JMB

Probing Activation of the Prokaryotic Arginine Transcriptional Regulator Using Chimeric Proteins

Carol A. M. Holtham[1], Kornelia Jumel[2], Coleen M. Miller[1] Stephen E. Harding[2], Simon Baumberg[1] and Peter G. Stockley[1]*

[1]*Astbury Centre for Structural Molecular Biology and North of England Structural Biology Centre, Faculty of Biological Sciences, University of Leeds Leeds, LS2 9JT, UK*

[2]*N.C.M.H. Unit, School of Biological Sciences, University of Nottingham, Sutton Bonington, Leicestershire LE12 5RD, UK*

The major transcription factors controlling arginine metabolism in *Escherichia coli* and *Bacillus subtilis,* ArgR and AhrC, respectively, are homologous multimeric proteins that form L-arginine-dependent DNA-binding complexes capable of repressing transcription of the biosynthetic genes (both), activating transcription of catabolic genes (AhrC only) or facilitating plasmid dimer resolution (both). Multimerisation and L-arginine binding are associated with the C-terminal 70-80 residues; the N-terminal regions contain a winged helix-turn-helix DNA-binding domain. We have constructed chimeric genes in which the sequences for the N and C-terminal domains have been swapped. The resultant chimeric proteins and their corresponding native proteins have been analysed for their ability to multimerise and bind DNA operator sites in an L-arginine-dependent fashion.

Gel filtration and equilibrium sedimentation analysis are consistent with the formation of hexamers by all four proteins in the presence of L-arginine and at high protein concentrations (>100 nM monomer). The hexamer sedimentation coefficients suggest that there is a reduction in molecular volume upon binding L-arginine, consistent with a conformational change accompanying an allosteric activation of DNA-binding. In the absence of L-arginine or at lower protein concentrations, the hexamers are clearly in rapid equilibrium with smaller subunits, whose dominant species appear to be based on trimers, as expected from the crystal structure of the ArgR C-terminal fragment, with the exception of the ArgR-C chimera, which apparently dissociates into dimers, suggesting that in the intact protein the DNA-binding domains may have a significant dimeric interaction. The hexamer-trimer K_d is in the micromolar range, suggesting that trimers are the principal species at *in vivo* concentrations.

DNA binding by all four proteins has been probed by gel retardation and DNase I footprinting analysis using all three types of naturally occurring operators: biosynthetic sites encompassing two 18 bp ARG boxes separated by 2 bp; biosynthetic sites containing two such boxes and a third 18 bp ARG box at a distance of 100 bp downstream, i.e. within the structural gene; and finally a catabolic operator which contains a single ARG box site. The data show that all four proteins bind to the operators at the expected regions in an L-arginine-dependent fashion. From the apparent affinities of the chimeras for each target site, there is no obvious sequence-specificity associated with the N-terminal domains; rather the data can be interpreted in terms of differential allosteric activation, including DNA binding in the absence of L-arginine.

Remarkably, the proteins show apparent ''anti-competition'' in the presence of excess, specific DNA fragments in gel retardation. This appears to be due to assembly of an activated form of the protein, probably hexamers, on the operator DNA. The data are discussed in terms of the current models for the mode of action of both native proteins.

© 1999 Academic Press

Present addresses: C. A. M. Holtham, Molecular Toxicology, Covance Laboratories Ltd., Otley Road, Harrogate, HG3 1PY, UK; C. M. Miller, The Burnham Institute, 10901, N.Torrey Pines Road, La Jolla, CA 92037, USA.
E-mail address of the corresponding author: stockley@bmb.leeds.ac.uk

Introduction

Transcriptional control of gene expression in the arginine pathways of *Escherichia coli* and *Bacillus subtilis* is regulated by homologous DNA-binding proteins composed of subunits containing ~150 amino acid residues, ArgR for *E. coli* (Lim *et al.*, 1987) and AhrC for *B. subtilis* (North *et al.*, 1989). The protein sequences share 27% sequence identity, mostly clustered in the C-terminal region. Gel filtration, chemical cross-linking and gel electrophoresis experiments suggested that both proteins are hexameric at roughly physiological concentration (Czaplewski *et al.*, 1992; Lim *et al.*, 1987). Mutational analysis of the *argR* gene (Tian & Maas, 1994) and other data (Maas, 1994) have led to the proposal that both repressors contain two domains, an N-terminal DNA-binding domain and a C-terminal L-arginine-binding/multimerisation domain. This general picture has been confirmed by structural studies of ArgR fragments. Recently, Van Duyne *et al.* (1996) have reported the crystal structure of the C-terminal domain of ArgR (ArgRc) at 2.8 Å resolution. This has revealed that the oligomeric core consists of two trimers composed of ArgRc subunits. The trimers have tightly packed hydrophobic cores and are arranged back-to-back with a sparsely populated hydrophobic interface. Six L-arginine molecules bind at this interface effectively crosslinking the trimers.

Sherratt and co-workers (Burke *et al.*, 1994; Chen *et al.*, 1997) have isolated ArgR mutations deficient in trimer-trimer interactions, which appear still able to bind operator sites. The N-terminal region of ArgR, containing the DNA-binding determinants, has been reported to be monomeric (Grandori *et al.*, 1995), at least in isolation. A number of laboratories have reported solution structures for this region based on NMR data, which suggest that there is a winged helix-turn-helix motif within the domain (Sunnerhagen *et al.*, 1997; D. J. Sherratt, personal communication). Crystals of intact apo-AhrC have also been reported (Boys *et al.*, 1990) and the current progress in the structural studies of both systems holds the promise of a detailed comparison of these molecules at the atomic level.

ArgR binds to 18 bp palindromic sequences, ARG boxes, located adjacent to the promoters of the different *arg* operons (Glansdorff, 1987), with the consensus sequence a/tNTGAATa/ta/tt/at/aATTCANt/a, where a/t, for example, indicates that A or T occur with equal frequency, and N is any nucleotide (Maas, 1994). Biosynthetic operators usually contain two ARG boxes, *argG* being the exception with three such boxes. The affinities of these sites for the repressor have been estimated from quantitative DNase I footprinting as being between 0.1 and 1.0 nM ArgR (Maas, 1994). Adjacent ARG boxes are bound simultaneously and cooperatively by the repressor, which interacts along one face of the DNA duplex making both major and minor groove contacts, and bending the DNA by 70° (Tian *et al.*, 1992).

The *B. subtilis* homologue, AhrC, was first characterised in mutants resistant to the arginine analogue, arginine hydroxamate (Harwood & Baumberg, 1977; Mountain & Baumberg, 1980; Smith *et al.*, 1986a,b). Mutants mapping to *ahrC*, one of four arginine hydroxamate resistance loci, showed simultaneous loss of repression of the biosynthetic genes and activation of the catabolic genes, suggesting that the *ahrC* gene product was a direct regulatory link between the pathways. Limited *in vivo* cross-functioning of the gene products from *E. coli* and *B. subtilis* was also reported, suggesting that there was some conservation of regulatory interactions between the two organisms (Smith *et al.*, 1989). Subsequently, the interaction of the AhrC protein with one of the *Bacillus* biosynthetic operator regions adjacent to the *argC* promoter was examined by DNase I and hydroxyl radical footprinting. These experiments revealed two AhrC-binding sites. A higher-affinity site located within the *argC* promoter region, encompassing two ARG box-like regions, was termed $argC_{O1}$. The second, lower-affinity, binding site (called $argC_{O2}$) is located within the *argC* structural gene, an arrangement reminiscent of the *E. coli lac* operon, leading to the suggestion that repression loops might form (Czaplewski *et al.*, 1992). A similar situation has now been shown to exist in the other biosynthetic operator, *argGH* (Miller, 1997; C.M.M. *et al.*, unpublished results).

Recently, we reported characterisation of the complexes formed by AhrC at two catabolically induced promoters, *rocA* and *rocD* (Calogero *et al.*, 1994; Klingel *et al.*, 1995; Gardan *et al.*, 1995; Miller *et al.*, 1997). Both catabolic operators consist of a single copy of an ARG box-like site, located directly adjacent to the transcription start. They also have promoter elements at positions −12 and −24, which suggest that transcription at these sites is directed by the *B. subtilis* homologue of σ^{54}, namely σ^L (Débarbouillé *et al.*, 1991). Binding at these sites was shown to be cooperative with respect to protein concentration and appeared to be mediated by L-arginine, protein and DNA concentrations.

In order to understand the relationship between binding DNA operator sites containing roughly dyad-symmetric sequences and the co-repressor-induced multimerisation of trimeric units, we have produced chimeric proteins

between ArgR and AhrC. The chimeras were constructed such that the DNA encoding the N-terminal DNA-binding domain of one protein was fused in frame with the DNA encoding the C-terminal oligomerisation domain.

Results

Construction of the chimeric genes, over-expression and purification of the native and chimeric proteins

The amino acid sequences of AhrC and ArgR are shown in Figure 1. Grandori *et al.* (1995) noted that not only did the sequence similarity between ArgR and AhrC point to a domain boundary between amino acid residues 90-100, but also that there was a distinct charge residue boundary at the same place. In ArgR, the pI for the first 100 residues was calculated to be ∼9.5, whilst the remaining 50 residues had a calculated pI of ∼4.0. Proteolysis of ArgR suggested that residues 73-83 represented a flexible linker region between the two domains. Using this information as a guide, Grandori *et al.* (1995) cloned and purified the first 96 residues of ArgR. By including the linker region it was hoped that the N-terminal DNA-binding domain would be more stable. Knowledge of these results prior to publication facilitated our decision to splice the wild-type genes between the amino acids highlighted in Figure 1. This site lies C-terminal to the suspected linker region and is close to the pI boundary. More importantly the amino acids that flank the splice site are identical in both proteins, which it was hoped would reduce any perturbation in the tertiary structures. The resultant chimeric genes encode the N-terminal 87 residues of AhrC fused to the region 90-156 of ArgR *(ahrC-R)* and the N-terminal 89 residues of ArgR fused to the region 88-149 of AhrC *(argR-C)*.

The chimeric genes were synthesised using a two-step PCR protocol, detailed in Figure 2, and the products subsequently ligated into suitably prepared pET22b expression plasmids. For each recombinant plasmid a single clone was selected for automated DNA sequencing. The expected nucleotide sequences of all the genes was confirmed, with the exception that a single point mutation (T instead of C) at base-pair 275 was found in the pET*argR-C* clone. The change is translationally silent, both codons ATC and ATT specifying isoleucine, so this clone was used in all subsequent experiments to generate ArgR-C protein.

Expression of the chimeric and wild-type proteins was achieved by transforming BL21(DE3) *E. coli* cells to ampicillin resistance with the appropriate pET plasmids. Protein expression was induced by the addition of IPTG using standard protocols, followed by lysis and fractionation into insoluble pellet and soluble supernatant fractions. Like AhrC, both AhrC-R and ArgR-C were found in the pellet fractions at this stage and so the subsequent purification protocol was based on that for wild-type AhrC (Czaplewski *et al.*, 1992), final purification being by chromatography on S-Sepharose (Pharmacia). The progress of each purification was followed by SDS-PAGE (Figure 3). Both purified chimeras appeared as a single band on Coomasssie brilliant blue-stained gels. The identities of the purified chimeras were confirmed by N-terminal amino acid sequencing (AhrC-R only) and/or mass spectrometry. For AhrC-R, the observed molecular ion had a mass of 17,159.7 Da, within the error range for the expected molecular mass of 17,200 Da. For ArgR-C, the observed molecular mass of 16,664 Da was almost identical with the expected molecular mass of 16,666 Da and so N-terminal sequencing was not carried out.

```
AhrC    1 ...MNKGQRHIK.IREIITSNEIETQDELVDMLKQDGY.KVTQATVSRDI 45
            | :  :| :: ::   : :  |:|:|  | ::|:  : |  ||| :
ArgR    1 MRSSAKQEELVKAFKALLKEEKFSSQGEIVAALQEQGFDNINQSKVSRML 50

       46 KELHLVKVPTNNGSYKYSLPADQRFNPLSKLKRALMDAFVKIDSASHMIV 95
            :  |:       |:|||:      :   :| : : ||    ::|
       51 TKFGAVRTRNAKMEMVYCLPAEL...GVPTTSSPLKNLVLDIDYNDAVVV 97

       96 LKTMPGNAQAIGALMDNLDWDE.MMGTICGDDTILIICRTPEDTEGVKNR 144
            : | || || |: |:| |:  | ::||| ||||:        :: :
       98 IHTSPGAAQLIARLLDSLGKAEGILGTIAGDDTIFTTPANGFTVKDLYEA 147

      145 LLELL*
            :|||:
      148 ILELFDQEL*
```

Figure 1. Amino acid sequence comparison of AhrC and ArgR. The sequence comparison was compiled using a Gap alignment (GCG Package Version 7, Genetics Computer Group, Wisconsin), which maximises matches and minimises the number of gaps. The upper sequence is AhrC whilst the lower sequence is ArgR. Identical amino acid residues are indicated by a vertical line, whilst similar residues are indicated by two dots. *, Translational stop codon. The amino acids flanking the splice site used to construct the chimeric proteins are shown in italics.

A

```
                              NdeI ↓
                GG AATTCCAT→→→→→(oligo H3257)→→→→→
       1   GGAAATAGAG GTGCTTACAT GAACAAAGGC CAGAGGCATA
TTAAAATCAG
                              ↑
      51   AGAGATCATT ACAAGCAATG AAATCGAAAC GCAGGATGAA
TTGGTCGACA
     101   TGCTAAAGCA GGATGGATAC AAAGTAACGC AGGCCACGGT
ATCACGGGAT
     151   ATTAAAGAAC TTCACCTTGT AAAAGTGCCT ACAAATAATG
GTTCTTACAA
     201   ATACAGTCTT CCGGCGGACC AGCGTTTCAA TCCGCTGTCA
AAGCTAAAGC
                  ←←←(oligo G3280) ←←←←←←←C TGATGTTGCT ACGTCAACA
     251   GGGCATTGAT GGACGCATTT GTGAAAATAG ACTCAGCAAG
CCATATGATT
                                 →→→→→(oligo G3285)→→→→
     301   GTGCTGAAAA CGATGCCGGG CAATGCCCAG GCAATCGGGG
CGCTGATGGA
     351   CAATTTGGAC TGGGATGAAA TGATGGGGAC CATTTGCGGA
GATGATACGA
     401   TTTTAATTAT TTGCCGGACT CCTGAAGATA CAGAAGGCGT
AAAAAACAGG
     451   CTCCTTGAAC TGCTGTAACA GAAAATCTAA CAAAGA
                ←←←←(oligo G3286) ←←←←←TT CGAAGCGTAA
                              HindIII ↑
```

B

```
                              NdeI ↓
                GGAATT CCAT→→→(oligo H3256) →→→
   1   ATCAACCACC ATATCGGGTG ACTTATGCGA AGCTCGGCTA AGCAAGAAGA
                              ↑
  51   ACTAGTTAAA GCATTTAAAG CATTACTTAA AGAAGAGAAA TTTAGCTCCC
 101   AGGGCGAAAT CGTCGCCGCG TTGCAGGAGC AAGGCTTTGA CAATATTAAT
 151   CAGTCTAAAG TCTCGCGGAT GTTGACCAAG TTTGGTGCTG TACGTACACG
 201   CAATGCCAAA ATGGAAATGG TTTACTGCCT GCCAGCTGAA CTGGGTGTAC
                  ←←←←←← oligo G3284)←←←←←←←←TGAGTCGTT
 251   CAACCACCTC CAGTCCATTG AAGAATCTGG TGCTGGATAT CGACTACAAC
                                  →→→→(oligo
       CGGTATACTA
 301   GATGCAGTTG TCGTGATTCA TACCAGCCCT GGCGCGGCGC AGTTAATTGC
       G3281)→→→→→→→→→→
 351   TCGCCTGCTG GACTCACTGG GCAAAGCAGA AGGTATTCTG GGCACCATCG
 401   CTGGCGATGA CACCATCTTT ACCACCCCTG CTAACGGTTT CACAGTCAAA
 451   GACCTGTACG AAGCGATTTT AGAGCTGTTC GACCAGGAGC TTTAATCTCT
                                  ←←←← (oligo G3282)
 501   GCCCCGTCGT T
       ←←←←TTCGA AGCGTAA
          ↑HindIII
```

Figure 2. (*Legend shown on opposite page*)

Estimation of the relative, native molecular mass for the wild-type and chimeric arginine regulators

Samples of AhrC, ArgR, AhrC-R and ArgR-C were prepared from glycerol stocks by buffer exchange using gel filtration columns (Nap-5, Pharmacia) into 20 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 350 mM NaCl ± 10 mM L-arginine and stored at 4 °C.

The native molecular masses relative to protein standards in the presence or absence of L-arginine were then estimated on a Superdex-200 column (Pharmacia) connected to an FPLC system. Standard globular proteins were used to calibrate the column in ± L-arginine conditions. Experiments were carried out in triplicate over a number of weeks, calibrating the column each time. Representative elution profiles of the wild-type and chimeric
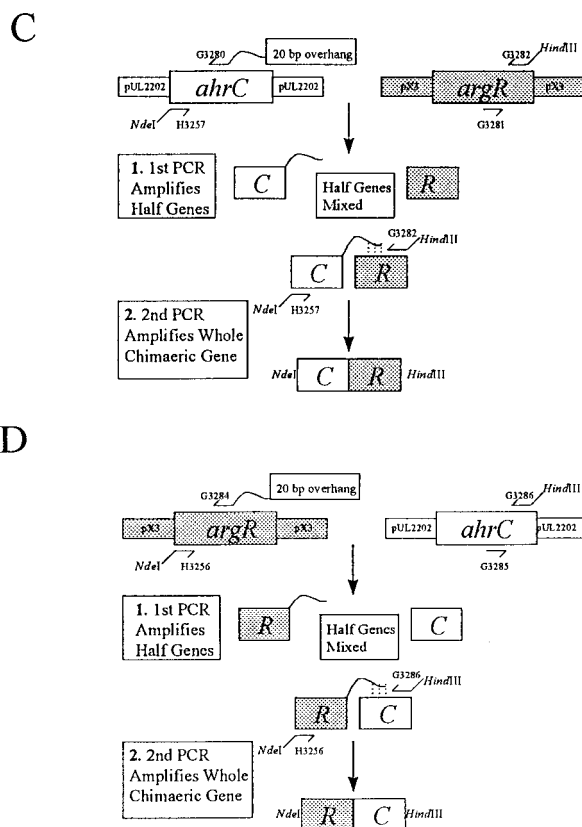
Figure 2. Binding sites of oligos and the PCR protocol used to construct the chimeras. (a) and (b) Nucleotide sequences of *ahrC* (Genbank accession number M27869) and *argR* (Genbank accession numbers M17532 and J02965), respectively. → → →, The binding sites and orientation of the eight oligos used to PCR amplify the four half genes. The sequence of non-complementary DNA at the 5′ end of some of the oligos are shown. Roman characters signify a restriction site at the end of the oligo (indicated above or below), whilst italics indicate the non-homologous regions used to "glue" half genes together in the secondary PCR. ↑, The start codon. (c), (d) PCR protocol used to synthesise *ahrC-R* (c) and *argR-C* (d). Oligos and their binding sites as seen in (a) and (b).

proteins in ± L-arginine buffers are shown in Figure 4. Table 1A lists the apparent molecular masses derived from the gel filtration experiments. The oligomeric state of ArgR was apparently unaffected by the presence of L-arginine and was consistent with the reported hexamer (Lim *et al.*, 1987). In contrast, two forms of AhrC were observed, a hexamer in the presence of L-arginine but a smaller oligomeric species, with an apparent molecular mass of 70(±9) kDa, was present in the absence of the co-factor. This is consistent with a tetrameric complex, although the resolution of the technique and the broadness of the peak of eluting protein, together with the assumptions used to construct the standard curve, mean that other closely related stoichiometries cannot be excluded.

Two similar but distinct forms of the AhrC-R chimera were also isolated under the same conditions, a hexamer being present in the presence of L-arginine, with a dimeric/trimeric species present in the absence of co-factor. Two forms of the ArgR-C chimera were also observed, the smaller species also being present even in the presence of L-arginine. Once again the largest species appeared to be hexameric with the smaller species being either

tetrameric or trimeric. In the absence of L-arginine only the smaller species was observed. SDS-PAGE analysis of all the species isolated confirmed that their differing apparent molecular masses were due to their oligomeric state and not to protein degradation (data not shown).

In order to examine the reversibility of the oligomerisation reaction, proteins which yielded smaller subunits in the absence of co-repressor were chromatographed in minus L-arginine buffers, as described above, the fractions corresponding to the smaller aggregates collected and L-arginine-containing buffer added to half the samples. After equilibration on ice for at least 60 minutes, the samples were rechromatographed. The apparent molecular masses of the eluting species in both ±-L-arginine buffers are listed in Table 1B. The addition of co-repressor to the AhrC sample resulted in the formation of the hexameric species. However, the chimeras showed distinct properties. Re-chromatography of both AhrC-R and ArgR-C in the absence of L-arginine resulted in species which were apparently smaller than the initial isolate, corresponding in the case of AhrC-R to a subunit monomer, consistent with further dis-

(a)

(b)
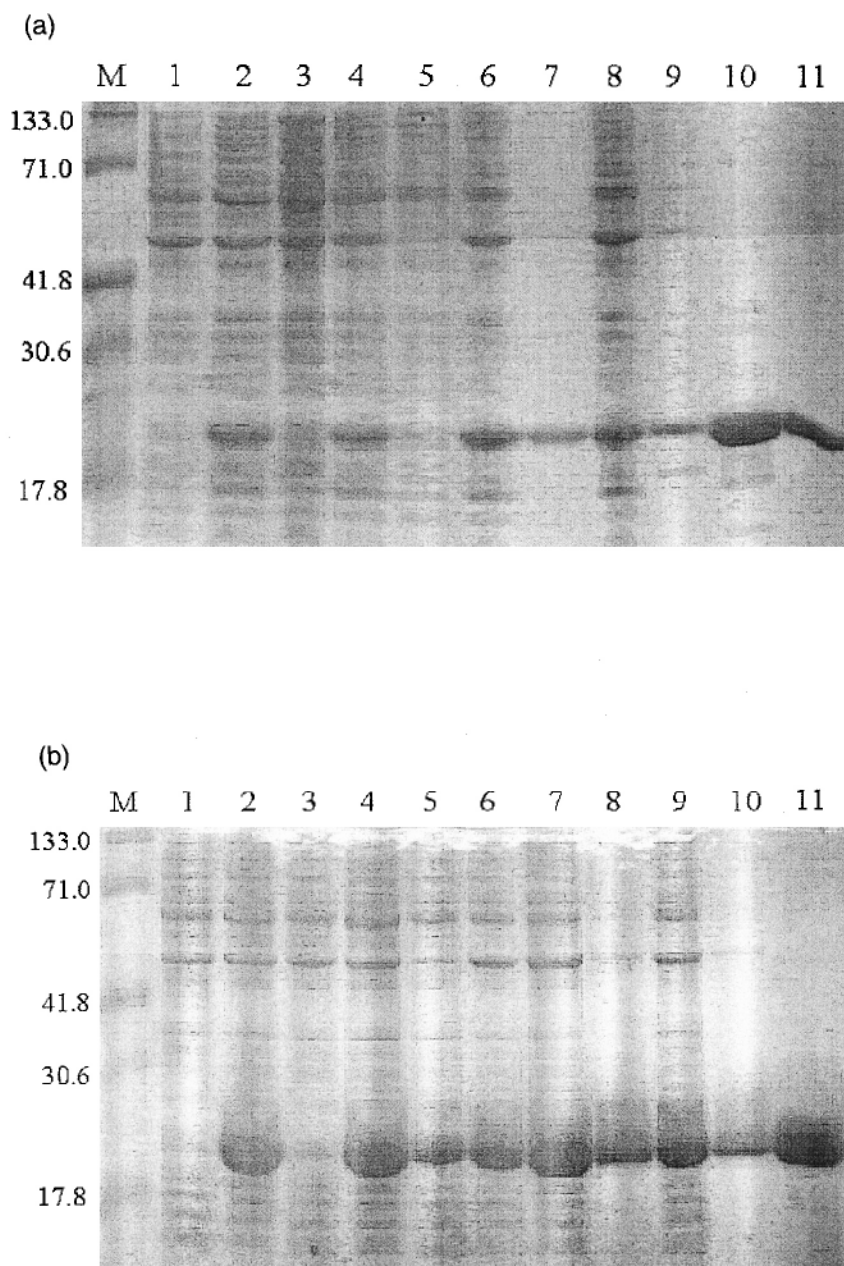
Figure 3. Purification of the chimeric proteins. (a) SDS-PAGE of AhrC-R samples. Lanes: M, protein molecular mass markers, expressed as kDa; 1, uninduced BL21(DE3)-pET*ahrC-R*; 2, induced BL21(DE3)-pET*ahrC-R*; 3, induced BL21(DE3) pET22b; 4, whole-cell sonicate; 5, sonicate supernatant; 6, sonicate pellet; 7, supernatant from high-salt incubation; 8, pellet from high-salt incubation; 9, supernatant from low-salt dialysis; 10, pellet from low-salt dialysis (resuspended and loaded on to the column); and 11, column fraction containing pure AhrC-R. (b) SDS-PAGE of ArgR-C. Lanes: M, protein molecular mass markers, expressed as kDa; 1, uninduced BL2I(DE3)pET*argR-C*; 2, induced BL2I(DE3)pET*argR-C*; 3, induced BL21(DE3)pET22b; 4, whole-cell sonicate; 5, sonicate supernatant; 6, sonicate pellet; 7, whole sample in high-salt; 8, supernatant from high-salt incubation; 9, pellet from high-salt incubation; 10, lane 8 supernatant after low-salt dialysis (loaded onto the column); and 11, column fraction containing pure ArgR-C.

sociation as the protein concentration was progressively lowered by passage across the column. In the presence of L-arginine, AhrC-R reformed hexamers, although with evidence of some monomer remaining. ArgR-C, in contrast, appeared to undergo a disproportionation reaction, generating three separable species which appear to correspond to tetramer, trimer and monomer.

The data are consistent with an L-arginine and protein concentration-dependent oligomerisation of the arginine regulatory proteins. Previously, we reported that AhrC was hexameric in the absence of L-arginine (Czaplewski *et al.*, 1992). This is clearly not the case and the earlier result could have been due to carry-over of L-arginine in our preparations. Strikingly, the AhrC-R chimera is unstable in the absence of L-arginine even though it contains the bulk of the ArgR C-terminal domain encompassing the co-repressor binding site, suggesting that both domains in the molecule might be involved in oligomer stabilisation. It is difficult, given the resolution of gel filtration experiments, to determine the accurate stoichiometries of the smaller subunits isolated unless they are obviously monomeric. However, there was no clear division into hexamers and trimers, which had been expected based on published data. We therefore carried out further measurements in the analytical ultracentrifuge in order to get a more accurate estimate of the nature of the smaller subunits.
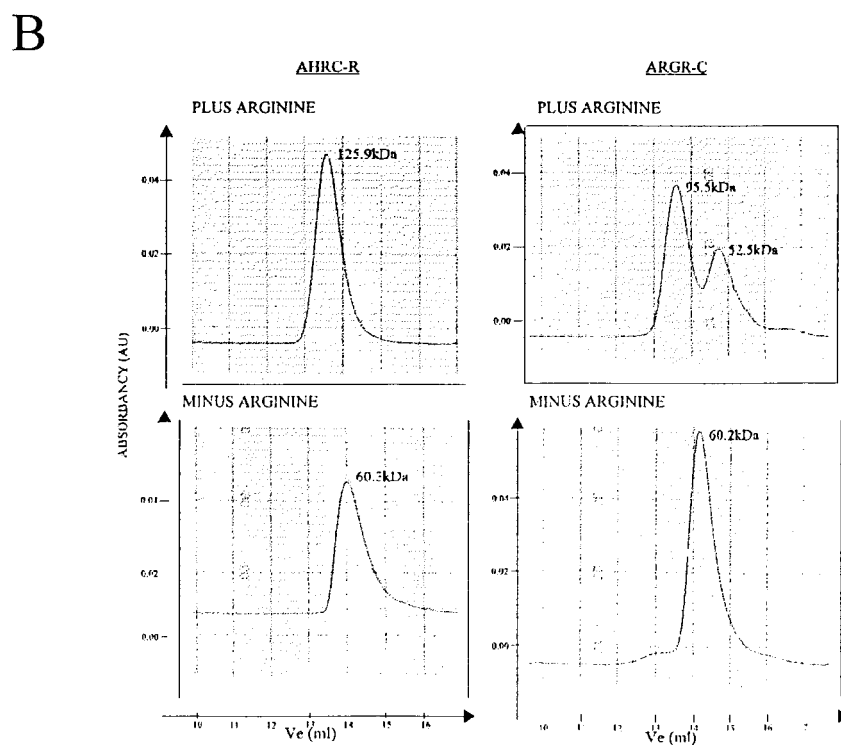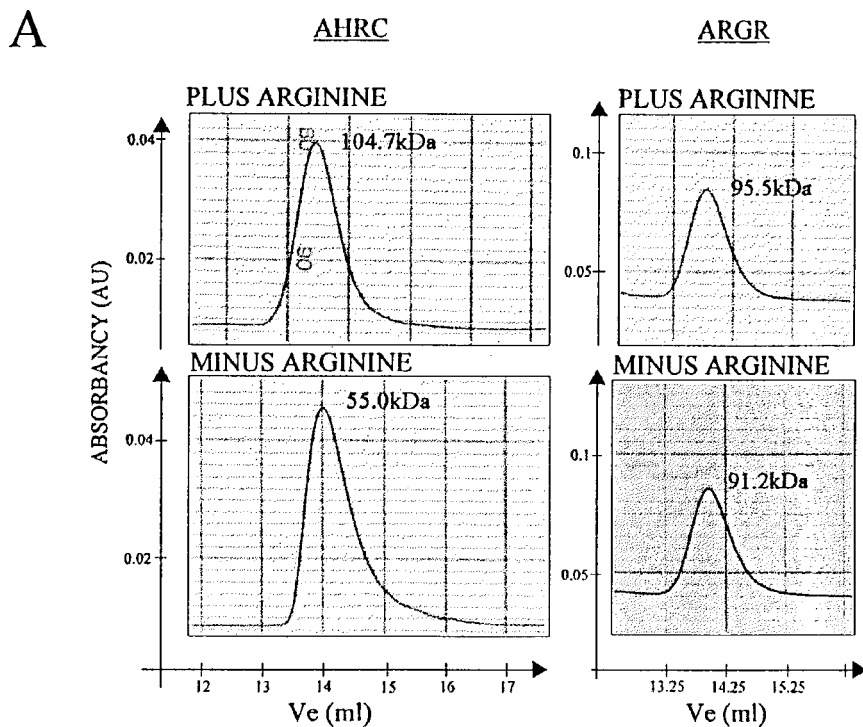
Figure 4. Size-exclusion chromatography. Representative elution profiles of the arginine repressors and the chimeras. Elution profiles are shown for each protein in elution buffer (20 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 350 mM NaCl, ±10 mM L-arginine). The 200 μl aliquots were loaded onto the column and eluted in 30 ml of buffer at 0.5 ml/minute. Protein concentrations ranged from 0.5 mg/ml to 1.2 mg/ml. The profiles have been aligned at $t = 0$. (a) Wild-type repressors; (b) chimeras. Apparent molecular masses are indicated in kDa.

Ultracentrifugation of the wild-type and chimeric proteins

Protein solutions for ultracentrifugation experiments, with or without 10 mM L-arginine, were prepared as for size-exclusion chromatography with the addition of 2 mM 2-mercaptoethanol to the buffer. This was necessary to increase the stability of the wild-type proteins, which aggregate slowly in the absence of thiol reagent. Its presence allowed extensive hydrodynamic and functional assays of the same batches of each protein. The solutions were stored at 4 °C for up to four weeks and diluted in the same buffer prior to use.

Sedimentation velocity measurements

Sedimentation velocity experiments were used to examine the homogeneity of the protein solutions

Table 1. Size-exclusion chromatography data

A. *±L-Arginine conditions*[a]

	Plus arginine			Minus arginine	
Protein	Average $M_{w,app}$	n		Average	n
AhrC	99.7 ± 3.7	6		70.7 ± 9.4	4/5
ArgR	112.7 ± 21.6	6		104.0 ± 11.0	6
AhrC-R	127.8 ± 2.8	6		39.3 ± 13.4	2/3
ArgR-C	a[b] 114.8 ± 13.7	6		63.1 ± 8.4	3/4

B. *Re-chromatography of ±L-arginine samples*[c]

	Initial gel filtration			Secondary gel filtration			
	Minus arginine			Plus arginine		Minus arginine	
Protein	$M_{w,app}$	µg/ml	n	$M_{w,app}$	n	$M_{w,app}$	n
AhrC	77.6	54.6	4	93.3	6	67.6	4
AhrC-R	41.7	32.0	2/3	120.2	6	21.4	1
				21.4	1		
				66.1	4		
ArgR-C	63.1	163.0	4	44.7	2/3	44.7	2/3
				25.1	1		

  [a] The Table lists the average ($n = 3$), apparent molecular masses ($M_{w,app}$) ± standard deviation (kDa), calculated from an appropriate standard curve, based on size-exclusion data for the four proteins of sub-unit molecular masses 16.7 (AhrC), 16.5 (ArgR), 17.2 (AhrC-R) and 16.7 (ArgR-C) kDa, respectively. Also shown is the most likely stoichiometry (n) for each species.
  [b] a and b refer to the two ArgR-C peaks seen under plus arginine conditions (see Figure 4).
  [c] Data listed are for one set of samples from (a) above. Peaks corresponding to aggregates smaller than hexamer were collected incubated ± 10 mM L-arginine and then re-chromatographed.

in the presence or absence of L-arginine. Each protein was adjusted to 0.5 mg/ml prior to centrifugation at 40,000 revs/minute and 20 °C. Single, sharp boundaries were observed for the ArgR wild-type protein irrespective of whether co-repressor was present. Similar single, sharp boundaries were also observed for the AhrC-R chimera in the presence of L-arginine. The other two solutions showed slightly less well defined boundaries, which could be a sign of heterogeneity, with the ArgR-C chimera being the worse.

The $s_{20,w}$ values for each protein in the presence or absence of 10 mM L-arginine are listed in Table 2. The sedimentation coefficient of AhrC-R in the presence of the L-arginine co-factor was similar to that of AhrC and ArgR, suggesting that this chimera has a similar size and shape to both wild-type proteins. ArgR-C in the presence of L-arginine had a significantly lower sedimentation coefficient compared to the other three proteins, suggesting that it may have a more extended shape (or much greater degree of dissociation). Additionally, the

$s_{20,w}$ value in the absence of co-factor was almost half that obtained in L-arginine-containing buffer, consistent with dissociation of the intact hexameric protein. Assuming that both the hexamer and dissociation product are roughly spherical, the standard Mark-Houwink-Kuhn-Sakurada relation (see Harding, 1995) can be used to predict the molecular mass of the dissociation product, i.e. $s \propto M^{2/3}$, so if $M(\text{hexamer}) = 100,000$, then if $s(\text{hexamer}) = 3.95$ S and $s_{20,w}$ (n-mer) = 2.22 S, this suggests a molecular mass of 42,000 Da, consistent with a trimer or even a dimer. However, smaller structures of this sort would be expected to have different molecular shapes from the hexamer: the data do nonetheless reflect the expected stability of the chimeric proteins based on the size-exclusion chromatography data.

Sedimentation equilibrium measurements

Sedimentation equilibrium measurements were then performed to obtain a more accurate estimate

Table 2. Sedimentation coefficients of AhrC, ArgR and their chimeras

	Plus L-arginine		Minus L-arginine	
	$s_{T,b}$ (S)	$s_{20,w}$ (S)	$s_{T,b}$ (S)	$s_{20,w}$ (S)
AhrC	4.5 ± 0.1	4.9 ± 0.1	4.3 ± 0.1	4.7 ± 0.1
ArgR	4.8 ± 0.1	5.2 ± 0.1	4.3 ± 0.1	4.7 ± 0.1
AhrC-R	4.3 ± 0.1	4.7 ± 0.1	n.d.	n.d.
ArgR-C	3.9 ± 0.3	4.2 ± 0.3	2.2 ± 0.1	2.4 ± 0.1

  Sedimentation velocity experiments were carried out as described in the text, in the presence (10 mM) or absence of L-arginine. The average distance between each scan was used to determine the s values listed; units = Svedberg (S): 1 S $\equiv 1 \times 10^{-13}$ s. n.d. indicates the sample was not run under those conditions or else the results were too noisy to analyze.

Table 3. Summary of sedimentation equilibrium data

Protein	Concn (mg/ml)	$M_{w,app}$ (+L-Arg) (kDa)	$M_{w,app}$ (−L-Arg) (kDa)
AhrC	0.80	122 ± 20	92 ± 2
	0.60	101 ± 2	99 ± 2
	0.50	65 ± 20	53 ± 10
	0.25	66 ± 3	64 ± 5[a]
ArgR	1.00	94 ± 8	85 ± 2
	0.80	103 ± 7	81 ± 2
	0.60	n.d.	60 ± 3
	0.50	63 ± 5[a]	n.d.
AhrC-R	1.00	94 ± 5	77 ± 5
	0.80	115 ± 5	78 ± 2
	0.60	60 ± 5	48 ± 3
ArgR-C	1.00	89 ± 2	56 ± 2
	0.80	105 ± 5	61 ± 3
	0.60	40 ± 3[b]	n.d.
	0.44	37 ± 6[b]	70 ± 6

Apparent mass average molecular masses of protein samples using data obtained at 12000 revs/min, 4 °C ([a] = 15,000 revs/min; [b] = 9000 revs/min). Apparent molecular masses are expressed in kDa. n.d., data were too noisy to analyze due to the presence of 2 mM 2-mercaptoethanol in the buffers.

of the molecular masses of the proteins, avoiding ambiguities due to molecular shape that affect the interpretation of gel filtration and sedimentation velocity measurements. Table 3 summarises the mass average molecular mass behaviour obtained using the data analysis program MSTARA (Harding *et al.*, 1992; Cölfen & Harding, 1997), showing a clear dissociation of the hexameric proteins as the loading concentration is decreased. The background UV noise from the solvent (due to the presence of 2 mM 2-mercaptoethanol) in the concentration distribution data prevented the evaluation of data sets for each loading concentration, *c*, of point mass average molecular mass, $M_{w,app}$ (*r*) as a function of local concentration in the ultracentrifuge cell, *c* (*r*) (Cölfen & Harding, 1997), and restricted the use of "global analysis" procedures. The trend in the whole distribution weight averages, $M_{w,app}$ with loading concentration, *c*, is however, quite clear.

DNA operator recognition by the wild-type and chimeric proteins

There are three different categories of arginine operators present in *B. subtilis* and *E. coli*. The catabolic operators of *B. subtilis* contain only a single ARG box (Miller *et al.*, 1997), whilst the *E. coli* autoregulatory and biosynthetic operators have two ARG boxes (Maas, 1994). The third category of operator has three ARG boxes and includes the *B. subtilis* biosynthetic (Czaplewski *et al.*, 1992) and the *E. coli* argG operators. An operator from each category was used to test the ability of each chimera to recognise specific DNA sequences. Table 4 lists the details of the operators used.

Gel retardation assays

Figure 5 shows representative mobility-shift assays for the *rocABC* operator with ArgR and the two chimeric proteins. The binding of AhrC to this fragment has already been reported (Klingel *et al.*,

Table 4. Operator fragments used and their apparent affinities for the repressor proteins

Operator	*rocABC*		*argR*		*argC*	
Species	*B. subtilis*		*E. coli*		*B. subtilis*	
Size (bp)	277		180		276	
Plasmid source	pABC277		pX3		pUL3000	
No. ARG boxes	1		2		3	
Operator type	Catabolic		Autoregulatory		Biosynthetic	
$K_{d'}$ (nM)	Plus L-arginine	Minus L-arginine	Plus L-arginine	Minus L-arginine	Plus L-arginine	Minus L-arginine
AhrC-R	40	100	10	80	10	75
AhrC	124[a]	-	187[b]	-	7.3[a]	-
ArgR-C	170	-	150	150	140	140
ArgR	120	-	23[b]	-	89[b]	-

K_d' values were estimated by eye as the point at which 50 % of the input DNA had been retarded. −, No retardation was observed.
[a] Data are taken from Miller (1996) and have been calculated by curve fitting the data from multiple binding assays, as described by Miller *et al.* (1997).
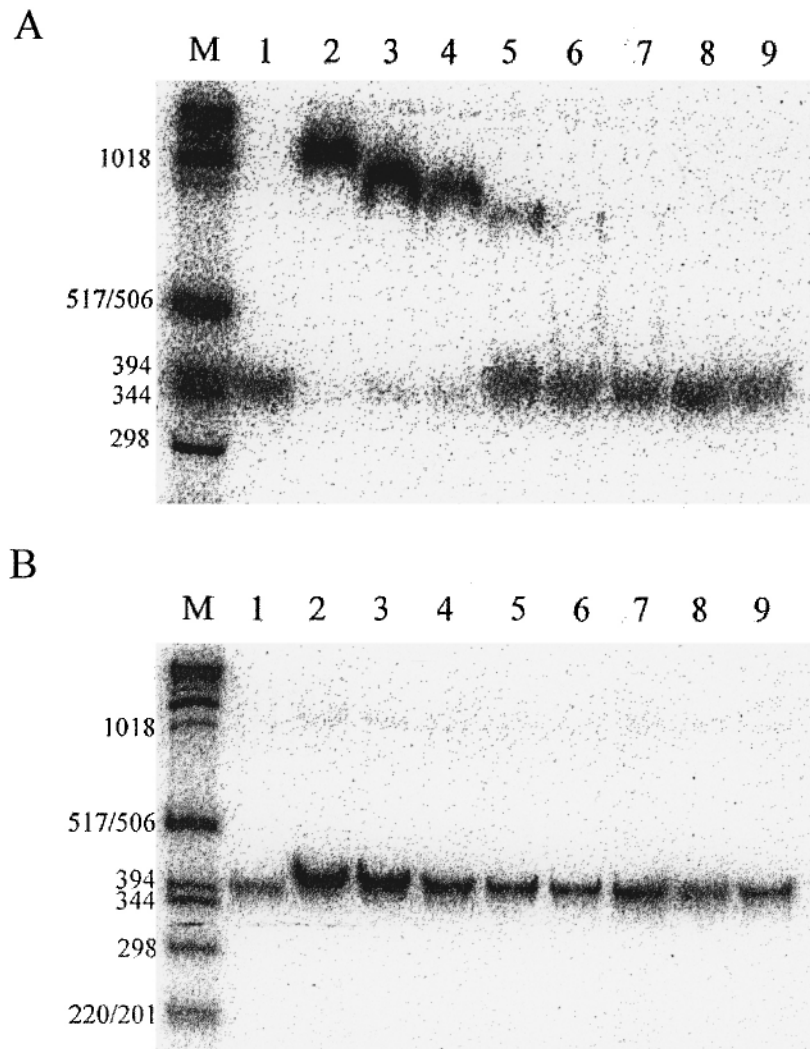[b] Data taken from Smith *et al.* (1989).

Figure 5. (*Legend shown on page 719*)

1995). For ArgR, there was a clear L-arginine-dependent retardation, yielding a species whose mobility was protein concentration dependent. The mobility of this species at the highest concentration tested was slightly lower than that of the complex formed with AhrC, perhaps suggesting that the complexes have somewhat different DNA architectures. Under the conditions of these assays, with protein in large molar excess over the target DNA, the point at which 50% of the input DNA is retarded approximates to the apparent equilibrium dissociation constant, K_d'. For ArgR binding to this *B. subtilis* operator, the $K_d' \sim 120$ nM (hexamer), is similar to the affinity of the homologous protein (K_d' for AhrC:*rocABC* ~124 nM hexamer), consistent with the known cross-functioning of these proteins for operator recognition (Smith *et al.*, 1989).

Similar experiments with the chimeras gave distinct recognition patterns. In the presence of L-arginine, AhrC-R apparently bound the *rocABC* operator with roughly three times the affinity of either wild-type protein, although the mobility of

the retarded species was markedly dependent on protein concentration. ArgR-C had an affinity somewhat lower than that of either wild-type protein for this fragment and produced a retarded species with greater mobility (Table 4). Remarkably, AhrC-R was also able to bind tightly to the operator in the absence of co-factor (AhrC does not bind at all in these conditions).

These properties were even more dramatically apparent in the binding of the chimeras to the two ARG box *argR* operator from *E. coli* (Figure 5(e) and (f)). The AhrC-R chimera binds very tightly to this operator, even showing higher affinity for the site in the absence of co-repressor than AhrC does in its presence. The affinity of ArgR-C is significantly lower and is similar to that of AhrC for this site (Table 4). However, it also shows significant binding in the absence of L-arginine, forming a retarded species with a significantly greater mobility relative to that formed with co-repressor.

In contrast, the binding of the chimeras to the *argC* fragment is more in line with expectations
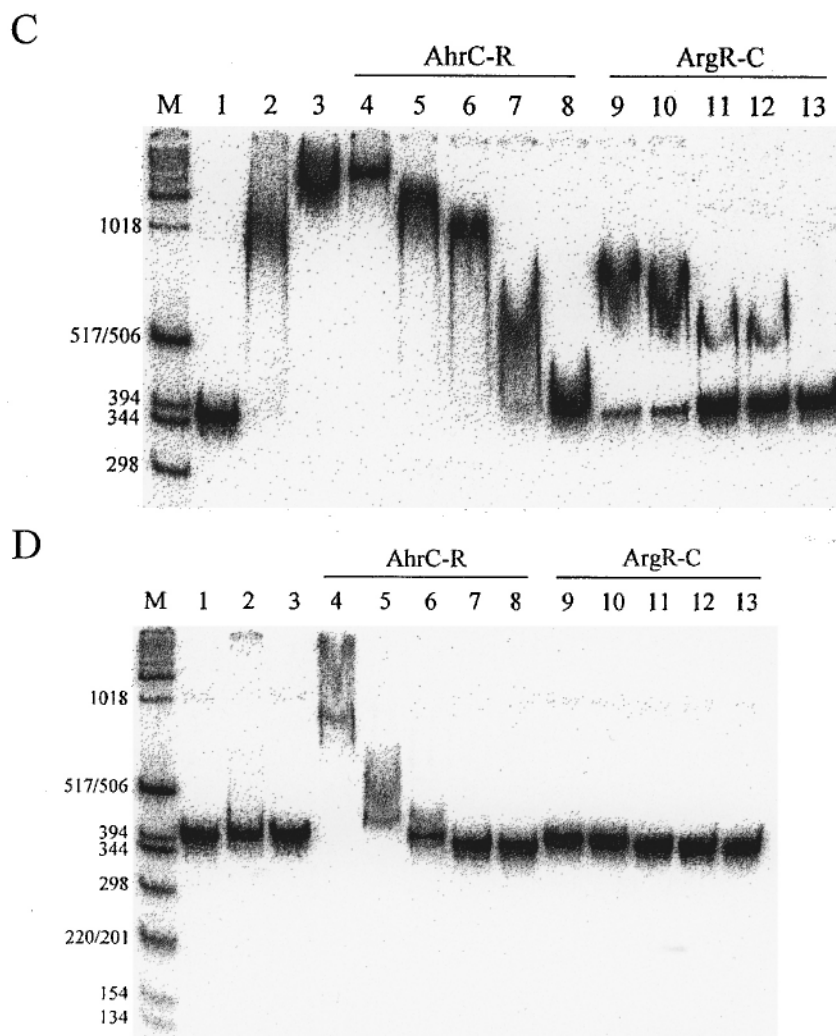
Figure 5. (*Legend shown on page 719*)

DNase I footprinting of the chimeras

In order to confirm the DNA-binding specificities of the chimeric proteins, DNase I footprinting experiments were performed (data not shown). Where the region of DNA protected was not obvious by eye, i.e. for the lower affinity cases, densitometry was used to define the footprint boundaries. The results are summarised in Figure 6. In most cases, the chimeric proteins protected the expected operator regions to within a few base-pairs of the sites defined by the native repressors, showing that they retained the ability to bind in a sequence-specific fashion. Although the footprint for AhrC-R binding to *argR* was identical with or without co-repressor, it was not possible to detect a footprint for ArgR-C under these conditions. No footprint for ArgR-C binding to *argC* was detect-

able under any conditions at the protein concentrations used.

DNA-mediated assembly of repression complexes

The affinities of the wild-type and chimeric proteins for the operator sites *in vitro*, together with their hydrodynamic behaviour, suggest that trimers or the other smaller molecular mass species below hexamer size would be the dominant form of the protein in solution able to bind to DNA. This is consistent with the reports by Burke *et al.* (1994) and Chen *et al.* (1997) that trimers of an ArgR mutant are capable of binding operator. One consequence of this is that operator binding appears to stimulate hexamer formation. A further curious effect of this pathway of formation of the repressor-operator complex is that these systems appear to show anti-competition behaviour in gel retardation assays, as we have already reported for AhrC (Miller *et al.*, 1997). We have proposed that
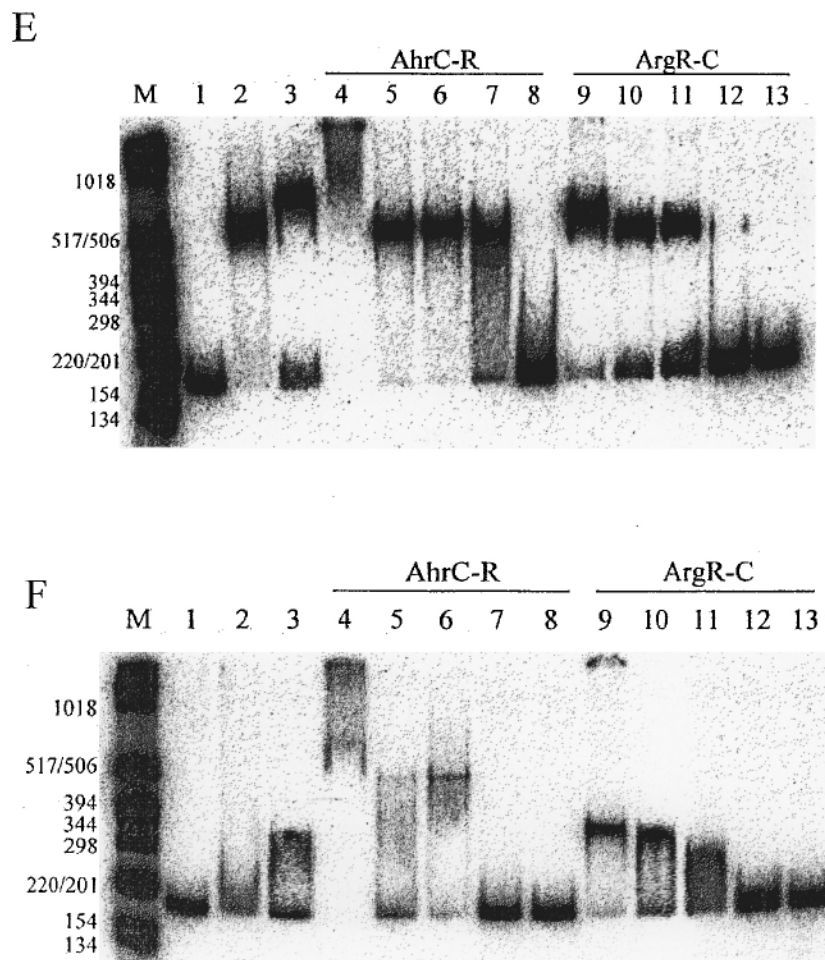
Figure 5. (*Legend shown on page 719*)

the repression complexes form from trimers, one of which must be transiently bound to the operator. The trimers can then assemble into hexamers which are more stable on the DNA (Figure 7), even in the absence of a second ARG box.

Dissociation of the hexameric complex is not simply the reverse of assembly, rather the hexameric proteins appear capable of dissociation directly from the DNA, resulting in an unbound, high-affinity form of the protein. When standard gel retardation assays, such as those shown in Figure 5, are challenged with unlabelled, specific DNA fragments, the paradoxical outcome is apparently an increased affinity for the radioactive fragments. We believe that this is due to the effect of DNA-mediated increases in the free hexamer concentration. Eventually, however, this effect reaches a plateau, presumably when all the available protein exists as hexamer, and normal competitive effects are seen. Note that simple mass action effects would not account for this behaviour.

In order to assess whether operator binding occurs *via* the same mechanism with both the chimeras and wild-type ArgR, similar gel retardation assays were carried out here. The results are

shown in Figure 8. In each case, the addition of molar excesses of cold competitor DNA fragments carrying operator sites to the binding reactions increased the apparent affinity of the proteins for the operator, suggesting that they were all binding the operators in the same way as AhrC binds to the catabolic site. In some cases, the mobility of the retarded species was also altered, suggesting that the high affinity form of the protein might result in dramatically different architectures for the DNA complexes being generated.

Discussion

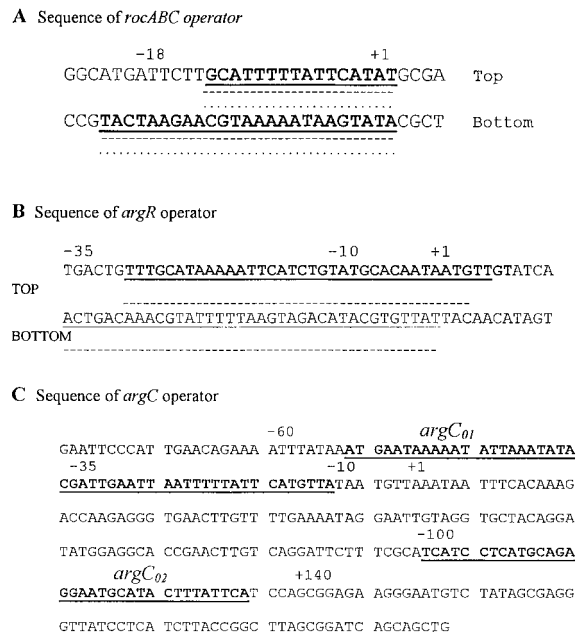### Characterisation of the chimeras

We have shown that N and C-terminal chimeras of ArgR and AhrC are stable and can be expressed to high levels in *E. coli*. Both chimeras showed solubility properties closer to wild-type AhrC than to ArgR and were therefore purified using a protocol based on that devised previously for the *Bacillus* protein (Czaplewski *et al.*, 1992). Gel filtration analysis suggested that quaternary structures of the chimeras were similar, each protein appearing
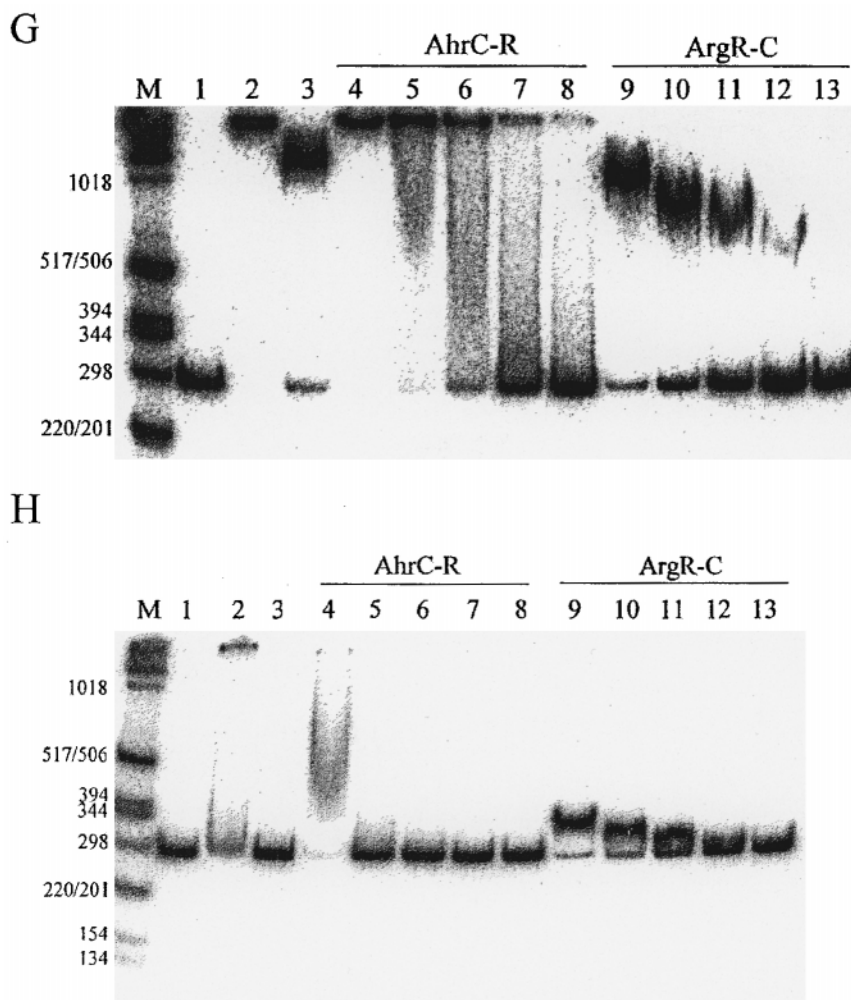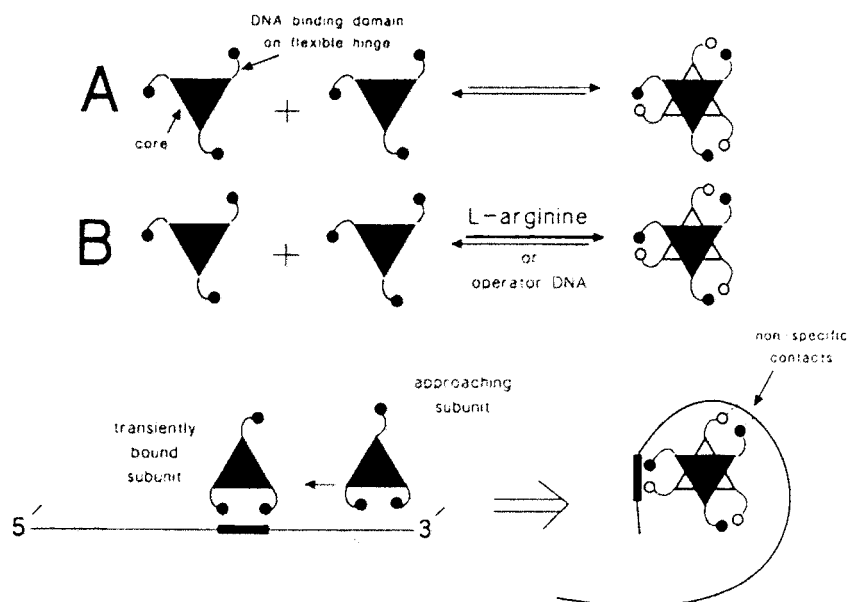
Figure 5. Gel shift assays. Radiolabelled operator DNAs (<pM) were incubated in 1 × FB binding buffer with differing concentrations of the various proteins, in the (a), (c), (e), (g) presence or (b), (d), (f), (h) absence of 10 mM L-arginine for 20 minutes, 37 °C, before being loaded on a 6 % (w/v, polyacrylamide (29:1, w/w, acrylamide:bisacrylamide), 1 × TB gel, and electrophoresed for three to four hours at a constant 30 mA. Protein concentrations are given for hexamers and DNA molecular size markers are labelled as bp alongside the marker lanes (M), both here and in Figure 8. (a), (b) Binding of ArgR to the *rocABC* operator. Lanes 1, DNA only control; 2-9, ArgR at 300 nM, 210 nM, 147 nM, 103 nM, 72 nM, 50 nM, 35 nM and 25 nM, respectively. (c), (d) Binding of AhrC-R and ArgR-C to the *rocABC* operator. Lanes 1, DNA only control; 2, 144 nM AhrC; 3, 347 nM ArgR; 4-8, AhrC-R at 300 nM, 120 nM, 84 nM, 58 nM and 23.3 nM, respectively; 9-13, ArgR-C at 241 nM, 200 nM, 140 nM, 116 nM and 47 nM, respectively. (e), (f) Binding of AhrC-R and ArgR-C to the *argR* operator. Lanes 1, DNA only control; 2, 112 nM AhrC; 3, 200 nM ArgR; 4-8, AhrC-R at 350 nM, 81 nm, 98 nm, 19 nM and 5 nM, respectively; 9-13, ArgR-C at 500 nM, 289 nM, 201 nM, 97 nM and 56 nM, respectively. Note: due to a loading error the concentration in lanes 6 is higher than that in lanes 5. (g), (h) Binding of AhrC-R and ArgR-C to the *argC* operator. Lanes 1, DNA only control; 2, 83 nM AhrC; 3, 200 nM ArgR; 4-8, AhrC-R at 100 nM, 46 nM, 19.4 nM, 11 nM and 8 nM, respectively; 9-13, ArgR-C at 347 nM, 241 nM, 167 nM, 116 nM and 81 nM, respectively. Note, the target fragment contains three ARG boxes, which constitute two binding sites for AhrC consistent with the decreased mobility of the native protein and the AhrC-R chimera with respect to ArgR.

capable of forming hexamers at high concentration in the presence of L-arginine. In the absence of L-arginine, all the proteins except ArgR underwent dissociation into smaller subunits, with apparent molecular masses of the dissociation products varying from the equivalent of a dimer (AhrC-R) to a tetramer (AhrC) (Table 1A). Re-chromatography showed that this dissociation was reversible at least for AhrC and AhrC-R. These results were not consistent with expectations based on the X-ray crystal structure of the ArgR C-terminal domain, which shows the hexamer to consist of a dimer of trimeric units (Van Duyne *et al.*, 1996) or experiments which suggest that trimers of ArgR mutants can bind operator DNA (Burke *et al.*, 1994; Chen *et al.*, 1997). Since gel filtration is a non-absolute probe of molecular mass, dependent on assumptions of identical molecular shape for the calibration of standards, we carried out detailed and absolute analytical ultracentrifugation exper-

A Sequence of *rocABC operator*

```
        -18                        +1
    GGCATGATTCTTGCATTTTTATTCATATGCGA   Top
            ..............................
    CCGTACTAAGAACGTAAAAATAAGTATACGCT   Bottom
            ..............................
```

B Sequence of *argR operator*

```
    -35                           -10       +1
    TGACTGTTTGCATAAAAATTCATCTGTATGCACAATAATGTTGTATCA
TOP
            -------------------------------------------------
    ACTGACAAACGTATTTTTAAGTAGACATACGTGTTATTACAACATAGT
BOTTOM  -------------------------------------------------
```

C Sequence of *argC operator*

```
                            -60          argC01
    GAATTCCCAT TGAACAGAAA ATTTATAAAT GAATAAAAAT ATTAAATATA
    -35                   -10        +1
    CGATTGAATT AATTTTTATT CATGTTATAA TGTTAAATAA TTTCACAAAG

    ACCAAGAGGG TGAACTTGTT TTGAAAATAG GAATTGTAGG TGCTACAGGA
                                                -100
    TATGGAGGCA CCGAACTTGT CAGGATTCTT TCGCATCATC CTCATGCAGA
          argC02                +140
    GGAATGCATA CTTTATTCAT CCAGCGGAGA AGGGAATGTC TATAGCGAGG

    GTTATCCTCA TCTTACCGGC TTAGCGGATC AGCAGCTG
```

Figure 6. Summary of the DNase I footprinting data. Data are shown for footprinting reactions carried out in the presence of 10 mM L-arginine. Continuous lines indicate the AhrC-R binding sites; broken lines the ArgR-C sites; dotted lines the ArgR sites and bases in bold indicate the AhrC sites. The protected bases on the top strand of *argR* are identical with all but the most 3' bases of the published ArgR site (Tian *et al.*, 1992).

iments to investigate the hydrodynamic behaviour of these proteins.

The hydrodynamic properties of the repressors and their chimeras

The results from the gel filtration studies (Table 1), the sedimentation velocity (Table 2) and sedimentation equilibrium (Table 3) experiments all clearly point to dissociative behaviour with decreasing concentration of the wild-type AhrC and ArgR protein hexamers and their chimeras, both in the presence and absence of L-arginine in the solvent. The stoichiometry of the dissociation equilibrium in the concentration range studied (between 0.2 - 1.0 mg/ml) cannot be unequivocally determined (between hexamer-trimer and hexamer-dimer) by the solution data alone, although there is some evidence for "bottoming out of the data" at the trimer (~50 kDa). An exception is the ArgR-C chimera with L-arginine present in the solvent, which shows a molecular mass approaching that expected for a dimer (33.6 kDa).

We know, however, from the X-ray crystal structure of the C-terminal domain of ArgR and from solution exchange experiments with truncated subunits (Van Duyne *et al.*, 1996) that the most likely form of the dissociation is that of hexamer-trimer. It would therefore be instructive to interpret the sedimentation equilibrium concentration distribution for the wild-type AhrC and ArgR systems
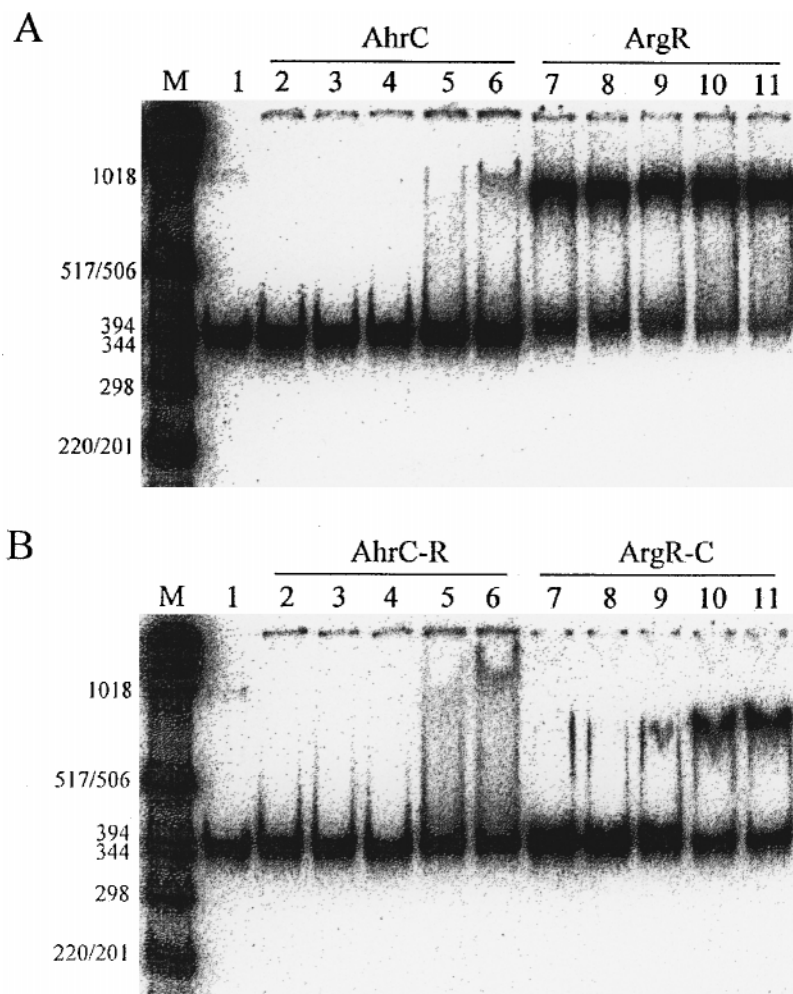

Figure 7. Model for operator-mediated assembly of high-affinity forms of the repressors. Each AhrC is shown as an initial subunit smaller than a hexamer, a trimer in this example, consisting of a multimerisation C-terminal core supporting a monomeric DNA-binding domain on a flexible hinge. Hexamerisation (reactions A and B) is favoured by increasing protein concentration or by the binding of L-arginine, which is known to bind in the interface between trimers (Van Duyne *et al.*, 1996). A plausible model for the DNA-mediated multimerisation detected is shown in reaction C. A transiently bound, low-affinity, trimer is bound at an operator site. A second, non-specifically bound trimer then collides with this complex and undergoes multimerisation, allowing DNA binding domains from the two trimers to interact with the DNA, resulting in a more symmetric (more stable?) arrangement of the protein hexamer.

Figure 8. (*Legend shown on page 723*)

and the AhrC-R chimera in terms of a molar dissociation constant, K_d, for a trimer-hexamer system. This can be done in two ways; firstly, by model fitting the absorbance *versus* radial position in the ultracentrifuge cell (Figure 9; for a given loading concentration chosen so that there is a mixture of both hexamers and trimers present) using the routine ASSOC4 provided by the ultracentrifuge manufacturer (see McRorie & Voelker, 1993). This software package essentially fits the data to equation (III-27) of Kim *et al.* (1977), and we use it here with the trimer as an "effective monomer" and the hexamer as an "effective dimer". ASSOC4 yields the association constant, k_{a2}, (in absorbance units^{-1}) as the primary parameter. This is converted to grams per litre using an extinction co-efficient calculated from the amino acid sequence (see Perkins, 1986), and then to a molar dissociation constant from the relationship, $K_d = 2/(M_1 \times k_{a2})$ (Williams, 1973), where in this case M_1 is the trimer molecular mass. For noisier data sets this analysis is not useful, and for these cases k_{a2} and K_d can be estimated using an alternative procedure from the molecular mass data of Table 3 and the conservation of mass equation: $c = c_1 + k_{a2}c_1^2$ (see Kim *et al.*,

1977), where c is the total cell loading concentration, c_1 in this case is the concentration of trimer, determined from $c_1 = [(M_2 - M_w)/(M_2 - M_1)]c$, with M_2 the molecular mass of the hexamer. Table 5 summarises the resultant K_d values. In all cases the concentrations are assumed small enough so that thermodynamic non-ideality can be neglected.

We can see from these data that the strength of the association is moderate to weak (see Silkowski *et al.*, 1997) with a K_d value between 10^{-5} and 10^{-4} M in all cases. The absence of L-arginine appears to have no major effect on the strength of the association, perhaps because the charged co-repressor binding sites are screened at the ionic strength of these experiments. Wild-type AhrC trimers appear to have a higher affinity for each other compared with both the AhrC-R chimera and the ArgR protein. The fact that the chimera's affinity is distinct from that of the parent C-terminal domain protein (ArgR) suggests a possible interaction between N and C-terminal domains. The X-ray crystal structure of intact apo-ArgR from *Bacillus stearothermophilus* (G.D. Van Duyne, personal communication) supports this
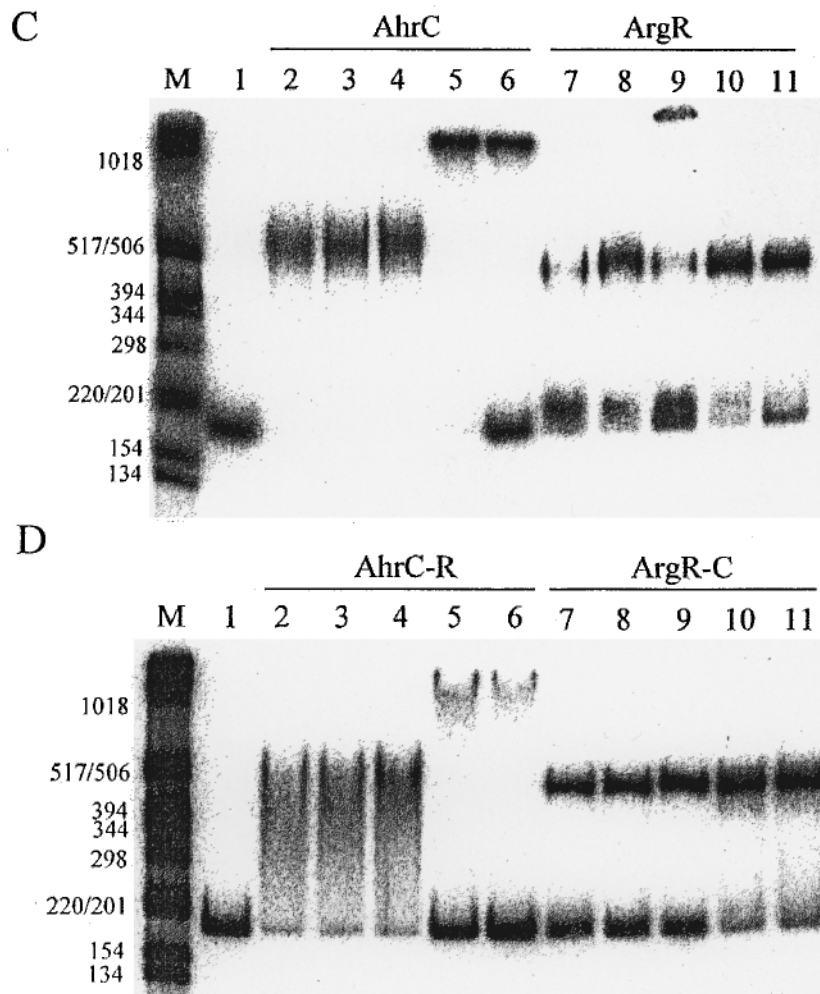
Figure 8. (*Legend shown on page 723*)

interpretation: the N-terminal DNA-binding domain of one trimer makes protein-protein inter-actions with the C-terminal oligomerisation domain of the neighbouring trimer. From Table 3, the monomer-trimer equilibrium is evidently much stronger, with little evidence of dissociation in the concentration range studied with the exception of the ArgR-C chimera, which appears to form dimers. Such dimers could be the result of the ''top-bottom'' contacts seen in the *B. stearothermophilus* ArgR structure and are consistent with a reported dimeric ArgR mutant, Gly123Asp (Tian & Maas, 1994). Probing the equilibrium below trimer level would require the ability to investigate the system at much lower protein concentrations than is possible with our current instruments.

Characterisation of operator recognition by the chimeras

The data described above suggest that at low concentrations, such as those used routinely in DNA-binding assays, all the proteins consist of the smaller species, mostly trimers, and that hexamer formation is facilitated by DNA-binding as pro-

posed for AhrC (Miller *et al.*, 1997). Complexes in the gel assays showed one of three mobilities: a species equivalent to a ~600 bp DNA marker frag-ment; a species equivalent to a ~1000 bp fragment; and a species which did not enter the gel. In some cases the mobility of the complex was greater at lower protein concentrations, suggesting that dis-sociation was occurring during the experiment. The two types of complex within the gel are sug-gestive of differing stoichiometries, especially in view of the hydrodynamic properties of these pro-teins, but may simply reflect differing degrees of DNA bending induced by protein binding. In gen-eral, complexes with proteins encompassing the AhrC N-terminal domain produced complexes of reduced mobility, especially with the *argC* oper-ator, which contains three ARG boxes and is unable to enter the gel, consistent with the binding of multiple repressors at $argC_{O1}$ and $argC_{O2}$ as pro-posed previously (Czaplewski *et al.*, 1992). Proteins encompassing the ArgR N-terminal domain pro-duce complexes with higher mobilities and it is noticeable that the ArgR-C protein always pro-duces a complex with the same mobility (equival-
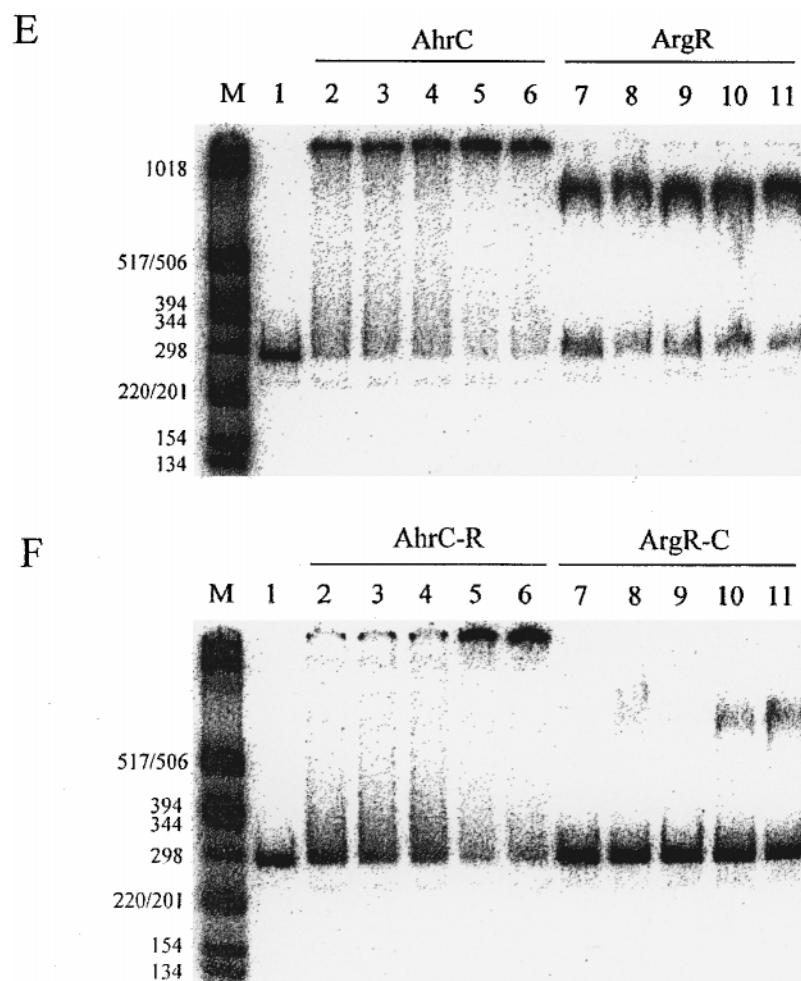
Figure 8. Competition gel mobility assays. Lanes 1, no protein control; 2 and 7, no unlabelled DNA added; 3 and 8, 25-fold molar excess unlabelled pUC DNA; 4 and 9, 50-fold molar excess unlabelled pUC DNA; 5 and 10, 25-fold molar excess unlabelled *rocABC* operator; 6 and 11, 50-fold molar excess unlabelled *rocABC* operator. Competition with the *rocABC* operator: (a) lanes 2-6, 90 nM AhrC; 7-11, 70 nM ArgR; (b) lanes 2-6, 70 nM AhrC-R; 7-11, 80 nM ArgR-C. Competition with the *argR* operator: (c) lanes 2-6, 65 nM AhrC; 7-11, 100 nM ArgR; (d) lanes 2-6, 10 nM AhrC-R; 7-11, 140 nM ArgR-C. Competition with the *argC* operator: (e) lanes 2-6, 25 nM AhrC; 7-11, 125 nM ArgR; (f) lanes: 2-6, 25 nM AhrC-R; 7-11, 125 nM ArgR-C.

ent to the 600 bp fragment) whatever the source of the DNA target. The competition experiments reveal even more complex behaviour and in some cases, such as for AhrC and AhrC-R binding to the *argR* target, the mobility of the complex changes dramatically whilst it remains constant for those proteins having an N-terminal domain from ArgR. The apparent anti-competition effect for AhrC has been reported previously (Miller *et al.*, 1997) and is confirmed here for ArgR and the chimeras (Figure 8). In principle, this behaviour could also arise from an association of monomers or dimers (cf. the ArgR-C protein) as well as the trimers shown in Figure 7, but since we are unable to investigate these low concentrations directly by ultracentrifugation this is purely speculative. The stoichiometry of the complex of ArgR bound to the *argR* target has been measured directly and shown to be equivalent to one hexamer per two ARG boxes (Tian *et al.*, 1992), although trimeric mutant

repressors can also bind (Chen *et al.*, 1997). Investigations of *argR*/ArgR complex formation using surface plasmon resonance (SPR) assays, however, are at least consistent with a stoichiometry of one trimer per operator fragment, whereas similar assays with AhrC suggest hexamers are being bound (Stockley *et al.*, 1998).

The gel retardation assays also reveal one unexpected property of both chimeras, namely that they appear to be able to bind operators in the absence of L-arginine. The one exception being ArgR-C with the *rocABC* target, which only contains a single ARG box site. DNase footprinting confirms that the sequence-specificity of the interaction with DNA has not been affected by formation of the chimeras. Indeed, the AhrC-R chimera appears to have a significantly higher affinity for all the operators than the wild-type proteins, with the exception of the *argC* target where it is almost identical with the AhrC affinity. These

Table 5. Estimates for the molar dissociation constants for a hexamer-trimer equilibrium for the AhrC, ArgR and AhrC-R proteins from sedimentation analysis

	c (mg/ml)	k_{a2} (absorbance^{-1})	ε_{280} (l g^{-1} cm^{-1})	k_{a2} (l/g)	K_d (μM)
AhrC (+Arg)	0.5	4.9 ± 2.7	0.586	3.5 ± 1.9[a]	11 ± 6
AhrC (−Arg)	0.5	3.9 ± 1.4	0.586	2.7 ± 1.0[a]	15 ± 6
ArgR (+Arg)	0.5	3.9 ± 1.0	0.240	1.1 ± 0.3[a]	36 ± 10
ArgR (−Arg)	0.6; 0.8; 1.0	-	0.242[c]	0.7 ± 0.3[b]	55 ± 25
AhrC-R (+Arg)	0.6	-	?	0.6 ± 0.3[b]	65 ± 35
AhrC-R (−Arg)	0.8; 1.0	-	?	1.1 ± 0.5[b]	35 ± 15

Values were calculated as described in the text.
[a] From modeling the absorbance as a function of radial position for a given loading concentration.
[b] Estimated from the molecular mass data of Table 3.
[c] Not used.

data for operator recognition can not simply be due to altered multimerisation of the chimeras since the solution data described above clearly indicate that the hexamer forms of both chimeras are less stable than the wild-type proteins in the absence of L-arginine. Rather, the data are consistent with an allosteric activation of the DNA-binding domains, although other explanations based on the quaternary structure of the chimeras could be possible. In the wild-type proteins this is presumably coupled to L-arginine binding and multimerisation, but the chimeras appear to have been

(b)

(a)

Figure 9. Analysis of the sedimentation equilibrium data. Sedimentation equilibrium ASSOC4 fit to the solute concentrations distribution data for AhrC protein in the absence of L-arginine in the solvent. The line fitted in (a) corresponds to a dissociation constant (hexamer-trimer), K_d 15($\pm$6) μM. (b) The plot of the residuals from this fit. Loading concentration, $c = 0.5$ mg/ml; UV absorbance was recorded at 278 nm.

"locked" in a conformation closer to the higher affinity form, hence they can bind with high affinity in the absence of co-repressor. This is consistent with the arguments by Sunnerhagen *et al.* (1997) that there is reasonably tight coupling between the N and C-terminal domains of ArgR.

In summary, we have established a remarkable conservation of the three-dimensional structure and function of these two transcriptional regulatory proteins from very divergent bacterial sources. The data suggest that arginine repressors bind target operator sites *via* a mechanism that possibly involves allosteric activation and a protein-protein multimerisation reaction(s), which is dependent on protein concentration, L-arginine concentration and operator DNA binding, making them examples of one the most complicated bacterial repressor systems yet characterised.
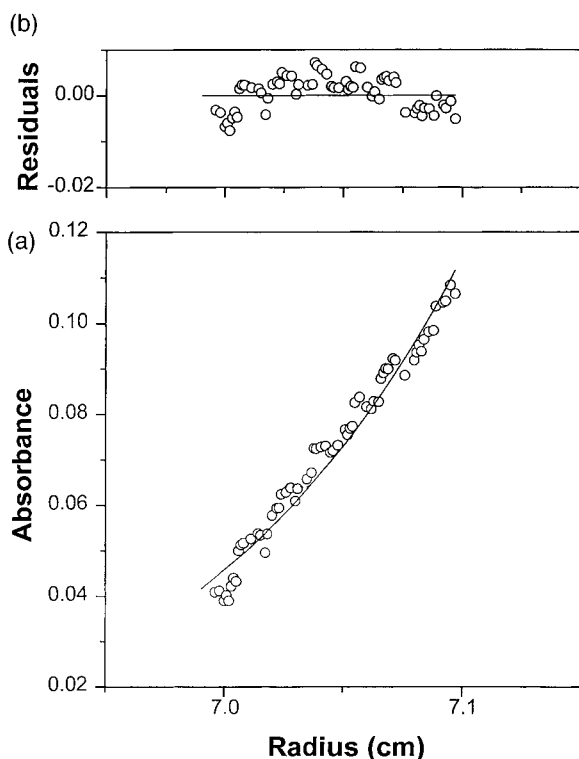
Materials and Methods

Protein purification

Over-expression of cloned genes

AhrC, ArgR, AhrC-R and ArgR-C were all expressed in a phage T7 promoter expression vector (pET22b) in *E. coli* BL21(DE3) cells, which contain an inducible T7 polymerase gene (Studier & Moffatt, 1990). Expression of the cloned genes was induced by the addition of IPTG (to a final concentration of 1 mM) when the culture A_{600} reached 0.6-1.03. Incubation was continued for a further two to three hours before chilling and pelleting the cells. The cell pellets were then stored at 4 °C overnight.

Purification of AhrC and the chimeric protein AhrC-R

The cell pellet was resuspended in 150 ml of ice-cold Arg buffer (20 mM Tris-HCl (pH 7.5), 10 mM MgCl$_2$, 2 mM 2-mercaptoethanol, 1 mM PMSF, 25 μM TPCK). The 30 ml sample was sonicated (two, 30 second periods) in an ethanol-ice bath, followed by centrifugation at 15,000 *g* for 30 minutes at 4 °C. The supernatant was discarded whilst the pellet was resuspended in 100 ml of ice-cold Arg buffer. The insoluble protein suspension was treated with 5 μg/ml DNase I at 37 °C for 30 minutes. Solid NaCl (to a final concentration of 0.5 M) was added and the protein solution incubated for another 30 minutes, before being centrifuged at 15,000 *g* for 30 minutes at 4 °C. The pellet was discarded and the supernatant was dialysed against two litres of Arg buffer

plus 75 mM NaCl for three hours. The precipitate was collected by centrifugation at 10,000 *g* for 30 minutes at 4 °C; the supernatant was discarded and the precipitated protein resuspended in Arg buffer plus 600 mM NaCl before dialysis overnight against Arg buffer plus 300 mM NaCl. The proteins were then purified on an S-Sepharose (Pharmacia) ion-exchange column equilibrated with Arg buffer plus 300 mM NaCl. The column was developed with a 300 mM-600 mM NaCl gradient in 800 ml of Arg buffer.

Purification of the chimeric protein, ArgR-C

The purification of ArgR-C was carried out as described above, except that ArgR-C remains soluble in 75 mM NaCl and was therefore dialysed against Arg buffer plus 75 mM NaCl for three hours, and any precipitated material discarded before dialysis against Arg buffer plus 300 mM NaCl overnight and the final ion-exchange chromatography step.

Purification of ArgR

ArgR was purified using a slight modification of the procedure described by Lim *et al.* (1987).

Characterisation and storage of the purified proteins

Samples of protein at different stages in the purification were analysed by SDS-PAGE (Figure 3). The concentration of pooled fractions was determined by measuring the A_{280} of each sample. Using the amino acid sequence to predict the extinction coefficients for each protein (Gill & von Hippel, 1989) the following values were calculated for a 1 mg/ml solution, A_{280} of 0.586 for AhrC, 0.24 for ArgR, 0.376 for AhrC-R and 0.429 for ArgR-C. Proteins were pooled and stored in column buffer plus 20 % (v/v) glycerol at -70 °C for up to a year.

Analysis of the oligomeric state of the proteins

Size-exclusion chromatography

Estimated molecular masses of the four proteins were determined using a Superdex-200 sizing column (Pharmacia) connected to a Pharmacia FPLC system. Glycerol was removed from the samples prior to use by gel filtration into 20 mM Tris-HCl (pH 7.5), 10 mM $MgC1_2$, 350 mM NaCl, $\pm$ 10 mM L-arginine. The 200 μl sample was injected onto the Superdex column, which was run at 0.5 ml/minute and 0.5 ml fractions were collected. A standard curve was produced with the following as molecular mass standards: blue dextran (2000 kDa); P-amylase (200 kDa); alcohol dehydrogenase (150 kDa); bovine serum albumun (66 kDa); carbonic anhydrase (29 kDa) and cytochrome *c* (12.4 kDa). There was no change in elution volumes when L-arginine was present.

Ultracentrifugation analysis

Samples were prepared from frozen stocks as described above and stored at 4 °C for up to a month. The partial specific volume ($\bar{v}$) of the wild-type and chimeric proteins was estimated from the amino acid sequence using the ''consensus'' formula described by Perkins (1986) (see also Laue *et al.*, 1992).

Sedimentation velocity and sedimentation equilibrium experiments were carried out using a Beckman Optima XL-A analytical ultracentrifuge (Beckman, Palo Alto, CA) fitted with a scanning absorption optical system. In order to determine the sedimentation coefficients of the wild-type and chimeric proteins, all proteins were diluted with the aqueous solvent described above to give a final concentration of 0.5 mg/ml. Double sector cells were filled with 280 μl of protein solution and 300 μl of reference solvent (buffer) in their respective channels. Solutions were run at 40,000 revs/minute at 20.0 °C and solute distributions scanned at 280 nm; because of noise due to the presence of 2-mercaptoethanol in the solvent, some scans were also recorded at 230 nm. A series of ten scans were taken at fixed time intervals. Scans were analysed using the routine XLAVEL (Cölfen *et al.*, 1996), giving the sedimentation coefficient, $s_{T,b}$ for each protein, where the subscript T,b defines the temperature and buffer conditions, the standard error and a radial dilution correction to the loading concentration. The $s_{T,b}$ values were corrected to standard conditions of solvent density, ρ, and viscosity, η, i.e. that of water at 20.0 °C, according to Tanford (1961) to give $s_{20,w}$ values:

$$s_{20,w} = \{(1 - \bar{v}\rho)_{20,w}/(1 - \bar{v}\rho)_{T,b}\}\{\eta_{T,b}/\eta_{20,w}\}\, s_{T,b}$$

Molecular mass determinations of the proteins were carried out using initial loading concentrations between 0.2 and 1.0 mg/ml for each protein. An 80 μl sample of protein solution and 100 μl of reference solvent were loaded into their respective channels. Since the sedimentation equilibrium experiments were of significantly longer duration ($\sim$24 hours) than the sedimentation velocity measurements, they were performed at 4 °C. Three different rotor speeds (9000, 12,000, and 15,000 revs/minute) were employed. Data were analysed initially using the MSTARA program (Cölfen and Harding, 1997). This procedure provides an estimate for the apparent mass average molecular mass, $M_{w,app}$ over the whole distribution of protein in the ultracentrifuge cell independently of any assumed model, and is particularly well suited for heterogeneous systems (Creeth & Harding, 1982). At low loading concentrations the assumption can reasonably be made that thermodynamic non-ideality effects are small, and the apparent $M_{w,app}$ is $\sim M_w$, the true or ''ideal'' mass average molecular mass. The data were then further analysed for the characterisation of self-association/dissociation molar dissociation constants using (i) direct analysis (where applicable) of the absorbance *versus* radial displacement profiles in the ultracentrifuge cell, using the routine ASSOC4 (McRorie & Voelker, 1993); (ii) where this is not applicable, from the relation of $M_{w,app}$ with the loading concentration, c, data. Global analysis of combined data sets at differing concentrations and rotor speeds was attempted using the programme NONLIN, but yielded no further information.

Protein:DNA interaction analysis

Gel retardation assays

Operator sites were PCR amplified from appropriate plasmids using 5′ radiolabelled primers. Proteins were diluted serially in $1 \times$ FB (10 mM Tris-HCl (pH 7.4), 5 mM $MgC1_2$, 250 mM KCl, 2.5 mM $CaCl_2$, 0.5 mM DTT, 50 μg/ml sheared calf thymus DNA and 2.5 % (w/v) glycerol) binding buffer; 10 mM L-arginine was added as required. A 15 μl sample of each dilution was added

to the well of a microtitre plate. The labelled operator DNA was also diluted in 1 × FB; a 15 µl sample of DNA solution (corresponding to 5000 counts) was mixed with the diluted protein and the microtitre plate incubated at 37 °C for 20 minutes. Samples were electrophoresed on a 6 % (w/v) polyacrylamide (29:l, w/w, acrylamide:bis-acrylamide), 1 × TB gel containing 5 mM MgC1$_2$ and 10 mM L-arginine as required. The gel was prerun at 30 mA for 20 minutes and the samples loaded directly (without tracking dyes) whilst the gel was running at 300 V; two minutes after loading the current was switched to 30 mA. The running buffer (1 × TB, 5 mM MgC1$_2$, 10 mM L-arginine (if required)) was recirculated.

Specific competitor DNA gel retardation assay procedure

Approximately ~0.03 pmol of radiolabelled ("hot") *rocABC*, *argR* or *argC* operator was mixed with either a 25 or 50-fold molar excess of an unlabelled 300 bp fragment of pUC DNA (PCR amplified from the backbone of pUC13 vector) or the *rocABC*, *argR* or *argC* operator fragment in 1 × FB binding buffer. A known concentration of wild-type or chimaeric protein was then added and the reactions were equilibrated at 37 °C for 20 minutes prior to electrophoresis alongside samples which did not contain the competitor DNA. Protein concentrations were chosen that gave less than 50 % shifting under these conditions so that any effects of the competitor DNA could readily be distinguished.

DNase I footprint analysis

Footprinting procedures were as described (Czaplewski *et al.*, 1992; Miller *et al.*, 1997).

Acknowledgements

References

Boys, C. W. G., Czaplewski, L. G., Phillips, S. E. V., Baumberg, S. & Stockley, P. G. (1990). Crystallization of the arginine-dependent repressor/activator AhrC from *Bacillus subtilis*. *J. Mol. Biol.* **213**, 227-228.

Burke, M., Merican, A. F. & Sherratt, D. J. (1994). Mutant *Escherichia coli* arginine repressor proteins that fail to bind L-arginine, yet retain the ability to bind their normal DNA-binding sites. *Mol. Microbiol.* **13**, 609-618.

Chen, S. H., Merican, A. F. & Sherratt, D. J. (1997). DNA binding of *Escherichia coli* arginine repressor mutants altered in oligomeric state. *Mol. Microbiol.* **24**, 1143-1156.

Calogero, S., Gardan, R., Glaser, P., Schweizer, J., Rapoport, G. & Débarbouillé, M. (1994). RocR, a novel regulatory protein controlling arginine utilization in *Bacillus subtilis*, belongs to the NtrC/NifA family of transcriptional activators. *J. Bacteriol.* **176**, 1234-1241.

Cölfen, H. & Harding, S. E. (1997). MSTARA and MSTARI: interactive PC algorithms for simple, model independent evaluation of sedimentation equilibrium data. *Eur. Biophys. J.* **25**, 333-346.

Cölfen, H., Harding, S. E., Wilson, E. K., Packman, L. C. & Scrutton, N. S. (1996). Homodimeric and expanded behaviour of trimethylamine dehydrogenase in solution at different temperatures. *Eur. Biophys. J.* **24**, 159-164.

Creeth, J. M. & Harding, S. E. (1982). Some observations on a new type of point average molecular weight. *J. Biochem. Biophys. Methods*, **7**, 25-34.

Czaplewski, L. G., North, A. K., Smith, M. C. M., Baumberg, S. & Stockley, P. G. (1992). Purification and initial characterization of AhrC: the regulator of arginine metabolism genes in *Bacillus subtilis*. *Mol. Microbiol.* **6**, 267-275.

Débarbouillé, M., Martin-Verstraete, I., Kunst, F. & Rapoport, G. (1991). The *Bacillus subtilis* sigL gene encodes an equivalent of σ54 from Gram-negative bacteria. *Proc. Natl Acad. Sci. USA,* **88**, 9092-9096.

Gardan, R., Rapoport, G. & Débarbouillé, M. (1995). Expression of the rocDEF operon involved in arginine catabolism in *Bacillus subtilis*. *J. Mol. Biol.* **249**, 843-856.

Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319-326.

Glansdorff, N. (1987). Biosynthesis of arginine and polyamines. In Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (Neidhardt., F. C., ed.), pp. 321-344, American Society for Microbiology, Washington, DC.

Grandori, R., Lavoie, T. A., Pflumm, M., Tian, G., Niersbach, H., Maas, W. K., Fairman, R. & Carey, J. (1995). The DNA-binding domain of the hexameric arginine repressor. *J. Mol. Biol.* **254**, 150-162.

Harding, S. E. (1995). On the hydrodynamic analysis of macromolecular conformation. *Biophys. Chem.* **55**, 69-93.

Harding, S. E., Horton, J. C. & Morgan, P. J. (1992). MSTAR: A FORTRAN program for the model independent molecular weight analysis of macromolecules using low speed or high speed sedimentation equilibrium. In *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J. & Horton, J. C., eds), pp. 275-294, Royal Society of Chemistry, Cambridge.

Harwood, C. R. & Baumberg, S. (1977). Arginine hydroxamate-resistant mutants of *Bacillus subtilis* with altered control of arginine metabolism. *J. Gen. Microbiol.* **100**, 177-188.

Kim, H., Deonier, R. C. & Williams, J. W. (1977). The investigation of self-association reactions by equilibrium ultracentrifugation. *Chem. Rev.* **77**, 659-690.

Klingel, U., Miller, C. M., North, A. K., Stockley, P. G. & Baumberg, S. (1995). A binding site for activation by the *Bacillus subtilis* AhrC protein, a repressor/activator of arginine metabolism. *Mol. Gen. Genet.* **248**, 329-340.

Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical*

Ultracentrifugation in Biochemistry and Polymer Science (Harding, S. E., Rowe, A. J. & Horton, J. C., eds), pp. 90-125, Royal Society of Chemistry, Cambridge.

Lim, D., Oppenheim, J. D., Eckhardt, T. & Maas, W. K. (1987). Nucleotide sequence of the *argR* gene of *Escherichia coli* K-12 and isolation of its product, the arginine repressor. *Proc. Natl Acad. Sci. USA,* **84**, 6697-6701.

Maas, W. K. (1994). The arginine repressor of *Escherichia coli. Microbiol. Rev.* **58**, 631-640.

McRorie, D. K. & Voelker, P. J. (1993). *Self-associating Systems in the Analytical Ultracentrifuge*, Beckmann Instruments Ltd., Fullerton.

Miller, C. M. (1996). Mechanisms and functions of the *Bacillus subtilis* arginine regulatory protein, AhrC. PhD thesis, Department of Biology, University of Leeds, UK.

Miller, C. M., Baumberg, S. & Stockley, P. G. (1997). Operator interactions by the *Bacillus subtilis* arginine repressor/activator, AhrC: novel positioning and DNA-mediated assembly of a transcriptional activator at catabolic sites. *Mol. Microbiol.* **26**, 37-48.

Mountain, A. & Baumberg, S. (1980). Map locations of some mutants conferring resistance to arginine hydroxamate in *Bacillus subtilis* 168. *Mol. Gen. Genet.* **178**, 691-701.

North, A. K., Smith, M. C. M. & Baumberg, S. (1989). Nucleotide sequence of a *Bacillus subtilis* arginine regulatory gene and homology of its product to the *Escherichia coli* arginine repressor. *Gene,* **80**, 29-38.

Perkins, S. J. (1986). Protein volumes and hydration effects. The calculation of partial specific volumes, neutron scattering matchpoints and 280-nm absorption coefficients for proteins and glycoproteins from amino acid sequences. *Eur. J. Biochem.* **57**, 169-180.

Silkowski, H., Davis, S. J., Barclay, A. N., Rowe, A. J., Harding, S. E. & Byron, O. (1977). Characterisation of the low affinity interaction between rat cell adhesion molecules CD2 and CD48 by analytical ultracentrifugation. *Eur. Biophys. J.* **25**, 455-462.

Smith, M. C. M., Mountain, A. & Baumberg, S. (1986a). Sequence analysis of the *Bacillus subtilis* argC promoter region. *Gene,* **49**, 53-60.

Smith, M. C. M., Mountain, A. & Baumberg, S. (1986b). Cloning in *Escherichia coli* of a *Bacillus subtilis* arginine repressor gene through its ability to confer structural stability on a fragment carrying genes of arginine biosynthesis. *Mol. Gen. Genet.* **205**, 176-182.

Smith, M. C. M., Czaplewski, L., North, A. K., Baumberg, S. & Stockley, P. G. (1989). Sequences required for regulation of arginine biosynthesis promoters are conserved between *Bacillus subtilis* and *Escherichia coli*. *Mol. Microbiol.* **3**, 23-28.

Stockley, P. G., Baron, A. J., Wild, C. M., Parsons, I. D., Miller, C. M., Holtham, C. A. M. & Baumberg, S. (1998). Dissecting the molecular details of prokaryotic transcriptional control: the methionine and arginine repressor proteins. *Biosensors. Bioelect.* **13**, 637-650.

Studier, F. W., Rosenburg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60-89.

Sunnerhagen, M., Nilges, M., Otting, G. & Carey, J. (1997). Solution structure of the DNA-binding domain and model for the complex of multifunctional hexameric arginine repressor with DNA. *Nature Struct. Biol.* **4**, 819-825.

Tanford, C. (1961). *Physical Chemistry of Macromolecules*, vol. 381, J. Wiley & Sons, New York.

Tian, G. & Maas, W. K. (1994). Mutational analysis of the arginine repressor of *Escherichia coli. Mol. Microbiol.* **13**, 599-608.

Tian, G., Lim, D., Carey, J. & Maas, W. K. (1992). Binding of the arginine repressor of *Escherichia coli* K12 to its operator sites. *J. Mol. Biol.* **226**, 387-397.

Van Duyne, G. D., Ghosh, G., Maas, W. K. & Sigler, P. B. (1996). Structure of the oligomerization and L-arginine binding domain of the arginine repressor of *Escherichia coli. J. Mol. Biol.* **256**, 377-391.

Williams, J. W. (1973). *Ultracentrifugation of Macromolecules: Modern Topics*, vol. 50, Academic Press, New York.

Edited by A. R. Fersht