Crystal Structure of the SarR Protein from Staphylococcus aureus

Yingfang Liu  
*University of Colorado, Denver*  

Adhar Manna  
*Dartmouth College*  

Ronggui Li  
*University of Colorado, Denver*  

Wesley E. Martin  
*University of Colorado, Denver*  

Robert C. Murphy  
*University of Colorado, Denver*  

See next page for additional authors  

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa  
Part of the Medical Microbiology Commons  

Recommended Citation  
Liu, Yingfang; Manna, Adhar; Li, Ronggui; Martin, Wesley E.; Murphy, Robert C.; Cheung, Ambrose L.; and Zhang, Gongyi, "Crystal Structure of the SarR Protein from Staphylococcus aureus" (2001). Open Dartmouth: Faculty Open Access Articles. 1558.  
https://digitalcommons.dartmouth.edu/facoa/1558  

This Article is brought to you for free and open access by Dartmouth Digital Commons. It has been accepted for inclusion in Open Dartmouth: Faculty Open Access Articles by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.
Crystal structure of the SarR protein from *Staphylococcus aureus*

Yingfang Liu*, Adhar Manna†, Ronggui Li*, Wesley E. Martin*, Robert C. Murphy*, Ambrose L. Cheung†, and Gongyi Zhang‡*

*National Jewish Medical and Research Center and Departments of Immunology and Pharmacology, School of Medicine, University of Colorado Health Science Center, 1400 Jackson Street, Denver, CO 80206; and †Departments of Microbiology and Immunology, Dartmouth Medical School, Hanover, NH 03755

Communicated by John W. Kappler, National Jewish Medical and Research Center, Denver, CO, January 5, 2001 (received for review December 15, 2000)

The expression of virulence determinants in *Staphylococcus aureus* is controlled by global regulatory loci (e.g., sarA and agr). The sar (Staphylococcus accessory regulator) locus is composed of three overlapping transcripts (sarA P1, P3, and P2, transcripts initiated from the P1, P3, and P2 promoters, respectively), all encoding the 124-aa SarA protein. The level of SarA, the major regulatory protein, is partially controlled by the differential activation of the sarA promoters. We previously partially purified a 13.6-kDa protein, designated SarR, that binds to the sarA promoter region to down-modulate sarA transcription from the P1 promoter and subsequently SarA expression. SarR shares sequence similarity to SarA, and another SarA homolog, SarS. Here we report the 2.3 Å-resolution x-ray crystal structure of the dimeric SarR-MBP (maltose binding protein) fusion protein. The structure reveals that the SarR protein not only has a classic helix-turn–helix module for DNA binding at the major grooves, but also has an additional loop region involved in DNA recognition at the minor grooves. This innovative mode could represent a new functional class of the “winged helix” family. The dimeric SarR structure could accommodate an unusually long stretch of ~27 nucleotides with two or four bending points along the course, which could lead to the bending of DNA by 90° or more, similar to that seen in the catabolite activator protein (CAP)–DNA complex. The structure also demonstrates the molecular basis for the stable dimerization of the SarR monomers and possible motifs for interaction with other proteins.

*S. aureus* is a versatile bacterium capable of causing a wide spectrum of pathology in humans, ranging from superficial abscesses to pneumonia, endocarditis, and sepsis (1). This versatility may be attributable to the impressive array of extracellular and cell-wall-associated virulence determinants coordinately expressed during the process of infection (2). The coordinate expression of many virulence determinants in *S. aureus* has been shown to be regulated by global regulatory elements such as sarA (Staphylococcus accessory regulator) and agr (accessory global regulator) (3, 4). These regulatory elements, in turn, control the transcription of a wide variety of unlinked genes, many of which have been implicated in pathogenesis.

The global regulatory locus agr encodes a two-component, quorum sensing system that is involved in the generation of two divergent transcripts, RNAII and RNAIII, from two distinct promoters, P2 and P3, respectively. RNAIII, initiated from the P3 promoter, is the regulatory molecule of the agr response, hence responsible for the up-regulation in extracellular protein production and the down-regulation of cell-wall-associated protein synthesis during the postexponential growth phase (5, 6). The RNAII molecule, driven by the P2 promoter, encodes a four-gene operon, agrBDCA. Additionally, agrD, in conjunction with agrB, participates in the generation of an octapeptide with quorum sensing functions (7, 8). This autoinducing peptide can stimulate the transcription of the agr regulatory molecule RNAII, which ultimately interacts with target genes to modulate transcription (6) and possibly translation (9).

In contrast to agr, the sarA locus activates the synthesis of both extracellular (e.g., α- and β-hemolysins) and cell-wall proteins (e.g., fibronectin binding protein A) in *S. aureus* (3). The sarA locus is composed of three overlapping transcripts [sarA P1 (0.56 kb), sarA P3 (0.8 kb), and sarA P2 (1.2 kb) transcripts]. Because of their overlapping nature, each of these transcripts encodes the major 372-bp sarA gene, yielding the 14.5-kDa SarA protein (10). DNA footprinting studies revealed that the SarA protein binds to the promoters of several target genes (11) including agr, hla (α hemolysin gene), spa (protein A gene), and fthA (fibronectin binding protein A gene), thus implicating SarA as a regulatory molecule that can modulate target gene transcription via both agr-dependent and agr-independent pathways (11–13).

The complexity of the sarA triple promoter system also hints at the possibility that regulatory proteins may bind to the sarA promoter region to modulate sarA transcription and, subsequently, SarA expression. Using a DNA-affinity column, we recently identified a 13.6-kDa protein, designated SarR (14), that binds to the sarA promoter region to modulate SarA expression (16). A mutation in sarR decreases transcription from the P1 promoter, the predominant sarA promoter, and hence the ensuing SarA expression. To provide a framework for understanding how SarR behaves as a DNA-binding protein, we have dual functions of repressor as well as activator, we solved the crystal structure of SarR.

**Materials and Methods**

Protein Expression, Purification, and Crystallization. The intact 345-bp sarR gene was amplified by PCR using chromosomal DNA from *S. aureus* strain RN6390 as the template and primers containing flanking restriction sites (NdeI and BamHI) to facilitate cloning into an expression vector pMAL-c2 (New England Biolabs) modified by truncating 21 residues from the linker region that connects SarR and maltose binding protein (MBP; R.L. and G.Z., unpublished data). The recombinant plasmid containing the sarR gene was transformed to *E. coli* BL21(DE3)pLysS. Enhanced expression of SarR-MBP fusion was induced by adding IPTG (isopropyl-1-thio-β-D-galactopyranoside) to a 4-L growing culture (37°C) at an OD600 of 0.7. After 4 h of additional growth, cells were harvested, resuspended in buffer (50 mM Tris-HCl, 1 mM EDTA, pH 7.4, 300 mM NaCl, 5% glycerol, and 1 mM DTT) and subjected to cell lysis through...
a continuous-flow French press. After a 20,000 g spin, the soluble fraction was loaded onto an amylose resin affinity column (10 ml) and the SarR-MBP fusion protein was eluted with 10 mM maltose. The protein was loaded onto a MonoQ (Amersham Pharmacia) ion-exchange column. After elution with a NaCl gradient (0.1 to 0.5 M), the fraction containing the protein was found to be homogeneous as determined by a Coomassie-stained SDS-polyacrylamide gel. The concentration of the purified protein was determined with the Bio-Rad Protein Assay solution, using BSA as the standard. The SarR-MBP fusion protein (15 mg/ml) was crystallized by vapor diffusion against a solution of 5 mM \( \beta \)-mercaptoethanol, 100 mM Na acetate, 100 mM NaCl, pH 4.6, and 18–22% PEG monomethyl ether 2000. For cryo-crystallography, crystals were soaked in steps of increasing glycerol concentration (5% each step every 30 min) and finally into 20% glycerol before flash-freezing.

**Structure Determination and Refinement.** Data were collected in the laboratory on a Rigaku R-axis IV system and beamline 5.0.2 at the Advanced Light Source (ALS) at Berkeley. Data processing was performed with DENZO and SCALEPACK (17). The initial phases were obtained by a molecular replacement solution using the available MBP structure (18) and the program AMORE (19). Fo-Fc and 2Fo-Fc electron density maps were calculated by CNS (20). Map interpretation and model building were done by using the program O (21) and aided by the secondary prediction results (22). The map was improved by cycles of refinement using CNS with Non-Crystallographic Symmetry (NCS) constraints. A final refinement was performed with relaxed NCS restraints.

![Diagram of the three-dimensional structure](image)

**Table 1. Summary of crystallographic analysis**

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Resolution, Å</th>
<th>( R_{merge} ), %</th>
<th>No. reflections</th>
<th>R-factor ( ^{a} )</th>
<th>Free R-factor ( ^{b} )</th>
<th>Resolution, Å</th>
<th>R-factor ( ^{c} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native I</td>
<td>15.27</td>
<td>5696</td>
<td>5656</td>
<td>5512</td>
<td>2.83</td>
<td>20–4.58</td>
<td>23.24</td>
</tr>
<tr>
<td>Native II</td>
<td>21.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.02</td>
<td></td>
</tr>
</tbody>
</table>

\( ^{a} \text{Crystal space group P1; } a = 64.7 \text{ Å, } b = 70.6 \text{ Å, } c = 75.5 \text{ Å, } \alpha = 65.7^{\circ}, \beta = 67.2^{\circ}, \gamma = 69.6^{\circ}, \text{ two SarR-MBP monomers/unit cell.} \)

\( ^{b} \text{R-factor} = \sum_{i} \frac{F_{obs} - F_{calc}}{F_{obs}} \text{ for all amplitudes with } F_{obs} \geq 2 \text{ measured in the indicated resolution bin.} \)

\( ^{c} \text{Free R-factor was calculated with 5% of the data in each bin.} \)

*Crystalreplacement solutions, CO, (correlation coefficients); and R, are defined in AMORE (19).*

---

**Fig. 1.** (A) A ribbon (27) diagram of the three-dimensional structure of the SarR-MBP fusion protein. The SarR dimer is at the top colored green and yellow for each monomer, respectively, and the two MBP molecules are at the bottom colored blue and pink, respectively. (B) A worm diagram of the three-dimensional structure of the SarR monomer. Starting from the N terminus, \( \alpha1 (6–24) \rightarrow \alpha2 (32–48) \rightarrow \beta1 (47–50) \rightarrow \alpha3 (51–58) \rightarrow \alpha4 (63–74) \rightarrow \beta2 (79–83) \rightarrow \beta3 (90–96) \rightarrow \alpha5 (97–113) \). \( \alpha2, \beta1, \alpha3, \alpha4, \beta2, \) and \( \beta3 \) are defined as a “winged helix motif.” The worm is colored according to the B factor of the side chain: red (low B, 25), white (middle B, 45), and blue (high B, 80).
The structure of SarR-MBP fusion protein shows that the SarR dimer is located at the top of the two individual MBP molecules, connected by two loops (with residues both from SarR and modified MBP) between SarR and MBP. The positions of the MBP molecules suggest that they did not influence the structure of SarR dimer (Fig. 1A). The overall structure of the SarR monomer consists of five α helices, three short β strands, and several loops (Fig. 1B). The α1 helix extends out from the remaining molecule, forming an “L”-like structure with a stretch formed from the N-terminal residues of the protein. The α2 helix follows a 7-residue loop from α1, and is almost perpendicular to α1 (85°). The three β strands, β1, β2, and β3, form an antiparallel bundle, which is slightly twisted. α5 follows immediately after β3. Between β1 and β2 is a long flexible region (residues 51–79), which had poor electron density in the initial 2Fo-Fc and Fo-Fc maps, containing two helices, α3 (residue 51–56) and α4 (residue 63–75), respectively, and a short turn (residue 56–58); these three elements build up a typical helix-turn-helix structural module existing in DNA binding proteins. Homology alignment (Dali; ref. 24) of the SarR monomer with all available structures shows that the SarR monomer is homologous to winged helix proteins (25), such as transcription regulatory protein motA fragment (PDB ID code 1bja) with a Z score of 8.0 and transcriptional repressor smtb activation domain (PDB ID code 1smt) with a Z score of 7.2. Compared with winged helix proteins, the “W2” is replaced by a helix (α5) and the “W1” extends much further in the SarR monomer (Fig. 1B). Therefore, we postulate that SarR and its family of proteins likely are new members of the classic winged helix protein family.

Two L-like structures of α1 and the stretch of the N-termini come together, forming a dimer that has an elongated, slightly bent structure with overall dimensions of 71 Å × 37 Å × 34 Å (Fig. 2A, B, and C). The monomers are related to each other by a noncrystallographic local 2-fold axis. On the concave side and middle of the SarR dimer, there is a canyon-like structure with a depth of ∼15 Å (Fig. 2A and B). The canyon is formed by a part of α1, the loop that connects α1 and α2 and a part of α2 as its bed, and with α4 from both monomers acting as the two banks (Fig. 2A and B). On the convex side, all four termini (C- and N-termini of both molecules) form a flat platform with the N-terminus of one monomer adjacent to the C terminus of the other (Fig. 2C). This feature may possibly explain how SarS, a 250-residue protein homologous to SarA (26), may function as a heterodimer-like monomer because it contains two sections, each highly similar to the SarR module and 125 residues long (26). The entire dimer can be described as three individual subdomain structures (Fig. 2A). Three β strands plus α3 and α4 from each molecule form subdomains 1 and 2, respectively. Those two subdomains were poorly defined in the initial electron density map, which also was reflected by their high temperature factors compared with the rest in the final model (Fig. 1B), and thus can be expected to have high mobility in the molecule in solution. Subdomain 3, consisting of the major α helices α1, α2, and α5 from both monomers, is relatively rigid because of the restraints of hydrophobic interactions. These helices in subdomain 3 are arranged in such a manner that the entire subdomain looks like a twisted letter Z with the relatively flexible subdomain 1 and subdomain 2 covering its ends. The Z-shaped scaffold, a unique structural feature for SarA family, could also represent a new functional protein fold (Fig. 2A and C).

Dimerization Interfaces. Several lines of evidence suggest that, with the exception of SarS, the active form of the SarA family of proteins may be a homodimer. First, there is strong evidence showing that SarA exists as a homodimer in vitro as well as in vivo (15). Second, on mixing partially unfolded (4 M urea) full-length SarR protein and SarR-MBP fusion, a heterodimer product containing one copy each of SarR and SarR-MBP fusion could be separated by gel-filtration chromatography (data not shown). Third, we used crystals of SarR-MBP fusion protein to solve the SarR structure (see Materials and Methods). The SarR exists as dimer even in the MBP fusion form and we postulate that the dimerization form may be the active form existing in vivo. Fourth, based on the homodimer SarR structure, the interactions between the two monomers are quite extensive, with most residues involved in hydrophobic interactions (Figs. 1A and 3). Residues Leu-10, Ile-7, and Ile-4 from one monomer and...
residues Phe-20, Ile-35, Leu-109, and Ile-113 from the other monomer form two hydrophobic cores. The dimer interface buries 1,500 Å² of solvent-accessible surface area (1.4-Å probe; ref. 28) per monomer. Considering this hydrophobic feature of SarR dimerization, we believe that the dimer of SarR in the fusion protein is also the functional dimer in vivo. Furthermore, the L-like structure of α1 and the stretch of the N terminus of SarR physically block the dissociation of one monomer from its partner. This blockage explains why an extremely harsh condition (4 M urea) is required to disrupt the dimerization of individual monomers. Fifth, deletion analysis showed that mutated SarA, with a 15-residue truncation at its N terminus, exists as monomers in vitro (in solution or crystal packing forms; data not shown). Finally, as predicted from the sequence alignment of SarA, SarR, and SarS, most residues that are involved in the dimerization process are highly conserved (Fig. 3).

**DNA Binding and Bending.** As reported earlier, SarA and SarR proteins contain a high percentage of the residue Lys (29). Remarkably, most of the Lys residues are highly conserved between these two proteins (Fig. 3). It was predicted that most of these Lys residues are involved in DNA binding (29). Although the Lys residues are distributed throughout the entire primary sequence, in the 3D structure most Lys residues and some Arg residues are located primarily on one surface of the SarR dimer (the concave side). The electrostatic potential on this surface of the SarR dimer, calculated by the GRASP program (28), revealed a positively charged track on this side (Fig. 4A).

Additionally, the two winged helix motifs (subdomain 1 plus α2 from one monomer and subdomain 2 plus α2 from another monomer) are located on this side. We conclude that this side is the most likely site for DNA binding.

Using the program DALI to search for proteins with a structure similar to SarR, we examined manually every structure with a Z score higher than 4. We found that the spatial arrangement of the two SarR helix–turn–helix modules is quite similar to those described for catabolite activator protein (CAP; ref. 30; PDB ID code 1cgp with Z score of 4.7). Superposing the SarR winged helix motifs 1 and 2 with the corresponding domains from CAP dimer, the root mean square deviation (rmsd) of the α-carbon backbone is 1.8 Å for motif 1, and 2.1 Å for motif 2 (Fig. 5A).

Interestingly, some of the residues involved in DNA binding and bending in the CAP dimer, which interacts with the major grooves, are conserved in the SarR dimer and in other members of the SarA family (Fig. 3). For example, SarR-charged residues that are predicted to interact with phosphate groups on the DNA backbone are Lys-52, Lys-56, Lys-71, and Arg-82. Lys-52 and Arg-82 are absolutely conserved. Charged residues predicted to contact DNA bases are Lys-61 and Lys-67. Lys-61 is also absolutely conserved. Therefore, we predict that the SarR dimer and other SarA family members have protein–DNA interaction similar to the CAP dimer.

The interaction could cause the DNA to bend at two points by ~90° (30). The longest direct distance for the CAP and SarR dimer surface is ~71 Å (Fig. 4), which can hold a stretch of bent DNA. The DNA binding and bending could further disrupt the dimerization process (30).
DNA with ~27 base pair nucleotides that has a length of ~92 Å for a normal B form (30). This result is consistent with the experimental DNA footprinting data, showing that ~29 nucleotides from the sarA promoter region were involved in binding to SarR (31). This predicted bending of the DNA when SarR binds to DNA may reflect a regulatory mechanism for the SarA family of proteins in controlling target gene transcription.

We superimposed the SarR dimer on the CAP–DNA complex to construct a model for a SarR–DNA complex (Fig. 5 B and C). From this model, we can predict that, in addition to interactions of the α4 helix with the DNA major groove, SarR may make contacts with the DNA minor groove. The loop region between β2 and β3 and part of the two β strands (W1, a β-hairpin) are predicted to be quite flexible in the free SarR structure. Only slight adjustments of their conformations are required to position them to interact intensively with the minor groove of the DNA. Several residues that could be involved in the interactions are highly conserved residues (Asp-86, Glu-87, and Arg-88), with perhaps the side chain of Arg-88 interacting with the DNA phosphate backbone and the side chains of Asp-86 and Glu-87 interacting with bases (Fig. 5 B and C). There is no data so far to show whether additional bending could be induced by this interaction. This loop is too short in CAP and other winged helix proteins to have this minor groove binding function (25). One new member of one of the classes of winged helix proteins (RFX) does make DNA minor groove contacts, but in this case, the wing contacts the major groove and the helix contacts the minor groove (32). Therefore, the predicted SarR type of wing-minor groove interaction may be a unique feature of the SarA family of proteins, thus establishing them as a third class of the winged helix family (25, 32). The structure of DNA complexed with a member of SarA family is needed to test this prediction.

The two subdomains that are involved in DNA binding and bending in CAP have different conformations in the absence of DNA, but are identical in the complex structure (25). The corresponding subdomains (1 and 2) in SarR are almost identical in the free protein, but most of the side chains are poorly defined. We believe that these two subdomains have high mobility in the absence of DNA, as has been described for the CAP protein. The mobility can be indicated by their high temperature factors (Fig. 1 B).

**Possible Regulation Mechanism.** One class of transcription-activating proteins bears two structural motifs, namely DNA binding and activation domains (33). The SarR protein was initially defined as a transcriptional repressor protein that binds
the sarA promoter region (29), thus leading to reduced transcription from the sar P1 promoter. Because the sarA P1 promoter is the predominant promoter in the sarA regulatory system, we predicted and recently confirmed that the expression of SarA was increased in a sarR mutant (29). Transcriptional fusion studies also indicated that the sarA locus may be auto-regulatory (31), possibly mediated by the binding of SarA to its own promoter. It was postulated that an activation motif might be present on the SarA protein but not on the SarR protein, and that SarR may repress by a simple competitive displacement mechanism. Indeed, we have found that the binding affinity of SarR to a sarA promoter fragment is higher than its SarA counterpart (16). A second possibility is that SarA and SarR may form a heterodimer to interfere with the function of the SarR homodimer. Because of the conservation of residues involved in the dimerization, this could happen in vivo, although we have so far failed to generate a heterodimer of SarA and SarR when using His-tag proteins in vitro. Nevertheless, attempts to yield heterodimers with native SarA and SarR proteins without any His-tags are in progress. Finally, SarR may function in a manner similar to the bacteriophage lambda repressor (which also has a helix–turn–helix DNA binding motif). In this case, a slight difference in DNA binding site (one base pair shift) could turn an activator into a repressor by affecting RNA polymerase binding (34).

The structure of the SarR protein, combined with the sequence alignment of additional SarA family members (Fig. 3), suggests that the regulatory mechanism may be much more complicated. Although there is no isolated activation domain in the SarR dimeric structure compared with the CAP, a calculation of the surface electronic potential revealed two negatively charged patches on the convex side of the SarR dimer (Fig. 4B). These patches include residues Asp-6 and Asp-9 from molecule 1 and Glu-108, Glu-110, and Glu-111 from molecule 2. Many transcription regulators work by binding DNA and then interacting with a component of the RNA polymerase machinery (33). For example, CAP regulates downstream protein expression mostly through its interactions with the C-terminal domains of the α subunit of RNA polymerase (35). It is tempting to speculate that these acidic patches on the surface of SarR may represent activation motifs that allow SarR to regulate gene expression in a similar way as SarA. Because Glu-108, Glu-110, and Glu-111 are not conserved among the SarA family (Fig. 3), this activity may have different specificities among the family members. In this regard, we have experimental evidence to suggest that SarR, besides interacting with the sarA promoter acting as a repressor, also directly binds to the hla promoter (e.g., α hemolysin gene) as an activator, thus bypassing the effect of SarA in controlling target gene expression (data not shown).

We thank John Kappler for discussions and advice; Howard Hughes Medical Institute, Zuckerman/Canyon Ranch, and Alan Lapore for support of our x-ray and computing facility; Keith Henderson for assistance at 5.0.2 of the Advanced Light Source (ALS) at Berkeley; Philippa C. Marrack, James D. Crapo, and other members at National Jewish for their kind support; and Elizabeth A. Campbell and Seth A. Darst for initial crystallization experiments with SarR protein. G.Z. is supported by a start-up fund from National Jewish Medical and Research center. A.L.C. is supported by a grant from the National Institutes of Health.