J. Mol. Biol. (1993) **231**, 205–218

AraC Protein Can Activate Transcription from Only One Position and When Pointed in Only One Direction

Thadd Reeder and Robert Schleif

Biology Department, Johns Hopkins University
34th and Charles Streets, Baltimore, MD 21218, U.S.A.

(Received 6 October 1992; accepted 6 January 1993)

At the *araBAD* promoter, the RNA polymerase-proximal half-site for AraC binding partially overlaps the −35 region. Random and explicit spacing experiments show that both this partial overlapping and AraC binding to the polymerase-proximal half-site are necessary and sufficient for strong transcriptional activation. Normally, this occupancy is generated by the presence of arabinose, which shifts AraC from a DNA looping interaction involving the polymerase-distal half-site and the $araO_2$ site 210 base-pairs away, to an interaction with the two half-sites adjacent to RNA polymerase. Changing the polymerase-proximal half-site to a higher affinity AraC binding site gives activation in the absence of arabinose. Thus, arabinose is not required to transform AraC into an activating conformation. Because the two half-sites of *araI* are direct repeats, the RNA polymerase proximal and distal surfaces of AraC are not identical. When the *araI* site was turned around, no spacings were found from which AraC could activate transcription. In light of the strict spacing and orientation requirements for AraC activation, the interactions between AraC and RNA polymerase are likely to be specific and inflexible.

Keywords: AraC; arabinose; transcriptional activation; promoter–RNA polymerase interactions

1. Introduction

Like most other prokaryotic and eukaryotic genes, the genes that enable *Escherichia coli* to catabolize L-arabinose are positively regulated at the transcriptional level (Englesberg *et al.*, 1965; Greenblatt & Schleif, 1971; Lee *et al.*, 1974). Transcription from the catabolic *araBAD* operon, as well as operons involved in the uptake of arabinose, is regulated by AraC, a transcriptional activator protein (Schleif, 1987). At the *araBAD* promoter, P_{BAD}, in the absence of arabinose, the dimeric AraC protein contacts both the promoter-distal half of the *araI* site, whose position is called I_D and whose sequence is called I_1, and the $araO_2$ half-site that is located 211 base-pairs away, thus forming a DNA loop (Fig. 1: Dunn *et al.*, 1984; Lobell & Schleif, 1990). In this state, the operon is uninduced and expresses only a low basal level of the *ara* enzymes. In the presence of arabinose, the monomer of AraC bound at $araO_2$ relocates to the promoter-proximal half of the *araI* site, whose position is called I_P and whose sequence is called I_2, and transcription of P_{BAD} is induced (Lee *et al.*, 1987; Lobell & Schleif, 1990). These facts raise two questions. Is arabinose necessary to drive AraC into an activating conformation, or might occupancy of the promoter-proximal half-site be sufficient for activation if it could be accomplished in the absence of arabinose? Second, must AraC activate transcription from a site immediately adjacent to RNA polymerase, or can it act like an enhancer-binding protein and activate from a wide variety of positions?

The *ara* promoters are also transcriptionally regulated by the catabolite activator protein (CAP). CAP stimulates transcription from the *araBAD* promoter in an AraC-dependent manner. Part of

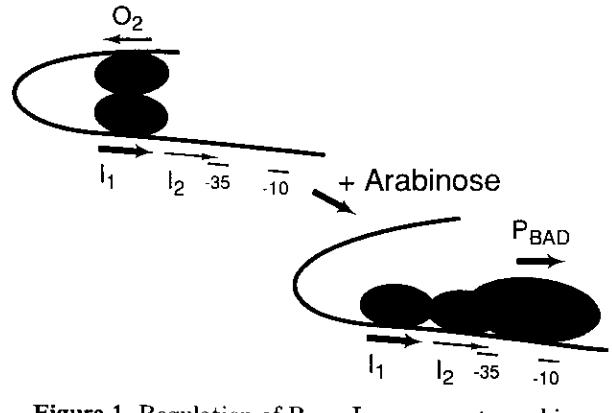

Figure 1. Regulation of P_{BAD}. In response to arabinose, AraC shifts occupancy from the O_2-I_1 half-sites to the I_1-I_2 half-sites, leading to transcriptional activation of P_{BAD}.

0022-2836/93/100205-14 $08.00/0

this stimulation is due to CAP breaking the repression loop generated between O_2 and I_1 (Hahn *et al.*, 1986; Lobell & Schleif, 1991), and part is independent of looping. The question of how CAP helps AraC to activate transcription was not addressed in these experiments.

Recent work has clarified the structure of the *araI* site (Lu *et al.*, 1992; Carra & Schleif, 1993). On the basis of DNase footprinting, contact experiments and sequence homology, *araI* was once thought to have an inverted repeat structure covering three major groove regions and to be located entirely upstream from the RNA polymerase site (Hendrickson & Schleif, 1985). We now know, however, that the *araI* site possesses a direct repeat structure of two 17 base-pair half-sites separated by four base-pairs, each half-site consisting of two major-groove regions (Carra & Schleif, 1993). The last four base-pairs of the promoter-proximal half-site, I_2, overlap the first four of the six base-pair −35 region of the *araBAD* promoter. This structure raises additional questions. Must the AraC binding site overlap the −35 region, and does activation by AraC occur only from one of the two possible orientations of the *araI* direct repeat?

We have exploited enzymatic and chemical tools to address the above questions. The AraC binding site was randomly placed at many different positions upstream from and overlapping with the P_{BAD} RNA polymerase binding site, also called the core promoter. Constructs that could be activated by AraC were then selected. We found the following: AraC can strongly activate transcription only when the promoter-proximal half-site, I_P, overlaps the −35 region of a promoter; the presence of arabinose is not necessary for AraC to induce transcription; the position of the promoter-distal half-site, I_D, is important in achieving full activation by AraC; and only one orientation of the *araI* direct repeat is active in induction. These results suggest that activation of RNA polymerase by AraC requires specific and relatively inflexible protein–protein interactions between a properly positioned and oriented AraC molecule and RNA polymerase. The amino acid residues that make these interactions are likely to be located in the monomer of AraC bound at I_P.

2. Materials and Methods

(a) *General methods, media and reagents*

General methods were as described (Schleif & Wensink, 1981; Sambrook *et al.*, 1989). Cells were grown in YT medium, or for enzyme assays, M10 medium containing 0·2% (v/v) glycerol, 10 μg thiamine/ml, 0·4% (v/v) Casamino acids, 40 μg leucine/ml and, when added, 0·2% (w/v) arabinose. Plates suitable for detection of high levels of β-galactosidase activity contained A+B salts (Schleif & Wensink, 1981) and 0·2% glycerol, 10 μg thiamine/ml, 0·3% Casamino acids, 0·04% (w/v) β-lactose, 40 μg leucine/ml, 30 μg 5-bromo-4-chloro-3-indolyl-β,D-galactopyranoside (X-gal)/ml and, when added, 0·2% arabinose. Fourteen hours after spreading transformed cells on these plates, only cells that express over 2000 Miller units yield solid blue colonies, and cells that express over 15,000 Miller units are dark blue. Plates as well as cultures, when appropriate, contain 100 μg streptomycin/ml, 15 μg tetracycline/ml and 100 μg ampicillin/ml. Cultures and plates were grown at 37 °C and β-galactosidase assays were done as described (Miller, 1972). Enzyme units are Miller units: $1000 \times A_{420}/(A_{600} \times t\,(\text{min}) \times \text{vol (ml)})$ and have a standard deviation of less than 15%.

Chemicals and reagents were purchased from standard sources. *Bal*31 exonuclease (slow, Wei *et al.*, 1983) was from International Biotechnologies, Inc. RNAzol (Chomczynski & Sacchi, 1987) was from Cinna/Biotecx and RNase inhibitor from U.S. Biochemical.

Oligonucleotides were synthesized on an Applied Biosystems 381A synthesizer using chemicals from Chruachem. Oligonucleotides were deprotected in 30% (w/v) NH_4OH for 15 min at room temperature, separated from the support beads, and incubated for another 6 h at 55 °C, precipitated with ethanol and purified on a Pharmacia FPLC Mono-Q anion exchange column (Cubellis *et al.*, 1985).

(b) *Strains and plasmids*

Isogenic AraC$^-$ (TR321) and AraC$^+$ (TR322) strains were made by crossing the F′ of XL1-Blue (Stratagene) into strains SH321 and SH322 (Hahn *et al.*, 1984). The resulting strains were TR321, F′ *proAB*, *lacI*^{q}ZΔM15, Tn*10*(*tet*^{r})/Δ*araC-leu* 1022, *araB*$^+$*A*$^+$*D*$^+$, Δ*lac74*, *galK*$^-$, *str*^{r} and TR322, F′ *proAB*, *lacI*^{q}ZΔM15, Tn*10*(*tet*^{r})/*araC*$^+$*B*$^+$*A*$^+$*D*$^+$, *leu*$^-$, Δ*lac74*, *galK*$^-$, *str*^{r}.

All promoters not produced by *Bal*31 digestion were constructed by cloning hybridized oligonucleotides at convenient restriction sites using standard techniques (Sambrook *et al.*, 1989). The starting plasmid for all the promoters was pRL41 (Lobell & Schleif, 1991) which contains a modified *araBAD* regulatory region with restriction enzyme cleavage sites flanking each protein binding site. The promoter is transcriptionally fused to the *lacZ* gene. To create pTAP2, the O_2 and O_1 regulatory sites in *araBAD* were deleted by digesting pRL41 with *Eco*RI and *Cla*I and replacing the deleted DNA with a 44 base-pair insert that contained unique *Sal*I, *Xho*I and *Kpn*I sites. pTAP2 was modifed to pTAP4 by creating a unique *Sac*I site at −20 (C to G at −20, A to G at −18) and a unique *Xba*I site at −5 (C to A at −2 and C to G at −1); neither of these restriction sites perturbs the normal regulation or strength of the *araBAD* promoter. To create pBAD-I_1I_2, the CAP site in pTAP4 was deleted by digesting with *Kpn*I and *Nhe*I, made blunt-ended with phage T4 DNA polymerase and the plasmid ligated closed. To make pTAP-up and pTAP-down, the CAP site, I_1 site and the first 4 base-pairs of I_2 were deleted from pTAP4 by digestion with *Kpn*I and *Bam*HI, the plasmids made blunt-ended with T4 DNA polymerase and ligated closed. Hybridized oligonucleotides containing the I_1-I_1 direct repeat sequence were then cloned into the upstream *Eco*RI site. Both orientations of the direct repeat were isolated to give pTAP-up and pTAP-down. All other promoters used in this study were derivatives of pTAP-up.

(c) Bal*31 exonuclease reactions*

All plasmids used in the *Bal*31 reactions were purified by CsCl equilibrium centrifugation as described (Schleif & Wensink, 1981). A total of 7·5 μg of supercoiled plasmid

was digested with various combinations of the appropriate restriction enzymes. Following extraction with phenol and precipitation with ethanol, the different digests were resuspended in a total volume of 47 μl of *Bal*31 buffer (Sambrook *et al.*, 1989) and the solution prewarmed to 30°C. Three units of *Bal*31 exonuclease (slow) were added and 5 μl of the reaction mixture was withdrawn every minute thereafter. The aliquots were immediately added to a single 150 μl quench solution at 65°C (final concentrations were 25 mM-EGTA (pH 8·0), 10 mM-Tris·HCl, pH 8·0). After the final aliquot was added, the quenched reactions were incubated another 10 min at 65°C and then extracted, precipitated and resuspended in T4 polymerase buffer (Sambrook *et al.*, 1989). To increase the variety of the deletion end-points, only 75% of the plasmid was made blunt-ended by sequential treatment with T4 DNA polymerase and then Klenow fragment (Sambrook *et al.*, 1989). The blunt-ended plasmid was recombined with the 25% of the plasmid that was not treated with the polymerases and ligated in a total volume of 350 μl for 24 h at 17°C in T4 ligase buffer (Sambrook *et al.*, 1989) with 17·5 units of T4 ligase. The ligated plasmids were extracted, precipitated and resuspended in 350 μl of TE buffer (10 mM-Tris·HCl (pH 8·0), 1 mM-EDTA). Typically, 1 μl gave 500 to 1500 transformants.

(d) *Primer extension reactions*

RNA was isolated from the same cultures that were assayed for β-galactosidase activity. Cells (3 ml) were spun down and vigorously resuspended in 1 ml of RNAzol and 100 μl of $CHCl_3$ was added. The cells were vortexed and left on ice for 15 min. After centrifuging for 10 min at 4°C, the aqueous layer was transferred to another tube and 1 vol. isopropyl alcohol was added. After at least 1 h at −20°C, the samples were spun and the pellet washed twice with cold 70°C (v/v) ethanol and dried.

RNA (4 μg) was mixed with 2 ng of ^{32}P-labeled primer JC44 (5′-GTGCGCGTGCAGCCCTTATTGCCC-GC-3′, approx. 0·25 pmol, approx. 100,000 cts/min), which binds 63 base-pairs downstream from the normal *araBAD* mRNA start site, and 1 μl of 5× hybridization buffer (5× = 2 M-NaCl, 0·2 M-Pipes·HCl (pH 6·5), 5 mM-EDTA) to give a total volume of 5 μl. The samples were heated to 90°C for 1 min and allowed to cool to room temperature over 20 min: 20 μl of reaction mixture was added to give final concentrations of 1 mM in each dNTP, 50 mM-Tris·HCl (pH 8·3), 10 mM-dithiothreitol, 75 mM-KCl, 3 mM-$MgCl_2$, 100 units of M-MLV reverse transcriptase, actinomycin D, when added, was 50 μg/ml, RNase inhibitor, when added, was 50 units. The reactions were incubated for 30 min at 37°C and then precipitated with ethanol. After drying, the samples were resuspended in 10 μl of a 1:1 (v/v) mix of TE buffer and USB stop solution (95% (v/v) formamide, 25 mM-EDTA, 0·05% (w/v) bromophenol blue and 0·05% (v/v) xylene cyanol FF), and run on a 6% denaturing polyacrylamide sequencing gel. A sequencing reaction using the same labeled primer as in the primer extension reactions generated markers.

(e) *Polymerase chain reaction mapping of deletions*

To determine whether the *Bal*31 reaction had generated a wide range of digestion products, 20 ng of the *Bal*31-treated and ligated plasmid was digested with *Hind*III, which cuts 52 base-pairs downstream from the normal *araBAD* transcriptional start site. The plasmid was added to 50 μl of PCR buffer (Saiki, 1989) containing 1·2 ng of ^{32}P-labeled primer #813 (5′-AATAGGCGTATCACGA-GGCCC-3′, approx. 0·2 pmol, approx. 100,000 cts/min), which binds 41 base-pairs upstream from the I_1-I_1 direct repeat. *Taq* polymerase (2·5 units) was added and the reactions cycled 20 times: 94°C for 1 min, 68°C for 2 min and 74°C for 30 s. After completion of the primer extensions, 150 μl of TE buffer was added, followed by extraction with phenol and precipitation with ethanol. The samples were washed with cold 70% ethanol, resuspended in 7 μl of the 1:1 mix of TE buffer and USB stop solution, and electrophoresed on a 6% denaturing polyacrylamide gel as before.

To determine whether individual *Bal*31-treated plasmids still contained a significant portion of the I_1-I_1 direct repeat, transformed colonies were directly assayed by polymerase chain reaction procedures as described (Sandhu *et al.*, 1989). Transformants containing *Bal*31-treated pTAP-up were amplified with primers JC44 and #834 (5′-CCCATAGCATTTTTATCCATAAGATTAG-CAT-3′), which hybridizes to all of the promoter-distal I_1 half-site and 6 base-pairs of the promoter-proximal I_1 half-site. Transformants containing the *Bal*31-treated pTAP-down were amplified with primers JC44 and #835 (5′-GATCTATGGATAAAAATGCTAATCTTATGGA-3′), which hybridizes to the same region of the I_1-I_1 direct repeat as #834, but to the opposite strand.

(f) In vivo *footprinting*

Cultures were grown as for the β-galactosidase assays, except in 50 ml volumes. Cells were treated with dimethyl sulfate and plasmid DNA was isolated as described (Martin *et al.*, 1986). The methylated guanine bases were cleaved by incubating the plasmid in 100 μl of freshly diluted 1 M-piperidine at 90°C for 30 min, cooled and extracted twice with 1 ml of *n*-butanol, once with 1 ml of 95% ethanol and resuspended in 20 μl of TE buffer. Multiple rounds of primer extension were performed using *Taq* DNA polymerase (Axelrod & Majors, 1989; Sasse-Dwight & Gralla, 1990), 4 μl of the dimethyl sulfate-treated plasmid and 1·2 ng of ^{32}P-labeled primer #813. The reaction conditions and subsequent gel electrophoresis were identical with those used to map the *Bal*31 deletions.

Hypermethylation at the half-sites in the various promoters was quantified using a Molecular Dynamics phosphoimager as follows (see Brenowitz *et al.*, 1986). The AraC-dependent dimethyl sulfate enhancement of the guanine at site k in lane N, $E_{Gk,N}$, was taken to be $(V_{Gk,N}/V_{Std,N})/V_{Gk',N'}/V_{Std',N'})$, where $V_{Gk,N}$ is the integrated volume (total counts) of the guanine at site k in lane N minus the background for lane N, and $V_{Std,N}$ is the integrated volume of the standard blocks of bands in lane N minus the background for lane N. $V_{Gk',N'}$ and $V_{Std',N'}$ are equivalent terms from a reaction lacking AraC protein.

3. Results

(a) *The promoter-proximal half-site,* I_P, *must overlap the −35 region*

*Bal*31 exonuclease was used to generate a large collection of DNA spacings between an AraC binding site and the RNA polymerase binding site of the *araBAD* promoter. This was done by placing an AraC binding site 56 base-pairs upstream from the RNA polymerase binding site, cleaving between

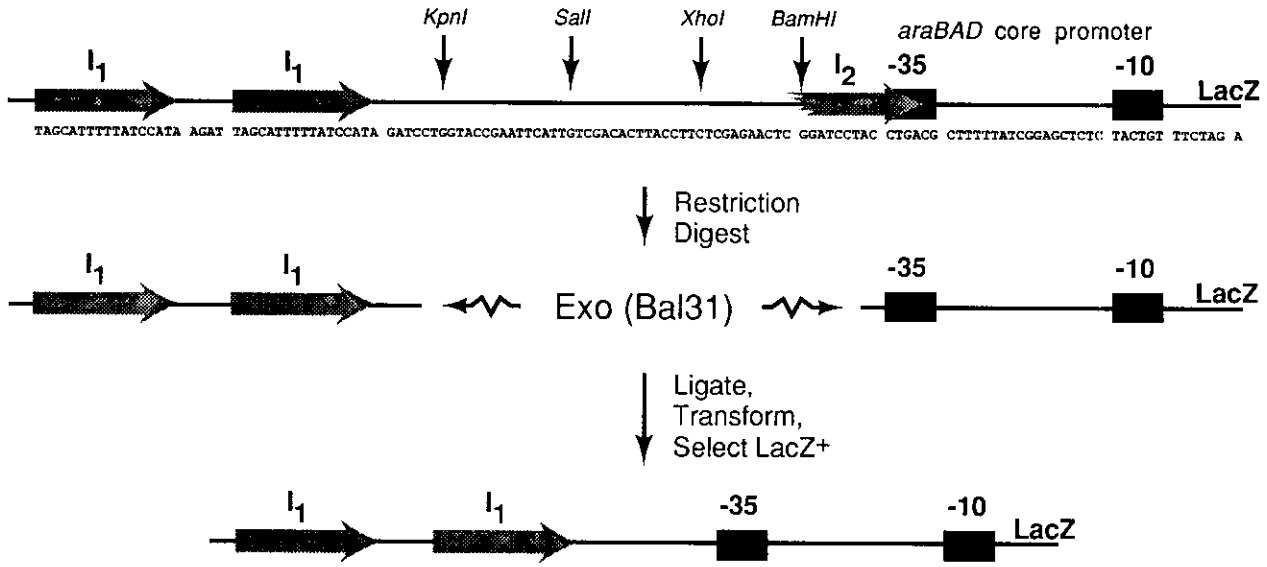

Figure 2. Diagram of the *Bal*31 spacing experiment using pTAP-up as the starting plasmid. The relevant features of pTAP-up are shown schematically on the first line and underneath is the DNA sequence from the 5′ end of the I_1-I_1 direct repeat (−130) to the *araBAD* promoter start site (+1). Breaks in the sequence correspond to the landmarks indicated above. pTAP-up contains the I_1-I_1 direct repeat separated by 56 base-pairs of spacer DNA from the start of the *araBAD* core promoter. The unique restriction sites within the spacer DNA are indicated. The last 13 base-pairs of the I_2 half-site, including the 4 base-pairs that overlap the −35 site, are also present. The core promoter is transcriptionally fused to the *LacZ* gene. Primer #813 binds 41 base-pairs upstream from the 5′ end of the I_1-I_1 direct repeat, and primer JC44 binds 63 base-pairs downstream from the *araBAD* promoter start site. The plasmid pTAP-down is identical with pTAP-up, except that the I_1-I_1 direct repeat is oriented in the opposite direction. Neither pTAP-up or pTAP-down, nor their derivatives, contain the O_2, O_1 or CAP regulatory sites present in *araBAD*.

the two sites, digesting with *Bal*31, and ligating the plasmid to regenerate a closed circle (Fig. 2). DNA thus treated was transformed into competent cells and colonies were selected in which the promoter was active.

The AraC binding site used in these experiments was a direct repeat of the 17 base-pair promoter-distal half-site of *araI*, called I_1. A direct repeat of two I_1 half-sites separated by four base-pairs binds AraC protein at least 100-fold more tightly than the wild-type *araI* direct repeat, denoted I_1-I_2 (Carra & Schleif, 1993). This dramatic difference in affinities apparently results from the fact that I_1 is a close fit to the ideal AraC binding site and closely approximates the consensus of AraC binding half-sites, whereas I_2 deviates significantly from the consensus.

The RNA polymerase binding site, or core promoter, used in these experiments is that of the wild-type P_{BAD}. Not only are the four bases of I_2 that overlap the RNA polymerase −35 region preserved, but an additional nine base-pairs of I_2 are retained (Fig. 2). Hence, among the digestion products should be a few where the wild-type *araI* site and core promoter have been regenerated. We reasoned that if these were found in the products we could conclude that the constructions and selections had worked as desired. The plasmid used in the *Bal*31 spacing experiment, pTAP-up, possessed a final necessary property. AraC protein was unable to activate transcription when bound at the starting location of the I_1-I_1 direct repeat. Therefore, deletion derivatives of pTAP-up that have moved the AraC binding site to positions that activate the promoter could easily be detected.

Primer extension was used to test whether the *Bal*31 reaction had generated a wide variety of deletions. Following *Bal*31 digestion and ligation, an aliquot of pTAP-up was digested at a unique *Hin*dIII restriction site located downstream from the normal *araBAD* transcription start site. The digested plasmid was used as a template for multiple rounds of primer extension with primer #813, which binds upstream from the I_1-I_1 direct repeat. Figure 3 shows that after *Bal*31 treatment and ligation, pTAP-up gave, within resolution of the sequencing gel, bands of all possible lengths between the primer and the *Hin*dIII site. Because the intensity of the bands at some positions was greater than at other positions, the distribution of end-points of the deletions following *Bal*31 digestion was not entirely random. This is consistent with previous observations that the rate of *Bal*31 digestion is sequence dependent (Wei *et al.*, 1983). Nevertheless, Figure 3 demonstrates that the *Bal*31 reactions created a wide variety of deletions, and thus a wide collection of possible spacings was generated between the I_1-I_1 direct repeat and the P_{BAD} core promoter.

The *Bal*31-treated pTAP-up plasmids were transformed into LacZ$^-$ AraC$^+$ strain TR322 and a total of 17,000 colonies were screened on X-gal indicator plates, both in the presence and absence of arabinose. Fourteen hours after plating, the transformed colonies had a wide range of phenotypes, ranging

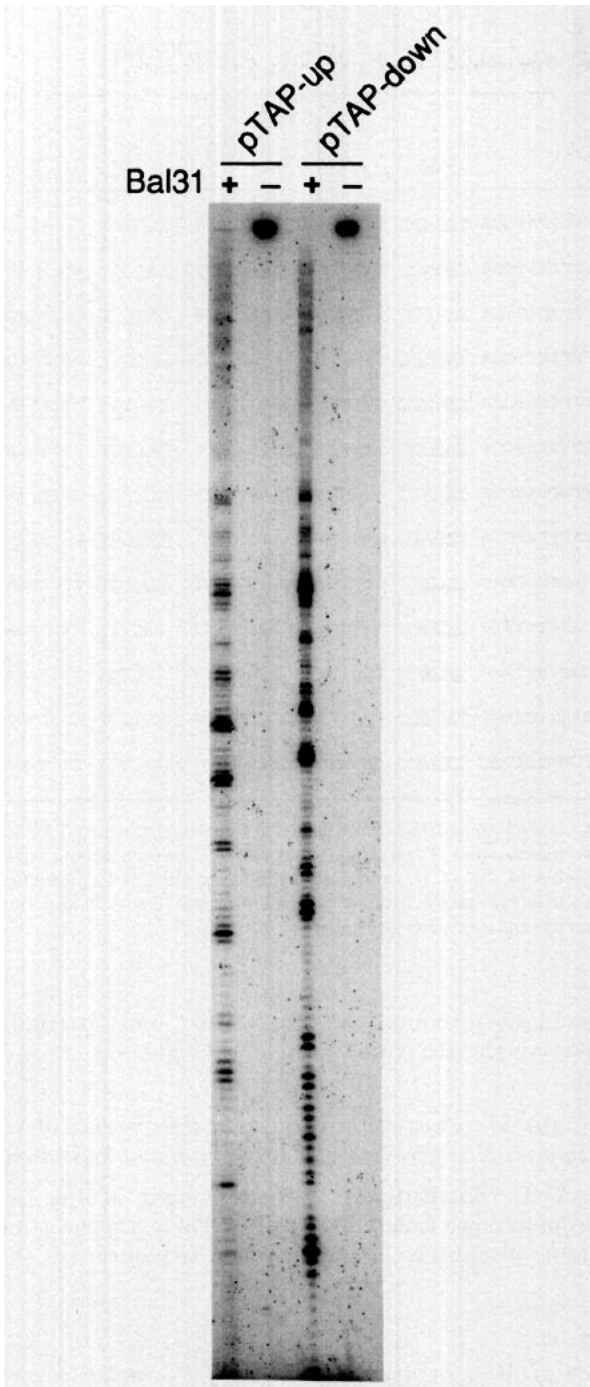

Figure 3. Primer extension analysis of the *Bal*31 reactions. After *Bal*31 digestion and ligation, pTAP-up and pTAP-down were cut at a unique *Hin*dIII site located 52 base-pairs downstream from the normal *araBAD* transcriptional start site and used as a template for multiple rounds of primer extension with primer #813. Control reactions, in which the plasmids were not treated with *Bal*31, show the full length of the DNA between the primer binding site and the *Hin*dIII site.

from white to intense blue. Twenty-eight of the darkest blue colonies were picked, 25 from plates containing arabinose and three from plates lacking arabinose. The plasmid DNA was purified from the selected colonies and retransformed into LacZ$^-$ AraC$^+$ (TR322) and LacZ$^-$ AraC$^-$ strains (TR321). Twenty-three of the 28 plasmids showed AraC-dependent β-galactosidase activity, 21 from the plates with arabinose and two from the plates without arabinose (Table 1). The remaining five promoters were weak AraC-independent promoters and will not be discussed further.

The *ara* regulatory region from the 23 plasmids that showed AraC-dependent expression were sequenced and found to contain a total of ten different deletions (P_1 to P_{10}, Table 1). Since the fusions could have generated entirely new promoters with new transcription start-points, the transcription start sites of these ten AraC-dependent promoters were mapped using primer extension assays, one of which is shown in Figure 4. From the transcription start sites thus determined, the -10 and -35 regions of each promoter were identified by inspection using the consensus -35 and -10 sequences, their separation, and their distance from the transcription start point (Hawley & McClure, 1983; Oliphant & Struhl, 1988). In eight of the ten different promoters, P_1 to P_8, the entire I_1-I_1 direct repeat is present and overlaps the RNA polymerase -35 region by two base-pairs. These eight promoters contain a total of two different -35 sequences and five different -10 sequences, including the P_{BAD} -10 sequence. The remaining two promoters, P_9 and P_{10}, possess the complete promoter-distal I_1 site, but replace 12 and four base-pairs, respectively, of the promoter-proximal I_1 site with the corresponding bases of the I_2 site that had been left next to the *araBAD* core promoter. Thus, these two promoters have the same core promoter as P_{BAD} and the same four base-pair overlap between the *araI* site and the RNA polymerase -35 region. Since only four of the ten promoters (P_5, P_6, P_9 and P_{10}) contain portions of the original P_{BAD} core promoter, the *Bal*31 digestion created six new core promoters that could be activated by AraC. In addition, P_9 differs from the wild-type *araI* site by only the change of a single base-pair, demonstrating that the *Bal*31 digestion and subsequent selection was also capable of essentially recreating P_{BAD}.

The *E. coli* OmpR and phage ϕ29 p4 proteins can significantly activate transcription when their binding sites are moved one helical turn upstream from their normal positions with respect to the -35 region (Serrano *et al.*, 1991; Maeda *et al.*, 1988). In light of the wide range of AraC activation ratios (5 to 120-fold) found among our deletions and the fact that most of these promoters were isolated multiple times (Table 1), it is unlikely that we missed isolating promoters that could be strongly (>5-fold) activated by AraC bound at a different location with respect to the -35 region. Nevertheless, we explicitly tested this possibility. Ten base-pairs of DNA was inserted into P_3 such that the I_1 half-sites were moved ten base-pairs upstream and the -35 sequence held constant. The new promoter, P_3-10-I_1I_1 possessed the same transcription start site as the parent P_3 promoter, but was activated

Table 1

β-Galactosidase activities and DNA sequence of P_1-P_{10}, P_{10}-I_1I_1, P_3-10-I_1I_1 and P_{BAD}-I_1I_2

Promoter (no. isolated)	β-Galactosidase activity (Miller units) −AraC	+AraC, −Ara	+AraC, +Ara	−35	−10
P_1 (2)	200	600	1000	***TAGCATTTTTATCCA*** TAGATC	CTGGTACCGAATTCATTG TCGGAT CCTACCTG
P_2 (2)	300	5300	13,000	***TAGCATTTTTATCCA*** TAGATC	CTGGTACCGAATTCATTG GATCCT ACCTGACG
P_3 (5)	400	20,600	34,500	***TAGCATTTTTATCCA*** TAGATC	CTGGTACCGAATTCATG GATCCT ACCTGACG
P_4 (1)	200	10,400	25,700	***TAGCATTTTTATCCA*** TAGATC	CTGGTACCGAATTCTCG GATCCT ACCTGACG
P_5 (1)	700	19,900	32,000	***TAGCATTTTTATCCA*** TAGATC	TTTTTATCGGAGCTCTC TACTGT TTCTAGA
P_6 (2)	400	16,500	32,700	***TAGCATTTTTATCCA*** TAGCGC	TTTTTATCGGAGCTCTC TACTGT TTCTAGA
P_7 (1)	900	15,100	40,700	***TAGCATTTTTATCCA*** TAGATC	CTGGTACCGAATTCATTG TCTACT GTTTCTA
P_8 (1)	400	7800	14,900	***TAGCATTTTTATCCA*** TAGATC	CTGGTACCGAATTCAT TACGCT TTTTATCG
P_9 (2)	200	200	9100	***TAGCAGATCCTAC*** CTGACG	CTTTTTATCGGAGCTCTC TACTGT TTCTAGA
P_{10} (6)	300	700	19,800	***TAGCATTTTTATC*** CTGACG	CTTTTTATCGGAGCTCTC TACTGT TTCTAGA
P_{10}-I_1I_1	300	3700	7400	***TAGCATTTTTATC*** CATACG	CTTTTTATCGGAGCTCTC TACTGT TTCTAGA
P_3-10-I_1I_1	500	900	1900	***TAGCATTTTTATCCATA***AGTCGAGA TAGATC	CTGGTACCGAATTCATG GATCCT ACCTGACG
P_{BAD}-I_1I_2 (wt)	200	200	6600	***TAGCGGATCCTAC*** CTGACG	CTTTTTATCGGAGCTCTC TACTGT TTCTAGA

β-Galactosidase activities and sequence of all AraC-dependent promoters (P_1-P_{10}) isolated from the *Bal*31 experiment using pTAP-up as the starting plasmid, along with the number of times each promoter was isolated. P_{10}-I_1I_1 and P_3-10-I_1I_1 were constructed using oligonucleotides, and P_{BAD}-I_1I_2 is the wild-type *araBAD* promoter deleted of the O_2, O_1 and CAP regulatory sites. The promoter-proximal half-site is shown in bold italics; the sequence of the distal half-site is I_1 in all the promoters and is omitted. The −35 and −10 sequences are spaced apart from the rest of the sequence, and the transcription start sites are in bold.

only fourfold by AraC (Table 1). In moving AraC one helical turn upstream, 95% of its transcriptional activation ability is lost. This is in contrast to OmpR and ϕ29 p4, which lose less than 50% of their activation when moved back one helical turn (Serrano *et al.*, 1991; Maeda *et al.*, 1988).

The conclusion from both the *Bal*31 and the explicit spacing experiments is that in order for AraC to strongly activate transcription, the promoter-proximal half-site of *araI* must partially overlap the RNA polymerase −35 site.

(b) I_P *occupancy by AraC, not the presence of arabinose, is necessary for transcriptional activation*

AraC constitutively activated eight of the ten promoters we isolated (P_1 to P_8). These are the same eight promoters in which the sequence of the

Table 2

β-Galactosidase activities and in vivo *dimethyl sulfate (DMS) enhancement values of P_{10} and P_3*

Promoter	−AraC, −arabinose β-Gal	DMS I_D	DMS I_P	+AraC, −arabinose β-Gal	DMS I_D	DMS I_P	+AraC, +arabinose β-Gal	DMS I_D	DMS I_P
P_{10}	300	1	1	700	15	1	19,800	55	9
P_3	400	1	1	20,600	21	10	34,500	60	27

Promoter sequence

P10 ***TAGCATTTTATGCATA****AGAT****TAGCATTTTTATGCTGA***CGCTTTTTATCGGAGCTCTCTACTGTTTCTAGA

P3 ***TAGCATTTTATGCATA****AGAT****TAGCATTTTTATGCATA***GATCCTGGTACCGAATTCATGGATCCTACCTGACG

β-Galactosidase activities and the *in vivo* fold-enhancement of dimethyl sulfate methylation at the I_D and I_P half-sites in the arabinose-dependent promoter P_{10} and the arabinose-independent promoter P_3 under the conditions indicated. The relevant sequences of the 2 promoters are shown with I_D and I_P in bold italic. The guanine base that is hypermethylated by dimethyl sulfate upon AraC binding lies on the strand opposite the cytosine base that is outlined. The −35 and −10 sequences are underlined and the mRNA start site is shown in bold.

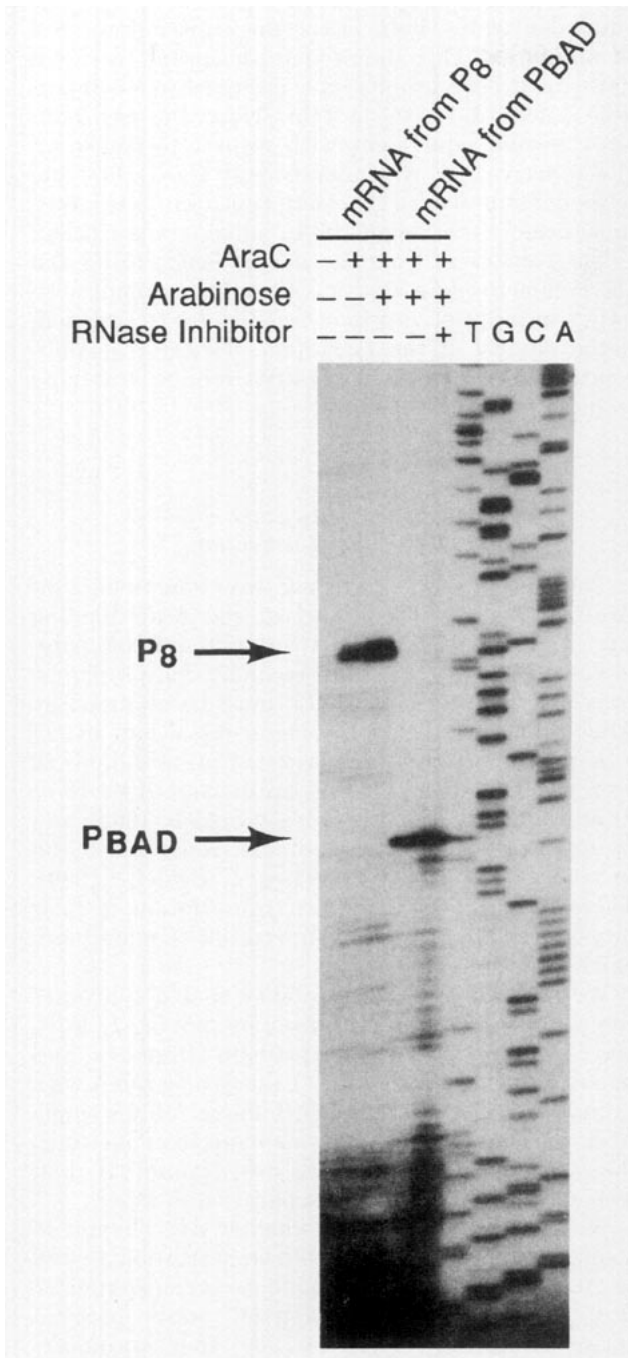

Figure 4. Mapping the transcriptional start site of the AraC-dependent promoter P_8. RNA was isolated from cultures containing the plasmid P_8 in TR321 ($AraC^-$) or TR322 ($AraC^+$) and treated as described in Materials and Methods. As a control, an identical primer extension reaction was performed with RNA from the wild-type *araBAD* promoter on the plasmid pRL41, both in the presence and absence of RNase inhibitor. The position of the 2 extension products are indicated by arrows. A sequencing reaction of pRL41 using primer JC44 serves as the marker.

promoter-proximal half-site, I_P, matched I_1 and in which I_P overlapped the −35 region by two base-pairs. Addition of arabinose to these constitutive promoters resulted in only a further approximately twofold induction. For convenience, we will refer to these as arabinose-independent promoters, despite the fact that they are still partly induced by arabinose. In contrast, the two promoters (P_9 and P_{10}) in which part of I_P matched the sequence of I_2 of the wild-type *araI* site and in which I_P overlapped the −35 region by four base-pairs, behaved like the wild-type P_{BAD}. They were not significantly activated by AraC unless arabinose is present (Table 1).

Either the extent of the −35–*araI* overlap or the exact sequence of I_P could be primarily responsible for arabinose-independence. To determine which, the I_P sequence of the inducible promoter P_{10} was changed to the I_1 sequence. The resulting promoter, P_{10}-I_1I_1, was constitutive in the presence of AraC, even though it has a four base-pair overlap between I_P and the −35 region (Table 1). Thus, arabinose-independence is primarily a consequence of the sequence of I_P, not the amount of overlap between I_P and the −35 region.

All our arabinose-independent promoters possess the I_1 sequence for their promoter-proximal half-site. Because this sequence binds AraC protein significantly more tightly than the normal promoter-proximal half-site, I_2 (Carra & Schleif, 1993), it seems possible that the arabinose-independent promoters have I_P occupied in the absence of arabinose and that occupancy of I_P is necessary and sufficient for transcriptional activation. This hypothesis is consistent with what is known about the *ara* system. Arabinose increases the affinity of AraC protein for *araI* (Hendrickson & Schleif, 1984) and, *in vitro*, arabinose addition causes I_2 to become occupied (Lee *et al.*, 1987; Lobell & Schleif, 1990).

The hypothesis predicts that an arabinose-independent promoter should have I_P occupied by AraC both in the presence and absence of arabinose, while an arabinose-dependent promoter should have I_P occupied only in the presence of arabinose. *In vivo* footprinting shows this prediction to be correct.

When occupied by AraC, the reactivity towards dimethyl sulfate of a conserved bottom strand guanine in the I_1 half-site, the O_2 half-site and both O_1 half-sites is increased dramatically (Martin *et al.*, 1986, see Table 2 for the specific guanine basis). *In vitro*, the degree of hypermethylation at the I_1 half-site is proportional to AraC occupancy, both in the presence and absence of arabinose (Huo, 1988). Thus, *in vivo* footprinting provides a reliable assay for AraC occupancy of its half-sites. In all the sites where AraC occupancy increases the guanine methylation rate, the guanine base is followed by an adenine base. In the wild-type P_{BAD} promoter, the promoter-proximal half-site, I_2, lacks this necessary adenine base and thus arabinose-induced occupancy of I_2 by AraC protein cannot be observed *in vivo*. Fortunately, the relevant guanine base in the promoter-proximal half-site of the arabinose-dependent promoter P_{10} contains the necessary adenine. Figure 5 shows the *in vivo* footprint of the arabinose-independent promoter P_3 and the arabinose-dependent promoter P_{10}, and Table 2

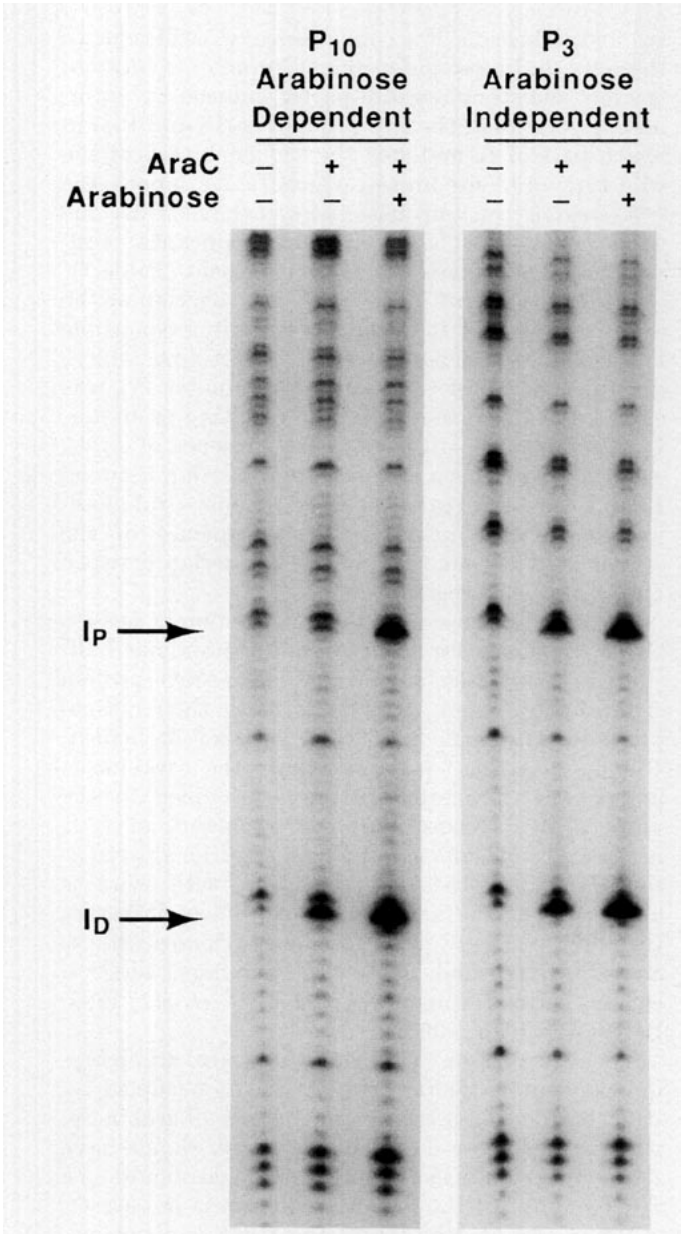

Figure 5. *In vivo* dimethyl sulfate footprint of the arabinose-dependent promoter P_{10} and the arabinose-independent promoter P_3. AraC binding to the I_D and I_P half-sites is reflected by the intensity of the methylation enhancement of the guanine bases indicated by arrows.

compares the methylation enhancements at the I_D and I_P half-sites of the two promoters with their corresponding β-galactosidase expression.

The arabinose-independent promoter, P_3, shows substantial methylation enhancement, and hence AraC occupancy, at the I_P half-site in the presence or absence of arabinose. As predicted, the I_P half-site of the arabinose-dependent promoter, P_{10}, did not become hypermethylated, and thus occupied by AraC protein, until arabinose was added and the promoter became active.

In contrast to the I_P half-sites, the promoter-distal half-sites of P_3 and P_{10} methylate at nearly equivalent rates, both in the presence and absence of arabinose. This shows that the promoter-distal half-site at both promoters is occupied to a comparable extent by AraC protein. Thus, the two half-sites are not equivalent with respect to transcriptional activation. AraC can occupy the I_D half-site without stimulating transcription, whereas AraC must occupy the I_P half-site to induce transcription.

The conclusion from the *in vivo* footprinting and the β-galactosidase assays is that AraC cannot activate transcription unless it occupies I_P. In addition, if the affinity of the I_P half-site for AraC is sufficiently high, then AraC does not require arabinose to induce transcription.

(c) *The location of* I_D *plays a role in transcriptional activation*

The preceding results reveal two requirements for AraC activation. First, part of the AraC binding site, *araI*, must properly overlap the RNA polymerase −35 region, and second, the promoter-proximal half-site of *araI*, I_P, must be occupied by AraC protein. Does the promoter-distal half-site of *araI* play any role in induction? AraC still binds *in vitro* when the two half-sites are separated by 10 or 21 additional base-pairs (Carra & Schleif, 1993), and *in vivo* AraC can bind to half-sites separated by 30 to about 400 base-pairs (Lee & Schleif, 1989). Therefore, can AraC induce transcription if I_P is properly located and I_D is moved ten or more base-pairs upstream?

We investigated the possibility that a split *araI* site could activate transcription by moving I_D in P_3 and P_6 150 base-pairs upstream from I_P and removing variable amounts of this spacer DNA with exonuclease *Bal*31 as described above. Only a single AraC-dependent promoter with more than one-third the activity of P_3 and P_6 was found. It had recreated the normal I_1-I_1 separation of P_6.

We also explicitly tested whether AraC bound to a split *araI* site would activate transcription as well as AraC activates when bound to an unsplit *araI* site. Ten additional base-pairs were inserted between the I_1 half-sites in the arabinose-independent promoter, P_3, to create P_3-I_1-*10*-I_1.

In the presence of arabinose, the P_3-I_1-*10*-I_1 promoter was activated only one-third as well as P_3 (Table 3). This reduction in activity could be the result of poor binding of AraC to the split *araI* site, or it could result from a reduced ability of the protein bound to the split *araI* site to activate transcription. To show it was not merely a lower affinity for the split I_1-*10*-I_1 half-sites, we constructed P_3-I_2-I_1, which has the promoter-distal I_1 half-site in P_3 replaced with the lower affinity I_2 half-site. *In vitro* experiments show that the affinity of AraC for I_1-*10*-I_1 is tenfold greater than its affinity for I_1-I_2 (Carra & Schleif, 1993). Nevertheless, in the presence of arabinose, AraC activated P_3-I_1-*10*-I_1 only one-third as well as P_3-I_2-I_1 (Table 3). To test this point more carefully,

Table 3
β-Galactosidase activities and in vivo *dimethyl sulfate (DMS) enhancement values of* P_3-I_1-10-I_1 *and* P_3-I_2-I_1

	−AraC, −arabinose			+AraC, −arabinose			+AraC, +arabinose		
		DMS			DMS			DMS	
Promoter	β-Gal	I_D	I_P	β-Gal	I_D	I_P	β-Gal	I_D	I_P
P_3-I_1-10-I_1	600	1	1	3400	17	13	12,100	30	23
P_3-I_2-I_1	600	·	1	3100	—	6	32,400	--	21

Promoter sequence

P_3-I_1-10-I_1 TAGCATTTTTATCCATAAGATAGATCTAGATTAGCAATTTTATCCTTAGATCCTGGTACCGAATTCATGGATCCTACCTGACG

P_3-I_2-I_1 TAGCGGATCCTACCTGAAGATTAGCAATTTTATCCATAGATCCTGGTACCGAATTCATGGATCCTACCTGACG

β-Galactosidase activities and the *in vivo* fold-enhancement of methylation at the I_D and I_P half-sites in the split P_3-I_1-10-I_1 promoter *versus* the P_3-I_2-I_1 promoter under the conditions indicated. The relevant sequences of the 2 promoters are shown with the same markings as in Fig. 2. The guanine base in the I_2 half-site in P_3-I_2-I_1 is not hypermethylated upon AraC binding, most likely because it lacks an adenine base following it.

we used dimethyl sulfate footprinting to monitor the *in vivo* occupancy of the I_P half-site at the two promoters. As shown in Table 3, the I_P half-site in P_3-I_1-10-I_1 was occupied by AraC to a greater extent than the I_P half-site in P_3-I_2-I_1. This demonstrates that mere occupancy of the I_P half-site by AraC protein is not sufficient to activate transcription fully. Full induction also requires that the promoter-distal half-site be properly located.

(d) *Only one orientation of AraC can activate transcription*

The AraC protein binding site in the wild-type *araBAD* operon and in the experiments described above is a direct repeat. Consequently, a different face of the AraC protein would be presented to RNA polymerase if the orientation of the direct repeat were reversed. Are there any positions from which AraC protein oriented in this opposite direction can activate transcription from the P_{BAD} promoter? This question was addressed using the *Bal*31 exonuclease approach described before.

The plasmid pTAP-down is identical with pTAP-up, except that the I_1-I_1 direct repeat is oriented in the opposite direction (Fig. 2). pTAP-down was restriction digested, treated with *Bal*31, ligated and transformed into TR322. A total of 24,000 colonies was screened without finding any colonies displaying AraC-dependent LacZ expression.

The failure to find AraC-dependent colonies could mean that the protein can activate transcription only when it is oriented in one direction. It could mean also that something is wrong with the plasmid or screening method. The controls described below suggest the experiment was properly performed. The *Bal*31 used to digest pTAP-down was active and produced many different spacings, as shown in Figure 3. We also tested explicitly that digestion had not been excessive, thereby deleting the I_1-I_1 direct repeat from all of the plasmids. Polymerase chain reaction studies showed that five out of ten randomly chosen colonies contained most of the I_1-I_1 direct repeat sequence. A similar proportion of *Bal*31-treated pTAP-up transformants (6 out of 10) contained the same amount of the I_1-I_1 direct repeat sequence. Finally, when a strong promoter was placed upstream from the *lacZ* gene in pTAP-down, β-galactosidase expression was normal, showing that the *lacZ* gene in pTAP-down is fully functional.

The observation that the *Bal*31 digest of pTAP-down gave no promoter that could be activated by AraC is a negative finding. Nevertheless, it strongly suggests that AraC protein cannot activate transcription when its orientation is reversed from the normal one.

4. Discussion

There are five main results from our experiments. These are: AraC can strongly activate transcription from essentially only one location with respect to RNA polymerase; AraC must occupy the promoter-proximal half-site of *araI* to activate transcription; the arabinose requirement for induction by AraC can be overcome with a strong AraC binding site; a properly positioned promoter-distal half-site of *araI* is necessary for full activation by AraC; and only one face of AraC can induce transcription. We will discuss these results and then present a picture for AraC activation.

(a) I_P *must overlap the RNA polymerase −35 region*

In the experiments described here, we used *Bal*31 exonuclease to place an AraC protein binding site, the I_1-I_1 direct repeat, at a large number of random locations with respect to the P_{BAD} promoter. This approach avoids several weaknesses inherent in

constructions in which a site is explicitly placed at a predetermined location. First, in such experiments, only a few spacings, all defined by the investigator, are examined. Additionally, the different activator spacings are all tested against the same core promoter. If from some positions the activator were capable of assisting a different step than that at which the promoter is kinetically limited, no activation would be observed. This is not a farfetched possibility, for the kinetic step at which CAP assists initiation can change with the binding position of CAP (Gaston *et al.*, 1990). The *Bal*31 exonuclease method does not suffer from these drawbacks. It creates a wide variety of random spacings, and because many different sequences are juxtaposed, a variety of promoter sequences are also created and screened. In our experiments, a total of nine different core promoters, including P_{BAD}, were isolated.

In all ten of the different AraC-dependent constructs we isolated using the *Bal*31 method, the promoter-proximal half-site for AraC protein binding, I_P, overlapped the RNA polymerase −35 site by two or four bases. In the natural AraC-regulated promoters, similar overlaps are found. At P_{BAD} and at P_E, I_P overlaps by four base-pairs, and at P_J the overlap appears to be two base-pairs (our unpublished results). In the fourth AraC-responsive promoter, P_{FGH}, the AraC binding site is not adjacent to the RNA polymerase binding and is oriented in the opposite direction (Hendrickson *et al.*, 1990). Unlike at the other *ara* promoters, AraC is incapable, by itself, of stimulating transcription at P_{FGH}, both *in vivo* and *in vitro* (Kolodrubetz & Schleif, 1981; Hendrickson *et al.*, 1990). It is likely that AraC regulates the activity of P_{FGH} by a fundamentally different mechanism from the mechanism used by the promoters studied in this work.

It seems unlikely that any spacings that would have yielded strong (>5-fold) activation were missed. Control experiments showed that virtually all deletion sizes were generated. The promoters that were selected had a wide range of activation levels (5 to 120-fold), and most of these promoters were isolated multiple times. It also seems unlikely that any spacing so greatly activated transcription that the cells did not survive and hence that spacing was not isolated. Two lines of evidence argue against this possibility. First, even our strongest inducers grew with doubling times within 50% of the weaker inducers. Second, we isolated variants of several of the strong inducers that are much weaker inducers. For instance, the sequence of P_2 is the same as that of P_3, except for one additional base-pair between the −10 and −35 region. This reduces the activity of the promoter from 34,500 units to 13,000 units. If inviable hyperinducers existed, weaker and viable variants of them should have been isolated.

Finally, when we explicitly moved I_P one helical turn of DNA upstream from its normal overlap with the −35 hexamer, transcription was reduced by 95%. From all of this information we conclude that AraC can directly and strongly activate transcription from only a single position with respect to RNA polymerase: the I_P half-site must overlap part of the −35 region.

Unlike AraC, many other activator proteins in *E. coli* are capable of activating transcription from multiple locations with respect to RNA polymerase. The best characterized of these proteins are the general nitrogen regulatory protein, NtrC, the porin synthesis regulator, OmpR and CAP. NtrC can activate *in trans* to RNA polymerase and appears to have a few positional constraints (Wedel *et al.*, 1990), while OmpR and CAP can activate transcription from multiple locations with respect to RNA polymerase, although the different locations must be in helical phase with each other (Gaston *et al.*, 1990; Maeda *et al.*, 1988). Not only do these proteins differ from AraC in their ability to activate transcription from multiple positions, they also differ from AraC in certain aspects of their transcriptional activation mechanisms. NtrC requires ATP hydrolysis for activation and activates an RNA polymerase that contains the sigma 54 subunit (Sasse-Dwight & Gralla, 1989; Wedel *et al.*, 1990). AraC activates an RNA polymerase containing the common sigma 70 subunit (Hahn *et al.*, 1986) and there is no discernible effect of ATP upon the activity of AraC (R. Schleif, unpublished results). OmpR is a member of the two-component regulatory family whose activity is regulated by phosphorylation (Igo *et al.*, 1989). There is no evidence that the activity of AraC is regulated by phosphorylation. Genetic and biochemical data suggest that CAP may, in fact, contain two activation domains (Gaston *et al.*, 1990; Bell *et al.*, 1990), while it appears unlikely that AraC contains more than one activation domain. Thus, the differing ability of NtrC, OmpR and CAP to activate transcription from multiple locations, and the inability of AraC to do so, may simply be a consequence of the different mechanisms each protein uses to activate transcription.

Is AraC unique among activator proteins in being confined to activating transcription from only a single location? A search of the National Center for Biotechnology Information databases at The National Library of Medicine (Nov. 92) reveals 22 prokaryotic relatives of AraC. These homologs possess greater than 17% identity over a span of at least 90 amino acid residues in the C-terminal half of AraC. Extended homology of this order in proteins of similar function almost certainly implies a similar tertiary structure (Sander & Schneider, 1991). The C-terminal half of AraC contains at least one of the DNA-containing areas (Brunelle & Schleif, 1989) and most likely contains the activation domain as well (Menon & Lee, 1988). If the structure and function of the C-terminal half of AraC is conserved among the members of the AraC family, there may be a large class of regulatory proteins that are confined to activating transcription from a single position. Another protein from *E. coli* that may also be restricted to activating transcription from only a

single position is the MalT protein. When MalT is bound to its highest affinity site at the *malK* promoter, it is unable to activate transcription. MalT provides significant activation only after it has been repositioned by CAP protein to a weaker binding site located three base-pairs closer to RNA polymerase (Richet *et al.*, 1991). This is the same position, within a base-pair, as is found at all the promoters activated by MalT (Vidal-Ingigliardi *et al.*, 1991). Finally, excluding the global activators CAP and FNR, most activator binding sites in *E. coli* are adjacent to, or overlap with, the −35 region of the promoter they regulate (Collado-Vides *et al.*, 1991). It is possible that many activators, at least in prokaryotes, may be limited to significantly activating transcription from only one location with respect to RNA polymerase.

(b) *AraC must occupy the* I_P *half-site to induce transcription*

Experiments presented in this paper show that AraC is incapable of activating transcription until it occupies the I_P half-site. This fact, coupled with the finding that I_P must also partially overlap the −35 RNA polymerase binding site, make it likely that the AraC subunit bound at I_P makes direct protein–protein contacts with RNA polymerase that are essential for transcriptional activation. Because moving I_P a helical turn away from the −35 region results in greatly reduced activation, these protein–protein contacts are likely to be inflexible.

(c) *A high-affinity AraC binding site overcomes arabinose dependence*

The existence of our arabinose-independent promoters demonstrates that arabinose is not required to transform AraC into an activating conformation. In our constructs, whenever I_P possessed the sequence of the high-affinity half-site, I_1, the I_P half-site was occupied by AraC as shown by *in vivo* footprinting, and the promoter was strongly activated in the absence of arabinose. Conversely, on the promoters that show little activation in the absence of arabinose, the I_P half-sites contain part of the weak-affinity I_2 sequence and are not occupied by AraC until arabinose is added. Thus, it is the occupancy of the promoter-proximal half-site by AraC that is relevant to transcriptional activation, not whether or not arabinose is present.

What role does arabinose play in transcriptional activation? *In vitro*, the major effect of arabinose on AraC protein is to change the affinity of AraC for different spatial arrangements of half-sites. In the absence of arabinose, AraC favors binding to half-sites separated by more than a helical turn of DNA, while in the presence of arabinose, AraC favors binding to half-sites separated by less than a helical turn of DNA (Carra & Schleif, 1993). At P_{BAD}, the addition of arabinose weakens the looping interaction of AraC when bound to the separated I_1 and O_2 half-sites and strengthens AraC binding to the adjacent I_1 and I_2 half-sites (Lobell & Schleif, 1990). Occupancy of I_2 by AraC then leads to transcriptional activation, as was suggested on the basis of DNase footprinting (Lee *et al.*, 1987). The change in affinity for, and hence occupancy of, different arrangements of half-sites in response to arabinose is sufficient to explain all the regulation phenomena that have been observed in the *araBAD* system.

(d) *The position of* I_D *plays a role in transcription*

The positional requirement for I_P places one subunit of AraC protein right up against RNA polymerase. In addition, because AraC can occupy the promoter-distal half-site, I_D, without activating transcription (Fig. 5; and see Martin *et al.*, 1986), it is plausible then that the subunit of AraC bound at I_D plays no role in transcriptional activation. Surprisingly, we found that moving the I_D half-site one helical turn away from I_P reduced activation in P_3 threefold.

There are three likely explanations for the dependence of activation upon the position of I_D. First, activation by AraC could involve DNA bending from both I_P and I_D; second, both AraC subunits could make direct and important contacts with RNA polymerase; and third, shifting the position of I_D could distort the conformation of the subunit bound at I_P. A number of activating proteins are known to bend DNA (Rojo *et al.*, 1990; Schultz *et al.*, 1991; Wang *et al.*, 1992) and a bend placed at the CAP binding site will substitute for CAP binding itself (Bracco *et al.*, 1989). AraC protein bends DNA about 50° (Lobell, 1990; R. Seabold, unpublished results) and thus moving a bend generated by AraC binding at I_D is a possibility. The AraC subunit bound at I_D could also contact RNA polymerase, and this contact could be disrupted when the I_D half-site is moved upstream. Finally, a distortion in the structure of the AraC subunit bound at the promoter-proximal half-site could be created by moving I_D further away, since the dimeric protein would then have to span a greater distance. This type of explanation has been invoked to explain why separating binding sites for the bacteriophage λ CI protein by helical turns of DNA interferes with transcriptional activation (Hochschild & Ptashne, 1988).

(e) *Only one face of AraC can activate transcription*

Using pTAP-up, in which the direct repeat AraC binding site has the native orientation, as the starting plasmid in the *Bal*31 spacing experiment, we isolated ten unique promoters that were activated by AraC. When the same experiment was done with pTAP-down, in which the AraC binding site has the opposite orientation, we found no promoter that could be activated by AraC. Control experiments indicated that the negative results with pTAP-down were not due to experimental error

such as inactive *Bal*31, excessive *Bal*31 digestion or a non-functional *lacZ* gene. Therefore, we interpret the results to mean that AraC protein can activate from only one orientation. This is not unreasonable, because two different faces of AraC are presented to RNA polymerase, depending on the orientation of the direct repeat half-sites. Several inferences can then be drawn concerning the mechanism whereby AraC activates transcription.

First, AraC cannot be activating transcription simply by being an inert "backstop" that prevents RNA polymerase from sliding across the −35 region. If that were the case, then it would not matter which face of AraC was presented to RNA polymerase. Second, if AraC-induced bends in the DNA helix activate transcription, then both the location and the direction of the bend are crucial for activation. This is because in the backward orientation, the AraC-induced bend in the DNA can be in either the same direction or the same position with respect to RNA polymerase as in the forward orientation, but not both simultaneously. Previous work has demonstrated that while intrinsically bent DNA can activate transcription, the bends must be properly located and phased (Bracco *et al.*, 1989; Gartenberg & Crothers, 1991). Finally, if protein–protein contacts between AraC and RNA polymerase are required for transcriptional activation, then only one face of AraC can make the requisite protein contacts. This strongly suggests that the putative protein–protein contacts are specific in nature and not simply the property of any collection of protruding amino acid residues on the surface on AraC.

Most transcriptional activators studied to date bind as dimers to half-sites with inverted repeat symmetry (Schleif, 1988). As expected, reversing the orientation of these inverted repeats with respect to RNA polymerase does not affect the ability of the activator to induce transcription (Gaston *et al.*, 1988). To our knowledge, our results with AraC are the first to suggest that activator proteins that bind to half-sites with direct repeat symmetry can function in only one orientation of the direct repeats. It will be interesting to see whether a similar constraint is placed on other activator proteins that bind direct repeats.

(f) *Picture of the AraC transcription complex*

Our current picture of the AraC–RNA polymerase transcription complex is shown in Figure 6. Each AraC subunit is depicted as being composed of two loosely connected domains, with one binding the DNA and the other responsible for AraC dimerization, as has been suggested (Carra & Schleif, 1993; Menon & Lee, 1990). The half-sites to which AraC binds are arranged as direct repeats. As is necessary for transcriptional activation, the I_P half-site overlaps the −35 hexamer and is occupied by AraC protein. This places the subunit of AraC bound at I_P close to the RNA polymerase bound at the −35 site and suggests that this AraC subunit and the RNA polymerase make specific protein–protein contacts that are necessary for activation. The findings that only one face of AraC can activate transcription and that the promoter-distal half-site can be moved without destroying activation support this hypothesis. The amino acid residues of AraC that make this putative contact to RNA polymerase cannot be very flexible as they cannot accommodate additional insertions of DNA between themselves and RNA polymerase. In addition, the half-sites are bent, as is observed when AraC is bound, and these bends may also play a role in activation by AraC.

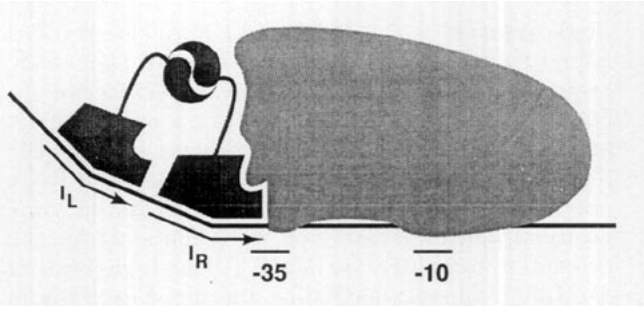

Figure 6. Current picture of AraC–RNA polymerase transcription complex. A dimer of AraC, shown in dark gray, binds to 2 half-sites organized as direct repeats. The RNA polymerase is shown in light gray and binds to the −35 and −10 sites of the core promoter.

We thank NIH grant GM18277 for support and John Carra, Silvia Bustos, Susan Egan, Tom Colonna, Becky Eustance, Casonya Johnson, Duc Nguyen, Beatrice Saviola, Robert Seabold and Jeff Withey for advice and encouragement.

References

Axelrod, J. D. & Majors, J. (1989). An improved method for photofootprinting yeast genes *in vivo* using Taq polymerase. *Nucl. Acids Res.* **17**, 171–183.

Bell, A., Gaston, K., Williams, R., Chapman, K., Kolb, A., Buc, H., Minchin, S., Williams, J. & Busby, S. (1990). Mutations that alter the ability of the *Escherichia coli* cyclic AMP receptor protein to activate transcription. *Nucl. Acids Res.* **18**, 7243–7250.

Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S. & Buc, H. (1989). Synthetic curved DNA sequences can act as transcriptional activators in *Escherichia coli*. *EMBO J.* **8**, 4289–4296.

Brenowitz, M., Senear, D. F., Shea, M. A. & Ackers, G. K. (1986). Quantitative DNase I footprint titration: a new method for studying protein-DNA interactions. *Methods Enzymol.* **130**, 132–181.

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

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

Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocynate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159.

Collado-Vides, J., Magasanik, B. & Gralla, J. D. (1991). Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* **55**, 371–394.

Cubellis, M. V., Marino, G., Mayol, L., Piccialli, G. & Sannia, G. (1985). Use of fast protein liquid chromatography for the purification of synthetic oligonucleotides. *J. Chromatogr.* **329**, 406–414.

Dunn, T., Hahn, S., Ogen, S. & Schleif, R. (1984). An operator at −280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Nat. Acad. Sci., U.S.A.* **81**, 5017–5020.

Englesberg, E., Irr, J. & Lee, N. (1965). Positive control of enzyme synthesis by gene C in the L-arabinose system. *J. Bacteriol.* **90**, 946–957.

Gartenberg, M. R. & Crothers, D. M. (1991). Synthetic DNA bending sequences increase the rate of *in vitro* transcription initiation at the *Escherichia coli lac* promoter. *J. Mol. Biol.* **219**, 217–230.

Gaston, K., Chan, B., Kolb, A., Fox, J. & Busby, S. (1988). Alterations in the binding site of the cyclic AMP receptor protein at the *Escherichia coli* galactose operon regulatory region. *Biochem. J.* **261**, 808–818.

Gaston, K., Bell, A., Kolb, A., Buc, H. & Busby, S. (1990). Stringent spacing requirements for transcriptional activation by CAP. *Cell*, **62**, 733–743.

Greenblatt, J. & Schleif, R. (1971). Arabinose C protein: regulation of the arabinose operon *in vitro*. *Nature New Biol.* **233**, 166–170.

Hahn, S., Dunn, T. & Schleif, R. (1984). Upstream repression and CRP stimulation of the *Escherichia coli* L-arabinose operon. *J. Mol. Biol.* **180**, 61–72.

Hahn, S., Hendrickson, W. & Schleif, R. (1986). Transcription of *Escherichia coli ara in vitro*. *J. Mol. Biol.* **188**, 355–367.

Hawley, D. K. & McClure, W. R. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucl. Acids Res.* **11**, 2237–2255.

Hendrickson, W. & Schleif, R. (1984). Regulation of the *Escherichia coli* L-arabinose operon studied by gel electrophoresis DNA binding assay. *J. Mol. Biol.* **178**, 611–628.

Hendrickson, W. & Schleif, R. (1985). A dimer of AraC protein contacts three adjacent major groove regions of the *araI* DNA site. *Proc. Nat. Acad. Sci., U.S.A.* **82**, 3129–3133.

Hendrickson, W., Stoner, C. & Schleif, R. (1990). Characterization of the *Escherichia coli araFGH* and *araJ* promoters. *J. Mol. Biol.* **215**, 487–510.

Hochschild, A. & Ptashne, M. (1988). Interaction at a distance between λ repressors disrupts gene activation. *Nature (London)*, **336**, 353–357.

Huo, L. (1988). Dissertation, Brandeis University, Waltham, MA.

Igo, M., Ninfa, A. J., Stock, J. B. & Silhavy, T. J. (1989). Phosphorylation and dephosphorylation of a bacterial activator by a transmembrane receptor. *Genes Develop.* **3**, 1725–1734.

Kolodrubetz, D. & Schleif, R. (1981). Regulation of the L-arabinose transport operons in *Escherichia coli*. *J. Mol. Biol.* **151**, 215–227.

Lee, D. & Schleif, R. (1989). *In vivo* DNA loops in *araCBAD*: size limit and helical repeat. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 476–480.

Lee, N., Wilcox, G., Gielow, W., Arnold, J., Cleary, P. & Englesberg, E. (1974). *In vitro* activation of the transcription of *araBAD* operon by *araC* activator. *Proc. Nat. Acad. Sci., U.S.A.* **71**, 634–638.

Lee, N., Francklyn, C. & Hamilton, E. (1987). Arabinose-induced binding of AraC protein to $araI_2$ activates the *araBAD* promoter. *Proc. Nat. Acad. Sci., U.S.A.* **84**, 8814–8818.

Lobell, R. B. (1990). Dissertation, Brandeis University, Waltham, MA.

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

Lobell, R. B. & Schleif, R. (1991). AraC–DNA looping: orientation and distance-dependent loop breaking by the cyclic AMP receptor protein. *J. Mol. Biol.* **218**, 45–54.

Lu, Y., Flaherty, C. & Hendrickson, W. (1992). AraC protein contacts asymmetric sites in the *Escherichia coli araFGH* promoter. *J. Biol. Chem.* **267**, 24848–24857.

Maeda, S., Ozawa, Y., Mizuno, T. & Mizushima, S. (1988). Stereospecific positioning of the *cis*-acting sequence with respect to the canonical promoter is required for activation of the *ompC* gene by a positive regulator, OmpR, in *Escherichia coli*. *J. Mol. Biol.* **202**, 433–441.

Martin, K., Huo, L. & Schleif, R. (1986). The DNA loop model for *ara* repression: AraC protein occupies the proposed loop sites *in vivo* and repression-negative mutations lie in these same sites. *Proc. Nat. Acad. Sci., U.S.A.* **83**, 3654–3658.

Menon, K. P. & Lee, N. (1990). Activation of *ara* operons by a truncated AraC protein does not require inducer. *Proc. Nat. Acad. Sci., U.S.A.* **87**, 3708–3712.

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

Oliphant, A. R. & Struhl, K. (1988). Defining the consensus sequences of *E. coli* promoter elements by random selection. *Nucl. Acids Res.* **16**, 7673–7683.

Richet, E., Vidal-Ingigliardi, D. & Raibaud, O. (1991). A new mechanism for coactivation of transcriptional activation: repositioning of an activator triggered by the binding of a second activator. *Cell*, **66**, 1185–1195.

Rojo, F., Zaballos, A. & Salas, M. (1990). Bend induced by the phage ϕ29 transcriptional activator in the viral late promoter is required for activation. *J. Mol. Biol.* **211**, 713–725.

Saiki, R. K. (1989). The design and optimization of the PCR. In *PCR Technology* (Erlich, H. A., ed.), pp. 7–16, Stockton Press, New York.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edit., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sander, C. & Schneider, R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, **9**, 56–68.

Sandhu, G., Precup, J. W. & Kline, B. C. (1989). The polymerase chain reaction colony miniprep. *Biotechnology*, **7**, 689–690.

Sasse-Dwight, S. & Gralla, J. D. (1990). Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor sigma 54. *Cell*, **62**, 945–954.

Schleif, R. (1987). The L-arabinose operon. In Escherichia coli *and* Salmonella typhimurium*: Cellular and Molecular Biology* (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaecter, M. & Umbarger, H. E., eds), vol. 2, pp. 1473–1481, American Society for Microbiology, Washington, DC.

Schleif, R. (1988). DNA binding by proteins. *Science*, **241**, 1182–1187.

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

Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991). Crystal structure of a CAP–DNA complex: the DNA is bent by 90°. *Science*, **253**, 1001–1007.

Serrano, M., Barthelemy, I. & Salas, M. (1991). Transcriptional activation at a distance by phage ϕ29 protein p4. *J. Mol. Biol.* **219**, 403–414.

Vidal-Ingigliadi, D., Richet, E. & Raibaud, O. (1991). Two MalT binding sites in direct repeat. *J. Mol. Biol.* **218**, 323–334.

Wang, L., Helmann, J. D. & Winans, S. C. (1992). The *A. tumefaciens* transcriptional activator OccR causes a bend at the target promoter, which is partially relaxed by a plant tumor metabolite. *Cell*, **69**, 659–667.

Wedel, A., Weiss, D. S., Popham, D., Drge, P. & Kustu, S. (1990). A bacterial enhancer functions to tether a transcriptional activator near a promoter. *Science*, **248**, 486–490.

Wei, C.-F., Alianell, G. A., Bencen, G. H. & Gray, H. B. (1983). Isolation and comparison of two molecular species of the *Bal*31 nuclease from *Alteromonas espejiana* with distinct kinetic properties. *J. Biol. Chem.* **258**, 13506–13512.

Edited by P. von Hippel