Crystal structure of *Enterococcus faecalis* SlyA-like transcriptional factor

Running title: Structure of SlyA reveals winged-helix DNA-binding motif

**Rui-ying Wu**\(^a\)*, Rong-guang Zhang\(^a\)*, Olga Zagnitko \(^b\), Irina Dementieva\(^a\), Natalia Maltzev\(^c\), James D. Watson\(^d\), Roman Laskowski\(^d\), Piotr Gornicki \(^b\) and **Andrzej Joachimiak**\(^a\)#

\(^a\) Biosciences Division and Structural Biology Center, Argonne National Laboratory, 9700 S. Cass Ave., Bldg 202, Argonne, IL 60439, USA.

\(^b\) Department of Molecular Genetics and Cell Biology, University of Chicago, 920 E. 58\(^{th}\) St., Chicago, IL 60637.

\(^c\) Mathematics and Computer Science Division, Argonne National Laboratory, 9700 S. Cass Ave., Bldg. 202, Argonne, IL 60439, USA.

\(^d\) European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK.

# Corresponding author: e-mail: andrzejj@anl.gov; phone: 630-252-3926; fax: 630-252-6126.

* These two authors contributed equally to this work.
Abstract
The crystal structure of a SlyA transcriptional regulator at 1.6 Å resolution is presented and structural relationships between members of the MarR/SlyA family are discussed. The SlyA family, which includes SlyA, Rap, Hor and RovA proteins, is widely distributed in bacterial and archaeal genomes. Current evidence suggests that SlyA-like factors act as repressors, activators, and modulators of gene transcription. These proteins have been shown to upregulate expression of molecular chaperones, acid-resistance proteins, and cytolysin, and to downregulate several biosynthetic enzymes. The crystal structure of SlyA from *Enterococcus faecalis*, determined as a part of an ongoing structural initiative (www.mcsg.anl.gov), revealed the same winged-helix DNA-binding motif that was recently found in the MarR repressor from *Escherichia coli* and the MexR repressor from *Pseudomonas aeruginosa*, a sequence homologue of MarR. Phylogenetic analysis of the MarR/SlyA family suggests that Sly is placed between the SlyA and MarR subfamilies and shows significant sequence similarity to members of both subfamilies.

Introduction
Comparisons of amino acid sequences and sequence patterns in prokaryotic genomes have identified many putative DNA-binding proteins and transcription factors. A sequence pattern common to several families of *Escherichia coli* regulatory proteins (GalR/LacI, DeoR, IclR, GntR, Crp, MarR, MerR, and AsnC) has been identified [1] and the structures of several transcriptional regulators have been determined [2]. Many of these proteins contain the classic helix-turn-helix (HTH) DNA-binding motif and exhibit an associated sequence pattern.

A large number of the putative transcription factors, however, reveal new DNA-binding motifs [2,3]. The MarR/SlyA family is one example. Members of this large family modulate gene expression and are widely distributed. The COG database (http://www.ncbi.nlm.nih.gov/cgi-bin/COG/palox?COG1846) includes 149 members of the MarR (multi-antibiotic resistance) family, among them 23 MarR homologues from *Bacillus subtilis*, MprA from *E. coli*, PecS from *Erwinia chrysanthemi*, and Hpr from *B. subtilis* [4,5]. The crystal structure of MarR revealed a new winged-helix (WH) DNA-binding motif [6]. This structural motif was recently found in the MexR repressor from *Pseudomonas aeruginosa*, a sequence homologue of MarR [7].
SlyA-like proteins were found in 118 species of bacteria and 15 species of archaea, with 16 SlyA homologues in *E. coli* alone (http://www.ncbi.nlm.nih.gov:80/cgi-bin/Entrez/blink). This family includes such proteins as SlyA of *Salmonella typhimurium* and *E. coli*, and the Rap, Hor, and RovA proteins from *Serratia marcescens*, *Erwinia carotovora*, and *Yersinia enterocolitica*, respectively [8] (Fig. 1). In *E. coli*, SlyA regulates expression of Cytolysin A, a cytotoxic protein that forms, stable, cation-selective transmembrane pores [9,10]. The *slyA* gene is implicated in virulence and environmental adaptation of *S. typhimurium*. It has been shown to upregulate expression of molecular chaperones, acid-resistance proteins, and cytolysin, and to downregulate several biosynthetic enzymes [11]. It has also been suggested that regulation of gene expression by SlyA is crucial for intracellular survival and/or replication of both enteroinvasive *E. coli* and *Salmonella serovar Typhimurium* in phagocytic host cells.

RovA, whose amino acid sequence is 77% identical to the *S. typhimurium* SlyA, plays a role in regulation of the invasion of mammalian cells by *Yersinia* and mediates regulation of invasin in response to environmental signals [12]. Invasin is a primary factor that allows efficient internalization of *Y. pseudotuberculosis* into eukaryotic cells. All the current evidence shows that SlyA-like factors act as repressors, activators, or modulators of gene transcription.

The SlyA protein from a pathogen *Enterococcus faecalis* (SlyA-Ef) was selected for this study as a part of an ongoing structural initiative (www.mcsg.anl.gov). Specific biochemical function, the signal molecule and the DNA-binding properties of this protein are unknown. Phylogenetic analysis of the MarR/SlyA family suggests that SlyA-Ef is placed between SlyA and MarR subfamilies. It shows significant sequence similarity to members of both subfamilies (Fig. 1). In this report we present the crystal structure of a SlyA transcriptional regulator at 1.6 Å resolution and discuss structural relationships between members of the MarR/SlyA family.

**Materials & Methods**

The *E. faecalis* SlyA ORF was amplified by PCR, with the *NdeI* and *BamHI* sites engineered at the translation start codon and immediately downstream of the translation stop codon, respectively, and cloned between the *NdeI* and *BamHI* sites of the pET15b vector (Novagen) in frame with the His-tag and rTEV cleavage site. Expression of the His-tagged fusion protein in *E.
coli strain BL21[DE3] carrying the pMAGIC vector was induced with isopropyl-β-D-thiogalactoside. Cells were harvested after 4 hours of culture at 37°C; suspended in 50 mM phosphate buffer pH 8.0, 300 mM NaCl, 10 mM imidazol, 10 mM β-mercaptoethanol, and 10% glycerol; and lysed by sonication. The fusion protein was purified by affinity chromatography using Ni-NTA Superflow resin (Qiagen). The His-tag was removed by digestion with rTEV and the resulting protein was purified by affinity chromatography using Ni-NTA Superflow resin (Qiagen). In this design three amino acid residues were added at the N-terminus of SlyA. The cleaved protein was further purified on an SP Sepharose Fast Flow column (Pharmacia) using 0.5 and 1.0 M NaCl two-step elution and concentrated with simultaneous buffer exchange using Centriplus-3 (Amicon) (3-kDa cutoff). A 2 mM protein stock solution in 10 mM Tris-HCl pH 7.4, 20 mM NaCl, and 1 mM DTT was used for crystallization. Selenomethionine-labeled SlyA protein was prepared using methionine biosynthesis inhibition method [13].

Equal volumes of SlyA protein stock solution and buffers were mixed in hanging droplets and equilibrated against 1-ml aliquots of solutions from the Hampton Research sparse matrix crystallization screening kit. SlyA was crystallized from 30% PEG 3350, 20% glycerol and 100 mM Hepes-Na (pH7.6) at 283ºK. Crystals (0.2*0.02*0.4 mm) were rinsed in cryoprotectant solution consisting of 25% glycerol in the crystallization solution and flash-frozen in liquid nitrogen. Diffraction data were collected at 100ºK at the 19BM beamline of the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory.

Crystals of native SlyA protein and its SeMet derivative diffracted to 1.8 Å and 2.0 Å, respectively. The space group was P2\(_{1}\)2\(_{1}\)2\(_{1}\) with cell dimensions of \(a = 43.512, b = 48.429, c = 95.911\). MAD data were collected to 2.0 Å resolution from a single crystal containing SeMet-labeled protein at three different x-ray wavelengths near the Se edge. Inverse beam strategy was used. The absorption edge was determined by fluorescent scan of the crystal as described previously [13]. The data were processed using HKL2000 suite [14]. Crystal characteristics and data collection statistics are presented in Table 1.

The structure of SlyA was determined using the MAD phasing method. All four Se sites were found in the asymmetric unit. MAD phases were calculated using the CNS suite [15] phases and improved using the density modification (DM) method as implemented by the CNS suite. Electron density maps were high quality and allowed autotracing of the amino acid chains using
the wARP program [16]. The procedure provided an initial model that included the majority of amino acid residues. The model was refined using 1.6 Å data and the CNS suite. Manual adjustment and model building was performed using program QUANTA [17]. The final R-factor and R-free are 24.8% and 27.5%, respectively, with excellent stereochemistry (Table 2). The final structure included 145 amino acid residues and 179 water molecules. The PROCHECK analysis showed that all residues are in the allowed region of the Ramachandran plot [18]. The electron density for the main chain is high quality, except for the N-terminal Met and the five C-terminal residues.

Results and Discussion

Structure of SlyA

The crystal of unliganded SlyA-Ef contains two monomers in an asymmetric unit related by a pseudo-two-fold dyad. The two-fold symmetry of the DNA-binding domains in the crystal structure (Fig. 2) is consistent with the palindromic character of the DNA sequences recognized by members of the SlyA family [9]. The SlyA dimer appears to have a “Swiss cheese”-like structure with numerous channels, cavities, and voids. We suggest that these features are used for interaction with as-yet-unknown signal molecules and other cofactors working together in transcription regulation (Fig. 2b).

The protein dimer consists of three function-related structural elements: a dimerization unit composed of two domains and two independent DNA-binding domains. The dimerization domain includes an N-terminal α-helix, H1, and two C-terminal α-helices, H6 and H7 (Fig. 3). The structure of the dimerization unit reveals domain-swapping. Helix H1 (subunit A) is inserted between H6’ and H7’ (subunit B) and forms a coil-coil with H7’. This interaction is mediated by a number of hydrophobic interactions and a salt bridge between conserved Glu120 and conserved Arg130’ from the opposite subunits. Arg130’ also forms a hydrogen bond to Asp126’. The C-terminus of H6 also contributes to the dimer interface. The interaction between subunits is very extensive, with a total surface of 5025 Å² buried between subunits, a very large area for such a small 150 residues protein.
The DNA-binding domain is formed by the central part of the polypeptide and includes α-helices H2, H3, H4, and H5, and a β-hairpin (residues 73–91). There is virtually no interaction between DNA-binding domains in the SlyA dimer. This domain contains a DNA-binding motif described previously as a winged-helix. In SlyA-Ef, it is composed of H5 and the 19-residue β-hairpin as a wing. This architecture closely resembles the WH motif of the well-established major-groove DNA-binding domain found in a number of transcriptional regulators [6, 19–22].

The WH motif in SlyA-Ef is preceeded by two short α-helices, H3 and H4, with a short β-strand and a small loop between them. The motif is followed by H6, which is a part of the dimerization domain. Two residues conserved in the SlyA family, Thr91 and Gly94, are located at the junction between the wing and H6 (Fig. 1). These residues are also conserved in the MarR family and may be important for the positioning of the DNA-binding domain on its DNA target.

Interactions within the DNA-binding domain are mostly hydrophobic and involve several conserved residues (Figs. 1 and 2). Interactions between the DNA-binding and dimerization domains includes four regions of the protein: (1) the C-terminus of H7 (Trp139 and Lys143); (2) the C-terminus of H1, short helix H2 (Phe25), and a loop between H2 and H3; (3) the N-terminus of H6 (Tyr98); and (4) the N-terminus of H1 of the opposite subunit, including a salt bridge between Glu41 and Arg6’ (these two residues are not strictly conserved in the SlyA family). It has been suggested that conserved glycine residues between secondary structure elements in the WH motif serve as flexible hinges to accommodate DNA targets with different half-site spacings [20]. In SlyA-Ef, glycines 31, 44, 72, and 94 may perform this function.

_SlyA family shows strong structural homology to MarR_

The SlyA-Ef has a number of structural homologues identified in the Protein Data Bank (PDB) both by the program available at the DALI server and by the Secondary Structure Matching Program at EMBL-EBI. The closest match identified by the program is the structure of MexR from _P. aeruginosa_ (PDB code 1LNW). 88% of the secondary structure elements match and the sequences are 17% identical. The r.m.s. deviation between structurally equivalent C-alpha positions of the monomer is 2.07 Å.

The next most significant matches (at 75% of the secondary structure elements matched) are to MarR from _E. coli_ (1JGS) and MotA from bacteriophage T4 (1BJA), with sequence identities of
16% and 12%, and r.m.s. deviations of 2.79 and 3.67 Å, respectively. The “Swiss cheese” characteristic of the SlyA dimer is also apparent in MarR and MexR proteins, although the precise nature of the accessible regions differs among them. This difference may reflect specificity for different signaling molecules or different modes of action.

**Interaction with DNA**

Crystal structures of the three MarR homologues as well as SlyA were determined without their DNA target or a signal molecules bound. The structure of MarR protein includes salicylate, a known inhibitor of DNA binding [6]. The structures of MexR and SlyA-Ef, discussed here, are of unliganded proteins, posing an important question of how these proteins interact with their specific DNA targets and whether these structures can provide new insight into DNA recognition.

The MexR structure shows three different dimers with varied spacing between their WH motifs and high degrees of conformational flexibility [7]. The largest space between Cα positions of Arg73/Arg73’ was 29.9 Å. In the SlyA-Ef structure, the equivalent spacing (Arg67/Arg67’) is shorter (24.63 Å). Analysis of the MarR structure using a program available at the PITA server to generate biological dimers [23] indicates that the spacing between the Cα positions of Arg77/Arg 77’ is 28.1 Å. These key arginine residues exhibit different arrangements in the three structures. In MexR, they are positioned on the DNA-binding surface and are expected to point toward the bases in the DNA major groove. In MarR and SlyA, the arginine side chains are also at the DNA binding surface but are expected to point away from the bases and the major groove. These differences suggest that the mode of binding of these proteins to DNA is not identical. Moreover, these data suggest that MarR/SlyA proteins may span a narrow minor groove and may recognize a rather short palindromic DNA sequence (14–16 bp).

The high-degree of flexibility of DNA-binding domains displayed in different crystals provides indirect evidence of the ability of this fold to adapt in order to recognize various DNA targets. Analysis of the crystal structure of the complex bound to DNA will test this hypothesis. It is also evident that the C-terminus of SlyA is in position to facilitate an interaction with DNA. The C-terminus of SlyA and its homologues contains several conserved positively charged residues that could make contacts with phosphate groups or bases.
In the SlyA-Ef dimer, the WH motifs are positioned to interact with a palindromic recognition sequence, with specific contacts expected to occur predominantly between H5 and the wing, and the major groove of DNA. The WH motif in SlyA-Ef is strongly positively charged due to the presence of five arginines and four lysines; however, these residues are not conserved between SlyA and MaR families, suggesting that different DNA targets or different sets of residues in the same DNA target are recognized.

A cavity between DNA-binding domains in the SlyA-Ef dimer and extending to the dimerization domain could serve as a binding site for a signal molecule. In MexR, a C-terminal tail of the C monomer is inserted into this cavity [7]. The residues lining the cavity are rather hydrophobic, but are not conserved, suggesting that these proteins respond to different signals. Interestingly, binding of a single ligand molecule to the dimer is expected to perturb the two-fold symmetry, allowing the molecule to detect differences in the DNA target sequence. A signal molecule could also control spacing between DNA-binding domains by arranging them differently. Another key difference between SlyA and MarR is that the entrance to the cavity in MarR is closed off by two salt bridges between Asp67 and Arg73. These interactions are not present in the SlyA or MexR proteins. Recent evidence suggests that SlyA-dependent regulation may involve other protein cofactors, such as cAMP receptor or nitrate reduction regulator [24], which may involve direct protein-protein interactions.

Overall, the SlyA structure is very similar to the structures of MarR and MexR, showing that the MarR/SlyA family shares a common fold and similar DNA-binding properties despite their low amino acid sequence similarity. However, these proteins show significant local divergence in sequence and structure that apparently allows them to respond to different signal molecules and to bind to diverse DNA targets by using alternative modes of action. The MarR/SlyA structure is an example of a highly adaptable protein fold and provides a plausible explanation for its wide use in transcription regulation in many different organisms.

Acknowledgments

Atomic coordinates have been deposited in the Protein Data Bank (PDB) with PDB-ID 1LJ9 and accession number RCSB015987. We wish to thank all members of the Structural Biology Center.
at Argonne National Laboratory for their help in conducting these experiments and Lindy Keller
for help in preparation of this manuscript. This work was supported by National Institutes of
Health Grant GM62414 and by the U.S. Department of Energy, Office of Biological and
Environmental Research, under contract W-31-109-Eng-38.
Table 1: Summary of Crystal and MAD Data

|                          |                |                |                |                |
|--------------------------|----------------|----------------|----------------|----------------|
| Unit cell                | a = 43.512 Å, b = 48.429 Å, c = 95.911 Å, α = β = γ = 90° |
| Space group              | P2(1)2(1)2(1)  |
| MW (Da) (145 residues)   | 15950          |
| Mol (AU)                 | 2              |
| SeMet (AU)               | 4              |

| MAD data collection      | Edge          | Peak          | Remote         |
|--------------------------|---------------|---------------|----------------|
| Wavelength (Å)           | 0.9795        | 0.9793        | 0.95354        |
| Resolution range (Å)     | 1.6           | 1.7           | 1.6            |
| No. of unique reflections| 35499         | 30018         | 32997          |
| Completeness (%)         | 98.1          | 99.1          | 91.1           |
| R merge (%)              | 7.3           | 7.1           | 7.3            |
Table 2: Crystallographic Statistics

| Resolution range (Å) | Centric Phasing | Centric FOM | Acentric Phasing | Acentric FOM | All No. | All FOM | All Phasing power |
|----------------------|-----------------|-------------|------------------|--------------|---------|---------|------------------|
| 30.0–1.8             | 0.8308          | 2.5720      | 0.6452           | 2.4129       | 42969   | 0.6571  | 2.4245           |

**Density modification**

0.80826

**Refinement**

| Resolution range (Å) | 30–1.6 |
|----------------------|--------|
| No. of reflections   | 32997  |
| σ cutoff             | 0.0    |
| R-value (%)          | 24.8   |
| Free R-value (%)     | 27.5 (2819) |
| Rms deviations from ideal geometry |
|  bond length (1-2) (Å) | 0.009 |
|  angle (°)           | 1.00   |
|  dihedral (°)        | 17.3   |
|  improper (°)        | 0.67   |
| No. of atoms         |
|  protein             | 2331   |
|  water               | 289    |
| Mean B-factor (Å²) for all atoms | 21.0 |
| Ramachandran plot statistics (%) |
|  residues in most favored regions | 99.2 |
|  residues in additional allowed regions | 0.8 |
|  residues in disallowed region      | 0.0   |
Figure Legends

Figure 1. Multiple sequence alignment and phylogenetic relationships among SlyA homologues. The threonine residue identical in all sequences is shown in dark gray; highly conserved amino acids (identical amino acids present in *E. fecalis* SlyA and at least 12 other proteins) and amino acids with similar properties are shown in gray. The amino acid sequence alignment was created using ClustalX and the neighbor-joining tree was created using PAUP. Bootstrap values (1000 trials) > 50% are shown.

Figure 2. Structure of SlyA-Ef. (a) SlyA-Ef dimer with subunits labeled in shades of blue and orange. Threonine residue present in all members of the MarR/SlyA family is shown in green and other conserved residues are shown in yellow. (b) SlyA-Ef dimer showing the solvent-accessible surface (drawn using 1.4 Å radius) and electrostatic potential. The SlyA-Ef reveals several channels and cavities filled with ordered water molecules. The positively charged surface associated with the HLH motif is likely to bind to a DNA duplex. Drawings prepared with WebLabViewerPro.

Figure 3. Secondary and domain structure of SlyA-Ef. (a) Secondary structure of SlyA-Ef. (b) SlyA-Ef dimerization domain showing domain swapping. (c) DNA-binding domain with winged-helix motif. Drawings prepared with WebLabViewerPro.
Figure 4. Comparison of the SlyA-Ef structure with the structures of MarR and MexR. (a) Superimposition of SlyA-Ef (blue) on MarR (red) subunits. (b) Superimposition of the SlyA-Ef (blue) and MexR (red) structures.
References

1. Perez-Rueda, E., and Collado-Vides, J. (2001) *J. Mol. Evol.* **53**(3), 172-179
2. Huffman, J.L., and Brennan, R.G. (2002) *Curr. Opin. Struct. Biol.* **12**(1), 98-106
3. Kenney, L.J. (2002) *Curr. Opin. Microbiol.* **5**(2), 135-141
4. Dehoux, P., and Cossart, P. (1995) *Mol. Microbiol.* **15**(3), 591
5. Thomson, N.R., Cox, A., Bycroft, B.W., Stewart, G.S., Williams, P., and Salmond, G.P. (1997) *Mol. Microbiol.* **26**(3), 531-544
6. Alekshun, M.N., Levy, S.B., Mealy, T.R., Seaton, B.A., and Head, J.F. (2001) *Nat. Struct. Biol.* **8**(8), 710-714
7. Lim, D., Poole, K., and Strynadka, N.C. (2002) *J. Biol. Chem.* **277**(32), 29253-29259
8. Revell, P.A., and Miller, V.L. (2001) *FEMS Microbiol. Lett.* **205**(2), 159-164
9. Ludwig, A., Bauer, S., Benz, R., Bergmann, B., and Goebel, W. (1999) *Mol. Microbiol.* **31**(2), 557-567
10. Oscarsson, J., Westermark, M., Lofdahl, S., Olsen, B., Palmgren, H., Mizunoe, Y., Wai, S.N., and Uhlin, B.E. (2002) *Infect. Immunol.* **70**(10), 5759-5769
11. Spory, A., Bosserhoff, A., von Rhein, C., Goebel, W., and Ludwig, A. (2002) *J. Bacteriol.* **184**(13), 3549-3559
12. Nagel, G., Lahrz, A., and Dersch, P. (2001) *Mol. Microbiol.* **41**(6), 1249-1269
13. Walsh, M.A., Dementieva, I., Evans, G., Sanishvili, R., and Joachimiak, A. (1999) *Acta Cryst. D Biol. Cryst.* **55**, 1168-1173
14. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
15. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. (1998) *Acta Cryst. D Biol. Cryst.* **54**, 905-921
16. Perrakis, A., Morris, R., and Lamzin, V.S. (1999) *Nat. Struct. Biol.* **6**(5), 458-463
17. QUANTA, San Diego: Molecular Simulations Inc., 2000
18. Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993) *J. Appl. Cryst.* **26**, 283-291
19. Gajiwala, K.S., and Burley, S.K. (2000) *Curr. Opin. Struct. Biol.* **10**(1), 110-116
20. Zhang, R.G., Kim, Y., Skarina, T., Beasley, S., Laskowski, R., Arrowsmith, C., Edwards, A., Joachimiak, A., and Savchenko, A. (2002) *J. Biol. Chem.* **277**(21), 19183-19190
21. Finnin, M.S., Cicero, M.P., Davies, C., Porter, S. J., White, S.W., and Kreuzer, K.N. (1997) *EMBO J.* **16**, 1992-2003

22. Littlewood, T.D. and Evan, G.I. (1994) *Protein Profile.* **1(6)**, 635-709

23. Hannes Listing, 2003, in preparation

24. Westermark, M., Oscarsson, J., Mizunoe, Y., Urbonaviciene, J., and Uhlin, B. E. (2000) *J. Bacteriol.* **182**, 6347-6357
Crystal structure of enterococcus faecalis SlyA-like transcriptional factor
Rui-ying Wu, Rong-guang Zhang, Olga Zagnitko, Irina Dementieva, Natalia Maltzev,
James D. Watson, Roman Laskowski, Piotr Gornicki and Andrzej Joachimiak

J. Biol. Chem. published online March 20, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M300292200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts