

**Baisakhee Saha, Somnath  
 Mukherjee, Debajyoti Dutta and  
 Amit Kumar Das\***

Department of Biotechnology, Indian Institute of  
 Technology, Kharagpur 721 302, India

Correspondence e-mail:  
 amitk@hijli.iitkgp.ernet.in

Received 23 October 2008  
 Accepted 22 December 2008

## Expression, purification, crystallization and preliminary X-ray diffraction analysis of the transcriptional repressor SirR from *Mycobacterium tuberculosis* H37Rv

SirR, a metal-dependent transcriptional repressor from *Mycobacterium tuberculosis* (Rv2788), was cloned in pQE30 expression vector with an N-terminal His<sub>6</sub> tag for heterologous overexpression in *Escherichia coli* M15 (pREP4) cells and purified to homogeneity using chromatographic procedures. The purified protein was crystallized using the sitting-drop vapour-diffusion technique. The crystals belonged to the tetragonal space group  $P4_12_1/P4_32_12$ , with unit-cell parameters  $a = 105.21$ ,  $b = 105.21$ ,  $c = 144.85$  Å. The X-ray diffraction data were processed to a maximum resolution of 2.5 Å. The Matthews coefficient suggests the presence of two ( $V_M = 4.01$  Å<sup>3</sup> Da<sup>-1</sup>) to four ( $V_M = 2.0$  Å<sup>3</sup> Da<sup>-1</sup>) molecules in the asymmetric unit. Calculation of the self-rotation function shows a crystallographic fourfold symmetry axis along the  $z$  axis ( $\chi = 90^\circ$ ) and also a twofold symmetry axis around the  $z$  axis ( $\chi = 180^\circ$ ).

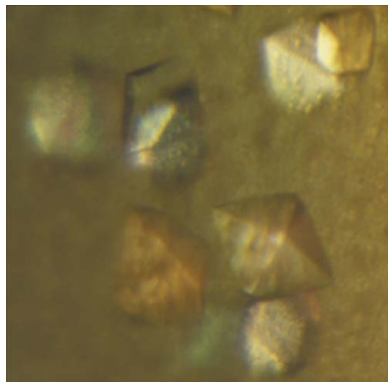
### 1. Introduction

Tuberculosis still continues to wreak havoc on mankind by threatening one third of the world's population annually. Without a coordinated control effort, tuberculosis will infect an estimated one billion more people by 2020, killing 70 million (<http://www.who.int>). To act as a successful pathogen, the tubercle bacillus *Mycobacterium tuberculosis* depends on iron acquisition. Specialized iron-regulatory proteins are key players in the process of iron homeostasis. These regulatory proteins aid the pathogen to respond to changing iron concentrations at the transcriptional level and belong to two distinct groups: the Fur and the DtxR families. Annotation of the *M. tuberculosis* genome sequence (Cole *et al.*, 1998) revealed two ferric uptake regulator-like (Fur-like) proteins, FurA and FurB, and two diphtheria-toxin repressor (DtxR) homologues: the iron-dependent regulator IdeR and the staphylococcal iron-regulator repressor SirR (Rodriguez & Smith, 2003).

FurA is the negative regulator of KatG (Zahrt *et al.*, 2001). In addition to being autoregulated (Sala *et al.*, 2003), its transcription is upregulated by oxidative stress (Milano *et al.*, 2001). Structural and functional characterization of FurB revealed it to be a Zn<sup>2+</sup>-dependent zinc-uptake regulator (Zur) that controls the genes responsible for bacterial zinc uptake (Lucarelli *et al.*, 2007).

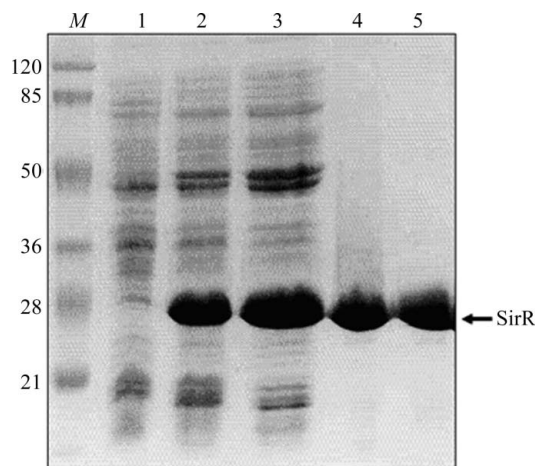
IdeR, one of the two DtxR homologues, binds to the operator region of siderophore- and virulence-factor-encoding genes in the presence of reduced iron (Gold *et al.*, 2001). This metalloregulator is essential since it cannot be disrupted in the mycobacterium in the absence of its second functional copy (Rodriguez *et al.*, 2002). IdeR binds metal ions such as iron, manganese, zinc, cobalt, nickel and magnesium and also regulates as many as 40 genes that are involved in vital physiological functions such as exochelin synthesis and transport, iron storage, siderophore biosynthesis and the activities of putative transporters, transcriptional regulators, general metabolic proteins, members of the PE/PPE family of conserved mycobacterial proteins and the virulence determinant MmpL4 (Rodriguez & Smith, 2003).

Metal-dependent transcriptional regulators are homodimers, with each subunit containing an N-terminal DNA-binding domain and an

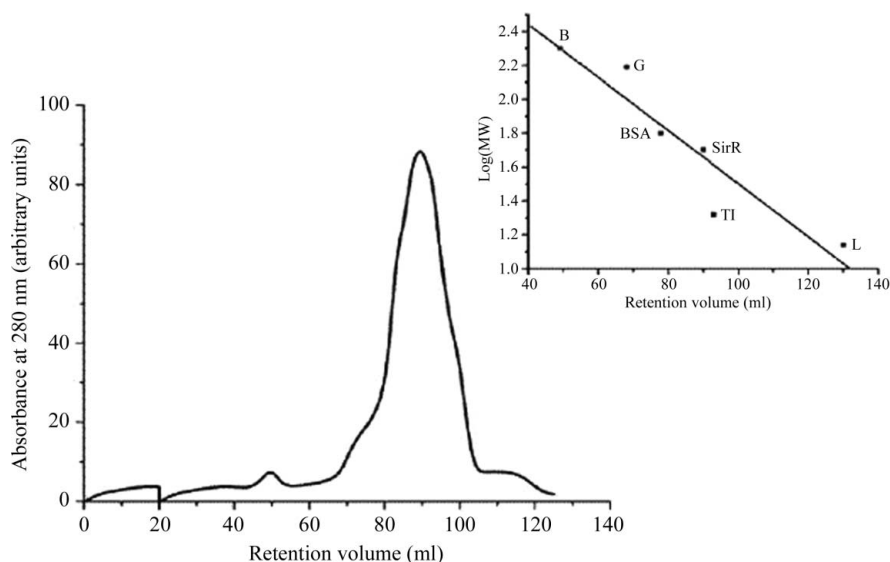


adjacent metal-binding domain. Their homodimeric nature allows interaction with DNA, mainly at palindromic or pseudo-palindromic regions, in which a helix–turn–helix protein domain is located in the major groove of the oligonucleotide (Huffman & Brennan, 2002). IdeR has a helix–turn–helix motif in the DNA-binding domain, metal-binding residues in the dimerization domain and a C-terminal SH3 (Src-homology domain 3) domain (Pohl *et al.*, 1999; Feese *et al.*, 2001). IdeR binds to the DNA as a double dimer, with each dimer on an opposite side of the DNA (Wisedchaisri *et al.*, 2004).

However, many iron-regulated genes in the pathogen escape modulation by IdeR. These are speculated to be under the control of SirR, another DtxR homologue (Manabe *et al.*, 2005). SirR was originally identified in *Staphylococcus epidermidis* as a 25 kDa protein that functions as a divalent metal cation-dependent transcriptional repressor that retards the migration of a 19 bp palindrome, the Sir box, in a metal-dependent ( $\text{Fe}^{2+}$  or  $\text{Mn}^{2+}$ ) manner. In *S. epidermidis*, SirR is located immediately upstream of a three-gene operon *sitABC*, an ABC transporter system that is engaged in iron



**Figure 1**  
15% SDS–PAGE analysis. Lane M, molecular-weight markers; lane 1, uninduced M15 cells; lane 2, induced M15 cells; lane 3, cell cytosol; lane 4, Ni-Sepharose eluate; lane 5, Purified SirR after gel-exclusion chromatography.



**Figure 2**  
Molecular-weight determination of SirR by size-exclusion chromatography on a Superdex 200 16/70C column. The protein was eluted as described in §2. The column was calibrated with globular protein molecular-weight markers:  $\beta$ -amylase (B),  $\gamma$ -globulin (G), bovine serum albumin (BSA), soybean trypsin inhibitor (TI) and lysozyme (L). The retention volume of the protein was 90 ml, corresponding to a molecular weight of  $\sim$ 50 kDa.

**Table 1**  
Data-collection and processing statistics.

Values in parentheses are for the highest resolution shell.

Wavelength (Å)	1.5418
Space group	$P4_12_1/P4_32_12$
Unit-cell parameters (Å)	
<i>a</i>	105.21
<i>b</i>	105.21
<i>c</i>	144.85
Unit-cell volume (Å <sup>3</sup> )	1603467.62
Resolution (Å)	31.56–2.5 (2.59–2.50)
No. of observed reflections	164715
No. of unique reflections	28735
Completeness (%)	99.8 (99.6)
Redundancy	5.73 (5.60)
$R_{\text{merge}}^{\dagger}$ (%)	0.061 (0.351)
Average $I/\sigma(I)$	14.2 (3.8)

$\dagger \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the observed intensity of a reflection and  $\langle I(hkl) \rangle$  is the mean of reflection  $hkl$ .

uptake and regulates its expression (Hill *et al.*, 1998). This regulator appears to control virulence genes in staphylococci, but its role in pathogenic mycobacteria is as yet unclear. However, the structural elucidation and biological characterization of SirR in *M. tuberculosis* remain to be performed (Lucarelli *et al.*, 2008). Hence, we have initiated structural and functional investigation of this important metal-dependent transcriptional repressor from this important pathogen. These investigations are not only of fundamental scientific interest, but may also provide a platform for the development of therapeutically useful compounds against the pathogen.

## 2. Materials and methods

### 2.1. Cloning

The gene *Rv2788* corresponding to SirR was amplified from *M. tuberculosis* H37Rv genomic DNA using the following gene-specific forward and reverse primer pair: 5'-CGGGATCCGTG-AGGCTGACGAGGAGCCT-3' (forward primer with a *Bam*HI recognition site) and 5'-CCCAAGCTTTCAGCTCACCACCCAG-ATTGC-3' (reverse primer with a *Hind*III recognition site). The PCR

product was purified and digested with *Bam*HI and *Hind*III; the amplicon was then ligated into the *Bam*HI/*Hind*III-digested pQE30 expression vector using T4 DNA ligase. The recombinant DNA was then transformed in *Escherichia coli* M15 (pREP4) cells. The clones were selected on LB agar plates containing 100 µg ml<sup>-1</sup> ampicillin and 25 µg ml<sup>-1</sup> kanamycin. Positive clones were verified by DNA sequencing.

## 2.2. Overexpression and purification

A positive clone harbouring the desired construct of SirR was grown in Luria broth with 100 µg ml<sup>-1</sup> ampicillin and 25 µg ml<sup>-1</sup> kanamycin at 310 K for 3 h, during which the *A*<sub>600</sub> reached 0.6; it was then induced with 100 µM IPTG and grown for a further 5 h at 310 K to maximize the overexpression of the recombinant protein in the cytosolic fraction. The cells from 1 l culture were resuspended in 20 ml buffer *A* (10 mM Tris-HCl pH 8.0, 300 mM NaCl and 10 mM imidazole) containing 0.1 mM each of leupeptin, pepstatin and aprotinin and 0.02 mM phenylmethylsulfonyl fluoride (PMSF). The suspension was lysed by ultrasonication on ice and the lysate was centrifuged at 14 000 rev min<sup>-1</sup> for 40 min. The supernatant was loaded onto Ni-Sepharose High Performance affinity matrix (GE Healthcare Biosciences) pre-equilibrated with buffer *A*. The column was then washed extensively with buffer *A* and the protein was eluted with buffer *B* (10 mM Tris-HCl pH 8.0, 300 mM NaCl and 300 mM imidazole). Size-exclusion chromatography was performed on Superdex 200 prep-grade matrix in a 16/70C column (GE Healthcare Biosciences) using an ÄKTAprime Plus system (GE Healthcare Biosciences). The Ni-Sepharose eluate was loaded onto the size-exclusion column pre-equilibrated with buffer *C* (20 mM Tris-HCl pH 8.0, 50 mM NaCl, 1 mM DTT). Fractions of 2 ml each were collected at a flow rate of 1 ml min<sup>-1</sup> and the absorbance at 280 nm was recorded during the run. To determine the molecular weight, the size-exclusion column was calibrated with the globular protein molecular-weight markers β-amylase (200 kDa), γ-globulin (158 kDa), bovine serum albumin (64 kDa), soybean trypsin inhibitor (20 kDa) and lysozyme (14 kDa), which had retention volumes of 49, 68, 78, 93 and 130 ml, respectively.

The fractions containing the desired protein were verified by SDS-PAGE and pooled together. The protein concentration was estimated by the method of Bradford (1976) and the purity was verified by 15% SDS-PAGE.

## 2.3. Crystallization

The purified protein was concentrated to 50 mg ml<sup>-1</sup> using Amicon Centriprep (10 kDa cutoff) and Amicon Centricon (10 kDa cutoff) concentrators. For crystallization purposes, the His tag was kept intact. Crystallization trials were performed by the sitting-drop vapour-diffusion method in a 96-well Corning CrystalX microplate (Hampton Research). Droplets containing 2 µl protein solution were mixed with an equal volume of mother liquor and equilibrated against 100 µl reservoir solution using commercially available sparse-matrix screens from Hampton Research (Crystal Screen and Crystal Screen II). Crystals were obtained from Hampton Crystal Screen II condition No. 8 (1.5 M sodium chloride, 10% ethanol) at 298 K after 10 d.

## 2.4. Data collection

The crystals obtained using 1.5 M sodium chloride, 10% ethanol were cryoprotected by dipping them into a solution consisting of mother liquor containing 15% glycerol. The diffraction data were

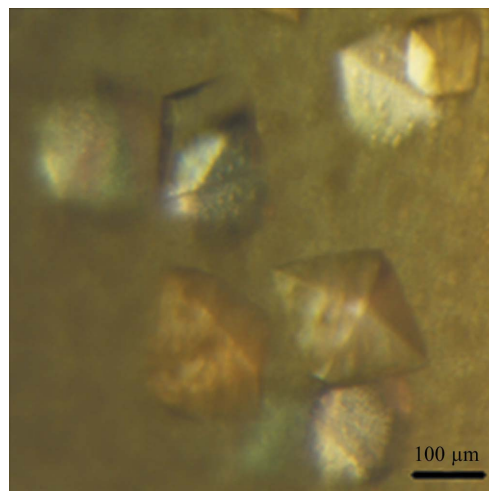
collected at our home source, which was equipped with a Rigaku R-Axis IV++ detector, using Cu Kα X-rays generated by a Rigaku Micromax HF007 Microfocus rotating-anode X-ray generator with a Varimax mirror system that was operated at 40 kV and 30 mA. Crystals were flash-cooled in a liquid-nitrogen stream at 100 K using a Rigaku X-stream 2000 and diffracted to a maximum resolution of 2.5 Å. A total of 180 frames of data were collected with an oscillation angle of 0.5°, an exposure time of 4 min per frame and a crystal-to-detector distance of 150 mm. Diffraction data were processed with *d\*TREK* v.9.8 (Pflugrath, 1999).

## 3. Results and discussion

SirR, a metal-dependent transcriptional repressor from *M. tuberculosis* H37Rv, was successfully cloned in pQE30 vector and overexpressed in *E. coli* M15 (pREP4) cells. From 1 l of culture, 15 mg protein was obtained in the soluble fraction. It was purified to homogeneity in two chromatographic steps: affinity chromatography on Ni-sepharose matrix and gel-exclusion chromatography. The molecular weight of the SirR monomer, ~25 kDa, as predicted from the sequence, was also confirmed by SDS-PAGE (Fig. 1). Size-exclusion chromatographic analysis indicates that the protein exists in a dimeric form in solution with a molecular weight of ~50 kDa (Fig. 2). This observation is consistent with that for IdeR (Pohl *et al.*, 1999). The pure protein was subjected to crystallization by the sitting-drop vapour-diffusion method.

The crystals obtained using 1.5 M sodium chloride, 10% ethanol were 0.25 × 0.17 × 0.12 mm in size (Fig. 3). Diffraction data were collected using a cryoprotected single crystal obtained using 1.5 M sodium chloride, 10% ethanol, which diffracted to a maximum of 2.5 Å resolution. Analysis of the symmetry and systematic absences in the recorded diffraction patterns indicated that the crystals belonged to the tetragonal space group *P*4<sub>1</sub>2<sub>1</sub>/*P*4<sub>3</sub>2<sub>1</sub>2, with unit-cell parameters *a* = 105.21, *b* = 105.21, *c* = 144.85 Å. A total of 164 715 observed reflections were merged to 28 735 unique reflections in the 31.56–2.5 Å resolution range. The overall completeness of the data set was 99.8%, with an *R*<sub>merge</sub> of 6.1%.

The Matthews coefficient suggests the presence of two (*V*<sub>M</sub> = 4.01 Å<sup>3</sup> Da<sup>-1</sup>) to four (*V*<sub>M</sub> = 2.0 Å<sup>3</sup> Da<sup>-1</sup>) molecules in the asymmetric unit. The solvent contents for two and four molecules in



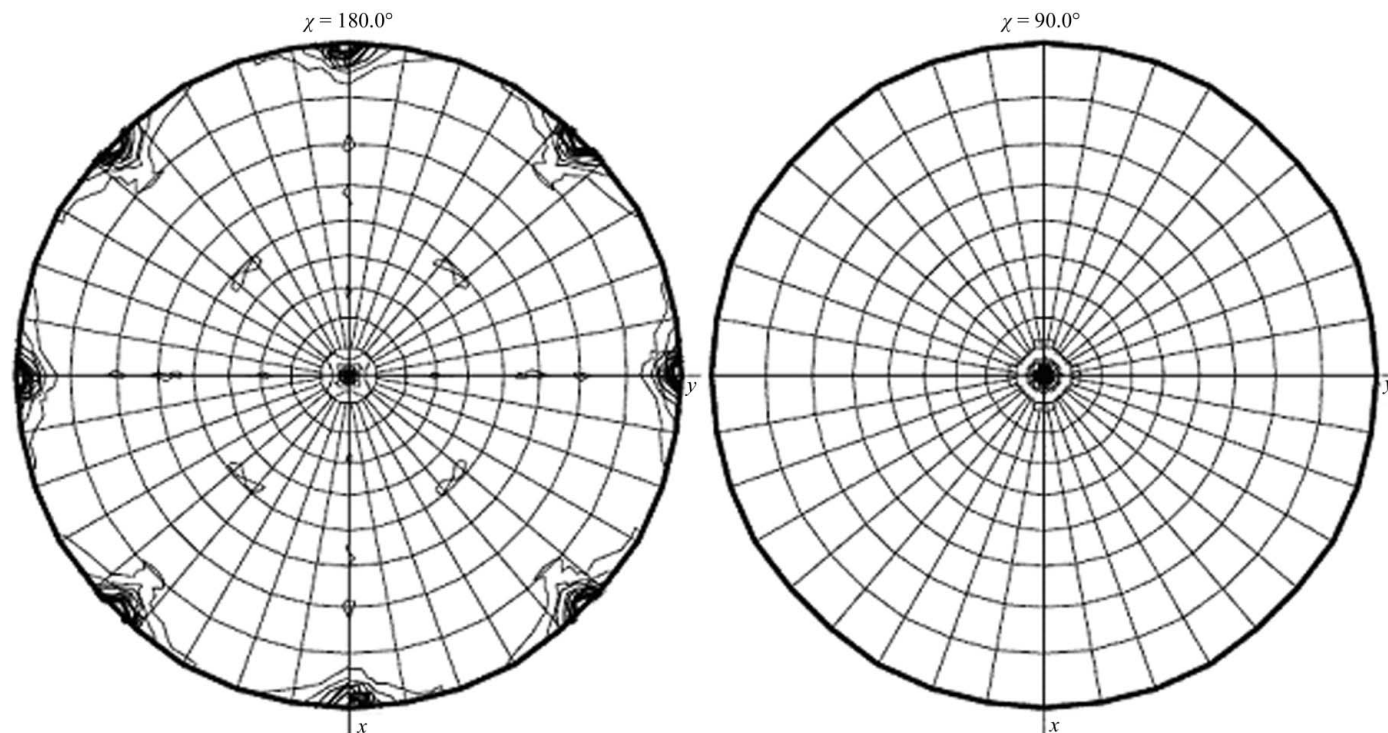
**Figure 3** Crystals of SirR from 1.5 M NaCl, 10% ethanol at 298 K measuring 0.25 × 0.17 × 0.12 mm.

the asymmetric unit are 69.3 and 38.6%, respectively (Matthews, 1968). This suggests the presence of one or two complete dimers in the asymmetric unit.

Calculation of the self-rotation function (Fig. 4) shows a crystallographic fourfold symmetry axis along the  $z$  axis ( $\chi = 90^\circ$ ) and a twofold symmetry axis around the  $z$  axis ( $\chi = 180^\circ$ ). Eight other peaks corresponding to crystallographic twofold axes can also be identified. No extra peaks arising from noncrystallographic symmetry are visible

in the self-rotation function. This indicates that either no noncrystallographic symmetry exists or that a noncrystallographic symmetry axis appears to run parallel to a crystallographic axis and hence is not visible as an additional peak. This can only be ascertained once the structure has been solved.

The multiple sequence alignment of SirR from *M. tuberculosis* with *S. epidermidis* SirR and *M. tuberculosis* IdeR demonstrates some interesting findings (Fig. 5). The probable metal-coordination resi-



**Figure 4** Self-rotation function of the data collected from the crystal of SirR, showing a crystallographic fourfold symmetry axis along the  $z$  axis ( $\chi = 90^\circ$ ), a twofold symmetry axis around the  $z$  axis ( $\chi = 180^\circ$ ) and eight other peaks corresponding to crystallographic twofold axes. This figure was generated using the program *MOLREP* (Collaborative Computational Project, Number 4, 1994).

MtbSirR	MRADDEPGDLSAVAQDYLVKIWIWAQEWSQDKVSTKMLAERIGVSASTASESIRKLAEQGL	60
StepSirR	-----MLTEEKEDYLKAILTN-DGDVSVFVSNKLSQFLNIKPPSVSEMVGRLKEGY	51
IdeR	-----MNELVDTTTEMYLRTIYDLEEEGVTPLRAR-IAERLDQSGPTVSVQTVSRMERDGL	53
	* : **.* : . : : : : . . . : . : . : *	
● ●		
MtbSirR	VDHEKYGAVTLTDSGRRAALAMVRRHRLLETFLVNELGYRWDEVHDEAEVLEHAVSDRLM	120
StepSirR	VETKHYKGARLTEEGKQTLDI IKRHRLLLEFLIEILQYNWEEVHQEAEIIEHRISDLFV	111
IdeR	LRVAGDRHLELTEKGRALATAVMRKHRLAERLLVDVIGLPWEEVHAEACRWEHVMSQEDVE	113
	: **.* : : : : ** * : * : : : * : * * * : * : * : *	
▲ ▲ ▲ ▲		
MtbSirR	ARIDAKLGFPQDRPHGDPIP-----GADGQVPTPPARQLWACRDGD--TGTVARISDA	171
StepSirR	ERLDKILNFPKTCPHGGVIP-----RGD-QYKEIFTSILNFEPGE--RVTVRRVRDK	161
IdeR	RRLVKVLNNPFTTSPFGNPIPLVELGVGPEPGADDANLVRITELPAGSPVAVVVRQLTEH	173
	* : * . * * * . * * : : : : : : * : * . * . * : : : *	
MtbSirR	DP---QMLRYFASIGISLDSRLRVLARREFAGMISVAIDSADGATVDLGSPPAAQAIWVVS	228
StepSirR	T----ELLVYLSKDIYIGNTVIEIVSKDDTN---KVIILKRNDIVTILSYENAMNIFAEK	214
IdeR	VQGDIDLITRLK DAGVVPNARVTVEVTPGGG---VTIVIPGHENVTLPHEMAHAVKVEK	229
	: : : : . . : . : : : : * * . . * * : : . .	
MtbSirR	-	
StepSirR	-	
IdeR	V 230	

**Figure 5** Sequence alignment of *M. tuberculosis* SirR (MtbSirR) with SirR from *S. epidermidis* (StepSirR) and IdeR from *M. tuberculosis*. Asterisks indicate fully conserved residues, colons denote strongly conserved residues and dots show weakly conserved residues. Probable metal-binding residues are denoted by triangles. Glu109 is represented by a square, while the DNA-binding residues are denoted by circles.

dues His86, Arg87, Glu90, His105, Glu112 and His113 are conserved. Glu109 of *M. tuberculosis* SirR corresponds to Glu103 of *S. epidermidis* SirR but to Cys105 of IdeR, an important metal-binding residue. This substitution is also observed in TroR, the DtxR homologue from *Treponema pallidum*. Comparison of the DNA-binding residues of IdeR with those of *M. tuberculosis* SirR reveals that Thr47 and Ser49 are conserved. However, a clear picture will only emerge once the structure has been solved.

Structure solution by the molecular-replacement technique using *MOLREP* (Vagin & Teplyakov, 1997) within the *CCP4* package (Collaborative Computational Project, Number 4, 1994) failed to provide any promising results, with a very low contrast value of 1.22. This is the consequence of poor sequence identity (31%) with the nearest homologue in the Protein Data Bank: *M. tuberculosis* IdeR (PDB code 1fx7; Feese *et al.*, 2001). Of 228 residues, only 58 residues of SirR could be aligned with 187 residues of IdeR. Therefore, the preparation of heavy-atom derivatives of SirR is in progress for structure solution by the multiple isomorphous replacement (MIR) method.

This work was carried out with financial assistance from the Department of Biotechnology, Government of India. SM thanks the Council of Scientific and Industrial Research, Government of India for an individual fellowship. The authors wish to thank Ms Linda Schuldt (European Molecular Biology Laboratory, Hamburg) for helpful suggestions and valuable discussions.

## References

- Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
- Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Feese, M. D., Ingason, B. P., Goranson-Siekierke, J., Holmes, R. K. & Hol, W. G. (2001). *J. Biol. Chem.* **276**, 5959–5965.
- Gold, B., Rodriguez, G. M., Marras, S. A., Pentecost, M. & Smith, I. (2001). *Mol. Microbiol.* **42**, 851–865.
- Hill, P. J., Cockayne, A., Landers, P., Morrissey, J. A., Sims, C. M. & Williams, P. (1998). *Infect. Immun.* **66**, 4123–4129.
- Huffman, J. L. & Brennan, R. G. (2002). *Curr. Opin. Struct. Biol.* **12**, 98–106.
- Lucarelli, D., Russo, S., Garman, E., Milano, A., Meyer-Klaucke, W. & Pohl, E. (2007). *J. Biol. Chem.* **282**, 9914–9922.
- Lucarelli, D., Vasil, M. L., Meyer-Klaucke, W. & Pohl, E. (2008). *Int. J. Mol. Sci.* **9**, 1548–1560.
- Manabe, Y. C., Hatem, C. L., Kesavan, A. K., Durack, J. & Murphy, J. R. (2005). *Infect. Immun.* **73**, 5988–5994.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Milano, A., Forti, F., Sala, C., Riccardi, G. & Ghisotti, D. (2001). *J. Bacteriol.* **183**, 6801–6806.
- Pflugrath, J. W. (1999). *Acta Cryst.* **D55**, 1718–1725.
- Pohl, E., Holmes, R. K. & Hol, W. G. (1999). *J. Mol. Biol.* **285**, 1145–1156.
- Rodriguez, C. M. & Smith, I. (2003). *Mol. Microbiol.* **47**, 1485–1494.
- Rodriguez, G. M., Voskuil, M. I., Gold, B., Schoolnik, G. K. & Smith, I. (2002). *Infect. Immun.* **70**, 3371–3381.
- Sala, C., Forti, F., Florio, E., Di Canneva, F., Milano, A., Riccardi, G. & Ghisotti, D. (2003). *J. Bacteriol.* **185**, 5357–5362.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
- Wisedchaisri, G., Holmes, R. K. & Hol, W. G. (2004). *J. Mol. Biol.* **342**, 1155–1169.
- Zahrt, T. C., Song, J., Siple, J. & Deretic, V. (2001). *Mol. Microbiol.* **39**, 1174–1185.