



# Research Journal of **Microbiology**

ISSN 1816-4935



Academic  
Journals Inc.

[www.academicjournals.com](http://www.academicjournals.com)

## Structural Insight into the Functionality of the Transcriptional Regulator SoxR from *Paracoccus pantotrophus* in Sulfur Oxidation Operon

<sup>1</sup>Angshuman Bagchi and <sup>2</sup>Tapash Ghosh

<sup>1</sup>Department of Biochemistry and Biophysics,

University of Kalyani, Kalyani, Nadia, West Bengal-741 235, India

<sup>2</sup>Bioinformatics Center, Bose Institute, AJC Bose Centenary Building,  
P1/12 CIT Scheme VIIM, Kolkata 700 054, India

**Abstract:** The redox reactions involving inorganic sulfur compounds in the environment are one of the major reactions of the global sulfur cycle mediated by prokaryotes of different phylogenetic groups. The Sulfur oxidizing gene cluster (Sox) of  $\alpha$ -Proteobacteria comprises of at least 15 genes, which form two transcriptional units, viz., *SoxSRT* and *SoxVWXYZABCDEFGH*. Biochemical studies reveal that SoxR belongs to the ArsR family of helix-turn-helix DNA binding proteins. Although SoxR proteins do not contain the conserved metal-binding box ELCVCDL as observed in case of other well studied ArsR family proteins, but there are a number of well conserved residues present throughout the sequence that are previously identified in other known ArsR family proteins collectively called atypical ArsR proteins. In the present study homology modeling is employed to construct the three-dimensional structure of the SoxR from  $\alpha$ -Proteobacteria *Paracoccus pantotrophus*. The predicted homology model of SoxR shows an overall structural similarity with winged helix-turn-helix family of DNA binding proteins. Since SoxR is a DNA binding protein we docked the modeled SoxR on to its promoter DNA sequence to predict the probable mode of interaction between them. Present studies are expected to contribute in designing experiments for future genetic and biochemical experiments to elucidate the role of different residues in classification of SoxR as an atypical ArsR family of proteins as well as DNA binding that would help to understand the regulation of *sox* gene expression in sulfur lithotrophs of  $\alpha$ -Proteobacteria family.

**Key words:** Homology modeling, DNA-protein interaction, sulfur oxidation, ArsR

### INTRODUCTION

Microbial redox reactions of sulfur contribute significantly in the cycling of this element in the environment. The range of oxidation states, +6 to -2, of sulfur is unique and as a result several important biological processes involving transformations of sulfur from one form to other have been evolved. Sulfur based chemo-or photolithotrophy is one of such processes in which electron transfer from reduced sulfur compounds is used by phylogenetically diverse bacterial and archaeal chemolithoautotrophic species (Friedrich, 1998). Sulfide, polysulfide, elemental sulfur, thiosulfate, polythionates and sulfites are the different forms of sulfur in the environment (Le Faou *et al.*, 1990), which serve as the electron donors in the process of respiration or photosynthesis of the sulfur oxidizing prokaryotes. Only little is understood about the molecular mechanism of this oldest known lithotrophic process.

Recent studies with both chemo-and photolithotrophic  $\alpha$ -proteobacteria such as *Paracoccus pantotrophus* (Para) and *Rhodovulum sulfidophilum* revealed that multiple-gene cluster,

**Corresponding Author:** Angshuman Bagchi, Department of Biochemistry and Biophysics, University of Kalyani, Kalyani, Nadia, West Bengal-741 235, India

*shxVW* (*soxVW*) and *soxXYZABCDEFGH* (Appia-Ayme *et al.*, 2001; Rother *et al.*, 2001) is associated with the sulfur lithotrophic metabolism. The first gene of the *sox* gene region, *soxR*, previously designated as ORF1, predicts a transcriptional regulator of the ArsR family and *soxS* (formerly ORF2) a periplasmic thioredoxin (Rother *et al.*, 2005). Both these genes are oriented divergently to the other genes of *sox* cluster (Friedrich *et al.*, 2001, 2005). Though a considerable progress in the genetics of sulfur lithotrophy is noted, the details of the molecular mechanism of regulation of *sox* gene expression have not yet been addressed. In the present study the goal is to understand the structural basis of the transcriptional regulation of *sox* operon by SoxR protein from Para. For that matter the three-dimensional structure of SoxR is described. Since other well known transcriptional regulators of the ArsR family proteins, having the distinct HTH motifs for DNA binding, functions as a dimer (for example SmtB (Cook *et al.*, 1998)), the dimeric model of SoxR protein have been generated and this model of the protein have been used to predict the probable mode of interactions responsible for the generation of the dimer. Both the monomeric and dimeric models of SoxR protein were docked on to the promoter DNA sequence of Para to elucidate the biochemical basis of the involvement of the protein in DNA binding. Generally the typical ArsR family proteins bear a signature sequence of a metal ion-binding box. In this case, the SoxR protein does not contain the conserved signature sequence of the metal ion-binding region. In order to reveal the structural basis of the inclusion of SoxR in the ArsR family proteins, the sequence of SoxR protein was analyzed and compared to other well-characterized atypical ArsR family proteins. The so-called atypical ArsR family proteins do not bear the standard signature sequence of the metal binding box. This study would help to elucidate the probable biochemical mechanism of the binding of SoxR on to its promoter DNA as well as the mechanism of regulation of *sox* operon in Para.

## MATERIALS AND METHODS

### Sequence Analysis

Biochemical and sequence analysis of SoxR protein showed that SoxR bears a significant sequence homology with a large numbers other pro and eukaryotic metalloregulatory proteins of ArsR family having conserved HTH motif. Moreover the HTH motif of Zn binding repressor protein SmtB from *Synechococcus* PCC7942 (PDB code: 1SMT) shows 40% sequence similarity with the similar motif of SoxR protein. The active form of this family protein is a dimer, which binds to their specific promoter sequence (Chun and Rosen, 1997; Cook *et al.*, 1998). The sequence analysis was performed using the software BLAST.

### Molecular Modelling

The monomeric model of SoxR was first built using one monomer of SmtB as a template using the program Modeller (Šali *et al.*, 1993). Two monomeric models of SoxR were superimposed separately on the corresponding monomeric units of the crystal structure and subsequently merged together to form the dimeric model of SoxR. The r.m.s deviations for superimposition for each monomer were 0.7Å on the corresponding monomeric templates of SmtB.

### Energy Refinement of the Predicted Structure

The dimeric model of SoxR thus built was subjected to energy minimization using the DISCOVER program (MSI/Accelrys) with consistent valence force field (cvff) (Dauber-Osguthorpe *et al.*, 1988) until the structure reached the derivative of 0.001 kcal mol<sup>-1</sup>. Procheck (Laskowski *et al.*, 1993) analysis was performed in order to identify the stereo chemical qualities of the final model and Ramachandran plots were drawn. Residue profiles were checked with Verify 3D (Eisenberg *et al.*, 1997). All these indicated a good quality of the model.

### Modelling of DNA-Protein Interaction

The promoter regions of the *sox* gene cluster were detected in the 69 bp intergenic region between *soxS* and *soxV* and in a second 114 bp intergenic region between *soxW* and *soxX* (GenBank Accession No. X79242) (Rother *et al.*, 2005). The *soxSRT* genetic organization in the *sox* locus is typical of regulatory function to control the upstream *sox* structural gene cluster. This region is analogous with the DtxR-recognized consensus DNA sequence. The 19 bp core promoter region covering the 20-residue HTH motif in DtxR was extracted from the crystal structure (PDB code: 1F5T) (Chen *et al.*, 2000). HTH motif of SoxR was superimposed on the HTH motif of DtxR with r.m.s deviation 0.6Å. The promoter sequence of Diphtheria toxin repressor was replaced by the promoter sequence of SoxR. The docking was performed using both the dimeric as well as the monomeric models of the SoxR protein. SoxR-DNA models thus formed were then subjected to constrained energy minimization. In the first step of the minimization process the DNA of the DNA-protein complex was kept fixed and the protein molecule was allowed to move. In the next step both the DNA and the protein were allowed to move. The energy minimization processes were carried out using the program DISCOVER (MSI/Accelrys) with the cvff force field.

## RESULTS AND DISCUSSION

### Description of the Structure

The modeled SoxR is an alpha beta protein. The monomeric model of SoxR starts with an antiparallel  $\beta$ -sheets (amino acid residues 11-14 and 26-28) joined by loops and bends. Then there are five successive  $\alpha$ -helices (amino acid residues 52-60, 65-69, 73-84, 89-96 and 100-112), which terminate in another antiparallel  $\beta$ -sheets (amino acid residues 116-119 and 126-129) (Fig. 1). Two of the helices (residues 89-96 and 100-112) form the standard helix-turn-helix (HTH) motif seen in many prokaryotic and eukaryotic DNA binding proteins. There is a three-residue loop between the two recognition helices also common for DNA binding proteins (Cook *et al.*, 1998). The HTH motif together with the anti-parallel beta sheet forms a structure known as beta hairpin. This hairpin structure occurs immediately after the recognition helix and is anchored by hydrophobic and hydrogen bonding interactions.

The dimeric model of SoxR is an elongated homodimer (Fig. 2). The structure shares similarity to both eukaryotic and prokaryotic winged helix DNA-binding proteins (Cook *et al.*, 1998). In the



Fig. 1: Ribbon representation of modeled SoxR monomer.  $\alpha$ -helices and  $\beta$ -sheets are shown as helices (red) and ribbon (cyan), respectively. The rest are shown as loops (white)



Fig. 2: Ribbon representation of modeled SoxR dimer.  $\alpha$ -helices and  $\beta$ -sheets are shown as helices (red) and ribbon (cyan), respectively. The rest are shown as loops (white)

dimeric structure, the residues of the other three helices from each monomer form a hydrophobic core. This serves primarily as scaffolding to correctly orient the HTH motif. The residues involved in the monomeric interface between the SoxR proteins in the dimeric model are primarily from the six scaffold helices. These hydrophobic residues at the beginning and end of the sheet (Leu115, Ile125, Leu130) are present in the hydrophobic core of the dimer. There are a total of 15 residues from each monomer that is 30 residues are present in the monomer interface. There are extensive hydrophobic as well as hydrogen bonding interactions between the monomers.

#### DNA-Protein Interaction

In order to study the interactions between the promoter DNA and the SoxR protein, both the monomeric and dimeric models of the SoxR have been docked onto the DNA. The interactions are found to be similar in both the cases, though the number of interacting residues are found to be more in case when the dimeric model of SoxR have been docked to DNA than in case of the monomeric SoxR. Most of the DNA binding interactions of the modeled SoxR protein occur via the amino acid residues of the HTH motif (amino acid residues 89-96 and 100-112). The interactions occurring with the DNA are mediated either by the phosphate backbone atoms or through the specific DNA bases. Residues Gln100, Arg111, Gln 100 of the first helix and Leu95 and Ser97 of the second helix of HTH motif of each subunit are involved in H-bonding. The side chains of Gln100 from each subunit forms H-bonds with Adenine 32 and Adenine 33, respectively. Side chains of Ser97 and Tyr127 are involved in hydrogen bonding with Adenine 32 and Thymine 37, respectively. The main chains of Leu95 and Leu96 form hydrogen bonding with Cytosine 31 and Adenine 32 of DNA. Gln105 and Gln100 are involved in hydrogen bonding with the phosphate groups of DNA backbone.

#### Comparison with Winged HTH Proteins

According to the structural classification of the HTH-DNA binding domains by Wintjens and Rooman (1996), the proteins are grouped according to the spatial arrangements of the helix triplets  $H_{R,2}$  to  $H_R$  or  $H_{R-1}$  to  $H_{R+1}$  (Fig. 3). When the coordinates of SoxR were used to search for structural homologs using DALI, the results showed that SoxR had strong structural similarities to all three families: Cyanobacterial metalloregulatory repressor SmtB, diphtheria toxin repressor DtxR, catabolite gene activator protein CAP (Cossart and Gicquel-Sanzey, 1982) although the sequence homology is less

```

---HR-3---          -----HR-2-----          -----HR-1-
smt      F F A V L A D P N R L R L L S L L A R S E L C V G D L A A
SoxR     F L K A L G H D G R L M I L C Y L M S G P K S V T E L E N

-----HR-----
smt      Q A I G V S E S A V S H Q L R S
SoxR     - L L S S R Q A V V S Q Q L A L

-----B1-----          -----B2---
smt      L R N L R L V S Y R K Q G R H V Y Y Q L Q
SoxR     R R H E G L V S A R R D G Q T I F Y S I -
    
```

Fig. 3: Sequence alignment of SoxR with SmtB. Bars show the alpha helices and beta strands in the proteins over sequence. The helices are labeled according to their position in the sequence relative to the putative recognition helix in the HTH motif ( $H_R$ )



Fig. 4: Superimposition of the  $\alpha$ -carbon backbones of HTH DNA binding motif of SoxR with SmtB, DtxR and CAP

among representatives of the HTH family proteins with SoxR. Superimpositions of the three helices  $H_{R-2}$  to  $H_R$  in SoxR with the corresponding three helices in SmtB, DtxR and CAP give RMS values of 0.3, 0.6, 0.9 Å respectively (Fig. 4). Furthermore, the angles between helix pairs and distances of closest approach between helix axis are most similar for SoxR, SmtB and toxin repressor. This suggests that SoxR belongs to the toxin repressor family in the classification by Wintjens and Rooman (1996).

#### Comparison of SoxR with the Well Characterized ArsR Sequences

ArsR proteins are a class of repressor that controls the basal level of expression of *ars* structural genes in response to the presence of arsenite. ArsR proteins contain a well-conserved metal-binding box, ELCVCDL to which the arsenite inducer binds (Butcher and Rawlings, 2002). This metal binding signature sequence is also present in SmtB, which precedes the first helix of the HTH motif of SmtB (Cook *et al.*, 1998). However this conserved signature is not present in the amino acid sequence of SoxR. Among the five most homologous proteins of SoxR belonging to ArsR family none contained the conserved cysteine residues in its metal-binding domain (ELCVCDL). These proteins are therefore called atypical ArsR family proteins. However there were a number of residues that were conserved in these proteins including well-studied ArsR family protein of *Acidithiobacillus ferrooxidans* (*At. ferrooxidans*) (Fig. 5). These proteins including SoxR contain some conserved amino acid residues (Butcher and Rawlings, 2002). In order to function, as a transcriptional regulator SoxR should bind to some inducer as is evident from the other members of this ArsR family proteins (as for an example SmtB which requires Zn for its activity). Though the actual inducer or the inducer-binding site in SoxR is not known, that conserved region could be the potential binding site of *sox* operon inducer.

```

NP522      ME-----ETDAIRSLAALAHSLRLRVFRMLVVAGPAGLTPGAIAEQLDVPG--ATLSFH
NP437      ME-----ERQARSSFAALSQETRLGIVRTLVVAGPVGMAAGAIAEKMGVSP--SNVSFH
NP354      MD-----NIGAIAALSALAOTTRLETFRLLVQHEPEGIPAGELARLIHVPQ--NTMSAH
At.        MEP--LQDPAQIVARLEALASPVRLEIFRLLVEQEPTGLVSGDIAEHLGQPH--NGISFH
NP486      MQTPTATVPHFIAAGFHALSDPIRINVLELLRQRE---LCVCDLCEALGVSQ--SKLSFH
SoxR      MIP----APWIRDKLAQSSGWGCRAPAYPAFRKEGNGDEPLGGIENIRQEPEQEADFDAL

NP522      LKELMNA--LVTQERDGRHLIYRAAFD---HMNGLLGFLTENCCQGEACPEFS-AAG--
NP437      LKELERTG--LITQRRESRSIVYSASYD---ALADLVKFLMEDCCVGHASVRAS-LNEQN
NP354      LATLARAG--LVKSERQSRSIYRADLE---GLRALTLFLLKDCCGGATELCAPLIAELT
At.        LKNLQHAG--LVTVQREGRYQRYRAAMP---VVRALVAYLTENCCHGTRDCALS--GETR
NP486      LKTLKEAS--LVNSRQQRWIYYSLNLP---QFAVLEKYLTDYQHFQVLPARCCDNK--
SoxR      IKAANDASTFLKALGHDGRLMILCYLMSGPKSVTELENLLSSRQAVVSQQLARLRHEGLV

NP522      -CKC-----
NP437      PCRVASSTPAEGGNG-----
NP354      PCCCQEAEAL-----
At.        SPSVQEGNQ-----
NP486      -----
SoxR      SARRDGGQTIFYSILDPKVVDLLAVLKKLFETDC
    
```

Fig. 5: Multiple-sequence alignment of atypical Ars proteins. Conserved amino acids are presented in bold face. The alignment was performed using CLUSTALW (19). At represents *At. ferrooxidans* (GenBank accession no. AAF69241). NP\_522 from *Ralstonia solanacearum* (GenBank accession No. NP\_522690), NP\_486 from *Nostoc* sp. (Gen Bank accession No. NP\_486806), NP\_437 from *Sinorhizobium meliloti* pSymB (GenBank accession No. NP\_437556), NP\_430 from *Caulobacter crescentus* (GenBank accession No. NP\_420316), NP\_354 from *Agrobacterium tumefaciens* (GenBank accession No. NP\_354498)

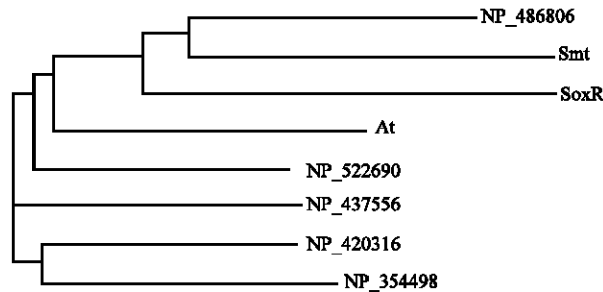


Fig. 6: Dendrogram of different ArsR family proteins along with SoxR and SmtB. This represents the possible grouping of *At. ferrooxidans* and SoxR protein in the same subclass of atypical ArsR regulators

The sequences of the SoxR protein and its five closest matches clustered as a subgroup upon homology analysis (Fig. 6). All members of this subgroup shared conserved regions among themselves and with SmtB in the absence of a typical metal-binding domain. However, that atypical *At. ferrooxidans* ArsR protein is able to regulate the arsenic-resistance genes in response to arsenic (Butcher and Rawlings, 2002), owing to the similarity with the *At. ferrooxidans* ArsR protein we predict SoxR as an atypical ArsR protein.

## CONCLUSIONS

In this study we have made an attempt to identify the structural basis of involvement of SoxR in the regulation of *sox* operon comparing with previously enumerated structural features for metalloregulatory proteins. We describe the three dimensional structure of the monomeric as well as

homodimeric SoxR. We investigated the mechanistic details of the possible dimerization of SoxR as well as the SoxR-DNA interactions. We identified the putative inducer-binding domain in order to properly classify the SoxR as a metalloregulatory protein

Finally we note that our studies will help in designing experiments for future genetic and biochemical work in order to elucidate the roles of the different residues in dimerization as well as DNA binding of SoxR that would enable in understanding the transcription regulation of *sox* gene expression in sulfur-oxidizing  $\alpha$ -Proteobacteria.

## REFERENCES

- Appia-Ayme, C., P.J. Little, Y. Matsumoto, A.P. Leech and B.C. Berks, 2001. Photosynthetic oxidation of both thiosulfate and sulfide in *Rhodovulum sulfidophilum*. J. Bacteriol., 183: 6107-6118.
- Butcher, B.G. and D.E. Rawlings, 2002. The divergent chromosomal ars operon of *Acidithiobacillus ferrooxidans* is regulated by an atypical ArsR protein. Microbiology, 148: 3983-3992.
- Chen, S., A. White, J. Love, J.R. Murphy and D. Ringe, 2000. Methyl groups of thymine bases are important for nucleic acid recognition by DtxR. Biochemistry, 39: 10397-10407.
- Chun, Xu and B.P. Rosen, 1997. Dimerization is essential for DNA binding and repression by the ArsR metalloregulatory protein of *Escherichia coli*. J. Biol. Chem., 272: 15734-15738.
- 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.
- Cossart, P. and B. Gicquel-Sanzey, 1982. Cloning and sequence of the *crp* gene of *Escherichia coli* K 12. Nucleic. Acids Res., 10: 1363-1378.
- Dauber-Osguthorpe, P., V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest and A.T. Hagler, 1988. Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase trimethoprim, a drug receptor system. Proteins, 4: 31-47.
- Eisenberg, D., R. Luthy and J.U. Bowie, 1997. VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol., 277: 396-404.
- Friedrich, C.G., 1998. Physiology and genetics of sulfur-oxidizing bacteria. Adv. Microbiol. Physiol., 39: 235-289.
- Friedrich, C.G., D. Rother, F. Bardischewsky, A. Quentmeier and J. Fischer, 2001. Oxidation of reduced inorganic sulfur compounds by bacteria: Emergence of a common mechanism? Applied Environ. Microbiol., 67: 2873-2882.
- Friedrich, C.G., F. Bardischewsky, D. Rother, A. Quentmeier and J. Fischer, 2005. Prokaryotic sulfur oxidation. Curr. Opin. Microbiol., 8: 253-259.
- Laskowski, R.A., M.W. McArthur, D.S. Moss and J.M. Thornton, 1993. PROCHECK: A program to check the stereochemistry of protein structures. J. Applied Crystallogr., 26: 283-291.
- Le Faou, A., B.S. Rajagopal, L. Daniels and G. Fauque, 1990. Thiosulfate, polythionates and elemental sulfur assimilation and reduction in the bacterial world. FEMS Microbiol. Rev., 6: 351-381.
- Rother, D., H.J. Henrich, A. Quentmeier, F. Bardischewsky and C.G. Friedrich, 2001. Novel genes of the *sox* gene cluster, mutagenesis of the flavoprotein SoxF and evidence for a general sulfur-oxidizing system in *Paracoccus pantotrophus* GB17. J. Bacteriol., 183: 4499-4508.
- Rother, D., G. Orawski, F. Bardischewsky and C.G. Friedrich, 2005. SoxRS-mediated regulation of chemotrophic sulfur oxidation in *Paracoccus pantotrophus*. Microbiology, 151: 1707-1716.
- Šali, A. and T.L. Blundell, 1993. Comparative protein modelling by satisfaction of spatial constraints. J. Mol. Biol., 234: 779-815.
- Wintjens, R. and M. Rooman, 1996. Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. J. Mol. Biol., 262: 294-313.