

Identification of SmtB/ArsR *cis* elements and proteins in archaea using the Prokaryotic InterGenic Exploration Database (PIGED)

MICHAEL BOSE,¹ DAVID SLICK,¹ MICKEY J. SARTO,¹ PATRICK MURPHY,² DAVID ROBERTS,² JACQUELINE ROBERTS² and ROBERT D. BARBER^{1,3}

¹ Biological Sciences Department, University of Wisconsin-Parkside, Kenosha, WI 53141, USA

² Department of Chemistry, DePauw University, Greencastle, IN 46135, USA

³ Corresponding author (barber@uwp.edu)

Received September 21, 2005; accepted January 5, 2006; published online April 28, 2006

Summary Microbial genome sequencing projects have revealed an apparently wide distribution of SmtB/ArsR metal-responsive transcriptional regulators among prokaryotes. Using a position-dependent weight matrix approach, prokaryotic genome sequences were screened for SmtB/ArsR DNA binding sites using data derived from intergenic sequences upstream of orthologous genes encoding these regulators. Sixty SmtB/ArsR operators linked to metal detoxification genes, including nine among various archaeal species, are predicted among 230 annotated and draft prokaryotic genome sequences. Independent multiple sequence alignments of putative operator sites and corresponding winged helix-turn-helix motifs define sequence signatures for the DNA binding activity of this SmtB/ArsR subfamily. Prediction of an archaeal SmtB/ArsR based upon these signature sequences is confirmed using purified *Methanosarcina acetivorans* C2A protein and electrophoretic mobility shift assays. Tools used in this study have been incorporated into a web application, the Prokaryotic InterGenic Exploration Database (PIGED; <http://bioinformatics.uwp.edu/~PIGED/home.htm>), facilitating comparable studies. Use of this tool and establishment of orthology based on DNA binding signatures holds promise for deciphering potential cellular roles of various archaeal winged helix-turn-helix transcriptional regulators.

Keywords: DNA binding, regulation, transcription, winged helix-turn-helix.

Introduction

Maintenance of intracellular metal concentrations is a vital aspect of cellular physiology, which employs an array of mechanisms for both sensing and controlling intracellular concentrations of metal ions, such as Co⁺², Zn⁺², Cd⁺², Ni⁺², Ag⁺², Hg⁺² and Pb⁺². Two component regulatory systems and distinct families of winged helix-turn-helix (HTH) transcriptional regulators have been shown to function in prokaryotic metal homeostasis (Xu and Rosen 1999, Brown et al. 2003, Busenlehner et al. 2003, Mergeay et al. 2003, Perron et al. 2004). Members of

the SmtB/ArsR winged-HTH protein family, which are widely annotated among bacterial and archaeal genome sequencing projects, act as metal-sensitive DNA binding repressors that control expression of various metal detoxification genes. Metal binding specificity within this family ranges between cobalt and lead with some SmtB/ArsR exhibiting affinity for multiple metals, offering an insightful model system for understanding the evolution of metal binding sites within proteins (Huckle et al. 1993, Wu and Rosen 1993, Cavet et al. 2002, Cavet et al. 2003, Pennella et al. 2003, Wang et al. 2005). Genes encoding SmtB/ArsR are often observed in chromosomal or episomal operons with genes encoding mechanisms for active transport, sequestration or chemical reduction of metals, and several reports cite these loci as examples of lateral gene transfer among diverse prokaryotes (Xu et al. 1996, Schirawski et al. 2002, Coombs and Barkay 2004).

While many SmtB/ArsR studies have focused on eubacterial proteins, recent investigations in the archaeal species *Halobacterium* sp. NRC-1 and *Sulfolobus solfataricus* P2 implicate SmtB/ArsR family regulators in archaeal cellular responses to elevated concentrations of arsenic or mercury, respectively (Schelert et al. 2004, Wang et al. 2004). Identification of SmtB/ArsR protein sequences within a diverse group of prokaryotes offers the opportunity to use “natural variants” to establish structure–function relationships and predict potential physiological roles for these proteins. To date, the amino acid sequence determinants for metal specificity are insufficiently characterized to allow clear predictions of metal binding properties from a primary sequence. However, comparative studies offer substantial impact in understanding DNA binding specificities within this protein family, which in turn provides insight into potential regulons controlled by these transcriptional regulators.

The DNA binding activity of SmtB/ArsR proteins is presumably mediated utilizing a well-characterized mechanism involving one α -helix, the “recognition” helix making nucleotide-specific contacts in the major groove of DNA and other protein determinants interacting in a non-specific manner to stabilize site-specific interactions (Gajiwala and Burley 2000).

In many cases, the “wing” appears to stabilize site-specific DNA binding by contacting the phosphate backbone and, in some instances, the minor groove. A distinct mode of DNA binding for winged-HTH proteins, with the wing-making site-specific contacts, was recently identified in a structural study (Gajiwala et al. 2000). Regardless of binding mode, winged-HTH proteins typically recognize DNA sequences exhibiting dyad symmetry. Examination of eubacterial SmtB/ArsR regulons, including those associated with the related CadC transcriptional regulator, reveals distinct inverted repeat sequences with 12-2-12 character often near the transcriptional start site, which act as regulatory DNA binding sites (Busenlehner et al. 2003). Mapping similar regulatory sequences in other prokaryotic genome sequences provides an opportunity to identify orthologous SmtB/ArsR based upon DNA binding properties and to assess relationships regarding gene regulation and gene family expansion.

Comparison of orthologous sequences has proven useful for analysis of prokaryotic intergenic sequences, especially regulatory element prediction (Gelfand et al. 2000, McCue et al. 2001, Tan et al. 2001, Rajewsky et al. 2002). With the rapidly growing wealth of microbial genome sequences representing numerous taxonomic groups, these approaches promise to become even more powerful. Recent comparative sequence studies have generously provided and demonstrated the utility of open source programs for use of position-dependent weight-matrices in searching intergenic spaces for conserved sequences (Studholme and Pau 2003, Studholme et al. 2004, Studholme and Dixon 2004, Tucker et al. 2004). A web application, the Prokaryotic InterGenic Exploration Database (PIGED; <http://bioinformatics.uwp.edu/~PIGED/home.html>) has been developed to facilitate investigation of prokaryotic intergenic regions using these programs and data derived from annotated microbial genome sequences at the National Center for Biotechnology Information (NCBI). Although analogous web applications exist with functionality for exploring intergenic sequences (Munch et al. 2003, Ray and Daniels 2003, van Helden 2003), PIGED has been designed with an emphasis on comparison of orthologous upstream intergenic regions found in archaeal and eubacterial species.

Using PIGED, 60 SmtB/ArsR DNA binding sites or operators, linked to metal detoxification genes, are predicted among 230 annotated prokaryotic genome sequences. Nine archaeal SmtB/ArsR and associated regulons are predicted from this analysis and the binding activity of one of these proteins, *Methanosarcina acetivorans* C2A MA4344, is confirmed using electrophoretic mobility shift assays. Relationships among proteins and *cis* elements are established through sequence comparisons and phylogenetic analysis, which indicates a potential for acquisition via lateral gene transfer for archaeal *smtB/arsR* and highlights potentially novel metal binding proteins with the DNA recognition signature of this SmtB/ArsR subfamily.

Materials and methods

Development and implementation of the Prokaryotic InterGenic Exploration Database

A MySQL database containing upstream prokaryotic intergenic sequences based on annotations submitted to NCBI is available at The Prokaryotic InterGenic Exploration Database (PIGED). This database was generated using a Perl program (IGSpy2.3), which uses coordinates from respective genome .ptt files to extract sequences greater than 5 bp in length upstream of individual genes oriented either in a tandem or divergent manner. Extracted upstream intergenic sequences are associated with the unique gene identifiers or “synonym” for the gene downstream of each respective intergenic sequence, which in turn is associated with individual genome sequences. Intergenic sequences of convergently oriented genes, typically representing approximately one percent of the total genome sequence, have not been included in this searchable database, which presently focuses on comparisons for upstream transcriptional regulatory elements.

PIGED allows investigators to access prokaryotic upstream intergenic sequences by several avenues: (1) SynSearch: upstream intergenic sequences can be obtained for any gene of interest by entering the desired gene synonym preceded by “>” (i.e., >MA4344, >MM1040, etc.). Gene names are also available as searchable values for *B. subtilis* and various *E. coli* strains, with a requirement for a preceding “>”. (2) COG-Search: gene synonyms can be obtained for orthologous genes as determined by Cluster of Orthologous Genes (COG) analysis (Tatusov et al. 2003). COG numbers associated with individual gene synonyms have been parsed from .ptt files and incorporated into the database, allowing investigators to retrieve orthologous synonyms and intergenic sequences based upon distinct taxonomic groups. Although paralogous relationships can present initial hurdles for this approach, a recent study used COG relationships to identify putative riboswitches in prokaryotic intergenic sequences (Abreu-Goodger et al. 2004). (3) SeqSearch: the PIGED upstream intergenic sequence database can be searched using nucleotide sequences or regular expression derivations to identify intergenic regions containing conserved or related sequence elements. Intergenic sequences retrieved via SeqSearch will be in lower case with the exception of the specific sequence that was used for the search, which is capitalized for ease of viewing and alignment. (4) Microarray Data: PIGED contains links to prokaryotic data available at the NCBI Gene Expression Omnibus (GEO). Investigators can utilize the clustering tools at GEO to analyze expression data, acquire gene synonyms for those genes exhibiting similar expression profiles and then access the upstream intergenic sequences of interest using SynSearch.

Output from SynSearch or Seqsearch provides the requested sequences in a series of text boxes that offer an editing option; sequences can be trimmed to facilitate subsequent multiple sequence alignment. Once sequences have been edited, links to ClustalW analysis (Thompson et al. 1994) or to output selected sequences as a text file are available. A separate link on the homepage allows users to input their own sequences of

interest for ClustalW analysis. ClustalW was chosen for comparative sequence analysis due to the portability of output into two previously developed programs that perform a combined position-dependent weight matrix determination and search of genome sequences (Studholme and Pau 2003).

Following ClustalW analysis, users can follow a link leading to a page requiring user-entered parameters to perform a "consensus" scan of prokaryotic genomic sequences. The first step in performing this scan is to choose a gap-free region of the ClustalW alignment output (< 60 characters in length) to be converted into a position-dependent weight matrix by entering the appropriate start and stop position information. Next, the user chooses an organism's genome sequence for scanning and whether the entire genome sequence or simply intergenic sequences, parsed based upon respective .ppt files, will be scanned. Scores generated in the scanning process reflect a normalized log-likelihood ratio statistic, known as the Kullback-Liebler distance, which can be used to estimate the statistical significance of the pattern (Stormo 2000). To determine output range, the user enters a cut-off value prior to submitting the scan. Cut-off values are empirically determined and dependent upon the weight matrix and genome sequence used in the scan. The authors found values of 70 to 90 to be useful over a range of searches. If scanning intergenic regions only, the output will exhibit the score for the match, strand, genomic sequence position, sequence match, proximate gene synonym and proximate gene description. Scans of entire genome sequences will generate output with score, strand, genomic sequence position and sequence match. An option to receive results via email is available and required for scanning most genome sequences when using Internet Explorer.

Protein purification and electrophoretic mobility shift assays

The *smtB/arsR* (MA4344) coding region was amplified by PCR from *M. acetivorans* C2A DNA using the upstream primer (5'-CGGGCATATGCAAGAAAATGCGATCG-3') and the downstream primer (5'-CCCCCTCGAGTCATATTTTTCCTCCAC-3'). The resulting PCR product was ligated into pGEM-T (Promega) prior to cloning into the *NdeI/EcoRI* sites of plasmid, pET-30b (Stratagene). Restriction enzyme analysis and DNA sequencing were performed to confirm clones at each step. After optimizing expression conditions, recombinant protein was produced in *E. coli* BL21(DE3) and crude lysates were generated by two passes through a French press cell at 124 MPa. The archaeal SmtB/ArsR was purified by binding to heparin-Sepharose (HiPrep 16/10 Heparin FF column, Amersham Biosciences) equilibrated with 50 mM Tris-HCl (pH 7.2), 1 mM EDTA, 1 mM dithiothreitol and elution was performed in the same buffer using a linear gradient to 0.5 M NaCl in a total volume of 200 ml. Further fractionation of pooled eluate was performed on a Source 15Q (Amersham Biosciences) anion exchange column in 50 mM Tris-HCl (pH 7.2), 1 mM EDTA and 1 mM dithiothreitol using a linear gradient to 1.0 M NaCl in a total volume of 200 ml. A single prominent band of the anticipated size (13.6 kDa) was

detected using denaturing polyacrylamide gel electrophoresis (PAGE).

Electrophoretic mobility shift assays (EMSA) were performed using the 2nd Generation DIG Gelshift and Detection Kit (Roche Applied Science). Two duplex 35 oligonucleotide DNA fragments (Integrated DNA Technologies), wild type (5'-AGTAATATATGAACAACACTGTTTCATACATTAATG-3') and mutant (5'-AGATTAATATAACAACACTACAGTTGTTCATTTAATG-3'), were end-labeled with digoxigenin-11-ddUTP following manufacturer instructions. Similarly, EMSA were performed following manufacturer instructions with the following additional conditions: final concentrations were adjusted to 5 mM EDTA and 2 mM dithiothreitol for the binding buffer, 75 fmole of either wild type or mutant DIG-labeled duplex DNA were used and 0.5 μ M was the final concentration of purified SmtB/ArsR in each assay. Binding reactions (20 μ l + 5 μ l of loading dye) were analyzed using 8% non-denaturing PAGE. Nucleic acid transfer to a nylon membrane (GeneScreenPlus; PerkinElmer) was performed via electroblotting (Millipore). Probe detection and development were performed according to manufacturer instructions.

Results and discussion

Prediction of chromosomal SmtB/ArsR operator binding sites

Annotation of archaeal genome sequences predicts numerous genes encoding proteins with significant sequence similarity to eubacterial SmtB/ArsR proteins. Functional orthologs within this gene family were initially assigned based upon synteny as determined by gene neighborhood analysis at STRING (Snel et al. 2000, Snel et al. 2002). In archaea, gene neighborhood relationships exhibited by *smtB/arsR* (COG-0640) appear restricted to the kingdom Euryarcheota, and reveal associations with various potential membrane-bound transport systems (COG0701, COG0477 and COG2217), which could function in metal efflux and homeostasis. Considering both the autoregulatory nature of many eubacterial SmtB/ArsR proteins and the propensity for archaeal winged-HTH proteins to act as autoregulators (Geiduschek and Ouhammouch 2005), orthologous intergenic sequences upstream of euryarchaeotal COG0640 genes were compared using PIGED.

Identical hexameric inverted repeats with a two base pair spacer (6-2-6) positioned between 7 and 21 nucleotides upstream of the annotated initiation codon were revealed through comparison of orthologous *smtB/arsR* intergenic sequences of *Methanosarcina acetivorans* C2A (MA4344), *Methanosarcina mazei* Goe1 (MM1040) and *Methanococcus maripaludis* S2 (MMP0217). A survey of unfinished euryarchaeotal genome sequences for *Methanosarcina barkeri fusaro* (Meth-4749) and *Methanococcoides burtonii* (Mbur1403) revealed orthologous *smtB/arsR* exhibiting similar hexameric inverted repeats, as well. This genetic arrangement appears absent within the draft genome sequence of the archaeon *Ferroplasma acidarmanus*, a microbe isolated from a metal-rich en-

vironment (Bond et al. 2000). Using the Seqsearch option at PIGED and regular expressions with degeneracy at different positions within the 6-2-6 operator, several additional potential binding sites were identified within intergenic sequences upstream of genes encoding SmtB/ArsR regulators in various prokaryotes. Sequences from 13 of these intergenic sequences (TTE2468, CAC2242, MM1040, MMP0217, MA4344, Mbur1403, Meth4749, CT0812, TDE1382, SSO2688, CPE-2304, sll0792 and PAB0625), were pared down to 6-2-6 inverted repeats and used to construct a position-dependent weight matrix for systematic scanning of all prokaryotic genome sequences available at PIGED.

Based on numerous searches using this scoring matrix, these putative operator sequences do not appear unique to genes encoding SmtB/ArsR among prokaryotic genome sequences. For instance, a high scoring sequence match is observed in various *Salmonella* sp. located between divergently transcribed genes encoding a putative mandelate racemase and a LacI family transcriptional regulator. Although results such as this may suggest a role for a SmtB/ArsR encoded by a distal gene in controlling expression of these genes, it is important to consider that slight variations in amino acid sequences comprising the recognition helix of SmtB/ArsR proteins can seemingly produce large differences in DNA specificity as observed in comparisons of SmtB and CadC (Busenlehner et al. 2003) or NmtR and CmtR (Cavet et al. 2003). As a result, the presence of one or more SmtB/ArsR within a genome sequence does not necessarily link these particular proteins to this operator. Results similar to those observed in *Salmonella* sp. may indicate that distinct transcriptional regulators use this operator as a binding site or the inverted repeat may act in another regulatory capacity, such as a riboswitch, depending on the organism.

These divergent possibilities pose clear obstacles for functional inference based upon identifying putative operator sequences. As a result, an initial constraint for this analysis was identification of high scoring sequence matches linked to *smtB/arsR* or other putative metal detoxification genes with appropriate positioning to serve as a *cis* regulatory element. Sequence matches meeting this linkage criterion, often exhibiting the highest score and well separated from other sequence matches, were included in the scoring matrix for subsequent scans of both intergenic sequences and whole genome sequences. Using this approach, 60 chromosomal SmtB/ArsR operators were predicted in 34 distinct eubacterial species (47 eubacterial strains) and nine archaeal species (Table 1). One eubacterial intergenic sequence (syc0262_c) was found to contain two high scoring sequence matches, while high scoring sequence matches were observed in distinct, multiple intergenic regions linked to metal detoxification genes for three eubacterial species (*B. subtilis*, *B. licheniformis* and *G. kaustophilus*). Intergenic and whole genome sequences were scanned at PIGED using scoring matrices derived from both 6-2-6 and 12-2-12 versions of these 60 predicted operator sequences to ensure a complete data set from this analysis.

Three sequences among the identified DNA elements have

been biochemically characterized as DNA binding sites or operators for SmtB/ArsR family transcriptional regulatory proteins, *Synechocystis* PCC 6803 SmtB (sll0792), *Staphylococcus aureus* CzrA (SAR2233) and *Mycobacterium tuberculosis* H37Rv NmtR (RV3744) (Busenlehner et al. 2003). Similar positioning relative to downstream genes, together with high sequence conservation within these putative binding sites, suggests an analogous mechanism for metal-dependent transcriptional repression in the additional forty prokaryotic species. Comparison of the 60 putative operator sequences within this data set reveals clear sequence conservation within SmtB/ArsR operator sequences, which is not surprising given the approach for identifying these sequences (Figure 1A). The consensus sequence for the putative SmtB/ArsR *cis* elements from diverse prokaryotes is consistent with previous in vitro DNA protection experiments and mutational studies of the binding site (Erbe et al. 1995, Turner et al. 1996). The inverted repeat nature of this operator sequence is obvious and demonstrates strong preference for specific nucleotides at several positions, with the exception of positions 3, 13, 14 and 24. In particular, nucleotides at positions 7 (T), 8 (G), 19 (C) and 20 (A) are completely conserved, implicating an importance for these bases in site-specific DNA binding of SmtB/ArsR proteins to this operator sequence.

Predicted operators linked to smtB/arsR

These results confirm a tendency for autoregulation within the SmtB/ArsR family, with approximately two-thirds of *smtB/arsR* loci exhibiting operators positioned for self-regulation. Thirty-two predicted operator sequences are located upstream of tandem-oriented genes, presumably co-transcribed, encoding SmtB/ArsR protein and putative efflux pumps (COG2217, COG1230 and COG0701; Table 1). In this orientation, the predicted operator sequences are positioned, depending on the loci, 7 nucleotides downstream or 31 nucleotides upstream of the annotated initiation codon for SmtB/ArsR. Positioning of this operator relative to downstream genes suggests an autoregulatory model to maintain metal homeostasis where SmtB/ArsR represses expression of both itself and the downstream efflux pump until elevation of cellular concentrations of a certain metal(s) increase sufficiently to result in SmtB/ArsR metal binding, release of DNA and gene expression. Subsequent decreases in intracellular metal concentration due to production of the efflux pump and potential sequestration by SmtB/ArsR lead to reassociation of non-metal bound SmtB/ArsR to the operator turning down expression.

Eleven predicted operators are present between genes oriented in a divergent manner encoding SmtB/ArsR and either a predicted efflux pump (COG2217 and COG1230), metallothionein or a reductase (COG1249) (Table 1). In most instances, the operator appears proximal to both genes, and as a result, could regulate expression of both regulator and detoxification gene(s). *Sulfolobus solfataricus* P2, *Sulfolobus tokodaii* str. 7 and *Thermoanaerobacter tengcongensis* MB4 genome sequences offer exceptions, where operators are proximal to *smtB/arsR*, but reside ~200 bp upstream of the

Table 1. Chromosomal signature of SmtB/ArsR and regulons based on predicted operator sequences.

Organism	SmtB/ArsR synonym	α^3 ^a	α^5	Operator position ^b	Predicted regulon ^c
<i>Bacillus anthracis</i> strains	GBAA0594	+	+	-28	COG0640 →(25)COG2217
	BA0594	+	+	-28	COG0640 →(25)COG2217
	BAS0563	+	+	-28	COG0640 →(25)COG2217
<i>Bacillus cereus</i> strains	BCZK0507	+	+	-28	COG0640 →(25)COG2217
	BCE0662	+	+	-28	COG0640 →(25)COG2217
	BC0595	+	+	-28	COG0640 →(25)COG2217
<i>Bacillus clausii</i> KSM-K16	ABC3390	-	+	-18	COG0640 →(12)COG1230
<i>Bacillus licheniformis</i> ATCC 14580	BLI02201	-	+	-95	COG1230(BLI03452)
	BLI00472	+	-	-41	COG2217(BLI03539)
<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	BSU19120	-	+	-85	COG1230(BSU26650)
	BSU03880	+	-	-70	COG2217(BSU33490)
<i>Bacillus thuringiensis</i> serovar konkukian str. 97-27	BT9727-0505	+	+	-28	COG0640 →(25)COG2217
<i>Chlorobium tepidum</i> TLS	CT0812	+	+	-8	COG0640 →(74)COG0701
<i>Clostridium acetobutylicum</i> ATCC 824	CAC2242	+	+	-27	COG0640 →(16)COG2217
<i>Clostridium perfringens</i> str. 13	CPE2304	+	+	-31	COG0640 →(-16)COG2217
	CPE0837	+	-		
<i>Deinococcus radiodurans</i> R1	DRA0071	+	+	-13	COG2217 →(-7) COG0695→(-3)COG0640
<i>Desulfotalea psychrophila</i> LSV54	DP0133	+	+	-27	COG0640 →(28)COG0701
<i>Desulfovibrio vulgaris</i> subsp. <i>vulgaris</i> str. Hildenborough	DVU2788	+	+	-2	COG0701(DVU3386)
<i>Enterococcus faecalis</i> V583	EF2423	-	+	-37	COG2217(EF0758)
<i>Fusobacterium nucleatum</i> subsp. <i>nucleatum</i> ATCC 25586	FN0260	+	+	-22	COG0640 →(-3)COG2217
<i>Geobacillus kaustophilus</i> HTA426	GK0580	-	+	-31	COG0640 →(15)COG1230
	GK1406	-	+	-26	COG0640 →(18)COG1230
<i>Geobacter sulfurreducens</i> PCA	GSU2149	+	-	-55	COG1609(GSU0950)
<i>Gloeobacter violaceus</i> PCC 7421	gll3429	+	+	-55	COG0640(gsr0877)
<i>Lactobacillus plantarum</i> WCFS1	LP_3287	-	+	-51	COG0640←→(139) COG1230
<i>Lactococcus lactis</i> subsp. <i>lactis</i> II1403	L133446	-	-	-17	COG0532(L0372)
<i>Listeria innocua</i> Clip11262	lin0149	-	-	-24	COG0640(lin2636)
<i>Listeria monocytogenes</i> str.	lmo0101	-	-	-25	COG0640(lmo2493)
	Lmof2365_0119	-	-	-25	COG0640(Lmof2365_2466)
<i>Methanococcoides burtonii</i>	Mbur1403	+	-	-2	COG0640 →(3)COG0701
<i>Methanococcus maripaludis</i> S2	MMP0217	+	-	-15	COG0640 →(3)COG2217
<i>Methanosarcina acetivorans</i> C2A	MA4344	+	+	-8	COG0640 →(3)COG0701
<i>Methanosarcina barkeri</i>	Meth4749	+	+	-7	COG0640 →(3)COG0701
<i>Methanosarcina mazei</i> Go1	MM1040	+	+	-8	COG0640 →(3)COG0701
<i>Methanothermobacter</i> <i>thermautotrophicus</i> str. Delta H	MTH1795	+	-	-18	COG2217(MTH411)

Continued overleaf

Table 1 Cont'd. Chromosomal signature of SmtB/ArsR and regulons based on predicted operator sequences.

Organism	SmtB/ArsR synonym	$\alpha 3^a$	$\alpha 5$	Operator position ^b	Predicted regulon ^c
<i>Mycobacterium bovis</i> AF2122/97	Mb3770	-	+	-22	COG0640 $\leftarrow\rightarrow$ (67) COG2217
<i>Mycobacterium tuberculosis</i> sp.	MT3852	-	+	-22	COG0640 $\leftarrow\rightarrow$ (22) COG2217
	RV3744	-	+	-22	COG0640 $\leftarrow\rightarrow$ (67) COG2217
<i>Nocardia farcinica</i> IFM 10152	Nfa31050	-	+	-51	COG0640 $\leftarrow\rightarrow$ (113) COG2217
<i>Nostoc (Anabena)</i> sp. PCC 7120	Alr0831	-	+	-42	NmtA (metallothionein) ^d
<i>Oceanobacillus iheyensis</i> HTE831	OB1400	-	+	-21	COG0640 \rightarrow (-3)COG1230
<i>Pyrococcus abyssi</i> GE5	PAB0625	+	+	-20	COG0640 \rightarrow (-3)COG2217
<i>Rhodospseudomonas palustris</i> CGA009	RPA3561	-	-	-42	COG1309 (RPA3696)
<i>Staphylococcus aureus</i>	SAR2233	-	+	-20	COG0640 \rightarrow (2)COG1230
	SACOL2137	-	+	-20	COG0640 \rightarrow (2)COG1230
	SAV2145	-	+	-20	COG0640 \rightarrow (2)COG1230
	SA1947	-	+	-20	COG0640 \rightarrow (2)COG1230
	SAS2048	-	+	-20	COG0640 \rightarrow (2)COG1230
	MW2069	-	+	-20	COG0640 \rightarrow (2)COG1230
<i>Staphylococcus epidermidis</i> sp.	SE1746	-	+	-23	COG0640 \rightarrow (-31)COG1230
	SERP1755	-	+	-23	COG0640 \rightarrow (-13)COG1230
<i>Streptomyces coelicolor</i> A3(2)	SCO6459	-	+	-23	COG0640 $\leftarrow\rightarrow$ (70) COG2217
<i>Sulfolobus solfataricus</i> P2	SSO2688	-	-	-246	COG0640 $\leftarrow\rightarrow$ (299) COG1249
<i>Sulfolobus tokodaii</i> str. 7	ST1076	-	-	-198	COG0640 $\leftarrow\rightarrow$ (206) COG1249
<i>Synechococcus elongatus</i> PCC 6301	Syc0262_c	-	+	-54	COG0640 $\leftarrow\rightarrow$ (101)
		-22			Syc0263_d (metallothionein)
<i>Synechocystis</i> sp. PCC 6803	Sll0792	+	+	-77	COG0640 \rightarrow (152) COG2217
<i>Thermoanaerobacter tengcongensis</i> MB4	TTE2468	+	+	-198	COG0640 \rightarrow (238) COG2217
<i>Thermosynechococcus elongatus</i> BP-1	Tlr1018	-	+	-56	tsl1016 (metallothionein)
<i>Thermus thermophilus</i> sp.	TTC0353	+	+	-20	COG0640 \rightarrow (32) COG2217
	TTHA0705	+	+	-20	COG0640 \rightarrow (32) COG2217
<i>Treponema denticola</i> ATCC 35405	TDE1382	-	+	-20	COG0640 \rightarrow (412)COG2217
<i>Wolinella succinogenes</i> DSM 1740	WS1120	+	+	-26	COG0640 \rightarrow (-13)COG2217

^a $\alpha 3$ and $\alpha 5$ denote metal binding sites associated with the SmtB/ArsR protein family based upon sequence similarity (Bushlehner et al. 2003).

^b Operator position represents the position of the highly conserved 3' CA dinucleotide within the operator relative to the predicted translational start site of bold-faced genes shown under column heading "Predicted regulon."

^c Predicted regulon describes the relationship between the predicted operator and chromosomal locus. Genes are represented by their respective COG assignment. Gene synonyms are included for operators unlinked to *smtB/arsR*. Bold-faced type denotes the gene used to measure distance to the operator sequence. \rightarrow indicates tandem orientation of genes, while $\leftarrow\rightarrow$ denotes divergent orientation. The number in parentheses represents the distance in nucleotides between genes.

^d *NmtA* is located between positions 3938083 \rightarrow 3937925 within the *Nostoc (Anabena)* sp. PCC 7120 genome sequence.

linked gene encoding the metal detoxification mechanism. Primer extension analysis shows the predicted SmtB/ArsR operator is located 240 bp upstream of the transcriptional start site for *merA* gene in *S. solfataricus*, which encodes a mercury reductase (Schelert et al. 2004). The functional significance of this predicted SmtB/ArsR operator for *S. solfataricus* MerR-mediated regulation of *merA* remains unclear. In addition, although predicted operators within *Gloeobacter violaceus* PCC

7421 and *Listeria* sp. are adjacent to COG0640 loci as observed in these other instances, these particular SmtB/ArsR exhibit no $\alpha 3$ or $\alpha 5$ metal binding sites within their respective amino acid sequences and presumably are not involved in metal sensing and tolerance.

Predicted operators unlinked to *smtB/arsR*

While a trend for SmtB/ArsR operators to be positioned to act

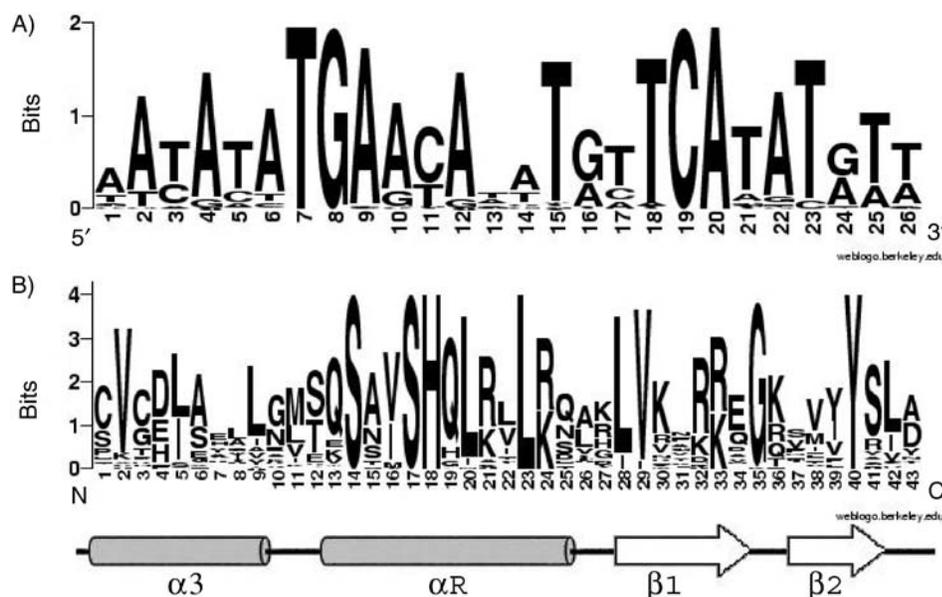


Figure 1. Signature sequences for SmtB/ArsR protein-DNA interactions. WebLogo (Crooks et al. 2004) was used to generate consensus profiles for (A) sixty SmtB/ArsR operator sites identified in this study and (B) forty-three amino acid sequences comprising the winged helix-turn-helix DNA binding motif of predicted autoregulatory SmtB/ArsR proteins. Secondary structural elements of the DNA binding motif are shown beneath the respective amino acid sequences comprising the alpha helices and beta sheets. αR represents the recognition helix of the DNA binding motif.

in an autoregulatory capacity is evident in these results, several instances of operators distal to *smtB/arsR* (COG0640) loci are identified as well. Twelve occurrences of *smtB/arsR*-distal predicted operator sequences both matching consensus (Figure 1A) and linked to putative metal detoxification genes are detected in this analysis. Genes encoding efflux pumps (COG1230, COG2217 or COG0701), which are unlinked to COG0640 genes in *Bacillus licheniformis* ATCC 14580, *Bacillus subtilis* subsp. *subtilis* str. 168, *Desulfovibrio vulgaris* subsp. *vulgaris* str. Hildenborough, *Enterococcus faecalis* V583 and *Methanothermobacter thermoautotrophicus* str. ΔH , exhibit predicted SmtB/ArsR operators between 2 and 95 nucleotides upstream of their respective annotated initiation start site. Recent transcriptome analyses examining the metal ion stress response of *B. subtilis* have shown CzrA (BSU19120) controls expression of both CadA (BSU33490) and CzcD (BSU26650) in response to a wide range of metal ions, establishing a precedent for SmtB/ArsR regulation of distal genes (Moore et al. 2005, Table 1).

Unique tandem orientations of genes and the predicted operator are observed in *Deinococcus radiodurans* R1 and strains of *Thermus thermophilus*, with the operator proximal to the gene encoding the efflux pump, rather than a linked COG0640. In addition, predicted genes encoding metallothioneins unlinked to COG0640 genes in *Nostoc* (*Anabena*) sp. PCC7120 and *Thermosynechococcus elongatus* BP-1 exhibit predicted SmtB/ArsR operator sequences 42 and 56 nucleotides upstream of their annotated translational start sites, respectively. The metallothionein found in *Nostoc* (*Anabena*) sp. PCC7120, NmtA, has not been annotated in this organism's genome sequencing project (Table 1).

A SmtB/ArsR subfamily DNA recognition sequence signature

With such a well conserved operator sequence, one might expect to observe comparable sequence conservation in amino

acid sequences comprising the winged-HTH DNA binding domain of SmtB/ArsR proteins, which is predicted to bind these sequences. Sequence comparison of the forty-three SmtB/ArsR variants that appear to be autoregulatory indicates strong sequence conservation in the winged-HTH domain (Figure 1B). The recognition helix of these proteins contains completely conserved histidine, leucine and serine residues at specific positions (Positions 14, 17, 18 and 23), while a tyrosine (Position 40) is completely conserved within the beta sheet wing (Figure 1B). Mapping this conserved DNA recognition signature sequence on both the SmtB and CadC protein structures (Cook et al. 1998, Ye et al. 2005) reveals these residues are oriented in a manner to interact with DNA, with the exception of the recognition helix leucine (Position 23 in Figure 1B), which is a key component of a hydrogen bonding network that modulates interactions between the DNA binding domain and the rest of the protein (Eicken et al. 2003).

A survey of the 1027 amino acid sequences comprising the ArsR family of winged-HTH proteins within the PFAM database (Bateman et al. 2004) identifies 63 SmtB/ArsR exhibiting this DNA recognition signature sequence (Figure 1B) in completed prokaryotic genome sequences. The *H. salinarium* SmtB/ArsR (VNG7125) was not identified in our initial analysis due to the episomal location of the gene encoding this protein and associated metal detoxification genes (Wang et al. 2004). However, subsequent sequence analysis confirms this locus exhibits a high scoring sequence match to the operator consensus with appropriate positioning (data not shown). DNA recognition signature-containing SmtB/ArsR proteins were not found in *L. lactis*, *Pseudomonas* sp. and *R. palustris* in our initial analysis as no linkage is observed between operator sequences and presumed metal-responsive genes in these genome sequences. This result is perhaps not surprising considering the absence of predicted metal-binding motifs ($\alpha 3$ or $\alpha 5$) for these winged HTH proteins present in PFAM, suggest-

ing that these SmtB/ArsR homologues may function in metal-independent regulation. The *Pseudomonas* sp. genes (PA0547, PSPTO0384 and PP4966) were excluded from subsequent analysis due to an apparent domain fusion event that has significantly altered the sequence, and possibly the activity, of these gene products. However, potential binding sites for *L. lactis* (L133446) and *R. palustris* (RPA3561) proteins based on sequence scanning and positioning are predicted (Table 1).

Candidate SmtB/ArsR regulators of genes distal from COG0640 with predicted operators in various organisms, such as *M. thermoautotrophicus*, *Desulfovibrio vulgaris* and *E. faecalis*, are found within these 63 amino acid sequences (Table 1). However, a few vague relationships between SmtB/ArsR proteins and this operator sequence (Figure 1A) are observed by identifying DNA recognition signature-containing proteins in PFAM, as well. Whereas this operator sequence (Figure 1A) has been linked to genes encoding SmtB/ArsR that lack both the DNA recognition signature and metal binding sites in *Gloeobacter violaceus* PCC 7421 (gsr0877) and *Listeria* sp. (lin2636, lmo2493 and Lmof2365_2466), these species contain distal genes encoding DNA recognition signature-containing SmtB/ArsR. Although the *G. violaceus* DNA recognition signature-containing SmtB/ArsR (gll3429) contains both $\alpha 3$ and $\alpha 5$ metal binding sites, the *Listeria* sp. proteins (lin0149, lmo0101 and Lmof23652365_0119) lack both metal binding sites. The relationships between these proteins and their putative operator sequences provide little insight into the potential physiological roles of these genes.

In addition, multiple DNA recognition signature-containing SmtB/ArsR proteins are encoded within the genome sequences of *Bacillus licheniformis* ATCC 14580, *Bacillus subtilis* str. 168, *Clostridium perfringens* str. 13 and *Geobacillus kaustophilus* HTA426. Both *G. kaustophilus* proteins exhibit $\alpha 5$ metal binding sites and appear to regulate independent genes encoding metal efflux pumps, based on gene neighborhood associations (Table 1). In contrast, only individual DNA recognition signature-containing SmtB/ArsR in *B. licheniformis* (BLI02201), *B. subtilis* (BSU19120) and *C. perfringens* (CPE2304) exhibit metal binding sites typically found within this protein family. The corresponding predicted DNA binding paralogs (BLI00472, BSU03880 and CPE0837) within each genome sequence lack the $\alpha 3$ or $\alpha 5$ metal binding sites. Although *B. subtilis* CzirA (BSU19120) has been shown to be involved in metal ion stress response, mutation of *yczG* (BSU03880) had no effect on metal tolerance (Moore et al. 2005). As a result, roles for these potential metal-independent regulators with presumed conserved DNA binding activities also remain unclear.

Despite these apparent metal-independent exceptions, correlation of the consensus operator sequence (Figure 1A) and numerous genes encoding proteins involved in metal sensing or detoxification indicates the indirect repeat sequence serves as a regulatory element in these diverse microbes, potentially as a *cis*-acting element for a DNA binding protein as observed in certain eubacteria (Busenlehner et al. 2003) or an RNA secondary structure functioning in post-transcriptional control.

To test this sequence as a potential DNA binding site in archaea and whether the sequence signatures are sufficient to predict SmtB/ArsR site-specific binding, an archaeal SmtB/ArsR homolog, *M. acetivorans* C2A protein MA4344, was heterologously produced, purified and evaluated for DNA binding activity using electrophoretic mobility shift assays. The wild type oligonucleotide substrate used in these assays correlates to DNA sequence from -32 to +3 of the MA4344 coding sequence. MA4344 binds this DNA sequence, supporting the autoregulatory model for regulation of this locus (Figure 2). In addition, specificity of MA4344 for this particular operator sequence was examined by constructing an oligonucleotide substrate with substitutions at conserved positions 7, 8, 12, 15, 19 and 20. The inability of MA4344 to bind this mutant operator sequence (Figure 2) indicates the wild type protein-DNA interaction is specific and requires certain bases at the conserved positions.

Prediction and characterization of *M. acetivorans* SmtB/ArsR site-specific binding illustrates interdomain conservation of this DNA binding activity, supports the prediction of similar activities in other archaeal species, and can implicate distal genes as potential regulons, as observed in *M. thermolithotrophicus*. Although the notion of specific amino acids correlating with specific bases can be viewed as an oversimplification due to the complexity of protein-DNA binding energies (Stormo and Fields 1998), derivation of signatures for both nucleotide and amino acid sequences involved in DNA binding offers a strategy to sort and examine large transcriptional regulatory families for potential structure, function, and physiological role inference. Similar compilation of other orthologous, perhaps autoregulatory, transcriptional regulators in archaea may facilitate development of experimental frameworks for elucidation of protein function and regulatory connections within these cells.

DNA	Wild type	Wild type	Mutant	Mutant
Protein	-	+	-	+

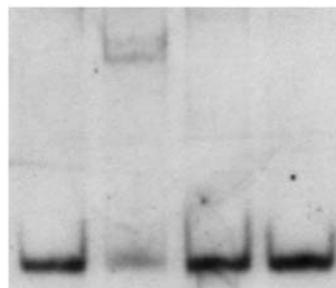


Figure 2. Electrophoretic mobility shift assays using purified archaeal SmtB. Electrophoretic mobility shift assays were performed using 75 fmole of either wild type (5'-AGATTAATATATGAACAACGTTCATACATTAATG-3') or mutant (5'-AGATTAATATAACAAC-TACAGTTGTTACATTAATG-3') digoxigenin-11-ddUTP-labeled duplex DNA and 0.5 μ M purified SmtB. Nucleotide substitutions are noted in the mutant sequence by italics.

lution of the SmtB/ArsR/CadC protein family. One model suggests CadC represents a link to a common ancestor that exhibited both metal binding sites ($\alpha 3$ and $\alpha 5$), while SmtB and ArsR have diverged through selection for one or the other metal binding site (Busenlehner et al 2002). A second model posits CadC as an intermediate between an ArsR-like ancestor and SmtB, where ArsR exhibits a form of the ancestral $\alpha 3$ binding site, CadC has acquired an additional metal binding site $\alpha 5$ and SmtB maintains a functional $\alpha 5$ binding site, but not a functional $\alpha 3$ site (Ye et al. 2005). The phylogenetic tree of DNA recognition signature proteins does little to differentiate between these two models, but exhibits distinct partitions of ArsR and SmtB/CadC proteins.

Despite the different DNA binding properties of SmtB, ArsR, and CadC proteins, this partitioning appears to be more related to the metal-binding attributes of these sequences. Sequences with the DNA recognition signature are found within both ArsR and SmtB/CadC partitions and outside of these partitions, suggesting the winged-HTH motifs exhibit a relatively high level of sequence conservation throughout this protein family. For instance, the most significant substitution in SmtB and CadC sequence comparisons of the DNA recognition signature, appears at Position 14 (Figure 1B), which is an alanine for all characterized CadC proteins. It is not known if this single substitution is sufficient to differentiate between the SmtB/ArsR operator and the CadC operator (Endo and Silver 1995, Wong et al. 2002), which exhibits transition mutations at Positions 7, 8, 19 and 20 (Figure 1A), or if other substitutions have an indirect effect altering DNA sequence binding specificity. However, phylogenetic analysis supports the notion that the SmtB/ArsR/CadC protein family has seemed to accumulate incremental changes in the DNA recognition motif, such as the substitution at Position 14, that produce affinities for distinct operator sites among these transcriptional regulators.

Acknowledgments

The authors would like to thank Dr. Suresh Chalasani for assistance with PHP programming and database development, Jacob Tatay for server support, Dr. David Higgs for use of materials in EMSA, and the Faculty Development committees at both University of Wisconsin-Parkside and DePauw University for their support. The authors greatly appreciate the efforts of numerous individuals for testing and providing critiques of PIGED. D.J.S. received funding from the National Science Foundation-REU program (Award # 0097493). M.J.S. is supported by a grant from the National Science Foundation (Award # 0333210).

References

- Abreu-Goodger, C., N. Ontiveros-Palacios, R. Ciria and E. Merino. 2004. Conserved regulatory motifs in bacteria: riboswitches and beyond. *Trends Genet.* 20:475–479.
- Bateman, A., L. Coin, R. Durbin, R.D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E.L. Sonnhammer, D.J. Studholme, C. Yeats and S.R. Eddy. 2004. The Pfam protein families database. *Nucleic Acids Res.* 32:D138–141.
- Bond, P.L., S.P. Smruga and J.F. Banfield. 2000. Phylogeny of microorganisms populating a thick, subaerial, predominantly lithotrophic biofilm at an extreme acid mine drainage site. *Appl. Environ. Microbiol.* 66:3842–3849.
- Brown, N.L., J.V. Stoyanov, S.P. Kidd and J.L. Hobman. 2003. The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* 27:145–163.
- Brunker, P., D. Rother, R. Sedlmeier, J. Klein, R. Mattes and J. Altenbuchner. 1996. Regulation of the operon responsible for broad-spectrum mercury resistance in *Streptomyces lividans* 1326. *Mol. Gen. Genet.* 251:307–315.
- Busenlehner, L.S., T.C. Weng, J.E. Penner-Hahn and D.P. Giedroc. 2002. Elucidation of primary ($\alpha 3N$) and vestigial ($\alpha 5$) heavy metal-binding sites in *Staphylococcus aureus* pI258 CadC: evolutionary implications for metal ion selectivity of SmtB/ArsR metal sensor proteins. *J. Mol. Biol.* 319:685–701.
- Busenlehner, L.S., M.A. Pennella and D.P. Giedroc. 2003. The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiol. Rev.* 27:131–143.
- Cavet, J.S., W. Meng, M.A. Pennella, R.J. Appelhoff, D.P. Giedroc and N.J. Robinson. 2002. A nickel-cobalt-sensing ArsR-SmtB family repressor. Contributions of cytosol and effector binding sites to metal selectivity. *J. Biol. Chem.* 277:38441–38448.
- Cavet, J.S., A.I. Graham, W. Meng and N.J. Robinson. 2003. A cadmium-lead-sensing ArsR-SmtB repressor with novel sensory sites. Complementary metal discrimination by NmtR and CmtR in a common cytosol. *J. Biol. Chem.* 278:44560–44566.
- Cook, W.J., S.R. Kar, K.B. Taylor and L.M. Hall. 1998. Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins. *J. Mol. Biol.* 275:337–346.
- Coombs, J.M. and T. Barkay. 2004. Molecular evidence for the evolution of metal homeostasis genes by lateral gene transfer in bacteria from the deep terrestrial subsurface. *Appl. Environ. Microbiol.* 70:1698–1707.
- Crooks, G.E., G. Hon, J.M. Chandonia and S.E. Brenner. 2004. WebLogo: a sequence logo generator. *Genome Res.* 14:1188–1190.
- Eicken, C., M.A. Pennella, X. Chen, K.M. Koshlap, M.L. VanZile, J.C. Sacchettini and D.P. Giedroc. 2003. A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ArsR zinc sensor proteins. *J. Mol. Biol.* 333:683–695.
- Endo, G. and S. Silver. 1995. CadC, the transcriptional regulatory protein of the cadmium resistance system of *Staphylococcus aureus* plasmid pI258. *J. Bacteriol.* 177:4437–4441.
- Erbe, J.L., K.B. Taylor and L.M. Hall. 1995. Metalloregulation of the cyanobacterial *smt* locus: identification of SmtB binding sites and direct interactions with metals. *Nucleic Acids Res.* 23:2472–2478.
- Gajiwala, K.S. and S.K. Burley. 2000. Winged helix proteins. *Curr. Opin. Struct. Biol.* 10:110–116.
- Gajiwala, K.S., H. Chen, F. Cornille, B.P. Roques, W. Reigh, B. Mach and S.K. Burley. 2000. Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403:916–921.
- Geiduschek, E.P. and M. Ouhammouch. 2005. Archaeal transcription and its regulators. *Mol. Microbiol.* 56:1397–1407.
- Gelfand, M.S., E.V. Koonin and A.A. Mironov. 2000. Prediction of transcription regulatory sites in archaea by a comparative genomic approach. *Nucleic Acids Res.* 28:695–705.
- Huckle, J.W., A.P. Morby, J.S. Turner and N.J. Robinson. 1993. Isolation of a prokaryotic metallothionein locus and analysis of transcriptional control by trace metal ions. *Mol. Microbiol.* 7:177–187.

- Kumar, S., K. Tamura and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform.* 5:150–163.
- Lawrence, J.G. and H. Ochman. 1997. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* 44:383–397.
- McCue, L.A., W. Thompson, C.S. Carmack, M.P. Ryan, J.S. Liu, V. Derbyshire and C.E. Lawrence. 2001. Phylogenetic footprinting of transcription factor binding sites in proteobacterial genomes. *Nucleic Acids Res.* 29:774–782.
- Mergeay, M., S. Monchy, T. Vallaey, V. Auquier, A. Benotmane, P. Bertin, S. Taghavi, J. Dunn, D. van der Lelie and R. Wattiez. 2003. *Ralstonia metallidurans*, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. *FEMS Microbiol. Rev.* 27:385–410.
- Moore, C.M., A. Gaballa, M. Hul, R.W. Ye and J.D. Helmann. 2005. Genetic and physiological responses of *Bacillus subtilis* to metal ion stress. *Mol. Microbiol.* 57:27–40.
- Munch, R., K. Hiller, H. Barg, D. Heldt, S. Linz, E. Wingender and D. Jahn. 2003. PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res.* 31:266–269.
- Pennella, M.A., J. E. Shokes, N.J. Cosper, R.A. Scott and D.P. Giedroc. 2003. Structural elements of metal selectivity in metal sensor proteins. *Proc. Natl. Acad. Sci. USA* 100:3713–3718.
- Perron, K., O. Caille, C. Rossier, C. van Delden, J. Dumas and T. Kohler. 2004. CzrC-CzcS, a two-component system involved in heavy metal and carbapenem resistance in *Pseudomonas aeruginosa*. *J. Biol. Chem.* 279:8761–8768.
- Rajewsky, N., N.D. Succi, M. Zapotocky and E.D. Siggia. 2002. The evolution of DNA regulatory regions for proteo-gamma bacteria by interspecies comparisons. *Genome Res.* 12:298–308.
- Ray, W.C. and C.J. Daniels. 2003. PACRAT: a database and analysis system for archaeal and bacterial intergenic sequence features. *Nucleic Acids Res.* 31:109–113.
- Schelert, J., V. Dixit, V. Hoang, J. Simbahan, M. Drozda and P. Blum. 2004. Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption. *J. Bacteriol.* 186:427–437.
- Schirawski, J., W. Hagens, G.F. Fitzgerald and D. Van Sinderen. 2002. Molecular characterization of cadmium resistance in *Streptococcus thermophilus* strain 4134: an example of lateral gene transfer. *Appl. Environ. Microbiol.* 68:5508–5516.
- Snel, B., G. Lehmann, P. Bork and M.A. Huynen. 2000. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* 28:3442–3444.
- Snel, B., P. Bork and M.A. Huymen. 2002. The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci. USA* 99:5890–5895.
- Stormo, G.D. 2000. DNA binding sites: representation and discovery. *Bioinformatics* 16:16–23.
- Stormo, G.D. and D.S. Fields. 1998. Specificity, free energy and information content in protein-DNA interactions. *Trends Biochem. Sci.* 23:109–113.
- Studholme, D.J. and R.N. Pau. 2003. A DNA element recognized by the molybdenum-responsive transcription factor ModE is conserved in proteobacteria, green sulphur bacteria and archaea. *BMC Microbiol.* 3:24–33.
- Studholme, D.J., S.D. Bentley and J. Kormanec. 2004. Bioinformatic identification of novel regulatory DNA sequence motifs in *Streptomyces coelicolor*. *BMC Microbiol.* 4:14–25
- Studholme, D.J. and R. Dixon. 2004. In silico analysis of the sigma54-dependent enhancer-binding proteins in *Pirellula* sp. Strain 1. *FEMS Microbiol. Lett.* 230:215–225.
- Tan, K., G. Moreno-Hagelsieb, J. Collado-Vides and G.D. Stormo. 2001. A comparative genomics approach to prediction of new members of regulons. *Genome Res.* 11:566–584.
- Tatusov, R.L., N.D. Fedorova, J.D. Jackson, A.R. Jacobs, B. Kiryutin, E.V. Koonin, D.M. Krylov, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41–54.
- Thompson, J.D., D.G. Higgins and T.J. Gibson. 1994. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
- Tucker, N.P., B. D'Autreaux, D.J. Studholme, S. Spiro and R. Dixon. 2004. DNA binding activity of the *Escherichia coli* nitric oxide sensor NorR suggests a conserved target sequence in diverse proteobacteria. *J. Bacteriol.* 186:6656–6660.
- Turner, J.S., P.D. Glands, A.C. Samson and N.J. Robinson. 1996. Zn²⁺-sensing by the cyanobacterial metallothionein repressor SmtB: different motifs mediate metal-induced protein-DNA dissociation. *Nucleic Acids Res.* 24:3714–3721.
- van Helden, J. 2003. Regulatory sequence analysis tools. *Nucleic Acids Res.* 31:3593–3596.
- Wang, G., S.P. Kennedy, S. Fasiludeen, C. Rensing and S. DasSarma. 2004. Arsenic resistance in *Halobacterium* sp. Strain NRC-1 examined by using an improved gene knockout system. *J. Bacteriol.* 186:3187–3194.
- Wang, Y., L. Hemmingsen and D.P. Giedroc. 2005. Structural and functional characterization of *Mycobacterium tuberculosis* CmtR, a PbII/CdII-sensing SmtB/ArsR metalloregulatory repressor. *Biochemistry* 44:8976–8988.
- Wong, M.D., Y.F. Lin and B.P. Rosen. 2002. The soft metal ion binding sites in the *Staphylococcus aureus* pI258 CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor are formed between subunits of the homodimer. *J. Biol. Chem.* 277:40930–40936.
- Wu, J. and B.P. Rosen. 1993. Metalloregulated expression of the *ars* operon. *J. Biol. Chem.* 268:52–58.
- Xu, C., W. Shi and B.P. Rosen. 1996. The chromosomal *arsR* gene of *Escherichia coli* encodes a trans-acting metalloregulatory protein. *J. Biol. Chem.* 271:2427–2432.
- Xu, C. and B. P. Rosen. 1999. Metalloregulation of Soft Metal Resistance Pumps. *In* Metals and Genetics. Ed. B. Sarkar. Plenum Press, New York, pp 5–19.
- Ye, J. A. Kandegedara, P. Martin and B.P. Rosen. 2005. Crystal structure of the *Staphylococcus aureus* pI258 CadC Cd(II)/Pb(II)/Zn(II)-responsive repressor. *J. Bacteriol.* 187:4214–4221.