

The DNA-binding Domain of the Hexameric Arginine Repressor

Rita Grandori¹, Teresa A. Lavoie¹, Michelle Pflumm², Guoling Tian³
Helmut Niersbach³, Werner K. Maas³, Robert Fairman⁴ and
Jannette Carey^{1*}

¹Chemistry Department
Princeton University
Princeton, NJ 08544-1009
USA

²Chemistry Department
Haverford College
370 Lancaster Ave.
Haverford, PA 19041-1392
USA

³Microbiology Department
NYU Medical School
550 First Avenue
New York, NY 10020, USA

⁴Bristol-Myers Squibb
Pharmaceutical Research
Institute, P.O. Box 4000
Princeton, NJ 08543-4000
USA

The arginine repressor of *Escherichia coli* is a classical feedback regulator, signalling the availability of L-arginine inside the cell. It differs from most other bacterial repressors in functioning as a hexamer, but structural details have been lacking and it shares no clear sequence homologies with other transcriptional regulators. Analysis of the amino acid residue sequence and proteolytic cleavage pattern of the repressor was used to identify a region predicted to house the DNA-binding function. When this protein fragment is overexpressed from a clone of the corresponding gene fragment, it represses ornithine transcarbamylase levels *in vivo*, and binds to the operator DNA *in vitro*, both in an arginine-independent manner. Sedimentation equilibrium and gel filtration indicate that the purified protein fragment is a monomer in solution. The results thus define the domain organization of the repressor at low resolution, suggesting that the N and C-terminal portions of the polypeptide chain are separated by a structural and functional border that decouples hexamerization and arginine binding from DNA binding.

© 1995 Academic Press Limited

*Corresponding author

Keywords: proteolysis; allosteric repressors; structural domains; functional domains; oligomeric structure

Introduction

The repressor of the arginine regulon in *Escherichia coli*, ArgR, functions as a direct sensor of L-arginine, which binds as a co-repressor (for a recent review, see Maas, 1994). ArgR is thought to function as a hexamer of identical 16,500 Da polypeptides (Lim *et al.*, 1987). It binds to operator sites (so-called "ARG boxes") in the promoter regions of several operons encoding enzymes for L-arginine biosynthesis. Two adjacent 18 base-pair ARG boxes are bound cooperatively by one ArgR hexamer, inducing a bending angle of ~70° in the DNA with the locus of bending centered between boxes (Tian *et al.*, 1992). In addition to its regulatory function, ArgR has an unexpected role as an

accessory factor in resolution of ColEI multimers during plasmid replication (Stirling *et al.*, 1988), a property reminiscent of the role of *E. coli* integration host factor (IHF) in the excision and packaging of lambda prophage (Kosturko *et al.*, 1989); IHF also functions as a specific transcription factor (Craig & Nash, 1984). ArgR and IHF thus typify a class of proteins having some of the features of both gene organizers and gene regulators (D'Ari *et al.*, 1993), making the mechanisms of their interaction with DNA particularly intriguing.

ArgR has two known homologs, the nearly identical sequence from *Salmonella typhimurium* (Lu *et al.*, 1992), and the very divergent repressor-activator (AhrC) from *Bacillus subtilis* (North *et al.*, 1989). None of the three proteins shares any obvious sequence similarity with other DNA-binding proteins, nor matches the consensus sequence of any known classes of DNA-binding motifs as reported in the PROSITE database (Bairoch, 1992). No detailed structural information has been available to date that could shed light on the unique aspects of

Abbreviations used: IHF, integration host factor; IPTG, isopropyl-β-D-thiogalactopyranoside; OTCase, ornithine transcarbamylase; PVDF, polyvinylidene difluoride; PMSF, phenylmethylsulfonyl fluoride; PCR, polymerase chain reaction.

DNA binding by hexameric regulatory proteins, although a preliminary crystallization report for AhrC has been published (Boys *et al.*, 1990). Mutant analyses with ArgR recently identified two positions in the N-terminal region that affect DNA binding and four positions in the C-terminal region that affect both DNA- and L-arginine-binding, some of which also affect oligomerization (Tian & Maas, 1994; Burke *et al.*, 1994).

We present here the results of biochemical experiments that allow us to map the DNA-binding function of ArgR to an N-terminal domain which by itself is neither hexameric nor arginine dependent, suggesting that these functions may be housed in the C-terminal domain of ArgR. Thus, the structural and functional organization of ArgR appears to be similar in general outline to the well-known allosteric repressor, LacI, which has an N-terminal DNA-binding domain that can function autonomously with retention of specificity despite loss of the ability to form tetramers, and a C-terminal domain in which subunit assembly and effector-binding functions are housed (Chakerian & Matthews, 1992; Schumacher *et al.*, 1994; Sauer, 1995).

Results

AhrC, the arginine repressor of *B. subtilis* (North *et al.*, 1989), is the only homolog of ArgR among known protein sequences that is clearly related but divergent enough to be potentially informative in a sequence comparison. Alignment of the two sequences (see Figure 4 of North *et al.*, 1989) had revealed that these proteins have a highly related (~40% identity) C-terminal segment of ~60 residues, and a considerably more divergent (~20% identity) N-terminal segment of ~90 residues, leading to an overall identity of only 28%. The ability of AhrC to provide ArgR function in *E. coli* (Smith *et al.*, 1989) had naturally focused attention on the more related C-terminal region as a possible locus of DNA-binding activity. However, we noted that the aligned sequences reveal a striking difference in clustering of charged residues in the two regions. The pI values we calculated for the N- and C-terminal regions are drastically different, averaging ~9.5 for the first ~100 residues and ~4 for the last ~50 (data not shown). Thus, both sequence similarity and charge-residue distribution point to a boundary in the vicinity of residues ~90 to 100 that could reflect structural and/or functional compartmentalization within the protein. Sequence comparison with the *S. typhimurium* protein (Lu *et al.*, 1992) does not add any information since it is identical to the *E. coli* protein in 95% of its amino acid sequence.

Proteolytic analysis of ArgR

Proteolytic sensitivity has been used to probe the domain organization and dynamics of many proteins in solution (for a recent review see Fontana

et al., 1993). Regions that are consistently sensitive to different proteases having distinct sequence specificities typically are found to correspond to flexible and/or unstructured regions, such as active sites or linkers between domains (Fontana *et al.*, 1993; Jaenicke, 1991). The time course of digestion by trypsin and by chymotrypsin was used to identify the labile regions of ArgR; to locate the cleavage sites in the primary sequence, digestion products resolved by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) were subjected to several cycles of Edman degradation (see Materials and Methods for sequences). Typical results are shown in Figure 1. The C-terminal ends of the fragments shown in Figure 1 (right) are those most consistent with amino acid analysis of the corresponding bands.

The results with chymotrypsin (A panels) show that the first detected event produces two fragments, one having the same N-terminal sequence as intact ArgR (band a) and a second (band b) beginning with Gly74. Taken together with the sizes of these fragments and the sequence specificity of the enzyme, this result suggests that the protein is bisected by cleavage at a very sensitive site in the region of Leu73. Fragment a is then shortened from its N terminus, yielding fragment a' whose N-terminal sequence starts at Lys15. Analogously, fragment b is shortened at its N terminus by cleavage after Leu82, yielding fragment b'; the amino acid composition of fragment b' suggests that little or no degradation occurs at its C terminus. Both fragments a and a' are degraded within the first hour of treatment, leaving fragment b' as the only fragment detected after prolonged chymotryptic treatment.

The results with trypsin digestion show a generally similar pattern (B panels). The bond between Lys62 and Met63 is cleaved early, generating a fragment (band b) that contains a unique N-terminal sequence beginning with Met63. Cleavage at Arg2 also takes place in the first two minutes of the reaction, generating a very short-lived fragment (band a) which migrates slightly faster than native ArgR. Although the N terminus of intact ArgR is readily sequenced and therefore is not blocked, no other fragments bearing sequences from the N-terminal region were identified during the time-course, implying that this region is rapidly degraded to fragments small enough to be lost during electrophoresis or staining, and suggesting that the N-terminal portion either contains a site(s) more labile than Lys62/Met63, or acquires a labile site(s) upon cleavage of the Lys62/Met63 bond; the large number of basic residues in the 1 to 62 segment may facilitate its rapid tryptic degradation. These results indicate that no N-terminal region is resistant to further proteolysis following cleavage of the 62/63 bond, suggesting that fragments truncated at residue 62 might not form a stably folded unit independently. The next proteolytic event shortens band b from its N-terminal end, yielding fragment b' with its sequence beginning at Asn84,

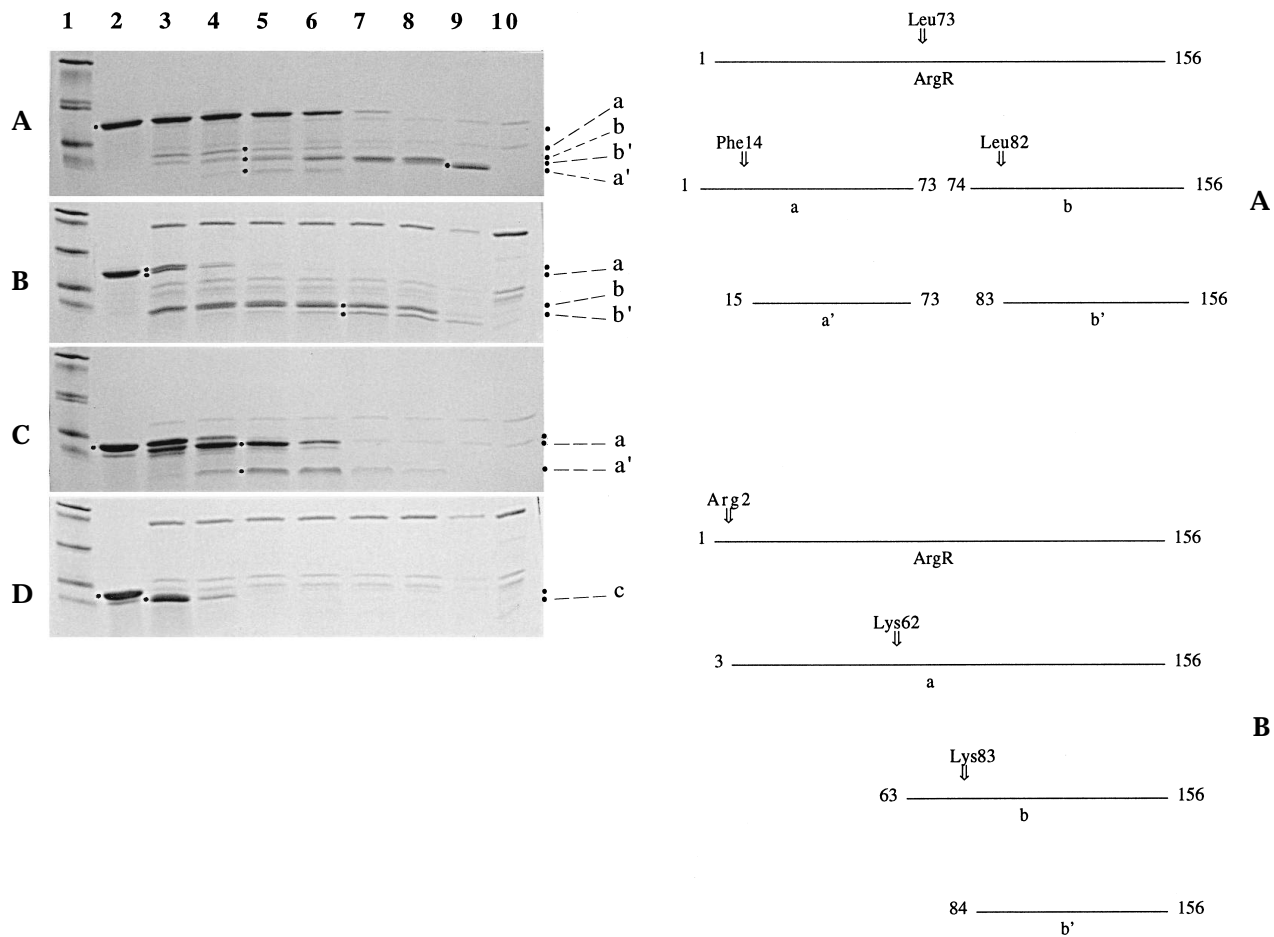


Figure 1. (Left) Proteolytic digestion of ArgR (panels A and B) and ArgRN (panels C and D), analyzed by Coomassie-stained SDS-PAGE gels. Time course of digestion by chymotrypsin (panels A and C) and trypsin (panels B and D). Lane 1, molecular mass markers (Da, from top to bottom in each panel: carbonic anhydrase, 29,000; trypsinogen, 24,000; trypsin inhibitor, 20,000; α -lactalbumin, 14,200; aprotinin, 6500); lane 2, starting sample of pure ArgR or ArgRN (see the text for discussion of degradation product in pure ArgRN); lanes 3 to 9, respectively: digestion for 2, 5, 10, 20, 60, 120 minutes, 14 hours; lane 10, chymotrypsin (panels A and C) or trypsin (panels B and D). Dots to the left of some bands and to the right of the gel mark the positions of intact ArgR or ArgRN (unlabeled) and proteolytic fragments discussed in the text. (Right) Interpretation of fragmentation patterns, shown schematically on the linearized protein chain.

presumably due to cleavage after Lys83. Even at very late stages of the time course, fragment b' persists without being detectably further shortened, similarly to the corresponding chymotryptic fragment b' (Figure 1A).

These results suggest that the region around residues ~70 to 80 of ArgR represents a linker between structural domains. The slightly greater resistance of the N-terminal portion to chymotrypsin (fragments a and a' in Figure 1A) than to trypsin presumably reflects the different sequence specificity of the two enzymes, consistent with the expectation that the hydrophobic substrate residues of chymotrypsin would be buried in a folded structure. The resistance of the C-terminal ~70 residues of the protein to prolonged proteolysis suggests that fragments of this length contain a segment of the chain that can remain folded independently. The central region (Lys15 to Leu73) of the N-terminal portion of the protein is also

somewhat protected from proteolysis, but its degradation within the first hour of treatment suggests the presence of flexible and/or exposed regions. Differential proteolytic sensitivity thus suggests a region near residues 70 to 80 with high accessibility that divides the chain into two portions, each of which seems to have some potential for folding independently.

Expression of the N-terminal domain

A gene fragment was designed to express the N-terminal moiety of ArgR in order to test the inference that this region of the protein might fold independently and contain the DNA-binding domain. Position 96 was chosen as the boundary for subcloning mainly on the basis of the discontinuities in pI and in sequence similarity to AhrC. In addition, we wanted to avoid truncating the designed fragment in the region of residues ~70 to

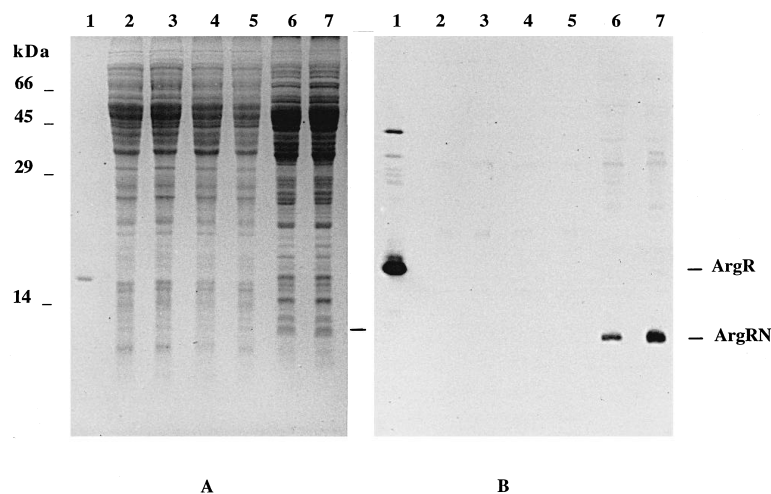


Figure 2. Expression of ArgRN from plasmid pNRG169, analyzed on duplicate SDS-PAGE gels by Coomassie-staining (panel A) and immunoblot (panel B). Lane 1, pure preparation of ArgR; lanes 2 to 7, total protein extracts from cultures of cells uninduced (lanes 2, 4, and 6) and induced (lanes 3, 5, and 7) by 1 mM IPTG; lanes 2 and 3, parental strain JM101; lanes 4 and 5, host strain JM101*argR*; lanes 6 and 7, host strain JM101*argR* transformed with plasmid pNRG169 expressing ArgRN. The positions of molecular mass markers are indicated (from top to bottom: bovine serum albumin, ovalbumin, carbonic anhydrase, α -lactalbumin). The alignment of the two panels is imperfect due to a slight change in the size of the gel during staining.

80 to allow for the possibility that proteolytic accessibility there could reflect a flexible or exposed active site, rather than a domain boundary, in that region. We also reasoned that, if N-terminal proteolytic fragments of ArgR truncated in the region of residues ~ 70 to 80 are rapidly degraded, perhaps the slightly longer fragment incorporating the entire basic region would be more stable. The ArgR gene in plasmid pDB169 (Lim *et al.*, 1987) was truncated by PCR cloning to yield a gene fragment encompassing amino acid residues 1 to 96 under the control of the *tac* promoter, and the plasmid was transformed into an ArgR⁻ host strain. Transformants were colony-purified, and the orientation and integrity of the insert were confirmed by restriction mapping and its sequence determined. The transformed strain was grown in LB to mid-log phase and then induced with isopropyl- β -D-thiogalactopyranoside (IPTG). Figure 2 shows total protein extracts from induced and uninduced cultures, resolved by denaturing gel electrophoresis and analyzed by Coomassie staining (Figure 2A) and immunoblotting (Figure 2B). Induction is correlated with a prominent band of the expected size (~ 10 kDa; lane 7), suggesting that the fragment is stable to intracellular proteolysis when overproduced. The protein in this band cross-reacts with antibodies to intact ArgR, and five cycles of Edman degradation on the protein in this band from a pure preparation (see below) confirmed its identity. We refer to this protein fragment as ArgRN.

In vivo activity

To assess the ability of ArgRN to regulate genes of the arginine regulon, ornithine transcarbamylase (OTCase) activity was assayed in cultures expressing the 10 kDa fragment. Cultures were grown in minimal medium A (Davis & Mingioli, 1950) with or without addition of 1 mM L-arginine, and the effect of IPTG induction was tested. Full derepres-

sion of the ArgF gene from which OTCase is expressed is observed in an ArgR⁻ host, and corresponds to ~ 350 units of OTCase specific activity (Table 1). Full repression, observed in the wild-type host grown in the presence of 1 mM L-Arg, corresponds to ~ 0.2 unit. Upon induction of ArgRN, OTCase levels are reduced from ~ 400 units to ~ 50 units, independently of the presence of 1 mM L-Arg. These results indicate that the fragment acts as a weak, L-Arg-independent repressor under these growth conditions. In repressing ArgF, the fragment gives approximately 90% of the repression given by intact, chromosomally encoded ArgR in the absence of L-Arg. Control experiments using cells transformed with the vector alone or with a plasmid clone of TrpR (Paluh & Yanofsky, 1986) indicate that the effect is not due to any other plasmid-encoded product, nor can it be observed by overexpressing a DNA-binding protein with different specificity (data not shown). As a second independent test of ArgRN activity, plasmid pNRG169 was also transformed into the *E. coli* strain EC146(λ AZ-7) (Eckhardt, 1980), which carries the *lacZ* gene under the control of the ArgA promoter in an ArgR⁻ recombinant λ lysogen.

Table 1. Repressibility of OTCase activity by ArgRN

Supplements	OTCase specific activity ^a			
	MG1655	MG1655 <i>argR</i>	MG1655/ pNRG169	MG1655 <i>argR</i> / pNRG169
None	13.7	351	22.9	406
Arg	0.2	324	0.5	390
IPTG	ND	ND	8.6	55
Arg + IPTG	ND	ND	1.8	46

Cells were grown to mid to late-log phase in arginine-free minimal medium A containing 1 mM L-arginine (Arg) or IPTG or both, and subjected to OTCase assay as described in Materials and Methods.

^a Specific activity is given as μ mol citrulline/h/mg of dry weight of the cells. ND, not determined.

Colony-purified transformants were plated onto MacConkey agar, where they gave a phenotype (pink color) intermediate between that of ArgR⁺ transformants and the untransformed host (data not shown). To test the specificity of the effect of ArgRN overproduction, we measured β -galactosidase activity in cells grown in the presence of IPTG. The levels of this enzyme were essentially the same (~2000 Miller units; Miller, 1972) in strains MG1655ArgR, MG1655ArgR/pNRG169, and MG1655ArgR/pDB169. Thus, overproduced ArgRN does not promiscuously regulate β -galactosidase.

The ability of ArgRN to regulate OTCase suggests that the ArgRN protein fragment is present in a soluble form, implying that its stability *in vivo* is not due trivially to production of inclusion bodies. Together these results indicate that fragment 1 to 96 of ArgR may fold independently of the C-terminal domain inside the cell and contains L-arginine-independent DNA-binding activity.

Purification of ArgRN

To improve the expression level of ArgRN for purification, we subcloned the corresponding coding region under the control of a T7 RNA polymerase promoter in the vector pET3a (Studier *et al.*, 1990) to generate plasmid pMP5O. This plasmid was transformed into strain BL21(λ DE3)ArgR, which carries an IPTG-inducible gene for T7 RNA polymerase (Studier *et al.*, 1990) and a disrupted chromosomal copy of the ArgR gene. We found that plasmid pMP5O is highly unstable in this strain during growth in rich medium, but that the plasmid can be maintained well by growth in minimal medium. Upon induction, the level of ArgRN overproduction is approximately 15% of the total cell protein (Figure 3, lane 4), as measured by SDS-PAGE densitometry (not shown).

To attempt to purify the basic ArgRN, we tested its behavior on a phosphocellulose cation exchange column (P-11) under different conditions. When a crude extract from an induced cell culture is applied to a P-11 column equilibrated in Arg buffer and then washed with the same buffer, elution of ArgRN is delayed slightly compared to the bulk of non-interacting proteins that are washed out with the first column-volume of buffer. The weak interaction of ArgRN with the P-11 resin under these conditions represents a powerful tool for purification. We reasoned that the magnesium cations contained in Arg buffer might compete in the interaction between ArgRN and the resin; thus, the behavior of ArgRN was tested on a P-11 column equilibrated in P-11 buffer, which contains no Mg²⁺. Indeed, the fragment is retained on the column under these conditions and is eluted at approximately 0.4 M NaCl in a linear salt concentration gradient. The elution behaviors are reproducible and independent of the sequence of the two column steps.

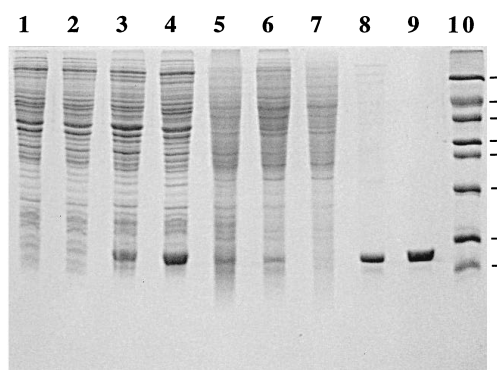


Figure 3. Expression and purification of ArgRN from plasmid pMP5O in host strain BL21(λ DE3)argR, analyzed by Coomassie-stained SDS-PAGE. Lanes 1 to 4, total protein extracts from cells grown in minimal medium, uninduced (lanes 1 and 3) or induced by 1 mM IPTG (lanes 2 and 4). Lanes 1 and 2, host strain alone; lanes 3 and 4, host strain transformed with pMP5O. Lanes 5 to 9, samples from ArgRN purification. Lane 5, soluble fraction from French press extract; lane 6, supernatant of streptomycin sulfate precipitation (this sample was loaded onto the first P-11 column); lane 7, P-11 buffer wash of first P-11 column; lane 8, concentrated pool of fractions containing ArgRN eluted from the first P-11 column with a salt gradient in P-11 buffer; lane 9, concentrated pool of fractions containing ArgRN eluted from the second P-11 column with Arg buffer. Lane 10, molecular mass markers (Da, top to bottom: bovine serum albumin, 66,000; ovalbumin, 45,000; glyceraldehyde-3-phosphate dehydrogenase, 36,000; carbonic anhydrase, 29,000; trypsinogen, 24,000; trypsin inhibitor, 20,000; α -lactalbumin, 14,200; aprotinin, 6500).

We obtained pure preparations of ArgRN in approximately 10% overall yield from crude protein extracts of induced cells by applying two sequential steps of cation exchange chromatography, first in P-11 buffer and then in Arg buffer. Figure 3 shows the intermediate and final samples (lanes 5 to 9) from a typical purification of ArgRN. The final pool was run on SDS-PAGE and transferred to polyvinylidene difluoride (PVDF) membrane for Western blotting and for N-terminal sequencing by five cycles of Edman degradation, both of which support its identity. The purity of the final pool was approximately 97% as judged by N-terminal sequencing carried out directly on the pool and by quantitative densitometry of overloaded SDS-PAGE gels. A faint band of approximately 20 kDa is detectable in the final ArgRN preparation (Figure 3, lane 9). It bears the same N-terminal sequence as ArgRN, and when this band is excised, re-boiled in fresh SDS sample buffer, and run on a second gel, both 20 kDa and 10 kDa bands are detected (not shown). Thus, the 20 kDa species probably represents a covalent dimeric form produced by oxidation of the unique cysteine at position 68 of ArgRN.

Proteolytic analysis of ArgRN

We used a variety of biochemical techniques to probe the structural properties of ArgRN and to test

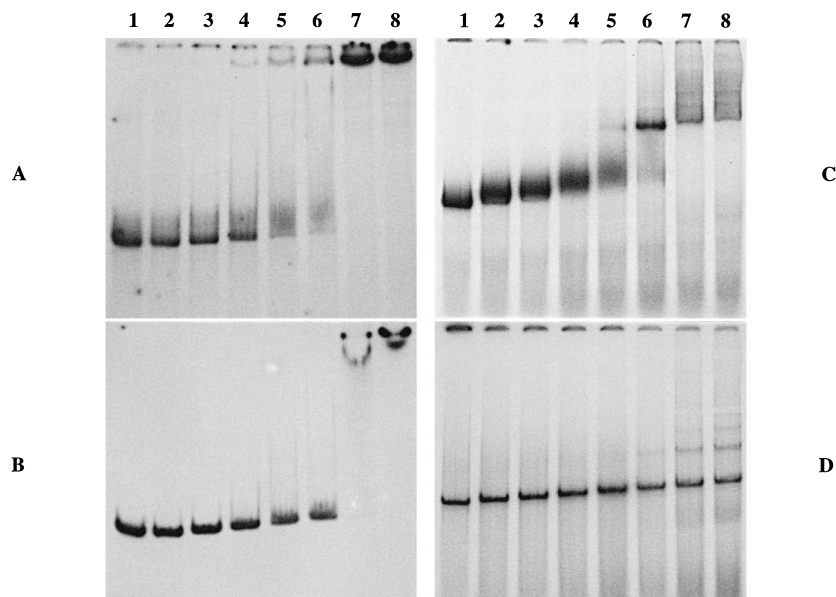


Figure 4. DNA-binding activity of ArgRN (A and B) and ArgR (C and D) by band-shift titration; an autoradiogram of ^{32}P -labeled DNAs is shown. Panels A and C: ArgF operator on a ~ 120 bp fragment. Panels B and D: *trpEDCBA* DNA (non-operator control) on a ~ 100 bp fragment. Lane 1, no protein; lanes 2 to 8, ArgRN or ArgR at final concentrations of 0.3, 0.6, 1.5, 3, 6, 15, and 30 μM .

whether the purified fragment is folded into a stable, compact structure. An important criterion to answer this question is that the truncated product displays susceptibility to limited proteolysis in a similar way to the corresponding region of the intact protein. We therefore subjected pure preparations of ArgRN to digestion by chymotrypsin and trypsin under the same conditions employed for intact ArgR. As in the case of intact ArgR, the identity of the intermediate products resolved by SDS-gel electrophoresis was assessed by N-terminal sequencing of the corresponding bands after transfer onto PVDF membranes. During digestion with chymotrypsin, ArgR generates two fragments from within the region 1 to 96 (bands a and a' in Figure 1, A panels). As shown in Figure 1C, ArgRN produces equivalent chymotryptic fragments (bands a and a') with similar kinetics. Also similarly to ArgR, bands a and a' of ArgRN are detectable throughout one hour of incubation under the conditions employed. The relatively slow rate of proteolysis of ArgRN, which is comparable to that of ArgR, implies the presence of folded structure in ArgRN, while the similarity of fragment identities suggests that this structure resembles that of the corresponding N-terminal portion of intact ArgR.

Because no fragment corresponding to the N-terminal region is detected during trypsin digestion of ArgR, we did not expect to observe intermediate products from ArgRN digestion with trypsin. Surprisingly, a short-lived fragment is detected in the first five minutes of digestion (band c in Figure 1D). Its N-terminal sequence starts at Phe23, deriving from cleavage at Lys22. No tryptic fragment starting at Phe23 was detected upon treating ArgR with lower enzyme-to-substrate ratios than those shown in Figure 1B (data not shown). These observations might indicate slight differences in the conformation of the N-terminal

domain in ArgR and ArgRN that we are not able to interpret with the presently available data.

As always observed in older preparations of ArgRN, an additional band migrating slightly faster than ArgRN is detected in the starting material employed in Figure 1 (lane 2, panels C and D). This fragment bears the same N-terminal sequence as ArgRN and is therefore likely to be generated by degradation of ArgRN from its C terminus. N-Terminal sequencing of the unresolved sample reveals the presence of a minor component corresponding to the ArgRN sequence starting at Leu85 (secondary sequence: LVLDI), with a ratio of primary-to-secondary peaks of ~ 30 . These observations, together with the mobility of the degradation product, are consistent with the inference that the region around residue 80 corresponds to a boundary between structural domains.

In vitro activity

We analyzed the DNA-binding activity of ArgRN *in vitro* by band-shift assay using Arg operator and non-operator DNAs. Reactions contained a ~ 120 bp fragment from plasmid pGT101 (Tian *et al.*, 1992) with two ARG boxes of the ArgF operator, or a ~ 100 bp fragment from the *trpEDCBA* regulatory region of *Serratia marcescens* (Figure 4). Consistent with the results of the *in vivo* experiments, addition of 5 mM L-arginine had no effect on DNA binding by ArgRN, although it markedly increased the affinity of ArgR, as has been reported (Lim *et al.*, 1987; Tian *et al.*, 1992); thus, only results obtained in the absence of L-arginine are described here. ArgRN binds to operator (panel A) and non-operator (panel B) DNAs with an apparent affinity (K_d , the protein concentration at which half the free DNA becomes bound) of between 3 and 6 μM under these conditions. ArgRN exhibits a slight preference for operator over non-operator DNA, as

judged by the slightly greater depletion of the free DNA band in lanes 5 and 6 of panel A than in the corresponding lanes of panel B, consistent with the *in vivo* results. Though small, this slight preference is highly reproducible, and is also clear when the band-shift titration is carried out on a mixture of operator and non-operator DNAs (not shown). In addition, the complex bands formed on the operator DNA are more discrete than those formed on the non-operator DNA. The binding behavior of ArgR on the operator DNA is quantitatively similar in the absence of L-arginine (Figure 4C), with an apparent K_d value of near $3 \mu\text{M}$, although the selectivity is slightly higher than that of ArgRN (tenfold *versus* fivefold), due to the slightly higher affinity of ArgRN for non-operator DNA.

Native molecular weight of ArgRN

The availability of pure, truncated active protein can be exploited to investigate whether the hexameric structure is an intrinsic feature of the DNA-binding domain of ArgR. We therefore analyzed the hydrodynamic size of ArgRN by gel filtration chromatography on a Superdex 75 column in Arg buffer. Figure 5A and B show elution profiles of ArgRN compared to molecular weight standards. ArgRN elutes between cytochrome *c* (12,400 Da) and carbonic anhydrase (29,000 Da). A plot of log molecular weight *versus* elution volume (Figure 5C) for the standards is linear, and the elution position of ArgRN interpolated onto this plot indicates an apparent molecular weight of 20,700 Da (expected molecular weight for the monomer, $\sim 10,700$ Da). The integrity of the eluted protein was confirmed by SDS-PAGE of the collected fractions (not shown). No additional peaks were detected in the elution profile of ArgRN by absorbance at 280 nm when the pure protein was run alone on the column. The elution profile of ArgRN was unaffected by the addition of 1 M salt or 20% glycerol in the running buffer. The narrowness and symmetry of the ArgRN elution profile suggests the presence of a single species under these conditions. The observed elution volume of ArgRN excludes a hexameric oligomerization state of the fragment and would be compatible with a dimeric assembly. Nevertheless, since the shape of a protein affects its hydrodynamic behavior, we tested the sedimentation properties of ArgRN by analytical centrifugation.

Sedimentation equilibrium ultracentrifugation at loading concentrations of 7, 20, or $50 \mu\text{M}$ ArgRN shows that the protein sediments as a monomer. Figure 6 (lower) shows the distribution of absorbance for a sample of $20 \mu\text{M}$ ArgRN centrifuged at three speeds at 4°C . The data are fitted well by the distribution calculated (continuous lines) for a single species with a molecular mass equal to that of the ArgRN monomer (10,667 Da), as judged by the randomness of the residuals from the global three-speed curve-fit (upper). Similar results were obtained in independent runs performed at 25°C or in the presence of 1 M NaCl. The results were

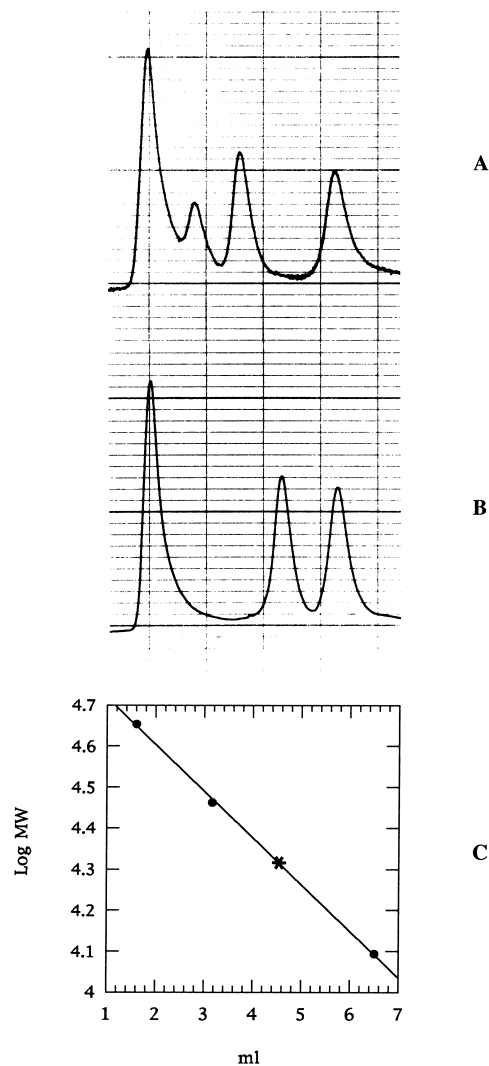


Figure 5. Size exclusion chromatography in Arg buffer, analyzed by absorbance at 280 nm of the eluting material. A and B, elution profiles. In A, peaks from left to right correspond to blue dextran, ovalbumin, carbonic anhydrase, and cytochrome *c*, and in B to blue dextran, ArgRN ($60 \mu\text{M}$), and cytochrome *c*. Vertical lines mark 2 ml intervals and horizontal lines mark 0.002 A intervals. C, semilog plot of molecular mass *versus* elution volume. Elution volumes (measured at peak) of calibration standards are averaged from four to five independent runs, with error estimated to be approximately ± 0.2 ml. The elution position of ArgRN (average of two runs in these conditions) is indicated by an asterisk.

affirmed by an analysis of fitted data in which molecular mass is treated as an adjustable parameter (Table 2). The second column shows the best-fit molecular mass from global five-speed analysis at each protein concentration. For the three lower concentrations, the predicted molecular mass is close to that of the ArgRN monomer. The square root of variance, which is a measure of the goodness of fit, is shown in columns 3 and 4 for dimeric and monomeric single-species models, respectively, providing strong evidence that ArgRN is a monomer in this concentration range. These results

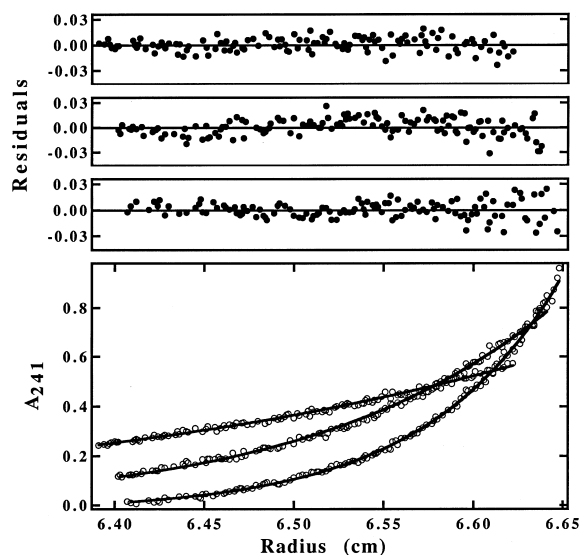


Figure 6. Sedimentation equilibrium analysis of ArgRN. A 20 μM sample of ArgRN was centrifuged at 4°C in 20 mM Tris-HCl (pH 7.5), 10 mM MgCl_2 , 100 mM NaCl, 1 mM β -mercaptoethanol. Data collected at rotor speeds of 20,000, 30,000, and 40,000 rpm (top to bottom, respectively, at lower left) are shown fitted to a single-species model using a fixed molecular mass corresponding to that of the monomer. The residuals from the curve-fits are shown in the corresponding upper panels.

are consistent with the gel filtration results in suggesting that ArgRN behaves as a single species. The larger apparent molecular mass found by gel filtration therefore reflects a larger hydrodynamic radius than that expected for a compact sphere of this molecular mass. Because the sedimentation results could be consistent with a random coil, which would be spherical but not compact, we used circular dichroism (CD) to examine the secondary structure content of ArgRN. ArgRN contains secondary structure as indicated by the spectrum obtained by CD on a pure protein preparation (Figure 7). The spectrum shows the typical features of mixed α/β structures, with a shoulder at 222 nm, a negative minimum at 208 nm, and a positive

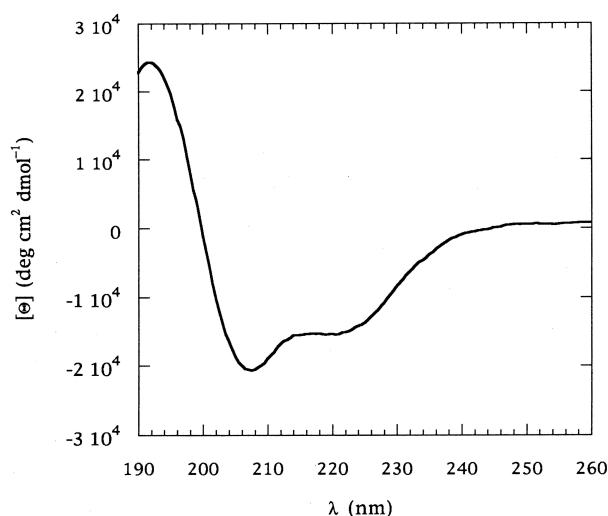


Figure 7. CD spectrum of 68 μM ArgRN at 4°C in 20 mM Tris-HCl (pH 7.5), 10 mM MgCl_2 , 100 mM NaCl, 1 mM β -mercaptoethanol. The spectrum was corrected by subtracting the spectrum of the same buffer as blank. Data were recorded on an AVIV model 60DS spectrometer with a 0.1 mm path-length cuvette.

maximum at 192 nm. Therefore, the data are most consistent with a non-spherical shape (either an oblate or prolate ellipsoid; Laue & Rhodes, 1990), or a molten globule (Dolgikh *et al.*, 1981; Ohgushi & Wada, 1983).

Interestingly, at concentrations of 100 and 157 μM ArgRN some higher-order oligomerization is evident. In single-species analysis in which molecular mass is treated as an adjustable parameter, the molecular mass increases with protein concentration (Table 2, column 2). Although the calculated molecular masses are within error of those determined at lower protein concentrations, the residuals from the single-species analysis (not shown) are clearly non-random. Therefore, an attempt was made to better define the oligomerization state at these protein concentrations by fitting the data to various monomer \leftrightarrow n -mer equilibria up to $n = 6$. As can be seen from Table 2, the square root of variance at the two highest protein

Table 2. Analysis of data from sedimentation equilibrium ultracentrifugation

[Protein] ^a (μM)	Square root of variance ($\times 10^4$)							
	M_r^b	2 ^c	1 ^c	1 \leftrightarrow 2 ^d	1 \leftrightarrow 3 ^d	1 \leftrightarrow 4 ^d	1 \leftrightarrow 5 ^d	1 \leftrightarrow 6 ^d
7	11,400 \pm 1500	124	65	64	64	63	63	63
20	11,500 \pm 500	366	109	101	100	100	100	100
50	10,900 \pm 600	657	134	132	130	129	129	128
100	12,700 \pm 1000	234	102	80	78	77	76	76
157	14,100 \pm 1200	439	289	177	151	138	137	143

Data were collected at 4°C using detection at 241 nm for the 7, 20, and 50 μM samples and at 251 nm for the 100 and 157 μM samples. Data were analyzed using global fits to 20,000, 25,000, 30,000, 35,000, and 40,000 rpm data using the HID program from the Analytical Ultracentrifugation Facility at the University of Connecticut.

^a Concentration of protein loaded.

^b Molecular mass in Daltons calculated from single-species analysis.

^c Dimer and monomer species, respectively.

^d Monomer \leftrightarrow n -mer equilibria of increasing order up to $n = 6$.

concentrations is considerably smaller for all monomer \leftrightarrow n -mer models (columns 5 to 9) than for any single-species model (columns 3 and 4). The results suggest that the oligomerization state at high concentration is probably larger than a trimer, as judged by the drop in the square root of variances between the 1 \leftrightarrow 3 and 1 \leftrightarrow 4 models for the 157 μ M data. Although it is not possible to determine the higher oligomerization state with confidence, the association constants (not shown) calculated from the monomer \leftrightarrow n -mer fits predict up to 10% oligomerized protein at a total protein concentration of 100 μ M and up to 20% at 157 μ M.

Discussion

The arginine repressor is a multifunctional protein of unknown structure. The data reported here allow us to describe the domain organization of this protein at low resolution and to map its DNA-binding activity to an N-terminal domain. This conclusion is in agreement with recent evidence for some N-terminal mutations that affect DNA-binding activity of the intact protein but not L-arginine binding or hexamerization (Tian & Maas, 1994). The fact that the N-terminal fragment displays activity *in vivo* implies that it is sufficient for both binding and repression, though it is somewhat less efficient than the intact protein. Our results imply that the multiple binding sites of the native ArgR hexamer can bind DNA independently, though further work will be needed to define the minimal DNA-binding unit of this protein and to characterize the affinity and specificity of DNA binding by ArgRN compared with ArgR.

The DNA-binding activity of ArgRN indicates that this protein fragment is able to fold independently of the C-terminal domain in the presence of DNA. However, on the basis of the observed hydrodynamic and sedimentation properties a compact folded structure cannot be definitively assigned to pure ArgRN in the absence of DNA. Our data are compatible with either a non-spherical folded structure or a molten globule. Nevertheless, the first possibility is strengthened by the highly similar patterns and rates of proteolysis of ArgRN and the corresponding region of ArgR. Limited proteolysis on other DNA-binding proteins sometimes shows the presence of hypersensitive sites in their DNA-recognition regions (Tasayco & Carey, 1992), consistent with the role of these regions as active sites and with the relatively superficial location, and sometimes disordered structures (Spolar & Record, 1994), of the DNA-binding motif. The data for ArgR suggest that residues ~70 to 80 correspond to a linker between structural domains, although the results do not imply that the ArgRN fragment necessarily folds as a single globular domain. The relative stability of fragments 83/84 to 156 of ArgR, and the slight shortening of ArgRN from its C terminus, suggest that residues 83/84 to 96 might properly be part of the C-terminal domain. The extent of the linker

region toward the N-terminal side is not well defined by our results.

The monomeric state of ArgRN suggests, by difference, that the C-terminal domain plays a critical role in promoting hexamerization. In addition, the considerable resistance of the short C-terminal segment to prolonged proteolysis is likely to reflect association into a higher-order assembly, suggesting the probable independence of its putative hexamerization function. Another function that is suggested to reside in the C-terminal domain is arginine binding, since ArgRN displays arginine-independent DNA-binding activity. Consistent with this implication, mutations at positions 105, 123 (Tian & Maas, 1994), 128 and 129 (Burke *et al.*, 1994) affect arginine binding. While this paper was in preparation, crystallization and preliminary X-ray diffraction analysis was reported (G. Van Duyne & P. B. Sigler, personal communication to W.K.M., cited by Maas, 1994) for a cloned C-terminal fragment of ArgR corresponding to a fragment generated originally by spontaneous proteolysis between residues 78 and 79 during crystallization attempts. The preliminary models indicate a hexameric assembly of polypeptides comprising residues 80 to 152. The complex exhibits both 2-fold and 3-fold symmetry, with two donut-shaped rings of three monomers each stacked together, and six molecules of L-arginine at the interface between rings.

Cross-talk between the N-terminal and C-terminal domains of ArgR could be achieved by changes in conformation and/or in multimeric state coupled to L-arginine binding. From the fact that ArgRN exhibits arginine-independent binding to DNA, we can infer that such changes affect the details of operator binding. Thus, it will be important to compare the detailed features of the DNA complex formed by ArgRN with those of intact ArgR.

Due to the homology between ArgR and AhrC, and functional complementation by AhrC in *E. coli* (Smith *et al.*, 1989), our inferences are likely to apply also to the repressor from *Bacillus subtilis*. A proteolytic pattern that could be consistent with our findings has been reported for AhrC (Czaplewski *et al.*, 1992). The sequence similarity between the two proteins in their N-terminal region, though very low, is not due merely to the abundance of basic residues (not shown). It would be of particular interest to identify the structural features that allowed conservation of DNA recognition specificity in such divergent sequence contexts.

None of the consensus motifs defined by known categories of transcription factors has been identified in the primary sequence of ArgR or AhrC (Lim *et al.*, 1987; North *et al.*, 1989). We were intrigued by the functional analogy to IHF and thus investigated in more detail the possibility of sequence similarity. The fragment that we demonstrated here to contain the DNA-binding activity of ArgR has the same size and pI value as the members of the histone-like protein family to which IHF belongs (Drlica & Rouviere-Yaniv, 1987). We

found that the N-terminal ~100 residues of ArgR and AhrC could each be aligned to the histone-like proteins, but no single common frame could be found for simultaneous alignment of both proteins with the family. Furthermore, the best values of identity and similarity found for individual alignments with members of the HU family are well below the threshold (33% identity) for structural homology in sequences of this length (Sander & Schneider, 1991). In any case, a putative relationship to the HU family would not illuminate the details of DNA binding, since footprinting studies reveal minor groove contact from both "faces" of DNA for IHF (Yang & Nash, 1989), but major and minor groove contact from only one face for ArgR and AhrC (Charlier *et al.*, 1992; Lu *et al.*, 1992; Tian *et al.*, 1992; Czaplewski *et al.*, 1992), implying at least a difference in the local structures of the protein regions that contact DNA (Yang & Carey, 1995). The protein fragment we describe here provides a useful experimental system that might help to understand the mechanism of DNA recognition by the arginine repressor.

Materials and Methods

Proteolysis

Reaction mixtures contained ~150 µg/ml of pure ArgR (Lim *et al.*, 1987) or ArgRN, 15 mM Tris-HCl (pH 7.5), 5 mM MgCl₂, and 65 µg/ml trypsin or 13 µg/ml chymotrypsin. The enzyme was diluted from a freshly prepared stock (1 mg/ml in 10 mM Tris-HCl (pH 8)). The reactions were carried out at room temperature. Aliquots were removed over a time course, and the reaction was stopped by addition of SDS-PAGE loading buffer and immediate boiling.

Protein analysis

Discontinuous SDS-PAGE was performed according to Laemmli (1970) in a Bio-Rad minigel apparatus at 15% (w/v) acrylamide concentration or, for analysis of proteolytic fragments, in 10% to 20% acrylamide gradient gels (Bio-Rad). After electrophoresis, gels were either stained with Coomassie R (1 g/l in 50% (v/v) methanol, 10% (v/v) acetic acid), or blotted to polyvinylidene difluoride (PVDF) membranes in 10 mM 3-[cyclohexylamino]-1-propanesulfonic acid (pH 11), 10% methanol for sequencing, amino acid analysis, or detection by antibodies. Western blotting was performed according to the protocol provided by Applied Biosystems using a 1:100 dilution of total anti-ArgR rabbit antiserum (Lim *et al.*, 1987) in 25 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 3% (w/v) bovine serum albumin. Peroxidase-conjugated goat anti-IgG antibodies from Sigma at 1:8000 dilution were used with 3,3'-diaminobenzidine and hydrogen peroxide for detection. All computer analysis of protein sequences was done with the programs of the GCG package of the Wisconsin Genetics Group (Devereux *et al.*, 1984).

The ArgR sequence in one-letter amino acid code is: MRSSAKQEEL VKAFKALLKE EKFSSQGEIV AALQE-QGFDN INQSKVSRML TKFGAVRTRN AKMEMVYCLP AELGVPTTSS PLKNLVLDDID YNDAVVVIHT SP-GAAQLIAR LLDLSLGAEG ILGTIAGDDT IFTTPANGFT VKDLYEAILE LFDQEL. Sequential rounds of Edman

degradation on the proteolytic products identified in Figure 1 were as follows. Panels A, C: band a (MRSSA), band a' (KALLK), band b (GVPTT), band b' (KNLVL). Panels B, D: band a (SSAKQ), band b (MEMVY), band b' (NLVLD), band c (FSSQG). N-Terminal sequencing and amino acid analysis of proteins and peptides were performed by the staff of the Princeton University Molecular Biology Department sequencing facility using standard methods.

Strains and plasmids

The following strains of *E. coli* were utilized in this work; JM101 (Δlac pro, *supE thi/F' traD36, proAB⁺, lacZ⁺ ZAM15*) from Pharmacia; JM101 *argR* (Tian & Maas, 1994); EC146 (λ AZ-7) (*argD, argR, argA::lacZ*) (Eckhardt, 1980); the wild-type strain MG1655 (ATCC 47076) and MG1655 *argR* (Tian & Maas, 1994); BL21 (λ DE3) *argR* (Studier *et al.*, 1990; and this work). Strain BL21 (λ DE3) was made *argR* by phage P1 transduction as described by Tian *et al.* (1994). Plasmid pNRG169 was constructed by replacing the 1147 bp *HindIII-SalI* fragment of pDB169 (Lim *et al.*, 1987) with a 393 bp *HindIII-SalI* PCR product. This fragment was designed to span the sequence of pDB169 from the unique *HindIII* site upstream of the *tac* promoter to nucleotide 286 of the ArgR coding sequence, and to introduce a TAA stop codon adjacent to codon 96 of ArgR, a substitution of Lys for Val96, and a *SalI* site 3' to the stop codon. The sequence of the upstream oligonucleotide was 5' CACAAAGCTTACTCCCCATCC-CCCTGTTG3' (*HindIII* site underlined), and the sequence of the downstream (antisense) oligonucleotide was 5' GTGCGTCCGACTTACTTAAGTGCATCGTTGTAGTCGAT-ATC3' (*SalI* site underlined, stop codon indicated in boldface, and codon 96 in lower case letters). The substitution of Lys for Val96 as the final residue of the fragment was influenced by a putative homology with IHF (see Discussion), which is thought to have a C-terminal helical segment. A T7 RNA polymerase-expressible clone was derived from pDB169 using the same strategy as above, with the upstream oligonucleotide 5' GGAATTCCATATGCGAAGCTCGGC-TAAGCAA3' (*NdeI* site underlined; start codon in boldface) and the downstream (antisense) oligonucleotide 5' GTGCGGATCCTTAAGACTGCATCGTTGTAGTCGAT-ATC3' (*BamHI* site underlined, stop codon in boldface, and codon 96 in lower case letters). Note that in this construct the codon for Val96 is retained. The PCR product was digested with *NdeI* and *BamHI* and ligated into the corresponding sites of vector pET3a (New England Biolabs) to create plasmid pMP50. The entire ArgRN gene region of each plasmid was sequenced by the chain-terminator method (Sanger *et al.*, 1977) using oligonucleotide primers complementary to each strand near the center of the ArgR gene.

Total protein extracts

Overnight cultures of *E. coli* cells were diluted 100-fold in fresh LB medium (Sambrook *et al.*, 1989) in the presence of the appropriate antibiotic(s) and incubated for two hours at 37° C. IPTG (1 mM final concentration) was added to the culture in exponential growth, and cells were incubated at 37° C for four to six hours. Cells were collected, washed with water, lysed by resuspension in the electrophoresis loading buffer (Laemmli, 1970), boiled for five minutes, and microfuged at 14,000 rpm 13,000 g. The soluble material was used for electrophoresis.

Ornithine transcarbamylase (OTCase) assay

The assay was performed following a modification of the method of Guthöhrlein & Knappe (1968) as described by Tian *et al.* (1994). Cell samples (0.5 ml) in mid-log phase of growth in the indicated conditions were permeabilized by incubating with 2% (v/v) toluene. From 1 to 50 μ l of cell sample was employed for the reaction. The amount of citrulline produced during the reaction was determined by color development (λ_{max} 490 nm) with diacetyl monoxime in redox buffer using citrulline as a standard. The specific activity of OTCase is defined as μ mol of citrulline produced in the assay per hour per mg total cellular protein, where 1 A_{490} corresponds to 0.64 mg bacterial dry weight (Jacoby & Gorini, 1967).

Protein purification

IPTG-induced cells (12 g wet weight) expressing ArgRN (BL21(λ DE3)*argR*/pMP5O), collected from six liters of culture in minimal E medium (Davis & Mingioli, 1950) containing 0.2% (w/v) glucose, 1.7% (w/v) thiamine, and 100 μ g/ml ampicillin, were washed with Arg buffer (20 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 10 mM β -mercaptoethanol; Lim *et al.*, 1987), resuspended in 35 ml of the same buffer containing 1 mM phenylmethylsulfonyl fluoride (PMSF), and broken using a French press. After centrifugation at 30,000 *g* for 30 minutes, DNA was precipitated from the crude supernatant by adding streptomycin sulfate from a 20% (w/v) stock in water to 1% and incubating for one hour at 4°C with stirring. The sample was then centrifuged at 30,000 *g* for 30 minutes and the supernatant dialyzed against P-11 buffer (10 mM sodium phosphate (pH 7.6), 0.1 mM EDTA, 100 mM NaCl). The sample was applied to a P-11 (cellulose phosphate; Whatman) column prepared according to the manufacturer's protocol and equilibrated at 4°C in P-11 buffer. The column was washed with 100 ml of P-11 buffer and developed with a linear gradient of NaCl from 100 mM to 1 M in P-11 buffer with a flux of 2 ml/min, and fractions of 4 ml were collected and analyzed by SDS-PAGE. The pool of fractions containing ArgRN was concentrated by Centriprep3 concentrators (Amicon), dialyzed against Arg buffer lacking β -mercaptoethanol, clarified by centrifugation at 30,000 *g* for 30 minutes, and loaded onto a 100 ml P-11 column equilibrated in Arg buffer lacking β -mercaptoethanol. Proteins were eluted with 300 ml of the same buffer at a flux of 2 ml/min, and 4 ml fractions were collected and analyzed by SDS-PAGE. The concentrated pool of fractions containing pure ArgRN was brought to 20% (v/v) glycerol, 10 mM β -mercaptoethanol, and stored at -20°C. Protein concentration was determined by quantitative amino acid analysis.

Band-shift assay

Band-shift titration experiments were performed as described by Tian *et al.* (1992) with some modifications. A ³²P end-labeled *Bam*HI-*Eco*RV 120 bp DNA fragment containing the two adjacent ARG boxes of the *argF* operator was prepared from pGT101, a pBend2 derivative (Tian *et al.*, 1992), and purified from an acrylamide gel by the crush-and-soak method of Sambrook *et al.* (1989). DNA concentration was estimated visually from ethidium bromide-stained gels. A ~100 bp control DNA fragment lacking ARG boxes was similarly prepared by *Bam*HI and *Eco*RI digestion of plasmid pRK9; this

fragment contains the *trp*EDCBA promoter/operator of *Serratia marcescens* (Nichols & Yanofsky, 1983). Approximately 0.3 nM DNA fragment was incubated with different amounts of protein in 20 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 10 mM CaCl₂, 100 mM KCl, 10% glycerol for 15 minutes, the conditions used historically for the DNA-binding assay of ArgR (Tian *et al.*, 1992). Incubation and electrophoresis with ArgRN were routinely performed at 4°C, since better detection of complex formation by ArgRN was achieved under these conditions than at the higher temperatures usually employed for intact ArgR (Tian *et al.*, 1992). A total volume of 5 μ l was loaded onto a running 7.5% polyacrylamide gel. Electrophoresis in 100 mM Tris-borate-EDTA buffer (pH 8.8) was performed at a constant voltage (200 V) for one hour.

Size-exclusion chromatography

Samples of pure ArgRN in 200 μ l of Arg buffer were loaded onto a prepacked 30 \times 1 cm Superdex 75 column (Pharmacia) equilibrated at room temperature with Arg buffer. The column was calibrated with molecular weight standards (Sigma): horse heart cytochrome *c* (12,400 Da), carbonic anhydrase (29,000 Da), ovalbumin (45,000 Da), and Blue Dextran (void volume). Proteins were eluted with Arg buffer at a flux of 0.5 ml/min and detected by absorbance at 280 nm. Eluted proteins were collected in 1 ml fractions and aliquots were analyzed by SDS-PAGE.

Analytical ultracentrifugation

All experiments were performed in a Beckman model XLA ultracentrifuge using an An 60 Ti rotor. Data were collected using six-channel Epon charcoal-filled centerpieces with a 12 mm path length containing 110 μ l samples or 125 μ l buffer references. Samples were centrifuged at 20,000, 25,000, 30,000, 35,000, and 40,000 rpm and the protein distribution was monitored at either 241 or 251 nm. Twenty successive radial scans were obtained using a 0.001 cm step-size, and achievement of equilibrium was assumed if no change in distribution was observed after 12 and 24 hours. Global three-speed analysis using a single-species model was performed using software running under Igor (Wavemetrics, Lake Oswego, OR) and incorporating the algorithm of Michael L. Johnson (Johnson *et al.*, 1981). All other analyses, including five-speed and multi-concentration global analyses of single species and monomer \leftrightarrow *n*-mer equilibrium models were performed using the HID program from the Analytical Ultracentrifugation Facility at the University of Connecticut. The partial specific volume for ArgRN (0.744 ml/g) was calculated from the weight average of the partial specific volumes of the individual amino acids (Cohn & Edsall, 1943). The density of the solvent ($r = 1.0125$) was determined gravimetrically.

Acknowledgements

We thank Dongbin Lim for insightful discussions and advice in planning the PCR cloning; Renata Maas for sharing her unpublished innovations for improving the yield of PCR reactions; Joel Oppenheim for the generous gift of anti-ArgR antibodies; Cristina Iftode for early

contributions to protease-sensitivity studies; Heather Morris for providing pure preparations of ArgR and for always useful discussions; Barbara Devlin, Curtis Krier and Saw Kyin from the campus sequencing facility for their expedient assistance. We thank Howard Nash for critical reading of the manuscript and his continuing interest. This project was supported by grants (GM43558 to J.C. and GM06048 to W.K.M.) from the NIH. R.G. was supported by a postdoctoral fellowship of the Ministero Italiano della Pubblica Istruzione.

References

- Bairoch, A. (1992). PROSITE: a dictionary of sites and patterns in proteins. *Nucl. Acids Res.* **20**, 2013–2018.
- 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 ability to bind their normal DNA-binding sites. *Mol. Microbiol.* **13**, 609–618.
- Chakerian, A. T. & Matthews, K. S. (1992). Effect of lac repressor oligomerization on regulatory outcome. *Mol. Microbiol.* **6**, 963–968.
- Charlier, D., Roovers, M., Van Viïet, F., Boyen, A., Cunin, R., Nakamura, Y., Glansdorff, N. & Pierard, A. (1992). Arginine regulon of *Escherichia coli* K-12. *J. Mol. Biol.* **226**, 367–386.
- Cohn, E. J. & Edsall, J. T. (1943). In *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions*, pp. 370–377, Reinhold Publishing Corporation, New York.
- Craig, N. L. & Nash, H. A. (1984). *E. coli* integration host factor binds to specific sites in DNA. *Cell*, **39**, 707–716.
- Czaplewski, L., 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.
- D'Ari, R., Lin, R. T. & Newman, E. B. (1993). The leucine-responsive regulatory protein: more than a regulator? *Trends Biochem. Sci.* **18**, 260–263.
- Davis, B. D. & Mingioli, E. S. (1950). Mutants of *Escherichia coli* requiring methionine or vitamin B12. *J. Bacteriol.* **60**, 17–28.
- Devereux, J., Haeblerli, P. & Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387–395.
- Dolgikh, D. A., Gilmanishin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Venyaminov, S. Y. & Ptitsyn, O. B. (1981). α -Lactalbumin: compact state with fluctuating tertiary structure? *FEBS Letters*, **136**, 311–315.
- Drlica, K. & Rouviere-Yaniv, J. (1987). Histone-like proteins of bacteria. *Microbiol. Rev.* **51**, 301–319.
- Eckhardt, T. (1980). Isolation of plasmids carrying the arginine repressor gene *argR* of *Escherichia coli* K-12. *Mol. Gen. Genet.* **178**, 447–452.
- Fontana, A., de Laureto, P. P. & De Filippis, V. (1993). Molecular aspects of proteolysis of globular proteins. In *Stability and Stabilization of Enzymes* (Van den Tweel, J. J., Harder, A. & Buitelaar, R. M., eds), pp. 101–110, Elsevier Science, Maastricht, The Netherlands.
- Guthöhrlein, G. & Knappe, J. (1968). Modified determination of citrulline. *Anal. Biochem.* **26**, 188–191.
- Jacoby, G. A. & Gorini, L. (1967). A unitary account of the repression mechanism of arginine biosynthesis in *Escherichia coli*. 1. The genetic evidence. *J. Mol. Biol.* **24**, 41–50.
- Jaenicke, R. (1991). Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry*, **30**, 3147–3161.
- Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halverson, H. R. (1981). Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* **36**, 575–588.
- Kosturko, L. D., Daub, E. & Murialdo, H. (1989). The interaction of *E. coli* integration host factor and λ cos DNA: multiple complex formation and protein-induced bending. *Nucl. Acids Res.* **17**, 317–334.
- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.
- Laue, T. M. & Rhodes, D. (1990). Determination of size, molecular weight, and presence of subunits. *Methods Enzymol.* **182**, 566–587.
- Lim, D., Oppenheim, J., 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.
- Lu, C. D., Houghton, J. E. & Abdelal, A. T. (1992). Characterization of the arginine repressor from *Salmonella typhimurium* and its interaction with the *carAB* operator. *J. Mol. Biol.* **225**, 11–24.
- Maas, W. K. (1994). The arginine repressor of *Escherichia coli*. *Microbiol. Rev.* **58**, 631–640.
- Miller, J. H. (1972). *Experiments in Molecular Genetics*, pp. 352–355, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Nichols, B. P. & Yanofsky, C. (1983). Plasmids containing the *trp* promoters of *Escherichia coli* and *Serratia marcescens* and their use in expressing cloned genes. *Methods Enzymol.* **101**, 155–164.
- 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.
- Ohgushi, M. & Wada, A. (1983). Molten-globule state: a compact form of globular proteins with mobile side-chains. *FEBS Letters*, **164**, 21–24.
- Paluh, J. L. & Yanofsky, C. (1986). High level production and rapid purification of the *E. coli trp* repressor. *Nucl. Acids Res.* **14**, 7851–7860.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (Nolan, C., ed.), 2nd edit., pp. 6.46–48 and A.1, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Sander, C. & Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins: Struct. Funct. Genet.* **9**, 56–68.
- Sanger, F., Niklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
- Sauer, R. T. (1995). Minor groove DNA-recognition by α -helices. *Nature Struct. Biol.* **2**, 7–9.
- Schumacher, M. A., Yell Choi, K., Zalkin, H. & Brennan, R. G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by α helices. *Science*, **266**, 763–770.
- 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.
- Spolar, R. S. & Record, M. T., Jr (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.
- Stirling, C. J., Szatmari, G., Stewart, G., Smith, M. C. M. & Sherratt, D. J. (1988). The arginine repressor is essential for plasmid stabilizing site-specific recombination at the ColE1 *cer* locus. *EMBO J.* **7**, 4389–4395.
- Studier, F. W., Rosenberg, 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.
- Tasayco, M. L. & Carey, J. (1992). Ordered self-assembly of polypeptide fragments to form natively like dimeric *trp* repressor. *Science*, **255**, 594–597.
- 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* K-12 to its operator sites. *J. Mol. Biol.* **226**, 387–397.
- Tian, G., Lim, D., Oppenheim, J. D. & Maas, W. K. (1994). Explanation of different types of regulation of arginine biosynthesis in *Escherichia coli* B and *Escherichia coli* K-12 caused by a difference between their arginine repressors. *J. Mol. Biol.* **235**, 221–230.
- Yang, C. & Nash, H. A. (1989). The interaction of *E. coli* IHF protein with its specific binding sites. *Cell*, **57**, 869–880.
- Yang, J. & Carey, J. (1995). Footprint phenotypes: structural models of DNA-binding proteins from chemical-modification analysis of DNA. *Methods Enzymol.* **259**, 452–468.

Edited by R. Schleif

(Received 10 May 1995; accepted 31 August 1995)