

Acta Crystallographica Section D

**Biological
Crystallography**

ISSN 0907-4449

The structure of AhrC, the arginine repressor/activator protein from *Bacillus subtilis*

Caitríona A. Dennis, Nicholas M. Glykos, Mark R. Parsons and Simon E. V. Phillips

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site provided that this cover page is retained. Republication of this article or its storage in electronic databases or the like is not permitted without prior permission in writing from the IUCr.

The structure of AhrC, the arginine repressor/activator protein from *Bacillus subtilis*

Caitríona A. Dennis, Nicholas M. Glykos,[†] Mark R. Parsons and Simon E. V. Phillips*

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

[†] Current address: Institute of Molecular Biology
and Biotechnology, Foundation for Research
and Technology Hellas, PO Box 1527,
71110 Heraklion, Crete, Greece.

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

In the Gram-positive bacterium *Bacillus subtilis* the concentration of the amino acid L-arginine is controlled by the transcriptional regulator AhrC. The hexameric AhrC protein binds in an L-arginine-dependent manner to pseudo-palindromic operators within the promoter regions of arginine biosynthetic and catabolic gene clusters. AhrC binding results in the repression of transcription of biosynthetic genes and in the activation of transcription of catabolic genes. The crystal structure of AhrC has been determined at 2.7 Å resolution. Each subunit of the protein has two domains. The C-terminal domains are arranged with 32 point-group symmetry and mediate the major intersubunit interactions. The N-terminal domains are located around this core, where they lie in weakly associated pairs but do not obey strict symmetry. A structural comparison of AhrC with the arginine repressor from the thermophile *B. stearothermophilus* reveals close similarity in regions implicated in L-arginine binding and DNA recognition, but also reveals some striking sequence differences, especially within the C-terminal oligomerization domain, which may contribute to the different thermostabilities of the proteins. Comparison of the crystal structure of AhrC with a 30 Å resolution model obtained by combining X-ray structure-factor amplitudes with phases derived from electron-microscopic analyses of AhrC crystals confirms the essential accuracy of the earlier model and suggests that such an approach may be more widely useful for obtaining low-resolution phase information.

Received 19 June 2001

Accepted 18 December 2001

PDB Reference: AhrC, 1f9n,
r1f9nsf.

1. Introduction

Transcriptional control of arginine pathways in *B. subtilis* is exerted through a repression/activation function carried out by a single DNA-binding protein, AhrC, the product of the *ahrC* gene. AhrC was first characterized in mutants resistant to the arginine analogue arginine hydroxamate (Harwood & Baumberg, 1977; Mountain & Baumberg, 1980; Smith *et al.*, 1986*a,b*). Mutants mapping to *ahrC*, one of four arginine hydroxamate resistant loci, showed simultaneous loss of repression of the biosynthetic genes and activation of the catabolic genes, suggesting the *ahrC* gene product was a direct regulatory link between the pathways.

Biosynthesis of arginine within *B. subtilis* occurs in a pathway containing seven steps and all of the genes encoding the enzymes in the pathway are located in two clusters within the genome (*argCJBD-cpa-F* and *argGH*). In the presence of L-arginine, AhrC interacts with operators within the promoter regions of *argC* and *argG* and represses transcription of these gene clusters (Smith *et al.*, 1986*a*; Miller, 1997). These operators contain DNA sequences which are similar to the 18 bp

pseudo-palindromic ARG boxes which are the binding sites for the *E. coli* arginine repressor ArgR (ArgREc; Glansdorff, 1987; Smith *et al.*, 1989). Examination of the *argC* promoter and gene using DNase I and hydroxyl radical footprinting revealed two AhrC-binding sites, which were named *argC*₀₁ and *argC*₀₂. The higher affinity binding site, *argC*₀₁, contains two ARG boxes separated by 11 bp and lies within the *argC* promoter, whilst the lower affinity site, *argC*₀₂, contains a single ARG box and is located within the *argC* structural gene (Czaplewski *et al.*, 1992). The *argG* promoter contains two ARG boxes, separated by 2 bp, upstream of the transcription start site (Miller, 1997).

The *B. subtilis* arginine catabolic pathway contains six enzymes encoded by genes within the two clusters *rocABC* and *rocDEF*. AhrC has been shown to interact in an L-arginine-dependent manner with operators within the promoter regions of *rocA* and *rocD*. Each of these operators consists of a single ARG box which is located directly adjacent to the transcription start site (Calogero *et al.*, 1994; Klingel *et al.*, 1995; Gardan *et al.*, 1995; Miller *et al.*, 1997). The affinity of AhrC for the catabolic operators is 10- to 20-fold less than for the biosynthetic gene promoters (Miller *et al.*, 1997), a finding consistent with the notion of cooperative binding of AhrC to the tandem repeats of ARG boxes within the upstream regulatory regions of *argC* and *argG*. The mechanism by which AhrC activates transcription from *rocA* and *rocD* remains unclear, although AhrC binding has been shown to increase the natural bend of the *rocA* promoter (Miller *et al.*, 1997), which may facilitate interactions with RNA polymerase.

Arginine-regulatory proteins, which are usually called ArgR in organisms other than *B. subtilis*, have been identified in *E. coli* (Lim *et al.*, 1987), *B. stearothermophilus* (Dion *et al.*, 1997) and *Salmonella typhimurium* (Lu *et al.*, 1992). These proteins have been biochemically characterized and shown to act as repressors of arginine biosynthesis in their respective hosts, although any roles in the activation of arginine catabolism remain to be confirmed experimentally. Sequences are also known for probable AhrC/ArgR homologues from *Clostridium perfringens* (Ohtani *et al.*, 1997), *Haemophilus influenzae* (Fleischmann *et al.*, 1995), *Mycobacterium tuberculosis* (Cole *et al.*, 1998), *Streptomyces clavuligerus* (Rodriguez-Garcia *et al.*, 1997) and *Streptococcus pneumoniae* (Priebe *et al.*, 1988).

The best characterized of the AhrC homologues is ArgR of *E. coli* (ArgREc; Lim *et al.*, 1987; Maas, 1994). AhrC and ArgREc share 27% identity (North *et al.*, 1989) and can cross-function to some extent *in vivo*. AhrC can repress *E. coli* arginine genes and complement for ArgR as an essential accessory protein in the resolution of plasmid ColE1 multimers (Stirling *et al.*, 1988); however, ArgR cannot repress the *B. subtilis argC* promoter (Smith *et al.*, 1989).

AhrC and ArgR both require the binding of L-arginine for high-affinity binding to their operators and in the presence of L-arginine both proteins exist as hexamers (Czaplewski *et al.*, 1992; Lim *et al.*, 1987). In the absence of L-arginine, hexamers are in rapid equilibrium with trimers (Holtham *et al.*, 1999). The position of equilibrium seems to differ for different

homologues, with trimers being dominant for *B. stearothermophilus* ArgR (ArgRBst; Dion *et al.*, 1997; Ni *et al.*, 1999), but hexamers favoured for AhrC (Czaplewski *et al.*, 1992), ArgREc (Lim *et al.*, 1987) and *S. typhimurium* ArgR (Lu *et al.*, 1992).

The sensitivity of AhrC to proteolytic digestion, yielding two fragments of similar size (Czaplewski *et al.*, 1992), suggested that the 16.7 kDa protein subunit had a two-domain structure. Mutational analysis of the *argR* gene (Tian & Maas, 1994) identified the roles of the N-terminal domain in DNA binding and of the C-terminal domain in oligomerization and arginine binding. These findings were supported by structural studies of the isolated domains of ArgREc. The crystal structure of the C-terminal domain revealed it to form a compact hexamer with 32 symmetry and identified the binding sites for six L-arginine molecules (Van Duyne *et al.*, 1996), whilst NMR studies of the N-terminal 78-residue portion (Sunnerhagen *et al.*, 1997) showed it to be monomeric and to have structural homology to the winged helix–turn–helix family of DNA-binding domains (Brennan, 1993).

More recently, the crystal structure of an intact arginine repressor, ArgR from *B. stearothermophilus* (ArgRBst), has been determined (Ni *et al.*, 1999). As anticipated, the structure reveals that the protein oligomerizes through its C-terminal domain and that the hexameric core has 32 symmetry. The DNA-binding domains are located surrounding the core and do not obey strict 32 symmetry. In the structure of intact ArgRBst, the protein is without bound arginine, but comparison with the structure of the arginine-bound ArgRBst core suggests that arginine binding induces a tightening of the trimer–trimer interface through the exclusion of water molecules and a rotation of one trimer by about 15° with respect to the other (Ni *et al.*, 1999). Model-building studies suggest that this rotation may move the DNA-binding domains into positions that are more favourable for interaction with DNA.

AhrC and ArgRBst share 72% identity throughout their sequence. Although there has been no direct evidence that ArgRBst can function as an activator of catabolic genes owing to its sequence similarity to AhrC and the retention of conservation with the DNA-binding region, especially the residues reported to be involved in contacting the ARG box (Miltcheva Karaivanova *et al.*, 1999), it seems likely that an activation function occurs. The main differences between the two protein sequences reside in the C-terminal domain within the area responsible for multimerization. As *B. stearothermophilus* is a thermophile, it is expected that in order to provide additional thermal stability intersubunit interactions in ArgRBst would be more substantial than those required by AhrC from the mesophile *B. subtilis*.

Here, we report the crystal structure of apo-AhrC solved to a resolution of 2.7 Å. We compare this structure with those of ArgRBst and of the two isolated domains of ArgREc. In particular, differences between AhrC and ArgRBst have been examined to try to identify the structural basis of the different thermal stabilities of the two proteins.

An earlier electron-microscopic study of crystals of AhrC had allowed the calculation of a low-resolution model of the

protein, which showed a number of clear lobes arranged around a central core (Glykos *et al.*, 1998). The structure presented here validates the results of this earlier analysis and suggests that such a use of electron microscopy to obtain low-resolution phases to be used in conjunction with X-ray amplitudes may be a valuable approach for determining limited structural information in cases where heavy-atom derivatives are not available.

2. Materials and methods

2.1. Purification, crystallization and data collection

The purification of AhrC from an overexpressing strain of *E. coli* was carried out using a protocol described previously (Holtham *et al.*, 1999). The crystallization conditions were modified from those previously published (Boys *et al.*, 1990). Crystals were grown using the hanging-drop method of vapour diffusion. Protein at 10 mg ml⁻¹ was mixed with equal volumes of well solution containing 60 mM ammonium sulfate, 100 mM phosphate buffer pH 4.9, 3% (w/v) PEG 4K. A cocktail of protease inhibitors was present throughout all experiments. Crystals were obtained after 7 d incubation at 291 K and belong to the orthorhombic space group *C222*₁, with unit-cell parameters $a = 230.31$, $b = 73.86$, $c = 138.60$ Å and one AhrC hexamer per asymmetric unit.

X-ray diffraction data were recorded at room temperature using a 30 cm MAR area detector at station 9.6 of the Synchrotron Radiation Source at the CLRC Daresbury Laboratory. Data were processed using *MOSFLM* (Leslie, 1992) and the *CCP4* program suite (Collaborative Computational Project, Number 4, 1994). The merged data set was 88% complete to 2.7 Å resolution with 1.6-fold multiplicity and had a merging *R* factor (R_{sym}) of 6.6%.

2.2. Structure determination

Initial phases were obtained from molecular replacement using *AMoRe* (Navaza, 1997). The structure of the core hexamer from *E. coli* (PDB code 1xxc) was used as a trial model. Using a resolution range of 15–5 Å in the rotation function, six peaks with a height of 5.5 were obtained, consistent with one hexamer in the asymmetric unit. Unfortunately, the translation function could not be solved using molecular replacement. However, using information about crystal packing obtained from electron microscopy (Glykos *et al.*, 1998), the translation vector was applied to the rotated model. This partial model was subject to low-resolution rigid-body refinement using *CNS* (Brünger *et al.*, 1998). Although the map from the core hexamer was of sufficient quality to allow rebuilding, there was no evidence for the DNA-binding domains. Phases from this partial model were then used to help solve a low-resolution (6 Å) data set from a heavy-atom derivative prepared by soaking crystals in a solution containing Nb₆Cl₁₂²⁺ ions (Glykos, 1995). This proved to contain a single cluster bound at the centre of the AhrC hexamer. Using combined model and heavy-atom phases, a low-resolution map was calculated and density for four of the

Table 1

Refinement statistics.

Values in parentheses are for the highest resolution shell (2.87–2.7 Å).

Resolution (Å)	30.0–2.7
No. of reflections	28055 (4440)
Completeness (%)	84.9 (85.7)
R_{work}	0.22 (0.27)
R_{free}	0.27 (0.33)
R.m.s. angles (°)	1.0
R.m.s. bonds (Å)	0.004
Average B (Å ²)	
Main-chain atoms (chains A–F)	72.2, 50.5, 74.5, 60.4, 43.9, 83.8
Side-chain atoms (chains A–F)	79.6, 56.4, 79.4, 66.2, 50.8, 88.3

DNA-binding domains could be identified. These domains were rebuilt in *O* (Jones *et al.*, 1991) using the NMR structure of the *E. coli* ArgR DNA-binding domain as a model. After several rounds of NCS averaging using *DM* (Collaborative Computational Project, Number 4, 1994), density appeared for the remaining two DNA-binding domains. The whole structure was subjected to refinement by *CNS*. Restrained NCS was applied to the core hexamer and the helices of the DNA-binding domains until the final stages of refinement. Statistics of refinement are shown in Table 1.

The stereochemical quality of the model was assessed with the program *PROCHECK* (Laskowski *et al.*, 1993). A Ramachandran plot indicated that 89.0% of amino-acid resi-

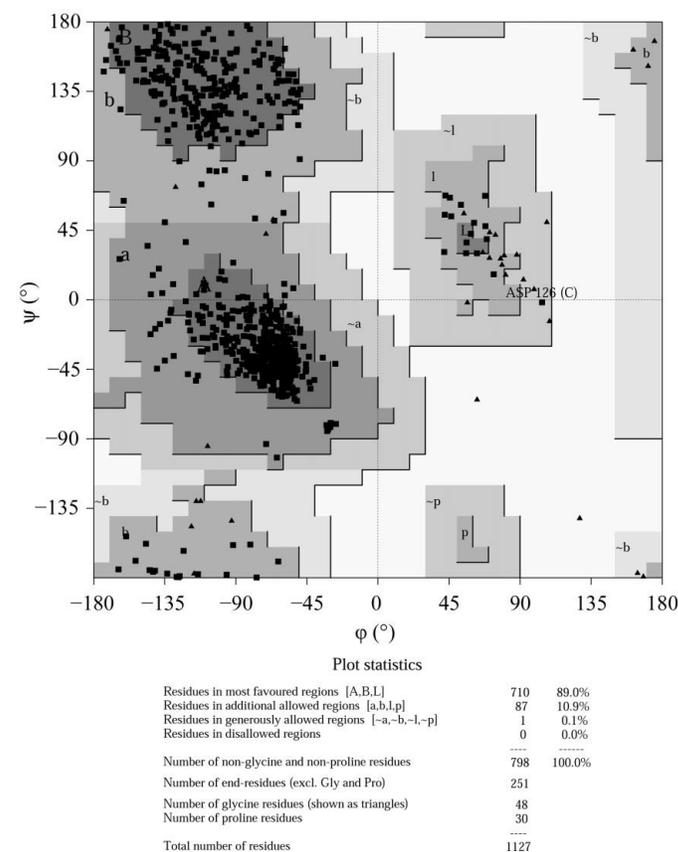


Figure 1

A Ramachandran plot for the final refined structure of AhrC.

dues were in most favoured regions, 10.9% were in additionally allowed regions and 0.15% were in generously allowed regions (Fig. 1). In comparisons of AhrC with ArgRBst and ArgREc, the protein structures were aligned using their C α atoms in *LSQKAB* and intermolecular and intramolecular contacts were determined using *ACT* (both from the Collaborative Computational Project, Number 4, 1994).

2.3. EM reconstruction

Using the combined EM/X-ray amplitudes and phases determined previously (Glykos *et al.*, 1998), an electron-density map was calculated using *FFT* (Collaborative Computational Project, Number 4, 1994). The electron-density map was converted to a mask at a threshold of 0.9 r.m.s.d. using *XDLMAPMAN* (Collaborative Computational Project, Number 4, 1994) and the resulting mask was edited with *O* (Jones *et al.*, 1991) to remove density from crystallographically related molecules. The edited mask was converted back to a *CCP4* map with *XDLMAPMAN*; the program *MAPMAN* from the *RAVE* package (Kleywegt & Jones, 1994) was then used to prepare a file of coordinates of dummy atoms at all grid points lying within the mask. Spherical harmonics up to and including $L = 8$ and the corresponding surface were calculated using *CRY SOL* and *TRANS3* (Svergun *et al.*, 1995). The same two programs were used to produce a surface from the atomic coordinates of the

refined AhrC structure and both surfaces were displayed using *ASSA* (Kozin *et al.*, 1997).

3. Results

3.1. The structure of the AhrC subunit

Crystals of AhrC contain one complete hexamer per asymmetric unit. Each subunit within the hexamer has a two-domain structure. The N-terminal domain, residues 1–65, comprises three α -helices and two β -strands which fold into a compact winged helix–turn–helix DNA-binding motif (Brennan, 1993) (Fig. 2). The wings in the motif are composed of two short β -strands and loop regions, which follow in sequence and lie adjacent to the presumed recognition helix (helix 3). The C-terminal domain (residues 72–149) has an α/β topology in which three α -helices are packed against a four-stranded antiparallel β -sheet (Fig. 2). Connecting the two domains is a linker region comprising residues 66–71, which consists of a region of random coil. This linker region seems to be inherently flexible, as it adopts a range of somewhat different conformations in the six subunits of AhrC. It is also the presumed site of cleavage by a number of proteolytic enzymes (Czaplewski *et al.*, 1992).

3.2. The structure of the AhrC hexamer

The monomers of AhrC associate *via* their C-terminal (core) domains to form a hexamer, which may be thought of as resulting from the face-to-face association of a pair of trimers. The six C-terminal domains obey strict 32 non-crystallographic symmetry (NCS), with domains from each trimer positioned above one another, giving the hexameric core a stacked configuration when viewed along the threefold axis (Fig. 3*a*). The r.m.s. deviation for all main-chain atoms within the core domains is 0.5 Å. Presumably as a result of the inherent flexibility of the interdomain linkers and of their different crystal packing environments, the DNA-binding domains (DBD) adopt slightly different positions around the periphery of the core and deviate from strict NCS (Fig. 3*a*). The structures of the five DBDs themselves are well conserved, with a main-chain r.m.s. deviation of 0.8 Å. As a result of interactions made with symmetry-related AhrC molecules in the crystal, the DBD of subunit *C* adopts a slightly different conformation and has an average r.m.s. deviation from the other DBDs of 1.8 Å.

Because of the lengthy linker between the domains, each DBD is not tethered adjacent to the core domain from the same subunit, but rather interacts with the core domain of a subunit from the opposite trimer of the hexamer (Fig. 3*b*). As a result

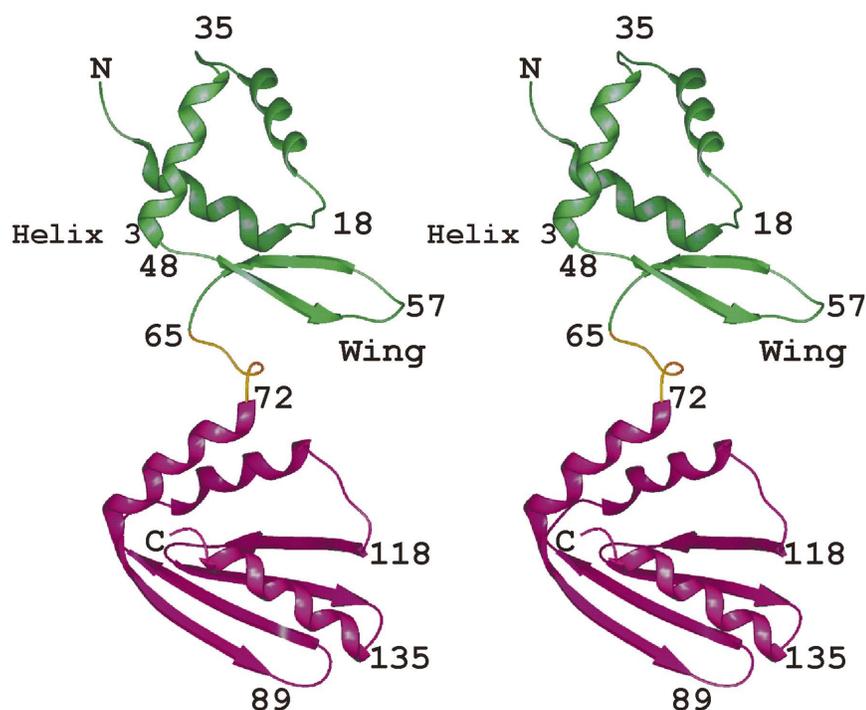
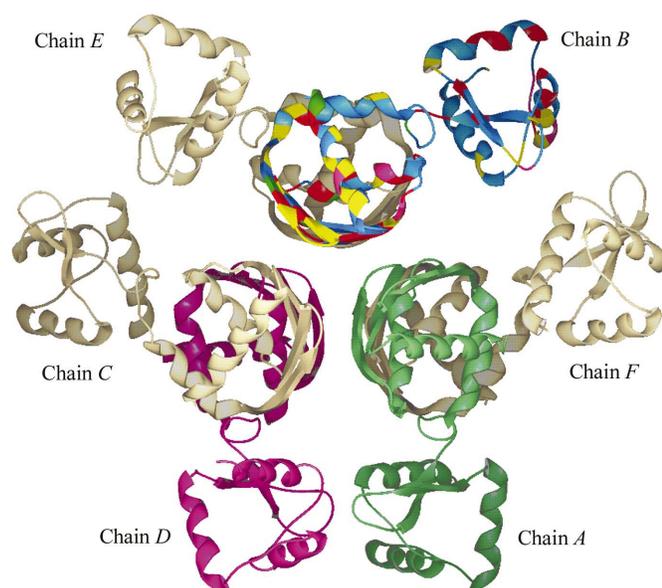


Figure 2

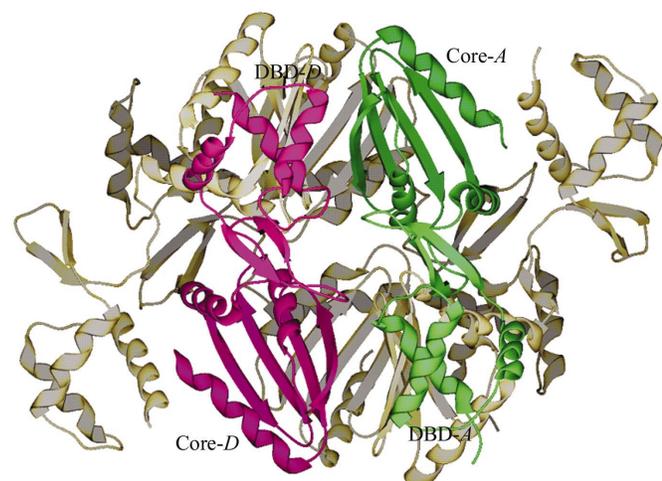
A stereoview of the AhrC monomer. The N-terminal DNA-binding domain (DBD) is coloured green, the C-terminal oligomerization domain is coloured magenta and the loop region which connects the domains is yellow. A number of residues are numbered and within the DBD the 'wing' and recognition helix are indicated. This figure was produced using *SPOCK* (Christopher, 1998).

of this crossing over, the three core domains of each trimer are surrounded by the three DBDs of the opposite trimer. The interactions between DBDs and core domains are rather few and differ somewhat for each subunit, as a result of the non-symmetric arrangement of the DBDs around the core. The principal common interaction observed in all six subunits occurs between Arg11 of the DBD and Met81 from the adjacent core domain. The lack of strong interactions between core and DNA-binding domains permits significant flexibility in the latter domains. The mobility of the domains is reflected in the high average atomic temperature factors (*B* factors) of these regions of the refined crystal structure (Table 1). Some individual *B* factors exceed 100 Å², especially in the loop regions attributed to the wing of the DNA-binding motif. This flexibility observed in the domain arrangement then gives support to the hypothesis of domain movement to provide a high-affinity DNA-binding site (Miller *et al.*, 1997).

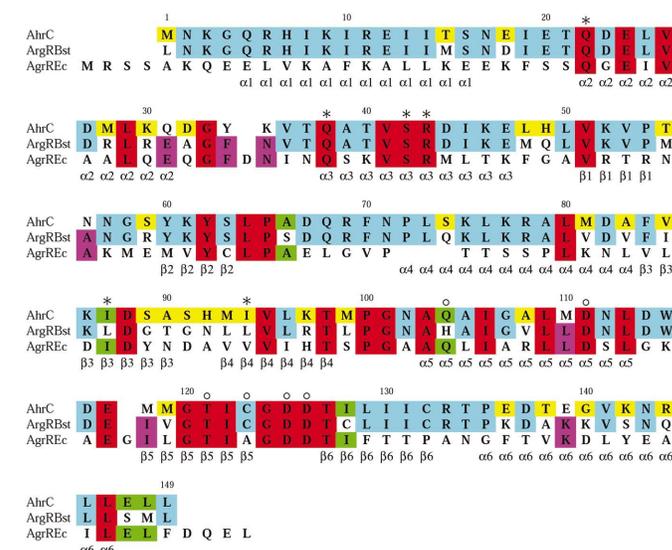
Examination of the arrangement of the DBDs around the core shows them to lie in pairs, which are presumed to be functionally significant for binding to approximately twofold-symmetric DNA sequences. Following the subunit labelling used in Fig. 3(*a*), the DBDs pair up as follows: *A* with *D*, *C* with *E* and *B* with *F*. One striking observation is that this pairing of DBDs is different from the pairing of C-terminal core domains which results from the association of trimers, as this aligns *A* with *F*, *B* with *E* and *C* with *D*. Residues making contacts between the adjacent DBDs are located in the loop region forming one of the 'wings' in the DNA-binding motif (residues 56 and 57) and they contact residues towards the end of the first helix (15 and 16). It is anticipated that upon L-arginine binding these pairs of DBDs form twofold-symmetric DNA-binding units in which helix 3 from each is involved in binding to the approximately palindromic DNA sequence of the ARG box.



(a)



(b)



- Residues conserved between all three species.
- Residues conserved between *B. subtilis* and *B. stearothermophilus* but not *E. coli*.
- Residues conserved between *B. subtilis* and *E. coli* but not *B. stearothermophilus*.
- Residues conserved between *B. stearothermophilus* and *E. coli* but not *B. subtilis*.
- Residues not conserved amongst the three species.
- * Residue mutated in the study of Miltcheva Karavanova *et al.* (1999).
- Residue implicated in the binding of L-arginine (Van Duyn *et al.*, 1996).

(c)

Figure 3

The structure of the AhrC hexamer. The association of the hexamer occurs *via* interactions involving the C-terminal domains; the N-terminal DNA-binding domains are arranged around the periphery, where they lie in pairs but do not obey strict symmetry. (a) The AhrC hexamer viewed along the molecular threefold axis. For clarity, two of the subunits (chains *A* and *D*) have been coloured (green and magenta, respectively) and the remainder are grey, apart from chain *B* which has been colour-coordinated with the sequence shown in (c). (b) The AhrC hexamer viewed perpendicular to the molecular threefold axis and along a molecular dyad. N-terminal (DBD) and C-terminal (core) domains of subunits *A* and *D* are labelled. This figure was produced using SPOCK. (c) An alignment of the amino-acid sequences of AhrC, ArgRBst and ArgREc. Residue colouring is according to conservation as described in the key. Elements of secondary structure are shown below the sequences and residue numbers (of AhrC) and the location of several significant residues are given above the sequences.

AhrC can be considered to have a hierarchical structure, with the hexamer being made up of two trimers. Solution studies have shown that in the absence of L-arginine AhrC exists in a rapidly exchanging equilibrium of hexamers and trimers, but that smaller species, such as monomers or dimers, are not seen (Holtham *et al.*, 1999). The associations which stabilize the hexameric core of AhrC can therefore be thought of as being of two types: strong intra-trimer interactions and more labile inter-trimer interactions.

The intra-trimer interface is mostly composed of three of the four β -strands of the core domain of each subunit. Around the threefold axis there is a hydrophilic pore which is surrounded by a ring made up of Ser89, Ser91 and His92 from each subunit. Each Ser91 is hydrogen bonded to His92 of the adjacent subunit and these residues are also involved in binding a number of water molecules within the pore. These interactions, along with interactions between hydrophobic residues around the threefold axis, serve to stabilize the tight interface of the stable trimer.

The hexamer is then formed through the interactions of two trimers. The resulting inter-trimer interface is considerably more labile than those within each trimer and this is reflected in the rather few strong intersubunit interactions. In fact, much of the interaction is mediated by a network of water molecules within the interface. The arginine-binding pocket is located in this area and is formed from residues Gln104, Asp111, Thr121, Cys123, Asp125 and Asp126 (Van Duyne *et al.*, 1996). A pattern of direct interactions are observed between Gln104 of each subunit and Asp125 of the subunit facing it across the trimer-trimer interface. By analogy with ArgREc (Van Duyne *et al.*, 1996) and ArgRBst (Ni *et al.*, 1999), it is anticipated that

significant changes to the inter-trimer interface occur upon the binding of L-arginine, resulting in a rotation of one trimer with respect to the other and the expulsion of some of the water molecules.

3.3. Comparison of the crystal structure of AhrC with other arginine repressors

The crystal structure of AhrC described here is the second structure of an intact arginine transcriptional regulator to be described, the other being that of ArgR from *B. stearo-thermophilus* (ArgRBst; Ni *et al.*, 1999). Both of these structures are of apo protein, *i.e.* without the bound effector L-arginine. In addition to these intact proteins, crystal structures are also known of the core hexamers from both ArgRBst with bound L-arginine (Ni *et al.*, 1999) and ArgREc both with and without L-arginine (Van Duyne *et al.*, 1996). The solution structure of the isolated DBD of ArgREc was determined by NMR (Sunnerhagen *et al.*, 1997). A detailed comparison has been made of the above structures, focusing primarily on the two intact proteins AhrC and ArgRBst.

The crystal packing arrangements of AhrC and ArgRBst are quite different, although it is of interest to note that both crystallize in space group $C222_1$ with a complete hexamer per asymmetric unit and that the two crystals have some similar unit-cell parameters (AhrC, $a = 230.3$, $b = 73.9$, $c = 138.6$ Å; ArgRBst, $a = 72.8$, $b = 121.9$, $c = 227.4$ Å). As anticipated from the high sequence identity (>70%) between AhrC and ArgRBst, their crystal structures are very similar (Fig. 4). The hexameric cores of the two proteins superimpose very closely, with an r.m.s.d. of 0.5 Å for the main-chain atoms. As a result of the flexibility of the inter-domain linker and the different crystal packing environments of different subunits in the two structures, the r.m.s.d. of the whole hexamers is much larger (10 Å), although individual DBDs superimpose with an r.m.s.d. of 0.9 Å. The fact that the tertiary and quaternary structures of AhrC and ArgRBst from different crystal forms are so similar gives weight to the argument that these structures are representative of the structures of the proteins in solution.

ArgREc has a considerably lower sequence similarity with AhrC than does ArgRBst. The overall sequence identity is only 27%, with many of the identical residues being located within the C-terminal oligomerization domain, where the identity rises to 45%. The success of molecular replacement using the ArgREc core structure as a trial model (see §2) confirms that both proteins share similar structures within the 32 symmetric core region.

With the aid of the structural information available on the arginine repressors, a detailed analysis of sequence conservation has been carried out. An alignment of the sequences of AhrC, ArgRBst and ArgREc is presented in Fig. 3(c). The positions of conserved residues have been mapped onto the three-dimensional structure of AhrC (Fig. 3a, chain B) to show their distribution. There are 33 residues conserved among all three proteins and these are shown in red in Fig. 3(c). The N-terminal 25 amino acids of ArgREc show no identical

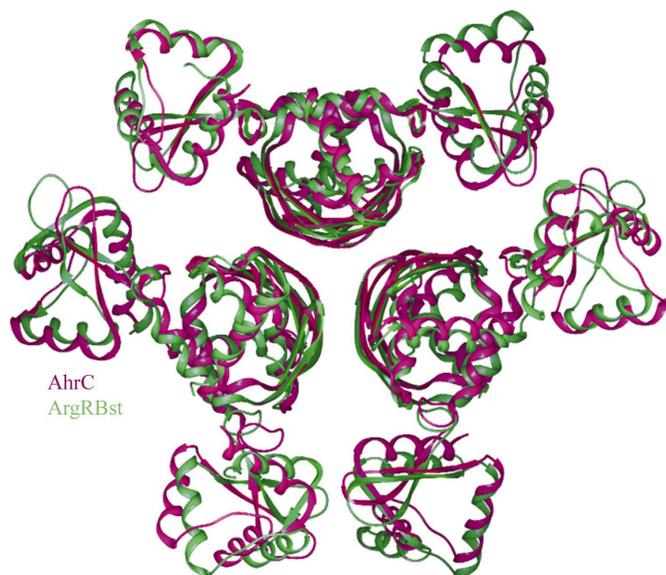


Figure 4
A superposition of the structures of *B. subtilis* AhrC (green) with its *B. stearo-thermophilus* homologue ArgR (ArgRBst, magenta). The core domains superimpose closely, whilst the DNA-binding domains are seen to adopt a variety of slightly different orientations. This figure was produced using *SPOCK*.

residues and little sequence similarity to either of the *Bacillus* proteins. Throughout the remainder of the sequences conserved residues are seen to be quite widely distributed, with notable clustering in several structural regions.

Within the DBD, conserved residues are found in helices 2 and 3 and at the base of the wing. Four of these residues in ArgRBst (Gln22, Gln38, Ser42 and Arg43) have been subjected to mutagenesis and, in an assay monitoring loss of repression of a reporter gene, shown to be important in DNA binding (Miltcheva Karaivanova *et al.*, 1999). These data, coupled with modelling of the putative ArgRBst–DNA complex (Ni *et al.*, 1999), suggest that helix 3 plays the central role in DNA binding. The arginine transcriptional regulators would therefore seem to be members of the classical HNF-3/fork-head winged helix family of DNA-binding proteins (Clark *et al.*, 1993), rather than resembling the recently characterized hRFX1 protein which interacts with DNA primarily *via* residues in the wing (Gajiwala *et al.*, 2000).

Conserved residues within the C-terminal core domain of AhrC include those in a particularly striking cluster (residues 120–127) encompassing the C-terminal portion of β -strand 5, the connecting loop and the first residue of β -strand 6. This region contains several residues implicated in the binding of L-arginine: Thr121, Asp125 and Asp126.

3.4. Structural and functional differences between AhrC and ArgRBst

The sequence identity between AhrC and ArgRBst is greater than 70% and despite this high degree of sequence similarity it is known that the two proteins have some different molecular properties. One important functional difference is in the thermal stability of the proteins, as unlike *B. subtilis* which is a mesophile, *B. stearothermophilus* is a thermophile and so all proteins within the organism must be capable of functioning at elevated temperatures. Through an analysis of the locations of some of the differing amino acids in the structures of AhrC and ArgRBst, it may be possible to explain some of the functional differences between the proteins.

Examination of the amino-acid sequences of the C-terminal oligomerization domains of AhrC and ArgRBst reveals only two main regions of difference. The first of these regions, located in the loop close to the molecular threefold axis, is residues 89–94 (shown in yellow in Figs. 3*a* and 3*c*). Two of these residues, Ser91 and His92, are involved in a symmetry-related network of hydrogen-bond interactions (as described above), which cannot be mimicked in ArgRBst, in which the two residues are glycine and asparagine, respectively. The second region comprises six residues within helix 6 (residues 136–144). Five substitutions out of the six amino acids results in changes in charge between the proteins. In AhrC, there are two additional salt bridges that may be formed between residues Thr138 and His92 and between Lys142 and Ser89. These salt bridges are not observed in ArgRBst, although Lys140 makes an interaction with Glu117. These small differences in charge and hydrophobicity render AhrC slightly more hydrophilic than ArgRBst.

As part of their mutational analysis of ArgRBst, Miltcheva Karaivanova and coworkers made a series of single and double mutations at amino acids 87 and 94 within the core domain of the repressor (Miltcheva Karaivanova *et al.*, 1999). These are residues which differ between AhrC and ArgRBst

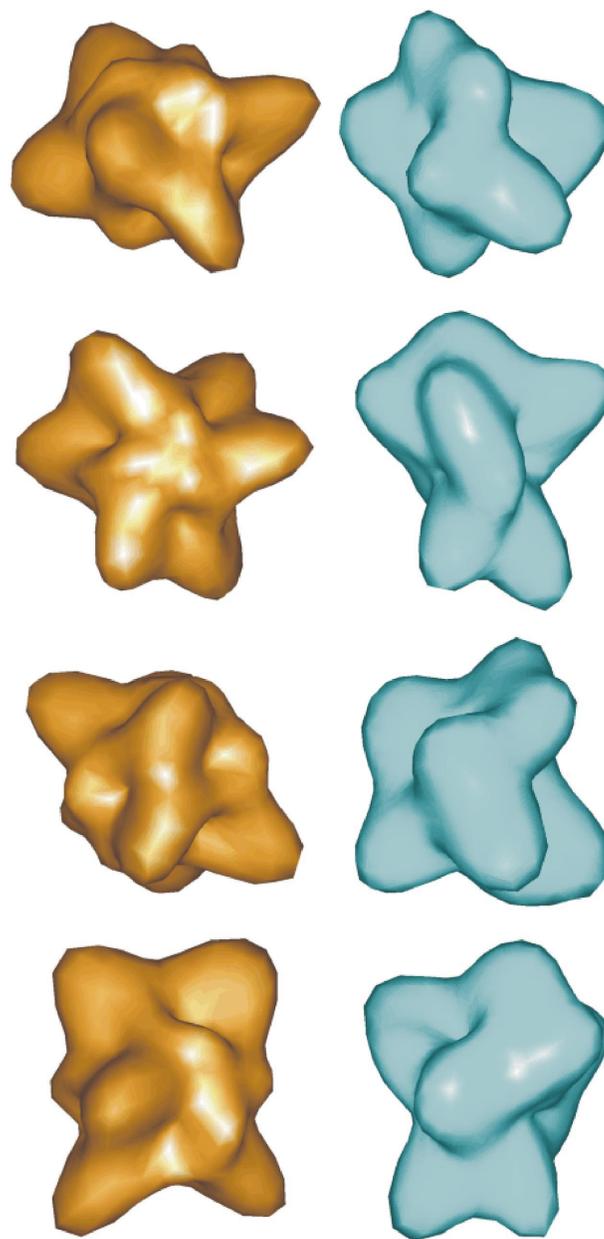


Figure 5
A comparison of surface representations of the crystal structure of AhrC (left) with the 30 Å resolution structure determined using phases derived from electron microscopy (right; Glykos *et al.*, 1998). Four different views of each of the two structures are shown. The second and third pairs are views along the molecular threefold and twofold axes, respectively. The top pair shows views along a direction which approximately bisects these axes and the bottom row is approximately 30° away from this view. The molecular surfaces were calculated using CRYSOLO and TRANS3 (Svergun *et al.*, 1995) and displayed using ASSA (Kozin *et al.*, 1997) as described in §2.

and two of the mutants produced in fact replace the ArgRBst residue(s) with the side chain(s) found in AhrC (Leu94Ile and the double mutant Leu87Ile/Leu94Ile). In an assay of repression of an *argC-lacZ* reporter gene, the Leu94Ile mutant showed an increase in repression of about 13% compared with the wild type, whilst the double mutant showed no significant difference. Miltcheva Karaivanova and coworkers went on to analyse the temperature dependence of DNA binding of wild type ArgRBst and one of their double mutants (Leu87Phe/Leu94Val). The wild-type repressor was found to be more efficient in binding to a 137 bp *argC* operator fragment at elevated temperatures, showing an apparent K_d of 78 nM at 328 or 343 K, compared with 104 nM at 310 K. Such an effect was not observed for the double mutant (Miltcheva Karaivanova *et al.*, 1999). Whilst it would be inappropriate to extrapolate these data from one mutant too far, they do suggest that even quite small differences in critical regions of the proteins might have substantial effects on the thermal characteristics of the proteins.

3.5. Comparison of the crystal structure of AhrC with its EM reconstruction

For a number of years following the initial crystallization of AhrC (Boys *et al.*, 1990), our crystallographic analysis of AhrC was hindered by difficulties in preparing useful isomorphous heavy-atom derivatives and in identifying a correct molecular-replacement solution using the *E. coli* ArgR core fragment (van Duyne *et al.*, 1996) when it became available. Because of the anticipated non-crystallographic symmetry of the AhrC hexamer, even a low-resolution model of the protein, in conjunction with real-space averaging and solvent flattening, might have allowed us to solve our heavy-atom or molecular-replacement problems. In order to produce a low-resolution model, an electron-microscopic study of AhrC crystals was carried out to complement our X-ray studies (Glykos *et al.*, 1998). From electron micrographs of thin fragments of AhrC crystals, the [001], [010] and [130] projections could be identified and the phases (signs) of centrosymmetric reflections calculated by Fourier transformation of the micrographs. 22 of the 35 unique reflections to 30 Å resolution can be measured from these three projections. The phases of two additional low-resolution centric reflections from the [0*kl*] plane were determined through permutation syntheses. The phases of acentric reflections were calculated from a simple crystal packing model comprising a pseudo-atom with a large temperature factor placed at the molecular centre, which was estimated from heavy-atom isomorphous difference Patterson maps, from low-resolution permutation syntheses and from electron micrographs of negatively stained crystal fragments. This resulting phase set was used in conjunction with experimentally measured X-ray structure-factor amplitudes to calculate an electron-density map at 30 Å resolution (Glykos *et al.*, 1998).

This map (Fig. 5, right-hand side) showed the density to be organized into a number of lobes around a central core. In order to evaluate the accuracy of this earlier model, it has

Table 2

Comparison of the phases of low-resolution reflections ($d \geq 30$ Å) determined in the previous combined electron-microscopic/X-ray study with those calculated from the refined crystal structure.

Reflections (*hkl*) are indicated as being either centric (C) or acentric (A). $|F_c|$ and φ_{calc} are the structure-factor amplitude and phase calculated from the refined 27 Å crystal structure of AhrC. $|F_{\text{EM-X}}|$ and $\varphi_{\text{EM-X}}$ are the structure-factor amplitude and phase determined in the combined electron-microscopic/X-ray crystallographic study of Glykos *et al.* (1998). The phases of the two [0*kl*] reflections [021] and [022] were correctly determined by permutation syntheses. The [311] reflection was too weak to measure and its phase was not determined (n.d.). $\Delta\varphi$ is the difference between φ_{calc} and $\varphi_{\text{EM-X}}$.

Flag for centrics	<i>h</i>	<i>k</i>	<i>l</i>	Resolution $1/d^2$ (Å ⁻²)	$ F_c $	φ_{calc} (°)	$ F_{\text{EM-X}} $	$\varphi_{\text{EM-X}}$ (°)	$\Delta\varphi$ (°)
C	2	0	0	0.00008	8571.9	180.0	560	180	0
C	2	0	1	0.00013	32020.4	90.0	8120	90	0
C	1	1	0	0.00020	66871.6	0.0	8960	0	0
C	0	0	2	0.00021	60106.9	180.0	12000	180	0
A	1	1	1	0.00025	19311.4	194.8	6560	163	32
C	2	0	2	0.00028	17688.8	0.0	6160	0	0
C	4	0	0	0.00030	6274.2	180.0	280	180	0
C	3	1	0	0.00035	8601.6	180.0	2800	180	0
C	4	0	1	0.00035	314.1	90.0	140	90	0
A	3	1	1	0.00040	2909.0	16.5	Weak	n.d.	n.d.
A	1	1	2	0.00041	8038.1	319.5	4200	202	118
C	4	0	2	0.00051	25148.8	0.0	7280	0	0
C	2	0	3	0.00054	386.1	90.0	338	90	0
A	3	1	2	0.00056	9948.4	108.8	4074	169	60
C	5	1	0	0.00065	20133.6	0.0	5607	0	0
A	1	1	3	0.00067	15390.1	354.8	4359	137	142
C	6	0	0	0.00068	3612.2	0.0	280	0	0
A	5	1	1	0.00071	13149.6	265.3	3798	139	126
C	0	2	0	0.00073	17950.8	180.0	4480	180	0
C	6	0	1	0.00073	18975.1	90.0	7000	90	0
C	4	0	3	0.00077	12281.8	270.0	5948	270	0
C	0	2	1	0.00079	21025.7	0.0	6000	0	0
C	2	2	0	0.00081	18984.7	180.0	6440	180	0
A	3	1	3	0.00082	10729.7	49.2	3550	35	14
C	0	0	4	0.00083	2024.1	0.0	3228	0	0
A	2	2	1	0.00086	8899.3	227.2	2160	267	40
A	5	1	2	0.00086	15427.9	299.1	3661	207	92
C	6	0	2	0.00089	5795.1	0.0	582	0	0
C	2	0	4	0.00090	13581.3	180.0	4768	180	0
C	0	2	2	0.00094	6323.3	180.0	3135	180	0
A	2	2	2	0.00102	13436.0	333.0	3009	84	111
A	1	1	4	0.00103	27391.9	267.1	7000	205	62
C	4	2	0	0.00103	13006.2	0.0	2077	0	0
A	4	2	1	0.00109	404.2	288.7	2217	177	112
C	7	1	0	0.00111	2150.1	0.0	3921	180	180

been compared with a 30 Å resolution surface calculated from the refined crystal structure. The two models are presented in Fig. 5. As can be seen, the combined EM/X-ray model shows many of the features of the final crystal structure. Lobes corresponding to five of the six DNA-binding domains are correctly located, as seen most clearly in Fig. 5 (right-hand side). The molecular 32 symmetry is less clearly seen in the earlier model and the directions of the molecular threefold and twofold axes were not known with certainty at the time that the model was calculated, so the mask could not be improved by NCS averaging.

To examine the accuracy of the phases derived in this earlier study, they were compared with phases calculated from the refined crystal structure (Table 2). All of the centrosymmetric phases have the correct sign, with the exception of the highest resolution reflection [710]. The amplitude-weighted mean

phase difference $[\sum(|F_c| \times \Delta\phi) / \sum |F_c|]$ for all reflections to 30 Å resolution is 25° and for just the acentric reflections it is 62° to 35 Å resolution, increasing to 88° to 30 Å resolution.

Considering that only 34 terms were used in the calculation of the EM/X-ray model, it shows remarkable similarity to the final crystal structure. The general agreement between the structures both validates the earlier low-resolution model and suggests that the electron-microscopic examination of thin fragments of protein crystals may prove a valuable method for obtaining low-resolution phases in other favourable cases.

4. Discussion

The crystal structure of AhrC, the arginine activator/repressor protein from *B. subtilis*, has been solved and refined to a resolution of 2.7 Å. As anticipated, the protein folds into a two-domain structure and oligomerizes *via* its C-terminal domain into a hexamer with 32 symmetry, effectively a dimer of trimers, first seen in the *E. coli* ArgR core structure (Van Duyne *et al.*, 1996). The N-terminal domains which are responsible for binding DNA are located around the periphery of the core hexamer and, as was seen in the structure of ArgR from *B. stearothermophilus* (Ni *et al.*, 1999), the polypeptide chains cross over, with the result that the DNA-binding domains from one trimer are located around the core domains of the other trimer and *vice versa*. The DNA-binding domains are loosely associated in pairs and it is anticipated that each DBD pair will bind to an 18 bp approximately palindromic DNA operator sequence known as an ARG box. As a result of differences in their crystal packing environments and of the inherent flexibility of the loop connecting the DNA-binding and core domains, the DNA-binding domains do not obey strict NCS. This flexibility is thought to be functionally significant in allowing pairs of DNA-binding domains to adopt the necessary arrangement for operator recognition. AhrC becomes activated for DNA binding when L-arginine binds to the core domains. By analogy with the crystal structures of L-arginine-bound core hexamers of ArgREc and ArgRBst, arginine binding results in the tightening of the trimer–trimer interface through the expulsion of solvent molecules and in a relative rotation of the trimers, which is thought to bring the N-terminal domains into their preferred orientation for DNA binding.

AhrC and ArgRBst originate from different strains of *Bacillus* and share 70% sequence identity, which inevitably confers great structural similarity. Although very similar in sequence and structure, some important differences must exist, as ArgRBst must be able to function at the elevated temperatures in which the thermophilic organism *B. stearothermophilus* can survive. An examination of the locations of different amino acids in the two structures revealed a run of six amino acids in a loop close to the trimer axis, which is observed to involve more hydrogen bonding in AhrC than in ArgRBst. The second region within helix 6 of AhrC contains amino acids with a different charge to the corresponding residues in ArgRBst. Two different salt bridges occur linking helix 6 with the loop region described above. A mutational

analysis of residues of ArgRBst within and adjacent to this region suggests that even quite conservative changes can alter the thermal properties of the protein (Miltcheva Karaivanova *et al.*, 1999). In common with examinations of other proteins, it has not proven possible to identify specific structural differences between a mesophile protein and a thermophile protein on the basis of their crystal structures alone (Creighton, 1993).

Prior to the solution of the crystal structure of AhrC, a model of the protein had been determined by combining phases derived from electron microscopy of crystal fragments with X-ray amplitudes for 34 structure factors to a maximum resolution of 30 Å (Glykos *et al.*, 1998). This model has been compared with a low-resolution envelope constructed around the refined structure. Considering how few independent data were used in the calculation and that there was no imposition of molecular symmetry, the two models show remarkable similarities; for instance, lobes representing five of the six DNA-binding domains are clearly present in the earlier model. The high quality of the earlier model is due to a significant degree to the fact that 23 of the 34 measured reflections lie in centrosymmetric zones and the signs of the phases of all but one of these agree with those calculated from the refined crystal structure. This result suggests that in favourable cases electron microscopy of crystal fragments may prove a valuable aid in determining crystal packing and calculating a low-resolution molecular envelope for use in solvent flattening and NCS averaging in crystal structure determination.

We thank the staff of Daresbury SRS station 9.6 for assistance with data collection. We acknowledge helpful discussions with Professor S. Baumberg, Professor P. G. Stockley and Dr J. Jäger. This work was supported by grants from the BBSRC and by a BBSRC studentship (to NMG).

References

- Boys, C. W. G., Czaplowski, L. G., Phillips, S. E. V., Baumberg, S. & Stockley, P. G. (1990). *J. Mol. Biol.* **213**, 227–228.
- 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, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Calogero, S., Gardan, R., Glaser, P., Schweizer, J., Rapoport, G. & Débarbouillé, M. (1994). *J. Bacteriol.* **176**, 1234–1241.
- Christopher, J. A. (1998). *SPOCK*. <http://quorum.tamu.edu/spock/>.
- Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. (1993). *Nature (London)*, **364**, 1903–1906.
- Cole, S. T. *et al.* (1998). *Nature (London)*, **393**, 537–544.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Creighton, T. E. (1993). *Proteins: Structures and Molecular Properties*, 2nd ed. New York: W. H. Freeman & Co.
- Czaplowski, L. G., North, A. K., Smith, M. C. M., Baumberg, S. & Stockley, P. G. (1992). *Mol. Microbiol.* **6**, 267–275.
- Dion, M., Charlier, D., Wang, H., Gigot, D., Savchenko, A., Hallet, J. N., Glansdorff, N. & Sakanyan, V. (1997). *Mol. Microbiol.* **25**, 385–398.
- Fleischmann, R. D. *et al.* (1995). *Science*, **269**, 496–512.

- Gajiwala, K. S., Chen, H., Cornille, F., Roques, B. P., Reith, W., Mach, B. & Burley, S. K. (2000). *Nature (London)*, **403**, 916–921.
- Gardan, R., Rapoport, G. & Débarbouillé, M. (1995). *J. Mol. Biol.* **249**, 843–856.
- Glansdorff, N. (1987). *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology*, edited by F. C. Neidhardt, pp. 321–344. Washington: American Society for Microbiology.
- Glykos, N. M. (1995). PhD thesis. University of Leeds, UK.
- Glykos, N. M., Holzenburg, A. & Phillips, S. E. V. (1998). *Acta Cryst.* **D54**, 215–225.
- Harwood, C. R. & Baumberg, S. (1977). *J. Gen. Microbiol.* **100**, 177–188.
- Holtham, C. A. M., Jumel, K., Miller, C. M., Harding S. E., Baumberg, S. & Stockley, P. G. (1999). *J. Mol. Biol.* **289**, 707–727.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Kleywegt, G. J. & Jones, T. A. (1994). *Proceedings of the CCP4 Study Weekend. From First Map to Final Model*, edited by S. Bailey, R. Hubbard & D. Waller, pp. 59–66. Warrington: Daresbury Laboratory.
- Klingel, U., Miller, C. M., North, A. K., Stockley, P. G. & Baumberg, S. (1995). *Mol. Gen. Genet.* **248**, 329–340.
- Kozin, M. B., Volkov, V. V. & Svergun, D. I. (1997). *J. Appl. Cryst.* **30**, 811–815.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Leslie, A. G. W. (1992). *Int CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**, 83–91.
- Lim, D., Oppenheim, J. D., Eckhardt, T. & Maas, W. K. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 6697–6701.
- Lu, C.-D., Houghton, J. E. & Abdelal, A. T. (1992). *J. Mol. Biol.* **225**, 11–24.
- Maas, W. K. (1994). *Microbiol. Rev.* **58**, 631–640.
- Miller, C. M. (1997). PhD thesis. University of Leeds, UK.
- Miller, C. M., Baumberg, S. & Stockley, P. G. (1997). *Mol. Microbiol.* **26**, 37–48.
- Miltcheva Karaivanova, I., 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**, 843–855.
- Mountain, A. & Baumberg, S. (1980). *Mol. Gen. Genet.* **178**, 691–701.
- Navaza, J. (1997). *Methods Enzymol.* **276**, 581–594.
- 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.
- Ohtani, K., Bando, M., Swe, T., Banu, S., Oe, M., Hayashi, H. & Shimizu, T. (1997). *FEMS Microbiol. Lett.* **146**, 155–159.
- Priebe, S. D., Hadi, S. M., Greenberg, B. & Lacks, S. A. (1988). *J. Bacteriol.* **170**, 190–196.
- Rodriguez-Garcia, A., Ludovice, M., Martin, J. F. & Liras, P. (1997). *Mol. Microbiol.* **25**, 219–228.
- Smith, M. C. M., Czaplewski, L., North, A. K., Baumberg, S. & Stockley, P. G. (1989). *Mol. Microbiol.* **3**, 23–28.
- Smith, M. C. M., Mountain, A. & Baumberg, S. (1986a). *Gene*, **49**, 53–60.
- Smith, M. C. M., Mountain, A. & Baumberg, S. (1986b). *Mol. Gen. Genet.* **205**, 176–182.
- Stirling, C. J., Szatmari, G., Stewart, G., Smith, M. C. M. & Sherratt, D. J. (1988). *EMBO J.* **4**, 4389–4395.
- Sunnerhagen, M., Nilges, M., Otting, G. & Carey, J. (1997). *Nature Struct. Biol.* **4**, 819–826.
- Svergun, D. I., Barberato, C. & Kock, M. H. J. (1995). *J. Appl. Cryst.* **28**, 768–773.
- Tian, G. & Maas, W. K. (1994). *Mol. Microbiol.* **13**, 599–608.
- Van Duyne, G. D., Ghosh, G., Maas, W. K. & Sigler, P. B. (1996). *J. Mol. Biol.* **256**, 377–391.