Structure
Article

Structure of the *Staphylococcus aureus* AgrA LytTR Domain Bound to DNA Reveals a Beta Fold with an Unusual Mode of Binding

David J. Sidote,[1] Christopher M. Barbieri,[1] Ti Wu,[1] and Ann M. Stock[1,*]
[1]Center for Advanced Biotechnology and Medicine and Department of Biochemistry, Howard Hughes Medical Institute, University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School, 679 Hoes Lane, Piscataway, NJ 08854, USA
*Correspondence: stock@cabm.rutgers.edu
DOI 10.1016/j.str.2008.02.011

SUMMARY

The LytTR domain is a DNA-binding motif found within the AlgR/AgrA/LytR family of transcription factors that regulate virulence factor and toxin gene expression in pathogenic bacteria. This previously uncharacterized domain lacks sequence similarity with proteins of known structure. The crystal structure of the DNA-binding domain of *Staphylococcus aureus* AgrA complexed with a DNA pentadecamer duplex has been determined at 1.6 Å resolution. The structure establishes a 10-stranded β fold for the LytTR domain and reveals its mode of interaction with DNA. Residues within loop regions of AgrA contact two successive major grooves and the intervening minor groove on one face of the oligonucleotide duplex, inducing a substantial bend in the DNA. Loss of DNA binding upon substitution of key interacting residues in AgrA supports the observed binding mode. This mode of protein-DNA interaction provides a potential target for future antimicrobial drug design.

INTRODUCTION

Bacterial response regulators of two-component signal transduction systems are used for transcriptional regulation of a wide range of genes in response to cellular and environmental signals. An analysis of 330 bacterial and archaeal genomes identified 5589 DNA binding response regulators (Galperin, 2006). While all characterized response regulators contain a conserved receiver domain necessary for phosphorylation dependent activation and dimerization, 95.3% interact with DNA through a helix-turn-helix DNA-binding domain or a variation of this domain (e.g., winged helix-turn-helix, helix-ribbon-helix). The remaining 4.7% of proteins annotated as DNA-binding response regulators contain a conserved, uncharacterized DNA-binding domain named the LytTR domain, after the *Bacillus subtillus* LytT and *Staphylococcus aureus* LytR response regulators (Nikolskaya and Galperin, 2002).

The LytTR domain is a ~105-residue bacterial DNA-binding domain found most commonly as the effector module in response regulators, although it can also be found as a stand-alone domain and in combination with membrane-associated MHYT (Met-His-Tyr-Thr), ABC (adenosine triphosphate-binding cassette) transporter, and PAS (Per-Arnt-Sim) domains (Nikolskaya and Galperin, 2002). LytTR domains exist mostly in the genomes of γ-proteobacteria and firmicutes, where they are encoded in only one or two genes per genome—substantially less prevalent than transcription factors of the more populated response regulator subfamilies. A common function associated with response regulators that contain LytTR domains is the regulation of virulence factor and toxin production. Examples of such proteins include *Pseudomonas aeruginosa* AlgR, the response regulator that modulates the production of the exopolysaccharide alginate involved in chronic pneumonia (Lizewski et al., 2002; Mohr et al., 1991); *Clostridium perfringens* VirR, which is required for production of toxins implicated in gas gangrene (Rood, 1998; Shimizu et al., 2002); *Streptococcus pneumoniae* BlpR, a transcriptional regulator of bacteriocin production and the only response regulator required for growth in that organism (Dawid et al., 2007; de Saizieu et al., 2000); *Lactobacillus plantarum* PlnC, which is involved in bacteriocin regulation (Diep et al., 2003; Risøen et al., 2001); and *S. aureus* AgrA, the transcriptional component of a quorum sensing system and global regulator of virulence that up-regulates secreted virulence factors and down-regulates cell wall-associated proteins (Abdelnour et al., 1993; Novick et al., 1995; Novick, 2003). The *agr* locus contains two promoter regions. The first, designated P2, contains two high-affinity LytTR domain binding sites. The second promoter region, P3, is located on the opposite strand, 40 bp downstream of the P2 region, and contains both a high-affinity LytTR binding site and a low-affinity binding site (Koenig et al., 2004).

In most systems studied, LytTR domain-containing response regulators dimerize and bind to direct 9 bp repeats, often imperfect, which are separated by 12 bp and are located just upstream of the transcription start site (Cheung and Rood, 2000; de Saizieu et al., 2000; Diep et al., 1996; Koenig et al., 2004; Mohr et al., 1991, 1992; Risøen et al., 1998; Ween et al., 1999). The number of base pairs separating the repeats has been reported to be crucial, as reducing the number by a single bp renders the promoter nonfunctional (Knutsen et al., 2004; Risøen et al., 2001). An exception has been observed in the AlgR system, in which two binding sites separated by 66 bp are located far upstream and a third binding site is located just upstream of the transcriptional start site. This arrangement has been implicated in a DNA looping mechanism required for transcriptional activation (Mohr et al., 1990).

Although the role of LytTR domain transcription factors in the regulation of bacterial virulons has been investigated in many pathogenic organisms, structural information is lacking. The absence of sequence similarity to any characterized DNA-binding motif suggests that the LytTR domain might be a unique fold, and little is known about the molecular interaction between LytTR domains and DNA. To address these questions, we have determined the X-ray crystal structure of the C-terminal DNA-binding domain of the *S. aureus* response regulator AgrA in complex with a 15 bp DNA duplex containing a 9 bp consensus binding sequence. The LytTR domain consists of a 10 stranded elongated β-β-β fold that appears to have arisen from duplication of a 5 stranded β motif. Residues within three loops protruding from an edge of the β sheets make base-specific contacts to two adjacent major grooves and the intervening minor groove of the DNA. The AgrA-DNA interface is sterically and electrostatically complementary, with a buried surface of 1500 $Å^2$, and the DNA is significantly bent and distorted as it conforms to the protein surface. This novel protein-DNA interaction provides a unique target in pathogenic bacteria that might be exploited for antimicrobial drug design.

RESULTS AND DISCUSSION

Crystallization of an $AgrA_C$-DNA Complex

Initial attempts to produce quantities of full-length *S. aureus* AgrA suitable for structure determination were unsuccessful due to insolubility. To circumvent this problem, the C-terminal DNA-binding domain, termed $AgrA_C$ (residues 137–238 with an initiator methionine), was expressed and purified. We were unsuccessful in crystallizing $AgrA_C$ alone, and so pursued investigation of $AgrA_C$-DNA complexes.

We first determined the minimum sequence necessary to obtain a stable interaction between $AgrA_C$ and DNA oligonucleotides. Because we were investigating the isolated C-terminal domain, which migrates as an apparent monomer during gel filtration chromatography (data not shown), a single 9 bp repeat was used as the basis for construction of a set of oligonucleotides for DNA-binding studies. Beginning with a 19 bp oligonucleotide that contained the 9 bp consensus binding sequence flanked by 5 bp on each side, corresponding to the sequence of the upstream binding site in the P2 promoter region of the *agr* locus, we systematically shortened the flanking regions 1 bp at a time, while simultaneously adding 1 bp to the opposite flanking region to maintain a constant length of 19 bp, until we could no longer detect a stable interaction by using electrophoretic mobility shift assays. We determined that ~3 bp on either side of the 9 bp consensus binding sequence were necessary for optimal $AgrA_C$-DNA interaction.

Using this 15 bp minimum sequence as a starting point, oligonucleotides of various lengths, with and without 1-nucleotide overhangs, were designed for crystallization trials. Diffraction quality crystals were obtained by cocrystallizing $AgrA_C$ with a 15 bp duplex containing 1 nucleotide overhangs at the 5′ end of each strand, and centered on the 9 bp consensus binding sequence.

Structure Determination and Analysis

The X-ray crystal structure of the $AgrA_C$-DNA complex was determined at 1.6 Å resolution using single-wavelength anomalous diffraction phasing with oligonucleotides containing two 5-bromo deoxyuridine bases at positions 14 on strand A and 15′ on strand B (Figure 1A). The crystals belonged to space group $P4_1$ with one $AgrA_C$-DNA complex per asymmetric unit and a solvent content of 52%. The final electron density maps were of excellent quality and allowed for unambiguous construction of a model containing residues 136–238 of $AgrA_C$ as well as the complete DNA molecule. The crystallographic R and R_{free} were 0.195 and 0.222, respectively, and the final model exhibited excellent geometry (Table 1). Within the crystal lattice, the DNA forms a pseudo-continuous helix along the crystallographic *c* axis through a reverse Hoogsteen interaction between complementary overhangs with symmetry-related molecules. Other packing interactions occur between strand β8 of one $AgrA_C$ and β5 of an adjacent molecule and between the extended N-terminal 6 residues of $AgrA_C$ and the DNA major groove of a symmetry-related complex.

Elongated β-β-β Fold in the LytTR Domain

The structure of $AgrA_C$ reveals a novel topology, having 10 β strands arranged into three antiparallel β sheets and a small two-turn α helix that is not involved in DNA binding (Figures 1B–1E). The sheets are arranged roughly parallel to each other in an elongated β-β-β sandwich. A hydrophobic five-stranded β sheet (sheet 2, β3–β7) is at the center of the domain with two smaller amphipathic β sheets (sheet 1, β1–β2; sheet 3, β8–β10) positioned on either side (Figures 1C and 1D).

The α helix is located between strands β5 and β6 and is packed along the edges of sheet 1 and sheet 2. Each sheet is linked to the next by a 3_{10} turn followed by a buried isoleucine that anchors the turn. A salt bridge interaction between residues D157 and H208, each located in a turn of 3_{10} helix, act to further stabilize the connections between the sheets. Salt bridges are also observed between residues R195, located in the loop between helix 1 and strand β6, and residue E141 located in the beginning of strand β1 and D157. Salt bridges between residues D176 and K223 and between H174 and E226 also stabilize the interaction between sheets 2 and 3 (Figure 2). A single turn of 3_{10} helix at the C terminus packs between sheets 2 and 3. The 6 residues at the N terminus of $AgrA_C$ adopt an extended conformation that is likely to be different in the context of the full-length protein (Figures 1D and 1E).

An interesting feature of this fold is the two-fold symmetry between strands β1–β5 and β6–β10. A least-squares alignment of residues 136–193 and 194–238 shows that the connectivity and spatial positioning of the strands is remarkably similar, suggesting that this fold is derived from duplication of a smaller domain (Figure 1C).

Structural Homology between $AgrA_C$ and Sac7d

DALI (Holm and Sander, 1995) and VAST (Gibrat et al., 1996) structure homology searches revealed that $AgrA_C$ is similar to the well-characterized 66 residue archaeal DNA-binding protein Sac7d from *Sulfolobus acidocaldarius* (Sso7d in *Sulfolobus solfataricus*) and a 74 residue protein of unknown function from *P. aeruginosa* (Lin et al., 2006). Sac7d/Sso7d are basic, highly abundant 7 kDa proteins that bind nonspecifically to DNA conferring thermo, chemical, and acid stability. The structure of Sac7d/Sso7d is comprised of five antiparallel β strands

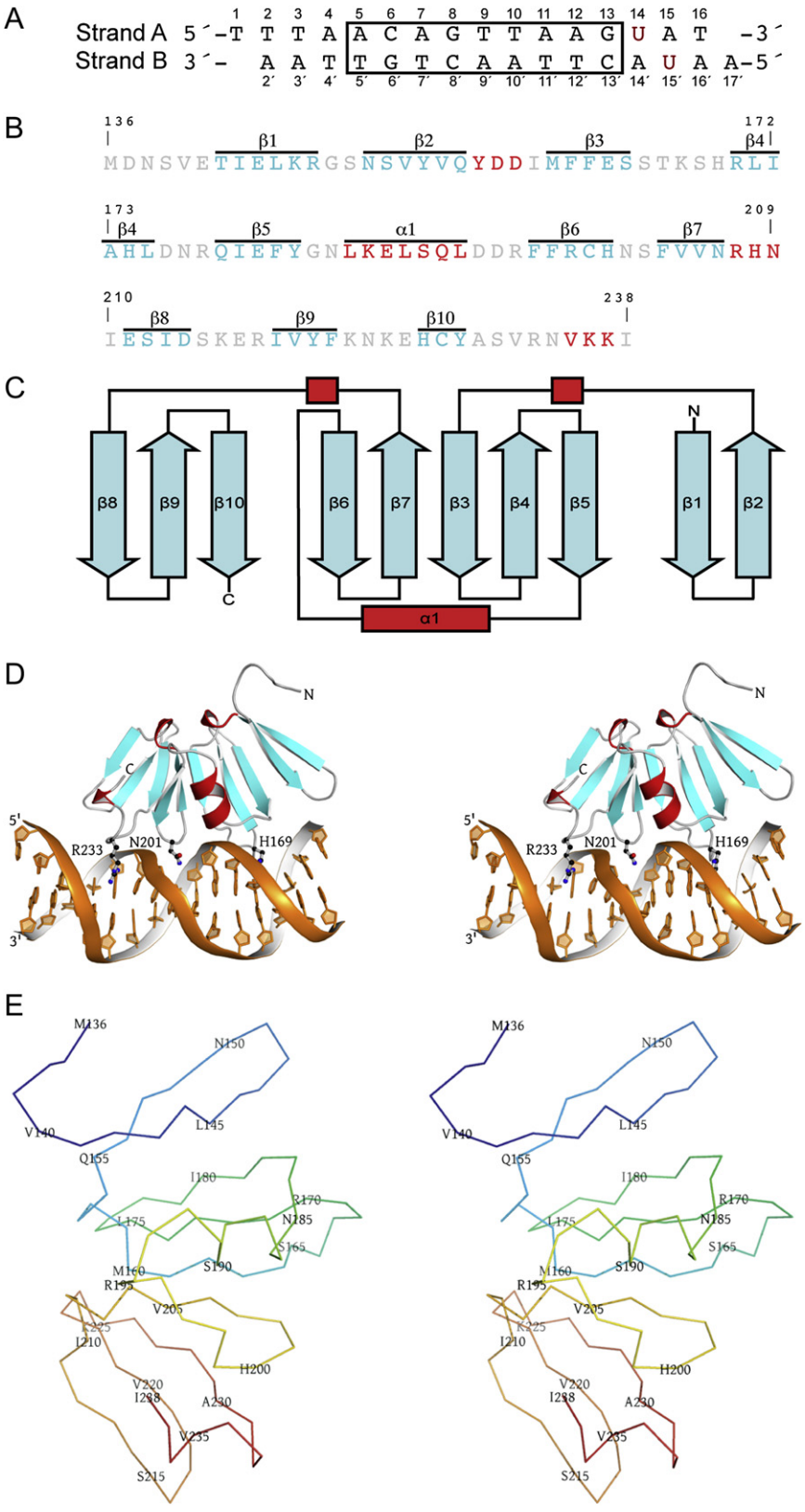

Figure 1. Structure of the AgrA$_C$-DNA Complex

(A) Sequence of the DNA oligonucleotide crystallized in complex with AgrA$_C$. The highly conserved LytTR consensus sequence is boxed. A red U indicates each position substituted with 5′-bromouracil. The coding strand is labeled as Strand A and the noncoding strand as Strand B.

(B) Sequence of the C-terminal DNA-binding domain of AgrA (residues 137–238 and initiator methionine). Loop regions are colored gray, helices are colored red, and strands are colored cyan. The secondary structure elements are indicated above each region.

(C) Topology diagram of AgrA$_C$. Secondary structure elements are colored as described in (B).

(D) A stereo ribbon representation of the AgrA$_C$-DNA complex. AgrA$_C$ is shown in an orientation and color scheme similar to that in (B). Residues that make base-specific contacts with the DNA are shown in ball-and-stick representation.

(E) Stereo view of an alpha carbon trace of AgrA$_C$. The chain is color-ramped from blue at the N terminus to red at the C terminus, with residue numbers indicated at every fifth Cα.

Table 1. Data Collection and Refinement Statistics

Data collection	
Space group	$P4_1$
Cell dimensions	
a, *b*, *c* (Å)	47.9, 47.9, 100.1
α, β, γ (°)	90, 90, 90
Resolution (Å)	17.3–1.60 (1.66–1.60)[a]
R_{merge}[b] (%)	1.7 (15.5)
I/σI	41.7 (5.7)
Completeness (%)	98.1 (97.3)
Redundancy	3.8 (3.7)
Refinement	
Resolution (Å)	17.3–1.6
No. reflections	27,747
R_{work}[c]/R_{free}[d] (%)	19.5/22.2
No. atoms	
Protein	872
DNA	650
Water	243
Ions	2
Average B-factors (Å^2)	
Protein	16.5
DNA	19.6
Water	21.9
Ions	19.9
Rmsd[e]	
Bond lengths (Å)	0.007
Bond angles (°)	1.38

[a] Values in parentheses are for the highest-resolution shell.
[b] $R_{merge} = (\Sigma_h\Sigma_i | < I(h) > - I(h)_i|)/\Sigma_h\Sigma_i I(h)_i$, where $I(h)_i$ is the i^{th} observation of reflection h, and < I(h) > is the mean intensity of all observations of reflection h.
[c] $R_{work} = (\Sigma||F_o| - |F_c||)/\Sigma|F_o|$, where $|F_o|$ and $|F_c|$ are observed and calculated structure factor amplitudes, respectively.
[d] R_{free} was calculated for 5% of the randomly selected reflections of data sets that were not used in the refinement.
[e] Rmsd, root-mean-square deviation.

arranged into a two-stranded β sheet connected to a three-stranded β sheet by way of a single turn of 3_{10} helix, remarkably similar to strands 1–5 and 6–10 of $AgrA_C$ (Figure 3A; McCrary et al., 1996; Robinson et al., 1998). Sac7d/Sso7d contact DNA with the β sheet of Sac7d/Sso7d interacting in the minor groove by intercalating two hydrophobic residues between the DNA bases (Figure 3B; Peters et al., 2004; Robinson et al., 1998). This interaction causes significant unwinding and severely kinks the DNA by 72°.

The mode of interaction of Sac7d/Sso7d with DNA is distinctly different than that of $AgrA_C$, which binds specifically to a conserved DNA sequence through loops that act as fingers, inserting themselves into successive major grooves and gently bending the DNA (Figure 3B). Although Sac7d/Sso7d share identical secondary structure topology and nearly identical tertiary structure with the subdomains of $AgrA_C$, these proteins have evolved completely different mechanisms for interacting with DNA.

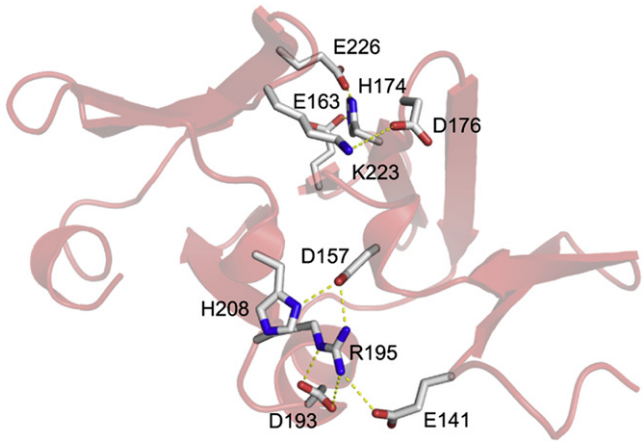

Figure 2. Salt Bridge Interactions in $AgrA_C$
Ribbon representation of $AgrA_C$ displaying salt bridges that stabilize the fold. Side chains of residues involved in salt bridge interactions are shown in stick representation with carbon in white, oxygen in red, and nitrogen in blue. The salt bridges are concentrated on the surface opposite of the surface that contacts DNA.

Binding of $AgrA_C$ Significantly Distorts DNA

The majority of the DNA within the $AgrA_C$-DNA complex adopts the B form, although local distortions demonstrate characteristics of both C and D forms. The DNA adopts a global bend of ~38° as it conforms to the concave DNA-binding surface of $AgrA_C$ (Figure 1D). Previously reported footprinting experiments performed with full-length AgrA bound to the P2 region demonstrated an increased sensitivity to DNase I along a single face of the DNA, indicating that AgrA binds along a single face of the DNA (Koenig et al., 2004), consistent with the structure presented here. The binding interaction causes a compression of the two major grooves and the minor groove between them. Strong distortion is observed for bp A7′–T7′, located just after the specific interaction with residue R233 in the major groove, and A12′–T12′, located just before the specific H169 interaction in the succeeding major groove (Figures 4A and 4B). In both base pairs there is a high degree of stagger, buckle, and opening as well as base-step distortions in the tilt and roll. The base pairs T9′–A9′ and T10′–A10′, absolutely conserved in all LytTR domain consensus binding sequences, demonstrate a high degree of buckle and tilt as well as a large degree of twist, thus acting as a hinge that enables the DNA to conform to the protein surface.

$AgrA_C$ Contacts Successive Major Grooves

Unlike most other major groove DNA-binding proteins, $AgrA_C$ does not interact through a recognition helix. Instead, it sits on the DNA and inserts long loops into successive major grooves, making single direct base contacts in each major groove (Figure 1D). $AgrA_C$ binds to a single face of the DNA with its long axis running along the DNA backbone, covering nearly 16 bp or a length of ~50 Å, with a buried surface area of 1516 Å^2. The N terminus of the domain is positioned at the edge of the β sheet that lies opposite to the edge contacting DNA. Thus the N-terminal regulatory domain attached to this terminus would presumably be positioned in a location sterically compatible with DNA binding by the LytTR domain. The C terminus of the

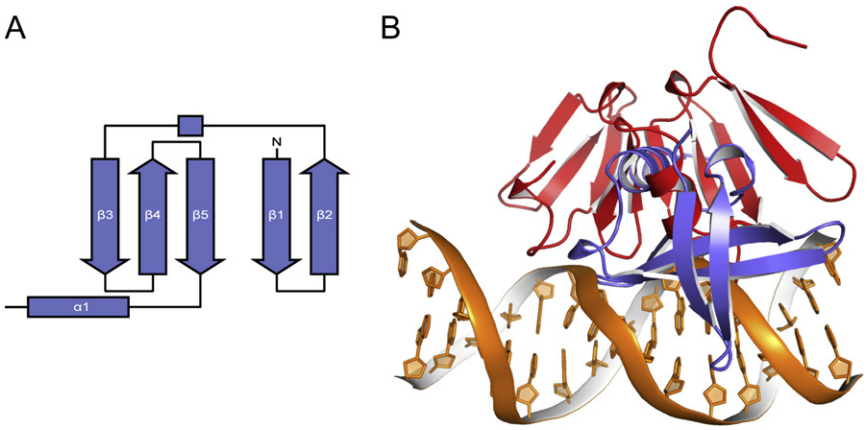

Figure 3. Structural Comparison of $AgrA_C$ and Sac7d

(A) Topology diagram of *Sulfolobus acidocaldarius* Sac7d.

(B) Different DNA-binding orientations of $AgrA_C$ and Sac7d. Structures of the DNA molecules within the $AgrA_C$-DNA and Sac7d-DNA (Robinson et al., 1998) complexes were superimposed using the program SuperPose (Maiti et al., 2004). For clarity in illustrating the different orientations of the two proteins when interacting with DNA, only a single DNA duplex, that of the $AgrA_C$-DNA complex, is shown.

domain also lacks contacts with DNA. This is interesting in regard to mutations at the C terminus of AgrA that have been identified in both laboratory and clinical strains of nonhemolytic *S. aureus* (Traber and Novick, 2006). A frameshift mutation that lengthens AgrA by 3 residues results in partially defective Agr-mediated transcription, as evidenced by delayed and lower production of RNAIII, while a mutation that adds 21 amino acids results in complete loss of RNAIII production. The structure provides no explanation for how these altered C-terminal residues would directly interfere with DNA binding. Thus it seems likely that the transcriptional defects in these variant AgrA proteins result from either decreased protein stability or loss of other intramolecular or intermolecular interactions, such as contact with the regulatory domain, polymerase, or other transcription factors.

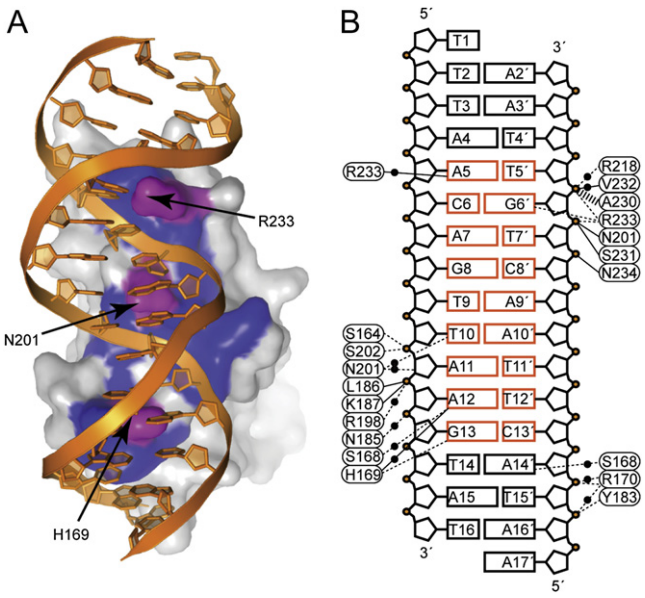

Figure 4. Interactions between $AgrA_C$ and the LytTR Consensus Sequence

(A) Surface representation of $AgrA_C$ bound to DNA showing the DNA contact region of $AgrA_C$. The DNA, shown in a ribbon representation, is colored orange. The surface of $AgrA_C$ is shown in white with residues that make base-specific contacts colored magenta and those involved in nonspecific contacts colored blue.

(B) Schematic diagram of $AgrA_C$-DNA interactions. Amino acid residues of $AgrA_C$ are identified as ovals and bases are identified as rectangles with the consensus DNA-binding sequence colored orange. Solid lines indicate contacts between backbone atoms of amino acid residues and DNA. Dashed lines indicate contacts between side-chain atoms and DNA. Black circles indicate water-mediated interactions. The wide dashed line indicates a van der Waals interaction.

There are several direct and water-mediated nonspecific stabilizing contacts between the protein and the DNA (Figure 4). Indeed, the ratio of nonspecific DNA backbone interactions (10) to direct base contacts (2) made by $AgrA_C$ is similar to that observed for nonspecific DNA-binding proteins (Luscombe and Thornton, 2002). However, binding of AgrA is specific for the highly conserved 9 bp consensus sequence. Because many of the base pairs within this consensus make neither direct nor water-mediated contacts to $AgrA_C$, it is likely that the strict DNA sequence conservation reflects a role other than that in protein-DNA interactions, such as allowing for specific DNA conformation. The regions flanking the 9 bp consensus binding sequence are involved in nonspecific DNA backbone interactions with $AgrA_C$ and appear to contribute primarily to stability rather than specificity.

The protein has a calculated pI of 7.8 and an asymmetric distribution of surface charge. An electrostatic representation shows that $AgrA_C$ contains the highest concentration of positively charged residues along the DNA-binding surface, while the remainder of the protein is predominately negatively charged (data not shown). The shape and charge complementarity between $AgrA_C$ and DNA indicate a snug interaction within the complex.

H169 and R233 Make Specific DNA Contacts

Surprisingly, only two amino acid side chains are involved in direct base-specific interactions within the $AgrA_C$-DNA complex. Residue H169, located in the loop between β4 and β5, interacts with G13 on strand A, forming a Nε2-O6 hydrogen bond interaction. The backbone amide of H169 also forms a water-mediated hydrogen bond with A12 N7 (Figure 5A). In the succeeding major groove, residue R233, located in the loop between β10 and the C-terminal 3_{10} helix, forms a bidentate NH1-N7 and NH2-O6 hydrogen bond interaction with G6′ on strand B, as well as a water-mediated interaction between NH_2 and N6 of A5 (Figure 5A). These two specific interactions involve G-C base pairs separated by 6 intervening base pairs, providing sequence specificity. In the minor groove, N201 makes side-chain water-mediated contacts with T10 and the ribose of A11.

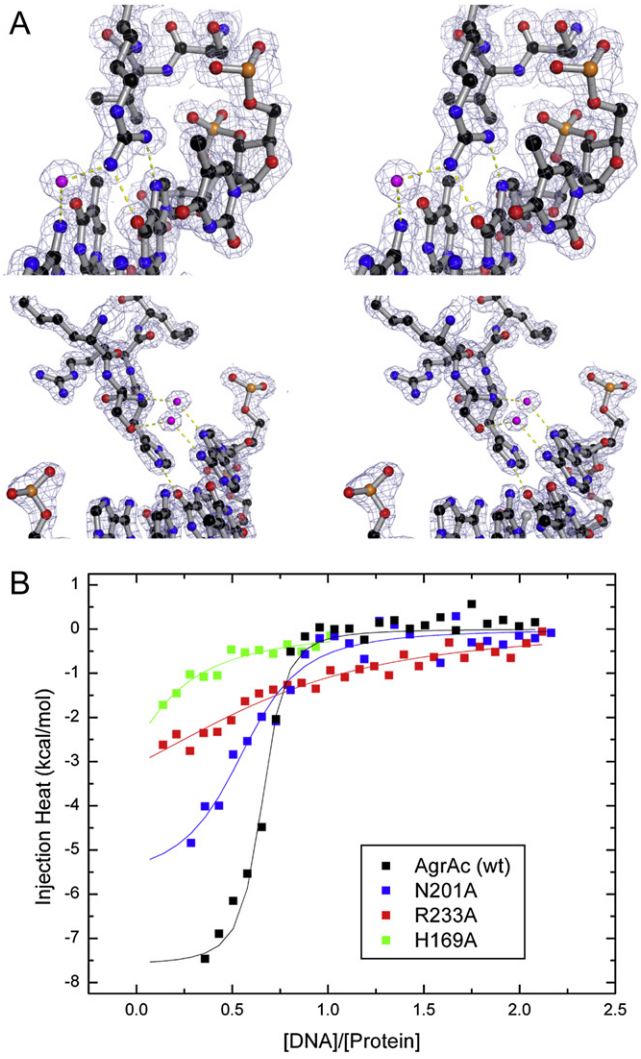

Figure 5. $AgrA_C$ Residues Required for Specific Interaction with DNA

(A) Stereo views of $2F_O - F_C$ electron density contoured at 1.5σ (gray mesh) for the residues of $AgrA_C$ that make base-specific contacts with DNA. Residue R233 (top) and H169 (bottom) are shown in ball-and-stick representation with carbon in black, oxygen in red, nitrogen in blue, and phosphorus in gold. (B) Isothermal titration calorimetry analyses of wild-type and mutant $AgrA_C$ proteins binding to DNA. Titrations were performed as described in Experimental Procedures. Data were fit with a model for one binding site and yielded K_d values of 0.08, 3, 0.7, and 7 μM for wild-type, H169A, N201A, and R233A $AgrA_C$, respectively.

Although the LytTR domain DNA consensus binding sequence is highly conserved, the residues that make base-specific contacts in AgrA (H169, N201, and R233) are not strictly conserved in the LytTR family. However, assessment of conservation is hindered by several factors. A relatively low level of sequence identity between family members, short β strands with poorly defined end points and register, and a relatively large percentage of residues in variable loop regions preclude unambiguous sequence alignments. Furthermore, it is the loop regions of the LytTR domain, rather than repetitive secondary structure elements, that contact DNA. It is likely that many residues within loops will not superimpose in different three-dimensional structures, thus similar functional roles might be provided by residues with slightly different positions in primary sequence. This might be especially true for residues corresponding to H169 and R233, which lie within relatively large loops. Additionally, the relatively large range of geometries available to residues within loops and the added versatility of water-mediated contacts would potentially allow for similar protein-DNA contacts to be mediated by different amino acid side chains.

H169 and R233 are Essential for DNA Binding

In order to assess the importance of side chains that were observed to make base-specific contacts in the $AgrA_C$-DNA complex, site-specific substitutions of alanine were introduced into $AgrA_C$ at positions H169, N201, and R233. The DNA-binding affinities of the wild-type and alanine-subsituted $AgrA_C$ proteins were compared by isothermal titration calorimetry. Wild-type $AgrA_C$ binds specifically to a 19 bp duplex containing the 9 bp consensus binding sequence with a K_d of ~80 nM (Figure 5B).

When H169 or R233, residues that make direct contacts to DNA bases, are substituted with alanines, the K_d values are ~40- to 90-fold higher (Figure 5B). These data are consistent with results reported for VirR, the *C. perfringens* LytTR domain-containing response regulator, which showed that mutating a residue corresponding to R233 in $AgrA_C$ abolished DNA-binding activity (McGowan et al., 2003). Moreover, an N201A substitution in $AgrA_C$ results in a significant decrease in binding efficiency, with an almost 10-fold increase in K_d, suggesting that this interaction might be important for stabilizing the $AgrA_C$-DNA interaction at the point where the greatest bending of DNA occurs. This region in *C. perfringens* VirR was previously investigated by site-specific substitution, and consistent with the results observed with $AgrA_C$, alterations of this residue caused diminished DNA binding (McGowan et al., 2002).

The DNA-binding affinity observed for $AgrA_C$ and a DNA duplex containing a single binding site is significantly weaker than affinities previously reported for full-length AgrA. Using electromobility shift assays and the dimeric binding site within the P2 promoter, K_d values of 3.8 nM and 0.16 nM were determined for unphosphorylated and phosphorylated full-length AgrA, respectively (Koenig et al., 2004). The 500-fold higher binding affinity of phosphorylated AgrA relative to that of the isolated DNA-binding domain likely reflects either the enhanced affinity of a dimeric binding site, contributions of the N-terminal domain to binding, or a combination of both.

Insights from a Model of the $AgrA_C$ Dimer

The characteristics of DNA binding sites for LytTR domain transcription factors have been studied for *S. aureus* AgrA (Koenig et al., 2004), *C. perfringens* VirR (Cheung and Rood, 2000), *L. plantarum* PlnC and PlnD (Risøen et al., 1998, 2001; Straume et al., 2006), *S. pneumoniae* BlpR and ComE (Knutsen et al., 2004), and *P. aeruginosa* AlgR (Kato and Chakrabarty, 1991; Mohr et al., 1990, 1991, 1992). In all cases except the AlgR promoter, which consists of two tandem repeats and another repeat located far upstream of the transcriptional start site, the LytTR domain binding region consists of two direct repeats separated by exactly 12 bp. The exact spacing has been found to be

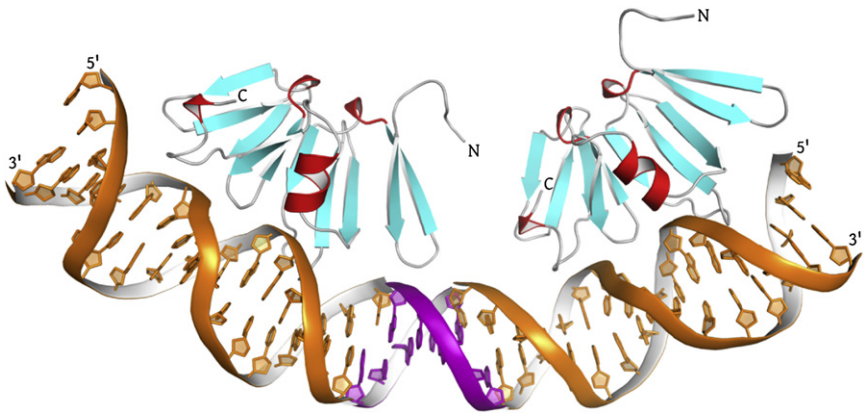

Figure 6. Model of the $AgrA_C$ Dimer Bound to DNA

An $AgrA_C$ dimer bound to direct repeat recognition elements was modeled by aligning two $AgrA_C$-DNA complexes together with 3 bp of B-form DNA (purple) in order to achieve the proper 12 bp spacer between the consensus binding sequences. The molecules were manually aligned in PyMol (Delano, 2002) using the termini of the DNA strands as guides.

important for binding, but the molecular basis of this requirement has not been determined.

To investigate this issue, we constructed a model of $AgrA_C$ domains bound to two direct repeats with the intervening 12 bp spacer modeled as B-form DNA (Figure 6). When the $AgrA_C$-DNA complex is positioned in a tandem orientation with the 12 bp intervening sequence, the two $AgrA_C$ molecules sit along the same face of the DNA. The distance between the two protein molecules (10.1 Å at their closest point of contact) suggests that they do not interact, and that the dimeric interaction must be mediated solely by the N-terminal domains. In this scenario, the stringent spacing between the DNA binding sites would propagate from the orientation and spacing of the N-terminal domains within the dimer through rigid interfaces between the N-terminal and C-terminal domains of each protomer. This model implies that the N-terminal domains dimerize with translational rather than rotational symmetry, paralleling the symmetry of the direct repeat DNA binding sites.

EXPERIMENTAL PROCEDURES

Plasmid Construction

A T7-based inducible expression vector for the C-terminal domain of *S. aureus* AgrA ($AgrA_C$, residues 137–238) was constructed by polymerase chain reaction amplification of the coding region using plasmid pJR1 DNA (Koenig et al., 2004) as template. The forward primers 5′-GATCCATATGGATAATAGC GTTGAAACGATTGAATTAAAACG-3′ and reverse primer 5′-GATCAAGCTTTT ATATTTTTTTAACGTTTCTCACCGATG-3′ introduced a unique NdeI restriction site and initiator methionine codon immediately prior to the codon for Asp137 and a unique BamHI restriction site immediately following the stop codon. The polymerase chain reaction product was digested with NdeI and BamHI and ligated into pET9a (Novagen), generating plasmid pDS3. The nucleotide sequence of the coding region of pDS3 was determined by automated sequencing and verified (GenBank 78172216). Expression vectors for $AgrA_C$ mutant proteins (H169A, N201A, R233A, and H169A/R233A) were produced from plasmid pDS3 using the Quick-change Mutagenesis kit (Stratagene), and sequences of the coding region were determined.

Expression, Purification, and Crystallization

Recombinant $AgrA_C$ was expressed from pDS3 in *Escherichia coli* BL21(DE3) (Novagen). Cells were grown with shaking in 6 l Terrific Broth with 30 mg/l kanamycin at 37°C to an optical density at 600 nm of ~0.5, then cooled to 18°C. Protein expression was induced by addition of 0.3 mM isopropyl-β-D-thiogalactopyranoside and incubation was continued with shaking at 18°C for 24 hr. Cells were collected by centrifugation at 6400 × g for 10 min and frozen at −20°C. Frozen cells were resuspended in 20 mM sodium phosphate (pH 7.0), 0.1 M NaCl, 5 mM phenylmethylsulfonyl fluoride at a ratio of 3 ml/g cells and then lysed by sonication at 0°C. All subsequent steps were performed at 4°C, unless indicated otherwise. All protein solutions were filtered through 0.2 μm filter units prior to loading onto columns. Unbroken cells, cell debris, and membranes were removed by centrifugation at 100,000 × g for 1 hr. $(NH_4)_2SO_4$, 51.6 g per 100 ml lysate, was added with constant stirring on ice for > 30 min, and precipitated protein was collected by centrifugation at 26,900 g for 20 min. The pellet was resuspended in 20 mM sodium potassium phosphate (pH 7.0), 0.1 M NaCl, dialyzed against the same buffer, and applied to 2 × 5 ml HiTrap SP HP columns (GE Healthcare) equilibrated with 20 mM sodium potassium phosphate (pH 7.0), 0.1 M NaCl. Bound $AgrA_C$ was eluted using a 10 column volume 0.3–0.6 M NaCl gradient in 20 mM sodium potassium phosphate. Fractions containing $AgrA_C$ were pooled, and 4.0 M $(NH_4)_2SO_4$ was added to a final concentration of 1.0 M, then applied to a HiLoad 16/10 Phenyl Sepharose HP column (GE Healthcare) equilibrated with 20 mM sodium postassium phosphate (pH 7.0), 1.0 M $(NH_4)_2SO_4$. $AgrA_C$ was eluted in the column flow-through. Fractions containing $AgrA_C$ were pooled, concentrated to ~4 ml using an Amicon Ultra-15 (Ultracel10) centrifugal filter unit (Millipore), and applied to a HiLoad 26/60 Superdex 75 gel filtration column (GE Healthcare) equilibrated with 20 mM Bis-Tris (pH 6.0), 0.1 M NaCl. $AgrA_C$ migrated as a single peak with a mobility corresponding to ~12 kDa. The identity of the protein was confirmed by N-terminal amino acid sequence analysis. Yields of soluble $AgrA_C$ were typically 5–10 mg/l cells. Alanine-substituted $AgrA_C$ proteins were purified by similar procedures.

Oligonucleotides used in crystallization, 5′-TTTAACAGTTAAGTAT-3′ (designated strand A, nucleotides 1-16) and 5′-AATACTTAACTGTTAA-3′ (designated strand B, nucleotides 17′–2′), were purchased from Integrated DNA Technologies with standard desalting. Duplex DNA was formed by resuspending the single-stranded oligonucleotides in annealing buffer (10 mM Tris-Cl [pH 7.5], 150 mM NaCl, 5 mM $MgCl_2$), mixing equimolar amounts of each strand and heating to 70°C for 10 min followed by slow cooling to 20°C overnight. Modified DNA containing two covalently bound bromines (5′-bromouracil in place of thymine at nucleotides 14′ and 15′) was used for crystallographic phasing. The $AgrA_C$-DNA complex was formed by mixing 1.0 mM $AgrA_C$ with 1.2 mM DNA, incubated for 30 min, and passed through a 0.2 μm filter. Crystals of the $AgrA_C$-DNA complex were grown at 20°C using the hanging drop vapor diffusion method. Crystals were obtained in 2 days using 40% PEG 400, 0.1 M Bis-Tris (pH 5.5) as the reservoir solution and equal volumes of reservoir and $AgrA_C$-DNA solution in the drop. No additions were required for cryoprotection.

Structure Determination

A single-wavelength anomalous diffraction dataset was collected at 100 K using an ADSC Q4R CCD detector at the National Synchrotron Light Source beamline X4A. The peak wavelength was determined to be 0.9204 Å by a fluorescence scan. Data were collected using the inverse beam method with an oscillation angle of 1.5° per frame. Diffraction intensities were integrated and scaled using HKL2000 (Table 1; Otwinowski and Minor, 1997). The asymmetric unit contains one $AgrA_C$-DNA complex, with a solvent content of 52%.

The positions of the two Br atoms and the phases were calculated and refined using SHELX (Schneider and Sheldrick, 2002). Following solvent

flattening with DM (Cowtan, 1994), the correct hand was interpretable and an initial model was built into electron density using Coot (Emsley and Cowtan, 2004). Phases were further improved by positional refinement using REFMAC (Murshudov et al., 1997) with data to 1.8 Å, followed by multiple cycles of manual rebuilding and positional refinement and refinement of atomic B-factors using data to 1.6 Å resolution. Water molecules were added where supported by both chemistry and geometry, and the difference electron density was >3σ and the $2|F_{obs}| - |F_{calc}|$ density was >1σ. The quality and stereochemistry of the model were evaluated using PROCHECK (Laskowski et al., 1993). Refinement statistics are shown in Table 1. Structural images were generated using PyMol (Delano, 2002).

Analysis of DNA Binding by Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) measurements were conducted at 25°C on a MicroCal VP-ITC (MicroCal, Inc., Northampton, MA). The solution conditions for all ITC measurements were 10 mM potassium phosphate (pH 7.0), 0.1 M NaCl. In each experiment, 10 μl aliquots of a 150 μM solution of 19-mer duplex DNA (ATTTAACAGTTAAGTATTT and complement) were sequentially injected from a 300 μl rotating syringe (300 rpm) into an isothermal sample chamber containing 1.42 ml of 15 μM protein. The duration of each injection was 10 s, with a 60 s initial delay prior to the first injection and 300 s delays between injections. Each DNA-protein experiment was accompanied by the corresponding control experiment, in which the 19-mer duplex DNA was injected into a solution of buffer alone. Each injection generated a heat burst curve (μcal/s versus s), the area under which was determined by integration (using Origin version 7.0 software [MicroCal, Inc., Northampton, MA]), to obtain a measure of the heat associated with that injection. The measure of the heat associated with each DNA-buffer injection was subtracted from that of the corresponding heat associated with each DNA-protein injection to yield the heat of DNA binding for that injection. The buffer-corrected ITC profiles for the binding of each protein to 19-mer duplex DNA were fit with a model for one set of binding sites.

ACCESSION NUMBERS

The atomic coordinates and structure factors for $AgrA_C$ have been deposited in the Protein Data Bank with the accession code 3BS1.

ACKNOWLEDGMENTS

We thank Robin Koenig and Barry Hurlburt for plasmid pJR1 and Randy Abramowitz and John Schwanof at the NSLS X4 beamlines for technical assistance during data collection. We thank Michael Galperin for valuable discussions and insight regarding LytTR domains. This study was supported in part by National Institutes of Health Grant 2R37GM47958. A.M.S. is an investigator of the Howard Hughes Medical Institute. The authors declare no competing financial interests.

Received: January 2, 2008
Revised: February 8, 2008
Accepted: February 11, 2008
Published: May 6, 2008

REFERENCES

Abdelnour, A., Arvidson, S., Bremell, T., Ryden, C., and Tarkowski, A. (1993). The accessory gene regulator (*agr*) controls *Staphylococcus aureus* virulence in a murine arthritis model. Infect. Immun. *61*, 3879–3885.

Cheung, J.K., and Rood, J.I. (2000). The VirR response regulator from *Clostridium perfringens* binds independently to two imperfect direct repeats located upstream of the *pfoA* promoter. J. Bacteriol. *182*, 57–66.

Cowtan, K. (1994). "DM": an automated procedure for phase improvement by density modification. In Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, *Volume 31*, S. Bailey and M. Winn, eds. (Daresbury, UK: Daresbury Laboratory), pp. 34–38.

Dawid, S., Roche, A.M., and Weiser, J.N. (2007). The *blp* bacteriocins of *Streptococcus pneumoniae* mediate intraspecies competition both *in vitro* and *in vivo*. Infect. Immun. *75*, 443–451.

de Saizieu, A., Gardes, C., Flint, N., Wagner, C., Kamber, M., Mitchell, T.J., Keck, W., Amrein, K.E., and Lange, R. (2000). Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. J. Bacteriol. *182*, 4696–4703.

Delano, W.L. (2002). The Pymol Molecular Graphics System (San Carlos, CA: DeLano Scientific).

Diep, D.B., Havarstein, L.S., and Nes, I.F. (1996). Characterization of the locus responsible for the bacteriocin production in *Lactobacillus plantarum* C11. J. Bacteriol. *178*, 4472–4483.

Diep, D.B., Myhre, R., Johnsborg, O., Aakra, A., and Nes, I.F. (2003). Inducible bacteriocin production in *Lactobacillus* is regulated by differential expression of the *pln* operons and by two antagonizing response regulators, the activity of which is enhanced upon phosphorylation. Mol. Microbiol. *47*, 483–494.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132.

Galperin, M.Y. (2006). Structural classification of bacterial response regulators: diversity of output domains and domain combinations. J. Bacteriol. *188*, 4169–4182.

Gibrat, J.F., Madej, T., and Bryant, S.H. (1996). Surprising similarities in structure comparison. Curr. Opin. Struct. Biol. *6*, 377–385.

Holm, L., and Sander, C. (1995). Dali: a network tool for protein structure comparison. Trends Biochem. Sci. *20*, 478–480.

Kato, J., and Chakrabarty, A.M. (1991). Purification of the regulatory protein AlgR1 and its binding in the far upstream region of the *algD* promoter in *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. USA *88*, 1760–1764.

Knutsen, E., Ween, O., and Havarstein, L.S. (2004). Two separate quorum-sensing systems upregulate transcription of the same ABC transporter in *Streptococcus pneumoniae*. J. Bacteriol. *186*, 3078–3085.

Koenig, R.L., Ray, J.L., Maleki, S.J., Smeltzer, M.S., and Hurlburt, B.K. (2004). *Staphylococcus aureus* AgrA binding to the RNAIII-*agr* regulatory region. J. Bacteriol. *186*, 7549–7555.

Laskowski, R.A., McArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. *26*, 282–291.

Lin, Y.C., Liu, G., Shen, Y., Bertonati, C., Yee, A., Honig, B., Arrowsmith, C.H., and Szyperski, T. (2006). NMR structure of protein PA2021 from *Pseudomonas aeruginosa*. Proteins *65*, 767–770.

Lizewski, S.E., Lundberg, D.S., and Schurr, M.J. (2002). The transcriptional regulator AlgR is essential for *Pseudomonas aeruginosa* pathogenesis. Infect. Immun. *70*, 6083–6093.

Luscombe, N.M., and Thornton, J.M. (2002). Protein-DNA interactions: amino acid conservation and the effects of mutations on binding specificity. J. Mol. Biol. *320*, 991–1009.

Maiti, R., Van Domselaar, G.H., Zhang, H., and Wishart, D.S. (2004). SuperPose: a simple server for sophisticated structural superposition. Nucleic Acids Res. *32*, W590–W594.

McCrary, B.S., Edmondson, S.P., and Shriver, J.W. (1996). Hyperthermophile protein folding thermodynamics: differential scanning calorimetry and chemical denaturation of Sac7d. J. Mol. Biol. *264*, 784–805.

McGowan, S., Lucet, I.S., Cheung, J.K., Awad, M.M., Whisstock, J.C., and Rood, J.I. (2002). The FxRxHrS motif: a conserved region essential for DNA binding of the VirR response regulator from *Clostridium perfringens*. J. Mol. Biol. *322*, 997–1011.

McGowan, S., O'Connor, J.R., Cheung, J.K., and Rood, J.I. (2003). The SKHR motif is required for biological function of the VirR response regulator from *Clostridium perfringens*. J. Bacteriol. *185*, 6205–6208.

Mohr, C.D., Martin, D.W., Konyecsni, W.M., Govan, J.R., Lory, S., and Deretic, V. (1990). Role of the far-upstream sites of the *algD* promoter and the *algR* and *rpoN* genes in environmental modulation of mucoidy in *Pseudomonas aeruginosa*. J. Bacteriol. *172*, 6576–6580.

Mohr, C.D., Hibler, N.S., and Deretic, V. (1991). AlgR, a response regulator controlling mucoidy in *Pseudomonas aeruginosa*, binds to the FUS sites of the *algD* promoter located unusually far upstream from the mRNA start site. J. Bacteriol. *173*, 5136–5143.

Mohr, C.D., Leveau, J.H., Krieg, D.P., Hibler, N.S., and Deretic, V. (1992). AlgR-binding sites within the *algD* promoter make up a set of inverted repeats separated by a large intervening segment of DNA. J. Bacteriol. *174*, 6624–6633.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. *53*, 240–255.

Nikolskaya, A.N., and Galperin, M.Y. (2002). A novel type of conserved DNA-binding domain in the transcriptional regulators of the AlgR/AgrA/LytR family. Nucleic Acids Res. *30*, 2453–2459.

Novick, R.P., Projan, S.J., Kornblum, J., Ross, H.F., Ji, G., Kreiswirth, B., Vandenesch, F., and Moghazeh, S. (1995). The *agr* P2 operon: an autocatalytic sensory transduction system in *Staphylococcus aureus*. Mol. Gen. Genet. *248*, 446–458.

Novick, R.P. (2003). Autoinduction and signal transduction in the regulation of staphylococcal virulence. Mol. Microbiol. *48*, 1429–1449.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *276*, 307–326.

Peters, W.B., Edmondson, S.P., and Shriver, J.W. (2004). Thermodynamics of DNA binding and distortion by the hyperthermophile chromatin protein Sac7d. J. Mol. Biol. *343*, 339–360.

Risøen, P.A., Håvarstein, L.S., Diep, D.B., and Nes, I.F. (1998). Identification of the DNA-binding sites for two response regulators involved in control of bacteriocin synthesis in *Lactobacillus plantarum* C11. Mol. Gen. Genet. *259*, 224–232.

Risøen, P.A., Johnsborg, O., Diep, D.B., Hamoen, L., Venema, G., and Nes, I.F. (2001). Regulation of bacteriocin production in *Lactobacillus plantarum* depends on a conserved promoter arrangement with consensus binding sequence. Mol. Genet. Genomics *265*, 198–206.

Robinson, H., Gao, Y.G., McCrary, B.S., Edmondson, S.P., Shriver, J.W., and Wang, A.H. (1998). The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. Nature *392*, 202–205.

Rood, J.I. (1998). Virulence genes of *Clostridium perfringens*. Annu. Rev. Microbiol. *52*, 333–360.

Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. Acta Crystallogr. D Biol. Crystallogr. *58*, 1772–1779.

Shimizu, T., Shima, K., Yoshino, K., Yonezawa, K., Shimizu, T., and Hayashi, H. (2002). Proteome and transcriptome analysis of the virulence genes regulated by the VirR/VirS system in *Clostridium perfringens*. J. Bacteriol. *184*, 2587–2594.

Straume, D., Axelsson, L., Nes, I.F., and Diep, D.B. (2006). Improved expression and purification of the correctly folded response regulator PlnC from lactobacilli. J. Microbiol. Methods *67*, 193–201.

Traber, K., and Novick, R. (2006). A slipped-mispairing mutation in AgrA of laboratory strains and clinical isolates results in delayed activation of *agr* and failure to translate delta- and alpha-haemolysins. Mol. Microbiol. *59*, 1519–1530.

Ween, O., Gaustad, P., and Havarstein, L.S. (1999). Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. Mol. Microbiol. *33*, 817–827.