Recognition of Overlapping Nucleotides by AraC and the Sigma Subunit of RNA Polymerase

ANJALI DHIMAN AND ROBERT SCHLEIF*

Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218

Received 3 September 1999/Accepted 14 June 2000

The Escherichia coli promoter pBAD, under the control of the AraC protein, drives the expression of mRNA encoding the AraB, AraA, and AraD gene products of the arabinose operon. The binding site of AraC at pBAD overlaps the RNA polymerase −35 recognition region by 4 bases, leaving 2 bases of the region not contacted by AraC. This overlap raises the question of whether AraC substitutes for the sigma subunit of RNA polymerase in recognition of the −35 region or whether both AraC and sigma make important contacts with the DNA in the −35 region. If sigma does not contact DNA near the −35 region, pBAD activity should be independent of the identity of the bases in the hexamer region that are not contacted by AraC. We have examined this issue in the pBAD promoter and in a second promoter where the AraC binding site overlaps the −35 region by only 2 bases. In both cases promoter activity is sensitive to changes in bases not contacted by AraC, showing that despite the overlap, sigma does read DNA in the −35 region. Since sigma and AraC are thus closely positioned at pBAD, it is possible that AraC and sigma contact one another during transcription initiation. DNA migration retardation assays, however, showed that there exists only a slight degree of DNA binding cooperativity between AraC and sigma, thus suggesting either that the normal interactions between AraC and sigma are weak or that the presence of the entire RNA polymerase is necessary for significant interaction.

The sigma subunit of RNA polymerase (referred to here as sigma) is responsible for the binding of the holoenzyme to promoters during transcription initiation (2, 46). It does this by making sequence-specific contacts with bases in hexameric sequences centered at 10 and 35 bases upstream of the transcription start site on promoters (3, 13, 18, 32, 45, 50). At the −10 hexamer, sigma makes base-specific contacts with the nontranscribed strand (23, 34, 41, 42). In addition to sigma-DNA interactions during initiation, protein-protein contacts also occur between transcriptional activators and subunits of RNA polymerase (1, 11, 14, 21, 22, 36, 43).

At many promoters, the recognition sequences of transcriptional activator proteins partly overlap the 6 bases of the −35 region that are contacted by the sigma subunit of RNA polymerase (4). In these cases, does the activator substitute for sigma in the recognition of the −35 region; do both proteins read the −35 region, necessitating overlapped reading by both proteins; or does sigma read an adjacent sequence?

On one hand, direct protein-protein contacts between sigma and upstream transcriptional activators seem to occur. At the λ pRN promoter, the binding site of λcI overlaps the −35 region for sigma by 2 nucleotides, and genetic experiments suggest an interaction between the λcI protein and the −35 recognition motif of sigma 70 (25, 31). Recently, interactions between sigma and Ada, an AraC homologue from the XylS family of proteins, have been demonstrated genetically at the ada, alkA, and aidB promoters (27, 28). A direct sigma-Ada interaction at the ada and aidB promoters has also been revealed biochemically with DNA migration retardation assays similar to those presented in this paper (27). On the other hand, at the PhoB-dependent PpstS and the CRP-dependent Pgal promoters, where the activator binding site completely overlaps the −35 hexamer, it appears possible that the activator can substitute entirely for recognition by sigma in the −35 region (26).

We studied the ara promoter, pBAD, which is under the control of two activators, CRP (29, 30) and AraC (12, 15) (Fig. 1). The binding of AraC to the I1 and I2 half-sites is stimulated by the presence of arabinose. When these sites are occupied by AraC, and if they overlap the −35 hexamer by 2 or 4 bases, transcription is actively initiated from pBAD (39).

At pBAD, it is likely that the C-terminal domain of the α subunit of polymerase interacts both with CRP and with AraC (49). Two lines of reasoning suggest that AraC may also interact with the sigma subunit of RNA polymerase. First, the R506H mutation in the sigma subunit allows AraC to stimulate pBAD to high levels in the absence of the normally required CRP (19). Second, although AraC can activate transcription from its position partially overlapping the −35 hexamer, it cannot activate (39) as CRP (14, 47) or OmpR (33) can when they are moved upstream by one or more helical turns.

We have examined whether sigma reads that part of the −35 region that lies outside the AraC-contacting region. If it does read this region, then AraC is not substituting for the contacts made by sigma in the region, and either sigma reads the −35 region as before, or it is only slightly displaced by the presence of AraC. We also analyzed sigma binding at the −35 hexamer at a second promoter where the AraC binding site overlaps the hexamer by only 2 bases.

Our results showed that sigma contacts the nonoverlapped bases of the −35 hexamer. Because of the close spatial placement of AraC and sigma on the promoter DNA, we then looked for an interaction between AraC and sigma that would reveal itself as cooperativity in the binding of AraC and sigma to DNA. To avoid the difficulties that would arise from the known interactions between AraC and the alpha subunit of RNA polymerase (49), we used purified sigma in the absence of the other RNA polymerase subunits. Also, to enhance weak DNA binding affinity of sigma in the absence of core polymerase, we used a truncated variant of sigma (Δ133). This truncation rid the protein of the N-terminal acidic domain that

* Corresponding author. Mailing address: Department of Biology, The Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218. Phone: (410) 516-5207. Fax: (410) 516-5213. E-mail: bob@gene.bio.jhu.edu.
in PCR buffer for 29 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min.

**Materials and Methods**

**Strains and plasmids.** The plasmid used for the initial construction of the 1T_pBAD mutants contained an 1T_pGAL-golK UG in pESS1 (20). The promoter region of the p7 plasmid, which carries an 1T_pGAL-golK UG (39), was replaced with the 1I_pGAL promoter region, resulting in an 1T_pBAD-galK fusion. Promoter activity was assayed in TR322 cells (arcA B- A' D- golK Strr) (only relevant markers are shown) (16). The plasmid used for overexpression of the 1T_pGAL promoters was pQE30 (QIAGEN), in which the p60 gene is under the control of the T5 promoter (48). This was a kind gift from Alicia Dombroski. Protein was overexpressed in XL1 Blue cells (recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1) and purified from inclusion bodies by using nickel columns under denaturing conditions, and renatured as described previously (9, 48).

**Construction of mutant 1T_pBAD templates.** Site-directed PCR mutagenesis was performed to modify the promoter-proximal _araI_ site in pBAD2 and to randomize the nonoverlapping bases (X) in the 35 box. The 1T-half site (TAGCCGATCCATCATCA) contained the beginning sequence of the _rpoD_ region to a _lacZ_ (only relevant markers are shown) (16). The promoter region was amplified from pES51 with two oligo

**Assays.** The promoter activity of the pBAD promoters was quantitated in *Escherichia coli* TR322 cells (16) with either β-galactosidase or galactokinase levels. The cells were grown to an optical density at 600 nm of 0.6 in M10 minimal salts, 0.4% glyceral, 10 μg of vitamin B, per ml, 0.4% Casamino Acids, 1 mM MgSO4, and 0.2% arabino (44). 1 ml was withdrawn, and promoter activity was assayed for β-galactosidase, as described by Miller (37), or for galactokinase (10, 35).

**Construction of promoter templates for the DNA migration retardation assay.** End-labeled DNA fragments were generated by PCR using two oligonucleotides such that the 1T-site was centrally located on the 100-bp product. PCR was performed using 100 ng of γ²P-end-labeled oligonucleotide (ATTGCCAGCG CTCCACCA) at 10² cpm/pg, 300 ng of unlabeled oligonucleotide (CTGGTCA CTCCTACCCA), and 10 ng of template plasmid with 0.4 U of Taq polymerase in PCR buffer containing 29 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min. The 1T_pGAL variant promoter fragments used to test for sigma binding were generated by PCR. The 1I_pGAL bubble (Fig. 2d) was constructed by annealing two oligonucleotides. For the 1I_pGAL consensus -10 template, the TATAAT sequence at the -10 box was introduced into pBAD, in vitro mutagenesis reaction, as described below before PCR amplification. For the in vitro mutagenesis reaction, 50 ng of double-stranded-DNA template was mixed with 125 ng each of the two complementary oligonucleotides containing the desired mutation. Ten units of DpnI endonuclease was added for 1 h at 37°C to digest the original methylated DNA template present in the reaction. This is the site-directed mutagenesis technique of the QuickChange protocol of Stratagene.

**End-labeled DNA templates for the in vitro DNA migration retardation assay were prepared by PCR amplification.** For PCR, 100 ng of γ²P-end-labeled oligonucleotide at 10² cpm/pg, 300 ng of unlabeled oligonucleotide(s), and 25 ng of template plasmid containing the required promoter were mixed in 100 μl of PCR buffer. The PCR cycle parameters were 95°C for 1 min, 55°C for 1 min, and 72°C for 2 min for 29 cycles.

**Purification of the sigma subunit.** The R596H mutation was introduced into the Δ333 sigma-encoding DNA template by in vitro site-directed mutagenesis (QuickChange protocol of Stratagene). The hexahistidine-containing Δ333 variants were overexpressed, purified from inclusion bodies by using nickel columns under denaturing conditions, and renatured as described previously (9, 48).

**DNA migration retardation assay.** The DNA migration retardation assay was used to measure dissociation rates of AraC from mutant 1I_pGAL templates previously described (17). AraC was bound to the mutant 1T_pBAD templates in containing 10 mM Tris-Cl (pH 7.4), 1 mM K-EDTA, 75 mM or 150 mM KCl (depending on the salt concentration required), 1 mM diethiothreitol, 0.5% glycerol, and 0.05% NP-40. The higher salt concentration was used when the binding reactions were performed in the presence of arabino because AraC binds more tightly to DNA in the presence of its ligand and does not show any significant dissociation at lower salt concentrations. For the binding

**FIG. 2. DNA template variants used to test sigma binding.** The two AraC binding half-sites are shown with arrows (underlined in the sequences), and the 35 and -10 hexamers for sigma are shown with solid or broken lines (boldface in the sequences). (a) 1I_pGAL with the wild-type _araI_ half-sites for AraC and the 4-nucleotide overlap at the -35 hexamer. The sequence of the -10 hexamer was not changed. (b) 1I_pGAL -consensus -10. The -10 hexamer was changed to 5' GCCCATAGCATTTTTATCCATAAGATTAGCGGATCCTAC 3'. (c) 1I_pGAL consensus -10. The 1I_pGAL was changed to the consensus sequence TATAAT to generate a stronger σ70 binding site, 5' GCCCATAGCATTTTTATCCATAAGATTAGCGGATCCTAC 3'. (d) 1I_pGAL containing the 1T site in place of the 1I and half site in pBAD. The sequence of the nonoverlapping nucleotides in the 35 hexamer was the same as the parental 5' σ70. The 1T site is not as tight a binding site for AraC as 1I, but it is tighter than the 1I site. 5' GCCCATAGCATTTTTATCCATAAGATTAGCGGATCCTAC 3'. (e) 1I_pGAL bubble is the same as the wild-type 1I_pGAL with a heteroduplex mismatched stretch of DNA (lowercase letters) at the -10 region. 5' GCCCATAGCATTTTTATCCATAAGATTAGCGGATCCTAC 3'. (f) 1I_pGAL bubble is the same as the wild-type 1I_pGAL with a heteroduplex mismatched stretch of DNA (lowercase letters) at the -10 region. 5' GCCCATAGCATTTTTATCCATAAGATTAGCGGATCCTAC 3'.
reaction, purified AraC was added so that just 100% of 1 ng (~10^6 cpm) of end-labeled DNA was bound. Binding of AraC to DNA was allowed to proceed for 10 min, after which an excess of a competitor containing four tandem end-labeled DNA was bound. Binding of AraC to DNA was allowed to proceed free DNA and the various protein-bound species.

I

For normal activation of RNA polymerase, whereas in the 2

G are equivalent at position 2

in the system, no auxiliary activators are needed by RNA polymerase. Changes occur. In the results from the very different contexts in which the sequence 2

ant

that sigma does read the sequence of the two bases. In the P22 system, the dramatic change to 10% in the difference between the modest change to 90% activity in our 2

pBAD

template, we designed a promoter variant, I1-1*pBAD (Fig. 3), in which the AraC binding site has been moved upstream by 2 bases. The *I site contains the beginning sequence of the L2 site and the later sequence of the I1 site. This promoter is still AraC dependent and is 2.3 times as active as the wild-type pBAD promoter. Making such a change in the promoter permits four bases in the −35 region to be altered without affecting AraC binding. We chose to alter these four nucleotides in two ways—by directed and by random mutagenesis. If, despite its requirement for AraC and CRP for full stimulation, and despite the results obtained from the 2-base overlap, pBAD possesses the same promoter sequence dependence as “bare” promoters like P22 ant, then a change to tagGAC should have a particularly dramatic stimulatory effect because of increased homology to the consensus −35 hexamer. Table 1 shows that the activity of tagGAC was not significantly different from that of the parental sequence, tagGAC. The activities of most of the entries shown in Table 1 that resulted from random mutagenesis, however, were strongly dependent upon the sequence of the −35 region outside the *I half-site, thereby indicating that sigma does contact the four bases.

Two potential factors could invalidate the conclusion that I1-1*pBAD activity is dependent upon the identity of the −35 region nucleotides outside the *I half-site. First, introduction of the altered nucleotides might have inadvertently altered nucleotides elsewhere in the plasmid as well, for example, within the β-galactosidase gene. To verify that the decreased levels of β-galactosidase activity we observed with some of the variants were indeed due to changes in the −35 region of the promoter and not due to extraneous mutations elsewhere on the plasmids, we changed the four randomized bases in three mutant templates back to the parental sequence by oligonucleotide-directed mutagenesis. The activities of most of the entries shown in Table 1 that resulted from random mutagenesis, however, were strongly dependent upon the sequence of the −35 region outside the *I half-site, thereby indicating that sigma does contact the four bases.

Altering the two bases of the natural ara pBAD promoter −35 region by 4 bp, it is conceivable that the sigma subunit does not contact the −35 region at all and that AraC assumes the role normally taken by part of sigma. One way to determine whether sigma makes DNA contacts in the −35 region is to vary the sequence of that part of the −35 region that is not contacted by AraC. If promoter activity is insensitive to such sequence changes, we could reasonably infer that sigma does not contact DNA in the region.

Altering the two bases of the natural ara pBAD promoter −35 region (Fig. 3) that are not part of L2 from CG→TT and CG→CC decreased promoter activity to 90 and 50%, respectively, of that of the parental sequence. These results suggest that sigma does read the sequence of the two bases. In the P22 ant promoter however, Moyle and coworkers found that C and G are equivalent at position −30 and that a C-to-T change at position −31 reduces activity to 10% (38). Most likely the difference between the modest change to 90% activity in our system and the dramatic change to 10% in the ant promoter results from the very different contexts in which the sequence changes occur. In the ara system, AraC and CRP are required for normal activation of RNA polymerase, whereas in the ant system, no auxiliary activators are needed by RNA polymerase.

To increase the number of bases in the −35 region that are not contained within the promoter-proximal AraC binding site,
to altered AraC binding at the −35 region. The results suggest that altered sigma binding is the cause of the reduction.

Does sigma directly interact with AraC? On one hand, the partial interdigitation of the AraC and RNA polymerase sigma subunit binding sites on DNA suggests that the two proteins could be located very close to each other and hence might have critical interactions with one another. On the other hand, the fact that the binding site of AraC can be moved 2 bases up-stream without strongly affecting promoter activity, as in I1-I2pBAD (39) and I1-I2pBAD, suggests that perhaps AraC and sigma do not make specific contacts with each other. To test if AraC and sigma do interact with one another, we looked for cooperativity in their binding to DNA.

Purified sigma factor does not detectably bind to promoters by itself, but truncation of its acidic N-terminal domain reveals a weak promoter binding specificity (6, 7, 48). Therefore in looking for cooperativity between AraC and sigma factor in binding at AraC-activated promoters, we used a sigma variant with its N-terminal 133 amino acids deleted. The binding of sigma was examined on I1-I2pBAD and parental I1-I2pBAD templates, but no binding was observed on either template in the presence or absence of bound AraC protein (see Materials and Methods for a description of the DNA templates). Changing the conserved arginine at position 596 to a histidine in the sigma subunit enables RNA polymerase to be active on pBAD in the absence of CRP (19). Possibly the increased activation results from a sigma-AraC interaction, either an interaction where none existed before or a stronger interaction. Therefore, we introduced the R596H mutation into the 133 sigma variant. We were still unable to observe sigma binding to either the I1-I2pBAD or parental I1-I2pBAD DNA in the presence or absence of AraC. To create a stronger sigma binding site on pBAD, we changed the −10 hexamer to the consensus −10 sequence (see Materials and Methods), but still no binding was observed for either the Δ133 or the R596HΔ133 sigma variant on I1-I2pBAD consensus −10.

In a further effort to increase sigma binding, we used a template that mimics the DNA present in the open complex (RPc) during transcription initiation (6, 8). Such a bubble sequence provides a significant advantage for sigma binding, as shown by the preference of RNA polymerase holoenzyme for binding to premelted sequences (5). We used a heteroduplex mismatch bubble-containing template, I1-I2pBAD bubble, that contained the AraC binding sites, I1 and I2, with a mismatch region spanning the −10 region (see Materials and Methods).
With such DNA, we observed some AraC-dependent DNA binding by the $\Delta 133$ and R596H$\Delta 133$ sigma variants (Fig. 5). Using another bubble template with binding sites for AraC and the consensus $-10$ region on the nontemplate strand of the bubble (34) did not enhance binding by sigma in the presence or absence of AraC. We note that the AraC-dependent binding by the truncated sigma protein in all these experiments was not completely reliable, and occasional experiments failed to demonstrate any cooperativity in the binding of AraC and sigma to DNA.

**DISCUSSION**

Our experiments yield the following conclusions: AraC and the sigma subunit of RNA polymerase both make contacts with DNA in the $-35$ region of the $p_{BAD}$ promoter, and interactions between AraC and a truncated form of sigma can be observed in their binding in vitro to DNA, but these interactions are not strong.

The $-35$ hexamer of $p_{BAD}$ shares homology of four bases with the consensus $-35$ sequence, making it a potentially tight binding site for the sigma subunit of RNA polymerase (Fig. 3). On the other hand, because the polymerase-proximal half-site of AraC overlaps the $-35$ region by 4 bp, it is possible that AraC substitutes for the role taken by the domain of sigma that normally contacts the $-35$ region. In the experiments reported here, we found that $p_{BAD}$ activity is strongly dependent on the identities of the bases of the $-35$ hexamer that are not contacted by AraC. We presume, then, that these bases are contacted by sigma. Consequently, we further presume that either the remaining bases of the $-35$ region are also contacted by the sigma subunit or sigma is displaced and reads the bases immediately adjacent to the AraC binding site.

It is possible that AraC and the sigma subunit sequentially contact the common four bases in their partially overlapping binding sites at $p_{BAD}$. We note, however, that simultaneous contact of these bases by alpha helices is also geometrically possible (Fig. 6). Our detection of cooperativity, albeit weak, in the binding of sigma and AraC to the DNA indicates a direct interaction between the two, suggesting simultaneous DNA binding. We must add that our in vitro binding studies utilized AraC and the sigma subunit alone but that the normal binding involves AraC and the RNA polymerase holoenzyme. The additional subunits of RNA polymerase could also interact with AraC or alter the structure of sigma and alter the interaction.

We first used a truncated variant of sigma ($\Delta 133$) that lacked the N-terminal acidic domain to enhance the weak DNA binding affinity of sigma in the absence of core polymerase. Because we observed no binding by the truncated sigma factor and no binding cooperativity between AraC and sigma, we then tried DNA templates that should have higher affinity than double-stranded DNA. Ultimately, we did observe binding cooperativity between AraC and truncated sigma, but this required the use of a DNA template possessing a single-stranded region in the $-10$ region. We are aware of only one other experiment to examine by direct biochemical means an interaction between an upstream activator and sigma factor (27). This work used the Ada protein, double-stranded DNA, and intact sigma factor. As significant binding cooperativity was observed with the Ada and sigma proteins at the $ada$ and $aidB$ promoters, it is possible that the Ada protein interacts significantly more strongly with sigma than does AraC. Alternatively, it is possible that in removing the N-terminal portion of sigma to enhance its DNA binding abilities, we also removed important regions for the AraC-sigma protein interaction.

Promoter recognition at $p_{BAD}$ by sigma is intriguing because,

FIG. 5. DNA migration retardation assay showing weak cooperativity in the binding of AraC and $\sigma^{70}$ to $l_{1}$-$l_{2}$ bubble DNA. The ingredients present (+) are indicated, and except in lanes 7 and 8, where the concentrations of R596H$\Delta 133$ $\sigma^{70}$ were 190 nM and 130 nM, respectively, the concentrations used were 420 pM AraC, 60 nM $\Delta 133$ $\sigma^{70}$, and 380 nM R596H$\Delta 133$ $\sigma^{70}$.

FIG. 6. Model showing simultaneous recognition of four overlapping bases by two $\alpha$-helices fitted into the major groove visualized from the top (left), front (middle), and side looking down one DNA major groove directly at one $\alpha$-helix (right).
while any deviation from the parental sequence causes a reduction in promoter activity, we could see no correlation between specific sequence changes and promoter activity. Similarly, at the pmelR promoter, positions 3 to 6 of the ~35 region lie outside the CRP binding site and play an important role in activation by sigma. Some ~35 hexamer sequences at pmelR are more tolerant of substitutions than others, and mutations that change nonconsensus bases to consensus do not necessarily increase promoter activity (40). Similar observations have been noted with the melAB promoter (24) and the P22 ant promoter (38).

ACKNOWLEDGMENTS

We thank Alicia Dombroski for helpful advice and for providing the Δ133 sigma template and the purification protocol and members of our laboratory for ongoing discussions. This work was supported by NIH grant GM18277 to R.S.

REFERENCES


