ribosomal subunit protein uS11	aq_073	13.518	10	13	5	2

O67227	Uncharacterized protein	aq_1163	14.885	32	11	5	4
O67562	Large ribosomal subunit protein ul30	aq_1644	14.454	37	11	4	4
O66541	Citrate synthase	aq_150	29.057	14	11	5	3
O66582	KaiC domain-containing protein	aq_204	32.931	16	10	4	4
O67534	Rhodanese domain-containing protein	aq_1599	17.3	30	10	4	4
O66705	CRISPR type III-B/RAMP module-associated protein Cmr5	aq_384	15.66	25	7	3	3
O66901	Hydrogenase expression/formation protein B	aq_671	28.997	16	7	3	3
P0A466	Large ribosomal subunit protein bl12	aq_1937	13.551	9	7	3	2
<b>O66719</b>	<b>DsrE3B</b>	<b>aq_401</b>	<b>14.887</b>	<b>28</b>	<b>7</b>	<b>3</b>	<b>2</b>
O66910	Arsenate reductase ArsC	aq_685	16.233	20	6	3	3
O67706	Uncharacterized protein	aq_1854	17.702	11	6	2	2
O67424	Minor pilin	aq_1433	12.83	17	5	2	2
O67548	Metallo-beta-lactamase domain-containing protein	aq_1625	28.245	9	5	2	2
O67570	Large ribosomal subunit protein ul14	aq_1654	13.263	25	5	2	2
O67073	NADPH-dependent 7-cyano-7-deazaguanine reductase	aq_931	15.321	18	5	2	2
O67054	Phosphoheptose isomerase	aq_908	20.693	16	4	2	2
O66433	Large ribosomal subunit protein ul23	aq_012	12.363	23	5	2	2
O67234	Phosphohistidine phosphatase SixA	aq_1173	16.997	18	4	2	2
O66862	Acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase	aq_604	28.398	10	4	2	2

Proteins from bands 2 (about 18 kDa) and 3 (about 15 kDa) of the SDS gel (Figure 8b) were analyzed. Accession: accession number in UniProt database. MW: theoretical molecular weight of protein in kDa. Coverage: percent protein sequence coverage by the matching peptides. # Peptides: number of distinct peptides matching to protein sequence and unique to this protein. # PSMs: peptide spectrum match number (given by the algorithm corresponding to the total number of identified peptide sequences for the protein, including those redundantly identified). The *Aq. aeolicus* TusA protein (WP\_024015099) is not referenced in Uniprot and the corresponding sequence was manually added in the database.

Table S7. Strains, plasmids and primers

Strains primers or plasmids	Relevant genotype, description or sequence	Reference or source
<b>Strains</b>		
<i>E. coli</i> 10-beta	$\Delta(\text{ara-leu})$ 7697 <i>araD139 thuA</i> $\Delta(\text{lacX74 galK16 galE15 e14-}\phi 80\text{dlacZ}\Delta\text{M15 recA1 relA1 endA1 nupG rpsL (Str}^{\text{R}}) \text{rph spoT1 } \Delta(\text{mrr-hsdRMS-mcrBC})$	New England Biolabs
<i>E. coli</i> DH5 $\alpha$	F <sup>-</sup> $\phi 80\text{lacZ}\Delta\text{M15 } \Delta(\text{lacZYA-argF})\text{U169 recA1 endA1 hsdR17}(\text{rK}^-, \text{mK}^-) \text{phoA supE44 } \lambda\text{-thi-1 gyrA96 relA1}$	New England Biolabs
<i>E. coli</i> BL21 (DE3)	F <sup>-</sup> <i>ompT hsdS<sub>B</sub> (r<sub>B</sub><sup>-</sup>, m<sub>B</sub><sup>-</sup>) gal dcm</i> (DE3)	Novagen
<i>Hyphomicrobium denitrificans</i> $\Delta\text{tsdA}$	Sm <sup>r</sup> , in-frame deletion of <i>tsdA</i> in <i>H. denitrificans</i> Sm200	(1)
<i>Hyphomicrobium denitrificans</i> $\Delta\text{tsdA } \Delta\text{dsrE3C}$	Sm <sup>r</sup> , in-frame deletion of <i>dsrE3C</i> (Hden_0688) in <i>H. denitrificans</i> $\Delta\text{tsdA}$	(2)
<i>Hyphomicrobium denitrificans</i> $\Delta\text{tsdA dsrE-C83S}$	Sm <sup>r</sup> , mutation of <i>dsrE3C</i> (Hden_0688) C83 to S in <i>H. denitrificans</i> $\Delta\text{tsdA}$	This work
<i>Hyphomicrobium denitrificans</i> $\Delta\text{tsdA dsrE-C84S}$	Sm <sup>r</sup> , mutation of <i>dsrE3C</i> (Hden_0688) C83 to S in <i>H. denitrificans</i> $\Delta\text{tsdA}$	This work
<b>Primers</b>		
Hden0688 ( <i>dsrE3C</i> )_Up_Fw_PstI	ATATCTGCAGCCAATCTGCGTGGCGTTCCG	This work
Hden0688 ( <i>dsrE3C</i> )_Up_Rev	CCCTGCCGTCGAAAAATTCATGCCACCTCCCCGATATG	This work
Hden0688 ( <i>dsrE3C</i> )_Down_Fw	CATATCGGGGAGGTGGCATTGAATTTTCGGACGGCAGGG	This work
Hden0688 ( <i>dsrE3C</i> )_Down_Rev_XbaI	CATGCTCAGATCGCGCTCGGTGATGCGATG	This work
Hden0688 ( <i>dsrE3C</i> )_C83S_Fw	GTGAAATTTTTCTCCTGTCTCCCAATCTC	This work
Hden0688 ( <i>dsrE3C</i> )_C83S_Rev	GAGATTGGGAGAACAGGAGAAAAATTCAC	This work
Hden0688 ( <i>dsrE3C</i> )_C84S_Fw	GTGAAATTTTTCTGCTCTTCTCCCAATCTC	This work
Hden0688 ( <i>dsrE3C</i> )_C84S_Rev	GAGATTGGGAGAGAGCAGAAAAATTCAC	This work
Hden0688 ( <i>dsrE3C</i> )_NdeI_fw	CACGCATATGTTGGCCGAAAAACTTCTG	This work
Hden0688 ( <i>dsrE3C</i> )_BamHI_rev	GCGTGGATCCTCAGTATGAAAGCACTTTG	This work
TK90_0639 ( <i>dsrE3B</i> )_NdeI_fw	AGAGCATATGATGGCTGAAGTGGGC	This work
TK90_0639 ( <i>dsrE3B</i> )_BamHI_rev	CGAGGATCCAAAATTAACAGCTTATCGT	This work
ThisiDRAFT_1818 ( <i>dsrE3B</i> )_NdeI_fw	AGACATATGACTGATGCAAGTC	This work
ThisiDRAFT_1818 ( <i>dsrE3B</i> )_XhoI_rev	TTTTTCTCGAGTTAGAAGTTATGAT	This work
HdenTusA_fw_NdeI	CGACCACATATGCCCATCTGACAGTTGAT	This work
HdenTusA_rev_EcoRI	GCAAGCGAATTCCTTATTTTTCGAACTGCGGGTGGCTCCAAGCGCTGGCCGCCGTG TGCTTGATCA	This work
Tk90TusA_fw_NdeI	AGACACCATATGCCAACTTTGACCAAGA	This work
TK90TusA_rev_EcoRI	CCTAGAGAATTCCTTATTTTTCGAACTGCGGGTGGCTCCAAGCGCTGGACTTCTTAC GAGGAAGTAGAAGT	This work
ThisiTusA_fw_NdeI	ACACACGCATATGGCAAATTTTGACCTAGAAGT	This work
ThisiTusA_rev_EcoRI	GCCAGTGAATTCCTTATTTTTCGAACTGCGGGTGGCTCCAAGCGCTGCTCTTGGGA TC	This work
HdenTusA_C13S_fw	GGCACGAACTCTCCTATCCCGATTTTGAAG	This work
HdenTusA_C13S_rev	CTTCAAAATCGGGATAGGAGAGTTCGTGCC	This work

Aq388a_Ndel_fw	TAGTTCATATGGCTACAATAACACCTGACAAGG	This work
Aq388a_XhoI_rev	AGTCTCGAGTCCTTTTTTCCTTATGTAGTAGTACTTACC	This work
AqT388a_C17_fw	CGATACTCCGGACTTAACCTCCTCTGCCCGTG	This work
Aq388a_C17_rev	CACGGGGCAGAGGAGAGTTAAGTCCGGAAGTATCG	This work
Aq388a_C54_fw	GATATTCACGCGTTCTCTCAAAGGACTGGACAC	This work
Aq388a_C54_rev	GTGTCCAGTCCCTTTGAGAGAACGCTGGAATATC	This work
HdRhd442_fw_Ndel	CGATCATATGAGTCAAAGAACTGCG	This work
HdRhd442_rev_HindIII	GATCAAGCTTCTCGTCGCATCTTTCAGC	This work
<b>Plasmids</b>		
pET-22b(+)	Ap <sup>r</sup>	Novagen
pET-15a(+)	Ap <sup>r</sup>	Novagen
pET-15a-Hd-DsrE3C	Ap <sup>r</sup> , pET-15a with dsrE3C (Hden_0688) insertion between NdeI and BamHI	This work
pET-15a-Hd-DsrE3C-C83S	Ap <sup>r</sup> , pET-15a-Hd-DsrE3C with Cys <sup>83</sup> Ser exchange	This work
pET-15a-Hd-DsrE3C-C84S	Ap <sup>r</sup> , pET-15a-Hd-DsrE3C with Cys <sup>84</sup> Ser exchange	This work
pET-15a-TK90-DsrE3B	Ap <sup>r</sup> , pET-15a with dsrE3B (TK90_0639) insertion between NdeI and BamHI	This work
pET-15a-Ts-DsrE3B-	Ap <sup>r</sup> , pET-15a with dsrE3B (ThisIDRAFT_2311) insertion between NdeI and BamHI	This work
pET22b-HdenTusA	Ap <sup>r</sup> , pET-22b with tusA (Hden_0698) insertion between NdeI and BamHI	This work
pET22b-HdenTusA-C13S	Ap <sup>r</sup> , pET22b-HdenTusA with Cys <sup>13</sup> Ser exchange	This work
pET22b-TK90TusA	Ap <sup>r</sup> , pET-22b with tusA insertion between NdeI and BamHI	This work
pET22b-ThisiTusA	Ap <sup>r</sup> , pET-22b with tusA insertion between NdeI and XhoI	This work
	Ap <sup>r</sup> , pET-22b-SoxR-N-Strep with Cys <sup>50</sup> Ser and Cys <sup>116</sup> Ser exchanges	This work
pk18mobsacB-Tc	Km <sup>r</sup> , Tc <sup>r</sup> pHP45ΩTc tetracycline cassette inserted into pk18mobsacB using SmaI	(2)
pk18mobsacB-Tc	Km <sup>r</sup> , Tc <sup>r</sup> pHP45ΩTc tetracycline cassette inserted into pk18mobsacB using SmaI	(2)
pk18mobsacB_Tc_dsrE3C-C83S	Km <sup>r</sup> , Tc <sup>r</sup> , 2.08 kb SOE PCR fragment implementing mutation C83S of <i>dsrE3C</i> cloned into pk18mobsacB-Tc using PstI and XbaI restriction sites	(2)
pk18mobsacB_Tc_dsrE3C-C84S	Km <sup>r</sup> , Tc <sup>r</sup> , 2.08 kb SOE PCR fragment implementing mutation C84S of <i>dsrE3C</i> cloned into pk18mobsacB-Tc using PstI and XbaI restriction sites	(2)
pk18mobsacB_Tc_ΔdsrE (Hden0688)	Km <sup>r</sup> , Tc <sup>r</sup> , 2.08 kb SOE PCR fragment implementing deletion of <i>dsrE3C</i> cloned into pk18mobsacB-Tc using PstI and XbaI restriction sites	This work

## References

1. Koch T, Dahl C (2018) A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds. *ISME J* 12:2479-2491.
2. Li J, Koch J, Flegler W, Garcia Ruiz L, Hager N, Ballas A, Tanabe TS, Dahl C (2023) A metabolic puzzle: consumption of C<sub>1</sub> compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>. *Biochim Biophys Acta Bioenerget* 1864:148932.



---

## Fe/S proteins in microbial sulfur oxidation

Kümpel, C., Grosser, M., Tanabe, T. S., Dahl, C.

---

Iron-sulfur clusters are of fundamental importance for enzymes involved in such as catalysis, iron-sulfur storage, electron transport or regulation of enzyme activity and gene expression (Beinert *et al.* 1997, Johnson *et al.* 2005). They are widespread in all domains of life and their functional diversity is comparable to prosthetic groups such as hemes and flavins (Beinert *et al.* 1997). Most described iron-sulfur clusters have a rhombic [2Fe-2S], cubane-like [3Fe-4S] or [4Fe-4S] structure (Beinert *et al.* 1997), but the number of proteins coordinating non-conventional and non-cubane iron-sulfur clusters is increasing (Wagner *et al.* 2017, Caserta *et al.* 2022). Here, the role of iron-sulfur clusters in microbial sulfur oxidation is reviewed. In particular, recent progress on the function of iron-sulfur clusters in dissimilatory sulfur oxidation, lipoate biosynthesis, and the associated iron-sulfur cluster biosynthesis pathways is summarized (Kümpel *et al.* 2024).

Two different types of gene clusters associated with sulfur dissimilation via the sHdr pathway can be distinguished. Both gene clusters and their particular differences were described, as well as their prevalence in the representative genome dataset of the Genome Taxonomy Database (Release R207). The the core genes of the type I gene cluster *shdrC1B1AHC2B2* are often accompanied by up to three genes for lipoate-binding proteins, genes encoding for a lipoate biosynthesis pathway and genes for multiple sulfur transferases. In the Pseudomonadota (formerly Proteobacteria), Aquificota and the archaeal phylum Thermoproteota (formerly Crenarchaeota and other phyla (Rinke *et al.* 2021)) studied here, the sHdr pathway is always encoded by a type I gene cluster. The majority of the genomes analyzed that code for the type I gene cluster also belong to one of these three phyla. Several studies have investigated the function of these genes through comparative genomic, transcriptomic, and proteomic approaches, supporting their essential role in dissimilatory sulfur oxidation (Quatrini *et al.* 2009, Ehrenfeld *et al.* 2013, Christel *et al.* 2016, Koch & Dahl 2018, Ernst *et al.* 2021, Li *et al.* 2023,b, Tanabe *et al.* 2023c).

The type II *shdr* gene clusters are very similar to the type I cluster and have been linked to dissimilatory sulfur oxidation based on comparative genomics (Justice *et al.* 2014, Cao *et al.* 2018, Tanabe *et al.* 2023c). Several genes were recognized to commonly appear in the type II gene clusters, but not in the type I gene clusters. In the following article, two of these genes coding for sHdrB3 and ETF:menaquinone oxidoreductase were named accordingly. The sHdrB3 polypeptide is a fusion of sHdrC2 and sHdrB2 and replaces them in the type II gene cluster. Type II gene clusters therefore have the core genes *shdrC1B1AHB3etfAB* and are usually accompanied by the genes for a ETF:menaquinone oxidoreductase, sulfur transferases, lipoate-binding proteins, lipoate synthases and lipoate:protein ligases. The core genes were present in a wide array of diverse phyla among the bacteria and the archaeal phylum Thermoplasmata.

The coordinating residues and structures of the cubane [4Fe-4S] and non-cubane iron-sulfur clusters are described for the sHdr subunits encoded in these two gene sets. Based on the postulated function of the sHdr complex, the iron-sulfur clusters in these units serve as both electron-conducting units and catalytically active centers. There are currently five pathways described for the biosynthesis of iron-sulfur clusters. Two of these pathways can be traced back to the last universal common ancestor, while the other three arose from these two ancestral pathways (Garcia *et al.* 2022b). It was determined that all of these pathways are capable of providing iron-sulfur clusters for sulfur oxidation by correlating the corresponding iron-sulfur cluster biosynthesis machinery with the respective sulfur oxidizers. This also applies to the reverse Dsr system, which was also analyzed.

Furthermore, the recent advances on the lipoic acid cofactor biosynthesis via the lipoate synthases LipS1, LipS2 and the lipoate:protein ligases sLpl(AB) are summarized as this pathway is highly relevant for the lipoate-binding protein and is commonly encoded in the *shdr* gene clusters (Tanabe *et al.* 2023b, Kümpel *et al.* 2024). For each of the lipoate synthases the residues coordinating the two cubane [4Fe-4S] clusters are analyzed. During the lipoate assembly reaction one of these iron-sulfur clusters acts as sulfur donor for the insertion at C6 or C8, while the other is required for the radical formation from S-adenosyl methionine. LipS2 then replaces a hydrogen atom at C8 with a sulfur atom followed by action of LipS1 that inserts the sulfur atom at C6 thus completing lipoate formation (Neti *et al.* 2022).

Additional aspects of this review are the general importance of iron-sulfur clusters, pathways in microbial sulfur oxidation and the iron-sulfur clusters in the reverse dissimilatory sulfite reductase system (Kümpel *et al.* 2024).

T.S.T. contributed to this review by conceptualization, investigation, figure design and writing: The analysis of the correlation of sHdr/rDsr system and the iron-sulfur cluster biosynthesis pathways was conceptualized and carried out by T.S.T. who also designed the associated figure. T.S.T. also contributed to the visualization of the type I and type II gene clusters and the recognition of the sHdrB3 as fusion protein. T.S.T. contributed to writing the text of the mentioned topics.



1 **Title:** Fe/S proteins in microbial sulfur oxidation

2

3 **Authors:** Kümpel, Carolin<sup>§</sup>; Grosser, Martina<sup>§</sup>; Tanabe, Tomohisa Sebastian; Dahl,  
4 Christiane\*

5 <sup>§</sup> Both authors contributed equally to this work.

6 Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich-Wilhelms-Universität Bonn,  
7 Bonn, Germany

8 **\*Correspondence:** Christiane Dahl, Institut für Mikrobiologie & Biotechnologie, Rheinische  
9 Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 168, D-53115  
10 Bonn, Germany, Phone +49 228 732119, Fax: +49 228 737576,  
11 E-Mail: ChDahl@uni-bonn.de

12 **Highlights**

- 13 • Fe/S clusters play important roles in bacterial dissimilatory sulfur oxidation
- 14 • Fe/S proteins serve as electron carriers and in the active sites of key sulfur-oxidizing  
15 enzymes
- 16 • Unusual noncubane Fe/S clusters catalyze the conversion of sulfane sulfur to sulfite
- 17 • Assembly of lipoate in sHdr-containing sulfur oxidizers involves novel radical SAM  
18 enzymes

19 **Abstract**

20 Iron-sulfur clusters serve as indispensable cofactors within proteins across all three domains  
21 of life. Fe/S clusters emerged early during the evolution of life on our planet and the  
22 biogeochemical cycle of sulfur is one of the most ancient and important element cycles. It is  
23 therefore no surprise that Fe/S proteins have crucial roles in the multiple steps of microbial  
24 sulfur metabolism. During dissimilatory sulfur oxidation in prokaryotes, Fe/S proteins not only  
25 serve as electron carriers in several steps, but also perform catalytic roles, including  
26 unprecedented reactions. Two cytoplasmic enzyme systems that oxidize sulfane sulfur to  
27 sulfite are of particular interest in this context: The rDsr pathway employs the reverse acting  
28 dissimilatory sulfite reductase rDsrAB as its key enzyme, while the sHdr pathway utilizes  
29 polypeptides resembling the HdrA, HdrB and HdrC subunits of heterodisulfide reductase from  
30 methanogenic archaea. Both pathways involve components predicted to bind unusual  
31 noncubane Fe/S clusters acting as catalysts for the formation of disulfide or sulfite. Mapping  
32 of Fe/S cluster machineries on the sulfur-oxidizing prokaryote tree reveals that ISC, SUF, MIS  
33 and SMS are all sufficient to meet the Fe/S cluster maturation requirements for operation of  
34 the sHdr or rDsr pathways. The sHdr pathway is dependent on lipoate-binding proteins that  
35 are assembled by a novel pathway, involving two Radical SAM proteins, namely LipS1 and  
36 LipS2. These proteins coordinate sulfur-donating auxiliary Fe/S clusters in atypical patterns by  
37 three cysteines and one histidine and act as lipoyl synthases by jointly inserting two sulfur  
38 atoms to an octanoyl residue.

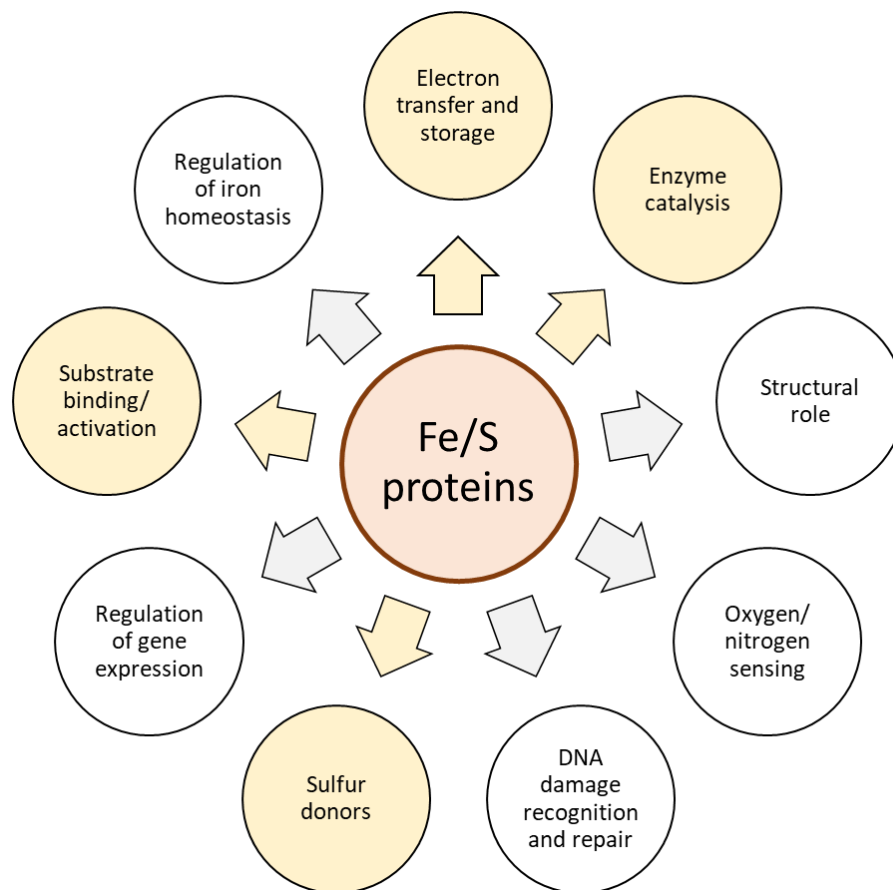
39 This article is part of a Special Issue entitled: Biogenesis and Function of Fe/S proteins.

40 **KEYWORDS:**

41 Dissimilatory sulfur oxidation, Dsr system, sulfur-oxidizing heterodisulfide reductase, Radical  
42 SAM proteins, Fe/S clusters, Lipoate assembly

43 **1. An introduction into the role of Fe/S clusters and their importance in the**  
44 **natural sulfur cycle**

45 Iron–sulfur (Fe/S) clusters are ancient inorganic cofactors found in all domains of life and  
46 are considered to be among the first catalysts in living organisms [1-5]. They exhibit remarkably  
47 versatile chemical/electronic properties and structural plasticity and can therefore assume a  
48 wide range of functions [2, 6, 7]. The first Fe/S proteins isolated were small electron carriers.  
49 While electron transfer is indeed still the predominant function for the very large number of  
50 Fe/S proteins known to date, additional roles such as regulation of gene expression, substrate  
51 binding/activation, and enzyme catalysis have emerged [8-12] (Fig. 1).



52

53 **Fig. 1.** Functions of Fe/S proteins. Roles of Fe/S proteins in microbial oxidative sulfur  
54 metabolism discussed in this review are highlighted in yellow.

55 Just as Fe/S clusters emerged early during the evolution of life on our planet, the  
56 biogeochemical cycle of sulfur is one of the most ancient and important element cycles [13].

57 Accordingly, dissimilatory sulfite reduction and sulfide oxidation are processes that can be  
58 traced back to early Earth history, 3.5 Ga ago [14, 15], and microbial sulfur metabolism is  
59 considered one of the earliest types of microbial activity. It is therefore not surprising that Fe/S  
60 proteins play an outstandingly important role during the multifaceted steps of microbial sulfur  
61 metabolism. In addition to serving as electron carriers in several steps, catalytic roles of Fe/S  
62 proteins in sulfur metabolism, including unprecedented reactions, are beginning to emerge  
63 (Fig. 1) [16-18]. The active site of sulfite reductase presents an enduringly fascinating example,  
64 where a cysteine coordination is shared by siroheme and a [4Fe-4S] cluster [19-23].

65 This review strictly focuses on the oxidative side of the sulfur cycle. We provide brief  
66 overviews of Fe/S cluster function and microbial sulfur oxidation. We highlight in particular  
67 the metabolic steps involving Fe/S proteins in the following section. The rest of the article  
68 presents examples of how Fe/S proteins contribute to novel and unexpected reactions in sulfur  
69 oxidation.

## 70 **2. Functions of Fe/S clusters**

71 Fe/S proteins facilitate one-electron redox reactions via Fe<sup>2+</sup> or Fe<sup>3+</sup> oxidation states of  
72 iron. Due to this functionality, they serve as critical components in respiratory and  
73 photosynthetic electron transfer chains [1], or as redox enzymes involved in carbon, oxygen,  
74 hydrogen, and nitrogen metabolism. Differences in the ligands and in the protein environment  
75 of the Fe/S clusters allow a wide range of redox potentials, from -600 to +400 mV [5]. Fe/S  
76 clusters exhibit a significant preference for thiolate ligation and consequently, cysteinyl sulfur  
77 is predominantly utilized as the ligand for Fe/S active sites. Nonetheless, ligation by histidine  
78 or to a lesser extent glutamine, glutamate, aspartate, threonine, serine, tyrosine, lysine or  
79 arginine has also been documented [4, 24-27].

80 Functions for Fe/S proteins beyond electron transfer include participation in  
81 oxygen/nitrogen sensing [28, 29] or DNA damage recognition and repair [4, 30, 31] (Fig. 1). In  
82 addition, Fe/S clusters can serve as the active sites of catalytic enzymes. For instance, the [4Fe-  
83 4S] cluster found in radical S-adenosylmethionine proteins is capable of cleaving S-  
84 adenosylmethionine [32, 33] and aconitase isomerizes citrate to isocitrate, with a non-  
85 liganded iron atom of an [4Fe-4S] cluster serving as a Lewis acid [9]. Additionally, two distinct  
86 disulfide reductases, ferredoxin:thioredoxin reductase from chloroplasts and heterodisulfide  
87 reductase from methanogenic archaea, were observed to employ Fe/S clusters in their active

88 sites for disulfide cleavage [8, 34-37]. Moreover, the sulfur donors for the biosynthesis of biotin  
89 and lipoyl, two sulfur-containing cofactors, have been identified as Fe/S clusters in biotin  
90 synthases and lipoyl synthases [32, 38, 39]. The corresponding Fe/S clusters must be  
91 reassembled during each catalytic cycle, indicating that degradation and reassembly of the  
92 Fe/S clusters are integral to the catalytic cycle [27, 40, 41].

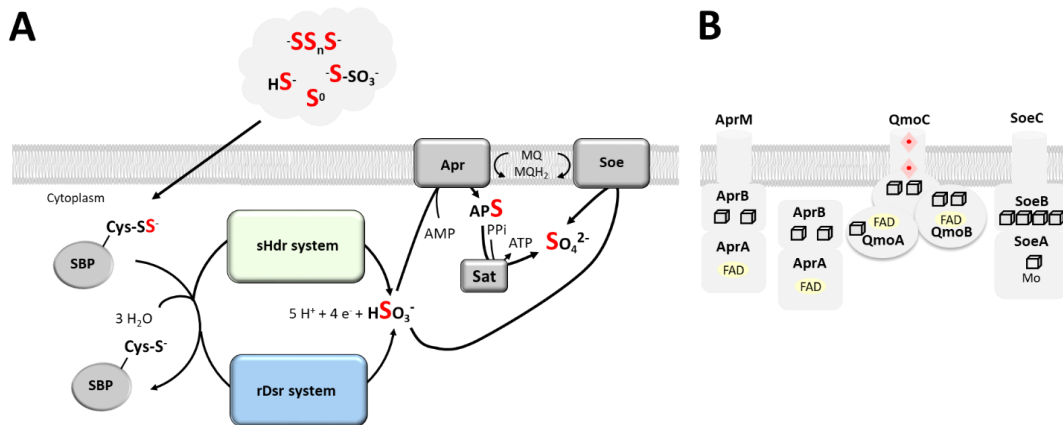
93 While most Fe/S clusters are composed of one to four iron atoms, there are also larger  
94 polynuclear clusters that contain additional metals such as Ni, V or Mo. These clusters are  
95 frequently involved in catalyzing multi-electron redox reactions [10]. Examples include carbon  
96 monoxide dehydrogenases and nitrogenases.

### 97 **3. Pathways of microbial sulfur oxidation**

98 The biogeochemical sulfur cycle is primarily driven by prokaryotes that reduce sulfate or  
99 sulfite [42, 43]. Sulfur-oxidizing microorganisms recycle the reduced sulfur (most reduced  
100 state: sulfide, -II) produced by the aforementioned group back to the oxidized state (most  
101 oxidized state: sulfate, +VI) [17]. Two distinct physiological groups of sulfur oxidizers exist:  
102 chemotrophs utilize electrons obtained from sulfur compounds to conserve energy via  
103 respiration with oxygen, nitrate or Mn(IV) [17, 44, 45], while phototrophs use sulfur  
104 compounds as electron donors for light-driven photosynthetic growth [46]. Sulfur oxidizers are  
105 not only phylogenetically extremely diverse and prevalent across a range of environments such  
106 as hydrothermal vents, sulfur-rich soils, or marine ecosystems including low-temperature  
107 Arctic sediments but also exhibit significant variability in their substrate utilization [17, 47].  
108 Some species specialize in oxidizing hydrogen sulfide (H<sub>2</sub>S), while others can metabolize  
109 thiosulfate (S<sub>2</sub>O<sub>3</sub><sup>2-</sup>), elemental sulfur, or sulfite (SO<sub>3</sub><sup>2-</sup>). Given this diversity, it is unsurprising  
110 that there is no universal mechanism for oxidizing reduced sulfur compounds. Different  
111 modules for oxidizing sulfur compounds can operate alone or in combination, enabling  
112 complete oxidation of the reduced sulfur substrate to sulfate and occasionally even providing  
113 redundant pathways. In view of this complexity, we will refrain from providing a  
114 comprehensive overview and instead guide the reader towards several relevant reviews on  
115 the subject [17, 46, 48-52]. Here, we focus on the steps involving intriguing Fe/S proteins.

116 The initial oxidation of sulfide and thiosulfate is typically catalyzed outside of the  
117 cytoplasm (if present, in the periplasm) (Fig. 2A). Sulfide oxidation involves flavoproteins, i.e.  
118 sulfide:quinone oxidoreductases [53, 54] and/or flavocytochrome *c* sulfide dehydrogenases

119 [55]. Polysulfides are formed as reaction products and are converted into extracellular or  
 120 intracellular sulfur globules by specific groups [48]. Many organisms are capable of mobilizing  
 121 this sulfur and transferring it to the cytoplasm for further oxidation (Fig. 2A). Thiosulfate can  
 122 undergo oxidation to tetrathionate through either a heme *c*-containing dehydrogenase [56,  
 123 57] or thiosulfate:quinone oxidoreductase [58]. In many cases, tetrathionate is not further  
 124 metabolized.



125

126 **Fig. 2. (A)** Sulfur oxidation pathways in the cytoplasm of dissimilatory sulfur-oxidizing  
 127 prokaryotes. Initial steps take place outside of the cytoplasm and are explained in more detail  
 128 in the text. SBP, sulfur trafficking and sulfur binding protein(s). The electrons released by  
 129 formation of hydrogen sulfite by the rDsr or sHdr systems can be fed into respiratory or  
 130 photosynthetic electron transport. The protons involved in quinone/quinol turnover are  
 131 omitted for simplification. sHdr, sulfur-oxidizing heterodisulfide reductase-like; rDsr, reverse-  
 132 acting dissimilatory sulfite reductase; Apr, APS reductase; MQ, menaquinone, Qmo, quinone-  
 133 interacting membrane-bound oxidoreductase; Sat, sulfate adenylyltransferase; Soe, sulfite-  
 134 oxidizing enzyme. **(B)** Schematic representation of cytoplasmic sulfite-oxidizing modules. Mo,  
 135 molybdenum *bis*-pterin guanosine dinucleotide cofactor;  $\square$ , cubane [4Fe-4S] clusters,  $\blacklozenge$ ,  
 136 heme cofactor; FAD, flavin adenine dinucleotide.

137 Alternatively, thiosulfate can be fully oxidized to sulfate via the so-called complete Sox  
 138 system. The *c*-type cytochrome SoxXA catalyzes the oxidative formation of a disulfide bond  
 139 between the sulfane sulfur of thiosulfate and the persulfurated active site cysteine residue of  
 140 SoxY [59]. Then, SoxB catalyzes the hydrolytic release of the sulfone group as sulfate, leaving  
 141 the original sulfane sulfur of thiosulfate bound to SoxY [60, 61]. The reaction cycle can be  
 142 completely executed in the periplasm of organisms that possess the hemomolybdo-protein  
 143 SoxCD. This protein catalyzes the oxidation of the sulfane sulfur bound to SoxY to sulfone,  
 144 which is followed by the hydrolytic release of sulfate catalyzed by SoxB [62]. A second group

145 of organisms possesses a “truncated” Sox system without SoxCD. As a result, the sulfane sulfur  
146 on SoxY cannot be further metabolized in the periplasm. Instead, it is transported by unknown  
147 means into the cytoplasm, where it then becomes a substrate for cytoplasmic oxidation  
148 systems (Fig. 2). Currently, there are no identified enzymes containing Fe/S clusters that  
149 participate in the described reactions outside of the cytoplasm.

150 After sulfur is imported into the cytoplasm, a persulfide shuttling system (SBP in Fig. 2A)  
151 made up of rhodanese-, DsrE- and TusA-like sulfurtransferases specifically delivers it to the  
152 enzymes catalyzing its oxidation to sulfite [50, 63-65]. Two enzyme systems can accomplish  
153 this task: the rDsr pathway, which utilizes the reverse acting dissimilatory sulfite reductase  
154 rDsrAB as its key enzyme [66-69], and the sulfur-oxidizing heterodisulfide reductase-like sHdr  
155 pathway that involves polypeptides resembling the HdrA, HdrB and HdrC subunits of  
156 heterodisulfide reductase from methanogenic archaea [70, 71]. A common feature in both  
157 pathways is the ability to shuttle at least part of the released electrons onto NAD<sup>+</sup>. This  
158 reduces the need for energy-demanding reverse electron flow in lithoautotrophic sulfur  
159 oxidizers. The rDsr and sHdr pathways occur almost exclusively, with only a handful of  
160 organisms possessing the genetic capacity for both oxidation routes [71, 72] (Fig. 3). While the  
161 rDsr pathway is restricted to only nine bacterial phyla (Nitrospinota, Bacteroidota,  
162 Spirochaetota, Actinobacteriota, Myxococcota, three unnamed phyla, and the classes  
163 Magnetococcia, Alphaproteobacteria and Gammaproteobacteria within the phylum  
164 Pseudomonadota), *shdr* gene sets, which can be categorized into two distinct types (refer to  
165 section 6), are broadly distributed (Fig. 3). They occur not only in eighteen of the bacterial  
166 phyla currently listed in the Genome Taxonomy Database (GTDB) but also in two phyla of the  
167 Archaea (Fig. 3) [18]. Both, the rDsr and the sHdr pathways, involve new reactions catalyzed  
168 at uncommon, likely noncubane Fe/S clusters. Details regarding these pathways will be further  
169 discussed below.

170 The conversion of sulfite to sulfate is the final step in sulfur oxidation. While the majority  
171 of bacterial sulfite-oxidizing enzymes are located in the periplasm, use *c*-type cytochromes as  
172 acceptors and do not contain Fe/S clusters [49], some organisms can perform this step in the  
173 cytoplasm (Fig. 2A,B). The membrane-bound, cytoplasmically oriented Fe/S molybdoprotein  
174 SoeABC catalyzes the direct one-step oxidation of sulfite to sulfate. [51, 73], which is thought  
175 to transfer electrons to the quinone pool (Fig. 2B). A number of sulfur-oxidizing bacteria can  
176 also use an indirect pathway via adenosine-5'-phosphosulfate (APS) catalyzed by APS

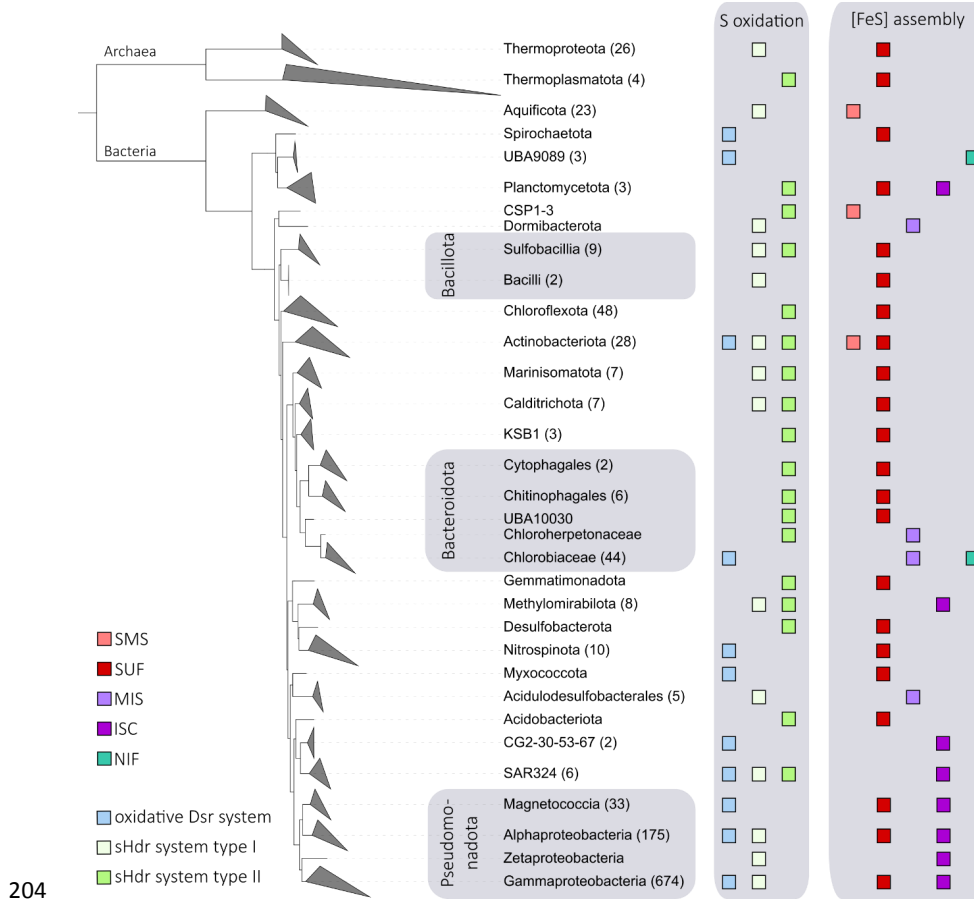
177 reductase (AprBA) and ATP sulfurylase (Sat) (Fig. 2A). Depending on the organism, two  
178 different modules are used to supply electrons to the quinone pool, either AprM [73, 74] or  
179 the QmoABC complex [75, 76] (Fig. 2B). Cubane Fe/S clusters are present in the AprB, QmoA,  
180 QmoB and SoeB polypeptides (Fig. 2B). Since they all have a mere electron-transfer function,  
181 they will not be discussed further.

#### 182 **4. Distribution of Fe/S cluster biogenesis systems in sulfur oxidizers**

183 Fe/S cluster biosynthesis and regeneration mechanisms are vital processes in cells that  
184 guarantee the accessibility of operational Fe/S clusters for diverse biological functions,  
185 including electron transfer and enzyme catalysis. These mechanisms necessitate the  
186 coordinated interplay of diverse proteins and enzymes to ensure the availability and  
187 functionality of Fe/S clusters. Three comparatively well-studied multiprotein Fe/S cluster  
188 biogenesis machineries, ISC, SUF, and NIF [77-80], were recently supplemented by two  
189 “minimal” Fe/S assembly systems, MIS (minimal iron-sulfur) and SMS (SUF-like minimal  
190 system) [81]. MIS consists of a cysteine desulfurase and a scaffold protein for iron-sulfur cluster  
191 assembly, while SMS is formed by a scaffold protein that assembles iron-sulfur clusters. The  
192 more complex ISC, SUF and NIF machineries evolved from the rudimentary MIS and SMS  
193 systems, which can be traced back to the last universal common ancestor, LUCA [81].

194 In contrast to Fe/S cluster biogenesis, the capacity for dissimilatory sulfur metabolism  
195 was not yet established in LUCA. The Dsr system arose after the split of the Bacteria and the  
196 Archaea and appeared first in the domain Archaea. Initially, it operated in a reductive direction  
197 to reduce sulfite to sulfide [82]. The sulfur oxidizing rDsr system has evolved at least twice,  
198 once in the family Chlorobiaceae (phylum Bacteroidota) and once within the phylum  
199 Pseudomonadota (formerly Proteobacteria) [82]. Classical heterodisulfide reductases are  
200 enzymes essential for methanogenesis, a pathway that can be traced back to the last common  
201 ancestor of the Archaea, LACA [83, 84]. Hdr proteins from methanogens probably gave rise to  
202 sHdr components, suggesting that sHdr-dependent sulfur oxidation, similar to the rDsr  
203 pathway, evolved after the divergence of Archaea and Bacteria.





**Fig. 3.** Distribution of enzyme systems for sulfane sulfur oxidation in the cytoplasm. Numbers indicate the number of assemblies positive for rDsr and/or sHdr components within a clade. Archaeal and bacterial genomes were downloaded from GTDB (release R207). One representative of each of the current 65,703 species clusters was analyzed. Open reading frames were determined using Prodigal [85] and subsequently annotated for sulfur-related proteins via HMSS2 [86]. rDsr was set to positive when at least 50% of the genes *rdsrABCEFH* or *rdsrMKJOP* were present in a syntenic block, type I and type II sHdr systems were counted positive when at least 50% of the genes *shdrC1B1AHC2B2* or *sHdrC1B1AHB3etfAB* were present in a syntenic block, respectively. Fe/S biogenesis systems were searched and annotated with Hidden Markov Models (HMM) from TIGRFAMs [87] and Pfam [88] databases with trusted cutoffs. MIS and SMS were identified as syntenic *iscS/iscU*-like homologs and *sufBC* homologs with the same HMMs with a score cutoff of 10. Ribosomal proteins from all identified rDsr and sHdr-containing species were individually aligned, trimmed, concatenated and used for phylogenetic tree inference as described previously [18].

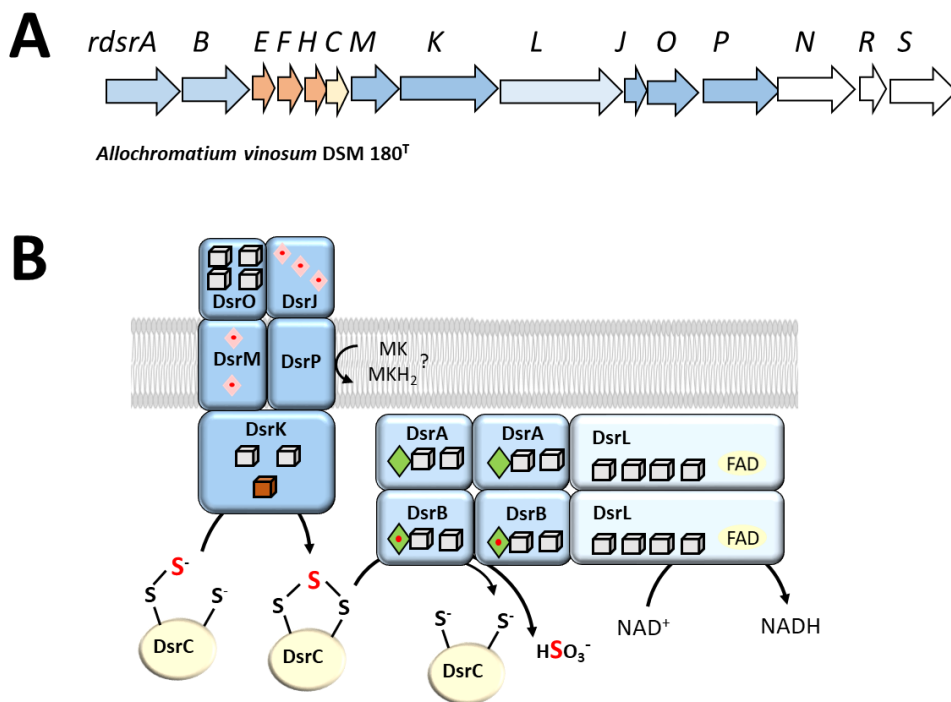
Here, we investigated whether specific Fe/S biogenesis mechanisms are associated with a particular sulfur oxidation pathway. To answer this question, we plotted the presence of all five Fe/S biogenesis machineries onto the tree of sulfur oxidizers that contain rDsr and/or sHdr. As evident from Fig. 3, there is no correlation between the sulfur oxidation pathway and the

223 machinery for Fe/S cluster biosynthesis. The lineages in which the rDsr system originated may  
 224 serve as examples: Within the Chlorobiaceae, with such well-studied representatives as  
 225 *Chlorobaculum tepidum* [89], MIS is the predominant system, while the rDsr-containing  
 226 Pseudomonadota employ SUF and/or ISC but not MIS. Archaeal sulfur oxidizers, including  
 227 *Sulfolobus acidophilus*, *Acidianus hospitalis* and *Metallosphaera cuprina* oxidize sulfur via sHdr  
 228 and possess only the SUF Fe/S biogenesis machinery, while the bacterial members of the  
 229 phylum Aquificae, the order Acididesulfobacterales or the class Zetaproteobacteria contain  
 230 sHdr/SMS, sHdr/MIS or sHdr/ISC combinations, respectively. In summary, our analysis shows  
 231 that ISC, SUF, MIS and SMS are all sufficient to meet the Fe/S cluster maturation needs of sulfur  
 232 oxidizers operating the sHdr or rDsr pathways.

## 233 **5. The role of Fe/S-clusters in rDsr-driven conversion of sulfane sulfur to** 234 **sulfite**

235 A typical *rdsr* gene cluster can consist of up to 15 genes (e.g. *dsrABEFHCMKLJOPNRS* in  
 236 the model organism *Allochromatium vinosum*, Fig. 4A) and includes the genes for the  
 237 sulfurtransferases (SBP in Fig. 2A) which provide rDsrAB with oxidizable sulfur [68, 69, 90-93].  
 238 Sulfane sulfur bound to the sulfur-binding protein DsrC serves as the substrate for reverse  
 239 acting dissimilatory sulfite reductase, rDsrAB (Fig. 4B). The cytoplasmatic enzyme rDsrAB in  
 240 sulfur oxidizers is homologous to DsrAB enzymes that function in sulfite reduction [94] and  
 241 forms a heterotetramer. DsrA and DsrB both carry an iron-tetrahydrophorphyrin that is  
 242 coupled to a [4Fe-4S] cluster through its cysteine heme axial ligand [69, 95]. The siroheme-  
 243 [4Fe-4S] sites in the DsrB subunits are catalytically active, while the ones in DsrA seem to have  
 244 a merely structural role [23, 94]. Both, DsrA and DsrB, also contain a ferredoxin domain binding  
 245 a second [4Fe-4S] cluster that most likely conducts electron transfer between the physiological  
 246 electron donor and the catalytic site. For sulfate reducers, which operate the Dsr system in the  
 247 reductive direction, it has been shown that a DsrC trisulfide, in which a sulfur atom is bridging  
 248 two strictly conserved cysteine residues, is formed as the reaction product of DsrAB [96].  
 249 Furthermore, a recent preprint describes the reduction of the DsrC trisulfide with electrons  
 250 from the menaquinol pool mediated by the electron-transporting membrane-bound  
 251 DsrMKJOP complex from a sulfate reducer [97]. This is fully in line with the previously reported  
 252 interaction of DsrC and DsrK from the phototrophic sulfur oxidizer *Allochromatium vinosum*  
 253 [90]. On this basis, we suggest that DsrMKJOP oxidizes persulfurated DsrC and generates a

254 DsrC trisulfide from which rDsrAB then releases sulfite in a further oxidative step (Fig. 4B).  
 255 Here,  $\text{NAD}^+$  serves as the electron acceptor and is reduced by the iron-sulfur flavoprotein DsrL  
 256 that tightly interacts with rDsrAB [66, 67].



257

258 **Fig. 4.** (A) Gene cluster in the purple sulfur bacterium *Allochromatium vinosum* encoding the  
 259 rDsr system (Alvin\_1251-1265). (B) Model of rDsr mediated sulfane sulfur oxidation in *A.*  
 260 *vinosum*. □, cubane [4Fe-4S] clusters; ▣, noncubane Fe/S center; ◇□, siroheme-[4Fe-4S]  
 261 cofactor; ◇◇, sirohydrochlorine-[4Fe-4S] cofactor; ♦, heme cofactor; FAD, flavin adenine  
 262 dinucleotide; MK, menaquinone.

263 While DsrA, DsrB, DsrO and DsrL contain regular cubane Fe/S clusters with functions in  
 264 electron transport, DsrK presents a very interesting case. As outlined above, it is suggested to  
 265 be the catalytic subunit of the DsrMKJOP complex and to act as a disulfide reductase with  
 266 persulfurated DsrC as its substrate (Fig. 4B). It contains two classical  $\text{CX}_2\text{CX}_2\text{CX}_3\text{C}$  binding sites  
 267 for canonical [4Fe-4S] clusters and one C-terminal cysteine-rich CCG domain with the potential  
 268 for binding a noncubane [4Fe-4S] cluster on the basis of the motif  $\text{CX}_n\text{CCGX}_n\text{CX}_2\text{C}$   
 269 ( $\text{CX}_{43}\text{CCGX}_{40}\text{CX}_2\text{C}$  in *A. vinosum* [68]). The cysteines in the CCG domain are arranged in the  
 270 same way as in the catalytic subunits HdrB or HdrD of heterodisulfide reductases from  
 271 methanogenic archaea [36, 68, 90, 98, 99]. Notably, HdrB and HdrD feature two CCG domains,  
 272 with each domain having five conserved cysteines. Each of these domains ligates one

273 noncubane [4Fe-4S] cluster, which is composed of highly distorted [3Fe-4S] and [2Fe-2S]  
274 subclusters that share an iron and an inorganic sulfur. One of the bridging sulfurs of the [3Fe-  
275 4S] moiety is occupied by a cysteine sulfur [36]. The two noncubane clusters in heterodisulfide  
276 reductases work together in hemolytic cleavage of the CoM-S-S-CoB disulfide bond. The  
277 reaction mechanism involves binding of each sulfur-containing coenzyme to one of the clusters  
278 [36]. As pointed out by Barbosa et al. (2023), an analogous mechanism cannot be applied by  
279 DsrK, because it can only bind a single noncubane Fe/S cluster. Instead, a strictly conserved  
280 cysteine from the missing noncubane cluster is proposed to be involved in the catalytic  
281 mechanism of DsrK-mediated DsrC trisulfide reduction in sulfate reducers [97]. In fact, this  
282 cysteine is also present in DsrK from sulfur oxidizers, supporting this notion.

283 Even though biochemical evidence for some of the assigned functions of Fe/S centers in  
284 the rDsr system is not yet available, they display a variety of different roles, from having a  
285 catalytic function in DsrK and DsrB to electron transfer between physiological electron donors  
286 and acceptors in DsrO and DsrL or even a structural role in DsrA.

## 287 **6. The role of Fe/S-clusters in sHdr-driven conversion of sulfane sulfur to** 288 **sulfite**

289 Clusters of genes encoding the sHdr pathway for sulfane sulfur oxidation in the  
290 cytoplasm fall into two distinct categories (Fig. 5). In both cases, the *shdr* genes are usually  
291 accompanied by genes for sulfur-trafficking proteins, one to three genes for lipolate-binding  
292 proteins (LbpA) and genes for enzymes of lipolate assembly [18, 70, 71]. Like the sHdr proteins  
293 themselves, LbpA2 from the alphaproteobacterial model organism *Hyphomicrobium*  
294 *denitrificans* is indispensable for sulfur oxidation [70, 71].



321 incompatible with basic principles of FBEB [105-107]. The experimentally determined redox  
322 potentials for the [4Fe-4S] cluster and the FAD bound in sHdrA indicate flow of single electrons  
323 from the [4Fe-4S] cluster to FAD, which is the inverse direction of the electron flow reported  
324 for mHdrA [16, 36]. The isoalloxazine ring of FAD is buried inside the sHdrA protein, implying  
325 no access for an external hydride-transferring compound [16]. sHdrA is therefore currently  
326 seen as an electron carrier/storage unit (Fig. 6).

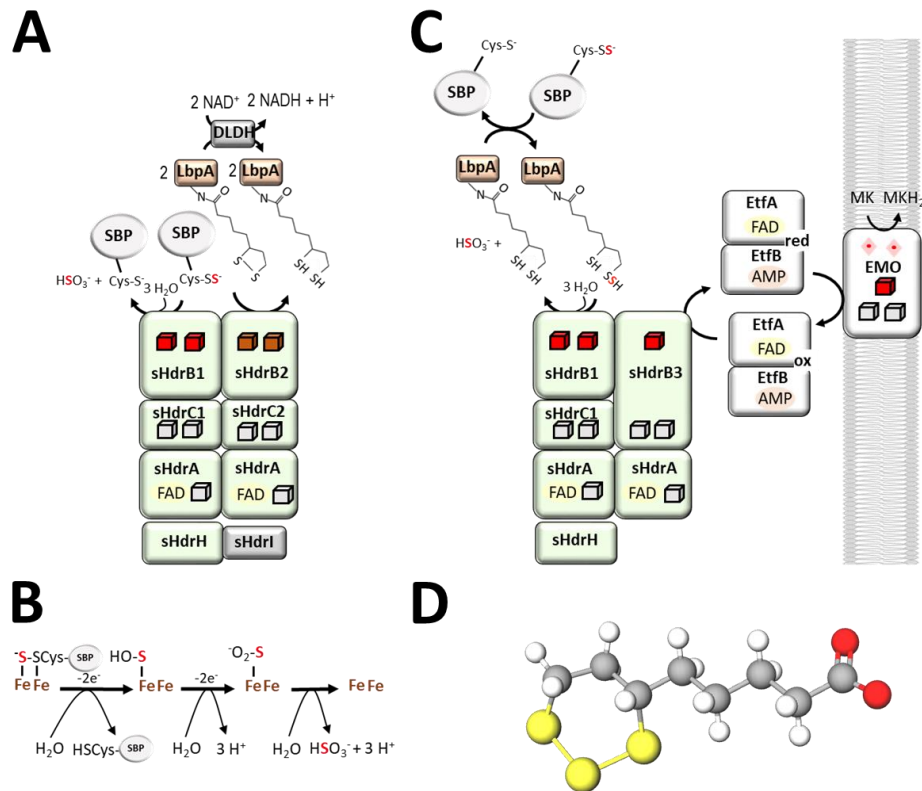
327 sHdrC1 has the capacity to bind two [4Fe-4S] clusters and likely functions as an electron  
328 transfer unit. sHdrB1 is of particular interest due to the noticeable deviations in sequence for  
329 its two noncubane iron-sulfur cluster-binding sites when compared to mHdrB [16, 71]. One of  
330 the five cysteines that ligate the proximal Fe/S cluster in mHdrB has been substituted by an  
331 aspartate, while two of the five cysteines for the distal noncubane Fe/S cluster have been  
332 replaced by serine. Although aspartate and serine are reported to have Fe-ligating capacity [4,  
333 26, 108], it is probable that the properties and catalytic capacities of the noncubane Fe/S  
334 cluster-containing active site of sHdrB1 differ significantly from those of classical mHdrB. We  
335 propose that the oxidation of protein-bound sulfane sulfur to sulfite occurs at this site (Fig. 6A)  
336 and that this represents a novel reaction carried out by Fe/S clusters. The sulfane sulfur and  
337 the protein cysteine sulfur may each be bound to an iron of the two Fe/S centers of sHdrB1  
338 (Fig. 6B). Subsequently, the Fe-ligated sulfane sulfur is oxidized to sulfite, probably with Fe-  
339 ligated sulfenate and sulfinates intermediates similar to what has been described for siroheme  
340 iron [96].

341 The sHdrC2 polypeptide is solely encoded in type I *shdr* gene clusters. It has the capacity  
342 to bind two cubane [4Fe-4S] clusters and is proposed to conduct electrons stored in sHdrA  
343 towards sHdrB2 (Fig. 6A). This protein can be reliably classified as a disulfide reductase due to  
344 the conservation of all ten cysteines ligating the two noncubane [4Fe-4S] clusters in mHdrB.  
345 The exact function of lipoate within the catalytic cycle of sHdr-LbpA-based sulfur oxidation is  
346 currently unclear. In the model put forward by Ernst et al. (2021), it was proposed, that lipoate  
347 may exclusively function as a redox switch in type I sHdr systems (Fig. 6A), accepting the  
348 electrons released at the sHdrB1 active site. The resulting dihydroliponamide would then be  
349 regenerated for the next reaction cycle by NAD<sup>+</sup>-reducing dihydroliponamide dehydrogenase.

350 In type II sHdr systems, genes are absent that encode sHdrC2 and sHdrB2 as separate  
351 entities. Instead, they are substituted by a gene that we term *shdrB3* (Fig. 5). It essentially  
352 encodes a fusion of the two proteins with sHdrC2 with its two cubane Fe/S clusters as the N-

353 terminal and a sHdrB2-like polypeptide as the C-terminal domain. However, there is a  
354 remarkable difference between the sHdrB3 carboxy-terminal domain and sHdrB2. sHdrB3  
355 contains only a single noncubane Fe/S cluster binding motif (CX<sub>35</sub>CCGX<sub>40</sub>CX<sub>2</sub>E in the proteins  
356 from *Kyrpidia tusciae* and *Sulfobacillus acidophilus*, Fig. 5), in which the fifth cysteine is  
357 replaced by glutamate, a residue with reported iron-ligating capacity [4]. Only two cysteines  
358 are retained of the second noncubane Fe/S cluster present in mHdrB. DsrK (see section 5) has  
359 the same overall architecture as sHdrB3. The BamD subunit of the benzoyl CoA reductase from  
360 *Geobacter metallireducens* is an additional example. However, BamD is capable of binding two  
361 noncubane Fe/S clusters, with one of ten cysteines substituted by serine [109].

362 It is feasible that sHdrB3 transfers electrons onto EtfAB, the electron transfer  
363 flavoprotein encoded adjacent to the *hdrB3* gene (Figs. 5 and 6C). Etf's are ubiquitous in nature  
364 and found in all three domain of life, Archaea, Bacteria, and Eukarya [110, 111]. They are  
365 heterodimers of a large  $\alpha$ -subunit (EtfA) and a small  $\beta$ -subunit (EtfB). Nonbifurcating Etf's  
366 contain one FAD and one AMP per heterodimer. In contrast, bifurcating EtfABs contain two  
367 FAD molecules, with the additional FAD replacing the AMP of the canonical Etf [112, 113]. The  
368 residues predicted to coordinate FAD and NADH in bifurcating Etf's [111] are not conserved in  
369 the EtfBs encoded in the type II *shdr* clusters shown in Fig. 5 and we conclude that these Etf's  
370 cannot bifurcate. Furthermore, we propose that the product of an adjacent gene carries out  
371 the electron transfer from Etf to menaquinone (Figs. 5 and 6C). This ~75-kDa protein consists  
372 of an N-terminal domain with five transmembrane helices and the capacity for binding two  
373 hemes *b*, a midsection containing two cubane [4Fe-4S] clusters and a heterodisulfide  
374 reductase-like CCG domain. Here, four of the five cysteines that coordinate the distal  
375 noncubane cluster in HdrB from *Methanothermococcus thermolithotrophicus* [36] are  
376 conserved, while the fifth is replaced by aspartate. This suggests that one noncubane Fe/S  
377 cluster may still be coordinated. All described features match those of  
378 Etf:(methyl)menaquinone oxidoreductases, EMO [114-118]. Many type I sHdr-containing  
379 sulfur oxidizers, such as *Thioalkalivibrio* or *Acidithiobacillus* species, do not possess genes for  
380 EtfAB and EMO.



381

382 **Fig. 6.** Suggestions for reaction mechanisms and molecule structures involved in sHdr-based  
 383 sulfur oxidation **(A)** Model of cytoplasmic sulfane sulfur oxidation by type I sHdr systems. A  
 384 scenario is proposed where protein (SBP) bound persulfide is converted into sulfite in a  
 385 disulfide/thiolate redox reaction. LbpA-bound lipoate would merely act as redox switch and is  
 386 reduced at the two noncubane Fe/S clusters of sHdrB1. The heterohexameric architecture of the  
 387 Hdr(ABC)<sub>2</sub> complex from the methanogen *Methanothermococcus thermolithotrophicus*  
 388 [36] is the basis for suggesting a five subunit sHdrAA'C1B1C2B2 complex. **(B)** Details of the  
 389 proposed four-electron oxidation at the Fe/S clusters of sHdrB1. The electrons are conducted  
 390 away via the ligating irons. **(C)** Model of cytoplasmic sulfane sulfur oxidation by type II sHdr  
 391 systems. A complex composition of sHdrAA'C1B1B3 is suggested because sHdrB3 essentially  
 392 presents a sHdrC2B2 fusion. Here, the possibility is taken into account that the lipoic acid  
 393 moiety serves as a sulfur-binding agent and becomes a sHdrB1 substrate in its persulfurated  
 394 form. In the depicted scenario, LbpA-bound lipoate would not act as a redox switch. Instead,  
 395 the four electrons released in the disulfide/thiolate reaction catalyzed at sHdrB1 would be  
 396 shuttled via EtfAB and EMO into the quinone pool. The reaction at sHdrB1 would follow the  
 397 same principle as illustrated in **(B)**. **(D)** Structure of 1,2,3-trithiane-4-pentanoic acid, which may  
 398 also be an intermediate of sHdr-driven sulfate sulfur oxidation. Color code for atoms: yellow,  
 399 sulfur; grey, carbon; white, hydrogen; red, oxygen. □, cubane [4Fe-4S] clusters; ▢, noncubane  
 400 Fe/S center; ◻, predicted noncubane Fe/S clusters with non-cysteinyll ligands; ◊, heme  
 401 cofactor; FAD, flavin adenine dinucleotide; DLDH, dihydrolipoamide dehydrogenase; EMO,  
 402 ETF:menaquinone oxidoreductase; Etf, electron-transferring flavoprotein; MK, menaquinone.



403 We would like to emphasize here, that the current very limited data basis allows different  
404 mechanistic proposals. For the type II sHdr system, we consider the possibility that LbpA-  
405 bound lipoate might serve as a sulfur substrate binding entity and present sulfur to the sHdrB1  
406 active site (Fig. 6C). Similar to the scenario illustrated in Fig. 6A, the released electrons would  
407 then flow through sHdrC1 and sHdrA, finally arriving at the electron output module, sHdrB3.  
408 We would also like to note that lipoamide trisulfide (Fig. 6D) could potentially serve as an  
409 intermediate in sHdr-based sulfur oxidation. In addition, the possibility of electron bifurcation  
410 at sHdrA should not be disregarded in future investigations, even if it was not considered in  
411 this review.

## 412 **7. Substrate-binding and sulfur-donating role of Fe/S proteins during** 413 **assembly of lipoic acid on its target protein LbpA**

414 The two previous sections discussed the function of Fe/S clusters as electron carriers and  
415 catalytically active sites during sulfur oxidation. We will now show that this metabolic pathway  
416 also involves Fe/S clusters acting as substrate-activating proteins and sulfur-donating entities  
417 (Fig. 1). These two functions are performed by special Fe/S proteins during the assembly of  
418 LbpA-bound lipoic acid, which is crucial for sHdr-based sulfur oxidation (Fig. 6) [18, 70, 71].

419 Lipoic acid is an eight-carbon saturated fatty acid with sulfur atoms at positions 6 and 8  
420 of the acyl chain (Fig. 6). In addition to its importance for sulfur oxidation, it plays a significant  
421 and well-established role in central carbon metabolism [18, 70, 119-121]. Here, lipoic acid-  
422 dependent enzyme systems include three  $\alpha$ -ketoacid dehydrogenases (e.g. pyruvate  
423 dehydrogenase), acetoin dehydrogenase and the glycine cleavage complex [70, 120, 122]. In  
424 all these cases, lipoate functions as an electrophile binding reaction intermediates and  
425 alternates between its reduced and oxidized states.

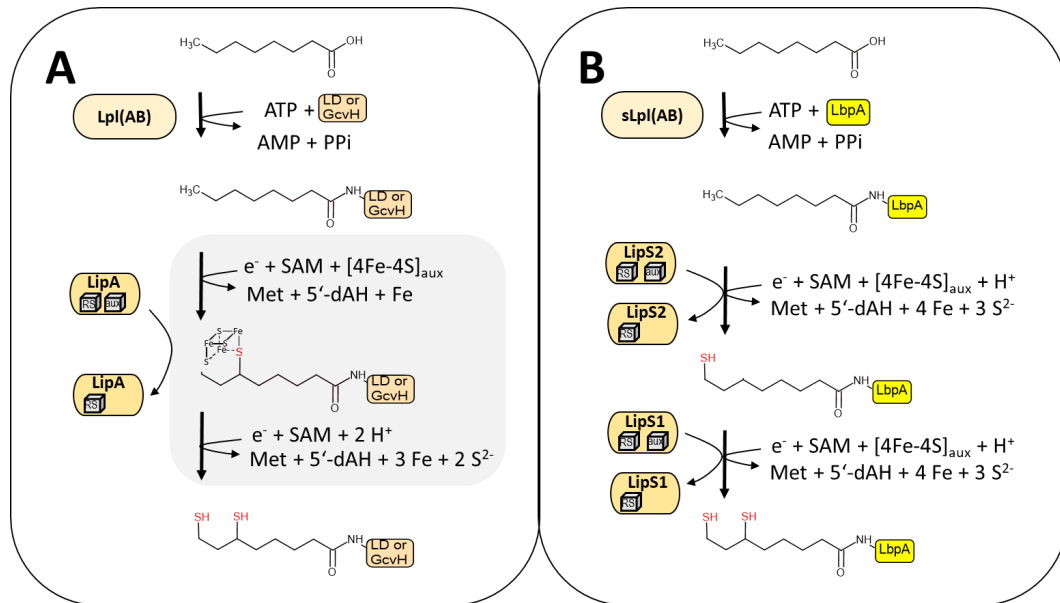
426 Lipoyl moieties are always bound to the  $\epsilon$ -amino groups of conserved lysine residues  
427 within the relevant enzymes [119]. If lipoic acid is not available externally for uptake and  
428 incorporation, it must be synthesized from its precursor octanoic acid. Lipoate assembly starts  
429 either from endogenous fatty acid biosynthesis, utilizing acyl carrier protein (ACP)-bound  
430 octanoyl residues, or from free octanoate [18, 119]. There are two different pathways for  
431 posttranslational assembly of lipoate starting with free octanoate (Fig. 7).

432 The first is outlined in Fig. 7A: In *E. coli* and many other organisms, using free octanoate  
433 involves the enzyme lipoate:protein ligase, Lpl(AB) or in the permutated case Lpl(BA) [18, 123,


434 124]. The ligase consists of two fused domains, the catalytic domain LpIA and the accessory  
435 domain LpIB. Lpl(AB) activates the precursor to octanoyl-AMP at the expense of ATP before  
436 transfer to the target protein [119]. Until recently, only one enzyme had been recognized as  
437 capable of introducing sulfur atoms to the octanoyl residues after their arrival on target  
438 proteins. This enzyme is lipoyl synthase, LipA. Lipoyl synthases constitute one of several classes  
439 of Radical S-adenosyl methionine (SAM) domain-containing enzymes (RS enzymes) [21] that  
440 catalyze sulfur attachment to unactivated carbon centers [125, 126]. Others include biotin  
441 synthase (BioB) which catalyzes the last step in the biosynthesis of biotin. This is the insertion  
442 of a sulfur atom between C<sub>6</sub> and C<sub>9</sub> of dethiobiotin [127]. LipA sequentially adds two sulfur  
443 atoms in a single reaction, first at position C<sub>6</sub> and then at C<sub>8</sub> of the octanoyl moiety [119, 128,  
444 129] (Fig. 7A). RS superfamily members all contain at minimum one radical SAM [4Fe-4S]<sup>2+</sup><sub>RS</sub>  
445 cluster that is ligated by three cysteine residues, usually occurring in a C<sub>X</sub><sub>3</sub>C<sub>X</sub><sub>2</sub>C pattern (Fig.  
446 8A). This ligand coordination is critical for the structural and functional stability of the proteins  
447 belonging to this family [130]. SAM binds to the fourth iron ion in a bidentate manner via its  
448 amino and carboxylate groups. When the cluster is in the [4Fe-4S]<sup>+</sup> state, it causes the  
449 fragmentation of SAM, resulting in the production of a 5'-deoxyadenosyl 5'-radical (5'dA•)-and  
450 methionine [131] (Fig. 7A). In almost all RS reactions, the function of the 5'-dA• is to extract  
451 hydrogen atoms (H) from a substrate. BioB contains an auxiliary [2Fe-2S] cluster in addition to  
452 the [4Fe-4S]<sub>RS</sub> cluster, while LipA contains an additional [4Fe-4S] cluster ([4Fe-4S]<sub>aux</sub>). The  
453 auxiliary clusters within lipoyl synthases exhibit atypical coordination patterns by three  
454 cysteines and a serine residue [39]. In LipA and BioB, the auxiliary cluster undergoes  
455 degradation as part of the reaction mechanism and serves as the source of sulfur atoms. This  
456 sacrificial process results in a limitation imposed upon the enzymes, allowing only a single  
457 complete catalytic turnover. The reconstitution of the LipA auxiliary cluster can be achieved  
458 through the involvement of a specific Fe/S cluster carrier protein known as NfuA [27, 40, 41,  
459 131].

460 The more recently discovered second pathway for lipoate assembly from free octanoate  
461 is depicted in Fig. 7B: A unique type of lipoate:protein ligase, named sLpl(AB), is frequently  
462 encoded in *shdr* genes clusters of sulfur-oxidizing bacteria but not limited to them [18, 70] (Fig.  
463 5). sLpl(AB) ligases not only modify LbpA proteins from the same organism but also show cross-  
464 species functionality among sulfur oxidizers [70]. On the other hand, the canonical Lpl(AB)  
465 ligase from *Streptomyces coelicolor* [124] fails to octanoylate LbpA proteins *in vitro* [70].

466 sLpl(AB) cannot replace *E. coli* Lpl(AB) *in vivo* and it has no activity upon an *E. coli* lipoyl domain  
 467 [70]. So far, nothing is known about the crucial elements that direct the different protein  
 468 substrates into the appropriate octanoylation pathway.



469

470 **Fig. 7. (A)** LipA-involving pathway for lipoate assembly starting with free octanoic acid. See text  
 471 for references. **(B)** LipS1/S2-involving pathway for lipoyl cofactor assembly starting with free  
 472 octanoic acid as established for *H. denitrificans* [18, 70] and *Thermococcus kodakarensis* [132,  
 473 133]. For archaea, bipartite ligases functioning as LplAB heterodimers have been reported  
 474 [134, 135]. See text for further references. , cubane [4Fe-4S] clusters.

475 Conspicuously, *shdr* gene clusters usually contain two genes for proteins of the Radical  
 476 SAM superfamily, LipS1 and LipS2 [18, 70] (Fig. 5). Recently, closely related proteins from the  
 477 thermophilic archaeon *Thermocoococcus kodakarensis* were shown to exert a LipA-like lipoyl  
 478 synthase function on chemically synthesized peptide substrates *in vitro* [132, 133]. Genetic  
 479 and phylogenetic analyses provided further evidence that these proteins are involved not only  
 480 in archaeal but also in bacterial lipoyl biosynthesis [18, 132]. LipS1 and LipS2 bear resemblance  
 481 to each other and also to LipA. This resemblance is confined to the N-terminal part of the  
 482 proteins where an RS Fe/S cluster is bound by cysteines in the characteristic CX<sub>3</sub>CX<sub>2</sub>C  
 483 arrangement (Fig. 8A). A motif ligating the auxiliary cluster in LipA is not present in LipS1 and  
 484 LipS2. Instead the LipS1 and LipS2 proteins possess other conserved motifs including cysteine  
 485 residues in the C-terminal domains, respectively (Fig. 8B,C).

<b>A</b>	MtLipA	74	FLIGGDQCTRRCDFCQIDT		
	HdLipS1	64	FSVTGGACALDCHCQAKI		
	TK90_LipS1	66	MSITGGDCKLACDHCQAKI		
	TkLipS1	20	VSLTGNAALNCAHCGKHY		
	HdLipS2	106	----PEGCRANCAYCGLAR		
	TK90_LipS2	65	----PEGCRANCAYCGLAR		
TkLipS2	36	----PGRCSNDCAFCQAR			
<b>B</b>	HdLipS1	MKIVPHIVIGL 65 aa	LLGCARPAG 34 aa	TYFQEHA	CCSI 330
	TK90_LipS1	MKVVPHIVIGL 66 aa	LLGCARPPG 34 aa	KVRVTPACCSI 342	
	TkLipS1	VRVAPHITIGL 66 aa	SIGCMRPLG 33 aa	DVEIIECCVM 279	
<b>C</b>	HdLipS2	FHRMCIS 98 aa	GAFIIGMGE 96 aa	TSGCPGKFRTDISACDRPY 385	
	TK90_LipS2	FERMCIS 96 aa	GAFIIGMGE 96 aa	TSGCPGKSDEEVSA	CNRPY 342
	TkLipS2	FARICLQ 89 aa	LLVIVVLGE 91 aa	THGCPG-----	CNRPY 283

486

487 **Fig. 8.** Partial alignments of selected LipA, LipS1 and LipS2 proteins. MtLipA, structurally  
 488 characterized *Mycobacterium tuberculosis* LipA (Rv2218) [39]; HdLipS1, *H. denitrificans* LipS1  
 489 (Hden\_0683); TK90\_LipS1, LipS1 from *Thioalkalivibrio* sp. K90mix (TK90\_0641); TkLipS1, LipS1  
 490 from *Thermococcus kodakarensis* (Tk2109); HdLipS2, *H. denitrificans* LipS2 (Hden\_0685);  
 491 TK90\_LipS2, LipS2 from *Thioalkalivibrio* sp. K90mix (TK90\_0644); TkLipS2, LipS2 from  
 492 *Thermococcus kodakarensis* (Tk2248). **(A)** RS-cluster ligating cysteines in LipA, LipS1 and LipS2  
 493 are highlighted in yellow. **(B)** Conserved cysteines and a histidine ligating an auxiliary Fe/S  
 494 cluster in LipS1 proteins are marked in yellow and green, respectively. **(C)** Conserved cysteines  
 495 and a histidine ligating an auxiliary Fe/S cluster in LipS2 proteins are marked in yellow and  
 496 green, respectively.

497 In fact, the cysteine and histidine residues highlighted in Fig. 8, parts B and C, have recently  
 498 been experimentally established as ligating Fe/S clusters in each, LipS1 and LipS2, from *T.*  
 499 *kodakarensis* [133]. In contrast to the canonical lipoyl assembly pathway (Fig. 7A), sulfur  
 500 insertions catalyzed by LipS1/S2 lipoyl synthase take place in two distinct reactions with the  
 501 two structurally different lipoyl synthases, LipS1 and LipS2 (Fig. 7B). During the conversion of  
 502 the octanoyl to a lipoyl moiety, Fe/S clusters are sacrificed. The means by which these clusters  
 503 can be regenerated currently remains a mystery. LipS1 and LipS2 each contribute only one  
 504 sulfur atom to the process [133]. LipS2 acts on an octanoyllysyl substrate to produce an 8-  
 505 mercaptooctanoyllysyl product, which is further transformed into the lipoyl product by the  
 506 action of LipS1 [133] (Fig. 7B).

507 A fascinating issue that remains to be clarified is the origin of electrons needed for  
 508 propelling the reactions catalyzed by the lipoyl synthases LipS1 and LipS2. The cleavage of S-  
 509 adenosylmethionine to generate a 5-dA• requires the input of an electron. In *E. coli* cells, the  
 510 electron for the LipA catalyzed reaction is supplied by NADPH via the flavodoxin/flavodoxin  
 511 reductase reducing system [136]. In many bacteria, the LipS1/S2 lipoyl synthase system

512 includes LipT, a predicted FAD-binding oxidoreductase with the potential to carry out an  
513 analogous function (Fig. 7) [18].

## 514 **8. Perspectives and outlook**

515 Despite significant progress in understanding sulfur oxidation in bacteria, numerous  
516 aspects still require further research. Elucidating the precise mechanisms and evolutionary  
517 history of sulfur oxidation pathways continues to be a vibrant area of scientific exploration.  
518 The exact composition of the functionally active sHdr complex in sulfur oxidizers is still  
519 unresolved and the nature of the Fe/S clusters driving sulfane sulfur oxidation to sulfite has  
520 not been experimentally clarified. The redox properties of the other Fe/S clusters in sHdr  
521 complexes have not been elucidated through biochemical and biophysical methods yet.  
522 Further knowledge gaps include the substrate of the putative disulfide reductase subunit  
523 sHdrB2, the role of the sHdrB3 subunit in type II sHdr systems, the flow of electrons through  
524 the enzyme complex and the possibility of electron bifurcation at the sHdrA FAD. Although the  
525 rDsr system is well studied compared to sulfur oxidation via the sHdr system, there are still  
526 unresolved details regarding this system, particularly concerning the function and mechanism  
527 of the DsrMKJOP complex. Many questions also remain regarding the novel lipoyl assembly  
528 pathway established in sHdr-containing sulfur oxidizers. These include its substrate range, the  
529 mechanisms of differentiation between substrates for lipoylation pathways running in parallel  
530 in the same organisms and the functions of additional proteins encoded in the respective gene  
531 clusters.

## 532 **CRedit authorship contribution statement**

533 Christiane Dahl: Funding acquisition, Supervision.

534 All authors: Conceptualization, Writing, Figure design, Investigation

## 535 **Declaration of competing interest**

536 The authors are not aware of any affiliations, memberships, funding, or financial  
537 holdings that might be perceived as affecting the objectivity of this review.

## 538 **Acknowledgements**

539 We thank the former and current members of our group for their great work and helpful  
540 discussions. CD acknowledges financial support from the Deutsche Forschungsgemeinschaft

541 (grants Da 351/8-2, 13-1 and 14-1). TST obtained a scholarship from the Studienstiftung des  
542 Deutschen Volkes.

### 543 References

- 544 [1] R. Lill, Function and biogenesis of iron-sulphur proteins, *Nature*, 460 (2009) 831-838. doi  
545 10.1038/nature08301
- 546 [2] H. Beinert, R.H. Holm, E. Münck, Iron-sulfur clusters: Nature's modular, multipurpose structures,  
547 *Science*, 277 (1997) 653-659. doi 10.1126/science.277.5326.653
- 548 [3] M. Fontecave, Iron-sulfur clusters: ever-expanding roles, *Nat. Chem. Biol.*, 2 (2006) 171-174. doi  
549 10.1038/nchembio0406-171
- 550 [4] G. Caserta, L. Zuccarello, C. Barbosa, C.M. Silveira, E. Moe, S. Katz, P. Hildebrandt, I. Zebger, S.  
551 Todorovic, Unusual structures and unknown roles of FeS clusters in metalloenzymes seen from  
552 a resonance Raman spectroscopic perspective, *Coord. Chem. Rev.*, 452 (2022) 214287. doi  
553 10.1016/j.ccr.2021.214287
- 554 [5] H. Beinert, Iron-sulfur proteins: ancient structures, still full of surprises, *J. Biol. Inorg. Chem.*, 5  
555 (2000) 2-15. doi 10.1007/s007750050002
- 556 [6] M.K. Johnson, Iron-sulfur proteins: new roles for old clusters, *Curr. Opin. Chem. Biol.*, 2 (1998)  
557 173-181. doi 10.1016/s1367-5931(98)80058-6
- 558 [7] D.C. Johnson, D.R. Dean, A.D. Smith, M.K. Johnson, Structure, function, and formation of  
559 biological iron-sulfur clusters, *Ann. Rev. Biochem.*, 74 (2005) 247-281. doi  
560 10.1146/annurev.biochem.74.082803.133518
- 561 [8] S. Dai, R. Friemann, D.A. Glauser, F. Bourquin, W. Manieri, P. Schurmann, H. Eklund, Structural  
562 snapshots along the reaction pathway of ferredoxin-thioredoxin reductase, *Nature*, 448 (2007)  
563 92-96. doi 10.1038/nature05937
- 564 [9] W.H. Tong, T.A. Rouault, Metabolic regulation of citrate and iron by aconitases: role of iron-sulfur  
565 cluster biogenesis, *Biometals*, 20 (2007) 549-564. doi 10.1007/s10534-006-9047-6
- 566 [10] D.C. Rees, Great metalloclusters in enzymology, *Ann. Rev. Biochem.*, 71 (2002) 221-246. doi  
567 10.1146/annurev.biochem.71.110601.135406
- 568 [11] P.J. Kiley, H. Beinert, The role of Fe-S proteins in sensing and regulation in bacteria, *Curr. Opin.*  
569 *Microbiol.*, 6 (2003) 181-185. doi 10.1016/s1369-5274(03)00039-0
- 570 [12] R. Rohac, J.C. Crack, E. de Rosny, O. Gigarel, N.E. Le Brun, J.C. Fontecilla-Camps, A. Volbeda,  
571 Structural determinants of DNA recognition by the NO sensor NsrR and related Rrf2-type [FeS]-  
572 transcription factors, *Commun. Biol.*, 5 (2022) 769. doi 10.1038/s42003-022-03745-7
- 573 [13] B.B. Jørgensen, A.J. Findlay, A. Pellerin, The biogeochemical sulfur cycle of marine sediments,  
574 *Front. Microbiol.*, 10 (2019) 849. doi 10.3389/fmicb.2019.00849
- 575 [14] Y. Shen, R. Buick, D.E. Canfield, Isotopic evidence for microbial sulphate reduction in the early  
576 Archaean era, *Nature*, 410 (2001) 77-81. doi 10.1038/35065071
- 577 [15] J.R. Havig, T.L. Hamilton, A. Bachan, L.R. Kump, Sulfur and carbon isotopic evidence for metabolic  
578 pathway evolution and a four-stepped Earth system progression across the Archean and  
579 Paleoproterozoic, *Earth Sci. Rev.*, 174 (2017) 1-21. doi 10.1016/j.earscirev.2017.06.014
- 580 [16] C. Ernst, K. Kayashta, T. Koch, S.S. Venceslau, I.A.C. Pereira, U. Demmer, U. Ermler, C. Dahl,  
581 Structural and spectroscopic characterization of a HdrA-like subunit from *Hyphomicrobium*  
582 *denitrificans*, *FEBS J.*, 288 (2021) 1664-1678. doi 10.1111/febs.15505
- 583 [17] C. Dahl, A biochemical view on the biological sulfur cycle, in: P. Lens (Ed.) *Environmental*  
584 *technologies to treat sulfur pollution: principles and engineering*, IWA Publishing, Place  
585 Published, 2020, pp. 55-96.
- 586 [18] T.S. Tanabe, M. Grosser, L. Hahn, C. Kümpel, H. Hartenfels, E. Vtulkin, W. Flegler, C. Dahl,  
587 Identification of a novel lipic acid biosynthesis pathway reveals the complex evolution of lipoate  
588 assembly in prokaryotes, *PLoS Biol.*, 21 (2023) e3002177. doi 10.1371/journal.pbio.3002177

- 589 [19] S.N. Khan, A. Griffith, F. De Proft, E. Miliordos, R.W.A. Havenith, D. Bykov, A.V. Cunha, [Fe<sub>4</sub>S<sub>4</sub>]  
590 cubane in sulfite reductases: new insights into bonding properties and reactivity, *Phys Chem*  
591 *Chem Phys*, 24 (2022) 18543-18551. doi 10.1039/d2cp02124b
- 592 [20] D.R. Colman, G. Labesse, G.V.T. Swapna, J. Stefanakis, G.T. Montelione, E.S. Boyd, C.A. Royer,  
593 Structural evolution of the ancient enzyme, dissimilatory sulfite reductase, *Proteins*, (2022). doi  
594 10.1002/prot.26315
- 595 [21] M. Jespersen, A.J. Pierik, T. Wagner, Structures of the sulfite detoxifying F(420)-dependent  
596 enzyme from Methanococcales, *Nat. Chem. Biol.*, (2023). doi 10.1038/s41589-022-01232-y
- 597 [22] T.F. Oliveira, E. Franklin, J.P. Afonso, A.R. Khan, N.J. Oldham, I.A.C. Pereira, M. Archer, Structural  
598 insights into dissimilatory sulfite reductases: structure of desulforubidin from *Desulfomicrobium*  
599 *norvegicum*, *Front. Microbiol.*, 2 (2011) 71. doi 10.3389/fmicb.2011.00071
- 600 [23] T.F. Oliveira, C. Vonrhein, P.M. Matia, S.S. Venceslau, I.A.C. Pereira, M. Archer, The crystal  
601 structure of *Desulfovibrio vulgaris* dissimilatory sulfite reductase bound to DsrC provides novel  
602 insights into the mechanism of sulfate respiration, *J. Biol. Chem.*, 283 (2008) 34141-34149. doi  
603 10.1074/jbc.M805643200
- 604 [24] A. Volbeda, M.T.P. Martinez, J.C. Crack, P. Amara, O. Gigarel, J.T. Munnoch, M.I. Hutchings, C.  
605 Darnault, N.E. Le Brun, J.C. Fontecilla-Camps, Crystal structure of the transcription regulator RsrR  
606 reveals a [2Fe-2S] cluster coordinated by Cys, Glu, and His residues, *J. Am. Chem. Soc.*, 141 (2019)  
607 2367-2375. doi 10.1021/jacs.8b10823
- 608 [25] D.W. Bak, S.J. Elliott, Alternative FeS cluster ligands: tuning redox potentials and chemistry, *Curr.*  
609 *Opin. Chem. Biol.*, 19 (2014) 50-58. doi 10.1016/j.cbpa.2013.12.015
- 610 [26] L. Calzolari, C.M. Gorst, Z.H. Zhao, Q. Teng, M.W.W. Adams, G.N. Lamar, <sup>1</sup>H NMR investigation of  
611 the electronic and molecular structure of the four-iron cluster ferredoxin from the  
612 hyperthermophile *Pyrococcus furiosus*. Identification of Asp 14 as a cluster ligand in each of the  
613 four redox states, *Biochemistry*, 34 (1995) 11373-11384. doi
- 614 [27] E.L. McCarthy, S.J. Booker, Destruction and reformation of an iron-sulfur cluster during catalysis  
615 by lipoyl synthase, *Science*, 358 (2017) 373-377. doi 10.1126/science.aan4574
- 616 [28] M.N. Pinto, J. Ter Beek, L.A. Ekanger, E. Johansson, J.K. Barton, The [4Fe4S] cluster of yeast DNA  
617 polymerase epsilon is redox active and can undergo DNA-mediated signaling, *J. Am. Chem. Soc.*,  
618 143 (2021) 16147-16153. doi 10.1021/jacs.1c07150
- 619 [29] R.M.B. Silva, M.A. Grodick, J.K. Barton, UvrC coordinates an O<sub>2</sub>-sensitive [4Fe4S] cofactor, *J. Am.*  
620 *Chem. Soc.*, 142 (2020) 10964-10977. doi 10.1021/jacs.0c01671
- 621 [30] O.A. Lukianova, S.S. David, A role for iron-sulfur clusters in DNA repair, *Curr. Opin. Chem. Biol.*, 9  
622 (2005) 145-151. doi 10.1016/j.cbpa.2005.02.006
- 623 [31] M.F. White, M.S. Dillingham, Iron-sulphur clusters in nucleic acid processing enzymes, *Curr Opin*  
624 *Struct Biol*, 22 (2012) 94-100. doi 10.1016/j.sbi.2011.11.004
- 625 [32] N.B. Ugulava, K.K. Frederick, J.T. Jarrett, Control of adenosylmethionine-dependent radical  
626 generation in biotin synthase: a kinetic and thermodynamic analysis of substrate binding to  
627 active and inactive forms of BioB, *Biochemistry*, 42 (2003) 2708-2719. doi 10.1021/bi0261084
- 628 [33] G. Layer, D.W. Heinz, D. Jahn, W.D. Schubert, Structure and function of radical SAM enzymes,  
629 *Curr. Opin. Chem. Biol.*, 8 (2004) 468-476. doi 10.1016/j.cbpa.2004.08.001
- 630 [34] S. Dai, C. Schwendtmayer, P. Schurmann, S. Ramaswamy, H. Eklund, Redox signaling in  
631 chloroplasts: cleavage of disulfides by an iron-sulfur cluster, *Science*, 287 (2000) 655-658. doi  
632 10.1126/science.287.5453.655
- 633 [35] E.C. Duin, S. Madadi-Kahkesh, R. Hedderich, M.D. Clay, M.K. Johnson, Heterodisulfide reductase  
634 from *Methanothermobacter marburgensis* contains an active-site [4Fe-4S] cluster that is directly  
635 involved in mediating heterodisulfide reduction, *FEBS Lett.*, 512 (2002) 263-268. doi  
636 10.1016/S0014-5793(02)02281-0
- 637 [36] T. Wagner, J. Koch, U. Ermler, S. Shima, Methanogenic heterodisulfide reductase (HdrABC-  
638 MvhAGD) uses two noncubane [4Fe-4S] clusters for reduction, *Science*, 357 (2017) 699-703. doi  
639 10.1126/science.aan0425

- 640 [37] E.M. Walters, M.K. Johnson, Ferredoxin:thioredoxin reductase: Disulfide reduction catalyzed via  
641 novel site-specific [4Fe-4S] cluster chemistry, *Photosynth. Res.*, 79 (2004) 249-264. doi  
642 10.1023/B:PRES.0000017195.05870.61
- 643 [38] F. Berkovitch, Y. Nicolet, J.T. Wan, J.T. Jarrett, C.L. Drennan, Crystal structure of biotin synthase,  
644 an S-adenosylmethionine-dependent radical enzyme, *Science*, 303 (2004) 76-79. doi  
645 10.1126/science.1088493
- 646 [39] M.I. McLaughlin, N.D. Lanz, P.J. Goldman, K.H. Lee, S.J. Booker, C.L. Drennan, Crystallographic  
647 snapshots of sulfur insertion by lipoyl synthase, *Proc. Natl. Acad. Sci. USA*, 113 (2016) 9446-9450.  
648 doi 10.1073/pnas.1602486113
- 649 [40] E.L. McCarthy, S.J. Booker, Biochemical approaches for understanding iron-sulfur cluster  
650 regeneration in *Escherichia coli* lipoyl synthase during catalysis, *Meth. Enzymol.*, 606 (2018) 217-  
651 239. doi 10.1016/bs.mie.2018.06.006
- 652 [41] E.L. McCarthy, A.N. Rankin, Z.R. Dill, S.J. Booker, The A-type domain in *Escherichia coli* NfuA is  
653 required for regenerating the auxiliary [4Fe-4S] cluster in *Escherichia coli* lipoyl synthase, *J. Biol.*  
654 *Chem.*, 294 (2019) 1609-1617. doi 10.1074/jbc.RA118.006171
- 655 [42] R. Rabus, S.S. Venceslau, L. Wohlbrand, G. Voordouw, J.D. Wall, I.A. Pereira, A post-genomic view  
656 of the ecophysiology, catabolism and biotechnological relevance of sulphate-reducing  
657 prokaryotes, *Adv. Microb. Physiol.*, 66 (2015) 55-321. doi 10.1016/bs.ampbs.2015.05.002
- 658 [43] K. Wasmund, M. Mussmann, A. Loy, The life sulfuric: microbial ecology of sulfur cycling in marine  
659 sediments, *Environ. Microbiol.*, 9 (2017) 323-344. doi 10.1111/1758-2229.12538
- 660 [44] J.V. Henkel, O. Dellwig, F. Pollehne, D.P.R. Herlemann, T. Leipe, H.N. Schulz-Vogt, A bacterial  
661 isolate from the Black Sea oxidizes sulfide with manganese(IV) oxide, *Proc. Natl. Acad. Sci. USA*,  
662 116 (2019) 12153-12155. doi 10.1073/pnas.1906000116
- 663 [45] J.V. Henkel, H.N. Schulz-Vogt, O. Dellwig, F. Pollehne, T. Schott, C. Meeske, S. Beier, K. Jürgens,  
664 Biological manganese-dependent sulfide oxidation impacts elemental gradients in redox-  
665 stratified systems: indications from the Black Sea water column, *ISME J.*, (2022). doi  
666 10.1038/s41396-022-01200-3
- 667 [46] C. Dahl, Sulfur metabolism in phototrophic bacteria, in: P.C. Hallenbeck (Ed.) *Modern topics in*  
668 *the phototrophic prokaryotes: Metabolism, bioenergetics and omics*, Springer International  
669 Publishing, Place Published, 2017, pp. 27-66.
- 670 [47] C. Dahl, C.G. Friedrich, A. Kletzin, Sulfur oxidation in prokaryotes, *Encyclopedia of Life Sciences*,  
671 John Wiley & Sons, Ltd., Place Published, 2008.
- 672 [48] C. Dahl, Bacterial intracellular sulfur globules, in: D. Jendrossek (Ed.) *Bacterial organelles and*  
673 *organelle-like inclusions*, Springer, Place Published, 2020, pp. 19-51.
- 674 [49] U. Kappler, J.H. Enemark, Sulfite-oxidizing enzymes, *J. Biol. Inorg. Chem.*, 20 (2014) 253-264. doi  
675 10.1007/s00775-014-1197-3
- 676 [50] T.S. Tanabe, S. Leimkühler, C. Dahl, The functional diversity of the prokaryotic sulfur carrier  
677 protein TusA, *Adv. Microb. Physiol.*, 75 (2019) 233-277. doi 10.1016/bs.ampbs.2019.07.004
- 678 [51] R. Wang, J.-Q. Lin, X.-M. Liu, X. Pang, C.-J. Zhang, C.-L. Yang, X.-Y. Gao, C.-M. Lin, Y.-Q. Li, Y. Li, J.-  
679 Q. Lin, L.-X. Chen, Sulfur oxidation in the acidophilic autotrophic *Acidithiobacillus* spp, *Front.*  
680 *Microbiol.*, 9 (2019) 3290. doi 10.3389/fmicb.2018.03290
- 681 [52] T.S. Tanabe, C. Dahl, HMS-S-S: a tool for the identification of sulphur metabolism-related genes  
682 and analysis of operon structures in genome and metagenome assemblies, *Mol. Ecol. Resour.*,  
683 22 (2022) 2758-2774. doi 10.1111/1755-0998.13642
- 684 [53] M. Marcia, U. Ermler, G.H. Peng, H. Michel, A new structure-based classification of  
685 sulfide:quinone oxidoreductases, *Proteins: Structure Function and Bioinformatics*, 78 (2010)  
686 1073-1083. doi 10.1002/prot.22665
- 687 [54] M. Marcia, U. Ermler, G.H. Peng, H. Michel, The structure of *Aquifex aeolicus* sulfide:quinone  
688 oxidoreductase, a basis to understand sulfide detoxification and respiration, *Proc. Natl. Acad. Sci.*  
689 *USA*, 106 (2009) 9625-9630. doi 10.1073/pnas.0904165106
- 690 [55] M.A. Cusanovich, T.E. Meyer, R.G. Bartsch, Flavocytochrome *c*, in: F. Muller (Ed.) *Chemistry and*  
691 *Biochemistry of Flavoenzymes*, CRC Press, Place Published, 1991, pp. 377-399.



- 692 [56] K. Denkmann, F. Grein, R. Zigann, A. Siemen, J. Bergmann, S. van Helmont, A. Nicolai, I.A.C.  
693 Pereira, C. Dahl, Thiosulfate dehydrogenase: a wide-spread unusual acidophilic c-type  
694 cytochrome, *Environ. Microbiol.*, 14 (2012) 2673-2688. doi 10.1111/j.1462-2920.2012.02820.x
- 695 [57] J.M. Kurth, J.A. Brito, J. Reuter, A. Flegler, T. Koch, T. Franke, E.M. Klein, S.F. Rowe, J.N. Butt, K.  
696 Denkmann, I.A.C. Pereira, M. Archer, C. Dahl, Electron accepting units of the diheme cytochrome  
697 c TsdA, a bifunctional thiosulfate dehydrogenase/tetrathionate reductase, *J. Biol. Chem.*, 291  
698 (2016) 24804-24818. doi 10.1074/jbc.M116.753863
- 699 [58] F.H. Müller, T.M. Bandejas, T. Urich, M. Teixeira, A. Kletzin, Coupling of the pathway of sulphur  
700 oxidation to dioxygen reduction: characterization of a novel membrane-bound  
701 thiosulphate:quinone oxidoreductase, *Mol. Microbiol.*, 53 (2004) 1147-1160. doi  
702 10.1111/j.1365-2958.2004.04193.x
- 703 [59] D.B. Grabarczyk, B.C. Berks, Intermediates in the Sox sulfur oxidation pathway are bound to a  
704 sulfane conjugate of the carrier protein SoxYZ, *PLoS One*, 12 (2017) e0173395. doi  
705 10.1371/journal.pone.0173395
- 706 [60] V. Sauvé, P. Roversi, K.J. Leath, E.F. Garman, R. Antrobus, S.M. Lea, B.C. Berks, Mechanism for the  
707 hydrolysis of a sulfur-sulfur bond based on the crystal structure of the thiosulfohydrolase SoxB,  
708 *J. Biol. Chem.*, 284 (2009) 21707-21718. doi 10.1074/jbc.M109.002709
- 709 [61] D.B. Grabarczyk, P.E. Chappell, S. Johnson, L.S. Stelzl, S.M. Lea, B.C. Berks, Structural basis for  
710 specificity and promiscuity in a carrier protein/enzyme system from the sulfur cycle, *Proc. Natl.*  
711 *Acad. Sci. USA*, 112 (2015) E7166-E7175. doi 10.1073/pnas.1506386112
- 712 [62] U. Zander, A. Faust, B.U. Klink, D. de Sanctis, S. Panjikar, A. Quentmeier, F. Bardischewsky, C.G.  
713 Friedrich, A.J. Scheidig, Structural basis for the oxidation of protein-bound sulfur by the sulfur  
714 cycle molybdohemo-enzyme sulfane dehydrogenase SoxCD, *J. Biol. Chem.*, 286 (2010) 8349-  
715 8360. doi 10.1074/jbc.M110.193631
- 716 [63] C. Dahl, Cytoplasmic sulfur trafficking in sulfur-oxidizing prokaryotes, *IUBMB Life*, 67 (2015) 268-  
717 274. doi 10.1002/iub.1371
- 718 [64] L.J. Liu, Y. Stockdreher, T. Koch, S.T. Sun, Z. Fan, M. Josten, H.G. Sahl, Q. Wang, Y.M. Luo, S.J. Liu,  
719 C. Dahl, C.Y. Jiang, Thiosulfate transfer mediated by DsrE/TusA homologs from acidothermophilic  
720 sulfur-oxidizing archaeon *Metallosphaera cuprina*, *J. Biol. Chem.*, 289 (2014) 26949-26959. doi  
721 10.1074/jbc.M114.591669
- 722 [65] S.S. Venceslau, Y. Stockdreher, C. Dahl, I.A.C. Pereira, The "bacterial heterodisulfide" DsrC is a key  
723 protein in dissimilatory sulfur metabolism, *Biochim. Biophys. Acta*, 1837 (2014) 1148-1164. doi  
724 10.1016/j.bbabi.2014.03.007
- 725 [66] M. Löffler, J. Feldhues, S.S. Venceslau, L. Kammler, F. Grein, I.A.C. Pereira, C. Dahl, DsrL mediates  
726 electron transfer between NADH and rDsrAB in *Allochromatium vinosum*, *Environ. Microbiol.*, 22  
727 (2020) 783-795. doi 10.1111/1462-2920.14899
- 728 [67] M. Löffler, K. Wallerang, S.S. Venceslau, I.A.C. Pereira, C. Dahl, The iron-sulfur flavoprotein DsrL  
729 as NAD(P)H:acceptor oxidoreductase in oxidative and reductive dissimilatory sulfur metabolism,  
730 *Front. Microbiol.*, 11 (2020) 578209. doi 10.3389/fmicb.2020.578209
- 731 [68] C. Dahl, S. Engels, A.S. Pott-Sperling, A. Schulte, J. Sander, Y. Lübbe, O. Deuster, D.C. Brune, Novel  
732 genes of the *dsr* gene cluster and evidence for close interaction of Dsr proteins during sulfur  
733 oxidation in the phototrophic sulfur bacterium *Allochromatium vinosum*, *J. Bacteriol.*, 187 (2005)  
734 1392-1404. doi 10.1128/JB.187.4.1392-1404.2005
- 735 [69] A.S. Pott, C. Dahl, Sirohaem-sulfite reductase and other proteins encoded in the *dsr* locus of  
736 *Chromatium vinosum* are involved in the oxidation of intracellular sulfur, *Microbiology*, 144  
737 (1998) 1881-1894. doi 10.1099/00221287-144-7-1881
- 738 [70] X. Cao, T. Koch, L. Steffens, J. Finkensieper, R. Zigann, J.E. Cronan, C. Dahl, Lipoate-binding  
739 proteins and specific lipoate-protein ligases in microbial sulfur oxidation reveal an atypical role  
740 for an old cofactor, *eLife*, 7 (2018) e37439. doi 10.7554/eLife.37439
- 741 [71] T. Koch, C. Dahl, A novel bacterial sulfur oxidation pathway provides a new link between the  
742 cycles of organic and inorganic sulfur compounds, *ISME J.*, 12 (2018) 2479-2491. doi  
743 10.1038/s41396-018-0209-7

- 744 [72] T. Berben, L. Overmars, D.Y. Sorokin, G. Muyzer, Diversity and distribution of sulfur oxidation-  
745 related genes in *Thioalkalivibrio*, a genus of chemolithoautotrophic and haloalkaliphilic sulfur-  
746 oxidizing bacteria, *Front. Microbiol.*, 10 (2019) 160. doi 10.3389/fmicb.2019.00160
- 747 [73] C. Dahl, B. Franz, D. Hensen, A. Kesselheim, R. Zigann, Sulfite oxidation in the purple sulfur  
748 bacterium *Allochromatium vinosum*: Identification of SoeABC as a major player and relevance of  
749 SoxYZ in the process, *Microbiology*, 159 (2013) 2626-2638. doi 10.1099/mic.0.071019-0
- 750 [74] B. Meyer, J. Kuever, Phylogeny of the alpha and beta subunits of the dissimilatory adenosine-5'-  
751 phosphosulfate (APS) reductase from sulfate-reducing prokaryotes - origin and evolution of the  
752 dissimilatory sulfate-reduction pathway, *Microbiology*, 153 (2007) 2026-2044. doi  
753 10.1099/mic.0.2006/003152-0
- 754 [75] A.G. Duarte, A.A. Santos, I.A. Pereira, Electron transfer between the QmoABC membrane  
755 complex and adenosine 5'-phosphosulfate reductase, *Biochim. Biophys. Acta*, 1857 (2016) 380-  
756 386. doi 10.1016/j.bbabi.2016.01.001
- 757 [76] L. Appel, M. Willistein, C. Dahl, U. Ermler, M. Boll, Functional diversity of prokaryotic HdrA(BC)  
758 modules: role in flavin-based electron bifurcation processes and beyond, *Biochim. Biophys. Acta*  
759 *Bioenerget.*, 1862 (2021) 148379. doi 10.1016/j.bbabi.2021.148379
- 760 [77] F.W. Outten, Recent advances in the Suf Fe-S cluster biogenesis pathway: Beyond the  
761 Proteobacteria, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853 (2015)  
762 1464-1469. doi 10.1016/j.bbamcr.2014.11.001
- 763 [78] J.H. Kim, J.R. Bothe, T.R. Alderson, J.L. Markley, Tangled web of interactions among proteins  
764 involved in iron-sulfur cluster assembly as unraveled by NMR, SAXS, chemical crosslinking, and  
765 functional studies, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1853 (2015)  
766 1416-1428. doi 10.1016/j.bbamcr.2014.11.020
- 767 [79] B. Roche, L. Aussel, B. Ezraty, P. Mandin, B. Py, F. Barras, Reprint of: Iron/sulfur proteins  
768 biogenesis in prokaryotes: formation, regulation and diversity, *Biochim. Biophys. Acta -*  
769 *Bioenergetics*, 1827 (2013) 923-937. doi 10.1016/j.bbabi.2013.05.001
- 770 [80] J.J. Braymer, S.A. Freibert, M. Rakwalska-Bange, R. Lill, Mechanistic concepts of iron-sulfur  
771 protein biogenesis in Biology, *Biochimica et Biophysica Acta (BBA) Molecular Cell Research*, 1868  
772 (2021) 118863. doi 10.1016/j.bbamcr.2020.118863
- 773 [81] P.S. Garcia, F. D'Angelo, S. Ollagnier de Choudens, M. Dussouchaud, E. Bouveret, S. Gribaldo, F.  
774 Barras, An early origin of iron-sulfur cluster biosynthesis machineries before Earth oxygenation,  
775 *Nat Ecol Evol*, 6 (2022) 1564-1572. doi 10.1038/s41559-022-01857-1
- 776 [82] S. Neukirchen, I.A.C. Pereira, F.L. Sousa, Stepwise pathway for early evolutionary assembly of  
777 dissimilatory sulfite and sulfate reduction, *ISME J.*, 17 (2023) 1680-1692. doi 10.1038/s41396-  
778 023-01477-y
- 779 [83] R. Mei, M. Kaneko, H. Imachi, M.K. Nobu, The origin and evolution of methanogenesis and  
780 Archaea are intertwined, *PNAS Nexus*, 2 (2023) pgad023. doi 10.1093/pnasnexus/pgad023
- 781 [84] W.F. Martin, M.C. Weiss, S. Neukirchen, S. Nelson-Sathi, F.L. Sousa, Physiology, phylogeny, and  
782 LUCA, *Microbial Cell*, 3 (2016) 582-587. doi 10.15698/mic2016.12.545
- 783 [85] D. Hyatt, G.L. Chen, P.F. Locascio, M.L. Land, F.W. Larimer, L.J. Hauser, Prodigal: prokaryotic gene  
784 recognition and translation initiation site identification, *BMC Bioinformatics*, 11 (2010) 119. doi  
785 10.1186/1471-2105-11-119
- 786 [86] T.S. Tanabe, C. Dahl, HMSS2: an advanced tool for the analysis of sulfur metabolism, including  
787 organosulfur compound transformation, in genome and metagenome assemblies, *Mol. Ecol.*  
788 *Resour.*, 00 (2023) 1-16. doi 10.1111/1755-0998.13848
- 789 [87] W. Li, K.R. O'Neill, D.H. Haft, M. DiCuccio, V. Chetvernin, A. Badretdin, G. Coulouris, F. Chitsaz,  
790 M.K. Derbyshire, A.S. Durkin, N.R. Gonzales, M. Gwadz, C.J. Lanczycki, J.S. Song, N. Thanki, J.  
791 Wang, R.A. Yamashita, M. Yang, C. Zheng, A. Marchler-Bauer, F. Thibaud-Nissen, RefSeq:  
792 expanding the prokaryotic genome annotation pipeline reach with protein family model  
793 curation, *Nucleic Acids Res.*, 49 (2021) D1020-D1028. doi 10.1093/nar/gkaa1105

- 794 [88] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto,  
795 L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman, Pfam: The protein families database in  
796 2021, *Nucleic Acids Res.*, 49 (2021) D412-D419. doi 10.1093/nar/gkaa913
- 797 [89] L.G. Falkenby, M. Szymanska, C. Holkenbrink, K.S. Habicht, J.S. Andersen, M. Miller, N.U. Frigaard,  
798 Quantitative proteomics of *Chlorobaculum tepidum*: insights into the sulfur metabolism of a  
799 phototrophic green sulfur bacterium, *FEMS Microbiol. Lett.*, 323 (2011) 142-150. doi
- 800 [90] F. Grein, I.A. Pereira, C. Dahl, Biochemical characterization of individual components of the  
801 *Allochromatium vinosum* DsrMKJOP transmembrane complex aids understanding of complex  
802 function *in vivo*, *J. Bacteriol.*, 192 (2010) 6369-6377. doi 10.1128/JB.00849-10
- 803 [91] F. Grein, S.S. Venceslau, L. Schneider, P. Hildebrandt, S. Todorovic, I.A.C. Pereira, C. Dahl, DsrJ, an  
804 essential part of the DsrMKJOP complex in the purple sulfur bacterium *Allochromatium vinosum*,  
805 is an unusual triheme cytochrome c, *Biochemistry*, 49 (2010) 8290-8299. doi 10.1021/bi1007673
- 806 [92] Y. Stockdreher, M. Sturm, M. Josten, H.G. Sahl, N. Dobler, R. Zigann, C. Dahl, New proteins  
807 involved in sulfur trafficking in the cytoplasm of *Allochromatium vinosum*, *J. Biol. Chem.*, 289  
808 (2014) 12390-12403. doi 10.1074/jbc.M113.536425
- 809 [93] Y. Stockdreher, S.S. Venceslau, M. Josten, H.G. Sahl, I.A.C. Pereira, C. Dahl, Cytoplasmic  
810 sulfurtransferases in the purple sulfur bacterium *Allochromatium vinosum*: evidence for sulfur  
811 transfer from DsrEFH to DsrC, *PLoS One*, 7 (2012) e40785. doi 10.1371/journal.pone.0040785
- 812 [94] A. Schiffer, K. Parey, E. Warkentin, K. Diederichs, H. Huber, K.O. Stetter, P.M.H. Kroneck, U. Ermler,  
813 Structure of the dissimilatory sulfite reductase from the hyperthermophilic archaeon  
814 *Archaeoglobus fulgidus*, *J. Mol. Biol.*, 379 (2008) 1063-1074. doi 10.1016/j.jmb.2008.04.027
- 815 [95] C. Dahl, N.M. Kredich, R. Deutzmann, H.G. Trüper, Dissimilatory sulphite reductase from  
816 *Archaeoglobus fulgidus*: physico-chemical properties of the enzyme and cloning, sequencing and  
817 analysis of the reductase genes, *J. Gen. Microbiol.*, 139 (1993) 1817-1828. doi
- 818 [96] A.A. Santos, S.S. Venceslau, F. Grein, W.D. Leavitt, C. Dahl, D.T. Johnston, I.A. Pereira, A protein  
819 trisulfide couples dissimilatory sulfate reduction to energy conservation, *Science*, 350 (2015)  
820 1541-1545. doi 10.1126/science.aad3558
- 821 [97] A.C. Barbosa, S.S. Venceslau, I.C. Pereira, DsrMKJOP is the terminal reductase complex in  
822 anaerobic sulfate respiration, *bioRxiv*, (2023). doi 10.1101/2023.08.03.551783
- 823 [98] R.H. Pires, S.S. Venceslau, F. Morais, M. Teixeira, A.V. Xavier, I.A.C. Pereira, Characterization of the  
824 *Desulfovibrio desulfuricans* ATCC 27774 DsrMKJOP complex - a membrane-bound redox complex  
825 involved in the sulfate respiratory pathway, *Biochemistry*, 45 (2006) 249-262. doi  
826 10.1021/bi0515265
- 827 [99] G.J. Mander, E.C. Duin, D. Linder, K.O. Stetter, R. Hedderich, Purification and characterization of  
828 a membrane-bound enzyme complex from the sulfate-reducing archaeon *Archaeoglobus*  
829 *fulgidus* related to heterodisulfide reductase from methanogenic archaea *Eur J Biochem*, 269  
830 (2002) 1895-1904. doi 10.1046/j.1432-1033.2002.02839.x
- 831 [100] M.S. Finnin, J.R. Donigian, A. Cohen, V.M. Richon, R.A. Rifkind, P.A. Marks, R. Breslow, N.P.  
832 Pavletich, Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors,  
833 *Nature*, 401 (1999) 188-193. doi 10.1038/43710
- 834 [101] M. Winkler, K. Dokulil, H. Weber, T. Pavkov-Keller, B. Wilding, The nitrile-forming enzyme 7-  
835 cyano-7-deazaguanine synthase from *Geobacillus kaustophilus*: a reverse nitrilase?,  
836 *ChemBioChem*, 16 (2015) 2373-2378. doi 10.1002/cbic.201500335
- 837 [102] V. Anantharaman, L. Aravind, MOSC domains: ancient, predicted sulfur-carrier domains, present  
838 in diverse metal-sulfur cluster biosynthesis proteins including Molybdenum cofactor sulfurases,  
839 *FEMS Microbiol. Lett.*, 207 (2002) 55-61. doi 10.1111/j.1574-6968.2002.tb11028.x
- 840 [103] S. Weinitschke, K. Denger, A.M. Cook, T.H.M. Smits, The DUF81 protein TauE in *Cupriavidus*  
841 *necator* H16, a sulfite exporter in the metabolism of C<sub>2</sub> sulfonates, *Microbiology*, 153 (2007)  
842 3055-3060. doi 10.1099/mic.0.2007/009845-0
- 843 [104] A.K. Kaster, J. Moll, K. Parey, R.K. Thauer, Coupling of ferredoxin and heterodisulfide reduction  
844 via electron bifurcation in hydrogenotrophic methanogenic archaea, *Proc. Natl. Acad. Sci. USA*,  
845 108 (2011) 2981-2986. doi 10.1073/pnas.1016761108

- 846 [105] W. Nitschke, M.J. Russell, Redox bifurcations: mechanisms and importance to life now, and at its  
847 origin: a widespread means of energy conversion in biology unfolds, *Bioessays*, 34 (2012) 106-  
848 109. doi 10.1002/bies.201100134
- 849 [106] W. Buckel, R.K. Thauer, Energy conservation via electron bifurcating ferredoxin reduction and  
850 proton/ $\text{Na}^+$  translocating ferredoxin oxidation, *Biochim. Biophys. Acta*, 1827 (2013) 94-113. doi  
851 10.1016/j.bbabi.2012.07.002
- 852 [107] C.E. Lubner, D.P. Jennings, D.W. Mulder, G.J. Schut, O.A. Zadvornyy, J.P. Hoben, M. Tokmina-  
853 Lukaszewska, L. Berry, D.M. Nguyen, G.L. Lipscomb, B. Bothner, A.K. Jones, A.F. Miller, P.W. King,  
854 M.W.W. Adams, J.W. Peters, Mechanistic insights into energy conservation by flavin-based  
855 electron bifurcation, *Nat. Chem. Biol.*, 13 (2017) 655-659. doi 10.1038/nchembio.2348
- 856 [108] J.E. Harmer, M.J. Hiscox, P.C. Dinis, S.J. Fox, A. Iliopoulos, J.E. Hussey, J. Sandy, F.T. Van Beek, J.W.  
857 Essex, P.L. Roach, Structures of lipoyl synthase reveal a compact active site for controlling  
858 sequential sulfur insertion reactions, *Biochem. J.*, 464 (2014) 123-133. doi 10.1042/BJ20140895
- 859 [109] S.G. Huwiler, C. Löffler, S.E.L. Anselmann, H.J. Stark, M. von Bergen, J. Flechsler, R. Rachel, M.  
860 Boll, One-megadalton metalloenzyme complex in *Geobacter metallireducens* involved in  
861 benzene ring reduction beyond the biological redox window, *Proc. Natl. Acad. Sci. USA*, (2019).  
862 doi 10.1073/pnas.1819636116
- 863 [110] M.H. Tsai, M.H. Saier, Jr., Phylogenetic characterization of the ubiquitous electron transfer  
864 flavoprotein families ETF- $\alpha$  and ETF- $\beta$ , *Research in Microbiology*, 146 (1995) 397-404. doi  
865 10.1016/0923-2508(96)80285-3
- 866 [111] A.M. Garcia Costas, S. Poudel, A.F. Miller, G.J. Schut, R.N. Ledbetter, K.R. Fixen, L.C. Seefeldt,  
867 M.W.W. Adams, C.S. Harwood, E.S. Boyd, J.W. Peters, Defining electron bifurcation in the  
868 electron-transferring flavoprotein family, *J. Bacteriol.*, 199 (2017) e00440-00417. doi  
869 10.1128/JB.00440-17
- 870 [112] N.P. Chowdhury, A.M. Mowafy, J.K. Demmer, V. Upadhyay, S. Koelzer, E. Jayamani, J. Kahnt, M.  
871 Hornung, U. Demmer, U. Ermler, W. Buckel, Studies on the mechanism of electron bifurcation  
872 catalyzed by electron transferring flavoprotein (Etf) and butyryl-CoA dehydrogenase (Bcd) of  
873 *Acidaminococcus fermentans*, *J. Biol. Chem.*, 289 (2014) 5145-5157. doi  
874 10.1074/jbc.M113.521013
- 875 [113] G.J. Schut, N. Mohamed-Raseek, M. Tokmina-Lukaszewska, D.W. Mulder, D.M.N. Nguyen, G.L.  
876 Lipscomb, J.P. Hoben, A. Patterson, C.E. Lubner, P.W. King, J.W. Peters, B. Bothner, A.-F. Miller,  
877 M.W.W. Adams, The catalytic mechanism of electron-bifurcating electron transfer flavoproteins  
878 (ETFs) involves an intermediary complex with  $\text{NAD}^+$ , *J. Biol. Chem.*, 294 (2019) 3271-3283. doi  
879 10.1074/jbc.RA118.005653
- 880 [114] M. Agne, L. Appel, C. Seelmann, M. Boll, Enoyl-coenzyme A respiration via formate cycling in  
881 syntrophic bacteria, *mBio*, 13 (2021) e0374021. doi 10.1128/mbio.03740-21
- 882 [115] M. Agne, S. Estelmann, C.S. Seelmann, J. Kung, D. Wilkens, H.G. Koch, C. van der Does, S.V. Albers,  
883 C. von Ballmoos, J. Simon, M. Boll, The missing enzymatic link in syntrophic methane formation  
884 from fatty acids, *Proc. Natl. Acad. Sci. USA*, 118 (2021). doi 10.1073/pnas.2111682118
- 885 [116] J. Heider, M. Szalaniec, B.M. Martins, D. Seyhan, W. Buckel, B.T. Golding, Structure and function  
886 of benzylsuccinate synthase and related fumarate-adding glycy radical enzymes, *Journal of*  
887 *Molecular Microbiology and Biotechnology*, 26 (2016) 29-44. doi 10.1159/000441656
- 888 [117] M.S. Vogt, K. Schuhle, S. Kolzer, P. Peschke, N.P. Chowdhury, D. Kleinsorge, W. Buckel, L.O. Essen,  
889 J. Heider, Structural and functional characterization of an electron transfer flavoprotein involved  
890 in toluene degradation in strictly anaerobic bacteria, *J. Bacteriol.*, 201 (2019). doi  
891 10.1128/JB.00326-19
- 892 [118] L. Wöhlbrand, J.H. Jacob, M. Kube, M. Mussmann, R. Jarling, A. Beck, R. Amann, H. Wilkes, R.  
893 Reinhardt, R. Rabus, Complete genome, catabolic sub-proteomes and key-metabolites of  
894 *Desulfobacula toluolica* Tol2, a marine, aromatic compound-degrading, sulfate-reducing  
895 bacterium, *Environ. Microbiol.*, 15 (2013) 1334-1355. doi 10.1111/j.1462-2920.2012.02885.x

- 896 [119] J.E. Cronan, Assembly of lipoic acid on its cognate enzymes: an extraordinary and essential  
897 biosynthetic pathway, *Microbiol. Mol. Biol. Rev.*, 80 (2016) 429-450. doi 10.1128/MMBR.00073-  
898 15
- 899 [120] M.D. Spalding, S.T. Prigge, Lipoic acid metabolism in microbial pathogens, *Microbiol. Mol. Biol.*  
900 *Rev.*, 74 (2010) 200-228. doi 10.1128/MMBR.00008-10
- 901 [121] M.F. Dunn, Vitamin formation from fatty acid precursors, in: O. Geiger (Ed.) *Biogenesis of Fatty*  
902 *Acids, Lipids and Membranes*, Springer Nature Switzerland AG 2019, Place Published, 2019, pp.  
903 259-271.
- 904 [122] R.N. Perham, Swinging arms and swinging domains in multifunctional enzymes: catalytic  
905 machines for multistep reactions, *Ann. Rev. Biochem.*, 69 (2000) 961-1004. doi  
906 10.1146/annurev.biochem.69.1.961
- 907 [123] D.E. Green, T.W. Morris, J. Green, J.E. Cronan, Jr., J.R. Guest, Purification and properties of the  
908 lipoate protein ligase of *Escherichia coli*, *Biochem. J.*, 309 ( Pt 3) (1995) 853-862. doi  
909 10.1042/bj3090853
- 910 [124] X. Cao, J.E. Cronan, The *Streptomyces coelicolor* lipoate-protein ligase is a circularly permuted  
911 version of the *Escherichia coli* enzyme composed of discrete interacting domains, *J. Biol. Chem.*,  
912 290 (2015) 7280-7290. doi 10.1074/jbc.M114.626879
- 913 [125] N.D. Lanz, S.J. Booker, Auxiliary iron-sulfur cofactors in radical SAM enzymes, *Biochimica et*  
914 *Biophysica Acta (BBA) - Molecular Cell Research*, 1853 (2015) 1316-1334. doi  
915 10.1016/j.bbamcr.2015.01.002
- 916 [126] P.A. Frey, S.J. Booker, Radical mechanisms of S-adenosylmethionine-dependent enzymes, *Adv.*  
917 *Prot. Chem.*, 58 (2001) 1-45. doi 10.1016/s0065-3233(01)58001-8
- 918 [127] C.J. Fugate, J.T. Jarrett, Biotin synthase: insights into radical-mediated carbon-sulfur bond  
919 formation, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1824 (2012) 1213-  
920 1322. doi 10.1016/j.bbapap.2012.01.010
- 921 [128] P. Douglas, M. Kriek, P. Bryant, P.L. Roach, Lipoyl synthase inserts sulfur atoms into an octanoyl  
922 substrate in a stepwise manner, *Angew Chem Int Ed Engl*, 45 (2006) 5197-5199. doi  
923 10.1002/anie.200601910
- 924 [129] N.D. Lanz, M.E. Pandelia, E.S. Kakar, K.H. Lee, C. Krebs, S.J. Booker, Evidence for a catalytically  
925 and kinetically competent enzyme-substrate cross-linked intermediate in catalysis by lipoyl  
926 synthase, *Biochemistry*, 53 (2014) 4557-4572. doi 10.1021/bi500432r
- 927 [130] G.L. Holliday, E. Akiva, E.C. Meng, S.D. Brown, S. Calhoun, U. Pieper, A. Sali, S.J. Booker, P.C.  
928 Babbitt, Atlas of the radical SAM superfamily: divergent evolution of function using a "Plug and  
929 Play" domain, *Meth. Enzymol.*, 606 (2018) 1-71. doi 10.1016/bs.mie.2018.06.004
- 930 [131] S.J. Booker, C.T. Lloyd, Twenty years of Radical SAM! The genesis of the superfamily, *ACS Bio &*  
931 *Med Chem Au*, 2 (2022) 538-547. doi 10.1021/acsbiochem.2c00078
- 932 [132] J.Q. Jin, S.I. Hachisuka, T. Sato, T. Fujiwara, H. Atomi, A structurally novel lipoyl synthase in the  
933 hyperthermophilic archaeon *Thermococcus kodakarensis*, *Appl. Environ. Microbiol.*, 86 (2020).  
934 doi 10.1128/AEM.01359-20
- 935 [133] S.S. Neti, D. Sil, D.M. Warui, O.A. Esakova, A.E. Solinski, D.A. Serrano, C. Krebs, S.J. Booker,  
936 Characterization of LipS1 and LipS2 from *Thermococcus kodakarensis*: Proteins annotated as  
937 biotin synthases, which together catalyze formation of the lipoyl cofactor, *ACS Bio Med Chem*  
938 *Au*, 2 (2022) 509-520. doi 10.1021/acsbiochem.2c00018
- 939 [134] Q.H. Christensen, J.E. Cronan, The *Thermoplasma acidophilum* LplA-LplB complex defines a new  
940 class of bipartite lipoate-protein ligases, *J. Biol. Chem.*, 284 (2009) 21317-21326. doi  
941 10.1074/jbc.M109.015016
- 942 [135] J.Q. Jin, T. Sato, S.I. Hachisuka, H. Atomi, A lipoate-protein ligase is required for de novo lipoyl-  
943 protein biosynthesis in the hyperthermophilic archaeon *Thermococcus kodakarensis*, *Appl.*  
944 *Environ. Microbiol.*, 88 (2022) e0064422. doi 10.1128/aem.00644-22
- 945 [136] R.M. Cicchillo, D.F. Iwig, A.D. Jones, N.M. Nesbitt, C. Baleanu-Gogonea, M.G. Souder, L. Tu, S.J.  
946 Booker, Lipoyl synthase requires two equivalents of S-adenosyl-L-methionine to synthesize one  
947 equivalent of lipoic acid, *Biochemistry*, 43 (2004) 6378-6386. doi 10.1021/bi049528x



---

## A metabolic puzzle: Consumption of C1 compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>

Li, J., Koch, J., Flegler, W., Garcia Ruiz, L., Hager, N., Ballas, A., Tanabe, T. S., & Dahl, C.

---

*Hyphomicrobium denitrificans* X<sup>T</sup> (ATCC 51888) is an Alphaproteobacterium commonly found in various environments, including soils, brackish water, sewage water, and fresh water. Several carbon compounds can be utilized by *H. denitrificans* X<sup>T</sup> including methanol, formate, methylamine, dimethylamine and dimethylsulfide during chemoorganoheterotrophic growth. *H. denitrificans* cannot utilize carbon sources with more than two carbon atoms, nor can it grow autotrophically with carbon dioxide. It is an excellent model organism for studying the sHdr system, as it does not rely on dissimilatory sulfur oxidation for energy conversion, but uses reduced sulfur compounds as an additional electron donor during methylotrophic growth (Koch & Dahl 2018, Li et al. 2023). Although thiosulfate is oxidized to sulfate by *H. denitrificans*, this additional electron source did not increase the growth rate (Koch & Dahl 2018). Instead the addition of thiosulfate caused a significant growth retardation. The following article aimed to explain.

The initial decomposition of thiosulfate to sulfate and protein bound sulfane sulfur is proposed to be catalyzed by an truncated Sox system (Koch & Dahl 2018). This hypothesis was tested through a reverse genetic approach. The thiosulfate oxidation through the Sox system is initiated by the c-type cytochrome SoxXA that forms a disulfide between the active site cysteine residue of SoxYZ and the thiosulfate sulfane sulfur (Sauvé et al. 2007). Then, the protein bound sulfone group (SoxZY-SSO<sub>3</sub>) is then hydrolytically cleaved by SoxB releasing sulfate and SoxYZ-bound sulfane sulfur (Sauvé et al. 2009). Truncated Sox systems lack the sulfur dehydrogenase SoxCD, which is required for the oxidation to the sulfonate state (Zander et al. 2011). In *H. denitrificans* the SoxY-bound sulfane sulfur is transferred to the cytoplasm by an unknown mechanism and oxidized via the cytoplasmic sHdr pathway. The involvement of the truncated Sox system in the early stages of thiosulfate oxidation was confirmed by individual in-frame deletion of *soxXA* and *soxYZ* (Li et al. 2023). The generated strains *H. denitrificans*  $\Delta$ *tsdA*  $\Delta$ *soxXA* and *H. denitrificans*  $\Delta$ *tsdA*  $\Delta$ *soxYZ* were unable to oxidize thiosulfate proving the essential function of the truncated Sox system. Furthermore, the occurrence of the thiosulfate oxidation in the genus *Hyphomicrobium* was assessed and described (Li et al. 2023).

In a second approach the regulation of the sHdr pathway was investigated. The transcription of the *shdr* gene cluster in *H. denitrificans* was proposed to be regulated by ArsR-type regulator sHdrR (Koch & Dahl 2018). Here, the thiosulfate consumption of the strains *H. denitrificans*  $\Delta$ *tsdA*  $\Delta$ *shdrR* and strains *H. denitrificans*  $\Delta$ *tsdA* were compared. When grown on methanol and thiosulfate, *H.*

---

Li, J., Koch, J., Flegler, W., Garcia Ruiz, L., Hager, N., Ballas, A., Tanabe, T. S., & Dahl, C. (2022). A metabolic puzzle: Consumption of C1 compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>. *Biochim. Biophys. Acta Bioenerg.*, 1864(1), 148932; doi.org/10.1016/j.bbabo.2022.148932

*denitrificans*  $\Delta tsdA \Delta shdrR$  consumed thiosulfate with a much higher specific oxidation rate and showed a significantly reduced growth rate compared to the reference strain *H. denitrificans*  $\Delta tsdA$ . The specific oxidation rate was even higher when *H. denitrificans*  $\Delta tsdA \Delta shdrR$  was exposed to thiosulfate prior to the growth experiment, indicating the existence of an additional regulatory process (Li *et al.* 2023).

The additional electron source was considered to inhibit the assimilation of carbon from methanol. As a reduced carbon compound methanol is oxidized by *H. denitrificans* in a five step reaction to formate, which is then assimilated through the serine pathway. It has been hypothesized that the thiosulfate oxidation over-reduces the cellular nicotinamide dinucleotide and cytochrome *c* pools. Therefore, the electron flow from methanol to the cytochrome *c* pool and ultimately the oxidation and assimilation of methanol for biomass production is inhibited. Indeed a growth inhibition was not observed when the carbon source was changed from methanol to formate, which can be directly assimilated via the serine pathway and does not require initial oxidation. However, the  $NAD^+/NADH$  ration was significantly higher when *H. denitrificans*  $\Delta tsdA \Delta shdrR$  was grown on either formate or methanol without thiosulfate, while the addition of thiosulfate decreased the  $NAD^+/NADH$  ratio. This contradicted the hypothesis of an inhibition of carbon assimilation by over-reduction. Therefore the formation of an inhibitory intermediate was investigated, that interferes with the methanol but not with the formate assimilation. This approach revealed a previously unreported toxic sulfite formation and excretion in the periplasm (Koch & Dahl 2018). This sulfur compounds likely inhibits the periplasmic methanol dehydrogenase and consequently methanol assimilation. It was rationalized that the cytoplasmic formate assimilation is protected from the periplasmic sulfite through the cell membrane and sulfite exporters. Thus, growth inhibition could be attributed to cytoplasmically generated toxic sulfite that is excreted into the periplasm (Li *et al.* 2023).

T.S.T. contributed to this by conceptualization and investigation: T.S.T. contributed to the conceptualization of the *soxXA* and *soxYZ* in-frame deletion and the investigation of thiosulfate oxidation systems in *Hyphomicrobium* *sp.*





## A metabolic puzzle: Consumption of C<sub>1</sub> compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>

Jingjing Li, Julian Koch, Wanda Flegler, Leon Garcia Ruiz, Natalie Hager, Alina Ballas, Tomohisa S. Tanabe, Christiane Dahl\*

Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany

### ARTICLE INFO

#### Keywords:

*Hyphomicrobium denitrificans*

Sulfur oxidation

Thiosulfate

Serine pathway

Methanol degradation

Formate assimilation

C<sub>1</sub> metabolism

sHdr pathway

### ABSTRACT

Many obligately heterotrophic methylotrophs oxidize thiosulfate as an additional electron source during growth on C<sub>1</sub> compounds. Although two different pathways of thiosulfate oxidation are implemented in *Hyphomicrobium denitrificans* X<sup>T</sup>, a pronounced negative effect on growth rate is observed when it is cultured in the simultaneous presence of methanol and thiosulfate. In this model organism, periplasmic thiosulfate dehydrogenase TsdA catalyzes formation of the dead-end product tetrathionate. By reverse genetics we verified the second pathway that also starts in the periplasm where SoxXA catalyzes the oxidative fusion of thiosulfate to SoxYZ, from which sulfate is released by SoxB. Sulfane sulfur is then further oxidized in the cytoplasm by the sulfur-oxidizing heterodisulfide reductase-like system (sHdr) which is produced constitutively in a strain lacking the transcriptional repressor sHdrR. When exposed to thiosulfate, the  $\Delta$ sHdrR strain exhibited a strongly reduced growth rate even without thiosulfate in the pre-cultures. When grown on methanol, cells exhibit significantly increased NAD<sup>+</sup>/NADH ratios in the presence of thiosulfate. In contrast, thiosulfate did not exert any negative effect on growth rate or increase NAD<sup>+</sup> levels during growth on formate. On both C<sub>1</sub> substrates, excretion of up to 0.5 mM sulfite as an intermediate of thiosulfate (2 mM) oxidation was recorded. Sulfite is known to form adducts with pyrroloquinoline quinone, the cofactor of periplasmic methanol dehydrogenase. We rationalize that this causes specific inhibition of methanol degradation in the presence of thiosulfate while formate metabolism in the cytoplasm remains unaffected.

### 1. Introduction

In respiratory organisms, energy can be conserved from electrons flowing from reduced inorganic or organic compounds (litho- vs. organotrophy) to more oxidized acceptor molecules. While electron donors of autotrophic organisms – be it organic or inorganic – are oxidized but not assimilated, electron donors of heterotrophic organisms (e.g. sugars) are assimilated into biomass and serve as carbon sources. In nature, where bacteria encounter multiple energy and/or carbon sources at the same time, separation between the different metabolic modes is not strict. Instead, many bacteria can draw on different sources of electrons in parallel and may not thrive as exclusive organohetero- or lithoautotrophs, especially when their substrates are present only in low concentrations [1,2]. A number of these organisms can use chemolithoheterotrophy, a mixed metabolic mode in which an organocarbon compound is used with simultaneous oxidation of an inorganic

species such as thiosulfate, Mn<sup>2+</sup> or molecular hydrogen as an auxiliary electron donor. True chemolithoheterotrophs can thereby generate additional proton-motive force ( $\Delta p$ ) or sodium motive force ( $\Delta Na^+$ ) used in turn to generate ATP [3–7].

Based on observations of phenotypes, two physiological groups of obligately heterotrophic sulfur-oxidizing bacteria have classically been distinguished [8]: the long-known organisms mostly encountered in marine or saline environments which oxidize sulfur compounds incompletely to tetrathionate [7,9–13] and the increasing number of those organisms that oxidize sulfur compounds all the way to sulfate [14–18]. In 2018, we described that the Alphaproteobacterium *Hyphomicrobium denitrificans* X<sup>T</sup> (DSM 1869<sup>T</sup>, ATCC 51888<sup>T</sup>) is capable of both, formation of tetrathionate and sulfate from thiosulfate (Fig. 1) and that the product formed is dependent on the initial substrate concentration in batch culture [19]. The organism is a representative of the Hyphomicrobiaceae (order Hyphomicrobiales), a family of

\* Corresponding author at: Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 168, D-53115 Bonn, Germany.

E-mail address: [ChDahl@uni-bonn.de](mailto:ChDahl@uni-bonn.de) (C. Dahl).

<https://doi.org/10.1016/j.bbabio.2022.148932>

Received 24 February 2022; Received in revised form 17 October 2022; Accepted 24 October 2022

Available online 30 October 2022

0005-2728/© 2022 Elsevier B.V. All rights reserved.

phenotypically quite diverse, mostly aerobic bacteria [20]. Like other *Hyphomicrobium* species, the appendaged, budding *H. denitrificans* is ubiquitous in soils as well as fresh and brackish waters. As a restricted facultative methylotroph it can neither grow autotrophically nor on compounds with three or more carbon atoms [21]. Highest growth yields are reached with methanol or methylamine(s) [22,23]. The substrate range of *Hyphomicrobium* spp. further includes formate, acetate, ethanol, dimethyl sulfoxide or dimethyl sulfide and can vary even within strains of the same species [21].

It is well established that selected *Hyphomicrobium* strains and also representatives of other families of the Hyphomicrobiales use inorganic sulfur compounds like thiosulfate and sulfide as additional electron donors during methylotrophic growth, thereby increase their growth yield and thus appear as true chemolithoheterotrophs [15,26–28]. However, although even two different pathways of thiosulfate oxidation are implemented in *H. denitrificans* X<sup>T</sup> (Fig. 1), no growth benefits from the inorganic sulfur compound could be detected in batch culture [19]. To the contrary, a significant reduction in growth rate was observed in the simultaneous presence of thiosulfate and methanol or methylamine [19,25]. Here, we aimed to explain this puzzling observation by combining several different experimental strategies. First, the involvement of Sox proteins in the initial degradation of thiosulfate in the periplasm was rigorously verified by reverse genetics. Second, information was collected on the regulation of the sHdr pathway. Third, growth experiments were performed with informative mutant strains on different C<sub>1</sub> compounds. In this context, the genetic basis of C<sub>1</sub> metabolism in *H. denitrificans* was analyzed in detail. To shed first light on the bioenergetics of the interplay between sulfur compound oxidation and

C<sub>1</sub> carbon compound oxidation and assimilation, NAD<sup>+</sup>/NADH ratios were determined in cells growing on C<sub>1</sub> compounds with or without thiosulfate. Increased levels were observed on thiosulfate/methanol but not on thiosulfate/formate and can be explained by excretion of sulfite as an intermediate of thiosulfate oxidation. Sulfite thus appears to inhibit methanol but not formate consumption.

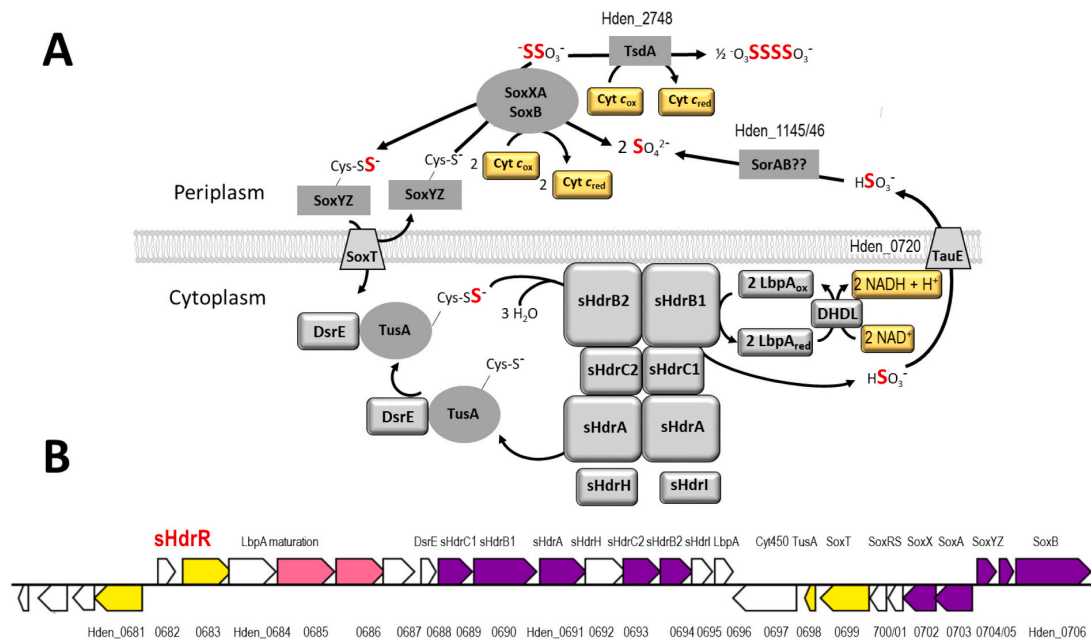
## 2. Materials and methods

### 2.1. Bacterial strains, plasmids, primers, and growth conditions

Table S1 lists the bacterial strains, and plasmids that were used for this study. *Escherichia coli* strains were grown on lysogeny broth (LB) media under aerobic conditions [29] at 37 °C unless otherwise indicated. *Escherichia coli* BL21(DE3) was used for recombinant protein production. *E. coli* 10β was used for molecular cloning. *H. denitrificans* strains were cultivated in minimal media kept at pH 7.2 with 100 mM 3-(N-Morpholino)propanesulfonic acid buffer as described before [19]. Media contained either 24.4 mM methanol or 25 mM formate and were supplemented with thiosulfate in the indicated concentrations when needed. Antibiotics for *E. coli* and *H. denitrificans* were used at the following concentrations (in μg ml<sup>-1</sup>): ampicillin, 100; kanamycin, 50; streptomycin, 200; chloramphenicol, 25.

### 2.2. Recombinant DNA techniques

Standard techniques for DNA manipulation and cloning were used unless otherwise indicated [30]. Restriction enzymes, T4 ligase and Q5



**Fig. 1.** (A) Current model of thiosulfate oxidation in *H. denitrificans*. The diheme cytochrome *c* thiosulfate dehydrogenase, TsdA (EC 1.8.2.2, thiosulfate:ferricytochrome *c* oxidoreductase), resides in the periplasm and catalyzes oxidative condensation of two thiosulfate molecules to tetrathionate, which is a dead-end product and not metabolized any further [19]. Complete oxidation of thiosulfate to sulfate also starts in the periplasm where enzymes SoxXA (EC 2.8.5.2, L-cysteine S-thiosulfotransferase) and SoxB (EC 3.1.6.20, S-sulfosulfanyl-L-cysteine sulfohydrolase) act together in oxidative attachment of thiosulfate to the sulfur carrier protein SoxYZ and subsequent hydrolytic release of sulfate from the thiosulfonate adduct. The sulfane sulfur stemming from thiosulfate is then transferred to the cytoplasm and further oxidized by the proteins of the sulfur-oxidizing heterodisulfide reductase-like enzyme system, sHdr, in conjunction with the lipoate-binding protein LbpA [19,24,25]. The resulting sulfite is transported back to the periplasm and oxidized to sulfate in a so far unresolved manner that may involve the products of genes Hden\_1145/45. Locus tags are given for those proteins that are not encoded by any of the genes shown in (B). DHDL, dihydroliipoamide dehydrogenase (EC 1.8.1.4, dihydroliipoyl dehydrogenase): four different genes encoding this enzyme are present (Hden\_0791, Hden\_0896, Hden\_1300, Hden\_3225) (B) The *sHdr* gene cluster and its vicinity in *H. denitrificans*. Encoded proteins or functions as well as locus tags are given. Color codes of genes are according to KEGG (<https://www.kegg.jp/kegg/kegg1c.html>).

polymerase were obtained from New England Biolabs (Ipswich, UK) and used according to the manufacturer's instructions. Oligonucleotides for cloning were obtained from Eurofins MWG (Ebersberg, Germany). Plasmid DNA from *E. coli* was purified using the GeneJET Plasmid Miniprep kit (Thermo Scientific, Waltham, USA). Chromosomal DNA from *H. denitrificans* strains was prepared using the First-DNA all-tissue Kit (GEN-IAL GmbH, Troisdorf, Germany).

### 2.3. Construction of *H. denitrificans* mutant strains

For markerless *in frame* deletion of the *H. denitrificans* *shdrR* (Hden\_0682), *soxYZ* (Hden\_0704/05) and *soxXA* (Hden\_0702/03) genes by splicing overlap extension (SOE) [31], PCR fragments were constructed using primers listed in Table S1. The  $\Delta$ *shdrR* fragment was inserted into *pk18mobsacB* [32] using BamHI and XbaI restriction sites. The SmaI-excised tetracycline cassette from pHP45 $\Omega$ -Tc [33] was inserted into the SmaI site of the resulting plasmid *pk18mobsacB* $\Delta$ *shdrR*. The  $\Delta$ *soxYZ* and  $\Delta$ *soxXA* fragments were inserted into the XbaI site of *pk18mobsacB*-Tc which had been constructed by insertion of the SmaI-excised tetracycline cassette from pHP45 $\Omega$ -Tc into the SmaI site of *pk18mobsacB*. The final constructs *pk18mobsacB* $\Delta$ *shdrR*-Tc, *pk18mobsacB* $\Delta$ *soxYZ*-Tc and *pk18mobsacB* $\Delta$ *soxXA*-Tc were electroporated into *H. denitrificans*  $\Delta$ *tsdA* and transformants were selected using previously published procedures [19,25]. Single crossover recombinants were Cm<sup>r</sup> and Tc<sup>r</sup>. Double crossover recombinants were Tc<sup>s</sup> and survived in the presence of sucrose due to loss of both, the vector-encoded levansucrase (SacB) and the tetracyclin resistance gene. The genotypes of the *H. denitrificans* mutant strains generated in this study were confirmed by PCR.

### 2.4. Characterization of phenotypes, quantification of sulfur compounds and protein content

Growth was followed by turbidity measurements at either 430 nm or 600 nm. A factor of 1.5947 ( $R^2 = 0.9994$ ) was determined for conversion of OD<sub>600</sub> into OD<sub>430</sub> (Fig. S1). In general, a small wavelength is preferable for OD measurements especially when cell density is low, because the sensitivity of the measurement is then higher [34]. For growth experiments, media with 24.4 mM methanol and varying concentrations of thiosulfate were inoculated to a start OD<sub>430</sub> of 0.008 with pre-cultures in late-exponential growth phase cultured on the same medium (50 ml culture in 100 ml-Erlenmeyer flasks). For phenotypic characterization, main cultures of 200 ml in 500 ml Erlenmeyer flasks were shaken at 200 rpm and 30 °C. Samples were taken in regular intervals and optical densities as well as thiosulfate, methanol, sulfite and sulfate were determined as necessary. Alternatively, growth experiments were run in 48-well microtiter plates. Plates were continuously shaken at 200 rpm and growth was followed by measuring optical density at 600 nm every 5 min using an Infinite 200Pro (Tecan, Crailsheim, Germany) plate reader. Each well contained 1 ml medium inoculated to an OD<sub>600</sub> of 0.025. For each set of experiments, five wells were run in parallel, one of which remained untouched throughout and served exclusively for following growth. From the other 1-ml parallel cultures, samples were taken for thiosulfate measurements at regular time intervals. Sampling of the same well was restricted to twice 100  $\mu$ l because sampling would otherwise have substantially reduced the culture volume and changed growth conditions, i.e. the ratio culture volume/surface area and thus aeration of the cultures.

Thiosulfate, sulfite and sulfate were determined by previously described colorimetric and turbidometric methods [35]. Experiments run in microtiter plates were restricted to quantification of thiosulfate which was performed with technical triplicates in a miniaturized format. Fifty  $\mu$ l H<sub>2</sub>O and 40  $\mu$ l 200 mM sodium acetate (pH 4.8) were added to 40  $\mu$ l culture supernatants, mixed, followed by addition of 10  $\mu$ l 200 mM NaCN and 10  $\mu$ l 40 mM CuCl<sub>2</sub>. After mixing thoroughly again, 10  $\mu$ l iron nitrate solution (30 g l<sup>-1</sup> Fe(NO<sub>3</sub>)<sub>3</sub>  $\times$  9 H<sub>2</sub>O and 40 ml 55% (v/v) HNO<sub>3</sub>

made up to 100 ml with distilled water) were added, mixed again and absorption at 460 nm was read in 96-well microtiter plates against a reagent blank with a Sunrise Tecan microplate reader. The same instrument was used for miniaturized sulfite determinations that were performed as follows: 175  $\mu$ l sample containing varying amounts of culture supernatant were mixed with 50  $\mu$ l 2% zinc acetate and 25  $\mu$ l 0.04% fuchsin (in 10% (v/v) H<sub>2</sub>SO<sub>4</sub>), incubated at room temperature for 10 min and measured against a reagent blank at 570 nm.

For the *H. denitrificans* type strain, the relationship between dry weight and turbidity measured at 430 nm has been reported to be linear to an optical density of 2.0, with OD<sub>430</sub> 2.0 = 0.60 mg dry wt ml<sup>-1</sup> [36–38]. Biomass values given in this work are all based on this conversion factor. We also calculated the protein content of cultures in Erlenmeyer flasks from OD<sub>430</sub>: A linear correlation between OD<sub>430</sub> and protein in cultures was found for cells growing on methanol up to an OD<sub>430</sub> of 1.4 and for cells growing on formate up to an OD<sub>430</sub> of 0.9 (Fig. S2). Conversion factors of 0.1476 ( $R^2 = 0.9883$ ) and 0.1322 ( $R^2 = 0.9802$ ) were determined for growth on methanol and formate, respectively, i.e. an OD<sub>430</sub> of 1.0 amounts to 0.146 mg protein ml<sup>-1</sup> on methanol and to 0.132 mg protein ml<sup>-1</sup> on formate. Protein thus constitutes 48.7% and 44.1%, respectively of the dry mass (ratios dry weight/protein 2.01 and 2.27), which is well within the range of values available for other species. For *Hyphomicrobium* strain EG, ratios of dry weight/protein between 2.3 and 2.8 were reported on different substrates [28]. The protein content was determined by the Biuret method as follows. One or two ml of cell suspension was centrifuged for 15 min at 16,100  $\times$ g and room temperature. The cell pellet was washed once with 1 ml 1% NaCl and resuspended in 0.5 ml dH<sub>2</sub>O. After addition of 50  $\mu$ l 5 M NaOH, the samples were incubated at 95 °C for 5 min and cooled down to room temperature. Then, 200  $\mu$ l of copper sulfate reagent (6.25 g Na–K tartrate, 1.25 g CuSO<sub>4</sub>  $\times$  5 H<sub>2</sub>O, 3.12 g KI and 5 g NaOH in a final volume of 500 ml H<sub>2</sub>O) were added, followed by incubation for 30 min at room temperature and centrifugation at 16,100  $\times$ g for 10 min. The absorbance at 546 nm was measured against a reagent blank.

Specific thiosulfate oxidation rates were calculated as follows: Thiosulfate concentrations determined in growth experiments were plotted graphically against time and fitted by a polynomial trend line between the second and fifth degree. The coefficient of determination helped to determine the correct degree of the polynomial. The first derivative of the polynomial function equation was then calculated and corresponded to the slope, i.e. thiosulfate oxidation rate in  $\mu$ M h<sup>-1</sup> at each time point, from which the specific oxidation rate [ $\mu$ mol thiosulfate h<sup>-1</sup> (mg protein)<sup>-1</sup>] was derived.

All growth experiments were repeated three to five times. Representative experiments with two biological replicates for each strain are shown. All quantifications are based on at least three technical replicates.

### 2.5. Quantification of methanol

Methanol was determined with an analytical high performance liquid chromatography system (Knauer, Germany) equipped with a refractive index detector and an Eurokat H (Knauer, Germany) column system consisting of a pre-column (Eurokrat H, 10  $\mu$ m, 30  $\times$  8 mm) and a main column (Eurokat H, 10  $\mu$ m, 300  $\times$  8 mm). The system was run with 5 mM sulfuric acid in ultrapure water at a flow rate of 0.6 ml/min and a temperature of 65 °C. Results were evaluated using the ClarityChrom program.

### 2.6. Overproduction and purification of recombinant sHdrR

The 411-bp *shdrR* gene was amplified from *H. denitrificans* genomic DNA and cloned between the NdeI and NotI sites of pET22b (+), resulting in pET22b-*shdrR*. Recombinant sHdrR was overproduced in *E. coli* BL21(DE3). The cells were grown at 37 °C in 200 ml LB medium

containing ampicillin up to an OD<sub>600</sub> of 0.6. Expression of *shdrR* was induced by adding 0.5 mM IPTG. IPTG-induced *E. coli* cells were grown over night at 20 °C. Cells were harvested at 14,000 ×g for 20 min and the pellet was washed with 50 mM Tris-HCl, pH 7.4. Three ml resuspending buffer (20 mM sodium phosphate buffer pH 7.4, 500 mM NaCl, 20 mM imidazole containing a spatula tip of deoxyribonuclease I and protease inhibitor) were added per g wet weight for homogenization. Cell lysis was achieved by sonification and followed by centrifugation (16,100 ×g, 30 min, and 4 °C) and ultracentrifugation (145,000 ×g, 1 h, 4 °C). The supernatant was applied to a Ni<sup>2+</sup>-NTA column equilibrated with lysis buffer. The column was washed with six volumes of lysis buffer and eluted with 50–500 mM imidazole in the same buffer. The protein was assessed for its purity by 12.5 % SDS-PAGE.

### 2.7. Electrophoretic mobility shift assay (EMSA)

Gel electrophoretic mobility shift assays are used to detect interactions between proteins and nucleic acids. In the assay, solutions of protein and nucleic acid are combined and the resulting mixtures are subjected to polyacrylamide under native conditions. After electrophoresis, the distribution of nucleic acid species is determined. In general, protein-nucleic acid complexes migrate more slowly than the corresponding free nucleic acid [39]. The binding reaction mixture (15 µl final volume), contained purified sHdrR protein (52 nM), 2 µl 50 % glycerol and 1.5 µl 10 × binding buffer (100 mM Tris-HCl, 500 mM KCl, 10 mM DTT, 5 % glycerol, pH 8.0). Reaction mixtures were pre-incubated for 20 min at room temperature followed by a further 30 min incubation at 30 °C after adding the DNA probe to a final concentration of 17 nM. The DNA probe consisted of a 362-bp fragment situated between the *shdrR* gene and the gene (Hden\_0681) upstream and was generated by PCR using primers EMSA-Fr and EMSA-Rev. The reaction mixtures were loaded onto 6 % native polyacrylamide gels after these had been pre-run at 100 V for 1 h at 4 °C with 0.25 × TBE buffer (25 mM Tris/borate, 0.5 mM EDTA). The loaded gels were electrophoresed in 0.25 × TBE with 0.5 % glycerol at 180 V for 2 h at 4 °C. Gels were subsequently stained for 20 min with SYBR green I. The bands corresponding to sHdrR-bound and free DNAs were visualized with a ChemiDoc Imaging System (BioRad).

### 2.8. Immunoblot analysis

*H. denitrificans* cell extracts were prepared as described before [19]. Western analysis was performed using the Trans-Blot Turbo Transfer system (Bio-Rad Laboratories, Munich, Germany) and nitrocellulose membranes (Amersham Protran 0.2 µm NC, GE Healthcare). sHdrA antigens were detected with antisera raised in rabbits (Eurogentec) against recombinant *H. denitrificans* sHdrA [19]. Antisera were used at 1:500 dilution. Binding of α-HdrA was detected with the SignalFire™ ECL reagent system (Cell Signaling Technology) and visualized with a ChemiDoc Imaging System (BioRad).

### 2.9. Quantification of the intracellular NAD(H<sup>+</sup>) ratio

Four milliliter culture broth were first cooled in an ice bath for 10 min to retard cell metabolism, collected by centrifugation (16,100 ×g at 4 °C for 5 min) and resuspended in 400 µl of 0.4 M HCl (for NAD<sup>+</sup>) or 0.4 M KOH (for NADH). The mixtures were then incubated at 30 °C (NADH) or 50 °C (NAD<sup>+</sup>) for 10 min and centrifuged at 16,100 ×g for 10 min at 4 °C. Then, 300 µl supernatant was neutralized by adding 300 µl of 0.4 M HCl or 0.4 M KOH, respectively. The neutralized samples were immediately used in a microcycling assay for NADH and NAD<sup>+</sup> determination [40] performed in 96 well plates and followed using a Sunrise Tecan microplate reader. Samples of 30 µl were combined with 120 µl of 125 mM Tris-HCl, pH 8.0 containing 2.5 mM phenazine methosulfate, 0.65 mM 3-(4,5-dimethyl-thiazoyl-2)-2,5-diphenyltetrazolium bromide (MTT), 778 mM ethanol and 33 U/ml yeast alcohol dehydrogenase

(A3263, Sigma-Aldrich, St. Louis, MO). The rate of MTT reduction was followed at 570 nm and 25 °C. Calibration was performed by measuring samples containing 30, 60, and 90 pmol NAD<sup>+</sup> or NADH. To ensure reliability of the assay, samples with 30, 60, and 90 pmol of the nicotinamide dinucleotides underwent the same procedure as described for the cells.

## 3. Results and discussion

### 3.1. Thiosulfate oxidation in *H. denitrificans* X<sup>T</sup>

As already pointed out, *H. denitrificans* X<sup>T</sup> has the genetic potential for two distinct pathways of thiosulfate oxidation, both of which occur or begin in the periplasm and use c-type cytochromes as electron acceptors [19,41,42] (Fig. 1A). We reported earlier that at 5 mM, all thiosulfate is stoichiometrically converted to tetrathionate. Thus, thiosulfate dehydrogenase (TsdA) causing oxidative condensation of two thiosulfate molecules to tetrathionate is the predominant catalyst under these conditions. A mutant strain lacking the *tsdA* gene is unable to form tetrathionate [19].

When exposed to thiosulfate concentrations of 2.5 mM or less, *H. denitrificans* X<sup>T</sup> wildtype switches gears and the majority of the sulfur compound is no longer transformed to tetrathionate but completely oxidized to sulfate [19]. It is not surprising that the same holds true for the strain lacking *tsdA*. In Fig. 2A growth and methanol consumption in the absence and presence of thiosulfate are compared for the *H. denitrificans* Δ*tsdA* strain. Just as expected for an organism incapable of autotrophic growth, the increase in biomass was inversely proportional to the decrease of the carbon source methanol. Thiosulfate was completely oxidized to sulfate (data not shown). In full agreement with earlier results, growth of the Δ*tsdA* strain was significantly retarded in the presence of thiosulfate (compare growth curves with open and filled circles in Fig. 2A), fully in line with a much slower consumption of methanol in the presence of thiosulfate (compare open and filled boxes in Fig. 2A).

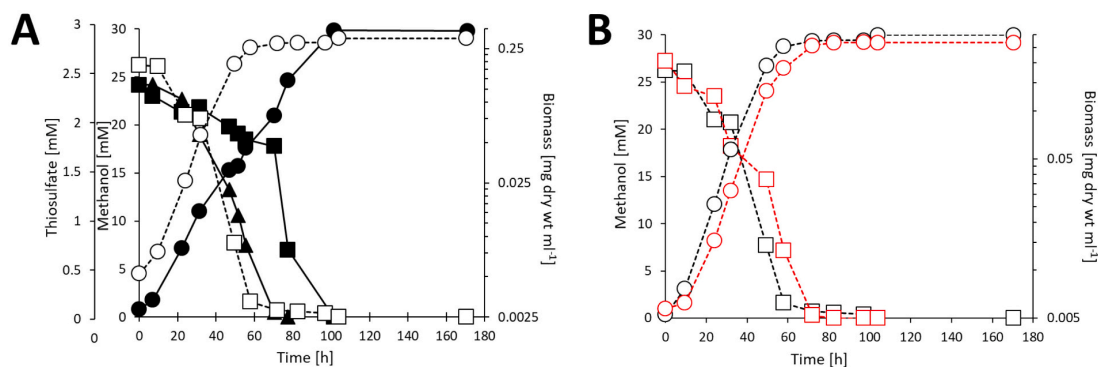
The genes encoding the enzymes acting together in the oxidation of thiosulfate to sulfate are all situated in the same genomic *shdr-lbpA-sox* island (Fig. 1B). Here, we show by individual markerless deletion of *soxXA* and *soxYZ* that the products of these genes are absolutely essential for thiosulfate oxidation in the Δ*tsdA* strain (Fig. S3). We can thus state with confidence that a typical incomplete Sox pathway without involvement of a sulfane sulfur dehydrogenase, SoxCD, [43,44] is operated in *H. denitrificans* X<sup>T</sup> (Fig. 1A). SoxXA catalyzes the oxidative fusion of thiosulfate to a conserved cysteine of the sulfur carrier protein SoxYZ encoded by Hden\_0704/05. The other four SoxYZ homologs encoded in *H. denitrificans* (Hden\_1399/1400 and three *soxYZ* fusions, Hden\_0338, Hden\_1000 and Hden\_1147) cannot functionally replace the protein encoded in immediate vicinity of *shdr-lbpA*. Just as in other organisms lacking bona fide SoxCD, sulfane sulfur still bound to SoxYZ cannot be oxidized in the periplasm but has to be transferred into the cytoplasm in a so far unresolved manner. Here, oxidation of sulfur to sulfite is strictly dependent on the proteins of the sulfur-oxidizing heterodisulfide reductase-like (sHdr) system (Fig. 1A) [19]. The four electrons released in this step are in all probability transferred to NAD<sup>+</sup> [24,45] (Fig. 1A). The lipoate-binding protein LpbA is essential for this step [25]. In *H. denitrificans* strains lacking functional sHdr or LpbA proteins the strong reduction of growth rate caused by thiosulfate in the wildtype was released [19,25]. It reappeared when the Δ*shdr* strain latter was re-equipped with the *shdr* genes *in trans* [19].

### 3.2. Regulation of *shdr* genes in *Hyphomicrobium denitrificans*

To even more clearly attribute the observed effects of thiosulfate on methanol assimilation and thus growth rate to the sHdr system, we first collected information on its regulation. A previous comparative proteomics approach had already shown that the proteins encoded in the

J. Li et al.

BBA - Bioenergetics 1864 (2023) 148932



**Fig. 2.** Part A shows *H. denitrificans*  $\Delta tsdA$  growing on 24.4 mM methanol (boxes). Precultures contained 2 mM thiosulfate. Cultures without (open symbols, dashed lines) and with thiosulfate (filled symbols, bold lines) are compared. Biomass is given as mg dry weight per ml ( $\circ$  no thiosulfate,  $\bullet$  with thiosulfate). Methanol concentrations are indicated in the absence ( $\square$ ) and in the presence ( $\blacksquare$ ) of thiosulfate. Thiosulfate is given as black triangles ( $\blacktriangle$ ). In part B growth and methanol consumption in thiosulfate-free minimal medium are compared for *H. denitrificans*  $\Delta tsdA$  (black symbols and lines) and *H. denitrificans*  $\Delta tsdA \Delta shdrR$  (red symbols and lines). Methanol concentrations ( $\square$ ,  $\square$ ) and biomass content ( $\circ$ ,  $\circ$ ) are displayed. Error bars indicating SD are too small to be visible. Sulfate was also quantified but is not shown for clarity. Sulfate concentrations increased by 5 mM in cultures initially containing 2.5 mM thiosulfate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

hypomicrobial *shdr* cluster are much more abundant in the presence of dimethylsulfide (DMS) than in the absence of an oxidizable sulfur compound [19]. Thiosulfate is an intermediate of DMS degradation. We postulated that the transcription of the *shdr* genes is regulated by the ArsR-type regulator encoded by the first gene in the operon, *shdrR* (Hden\_0682) (Fig. 1B) [19]. ArsR-family regulators function as transcriptional repressors and include a wide range of metal-, metalloid- and non-metal-sensing proteins [46]. Here, we proved the function of sHdrR as a repressor by construction of *H. denitrificans*  $\Delta tsdA \Delta shdrR$ , a mutant strain with a markerless deletion of *shdrR* in a  $\Delta tsdA$  background. In the absence of thiosulfate, this strain uses methanol just as the  $\Delta tsdA$  reference strain (Fig. 2B) but produces the sHdr system constitutively as shown by immunoblot analysis with an antibody directed against sHdrA (Fig. 3A). *H. denitrificans* sHdrR was produced as a His-tagged recombinant protein in *E. coli*. Electrophoretic mobility shift assays verified that it specifically binds to the 362 bp DNA fragment between its own gene and the divergently transcribed gene Hden\_0681 (Fig. 3B).

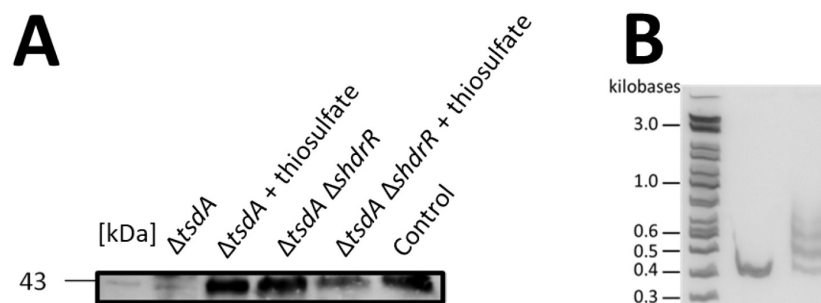
### 3.3. Growth and thiosulfate oxidation on a reduced $C_1$ carbon source: Methanol

In the next step, the effect of thiosulfate during growth on the reduced  $C_1$ -carbon compound methanol was assessed in detail for the *H. denitrificans*  $\Delta tsdA$  reference strain. The growth rate-decreasing effect of thiosulfate became stronger the more thiosulfate was added to the cultures (Figs. 4 and S4). The effect of thiosulfate was particularly impressive when cultures were inoculated with thiosulfate-induced cells (Figs. 4 and S4). Specific thiosulfate oxidation rates rose with the

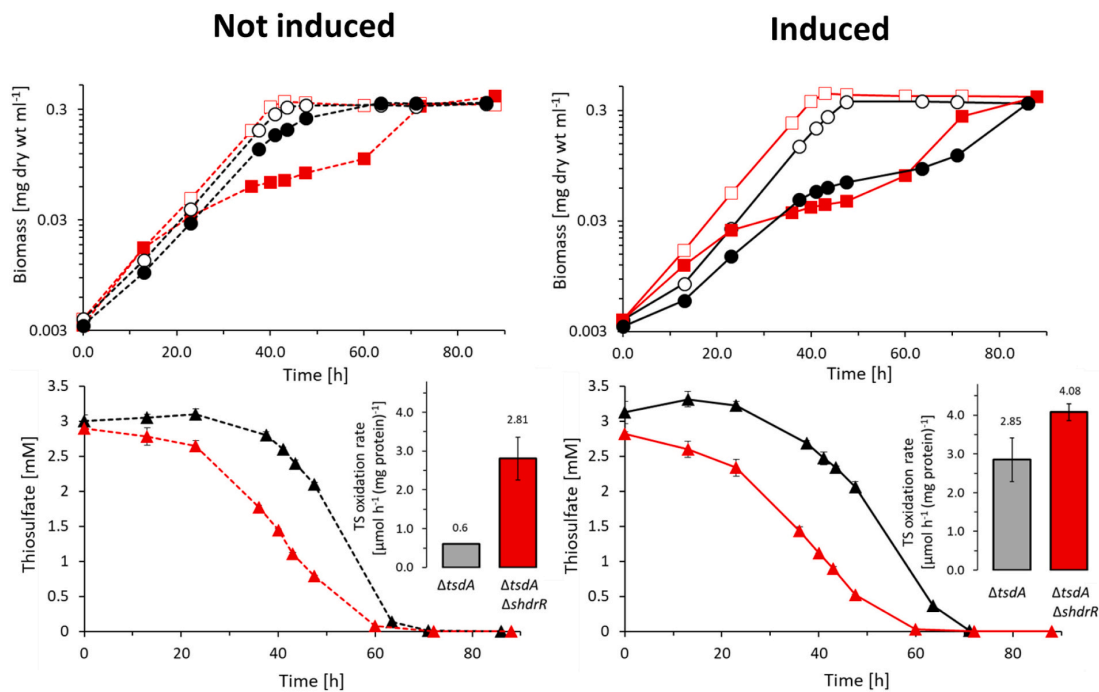
thiosulfate concentration and were about four-fold higher when pre-cultures had already been exposed to thiosulfate (Figs. 4 and S4).

The picture substantially changed for the  $\Delta tsdA \Delta shdrR$  strain. This strain, that constitutively produces the sHdr complex, exhibited a very strongly reduced growth rate and a high specific thiosulfate oxidation rate even without induction of pre-cultures. As soon as thiosulfate was consumed, the growth rate increased substantially (Fig. 4). The *H. denitrificans*  $\Delta tsdA \Delta shdrR$  strain exhibited somewhat higher specific thiosulfate consumption rates when pre-cultures had been exposed to thiosulfate than for the non-induced case (Figs. 4 and S4), possibly indicating an additional regulator involved in the overall process.

To further demonstrate the negative effect of thiosulfate on growth rate, thiosulfate was added in early exponential growth phase to cultures of *H. denitrificans*  $\Delta tsdA \Delta shdrR$  growing on 25 or 50 mM methanol. As soon as thiosulfate consumption set in, a drastic decrease in growth rate was observed (Fig. S5). To exclude with certainty that the growth rates observed for the *H. denitrificans*  $\Delta tsdA$  and  $\Delta tsdA \Delta shdrR$  strains in the presence and absence of thiosulfate were negatively influenced by insufficient oxygen supply, growth was followed in 500 ml Erlenmeyer flasks filled with different culture volumes (100 and 200 ml). Consistent with earlier reports of reduced growth rates at oxygen tensions above 65 % air saturation [28], the better oxygenated 100 ml cultures grew significantly slower, irrespective of whether thiosulfate was present (Fig. S6). Slowest growth was observed for the  $\Delta tsdA \Delta shdrR$  strain in 100 ml medium containing thiosulfate.



**Fig. 3.** (A) Western blot analysis with antiserum against sHdrA [19] performed with crude extracts of *H. denitrificans* strains  $\Delta tsdA$  and  $\Delta tsdA \Delta shdrR$  grown either on 24.4 mM methanol alone or on 24.4 mM methanol plus 2 mM thiosulfate. Ten  $\mu$ g protein were loaded per lane. Recombinant sHdrA produced in *E. coli* [19,24] was used as the control. (B) Binding of sHdrR to the promoter region of the *shdr* gene cluster assessed by EMSA. Lane 1: DNA size marker; lane 2: 17 nM of DNA fragment; lane 3: Combination of 17 nM of DNA fragment and 52 nM recombinant sHdrR. Reaction mixtures were incubated at 30 °C for 30 min.



**Fig. 4.** Growth and thiosulfate consumption of *H. denitrificans*  $\Delta tsdA$  (black circles and lines) and  $\Delta tsdA \Delta shdrR$  (red boxes and lines). Cultures were grown on methanol-containing medium (24.4 mM methanol) without (open symbols) or with 3 mM thiosulfate (filled symbols). Precultures contained either no thiosulfate (not induced, broken lines) or 2 mM thiosulfate (induced, solid lines). In the lower panels, triangles indicate thiosulfate concentrations for *H. denitrificans*  $\Delta tsdA$  (black) and for *H. denitrificans*  $\Delta tsdA \Delta shdrR$  (red). Specific thiosulfate oxidation (TS) rates are depicted in the same color code. Error bars indicating SD are too small to be visible for determination of biomass. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

### 3.4. Growth and thiosulfate oxidation on an oxidized C<sub>1</sub> carbon source: formate

We set out to collect more information on the interplay between sulfur oxidation and carbon metabolism in *H. denitrificans* by replacing methanol with a more oxidized C<sub>1</sub>-carbon source. While externally added formaldehyde does not appear to be a suitable substrate for this species [21], we confirmed previous reports that the type strain studied here can grow on formate as the sole source of carbon and electrons [47].

The simultaneous presence of thiosulfate in formate cultures exerted a completely different effect than observed on methanol. Growth rates appeared essentially independent of the presence of thiosulfate in precultures and in main cultures and also independent of the constitutive presence of the sHdr system in the *H. denitrificans*  $\Delta tsdA \Delta shdrR$  strain (Fig. 5). Whether the presence of thiosulfate as accessory electron donor during growth on formate causes significant increases in growth yield of the *Hyphomicrobium* strain studied here, has to await further studies, at best in continuous culture.

### 3.5. Respiratory electron transport and C<sub>1</sub>-metabolism in *Hyphomicrobium denitrificans* X<sup>T</sup>

From the reported growth experiments, it appeared that in the *H. denitrificans* type strain thiosulfate as an auxiliary electron donor prevents effective assimilation of a reduced C<sub>1</sub> carbon source, i.e. methanol, but does not negatively influence the rate of biomass production from an oxidized C<sub>1</sub> carbon compound. In order to understand this finding, we performed a detailed analysis of respiratory electron transport and C<sub>1</sub> metabolism and in *H. denitrificans* X<sup>T</sup> on the basis of previously published results in combination with an analysis of the strain's genome sequence using the KEGG pathways database ([www.genome.jp/kegg/pathway.html](http://www.genome.jp/kegg/pathway.html)) [48].

BLASTP [49] was used to confirm the absence or presence of genes. The results are depicted in Figs. 6 and S7. Additional information is provided in Table S2.

Briefly, electrons from reduced carriers like NADH/FADH<sub>2</sub>, ubiquinol or reduced cytochrome c can enter a respiratory chain composed of complex I, II and III and an array of terminal reductases that can mediate both oxygen respiration and anaerobic respiration on nitrate. In methyloprots like *H. denitrificans*, substrates like methanol serve two purposes: first they are oxidized to CO<sub>2</sub> producing reducing equivalents for energy conservation via respiration and second they are assimilated for biomass production (Fig. 6A, B). The ability of *H. denitrificans* to oxidize and assimilate methanol is based on a combination of pathways and enzymes similar but not identical to those of the model methyloprot, *Methylobacterium extorquens* (formerly *Methylobacterium extorquens* [50]) (reviewed in [23,51]).

Methanol is first oxidized to formaldehyde by periplasmic pyrrolo-quinoline quinone (PQQ)-containing methanol dehydrogenase [52]. Formaldehyde is further processed in the cytoplasm by an elaborate pathway employing tetrahydromethanopterin (THMPT) as a cofactor [23]. Formate release from formyl-THMPT is catalyzed by the formyltransferase/hydrolase complex. The step comprises two consecutive reactions where methyloluran acts as an intermediate carrier of the one-carbon unit [53,54]. Formate is finally oxidized to CO<sub>2</sub> by formate dehydrogenase [55]. In many methyloprots, an oxidative C<sub>1</sub> transfer pathway analogous to the THMPT pathway is proposed that is linked to tetrahydrofolate (THF) instead of THMPT. The first step in this pathway would be the reaction of formaldehyde with THF. An enzyme catalyzing this reaction has not been found in any organism so far and it has been proposed that a spontaneous chemical reaction may be a sufficient source of methylene-THF. However, recent result severely questions this concept [23,56].

Assimilation of C<sub>1</sub> carbon in *H. denitrificans* X<sup>T</sup> occurs via the serine



cycle [57] as briefly depicted in Fig. 6. Formation of serine from methylene-THF and glycine is catalyzed by glycine hydroxymethyl-transferase [58]. Serine is then converted to glycerate-2-phosphate and phosphoenolpyruvate (PEP). CO<sub>2</sub> is co-assimilated by carboxylation of PEP to oxaloacetate. In addition to the serine cycle, *H. denitrificans* X<sup>T</sup> harbors the genetic equipment for the ethylmalonyl-CoA pathway and a functional glyoxylate shunt [23]. For *H. denitrificans* it is most likely that THF-based C<sub>1</sub> metabolism is run exclusively in the reductive direction and serves purely assimilatory purposes (Fig. 6B). Non-reversible formyl-THF hydrolase (PurU), that is usually present in organisms running an oxidative THF-linked pathway, is not encoded in *H. denitrificans* [23]. In summary, biomass formation from methanol in *H. denitrificans* appears to involve a long THMPT-dependent oxidative route yielding formate that is then hooked up to THF in an ATP-dependent manner and re-reduced up to the level of formaldehyde before delivery into the serine pathway (Fig. 6B).

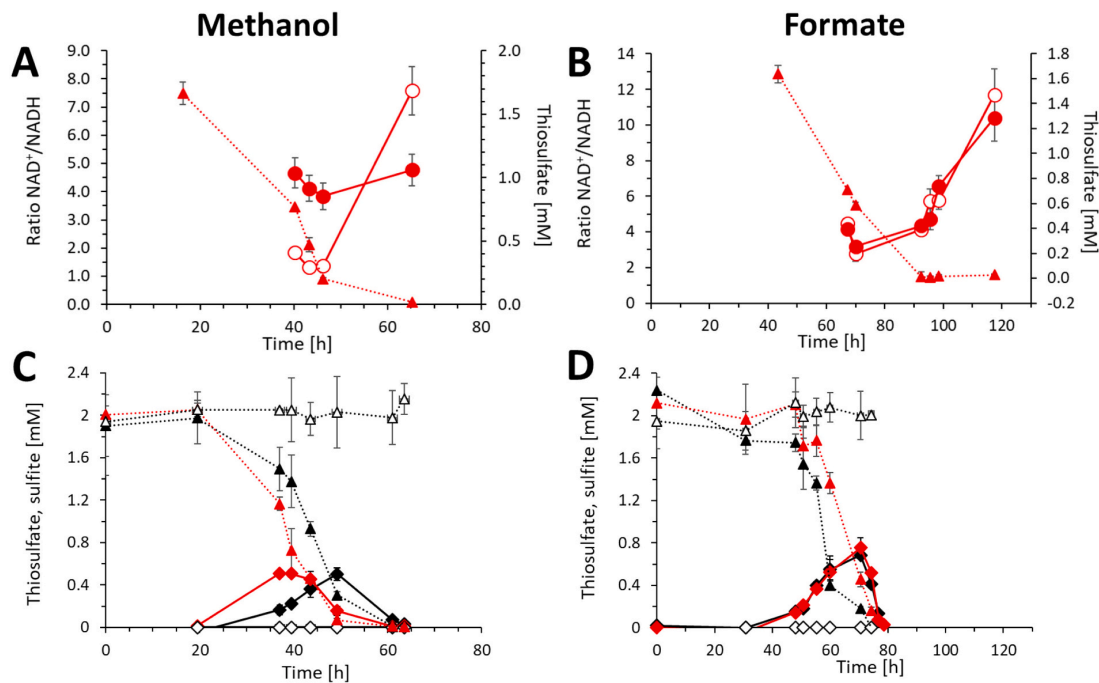
### 3.6. Bioenergetic status of thiosulfate-grown *H. denitrificans* and formation of a sulfur intermediate

We rationalized that the observed negative effect of thiosulfate on biomass production from methanol but not from formate can in principle have two different reasons: First, thiosulfate oxidation may cause an over-reduction of the cellular nicotinamide dinucleotide and cytochrome *c* pools. Such over-reduction would then prevent effective assimilation of methanol into biomass, as methanol must first be oxidized all the way to formate before it can be hooked up to tetrahydrofolate, re-reduced up to the level of formaldehyde and finally delivered into the serine pathway for assimilation. Second, thiosulfate oxidation may lead to production of an intermediate inhibiting methanol but not formate metabolism.

To solve these questions, we first assessed the bioenergetic status of

*H. denitrificans* during growth on methanol or formate in the absence versus the presence of thiosulfate by measuring the NAD<sup>+</sup>/NADH ratio of the  $\Delta tsdA \Delta shdrR$  strain. As apparent by the comparison of Fig. 7A and B the NAD<sup>+</sup>/NADH ratio drastically rose in methanol-grown cultures in exponential phase from  $1.8 \pm 1.3$  to  $4.6 \pm 0.5$  when cells were actively oxidizing thiosulfate, while it remained unchanged at a ratio of about 4 on formate. As expected, the ratio further increased in all cases when cells reached stationary phase (Figs. 7A, B and S8) and oxidizable substrates were no longer available. These results clearly discounted the idea that thiosulfate may cause an over-reduction of the nicotinamide dinucleotide pool.

In the next step, we assessed the possibility of the formation of potentially inhibitory intermediates during the oxidation of thiosulfate. Indeed, *H. denitrificans*  $\Delta tsdA \Delta shdrR$  as well as the  $\Delta tsdA$  reference strain excreted up to 0.5 mM sulfite into the medium when exposed to 2 mM thiosulfate, irrespective of the C<sub>1</sub> compound present (Figs. 7C and D, S8). The highly reactive and potentially toxic sulfite is widely used as a food preservative [59]. Sulfite formed in the cytoplasm of *H. denitrificans* via the sHdr pathway is most probably shuttled across the cytoplasmic membrane by a TauE like-sulfite exporter (Fig. 1). An enzyme catalyzing efficient sulfite oxidation is obviously not present in the periplasm of *H. denitrificans*. The sulfite oxidation rates observed in cultures of about 0.1 mM h<sup>-1</sup> are in the same range as those we determined for 0.5 mM and 1 mM sulfite dissolved in cell-free medium and incubated at the same temperature and shaking frequency ( $\sim 0.12$  mM h<sup>-1</sup> and  $\sim 0.18$  mM h<sup>-1</sup>, respectively). While growth on formate is driven completely by cytoplasmic enzymes including formate dehydrogenase, growth on methanol is initiated in the periplasm by methanol dehydrogenase. We rationalize that formate utilization is shielded from effects exerted by sulfite through its efficient export, while periplasmic PQQ-containing methanol dehydrogenase is negatively affected. It should be noted that the formation of PQQ-sulfite adducts is well established [60,61] and



**Fig. 7.** NAD<sup>+</sup>/NADH ratio in *H. denitrificans*  $\Delta tsdA \Delta shdrR$  on 50 mM methanol (A) and 50 mM formate (B) in the absence (open circles) and presence of 2 mM thiosulfate (filled circles). Thiosulfate is indicated by red filled triangles and was added at a biomass of 0.033 mg dry wt ml<sup>-1</sup>. Precultures did not contain thiosulfate. In the lower panels, triangles indicate thiosulfate concentrations for *H. denitrificans*  $\Delta tsdA$  (black), for *H. denitrificans*  $\Delta tsdA \Delta shdrR$  (red) and the thiosulfate-oxidation negative strain (open). Sulfite concentrations are given by diamonds following the same color code. Cultures shown in panels C and D were grown on methanol and formate, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



may be the basis for inhibition of methanol dehydrogenase.

### 3.7. Occurrence of *tsdA*, *sox* and *shdr* genes in *Hyphomicrobium* species

The negative influence of thiosulfate on the assimilation of reduced C<sub>1</sub>-carbon compounds in a *Hyphomicrobium* strain that has two different thiosulfate oxidation pathways, one of which involves the sHdr complex and results in excretion of inhibitory concentrations of sulfite, led us to the question whether this is a general feature within the genus or an adaptation of a specific strain. We searched all genomes available for strains of the genus *Hyphomicrobium* for the presence of *tsdA* encoding periplasmic tetrathionate-forming thiosulfate dehydrogenase, the genes for the periplasmic Sox system and the genes for the cytoplasmic sHdr proteins (Tables 1 and S3). Even when only strains with complete genomes or assemblies on the scaffold level are considered, it is apparent that the equipment with sulfur oxidation pathways is strain specific (Table 1). While some strains do not contain any gene encoding proteins for thiosulfate oxidation, others contain solely *tsdA*, a combination of *sox* and *shdr* genes or all three of the latter. This observation is fully in line with reports for other methylotrophs of the genus *Methylobacterium*/*Methylorubrum* (Methylobacteriaceae, order Hyphomicrobiales) [26].

## 4. Conclusions

Here, we provide an experiment-based explanation for growth retardation of *H. denitrificans* X<sup>T</sup> by thiosulfate during growth on methanol and shed first light on the bioenergetics underlying simultaneous use of an inorganic sulfur compound and C<sub>1</sub> compounds as electron donors. A first explanation model hypothesized that the oxidation of the auxiliary inorganic electron donor leads to an over-reduction of

electron acceptors, i.e. cytochromes *c* and NAD(P)<sup>+</sup> that are thus not available in sufficient concentrations to allow efficient oxidation of methanol. This reduced substrate has to be oxidized to formate before it can serve as a substrate for assimilation into biomass. This idea was experimentally discounted and instead sulfite was detected as an exported intermediate of thiosulfate oxidation. While formate metabolism in the cytoplasm remained unaffected, the strong negative effect on biomass formation from methanol is probably attributed to inhibition of the periplasmic methanol dehydrogenase, which is likely due to formation of a sulfite-adduct of its PQQ cofactor.

Whether thiosulfate can be oxidized at all and, if so, by which pathways appears to be strain-specific among the genus *Hyphomicrobium* as well as among other methylotrophs. At elevated thiosulfate concentrations, TsdA catalyzing tetrathionate formation appears as the enzyme of choice, substantiated by the observation that *H. denitrificans* X<sup>T</sup> exclusively forms tetrathionate at thiosulfate concentrations above 2.5 mM [19]. Possessing the sHdr pathway may be advantageous at low thiosulfate concentrations in a range not leading to accumulation of inhibitory amounts of excreted sulfite. This can probably occur in oxygenated environments, where the toxic compound is effectively chemically oxidized, or in habitats, where sulfite is removed rapidly by other members of the community.

Different equipment with thiosulfate oxidation pathways thus probably allows fine-tuned adaptation to environmental conditions. It is well established that in soils, a major habitat for *Hyphomicrobia* [21], methanol is present and its concentrations can be spatially and heterogeneously distributed. In proximity to plant material, hot spots of methanol might exist that are not detectable in mixtures with bulk soil due to a dilution effect [69]. While thiosulfate is typical for certain marine environments, particularly in aerobic/anaerobic interfaces, it

**Table 1**

Occurrence of genes for enzymes involved in thiosulfate and cytoplasmic sulfur oxidation in the genus *Hyphomicrobium*. Only those strains are shown whose genome assembly is either complete or on the scaffold level. Information on genomes available on the level of contigs is provided in Table S2. The gene *soxX* is put in brackets, because it is not always present. In classical heterodimeric SoxAX proteins, SoxX serves as the site of electron storage and transfer, while SoxA harbors the catalytically active site. It is therefore well conceivable that SoxA alone is active and transfers electrons directly to a separate cytochrome *c* acceptor encoded elsewhere in the genome.

<i>Hyphomicrobium</i> species/strain	<i>tsdA</i>	<i>soxA(X)BYZ</i>	<i>shdrC1B1AHC2B2</i>	Accession	Assembly level	Reference
<i>H. denitrificans</i> X <sup>T</sup> SM1869	Hden_2748	Hden_0703-0706	Hden_0689-0694	GCA_000143145.1	Complete	[47,62]
<i>H. denitrificans</i> 1NES1	–	–	–	GCA_000230975.3	Complete	[63]
<i>H. denitrificans</i> SCN18_30_10_14_R2_B_61_9	J0H36_02460	J0H36_02935 06390, 07480-5	J0H36_02865-02890	GCA_017304115.1	Scaffold	[64]
<i>H. denitrificans</i> SCN18_30_10_14_R3_B_61_7	J0H04_09105	–	–	GCA_017307215.1	Scaffold	[64]
<i>H. denitrificans</i> SCN18_30_10_14_R1_P_61_7	–	J0H37_09715-30	J0H37_09135-40 <sup>a</sup> J0H37_09200-05	GCA_017305615.1	Scaffold	[64]
<i>H. nitrativorans</i> NL23 <sup>T</sup> (ATCC BAA-2476)	–	–	–	GCA_000503895.1	Complete	[65,66]
<i>H. sulfonivorans</i> WDL6	–	–	–	GCA_001541235.1	Scaffold	Albers, P., unpublished
<i>H. zavarzini</i> ATCC 27496 <sup>T</sup>	–	F812_RS23075, 0107895, 930-940	–	GCA_000383415.1	Scaffold	[21]
<i>H. sp.</i> AWTP1-2	EKK30_01635	–	–	GCA_003987855.1	Scaffold	[67]
<i>H. sp.</i> AWTP1-10	EKK38_19120	–	–	GCA_003987725.1	Scaffold	[67]
<i>H. sp.</i> MC1	HYPMC_2182	–	–	GCA_000253295.1	Complete	Ge-scope
<i>H. sp.</i> ghe19	HYPP_04503	–	–	LR743509	Complete	Creemers, G. unpublished
<i>H. sp.</i> 99	G359_RS04385	–	–	GCA_000384335.2	Scaffold	Chistoserdova, L. et al, unpublished
<i>H. sp.</i> 12-62-95	–	–	–	GCA_002279935.1	Scaffold	Kantor, R.S. et al., unpublished
<i>H. sp.</i> 32-62-53	–	–	–	GCA_002280885.1	Scaffold	Kantor, R.S. et al., unpublished
<i>H. sp.</i> SCN 65-11	–	–	–	GCA_001724295.1	Scaffold	[68]
<i>H. sp.</i> SCN18_10_11_15_R1_B_65_8	–	J0J14_11055-70 <sup>a</sup> J0J14_14585 <sup>a</sup>	J0J14_15020-040 <sup>a</sup>	GCA_017306765.1	Scaffold	[64]
<i>H. sp.</i> SCN18_10_11_15_R2_B_65_9	–	J0I57_14080-90 <sup>a</sup>	J0I57_20835-50 <sup>a</sup>	GCA_017306735.1	Scaffold	[64]
<i>H. sp.</i> SCN18_30_10_14_R3_B_64_9	–	J0I75_12920-30 <sup>a</sup>	J0I75_04305-030	GCA_017306765.1	Scaffold	[64]
<i>H. sp.</i> SCN18_26_2_15_R2_B_61_8	J0I81_01825	–	–	GCA_017306805.1	Scaffold	[64]

<sup>a</sup> Partly present.

has also been reported to be present in most soils except in very humid regions [70]. Low-molecular-weight organic acids such as formate, are also ubiquitous in soils, as they are important root exudates [71] and are intermediates and by-products of anaerobic carbon metabolism [72]. For a facultative denitrifier like *H. denitrificans* X<sup>T</sup>, formate as a carbon source may be particularly important in the absence of oxygen and simultaneous oxidation of thiosulfate under anaerobic conditions may provide significant growth advantages. Clearly, future research has to address this issue more comprehensively and in more detail.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (Grant Da 351/13-1). Jingjing Li was financed by a Scholarship of the China Scholarship Council and Tomohisa Sebastian Tanabe received a scholarship from the Studienstiftung des Deutschen Volkes. We thank Franziska Wienberg and Andre Neff for help with methanol quantification.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbabi.2022.148932>.

#### References

- [1] A. Eiler, Evidence for the ubiquity of mixotrophic bacteria in the upper ocean: implications and consequences, *Appl. Environ. Microbiol.* 72 (2006) 7431–7437, <https://doi.org/10.1128/AEM.01559-06>.
- [2] Z. Kolber, Energy cycle in the ocean: powering the microbial world, *Oceanography* 20 (2007) 79–88, <https://doi.org/10.5670/oceanog.2007.51>.
- [3] P.A. Trudinger, Metabolism of thiosulfate and tetrathionate by heterotrophic bacteria from soil, *J. Bacteriol.* 93 (1967) 550–559, <https://doi.org/10.1128/jb.93.2.550-559.1967>.
- [4] R. Boden, D.P. Kelly, J.C. Murrell, H. Schäfer, Oxidation of dimethylsulfide to tetrathionate by *Methylophaga thiooxidans* sp. nov.: a new link in the sulfur cycle, *Environ. Microbiol.* 12 (2010) 2688–2699, <https://doi.org/10.1111/j.1462-2920.2010.02238.x>.
- [5] R. Boden, L.P. Hutt, Bacterial metabolism of C1 sulfur compounds, in: F. Rojo (Ed.), *Aerobic Utilization of Hydrocarbons, Oils And Lipids. Handbook of Hydrocarbon And Lipid Microbiology*, Springer Nature Switzerland AG, Cham, 2019, pp. 1–43.
- [6] R. Boden, L.P. Hutt, Chemolithoheterotrophy: means to higher growth yields from this widespread metabolic trait, in: F. Rojo (Ed.), *Aerobic Utilization of Hydrocarbons, Oils And Lipids*, Springer Nature Switzerland AG, Cham, 2019, pp. 493–517.
- [7] L.P. Hutt, G.M. Harper, A.J. Moody, R. Boden, Insights into growth kinetics and roles of enzymes of Krebs' cycle and sulfur oxidation during exochemolithoheterotrophic growth of *Achromobacter aegrifaciens* NCCB 38021 on succinate with thiosulfate as the auxiliary electron donor, *Arch. Microbiol.* (2020), <https://doi.org/10.1007/s00203-020-02028-1>.
- [8] D.Y. Sorokin, T.P. Tourova, G. Muyzer, *Citricella thiooxidans* gen. nov., sp. nov., a novel lithoheterotrophic sulfur-oxidizing bacterium from the Black Sea, *Syst. Appl. Microbiol.* 28 (2005) 679–687, <https://doi.org/10.1016/j.syapm.2005.05.006>.
- [9] J.H. Tuttle, P.E. Holmes, H.W. Jannasch, Growth rate stimulation of marine pseudomonads by thiosulfate, *Arch. Microbiol.* 99 (1974) 1–14, <https://doi.org/10.1007/BF00696218>.
- [10] J.H. Tuttle, H.W. Jannasch, Occurrence and types of Thiobacillus-like bacteria in sea, *Limnol. Oceanogr.* 17 (1972) 532–543, <https://doi.org/10.4319/lo.1972.17.4.0532>.
- [11] J.H. Tuttle, H.W. Jannasch, Sulfide and thiosulfate oxidizing bacteria in anoxic marine basins, *Mar. Biol.* 20 (1973) 64–70, <https://doi.org/10.1007/BF00387676>.
- [12] D.Y. Sorokin, L.A. Robertson, J.G. Kuenen, Sulfur cycling in *Catenococcus thioocyclus*, *FEMS Microbiol. Ecol.* 19 (1996) 117–125, <https://doi.org/10.1111/j.1574-6941.1996.tb00204.x>.
- [13] L. Wang, Z. Shao, Aerobic denitrification and heterotrophic sulfur oxidation in the genus *Halomonas* revealed by six novel species characterizations and genome-based analysis, *Front. Microbiol.* 12 (2021), 652766, <https://doi.org/10.3389/fmicb.2021.652766>.
- [14] S. Spring, U. Jäckel, M. Wagner, P. Kämpfer, *Ottowia thiooxydans* gen. nov., sp. nov., a novel facultatively anaerobic, N<sub>2</sub>O-producing bacterium isolated from activated sludge, and transfer of *Aquaspirillum gracile* to *Hylemonella gracilis* gen. nov., comb. nov, *Int. J. Syst. Evol. Microbiol.* 54 (2004) 99–106, <https://doi.org/10.1099/ijs.0.02727-0>.
- [15] S.K. Das, A.K. Mishra, B.J. Tindall, F.A. Rainey, E. Stackebrandt, Oxidation of thiosulfate by a new bacterium, *Bosea thiooxidans* (strain BI-42) gen. nov., sp. nov.: analysis of phylogeny based on chemotaxonomy and 16S ribosomal DNA sequencing, *Int. J. Syst. Microbiol.* 46 (1996) 981–987, <https://doi.org/10.1099/00207713-46-4-981>.
- [16] J.M. Gonzalez, J.S. Covert, W.B. Whitman, J.R. Henriksen, F. Mayer, B. Scharf, R. Schmitt, A. Buchan, J.A. Fuhrman, R.P. Kiene, M.A. Moran, *Silicibacter pomeroyi* sp. nov. and *Roseovarius nubinhibens* sp. nov., dimethylsulfoniopropionate-demethylating bacteria from marine environments, *Int. J. Syst. Evol. Microbiol.* 53 (2003) 1261–1269, <https://doi.org/10.1099/ijs.0.02491-0>.
- [17] D.Y. Sorokin, T.P. Tourova, B.B. Kuznetsov, I.A. Bryantseva, V.M. Gorlenko, *Roseinatronobacter thiooxidans* gen. nov., sp. nov., a new alkaliphilic aerobic bacteriochlorophyll a - containing bacterium isolated from a soda lake, *Microbiology* 69 (2000) 75–82, <https://doi.org/10.1007/BF02757261>.
- [18] Y. Xin, R. Gao, F. Cui, C. Lu, H. Liu, H. Liu, Y. Xia, L. Xun, The heterotrophic bacterium *Cupriavidus pinatubonensis* JMP134 oxidizes sulfide to sulfate with thiosulfate as a key intermediate, *Appl. Environ. Microbiol.* 86 (2020), <https://doi.org/10.1128/AEM.01835-20>.
- [19] T. Koch, C. Dahl, A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds, *ISME J.* 12 (2018) 2479–2491, <https://doi.org/10.1038/s41396-018-0209-7>.
- [20] A. Oren, X.-W. Xu, The family *Hyphomicrobiaceae*, in: E. Rosenberg, E.F. DeLong, E. Stackebrandt, F. Thompson (Eds.), *The Prokaryotes*, Springer, Berlin Heidelberg, 2014, pp. 247–281.
- [21] C. Glesche, A. Fesefeldt, P. Hirsch, *Hyphomicrobium* Stutzer and Hartleb 1898, 76<sup>AL</sup>, in: *Bergey's Manual of Systematics of Archaea and Bacteria*, John Wiley & Sons, Inc. in association with Bergey's Manual Trust, Hoboken, New Jersey, 2015, pp. 1–34.
- [22] A.C. Anthony, *The Biochemistry of Methylophilic*, Academic Press, London, 1982.
- [23] L. Chistoserdova, Modularity of methylophilic, revisited, *Environ. Microbiol.* 13 (2011) 2603–2622, <https://doi.org/10.1111/j.1462-2920.2011.02464.x>.
- [24] C. Ernst, K. Kayashta, T. Koch, S.S. Venceslau, I.A.C. Pereira, U. Demmer, U. Ermler, C. Dahl, Structural and spectroscopic characterization of a HdrA-like subunit from *Hyphomicrobium denitrificans*, *FEBS J.* 288 (2021) 1664–1678, <https://doi.org/10.1111/febs.15505>.
- [25] X. Cao, T. Koch, L. Steffens, J. Finkensieper, R. Zigann, J.E. Cronan, C. Dahl, Lipoate-binding proteins and specific lipoate-protein ligases in microbial sulfur oxidation reveal an atypical role for an old cofactor, *eLife* 7 (2018), e37439, <https://doi.org/10.7554/eLife.37439>.
- [26] R. Anandham, P. Indiraghandi, M. Madhaiyan, J. Chung, K.Y. Ryu, H.J. Jee, T.-M. Sa, Thiosulfate oxidation, mixotrophic growth of *Methylobacterium goeingense* and *Methylobacterium fujisawaense*, *J. Microbiol. Biotechnol.* 19 (2009) 17–22, <https://doi.org/10.4014/jmb.0802.127>.
- [27] R. Anandham, P. Indiraghandi, M. Madhaiyan, K. Kim, W. Yim, V.S. Saravanan, J. Chung, T. Sa, Thiosulfate oxidation and mixotrophic growth of *Methylobacterium oryzae*, *Can. J. Microbiol.* 53 (2007) 869–876, <https://doi.org/10.1139/W07-057>.
- [28] G.M.H. Suylen, G.C. Steffens, J.G. Kuenen, Chemolithotrophic potential of *Hyphomicrobium* species capable of growth on methylated sulphur compounds, *Arch. Microbiol.* 146 (1986) 192–198, <https://doi.org/10.1007/BF00402350>.
- [29] J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989.
- [30] F.A. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl, *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1997.
- [31] R.M. Horton, PCR mediated recombination and mutagenesis: SOEing together tailor-made genes, *Mol. Biotechnol.* 3 (1995) 93–99, <https://doi.org/10.1007/BF02789105>.
- [32] A. Schäfer, A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach, A. Pühler, Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*, *Gene* 145 (1994) 69–73, [https://doi.org/10.1016/0378-1119\(94\)90324-7](https://doi.org/10.1016/0378-1119(94)90324-7).
- [33] R. Fellay, J. Frey, H.M. Krisch, Interposon mutagenesis of soil and water bacteria: a family of DNA fragments designed for in vivo insertional mutagenesis of Gram-negative bacteria, *Gene* 52 (1987) 147–154, [https://doi.org/10.1016/0378-1119\(87\)90041-2](https://doi.org/10.1016/0378-1119(87)90041-2).
- [34] R. Boden, L.P. Hutt, Determination of kinetic parameters and metabolic modes using the chemostat, in: R. Steffan (Ed.), *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Biodegradation and Bioremediation*, Springer, Cham, 2018, pp. 1–42.
- [35] C. Dahl, Insertional gene inactivation in a phototrophic sulphur bacterium: APS-reductase-deficient mutants of *Chromatium vinosum*, *Microbiology* 142 (1996) 3363–3372, <https://doi.org/10.1099/1350872-142-12-3363>.
- [36] A.G. Brooke, M.M. Attwood, Methylamine uptake by the facultative methylophilic *Hyphomicrobium* X, *Microbiology* 130 (1984) 459–463, <https://doi.org/10.1099/00221287-130-3-459>.

J. Li et al.

BBA - Bioenergetics 1864 (2023) 148932

- [37] A.G. Brooke, M.M. Attwood, Regulation of enzyme synthesis in *Hyphomicrobium X*: growth on mixtures of methylamine and ethanol in continuous cultures, *FEMS Microbiol. Lett.* 29 (1985) 251–256, <https://doi.org/10.1111/j.1574-6968.1985.tb00871.x>.
- [38] A.G. Brooke, M.M. Attwood, Regulation of enzyme synthesis during the growth of *Hyphomicrobium X* on mixtures of methylamine and ethanol, *Microbiology* 129 (1983) 2399–2404, <https://doi.org/10.1099/00221287-129-8-2399>.
- [39] L.M. Hellman, M.G. Fried, Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions, *Nat. Protoc.* 2 (2007) 1849–1861, <https://doi.org/10.1038/nprot.2007.249>.
- [40] C. Bernofsky, M. Swan, An improved cycling assay for nicotinamide adenine dinucleotide, *Anal. Biochem.* 53 (1973) 452–458.
- [41] C.G. Friedrich, F. Bardischewsky, D. Rother, A. Quentmeier, J. Fischer, Prokaryotic sulfur oxidation, *Curr. Opin. Microbiol.* 8 (2005) 253–259, <https://doi.org/10.1016/j.mib.2005.04.005>.
- [42] J.M. Kurth, J.A. Brito, J. Reuter, A. Flegler, T. Koch, T. Franke, E.M. Klein, S. F. Rowe, J.N. Butt, K. Denkmann, I.A.C. Pereira, M. Archer, C. Dahl, Electron accepting units of the dihemecytochrome c TsdA, a bifunctional thiosulfate dehydrogenase/tetrathionate reductase, *J. Biol. Chem.* 291 (2016) 24804–24818, <https://doi.org/10.1074/jbc.M116.753863>.
- [43] C. Dahl, A biochemical view on the biological sulfur cycle, in: P. Lens (Ed.), *Environmental Technologies to Treat Sulfur Pollution: Principles and Engineering*, IWA Publishing, London, 2020, pp. 55–96.
- [44] C. Dahl, C.G. Friedrich, A. Kletzin, Sulfur oxidation in prokaryotes, in: *Encyclopedia of Life Sciences*, John Wiley & Sons Ltd, Chichester, 2008.
- [45] L. Appel, M. Willistein, C. Dahl, U. Ermler, M. Boll, Functional diversity of prokaryotic HdrA(BC) modules: role in flavin-based electron bifurcation processes and beyond, *Biochim. Biophys. Acta Bioenerg.* 1862 (2021), 148379, <https://doi.org/10.1016/j.bbabi.2021.148379>.
- [46] R.P. Saha, S. Samanta, S. Patra, D. Sarkar, A. Saha, M.K. Singh, Metal homeostasis in bacteria: the role of ArsR-SmtB family of transcriptional repressors in combating varying metal concentrations in the environment, *Biomaterials* 30 (2017) 459–503, <https://doi.org/10.1007/s10534-017-0020-3>.
- [47] T. Urakami, J. Sasaki, K.I. Suzuki, K. Komagata, Characterization and description of *Hyphomicrobium denitrificans* sp. nov., *Int. J. Syst. Bacteriol.* 45 (1995) 528–532, <https://doi.org/10.1099/00207713-45-3-528>.
- [48] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, M. Tanabe, KEGG for integration and interpretation of large-scale molecular data sets, *Nucleic Acids Res.* 40 (2012) D109–D114, <https://doi.org/10.1093/nar/gkr988>.
- [49] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* 25 (1997) 3389–3402, <https://doi.org/10.1093/nar/25.17.3389>.
- [50] P.N. Green, J.K. Ardley, Review of the genus *Methylobacterium* and closely related organisms: a proposal that some *Methylobacterium* species be reclassified into a new genus, *Methylorubrum* gen. nov., *Int. J. Syst. Evol. Microbiol.* 68 (2018) 2727–2748, <https://doi.org/10.1099/ijsem.0.002856>.
- [51] H. Smejkalova, T.J. Erb, G. Fuchs, Methanol assimilation in *Methylobacterium extorquens* AM1: demonstration of all enzymes and their regulation, *PLoS One* 5 (2010), e13001, <https://doi.org/10.1371/journal.pone.0013001>.
- [52] J.A. Duine, J. Frank, J. Westerling, Purification and properties of methanol dehydrogenase from *Hyphomicrobium X*, *Biochim. Biophys. Acta* 524 (1978) 277–287.
- [53] J.L. Hemmann, T. Wagner, S. Shima, J.A. Vorholt, Methyloluran is a prosthetic group of the formyltransferase/hydrolase complex and shuttles one-carbon units between two active sites, *Proc. Natl. Acad. Sci. U. S. A.* 116 (2019) 25583–25590, <https://doi.org/10.1073/pnas.1911595116>.
- [54] J.L. Hemmann, O. Saurel, A.M. Ochsner, B.K. Stodden, P. Kiefer, A. Milon, J. A. Vorholt, The one-carbon carrier methyloluran from *Methylobacterium extorquens* AM1 contains a large number of  $\alpha$ - and  $\gamma$ -linked glutamic acid residues, *J. Biol. Chem.* 291 (2016) 9042–9051, <https://doi.org/10.1074/jbc.M116.714741>.
- [55] I.W. Marison, M.M. Attwood, Partial purification and characterization of a dye-linked formaldehyde dehydrogenase from *Hyphomicrobium X*, *Microbiology* 117 (1980) 305–313, <https://doi.org/10.1099/00221287-117-2-305>.
- [56] H. He, E. Noor, P.A. Ramos-Parra, L.E. Garcia-Valencia, J.A. Patterson, R.L. Diaz de la Garza, A.D. Hanson, A. Bar-Even, *In vivo* rate of formaldehyde condensation with tetrahydrofolate, *Metabolites* 10 (2020) 65, <https://doi.org/10.3390/metabo10020065>.
- [57] C. Anthony, How half a century of research was required to understand bacterial growth on C<sub>1</sub> and C<sub>2</sub> compounds: the story of the serine cycle and the ethylmalonyl-CoA pathway, *Sci. Prog.* 94 (2011) 109–137, <https://doi.org/10.3184/003685011X13044430633960>.
- [58] W. Harder, M.M. Attwood, J.R. Quayle, Methanol assimilation by *Hyphomicrobium* sp. *J. Gen. Microbiol.* 78 (1973) 155–163, <https://doi.org/10.1099/00221287-78-1-155>.
- [59] U. Kappler, C. Dahl, Enzymology and molecular biology of prokaryotic sulfite oxidation (minireview), *FEMS Microbiol. Lett.* 203 (2001) 1–9.
- [60] W.S. McIntire, Reaction of 2, 7, 9-Tricarboxy-PQQ with nucleophiles, in: J. A. Jongejans, J.A. Duine (Eds.), *PQQ and Quinoproteins*, Springer, Dordrecht, 1989, pp. 233–235.
- [61] A.R. Dewanti, J.A. Duine, Reconstitution of membrane-integrated quinoprotein glucose dehydrogenase apoenzyme with PQQ and the holoenzyme's mechanism of action, *Biochemistry* 37 (1998) 6810–6818, <https://doi.org/10.1021/bi9722610>.
- [62] C. Martineau, F. Mauffrey, R. Villemur, Comparative analysis of denitrifying activities of *Hyphomicrobium nitrativans*, *Hyphomicrobium denitrificans*, and *Hyphomicrobium zavarzinii*, *Appl. Environ. Microbiol.* 81 (2015) 5003–5014, <https://doi.org/10.1128/AEM.00848-15>.
- [63] R. Venkatramanan, O. Prakash, T. Woyke, P. Chain, L.A. Goodwin, D. Watson, S. Brooks, J.E. Kostka, S.J. Green, Genome sequences for three denitrifying bacterial strains isolated from a uranium- and nitrate-contaminated subsurface environment, *Genome Announc.* 1 (2013), <https://doi.org/10.1128/genomeA.00449-13>.
- [64] R.J. Huddy, R. Sachdeva, F. Kadzinga, R.S. Kantor, S.T.L. Harrison, J.F. Banfield, Thiocyanate and organic carbon inputs drive convergent selection for specific autotrophic *Afipta* and *Thiobacillus* strains within complex microbiomes, *Front. Microbiol.* 12 (2021), 643368, <https://doi.org/10.3389/fmicb.2021.643368>.
- [65] C. Martineau, C. Villeneuve, F. Mauffrey, R. Villemur, *Hyphomicrobium nitrativans* sp. nov., isolated from the biofilm of a methanol-fed denitrification system treating seawater at the Montreal Biodome, *Int. J. Syst. Evol. Microbiol.* 63 (2013) 3777–3781, <https://doi.org/10.1099/ijms.0.048124-0>.
- [66] C. Martineau, C. Villeneuve, F. Mauffrey, R. Villemur, Complete genome sequence of *Hyphomicrobium nitrativans* strain NL23, a denitrifying bacterium isolated from biofilm of a methanol-fed denitrification system treating seawater at the Montreal biodome, *Genome Announc.* 2 (2014), <https://doi.org/10.1128/genomeA.01165-13>.
- [67] R.S. Kantor, S.E. Miller, K.L. Nelson, The water microbiome through a pilot scale advanced treatment facility for direct potable reuse, *Front. Microbiol.* 10 (2019) 993, <https://doi.org/10.3389/fmicb.2019.00993>.
- [68] R.S. Kantor, A.W. van Zyl, R.P. van Hille, B.C. Thomas, S.T. Harrison, J.F. Banfield, Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics, *Environ. Microbiol.* 17 (2015) 4929–4941, <https://doi.org/10.1111/1462-2920.12936>.
- [69] R. Conrad, P. Claus, Contribution of methanol to the production of methane and its <sup>13</sup>C-isotopic signature in anoxic rice field soil, *Biogeochemistry* 73 (2005) 381–393, <https://doi.org/10.1007/s10533-004-0366-9>.
- [70] E.L. Barrett, M.A. Clark, Tetrathionate reduction and production of hydrogen sulfide from thiosulfate, *Microbiol. Rev.* 51 (1987) 192–205, <https://doi.org/10.1128/mr.51.2.192-205.1987>.
- [71] D.L. Jones, Organic acids in the rhizosphere – a critical review, *Plant Soil* 205 (1998) 25–44, <https://doi.org/10.1023/a:1004356007312>.
- [72] M. Bott, Anaerobic citrate metabolism and its regulation in enterobacteria, *Arch. Microbiol.* 167 (1997) 78–88.

## Supplementary Information

### **A metabolic puzzle: Consumption of C<sub>1</sub> compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>**

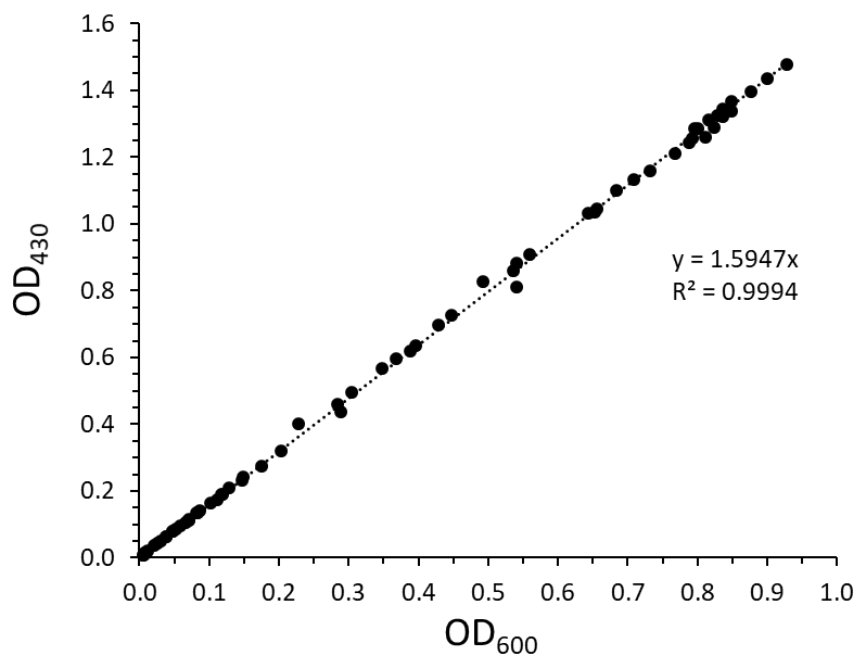
Jingjing Li, Julian Koch, Wanda Flegler, Leon Garcia Ruiz, Natalie Hager, Alina Ballas,  
Tomohisa S. Tanabe, Christiane Dahl

Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich-Wilhelms-Universität  
Bonn, Meckenheimer Allee 168, D-53115 Bonn, Germany

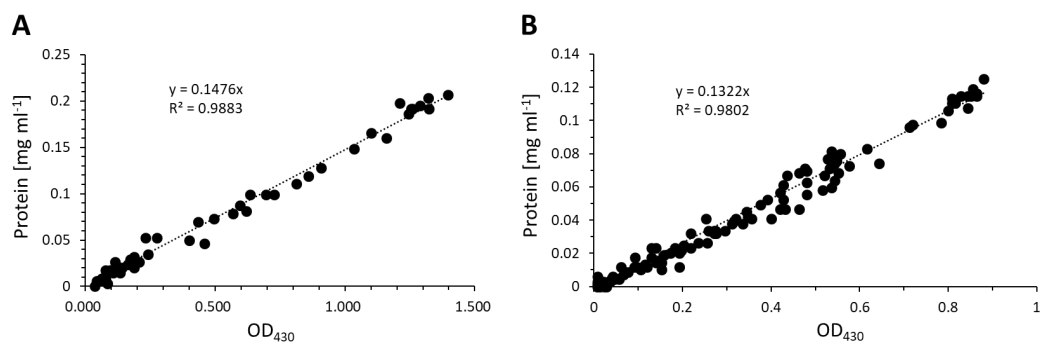
This document contains supplementary information including additional detail concerning bacterial strains, primers and plasmids (Supplementary Table S1), methods used in growth experiments (Supplementary Figures S1 and S2), growth experiments (Supplementary Figures S3-6, S8), electron input and output modules in *H. denitrificans* (Fig. S7), enzyme commission numbers, locus tags and references for enzymes of electron transport, respiration and C<sub>1</sub> metabolism in *H. denitrificans* (Table S2) and occurrence of the *tsdA* gene and *sox* and *shdr* gene clusters in *Hyphomicrobium* species/strains (Table S3).

### Supplementary Figures

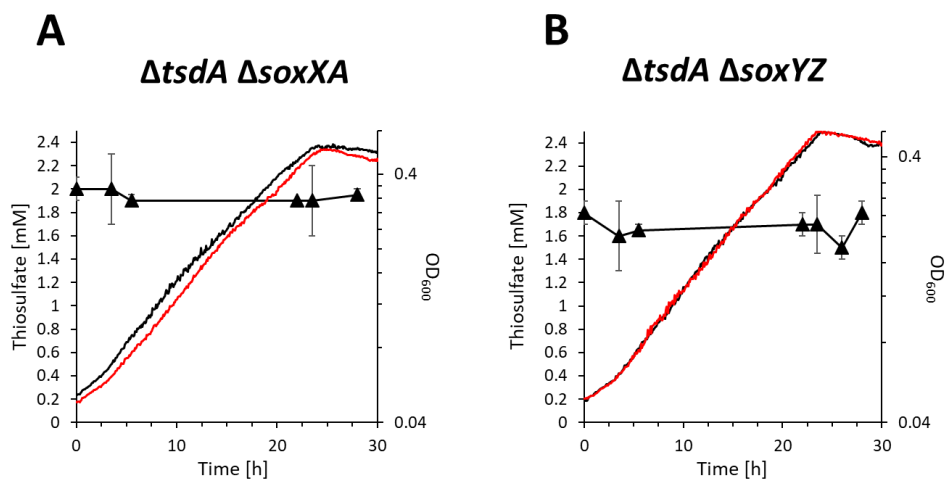
**Fig. S1.** Graph illustrating determination of the factor for OD<sub>600</sub> to OD<sub>430</sub> conversion for *H. denitrificans*. Samples with turbidities exceeding the linear range (here optical densities above 0.3) were diluted with culture medium.



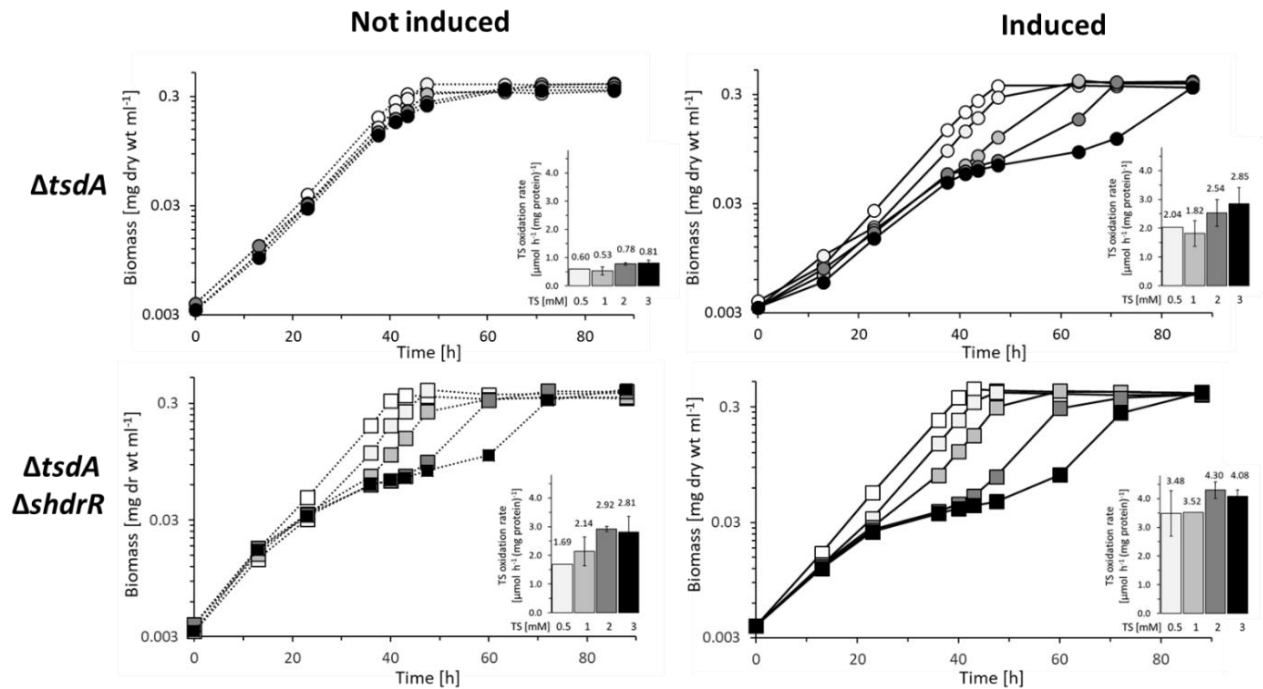
**Fig. S2.** Graph illustrating determination of the factors for OD<sub>430</sub> to protein conversion for *H. denitrificans* growing on methanol (A) and formate (B).



**Fig. S3.** Growth and thiosulfate oxidation by *Hyphomicrobium denitrificans*  $\Delta tsdA$  strains lacking either *soxXA* (A) or *soxYZ* (B). The experiments were performed in 48 well plates in a plate reader with continuous shaking. Solid lines show optical density measurements at 600 nm for cultures on 24.4 mM methanol and 2 mM thiosulfate inoculated with precultures not containing thiosulfate (black, not induced) or containing 2 mM thiosulfate (red, induced). Black triangles represent thiosulfate concentrations for the non-induced cultures. Thiosulfate also stayed constant in the other cases but concentrations are omitted for clarity. Error bars indicate standard deviation for three technical replicates.

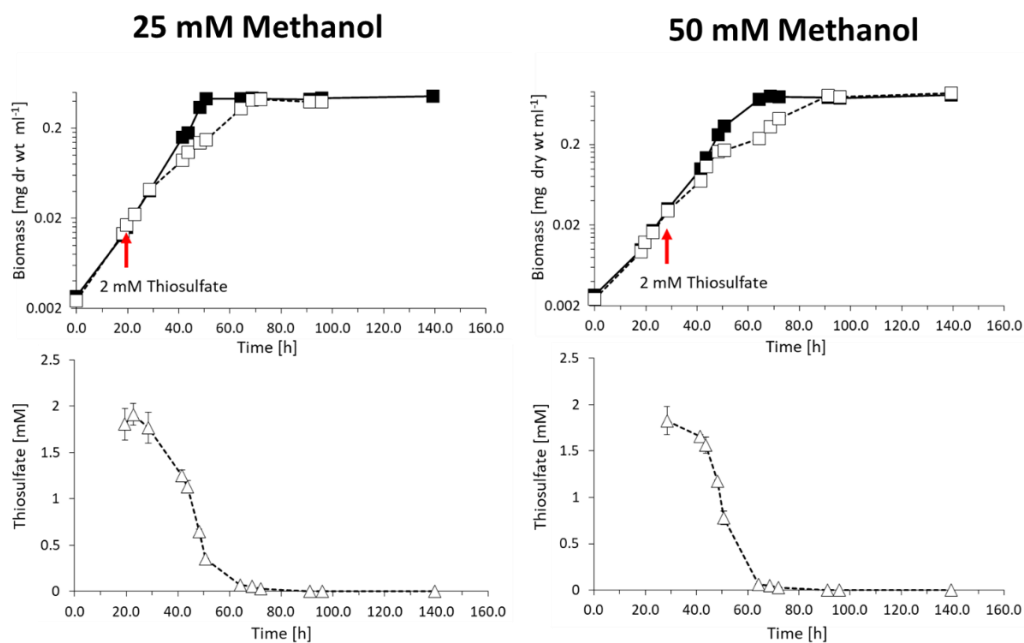


**Fig. S4.** Growth and thiosulfate consumption of *H. denitrificans*  $\Delta tsdA$  (upper panels, circles) and  $\Delta tsdA \Delta shdrR$  (lower panels, boxes). Cultures were grown on methanol-containing medium without (open symbols) or with increasing thiosulfate concentrations (filled symbols:  $\circ$ , 0.5 mM;  $\odot$ , 1 mM;  $\bullet$ , 2 mM;  $\bullet$ , 3 mM thiosulfate). Precultures contained either no thiosulfate (not induced, broken lines) or 2 mM thiosulfate (induced, solid lines). Specific thiosulfate oxidation (TS) rates are depicted as insets and follow the same gray shading code as stated earlier. Error bars indicating SD are too small to be visible.

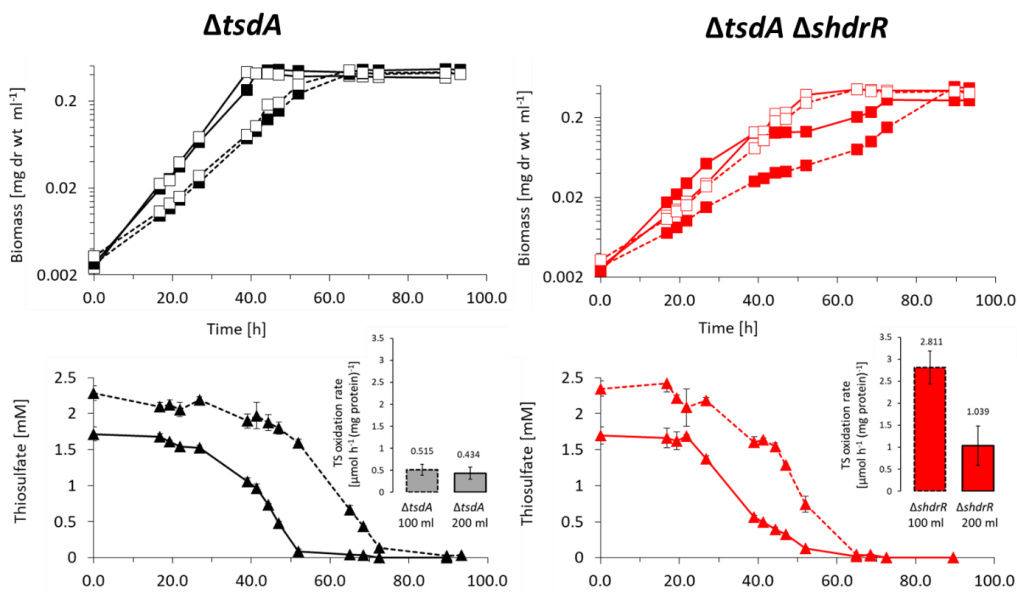




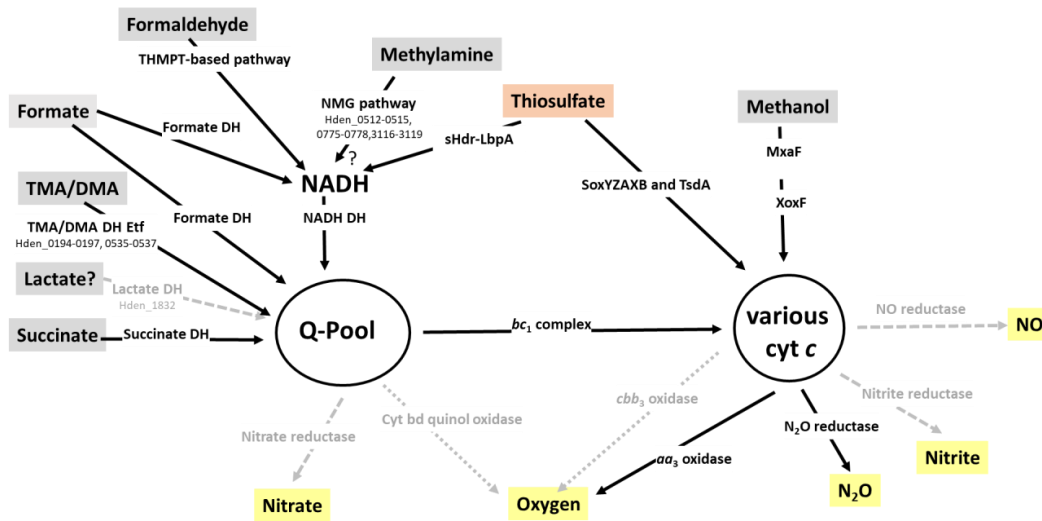
**Fig. S5.** Growth and thiosulfate consumption of *H. denitrificans*  $\Delta tsaA \Delta shdrR$  on 25 mM methanol (left panels) and 50 mM methanol (right panels). Cultures were grown without (filled boxes) or with 2 mM thiosulfate (open boxes). Thiosulfate was added to growing cultures as indicated by red arrows. Pre-cultures did not contain thiosulfate. In the lower panels, triangles indicate thiosulfate concentrations. Error bars indicating SD are too small to be visible for determination of biomass.



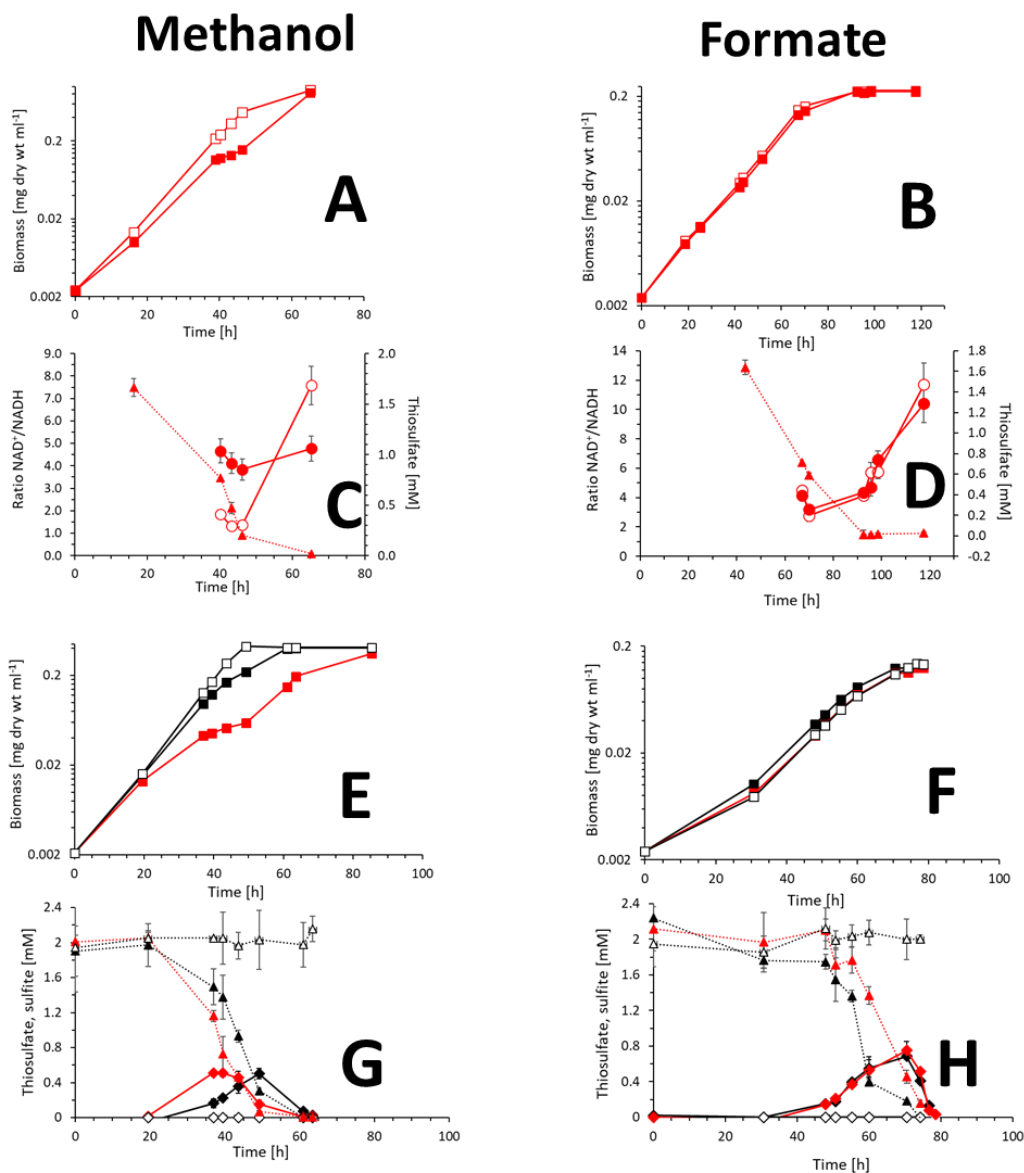
**Fig. S6.** Growth and thiosulfate consumption of *H. denitrificans*  $\Delta tsdA$  (left panels, black boxes) and  $\Delta tsdA \Delta shdR$  (right panels, red boxes). Cultures were grown on medium containing 25 mM methanol without (open symbols) or with 2 mM thiosulfate (filled black or red boxes). Pre-cultures did not contain thiosulfate. In the lower panels, triangles indicate thiosulfate concentrations for *H. denitrificans*  $\Delta tsdA$  (black) and for *H. denitrificans*  $\Delta tsdA \Delta shdR$  (red). Cultures were grown at 30°C and 200 rpm in 500 ml Erlenmeyer flasks containing either 100 ml (broken lines) or 200 ml (solid lines) medium. Specific thiosulfate oxidation rates are depicted as insets and follow the same code. Error bars indicating SD are too small to be visible for determination of biomass.



**Fig. S7.** Overview of electron input and output modules in *Hyphomicrobium denitrificans* ATCC 51888. Proteins that were not detected in a proteomic study during growth on dimethyl sulfide (DMS) at high oxygen concentration [1] are printed in light grey. DMS is transformed to methanethiol in a monooxygenase-catalyzed reaction and sulfide is released from methanethiol by a periplasmic oxidase. In both cases the electrons released from the organosulfur substrate are transferred directly onto oxygen and formaldehyde is formed as a product [1-3] which is then oxidized to formate along a tetrahydromethanopterin (THMPT)-based pathway yielding NADH as indicated in the figure. Methylamine is metabolized via *N*-methyl-glutamate (NMG) using an *N*-methyl-glutamate dehydrogenase [4]. Di- and trimethylamine dehydrogenases are located in the cytoplasm [5]. TMA/DMA, trimethyl-/dimethylamine; DH, dehydrogenase; MxaF and XoxF, methanol dehydrogenase. Locus tags are given for enzymes/pathways not listed in Table S2.



**Fig. S8.** Growth of *H. denitrificans* on 50 mM methanol (A) and 50 mM formate (B) in the absence (open boxes) and presence of thiosulfate (filled boxes).  $\text{NAD}^+/\text{NADH}$  ratios in *H. denitrificans*  $\Delta\text{tsdA } \Delta\text{shdrR}$  on methanol (C) and formate (D) are given for cells grown in the absence (open circles) or in the presence of thiosulfate (filled circles). Thiosulfate is indicated by red filled triangles and was added at a biomass of  $0.033 \text{ mg dry wt ml}^{-1}$ . Panels E and F show growth for *H. denitrificans*  $\Delta\text{tsdA}$  (black), for *H. denitrificans*  $\Delta\text{tsdA } \Delta\text{shdrR}$  (red) and a thiosulfate-oxidation negative strain (open) in the presence of 2 mM thiosulfate on 25 mM methanol or 25 mM formate, respectively. Consumption of thiosulfate and formation of sulfite are shown in panels G and H. Thiosulfate is given as triangles, sulfite concentrations are shown by diamonds following the same color code as in panels E and F.



## Supplementary Tables

Table S1. Strains, plasmids and primers

Strains primers or plasmids	Relevant genotype, description or sequence	Reference or source
<b>Strains</b>		
<i>E. coli</i> NEB 10 $\beta$	$\Delta(\text{ara-leu})$ 7697 <i>araD139 fhuA <math>\Delta</math>lacX74 galK16 galE15 e14-<math>\phi</math>80dlacZ<math>\Delta</math>M15 recA1 relA1 endA1 nupG rpsL (Str<sup>r</sup>) rph spoT1 <math>\Delta</math>(<i>mrr-hsdRMS-mcrBC</i>)</i>	New England Biolabs
<i>E. coli</i> BL21 (DE3)	F- <i>ompT hsdS<sub>B</sub>(r<sub>B</sub> m<sub>B</sub>) gal dcm met</i> (DE3)	Novagen
<i>H. denitrificans</i> $\Delta$ <i>tsdA</i>	Sm <sup>R</sup> , in-frame deletion of <i>tsdA</i> (Hden_2748) in <i>H. denitrificans</i> Sm200	[1]
<i>H. denitrificans</i> $\Delta$ <i>tsdA</i> $\Delta$ <i>shdrR</i>	In-frame deletion of <i>shdrR</i> (Hden_0682) in <i>H. denitrificans</i> $\Delta$ <i>tsdA</i>	This work
<i>H. denitrificans</i> $\Delta$ <i>tsdA</i> $\Delta$ <i>soxXA</i>	In-frame deletion of <i>soxX</i> (Hden_0702) and <i>soxA</i> (Hden_0703) in <i>H. denitrificans</i> $\Delta$ <i>tsdA</i>	This work
<i>H. denitrificans</i> $\Delta$ <i>tsdA</i> $\Delta$ <i>soxYZ</i>	In-frame deletion of <i>soxY</i> (Hden_0704) and <i>soxZ</i> (Hden_0705) in <i>H. denitrificans</i> $\Delta$ <i>tsdA</i>	This work
<b>Primers</b>		
Hden_0703_MaDel_upFW	ACGCTAGATTGAAGGACGCGGTGAACCTATTG (XbaI)	This work
Hden_0703_MaDel_upRV	GCGTAGAACGGTCTTAGCGCATGGGTCACCAAATCTCTGC	This work
Hden_0703_MaDel_dwnFW	GCAGAAATTTGGTGACCCATGCGCTAAGACCGTTCTACGC	This work
Hden_0703_MaDel_dwnRV	GTATTCTAGACTAACGACGACGAAGGTGGGCGTG (XbaI)	This work
Hden_0704_MaDel_upFW	AGGCTCTAGAGATGGGACATTGATCTCTCC (XbaI)	This work
Hden_0704_MaDel_upRV	GACGACGCCGAACCCGTCATTCCATCAGCCATCTCTC	This work
Hden_0704_MaDel_dwnFW	GAGAGATGGCGTGATGGAATGACGGGTTCCGGCGTCGTC	This work
Hden_0704_MaDel_dwnRV	CAACTCTAGACTTGTATGGCGCACGCGAC (XbaI)	This work
Fwd_deltaHden0682_BamHI	GCATGGATCCGCGAAAATGTGCACCGGAG (BamHI)	This work
Up_Rev_deltaHden0682	GCTGAAGACTTCGCTCTAATTAGCCATAGGAGTTGCATCCA	This work
Down_Fwd_deltaHden0682	TGGATGCAACTCCTATGGCTAATTAGAGCGAGTCTTCAGC	This work
Rev_deltaHden0682_XbaI	AAGCTCTAGATATGCGGCAGCCGTTGACGC (XbaI)	This work
Fr-pet22b-Hden0682-NdeI	GGCACATATGGCTGCTGTAAGCCACG (NdeI)	This work
Rev-pet22b-Hden0682-NotI	TTTTGCGGCCGCATTCCAGCGTTTTCCCGCAC (NotI)	This work
EMSA-Fr	TTCCCGCCCCGTCTTGGTTT	This work
EMSA-Rev	AGGAGTTGCATCCAAAAAAGCGTG	This work
<b>Plasmids</b>		
pET-22b (+)	Ap <sup>R</sup> , T7 promoter, lac operator, C-terminal His tag, pelB leader	Novagen
pHP45 $\Omega$ -Tc	Ap <sup>R</sup> , Tc <sup>R</sup>	[6]
pET-22bHdsHdrR	Ap <sup>R</sup> , NdeI-NotI fragment of PCR-amplified <i>shdrR</i> (Hden_0682) in NdeI-NotI of pET-22b (+)	This work
pk18 <i>mobsacB</i>	Km <sup>R</sup> , Mob <sup>+</sup> , <i>sacB</i> , <i>oriV</i> , <i>oriT</i> , <i>lacZ<math>\alpha</math></i>	[7]
pk18 <i>mobsacB</i> -Tc	pHP45 $\Omega$ Tc tetracycline cassette inserted into pk18 <i>mobsacB</i> using SmaI	This work
pk18 <i>mobsacB</i> $\Delta$ <i>soxYZ</i> -Tc	Km <sup>R</sup> , Tc <sup>R</sup> , 2 kb XbaI fragment of PCR-amplified genome region around <i>soxYZ</i> with deletion of <i>soxYZ</i> cloned into XbaI of pk18 <i>mobsacB</i> -Tc	This work
pk18 <i>mobsacB</i> $\Delta$ <i>soxXA</i> -Tc	Km <sup>R</sup> , Tc <sup>R</sup> , 1.82 kb XbaI fragment of PCR-amplified genome region around <i>soxXA</i> with deletion of <i>soxXA</i> cloned into XbaI of pk18 <i>mobsacB</i> -Tc	This work
pk18 <i>mobsacB</i> $\Delta$ <i>shdrR</i>	Km <sup>R</sup> , 2.2 kb BamHI/XbaI fragment of PCR-amplified genome region around <i>shdrR</i> with deletion of <i>shdrR</i> cloned into BamHI/XbaI of pk18 <i>mobsacB</i>	This work
pk18 <i>mobsacB</i> $\Delta$ <i>shdrR</i> -Tc	Km <sup>R</sup> , Tc <sup>R</sup> , pHP45 $\Omega$ Tc tetracycline cassette inserted into pk18 <i>mobsacB</i> $\Delta$ <i>shdrR</i> using SmaI	This work

**Table S2. Respiratory enzyme complexes and enzymes of C<sub>1</sub> metabolism in *H. denitrificans* X<sup>T</sup>**

Enzyme	EC number	Subunits	Locus tags in <i>H. denitrificans</i> X <sup>T</sup>	Reference enzyme activity/presence
Complex I: NADH:ubiquinone oxidoreductase	7.1.1.2	NuoABCDEFGHIJKLMN	Hden_1929-1946	
Complex II: succinate dehydrogenase	1.3.5.1	SdhABCD	Hden_3236, Hden_3238-3240	[8]
Complex III: ubiquinol: cytochrome c reductase	7.1.1.8	Cyt 1 CytB ISP	Hden_2526-2528	
Cytochrome c oxidase: cytochrome aa <sub>3</sub> -type	7.1.1.9	CoxABC	Hden_2903, 2907, 2908	[9, 10]
Cytochrome c oxidase: cytochrome cbb <sub>3</sub> -type	7.1.1.9	CcoNOQP	Hden_2047-2050	
Cytochrome <i>bd</i> ubiquinol oxidase	7.1.1.7	CydAB	Hden_3144-3145	
Nitrate reductase	1.7.5.1	NarGHI	Hden_0926, 0927, 0929	[11]
Nitrite reductase	1.7.2.1	NirK	Hden_0591	[12-14]
Nitric oxide (NO) reductase	1.7.2.5	cNorCBQD	Hden_0581-0584	[14]
Nitrous oxide (N <sub>2</sub> O) reductase	1.7.2.4	NosZ	Hden_1882	[15]
Methanol dehydrogenase PQQ-dependent	1.1.2.7	MxaFIG	Hden_1320, 1321, 1323	[16, 17]
Methanol dehydrogenase, lanthanide-dependent	1.1.2.10	XoxF	Hden_1305 Hden_1869 Hden_1617 Hden_2848	
Formaldehyde-activating enzyme, NAD-linked glutathione-independent, 5.6.7.8-THMPT hydrolase	4.2.1.147	Fae	Hden_1474 Hden_1875 Hden_2126	[18]
Methylene-THMPT dehydrogenase	1.5.1.-	MtdB	Hden_1479	[19]
Methenyl-THMPT cyclohydrolase	3.5.4.27	Mch	Hden_1477	
Formylmethanofuran-THMPT N-formyltransferase, formyltransferase/hydrolase complex	2.3.1.101	FhcABCD	Hden_1608-1611	
Formate dehydrogenase	1.17.1.9	FdoGHI FdhD FdwAB	Hden_2464-2486 Hden_0607, 0527	[20]
Formate-tetrahydrofolate ligase	6.3.4.3	FtfL	Hden_0104 Hden_3136	[21]
Bifunctional methylene tetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase	1.5.1.5/ 3.5.49	Fold	Hden_3211 Hden_0103	[19, 22]
Glycine hydroxymethyltransferase	2.1.2.1	GlyA	Hden_0960	[23]

**Table S3. Occurrence of genes for enzymes involved in thiosulfate and cytoplasmic sulfur oxidation in the genus *Hyphomicrobium*.** The gene *soxX* is put in brackets, because it is not always present. In classical heterodimeric SoxAX proteins, SoxX serves as the site of electron storage and transfer, while SoxA harbours the catalytically active site. It is therefore well conceivable that SoxA alone is active and transfers electrons directly to a separate cytochrome *c* acceptor encoded elsewhere in the genome.

<i>Hyphomicrobium</i> species	<i>tsdA</i>	<i>soxA(X)BYZ</i>	<i>shdrC1B1AHC2B2</i>	Genome assembly accession	Assembly level	Reference
<i>Hyphomicrobium album</i> XQ2 <sup>T</sup> (KCTC 82378)	-	-	-	GCA_009708035.1	Contig	[24]
<i>Hyphomicrobium denitrificans</i> X <sup>T</sup> (DSM1869)	Hden_2748	Hden_0703-0706	Hden_0689-0694	GCA_000143145.1	Complete	[14, 25]
<i>Hyphomicrobium denitrificans</i> 1NES1	-	-	-	GCA_000230975.3	Complete	[26]
<i>Hyphomicrobium denitrificans</i> SCN18_30_10_14_R2_B_61_9	J0H36_02460	J0H36_02935	J0H36_02865-02890	GCA_017304115.1	Scaffold	[27]
<i>Hyphomicrobium denitrificans</i> SCN18_30_10_14_R3_B_61_7	J0H04_09105	-	-	GCA_017307215.1	Scaffold	[27]
<i>Hyphomicrobium denitrificans</i> SCN18_30_10_14_R1_P_61_7	-	J0H37_09715-30	J0H37_09135-40*	GCA_017305615.1	Scaffold	[27]
<i>Hyphomicrobium facile</i> subsp. <i>facile</i> DSM 1565 <sup>T</sup>	SAMN04488557_3656	-	-	GCA_900116175	Contig	[28, 29]
<i>Hyphomicrobium methylivorans</i> Bras1	DLM45_14295	DLM45_12165-85	-	GCA_013626205.1	Contig	[30]
<i>Hyphomicrobium nitrativorans</i> NL23 <sup>T</sup> (ATCC BAA-2476)	-	-	-	GCA_000503895.1	Complete	[31, 32]
<i>Hyphomicrobium sulfonivorans</i> S1 <sup>T</sup> (DSM 13863)	-	-	-	GCA_013306565.1	Contig	[33]
<i>Hyphomicrobium sulfonivorans</i> WDL6	-	-	-	GCA_016125985.1	Contig	
<i>Hyphomicrobium zavarzinii</i> ATCC 27496 <sup>T</sup>	-	-	-	GCA_001541235.1	Scaffold	Albers, P., unpublished
<i>Hyphomicrobium zavarzinii</i> new MAG-140	-	F812_RS23075, 0107895, 930-940	-	GCA_000383415.1	Scaffold	[28]
<i>Hyphomicrobium</i> sp. AWTP1-2	EKK30_01635	JNN24_16950, 980-90	-	GCA_016793385.1	Contig	[34]
<i>Hyphomicrobium</i> sp. AWTP1-10	EKK38_19120	-	-	GCA_003987855.1	Scaffold	[35]
<i>Hyphomicrobium</i> sp. GJ21	HYPGJ_31387	-	-	GCA_003987725.1	Scaffold	[35]
<i>Hyphomicrobium</i> sp. MAG_27	-	HYPGJ_30398-402	HYPGJ_30410-15	GCA_001006785.1	Contig	[36]
<i>Hyphomicrobium</i> sp. MC1	HYPMC_2182	-	-	GCA_019912585.1	Contig	[37]
<i>Hyphomicrobium</i> sp. 802	-	-	-	GCA_000253295.1	Complete	Ge-scope
				GCA_000526135.1	Contig	Christoserdova, L. et al., unpublished

<i>Hyphomicrobium</i> sp. ghe19	HYPP_04503	-	-	LR743509	Complete	Creemers, G. unpublished
<i>Hyphomicrobium</i> sp. CS1BSMeth3	-	-	CS1BSM3_RS27400-45	GCA_900117415.1	Contig	Adelskov, J., Patel, K.C.B., unpublished
<i>Hyphomicrobium</i> sp. NBD2Meth4	-	-	-	GCA_900117445.1	Contig	Adelskov, J., Patel, K.C.B., unpublished
<i>Hyphomicrobium</i> sp. 99	G359_RS04385	-	-	GCA_000384335.2	Scaffold	Adelskov, J., Patel, K.C.B., unpublished
<i>Hyphomicrobium</i> sp. New MAG-139	-	-	-	GCA_016793445.1	Contig	[34]
<i>Hyphomicrobium</i> sp. RIG5714	-	-	-	GCA_017512425.1	Contig	[38]
<i>Hyphomicrobium</i> sp. SB8	-	-	-	GCA_009026145.1	Contig	unpublished
<i>Hyphomicrobium</i> sp. Hjor_18-Q3-R7-51_BAT3C.262	-	-	-	GCA_016716675.1	Contig	[39]
<i>Hyphomicrobium</i> sp. DS3.3.23	-	-	-	GCA_002928735.1	Contig	[40]
<i>Hyphomicrobium</i> sp. FW.3.32	-	CTY20_06750-65 CTY20_09450-65	-	GCA_002928955.1	Contig	[40]
<i>Hyphomicrobium</i> sp. PB1.3.35	-	-	-	GCA_002928515.1	Contig	[40]
<i>Hyphomicrobium</i> sp. WM.3.63	-	-	-	GCA_002928465.1	Contig	[40]
<i>Hyphomicrobium</i> sp. WM.3.50	-	C0511_04010-25 C0511_16275-85	-	GCA_013821915.1	Contig	[40]
<i>Hyphomicrobium</i> sp. WM.3.5	-	CTY40_030865-80 CTY40_10070-80	-	GCA_002928395.1	Contig	[40]
<i>Hyphomicrobium</i> sp. 12-62-95	-	-	-	GCA_002279935.1	Scaffold	Kantor, R.S. et al, unpublished
<i>Hyphomicrobium</i> sp. 32-62-53	-	-	-	GCA_002280885.1	Scaffold	Kantor, R.S. et al, unpublished
<i>Hyphomicrobium</i> sp. Time.spades.CONCOCT.2.5kb_061	-	-	-	GCA_019744875.1	Contig	[41]
<i>Hyphomicrobium</i> sp. co.spades.DASTOOL.2.5kb_064	-	-	-	GCA_019748215.1	Contig	[26]
<i>Hyphomicrobium</i> sp. co.spades.CONCOCT.2.5kb_134	-	-	-	GCA_019751875.1	Contig	[41]
<i>Hyphomicrobium</i> sp. co.spades.CONCOCT.1kb_066	K2Q04_10990	-	-	GCA_019753045.1	Contig	[41]
<i>Hyphomicrobium</i> sp. SJ665	-	-	-	GCA_020852195.1	Contig	[42]
<i>Hyphomicrobium</i> sp. SCN 65-11	-	-	ABS54_03385-03410	GCA_001724295.1	Scaffold	[43]
<i>Hyphomicrobium</i> sp. P-RSF-NP-07	-	-	-	GCA_014379315.1	Contig	[44]
<i>Hyphomicrobium</i> sp. SCN18_10_11_15_R1_B_65_8	-	J0J14_11055-70* J0J14_14585*	J0J14_15020-040*	GCA_017306765.1	Scaffold	[27]
<i>Hyphomicrobium</i> sp. SCN18_10_11_15_R2_B_65_9	-	J0I57_14080-90*	J0I57_20835-50*	GCA_017306735.1	Scaffold	[27]
<i>Hyphomicrobium</i> sp. SCN18_30_10_14_R3_B_64_9	-	J0I75_12920-30*	J0I75_04305-030	GCA_017306765.1	Scaffold	[27]
<i>Hyphomicrobium</i> sp. SCN18_26_2_15_R2_B_61_8	J0I81_01825	-	-	GCA_017306805.1	Scaffold	[27]

\*Partly present





## References

- [1] T. Koch, C. Dahl, A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds, *ISME J.*, 12 (2018) 2479-2491. doi 10.1038/s41396-018-0209-7
- [2] O. Eyice, N. Myronova, A. Pol, O. Carrión, J.D. Todd, T.J. Smith, S.J. Gurman, A. Cuthbertson, S. Mazard, M.A. Mennink-Kersten, T.D. Bugg, K.K. Andersson, A.W. Johnston, H.J. Op den Camp, H. Schäfer, Bacterial SBP56 identified as a Cu-dependent methanethiol oxidase widely distributed in the biosphere, *ISME J.*, 12 (2017) 145-160. doi 10.1038/ismej.2017.148
- [3] H. Schäfer, N. Myronova, R. Boden, Microbial degradation of dimethylsulphide and related C<sub>1</sub>-sulphur compounds: organisms and pathways controlling fluxes of sulphur in the biosphere, *J. Exp. Bot.*, 61 (2010) 315-334. doi 10.1093/jxb/erp355
- [4] J.B.M. Meiberg, W. Harder, Aerobic and anaerobic metabolism of trimethylamine, dimethylamine and methylamine in *Hyphomicrobium X*, *J. Gen. Microbiol.*, 106 (1978) 265-276. doi 10.1099/00221287-106-2-265
- [5] A.A. Kasprzak, D.J. Steenkamp, Localization of the major dehydrogenases in two methylotrophs by radiochemical labeling, *J. Bacteriol.*, 156 (1983) 348-353. doi 10.1128/jb.156.1.348-353.1983
- [6] R. Fellay, J. Frey, H.M. Krisch, Interposon mutagenesis of soil and water bacteria: a family of DNA fragments designed for in vivo insertional mutagenesis of Gram-negative bacteria, *Gene*, 52 (1987) 147-154. doi 10.1016/0378-1119(87)90041-2
- [7] A. Schäfer, A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach, A. Pühler, Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*, *Gene*, 145 (1994) 69-73. doi 10.1016/0378-1119(94)90324-7
- [8] M.M. Attwood, W. Harder, The oxidation and assimilation of C<sub>2</sub> compounds by *Hyphomicrobium* sp., *J. Gen. Microbiol.*, 84 (1974) 350-356. doi 10.1099/00221287-84-2-350
- [9] D. Widdowson, C. Anthony, The microbial metabolism of C<sub>1</sub> compounds. The electron-transport chain of *Pseudomonas* Aml, *Biochem. J.*, 152 (1975) 349-356. doi 10.1042/bj1520349
- [10] P.J. Large, J.B.M. Meiberg, W. Harder, Cytochrome *c<sub>oo</sub>* is not a primary electron acceptor for the amine dehydrogenases of *Hyphomicrobium X*, *FEMS Microbiol. Lett.*, 5 (1979) 281-286. doi 10.1016/0378-1119(79)90041-2
- [11] J.B.M. Meiberg, P.M. Bruinenberg, W. Harder, Effect of dissolved oxygen tension on the metabolism of methylated amines in *Hyphomicrobium X* in the absence and presence of nitrate: evidence for 'aerobic' denitrification, *Microbiology*, 120 (1980) 453-463. doi 10.1099/00221287-120-2-453
- [12] Deligeer, R. Fukunaga, K. Kataoka, K. Yamaguchi, K. Kobayashi, S. Tagawa, S. Suzuki, Spectroscopic and functional characterization of Cu-containing nitrite reductase from *Hyphomicrobium denitrificans* A3151, *Journal of Inorganic Biochemistry*, 91 (2002) 132-138. doi 10.1016/s0162-0134(02)00442-7
- [13] K. Yamaguchi, K. Kataoka, M. Kobayashi, K. Itoh, A. Fukui, S. Suzuki, Characterization of two type 1 Cu sites of *Hyphomicrobium denitrificans* nitrite reductase: a new class of copper-containing nitrite reductases, *Biochemistry*, 43 (2004) 14180-14188. doi 10.1021/bi0492657
- [14] C. Martineau, F. Mauffrey, R. Villemur, Comparative analysis of denitrifying activities of *Hyphomicrobium nitrativorans*, *Hyphomicrobium denitrificans*, and *Hyphomicrobium zavarzinii*, *Appl. Environ. Microbiol.*, 81 (2015) 5003-5014. doi 10.1128/AEM.00848-15
- [15] K. Yamaguchi, A. Kawamura, H. Ogawa, S. Suzuki, Characterization of nitrous oxide reductase from a methylotrophic denitrifying bacterium, *Hyphomicrobium denitrificans* A3151, *Journal of Biochemistry (Tokyo)*, 134 (2003) 853-858. doi 10.1093/jb/mvg211
- [16] M. Dijkstra, J. Frank, Jr., J.A. Duine, Studies on electron transfer from methanol dehydrogenase to cytochrome *c<sub>L</sub>*, both purified from *Hyphomicrobium X*, *Biochem. J.*, 257 (1989) 87-94. doi 10.1042/bj2570087
- [17] J.A. Duine, J. Frank, J. Westerling, Purification and properties of methanol dehydrogenase from *Hyphomicrobium X*, *Biochim. Biophys. Acta*, 524 (1978) 277-287. doi 10.1016/0014-7167(78)90041-2
- [18] P.A. Poels, J.A. Duine, NAD-linked, GSH- and factor-independent aldehyde dehydrogenase of the methylotrophic bacterium, *Hyphomicrobium X*, *Arch. Biochem. Biophys.*, 271 (1989) 240-245. doi 10.1016/0003-9861(89)90274-9

- [19] M. Goenrich, J. Bursy, E. Hubner, D. Linder, A.C. Schwartz, J.A. Vorholt, Purification and characterization of the methylene tetrahydromethanopterin dehydrogenase MtdB and the methylene tetrahydrofolate dehydrogenase FolD from *Hyphomicrobium zavarzinii* ZV580, *Arch. Microbiol.*, 177 (2002) 299-303. doi 10.1007/s00203-001-0394-y
- [20] I.W. Marison, M.M. Attwood, Partial purification and characterization of a dye-linked formaldehyde dehydrogenase from *Hyphomicrobium* X, *Microbiology*, 117 (1980) 305-313. doi 10.1099/00221287-117-2-305
- [21] M.M. Attwood, W. Harder, Formate assimilation by *Hyphomicrobium* X, *FEMS Microbiol. Lett.*, 3 (1978) 111-114. doi 10.1111/j.1574-6968.1978.tb01895.x
- [22] I.W. Marison, M.M. Attwood, A possible alternative mechanism for the oxidation of formaldehyde to formate, *Microbiology*, 128 (1982) 1441-1446. doi 10.1099/00221287-128-7-1441
- [23] W. Harder, M.M. Attwood, J.R. Quayle, Methanol assimilation by *Hyphomicrobium* sp., *J. Gen. Microbiol.*, 78 (1973) 155-163. doi 10.1099/00221287-78-1-155
- [24] Q. Xu, Y. Zhang, X. Wang, G. Wang, *Hyphomicrobium album* sp. nov., isolated from mountain soil and emended description of genus *Hyphomicrobium*, *Arch. Microbiol.*, 203 (2021) 5931-5936. doi 10.1007/s00203-021-02473-6
- [25] T. Urakami, J. Sasaki, K.I. Suzuki, K. Komagata, Characterization and description of *Hyphomicrobium denitrificans* sp. nov., *Int. J. Syst. Bacteriol.*, 45 (1995) 528-532. doi 10.1099/00207713-45-3-528
- [26] R. Venkatramanan, O. Prakash, T. Woyke, P. Chain, L.A. Goodwin, D. Watson, S. Brooks, J.E. Kostka, S.J. Green, Genome sequences for three denitrifying bacterial strains isolated from a uranium- and nitrate-contaminated subsurface environment, *Genome Announc.*, 1 (2013). doi 10.1128/genomeA.00449-13
- [27] R.J. Huddy, R. Sachdeva, F. Kadzina, R.S. Kantor, S.T.L. Harrison, J.F. Banfield, Thiocyanate and organic carbon inputs drive convergent selection for specific autotrophic *Afiplia* and *Thiobacillus* strains within complex microbiomes, *Front. Microbiol.*, 12 (2021) 643368. doi 10.3389/fmicb.2021.643368
- [28] C. Gliesche, A. Fesefeldt, P. Hirsch, *Hyphomicrobium* Stutzer and Hartleb 1898, 76<sup>AL</sup>, *Bergey's manual of systematics of Archaea and Bacteria*, John Wiley & Sons, Inc., in association with Bergey's Manual Trust, Place Published, 2015, pp. 1-34.
- [29] P. Hirsch, Genus *Hyphomicrobium* Stutzer and Hartleb 1898, 76<sup>AL</sup>, in: J.T. Staley, M.P. Bryant, N. Pfennig, J.G. Holt (Eds.) *Bergey's manual of systematic bacteriology*, The Williams & Wilkins Co., Place Published, 1989, pp. 1895-1904.
- [30] Y. Izumi, M. Takizawa, Y. Tani, H. Yamada, An obligate methylotrophic *Hyphomicrobium* strain identification, growth characteristics and cell composition, *Journal of Fermentation Technology*, 60 (1982) 371-375. doi
- [31] C. Martineau, C. Villeneuve, F. Mauffrey, R. Villemur, *Hyphomicrobium nitrativorans* sp. nov., isolated from the biofilm of a methanol-fed denitrification system treating seawater at the Montreal Biodome, *Int. J. Syst. Evol. Microbiol.*, 63 (2013) 3777-3781. doi 10.1099/ijs.0.048124-0
- [32] C. Martineau, C. Villeneuve, F. Mauffrey, R. Villemur, Complete genome sequence of *Hyphomicrobium nitrativorans* strain NL23, a denitrifying bacterium isolated from biofilm of a methanol-fed denitrification system treating seawater at the Montreal biodome, *Genome Announc.*, 2 (2014). doi 10.1128/genomeA.01165-13
- [33] E. Borodina, D.P. Kelly, P. Schumann, F.A. Rainey, N.L. Ward-Rainey, A.P. Wood, Enzymes of dimethylsulfone metabolism and the phylogenetic characterization of the facultative methylotrophs *Arthrobacter sulfonivorans* sp. nov., *Arthrobacter methylotrophus* sp. nov., and *Hyphomicrobium sulfonivorans* sp. nov., *Arch. Microbiol.*, 177 (2002) 173-183. doi 10.1007/s00203-001-0373-3
- [34] Y. Lin, L. Wang, K. Xu, K. Li, H. Ren, Revealing taxon-specific heavy metal-resistance mechanisms in denitrifying phosphorus removal sludge using genome-centric metaproteomics, *Microbiome*, 9 (2021) 67. doi 10.1186/s40168-021-01016-x
- [35] R.S. Kantor, S.E. Miller, K.L. Nelson, The water microbiome through a pilot scale advanced treatment facility for direct potable reuse, *Front. Microbiol.*, 10 (2019) 993. doi 10.3389/fmicb.2019.00993

- [36] F. Bringel, C.P. Postema, S. Mangenot, S. Bibi-Triki, P. Chaignaud, M. Farhan Ul Haque, C. Gruffaz, L. Hermon, Y. Louhichi, B. Maucourt, E.E.L. Muller, T. Nadalig, A. Lajus, Z. Rouy, C. Medigue, V. Barbe, D.B. Janssen, S. Vuilleumier, Genome sequence of the dichloromethane-degrading bacterium *Hyphomicrobium* sp. strain GJ21, *Genome Announc.*, 5 (2017) e00622-00617. doi 10.1128/genomeA.00622-17
- [37] P.D. Martins, M.J. Echeveste, A. Arshad, J. Kurth, H. Ouboter, M.S.M. Jetten, C.U. Welte, Unraveling nitrogen, sulfur and carbon metabolic pathways and microbial community transcriptional responses to substrate deprivation and toxicity stresses in a bioreactor mimicking anoxic brackish coastal sediment conditions, *BioRxiv*, (2021). doi 10.1101/2021.08.31.458400
- [38] F. Xie, W. Jin, H. Si, Y. Yuan, Y. Tao, J. Liu, X. Wang, C. Yang, Q. Li, X. Yan, L. Lin, Q. Jiang, L. Zhang, C. Guo, C. Greening, R. Heller, L.L. Guan, P.B. Pope, Z. Tan, W. Zhu, M. Wang, Q. Qiu, Z. Li, S. Mao, An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants, *Microbiome*, 9 (2021) 137. doi 10.1186/s40168-021-01078-x
- [39] C.M. Singleton, F. Petriglieri, J.M. Kristensen, R.H. Kirkegaard, T.Y. Michaelsen, M.H. Andersen, Z. Kondrotaitė, S.M. Karst, M.S. Dueholm, P.H. Nielsen, M. Albertsen, Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing, *Nat. Commun.*, 12 (2021) 2009. doi 10.1038/s41467-021-22203-2
- [40] Y. Zhang, M. Kitajima, A.J. Whittle, W.T. Liu, Benefits of genomic insights and CRISPR-Cas signatures to monitor potential pathogens across drinking water production and distribution systems, *Front. Microbiol.*, 8 (2017) 2036. doi 10.3389/fmicb.2017.02036
- [41] S. Vosloo, L. Huo, C.L. Anderson, Z. Dai, M. Sevilano, A. Pinto, Evaluating de novo assembly and binning strategies for time series drinking water metagenomes, *Microbiology Spectrum*, 9 (2021) e0143421. doi 10.1128/Spectrum.01434-21
- [42] C. Suarez, C.J. Sedlacek, D.J.I. Gustavsson, A. Eiler, O. Modin, M. Hermansson, F. Persson, Disturbance-based management of ecosystem services and disservices in partial nitrification anammox biofilms, *BioRxiv*, (2021). doi 10.1101/2021.07.05.451122
- [43] R.S. Kantor, A.W. van Zyl, R.P. van Hille, B.C. Thomas, S.T. Harrison, J.F. Banfield, Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics, *Environ. Microbiol.*, 17 (2015) 4929-4941. doi 10.1111/1462-2920.12936
- [44] L. Poghosyan, H. Koch, J. Frank, M.A.H.J. van Kessel, G. Cremers, T. van Alen, M.S.M. Jetten, H.J.M. Op den Camp, S. Lucker, Metagenomic profiling of ammonia- and methane-oxidizing microorganisms in a Dutch drinking water treatment plant, (2020). doi 10.1101/2020.05.19.103440
- [1] T. Koch, C. Dahl, A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds, *ISME J.*, 12 (2018) 2479-2491. doi 10.1038/s41396-018-0209-7
- [2] O. Eyice, N. Myronova, A. Pol, O. Carrión, J.D. Todd, T.J. Smith, S.J. Gurman, A. Cuthbertson, S. Mazard, M.A. Mennink-Kersten, T.D. Bugg, K.K. Andersson, A.W. Johnston, H.J. Op den Camp, H. Schäfer, Bacterial SBP56 identified as a Cu-dependent methanethiol oxidase widely distributed in the biosphere, *ISME J.*, 12 (2017) 145-160. doi 10.1038/ismej.2017.148
- [3] H. Schäfer, N. Myronova, R. Boden, Microbial degradation of dimethylsulphide and related C<sub>1</sub>-sulphur compounds: organisms and pathways controlling fluxes of sulphur in the biosphere, *J. Exp. Bot.*, 61 (2010) 315-334. doi 10.1093/jxb/erp355
- [4] J.B.M. Meiberg, W. Harder, Aerobic and anaerobic metabolism of trimethylamine, dimethylamine and methylamine in *Hyphomicrobium* X, *J. Gen. Microbiol.*, 106 (1978) 265-276. doi 10.1099/00221287-106-2-265
- [5] A.A. Kasprzak, D.J. Steenkamp, Localization of the major dehydrogenases in two methylotrophs by radiochemical labeling, *J. Bacteriol.*, 156 (1983) 348-353. doi 10.1128/jb.156.1.348-353.1983
- [6] R. Fellay, J. Frey, H.M. Krisch, Interposon mutagenesis of soil and water bacteria: a family of DNA fragments designed for in vivo insertional mutagenesis of Gram-negative bacteria, *Gene*, 52 (1987) 147-154. doi 10.1016/0378-1119(87)90041-2
- [7] A. Schäfer, A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach, A. Pühler, Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of

- defined deletions in the chromosome of *Corynebacterium glutamicum*, *Gene*, 145 (1994) 69-73. doi 10.1016/0378-1119(94)90324-7
- [8] Q. Xu, Y. Zhang, X. Wang, G. Wang, *Hyphomicrobium album* sp. nov., isolated from mountain soil and emended description of genus *Hyphomicrobium*, *Arch. Microbiol.*, 203 (2021) 5931-5936. doi 10.1007/s00203-021-02473-6
- [9] T. Urakami, J. Sasaki, K.I. Suzuki, K. Komagata, Characterization and description of *Hyphomicrobium denitrificans* sp. nov., *Int. J. Syst. Bacteriol.*, 45 (1995) 528-532. doi 10.1099/00207713-45-3-528
- [10] C. Martineau, F. Mauffrey, R. Villemur, Comparative analysis of denitrifying activities of *Hyphomicrobium nitratorans*, *Hyphomicrobium denitrificans*, and *Hyphomicrobium zavarzinii*, *Appl. Environ. Microbiol.*, 81 (2015) 5003-5014. doi 10.1128/AEM.00848-15
- [11] R. Venkatramanan, O. Prakash, T. Woyke, P. Chain, L.A. Goodwin, D. Watson, S. Brooks, J.E. Kostka, S.J. Green, Genome sequences for three denitrifying bacterial strains isolated from a uranium- and nitrate-contaminated subsurface environment, *Genome Announc.*, 1 (2013). doi 10.1128/genomeA.00449-13
- [12] R.J. Huddy, R. Sachdeva, F. Kadzinga, R.S. Kantor, S.T.L. Harrison, J.F. Banfield, Thiocyanate and organic carbon inputs drive convergent selection for specific autotrophic *Afipia* and *Thiobacillus* strains within complex microbiomes, *Front. Microbiol.*, 12 (2021) 643368. doi 10.3389/fmicb.2021.643368
- [13] C. Gliesche, A. Fesefeldt, P. Hirsch, *Hyphomicrobium* Stutzer and Hartleb 1898, 76<sup>AL</sup>, *Bergey's manual of systematics of Archaea and Bacteria*, John Wiley & Sons, Inc., in association with Bergey's Manual Trust, Place Published, 2015, pp. 1-34.
- [14] P. Hirsch, Genus *Hyphomicrobium* Stutzer and Hartleb 1898, 76<sup>AL</sup>, in: J.T. Staley, M.P. Bryant, N. Pfennig, J.G. Holt (Eds.) *Bergey's manual of systematic bacteriology*, The Williams & Wilkins Co., Place Published, 1989, pp. 1895-1904.
- [15] Y. Izumi, M. Takizawa, Y. Tani, H. Yamada, An obligate methylotrophic *Hyphomicrobium* strain identification, growth characteristics and cell composition, *Journal of Fermentation Technology*, 60 (1982) 371-375. doi
- [16] C. Martineau, C. Villeneuve, F. Mauffrey, R. Villemur, *Hyphomicrobium nitratorans* sp. nov., isolated from the biofilm of a methanol-fed denitrification system treating seawater at the Montreal Biodome, *Int. J. Syst. Evol. Microbiol.*, 63 (2013) 3777-3781. doi 10.1099/ijs.0.048124-0
- [17] C. Martineau, C. Villeneuve, F. Mauffrey, R. Villemur, Complete genome sequence of *Hyphomicrobium nitratorans* strain NL23, a denitrifying bacterium isolated from biofilm of a methanol-fed denitrification system treating seawater at the Montreal biodome, *Genome Announc.*, 2 (2014). doi 10.1128/genomeA.01165-13
- [18] E. Borodina, D.P. Kelly, P. Schumann, F.A. Rainey, N.L. Ward-Rainey, A.P. Wood, Enzymes of dimethylsulfone metabolism and the phylogenetic characterization of the facultative methylotrophs *Arthrobacter sulfonivorans* sp. nov., *Arthrobacter methylotrophus* sp. nov., and *Hyphomicrobium sulfonivorans* sp. nov., *Arch. Microbiol.*, 177 (2002) 173-183. doi 10.1007/s00203-001-0373-3
- [19] Y. Lin, L. Wang, K. Xu, K. Li, H. Ren, Revealing taxon-specific heavy metal-resistance mechanisms in denitrifying phosphorus removal sludge using genome-centric metaproteomics, *Microbiome*, 9 (2021) 67. doi 10.1186/s40168-021-01016-x
- [20] R.S. Kantor, S.E. Miller, K.L. Nelson, The water microbiome through a pilot scale advanced treatment facility for direct potable reuse, *Front. Microbiol.*, 10 (2019) 993. doi 10.3389/fmicb.2019.00993
- [21] F. Bringel, C.P. Postema, S. Mangenot, S. Bibi-Triki, P. Chaignaud, M. Farhan Ul Haque, C. Gruffaz, L. Hermon, Y. Louhichi, B. Maucourt, E.E.L. Muller, T. Nadalig, A. Lajus, Z. Rouy, C. Medigue, V. Barbe, D.B. Janssen, S. Vuilleumier, Genome sequence of the dichloromethane-degrading bacterium *Hyphomicrobium* sp. strain GJ21, *Genome Announc.*, 5 (2017) e00622-00617. doi 10.1128/genomeA.00622-17
- [22] P.D. Martins, M.J. Echeveste, A. Arshad, J. Kurth, H. Ouboter, M.S.M. Jetten, C.U. Welte, Unraveling nitrogen, sulfur and carbon metabolic pathways and microbial community transcriptional responses to substrate deprivation and toxicity stresses in a bioreactor mimicking anoxic brackish coastal sediment conditions, *BioRxiv*, (2021). doi 10.1101/2021.08.31.458400

- [23] F. Xie, W. Jin, H. Si, Y. Yuan, Y. Tao, J. Liu, X. Wang, C. Yang, Q. Li, X. Yan, L. Lin, Q. Jiang, L. Zhang, C. Guo, C. Greening, R. Heller, L.L. Guan, P.B. Pope, Z. Tan, W. Zhu, M. Wang, Q. Qiu, Z. Li, S. Mao, An integrated gene catalog and over 10,000 metagenome-assembled genomes from the gastrointestinal microbiome of ruminants, *Microbiome*, 9 (2021) 137. doi 10.1186/s40168-021-01078-x
- [24] C.M. Singleton, F. Petriglieri, J.M. Kristensen, R.H. Kirkegaard, T.Y. Michaelsen, M.H. Andersen, Z. Kondrotaitė, S.M. Karst, M.S. Dueholm, P.H. Nielsen, M. Albertsen, Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing, *Nat. Commun.*, 12 (2021) 2009. doi 10.1038/s41467-021-22203-2
- [25] Y. Zhang, M. Kitajima, A.J. Whittle, W.T. Liu, Benefits of genomic insights and CRISPR-Cas signatures to monitor potential pathogens across drinking water production and distribution systems, *Front. Microbiol.*, 8 (2017) 2036. doi 10.3389/fmicb.2017.02036
- [26] S. Vosloo, L. Huo, C.L. Anderson, Z. Dai, M. Sevilano, A. Pinto, Evaluating de novo assembly and binning strategies for time series drinking water metagenomes, *Microbiology Spectrum*, 9 (2021) e0143421. doi 10.1128/Spectrum.01434-21
- [27] C. Suarez, C.J. Sedlacek, D.J.I. Gustavsson, A. Eiler, O. Modin, M. Hermansson, F. Persson, Disturbance-based management of ecosystem services and disservices in partial nitrification anammox biofilms, *BioRxiv*, (2021). doi 10.1101/2021.07.05.451122
- [28] R.S. Kantor, A.W. van Zyl, R.P. van Hille, B.C. Thomas, S.T. Harrison, J.F. Banfield, Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics, *Environ. Microbiol.*, 17 (2015) 4929-4941. doi 10.1111/1462-2920.12936
- [29] L. Poghosyan, H. Koch, J. Frank, M.A.H.J. van Kessel, G. Cremers, T. van Alen, M.S.M. Jetten, H.J.M. Op den Camp, S. Lücker, Metagenomic profiling of ammonia- and methane-oxidizing microorganisms in a Dutch drinking water treatment plant, (2020). doi 10.1101/2020.05.19.103440

---

## In the Alphaproteobacterium *Hyphomicrobium denitrificans* SoxR Serves a Sulfane Sulfur-Responsive Repressor of Sulfur Oxidation

Li, J., Törkel, K., Koch, J., Tanabe, T. S., Hsu, H. Y. & Dahl, C.

---

Many sulfur-oxidizers utilize thiosulfate as additional or alternative electron source during chemolithotrophic or photolithotrophic growth. Thiosulfate is either converted to tetrathionate by the thiosulfate dehydrogenase or to oxidized to sulfate through the periplasmic Sox system. For the complete conversion of thiosulfate to sulfate SoxXA, SoxYZ, SoxB and SoxCD are required (Friedrich *et al.* 2001). However, several established sulfur-oxidizing chemolithotrophic and photolithotrophic bacteria encode a truncated Sox system that lacks the SoxCD component (Dahl 2020, Li *et al.* 2023b). Thus, the SoxY-bound sulfane sulfur cannot be oxidized to sulfone. These bacteria use the cytoplasmic sHdr pathway or reverse Dsr system to completely oxidize the sulfane sulfur to sulfite or sulfate in the cytoplasm. In *Hyphomicrobium denitrificans* a truncated Sox system is encoded in directly next to the *shdr* gene cluster (Koch & Dahl 2018, Tanabe *et al.* 2023c). This truncated Sox system was shown to be active during thiosulfate oxidation and to be coupled to the cytoplasmic sHdr system (Li *et al.* 2023). Transcription of genes required for sulfur oxidation was previously revealed to be regulated by ArsR-type repressor sHdrR and an additional unknown regulator that remained to be elucidated. Here properties, DNA binding sites and regulated genes of the ArsR-type repressor SoxR are described. This regulator is encoded in the *shdr/sox* gene cluster of *H. denitrificans* and interacts directly or indirectly with the regulator sHdrR to regulate the thiosulfate oxidation in *H. denitrificans* (Li *et al.* 2023b).

SoxR was shown to be of general importance in complete and truncated Sox systems. The occurrence of *soxR* in *sox* gene clusters was analyzed with HMS-S-S and HMSS2 in all representative species of the genome taxonomy database (GTDB release R.207). The Sox system was shown to be present in Pseudomonadota (formerly Proteobacteria), Aquificota and Bacteroidota. The *sox* gene cluster was not present in any sulfur-oxidizing archaeon and also absent in a number of proposed or established sulfur-oxidizers from other bacterial phyla encoding the reverse Dsr system and/or the sHdr pathway. This analysis showed that the SoxR regulator was more often encoded in complete Sox systems than in truncated Sox systems. Furthermore, the regulator SoxR was limited to *sox* gene clusters in the phylum Pseudomonadota (formerly Proteobacteria). Analysis of species encoding regulated and unregulated Sox systems showed that SoxR is often absent in truncated Sox systems of established chemolithoautotrophic sulfur oxidizers such as green and purple sulfur bacteria or Aquificota (Li *et al.* 2023b).

---

Li, J., Törkel, K., Koch, J., Tanabe, T. S., Hsu, H. Y. & Dahl, C. (2023) In the Alphaproteobacterium *Hyphomicrobium denitrificans* SoxR Serves a Sulfane Sulfur-Responsive Repressor of Sulfur Oxidation *Antioxidants*, 12(8), 1620; doi.org/10.3390/antiox12081620

For first genetic evidence on the function SoxR the strain *H. denitrificans*  $\Delta tsdA \Delta soxR$  generated. This strain oxidized thiosulfate with a much higher specific oxidation rate and showed a significantly decreased growth rate compared with the reference strain *H. denitrificans*  $\Delta tsdA$ . The specific thiosulfate oxidation rate was even increased when *H. denitrificans*  $\Delta tsdA \Delta soxR$  was exposed to thiosulfate prior to the measurement. A similar increase was observed for *H. denitrificans*  $\Delta tsdA \Delta sHdrR$ . Likely the induction of the regulator sHdrR or SoxR respectively causes this additional increase (Li *et al.* 2023b).

To elucidate the genes regulated by SoxR, RT-qPCR was performed. The transcript levels of twelve selected genes were compared between the *H. denitrificans*  $\Delta tsdA \Delta soxR$  strain and reference *H. denitrificans*  $\Delta tsdA$  strain during thiosulfate oxidation and in the absence of thiosulfate. The selected genes included several *sox* genes, *shdr* genes, transporters *soxT1A* and *soxT1B*, genes for sulfur transferases and *lbpA*. SoxR most significantly affected the transcription level of *sox* genes and to a lesser extent *shdr*, *lbpA* and *tusA* genes. For the latter, much higher transcription levels were observed in the reference strain during thiosulfate consumption, suggesting the existence of an additional regulator, which was considered to be sHdrR. Transcription of *soxT1B* and *soxR* was not influenced by thiosulfate indicating that SoxT1B is also involved in the sensing of thiosulfate and the regulation of the sHdr pathway and/or Sox system (Li *et al.* 2023b).

Recombinant SoxR was also used to determine the DNA binding sites of SoxR in the *shdr/sox* gene cluster of *H. denitrificans*. A total of four DNA binding sites were identified for SoxR. Recombinant SoxR appeared in a monomeric and dimeric state on size exclusion gel chromatography. The dimerization was found to be dependent on two cysteine residues that also influence the conformational state of SoxR. These cysteines were also important for the sulfur sensing and DNA binding properties of SoxR. As assessed through EMSA and MALPEG assays polysulfide and to a lesser extent thiosulfate modify SoxR. As thiosulfate is oxidized in the periplasm and probably does not reach high concentrations, SoxR was considered to be a sulfane sulfur sensing regulator. In summary the results allowed to propose a regulation mechanism. In this model SoxY-bound sulfane sulfur is transported into the cytoplasm through a SoxT transporter. The sulfur transferase TusA may then channel the sulfur to SoxR. An intramolecular SoxR disulfide is then formed, allowing a conformational change that terminally activates transcription of the *shdr/sox* gene cluster (Li *et al.* 2023b).

T.S.T. contributed to the study by investigation, bioinformatic analysis and writing: T.S.T. performed the analysis of the presence of Sox systems and the co-occurrence of SoxR. T.S.T. created the polysulfide solution for the investigation. T.S.T. contributed to writing the respective sections in the manuscript.





Article

# In the Alphaproteobacterium *Hyphomicrobium denitrificans* SoxR Serves a Sulfane Sulfur-Responsive Repressor of Sulfur Oxidation

Jingjing Li, Kaya Törkel, Julian Koch, Tomohisa Sebastian Tanabe , Hsun Yun Hsu and Christiane Dahl \*

Institut für Mikrobiologie & Biotechnologie, Rheinische Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 168, 53115 Bonn, Germany; s6jglii2@uni-bonn.de (J.L.); ktoerke@uni-bonn.de (K.T.); julikoch95@gmx.de (J.K.); s6totana@uni-bonn.de (T.S.T.); s6hshsuu@uni-bonn.de (H.Y.H.)

\* Correspondence: chdahl@uni-bonn.de

**Abstract:** In organisms that use reduced sulfur compounds as alternative or additional electron donors to organic compounds, transcriptional regulation of genes for enzymes involved in sulfur oxidation is needed to adjust metabolic flux to environmental conditions. However, little is known about the sensing and response to inorganic sulfur compounds such as thiosulfate in sulfur-oxidizing bacteria. In the Alphaproteobacterium *Hyphomicrobium denitrificans*, one strategy is the use of the ArsR–SmtB-type transcriptional regulator SoxR. We show that this homodimeric repressor senses sulfane sulfur and that it is crucial for the expression not only of *sox* genes encoding the components of a truncated periplasmic thiosulfate-oxidizing enzyme system but also of several other sets of genes for enzymes of sulfur oxidation. DNA binding and transcriptional regulatory activity of SoxR are controlled by polysulfide-dependent cysteine modification. The repressor uses the formation of a sulfur bridge between two conserved cysteines as a trigger to bind and release DNA and can also form a vicinal disulfide bond to orchestrate a response to oxidizing conditions. The importance of the sulfur bridge forming cysteines was confirmed by site-directed mutagenesis, mass spectrometry, and gel shift assays. In vivo, SoxR interacts directly or indirectly with a second closely related repressor, sHdrR.

**Keywords:** *Hyphomicrobium denitrificans*; sulfur oxidation; thiosulfate; SoxR; transcriptional regulation; reactive sulfur species; repressor



**Citation:** Li, J.; Törkel, K.; Koch, J.; Tanabe, T.S.; Hsu, H.Y.; Dahl, C. In the Alphaproteobacterium *Hyphomicrobium denitrificans* SoxR Serves a Sulfane Sulfur-Responsive Repressor of Sulfur Oxidation.

*Antioxidants* **2023**, *12*, 1620. <https://doi.org/10.3390/antiox12081620>

Academic Editor: Kenneth R. Olson

Received: 30 June 2023

Revised: 11 August 2023

Accepted: 14 August 2023

Published: 16 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Thiosulfate ( $S_2O_3^{2-}$ ) is a sulfur substrate that is oxidized by the majority of dissimilatory sulfur oxidizers. Its complete oxidation to sulfate is always initiated, and in many cases also completely performed, in the bacterial periplasm and involves the well-studied thiosulfate-oxidizing Sox multienzyme system [1–3] (Figure 1a). Three proteins, SoxYZ, SoxXA, and SoxB, are required for the initial steps. The *c*-type cytochrome SoxXA catalyzes the oxidative formation of a disulfide linkage between the sulfane sulfur of thiosulfate and the persulfurated active site cysteine residue of SoxY [4]. Then, SoxB catalyzes the hydrolytic release of the sulfone group as sulfate, leaving the original sulfane sulfur of thiosulfate bound to SoxY [5,6]. The reaction cycle can be fully completed in the periplasm of organisms containing the hemomolybdo-protein SoxCD, which catalyzes the oxidation of SoxY-bound sulfane sulfur to sulfone, followed again by SoxB-catalyzed hydrolytic release of sulfate [7].

Many sulfur oxidizers do not contain SoxCD and have a so-called “truncated” Sox system (Figure 1b) [2]. For complete oxidation to sulfate, truncated Sox systems can be combined with cytoplasmic sulfur oxidation systems. How the sulfur is transferred into the cytoplasm for further oxidation is still a mystery. The Alphaproteobacterium *Hyphomicrobium denitrificans* X<sup>T</sup> (DSM 1869<sup>T</sup>) is a representative of this group [8] (Figure 1b).

In this organism, two genes encoding predicted sulfur compound transporters (SoxT1A and SoxT1B) are located in close proximity to the *sox* genes and the genes for the cytoplasmic sulfane sulfur-oxidizing heterodisulfide reductase-like (sHdr) system (Figure 1b). While the *H. denitrificans* Sox and sHdr proteins have been shown experimentally to be essential for thiosulfate oxidation [8–10], evidence for the proposed sulfur transport has not been provided so far.

The obligately heterotrophic *H. denitrificans* oxidizes thiosulfate as an additional electron donor during growth on compounds like methanol [9]. In batch culture, substantial amounts of sulfite are excreted as the product of sHdr-catalyzed sulfur oxidation and accumulate because an enzyme catalyzing efficient sulfite oxidation is not present [9]. Accumulation of sulfite as an intermediate has also been described for some facultatively autotrophic sulfur oxidizing Alphaproteobacteria, e.g., *Rhodovulum* (previously *Rhodobacter sulfidophilum*) [16].

(Bi)sulfite ( $\text{HSO}_3^-$ ),  $\text{SO}_3^{2-}$  is a highly reactive, strong nucleophile and has many toxic effects. Its strong reducing capacity ( $E_0'$  for the sulfate/sulfite couple is  $-515$  mV) contributes to its toxicity and antimicrobial action, which have led to its widespread use as a food preservative [17,18]. Free sulfite can damage DNA through the formation of adducts [19–21]. Its toxic effect on mammalian cells has been attributed to the formation of sulfur- and oxygen-based free radicals [22,23] which can in turn react with lipids and proteins [24,25]. The full Sox pathway or the truncated Sox/sHdr combination may be advantageous, despite the intermediate release of sulfite, for organisms such as *H. denitrificans* or *R. sulfidophilum* at low thiosulfate concentrations if removal by other members of the community or chemical oxidation in oxygenated environments keeps sulfite concentrations below inhibitory levels. In any case, the formation of the toxic intermediate sulfite during the oxidation of sulfur compounds, as well as the switching between organic and inorganic electron donors, requires fine-tuning to the environmental conditions.

Accordingly, complex regulatory patterns have been reported for facultative sulfur oxidizers, with upregulation usually occurring only in the presence of metabolizable sulfur substrates, whereas the corresponding genes are thought to always be highly expressed in chemolithoautotrophs restricted to the oxidation of sulfur compounds. In *H. denitrificans* and other Alphaproteobacteria that are not restricted to sulfur oxidation, such as *R. sulfidophilum*, *Paracoccus pantotrophus* or *Pseudaminobacter salicylatoxidans*, the ability to oxidize thiosulfate and, depending on the organism, other reduced inorganic and organic sulfur compounds such as sulfide or dimethyl sulfide, is not constitutive but can be induced by the presence of oxidizable sulfur compounds [9,16,26,27]. While the transcriptional repressor sHdrR is involved in this process in *H. denitrificans* [9], genetic and biochemical studies have identified the related SoxR protein as a major regulator in *P. pantotrophus* and *P. salicylatoxidans* [26–28], both of which contain a complete Sox system and are unable to oxidize sulfane sulfur in the cytoplasm.

SoxR is a member of the arsenic repressor (ArsR–SmtB) family of prokaryotic repressors [29–32]. Members of the ArsR–SmtB family were originally recognized as metal-responsive transcriptional regulators, but there are also members in this family that have been shown to sense reactive oxygen or sulfur species [33]. SqrR from *Rhodobacter capsulatus* and BigR from *Xylella fastidiosa* belong to this group and control the transcription of genes involved in sulfide-dependent photosynthesis and the detoxification of  $\text{H}_2\text{S}$  derived from associated host plants, respectively [34–36]. Knowledge about SoxR is comparatively sparse. While binding regions for the transcriptional repressor have been identified in promoter–operator segments within the *sox* gene clusters of *P. denitrificans* and *P. salicylatoxidans* [26,27], no information is available on factors that control its DNA-binding capacity. It is therefore completely unclear how SoxR senses the presence of oxidizable sulfur compounds and how it then triggers the transcription of sulfur oxidation genes.



Here, we start to close this knowledge gap by first providing information on the general distribution of complete and truncated Sox systems and their co-occurrence with SoxR. Furthermore, we present genetic information for SoxR function in *H. denitrificans*, identify target genes and map its binding sites. The DNA-binding properties of the homodimeric repressor and its response to bridging of the sulfur atoms of two conserved cysteine residues by one to three sulfur atoms are characterized via site-directed mutagenesis, mass spectrometry, MalPEG assays, and electrophoretic mobility shift assays (EMSA).

## 2. Materials and Methods

### 2.1. Bacterial Strains, Plasmids, Primers, and Growth Conditions

Table S1 lists the bacterial strains, primers and plasmids that were used for this study. *Escherichia coli* strains were grown on complex lysogeny broth (LB) medium [37] under aerobic conditions at 37 °C unless otherwise indicated. *Escherichia coli* BL21 (DE3) was used for recombinant protein production. *E. coli* strains 10 beta and DH5 $\alpha$  were used for molecular cloning. *H. denitrificans* strains were cultivated in minimal media containing 24.4 mM methanol kept at pH 7.2 with 100 mM 3-(*N*-Morpholino)propanesulfonic acid (MOPS) buffer as described before [8]. Thiosulfate was added as needed. Antibiotics for *E. coli* and *H. denitrificans* were used at the following concentrations (in  $\mu\text{g mL}^{-1}$ ): ampicillin (Ap), 100; kanamycin (Km), 50; streptomycin (Sm), 200; tetracycline (Tc), 15; and chloramphenicol (Cm), 25.

### 2.2. Recombinant DNA Techniques

DNA manipulation and cloning were performed using standard techniques, unless otherwise indicated [38]. Restriction enzymes, T4 ligase and Q5 polymerase were purchased from New England Biolabs (Frankfurt, Germany) and used according to the manufacturer's instructions. Oligonucleotides for cloning were obtained from Eurofins MWG (Ebersberg, Germany). The GenJET Plasmid Miniprep kit (Thermo Scientific, Waltham, MA, USA) and the First-DNA all-tissue Kit (GEN-IAL GmbH, Troisdorf, Germany) were used for the purification of plasmid DNA from *E. coli* and chromosomal DNA from *H. denitrificans* strains, respectively.

### 2.3. Construction of Plasmid for Deletion of *soxR* in *H. denitrificans*

For markerless deletion of the *H. denitrificans soxR* (Hden\_0700) gene by splicing overlap extension (SOE) [39], PCR fragments were constructed using the primers P1 fwd up hden\_0700, P2 rev up hden\_0700, P3 fwd down hden\_0700 and P4 rev down hden\_0700 (Table S1). The resulting 1.04 kb SOE PCR fragment was cloned into the XbaI and PstI sites of pK18*mobsacB*-Tc [9]. The final construct, pK18*mobsacB*-Tc- $\Delta$ *soxR*, was electroporated into *H. denitrificans*  $\Delta$ *tsdA* and transformants were selected using previously published procedures [8,10]. Single crossover recombinants were Cm<sup>r</sup> and Tc<sup>r</sup>. Double crossover recombinants were Tc<sup>s</sup> and survived in the presence of sucrose due to the loss of the vector-encoded tetracyclin resistance and levansucrase (SacB) genes.

### 2.4. Characterization of Phenotypes, Quantification of Sulfur Compounds and Protein Content

Growth experiments with *H. denitrificans* were run in 200 mL medium with 24.4 mM methanol and varying concentrations of thiosulfate in 500 mL Erlenmeyer flasks, as described in [9]. Thiosulfate concentrations, protein content, and specific thiosulfate oxidation rates were determined by previously described methods [9,40]. All growth experiments were repeated three times. Representative experiments with two biological replicates for each strain are shown. All quantifications are based on at least three technical replicates.

### 2.5. RNA Preparation

Total RNA of *H. denitrificans* was isolated from cells harvested in mid-log phase. *H. denitrificans* strains  $\Delta$ *tsdA* and  $\Delta$ *tsdA*  $\Delta$ *soxR* were grown in 50 mL methanol-containing medium at 30 °C with shaking at 250 rpm in 100 mL Erlenmeyer flasks. Cells from 2 mL

were harvested by centrifugation at  $16,000 \times g$  for 5 min. The cell pellet was incubated with 500  $\mu\text{L}$  of 10% SDS (sodium dodecyl sulfate) containing  $1 \text{ mg mL}^{-1}$  lysozyme at room temperature for 5 min. Then 700  $\mu\text{L}$  of TRIzol [41] was added and the mixture was incubated for another 5 min. This step was followed by the addition of 1 mL ROTI<sup>®</sup>Aqua-P/C/I reagent (Carl Roth GmbH, Karlsruhe, Germany), 10 min of incubation and centrifugation at  $13,000 \times g$  for 5 min. RNA purification from the supernatant was achieved with the Monarch Total RNA Miniprep Kit (New England Biolabs, Frankfurt, Germany). gDNA was removed by treating 10  $\mu\text{L}$  samples with an absorption at 260 nm corresponding to  $\sim 1 \mu\text{g}$  RNA with 1 U of RNase-free DNase I (ThermoFisher, Waltham, MA, USA) in the  $\text{MgCl}_2$ -containing reaction buffer provided by the manufacturer. RNA concentrations were measured with an NanoDrop Biospectrometer (Eppendorf, Hamburg, Germany). The absence of gDNA was verified using the primers rpoB-denitf and rpoB-denitr [42], which bind only to gDNA and not to the corresponding RNA.

#### 2.6. Expression Studies Based on RT-qPCR

RNA samples of 100 ng were used for RT-qPCR analysis via the Luna Universal One-Step RT-qPCR Kit (New England Biolabs, Frankfurt, Germany) and the CFX Connect<sup>™</sup> real-time detection system (Bio-Rad, Munich, Germany) according to the instructions of the manufacturers. The level of *rpoB* mRNA was used as an internal standard [42]. Approximately 200 bp fragments were amplified (see Table S1 in the supplemental material) with an annealing temperature of 60 °C. The RT-qPCR conditions were as follows: 10 min at 55 °C (reverse transcription using random nonamer primers), 1 min at 95 °C (inactivation of the reverse transcriptase and activation of the polymerase), 40 cycles of 15 s at 95 °C, 30 s at 60 °C, followed by melting curve analysis, in which the temperature was increased every 10 s by 1 °C, from a start at 60 °C to 95 °C. Analyses of melting curves and calculations of  $C_t$  (calculated threshold) values were automatically quantified with the Bio-Rad CFX Manager 3.1 (3.1.1517.0823) software.  $C_t$  values for each point in time were run in triplicate. Relative expression ratios were calculated by the  $2^{-\Delta\Delta C_t}$  method [43].

#### 2.7. Cloning, Site-Directed Mutagenesis, Overproduction, and Purification of Recombinant SoxR Proteins

The *soxR* gene was amplified from *H. denitrificans* genomic DNA with primers adding a sequence for an N-terminal Strep-tag and cloned between the NdeI and HindIII sites of pET-22b (+), resulting in pET22b-SoxR-Strep. Cysteine-to-serine exchanges were implemented with the Q5 Site-Directed Mutagenesis Kit (New England Biolabs, Frankfurt, Germany) according to the manufacturer's instructions and using the primers listed in Table S1. Recombinant SoxR proteins were overproduced in *E. coli* BL21 (DE3) containing plasmids pET22b-SoxR-Strep, pET-22b-SoxR C<sup>50S</sup>, pET-22b-SoxR C<sup>116S</sup>, and pET-22b-SoxR C<sup>50S</sup> C<sup>116S</sup>. The cells were grown in 1 L Erlenmeyer flasks at 37 °C in 400 mL LB medium containing ampicillin up to an OD<sub>600</sub> of 0.5–0.6. Expression of *soxR* was induced by adding 0.5 mM IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside). IPTG-induced *E. coli* cells were grown overnight at 20 °C. Cells were harvested at  $14,000 \times g$  for 30 min. Three mL of lysis buffer (100 mM Tris-HCl buffer pH 7.0 and 5 mM EDTA (ethylenediaminetetraacetic acid containing a spatula tip of deoxyribonuclease I and protease inhibitor) were added per g of wet weight for homogenization. Cell lysis was achieved by sonification and followed by centrifugation ( $16,100 \times g$ , 30 min, and 4 °C) and ultracentrifugation ( $145,000 \times g$ , 1 h, 4 °C). The supernatant was applied to a Strep-tactin affinity chromatography column equilibrated with buffer W (100 mM Tris-HCl, pH 8.0, 150 mM NaCl). The column was washed with six volumes of buffer W and eluted with buffer E (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2.5 mM D-desthiobiotin). The protein was assessed for its purity by 12.5% SDS-PAGE (polyacrylamide gel electrophoresis). Pure SoxR proteins were stored on ice in buffer W. Buffer exchange was achieved with Amicon<sup>®</sup> Ultra-3K centrifugal filters (Merck Millipore, Darmstadt, Germany).

### 2.8. Electrophoretic Mobility Shift Assays (EMSA)

Interactions between proteins and nucleic acids are detected by gel electrophoretic mobility shift assays. In these, solutions of nucleic acid and protein are combined and analyzed for the distribution of nucleic acid species by native polyacrylamide gel electrophoresis. In general, the migration of protein–nucleic acid complexes is slower than that of the corresponding free nucleic acid. The binding reaction mixture (15  $\mu$ L final volume) contained purified SoxR wild-type or variant protein in various concentrations (up to 700 nM), 2  $\mu$ L 50% glycerol and 1.5  $\mu$ L  $10 \times$  binding buffer (100 mM Tris-HCl, 500 mM KCl, 10 mM DTT, 5% glycerol, pH 8.0). Reaction mixtures were pre-incubated for 20 min at room temperature, followed by a further 30 min incubation at 30 °C after adding the DNA probe to a final concentration of 17 nM. The DNA probes consisted of a 362-bp fragment covering the entire intergenic region between the *shdrR* (Hden\_0682) and the *soxT1A* (Hden\_0681) genes, a 180-bp fragment representing the central part of the first product (created with primers EMSA-Fr2-Fr and EMSA\_Fr3-Rev), a 177-bp fragment situated between the *shdrR* and the *lipS1* (Hden\_0683) genes, a 173-bp fragment situated between the *lipX* (Hden\_0687) and *dsrE3C* (Hden\_0688) genes, a 176-bp fragment located between the *tusA* (Hden\_0698) and *hyp* (Hden\_0697) genes, and a 151-bp fragment situated between the *soxA* (Hden\_0703) and *soxY* (Hden\_0704) genes. All primers used are listed in Table S1. Native 6% polyacrylamide gels were loaded with the reaction mixtures after pre-running the gels at 100 V for 1 h at 4 °C with  $0.25 \times$  TBE buffer (25 mM Tris-borate, 0.5 mM EDTA).  $0.25 \times$  TBE with 0.5% glycerol was used as the buffer for running the loaded gels for 1 h at 180 V and 4 °C. Gels were subsequently stained for 20 min with SYBR green I. The bands corresponding to SoxR-bound and free DNAs were visualized with a ChemiDoc Imaging System (BioRad, Munich, Germany).

### 2.9. Gel Permeation Chromatography

The size exclusion chromatography column Superdex™ 75 Increase 10/300 GL (Cytiva, Freiburg, Germany) was calibrated using Blue dextran (2000 kDa), conalbumin (75 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), lactoglobulin (35 kDa), carbonic anhydrase (29 kDa), chymotrypsin (23 kDa), and ribonuclease (13.7 kDa). The calibration curve was plotted using the gel-phase distribution coefficient ( $K_{av}$ ) versus the logarithm of molecular weight.  $K_{av} = (V_e - V_0) / (V_c - V_0)$ , where  $V_e$  = elution volume,  $V_0$  = column void volume (7.94 mL based on Blue dextran elution volume) and  $V_c$  geometric column volume (24 mL). The column was run in 50 mM Tris-HCl, pH 8.0 and 150 mM NaCl at a flow rate of 0.8 mL min<sup>-1</sup> using an Äkta FPLC system.

### 2.10. Preparation of Polysulfides

A polysulfide stock solution was prepared according to Ikeda et al. [44] by mixing 1.2 g NaHS  $\times$  H<sub>2</sub>O and 0.16 g sulfur powder with 3 mL oxygen-free water in a closed 10 mL serum bottle under a nitrogen atmosphere for 1 h at room temperature. Then, the volume was filled up to 10 mL with oxygen-free water. Based on the average length of the resulting polysulfides of four sulfur atoms, their concentration is 0.5 M in the final solution, which can be kept at room temperature for many months. If necessary, the polysulfide solution was diluted with oxygen-free water and immediately used for persulfuration reactions.

### 2.11. Redox Treatments, Persulfuration Reactions, MalPEG Gel-Shift Assays and Mass Spectrometry

A total of 5  $\mu$ g protein was treated with dithiothreitol (DTT, 1 mM and 5 mM for samples analyzed by mass spectrometry and EMSA, respectively) for reduction, 5 mM CuCl<sub>2</sub> for oxidation, 0.5 mM polysulfide for persulfuration, 1 mM MalPEG (methoxy-polyethylene glycol maleimide, MW 10,000 g mol<sup>-1</sup>) for PEGylation, or 5 mM iodoacetamide for carbamidomethylation in a final volume of 15  $\mu$ L containing 100 mM Tris-HCl, pH 8.0 and 150 mM NaCl. When polysulfide, MalPEG, and DTT were applied consecutively, concentrations were 0.5 mM, 5 mM, and 1 mM, respectively. When polysulfide and DTT

were applied consecutively, concentrations were 0.5 mM and 10 mM, respectively. Protein samples used in EMSA experiments were reacted with the reagents for 20 min at 25 °C. Samples analyzed by SDS–PAGE were incubated with each reagent for 15 min at 30 °C. Reactions were either stopped by the addition of 5 µL of 4 × non-reducing Roti®-Load2 (Carl Roth GmbH, Karlsruhe, Germany) and subjected to 15% SDS–PAGE without boiling the sample or analyzed by mass spectrometry. Samples of 20 µL were desalted by ZiptipC4 Pipette tips (Merck Millipore, Darmstadt, Germany) and measured by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry at the Core Facility Protein Synthesis & BioAnalytics, Pharmaceutical Institute, University of Bonn.

### 2.12. Distribution of Sox Systems and SoxR: Dataset Generation and Analysis

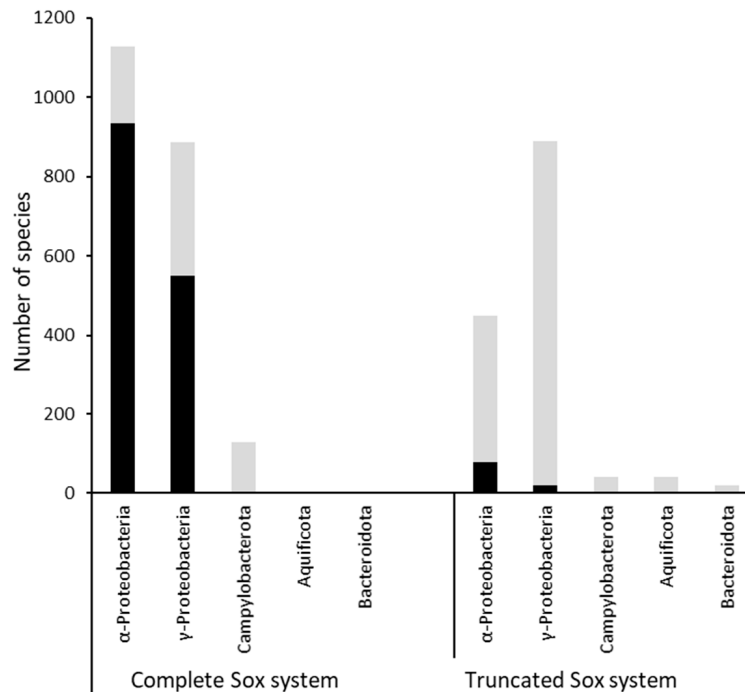
Archaeal and bacterial genomes were downloaded from the Genome Taxonomy Database (GTDB, release R207). In the GTDB, all genomes are pre-validated, sorted according to validly published taxonomies and are of high quality (completeness minus 5 \*contamination must be greater than 50%). One representative of each of the current 65,703 species clusters has been analyzed. The GTDB is based on recently standardized archaeal and bacterial taxonomies derived by normalizing the evolutionary distance between taxonomic levels [45,46]. For bacteria, the database currently lists 148 phyla. For the Archaea, GTDB distinguishes 16 phyla. Open reading frames were determined using Prodigal [47] and subsequently annotated for SoxR, other Sox proteins, and clustering of the respective genes via HMS-S-S with default conditions [48]. Chromatiaceae and Ectothiorhodospiraceae were treated as exceptions as they do not contain contiguous *sox* clusters, but the thiosulfate-oxidizing capabilities and functionality of the Sox proteins have been experimentally established for relevant species [49].

## 3. Results

### 3.1. Distribution of Sox Systems and the SoxR Regulator

We first asked how complete and truncated Sox systems (Figure 1) are distributed among the prokaryotes and analyzed the genomes available in the Genome Taxonomy Database (GTDB, release R207). In GTDB, all genomes are sorted according to validly published taxonomies. In addition, we asked which groups of these prokaryotes contain a *soxR* that is linked to the other *sox* genes. In order to accurately identify and discriminate between the Sox components, we used HMS-S-S, a tool that specifically finds sulfur metabolism-related proteins [48]. As shown in Figure 2 and Table S2, genes encoding Sox proteins are not found among the Archaea. They exclusively occur in 17 of the currently 169 bacterial phyla distinguished in the GTDB. The highest proportion of species with Sox in a phylum is observed for the Aquificota (54%), followed by the Campylobacterota (30.7%), the Deinococcota (24.3%), and the Proteobacteria (19.3%) (Table S2). The SoxR regulator is strictly confined to the Proteobacteria (Figure 2).

The Aquificota contain exclusively organisms with a truncated Sox system (Table S2), which are strictly chemolithoautotrophic sulfur oxidizers, with a few having additional organoheterotrophic capacity [50]. Among the Sox-containing Campylobacterota, about three quarters rely on a complete system. The type of Sox system varies within a family and even within a single genus. Many Campylobacterota species, e.g., members of the families Sulfurimonadaceae, Sulfurispirillaceae or Sulfurovovaceae, are established chemolithoautotrophic sulfur oxidizers [51–53]. In the Deinococcota, the complete Sox system is much more abundant than the truncated version (Table S2), with occurrences in *Thermus* and *Meiothermus* species known as sulfur-oxidizing mixotrophs [54,55]. Among the Bacteroidota, the general abundance of Sox is low, but here we find the obligately photolithoautotrophic sulfur oxidizers of the order Chlorobiales [56], all of which encode the truncated set of Sox proteins.



**Figure 2.** Occurrence of complete and truncated Sox systems among five bacterial phyla. The simultaneous presence of SoxR is indicated in black.

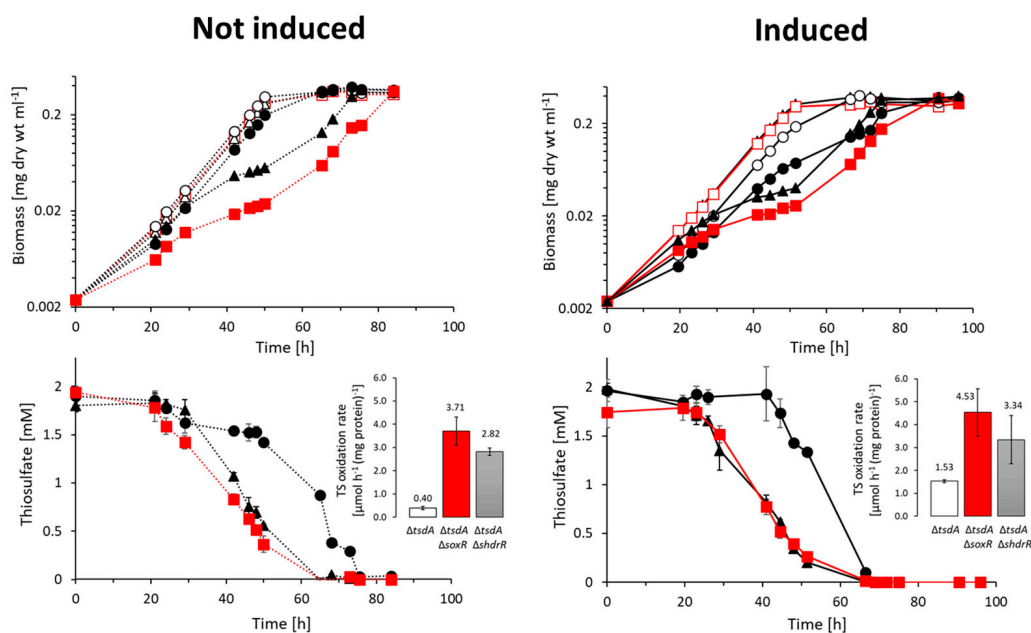
By far the highest absolute numbers of Sox-containing species are found among the Proteobacteria, here exclusively in the classes Alphaproteobacteria and Gammaproteobacteria. The complete Sox system appears more frequently than the truncated version in metabolically versatile members of the alphaproteobacterial families Rhizobiaceae [57] and Rhodobacteraceae [58], while the opposite is true for a number of gammaproteobacterial families, e.g., the Thioglobaceae, Chromatiaceae, Ectothiorhodospiraceae, Thiomicrospiraceae, and Thiotrichaceae (Table S2), all of which contain members with established chemo- or photolithotrophic sulfur-oxidizing capabilities [59–63]. On the other hand, families like the alphaproteobacterial Xanthobacteraceae or the gammaproteobacterial Burkholderiaceae contain species encoding complete or truncated Sox systems in almost equal numbers. The important general rule to emerge from our analysis is that the gene for the SoxR transcriptional regulator is more often linked to the genes for the complete Sox System than to those for the truncated Sox system (Figure 2).

### 3.2. Genetic Evidence for SoxR Function in *H. denitrificans*

Previously, we showed that the ArsR-type regulator encoded by the first gene of the *H. denitrificans* *shdr-lbpA* operon, sHDrR, functions as a repressor of *shdr* gene expression in the absence of oxidizable sulfur compounds [9]. The phenotypic characterization of a mutant strain lacking the *shdrR* gene indicated an additional regulator involved in the overall process. Indeed, a further candidate transcriptional repressor, SoxR, is encoded downstream of *soxXA* in *H. denitrificans* (Figure 1b). To assign a function for SoxR in transcriptional regulation of the hypomicrobial *sox* and possibly also the *shdr* and associated genes, we constructed *H. denitrificans*  $\Delta$ *tsdA*  $\Delta$ *soxR*, a mutant strain with a markerless deletion of *soxR* in a  $\Delta$ *tsdA* background. The reference strain *H. denitrificans*  $\Delta$ *tsdA* lacks thiosulfate dehydrogenase and oxidizes thiosulfate exclusively via the pathway combining Sox and sHDr–LbpA [8,9] (Figure 1b). When grown in the presence of methanol as a carbon



source and thiosulfate as an additional electron source, the  $\Delta tsdA$  reference strain excretes sulfite, which causes a growth retardation that is particularly impressive when cultures are inoculated with thiosulfate-induced cells ([9], also compare open and filled circles in the upper right panel of Figure 3). Like the *H. denitrificans* strain lacking the *shdrR* gene, the *soxR*-deficient strain exhibited a high specific thiosulfate oxidation rate and a significantly reduced growth rate even without the induction of pre-cultures (Figure 3). The growth rate increased significantly when thiosulfate was depleted. When pre-cultures were exposed to thiosulfate, both regulator-negative strains exhibited slightly higher specific thiosulfate consumption rates than in the non-induced case, fully in line with the finding that a second regulator is involved in the overall process.

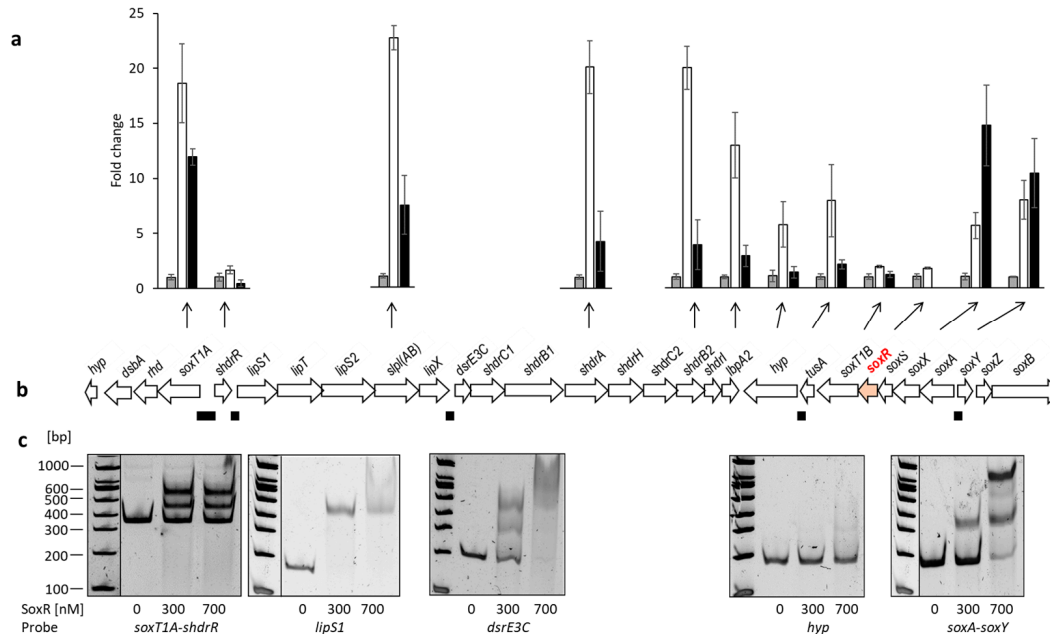


**Figure 3.** Growth and thiosulfate consumption of *H. denitrificans*  $\Delta tsdA$  (black circles and lines),  $\Delta tsdA \Delta shdrR$  (black triangles and lines), and  $\Delta tsdA \Delta soxR$  (red boxes and lines). All strains were grown in medium containing 24.4 mM methanol, either in the absence (open symbols) or in the presence of 2 mM thiosulfate (filled symbols). Pre-cultures contained either no thiosulfate (not induced, broken lines) or 2 mM thiosulfate (induced, solid lines). Thiosulfate concentrations for the different cultures are depicted in the lower panels. Symbol assignments and the color code for specific thiosulfate (TS) oxidation rates are the same as in the upper panels. Error bars indicating SD are too small to be visible for the determination of biomass.

### 3.3. Identification of Genes Controlled by SoxR by RT-qPCR for Different *H. denitrificans* Strains

To examine which genes are affected by the SoxR regulator protein, RT-qPCR experiments were performed, and the transcription levels of twelve genes in the *H. denitrificans* sulfur oxidation genome region were compared in the absence and presence of thiosulfate for the  $\Delta tsdA$  reference strain (Figure 4). In addition, transcription levels were determined for the same genes in the *H. denitrificans*  $\Delta tsdA \Delta soxR$  mutant in the absence of thiosulfate. All cultures were harvested in the exponential growth phase. The studied genes included *soxT1A*, the first of a set of genes transcribed in the opposite direction of *shdrR*, the gene for the sHdrR regulator, and two of the genes encoding proteins involved in LbpA2 assembly (*lipS1* and *slp1(AB)*). LbpA2 is a lipoate-binding protein essential for sulfur oxidation [10]. Four genes were chosen as examples for those encoding the *shdr-lbpA2* cytoplasmic sulfane sulfur oxidation system (*dsrE3C*, *shdrA*, *shdrB2*, and *lbpA2*). These genes are followed by

genes transcribed in the opposite direction and encoding part of the Sox system (SoxXA), the SoxR regulator, SoxS, which is a periplasmic thiol–disulfide oxidoreductase, as well as a second potential sulfur transporter, SoxT1B, the cytoplasmic sulfurtransferase TusA, and a predicted cytochrome P450 (Figures 1 and 4b). Except for *soxS*, all of these genes were included in the RT-qPCR analysis. Finally, the analysis was extended to *soxY* and *soxB*. These genes follow the previously described genes in the opposite direction in a *soxYZB* arrangement (Figures 1 and 4b).



**Figure 4.** (a) Relative mRNA levels of twelve genes located in the *shdr-sox* genetic region (depicted in panel; (b) from *H. denitrificans* for the  $\Delta$ *tsdA* reference strain in the absence (gray columns) and presence of thiosulfate (white columns), as assessed by RT-qPCR. Results for *H. denitrificans*  $\Delta$ *soxR* are shown by black columns. Results were adjusted using *H. denitrificans* *rpoB*, which encodes the  $\beta$ -subunit of RNA polymerase, as an endogenous reference, according to [42]. (b) DNA regions tested in EMSA assays for SoxR binding are indicated as black rectangles below the hypomicrobial *shdr-sox* genes. The *soxR* gene is highlighted in red for easier orientation. Fragment sizes: 362 bp for the *soxT1A-shdrR* intergenic region, 177 bp and 173 bp for the regions upstream of *lipS1* and *dsrE3C*, respectively. The fragments downstream of *tusA* and between *soxA* and *soxY* had sizes of 176 bp and 151 bp, respectively; (c) EMSA analysis of Strep-tagged SoxR with upstream promoter sequence probes of sulfur oxidation-related genes as specified in (b). DNA probes of 17 nM were incubated with different amounts of SoxR (300 and 700 nM). Vertical lines separate samples that were run on the same gel but were not directly adjacent.

With the exception of the genes for the two transcriptional regulators, *shdrR* and *soxR*, and *soxT1B*, which is located just downstream of *soxR*, all the genes tested were upregulated at least five-fold when the reference strain was exposed to thiosulfate, with the strongest responses for *lpl(AB)*, *shdrA*, and *shdrB2* (Figure 4a). In the strain lacking SoxR, transcription of various *sox* and *shdr* genes was much higher than in the reference strain, even in the absence of thiosulfate. The lack of *soxR* most strongly affected transcription of *soxA* and *soxY* but was also evident for *shdr* genes, *soxT1A*, *lipS1*, and *lpl(AB)* (Figure 4a). With the exception of the genuine *sox* genes tested, the reference strain showed a stronger response to the presence of thiosulfate than the  $\Delta$ *tsdA*  $\Delta$ *soxR* mutant in its absence. This observation

clearly points to the presence of at least one further regulatory element, most probably sHdrR [9]. On the other hand, the strong response of numerous genes, in addition to those for the genuine Sox system, shows that their transcription is either directly or indirectly affected by SoxR.

### 3.4. Identification of SoxR Target Sites by EMSA

The finding that SoxR affects transcription of genes outside the *sox* operons was unexpected and afforded a closer analysis. To that end, we inspected intergenic regions within the hyphomicrobial sulfur oxidation region and identified four with conspicuous inverted and direct repeats with the potential to serve as repressor binding sites and used them as probes for EMSA (Figure 4b). A 176-bp fragment located upstream of the hypothetical gene Hden\_0697 served as a control (Figure 4b). Indeed, SoxR bound to four of the five tested probes (Figure 4c). Among these is the intergenic region between *soxT1A* and the gene for the SoxR-related repressor sHdrR. This region had already been shown to serve as a binding site for sHdrR [9], further emphasizing the notion that the two repressors work intimately together.

### 3.5. Properties of the SoxR protein

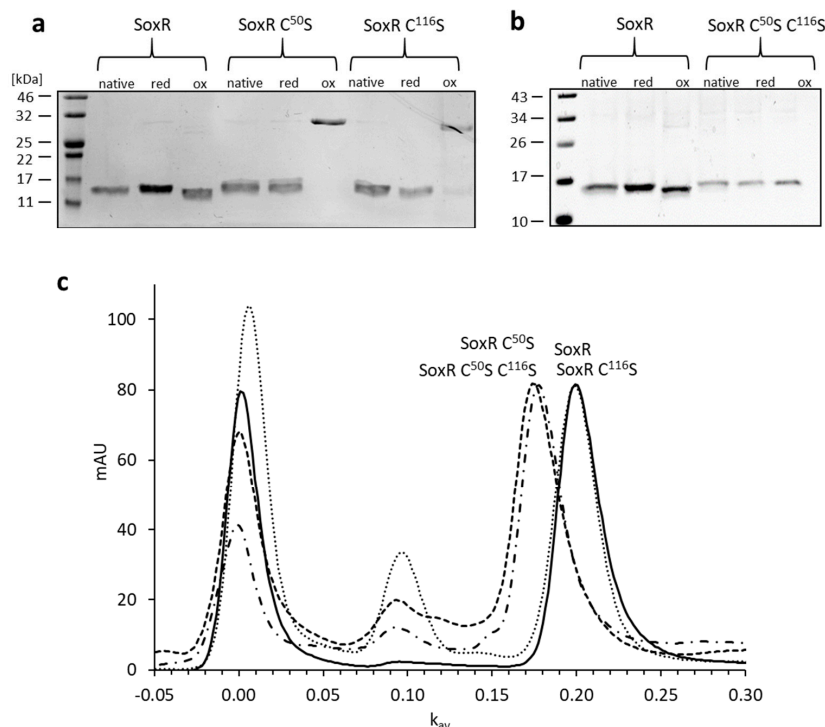
The *H. denitrificans* SoxR protein has a length of 124 amino acids, and a BlastP search (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>, accessed on 29 June 2023) identified *R. capsulatus* SqrR as the most similar structurally characterized protein. *H. denitrificans* SoxR shows 53%, 52%, 43%, and 42% amino acid identity to *R. capsulatus* SqrR, *P. salicylatoxidans* SoxR, *Xylella fastidiosa* BigR, and *H. denitrificans* sHdrR, respectively. All of these regulators share two conserved cysteine residues, Cys<sup>50</sup> and Cys<sup>116</sup>, in the hyphomicrobial protein (Figure 5). The equivalent residues in SqrR are required for sensing sulfide [35,64].

<i>H. denitrificans</i> SoxR	MSGILPNEVIAALEADEEVSPPELKRVLRLARKASDFLKALAHESRLILCLLAEKE-RSA	59
<i>H. denitrificans</i> sHdrR	MAVVKPRTNRPVAVKARTRQPALHSTDAIEQATALLRALGSPHRLAILCLLEGE-RTV	59
<i>Pseudaminobacter</i> SoxR	-----MNLPKLSADQS-PEEFKLLQARKASDLLKALSHEGRLLILCLLAEGE-KSV	51
<i>Xylella fastidiosa</i> BigR	-----MVNEMRDDTRPHMTREDMEKRANVANLLKTLSSHVRLMLVCTLVEGE-FSV	51
<i>Vibrio cholerae</i> HluY	-----MPYL---KGAPMNLQEMEKNSAKAVVLLKAMANERRLQILMLLDNE-LSV	47
<i>Rhodobacter capsulatus</i> SqrR	-----MGSDDERCAALDAEEMATRARAASNLKALAEGRIMMICYLASGE-KSV	50
<i>Escherichia coli</i> YgaV	-----MTELAQLQASAEQAALLKAMSHPKRLLILMLSGSPGTA	41
	. : * : : : . * * : * * :	..
<i>H. denitrificans</i> SoxR	GELENLLSINQPTVSSQQLARLRDLGLVQARREGKAVIYSLPDETRRFIGAIYDKFQREE	119
<i>H. denitrificans</i> sHdrR	SEICDKIGARQSLVSQLHLRRLDLGLVKS DRNGYFVSYSLTSAPAQEI IATLHKYCATS	119
<i>Pseudaminobacter</i> SoxR	SELESIMHMPQAAVSSQQLARLRDFDLVNTREGRVYIYSSIASSEVSSVISTLYGLFCAPV	111
<i>Xylella fastidiosa</i> BigR	GELEQQIGIGQPTLSQQLGVLRESGIVETRRNIKQIFRYRLTEAKAAQLVNALYTFICAQE	111
<i>Vibrio cholerae</i> HluY	GELSSRLELSQSALSQHLAWLRDLVNTREKAQTIVFYFLSSTEVKAMIELLHRLYCAQAN	107
<i>Rhodobacter capsulatus</i> SqrR	TELETRLSTRQAAVSSQQLARLRDLGLVQSRREGKTIYYSLSDPRAARVVQTVVEYQFCSGD	110
<i>Escherichia coli</i> YgaV	GELTRITGLSASATSQLARMRDEGLDSQRDAQRILYSIKNEAVNAI IATLKNVYCP--	99
	*: **:* :* . : : : : : : * : . . : : : *	

**Figure 5.** Amino sequence alignment of SoxR homologs. Accession numbers/locus tags and references in the order of appearance: (Hden\_0700 [8], Hden\_0682 [8,9], WP\_010893290 [34], HLYU\_VIBCH [65], WP\_019171658 [27], ADE85198 [35], and b2667 [66]). An \* (asterisk) indicates positions which have a single, fully conserved residue. Conserved cysteine residues are highlighted in yellow. Colons (:) and single dots (.) indicate conserved and semi-conserved amino acids, respectively.

We sought to obtain information about the oligomerization state and conformation of SoxR as well as about the reactivity of the two cysteine residues. To this end, Strep-tagged SoxR as well as variants with serine in place of either one (SoxR Cys<sup>50</sup>Ser and SoxR Cys<sup>116</sup>Ser) or both cysteines (SoxR Cys<sup>50</sup>Ser Cys<sup>116</sup>Ser) were overexpressed in *E. coli*, purified by affinity chromatography, and subjected to reducing and non-reducing SDS-PAGE analysis in the as-isolated state, after reduction with DTT, and after oxidation with CuCl<sub>2</sub>. The same single 15 kDa band was obtained in all cases under reducing conditions (not shown). The band for the oxidized wild-type protein migrated slightly further than those for the as-isolate and reduced proteins under non-reducing conditions (Figure 6a), indicating a more compact structure due to the formation of an intramolecular

disulfide bond between Cys<sup>50</sup> and Cys<sup>116</sup>. The oxidized SoxR Cys<sup>50</sup>Ser and SoxR Cys<sup>116</sup>Ser variants formed intermolecular dimers connected by the remaining cysteine on each of the monomers (Figure 6a). These observations indicated a homodimeric state for the native proteins that allows close contact between the Cys<sup>50</sup> and Cys<sup>116</sup> residues, respectively, of the monomers and thus the formation of disulfide bridges under oxidizing conditions.



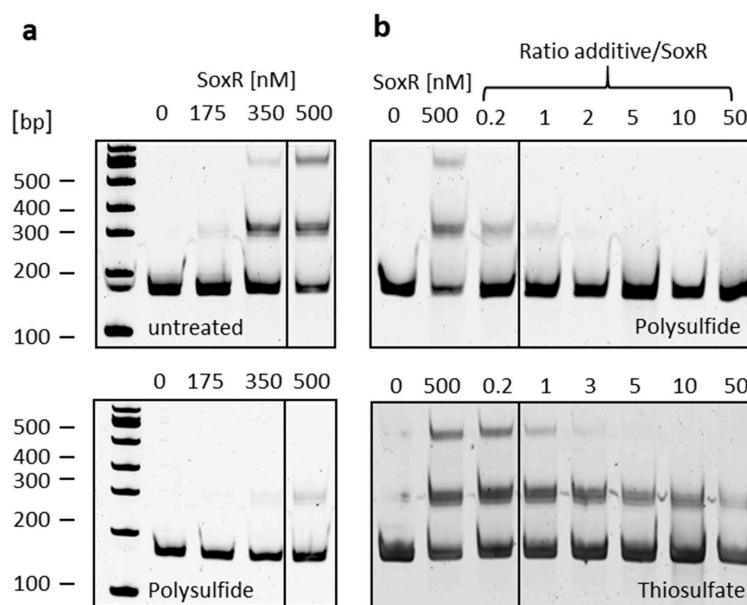
**Figure 6.** Conformation of SoxR and its variants as analyzed by non-reducing SDS–PAGE analysis (a,b) and gel permeation chromatography (c). For the experiments shown in (a) and (b), 5  $\mu$ g of SoxR or its variants were incubated in 15  $\mu$ L of 100 mM Tris–HCl, pH 8.0 and 150 mM NaCl with either 1 mM DTT or 5 mM CuCl<sub>2</sub> for 20 min at room temperature, mixed with 5  $\mu$ L of non-reducing Roti<sup>®</sup>-Load2 (Carl Roth GmbH, Karlsruhe, Germany) and run on 15% SDS polyacrylamide gels. The wild-type SoxR protein is shown twice (panels a,b) to allow direct comparison with protein variants on different gels. In (c), the elution profiles upon gel filtration on Superdex 75 Increase 10/300 are depicted for SoxR, solid line; SoxR Cys<sup>50</sup>Ser, dotted line; SoxR Cys<sup>116</sup>Ser, dashed line; SoxR Cys<sup>50</sup>Ser Cys<sup>116</sup>Ser, dashed-dotted line. SoxR and SoxR Cys<sup>116</sup>Ser dimers elute at a  $k_{av}$  of 0.2, corresponding to a molecular mass of 36.7 kDa, whereas SoxR Cys<sup>50</sup>Ser and SoxR Cys<sup>50</sup>Ser Cys<sup>116</sup>Ser elute earlier ( $k_{av}$  = 0.174, 41.9 kDa), indicating a more open conformation. The resolution of the column does not allow clear separation of the different tetrameric conformations ( $k_{av}$  0.086 to 0.093, corresponding to 65.9 to 63.6 kDa).

This conclusion was fully supported by size exclusion chromatography (Figure 6b). All variants, as well as wild-type SoxR, eluted with  $k_{av}$  values corresponding to molecular masses between 37.6 and 41.6 kDa, indicating dimerization of the 15.2 kDa monomers. Tetramers were also observed, with the highest abundance for the SoxR Cys<sup>116</sup>Ser variant. All proteins showed a tendency for the formation of higher oligomers in the void volume (Figure 6b). Notably, the Sox Cys<sup>50</sup>Ser single and the Cys<sup>50</sup>Ser Cys<sup>116</sup>Ser variant exchanges led to dimers eluting significantly earlier than those of wild-type SoxR and SoxR

Cys<sup>116</sup>Ser, indicating that the loss of Cys<sup>50</sup> but not that of Cys<sup>116</sup> leads to a more open, space-demanding conformation of the regulator protein.

### 3.6. SoxR Binding Properties

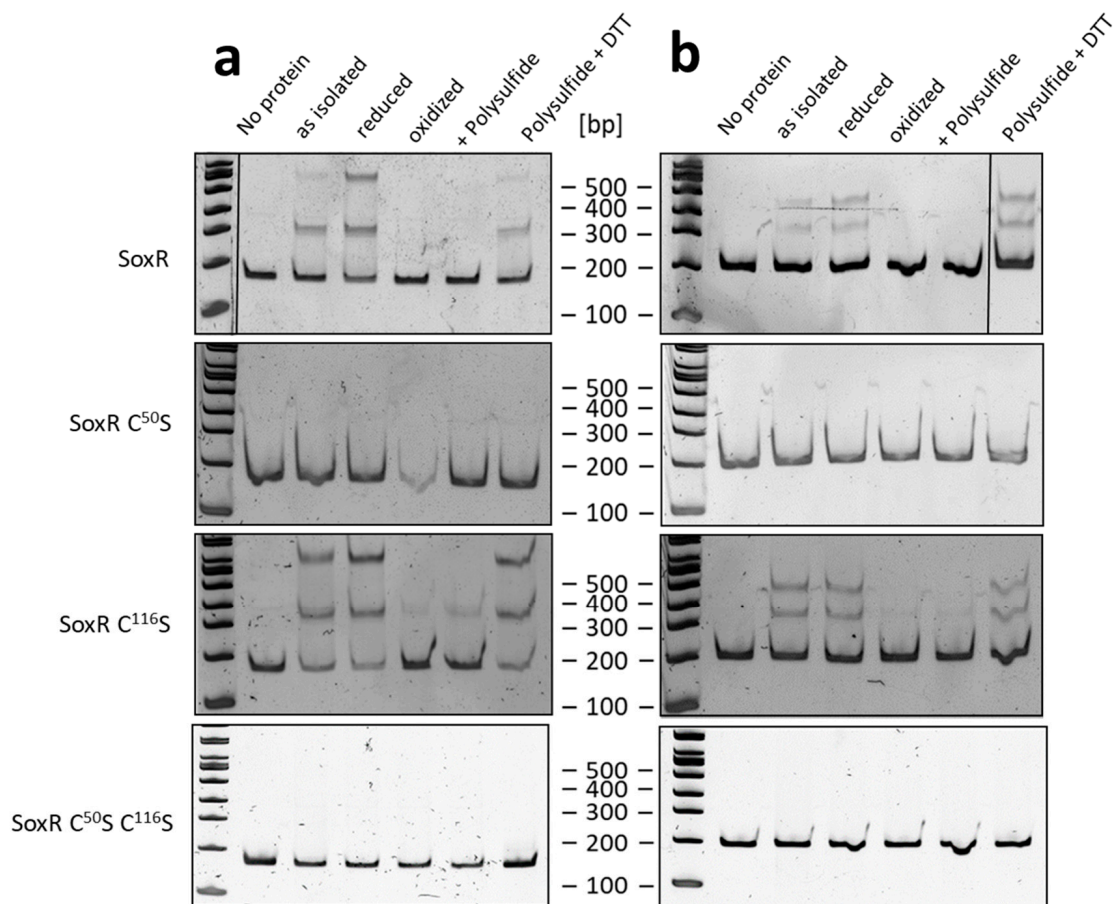
In the next step, EMSA assays were performed that allowed more detailed insights into the binding capacity of SoxR to the intergenic region between the divergently oriented *soxXA* and *soxYZB* genes (Figure 7). SoxR binds to the DNA probe in a concentration-dependent manner and leads to the appearance of two shifted bands indicating two different binding sites (Figure 7a, upper panel). As related proteins respond to persulfuration [35,64], we tested the response of SoxR to treatment with polysulfide, oxidized and reduced glutathione (GSH and GSSG), tetrathionate, sulfite, and thiosulfate in various molar ratios of protein and additive. Whereas GSH, GSSG, tetrathionate, and sulfite had no effect even when present in 50-fold excess compared to the protein (not shown), treatment with polysulfide above a molar ratio of 1 completely prevented binding of SoxR to the target DNA, and a shift was no longer observed (Figure 7a, lower panel, and Figure 7b, upper panel). Thiosulfate also had an effect, albeit a much milder one (Figure 7b, lower panel). The second shifted band disappeared at a ratio thiosulfate/SoxR of 5, and the first band still persisted at a ratio of 50, corresponding to a thiosulfate concentration of 0.2 mM. As outlined in the introduction, the initial steps of thiosulfate degradation occur in the periplasm, and it is therefore unlikely that thiosulfate would ever reach concentrations in the cytoplasm that would be required to elicit a response from SoxR.



**Figure 7.** (a) EMSA of the 151-bp *soxA–soxY* intergenic fragment with increasing amounts of untreated SoxR (**upper panel**) or SoxR pre-incubated with polysulfide in a molar ratio of SoxR/polysulfide of 1:1 (**lower panel**); (b) EMSA of the 151-bp *soxA–soxY* intergenic fragment with SoxR pre-incubated with increasing amounts of polysulfide (**upper panel**) or thiosulfate (**lower panel**). Vertical lines separate samples that were run on the same gel but were not directly adjacent.

EMSA assays were also performed with the as-isolated, reduced, oxidized, and polysulfide-treated SoxR variants and two different DNA probes (Figure 8). Reduction with DTT led to the same results as those obtained for the untreated proteins, indicating that they are fully reduced upon isolation and remain in this state during storage. Oxidation of wild-

type SoxR prevented binding to both tested DNA probes. While the SoxR Cys<sup>50</sup>Ser variant completely lost its DNA binding ability, the Cys<sup>116</sup>Ser variant bound effectively to the DNA probes. When polysulfide-treated wild-type SoxR was reduced with DTT in a second step, the protein regained its DNA-binding capacity, demonstrating that the modification caused by polysulfide was fully reversible by reduction. A response to oxidation or incubation with polysulfide was still observed for the Cys<sup>116</sup>Ser variant, albeit weaker than that of the wild-type protein. This behavior differs significantly from that of the related SqrR from *R. capsulatus*, where variants lacking one of the two conserved cysteines bind to their target DNA but do not show a loss of affinity upon persulfuration [35]. The SoxR variant lacking both cysteines was unable to bind DNA, regardless of the treatments applied.

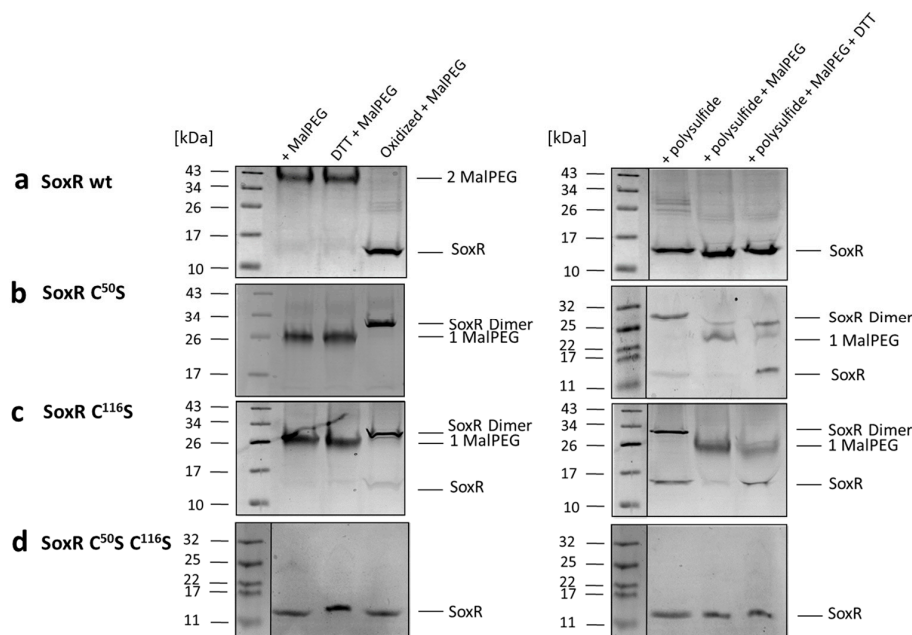


**Figure 8.** (a) EMSA of the 151-bp *soxA–soxY* intergenic fragment (17 nM) with 700 nM SoxR wild-type and variant proteins as isolated, reduced with DTT, oxidized with CuCl<sub>2</sub>, treated with polysulfide, and sequentially treated with polysulfide and DTT; (b) EMSA of the 180 bp central part of the *soxTIA–shdrR* intergenic fragment (17 nM) with 300 nM SoxR wild-type and variant proteins as isolated, reduced with DTT, oxidized with CuCl<sub>2</sub>, treated with polysulfide, and sequentially treated with polysulfide and DTT.

### 3.7. Redox State and Modification of SoxR

To clarify the chemical nature of the SoxR modifications by polysulfide and oxidation, gel-shift assays were performed using MalPEG, which selectively labels free thiol groups

covalently [67]. The modification can be detected by non-reducing SDS–PAGE since the molecular mass of the protein is increased by ~10 kDa per SH group modified. Treatment of the recombinant wild-type SoxR protein with MalPEG resulted in a single 20 kDa band shift, indicating that it contains two free cysteine residues, as expected (Figure 9a). In contrast, oxidized SoxR did not react with MalPEG (Figure 9a), demonstrating the existence of a disulfide bridge between Cys<sup>50</sup> and Cys<sup>116</sup>, as also suggested by non-reducing SDS–PAGE in the absence of MalPEG (Figure 6a). MalPEG labeling of the SoxR variants gave the expected results, with the variants carrying one cysteine showing a single 10 kDa shift (Figure 9b,c) and the double mutated variant not reacting with MalPEG as predicted (Figure 9d). After oxidation, the SoxR Cys<sup>50</sup>Ser variant produced exclusively dimers connected by Cys<sup>116</sup>–Cys<sup>116</sup> disulfide bridges and not reacting with MalPEG (Figure 9b), whereas the Cys<sup>116</sup>Ser variant showed incomplete dimerization. This observation is corroborated by the response of SoxR and its variants to polysulfide (Figure 9, right panels). While the wild-type protein stayed essentially monomeric, i.e., disulfide bonds between protein monomers were not formed, the SoxR Cys<sup>50</sup>Ser variant completely transformed into a dimer stable under denaturing conditions (Figure 9b). The Cys<sup>116</sup>Ser variant behaved differently, with a substantial fraction staying monomeric (Figure 9c). We note that the dimeric fraction of both variants obtained after treatment with polysulfide turned monomeric after incubation with MalPEG, possibly indicating a (poly)sulfur bridge between the remaining cysteine residues that is susceptible to cleavage by the thiol-binding agent. In conclusion, Cys<sup>50</sup> residues appear less prone to reaction than Cys<sup>116</sup> residues, just as has been reported for the corresponding cysteines in *R. capsulatus* SqrR [68], and/or they reside further apart from each other in the native SoxR dimer than Cys<sup>116</sup> residues.



**Figure 9.** Analysis of *H. denitrificans* SoxR cysteines with MalPEG gel-shift assays in non-reducing SDS–PAGE. Results are shown for the SoxR wild-type (wt) protein (a), single (Cys<sup>50</sup>Ser (b), and Cys<sup>116</sup>Ser (c), and double (Cys<sup>50</sup>Ser Cys<sup>116</sup>Ser (d)) variants after MalPEG treatment of the as-isolated, reduced, and oxidized states (left panels) as well as after pre-incubation with polysulfide (right panels). Polysulfide and MalPEG-reacted samples were further reduced with DTT. Vertical lines separate samples that were run on the same gel but were not directly adjacent.

The next set of reactions was the most revealing. When wild-type polysulfide-treated SoxR was reacted with MalPEG, it behaved just like the oxidized protein, i.e., MalPEG was not bound, indicating the absence of free cysteines (Figure 9a). Instead, one MalPEG was bound to the polysulfide-treated single cysteine replacement variants and could be released upon reduction with DTT (Figure 9b,c). We conclude that in the two latter cases, polysulfide led to the persulfuration of the single remaining cysteines, which then bound MalPEG. In the final step, MalPEG-sulfide conjugates were released by treatment with DTT, and the single cysteine SoxR variants reappeared in their unmodified monomeric form. The situation for wild-type SoxR is completely different. Either polysulfide merely leads to the formation of a Cys<sup>50</sup>–Cys<sup>116</sup> bridge, or one or more sulfur atoms are enclosed by the two cysteines.

Mass spectrometric analyses finally allowed a clear differentiation between these two possibilities (Table 1, Supplementary Figure S1). For these experiments, MalPEG was replaced by the thiol-modifying agent iodoacetamide, which leads to carbamidomethylation of free Cys sulfhydryl groups and thus adds a mass of 57 Da. As expected, wild-type SoxR gained 57 Da twice after iodoacetamide treatment, whereas the single Cys replacement variants were modified with only one carbamido group. Notably, polysulfide treatment led to persulfuration of all SoxR proteins except for the cysteine-less double replacement variant, which was measured as a control. The SoxR wild-type protein was modified with up to three sulfur atoms (+32 Da each) and did not react with iodoacetamide, demonstrating the formation of an intramolecular tri-, tetra-, or penta-sulfide bond between Cys<sup>50</sup> and Cys<sup>116</sup>. Although mass spectra do not provide exact quantitative information, peak heights indicate that bridges by two additional sulfur atoms are more abundant than one or three atom bridges for the SoxR wild-type protein, while the majority of the SoxR Cys<sup>50</sup>Ser and Cys<sup>116</sup>Ser variant polypeptides are modified by only one sulfur atom (Supplementary Figure S1).

**Table 1.** Mass spectrometry of SoxR and variants after treatment with modifying agents. CAM, carbamidomethylation; S, sulfur. Calculated masses for Strep-tagged SoxR and SoxR Cys<sup>50</sup>Ser, SoxR Cys<sup>116</sup>Ser, and SoxR Cys<sup>50</sup>Ser Cys<sup>116</sup> without the initiator methionine are 15,212.54 Da, 15,197.54 Da, 15,197.54 Da, and 15,182.54 Da, respectively.

Treatment	SoxR Mass (Da) (Addition: [Da])	SoxR C <sup>50</sup> S Mass (Da) (Addition: [Da])	SoxR C <sup>116</sup> S Mass (Da) (Addition: [Da])	SoxR C <sup>50</sup> S C <sup>116</sup> S Mass (Da) (Addition: [Da])
Native	15,212.8	15,197.3	15,198.2	15,182.3
DTT reduced	15,212.5	15,199.3	15,198.8	nd
CuCl <sub>2</sub> oxidized	15,210.5	15,196.7	15,196.9	15,182.0
Iodoacetamide	15,328.0 (2 CAM: 2 × 57.07)	15,255.2 (1 CAM: 57.07)	15,255.22 (1 CAM: 57.07)	nd
Polysulfide	15,212.5 15,244.7 (1 S: 32) 15,276.4 (2 S: 64) 15,308.0 (3 S: 96)	15,198.9 15,230.0 (1 S: 32)	15,198.0 15,230.3 (1 S: 32) 15,261.1 (2 S: 64)	15,182.0
Polysulfide + Iodoacetamide	15,212.9 15,244.3 (1 S: 32) 15,275.2 (2 S: 64) 15,306.0 (3 S: 96)	15,198.0 15,286.0 (1 S + 1 CAM: 90)	15,197.6 15,228.5 (1 S: 32) 15,285.4 (1 S + 1 CAM: 90)	15,180.5

#### 4. Discussion

In this study, we collected a wealth of new information on the transcriptional repressor SoxR. We show that among the more than 70,000 prokaryotic genomes investigated, bonafide *soxR* (i.e., genetically linked to the genes for the SoxYZ sulfur-binding protein and/or catalytic Sox components) occurs exclusively among the bacterial phylum Pro-



teobacteria, where it is more frequently found in gene clusters for complete Sox systems than for truncated Sox systems. Based on the available data, it is difficult to draw general conclusions from this observation. However, it appears that a number of bacteria that operate the truncated Sox system, such as the green and purple sulfur bacteria or members of the Aquificota, are dedicated sulfur-oxidizing chemolithoautotrophs without much need for sophisticated transcriptional regulation of the sulfur oxidation machinery.

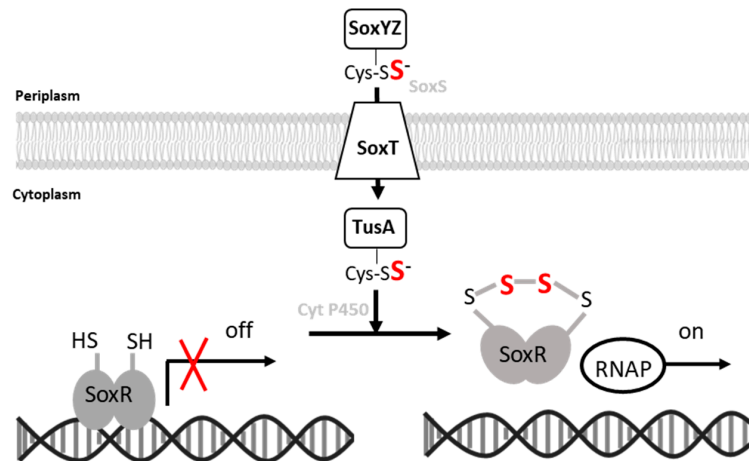
We show that in the model Alphaproteobacterium *H. denitrificans* SoxR is not only involved in the transcriptional regulation of true *sox* genes but that it also affects the transcription of a number of other genes. In particular, the *shdr* genes, which encode the cytoplasmic sulfur-oxidizing multi-enzyme system required for sulfane sulfur oxidation that cannot be achieved by the truncated hypomicrobial Sox system, are co-controlled by SoxR. How it interacts with a second, related repressor, sHdrR, that affects the transcription of the same genes [9] is an important research question for the future.

The expression levels of *sox* as well as of *shdr* and associated genes are increased by thiosulfate in wild-type cells and elevated in the *soxR*-deficient *H. denitrificans* mutant, irrespective of the presence of thiosulfate (Figure 4). DNA binding in vitro and probably also transcriptional repression in living cells involve thiol modifications. This can be concluded from the observation that the DNA-binding activity of recombinant SoxR is strongly reduced upon incubation with polysulfide, which leads to persulfuration of the regulator, as proven by reaction with MalPEG (Figure 9) and mass spectrometry (Table 1, Supplementary Figure S1). In polysulfide-treated SoxR, the two conserved cysteine residues can neither be modified by MalPEG nor by iodoacetamide. In addition, polysulfide treatment increases the mass of wild-type SoxR by 32, 64, or 96 Da. These findings can be fully explained by the formation of an intramolecular tri-, tetra-, or penta-sulfide bond formed upon interaction with reactive sulfane sulfur species. Thus, SoxR clearly is not a simple redox sensor switching between dithiol and disulfide states but has been identified as a transcriptional regulator sensing reactive sulfane sulfur species (Figure 10), similar to the related SqrR protein from *R. capsulatus* [35,68]. Notably, the substitution of the two crucial conserved cysteine residues leads to a different outcome for SoxR as compared to SqrR: The lack of Cys<sup>50</sup> causes complete loss of DNA binding in SoxR, whereas the lack of Cys<sup>116</sup> creates a variant that tightly binds to its target DNA and is less sensitive to persulfuration than the wild-type protein. In SqrR, both equivalent Cys–Ser variants are DNA-binding competent and do not respond to persulfuration as a signal [35]. Clearly, this difference inspires future research that should also include a detailed inspection of the conformational changes triggered by the formation of a sulfur bridge and resulting in the detachment of SoxR from its target DNA (Figure 10).

The physiological processes involving the various sulfane sulfur-responsive regulators characterized so far [34,35,69–72] differ fundamentally from those controlled by SoxR. The former mainly regulate stress responses, sense intracellular and extracellular reactive sulfur species, and ensure upregulation of H<sub>2</sub>S oxidation genes for the purpose of detoxification, i.e., they control the removal of excess sulfide and sulfane sulfur, thus contributing to cell survival in the presence of external reactive sulfur species. In contrast, SoxR regulates dissimilatory sulfur metabolism and enables the use of reduced sulfur compounds such as thiosulfate as electron donors for lithotrophic or mixotrophic growth.

As pointed out earlier, thiosulfate oxidation is initiated in the periplasm, and it is highly unlikely that thiosulfate itself serves as the signaling molecule. Instead, SoxR responds to the presence of low concentrations of sulfane sulfur, which was provided as polysulfide in our in vitro assays. A working hypothesis for how this signal reaches its destination, inspired by the arrangement of the respective genes in *H. denitrificans* (Figure 1b), is presented in Figure 10. It is conceivable that the sulfur bound to the sulfur carrier protein SoxYZ in the periplasm in the course of thiosulfate oxidation reaches the cytoplasm via a YedE-like SoxT transporter [73]. The periplasmic thiol–disulfide oxidoreductase SoxS [15] could be involved in the transfer of the sulfane sulfur to the transporter. Once in the cytoplasm, the sulfur transferase TusA [74] is a possible acceptor protein for the sulfur,

which could be passed on from there to SoxR, possibly involving the cytochrome P450 encoded by gene Hden\_0697.



**Figure 10.** Proposed signal transduction pathway and mode of action of the homodimeric SoxR repressor protein. Established sulfur-binding proteins are printed in black. Sulfane sulfur atoms that come originally from thiosulfate (cf. Figure 1) are highlighted in bold red.

## 5. Conclusions

In conclusion, our study shows that SoxR allows *H. denitrificans* to adapt to changes in thiosulfate availability via thiol persulfidation chemistry and the formation of an intramolecular sulfur bridge, which may involve transporters and sulfurtransferases encoded in the same genetic region. Clearly, much remains to be learned about this regulator, not only in terms of signal transduction but also in terms of crosstalk with its counterpart, sHdrR.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/antiox12081620/s1>, Figure S1: Mass spectra for SoxR and variants with Cys-Ser exchanges; Table S1: Strains, plasmids, and primers; Table S2: Distribution of Sox systems and SoxR. References [8,9,42] are cited in Table S1.

**Author Contributions:** C.D. and J.L. conceptualized the study; J.L., K.T., J.K., T.S.T. and H.Y.H. performed experiments; T.S.T. performed bioinformatic analyses. C.D. supervised the work and acquired funding. C.D. and J.L. wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was in part funded by the German Science Foundation, grant numbers Da 351/13-1 and Da 351/8-2. J.L. was financed by a Scholarship from the China Scholarship Council, and T.S.T. received a scholarship from the Studienstiftung des Deutschen Volkes.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article and Supplementary Materials.

**Acknowledgments:** We thank Laura Schiffer for helping with protein production. We gratefully acknowledge the support of the Core Facility "Protein Synthesis and Bioanalytics" of the University of Bonn for performing mass spectrometry.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

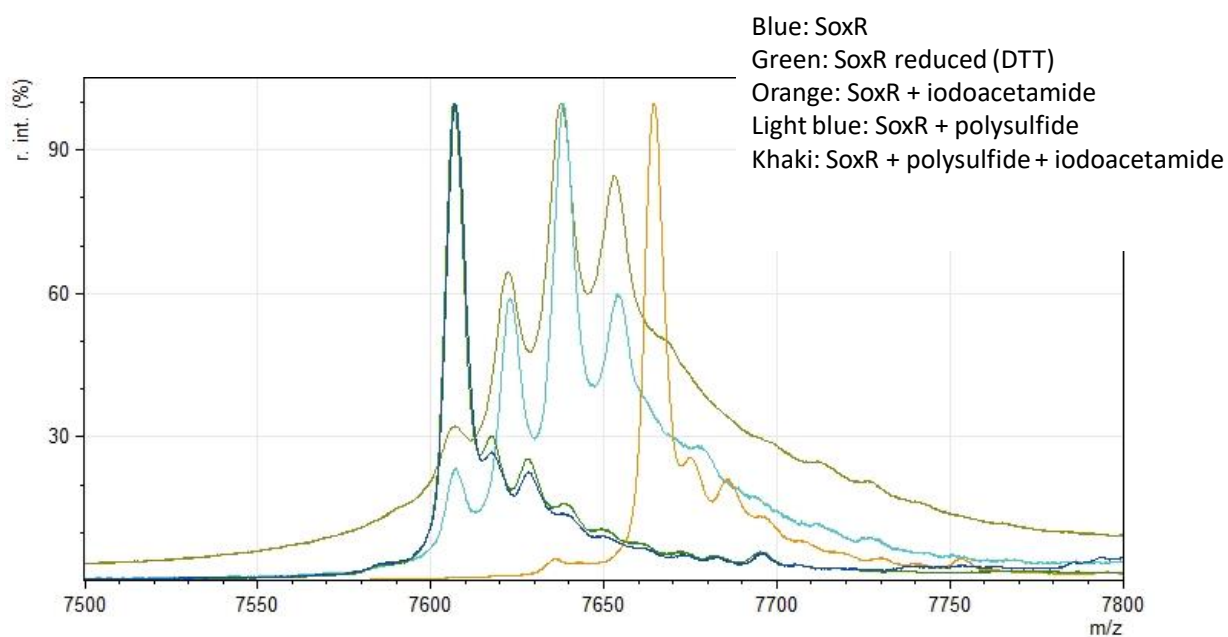
## References

- Friedrich, C.G.; Rother, D.; Bardischewsky, F.; Quentmeier, A.; Fischer, J. Oxidation of reduced inorganic sulfur compounds by bacteria: Emergence of a common mechanism? *Appl. Environ. Microbiol.* **2001**, *67*, 2873–2882. [[CrossRef](#)]
- Dahl, C. A biochemical view on the biological sulfur cycle. In *Environmental Technologies to Treat Sulfur Pollution: Principles and Engineering*, 2nd ed.; Lens, P., Ed.; IWA Publishing: London, UK, 2020; pp. 55–96.
- Appia-Ayme, C.; Little, P.J.; Matsumoto, Y.; Leech, A.P.; Berks, B.C. Cytochrome complex essential for photosynthetic oxidation of both thiosulfate and sulfide in *Rhodovulum sulfidophilum*. *J. Bacteriol.* **2001**, *183*, 6107–6118. [[CrossRef](#)]
- Grabarczyk, D.B.; Berks, B.C. Intermediates in the Sox sulfur oxidation pathway are bound to a sulfane conjugate of the carrier protein SoxYZ. *PLoS ONE* **2017**, *12*, e0173395. [[CrossRef](#)]
- Sauvé, V.; Roversi, P.; Leath, K.J.; Garman, E.F.; Antrobus, R.; Lea, S.M.; Berks, B.C. Mechanism for the hydrolysis of a sulfur-sulfur bond based on the crystal structure of the thiosulfohydrolase SoxB. *J. Biol. Chem.* **2009**, *284*, 21707–21718. [[CrossRef](#)]
- Grabarczyk, D.B.; Chappell, P.E.; Johnson, S.; Stelzl, L.S.; Lea, S.M.; Berks, B.C. Structural basis for specificity and promiscuity in a carrier protein/enzyme system from the sulfur cycle. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E7166–E7175. [[CrossRef](#)]
- Zander, U.; Faust, A.; Klink, B.U.; de Sanctis, D.; Panjikar, S.; Quentmeier, A.; Bardischewsky, F.; Friedrich, C.G.; Scheidig, A.J. Structural basis for the oxidation of protein-bound sulfur by the sulfur cycle molybdohemo-enzyme sulfane dehydrogenase SoxCD. *J. Biol. Chem.* **2010**, *286*, 8349–8360. [[CrossRef](#)] [[PubMed](#)]
- Koch, T.; Dahl, C. A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds. *ISME J.* **2018**, *12*, 2479–2491. [[CrossRef](#)] [[PubMed](#)]
- Li, J.; Koch, J.; Flegler, W.; Garcia Ruiz, L.; Hager, N.; Ballas, A.; Tanabe, T.S.; Dahl, C. A metabolic puzzle: Consumption of C<sub>1</sub> compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>. *Biochim. Biophys. Acta Bioenerget.* **2023**, *1864*, 148932. [[CrossRef](#)] [[PubMed](#)]
- Cao, X.; Koch, T.; Steffens, L.; Finkensieper, J.; Zigann, R.; Cronan, J.E.; Dahl, C. Lipoate-binding proteins and specific lipoate-protein ligases in microbial sulfur oxidation reveal an atypical role for an old cofactor. *eLife* **2018**, *7*, e37439. [[CrossRef](#)] [[PubMed](#)]
- Appia-Ayme, C.; Berks, B.C. SoxV, an orthologue of the CcdA disulfide transporter, is involved in thiosulfate oxidation in *Rhodovulum sulfidophilum* and reduces the periplasmic thioredoxin SoxW. *Biochem. Biophys. Res. Commun.* **2002**, *296*, 737–741. [[CrossRef](#)]
- Friedrich, C.G.; Quentmeier, A.; Bardischewsky, F.; Rother, D.; Orawski, G.; Hellwig, P.; Fischer, J. Redox control of chemotrophic sulfur oxidation of *Paracoccus pantotrophus*. In *Microbial Sulfur Metabolism*; Dahl, C., Friedrich, C.G., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 139–150.
- Orawski, G.; Bardischewsky, F.; Quentmeier, A.; Rother, D.; Friedrich, C.G. The periplasmic thioredoxin SoxS plays a key role in activation in vivo of chemotrophic sulfur oxidation of *Paracoccus pantotrophus*. *Microbiology* **2007**, *153*, 1081–1086. [[CrossRef](#)] [[PubMed](#)]
- Carius, Y.; Rother, D.; Friedrich, C.G.; Scheidig, A.J. The structure of the periplasmic thiol-disulfide oxidoreductase SoxS from *Paracoccus pantotrophus* indicates a triple Trx/Grx/DsbC functionality in chemotrophic sulfur oxidation. *Acta Crystallogr. D Biol. Cryst.* **2009**, *65*, 229–240. [[CrossRef](#)] [[PubMed](#)]
- Rother, D.; Ringk, J.; Friedrich, C.G. Sulfur oxidation of *Paracoccus pantotrophus*: The sulfur-binding protein SoxYZ is the target of the periplasmic thiol-disulfide oxidoreductase SoxS. *Microbiology* **2008**, *154*, 1980–1988. [[CrossRef](#)]
- Neutzling, O.; Pfeleiderer, C.; Trüper, H.G. Dissimilatory sulphur metabolism in phototrophic “non-sulphur” bacteria. *J. Gen. Microbiol.* **1985**, *131*, 791–798. [[CrossRef](#)]
- Kappler, U.; Enemark, J.H. Sulfite-oxidizing enzymes. *J. Biol. Inorg. Chem.* **2014**, *20*, 253–264. [[CrossRef](#)]
- Kappler, U.; Schwarz, G. The sulfite oxidase family of molybdenum enzymes. In *Molybdenum and Tungsten Enzymes*; Hille, R., Schulzke, C., Kirk, M.L., Eds.; Royal Society of Chemistry: Cambridge, UK, 2016; pp. 240–273.
- Meng, Z.; Qin, G.; Zhang, B.; Bai, J. DNA damaging effects of sulfur dioxide derivatives in cells from various organs of mice. *Mutagenesis* **2004**, *19*, 465–468. [[CrossRef](#)]
- Yi, H.; Liu, J.; Zheng, K. Effect of sulfur dioxide hydrates on cell cycle, sister chromatid exchange, and micronuclei in barley. *Ecotoxicol. Environ. Saf.* **2005**, *62*, 421–426. [[CrossRef](#)]
- Ozturk, O.H.; Kucukatay, V.; Yonden, Z.; Agar, A.; Bagci, H.; Delibas, N. Expressions of N-methyl-D-aspartate receptors NR2A and NR2B subunit proteins in normal and sulfite-oxidase deficient rat’s hippocampus: Effect of exogenous sulfite ingestion. *Arch. Toxicol.* **2006**, *80*, 671–679. [[CrossRef](#)]
- Kao, Y.T.; Tan, C.; Song, S.H.; Ozturk, N.; Li, J.; Wang, L.; Sancar, A.; Zhong, D. Ultrafast dynamics and anionic active states of the flavin cofactor in cryptochrome and photolyase. *J. Am. Chem. Soc.* **2008**, *130*, 7695–7701. [[CrossRef](#)]
- Ozawa, T.; Hanaki, A. Spin-trapping of sulfite radical anion, SO<sub>3</sub><sup>•-</sup>, by a water-soluble, nitroso-aromatic spin-trap. *Biochem. Biophys. Res. Commun.* **1987**, *142*, 410–416. [[CrossRef](#)]
- Inouye, B.; Morita, K.; Ishida, T.; Ogata, M. Cooperative effect of sulfite and vanadium compounds on lipid peroxidation. *Toxicol. Appl. Pharmacol.* **1980**, *53*, 101–107. [[CrossRef](#)]
- Yang, S.F. Destruction of tryptophan during the aerobic oxidation of sulfite ions. *Environ. Res.* **1973**, *6*, 395–402. [[CrossRef](#)] [[PubMed](#)]
- Rother, D.; Orawski, G.; Bardischewsky, F.; Friedrich, C.G. SoxRS-mediated regulation of chemotrophic sulfur oxidation in *Paracoccus pantotrophus*. *Microbiology* **2005**, *151*, 1707–1716. [[CrossRef](#)] [[PubMed](#)]

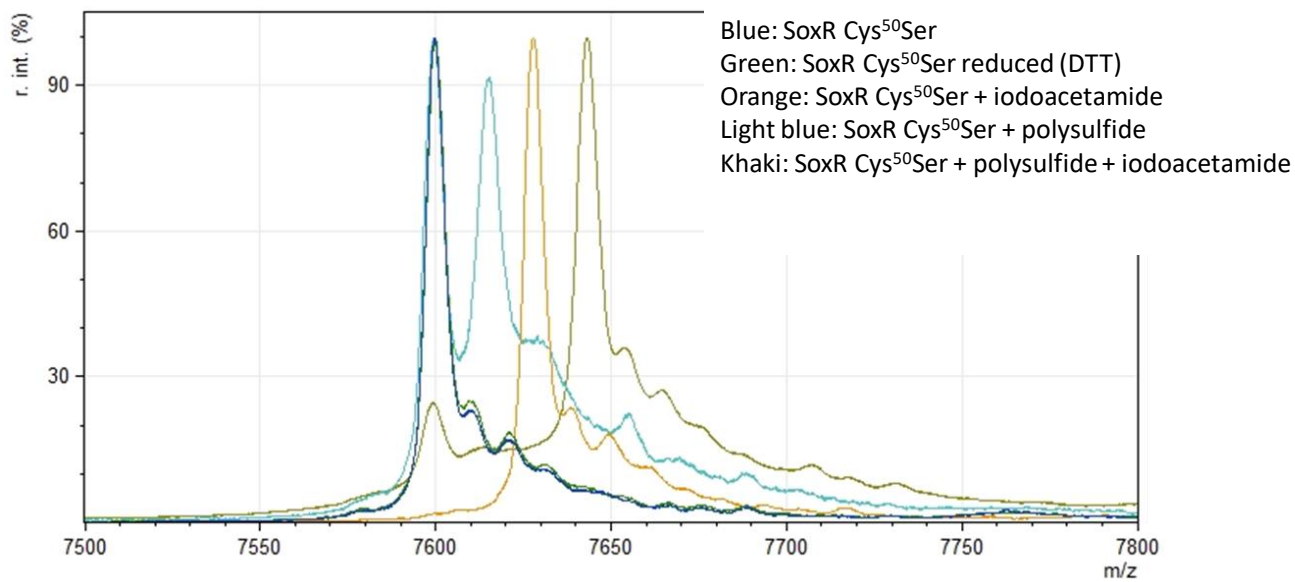
27. Mandal, S.; Chatterjee, S.; Dam, B.; Roy, P.; Das Gupta, S.K. The dimeric repressor SoxR binds cooperatively to the promoter(s) regulating expression of the sulfur oxidation (*sox*) operon of *Pseudaminobacter salicylatoxidans* KCT001. *Microbiology* **2007**, *153*, 80–91. [[CrossRef](#)]
28. Lahiri, C.; Mandal, S.; Ghosh, W.; Dam, B.; Roy, P. A novel gene cluster *soxSRT* is essential for the chemolithotrophic oxidation of thiosulfate and tetrathionate by *Pseudaminobacter salicylatoxidans* KCT001. *Curr. Microbiol.* **2006**, *52*, 267–273. [[CrossRef](#)] [[PubMed](#)]
29. Ma, Z.; Jacobsen, F.E.; Giedroc, D.P. Coordination chemistry of bacterial metal transport and sensing. *Chem. Rev.* **2009**, *109*, 4644–4681. [[CrossRef](#)] [[PubMed](#)]
30. Cook, W.J.; Kar, S.R.; Taylor, K.B.; Hall, L.M. Crystal structure of the cyanobacterial metallothionein repressor SmtB: A model for metalloregulatory proteins. *J. Mol. Biol.* **1998**, *275*, 337–346. [[CrossRef](#)]
31. Busenlehner, L.S.; Pennella, M.A.; Giedroc, D.P. The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiol. Rev.* **2003**, *27*, 131–143. [[CrossRef](#)]
32. Osman, D.; Cavet, J.S. Bacterial metal-sensing proteins exemplified by ArsR-SmtB family repressors. *Nat. Prod. Rep.* **2010**, *27*, 668–680. [[CrossRef](#)]
33. Capdevila, D.A.; Edmonds, K.A.; Giedroc, D.P. Metallochaperones and metalloregulation in bacteria. *Essays Biochem.* **2017**, *61*, 177–200.
34. Guimaraes, B.G.; Barbosa, R.L.; Soprano, A.S.; Campos, B.M.; de Souza, T.A.; Tonoli, C.C.; Leme, A.F.; Murakami, M.T.; Benedetti, C.E. Plant pathogenic bacteria utilize biofilm growth-associated repressor (BigR), a novel winged-helix redox switch, to control hydrogen sulfide detoxification under hypoxia. *J. Biol. Chem.* **2011**, *286*, 26148–26157. [[CrossRef](#)] [[PubMed](#)]
35. Shimizu, T.; Shen, J.; Fang, M.; Zhang, Y.; Hori, K.; Trinidad, J.C.; Bauer, C.E.; Giedroc, D.P.; Masuda, S. Sulfide-responsive transcriptional repressor SqrR functions as a master regulator of sulfide-dependent photosynthesis. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2355–2360. [[CrossRef](#)] [[PubMed](#)]
36. Shimizu, T.; Ida, T.; Antelo, G.T.; Ihara, Y.; Fakhoury, J.N.; Masuda, S.; Giedroc, D.P.; Akaike, T.; Capdevila, D.A.; Masuda, T. Polysulfide metabolizing enzymes influence SqrR-mediated sulfide-induced transcription by impacting intracellular polysulfide dynamics. *PNAS Nexus* **2023**, *2*, 1–11. [[CrossRef](#)] [[PubMed](#)]
37. Bertani, G. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J. Bacteriol.* **2004**, *186*, 595–600. [[CrossRef](#)]
38. Ausubel, F.A.; Brent, R.; Kingston, R.E.; Moore, D.D.; Seidman, J.G.; Smith, J.A.; Struhl, K. *Current Protocols in Molecular Biology*; John Wiley & Sons: New York, NY, USA, 1997.
39. Horton, R.M. PCR mediated recombination and mutagenesis: SOEing together tailor-made genes. *Mol. Biotechnol.* **1995**, *3*, 93–99. [[CrossRef](#)]
40. Dahl, C. Insertional gene inactivation in a phototrophic sulphur bacterium: APS-reductase-deficient mutants of *Chromatium vinosum*. *Microbiology* **1996**, *142*, 3363–3372. [[CrossRef](#)]
41. Chomczynski, P.; Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **1987**, *162*, 156–159. [[CrossRef](#)]
42. Martineau, C.; Mauffrey, F.; Villemur, R. Comparative analysis of denitrifying activities of *Hyphomicrobium nitrivorans*, *Hyphomicrobium denitrificans*, and *Hyphomicrobium zavarzinii*. *Appl. Environ. Microbiol.* **2015**, *81*, 5003–5014. [[CrossRef](#)]
43. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)]
44. Ikeda, S.; Satake, H.; Hisano, T.; Terazawa, T. Potentiometric argentimetric method for the successive titration of sulphide and dissolved sulphur in polysulphide solutions. *Talanta* **1972**, *19*, 1650–1654. [[CrossRef](#)]
45. Parks, D.H.; Chuvochina, M.; Waite, D.W.; Rinke, C.; Skarszewski, A.; Chaumeil, P.-A.; Hugenholtz, P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **2018**, *36*, 996–1004. [[CrossRef](#)] [[PubMed](#)]
46. Rinke, C.; Chuvochina, M.; Mussig, A.J.; Chaumeil, P.A.; Davin, A.A.; Waite, D.W.; Whitman, W.B.; Parks, D.H.; Hugenholtz, P. A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat. Microbiol.* **2021**, *6*, 946–959. [[CrossRef](#)] [[PubMed](#)]
47. Hyatt, D.; Chen, G.L.; Locascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **2010**, *11*, 119. [[CrossRef](#)] [[PubMed](#)]
48. Tanabe, T.S.; Dahl, C. HMS-S-S: A tool for the identification of sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies. *Mol. Ecol. Resour.* **2022**, *22*, 2758–2774. [[CrossRef](#)]
49. Dahl, C. Sulfur metabolism in phototrophic bacteria. In *Modern Topics in the Phototrophic Prokaryotes: Metabolism, Bioenergetics and Omics*; Hallenbeck, P.C., Ed.; Springer International Publishing: Cham, Switzerland, 2017; pp. 27–66.
50. Gupta, R.S. The phylum Aquificae. In *The Prokaryotes—Other Lineages of Bacteria and the Archaea*; Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2014.
51. Kodama, Y.; Watanabe, K. *Sulfuricurvum kujiense* gen. nov., sp. nov., a facultatively anaerobic, chemolithoautotrophic, sulfur-oxidizing bacterium isolated from an underground crude-oil storage cavity. *Int. J. Syst. Evol. Microbiol.* **2004**, *54*, 2297–2300. [[CrossRef](#)]
52. Xie, S.; Wang, S.; Li, D.; Shao, Z.; Lai, Q.; Wang, Y.; Wei, M.; Han, X.; Jiang, L. *Sulfurovum indicum* sp. nov., a novel hydrogen- and sulfur-oxidizing chemolithoautotroph isolated from a deep-sea hydrothermal plume in the Northwestern Indian Ocean. *Int. J. Syst. Evol. Microbiol.* **2019**, *71*, 004748. [[CrossRef](#)] [[PubMed](#)]

53. Labrenz, M.; Grote, J.; Mammitzsch, K.; Boschker, H.T.S.; Laue, M.; Jost, G.; Glaubitz, S.; Jürgens, K. *Sulfurimonas gotlandica* sp. nov., a chemoautotrophic and psychrotolerant epsilonproteobacterium isolated from a pelagic redoxcline, and an emended description of the genus *Sulfurimonas*. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 4141–4148. [\[CrossRef\]](#)
54. Skirnisdottir, S.; Hreggvidsson, G.O.; Holst, O.; Kristjansson, J.K. Isolation and characterization of a mixotrophic sulfur-oxidizing *Thermus scotoductus*. *Extremophiles* **2001**, *5*, 45–51. [\[CrossRef\]](#)
55. Bjornsdottir, S.H.; Petursdottir, S.K.; Hreggvidsson, G.O.; Skirnisdottir, S.; Hjorleifsdottir, S.; Arnfinnsson, J.; Kristjansson, J.K. *Thermus islandicus* sp. nov., a mixotrophic sulfur-oxidizing bacterium isolated from the Torfajökull geothermal area. *Int. J. Syst. Evol. Microbiol.* **2009**, *59*, 2962–2966. [\[CrossRef\]](#)
56. Imhoff, J.F. Phylogenetic taxonomy of the family *Chlorobiaceae* on the basis of 16S rRNA and *fmo* (Fenna-Matthews-Olson protein) gene sequences. *Int. J. Syst. Evol. Microbiol.* **2003**, *53*, 941–951. [\[CrossRef\]](#)
57. Carareto Alves, L.M.; de Souza, J.A.M.; Varani, A.d.M.; Lemos, E.G.d.M. The Family Rhizobiaceae. In *The Prokaryotes*; Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 419–437.
58. Pujalte, M.J.; Lucena, T.; Ruvira, M.A.; Arahal, D.R.; Macián, M.C. The Family Rhodobacteraceae. In *The Prokaryotes*; Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 439–512.
59. Spietz, R.L.; Lundeen, R.A.; Zhao, X.; Nicastro, D.; Ingalls, A.E.; Morris, R.M. Heterotrophic carbon metabolism and energy acquisition in *Candidatus* *Thioglobus singularis* strain PS1, a member of the SUP05 clade of marine Gammaproteobacteria. *Environ. Microbiol.* **2019**, *21*, 2391–2401. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Imhoff, J.F. Family I. Chromatiaceae Bavendamm 1924, 125 AL emend. Imhoff 1984b, 339. In *Bergey's Manual of Systematic Bacteriology*; Brenner, D.J., Krieg, N.R., Staley, J.T., Garrity, G.M., Eds.; Springer: New York, NY, USA, 2005; Volume 2, pp. 3–40.
61. Imhoff, J.F. Family II. Ectothiorhodospiraceae Imhoff 1984b, 339 VP. In *Bergey's Manual of Systematic Bacteriology*; Brenner, D.J., Krieg, N.R., Staley, J.T., Garrity, G.M., Eds.; Springer: New York, NY, USA, 2005; Volume 2, pp. 41–57.
62. Boden, R.; Scott, K.M.; Williams, J.; Russel, S.; Antonen, K.; Rae, A.W.; Hutt, L.P. An evaluation of *Thiomicrospira*, *Hydrogenovibrio* and *Thioalkalimicrobium*: Reclassification of four species of *Thiomicrospira* to each *Thiomicrothabdis* gen. nov. and *Hydrogenovibrio*, and reclassification of all four species of *Thioalkalimicrobium* to *Thiomicrospira*. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 1140–1151. [\[PubMed\]](#)
63. Boden, R.; Scott, K.M. Evaluation of the genus *Thiothrix* Winogradsky 1888 (Approved Lists 1980) emend. Aruga et al. 2002: Reclassification of *Thiothrix disciformis* to *Thiolinea disciformis* gen. nov., comb. nov., and of *Thiothrix flexilis* to *Thiofilum flexile* gen. nov., comb. nov., with emended description of *Thiothrix*. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 2226–2239. [\[PubMed\]](#)
64. Shimizu, T.; Masuda, S. Characterization of redox-active cysteine residues of persulfide-responsive transcriptional repressor SqrR. *Commun. Integr. Biol.* **2017**, *10*, e1329786. [\[CrossRef\]](#)
65. Mukherjee, D.; Datta, A.B.; Chakrabarti, P. Crystal structure of HlyU, the hemolysin gene transcription activator, from *Vibrio cholerae* N16961 and functional implications. *Biochim. Biophys. Acta* **2014**, *1844*, 2346–2354. [\[CrossRef\]](#)
66. Gueuné, H.; Durand, M.J.; Thouand, G.; DuBow, M.S. The *ygaVP* genes of *Escherichia coli* form a tributyltin-inducible operon. *Appl. Environ. Microbiol.* **2008**, *74*, 1954–1958. [\[CrossRef\]](#)
67. Lu, J.; Deutsch, C. Pegylation: A method for assessing topological accessibilities in Kv1.3. *Biochemistry* **2001**, *40*, 13288–13301. [\[CrossRef\]](#)
68. Capdevila, D.A.; Walsh, B.J.C.; Zhang, Y.; Dietrich, C.; Gonzalez-Gutierrez, G.; Giedroc, D.P. Structural basis for persulfide-sensing specificity in a transcriptional regulator. *Nat. Chem. Biol.* **2021**, *17*, 65–70. [\[CrossRef\]](#)
69. Barbosa, R.L.; Benedetti, C.E. BigR, a transcriptional repressor from plant-associated bacteria, regulates an operon implicated in biofilm growth. *J. Bacteriol.* **2007**, *189*, 6185–6194. [\[CrossRef\]](#)
70. Grosseohme, N.; Kehl-Fie, T.E.; Ma, Z.; Adams, K.W.; Cowart, D.M.; Scott, R.A.; Skaar, E.P.; Giedroc, D.P. Control of copper resistance and inorganic sulfur metabolism by paralogous regulators in *Staphylococcus aureus*. *J. Biol. Chem.* **2011**, *286*, 13522–13531. [\[CrossRef\]](#)
71. Balasubramanian, R.; Hori, K.; Shimizu, T.; Kasamatsu, S.; Okamura, K.; Tanaka, K.; Ihara, H.; Masuda, S. The sulfide-responsive SqrR/BigR homologous regulator YgaV of *Escherichia coli* controls expression of anaerobic respiratory genes and antibiotic tolerance. *Antioxidants* **2022**, *11*, 2359. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Tang, C.; Li, J.; Shen, Y.; Liu, M.; Liu, H.; Liu, H.; Xun, L.; Xia, Y. A sulfide-sensor and a sulfane sulfur-sensor collectively regulate sulfur-oxidation for feather degradation by *Bacillus licheniformis*. *Commun. Biol.* **2023**, *6*, 167. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Tanaka, Y.; Yoshikaie, K.; Takeuchi, A.; Ichikawa, M.; Mori, T.; Uchino, S.; Sugano, Y.; Hakoshima, T.; Takagi, H.; Nonaka, G.; et al. Crystal structure of a YeeE/YedE family protein engaged in thiosulfate uptake. *Sci. Adv.* **2020**, *6*, eaba7637. [\[CrossRef\]](#)
74. Tanabe, T.S.; Leimkühler, S.; Dahl, C. The functional diversity of the prokaryotic sulfur carrier protein TusA. *Adv. Microb. Physiol.* **2019**, *75*, 233–277. [\[PubMed\]](#)

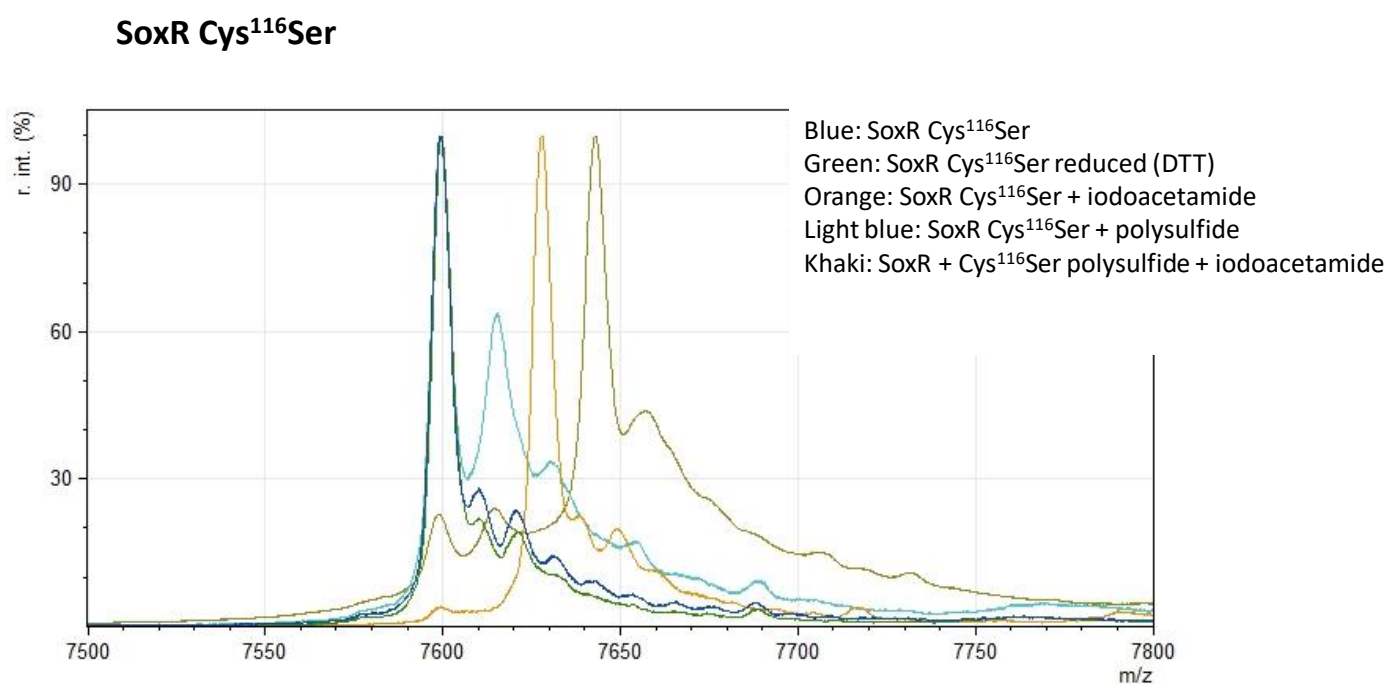
**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

**Figure S1. Mass spectra for SoxR and variants with Cys-Ser exchanges****SoxR wildtype**

Masses are shown for the  $mz/2$  species. Mass spectra obtained for the  $\text{CuCl}_2$  oxidized protein were almost identical to those for the as isolated protein and are not shown for clarity.

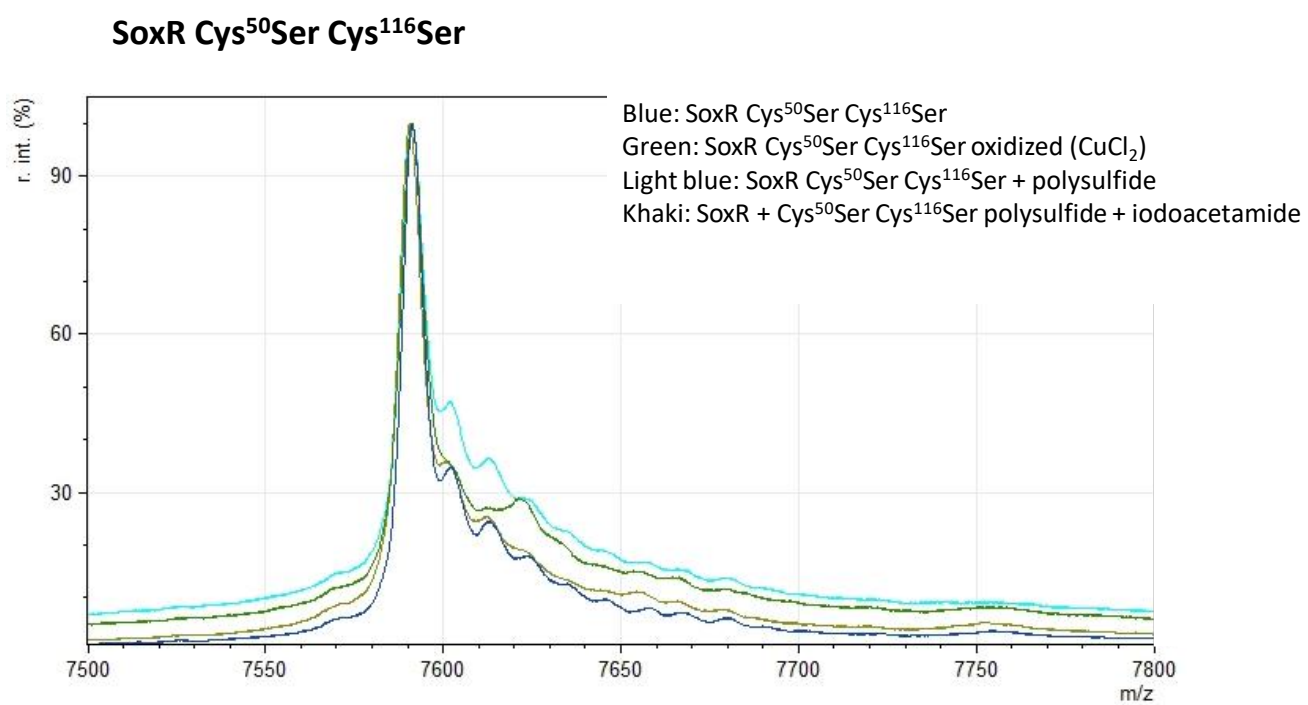
**SoxR Cys<sup>50</sup>Ser**

Masses are shown for the  $mz/2$  species. Mass spectra obtained for the  $\text{CuCl}_2$  oxidized protein were almost identical to those for the as isolated protein and are not shown for clarity.



Masses are shown for the  $mz/2$  species. Mass spectra obtained for the  $\text{CuCl}_2$  oxidized protein were almost identical to those for the as isolated protein and are not shown for clarity.





Masses are shown for the  $mz/2$  species.

Table S1. Strains, plasmids and primers

Strains primers or plasmids	Relevant genotype, description or sequence	Reference or source
<b>Strains</b>		
<i>E. coli</i> 10-beta	$\Delta(\text{ara-leu})$ 7697 <i>araD139 fhuA</i> $\Delta\text{lacX74 galK16 galE15 e14-}$ $\phi 80\text{dlacZ}\Delta\text{M15 recA1 relA1 endA1}$	New England Biolabs
<i>E. coli</i> DH5 $\alpha$	<i>nupG rpsL</i> (Str <sup>R</sup> ) <i>rph spoT1</i> $\Delta(\text{mrr-hsdRMS-mcrBC})$ F- $\phi 80\text{lacZ}\Delta\text{M15}$ $\Delta(\text{lacZYA-argF})\text{U169 recA1 endA1 hsdR17}(\text{r}_K, \text{m}_K)$ <i>phoA supE44</i> $\lambda\text{-thi-1 gyrA96}$ <i>relA1</i>	New England Biolabs
<i>E. coli</i> BL21 (DE3)	F- <i>ompT hsdS<sub>B</sub></i> ( <i>r<sub>B</sub></i> , <i>m<sub>B</sub></i> ) <i>gal dcm</i> (DE3)	Novagen
<i>Hyphomicrobium denitrificans</i> $\Delta\text{tsdA}$	Sm <sup>r</sup> , in-frame deletion of <i>tsdA</i> in <i>H. denitrificans</i> Sm200	[8]
<i>Hyphomicrobium denitrificans</i> $\Delta\text{tsdA}$ $\Delta\text{shdrR}$	Sm <sup>R</sup> , in-frame deletion of <i>shdrR</i> (Hden_0682) in <i>H. denitrificans</i> $\Delta\text{tsdA}$	[9]
<i>Hyphomicrobium denitrificans</i> $\Delta\text{tsdA}$ $\Delta\text{soxR}$	Sm <sup>R</sup> , deletion of <i>soxR</i> (Hden_0700) in <i>H. denitrificans</i> $\Delta\text{tsdA}$	This work
<b>Primers</b>		
EMSA-Fr	TTCCCGCCCCGCTTTGGTTT	This work
EMSA_Fr2_Fr	TCAGCGCTCGCCTGGAAGTC	[9]
EMSA_Fr3_Rev	TCTAAGCATCAACATATTCATATCTTTATATATTTTCG	This work
EMSA-Rev	AGGAGTTGCATCCAAAAAGCGTG	[9]
EMSA-Hden_0703/04-fw	GGGTCACCAAATTCGCAAGTCTC	This work
EMSA-Hden_0703/04-rev	ATCACGCCATCTCTCCCGGAA	This work
EMSA-Hden_0699/0698-fw	AATTCCACGGCTCCGCC	This work
EMSA-Hden_0699/0698-rev	TCGACAGCTTGCGGAAATCC	This work
EMSA-sHdrR-LipS1_F	TAGAGCGAGTCTTCAGC	This work
EMSA-sHdrR-LipS1_R	CGGCCCTCTGAGAAAAG	This work
EMSA-LipX-DsrE_F	GACTTCGCCGATCAATCGATC	This work
EMSA-LipX-DsrE_R	TGCCACCTCCCGATATG	This work
rpoB-denitf	AGGACGTGTTACCTCGATT	[42]
rpoB-denitr	CGGCTTCGTCAAGGTTCTTC	[42]
SoxT1A_0681_qPCR-Fr	CCCGAGTGATACGATTTCGCA	This work
SoxT1A_0681_qPCR-Rev	CTAAAATGCCGCGGTGATG	This work
LplA_qPCR-Fr	GGCCATGATCGATTTGCACC	This work
LplA_qPCR-Rev	CGAGATAAATTGCACCGCCG	This work
sHdrA_qPCR-Fr	CCGATCACCATTCGGTTCGA	This work
sHdrA_qPCR-Rev	CAATTGTTCCGGGCCGATC	This work

sHdrB2_qPCR-Fr	GACGTGGCCTACTATTCCGG	This work
sHdrB2_qPCR-Rev	CCGCGACGACAGATAGGTTT	This work
LbpA2_qPCR-Fr	GGTTCCAAGAGCAGCCTGAT	This work
LbpA2_qPCR-Rev	TCGTTGATCTCCAGAACCGC	This work
SoxXA_qPCR-Fr	CGGCGCTCATTACCTATCTC	This work
SoxXA_qPCR-Rev	TCGGGGTGTCTTTTTCAGTC	This work
TusA_qPCR-Fr	TCTGACAGTTGATGCCAAGG	This work
TusA_qPCR-Rev	CGTTTCCTCATGTTCAAGCA	This work
CytP450_qPCR-Fr	CAATACGGTTCTCGGACGTT	This work
CytP450_qPCR-Rev	CATTTCGTTTCCTGACGAGGT	This work
SoxT1B (0699)_qPCR-Fr	GCCGCGTCTCAGTAAATA	This work
SoxT1B (0699)_qPCR-Rev	AGCAGAAGACGGCAGATGAT	This work
SoxR_qPCR-Fr	TGAAGCGGACGAGGAAGTAT	This work
SoxR_qPCR-Rev	GAGACTGTGGGCTGGTTGAT	This work
sHdrR_qPCR-Fr	TTAGGAAGTCCGCATCGTCT	This work
sHdrR_qPCR-Rev	GCACTCGTTGCCAATAATA	This work
SoxY_qPCR-Fr	GTTACGCTTGGGACTTTTC	This work
SoxY_qPCR-Rev	GCCAATCGTCACCTTCACTT	This work
P1 fwd up hden_0700	TATACTGCAGGATCAAGGACGTGGTGGCG (PstI)	This work
P2 rev up hden_0700	CTCTCTATCGTTTGGGCTCCATTCTATCCCTCGGTCGC	This work
P3 fwd down hden_0700	GCGCACCAGGGATAGGAATGGAGCCGCAAACGATAGAGAG	This work
P4 rev down hden_0700	GTA <b>CTCTAG</b> AACGAACGCTGCCAGAAGCCC (XbaI)	This work
pET22 SoxR-Strep fw	TATA <b>CA</b> TATGTGGAGCCACCCGAGTTTCGAGAAAGCTAGCTCGGGCATCTTGCCAAAC (NdeI)	This work
pET22 SoxR-Strep rev	TGCT <b>AAGCTT</b> CTATCGTTTGGGCTCGGTT (HindIII)	This work
SoxR C(50)S_fwd	CTGATCCTCTCCCTGCTCGCTG	This work
SoxR C(50)S_rev	CAGGCGGGATTCTGTGAGC	This work
SoxR C(116)S_fwd	GATAAGTTTTCCCGCGAGGAAC	This work
SoxR C(116)S_rev	GTAGATGGCGCCGATGAA	This work
<b>Plasmids</b>		
pET-22b(+)	Ap <sup>r</sup>	Novagen
pET-22b-SoxR-N-Strep	Ap <sup>r</sup> , NdeI/HindIII fragment of amplified SoxR in Nde/HindIII of pET	This work
pET-22b-SoxR C <sup>50</sup> S	Ap <sup>r</sup> , pET-22b-SoxR-N-Strep with a Cys <sup>50</sup> Ser exchange	This work
pET-22b-SoxR C <sup>116</sup> S	Ap <sup>r</sup> , pET-22b-SoxR-N-Strep with a Cys <sup>116</sup> Ser exchange	This work
pET-22b-SoxR C <sup>50</sup> S C <sup>116</sup> S	Ap <sup>r</sup> , pET-22b-SoxR-N-Strep with Cys <sup>50</sup> Ser and Cys <sup>116</sup> Ser exchanges	This work

---

pK18 <i>mobsacB</i> -Tc	Km <sup>r</sup> , Tc <sup>r</sup> pHP45Q Tc tetracycline cassette inserted into pK18 <i>mobsacB</i> using SmaI	[9]
pK18 <i>mobsacB</i> _Tc_Δ <i>soxR</i>	Km <sup>r</sup> , Tc <sup>r</sup> , 1.04 kb SOE PCR fragment implementing deletion of nucleotides 4 to 362 of <i>soxR</i> cloned into pK18 <i>mobsacB</i> -Tc using PstI and XbaI restriction sites	This work

---

### References

8. Koch, T.; Dahl, C. A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds. *ISME J.* **2018**, *12*, 2479-2491.
9. Li, J.; Koch, J.; Flegler, W.; Garcia Ruiz, L.; Hager, N.; Ballas, A.; Tanabe, T.S.; Dahl, C. A metabolic puzzle: consumption of C<sub>1</sub> compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>. *Biochim. Biophys. Acta Bioenerget.* **2022**, *1864*, 148932.
42. Martineau, C.; Mauffrey, F.; Villemur, R. Comparative analysis of denitrifying activities of *Hyphomicrobium nitratorans*, *Hyphomicrobium denitrificans*, and *Hyphomicrobium zavarzinii*. *Appl. Environ. Microbiol.* **2015**, *81*, 5003-5014

---

## The type I sHdr system is co-occurring with thiosulfate oxidation via a truncated Sox system

Tanabe, T. S., & Dahl, C.

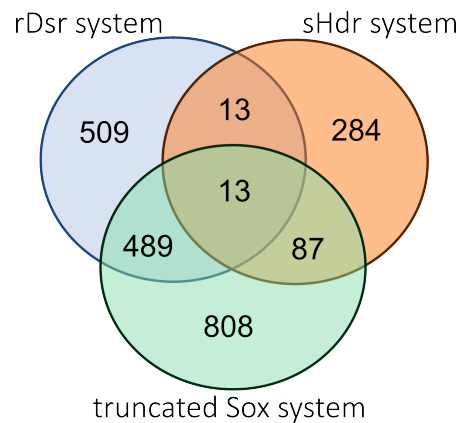
---

### Distribution of the truncated Sox system in sulfur-oxidizing bacteria

There are two different modules for the oxidation of thiosulfate to tetrathionate. Thiosulfate can be oxidized to tetrathionate by either a heme c-containing dehydrogenase (Kurth *et al.* 2016) or a thiosulfate:quinone oxidoreductase (Müller *et al.* 2004). For the complete degradation of tetrathionate the periplasmic tetrathionate hydrolase TetH is currently the only enzyme (Kanao *et al.* 2007, 2021). So far, the existence of this enzyme is only known in acidiphilic sulfur-oxidizing bacteria of the genera *Acidithiobacillus* (Kanao *et al.* 2021) and *Acidiphilium acidophilum* (de Jong *et al.* 1997) and the acidiphilic archaeon *Acidianus ambivalens* (Protze *et al.* 2011). Thus, tetrathionate is a dead-end product in sulfur-oxidizing bacteria such as *H. denitrificans* that do not possess TetH (Koch & Dahl 2018). A second module for the periplasmic thiosulfate oxidation is the Sox system. This periplasmic system is present in a large number of lithotrophic sulfur-oxidizing bacteria either in a complete or truncated form (Friedrich *et al.* 2008). The complete Sox system consists of SoxXA, SoxYZ, SoxB and SoxCD, which are required for the oxidation of thiosulfate to sulfite. In a first step, thiosulfate is bound to the SoxYZ protein by SoxXA in a reaction that covalently attaches the thiosulfate sulfane sulfur to a conserved cysteine of SoxY (Grabarczyk & Berks 2017). The terminal sulfone group is then hydrolyzed by SoxB, releasing sulfate (Sauvé *et al.* 2009). The remaining SoxY-bound sulfane sulfur is then oxidized by the sulfur dehydrogenase SoxCD to the sulfone state (Zander *et al.* 2011), which in turn is again cleaved by SoxB, completing the cycle. Many sulfur-oxidizing bacteria lack the SoxCD component and operate a truncated Sox system (Dahl 2020). Without SoxCD, the SoxY-bound sulfane sulfur is not oxidized to the sulfone state. To regenerate SoxY and to oxidize the sulfane sulfur, many sulfur oxidizing bacteria couple the periplasmic truncated Sox system to the cytoplasmic sHdr system or the reverse Dsr system (Dahl 2020). This coupling of the periplasmic thiosulfate oxidation module and the cytoplasmic sulfur oxidation via the sHdr system has been confirmed in *H. denitrificans* (Li *et al.* 2023). A similar mechanism for the thiosulfate oxidation is likely present in many sulfur oxidizing bacteria including several species of the genera *Aquifex*, *Hyphomicrobium*, *Thioalkalivibrio*, and *Acidithiobacillus*.

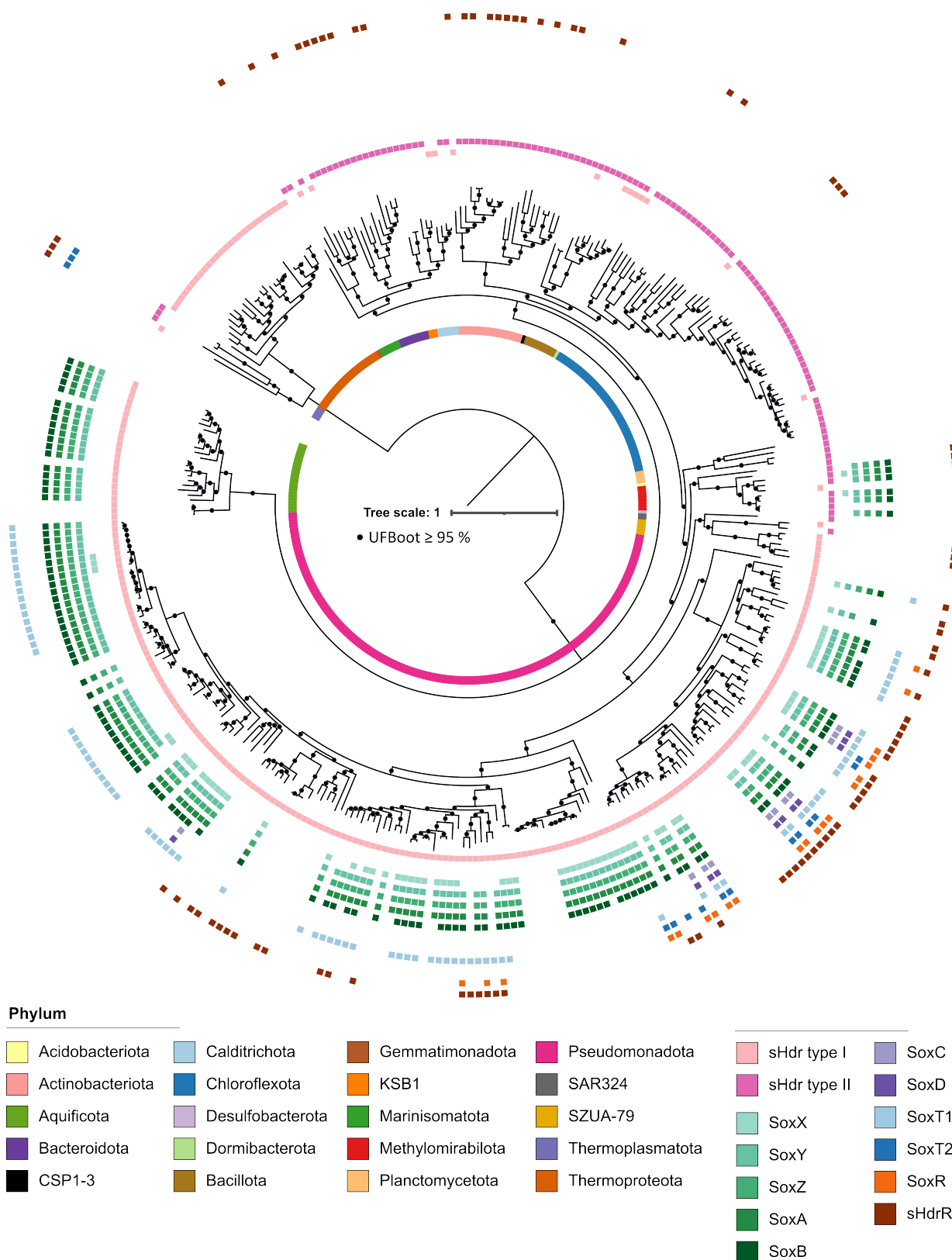
To estimate the number of species that operating a truncated Sox that is linked to an sHdr or reverse Dsr system, the presence of the required genes was analyzed in the Genome Taxonomy Database (GTDB release R207). Since multiple paralogous *sox* genes without a function in thiosulfate oxidation can be present in a single genome (Li *et al.* 2023), a Sox system was only considered as present if at least four of the genes *soxX*, *soxY*, *soxZ* and *soxB* were located in a single gene cluster. For a complete Sox system, *soxCD* was also required to be located in this gene cluster. Gene clusters

lacking *soxCD* were considered to encode for a truncated Sox system. The corresponding genes encoding the Sox, sHdr and Dsr systems were annotated with HMSS2 (Tanabe & Dahl 2023). Using this approach it was possible to assign the presence of each system to the assemblies in the GDTB. Surprisingly, an intersection of the prokaryotes revealed a high number of prokaryotes encoding only for a truncated Sox, sHdr or Dsr system. The large number of bacteria that encode only a truncated Sox system, but not for a reverse Dsr or sHdr system, is particularly conspicuous (Fig. 5). Without SoxCD or the oxidation of the SoxY-bound sulfane sulfur in the cytoplasm, it is puzzling how these bacteria oxidize the sulfane sulfur and regenerate SoxY. Conversely, the use of reduced sulfur compounds other than thiosulfate explains the absence of a truncated Sox system in sulfur oxidizers that use an sHdr system or reverse Dsr system. The alkaliphilic purple sulfur bacterium *Thiorhodospira sibirica* is an example of a bacterium that only encodes for a sHdr system (Tanabe *et al.* 2023c). *T. sibirica* oxidizes sulfide with sulfur globules as an intermediate but not thiosulfate (Bryantseva *et al.* 1999). Accordingly, a *sox* gene cluster is absent in this bacterium. The absence of *soxB* in sulfur-oxidizing bacteria encoding a reverse Dsr system has already been shown for a wide array of bacteria (Meyer *et al.* 2007). In this approach, the *soxB* genes were amplified from a huge number of reference strains (Meyer *et al.* 2007). The predictions of the genetic potential were consistent with the results of this gene amplification-based approach.



**Figure 5: Intersection of genomes encoding a sHdr-, reverse Dsr system and truncated Sox system.**

The Sox systems are mainly present in the phylum Pseudomonadota and rarely found in any other phylum (Li *et al.* 2023b). In addition, the type I and type II sHdr systems were not equally distributed among the bacterial and archaeal phyla (Tanabe *et al.* 2023c). Given the large number of prokaryotes only encoding a sHdr system, a possible correlation of the truncated Sox system with the type I or type II sHdr system was tested. For this approach, the presence of each *sox* gene (*soxX*, *soxY*, *soxZ*, *soxA*, *soxB*, *soxC*, *soxD*) in a single gene cluster was mapped to the presence of a type I and type II sHdr system. The presence of the transporters SoxT1 and SoxT2 and the regulators SoxR and sHdrR were added to this analysis. While the regulators had to be present in the *sox* or *shdr* gene cluster respectively, the genes *soxT1* and *soxT2* only had to be present in the genome to be detected. This was due to the different reliability of HMS-S-S in detecting of the corresponding proteins (Tanabe & Dahl 2022).



**Figure 6: Presence of *sox* genes and genes for the regulators SoxR and sHdrR in bacteria and archaea encoding a *shdr* gene cluster.** All Sox proteins, except SoxT1/SoxT2, had to be encoded in a syntenic gene cluster to be detected as present. The *sox* genes of Ectothiorhodospirales are not syntenic and were therefore assigned manually as described by Berben *et al.* (2019). The gene for the regulator sHdrR had to be located in the *shdr* gene cluster to be considered as present. The tree was calculated as described in Tanabe *et al.* (2023c).

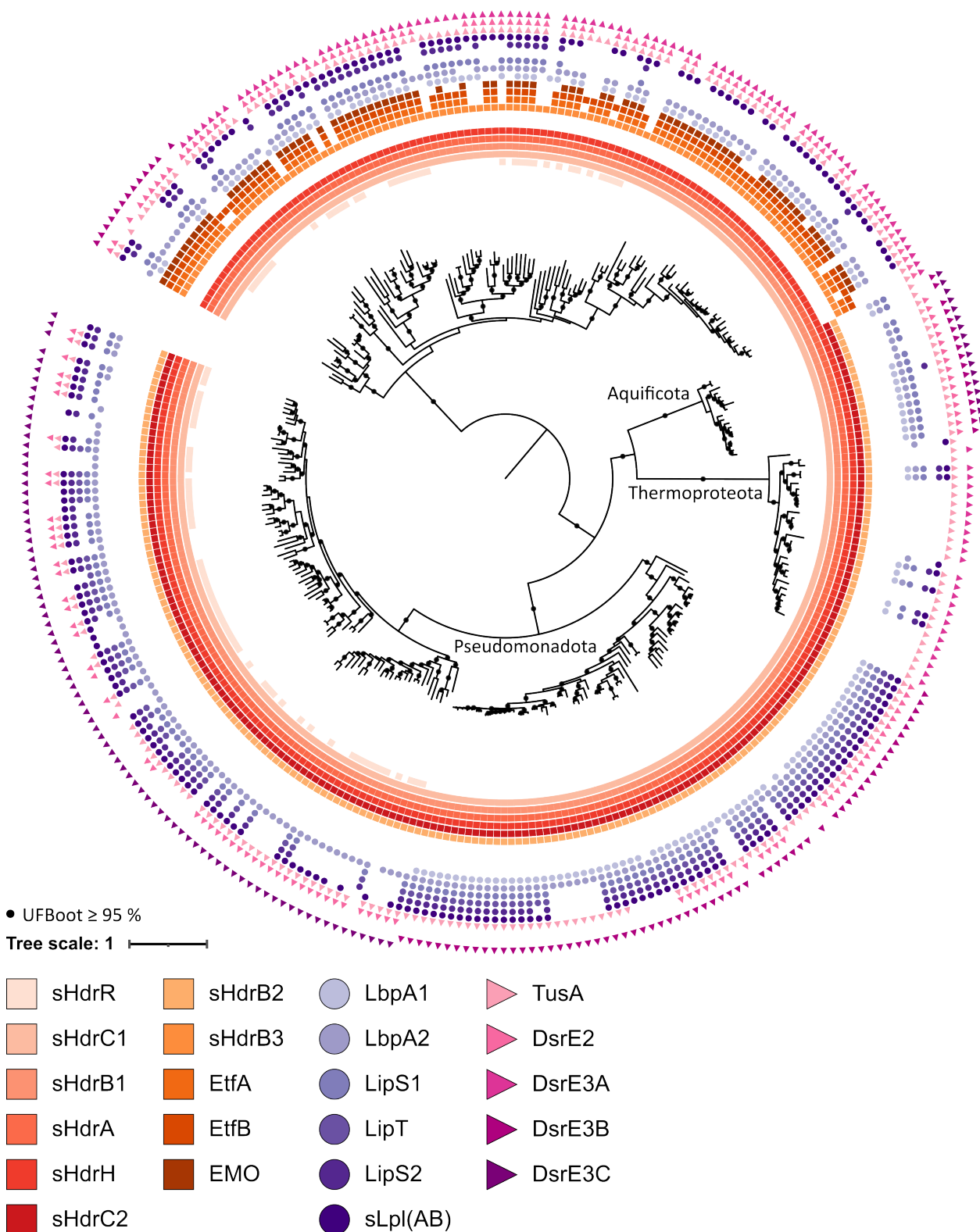
The analysis showed a clear correlation of the truncated Sox system with the type I sHdr system (Fig. 6). The majority of species encoding a type I gene cluster also encoded for a truncated Sox system. Phylogenetically, all these species were members of the Pseudomonadota and Aquificota. Conversely, only a minor fraction of species encoding a type II *shdr* gene cluster also encoded a truncated Sox system. SoxT transporters were only sparsely present in genomes with the former *shdr* genes and completely absent in the latter. At least SoxT1 or SoxT2 was present in genome when regulator SoxR was present. The same was not true for *shdrR*, which did not always occur with *soxT*. This supports a possible function in *sox* gene regulation, but contradicts its general a sulfur compound importing role for the sHdr system. Nevertheless, the sulfur import may be facilitated by this transporter in at least some sulfur oxidizers. SoxT was also not strictly co-occurring with the *shdr* gene. Thus, the regulator sHdr may not directly receive sulfur compounds from SoxT transporters but from another source.

### The phylogeny of the type I and type II *sHdr* proteins

The type I gene cluster is formed by the core genes *shdrC1B1AHC2B2*. The type II gene cluster differs by a fusion of *shdrC2* and *shdrB2* to *shdrB3*. Therefore the core genes of the type II cluster are *shdrC1B1AHB3*. In addition, two electron-transferring flavoproteins (EtfAB) and an ETF:menaquinone oxidoreductase (EMO) are usually encoded in the type II gene clusters. From an evolutionary perspective the type I system could have preceded the type II sHdr system. In this scenario, the type II system evolved through the gene fusion of *shdrB2C2* and the acquaintance of the genes for EtfAB and EMO. Alternatively, both types could have evolved from a common ancestor. To get a first insight into the evolution of these two gene clusters, the phylogeny of the concatenated core proteins was calculated. The proteins encoded by the *shdr* gene cluster were mapped onto the resulting phylogenetic tree. Thus, the tree also displays the composition of the gene cluster.

The resulting tree showed that the sHdr complexes fall into two very well distinct clades, that were separated by a long branch (Fig. 7). Each of these clades was formed by the all the representatives of the type I and the type II sHdr complex, respectively. Rooting with minimal ancestor deviation pointed towards a root between these two clades (Tria *et al.* 2017). This root positioning was confirmed with the concatenated mHdrABC sequences derived from methanogenic archaea and sulfate-reducing bacteria (Ramos *et al.* 2015, Wagner *et al.* 2017). The root position indicates a scenario in which the type I *shdr* cluster did not evolve from the type II *shdr* gene cluster or vice versa. Rather both evolved from a heterodisulfide reductase of a methanogenic archaeon or sulfate-reducing prokaryote. A scenario in which the sHdr system evolved twice from two completely different origins seems to be less likely as the lipoate-binding proteins and lipoate biosynthesis are present in both gene clusters. Therefore, the lipoate-binding protein could have been a component of the sHdr system before the differentiation into type I and type II. In this evolutionary scenario, the truncated Sox system was acquired after this differentiation.





**Figure 7: sHdr complex protein phylogeny** Proteins that are encoded in the *shdr* gene cluster are indicated as squares, triangles and circles. Functional annotation was performed with HMSS2 (Tanabe & Dahl 2023). Species that encoded for more than one *shdr* gene cluster and assemblies for which not all core components were available were removed from the analysis. Each sHdr protein type was first individually aligned with MAFFT (Katoh & Standley 2013). Alignments were then concatenated to create a supermatrix, which was trimmed with BMGE (Criscuolo & Gribaldo 2010). Phylogeny was calculated with iqTree (Minh *et al.* 2020).



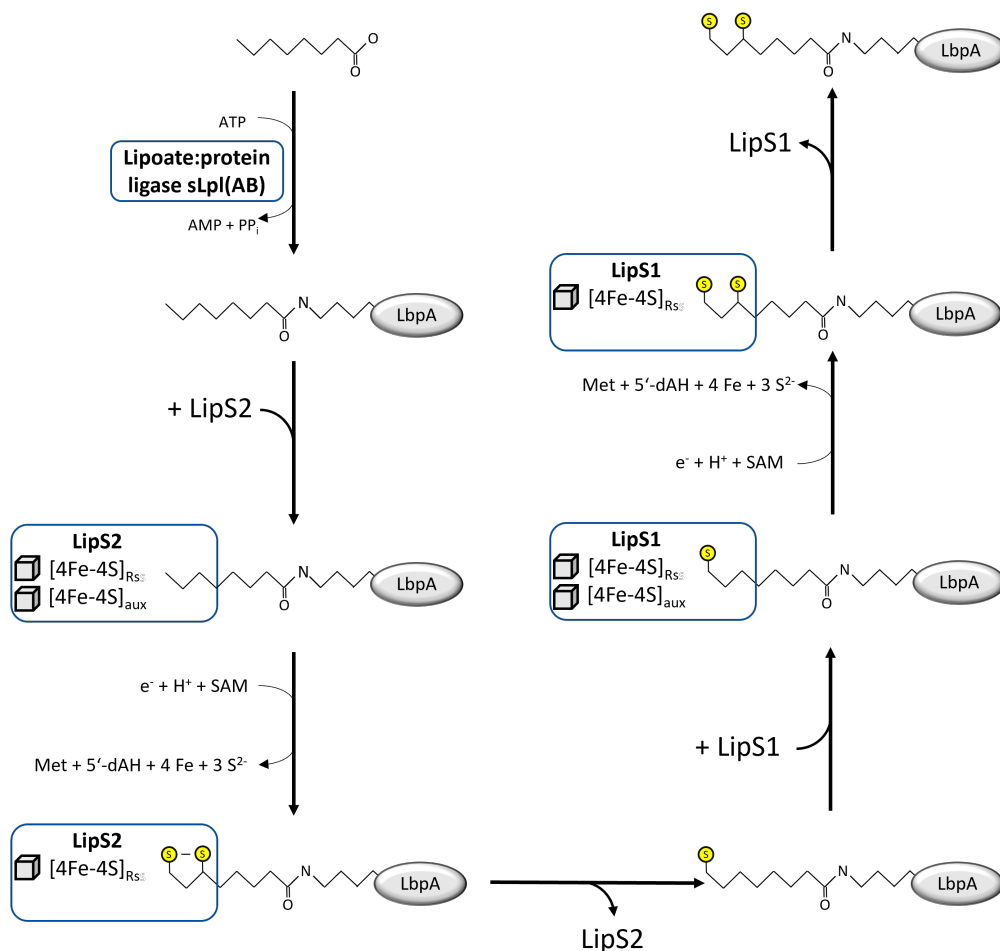
## V Discussion

The general aim of this work was to elucidate the lipoate assembly pathway for LbpA and the enzymatic mechanism of sulfur oxidation of the sHdr system. It has been possible to characterize a novel lipoate biosynthesis pathway that present in nearly all sulfur oxidizing prokaryotes encoding a sHdr system. Furthermore, this pathway was present in a large number of bacteria and archaea (Tanabe *et al.* 2023b). This was made possible by the development of several tools for the prediction of genes and proteins relates to sulfur metabolism (Tanabe & Dahl 2022, 2023). These tools were also applied to show the diversity of sulfur-oxidizing prokaryotes and the co-occurrence of the complex sulfur-oxidizing processes (Kümpel *et al.* 2024). The existence of two *shdr* gene cluster types was described and their association with the thiosulfate oxidizing Sox system was analyzed (Li *et al.* 2023b). The necessity of a truncated Sox system for the thiosulfate oxidation via the type I sHdr system was confirmed in *H. denitrificans* (Li *et al.* 2023). Biochemically, the proposed sulfane sulfur transfer to the sHdr system via several sulfur transferases including rhodanese Rhd442, TusA and DsrE3 proteins was elucidated. The transferases of *H. denitrificans*, *Thioalkalivibrio sp.* K90 mix, *T. sibirica* and *A. aeolicus* were biochemically characterized and their sulfur transferase activity was demonstrated. Sulfite production was confirmed for the reaction catalyzed by sHdr-LbpA. These new results extend the model of sulfur transfer for the sHdr systems these bacteria (Tanabe *et al.* 2023c).

### 1 A novel pathway for lipoate assembly in bacteria

The lipoate-binding protein is of crucial importance for the dissimilatory sulfur oxidation via the sHdr system in *H. denitrificans* (Cao *et al.* 2018). In contrast to all other known lipoate-requiring proteins LbpA was not lipoylated by the enzymes of the canonical lipoate synthesis machinery. The de novo lipoate assembly system remained enigmatic until it was possible to elucidate a new lipoate biosynthesis pathway formed by the lipoate synthases LipS1 and LipS2 and the *bona fide* lipoate:protein ligase sLpl(AB) (Fig. 8) (Tanabe *et al.* 2023b). Simultaneously, the catalytic mechanisms of the lipoate synthases LipS1 and LipS2 from *Thermococcus kodakarensis* became available complementing these results (Neti *et al.* 2022). The results from the *T. kodakarensis* proteins and *H. denitrificans* allowed the formulation of a model for the lipoate assembly on LbpA.

Unlike many other cofactors, lipoate is only functionally active when covalently attached to its cognate enzyme. Indeed, lipoylation of LbpA was possible in vitro with a *bona fide* lipoate:protein ligase sLpl(AB) encoded in the *shdr* gene cluster and free lipoate (Cao *et al.* 2018). Indeed, lipoylation via the *bona fide* sLpl(AB) with free lipoate can also result in lipoylated LbpA (Cao *et al.* 2018). A prerequisite for the efficient generation of holo-LbpA by the ligase activity alone would be an environment deficient of free octanoate and especially rich in free lipoate (Christensen & Cronan 2009). This is due to the broad substrate spectrum of lipoate:protein ligases that can catalyze the attachment of lipoate, octanoate and similar compounds (Green *et al.* 1995, Jin *et al.* 2022). The presence of octanoate would therefore result in non-functional octanoylated LbpA in the absence of an enzyme to catalyze lipoyl assembly. The lipoate biosynthetic pathway described here compensates for this deficiency by the presence of the lipoate synthases LipS1 and LipS2.



**Figure 8: Model of the LipS1/S2-involving lipoate assembly pathway starting with free octanoate as established for *H. denitrificans*** SAM, S-adenosylmethionine; 5'-dAH, 5'-deoxyadenosylhomocysteine; Met, methionine

For the novel lipoate biosynthesis pathway, sLpl(AB) is a key enzyme, because it has a specificity for LbpA proteins, but does not recognize the structurally and phylogenetically related GcvH proteins. Conversely, the target proteins of sLpl(AB) are also not recognized by ligases of the canonical pathway (Cao *et al.* 2018b, Tanabe *et al.* 2023b). De novo biosynthesis is initiated by the octanoylation of LbpA (Fig. 8) (Tanabe *et al.* 2023b). It therefore requires the octanoylation activity of sLpl(AB), making the lipoylation activity a beneficial side effect. The mechanism of this reaction has been elucidated by several structures of homologous ligases. The activation of octanoate with ATP and the transfer to the cognate enzyme are both catalyzed by the same enzyme (Cronan 2016). The same mechanism applies to the attachment of lipoate. In a first reaction, the free octanoate is activated with ATP to form octanoyl-AMP, releasing pyrophosphate. The hydrophobic octanoyl-AMP is then buried in a substrate-binding pocket of the LplA domain located at the interface between the LplA and LplB domains (Fujiwara *et al.* 2005). In this conformation only the phosphate and ribose of the AMP moiety are accessible, while the octanoyl moiety is shielded from the solvent (Kim *et al.* 2005). A conformational change between the two domains then enables the

transfer reaction (Cronan 2016). LplB does not participate in the in the transfer reaction, but is only required for the formation of octanoyl-AMP (Christensen & Cronan 2009). In a second, reaction the octanoyl moiety of the octanoyl-AMP intermediate is transferred to apoproteins, releasing AMP (Fujiwara *et al.* 2005). In the process, octanoate gets covalently attached to the  $\epsilon$ -amino group of a lysine residue via an amide bond (Spalding & Prigge 2010).

After octanoate is covalently attached to LbpA by sLpl(AB), the sulfur atoms are consecutively inserted by the two lipoate synthases LipS1 and LipS2. These two enzymes belong to the radical S-adenosylmethionine (SAM) protein family, like all known lipoate synthases (Broderick *et al.* 2014, Neti *et al.* 2022). Each of these enzymes contains two [4Fe-4S] iron-sulfur clusters. The coordinating residues and cysteine motifs of these proteins have been described previously (Kümpel *et al.* 2024). One of the [4Fe-4S] clusters is required to reductively cleave SAM to generate a 5'-deoxyadenosyl 5'-radical for the abstraction of hydrogen from the octanoyl moiety (Neti *et al.* 2022). After the hydrogen atom is abstracted, a single sulfur atom is donated from the second [4Fe-4S] to replace the cleaved hydrogen atom. For the lipoate assembly on the octanoyl moiety, LipS2 inserts the first sulfur atom at position C8. The protein bound 8-mercaptooctanoyl is then released from the lipoate synthase as an intermediate (Jin, Hachisuka, Sato, Fujiwara & Atomi 2020, Neti *et al.* 2022, Tanabe *et al.* 2023b). The lipoate synthase LipS1 binds to this intermediate and inserts the sulfur at position C6 completing the lipoate assembly. The electron for the radical formation is proposed to be donated by LipT, a putative FAD-binding oxidoreductase, in an NADPH-dependent reaction (Kümpel *et al.* 2024). This function was considered due common link between *lipT*, *lipS1* and *lipS2*, which are commonly found in a single collinear syntenic gene cluster (Tanabe *et al.* 2023b, Kümpel *et al.* 2024).

The described novel pathway was present in a large number of bacteria and archaea. With respect to LbpA, the novel biosynthesis pathway could be detected in almost all organisms with the sHdr system (Tanabe *et al.* 2023b). In most of these genomes, the LbpA and at least the ligases were located together in the sHdr gene cluster (Fig. 7). Especially in the bacterial domain, the canonical was present at the same time. These are probably not redundant machineries with respect to the proteins these pathway lipoylate. As demonstrated in *H. denitrificans* both pathways function in parallel without interfering or replacing each other (Tanabe *et al.* 2023b). In the absence of sLpl(AB), LbpA is not lipoylated in vivo by the canonical de novo biosynthesis pathway formed by LipB and LipA. The lack of octanoyltransferases to octanoylate LbpA is has also been observed for the octanoyltransferase LipM (Cao *et al.* 2018). This inability of the canonical pathway to modify LbpA is not limited to the *H. denitrificans* protein. The same substrate-specific activity was observed for the with LipA/LipM biosynthesis machinery of *B. subtilis*, which was unable to octanoylate three of the five GcvH-like proteins from *A. aeolicus*. In vitro experiments showed that this was due to a lack of octanoyl transfer from LipM to the three GcvH-like proteins (aq\_402, aq\_944, aq\_1657) (Cao *et al.* 2018b). These three proteins did not only lack the lipoate-relay activity, but also lacked glycine cleavage activity. Therefore, they were considered to be either evolutionary relics with no current function, or involved in another pathway (Cao *et al.* 2018b). In fact, comparative genomics and the analysis with HMS-S-S revealed that these three GcvH-like proteins proteins are actually lipoate-binding proteins (Tanabe & Dahl 2022). LbpA (aq\_402) was later co-purified with other

components of the sHdr system (Tanabe *et al.* 2023c). The novel lipoate assembly pathway is therefore likely to act specifically on all LbpA proteins in the sHdr systems I and II. Also, the LbpA is essentially depending on the biosynthesis via the sLpl(AB), LipS1 and LipS2.

## 2 The sulfur oxidation in the type I sHdr pathway

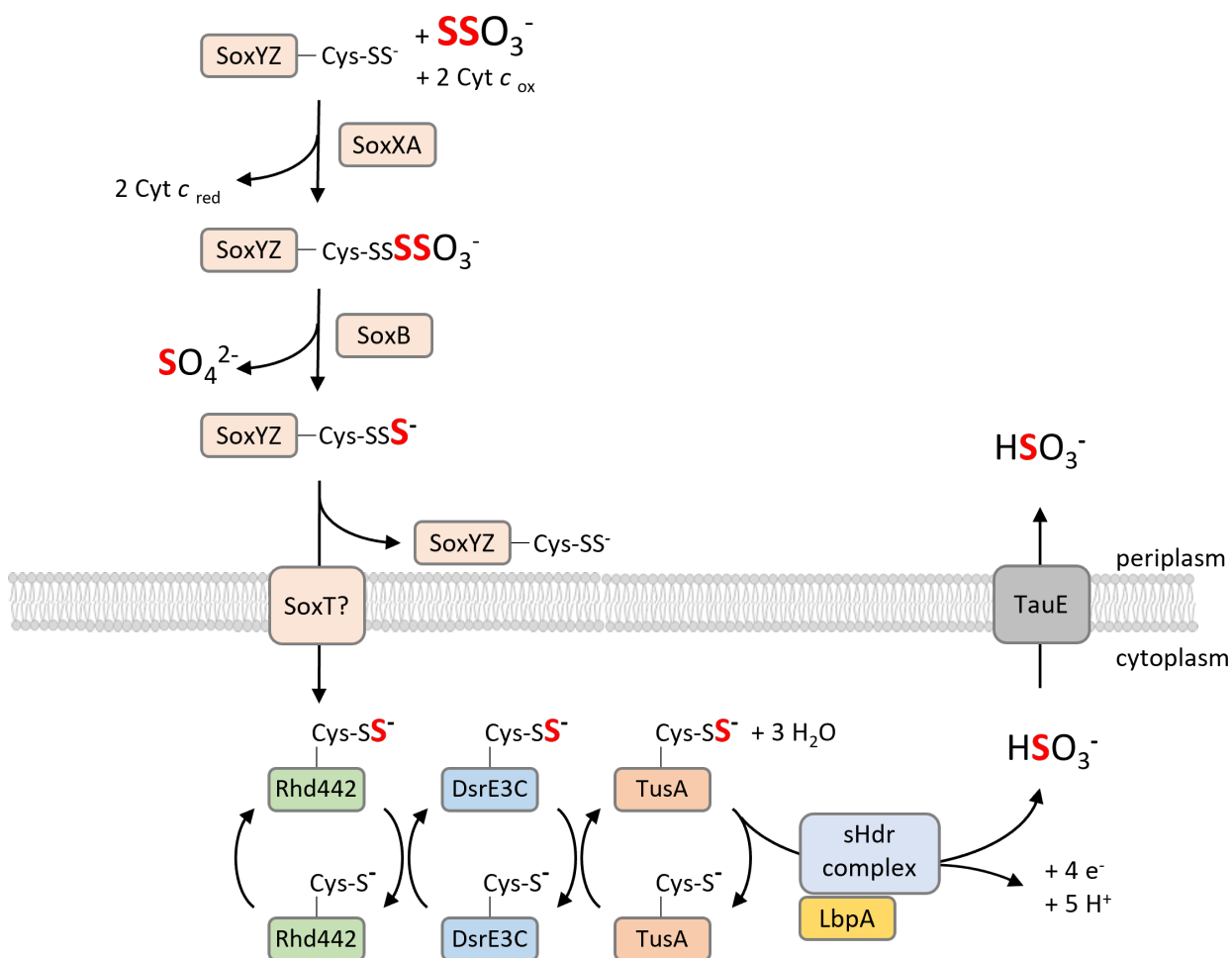
Since the postulation of the model for the sHdr pathway, it has been possible to characterize several of the processes, greatly expanding the knowledge of the dissimilatory sulfur oxidation. An extended model for the sHdr pathway of *H. denitrificans* has been formulated by combining these recent results (Fig. 9).

In the original postulation, thiosulfate is oxidized through the periplasmic truncated Sox system (Koch & Dahl 2018). In fact, a truncated Sox system has been shown to be present in the majority of bacteria encoding a type I sHdr system using HMS-S-S and HMSS2 (Tanabe & Dahl 2022, 2023). An actual involvement in thiosulfate oxidation was demonstrated for the truncated Sox system of *H. denitrificans* by the individual deletion of *soxXA* and *soxYZ*. The generated strain, was unable to oxidize thiosulfate, although the sHdr system was fully functional. This confirmed the function of the truncated Sox system in the early stages of the sHdr pathway. For the sulfur transfer between the periplasmic SoxY-bound sulfane sulfur and the cytoplasmic sHdr system the involvement of a YeeE/YedE class transporter was proposed (Koch & Dahl 2018). Two genes for transporters of this class were located in the *shdr* gene cluster (Li *et al.* 2023b). Comparative genomics and phylogenetic analysis revealed a significant sequence similarity to *soxT* transporters (Tanabe & Dahl 2022). These transporters participate in the regulation of the Sox system likely in association with the regulator SoxR (Lahiri *et al.* 2006). Analogously, at least one of the SoxT transporters and the regulators sHdr and SoxR regulate the transcription of the *shdr* genes in *H. denitrificans* (Li *et al.* 2023b).

Several sulfur transferases were thought to shuttle sulfur in protein-bound form to the sHdr complex of *H. denitrificans* (Koch & Dahl 2018, Tanabe *et al.* 2019). TusA and DsrE3C were shown to sequentially transfer sulfane sulfur (Tanabe *et al.* 2023c). In addition, a rhodanese Rhd442 that is encoded near the *shdr* gene cluster in *H. denitrificans* was characterized. In vitro, this transfer Rhd442 transferred sulfane sulfur to DsrE3C and with lower efficiency to TusA. Recombinant DsrE3C transferred sulfane sulfur unidirectionally to TusA. Characterization of the recombinant DsrE3B and TusA proteins from *Thioalkalivibrio sp.* K90mix and *T. sibirica* revealed a similar unidirectional sulfane sulfur transfer from DsrE3B to TusA (Tanabe *et al.* 2023c). Sulfur transfer from DsrE3 to TusA, but not in the opposite direction, has been reported for the in vitro transfer of *Metallosphaera cuprina* DsrE3A and TusA. The in vitro sulfur characterization of the sulfur transferases was complemented by a genetic approach in *H. denitrificans* which confirmed the essential function of DsrE3C in the sHdr system. Strains of *H. denitrificans* strain lacking *dsrE3C* or the catalytically active cysteine were demonstrated to have a significantly reduced thiosulfate oxidation rate. Co-purification experiments coupled to mass spectrometry in *A. aeolicus* cell extracts revealed a protein-protein interaction of TusA, DsrE3-type proteins. The likely acceptor protein of this sulfur relay system is LbpA, which was co-purified with TusA from *A. aeolicus* cell extracts (Tanabe *et al.* 2023c). This experiment also indicated an interaction of LbpA, TusA and the components

of the sHdr complex. Although these protein-protein interactions need to be confirmed with recombinant proteins, this is the first observation of an interaction of LbpA with the sHdr complex (Tanabe *et al.* 2023c).

The transferred sulfane sulfur is oxidized to sulfite, which is exported into the periplasm (Li *et al.* 2023). Despite this oxidation, little is known about the reaction catalyzed by the LbpA-sHdr complex. A first milestone towards this goal was reached when a sHdr complex with at least five domains was purified from *A. aeolicus* (Boughanemi *et al.* 2016). Crystallization of the sHdrA protein from *H. denitrificans* and modeling of the remaining sHdr proteins sHdrC1B1C2B2 resulted in a similar structural composition of the sHdr as it has been described for the mHdr complexes (Ernst *et al.* 2021). In contrast to the electron bifurcating mHdr complex, a flavin-based electron bifurcating reaction mechanism was not considered as plausible for the sHdr complex (Ernst *et al.* 2021). Based on this limited structural information, several preliminary reaction mechanisms have been hypothesized (Ernst *et al.* 2021, Kümpel *et al.* 2024). The lipoate cofactor acts in these models either as a sulfane sulfur presenting component that cycles between oxidized and reduced states (Kümpel *et al.* 2024) or simply as a redox switch component (Ernst *et al.* 2021).



**Figure 9: Model for the sHdr system of *H. denitrificans*** The sHdr complex is formed by the components sHdrC1, sHdrB1, sHdrA, sHdrH, sHdrC2 and sHdrB2. DsrE3C is depicted as a monomer, not as the homotrimer as isolated from heterologous production. LbpA, lipoate-binding proteins.

### 3 The type II sHdr system

Very little is known about the sHdr pathway encoded by the type II *shdr* gene cluster. It was first recognized as a variation of the type I sHdr gene cluster in *Sulfobacillus* sp., *Kyrpidia tusciae* and some other species. In these studies, a sulfur-oxidizing function was already suggested based on the similarity to the *shdr* cluster of *Acidithiobacillus* sp. and *H. denitrificans* (Justice *et al.* 2014, Cao *et al.* 2018). A broad assessment of the sulfur oxidation capacity in over 70000 genomes with HMS-S-S and HMSS2 showed the wide distribution of type II gene clusters (Tanabe & Dahl 2022, 2023). In total the type II system was detected in assemblies from eighteen bacterial and archaeal phyla (Tanabe *et al.* 2023c, Kumpel *et al.* 2024). Genomes encoding this system differed from the type I system in that they did not typically encode any Sox system (Fig. 6). The majority of prokaryotes with a type II system can probably not utilize thiosulfate. Several of the genomes that encoded for a type II system were derived from described sulfur-oxidizing bacteria. Their importance for dissimilatory sulfur oxidation can be inferred from the physiology of these species e.g. *Sulfobacillus acidophilus* or *Chloroherpeton thalassium*. *Sulfobacillus acidophilus* is a moderately thermophilic mineral sulfide oxidizing bacterium that was isolated from industrial mining sites and thermal springs. Growth occurs autotrophically and mixotrophically on pyrite (FeS), ferrous iron and elemental sulfur. Autotrophic growth on elemental sulfur acidified the environment, suggesting sulfuric acid or sulfuric acid production (Norris *et al.* 1996). Genetically *S. acidophilus* encodes a single type II *shdr* gene cluster, but no *sox* gene cluster, *soxT1*, *soxT2*, type I *shdr* gene cluster or *dsr* genes (Tanabe *et al.* 2023c). The green sulfur bacterium *Chloroherpeton thalassium* is another example of a well-known obligate phototrophic sulfur-oxidizing bacterium that depends on sulfide and carbon dioxide for growth (Gibson *et al.* 1984). Sulfide cannot be substituted by thio-sulfate for growth. External sulfur globules are produced as an intermediate and consumed at a low rate to generate sulfate (Gibson *et al.* 1984). Unlike the sulfur-oxidizing green sulfur bacteria of the genera *Prosthecochloris*, *Chlorobium*, and *Chlorobaculum*, *Chloroherpeton thalassium* does not encode for a dissimilatory sulfur-oxidizing system such as the Sox system or the reverse Dsr system (Frigaard & Dahl 2008). The sulfide:quinone oxidoreductases sulfide and periplasmic flavocytochrome c sulfide dehydrogenase FccAB are present in *C. thalassium* (Frigaard & Dahl 2008). Both enzymes catalyze the oxidation of sulfide oxidation to zerovalent sulfur, whereas oxidation to sulfate requires additional catalytic reactions. The ability of *C. thalassium* to oxidize sulfur to sulfate is probably due to a type II *shdr* gene cluster. The number of sulfur-oxidizing bacteria and archaea may therefore be much larger than previously thought.

### 4 The evolution of the lipocate biosynthesis and sHdr pathway

The novel lipocate biosynthesis pathway likely evolved in the archaeal domain (Tanabe *et al.* 2023b). Phylogeny of the key protein sLpl(AB) showed, that this enzyme evolved from a bipartite LplAB lipocate:protein ligase. These ligases have been recognized as an ancient types of ligases (Christensen & Cronan 2009). In support to the existence of this lipocate biosynthesis pathway in archaea the hyperthermophilic archaeon *Thermococcus kodakarensis* encodes for the proteins LipS1, LipS2 and a lipocate:protein ligase related to sLpl(AB) (Jin, Hachisuka, Sato, Fujiwara & Atomi 2020, Jin *et al.*



2022). De novo lipoate biosynthesis in *T. kodakarensis* was shown to depend on the presence of the lipoate:protein ligase, which had significant octanoylation activity (Jin *et al.* 2022). The activity of the lipoate synthases LipS1 and LipS2 was also demonstrated for the *T. kodakarensis* proteins, which inserted sulfur atoms on an artificial octanoylated polypeptide (Jin, Hachisuka, Sato, Fujiwara & Atomi 2020). A lipoate assembly pathway that initially scavenges free octanoate is consistent with a scenario in which lipoate:protein ligases evolved from octanoyltransferases (Braakman & Smith 2014, Cronan 2016, Tanabe *et al.* 2023b). As hypothesized by Cronan (2016), the early lipoate:protein ligases would indeed have scavenged octanoate before the lipoate cofactor biosynthesis, or the cofactor itself existed.

The knowledge of the evolution of lipoate biosynthesis and the gene cluster composition of the type I and type II gene clusters allows to hypothesize on the evolution of the sHdr system.

The lipoate assembly system probably evolved before the sHdr system in an ancient prokaryotic lineage. The comprehensive analysis of their distribution among archaea and bacteria and the phylogeny of the small transferase family also suggest the existence of lipoate-requiring proteins and lipoate assembly machinery in a lineage related to the last common ancestor (Tanabe *et al.* 2023b). Indeed, ancient prokaryotes are proposed to operate a glycine cleavage system (Braakman & Smith 2012, 2014). The lipoate synthases could have already existed, as the radical SAM family is one of the most versatile and most ancient protein families (Broderick *et al.* 2014). The required iron-sulfur cluster biosynthesis can even be traced back to the last universal common ancestor (Garcia *et al.* 2022b). With all prerequisites present, it cannot not be excluded that lipoate biosynthesis and lipoate are as old as methanogenesis. Comparing the geological evidence, methanogenesis already existed more than 3,46 Ga ago (Ueno *et al.* 2006), while the oldest evidence of an active sulfur cycle is dated to 3,2 Ga ago (Nabhan *et al.* 2020). The lipoate biosynthesis and the mHdr complex as it is required by all current forms of methanogenesis, is therefore likely predate the sHdr complex (Garcia *et al.* 2022).

The sHdr system may have evolved from the genetic combination of a lipoate biosynthesis machinery including LipS1, LipS2 and a lipoate:protein ligase, with a LbpA-like or GcvH-like protein and a methanogenic heterodisulfide reductase in an ancient methanogenic archaeon. During evolution, heterodisulfide reductases have undergone several architectural reorganization giving rise to long and short versions of the Hdr components and even fusion proteins (Appel *et al.* 2021). The composition of the sHdr complex could therefore be a result of a duplication of the B and C subunits combined with the loss of domains in the A subunit. Since all these components are essentially present in the type I and the type II, where *shdrC2* and *shdrB2* are fused, such a composition has likely existed before the differentiation into the two types (Fig. 7). An assumption that is supported by the rooting position of the sHdr phylogeny. When and from which lineage this evolution started remains to be determined, as it has not been possible to determine the closest relative of the sHdr complex.

## 5 Perspectives and outlook

Despite the remarkable progress that has been made in understanding bacterial sulfur oxidation, there are still several aspects of this process that warrant further investigation. These include the exploration of the evolutionary origin of sulfur oxidation and the discovery of new sulfur oxidation mechanisms.

Open questions about the novel lipoate biosynthesis pathway described here are the substrate range of the ligases and the general mechanism of differentiation between GcvH and LbpA that allows simultaneous activity of the canonical and novel lipoate biosynthesis pathways.

The architecture of the type I sHdr complex is still unresolved and the exact catalytic activity remains to be experimentally validated. In particular, the position of the catalytic active site, the role of non-cubane and cubane iron-sulfur clusters, and the composition of the complex require further investigation. For these questions, the purification of the active sHdr complex is the most urgent task and the greatest challenge. The components of the complex that interact with LbpA also need to be investigated. First indications of a physical interaction have been made, but need to be confirmed with purified proteins. Another question is how the sulfur transferred by the sulfur relay system is bound to LbpA. Knowledge of the type II sHdr complex is even more limited, and a function in dissimilatory sulfur oxidation remains to be confirmed.

The evolution of the sHdr system is another open question. The development of tools for the detection of the sHdr system is a first step towards solving this question. However, it is necessary to include the heterodisulfide reductases from several metabolic pathways to possibly determine the lineage in which the sHdr complex evolved.

The analysis of the distribution of the truncated Sox system also revealed many bacteria lacking the sHdr or rDsr system. It remains to be investigated whether and how these bacteria can completely oxidize thiosulfate to sulfite.

## References

- Anantharaman, K., Hausmann, B., Jungbluth, S. P., Kantor, R. S., Lavy, A., Warren, L. A., Rappé, M. S., Pester, M., Loy, A., Thomas, B. C. & Banfield, J. F. (2018), 'Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle', *ISME J.* **12**(7), 1715–1728.
- Appel, L., Willistein, M., Dahl, C., Ermler, U. & Boll, M. (2021), 'Functional diversity of prokaryotic HdrA(BC) modules: Role in flavin-based electron bifurcation processes and beyond', *Biochim. Biophys. Acta - Bioenerg.* **1862**(4), 148379.
- Ballesté-Delpierre, C., Fernandez-Orth, D., Ferrer-Navarro, M., Díaz-Peña, R., Odena-Caballol, A., Oliveira, E., Fàbrega, A. & Vila, J. (2017), 'First insights into the pleiotropic role of *vrf* (*yedF*), a newly characterized gene of *Salmonella Typhimurium*', *Sci. Rep.* **7**(1), 15291.
- Bateman, A. & Haft, D. H. (2002), 'HMM-based databases in InterPro', *Brief. Bioinformatics* **3**(3), 236–245.
- Battistuzzi, F. U., Feijao, A. & Hedges, S. B. (2004), 'A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land', *BMC Evol. Biol.* **4**, 44.
- Beinert, H., Holm, R. H. & Münck, E. (1997), 'Iron-sulfur clusters: nature's modular, multipurpose structures', *Science* **277**(5326), 653–659.
- Berben, T., Overmars, L., Sorokin, D. Y. & Muyzer, G. (2019), 'Diversity and Distribution of Sulfur Oxidation-Related Genes in Thioalkalivibrio, a Genus of Chemolithoautotrophic and Haloalkaliphilic Sulfur-Oxidizing Bacteria', *Front. Microbiol.* **10**, 160.
- Boden, R., Kelly, D. P., Murrell, J. C. & Schäfer, H. (2010), 'Oxidation of dimethylsulfide to tetrathionate by *Methylophaga thiooxidans* sp. nov.: a new link in the sulfur cycle', *Environ. Microbiol.* **12**(10), 2688–2699.
- Boughanemi, S., Lyonnet, J., Infossi, P., Bauzan, M., Kosta, A., Lignon, S., Giudici-Orticoni, M.-T. & Guiral, M. (2016), 'Microbial oxidative sulfur metabolism: biochemical evidence of the membrane-bound heterodisulfide reductase-like complex of the bacterium *Aquifex aeolicus*', *FEMS Microbiol. Lett.* **363**(15).
- Braakman, R. & Smith, E. (2012), 'The emergence and early evolution of biological carbon-fixation', *PLoS Comput. Biol.* **8**(4), e1002455.
- Braakman, R. & Smith, E. (2014), 'Metabolic evolution of a deep-branching hyperthermophilic chemoautotrophic bacterium', *PloS One* **9**(2), e87950.
- Brito, J. A., Denkmann, K., Pereira, I. A. C., Archer, M. & Dahl, C. (2015), 'Thiosulfate dehydrogenase (TsdA) from *Allochromatium vinosum*: structural and functional insights into thiosulfate oxidation', *J. Biol. Chem.* **290**(14), 9222–9238.
- Broderick, J. B., Duffus, B. R., Duschene, K. S. & Shepard, E. M. (2014), 'Radical S-adenosylmethionine enzymes', *Chem. Rev.* **114**(8), 4229–4317.
- Bryantseva, I., Gorlenko, V. M., Kompantseva, E. I., Imhoff, J. F., Süling, J. & Mityushina, L. (1999), '*Thiorhodospira sibirica* gen. nov., sp. nov., a new alkaliphilic purple sulfur bacterium from a Siberian soda lake', *Int. j. syst. bacteriol.* **49 Pt 2**, 697–703.

## References

- Buckel, W. & Thauer, R. K. (2013), 'Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation', *Biochim. Biophys. Acta* **1827**(2), 94–113.
- Burrichter, A., Denger, K., Franchini, P., Huhn, T., Müller, N., Spiteller, D. & Schleheck, D. (2018), 'Anaerobic Degradation of the Plant Sugar Sulfoquinovose Concomitant With H<sub>2</sub>S Production: *Escherichia coli* K-12 and *Desulfovibrio* sp. Strain DF1 as Co-culture Model', *Front. Microbiol.* **9**, 2792.
- Canfield, D. E. & Raiswell, R. (1999), 'The evolution of the sulfur cycle', *Am. J. Sci.* **299**(7-9), 697–723.
- Canfield, D. E., Rosing, M. T. & Bjerrum, C. (2006), 'Early anaerobic metabolisms', *Philos. Trans. R. Soc. Lond., B, Biol.* **361**(1474), 1819–34; discussion 1835–6.
- Cao, X. & Cronan, J. E. (2015), 'The *Streptomyces coelicolor* lipoate-protein ligase is a circularly permuted version of the *Escherichia coli* enzyme composed of discrete interacting domains', *J. Biol. Chem.* **290**(11), 7280–7290.
- Cao, X., Hong, Y., Zhu, L., Hu, Y. & Cronan, J. E. (2018b), 'Development and retention of a primordial moonlighting pathway of protein modification in the absence of selection presents a puzzle', *Proc. Natl. Acad. Sci. U.S.A.* **115**(4), 647–655.
- Cao, X., Koch, T., Steffens, L., Finkensieper, J., Zigann, R., Cronan, J. E. & Dahl, C. (2018), 'Lipoate-binding proteins and specific lipoate-protein ligases in microbial sulfur oxidation reveal an atypical role for an old cofactor', *eLife* **7**.
- Caserta, G., Zuccarello, L., Barbosa, C., Silveira, C. M., Moe, E., Katz, S., Hildebrandt, P., Zebger, I. & Todorovic, S. (2022), 'Unusual structures and unknown roles of FeS clusters in metalloenzymes seen from a resonance Raman spectroscopic perspective', *Coord. Chem. Rev.* **452**, 214287.
- Chernyh, N. A., Neukirchen, S., Frolov, E. N., Sousa, F. L., Miroshnichenko, M. L., Merkel, A. Y., Pimenov, N. V., Sorokin, D. Y., Ciordia, S., Mena, M. C., Ferrer, M., Golyshin, P. N., Lebedinsky, A. V., Cardoso Pereira, I. A. & Bonch-Osmolovskaya, E. A. (2020), 'Dissimilatory sulfate reduction in the archaeon '*Candidatus* Vulcanisaeta moutnovskia' sheds light on the evolution of sulfur metabolism', *Nat. Microbiol.* **5**(11), 1428–1438.
- Chicco, D. (2017), 'Ten quick tips for machine learning in computational biology', *BioData Min.* **10**, 35.
- Christel, S., Fridlund, J., Buetti-Dinh, A., Buck, M., Watkin, E. L. & Dopson, M. (2016), 'RNA transcript sequencing reveals inorganic sulfur compound oxidation pathways in the acidophile *Acidithiobacillus ferrooxidans*', *FEMS Microbiol. Lett.* **363**(7).
- Christensen, Q. H. & Cronan, J. E. (2009), 'The *Thermoplasma acidophilum* LplA-LplB complex defines a new class of bipartite lipoate-protein ligases', *J. Biol. Chem.* **284**(32), 21317–21326.
- Christensen, Q. H. & Cronan, J. E. (2010), 'Lipoic acid synthesis: a new family of octanoyltransferases generally annotated as lipoate protein ligases', *Biochemistry* **49**(46), 10024–10036.
- Christensen, Q. H., Hagar, J. A., O'Riordan, M. X. D. & Cronan, J. E. (2011b), 'A complex lipoate utilization pathway in *Listeria monocytogenes*', *J. Biol. Chem.* **286**(36), 31447–31456.
- Christensen, Q. H., Martin, N., Mansilla, M. C., de Mendoza, D. & Cronan, J. E. (2011), 'A novel amidotransferase required for lipoic acid cofactor assembly in *Bacillus subtilis*', *Mol. Microbiol.* **80**(2), 350–363.

- Cicchillo, R. M. & Booker, S. J. (2005), 'Mechanistic investigations of lipoic acid biosynthesis in *Escherichia coli*: both sulfur atoms in lipoic acid are contributed by the same lipoyl synthase polypeptide', *J. Am. Chem. Soc.* **127**(9), 2860–2861.
- Consortium, T. U. (2023), 'UniProt: the Universal Protein Knowledgebase in 2023', *Nucleic Acids Res.* **51**(D1), D523–D531.
- Criscuolo, A. & Gribaldo, S. (2010), 'BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments', *BMC Evol. Biol.* **10**(1), 210.
- Cronan, J. E. (2016), 'Assembly of Lipoic Acid on Its Cognate Enzymes: an Extraordinary and Essential Biosynthetic Pathway', *Microbiol. Mol. Biol. Rev.* **80**(2), 429–450.
- Cronan, J. E., Zhao, X. & Jiang, Y. (2005), 'Function, attachment and synthesis of lipoic acid in *Escherichia coli*', *Adv. Microb. Physiol* **50**, 103–146.
- Cruz, L. M., Trefflich, S., Weiss, V. A. & Castro, M. A. A. (2017), 'Protein Function Prediction', *Methods Mol. Biol.* **1654**, 55–75.
- Czaja, A. D., Beukes, N. J. & Osterhout, J. T. (2016), 'Sulfur-oxidizing bacteria prior to the Great Oxidation Event from the 2.52 Ga Gamahaan Formation of South Africa', *Geology* **44**(12), 983–986.
- Dahl, C. (2015), 'Cytoplasmic sulfur trafficking in sulfur-oxidizing prokaryotes', *IUBMB life* **67**(4), 268–274.
- Dahl, C. (2020), A biochemical view on the biological sulfur cycle, in P. N. L. Lens, ed., 'Environmental Technologies to Treat Sulphur Pollution: Principles and Engineering', IWA Publishing, pp. 55–96.
- Dahl, J.-U., Radon, C., Böhning, M., Nimtz, M., Leichert, L. I., Denis, Y., Jourlin-Castelli, C., Iobbi-Nivol, C., Méjean, V. & Leimkühler, S. (2013b), 'The sulfur carrier protein TusA has a pleiotropic role in *Escherichia coli* that also affects molybdenum cofactor biosynthesis', *J. Biol. Chem.* **288**(8), 5426–5442.
- de Jong, G. A., Hazeu, W., Bos, P. & Kuenen, J. G. (1997), 'Isolation of the tetrathionate hydrolase from *Thiobacillus acidophilus*', *Eur J Biochem* **243**(3), 678–683.
- de Zwart, J., Sluis, J. & Kuenen, J. G. (1997), 'Competition for Dimethyl Sulfide and Hydrogen Sulfide by *Methylophaga sulfidovorans* and *Thiobacillus thioparus* T5 in Continuous Cultures', *Appl. Environ. Microbiol.* **63**(8), 3318–3322.
- Douglas, P., Kriek, M., Bryant, P. & Roach, P. L. (2006), 'Lipoyl synthase inserts sulfur atoms into an octanoyl substrate in a stepwise manner', *Angew. Chem., Int. Ed.* **45**(31), 5197–5199.
- Eddy, S. R. (1998), 'Profile hidden Markov models', *Bioinformatics* **14**(9), 755–763.
- Ehrenfeld, N., Levicán, G. J. & Parada, P. (2013), 'Heterodisulfide Reductase from Acidithiobacilli is a Key Component Involved in Metabolism of Reduced Inorganic Sulfur Compounds', *Adv. Mat. Res.* **825**, 194–197.
- Emms, D. M. & Kelly, S. (2019), 'OrthoFinder: phylogenetic orthology inference for comparative genomics', *Genome Biol.* **20**(1), 238.

## References

- Ernst, C., Kayastha, K., Koch, T., Venceslau, S. S., Pereira, I. A. C., Demmer, U., Ermler, U. & Dahl, C. (2021), 'Structural and spectroscopic characterization of a HdrA-like subunit from *Hyphomicrobium denitrificans*', *FEBS J.* **288**(5), 1664–1678.
- Ferreira, D., Barbosa, A. C. C., Oliveira, G. P., Catarino, T., Venceslau, S. S. & Pereira, I. A. C. (2022), 'The DsrD functional marker protein is an allosteric activator of the DsrAB dissimilatory sulfite reductase', *Proc. Natl. Acad. Sci. U.S.A.* **119**(4).
- Fike, D. A., Bradley, A. S. & Rose, C. V. (2015), 'Rethinking the Ancient Sulfur Cycle', *Annu. Rev. Earth Planet. Sci.* **43**(1), 593–622.
- Filiatrault, M. J., Tomblin, G., Wagner, V. E., van Alst, N., Rumbaugh, K., Sokol, P., Schwingel, J. & Iglewski, B. H. (2013), '*Pseudomonas aeruginosa* PA1006, which plays a role in molybdenum homeostasis, is required for nitrate utilization, biofilm formation, and virulence', *PLoS One* **8**(2), e55594.
- Filipovic, M. R., Zivanovic, J., Alvarez, B. & Banerjee, R. (2018), 'Chemical Biology of H<sub>2</sub>S Signaling through Persulfidation', *Chem. Rev.* **118**(3), 1253–1337.
- Flieder, M., Buongiorno, J., Herbold, C. W., Hausmann, B., Rattei, T., Lloyd, K. G., Loy, A. & Wasmund, K. (2021), 'Novel taxa of Acidobacteriota implicated in seafloor sulfur cycling', *ISME J.* **15**(11), 3159–3180.
- Folz, J., Culver, R. N., Morales, J. M., Grembi, J., Triadafilopoulos, G., Relman, D. A., Huang, K. C., Shalon, D. & Fiehn, O. (2023), 'Human metabolome variation along the upper intestinal tract', *Nat. Metab.* **5**(5), 777–788.
- Friedrich, C. G., Quentmeier, A., Bardischewsky, F., Rother, D., Orawski, G., Hellwig, P. & Fischer, J. (2008), Redox Control of Chemotrophic Sulfur Oxidation of *Paracoccus pantotrophus*, in C. Dahl & C. G. Friedrich, eds, 'Microbial Sulfur Metabolism', Vol. 100, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 139–150.
- Friedrich, C. G., Rother, D., Bardischewsky, F., Quentmeier, A. & Fischer, J. (2001), 'Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism?', *Appl. Environ. Microbiol.* **67**(7), 2873–2882.
- Frigaard, N.-U. & Dahl, C. (2008), Sulfur Metabolism in Phototrophic Sulfur Bacteria, Vol. 54 of *Adv. Microb. Physiol.*, Elsevier, pp. 103–200.
- Fujiwara, K., Toma, S., Okamura-Ikeda, K., Motokawa, Y., Nakagawa, A. & Taniguchi, H. (2005), 'Crystal structure of lipoate-protein ligase A from *Escherichia coli*. Determination of the lipoic acid-binding site', *J. Biol. Chem.* **280**(39), 33645–33651.
- Gabaldón, T. & Koonin, E. V. (2013), 'Functional and evolutionary implications of gene orthology', *Nat. Rev. Genet.* **14**(5), 360–366.
- Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D. & Koonin, E. V. (2021), 'COG database update: focus on microbial diversity, model organisms, and widespread pathogens', *Nucleic Acids Res.* **49**(D1), D274–D281.
- Garber, A. I., Neelson, K. H., Okamoto, A., McAllister, S. M., Chan, C. S., Barco, R. A. & Merino, N. (2020), 'FeGenie: A Comprehensive Tool for the Identification of Iron Genes and Iron Gene Neighborhoods in Genome and Metagenome Assemblies', *Front. Microbiol.* **11**, 37.
- Garcia, P. S., D'Angelo, F., Ollagnier de Choudens, S., Dussouchaud, M., Bouveret, E., Gribaldo, S. & Barras, F. (2022b), 'An early origin of iron-sulfur cluster biosynthesis machineries before Earth oxygenation', *Nat. Ecol. Evol.* **6**(10), 1564–1572.

- Garcia, P. S., Gribaldo, S. & Borrel, G. (2022), 'Diversity and Evolution of Methane-Related Pathways in Archaea', *Annu. Rev. Microbiol.* **76**, 727–755.
- Gibson, J., Pfennig, N. & Waterbury, J. B. (1984), '*Chloroherpeton thalassium* gen. nov. et spec. nov., a non-filamentous, flexing and gliding green sulfur bacterium', *Arch. Microbiol.* **138**(2), 96–101.
- Goddard-Borger, E. D. & Williams, S. J. (2017), 'Sulfoquinovose in the biosphere: occurrence, metabolism and functions', *Biochem. J.* **474**(5), 827–849.
- Grabarczyk, D. B. & Berks, B. C. (2017), 'Intermediates in the Sox sulfur oxidation pathway are bound to a sulfane conjugate of the carrier protein SoxYZ', *PLoS One* **12**(3), e0173395.
- Green, D. E., Morris, T. W., Green, J., Cronan, J. E. & Guest, J. R. (1995), 'Purification and properties of the lipoate protein ligase of *Escherichia coli*', *Biochem. J.* **309** ( Pt 3)(Pt 3), 853–862.
- Gristwood, T., McNeil, M. B., Clulow, J. S., Salmond, G. P. C. & Fineran, P. C. (2011), 'PigS and PigP regulate prodigiosin biosynthesis in *Serratia* via differential control of divergent operons, which include predicted transporters of sulfur-containing molecules', *J. Bacteriol.* **193**(5), 1076–1085.
- Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K., Li, W., Chitsaz, F., Derbyshire, M. K., Gonzales, N. R., Gwadz, M., Lu, F., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Zheng, C., Thibaud-Nissen, F., Geer, L. Y., Marchler-Bauer, A. & Pruitt, K. D. (2018), 'RefSeq: an update on prokaryotic genome annotation and curation', *Nucleic Acids Res.* **46**(D1), D851–D860.
- Haft, D. H., Loftus, B. J., Richardson, D. L., Yang, F., Eisen, J. A., Paulsen, I. T. & White, O. (2001), 'TIGRFAMS: a protein family resource for the functional identification of proteins', *Nucleic Acids Res.* **29**(1), 41–43.
- Haft, D. H., Selengut, J. D. & White, O. (2003), 'The TIGRFAMS database of protein families', *Nucleic Acids Res.* **31**(1), 371–373.
- Hausmann, B., Pelikan, C., Herbold, C. W., Köstlbacher, S., Albertsen, M., Eichorst, S. A., Glavina Del Rio, T., Huemer, M., Nielsen, P. H., Rattei, T., Stingl, U., Tringe, S. G., Trojan, D., Wentrup, C., Woebken, D., Pester, M. & Loy, A. (2018), 'Peatland Acidobacteria with a dissimilatory sulfur metabolism', *ISME J.* **12**(7), 1729–1742.
- Hawkins, T. & Kihara, D. (2007), 'Function prediction of uncharacterized proteins', *J. Bioinform. Comput. Biol.* **5**(1), 1–30.
- Hermes, F. A. M. & Cronan, J. E. (2009), 'Scavenging of cytosolic octanoic acid by mutant LplA lipoate ligases allows growth of *Escherichia coli* strains lacking the LipB octanoyltransferase of lipoic acid synthesis', *J. Bacteriol.* **191**(22), 6796–6803.
- Hernández-Plaza, A., Szklarczyk, D., Botas, J., Cantalapiedra, C. P., Giner-Lamia, J., Mende, D. R., Kirsch, R., Rattei, T., Letunic, I., Jensen, L. J., Bork, P., von Mering, C. & Huerta-Cepas, J. (2023), 'eggNOG 6.0: enabling comparative genomics across 12 535 organisms', *Nucleic Acids Res.* **51**(D1), D389–D394.
- Herrmann, G., Jayamani, E., Mai, G. & Buckel, W. (2008), 'Energy conservation via electron-transferring flavoprotein in anaerobic bacteria', *J. Bacteriol.* **190**(3), 784–791.
- Higgins, K. A., Peng, H., Luebke, J. L., Chang, F.-M. J. & Giedroc, D. P. (2015), 'Conformational analysis and chemical reactivity of the multidomain sulfurtransferase, *Staphylococcus aureus* CstA', *Biochem* **54**(14), 2385–2398.

## References

- Hong, J., Luo, Y., Zhang, Y., Ying, J., Xue, W., Xie, T., Tao, L. & Zhu, F. (2020), 'Protein functional annotation of simultaneously improved stability, accuracy and false discovery rate achieved by a sequence-based deep learning', *Brief. Bioinformatics* **21**(4), 1437–1447.
- Hou, L., Tian, H.-Y., Wang, L., Ferris, Z. E., Wang, J., Cai, M., Older, E. A., Raja, M. R. K., Xue, D., Sun, W., Nagarkatti, P., Nagarkatti, M., Chen, H., Fan, D., Tang, X. & Li, J. (2022), 'Identification and Biosynthesis of Pro-Inflammatory Sulfonolipids from an Opportunistic Pathogen *Chryseobacterium gleum*', *ACS Chem. Biol.* **17**(5), 1197–1206.
- Huwiler, S. G., Löffler, C., Anselmann, S. E. L., Stärk, H.-J., von Bergen, M., Flechsler, J., Rachel, R. & Boll, M. (2019), 'One-megadalton metalloenzyme complex in *Geobacter metallireducens* involved in benzene ring reduction beyond the biological redox window', *Proc. Natl. Acad. Sci. U.S.A.* **116**(6), 2259–2264.
- Huynen, M. A. & Bork, P. (1998), 'Measuring genome evolution', *Proc. Natl. Acad. Sci. U.S.A.* **95**(11), 5849–5856.
- Ikei, M., Miyazaki, R., Monden, K., Naito, Y., Takeuchi, A., Takahashi, Y. S., Tanaka, Y., Mori, T., Ichikawa, M. & Tsukazaki, T. (2023), *Structure and function of a YeeE-YeeD complex for sophisticated thiosulfate uptake.*
- Ikeuchi, Y., Shigi, N., Kato, J.-I., Nishimura, A. & Suzuki, T. (2006), 'Mechanistic insights into sulfur relay by multiple sulfur mediators involved in thiouridine biosynthesis at tRNA wobble positions', *Mol. Cell* **21**(1), 97–108.
- Ishii, Y., Yamada, H., Yamashino, T., Ohashi, K., Katoh, E., Shindo, H., Yamazaki, T. & Mizuno, T. (2000), 'Deletion of the yhhP gene results in filamentous cell morphology in *Escherichia coli*', *Biosci. Biotechnol. Biochem.* **64**(4), 799–807.
- Jin, J., Chen, H., Wang, N., Zhu, K., Liu, H., Shi, D., Xin, J. & Liu, H. (2020), 'A Novel Lipoate-Protein Ligase, Mhp-LplJ, Is Required for Lipoic Acid Metabolism in *Mycoplasma hyopneumoniae*', *Front. Microbiol.* **11**, 631433.
- Jin, J.-Q., Hachisuka, S.-I., Sato, T., Fujiwara, T. & Atomi, H. (2020), 'A Structurally Novel Lipoyl Synthase in the Hyperthermophilic Archaeon *Thermococcus kodakarensis*', *Appl. Environ. Microbiol.* **86**(23).
- Jin, J.-Q., Sato, T., Hachisuka, S.-I. & Atomi, H. (2022), 'A Lipoate-Protein Ligase Is Required for De Novo Lipoyl-Protein Biosynthesis in the Hyperthermophilic Archaeon *Thermococcus kodakarensis*', *Appl. Environ. Microbiol.* **88**(13), e0064422.
- Johnson, D. C., Dean, D. R., Smith, A. D. & Johnson, M. K. (2005), 'Structure, function, and formation of biological iron-sulfur clusters', *Annu. Rev. Biochem.* **74**, 247–281.
- Jørgensen, B. B., Findlay, A. J. & Pellerin, A. (2019), 'The Biogeochemical Sulfur Cycle of Marine Sediments', *Front. Microbiol.* **10**, 849.
- Justice, N. B., Norman, A., Brown, C. T., Singh, A., Thomas, B. C. & Banfield, J. F. (2014), 'Comparison of environmental and isolate *Sulfobacillus* genomes reveals diverse carbon, sulfur, nitrogen, and hydrogen metabolisms', *BMC Genom.* **15**, 1107.
- Kanao, T., Hase, N., Nakayama, H., Yoshida, K., Nishiura, K., Kosaka, M., Kamimura, K., Hirano, Y. & Tamada, T. (2021), 'Reaction mechanism of tetrathionate hydrolysis based on the crystal structure of tetrathionate hydrolase from *Acidithiobacillus ferrooxidans*', *Protein Sci.* **30**(2), 328–338.



- Kanao, T., Kamimura, K. & Sugio, T. (2007), 'Identification of a gene encoding a tetrathionate hydrolase in *Acidithiobacillus ferrooxidans*', *J. Biotechnol.* **132**(1), 16–22.
- Kappler, U. & Dahl, C. (2001), 'Enzymology and molecular biology of prokaryotic sulfite oxidation', *FEMS Microbiol. Lett.* **203**(1), 1–9.
- Kaster, A.-K., Moll, J., Parey, K. & Thauer, R. K. (2011b), 'Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea', *Proc. Natl. Acad. Sci. U.S.A.* **108**(7), 2981–2986.
- Katoh, E., Hatta, T., Shindo, H., Ishii, Y., Yamada, H., Mizuno, T. & Yamazaki, T. (2000), 'High precision NMR structure of YhhP, a novel *Escherichia coli* protein implicated in cell division', *J. Mol. Biol.* **304**(2), 219–229.
- Katoh, K. & Standley, D. M. (2013), 'MAFFT multiple sequence alignment software version 7: improvements in performance and usability', *Mol. Biol. Evol.* **30**(4), 772–780.
- Kessler, D. (2006), 'Enzymatic activation of sulfur for incorporation into biomolecules in prokaryotes', *FEMS Microbiol. Rev.* **30**(6), 825–840.
- Kiene, R. P., Linn, L. J. & Bruton, J. A. (2000), 'New and important roles for DMSP in marine microbial communities', *J. Sea Res.* **43**(3-4), 209–224.
- Kim, D. J., Kim, K. H., Lee, H. H., Lee, S. J., Ha, J. Y., Yoon, H. J. & Suh, S. W. (2005), 'Crystal structure of lipoate-protein ligase A bound with the activated intermediate: insights into interaction with lipoyl domains', *J. Biol. Chem.* **280**(45), 38081–38089.
- Kjeldsen, K. U., Schreiber, L., Thorup, C. A., Boesen, T., Bjerg, J. T., Yang, T., Dueholm, M. S., Larsen, S., Risgaard-Petersen, N., Nierychlo, M., Schmid, M., Bøggild, A., van de Vossenbergh, J., Geelhoed, J. S., Meysman, F. J. R., Wagner, M., Nielsen, P. H., Nielsen, L. P. & Schramm, A. (2019), 'On the evolution and physiology of cable bacteria', *Proc. Natl. Acad. Sci. U.S.A.* **116**(38), 19116–19125.
- Klimke, W., Agarwala, R., Badretdin, A., Chetvernin, S., Ciufu, S., Fedorov, B., Kiryutin, B., O'Neill, K., Resch, W., Resenchuk, S., Schafer, S., Tolstoy, I. & Tatusova, T. (2009), 'The National Center for Biotechnology Information's Protein Clusters Database', *Nucleic Acids Res.* **37**(Database issue), D216–23.
- Koch, T. & Dahl, C. (2018), 'A novel bacterial sulfur oxidation pathway provides a new link between the cycles of organic and inorganic sulfur compounds', *ISME J.* **12**(10), 2479–2491.
- Koonin, E. V., Makarova, K. S. & Wolf, Y. I. (2021), 'Evolution of Microbial Genomics: Conceptual Shifts over a Quarter Century', *Trends Microbiol.* **29**(7), 582–592.
- Kümpel, C., Grosser, M., Tanabe, T. S. & Dahl, C. (2024), 'Fe/S proteins in microbial sulfur oxidation', *Biochim Biophys Acta Mol Cell Res.* .
- Kurth, J. M., Brito, J. A., Reuter, J., Flegler, A., Koch, T., Franke, T., Klein, E.-M., Rowe, S. F., Butt, J. N., Denkmann, K., Pereira, I. A. C., Archer, M. & Dahl, C. (2016), 'Electron Accepting Units of the Diheme Cytochrome c TsdA, a Bifunctional Thiosulfate Dehydrogenase/Tetrathionate Reductase', *J. Biol. Chem.* **291**(48), 24804–24818.
- Lahiri, C., Mandal, S., Ghosh, W., Dam, B. & Roy, P. (2006), 'A novel gene cluster soxSRT is essential for the chemolithotrophic oxidation of thiosulfate and tetrathionate by *Pseudaminobacter salicylatoxidans* KCT001', *Curr. Microbiol.* **52**(4), 267–273.

## References

- Lanz, N. D., Pandelia, M.-E., Kakar, E. S., Lee, K.-H., Krebs, C. & Booker, S. J. (2014), 'Evidence for a catalytically and kinetically competent enzyme-substrate cross-linked intermediate in catalysis by lipoyl synthase', *Biochem* **53**(28), 4557–4572.
- Letunic, I. & Bork, P. (2021), 'Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation', *Nucleic Acids Res.* **49**(W1), W293–W296.
- Li, F., Hinderberger, J., Seedorf, H., Zhang, J., Buckel, W. & Thauer, R. K. (2008), 'Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from *Clostridium kluyveri*', *J. Bacteriol.* **190**(3), 843–850.
- Li, J., Koch, J., Flegler, W., Garcia Ruiz, L., Hager, N., Ballas, A., Tanabe, T. S. & Dahl, C. (2023), 'A metabolic puzzle: Consumption of C1 compounds and thiosulfate in *Hyphomicrobium denitrificans* X<sup>T</sup>', *Biochim. Biophys. Acta - Bioenerg.* **1864**(1), 148932.
- Li, J., Törkel, K., Koch, J., Tanabe, T. S., Hsu, H. Y. & Dahl, C. (2023b), 'In the Alphaproteobacterium *Hyphomicrobium denitrificans* SoxR Serves a Sulfane Sulfur-Responsive Repressor of Sulfur Oxidation', *Antioxidants* **12**(8).
- Li, W., O'Neill, K. R., Haft, D. H., DiCuccio, M., Chetvernin, V., Badretdin, A., Coulouris, G., Chitsaz, F., Derbyshire, M. K., Durkin, A. S., Gonzales, N. R., Gwadz, M., Lanczycki, C. J., Song, J. S., Thanki, N., Wang, J., Yamashita, R. A., Yang, M., Zheng, C., Marchler-Bauer, A. & Thibaud-Nissen, F. (2021), 'RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation', *Nucleic Acids Res.* **49**(D1), D1020–D1028.
- Liu, L.-J., Stockdreher, Y., Koch, T., Sun, S.-T., Fan, Z., Josten, M., Sahl, H.-G., Wang, Q., Luo, Y.-M., Liu, S.-J., Dahl, C. & Jiang, C.-Y. (2014), 'Thiosulfate transfer mediated by DsrE/TusA homologs from acidothermophilic sulfur-oxidizing archaeon *Metallosphaera cuprina*', *J. Biol. Chem.* **289**(39), 26949–26959.
- Löffler, M., Wallerang, K. B., Venceslau, S. S., Pereira, I. A. C. & Dahl, C. (2020b), 'The Iron-Sulfur Flavoprotein DsrL as NAD(P)H:Acceptor Oxidoreductase in Oxidative and Reductive Dissimilatory Sulfur Metabolism', *Front. Microbiol.* **11**, 578209.
- Mahendrarajah, T. A., Moody, E. R. R., Schrempf, D., Szánthó, L. L., Dombrowski, N., Davín, A. A., Pisani, D., Donoghue, P. C. J., Szöllösi, G. J., Williams, T. A. & Spang, A. (2023), 'ATP synthase evolution on a cross-braced dated tree of life', *Nat. Commun.* **14**(1), 7456.
- Mander, G. J., Pierik, A. J., Huber, H. & Hedderich, R. (2004), 'Two distinct heterodisulfide reductase-like enzymes in the sulfate-reducing archaeon *Archaeoglobus profundus*', *Eur J Biochem* **271**(6), 1106–1116.
- Martin, N., Christensen, Q. H., Mansilla, M. C., Cronan, J. E. & de Mendoza, D. (2011), 'A novel two-gene requirement for the octanoyltransfer reaction of *Bacillus subtilis* lipoleic acid biosynthesis', *Mol. Microbiol.* **80**(2), 335–349.
- Mateos, K., Chappell, G., Klos, A., Le, B., Boden, J., Stüeken, E. & Anderson, R. (2023), 'The evolution and spread of sulfur cycling enzymes reflect the redox state of the early Earth', *Sci. Adv.* **9**(27), eade4847.
- McLoughlin, N., Grosch, E. G., Kilburn, M. R. & Wacey, D. (2012), 'Sulfur isotope evidence for a Paleoproterozoic subseafloor biosphere, Barberton, South Africa', *Geology* **40**(11), 1031–1034.
- Mei, R., Kaneko, M., Imachi, H. & Nobu, M. K. (2023), 'The origin and evolution of methanogenesis and Archaea are intertwined', *PNAS nexus* **2**(2), pgad023.

- Meyer, B., Imhoff, J. F. & Kuever, J. (2007), 'Molecular analysis of the distribution and phylogeny of the *soxB* gene among sulfur-oxidizing bacteria - evolution of the Sox sulfur oxidation enzyme system', *Environ. Microbiol.* **9**(12), 2957–2977.
- Miller, A. R., North, J. A., Wildenthal, J. A. & Tabita, F. R. (2018), 'Two Distinct Aerobic Methionine Salvage Pathways Generate Volatile Methanethiol in *Rhodospseudomonas palustris*', *MBio* **9**(2).
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A. & Lanfear, R. (2020), 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Mol. Biol. Evol.* **37**(5), 1530–1534.
- Mishanina, T. V., Libiad, M. & Banerjee, R. (2015), 'Biogenesis of reactive sulfur species for signaling by hydrogen sulfide oxidation pathways', *Nat. Chem. Biol.* **11**(7), 457–464.
- Mock, J., Wang, S., Huang, H., Kahnt, J. & Thauer, R. K. (2014), 'Evidence for a hexaheteromeric methylenetetrahydrofolate reductase in *Moorella thermoacetica*', *J. Bacteriol.* **196**(18), 3303–3314.
- Moran, M. A. & Durham, B. P. (2019), 'Sulfur metabolites in the pelagic ocean', *Nat. Rev. Microbiol.* **17**(11), 665–678.
- Morris, T. W., Reed, K. E. & Cronan, J. E. (1994), 'Identification of the gene encoding lipoate-protein ligase A of *Escherichia coli*. Molecular cloning and characterization of the *lplA* gene and gene product', *J. Biol. Chem.* **269**(23), 16091–16100.
- Morris, T. W., Reed, K. E. & Cronan, J. E. (1995), 'Lipoic acid metabolism in *Escherichia coli*: the *lplA* and *lipB* genes define redundant pathways for ligation of lipoyl groups to apoprotein', *J. Bacteriol.* **177**(1), 1–10.
- Müller, A. L., Kjeldsen, K. U., Rattai, T., Pester, M. & Loy, A. (2015), 'Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases', *ISME J.* **9**(5), 1152–1165.
- Müller, F. H., Bandejas, T. M., Urich, T., Teixeira, M., Gomes, C. M. & Kletzin, A. (2004), 'Coupling of the pathway of sulphur oxidation to dioxygen reduction: characterization of a novel membrane-bound thiosulphate:quinone oxidoreductase', *Mol. Microbiol.* **53**(4), 1147–1160.
- Müller, H., Marozava, S., Probst, A. J. & Meckenstock, R. U. (2020), 'Groundwater cable bacteria conserve energy by sulfur disproportionation', *ISME J.* **14**(2), 623–634.
- Nabhan, S., Marin-Carbonne, J., Mason, P. R. D. & Heubeck, C. (2020), 'In situ S-isotope compositions of sulfate and sulfide from the 3.2 Ga Moodies Group, South Africa: A record of oxidative sulfur cycling', *Geobiology* **18**(4), 426–444.
- Neti, S. S., Sil, D., Warui, D. M., Esakova, O. A., Solinski, A. E., Serrano, D. A., Krebs, C. & Booker, S. J. (2022), 'Characterization of LipS1 and LipS2 from *Thermococcus kodakarensis*: Proteins Annotated as Biotin Synthases, which Together Catalyze Formation of the Lipoyl Cofactor', *ACS bio & med chem Au* **2**(5), 509–520.
- Neukirchen, S., Pereira, I. A. C. & Sousa, F. L. (2023), 'Stepwise pathway for early evolutionary assembly of dissimilatory sulfite and sulfate reduction', *ISME J.* **17**(10), 1680–1692.
- Neukirchen, S. & Sousa, F. L. (2021), 'DiSCo: a sequence-based type-specific predictor of Dsr-dependent dissimilatory sulphur metabolism in microbial data', *Microb. Genom.* **7**(7).
- Nicholls, P. (1975), 'Inhibition of cytochrome c oxidase by sulphide', *Biochem. Soc. Trans.* **3**(2), 316–319.

## References

- Norris, P. R., Clark, D. A., Owen, J. P. & Waterhouse, S. (1996), 'Characteristics of *Sulfobacillus acidophilus* sp. nov. and other moderately thermophilic mineral-sulphide-oxidizing bacteria', *Microbiology* **142** ( Pt 4), 775–783.
- Parks, D. H., Chuvochina, M., Rinke, C., Mussig, A. J., Chaumeil, P.-A. & Hugenholtz, P. (2022), 'GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy', *Nucleic Acids Res.* **50**(D1), D785–D794.
- Peck, S. C., Denger, K., Burchard, A., Irwin, S. M., Balskus, E. P. & Schleheck, D. (2019), 'A glyceryl radical enzyme enables hydrogen sulfide production by the human intestinal bacterium *Bilophila wadsworthia*', *Proc. Natl. Acad. Sci. U.S.A.* **116**(8), 3171–3176.
- Peng, T., Lin, J., Xu, Y.-Z. & Zhang, Y. (2016), 'Comparative genomics reveals new evolutionary and ecological patterns of selenium utilization in bacteria', *ISME J.* **10**(8), 2048–2059.
- Pfeffer, C., Larsen, S., Song, J., Dong, M., Besenbacher, F., Meyer, R. L., Kjeldsen, K. U., Schreiber, L., Gorby, Y. A., El-Naggar, M. Y., Leung, K. M., Schramm, A., Risgaard-Petersen, N. & Nielsen, L. P. (2012), 'Filamentous bacteria transport electrons over centimetre distances', *Nature* **491**(7423), 218–221.
- Posner, M. G., Upadhyay, A., Bagby, S., Hough, D. W. & Danson, M. J. (2009), 'A unique lipoylation system in the Archaea. Lipoylation in *Thermoplasma acidophilum* requires two proteins', *FEBS J.* **276**(15), 4012–4022.
- Pott, A. S. & Dahl, C. (1998), 'Sirohaem sulfite reductase and other proteins encoded by genes at the *dsr* locus of *Chromatium vinosum* are involved in the oxidation of intracellular sulfur', *Microbiology* **144** ( Pt 7), 1881–1894.
- Protze, J., Müller, F., Lauber, K., Naß, B., Mentele, R., Lottspeich, F. & Kletzin, A. (2011), 'An Extracellular Tetrathionate Hydrolase from the Thermoacidophilic Archaeon *Acidianus Ambivalens* with an Activity Optimum at pH 1', *Front. Microbiol.* **2**, 68.
- Quatrini, R., Appia-Ayme, C., Denis, Y., Jedlicki, E., Holmes, D. S. & Bonnefoy, V. (2009), 'Extending the models for iron and sulfur oxidation in the extreme acidophile *Acidithiobacillus ferrooxidans*', *BMC Genom.* **10**, 394.
- Rabus, R., Venceslau, S. S., Wöhlbrand, L., Voordouw, G., Wall, J. D. & Pereira, I. A. C. (2015), 'A Post-Genomic View of the Ecophysiology, Catabolism and Biotechnological Relevance of Sulphate-Reducing Prokaryotes', *Adv. Microb. Physiol* **66**, 55–321.
- Rahn-Lee, L., Gorbatyuk, B., Skovgaard, O. & Losick, R. (2009), 'The conserved sporulation protein YneE inhibits DNA replication in *Bacillus subtilis*', *J. Bacteriol.* **191**(11), 3736–3739.
- Ramos, A. R., Grein, F., Oliveira, G. P., Venceslau, S. S., Keller, K. L., Wall, J. D. & Pereira, I. A. C. (2015), 'The FlxABCD-HdrABC proteins correspond to a novel NADH dehydrogenase/heterodisulfide reductase widespread in anaerobic bacteria and involved in ethanol metabolism in *Desulfovibrio vulgaris* Hildenborough', *Environ. Microbiol.* **17**(7), 2288–2305.
- Rinke, C., Chuvochina, M., Mussig, A. J., Chaumeil, P.-A., Davin, A. A., Waite, D. W., Whitman, W. B., Parks, D. H. & Hugenholtz, P. (2021), 'A standardized archaeal taxonomy for the Genome Taxonomy Database', *Nat. Microbiol.* **6**(7), 946–959.

- Roerdink, D. L., Mason, P. R., Farquhar, J. & Reimer, T. (2012), 'Multiple sulfur isotopes in Paleoproterozoic barites identify an important role for microbial sulfate reduction in the early marine environment', *Earth Planet. Sci. Lett.* **331-332**, 177–186.
- Ruff, J., Denger, K. & Cook, A. M. (2003), 'Sulphoacetaldehyde acetyltransferase yields acetyl phosphate: purification from *Alcaligenes defragrans* and gene clusters in taurine degradation', *Biochem. J.* **369**(Pt 2), 275–285.
- Sauvé, V., Bruno, S., Berks, B. C. & Hemmings, A. M. (2007), 'The SoxYZ complex carries sulfur cycle intermediates on a peptide swinging arm', *J. Biol. Chem.* **282**(32), 23194–23204.
- Sauvé, V., Roversi, P., Leath, K. J., Garman, E. F., Antrobus, R., Lea, S. M. & Berks, B. C. (2009), 'Mechanism for the hydrolysis of a sulfur-sulfur bond based on the crystal structure of the thio-sulfohydrolase SoxB', *J. Biol. Chem.* **284**(32), 21707–21718.
- Sharma, M., Lingford, J. P., Petricevic, M., Snow, A. J. D., Zhang, Y., Järvå, M. A., Mui, J. W.-Y., Scott, N. E., Saunders, E. C., Mao, R., Epa, R., da Silva, B. M., Pires, D. E. V., Ascher, D. B., McConville, M. J., Davies, G. J., Williams, S. J. & Goddard-Borger, E. D. (2022), 'Oxidative desulfurization pathway for complete catabolism of sulfoquinovose by bacteria', *Proc. Natl. Acad. Sci. U.S.A.* **119**(4).
- Shaw, F. L., Mulholland, F., Le Gall, G., Porcelli, I., Hart, D. J., Pearson, B. M. & van Vliet, A. H. M. (2012), 'Selenium-dependent biogenesis of formate dehydrogenase in *Campylobacter jejuni* is controlled by the fdhTU accessory genes', *J. Bacteriol.* **194**(15), 3814–3823.
- Shen, J., Keithly, M. E., Armstrong, R. N., Higgins, K. A., Edmonds, K. A. & Giedroc, D. P. (2015), 'Staphylococcus aureus CstB Is a Novel Multidomain Persulfide Dioxygenase-Sulfurtransferase Involved in Hydrogen Sulfide Detoxification', *Biochem* **54**(29), 4542–4554.
- Shen, J., Peng, H., Zhang, Y., Trinidad, J. C. & Giedroc, D. P. (2016b), 'Staphylococcus aureus sqr Encodes a Type II Sulfide:Quinone Oxidoreductase and Impacts Reactive Sulfur Speciation in Cells', *Biochem* **55**(47), 6524–6534.
- Shen, Y., Buick, R. & Canfield, D. E. (2001), 'Isotopic evidence for microbial sulphate reduction in the early Archaean era', *Nature* **410**(6824), 77–81.
- Shigi, N. (2014), 'Biosynthesis and functions of sulfur modifications in tRNA', *Front. Genet.* **5**, 67.
- Sirko, A., Hryniewicz, M., Hulanicka, D. & Böck, A. (1990), 'Sulfate and thiosulfate transport in *Escherichia coli* K-12: nucleotide sequence and expression of the cystTWAM gene cluster', *J. Bacteriol.* **172**(6), 3351–3357.
- Spalding, M. D. & Prigge, S. T. (2010), 'Lipoic acid metabolism in microbial pathogens', *Microbiol. Mol. Biol. Rev.* **74**(2), 200–228.
- Steinegger, M. & Söding, J. (2017), 'MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets', *Nat. Biotechnol.* **35**(11), 1026–1028.
- Steinegger, M. & Söding, J. (2018), 'Clustering huge protein sequence sets in linear time', *Nat. Commun.* **9**(1), 2542.
- Stockdreher, Y., Sturm, M., Josten, M., Sahl, H.-G., Dobler, N., Zigann, R. & Dahl, C. (2014), 'New proteins involved in sulfur trafficking in the cytoplasm of *Allochromatium vinosum*', *J. Biol. Chem.* **289**(18), 12390–12403.

## References

- Stockdreher, Y., Venceslau, S. S., Josten, M., Sahl, H.-G., Pereira, I. A. C. & Dahl, C. (2012), 'Cytoplasmic sulfurtransferases in the purple sulfur bacterium *Allochromatium vinosum*: evidence for sulfur transfer from DsrEFH to DsrC', *PLoS One* **7**(7), e40785.
- Svetlitsky, D., Dagan, T. & Ziv-Ukelson, M. (2020), 'Discovery of multi-operon colinear syntenic blocks in microbial genomes', *Bioinformatics* **36**(Suppl\_1), i21–i29.
- Tamames, J. (2001), 'Evolution of gene order conservation in prokaryotes', *Genome Biol.* **2**(6), RESEARCH0020.
- Tanabe, T. S., Bach, E., D'Ermo, G., Mohr, M. G., Hager, N., Pfeiffer, N., Guiral, M. & Dahl, C. (2023c), 'A cascade of sulfur transferases delivers sulfur to the sulfur-oxidizing heterodisulfide reductase-like complex', *bioRxiv* p. 2023.12.18.572138.
- Tanabe, T. S. & Dahl, C. (2022), 'HMS-S-S: A tool for the identification of Sulphur metabolism-related genes and analysis of operon structures in genome and metagenome assemblies', *Mol. Ecol. Resour.* .
- Tanabe, T. S. & Dahl, C. (2023), 'HMSS2: An advanced tool for the analysis of sulphur metabolism, including organosulphur compound transformation, in genome and metagenome assemblies', *Mol. Ecol. Resour.* **23**(8), 1930–1945.
- Tanabe, T. S., Grosser, M., Hahn, L., Kämpel, C., Hartenfels, H., Vtulkin, E., Flegler, W. & Dahl, C. (2023b), 'Identification of a novel lipoic acid biosynthesis pathway reveals the complex evolution of lipoate assembly in prokaryotes', *PLoS Biol.* **21**(6), e3002177.
- Tanabe, T. S., Leimkühler, S. & Dahl, C. (2019), 'The functional diversity of the prokaryotic sulfur carrier protein TusA', *Adv. Microb. Physiol* **75**, 233–277.
- Tanaka, Y., Yoshikaie, K., Takeuchi, A., Ichikawa, M., Mori, T., Uchino, S., Sugano, Y., Hakoshima, T., Takagi, H., Nonaka, G. & Tsukazaki, T. (2020), 'Crystal structure of a YeeE/YedE family protein engaged in thiosulfate uptake', *Sci. Adv.* **6**(35), eaba7637.
- Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997), 'A genomic perspective on protein families', *Science* **278**(5338), 631–637.
- Teoh, W. P., Resko, Z. J., Flury, S. & Alonzo, F. (2019), 'Dynamic Relay of Protein-Bound Lipoic Acid in *Staphylococcus aureus*', *J. Bacteriol.* **201**(22).
- Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L.-P. & Mi, H. (2022), 'PANTHER: Making genome-scale phylogenetics accessible to all', *Protein Sci.* **31**(1), 8–22.
- Thorup, C., Schramm, A., Findlay, A. J., Finster, K. W. & Schreiber, L. (2017), 'Disguised as a Sulfate Reducer: Growth of the Deltaproteobacterium *Desulfurivibrio alkaliphilus* by Sulfide Oxidation with Nitrate', *MBio* **8**(4).
- Thume, K., Gebser, B., Chen, L., Meyer, N., Kieber, D. J. & Pohnert, G. (2018), 'The metabolite dimethylsulfoxonium propionate extends the marine organosulfur cycle', *Nature* **563**(7731), 412–415.
- Tomblin, G., Schwingel, J. M., Lapek, J. D., Friedman, A. E., Darrah, T., Maguire, M., van Alst, N. E., Filiatrault, M. J. & Iglewski, B. H. (2013), '*Pseudomonas aeruginosa* PA1006 is a persulfide-modified protein that is critical for molybdenum homeostasis', *PLoS One* **8**(2), e55593.
- Tria, F. D. K., Landan, G. & Dagan, T. (2017), 'Phylogenetic rooting using minimal ancestor deviation', *Nat. Ecol. Evol.* **1**, 193.

- Trivedi, C. B., Stamps, B. W., Lau, G. E., Grasby, S. E., Templeton, A. S. & Spear, J. R. (2020), 'Microbial Metabolic Redundancy Is a Key Mechanism in a Sulfur-Rich Glacial Ecosystem', *mSystems* **5**(4).
- Ueno, Y., Yamada, K., Yoshida, N., Maruyama, S. & Isozaki, Y. (2006), 'Evidence from fluid inclusions for microbial methanogenesis in the early Archaean era', *Nature* **440**(7083), 516–519.
- Venceslau, S. S., Stockdreher, Y., Dahl, C. & Pereira, I. A. C. (2014), 'The "bacterial heterodisulfide" DsrC is a key protein in dissimilatory sulfur metabolism', *Biochim. Biophys. Acta* **1837**(7), 1148–1164.
- Vigneron, A., Cruaud, P., Culley, A. I., Couture, R.-M., Lovejoy, C. & Vincent, W. F. (2021), 'Genomic evidence for sulfur intermediates as new biogeochemical hubs in a model aquatic microbial ecosystem', *Microbiome* **9**(1), 46.
- Wagner, T., Koch, J., Ermler, U. & Shima, S. (2017), 'Methanogenic heterodisulfide reductase (HdrABC-MvhAGD) uses two noncubane 4Fe-4S clusters for reduction', *Science* **357**(6352), 699–703.
- Walker, A., Pfitzner, B., Harir, M., Schaubeck, M., Calasan, J., Heinzmann, S. S., Turaev, D., Rattei, T., Endesfelder, D., zu Castell, W., Haller, D., Schmid, M., Hartmann, A. & Schmitt-Kopplin, P. (2017), 'Sulfonolipids as novel metabolite markers of *Alistipes* and *Odoribacter* affected by high-fat diets', *Sci. Rep.* **7**(1), 11047.
- Wallace, I. M., O'Sullivan, O., Higgins, D. G. & Notredame, C. (2006), 'M-Coffee: combining multiple sequence alignment methods with T-Coffee', *Nucleic Acids Res.* **34**(6), 1692–1699.
- Wang, Y., Wegener, G., Williams, T. A., Xie, R., Hou, J., Tian, C., Zhang, Y., Wang, F. & Xiao, X. (2021), 'A methylotrophic origin of methanogenesis and early divergence of anaerobic multicarbon alkane metabolism', *Sci. Adv.* **7**(27).
- Watanabe, T., Kojima, H., Umezawa, K., Hori, C., Takasuka, T. E., Kato, Y. & Fukui, M. (2019), 'Genomes of Neutrophilic Sulfur-Oxidizing Chemolithoautotrophs Representing 9 Proteobacterial Species From 8 Genera', *Front. Microbiol.* **10**, 316.
- Weinitschke, S., Denger, K., Cook, A. M. & Smits, T. H. M. (2007), 'The DUF81 protein TauE in *Cupriavidus necator* H16, a sulfite exporter in the metabolism of C2 sulfonates', *Microbiology* **153**(Pt 9), 3055–3060.
- Wolf, P. G., Cowley, E. S., Breister, A., Matatov, S., Lucio, L., Polak, P., Ridlon, J. M., Gaskins, H. R. & Anantharaman, K. (2022), 'Diversity and distribution of sulfur metabolic genes in the human gut microbiome and their association with colorectal cancer', *Microbiome* **10**(1), 64.
- Wolfe, J. M. & Fournier, G. P. (2018), 'Horizontal gene transfer constrains the timing of methanogen evolution', *Nat. Ecol. Evol.* **2**(5), 897–903.
- Xiao, J., Zhang, Z., Wu, J. & Yu, J. (2015), 'A brief review of software tools for pangenomics', *Genomics, proteomics & bioinformatics* **13**(1), 73–76.
- Yamashino, T., Isomura, M., Ueguchi, C. & Mizuno, T. (1998), 'The yhhP gene encoding a small ubiquitous protein is fundamental for normal cell growth of *Escherichia coli*', *J. Bacteriol.* **180**(8), 2257–2261.

## References

- Ye, H., Borusak, S., Eberl, C., Krasenbrink, J., Weiss, A. S., Chen, S.-C., Hanson, B. T., Hausmann, B., Herbold, C. W., Pristner, M., Zwirzitz, B., Warth, B., Pjevac, P., Schleheck, D., Stecher, B. & Loy, A. (2023), 'Ecophysiology and interactions of a taurine-respiring bacterium in the mouse gut', *Nat. Commun.* **14**(1), 5533.
- Zander, U., Faust, A., Klink, B. U., de Sanctis, D., Panjikar, S., Quentmeier, A., Bardischewsky, F., Friedrich, C. G. & Scheidig, A. J. (2011), 'Structural basis for the oxidation of protein-bound sulfur by the sulfur cycle molybdohemo-enzyme sulfane dehydrogenase SoxCD', *J. Biol. Chem.* **286**(10), 8349–8360.
- Zhao, X., Miller, J. R., Jiang, Y., Marletta, M. A. & Cronan, J. E. (2003), 'Assembly of the covalent linkage between lipoic acid and its cognate enzymes', *Chem. Biol.* **10**(12), 1293–1302.
- Zhelezinskaia, I., Kaufman, A. J., Farquhar, J. & Cliff, J. (2014), 'Large sulfur isotope fractionations associated with Neoproterozoic microbial sulfate reduction', *Science* **346**(6210), 742–744.
- Zheng, Y., Wang, J., Zhou, S., Zhang, Y., Liu, J., Xue, C.-X., Williams, B. T., Zhao, X., Zhao, L., Zhu, X.-Y., Sun, C., Zhang, H.-H., Xiao, T., Yang, G.-P., Todd, J. D. & Zhang, X.-H. (2020), 'Bacteria are important dimethylsulfoniopropionate producers in marine aphotic and high-pressure environments', *Nat. Commun.* **11**(1), 4658.
- Zhou, Z., Tran, P. Q., Breister, A. M., Liu, Y., Kieft, K., Cowley, E. S., Karaoz, U. & Anantharaman, K. (2022), 'METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks', *Microbiome* **10**(1), 33.



## Acknowledgments

I would especially like to thank my supervisor PD. Dr. Christiane Dahl. Thank you for giving me the opportunity and the freedom to develop in your topic and to shape the project myself in so many areas. Your supervision of your doctoral students is outstanding. Without you, this work would not have been possible

I would like to thank the members of my promotion committee, Prof. Dr. Uwe Deppenmeier, Prof. Dr. Lukas Schreiber and Prof. Dr. Sigurd Höger for reading and evaluating the thesis.

I sincerely thank all the collaboration partners and co-authors of the included publications. Thanks to Alina Ballas, Carolin Kümpel, Elena Bach, Evelyn Vtulkin, Giulia D'Ermo, Hanna Hartenfels, Hsun Yu Hsu, Julian Koch, Kaya Törkel, Leah Hahn, Leon Garcia Ruiz, Marc Gregor Mohr, Martina Grosser, Natalie Hager, Niklas Pfeiffer and Wanda Flegler.

I thank Dr. Toni Köhl and all members of the Core Facility "Protein Synthesis and Bioanalytics" for the excellent cooperation and the really good explanations of mass spectrometry.

I thank Dr. Marianne Guiral and the members of the CNRS in Marseille for their work on the *Aquifex aeolicus* sulfur transferases.

I thank Dr. Malte Petersen from the High Performance Computing and Analytics Lab for the excellent personal advice and introduction to working with high performance computing.

I would like to express my sincere gratitude to my friends and colleagues at the Institut für Mikrobiologie und Biotechnologie who accompanied me in my daily work in the laboratory for the excellent working atmosphere, their moral support and the unforgettable events. I would especially like to thank my research group colleagues. Thanks to Maria Löffler, Julia Lofi, Corvin Ernst, Jingjing Li, Martina Grosser, Carolin Kümpel and Marc Mohr. I was also honored to supervise a large number of excellent undergraduate and graduate students, many of whom became co-authors on this paper. I am grateful for all their contributions and good work. I would also like to thank Benedikt Kniebel and Philipp Deibert for their helpful discussions on programming techniques.

My special thanks go to the Studienstiftung des deutschen Volkes. Thank you for the doctoral forums, academies, skills seminars, writing retreats and financial support. I would also like to thank my fellow doctoral scholarship holders and my friend Mirjam Ackermann for their support and discussion.