Title

A new family of transcriptional regulators of tungstoenzymes and molybdate/tungstate transport

Journal

Environmental Microbiology, 21(2)

ISSN

1462-2912

Authors

Rajeev, L

Garber, ME

Zane, GM

et al.

Publication Date

2019-02-01

DOI

10.1111/1462-2920.14500

Peer reviewed

Environmental Microbiology / Volume 21, Issue 2

Research article | Full Access

A new family of transcriptional regulators of tungstoenzymes and molybdate/tungstate transport

L. Rajeev, M. E. Garber, G. M. Zane, M. N. Price, I. Dubchak, J. D. Wall, P. S. Novichkov, A. Mukhopadhyay, A. E. Kazakov

First published: 10 December 2018
https://doi.org/10.1111/1462-2920.14500

UC-eLinks

Summary

Bacterial genes for molybdenum-containing and tungsten-containing enzymes are often differentially regulated depending on the metal availability in the environment. Here, we describe a new family of transcription factors with an unusual DNA-binding domain related to excisionases of bacteriophages. These transcription factors are associated with genes for various molybdate and tungstate-specific transporting systems as well as molybdo/tungsto-enzymes in a wide range of bacterial genomes. We used a combination of computational and experimental techniques to study a member of the TF family, named TaoR (for tungsten-containing aldehyde oxidoreductase regulator). In *Desulfovibrio vulgaris* Hildenborough, a model bacterium for sulfate reduction studies, TaoR activates expression of aldehyde oxidoreductase *aor* and represses tungsten-specific ABC-type transporter *tupABC* genes under tungsten-replete conditions. TaoR binding sites at *aor* promoter were identified by electrophoretic mobility shift assay and DNase I footprinting. We also reconstructed TaoR regulons in 45 Deltaproteobacteria by comparative genomics approach and predicted target genes for TaoR family members in other Proteobacteria and Firmicutes.

Introduction

Tungsten and molybdenum play an essential role as catalytic components of molybdoenzymes and tungstoenzymes (Johnson *et al*., 1996). Molybdenum-containing enzymes are found in all kingdoms of life, while tungsten-containing enzymes are restricted to prokaryotes, with most of them being found in anaerobic bacteria and archaea (Bevers *et al*., 2009). There are two major types of molybdoenzymes. The first type are molybdenum-iron nitrogenases that contain multinuclear iron (Fe)-Mo cofactor and catalyse the reduction of dinitrogen to ammonia. The second type include all other molybdoenzymes (and all tungstoenzymes) that contain a cofactor, where a single metal ion is coordinated by one or two pyranopterin dithiolene chelates called molybdenum cofactor (MoCo) or tungsten cofactor (TuCo) (Rothery and Weiner, 2015). These enzymes can use only tungsten (e.g. aldehyde oxidoreductase), only molybdenum (e.g. nitrate reductase) (Mendel, 2013), or either of these metals (e.g. dimethyl sulfoxide reductase). All metalloenzymes containing MoCo/TuCo are assigned to four enzyme families based on four different protein folds coordinating different forms of cofactor: sulfite oxidase (SO), xanthine oxidase (XO), dimethylsulfoxide reductase (DMSOR) and tungsten-containing aldehyde:ferredoxin oxidoreductase (AOR) families (Magalon *et al*., 2011).

The ability of prokaryotic organisms to utilize molybdenum and tungsten depends on transmembrane transport of molybdate and tungstate oxyanions into the cell. Various transporters are involved in this process, but specific uptake of molybdate and tungstate is mostly carried out by ATP-binding cassette (ABC) transporters (Zhang et al., 2011). These ABC-type transporters consist of periplasmic, transmembrane and ATPase components. The most widespread bacterial transporter for molybdate and tungstate is ModABC, which consists of ModA (molybdate/tungstate-binding protein), ModB (transmembrane channel subunit) and ModC (ATPase subunit) proteins. Other known tungsten-specific ABC-type transporters include WtpABC and TupABC. WtpABC was found mostly in archaea, and TupABC is present in both anaerobic bacteria and archaea. WtpA, a periplasmic component of Wtp transporter, binds both molybdate and tungstate, but has higher affinity for tungstate, while TupA, a periplasmic component of Tup transporter, has very strong specificity for tungstate (Zhang et al., 2011). Molybdate transport in many bacteria is regulated by the ModE repressor, which typically contains N-terminal DNA-binding domain and two TOBE (Transport-associated Oligonucleotide Binding) domains sensing molybdate. Variations in the domain composition of ModE protein, including an absence of the TOBE domains or their substitution with the PBP_like (Periplasmic Binding Protein like) domain, have been observed in several genomes by Zhang *et al*. (Zhang et al., 2011). They also found fusion proteins that have TOBE domains of ModE fused with various DNA-binding domains. Genomic context analyses of the genes encoding all those ModE-related domain fusions demonstrated their colocalization or even co-transcription with *modABC* genes in many genomes (Zhang et al., 2011). Recently, one such domain fusion, which has a DNA-binding domain similar to XerC/XerD recombinase and two C-terminal TOBE domains, was characterized both computationally and experimentally (Kazakov *et al*., 2013). This protein,

named TunR, activates transcription of the *modA* gene and *modBC* operon in the absence of tungstate in *Desulfovibrio vulgaris* Hildenborough. Molybdenum and tungsten serve dual roles in sulfate-reducing bacteria (SRB), since these metals are cofactors of essential enzymes but they also inhibit SRB growth (Taylor and Oremland, 1979). These oxyanions are adenylated by sulfate adenylyltransferase (SAT), the first enzyme of the sulfate reduction pathway, and this reaction leads to rapid depletion of cellular ATP. In addition, molybdate in SRB is proposed to be an inhibitor of sulfate transport (Newport and Nedwell, 1988). Besides being inhibitors, molybdate and tungstate are required for formate dehydrogenases (FDH) and other key enzymes of energy metabolism in sulfate-reducing Deltaproteobacteria (Zhang and Gladyshev, 2008). The dual role in SRB metabolism dictates a necessity for tight regulation of genes involved in transport, storage and utilization of molybdate and tungstate in SRB.

In this study, we describe a new family of transcriptional regulators of tungstate and molybdenum homeostasis. We have named this family TaoR, based on the novel regulator of tungsten-containing aldehyde:ferredoxin oxidoreductase in *Desulfovibrio alaskensis* G20. A distinctive feature of this family is a combination of Xis-like DNA-binding domain and PBP-like ligand-sensing domain. We have found members of the TaoR family in various groups of bacteria, mainly Proteobacteria and Firmicutes. We reconstructed TaoR regulons in anaerobic Deltaproteobacteria, and experimentally examined TaoR (DVU3193) in *D. vulgaris* Hildenborough, the model organism for sulfate reduction studies.

Results

Dde_0119 protein is a transcription factor

We first used the genome-wide fitness data available in *D. alaskensis* G20 (Kuehl *et al*., 2014) to explore genes involved in W and Mo response. Competitive fitness analysis of pooled barcoded transposon mutants is a powerful tool for high-throughput analysis of gene functions, and genome-wide fitness data across diverse conditions are available for the sulfate-reducing bacterium, *D. alaskensis* G20. We analysed that dataset in order to identify new proteins that may be involved in regulation of tungsten and molybdenum transport and metabolism. A search across a compendium of 98 fitness experiments publicly available in MicrobesOnline (Dehal *et al*., 2010) identified several genes co-fit with the *aor-4* (Dde_2460) gene encoding tungsten-containing aldehyde oxidoreductase (Hensgens *et al*., 1995). Six genes with fitness profiles most similar to that of Dde_2460 (Pearson's correlation coefficient > 0.6) include five enzymes of molybdenum cofactor biosynthesis (Dde_0230, Dde_0249, Dde_0709, Dde_1390, Dde_3228) and a gene encoding Dde_0119 protein, which is annotated as excisionase family DNA-binding protein or putative homologue of nitrogenase iron protein (Supporting Information Table S1). Growth on ethanol gave rise to the most pronounced phenotype of *aor-4* and *moaA* (encoding the first enzyme of MoCo biosynthesis pathway); and some of the lowest

fitness values for mutants in Dde_0119 were from growth on ethanol (Fig. 1). The observed co-fitness of MoCo biosynthesis genes with *aor-4* gene is not surprising, since aldehyde oxidoreductase enzyme uses the tungsten cofactor (Magalon *et al*., 2011). In contrast, a nitrogenase iron protein is required neither for aldehyde oxidoreductase activity nor for growth on ethanol. However, a DNA-binding domain is a typical feature of transcription factors, and transcription factors often have correlated fitness with one of their transcriptional targets (Sagawa *et al*., 2017). Thus, we hypothesize that Dde_0119 may be a regulatory protein and propose to name this regulator TaoR (tungsten-containing aldehyde oxidoreductase regulator).

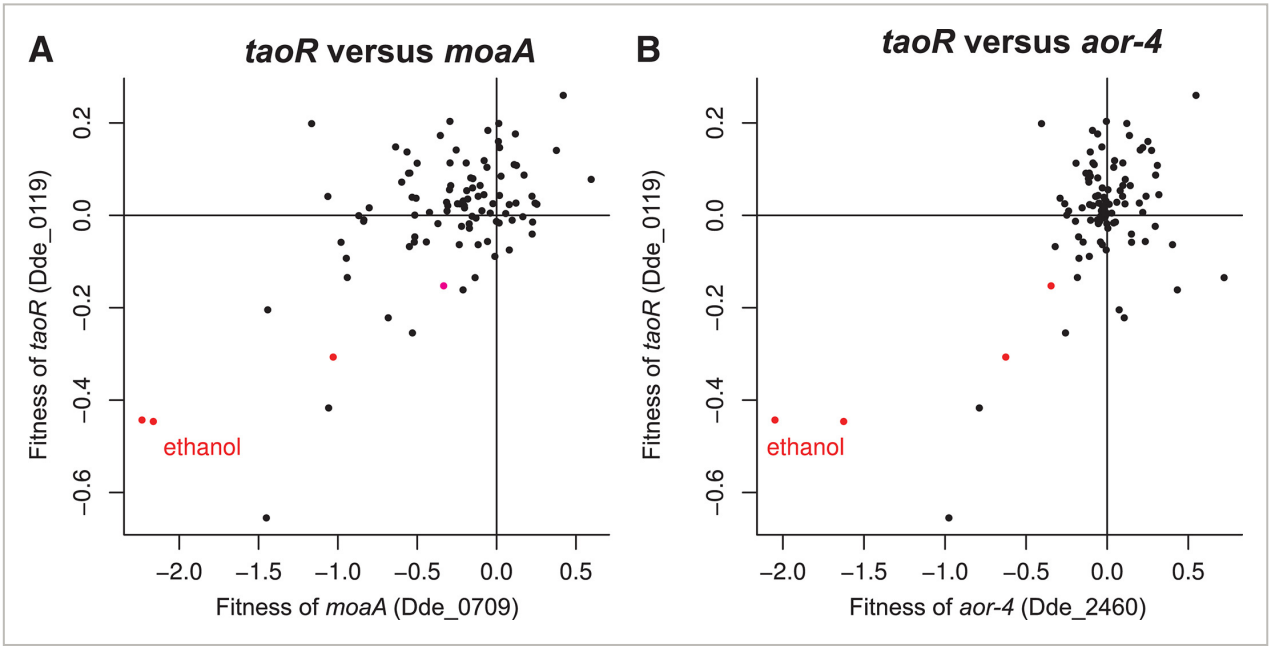

Figure 1

Open in figure viewer | PowerPoint

Comparison of gene fitness for *taoR* (*y* axis) versus *moaA* (panel A, *x* axis) and *aor-4* (panel B, *x* axis) across *D. alaskensis* G20 dataset from the MicrobesOnline database. Fitness values correspond to the normalized $\log_2$ change in abundance of the mutant strains during batch growth under various conditions.

TaoR represents a new family of transcriptional regulators

The putative *taoR* gene (Dde_0119) of *D. alaskensis* G20 is 930 bp long, it has a 427 bp long non-coding upstream region, and is not part of an operon. The TaoR protein (309 amino acids) contains two protein domains annotated in the Pfam database (Finn *et al*., 2016). The N-terminal domain (4–54 aa) is a helix-turn-helix DNA-binding domain (PF12728, also called HTH_17 domain in Pfam), and the C-terminal domain (85–275 aa) is the PBP (periplasmic binding protein) superfamily domain (PF12727 in Pfam database). Since single-component transcriptional regulators often contain both DNA-binding and ligand-binding domains (Balleza

et al., 2009), we speculated that the C-terminal PF12727 domain of the Dde_0119 protein is able to bind a ligand.

Pfam database describes DNA-binding domain PF12728 in several thousand proteins, and a vast majority (>80%) of them consist of only this single domain. There are few examples of such single domain proteins that have been characterized experimentally, and include excisionases Xis from phage P22 (Mattis *et al*., 2008) and Xis2d from the *Bacteroides* conjugative transposon CTnDOT (Wood and Gardner, 2015). These two proteins bind DNA in a sequence-specific manner, but their binding sites are strikingly different. While Xis2d binds a single region of dyad symmetry (Hopp *et al*., 2015), Xis protein from P22 binds cooperatively to four 6 bp direct repeats separated by 4 bp spacers. It was shown that both proteins bend DNA, and the bending induced by the Xis2d protein is important for transcriptional activation of nearby *mob* operon (Hopp *et al*., 2015).

Unlike the DNA-binding domain PF12728, the second domain of TaoR protein, PF12727, is usually a part of multidomain proteins. There are two most common domain architectures for PF12727-containing proteins in the Pfam database, each observed in about 30% of those proteins. The first of the architectures corresponds to the TaoR regulator, that is, the N-terminal DNA-binding domain is PF12728 and the C-terminal domain is PF12727. The second one consists of three domains typical for MoeA enzymes (PF00994, PF03453, PF03454) and the C-terminal domain PF12727. MoeA catalyses insertion of molybdenum or tungsten atom into molybdopterin, but the PF12727 domain seems to be dispensable for this function, as more than 90% of MoeA proteins in Pfam database lack the PF12727 domain.

Thus, putative TaoR protein from *D. alaskensis* G20 is not a solitary case of protein domain fusion, but a representative of a group of proteins. We define this group, which possess N-terminal PF12728 domain and C-terminal PF12727 domain, to be a new family of transcriptional regulators, and designate this family as TaoR.

Pfam database (Finn *et al*., 2016) (v. 29.0) contains 137 TaoR family proteins from the UniProt bacterial reference proteomes, and all but one protein are from bacteria. These proteins belong to 123 bacterial species, among them are 68 species of Firmicutes and 47 species of Proteobacteria. Other bacterial phyla are represented by one or two proteins. Bacterial genomes typically have only one TaoR family gene, except the 12 genomes that have two of them (10 of which are Firmicutes), and one genome, *Clostridium pasteurianum* BC1, that has three (Supporting Information Fig. S1).

We analysed gene neighbourhoods of TaoR family members and found that most of them (117 of 137) contain Mo- or W-associated genes, namely molybdenum or tungsten transporters, molybdoenzymes and molybdate/tungstate-binding proteins consistent with TaoR's putative

role in their regulation. For that purpose, we collected Pfam domain annotations for the surrounding genes (five upstream genes and five downstream genes) and found that 12 Pfam domains are present in 15 or more genome neighbourhoods (Table 1). Five of those domains are involved in transport, with permease (PF00528) and ATPase (PF00005) components of ABC-type transporting systems being the most frequent. Other three transport-specific domains included two periplasmic components of ABC-type transporters (PF13531 and PF12849) and transport-associated oligonucleotide binding domain (PF03459). ABC-type transporters colocalized with TaoR family members are annotated as tungstate-specific transporter Tup and molybdate-specific transporter Mod. The *tupABC* genes were observed most frequently in Proteobacteria and *modABC* genes were mostly observed in Firmicutes (Supporting Information Fig. S1). Short molybdate-binding proteins with a single TOBE domain are similar to *Eubacterium acidaminophilum* protein (Makdessi *et al*., 2004), which has been shown to bind molybdate, tungstate and chromate and implicated in molybdenum/tungsten storage and homeostasis.

Table 1. The most frequently observed protein domains in genome neighbourhoods of TaoR family genes. (Full list of domains is given in Supporting Information Table S2).

Pfam ID	Name	Description	Proteins	Associated functions
PF00528	BPD_transp_1	Permease of ABC-type transporters	93	Mod, Tup
PF00005	ABC_tran	ATPase of ABC-type transporter?	93	Mod, Tup
PF13531	SBP_bac_11	Bacterial extracellular solute-binding protein	63	Tup
PF12849	PBP_like_2	PBP superfamily domain	25	Mod
PF03459	TOBE	TOBE domain	25	
PF00543	P-II	Nitrogen regulatory protein P-II	23	Nif
PF00148	Oxidored_nitro	Nitrogenase component 1 type Oxidoreductase	23	Nif
PF02634	FdhD-NarQ	FdhD/NarQ family	17	Fdh
PF11749	DUF3305	Protein of unknown function (DUF3305)	16	Fdh
PF11748	DUF3306	Protein of unknown function (DUF3306)	16	Fdh
PF00072	Response_reg	Response regulator receiver domain	16	
PF00142	Fer4_NifH	4Fe-4S iron sulfur cluster binding proteins, NifH/frxC family	15	Nif

Functions associated with Pfam domains designated as Mod for ModABC transporter, Tup for TupABC transporter, Nif for dinitrogenase, Fdh for genes associated with formate dehydrogenase.

Other, less represented, Pfam domains are associated with molybdenum-containing enzymes and their accessory proteins: subunits of nitrogenase (PF00148 and PF00142), enzyme maturation protein FdhD (PF02634), regulatory protein NifI (PF00543) and proteins of unknown functions (PF11748 and PF11749). The PF11749 domain was found in putative molybdopterin-guanine dinucleotide biosynthesis protein A, but its function has not been experimentally confirmed.

We searched for putative binding motifs of TaoR family TFs in Proteobacteria and Firmicutes. MEME motif discovery tool identified tandem sequence repeats upstream of *mod* and *tup* operons colocalized with *taoR* genes (Supporting Information Fig. S2). A conserved 40 bp tandem repeat consisting of four 10 bp repeat units was found upstream of 40 of the 55 *mod* operons of Firmicutes. We also identified 20 bp tandem repeat consisting of two 10 bp repeat units upstream of 8 of the 9 *tup* operons from Alphaproteobacteria and Gammaproteobacteria. Similar to TaoR binding motif from Deltaproteobacteria (see below), predicted binding sites in Firmicutes tend to contain short poly-T stretches.

Based on the phylogenetic analysis of TaoR family members combined with domain mappings analysis in genome neighbourhood, we conclude that some of the regulatory genes (mostly the proteobacterial ones) are colocalized with tungstate transporters and FDH genes, while others (in Firmicutes and other taxa) are colocalized either with molybdate transporters or with nitrogenases. Such colocalization patterns are consistent in all studied taxonomic groups and are conserved across distant species, thus suggesting that TaoR family regulators may be involved in regulation of molybdate and tungstate transport genes and molybdoenzyme-encoding genes.

Whole-genome reconstruction of TaoR regulons in Deltaproteobacteria

Unlike other taxa, in Deltaproteobacteria, TaoR family genes have conserved neighbourhoods in 10 out of 18 genomes. The gene fitness profile for *D. alaskensis* G20 described above suggests that TaoR target genes may be scattered throughout the genome, beyond the proximal genes reviewed in the previous section. In order to explore additional targets for putative TaoR regulators, we performed a whole-genome reconstruction of regulons in deltaproteobacterial genomes. We identified a total of 52 TaoR family proteins in

deltaproteobacterial genomes. Phylogenetic analysis distributes these deltaproteobacterial regulators into two distant groups (Fig. 2). In the first group with 6 proteins from myxobacteria, *taoR* family genes are colocalized with *modAB* operon. In the second group with all remaining TaoR family proteins, mostly from sulfate-reducing bacteria, the *taoR* genes are colocalized with genes for a Tup transporter for 17 of the genomes. We reconstructed regulons for this second group of regulators as described in the Experimental procedures section.

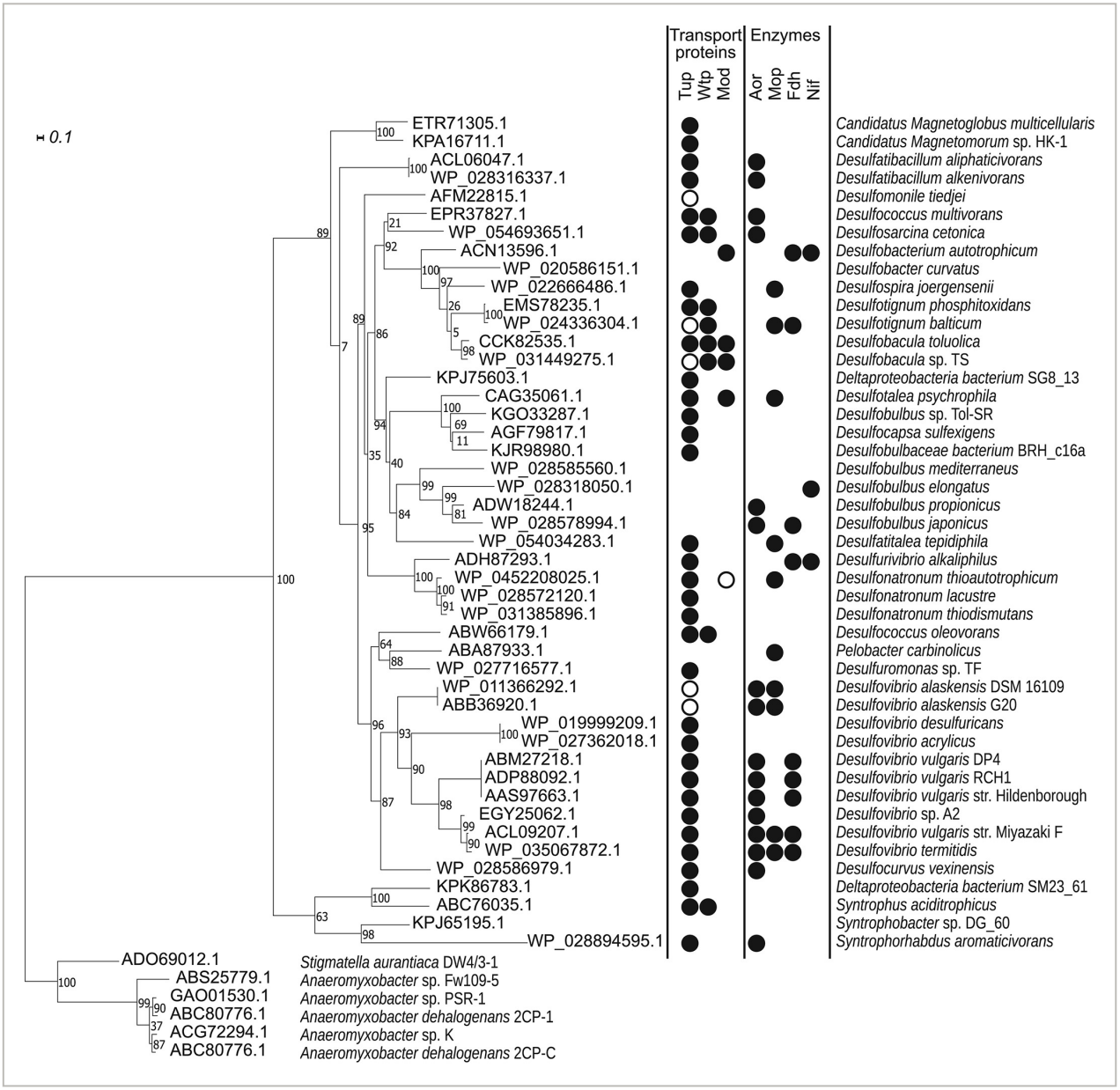

Figure 2

Open in figure viewer | PowerPoint

Phylogenetic tree of TaoR proteins from Deltaproteobacteria. Circles mark members of TaoR regulon. Open circles – separate genes encoding periplasmic components of ABC transporters. Filled circles – operons encoding enzymes or all components of transporters.

We predicted the TaoR binding motif to be a tandem repeat consisting of four 10 bp repeat units, with eight conserved nucleotides TTTGGTTT and two nonconserved positions. We found TaoR binding sites in 44 of the 46 genomes, exceptions being *Desulfobacter curvatus* and *Syntrophobacter* sp. DG_60. For these two organisms, only partial genome sequences are available in GenBank, and TaoR binding sites are possibly located in nonsequenced regions. Most upstream regions had a single predicted site; five exceptions had two predicted sites (Supporting Information Table S3). By searching with an additional position weight matrix for individual repeat units, we identified a variable number of repeat units in TaoR sites (from 3 to 10) (Fig. 3), with sites with 4 or 5 units being the most common (68% of all predicted sites). The highest number of repeat units, 10, was found only in one site upstream of the *tup* operon in *D. vulgaris* Miyazaki.

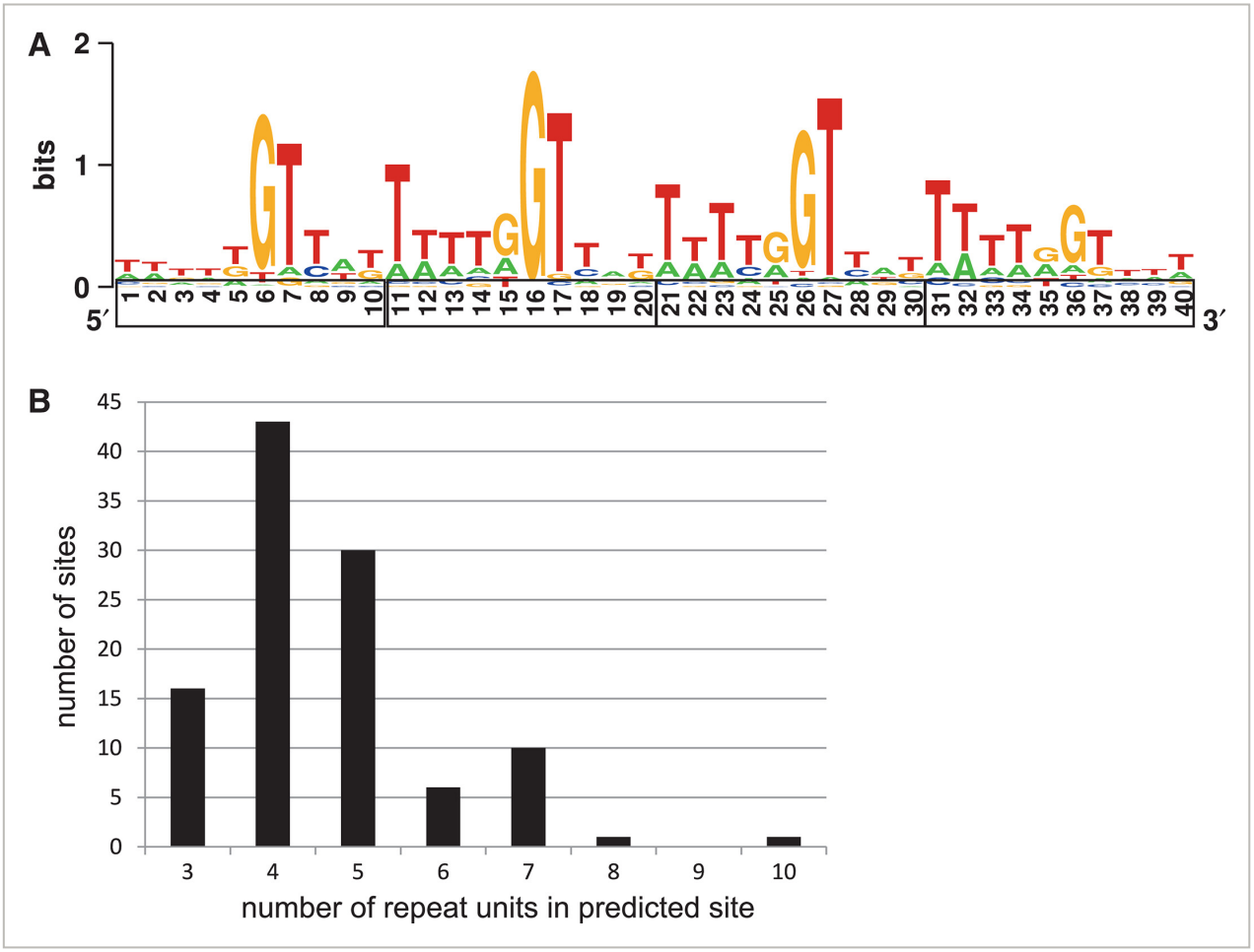

Figure 3

Open in figure viewer | PowerPoint

A. Predicted TaoR-binding motif in Deltaproteobacteria. B. Distribution of repeat unit numbers in predicted TaoR sites from Deltaproteobacteria.

Putative TaoR sites were found upstream of one to five operons per genome, thus TaoR regulons appear to be small (Table 2). The majority of target genes constitute ABC-type transporters and molybdo/tungstoenzymes, but no one target gene was found in all regulons. Instead, we observed a plasticity of TaoR-dependent regulation, with some target operons present in many regulons, and other operons found only in a few of them. Tungstate-specific transporter *tupABC*, the most common target operon, was observed in 33 TaoR regulons (and separate *tupA* genes in 7 other regulons). In 15 of the 33 genomes, the *taoR* gene is part of the *tup* operon. Six regulons included Wtp, another tungstate/molybdate-specific ABC transporter, and five regulons included the *modABC* operon or *modA* gene. Since ModABC transporters can transport both molybdate and tungstate (Aguilar-Barajas *et al*., 2011) it is not surprising that it is regulated by a putative tungstate-sensing regulator.

Table 2. TaoR regulons in Deltaproteobacteria (functional roles) '+' marks TaoR regulon members, '-' marks homologues without TaoR sites, '0' marks absence of corresponding gene/operon, '*' marks pseudogenes.

Genome	*taoR*	TUP	*tupA*	WTP	MOD	*modA*	AOR	MOP	FDH	NIF
Candidatus Magnetoglobus multicellularis str. Araruama	−	−	+	0	0	0	−	0	−	0
Candidatus Magnetomorum sp. HK-1 HK1c0779	−	−	+	0	0	0	−	0	−	0
Deltaproteobacteria bacterium SG8_13 WOR_8-12_2830	+	+	0	0	0	0	−	0	−	0
Deltaproteobacteria bacterium SM23_61 WORSMTZ_43995	+	+	+	−	0	0	−	0	−	0
Desulfatibacillum aliphaticivorans DSM 15576	−	+	0	−	−	0	+	0	−	−
Desulfatibacillum alkenivorans AK-01	−	+	0	−	−	0	+	0	−	−
Desulfatitalea tepidiphila S28bF	+	+	−	+	0	0	−	+	−	−
Desulfobacter curvatus DSM 3379	−	−	0	0	−	0	−	−	0	−
Desulfobacterium autotrophicum HRM2	−	−	0	−	+	0	−	−	+	+
Desulfobacula sp. TS	−	+*	+	+	+	0	−	0	−	0

Genome	*taoR*	TUP	*tupA*	WTP	MOD	*modA*	AOR	MOP	FDH	NIF
Desulfobacula toluolica Tol2	−	+	+	+	+	0	−	−	−	0
Desulfobulbaceae bacterium BRH_c16a BRHa_1007187	+	+	0	0	−	0	−	0	−	−
Desulfobulbus elongatus DSM 2908	−	0	0	0	−	0	0	−	−	+
Desulfobulbus japonicus DSM 18378	+	+	0	−	0	0	+	−	+	−
Desulfobulbus mediterraneus DSM 13871	+	0	0	0	0	0	−	0	−	−
Desulfobulbus propionicus DSM 2032	+	+	0	0	−	0	+	−	−	−
Desulfobulbus sp. Tol-SR	+	+	0	0	0	−	−	0	−	−
Desulfocapsa sulfexigens DSM 10523	+	+	0	0	−	0	−	0	0	−
Desulfococcus multivorans DSM 2059	−	+	0	+	−	0	+	0	−	0
Desulfococcus oleovorans Hxd3	−	+	0	+	0	0	−	0	−	0
Desulfocurvus vexinensis DSM 17965	−	+	0	0	−	0	+	−	−	−
Desulfomonile tiedjei DSM 6799	+	−	+	0	−	0	−	−	−	−
Desulfonatronum lacustre DSM 10312	+	+	0	0	−	−	−	0	−	−
Desulfonatronum thioautotrophicum strain ASO4-1	+	+	+	0	−	+	−	+	−	−
Desulfonatronum thiodismutans strain MLF-1	+	+	0	0	−	0	−	0	−	−
Desulfosarcina cetonica JCM 12296	+	+	+	+	−	0	+	0	−	−
Desulfospira joergensenii DSM 10085	−	+	0	−	−	0	−	+	−	−
Desulfotalea psychrophila LSv54	+	+	0	−	+	0	−	+	−	0
Desulfotignum balticum DSM 7044	+	−	+	+	−	0	−	+	+	−
Desulfotignum phosphitoxidans DSM 13687	+	−	+	+	−	0	−	0	−	0
Desulfovibrio acrylicus DSM 10141	−	+	0	0	−	0	0	−	−	−

Genome	*taoR*	TUP	*tupA*	WTP	MOD	*modA*	AOR	MOP	FDH	NIF
Desulfovibrio alaskensis DSM 16109	−	−	+	0	−	0	+	+	−	0
Desulfovibrio alaskensis G20	−	−	+	0	−	0	+	+	−	0
Desulfovibrio desulfuricans subsp. aestuarii DSM 17919	−	+	0	0	−	0	−	−	−	−
Desulfovibrio sp. A2	−	+	0	0	−	0	+	0	−	−
Desulfovibrio termitidis HI1	−	+	0	0	−	0	−	+	+	−
Desulfovibrio vulgaris DP4	−	+	−	0	−	0	+	−	+	0
Desulfovibrio vulgaris RCH1	−	+	−	0	−	0	+	−	+	0
Desulfovibrio vulgaris str. Miyazaki F	−	+	−	0	−	0	+	+	+	−
Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough	−	+	−	0	−	0	+	−	+	−
Desulfurivibrio alkaliphilus AHT2	+	+	0	0	−	0	0	0	+	+
Desulfuromonas sp. TF	+	+	0	0	−	0	−	0	−	0
Pelobacter carbinolicus DSM 2380	−	−	0	−	−	0	−	+	−	0
Syntrophobacter sp. DG_60	−	0	0	0	0	0	−	0	−	0
Syntrophorhabdus aromaticivorans UI	+	+	+	0	−	0	+	0	−	0
Syntrophus aciditrophicus SB	−	+	+	+	−	0	−	0	−	0

Locus tags for TF genes and regulated operons are given in Supporting Information Table S3.

Among the molybdo/tungstoenzyme -encoding genes regulated by TaoR, tungsten-containing aldehyde:ferredoxin oxidoreductase (*aor*) is the most common (15 regulons). Deltaproteobacteria may have several such enzymes (for example, *Desulfobacterium autotrophicum* genome contains eight *aor* genes), but only one of them (or two, in case of *Syntrophorhabdus aromaticivorans*) was predicted to be regulated by TaoR. Consistent with experimental data from gene fitness studies shown above, we found the predicted TaoR binding site upstream of *aor-4* (Dde_2460) gene in *D. alaskensis* G20.

We found operons encoding formate dehydrogenases in nine regulons. Formate dehydrogenases contain either Mo or W cofactors and play an important role in energy metabolism of anaerobic Deltaproteobacteria. Many sulfate reducing bacteria have two or more isoenzymes which differ in catalytic activity, metal cofactor specificity and membrane localization (Mota *et al*., 2011; Pereira *et al*., 2011). We also found TaoR binding sites upstream of the FDH maturation gene, *fdhE,* in *Desulfobacterium autotrophicum* and *Desulfocurvus vexinensis*.

Other enzyme encoding target genes include a *mop* gene encoding a molybdoenzyme from the xanthine dehydrogenase family (10 regulons) and operons encoding nitrogenase genes – *nifH, nifD, nifK* – and nitrogen regulatory proteins GlnB2 and GlnB3 (three regulons). The molybdoenzyme is homologous to aldehyde oxidoreductase from *Desulfovibrio gigas* (Thoenes *et al*., 1994), which is active against wide spectra of short- and long-chain aliphatic and aromatic aldehydes (Marangon *et al*., 2013).

Thus, our whole-genome regulon analysis in Deltaproteobacteria demonstrated that the TaoR family is linked to the regulation of tungstate transport and tungsto/molybdoenzymes genes. These results support our predictions from other divisions of Proteobacteria based on colocalization of TaoR family members with tungstate transporters. A distinctive feature of TaoR regulons in Deltaproteobacteria is a small number of target genes colocalized with regulatory genes from the *TaoR family*. In addition, we observed a significant variability between TaoR regulons, even in closely related genomes (like *Desulfovibrio* spp.). However, the most conserved members of TaoR regulons in Deltaproteobacteria are tungstate-specific transporter TupABC and a tungstoenzyme aldehyde:ferredoxin oxidoreductase Aor.

In *D. vulgaris*, TaoR is a tungstate responsive transcriptional regulator

To confirm TaoR regulon predictions, and to identify conditions under which TaoR is active, we experimentally analysed the TaoR regulon in *D. vulgaris* Hildenborough, a model sulfate-reducing bacterium. TaoR (DVU3193) was predicted to regulate three target genes/operons: DVU1179 (*aor*), DVU2812-DVU2809 (encoding Mo-dependent formate dehydrognease), and DVU0745-DVU0747 (*tupABC*). We examined expression of the first gene in these operons in wild-type *D. vulgaris* and a mutant strain containing a transposon insertion within the *taoR* gene, in the presence and absence of tungstate and molybdate (Fig. 4). Since *tupBC* was predicted to be in an operon separate from *tupA* in the MicrobesOnline database (Dehal *et al*., 2010) we tested both *tupA* and *tupB* expression by qRT-PCR.

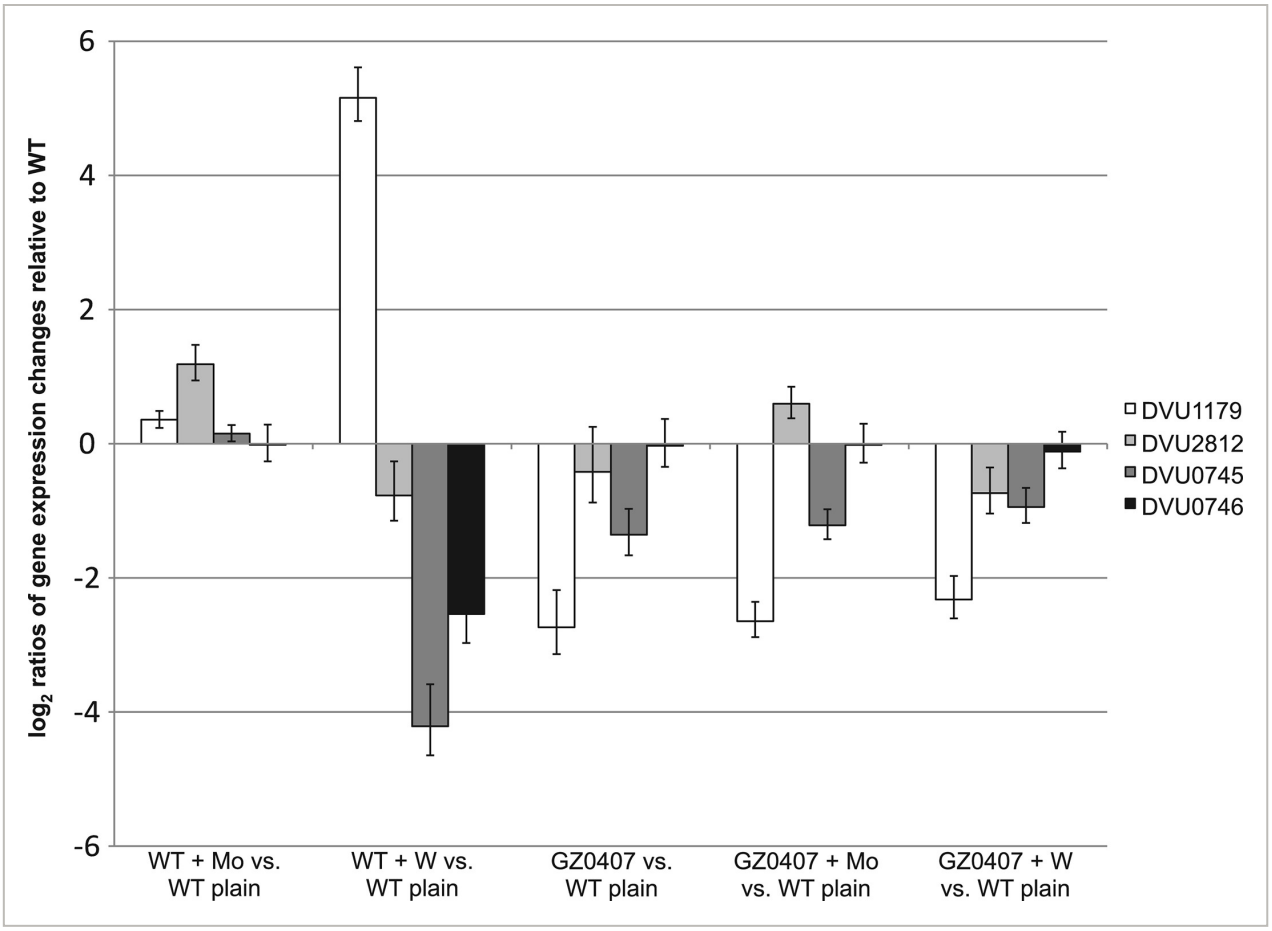

Figure 4

Open in figure viewer | PowerPoint

TaoR acts as a tungstate-responsive transcriptional regulator.Expression of predicted target genes *aor* (DVU1179), *fdnG-3* (DVU2812), *tupA* (DVU0745) and *tupB* (DVU0746) were tested by qRT-PCR in wild-type and *taoR* mutant strain GZ0407 in the presence and absence of molybdate and/or tungstate. The plots show $\log_2$ ratios of normalized gene expression relative to that of WT in the absence of molybdate or tungstate. Data are average of three biological replicates and error bars represent standard deviation.

We observed that, in the wild-type strain, *aor* gene expression was strongly induced in the presence of tungstate (35-fold) (Fig. 4). This transcription induction was also dependent on TaoR, since the *taoR* mutant was unable to induce the expression in the presence of tungstate. TaoR also appears to be necessary for *aor* expression in the absence of tungstate, since the *taoR* mutant had reduced expression (fivefold) in the absence of either molybdate or tungstate compared to wild-type strain.

We observed only a modest change in expression of the formate dehydrogenase gene DVU2812 in the wild-type strain (Fig. 4). The gene was slightly upregulated in the presence of molybdate (twofold) and slightly downregulated in the presence of tungstate (twofold). The

absence of TaoR did not affect DVU2812 gene expression. A similar effect of molybdate on transcription of this operon has been demonstrated earlier (da Silva *et al*., 2011). A significantly higher transcription of the operon has been observed under growth on formate and sulfate in comparison with growth on lactate and sulfate, suggesting additional regulatory mechanisms of formate dehydrogenase expression.

Expression changes for the *tupA* and *tupB* genes were similar to each other. The wild-type strain showed a repression of these genes (19-fold for *tupA*, and sixfold for *tupB*) in the presence of tungstate. The TaoR protein was required for this repression, as in the mutant strain there was no difference in expression in the presence or absence of tungstate (Fig. 4).

Purified TaoR binds to predicted sites upstream of the *aor* gene

To confirm that TaoR is a DNA-binding protein, we cloned the *taoR* gene from *D. vulgaris* Hildenborough into *Escherichia coli*, purified the protein and tested it by electrophoretic mobility shift assay (EMSA) and DNase I footprinting. We tested several DNA fragments of *aor* upstream region for TaoR binding by EMSAs (Fig. 5). TaoR shifted a 304 bp fragment i and this shift was specific (Supporting Information Fig. S3). When an excess of unlabeled *aor* upstream region was added to the reaction as specific competitor DNA, the shift was eliminated, but not when an excess of nonspecific competitor (empty vector DNA) was used (Supporting Information Fig. S3). Since expression studies indicated that TaoR responds to tungstate levels, we tested binding in the presence of sodium tungstate (0.1 to 10 μM). We did not observe any effect of tungstate on the binding *in vitro* (not shown).

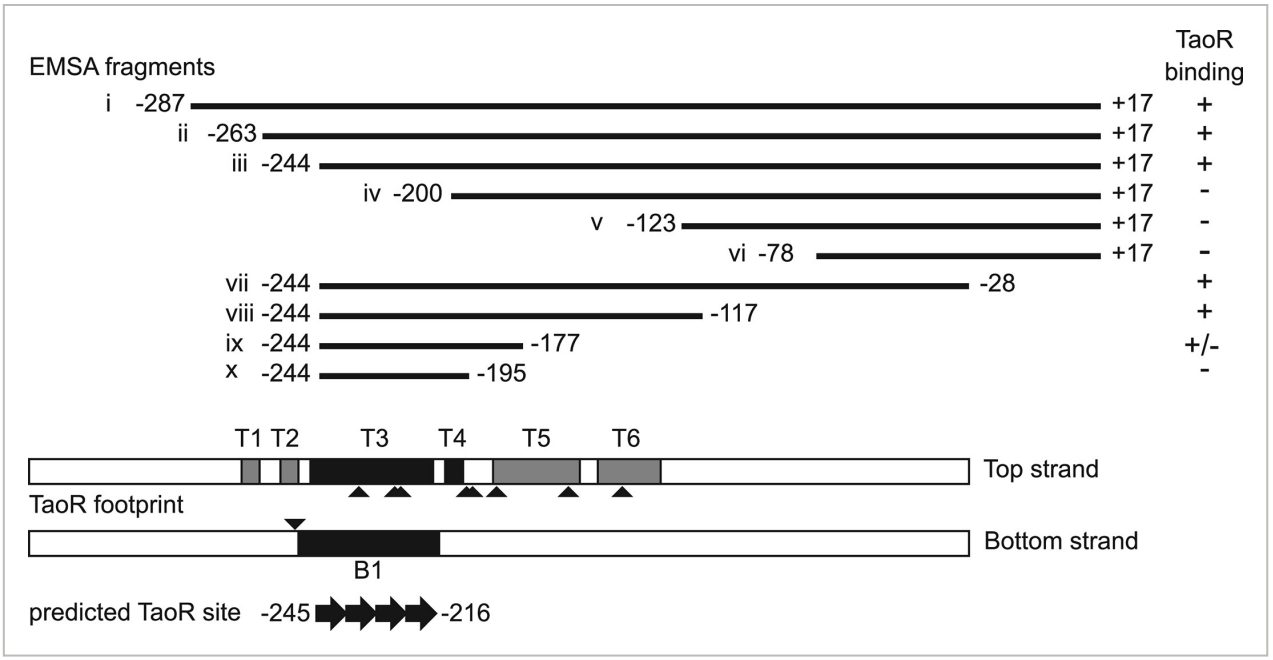

Figure 5

Open in figure viewer | PowerPoint

Mapping of TaoR binding site upstream of *aor* gene of *D. vulgaris*.Thick black lines indicate DNA fragments used in EMSA experiments. Numbers indicate positions relative to the start codon of *aor*. The black boxes indicate regions of strong DNase-I protection, the grey boxes indicate regions of weak DNase-I protection, and the black triangles indicate positions of enhanced cleavage in the presence of TaoR protein. Black arrows indicate four direct repeat units of the predicted TaoR site.

By truncating the original 304 bp fragment from the 5′ end, we observed that TaoR shifted fragments ii and iii, but not fragments iv, v and vi (Fig. 5 and Supporting Information Fig. S3). Thus, the absence of the predicted TaoR site in fragment iv resulted in a complete loss of binding. We then truncated fragment iii from the 3′ end. TaoR shifted fragments vii and viii, but only slightly shifted fragment ix (Fig. 5 and Supporting Information Fig. S3). TaoR also did not shift fragment x that corresponded to areas of strong protection in footprinting experiments (see below). Our results suggest that in addition to the predicted TaoR binding sites, the downstream area is also important for TaoR binding.

For precise mapping of the TaoR binding site upstream of the *aor* gene, we performed DNase I footprinting analysis. With a top strand-labelled substrate, TaoR strongly protected 41 bases at T3 (−247 to −207) and six bases at T4 (−205 to −199). TaoR weakly protected four additional areas: six bases each at T1 and T2, 29 bases at T5 (−189 to −161) and 21 bases at T6 (−154 to −134) (Fig. 5 and Supporting Information Fig. S4). DNase I cleavage was enhanced in six spots separated by either 10 or 20 bases (Fig. 5 and Supporting Information Fig. S4). With a bottom strand-labelled substrate, we only observed one strong large protected region B1 of 47 bases (−249 to −203) and one hypersensitive spot (−250; Fig. 5 and Supporting Information Fig. S5). The bottom strand footprint overlaps with the strong region of protection T3 seen on the top strand. Protein extract from a strain carrying the empty vector produced no regions of protection (data not shown). Consistent with our EMSA experiments, which localized the minimal region for TaoR binding at −244 to −177 bp upstream of *aor*, the DNase I footprinting revealed two binding sites within the *aor* upstream region. The upstream site corresponds to areas B1, T3 and T4 of strong protection and the downstream site corresponds to areas T5 and T6 of weak protection.

Discussion

We show that multidomain proteins with DNA-binding domain HTH_17 and C-terminal domain PBP_like constitute a new family of bacterial transcriptional regulators. This family, designated TaoR, is involved in the control of tungstate and molybdate transporters and tungsto/molybdoenzymes.

The DNA-binding domain HTH_17 is widespread among bacteria, and several thousands of HTH_17 family members are annotated in the Pfam database. However, only two transcriptional regulators with HTH_17 domains have been characterized experimentally: BldC, a transcription factor controlling early stage of aerial hyphae formation in *Streptomyces* spp. (Schumacher *et al*., 2018) and HnoC, a two-component system response regulator responding to nitric oxide in *Shewanella oneidensis* (Plate and Marletta, 2013). BldC binds sites with a variable number of direct repeat units and recognizes two sequence elements: 4 bp AT-rich region and a pyrimidine-guanine dinucleotide 4 bp downstream from the AT-rich region. HnoC binds a 20 bp imperfect inverted repeat, but footprints also contain direct repeat GACAaaT with AT-rich region and conserved pyrimidine-guanine dinucleotide. These two sequence elements also present in the predicted TaoR binding motifs from Deltaproteobacteria and Firmicutes (Supporting Information Fig. S2). The binding sites for non-regulatory DNA binding proteins with HTH_17 domains such as Xis excisionases from bacteriophages P22 (Mattis *et al*., 2008), L5 and Pukovnik (Singh *et al*., 2014) have similar tandem repeats with 4 units, which contain T-rich stretches. HTH_17 is considered to be part of the larger MerR family. However, TaoR, BldC and HnoC are distinct from canonical MerR regulators. Typically, MerR family proteins bind inverted repeated DNA sequences as symmetric head-to-head homodimers (Brown *et al*., 2003) close to regulated promoters. They activate suboptimal promoters having extended (19–20 bp) spacer between −35 and − 10 elements but can also repress promoters with 17 bp spacers. In contrast, TaoR binds far upstream of the *aor* transcription start site identified by 5′ RNA-seq in *D. vulgaris* Hildenborough (Price *et al*., 2011), and we could not find a promoter motif at the TaoR binding site. The mechanism of promoter activation by TaoR family regulators remains unclear.

The second characteristic feature of the TaoR family is the presence of a C-terminal ligand-binding domain (PBP_like), which is similar to periplasmic binding proteins. This domain defines the specific signal recognized by transcriptional factors. We determined that TaoR family members from Deltaproteobacteria (and, possibly, other Proteobacteria) respond to tungstate. Since TaoR family TFs from Firmicutes were predicted to regulate ModABC transporters, it is not unlikely that those TFs respond to molybdate (or both tungstate and molybdate). Phylogenetic analysis of PBP_like domain sequences from the Pfam database identified five separate clades with the domain structure of TaoR (Supporting Information Fig. S6). Three of them contain *tup*-associated regulators, while the other two groups contain *mod*-associated regulators. Thus, the TaoR domain organization may have formed independently two or more times.

In the Pfam database, the DNA-binding domain HTH_17 was found in combination with several C-terminal domains, besides the PBP_like domain (Fig. 6A), suggesting that more varieties of transcriptional regulators may exist. Putative TFs with a combination of HTH_17 and TOBE domains designated MerR-Mop (Zhang and Gladyshev, 2008) have been shown to be

colocalized with *modABC* genes in Actinobacteria and predicted to regulate molybdate transport. In contrast to most ModE proteins, MerR-Mop proteins have only one TOBE domain instead of two.

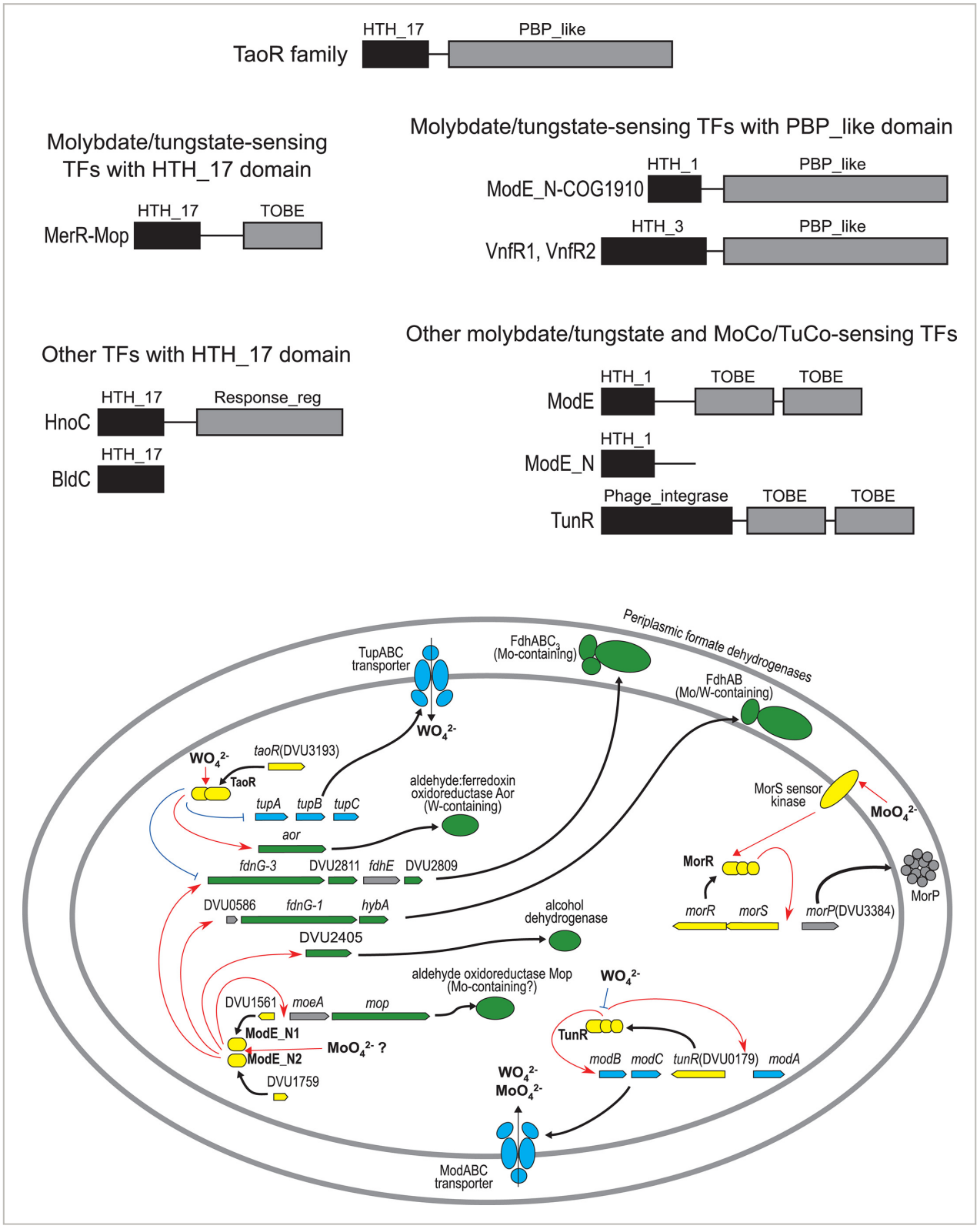

Figure 6

Open in figure viewer | PowerPoint

A. Domain architecture of regulatory proteins with HTH_17 and PBP_like domains and other molybdate/tungstate-responsive TFs. B. A model of molybdate/tungstate-dependent transcriptional regulation in *D. vulgaris* Hildenborough. Positive regulation is shown by red lines, and negative regulation is shown by blue lines. Regulatory genes and proteins are marked yellow, transport genes and proteins are marked blue, enzymes and their genes are marked green.

The PBP_like protein domain also exists in combination with various DNA-binding domains (Fig. 6A). Quite unsurprisingly, some of them are molybdate-responsive regulators, for example, VnfR1 and VnfR2 repressors of vanadium-dependent nitrogenase operon from *Anabaena variabilis* (Pratte *et al*., 2013). Similar proteins are present in some other Cyanobacteria and Deltaproteobacteria. Other predicted regulatory proteins with HTH_1 DNA-binding domain were designated as ModE_N-COG1910 in (Zhang and Gladyshev, 2008). One such regulator, FdsR from *Ralstonia eutropha* (Oh and Bowien, 1999), binds upstream of an operon encoding FDH and acts as an activator in the presence of formate or as a repressor in the absence of formate.

The current study sufficiently extends our knowledge about the regulation of molybdate and tungstate homeostasis in *D. vulgaris* Hildenborough (Fig. 6B). Previously, four transcriptional regulators were known to participate in this process, with only one of them experimentally characterized. Initial genomics studies identified two ModE_N proteins [as designated in (Zhang and Gladyshev, 2008)] (DVU1561, DVU1759) each with a single DNA-binding domain (Rodionov *et al*., 2004). ModE_N proteins lack a recognizable ligand-sensing domain, but they are present in multiple species and considered to be functional regulators of molybdate transport (Zhang and Gladyshev, 2008). There is some evidence that ModE_N proteins sense the cellular MoCo status rather than intracellular concentration of molybdate (Hoffmann *et al*., 2016). The predicted *D. vulgaris* regulon for the two ModE_N TFs from the RegPrecise database (Novichkov *et al*., 2010a,b) includes the *mop* gene encoding aldehyde oxidoreductase and two operons encoding formate dehydrogenases. The expression of these formate dehydrogenases depends on the metal content in the medium: FdhAB was predominant in the presence of tungstate, while $FdhABC_3$ was predominant in the presence of molybdate (da Silva *et al*., 2011). We propose that the ModE_N regulators activate one of the FDH operons and simultaneously repress the other operon. The third regulator is the tungstate-responsive TunR regulator that we recently characterized (Kazakov *et al*., 2013), TunR activates the molybdate/tungstate transporter *modABC* genes, but the activation is repressed in the presence of tungstate. TunR has an unusual DNA-binding domain similar to the domains of XerC and XerD recombinases and two ligand-sensing TOBE domains (Fig. 6A). The fourth regulator is a homologue of the

MorR response regulator from *D. alaskensis* G20 (Rivas *et al*., 2009). MorR belongs to the MorRS two-component regulatory system controlling the expression of the molybdenum response-associated periplasmic protein MorP. The current study adds the fifth regulatory protein, TaoR, which directly controls transcription of the *aor* gene, *tupABC* operon and one formate dehydrogenase operon in a tungstate-dependent manner.

Our work shows that sulfate-reducing bacteria have a complex regulatory network controlling the homeostasis of molybdenum and tungsten. This network orchestrates a switch in the expression of molybdenum-containing and tungsten-containing isoenzymes, as well as control of the expression of metal-specific transporting systems. This complex regulatory program ensures survival and proliferation of sulfate-reducing bacteria in diverse environmental niches with varying concentrations of molybdenum and tungsten. This network includes multiple transcriptional regulators, with some of them widespread in bacteria (ModE family), some present in a limited number of species (TaoR family), and some specific only for sulfate reducers (TunR family). Some of these transcription factors might evolve from non-regulatory DNA-binding proteins, like recombinases or excisionases. Future studies should aim at characterizing DNA-binding proteins with predicted molybdate- and tungstate-binding domains to discover new members of this regulatory network.

Experimental procedures

Genome and functional data sources

Genome sequences were downloaded from NCBI GenBank database (Benson *et al*., 2013). Protein domain mappings were downloaded from Pfam database (v. 29) (Finn *et al*., 2016).

Gene fitness dataset for *D. alaskensis* G20 (Kuehl *et al*., 2014) is available in MicrobesOnline database (Dehal *et al*., 2010) and was analysed by MicrobesOnline tools.

Phylogenetic and genome context analyses

Multiple alignment of protein sequences were built by T-COFFEE (Notredame *et al*., 2000). The phylogenetic trees were constructed by the maximum likelihood method implemented in the Fasttree software v. 2.1.10 (Price *et al*., 2010) and visualized by Dendroscope (Huson and Scornavacca, 2012).

Gene colocalization patterns were analysed by Pfam domain mappings. We selected all Pfam domains that occurred at least 15 times in the close genome neighbourhoods of TaoR family members (five genes upstream and five genes downstream).

Motif reconstruction and regulon prediction

We used the established approach of regulon reconstruction that is based on the comparative genomic analysis of candidate TF-binding sites in closely related bacterial genomes [reviewed in Rodionov (2007)]. First, we identified all TaoR family members in GenBank database by a combination of BLAST search (Camacho *et al*., 2009) in Deltaproteobacteria proteomes. To confirm the presence of two characteristic protein domains PF12727 and PF12728 in these proteins, we applied hmmscan (options: --cut_ga --domtblout) from HMMER v.3.1b2 (Eddy, 2009) to search against all HMMs from PFAM-A database, release 28 (Finn *et al*., 2016). Results of hmmscan search demonstrated that all deltaproteobacterial proteins from TaoR family have only PF12727 and PF12728 domains.

We searched for a conserved sequence motif upstream of *tupABC* operons that are typically colocalized or co-transcribed with *taoR* genes using MEME web service (http://meme-suite.org) (Bailey *et al*., 2009). The training set for motif search included entire upstream regions of 10 genes: AMJ54_14555, AMJ94_17745, DFT_RS24235, VR65_18055, Desti_0066, DESLA_RS0108485, LZ09_RS06005, GY33_RS0102910, JCM12296_RS13035 and DP0332. MEME was run with default parameters, except: minimal motif length set to 12 bp, maximal motif length set to 50 bp, motif occurrence set to 'any number of repetitions'. A motif with lowest e-value was selected as a candidate TaoR-binding motif.

For regulon reconstruction in Deltaproteobacteria by a comparative genomics approach we used the GenomeExplorer software package (Mironov *et al*., 2000). The 40 bp candidate motif identified by MEME was converted into position-weight matrix (PWM) using the SignalX program from GenomeExplorer package. This PWM was used for a whole-genome search in upstream regions of coding genes (from −400 to +50 with respect to the translation start) with a threshold equal to a minimal score among all sites in a training set of the matrix. Each genome encoding the TaoR family member was scanned with the PWM using the GenomeExplorer software, and orthologous genes having candidate regulatory sites in at least three genomes were selected. Orthologous proteins were defined as bidirectional best hits with 30% of identity threshold using the Smith–Waterman algorithm implemented in the GenomeExplorer and, if necessary, confirmed by analysis of phylogenetic trees.

Since initial 40 bp TaoR-binding motif produced overlapping hits in several genomes, it became evident that candidate TaoR sites may contain variable number of repeat units. To map individual repeat units precisely, we constructed a refined 10 bp motif based on individual repeat units extracted from the predicted 40 bp candidate sites. The first position of the repeat unit motif was selected with regard to TaoR footprinting data, in a way that 40 bp region of strong protection upstream of the *aor* gene corresponds to four complete repeat units. Individual repeat units were mapped in genome sequences of Deltaproteobacteria by custom scripts for PWM-based site search that use the search algorithm implemented in the RegPredict software (Novichkov *et al*., 2010a,b).

Bacterial strains

D. vulgaris Hildenborough strain ATCC 29579 was obtained from the American Type Culture Collection (Manassas, VA). The *taoR* transposon mutant was selected from a library of *D. vulgaris* Hildenborough transposon mutants constructed in a similar manner as the *D. alaskensis* G20 transposon library (Kuehl *et al*., 2014). Isolates from this library have been used in previous studies (Kazakov *et al*., 2013; Fels *et al*., 2013).

Expression assays

D. vulgaris was grown at 32 °C in 15 ml centrifuge tubes in an anaerobic chamber (Coy Laboratories, Grasslake, MI) in an atmosphere of 85% N_2/10% CO_2/5% H_2. WT and GZ0407 (Tn5::DVU3193) were grown from freezer stocks in 10 ml LS4D (Rajeev *et al*., 2015) supplemented with 0.1% yeast extract overnight. For the mutant, G418 (400 µg/ml) was added to the medium throughout the experiment. The cells were washed four times in a modified LS4D medium that was prepared with trace elements mixture lacking molybdate and tungstate ($LS4D_{mw}$). The washed cells were inoculated in $LS4D_{mw}$ (20 ml, 3% inoculum) and allowed to grow. The resulting culture was used to inoculate 10 ml of $LS4D_{mw}$, $LS4D_{mw}$ + 1 µM sodium molybdate, and $LS4D_{mw}$ + 0.3 µM sodium tungstate (in triplicate). The cells were harvested at mid-log phase.

RNA was extracted immediately with the Agilent Total RNA isolation mini kit (Agilent Technologies). The RNA was DNase digested to remove contaminating genomic DNA with the Turbo DNA-free DNase kit (Thermo Fisher Scientific, Waltham, MA). RNA concentrations were determined spectrophotometrically with the Nanodrop™ (Thermo Fisher Scientific). RNA integrity was determined visually by agarose gel electrophoresis. Five hundred nanograms of each RNA was reverse-transcribed with the iScript RT supermix (BioRad, Hercules, CA). Standards were prepared by pooling together 4 µl of each cDNA, followed by serial fivefold dilutions. The remaining cDNA samples were diluted 10-fold with RNase-free water. Primers for qPCR reactions were designed using Primer3 (Untergasser *et al*., 2012) (Supporting Information Table S4). qRT-PCR reactions were set up in triplicate, with each 20 µl reaction containing 1X SsoAdvanced Universal Sybr Green Supermix (BioRad, Hercules, CA), 500 nM of each primer, and 2 µl of cDNA in 96 well plates (white, BioRad). Reactions were cycled on a CFX96 Real-Time PCR Detection System (BioRad) as follows: 98 °C/2 min, and 40 cycles of 98 °C/10s, 60 °C/30 s, with a melt curve from 65 °C to 95 °C in 0.5 °C increments. *D. vulgaris* Hildenborough genes *bioD* (DVU2565) and *rpoH* (DVU1584) were used as reference genes. This study showed no change in expression for the chosen reference genes under the conditions tested. The fold change in expression was calculated by the CFX96 software based on qbase+ (Hellemans *et al*., 2007).

DVU3193 cloning in *E. coli,* protein expression and purification

DVU3193 (912 bp) was amplified from *D. vulgaris* Hildenborough genomic DNA and cloned into the vector pSKB3 (T7 promoter) with a cleavable N-terminal His-tag. Oligo design and PCR amplifications were carried out using the DiVA PCR service (Emery Station East, Lawrence Berkeley National Lab). The parts were assembled by the Gibson assembly method (Gibson *et al*., 2009), and transformed into *E. coli* DH10B. Transformants were selected on LB-Kan, and verified by colony PCR followed by sequencing. The sequence verified plasmid was then transformed into the expression strain *E. coli* BL21(DE3) and transformants were selected on LB-Kan.

Protein was purified in one of two ways. Initial electrophoretic mobility shift assays used protein that was purified as follows: a 5 ml overnight culture of the expression strain was used to inoculate 500 ml Terrific broth (TB) and grown at 37 °C until cells reached an OD_{600} of 1 to 1.5. Protein expression was induced with 0.25 mM IPTG and the cells were grown overnight at 25 °C. The cells were harvested by centrifuging at 15 317*g* at 4 °C for 15 min. Cells were resuspended in 30 ml 1X TBS (50 mM Tris-Cl, pH 7.6; 150 mM NaCl), and lysed by sonication. The lysate was spun down at 15 317 g for 10 min. The protein was purified by loading the lysate onto a Ni-resin column that was equilibrated in TBS. The lysate was allowed to bind for 10 min. The resin was washed with 300 ml of 40 mM imidazole in TBS, and the protein was eluted in 8 ml of 250 mM imidazole in TBS.

DNA footprinting assays and some EMSAs used protein that was purified with the following method: a 5 ml overnight culture of the expression strain was used to inoculate 10 ml TB. Growth, IPTG induction and harvesting of cells were as described above. Cell lysis and protein purification was performed with Talon Spin columns (Takara BioUSA, Mountain View, CA) and the HisTalon Buffer set (phosphate-buffer based) following the manufacturer's instructions.

Protein purified by either method was desalted using PD-10 disposable desalting column (GE Life Sciences, Pittsburgh, PA) according to the manufacturer's spin protocol, glycerol was added to a final concentration of 5% and the protein was flash frozen in liquid nitrogen and stored at −80 °C in 50 µl aliquots.

Electrophoretic mobility shift assay

Primers were ordered from IDT (Coralville, IA). DNA substrates longer than 60 bp were prepared by PCR amplifying the desired upstream region with one labelled primer (Cy3 or 6-carboxyfluorescein) and one unlabeled primer with *D. vulgaris* Hildenborough genomic DNA as template (Supporting Information Table S4), followed by gel purification. DNA substrates shorter than 60 bp were prepared by annealing a labelled and an unlabeled oligo by heating at 95 °C for 5 min followed by slow cooling in 10 mM Tris–HCl, pH 8, 50 mM NaCl, 1 mM EDTA. The

various DNA substrates and primers used are in Supporting Information Table S4. Binding reactions (20 µl) were set up with 100 fmol of labelled DNA substrate and protein (amount as specified) in either Buffer A (10 mM Tris HCl, pH 7.5, 50 mM KCl, 50 mM NaCl, 1 mM DTT, 10% glycerol, 1 mM EDTA) or Buffer B (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4 and 2 mM KH_2PO_4, 1 mM DTT, 1 mM EDTA) depending on if the protein was purified in a Tris-based buffer or a phosphate-based buffer. Poly dI-dC (1 µg) was added as a non-specific competitor to all binding reactions. Reactions were incubated at room temperature for 20 min. Reactions were mixed with EMSA loading buffer (without dye) and loaded on a 5% TBE polyacrylamide pre-cast mini-gel (BioRad. Hercules, CA) that was prerun for 30 min, and run at 90 V for 1.5 h. The gels were imaged on an Amersham Imager 600 (GE Life Sciences, Pittsburgh, PA).

DNase-I footprinting

A 314 bp region upstream of DVU1179 (−341 to −28) was amplified from *D. vulgaris* Hildenborough genomic DNA using one labelled and one unlabeled primer such that either the top or the bottom strand of the substrate was labelled on the 5′ end with 6-FAM (6-carboxyfluorescein). Binding reactions (100 µl) with TaoR and labelled DNA (8 ng) were set up in 10 mM sodium phosphate (pH 8), 50 mM NaCl, 10 mM KCl and 1 mM DTT. Reactions were incubated at room temperature in the dark for 30 min. The DNA was then digested with DNase I (0.08 U, New England Biolabs) for 5 min at room temperature in a total reaction volume of 120 µl. The reactions were stopped by addition of EDTA (15 mM). The reactions were immediately cleaned up with the Qiaquick PCR purification kit (Qiagen) and eluted in 30 µl water. Aliquots (10 µl) were mixed with 20 µl Hi-Di Formamide (Thermo Fisher Scientific) and 1 µl of LIZ 500 size standards (1/5 diluted) (Thermo Fisher Scientific). Sequencing reactions were also set up with the DNA substrate (5 ng) and the labelled primer with the Thermo sequenase cycle sequencing kit (Affymetrix, Santa Clara, CA). One microlitre of the sequencing reactions were mixed with 8.9 µl of Hi-Di formamide and 0.1 µl of LIZ 500 standards. All samples were analysed on a 3730 XL DNA Analyser (Thermo Fisher Scientific) at the UC Berkeley Sequencing Facility.

Acknowledgements

We thank Garima Goyal of the DiVA team at Emerystation East campus of LBNL for DNA synthesis service and Nathan Hillson (LBNL) for managing this service. We thank Adam Deutschbauer of EGSB division of LBNL for mapping of transposon insertion in GZ0407 strain.

This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231. The funders had no role in study design, data collection and interpretation, or the

decision to submit the work for publication. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide licence to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Supporting Information

Filename	Description

Filename	Description
emi14500-sup-0001-Figures.pdf PDF document, 905.1 KB	**Fig. S1**. Phylogenetic tree of TaoR family members. Colocalization patterns are marked as follows: squares (M: ModABC transporter), triangles (T: TupABC transporter), diamonds, blue organism names (F: formate dehydrogenase-associated genes), circles, red organism names (N: nitrogenase), crosses (P: Mop protein). **Fig. S2**. Predicted TF-binding motifs of TaoR family and binding sites of TFs and Xis proteins with HTH_1 7 domain. **Fig. S3**. Minimal aor promoter region that is shifted by TaoR. A. Upstream regions of aor that were used as DNA substrates in EMSAs. Numbers indicate start and stop positions relative to the ATG start codon of aor. The dark red boxes indicate regions of strong DNase-I protection, and the light red boxes indicate regions of weak DNase-I protection. B. TaoR specifically shifts the full length upstream region of aor (304 bp substrate i). EMSAs showing binding reactions with: Lane 1 . DNA only; lane 2. With empty vector extract; lane 3. With TaoR and an excess of unlabeled substrate i DNA (specific competitor); lane 4. With TaoR and an excess of unlabeled non-specific competitor DNA; lane 5. With TaoR. C. Minimal promoter region required for TaoR binding. Binding reactions are with or without TaoR and the respective DNA substrates as indicated in the figure. Lane 2 is a control with the empty vector extract added. DNA mobility shift by TaoR was observed for substrates i, ii, iii, vii and viii. **Fig. S4**. TaoR protects several regions in the promoter region of aor. DNase I protection assay was performed with a top-strand labelled 31 4 bp fragment upstream of aor that extended from −341 to −28. Fragment analysis of the region from −272 to −1 90 (A) and from −1 94 to −1 1 9 (B) show several regions of protection (red brackets, and red bases highlighted in the sequence), and a few areas of enhanced cleavage (green asterisks, green bases in sequence) in the presence of increasing amounts of TaoR. **Fig. S5**. TaoR protects a 40 bp region on the bottom strand of aor promoter. DNase I protection assay was performed with a bottom-strand labelled 31 4 bp fragment upstream of aor that extended from −341 to −28. Fragment analysis of the region from −281 to −1 56 shows a large footprint of 40 bases (red brackets, and red bases highlighted in the sequence), and a single point of enhanced cleavage (green asterisks, green bases in sequence) in the presence of increasing amounts of TaoR.
emi14500-sup-0002-FigureS6.pdf PDF document, 3.3 MB	**Fig. S6**. Phylogenetic tree of and PBP_like (PF1 2727) domain sequences. Protein sequences were downloaded from the PFAM database v. 29 website (Finn *et al*., 2016). Multiple sequence alignment was constructed with T-COFFEE v.11.0 (Notredame *et al*., 2000), phylogenetic tree was calculated with Fasttree v.2.1.10 (Price *et al*., 2010) and visualized with Dendroscope (Huson, Scornavacca, 2012). Proteins from TaoR, LysR and XRE TF families are marked by red, green and blue squares, respectively. The number on each interior branch is the bootstrap percentage.

Filename	Description
emi14500-sup-0003-Tables.xlsx Excel 2007 spreadsheet , 74.3 KB	**Table S1**. Gene fitness for selected Mo-related enzymes and TaoR regulator across *D. alaskensis* G20 dataset from the MicrobesOnline database. **Table S2**. Protein domains found in genome neighborhoods of TaoR family genes. **Table S3**. TaoR regulons and binding sites predicted in Deltaproteobacteria. **Table S4**. List of primers used for qPCR and EMSA.

Please note: The publisher is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing content) should be directed to the corresponding author for the article.

References

Aguilar-Barajas, E., Díaz-Pérez, C., Ramírez-Díaz, M. I., Riveros-Rosas, H., and Cervantes, C. (2011) Bacterial transport of sulfate, molybdate, and related oxyanions. *Biometals* **24**: 687– 707.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., *et al*. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202– W208.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Balleza, E., López-Bojorquez, L. N., Martínez-Antonio, A., Resendis-Antonio, O., Lozada-Chávez, I., Balderas-Martínez, Y. I., *et al*. (2009) Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiol Rev* **33**: 133– 151.
Wiley Online Library | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2013) GenBank. *Nucleic Acids Res* **41**: D36– D42.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Bevers, L. E., Hagedoorn, P.-L., and Hagen, W. R. (2009) The bioinorganic chemistry of tungsten. *Coord Chem Rev* **253**: 269– 290.
Crossref | CAS | Web of Science® | Google Scholar | UC-eLinks

Brown, N. L., Stoyanov, J. V., Kidd, S. P., and Hobman, J. L. (2003) The MerR family of transcriptional regulators. *FEMS Microbiol Rev* **27**: 145– 163.
Wiley Online Library | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Silva, S. M., Pimentel, C., Valente, F. M. A., Rodrigues-Pousada, C., and Pereira, I. A. C. (2011) Tungsten and molybdenum regulation of formate dehydrogenase expression in Desulfovibrio vulgaris Hildenborough. *J Bacteriol* **193**: 2909– 2916.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., *et al*. (2010) MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res* **38**: D396– D400.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Eddy, S. R. (2009) A new generation of homology search tools based on probabilistic inference. *Genome Inform* **23**: 205– 211.
Crossref | PubMed | Google Scholar | UC-eLinks

Fels, S. R., Zane, G. M., Blake, S. M., and Wall, J. D. (2013) Rapid transposon liquid enrichment sequencing (TnLE-seq) for gene fitness evaluation in underdeveloped bacterial systems. *Appl Environ Microbiol* **79**: 7510– 7517.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., *et al*. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**: D279– D285.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**: 343– 345.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Hellemans, J., Mortier, G., Paepe, A., Speleman, F., and Vandesompele, J. (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**: R19.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Hensgens, C. M., Hagen, W. R., and Hansen, T. A. (1995) Purification and characterization of a benzylviologen-linked, tungsten-containing aldehyde oxidoreductase from *Desulfovibrio gigas*. *J Bacteriol* **177**: 6195– 6200.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Hoffmann, M.-C., Ali, K., Sonnenschein, M., Robrahn, L., Strauss, D., Narberhaus, F., and Masepohl, B. (2016) Molybdate uptake by *Agrobacterium tumefaciens* correlates with the cellular molybdenum

cofactor status. *Mol Microbiol* **101**: 809– 822.
Wiley Online Library | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Hopp, C. M., Gardner, J. F., and Salyers, A. A. (2015) The Xis2d protein of CTnDOT binds to the intergenic region between the mob and tra operons. *Plasmid* **81**: 63– 71.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Huson, D. H., and Scornavacca, C. (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* **61**: 1061– 1067.
Crossref | PubMed | Web of Science® | Google Scholar | UC-eLinks

Johnson, M. K., Rees, D. C., and Adams, M. W. W. (1996) Tungstoenzymes. *Chem Rev* **96**: 2817– 2840.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Kazakov, A. E., Rajeev, L., Luning, E. G., Zane, G. M., Siddartha, K., Rodionov, D. A., *et al*. (2013) New family of tungstate-responsive transcriptional regulators in sulfate-reducing bacteria. *J Bacteriol* **195**: 4466– 4475.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Kuehl, J. V., Price, M. N., Ray, J., Wetmore, K. M., Esquivel, Z., Kazakov, A. E., *et al*. (2014) Functional genomics with a comprehensive library of transposon mutants for the sulfate-reducing bacterium Desulfovibrio alaskensis G20. *MBio* **5**: e01041– e01014.
Crossref | PubMed | Web of Science® | Google Scholar | UC-eLinks

Magalon, A., Fedor, J. G., Walburger, A., and Weiner, J. H. (2011) Molybdenum enzymes in bacteria and their maturation. *Coord Chem Rev* **255**: 1159– 1178.
Crossref | CAS | Web of Science® | Google Scholar | UC-eLinks

Makdessi, K., Fritsche, K., Pich, A., and Andreesen, J. R. (2004) Identification and characterization of the cytoplasmic tungstate/molybdate-binding protein (Mop) from Eubacterium acidaminophilum. *Arch Microbiol* **181**: 45– 51.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Marangon, J., Correia, H. D., Brondino, C. D., Moura, J. J. G., Romão, M. J., González, P. J., and Santos-Silva, T. (2013) Kinetic and structural studies of aldehyde oxidoreductase from *Desulfovibrio gigas* reveal a dithiolene-based chemistry for enzyme activation and inhibition by H(2)O(2). *PLoS One* **8**: e83234.
Crossref | PubMed | Web of Science® | Google Scholar | UC-eLinks

Mattis, A. N., Gumport, R. I., and Gardner, J. F. (2008) Purification and characterization of bacteriophage P22 Xis protein. *J Bacteriol* **190**: 5781– 5796.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Mendel, R. R. (2013) Metabolism of molybdenum. *Met Ions Life Sci* **12**: 503– 528.
Crossref | PubMed | Web of Science® | Google Scholar | UC-eLinks

Mironov, A. A., Vinokurova, N. P., and Gel'fand, M. S. (2000) Software for analyzing bacterial genomes. *Mol Biol (Mosk)* **34**: 253– 262.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Mota, C. S., Valette, O., González, P. J., Brondino, C. D., Moura, J. J. G., Moura, I., *et al*. (2011) Effects of molybdate and tungstate on expression levels and biochemical characteristics of formate dehydrogenases produced by *Desulfovibrio alaskensis* NCIMB 13491. *J Bacteriol* **193**: 2917– 2923.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Newport, P. J., and Nedwell, D. B. (1988) The mechanisms of inhibition of *Desulfovibrio* and *Desulfotomaculum* species by selenate and molybdate. *J Appl Microbiol* **65**: 419– 423.
Wiley Online Library | CAS | Google Scholar | UC-eLinks

Notredame, C., Higgins, D. G., and Heringa, J. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**: 205– 217.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Novichkov, P. S., Laikova, O. N., Novichkova, E. S., Gelfand, M. S., Arkin, A. P., Dubchak, I., and Rodionov, D. A. (2010a) RegPrecise: a database of curated genomic inferences of transcriptional regulatory interactions in prokaryotes. *Nucleic Acids Res* **38**: D111– D118.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Novichkov, P. S., Rodionov, D. A., Stavrovskaya, E. D., Novichkova, E. S., Kazakov, A. E., Gelfand, M. S., *et al*. (2010b) RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res* **38**: W299– W307.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Oh, J. I., and Bowien, B. (1999) Dual control by regulatory gene fdsR of the fds operon encoding the NAD+−linked formate dehydrogenase of Ralstonia eutropha. *Mol Microbiol* **34**: 365– 376.
Wiley Online Library | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Pereira, I. A. C., Ramos, A. R., Grein, F., Marques, M. C., Silva, S. M., and Venceslau, S. S. (2011) A comparative genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. *Front Microbiol* **2**: 69.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Plate, L., and Marletta, M. A. (2013) Phosphorylation-dependent derepression by the response regulator HnoC in the Shewanella oneidensis nitric oxide signaling network. *Proc Natl Acad Sci USA* **110**: E4648– E4657.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Pratte, B. S., Sheridan, R., James, J. A., and Thiel, T. (2013) Regulation of V-nitrogenase genes in *Anabaena variabilis* by RNA processing and by dual repressors. *Mol Microbiol* **88**: 413– 424.
Wiley Online Library | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Price, M. N., Deutschbauer, A. M., Kuehl, J. V., Liu, H., Witkowska, H. E., and Arkin, A. P. (2011) Evidence-based annotation of transcripts and proteins in the sulfate-reducing bacterium Desulfovibrio vulgaris Hildenborough. *J Bacteriol* **193**: 5716– 5727.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Rajeev, L., Chen, A., Kazakov, A. E., Luning, E. G., Zane, G. M., Novichkov, P. S., *et al*. (2015) Regulation of nitrite stress response in *Desulfovibrio vulgaris* Hildenborough, a model sulfate-reducing bacterium. *J Bacteriol* **197**: 3400– 3408.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Rivas, M. G., Carepo, M. S. P., Mota, C. S., Korbas, M., Durand, M.-C., Lopes, A. T., *et al*. (2009) Molybdenum induces the expression of a protein containing a new heterometallic Mo-Fe cluster in Desulfovibrio alaskensis. *Biochemistry* **48**: 873– 882.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Rodionov, D. A. (2007) Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem Rev* **107**: 3467– 3497.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Rodionov, D. A., Dubchak, I., Arkin, A., Alm, E., and Gelfand, M. S. (2004) Reconstruction of regulatory and metabolic pathways in metal-reducing delta-proteobacteria. *Genome Biol* **5**: R90.
Crossref | PubMed | Web of Science® | Google Scholar | UC-eLinks

Rothery, R. A., and Weiner, J. H. (2015) Shifting the metallocentric molybdoenzyme paradigm: the importance of pyranopterin coordination. *J Biol Inorg Chem* **20**: 349– 372.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Sagawa, S., Price, M. N., Deutschbauer, A. M., and Arkin, A. P. (2017) Validating regulatory predictions from diverse bacteria with mutant fitness data. *PLoS One* **12**: e0178258.
Crossref | PubMed | Web of Science® | Google Scholar | UC-eLinks

Schumacher, M. A., Hengst, C. D., Bush, M. J., Le, T. B. K., Tran, N. T., Chandra, G., *et al*. (2018) The MerR-like protein BldC binds DNA direct repeats as cooperative multimers to regulate Streptomyces development. *Nat Commun* **9**: 1139.
Crossref | PubMed | Web of Science® | Google Scholar | UC-eLinks

Singh, S., Plaks, J. G., Homa, N. J., Amrich, C. G., Héroux, A., Hatfull, G. F., and VanDemark, A. P. (2014) The structure of Xis reveals the basis for filament formation and insight into DNA bending within a mycobacteriophage intasome. *J Mol Biol* **426**: 412– 422.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Taylor, B. F., and Oremland, R. S. (1979) Depletion of adenosine triphosphate in Desulfovibrio by oxyanions of group VI elements. *Curr Microbiol* **3**: 101– 103.
Crossref | CAS | Web of Science® | Google Scholar | UC-eLinks

Thoenes, U., Flores, O. L., Neves, A., Devreese, B., Van Beeumen, J. J., Huber, R., *et al*. (1994) Molecular cloning and sequence analysis of the gene of the molybdenum-containing aldehyde oxido-reductase of *Desulfovibrio gigas*. The deduced amino acid sequence shows similarity to xanthine dehydrogenase. *Eur J Biochem* **220**: 901– 910.
Wiley Online Library | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**: e115.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Wood, M. M., and Gardner, J. F. (2015) The integration and excision of CTnDOT. *Microbiol Spectr* **3**: MDNA3–0020–2014.
PubMed | Web of Science® | Google Scholar | UC-eLinks

Zhang, Y., and Gladyshev, V. N. (2008) Molybdoproteomes and evolution of molybdenum utilization. *J Mol Biol* **379**: 881– 899.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks

Zhang, Y., Rump, S., and Gladyshev, V. N. (2011) Comparative genomics and evolution of molybdenum utilization. *Coord Chem Rev* **255**: 1206– 1217.
Crossref | CAS | PubMed | Web of Science® | Google Scholar | UC-eLinks