J. Mol. Biol. (1994) **243**, 821–829

DNA-Dependent Renaturation of an Insoluble DNA Binding Protein

Identification of the RhaS Binding Site at *rhaBAD*

Susan M. Egan and Robert F. Schleif

Biology Department
The Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, U.S.A.

Previous work has indicated that the RhaS protein directly activates the L-rhamnose catabolic operon, *rhaBAD*, and that the likely RhaS binding site lies downstream of position −84 relative to the *rhaBAD* transcription start point. Biochemical analysis of RhaS binding to this DNA site had not been possible due to the extreme insolubility of overproduced RhaS protein. Here we have been able to analyze directly the DNA binding properties of RhaS by developing a method to refold insoluble RhaS protein into a form with specific DNA binding activity. We found that active RhaS protein could be recovered only if the renaturation reaction was performed in the presence of DNA. We also found that the recovery of DNA-binding activity from the related AraC protein, after denaturation in urea, was dependent upon added DNA. To test the specificity of the recovered RhaS DNA-binding activity, and to define the binding site for comparison with other AraC family binding sites, we then investigated the details of the RhaS binding site. Using refolded RhaS protein in a DNase footprinting assay, we found that RhaS protects a region of the *rhaBAD* promoter from position −83 to −28. Analysis of the effects of single base mutations in the *rhaBAD* promoter region indicates that RhaS binds to an inverted repeat of two 17 bp half-sites separated by 16 bp, located between −81 and −32 relative to the *rhaBAD* transcription start site.

Keywords: rhamnose; RhaS; AraC family; renature; DNA binding protein

1. Introduction

The two regulators of the *Escherichia coli* L-rhamnose operons, RhaS and RhaR, are both members of the AraC family of transcription activator proteins (Tobin & Schleif, 1990*b*; Gallegos *et al.*, 1993; GenBank). RhaR directly activates *rhaSR* transcription (Tobin & Schleif, 1990*a*), while our previous results suggest that RhaS is the direct activator of the L-rhamnose catabolic operon, *rhaBAD* (Egan & Schleif, 1993, Figure 1). All of the AraC family proteins which have been analyzed to date pose serious solubility problems. Some of the family members, such as AraC, RhaR and MelR, are sufficiently soluble when overproduced that biochemical analysis can be directly performed (Schleif & Favreau, 1982; Tobin & Schleif, 1990*b*; Caswell *et al.*, 1992). The RhaS, XylS and SoxS proteins, on the other hand, are extremely insoluble, and DNA binding activity has not been detectable using ordinary methods (Egan & Schleif, 1993; deLorenzo *et al.*, 1993; D. Nguyen & R. Schleif, unpublished results).

The insolubility of the AraC family is not unique, in fact, many proteins form insoluble inclusion bodies *in vivo* when synthesized at high levels. This problem seems to be especially acute for regulatory proteins and other DNA binding proteins. Data from a number of DNA-binding proteins, including TrpR (Zhao *et al.*, 1993; Otwinowski *et al.*, 1988; Schevitz *et al.*, 1985), GCN4 (Talanian *et al.*, 1990; Weiss *et al.*, 1990; O'Neil *et al.*, 1991), λ repressor (Clarke *et al.*, 1991), and the homeodomain of *Antennapedia* (Otting *et al.*, 1990), has recently been interpreted to suggest that the binding motif of many DNA binding proteins may be fully or partially disordered when the proteins are not bound to DNA (Spolar & Record, 1994). The incomplete folding of DNA binding regions seems likely to result in increased solvent exposure of hydrophobic amino acids in some proteins, resulting in a general tendency for such proteins to aggregate when oversynthesized.

One would expect that DNA binding proteins would be fully folded when bound to DNA, resulting

Present address: S. M. Egan, Department of Microbiology, University of Kansas, Lawrence, KS 66045, U.S.A.

Correspondence: R. F. Schleif, Biology Department, The Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, U.S.A.

0022-2836/94/450821-09 $08.00/0

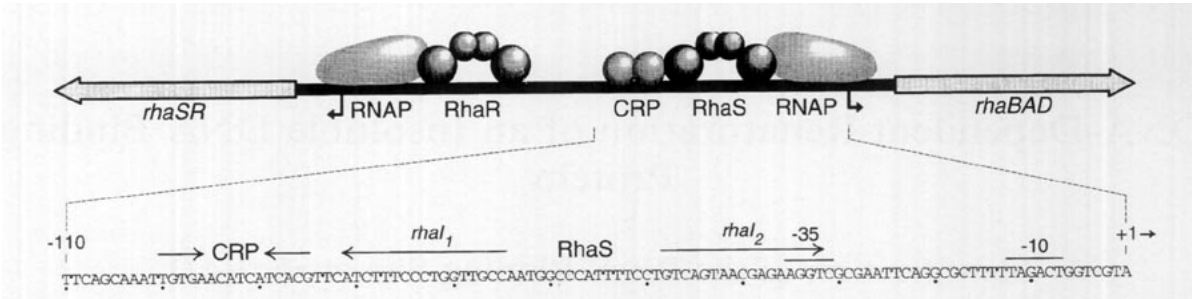

Figure 1. Schematic representation of the *E. coli rha* region. Top: the *rhaSR-rhaBAD* regulatory region with the activator proteins and RNA polymerase illustrated. The position of RhaS is based upon the results of this work. RhaS and RhaR are illustrated as dimers consisting of 2-domain monomers based on the structure of AraC (Bustos & Schleif, 1993). Bottom: DNA sequence of the *rhaBAD* promoter region illustrating the CRP and RhaS protein binding sites.

in the stabilization of the soluble form of the protein. Therefore, we thought that DNA might be capable of promoting the correct refolding of denatured DNA binding proteins. We found that the transcription activator protein RhaS could indeed be renatured from urea in the presence of DNA. DNA could also promote the correct refolding of AraC protein from urea, a reaction which had not been possible in the absence of DNA. It seems likely that this method of renaturation in the presence of DNA will be applicable to other DNA-binding proteins with similar solubility problems.

Based on genetic analysis and coupled transcription-translation assays, we had previously concluded that RhaS was the direct L-rhamnose activator of the operon encoding the L-rhamnose catabolic enzymes, *rhaBAD* (Egan & Schleif, 1993). This work had further suggested that the most likely binding site for RhaS was downstream of position −84 (Egan & Schleif, 1993) relative to the *rhaBAD* transcription start point (Moralejo *et al.*, 1993). Using the above-mentioned DNA-dependent renaturation method, we were able to detect specific binding of the RhaS protein to the predicted region of the *rhaBAD* promoter. In a DNase footprinting assay, RhaS protected a region upstream of *rhaBAD* from −83 to −28 relative to the transcription start site. We further localized the half-sites for RhaS binding by testing the effects of point mutations within the binding site region. This analysis indicates that the RhaS binding site extends from −81 to −32, and consists of two 17 bp inverted repeat half-sites separated by 16 bp.

2. Materials and Methods

(a) *General methods and materials*

General cloning methods were used as described (Sambrook *et al.*, 1989; Schleif & Wensink, 1981). Oligonucleotide primers were made on an Applied Biosystems 381A synthesizer, and treated as described (Egan & Schleif, 1993). They were deprotected at 55°C for 6 to 18 hours, and precipitated directly from this solution with ten volumes of *n*-butanol, not water-saturated (Sawadogo & Van Dyke, 1991). *E. coli* strains used were ECL116 *recA938*::Cm (Backman *et al.*, 1981; Egan & Schleif, 1993) for enzyme assays, and JM109 (*recA1 supE44 endA1 hsdR17 gyrA96 relA1 thi* Δ*(lac-proAB)*/F′*traD36 proAB*$^+$ *lacI*^{q} *lacZ*ΔM15) (Yanisch-Perron *et al.*, 1985) for protein overproduction.

(b) *β-Galactosidase assay*

For enzyme assays, cells were grown in M10 salts (Schleif & Wensink, 1981) plus 0.4% (v/v) glycerol, 50 μM thiamine, 0.2% (w/v) Casamino acids, and when added, 0.2% (w/v) L-rhamnose as previously described (Egan & Schleif, 1993). β-Galactosidase activities are expressed in terms of cell density (A_{600}), using the formula of Miller (1972). Each culture was assayed in duplicate. Reported values are averages of three independent experiments.

(c) *Promoter mutations*

Random mutations in the RhaS binding site region were made using a 93 bp oligonucleotide with degenerate sequence over the region to be mutagenized (Chiang *et al.*, 1993). The level of degeneracy was designed to create an average of 1.5 mutations within the 46 bp mutagenized region. Specific point mutations and the $rhaI_1I_1$ site were constructed using an upstream oligonucleotide which included the altered bases. The sequence of the binding site region in $rhaI_1I_1$ was: 5′-ATCTTTCCCTGGTTGCCAATG-GCCCATTTTCCTGGCAACCAGGGAAAGGT-3′, where the underlined sequences represent the half-sites for binding. DNA fragments for cloning were generated by polymerase chain reaction (PCR†) as described (Bustos & Schleif, 1993). The downstream oligonucleotide in each case allowed construction of a fusion of the first 15 codons of *rhaB* to the ninth codon of *lacZ* in pRS414 (Simons *et al.*, 1987), as previously described (Egan & Schleif, 1993). The template for the *rhaBAD* promoter region was plasmid pSME101 (Egan & Schleif, 1993), which carries the wild-type *rhaSR-rhaBAD* region. The DNA sequence of the entire cloned region was determined for every case. All fusions had an upstream endpoint at −110 from the *rhaBAD* transcription start point, and therefore included the CRP protein binding site.

(d) *RhaS protein preparation*

RhaS protein was overproduced in JM109 cells carrying pJTC108 (Tobin & Schleif, 1990*b*). The cells were grown in 100 ml of YT broth plus ampicillin (200 μg/ml) to mid-logarithmic phase, induced with 1 mM IPTG, and

† Abbreviations used: PCR, polymerase chain reaction; IPTG, isopropyl-thio-β-D-galactoside; CRP, cyclic AMP receptor protein.

allowed to grow overnight at 37°C. The insoluble protein was prepared from a crude fraction of sonicated cells. The sonicated material was centrifuged for 20 minutes at 17,000 *g*. The pellet was extracted with 100 ml of 1% (v/v) Triton X-100 in DNA mobility shift assay running buffer (10 mM Tris-acetate, 1 mM EDTA (pH 7.4); Hendrickson & Schleif, 1984) at room temperature for one hour. The insoluble fraction was repelleted, resuspended in 5 ml of DNA mobility shift assay running buffer and stored at − 70°C. Virtually all of the RhaS, as well as many other insoluble proteins, was recovered. The protein preparations had a total protein concentration on the order of 4 mg/ml, and consisted of approximately 50 to 70% RhaS protein. Assays with AraC were performed with purified protein (Schleif & Favreau, 1982).

(e) *DNA mobility shift assay*

DNA molecules for the mobility shift assays were generated *in vitro* using PCR with primers 744 and Δ153 for the RhaS binding site (Egan & Schleif, 1993) and as described by Carra & Schleif (1993) to make an *araI* site for AraC binding. Protein renaturation reactions were performed in 75 μl volumes containing buffer (10 mM Tris-acetate (pH 7.4), 1 mM EDTA 5% glycerol, 1 mM dithioerythritol), 8 M urea, and either purified AraC protein or an insoluble protein preparation containing RhaS. Reactions with the AraC protein also included 0.05% (v/v) Nonidet P40 (Shell Oil). As indicated, 50 mM L-rhamnose or L-arabinose, and either 50 or 150 mM KCl were added. For the basic renaturation protocol, 30 μg of sheared salmon sperm or calf thymus DNA and approximately 1 ng of ^{32}P-labeled DNA fragment were also included in the renaturation reaction. Otherwise, the ^{32}P-labeled DNA, or both the ^{32}P-labeled DNA and the salmon sperm or calf thymus DNA were added to the reaction after the renaturation (dialysis) was complete, and then incubated with the protein for 20 minutes at room temperature. A 25 μl aliquot of each sample was loaded on a DNA mobility shift assay gel. DNA mobility shift assays were performed as described (Hendrickson & Schleif, 1984). *In vitro* DNase footprinting was performed as described (Carra & Schleif, 1993) in 75 μl reactions.

Dialysis was performed for a total of one hour at room temperature against 1000 volumes of buffer with added KCl and L-rhamnose or L-arabinose, where indicated, with one buffer change after 30 minutes. Dialysis tubing was either Spectra/Por 2, M_r cut off 12,000 to 14,000 (Spectrum Medical Industries) or 20 mm dialysis tubing from Viscase (Chicago, IL). Either tubing could be used after boiling in distilled H_2O, but not after boiling in an EDTA solution. For reasons we do not understand, reactions performed with EDTA-boiled dialysis tubing resulted in a large amount of smearing above the position of the protein-DNA complex, and essentially no band at the expected position. The Spectra/Por tubing could also be used without any pretreatment.

3. Results

(a) *DNA-dependent renaturation of the RhaS and AraC proteins*

Due to the extreme insolubility of overproduced RhaS protein, we have previously been unable to perform biochemical experiments to explore its DNA binding properties. A variety of techniques including synthesis of the protein under lower growth temperatures and co-overproduction of GroESL did not result in sufficient active RhaS for assay. Since we had access to insoluble inclusion bodies which we assume are composed primarily of unfolded RhaS protein, we attempted to renature the protein by dissolving it in urea and dialyzing to allow refolding. Even in the presence of L-rhamnose, however, no detectable DNA binding activity could be recovered (Figure 2a, lanes 6 and 7). We reasoned that DNA might function as a ligand to promote the correct refolding of RhaS. Therefore, the insoluble RhaS protein was dissolved in urea and the dialysis reaction was performed with the protein in the presence of the specific DNA binding site region (labeled with ^{32}P), L-rhamnose, and a large excess of salmon sperm or calf

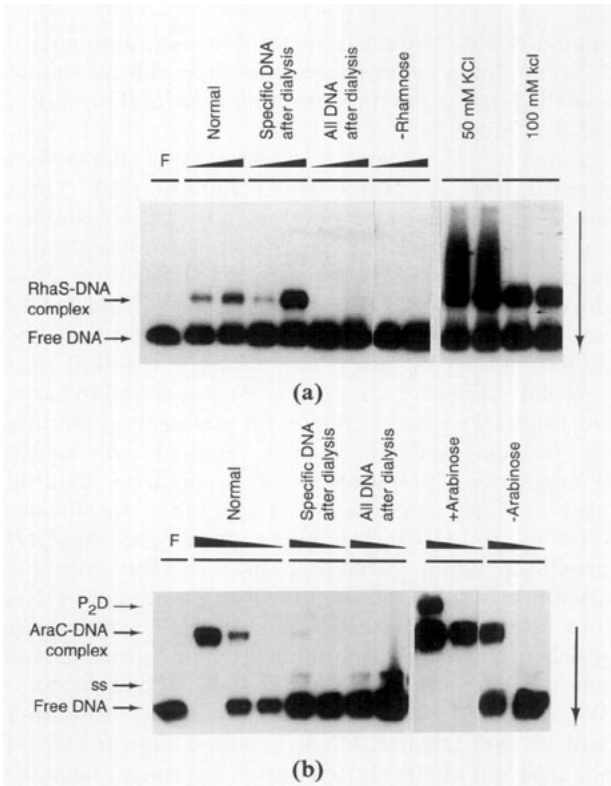

Figure 2.a, DNA mobility shift assays of RhaS binding to the *rhaBAD* promoter region. The vertical arrow indicates the direction of electrophoresis. Normal protocol: renaturation reaction contained RhaS protein preparation, L-rhamnose, 150 mM KCl, calf thymus DNA and the ^{32}P-labeled specific DNA fragment. All other reactions were based on the Normal protocol with the only changes listed above the lanes. Amount of RhaS protein preparation added: lane 1, none; lanes 2, 4, 6 and 8, 1 μl; lanes 3, 5, 7, 9 and 10 to 13, 2 μl. b, DNA mobility shift assays of AraC binding to the *araBAD* promoter region. The vertical arrow indicates the direction of electrophoresis. ss indicates single-stranded DNA. P_2D represents a complex consisting of 2 dimers of AraC protein bound to each DNA molecule. Approximate AraC concentration: lane 1, none; lane 2, 5×10^{-7} M; lanes 3, 5, 7, 9 and 11, 1×10^{-8} M; lanes 4, 6, 8, 10 and 12, 2×10^{-8} M. The AraC protein was estimated to be approximately 50% active prior to the renaturation reactions.

thymus DNA. Using this method in conjunction with DNA mobility shift assays, we were able to detect specific binding of RhaS protein to *rhaBAD* promoter DNA (Figure 2a).

Some of the features of the DNA dependent renaturation of RhaS were characterized. DNA binding activity was recovered only if the reactions contained crude RhaS protein, L-rhamnose, and a large excess of salmon sperm or calf thymus DNA during the dialysis (Figure 2a, lanes 6 to 9, and data not shown). The ^{32}P-labeled DNA fragment could be added before or after the dialysis reaction was complete (Figure 2a, lanes 2 to 5), but it had to carry the *rhaBAD* promoter region for any protein binding to be detected (data not shown). A nearly homogeneous preparation of a maltose binding protein-RhaS fusion was also capable of specifically binding to DNA fragments from the *rhaBAD* promoter region, indicating that accessory proteins are unlikely to be required for RhaS binding to DNA (our unpublished results). Nonidet P40, a detergent which increases the solubility of AraC protein, was found to decrease the recovery of active RhaS protein. With 50 mM KCl in the renaturation/binding buffer, the labeled DNA smeared upward from the position of the shifted band (Figure 2a, lanes 10 and 11). This smearing could be eliminated by performing the renaturation reactions in the presence of 100 to 150 mM KCl (Figure 2a, lanes 12 and 13), suggesting that the smearing was due to protein aggregation.

We tested whether DNA could also promote the refolding of AraC protein from a urea solution. Purification of AraC is difficult and its low solubility gives rise to very large losses during purification. The AraC protein which is purified is stable and soluble, and has been used to perform extensive DNA-binding studies (Brunelle & Schleif, 1989; Carra & Schleif, 1993; Hendrickson & Schleif, 1984). AraC protein could also be renatured from urea, and its activity also depended upon the presence of DNA (Figure 2b). When the renaturation was performed in the absence of L-arabinose, approximately tenfold more AraC protein was required to bind the same amount of DNA, consistent with the previous finding that AraC binds DNA with a higher affinity in the presence of L-arabinose (Hendrickson & Schleif, 1984).

The recovery of detectable DNA-binding activity from both the RhaS and AraC proteins was entirely dependent on the presence of DNA during the renaturation reaction. At 150 mM KCl, an excess of salmon sperm or calf thymus DNA seemed to promote the recovery of active RhaS protein as well as the specific ^{32}P-labeled binding site (Figure 2a). At 50 mM KCl, however, the addition of the specific binding site to the dialysis reaction increased the amount of binding of both RhaS and AraC to their respective binding sites by approximately fivefold (Figure 2b, and data not shown). We believe that under the lower salt conditions, protein that dissociated from the salmon sperm or calf thymus DNA had a greater tendency to aggregate and resulted in less binding to the specific DNA fragment.

(b) *The* in vitro *footprint of RhaS-DNA complex supports the predicted binding site*

Based on deletion analysis of the *rhaBAD* promoter region and inspection of the DNA sequence, we had previously predicted that the most likely binding site for RhaS was an inverted repeat consisting of 17 bp half-sites separated by 16 bp and overlapping the −35 region of the promoter by 4 bp (Egan & Schleif, 1993). Using DNA-dependent renaturation of RhaS to form a protein-DNA complex, we performed DNase footprinting of RhaS bound at *rhaBAD*. We found that RhaS protected a region of the DNA that included the entire predicted binding site, and extended several bases beyond the site in each direction (Figure 3). A DNA mobility shift assay of a portion of each sample showed that essentially all of the DNA was shifted in the samples with added protein (data not shown). The extent of the footprint on each of the two DNA strands was similar. In addition, there was a single base on the top strand that was hypersensitive to DNase cleavage with protein bound compared with free DNA. Interestingly, the 16 bp region predicted to lie between the two half binding sites was also largely protected from DNase cleavage. The extent of this footprint indicates that the RhaS binding site at *rhaBAD* lies between positions −83 and −30, but does not identify the exact half-sites for this binding.

(c) *Point mutations in the predicted RhaS binding site*

To identify further the half-sites for RhaS binding, we analyzed the induction from *rhaBAD* promoter fragments containing point mutations throughout the region. Each of the promoter fragments was fused with *lacZ* and assayed for β-galactosidase activity in the absence and presence of L-rhamnose. Since assay of the effects of these mutations depended upon transcription, we did not mutagenize the portion of the predicted binding site which overlaps the −35 region of the promoter. We isolated or constructed single base changes in the binding site region at 31 of the 46 possible positions. Many of the mutations had little or no effect on the promoter activity (Figure 4). Single base changes at eight positions had a significant effect on the level of RhaS-mediated transcription activation from the promoter (Figure 4). These bases are all located within the predicted 17 bp half-sites. When viewed from one face of the DNA helix, 17 bp consist of two major grooves and the intervening minor groove. The eight base changes that effected transcription all lie within the major grooves of the binding site when viewed from this face. This is consistent with the previous finding that the AraC and RhaR proteins contact their DNA binding sites primarily within two major grooves of each half-site, and that the contacts appear to be made along one face of the DNA (Brunelle & Schleif, 1989;

Tobin & Schleif, 1990*b*). Interestingly, one of the major groove regions of each RhaS half-site is nearly identical to the corresponding half-site for RhaR binding (Figure 5).

(d) *Spacing mutations between the two RhaS half-sites*

We isolated two different single base-pair deletions in the region between the RhaS half-sites (Table 1). These deletions had a more severe affect on activation from the promoter than any of the single base substitutions analyzed. This result suggested either that RhaS binding to DNA was intolerant of all changes in the half-site spacing, or that the wild-type spacing between the half-sites, 16 bp, was the lower limit of permissible spacings. To differentiate between these possibilities, we constructed binding sites with 1, 2, or 3 bp insertions between the half-sites. We found that binding sites with 1 or 2 bp insertions behaved essentially as the wild-type site (Table 1). Insertion of 3 bp resulted in reduced activation from the promoter, although not to the same extent as the 1 bp deletions. These results support the idea that 16 bp is the lower limit to the tolerable spacing between RhaS half-sites, and also suggest that 18 bp is the upper limit. Again, this is similar to the RhaR binding site where the two half-sites are separated by 17 bp.

(e) *Analysis of* rhaI$_1$I$_1$ *binding site*

The promoter distal half-site for AraC binding at *araBAD*, *araI*$_1$, binds much more strongly than the promoter proximal half-site, *araI*$_2$ (Carra & Schleif, 1993). Transformation of *araI*$_1$ to *araI*$_2$ results in a high level of L-arabinose independent transcription activation from the *araI*$_1$*I*$_1$ promoter (Reeder & Schleif, 1993). To characterize further the RhaS binding site, we constructed an almost perfect inverted repeat of the promoter-distal RhaS half-site, *rhaI*$_1$ (see Materials and Methods, and Figure 4), and tested the ability of this binding site to result in transcription activation. Seven of the eight mismatches in *rhaI*$_2$ relative to *rhaI*$_1$ were changed, while the single mismatch within the −35 region of the promoter was not changed. We found that transcription activation from the *rhaI*$_1$*I*$_1$ promoter, both in the absence and presence of L-rhamnose, was essentially the same as from the wild-type promoter (Figure 5). Thus, the *ara* and *rha* systems differ significantly, at least at this level.

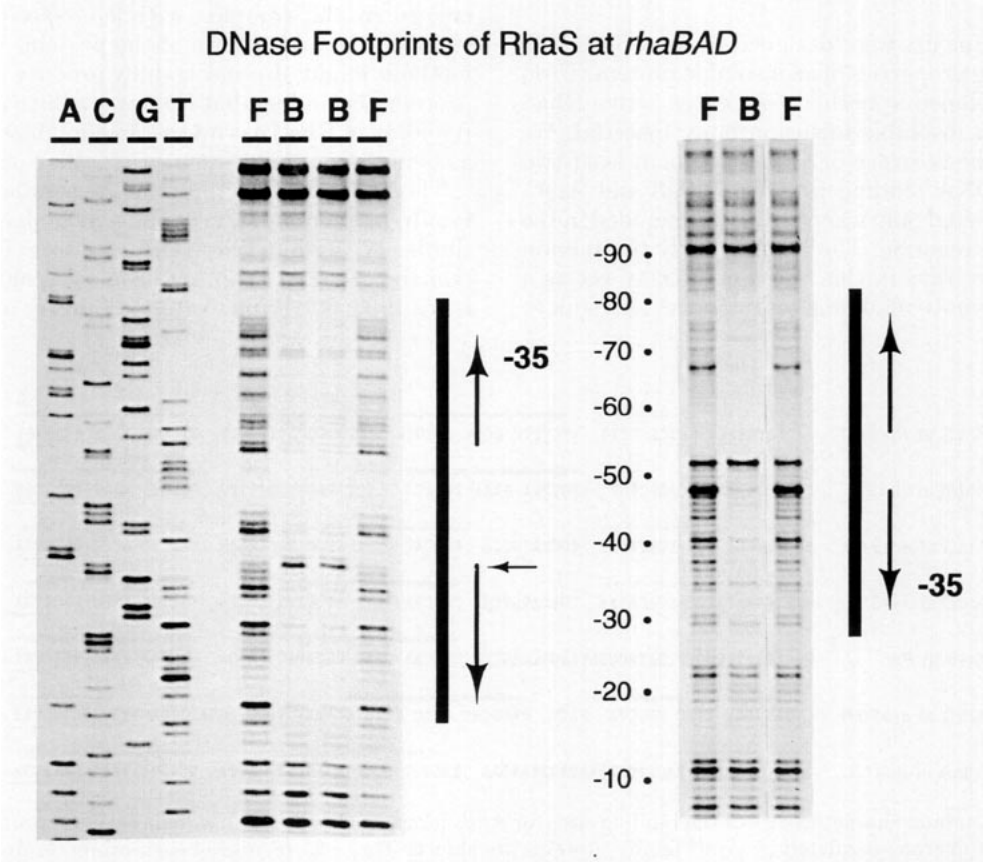

Figure 3. DNase footprint of the RhaS binding site at *rhaBAD*. The top strand is on the left and the bottom strand is on the right. Vertical bars represent the approximate limits of the RhaS protein-induced DNase protection. The horizontal arrow indicates the position of a RhaS protein-induced hypersensitive site. Vertical arrows indicate the position of the 2 RhaS half-sites for binding. Sequencing reactions were performed using the method of Sanger *et al.* (1977) employing the same ^{32}P-labeled oligonucleotide as used to generate the DNA fragment for footprinting. Dots indicate the position in the sequence relative to the *rhaBAD* transcription start point. F, no added RhaS preparation; B, added RhaS preparation.

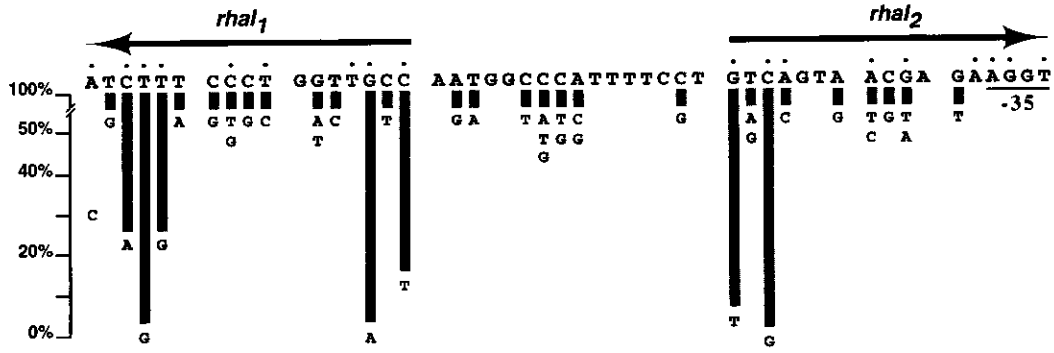

Figure 4. Effect of point mutations throughout the RhaS binding site region as assayed by β-galactosidase activity of *rhaB-lacZ* translational fusions to each construct. The DNA sequence of the region upstream of *rhaBAD* from − 81 to − 32 is shown. Single base changes analyzed are indicated by the bars below the wild-type sequence, with the specific nucleotides substituted indicated at the bottom of each bar. In the absence of L-rhamnose, the β-galactosidase activity from all of the substituted promoter regions was indistinguishable from that of wild-type. The relative β-galactosidase activity of each of the promoters in the presence of L-rhamnose is indicated by the length of the bar relative to the scale at the left. The wild-type promoter activity, 22,000 Miller units, was set to 100%. Promoters with activity between 50 and 100% of the wild-type are all indicated by short bars to the break-point in the scale. Horizontal arrows indicate the position of the half-sites for RhaS binding. Dots above the DNA sequence indicate the positions in the inverted repeat sequence in which there is a match between $rhaI_1$ and $rhaI_2$.

4. Discussion

(a) *Renaturation of RhaS and AraC is promoted by DNA*

Our experiments were designed to investigate the DNA binding properties of an insoluble protein. To do this, we developed a method to recover active RhaS protein from insoluble inclusion body material. We found that the recovery of activity from at least two denatured DNA binding proteins, RhaS and AraC, could be detected only when DNA was included in the renaturation reaction. The DNA is likely to function in one of two ways in this reaction. It may act as a ligand to promote refolding, or it may act as a sink to trap correctly folded protein to prevent unfolding and aggregation. The ability of an excess of non-specific DNA (calf thymus or salmon sperm) to promote the refolding suggests that there is sufficient binding energy in the complex with non-specific DNA to stabilize the folded form of the protein. While other methods might also successfully produce active RhaS protein, the finding that DNA was able to promote the refolding of RhaS has interesting implications, as well as being potentially useful with other proteins.

The RhaS and AraC proteins are members of a large family of regulatory proteins which share sequence similarity within the DNA binding (C-terminal) domain of AraC (Tobin & Schleif, 1990*b*; Gallegos *et al.*, 1993; GenBank, National Library of Congress).

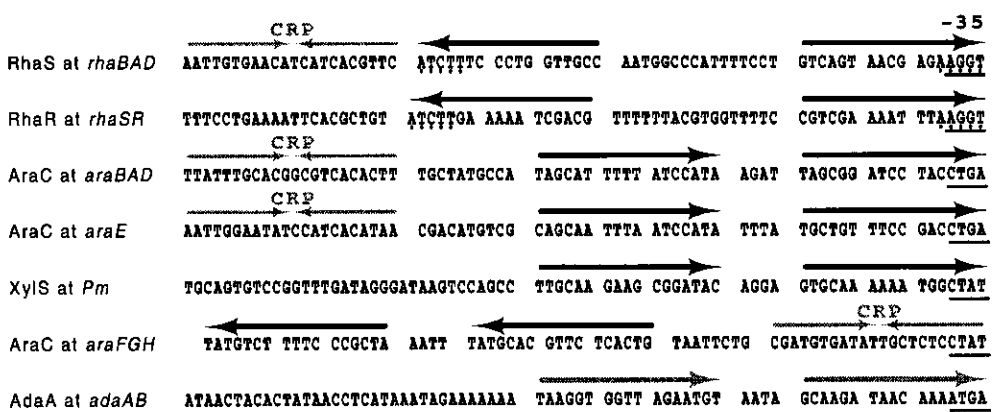

Figure 5. Common characteristics of the binding sites for AraC family proteins. DNA sequences of the promoter regions of a variety of operons regulated by AraC family proteins are shown. The − 35 regions of each promoter are underlined. Black arrows indicate the half-sites for binding the appropriate AraC family protein. The DNA sequence within each of these half-sites is divided into the approximate boundries of 2 major groove regions and the intervening minor groove region. Smaller gray arrows indicate the binding sites for the CRP protein. Larger gray arrows indicate a tentative binding site for the AdaA protein at *adaAB* based on the extent of the footprint (Morohoshi *et al.*, 1993) and the rules presented here. Dots below the DNA sequence indicate the positions of matching sequence between the similar major groove regions of the RhaS and RhaR protein binding sites.

Table 1

Effects of insertions and deletions between the RhaS half-sites, and $rhaI_1I_1$ *on* rhaB-lacZ *fusion expression*

Promoter	Nucleotide change‡	Position of change§	β-Galactosidase (spec. act.)†	
			− L-Rhamnose	+ L-Rhamnose
A. WT			1	22,000
18	− 1 (G)	− 61	0.4	65
3	− 1 (C)	− 59	0.6	50
343	+ 1 (A)	− 60.5	0.7	23,000
347	+ 2 (AT)	− 60.5	0.6	20,000
345	+ 3 (ATA)	− 60.5	0.9	1000
B. WT			14	11,000
$rhaI_1I_1$			3	13,000

† Determined as described in Materials and Methods and expressed in Miller units. Cultures were grown in a minimal glycerol medium with or without L-rhamnose, as indicated.

‡ Indicates the number of nucleotides either deleted or inserted relative to the wild-type sequence.

§ Indicates the position of the base(s) deleted or inserted relative to the *rhaBAD* transcription start site. The insertions were made between the bases at positions − 60 and − 61.

While limited solubility seems to be a general feature of the AraC family, severe insolubility has been found with a number of the members of this family, including RhaS, XylS and SoxS (Egan & Schleif, 1993; deLorenzo *et al.*, 1993; D. Nguyen & R. Schleif, unpublished results). It seems very likely that other members of the AraC family, as well as unrelated DNA-binding proteins, could also be renatured using a DNA-dependent method similar to that described here.

An unusual feature of DNA binding by the AraC protein and its family members concerns the large extent of the protein-DNA interaction. The dimeric AraC protein contacts a total of four major grooves of DNA (Lee *et al.*, 1987; Carra & Schleif, 1993), compared with the two contacted by most dimeric regulatory proteins. A protein-DNA interaction over four major grooves would be expected to result in such strong binding that the protein might be "glued" to the promoter region, eliminating the ability to regulate dynamically transcription. If the DNA-binding motif of AraC and its family members were partially unfolded when not bound to DNA then some of the binding energy could be used to complete the folding and thereby weaken binding to biologically manageable levels.

(b) *RhaS binding site at* rhaBAD

Once we determined that RhaS could bind to *rhaBAD* promoter DNA, we investigated the details of the RhaS binding site. This was done both to test the specificity of the recovered DNA binding activity, and to define the RhaS binding site for comparison with binding sites of other AraC family proteins. Using DNA-dependent renaturation of RhaS protein to test DNA binding, in combination with mutagenesis of the predicted RhaS binding site, we have determined that RhaS binds to a site with inverted repeat half-sites, separated by 16 bp and overlapping the − 35 region of the promoter by 4 bp. This binding site shares some features in common with the AraC protein DNA binding site, but differs in other ways. A consensus is beginning to emerge for the characteristics of AraC family binding sites, as described below, based on the sites for AraC, RhaR, RhaS and XylS (Carra & Schleif, 1993; Tobin & Schleif, 1990*b*; Kessler *et al.*, 1993).

Based on our results, we can draw some preliminary conclusions about the differences between transcription activation by RhaS versus AraC. AraC binds at the promoter distal half-site, $araI_1$, both in the presence and absence of L-arabinose, and transcription activation results when arabinose induces the protein to contact the promoter proximal half-site, $araI_2$ (Lee *et al.*, 1987; Lobell & Schleif, 1990). Reeder & Schleif (1993) found that the L-arabinose requirement for induction could be overcome if the promoter proximal half-site was changed to match the sequence of the distal half-site, $araI_1I_1$. A key component to this mechanism is that the distal *ara* half-site is a tighter binding site than the promoter proximal half-site. Our finding that the $rhaI_1I_1$ promoter did not show elevated L-rhamnose-independent expression argues against this type of mechanism for *rhaBAD* induction. Rather, it seems more likely that RhaS is only bound at the promoter region in the presence of L-rhamnose. This is consistent with the finding that both the amount and the DNA-binding activity of RhaS protein are higher in the presence than in the absence of L-rhamnose (Figure 2; Egan & Schleif, 1993).

(c) *DNA binding sites of AraC family proteins*

How is it that some AraC family proteins bind to DNA half-sites arranged as inverted repeats (Figure 4; Tobin & Schleif, 1990*b*), while others have half-sites arranged as direct repeats (Carra & Schleif, 1993; Kessler *et al.*, 1993; Caswell *et al.*, 1992)? Carra & Schleif (1993) found that AraC protein can bind to half-sites arranged as either direct or inverted repeats,

although the binding to the inverted site was weaker. They proposed that the dimerization and DNA binding domains of AraC are connected by a flexible linker region that allows the orientation of the DNA binding domain to be independent of that of the dimerization domain. Recent experiments have identified such a flexible linker between the domains of AraC (Bustos & Schleif, 1993; Eustance *et al.*, 1994). Preliminary evidence suggests that the RhaR protein is also flexible and can bind to a direct repeat in addition to its natural inverted repeat half-site orientation (B. Saviola & R. Schleif, unpublished results). Given this degree of flexibility it is easy to imagine that proteins related to AraC could bind strongly to either direct or inverted repeat sequences.

The currently available data suggest certain common characteristics in the DNA recognition sites of AraC family proteins (Figure 5). An understanding of these characteristics should begin to allow prediction of the half-sites for binding of new members of the family. In the binding sites studied thus far, the half-sites for binding are approximately 17 bp in length. As mentioned above, 17 bp half-sites include two major grooves and the intervening minor groove of DNA when viewed from one face of the helix. These half-sites can be arranged as either direct or inverted repeats, although the number of mismatches between the repeats can be fairly large. This is apparently due to the distribution of the binding energy over a large number of DNA contacts in each half-site. The spacing between the half-sites can vary, but the centers of the major grooves in adjacent half-sites are always separated by nearly integral turns of the DNA helix. When functioning as an activator, the AraC family protein binding site generally lies immediately upstream of the core promoter with a 4 bp overlap of the −35 region. Reeder & Schleif (1993) found that this was essentially the only promoter-proximal position from which AraC could activate transcription, the single exception being a 2 bp overlap with the −35 region.

There are two current examples of AraC family regulators with binding sites located further upstream such that they do not overlap the −35 region of the core promoter: AraC activation of *araFGH* and MelR activation of *melAB* (Hendrickson *et al.*, 1990; Caswell *et al.*, 1992). At *araFGH*, CRP protein binds at a site centered at −41.5, and AraC binds just upstream of CRP. Interestingly, this operon appears to be especially sensitive to catabolite repression and is fully activated only under conditions of extreme CRP activation (C. Johnson & R. F. Schleif, unpublished results). MelR binds at *melAB* between about −110 and −50, but it is not yet clear whether a second activator is involved in this case (Caswell *et al.*, 1992).

We thank the members of our laboratory and John Carra for critical discussions and comments on the manuscript, and Jeff Withey for technical assistance. This work was supported by National Institutes of Health grant GM18277 to R. S. S.M.E. is the recipient of post-doctoral fellowship GM14364 from the National Institutes of Health.

References

Backman, K., Chen, Y.-M. & Magasanik, B. (1981). Physical and genetic characterization of the *glnA-glnG* region of the *Escherichia coli* chromosome. *Proc. Nat. Acad. Sci., U.S.A.* **78**, 3743–3747.

Brunelle, A. & Schleif, R. (1989). Determining residue-base interactions between AraC protein and *araI* DNA. *J. Mol. Biol.* **209**, 607–622.

Bustos, S. A. & Schleif, R. F. (1993). Functional domains of the AraC protein. *Proc. Nat. Acad. Sci., U.S.A.* **90**, 5638–5642.

Carra, J. H. & Schleif, R. F. (1993). Variation of half-site organization and DNA looping by AraC protein. *EMBO J.* **12**, 35–44.

Caswell, R., Webster, C. & Busby, S. (1992). Studies on the binding of the *Escherichia coli* MelR transcription activator protein to operator sequences at the *melAB* promoter. *Biochem. J.* **287**, 501–508.

Chiang, L. W., Kovari, I. & Howe, M. (1993). Mutagenic oligonucleotide-directed PCR amplification (Mod-PCR): an efficient method for generating random base substitution mutations in a DNA sequence element. *PCR Methods Applicat.* **2**, 210–217.

Clarke, N. D., Beamer, L. J., Goldberg, H. R., Berkower, C. & Pabo, C. O. (1991). The DNA binding arm of λ repressor: critical contacts from a flexible region. *Science,* **254**, 267–270.

deLorenzo, V., Eltis, L., Kessler, B. & Timmis, K. N. (1993). Analysis of *Pseudomonas* gene products using *lacIq/P trp-lac* plasmids and transposons that confer conditional phenotypes. *Gene,* **123**, 17–24.

Egan, S. M. & Schleif, R. F. (1993). A regulatory cascade in the induction of *rhaBAD*. *J. Mol. Biol.* **234**, 87–98.

Eustance, R. J., Bustos, S. A. & Schleif, R. F. (1994). Reaching out: locating and lengthening the interdomain linker in AraC protein. *J. Mol. Biol.* **242**, 330–338.

Gallegos, M.-T., Michán, C. & Ramos, J. L. (1993). The XylS/AraC family of regulators. *Nucl. Acids Res.* **21**, 807–810.

Hendrickson, W. & Schleif, R. F. (1984). Regulation of the *Escherichia coli* L-arabinose operon studied by gel electrophoresis DNA binding assay. *J. Mol. Biol.* **174**, 611–628.

Hendrickson, W., Stoner, C. & Schleif, R. (1990). Characterization of the *Escherichia coli araFGH* and *araJ* promoters. *J. Mol. Biol.* **215**, 497–510.

Kessler, B., de Lorenzo, V. & Timmis, K. N. (1993). Identification of a *cis*-acting sequence within the *Pm* promoter of the TOL plasmid which confers XylS-mediated responsiveness to substituted benzoates. *J. Mol. Biol.* **230**, 699–703.

Lee, N., Francklyn, C. & Hamilton, E. (1987). Arabinose-induced binding of AraC protein to *araI2* activates the *araBAD* promoter. *Proc. Nat. Acad. Sci., U.S.A.* **84**, 8814–8818.

Lobell, R. B. & Schleif, R. F. (1990). DNA looping and unlooping by AraC protein. *Science,* **250**, 528–532.

Miller, J. H. (1972). *Experiments in Molecular Genetics.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Moralejo, P., Egan, S. M., Hidalgo, E. & Aguilar, J. (1993). Sequencing and characterization of a gene cluster encoding the enzymes for L-rhamnose metabolism in *Escherichia coli*. *J. Bacteriol.* **175**, 5585–5594.

Morohoshi, F., Hayashi, K. & Munakata, N. (1993). *Bascillus subtilis alkA* gene encoding inducible 3-methyladenine DNA glycosylase is adjacent to the *ada* operon. *J. Bacteriol.* **175**, 6010–6017.

O'Neil, K. T., Hoess, R. J. & DeGrado, W. F. (1991). DNA induced increase in the α-helical content of C/EBP and GCN4. *Biochemistry*, **30**, 9030–9034.

Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J.* **9**, 3085–3092.

Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). Crystal structure of *trp* repressor/operator complex at atomic resolution. *Nature (London)*, **335**, 321–329.

Reeder, T. & Schleif, R. (1993). AraC protein can activate transcription from only one position and when pointed in only one direction. *J. Mol. Biol.* **231**, 205–218.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Nat. Acad. Sci., U.S.A.* **174**, 5463–5467.

Sawadogo, M. & Van Dyke, M. W. (1991). A rapid method for the purification of deprotected oligodeoxynucleotides. *Nucl. Acids Res.* **19**, 674.

Schevitz, R. W., Otwinowski, Z., Joachimiak, A., Lawson, C. L. & Sigler, P. B. (1985). The three-dimensional structure of *trp* repressor. *Nature (London)*, **317**, 782–786.

Schleif, R. F. & Favreau, M. A. (1982). Hyperproduction of *araC* protein from *Escherichia coli*. *Biochemistry*, **21**, 778–782.

Schleif, R. F. & Wensink, P. (1981). *Practical Methods in Molecular Biology*. Springer-Verlag, New York.

Simons, R. W., Houman, F. & Kleckner, N. (1987). Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene*, **53**, 85–96.

Spolar, R. S. & Record, M. T. Jr (1994). Coupling of local folding to site-specific binding of proteins to DNA. *Science*, **263**, 777–784.

Talanian, R. V., McKnight, C. J. & Kim, P. S. (1990). Sequence-specific DNA binding by a short peptide dimer. *Science*, **249**, 769–771.

Tobin, J. F. & Schleif, R. F. (1990*a*). Transcription from the *rha* operon p_{SR} promoter. *J. Mol. Biol.* **211**, 1–4.

Tobin, J. F. & Schleif, R. F. (1990*b*). Purification and properties of RhaR, the positive regulator of the L-rhamnose operons of *Escherichia coli*. *J. Mol. Biol.* **211**, 75–89.

Weiss, M. A., Ellenberger, T., Wobbe, R., Lee, J. P., Harrison, S. C. & Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature (London)*, **347**, 575–587.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, **33**, 103–119.

Zhao, D., Arrowsmith, C. H., Jia, X. & Jardetzky, O. (1993). Refined solution structures of the *Escherichia coli trp* holo- and aporepressor. *J. Mol. Biol.* **229**, 735–746.

Edited by P. von Hippel

(Received 10 June 1994; accepted 23 August 1994)