Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

A high-resolution structure of the DNA-binding domain of AhrC, the arginine repressor/activator protein from *Bacillus subtilis*

James A. Garnett,‡ Simon Baumberg,† Peter G. Stockley and Simon E. V. Phillips*

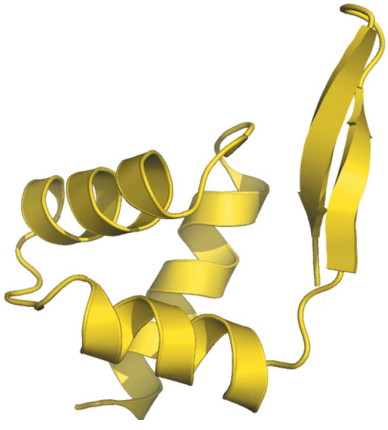
Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, England

‡ Current address: Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, England.
† Deceased 11th April 2007.

Correspondence e-mail: s.e.v.phillips@leeds.ac.uk

Received 29 August 2007
Accepted 1 October 2007

PDB Reference: NAhrC, 2p5k, r2p5ksf.



In *Bacillus subtilis* the concentration of L-arginine is controlled by the transcriptional regulator AhrC, which interacts with 18 bp DNA operator sites called ARG boxes in the promoters of arginine biosynthetic and catabolic operons. AhrC is a 100 kDa homohexamer, with each subunit having two domains. The C-terminal domains form the core, mediating intersubunit interactions and binding of the co-repressor L-arginine, whilst the N-terminal domains contain a winged helix–turn–helix DNA-binding motif and are arranged around the periphery. The N-terminal domain of AhrC has been expressed, purified and characterized and it has been shown that the fragment still binds DNA operators as a recombinant monomer. The DNA-binding domain has also been crystallized and the crystal structure refined to 1.0 Å resolution is presented.

1. Introduction

Regulation of the genes encoding arginine-metabolizing enzymes in many bacteria arises primarily *via* a conserved transcription factor called the arginine repressor (ArgR in *Escherichia coli*; Mountain & Baumberg, 1980; Mountain *et al.*, 1984; Lim *et al.*, 1987; North *et al.*, 1989). The *Bacillus subtilis* homologue was previously identified in a mutant strain resistant to arginine hydroxamate (Harwood & Baumberg, 1977) and thus termed the arginine hydroxamate resistant mutant C (AhrC). AhrC regulates the expression of biosynthetic genes, which are clustered into two operons that start with *argC* and *argG* (Smith *et al.*, 1989; Miller, 1996; Miller *et al.*, 1997). Additionally, AhrC functions as an activator of arginine catabolic genes (Klingel *et al.*, 1995; Gardan *et al.*, 1995; Miller, 1996).

AhrC forms hexamers of identical 149-amino-acid subunits, with multimerization being controlled by protein concentration and binding of the co-effector L-arginine (Czaplewski *et al.*, 1992). Proteolysis and sequence analysis established the presence of two domains in each subunit separated by an accessible linker (Czaplewski *et al.*, 1992). The N-terminal domains are involved in DNA binding, whilst the C-terminal domains control multimerization and L-arginine binding (Tian & Maas, 1994; Karaivanova *et al.*, 1999; Song *et al.*, 2002). The crystal structure of apo-AhrC was solved at 2.7 Å (Dennis *et al.*, 2002) and shows that the C-terminal domains create a compact hexameric core with strict 32 symmetry, whilst the DNA-binding domains lie around the periphery as loosely associated pairs.

Arginine repressors bind to operator sites containing imperfect 18 bp palindromes called ARG boxes (Cunin *et al.*, 1983; Lim *et al.*, 1987; Miller *et al.*, 1997) and the consensus sequence for *B. subtilis* operators is 5′-CATGAATAAAAATg/tCAAg/t-3′. The *argC* biosynthetic sites have two tandem repeats of ARG-box sites positioned 5′ to the transcript start separated by 11 bp; in addition, there is a single downstream ARG box within the coding region (Czaplewski *et al.*, 1992; Miller *et al.*, 1997). Single ARG-box sites are present at the *Bacillus* catabolic genes directly adjacent to the transcriptional start site (Miller *et al.*, 1997).

Here, we report the crystal structure of the free N-terminal DNA-binding domain of AhrC (NAhrC) at atomic resolution. We also discuss the structural comparisons with intact apo-AhrC (Dennis *et*

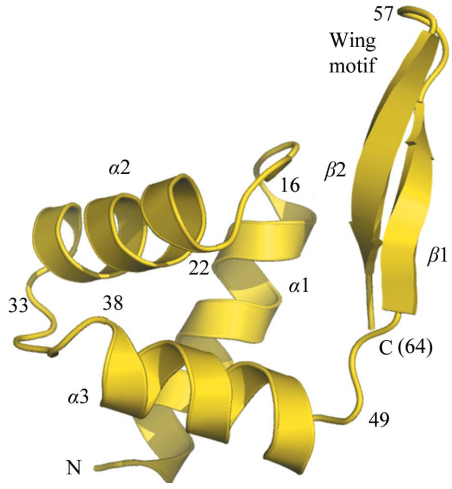


Figure 1
Secondary structure of the N-terminal DNA-binding domain of AhrC (NAhrC), with helices and sheets labelled and some residues numbered.

al., 2002) and arginine repressors from *E. coli* (Van Duyne *et al.*, 1996) and *B. stearothermophilus* (Ni *et al.*, 1999).

2. Materials and methods

2.1. Cloning and expression

The apo-AhrC crystal structure (Dennis *et al.*, 2002) was used to define the N-terminal domain of AhrC (residues 1–64; NAhrC). A construct encoding this domain was obtained by PCR from the *ahrC* gene with primers containing *Nde*I and *Bam*HI restriction sites, ligated into the vector pET22b and transformed into *E. coli* strain BL21 (DE3). Cells grown at 310 K in LB media were induced for 3 h with isopropyl β-D-1-thiogalactopyranoside at an $OD_{600\,nm}$ of 0.6.

2.2. Protein purification and crystallization

Cells were lysed by sonication in 20 m*M* Tris–HCl pH 7.5, 10 m*M* $MgCl_2$, 1 m*M* PMSF, 25 μ*M* TPCK and 1% 2-propanol, with an additional 6 μg ml^{-1} lysozyme. DNaseI was added to the supernatant at 1 μg ml^{-1} and a 30%(*w*/*v*) ammonium sulfate precipitation was performed. The supernatant was dialysed against lysis buffer, loaded onto an SP Sepharose High Performance XK 16 column (GE Healthcare) and equilibrated and washed with the same buffer. NAhrC eluted between 0.35 and 0.5 *M* NaCl as a single 7.4 kDa band on SDS–PAGE.

Crystallization experiments were carried out by hanging-drop vapour diffusion at 277 K. Crystals of NAhrC were obtained after ~7 d by mixing 2 µl NAhrC (2 mg ml^{-1}) in 20 m*M* Tris–HCl pH 7.5, 100 m*M* NaCl, 10 m*M* $MgCl_2$, 1 m*M* PMSF, 25 μ*M* TPCK and 1% 2-propanol with an equal volume of well solution [100 m*M* sodium cacodylate pH 6.5, 100 m*M* sodium acetate, 24%(*w*/*v*) PEG 8K].

2.3. Data collection, structure determination and refinement

Prior to data collection, crystals of NAhrC were soaked in cryoprotectant [well solution mixed in a 1:1 ratio with 30%(*w*/*v*) PEG 8K] for 60 s and then frozen immediately. Data were collected on station 14.1 at Daresbury Synchrotron Radiation Source (SRS) and on station ID14.4 at the European Synchrotron Radiation Facility (ESRF). All data were processed using *MOSFLM* (Leslie, 1992) and scaled using *SCALA* (Evans, 1993).

Phases were obtained by molecular replacement using *AMoRe* (Navaza, 1994). The AhrC coordinates (Dennis *et al.*, 2002) of residues 3–64 from chain *A* (PDB code 1f9n) were used as a trial model for the NAhrC data and the top solution after rigid-body refinement resulted in R = 0.42% and CC = 68.4%. Approximately 10% of the data were used to calculate R_{free}. Initial refinement was performed using *CNS* (Brünger *et al.*, 1998), with model building in *O* (Jones *et al.*, 1991). H atoms were added to NAhrC and a restrained refinement with individual anisotropic *B* factors was performed with *REFMAC*5 (Murshudov *et al.*, 1999); further model building was carried out with *Coot* (Emsley & Cowtan, 2004). The model shows good geometry as evaluated by *PROCHECK* (Laskowski *et al.*, 1993), with all residues in the allowed regions of the Ramachandran plot. The data-collection and refinement statistics are given in Table 1. Figures were generated with *PyMOL* (DeLano, 2002).

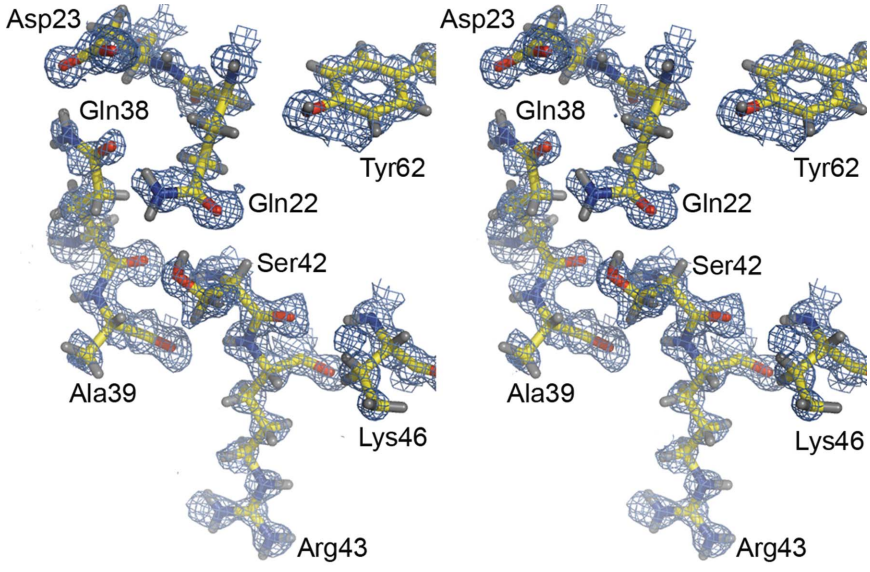


Figure 2
Stereoview of the suspected DNA-binding face of the NAhrC monomer: $2|F_o| - |F_c|$ map contoured at 1.5 r.m.s.

Table 1
Summary of data collection and refinement.

Values in parentheses are for the outermost resolution shell.

Data collection	
Beamline	ESRF ID14.4
Wavelength (Å)	0.939
Space group	$P2_12_12_1$
Unit-cell parameters (Å, °)	a = 25.11, b = 32.90, c = 66.23, α = 90, β = 90, γ = 90
Resolution (Å)	19.96–1.00 (1.05–1.00)
Unique observations	26874
R_{sym}	0.094 (0.216)
$\langle I/\sigma(I)\rangle$	12.8 (2.8)
Completeness (%)	89.4 (70.6)
Redundancy	4.2 (2.4)
Refinement	
R_{work}/R_{free} (%)	13.7/16.0
No. of residues	63
No. of water molecules	84
R.m.s. deviations from ideal†	
Angles (°)	0.016
Bonds (Å)	1.60
Ramachandran plot†	
Most favoured (%)	96.5
Allowed (%)	3.5
Disallowed (%)	—

† Determined using *PROCHECK* (Laskowski *et al.*, 1993).

3. Results and discussion

3.1. Characterization

Purified NAhrC was analyzed by mass spectrometry and shown to have a molecular weight of 7417 Da (expected molecular weight 7419 Da; Garnett, 2005). A 50 bp DNA fragment derived from the *B. subtilis argC* promoter [containing a single 18 bp ARG box (5′-CATGAATAAAAATTCAAG-3′) based on the *B. subtilis* consensus sequence; Miller *et al.*, 1997] and a randomized 50 bp DNA fragment were attached to a BIAcore SA chip *via* biotinylation of the 5′-sense strand of the DNA duplexes (Garnett, 2005). The DNA-binding properties of NAhrC were then assayed using surface plasmon resonance (SPR; Stockley *et al.*, 1998). The monomeric domain bound the ARG-box fragment with an apparent K_d of 21.3 μ*M* and the randomized fragment with an apparent K_d of 1 m*M*. This compares with a value of 1.4 n*M* for holo-AhrC binding to a 276 bp *rocA* fragment encompassing a natural single ARG box (Stockley *et al.*, 1998).

3.2. Overall structure

The crystal structure of NAhrC was solved using molecular replacement and refined to R = 13.7% and R_{free} = 16.0% at 1.0 Å resolution (Table 1). The crystal contains one monomer in the asymmetric unit and the electron-density maps, which were calculated with H atoms and anisotropic B factors, reveal high atomicity. Three α-helices pack against a two-stranded antiparallel β-sheet, creating a compact winged helix–turn–helix (wHTH) DNA-binding motif (Brennan, 1993; Fig. 1). Electron density for the α-helices is very good (Fig. 2), although some other regions of the structure are disordered and Met1 and the side chains of Gln5, Lys46 and Asn56 were left out of the final model. The wing displays the poorest electron density and this region has the highest average B factors, suggesting that this β-wing is highly mobile, which is consistent with the composite NMR structure from *E. coli* ArgR (Fig. 3*a*; Sunnerhagen *et al.*, 1997). Ordered solvent peaks were all assigned as water molecules and none appeared to correspond to bound ions.

The N-terminal domains from intact apo-AhrC (Dennis *et al.*, 2002) and intact *B. stearothermophilus* apo-ArgR (Ni *et al.*, 1999) superimpose very closely with NAhrC, with an average r.m.s.d. for C^{α} atoms of 0.65 Å for apo-AhrC and 0.58 Å for apo-ArgR (Fig. 3*b*). The major deviation within these structures is again observed within the wing region, with apo-AhrC displaying the largest deviations and high B factors.

4. Conclusions

A mutational study on residues within the N-terminal domain of *B. stearothermophilus* ArgR has demonstrated the importance of conserved residues on α3 (Gln38, Ser42 and Arg43), which is believed

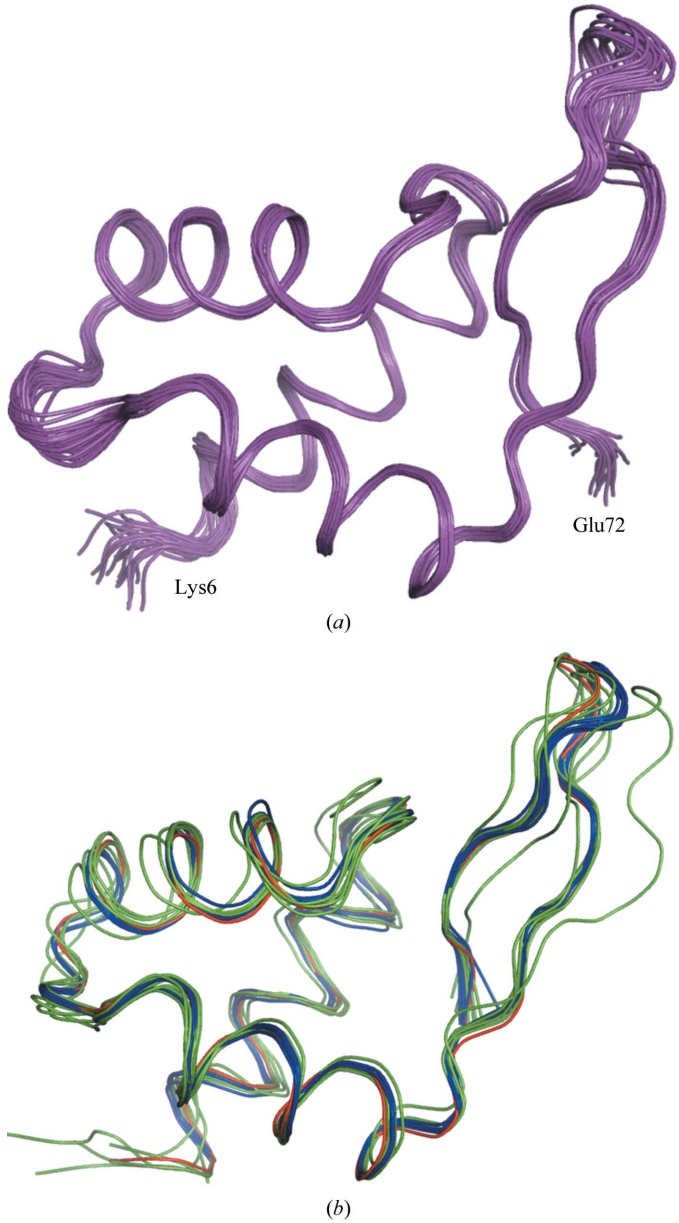


Figure 3
Structural comparison of the β-wing motif between homologous arginine repressors. (*a*) Composite NMR structure of *E. coli* ArgR N-terminal domain (Sunnerhagen *et al.*, 1997; PDB 1aoy), highlighting the flexible wing motif. The N- and C-termini (residues 1–5 and 73–78) are highly mobile and are not displayed. (*b*) Superimposition of the six N-terminal domains from *B. subtilis* apo-AhrC (Dennis *et al.*, 2002; PDB code 1f9n; green), the single *B. subtilis* NAhrC domain (red) and the six N-terminal domains from *B. stearothermophilus* ArgR (Ni *et al.*, 1999; PDB code 1b4a; blue). The highest r.m.s.d. values are found in the β-wing.

to be the recognition helix (Karaivanova *et al.*, 1999). A Gln22Arg mutation (on the α2 helix) also inhibited the capacity of the protein to bind DNA. In the NAhrC structure, Gln22 forms a network of hydrogen bonds with Gln38 and Ser42 and it has been suggested that this will make up the specific repressor–DNA interface (Ni *et al.*, 1999).

It is anticipated that upon binding of the co-repressor L-arginine and the subsequent conformational change within the C-terminal domains (Van Duyne *et al.*, 1996; Ni *et al.*, 1999; Jin *et al.*, 2004), the DNA-binding domains will come closer together allowing them to form dimers, where each monomer will then be able to occupy one half site of the ARG box. The β1 sheet is almost exclusively hydrophobic and is a good candidate for the dimer interface. In the crystal and NMR structures discussed here, it seems a common theme that the β-wing has dynamic properties whilst in a monomeric state. On co-repressor activation, dimerization could stabilize this flexible topology and, with the correct positioning of the recognition helices, allow Gln38, Ser42 and Arg43 to interact with adjacent major grooves of the ARG-box DNA, whilst the intervening minor groove is occupied by the stable β-wings.

We thank the SRS and the ESRF for providing synchrotron radiation facilities and the beamline scientists at stations 14.1 and ID-14-4 for their help with data collection. We thank Dr M. Parsons for correspondence, Dr A. Ashcroft for performing mass spectrometry and Dr A. Baron for interpretation of SPR data. This work was funded by the BBSRC.

References

Brennan, R. G. (1993). *Cell*, **74**, 773–776.
Brünger, 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. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Cunin, R., Eckhardt, T., Piette, J., Boyen, A., Piérard, A. & Glansdorff, N. (1983). *Nucleic Acids Res.* **11**, 5007–5019.
Czaplewski, L. G., North, A. K., Smith, M. C. M., Baumberg, S. & Stockley, P. G. (1992). *Mol. Microbiol.* **6**, 267–275.
DeLano, W. L. (2002). *The PyMol User's Manual.* DeLano Scientific, San Carlos, California, USA.
Dennis, C. A., Glykos, N. M., Parsons, M. R. & Phillips, S. E. V. (2002). *Acta Cryst.* D**58**, 421–430.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.
Gardan, R., Rapoport, G. & Débarbouillé, M. (1995). *J. Mol. Biol.* **249**, 843–856.
Garnett, J. A. (2005). PhD thesis. Department of Biochemistry and Molecular Biology, University of Leeds, England.
Harwood, C. R. & Baumberg, S. (1977). *J. Gen. Microbiol.* **100**, 177–188.
Jin, L., Xue, W.-F., Fukayama, J. W., Yetter, J., Pickering, M. & Carey, J. (2004). *J. Mol. Biol.* **346**, 43–56.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Karaivanova, I. M., Weigel, P., Takahashi, M., Fort, C., Versavaud, A., Van Duyne, G., Charlier, D., Hallet, J. N., Glansdorff, N. & Sakanyan, V. (1999). *J. Mol. Biol.* **291**, 843–845.
Klingel, U., Miller, C. M., North, A. K., Stockley, P. G. & Baumberg, S. (1995). *Mol. Gen. Genet.* **248**, 329–340.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Lim, D., Oppenheim, J. D., Eckhardt, T. & Maas, W. K. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 6697–6701.
Miller, C. M. (1996). PhD thesis. Department of Biology, University of Leeds, England.
Miller, C. M., Baumberg, S. & Stockley, P. G. (1997). *Mol. Microbiol.* **26**, 37–48.
Mountain, A. & Baumberg, S. (1980). *Mol. Gen. Genet.* **178**, 691–701.
Mountain, A., Mann, N. H., Manton, R. N. & Baumberg, S. (1984). *Mol. Gen. Genet.* **197**, 82–89.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Ni, J., Sakanyan, V., Charlier, D., Glansdorff, N. & Van Duyne, G. D. (1999). *Nature Struct. Biol.* **6**, 427–432.
North, A. K., Smith, M. C. M. & Baumberg, S. (1989). *Gene*, **80**, 29–38.
Smith, M. C. M., Czaplewski, L., North, A. K., Baumberg, S. & Stockley, P. G. (1989). *Mol. Microbiol.* **3**, 23–28.
Song, H., Wang, H., Gigot, D., Dimova, D., Sakanyan, V., Glansdorff, N. & Charlier, D. (2002). *J. Mol. Biol.* **315**, 255–274.
Stockley, P. G., Baron, A. J., Wild, C. M., Parsons, I. D., Miller, C. M., Holtham, C. A. M. & Baumberg, S. (1998). *Biosens. Bioelectron.* **13**, 637–650.
Sunnerhagen, M., Nilges, M., Otting, G. & Carey, J. (1997). *Nature Struct. Biol.* **4**, 819–826.
Tian, G. & Maas, W. K. (1994). *Mol. Microbiol.* **4**, 599–608.
Van Duyne, G. D., Ghosh, G., Maas, W. K. & Sigler, P. B. (1996). *J. Mol. Biol.* **256**, 377–391.