Location of close contacts between *Escherichia coli* RNA polymerase and guanine residues at promoters either with or without consensus −35 region sequences

Stephen MINCHIN and Stephen BUSBY
School of Biochemistry, University of Birmingham, P.O. Box 363, Birmingham B15 2TT, U.K.

Methylation-interference assays have been used to identify guanine residues that make important contacts with RNA polymerase during open-complex formation at two related *Escherichia coli* promoters. Methylation of lower-strand G-31 at a *gal* consensus promoter completely prevents complex formation, while modification of upper-strand G-33 has no detectable effect. At *galP1*, which lacks a consensus −35 region, modification of lower-strand G-33 and upper-strand G-14 reduces, but does not prevent, complex formation. G-33 is the only guanine residue in the −35 region of *galP1* where modification interferes with open-complex formation. Since this guanine residue is not protected in open complexes, we conclude that its modification causes alteration of, or interference with, a transient contact during the transcription initiation pathway.

INTRODUCTION

The major *Escherichia coli* RNA polymerase interacts with most promoters by recognizing two hexameric sequence elements located around 35 bp and 10 bp upstream of the transcription start (reviewed by McClure, 1985). Recognition is primarily due to interactions between these sequences and the sigma subunit (sigma-70): the '−35 region' is recognized by a helix–turn–helix structure near the C-terminus of sigma, while the '−10 region' is contacted by the '2.4' region, conserved between different sigma factors, located around residue 440 (see Waldburger et al., 1990, and references therein). Consensus −35 and −10 hexamer sequences have been deduced and, at most promoters, mutation of the actual −35 or −10 sequences away from the consensus leads to a reduction in the activity of that promoter (Hawley and McClure, 1983). However, a small number of cases have now been reported where sequence changes in the −35 hexamer have little or no effect, and where specific −35-region contacts appear unimportant for promoter activity (Ponnambalam et al., 1986; Keilty and Rosenberg, 1987; Ponnambalam et al., 1988; Peakman et al., 1990). Interestingly, in all these cases, the sequence motif 5′-TGN-3′ is found immediately upstream of the −10 hexamer, and it has been suggested that this motif creates an 'extended −10 region' that compensates for the poor −35 sequence (Keilty and Rosenberg, 1987; Kumar et al., 1992).

In previous work we demonstrated that transcription initiation at the *E. coli galP1* promoter is not dependent on specific −35-region sequences (Chan and Busby, 1989). Chan et al. (1990) then made a detailed comparison of the organization of open complexes at *galP1*, and a derivative promoter, $galP_{con}$, in which the *galP1* −35 region was replaced with the −35 hexamer consensus sequence, 5′-TTGACA-3′. In open complexes at $galP_{con}$, RNA polymerase makes close contact with bases in the −35 region and covers the DNA upstream to −45. In contrast, in open complexes at *galP1*, the −35 region is not protected and polymerase covers upstream bases as far as −55. Chan et al. (1990) suggested that the overall architecture of open complexes differs according to whether or not a −35 sequence resembling the consensus hexamer is present: in the absence of a correct −35-region contact, RNA polymerase appears to 'grope' further upstream to make compensatory contacts which involve significant bending and distortion of the DNA template. Interestingly, Hayward and collaborators recently demonstrated that the C-terminal segment of sigma, carrying the helix–turn–helix that normally recognizes the −35 hexamer, is dispensable for transcription from at least some promoters which lack −35-region consensus sequences (Kumar et al., 1992). Chan et al. (1990) used methylation-protection assays to locate guanine residues that make contacts with RNA polymerase in open complexes in *galP1* and $galP_{con}$. This is especially interesting since direct contacts between amino-acid side chains and guanine bases are crucial to many protein–DNA interactions (Steitz, 1990). In particular, arginine residues 584 and 588 in the C-terminus helix–turn–helix of sigma appear to interact directly with the two guanine residues in the −35 consensus hexamer, 5′TTGACA-3′ (Gardella et al., 1989; Siegele et al., 1989). In complexes at $galP_{con}$, RNA polymerase clearly protects methylation of lower-strand G-31 but, at *galP1*, there was no protection of any guanine residue in the −35 region, presumably because there are no specific contacts with this zone in the open complex. However, the guanine residue of the TGN motif just upstream of the −10 hexamer was protected, suggesting that RNA polymerase may contact this base directly. In this paper we described complementary methylation-interference assays at *galP1* and $galP_{con}$. These assays allow us to locate not only the guanine residues that make contact with polymerase in open complexes, but also any residues that make transient interactions during the formation of these complexes. We then measured the effects of changes at important bases on *galP1* expression *in vivo*.

EXPERIMENTAL

Figure 1(a) shows the upper-strand sequence of the two promoters, *galP1* and $galP_{con}$, used in this study. These two promoters, which were described in our previous reports (Chan et al., 1990; Grimes et al., 1991) were cloned on *Eco*RI–*Hin*dIII fragments into pBR322. The Δ*lac E. coli* strain, M182, and a Δ*crp* derivative were used as hosts throughout this work (Casadaban and Cohen, 1980; Busby and Dreyfus, 1983). The derivative of

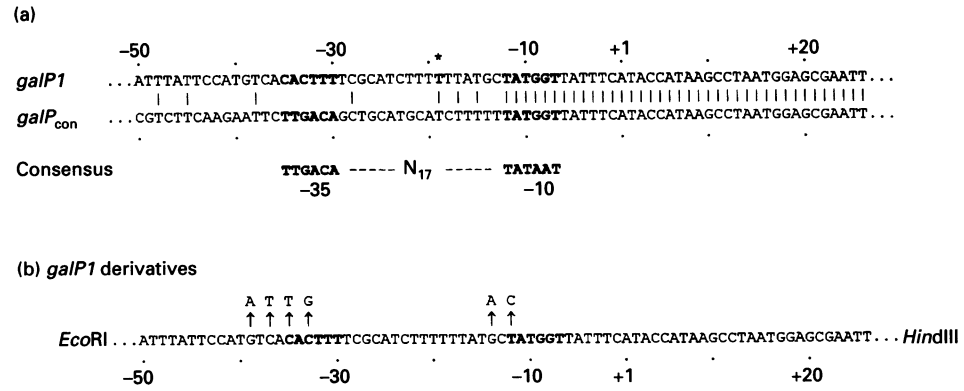

Figure 1 *galP1* and *galp*$_{con}$ promoter sequences

(**a**) The figure shows the upper-strand sequence of *galP1* and *galP*$_{con}$ from −50 to +26 with respect to the transcription start at +1. Each promoter was cloned with an *Eco*RI linker upstream of the promoter and a *Hin*dIII linker downstream. Identical bases are shown by vertical lines. The positions where the −10 and −35 hexamers would be found at consensus-type promoters are shown in bold type, together with the consensus for the two hexamers. For *galP1* the asterisk at position −19 denotes the GC to TA change present in all our constructions that inactivates the alternative *galP2* promoter (Bingham et al., 1986). (**b**) The figure shows the location of single bp changes in *galP1*.

Table 1 Effects of point mutations on *galP1* activity *in vivo*

β-Galactosidase was measured in M182Δ*crp* cells, containing the *lac* expression vector pAA182, into which different *galP1* derivatives had been cloned. Cells were grown in minimal medium plus 80 μg/ml ampicillin which contained fructose as a carbon source and assays were performed exactly as before (Busby and Dreyfus, 1983; Chan et al., 1990). The starting point was the *galP1* sequence shown in Figure 1(a): cells containing plasmid with this promoter contain 1000 standard (Miller, 1972) units of β-galactosidase and this can be taken as a measure of *galP1* activity *in vivo*. The Table lists the different mutations that were studied, together with the corresponding effects on promoter activity. Fragments carrying the different promoters were also cloned into pRW50, a low-copy-number broad-host-range vector encoding resistance to tetracycline (Lodge et al., 1992). The assays were repeated in the same medium containing 35 μg/ml tetracycline instead of ampicillin. The results (not shown) demonstrated that the hierarchy of promoter activities was identical to that found with pAA182 as a vector.

galP1 derivative	Promoter activity (Miller units of β-galactosidase)
galP1 with no mutation	1000
G:C to A:T at −14	150
C:G to G:C at −33	300
C:G to T:A at −35	1000
C:G to T:A at −37	1000
G:C to A:T at −39	1100
T:A to C:G at −12	20

galP$_{con}$ carrying base-pair changes T:A to A:T at −34 and T:A to G:C at −14 is promoter 'e' in Figure 1 of Chan et al. (1990). The mutations in *galP1* at −14, −35, −37, −39 and −12, listed in Table 1 and shown in Figure 1(b), had been isolated previously after random hydroxylamine mutagensis of the *gal* operon regulatory region (Busby and Dreyfus, 1983; Bingham et al., 1986). The mutation at −33 was made by site-directed mutagenesis using an Amersham kit (cat. no. RPN 1523).

For methylation-interference experiments, caesium chloride preparations of plasmids were made and *Hae*III–*Hin*dIII promoter DNA fragments were purified by electroelution from polyacrylamide gels (*Hae*III gives flush ends and cuts just upstream of the *Eco*RI site in pBR322). Purified fragments were end-labelled at the *Hin*dIII site either on the upper strand using [α-^{32}P]dATP and Klenow enzyme, or on the lower strand using [γ-^{32}P]ATP and polynucleotide kinase. In the experiment shown in Figure 5 the upper strand of *galP1* was labelled at the upstream *Eco*RI site with [γ-^{32}P]ATP. For methylation, the standard Maxam and Gilbert (1980) protocol was employed: 50 ng of labelled fragment in 10 μl was mixed with 200 μl of DMS buffer, 1 μl of dimethyl sulphate was added and the sample was incubated at 25 °C. After 90 s the reaction was stopped and the DNA was purified by alcohol precipitation.

To make open complexes, 5 nM-labelled and methylated DNA fragments were incubated for 30 min at 37 °C with 200 nm RNA-polymerase holoenzyme in 20 μl of standard transcription buffer [20 mM Tris/HCl, pH 8, 5% (w/v) glycerol, 100 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 1 mM dithiothreitol and 50 μg/ml BSA]. We used a high RNA polymerase concentration and long incubations to obtain maximum open-complex formation, and we checked that the results were not altered by doubling or halving the incubation times. After incubation, 160 μg/ml heparin was added and free and bound DNA was separated by immediately loading the sample on a 4% (w/v) polyacrylamide gel, run in standard gel-retardation assay conditions, as in our previous work (Gaston et al., 1989). Bands corresponding to free and bound DNA, as in Figure 2, were located by autoradiography, excised and the DNA was purified by electroelution. To identify the positions of guanine methylation, the DNA from 'free' and 'bound' bands was cleaved with piperidine, again using the Maxam and Gilbert (1980) protocol. Samples were analysed on 6% (w/v) sequencing gels calibrated using G-specific reactions.

*Eco*RI–*Hin*dIII fragments carrying different mutations in *galP1* were cloned into the *lac* expression vector, pAA182. Recombinant plasmids were transformed into M182Δ*crp* cells and promoter activities were assayed *in vivo* by the measurement of β-galactosidase activity, exactly as previously described (Chan et al., 1990). The same set of *galP1* fragments was also cloned into the low-copy-number broad-host-range *lac* expression vector pRW50 (Lodge et al., 1992) and the assays were repeated to check that copy number did not affect the hierarchy of promoter activities.

RESULTS AND DISCUSSION

Experimental strategy

The nucleotide sequences of *galP1* and $galP_{con}$ are shown in Figure 1: both promoters were cloned on *Eco*RI–*Hin*dIII fragments with the *Eco*RI site located upstream and the *Hin*dIII site downstream of the transcription start site. The *galP1* and $galP_{con}$ fragments differ in their sequence upstream of the −10 hexamer: $galP_{con}$ carries a consensus −35 hexamer while *galP1* lacks consensus −35 sequences. RNA polymerase binds at both promoters to form 1:1 open complexes that initiate transcripts at +1 (Chan et al., 1990; Grimes et al., 1991). Open complexes can be separated by electrophoresis on polyacrylamide gels: binding of RNA polymerase to promoter fragments results in their retardation during electrophoresis (Figure 2).

Dimethyl sulphate methylates the 7-N of guanine, which is located in the major groove of the DNA helix (Siebenlist et al., 1980, and references therein). ^{32}P-labelled fragments carrying either *galP1* or $galP_{con}$ were methylated and then pre-incubated with RNA polymerase. Fragments bound in open complexes with RNA polymerase could then be separated from free DNA by electrophoresis (Figure 2). After purification of labelled fragments from the 'bound' and 'free' bands, the DNA was cleaved at the sites of guanine methylation by treatment with piperidine, and sequence gels were run to identify the sites of methylation. Methylation of guanine residues at most positions does not affect open-complex formation and, thus, the corresponding bands in the sequence ladder appear in the 'bound' DNA samples. Modification of guanine residues at positions where methylation stops open-complex formation results in the total disappearance of the corresponding band from the 'bound' sample and an enhancement of the band in the 'free' sample. Modifications that decrease the stability of open complexes (or slow their formation) result in an enhancement of bands in the 'free' DNA and a reduction in the 'bound' bands.

Methylation interference at a consensus promoter

Purified fragments carrying $galP_{con}$ were end-labelled at the downstream *Hin*dIII site on either the upper or lower strand and methylated. The results of interference assays (Figure 3) clearly show that methylation of the lower-strand G at −31, corresponding to position 5 of the −35 consensus hexamer TTGACA, totally blocks open-complex formation. In contrast, modification of upper-strand G-33 (at position 3 of the −35 hexamer) has little or no effect on complex formation. Interestingly, both lower-strand G-31 and upper-strand G-33 are thought to interact with arginine side chains in the sigma helix–turn–helix (Gardella et al., 1989; Siegele et al., 1989). Our results are consistent with the suggestion that Arg-584 interacts directly with the 7-N of lower-strand G-31; however, upper-strand G-33 must make a different type of interaction with Arg-588. Note that mutations at both positions 3 and 5 of −35 hexamers lead to dramatic reductions in promoter strength and G:C is better conserved at position 3 than at position 5 (Hawley and McClure, 1983; Kobayashi et al., 1990). It is unlikely that our result is a peculiarity of using a synthetic consensus promoter since Siebenlist et al. (1980) found similar results with the *lacUV5* and the phage T7 A3 promoters.

Apart from position −31, as judged by the experiment shown in Figure 3, methylation at most positions in $galP_{con}$ has little or no effect on open-complex formation. In our previous work we showed that alteration of the $galP_{con}$ −35 hexamer from TTGACA to TAGACA causes a 20-fold reduction in promoter activity, but that this is reversed by a T:A to G:C transition at −14 (Chan et al., 1990) presumably because G:C at −14 provides a compensating contact with RNA polymerase. To investigate this point we performed methylation-interference assays with $galP_{con}$ carrying the base-pair changes T:A to A:T at

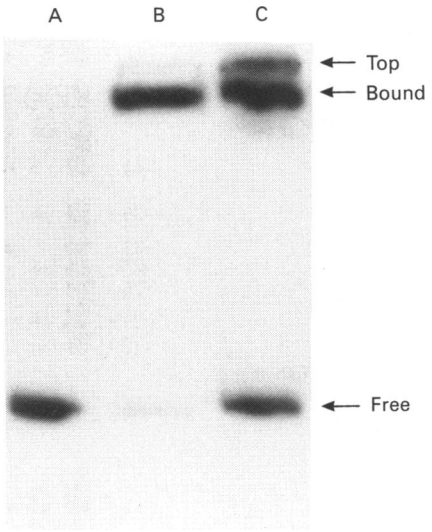

Figure 2 Gel retardation of promoter fragments by RNA polymerase

The figure shows an autoradiograph of a gel-retardation assay, run to separate promoter fragments bound to RNA polymerase from free fragments. RNA polymerase was pre-incubated with labelled fragments carrying $galP_{con}$ either before (lane B) or after (lane C) methylation as described in the Experimental section. Heparin was added and the sample was then loaded on to the gel. Labelled fragment alone was loaded in lane A.

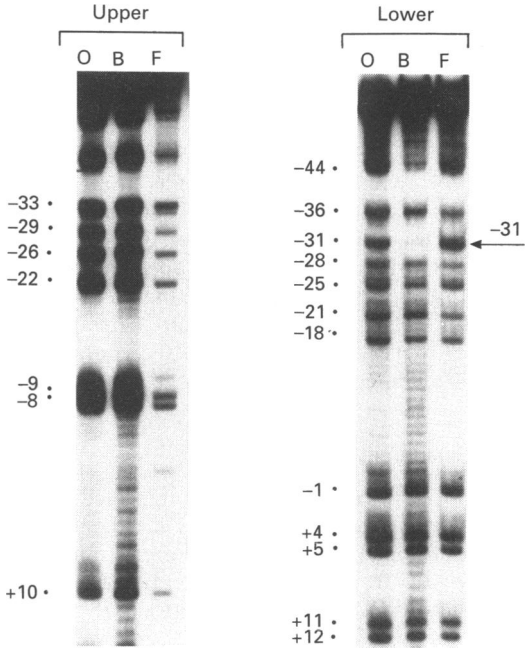

Figure 3 Methylation interference at $galP_{con}$

Free and bound labelled fragments were purified following gel-retardation assays such as that shown in Figure 2. The fragments were cleaved at methylated sites to generate ladders which were visualized by autoradiography after running a sequence gel. The fragments used carried $galP_{con}$ end-labelled at the *Hin*dIII site on either the upper or lower strand. The gel shows the positions of methylated bases in the bound (B) and free (F) fragment. G-sequence ladders were loaded in the lanes marked 'O' and used to calibrate the gel. The position of the missing band, lower-strand G-31, is indicated by an arrow.

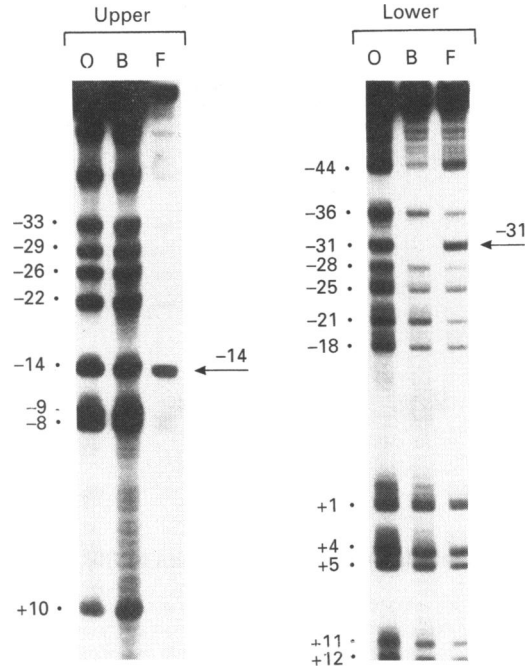

Figure 4 Methylation interference at mutated *galP*$_{con}$

Gel-retardation assays were performed with *galP*$_{con}$ carrying the base-pair changes T:A to A:T at −34 and T:A to G:C at −14. Methylation patterns in bound (B) and free (F) fragments were determined exactly as in Figure 2. The positions of lower-strand G-31 and upper-strand G-14 are indicated.

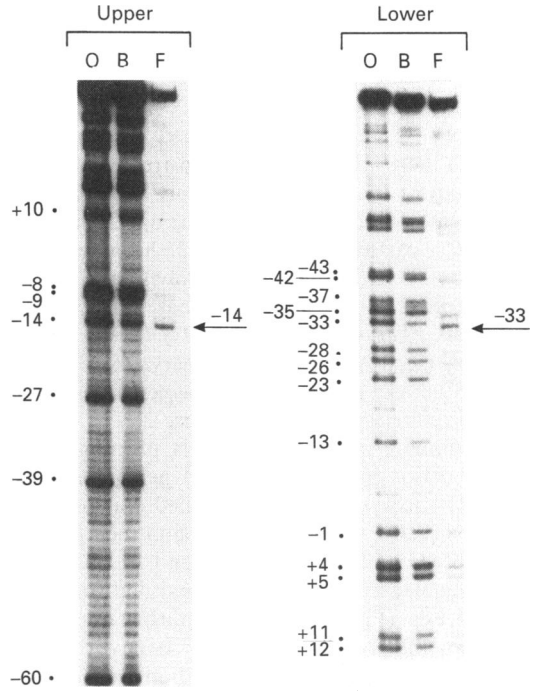

Figure 5 Methylation interference at mutated *galP1*

Gel-retardation assays were performed with *galP1*. Methylation patterns in bound (B) and free (F) fragments were determined as described in Figure 2. The positions of lower-strand G-33 and upper-strand G-14 that are enriched in the 'free' samples are indicated.

−34 and T:A to G:C at −14. The results in Figure 4 show that methylation of upper-strand G-14 results in an increase in the corresponding free labelled fragment but that some modified fragments are found in open complexes. Again, methylation of lower-strand G-31 appears to stop open-complex formation totally. From these results we conclude that G-14 makes a direct contact with RNA polymerase that helps open-complex formation. However, in contrast with the contact with lower-strand G-31, the interaction with G-14 is not essential.

Methylation interference at *galP1*

Methylation-interference assays were repeated with labelled fragments carrying *galP1*. The results in Figure 5 show that methylation of upper-strand G-14 or lower-strand G-33 results in an increase of 'free' labelled fragment. However, neither of these methylations results in disappearance of the corresponding band from the 'bound' DNA and, in contrast with the situation with *galP*$_{con}$, there are no positions at which guanine methylation totally stops open-complex formation.

The effect of G-14 methylation was expected, since we have previously shown that mutations at −14 reduce *galP1* activity and G-14 is protected in open complexes (Chan et al., 1990). However, the effect of methylation at G-33 was surprising since, in our previous work using a variety of probes, we concluded that the *galP1* −35 region in open complexes is exposed and makes no close contacts with polymerase: in methylation-protection assays on open complexes, whilst upper-strand G-14 was protected by polymerase, the reactivities of lower-strand G-33 and G-35 were not affected, and lower-strand G-37 became more reactive (Chan et al., 1990). To explain these results, we propose that modification of G-33 removes a contact or creates a change that hinders formation of one of the transients during open-complex formation, or destabilizes the open complex. Since increasing the incubation time does not increase the fraction of modified DNA in open complexes (results not shown), we can conclude that G-33 methylation does not simply reduce the forward rate constant for open-complex formation. However, it is impossible to provide a single unambiguous explanation for our results, since the amount of any band found in open complexes after heparin challenge and electrophoretic separation of free and bound DNA is a complex function of the rates of both making and breaking open complexes, and G-33 methylation is likely to affect both of these parameters. Notwithstanding, we can conclude that any contact with G-33 is transient and not maintained in the open complex. This is consistent with our previous suggestion that polymerase first attempts to make a 'correct' contact with the bases around −35 but, after failing to find bases resembling the consensus, then 'gropes' further upstream to make compensatory contacts (Chan et al., 1990). Note that, in contrast with G-33, the contact with G-14 is maintained in the open complex.

The importance of G-33 and G-14 at *galP1* was further investigated by mutational analysis. Point mutations were made by site-directed or local mutagenesis and *Eco*RI–*Hin*dIII fragments carrying mutated *galP1* derivatives were cloned into the *lac* expression vector, pAA182. Effects of the mutations on β-galactosidase expression were then measured: note that a Δ*crp* derivative of strain M182 was used as the host strain in this experiment, since expression of *galP1* can be activated by the *crp* gene product and we wanted to measure the effects of mutations on intrinsic promoter activity *in vivo*. Table 1 shows that mutation of the G:C bp at either −14 or −33 leads to a reduction in *galP1* activity. In contrast, mutation of the other G:C bp around the −35 region (at −35, −37 and −39) has no measurable effect, confirming that the involvement of the bp at −33 is specific. The

moderate reductions in promoter activity due to mutations at −33 or −14 are consistent with the suggestion that the guanine residues at these positions provide contacts that help, but are not essential, for open-complex formation. In contrast, mutation of the highly conserved T:A bp at the first position of the −10 hexamer (at −12) leads to a drastic loss of promoter activity (Table 1).

Conclusions

Methylation interference is a useful method for the identification of guanine bases that play a role in DNA–protein interactions. Surprisingly, for recognition of a consensus promoter by RNA polymerase, just one guanine residue (on the lower strand at −31 in the −35 hexamer) is absolutely essential. In contrast, at *galP1*, no guanine residue is essential although modification of lower-strand G-33 and upper-strand G-14 interferes with complex formation. These results underscore the fundamental differences between open complexes at *galP1* and promoters resembling the consensus.

It is apparent that *galP1* is typical of a class of promoters where recognition involves bending and distortion of upstream sequences: the role of these sequences has recently been identified by kinetic studies (Lavigne et al., 1992a). Interestingly, at least at *galP1*, upstream distortion results in DNA wrapping around the polymerase, and in a subsequent reduction in the thermal energy requirement for duplex unwinding around the transcription start (Grimes et al., 1991; Lavigne et al., 1992b). Thus, at *galP1*, in the absence of 'correct' −35-region contacts, the formation of transcriptionally competent complexes depends on both a contact provided by the extended −10 sequence carrying a TG motif, and wrapping of upstream sequences around polymerase. It is not clear whether both types of contacts are needed at all extended −10 promoters, or whether this type of promoter fulfils any particular function in the cell. However, it is striking that some transcription activators appear to act by promoting structures such as those found in open complexes at *galP1* (Busby and Buc, 1987; Zinkel and Crothers, 1991). Indeed the *E. coli* cyclic-AMP-receptor protein appears to accelerate transcription initiation at *galP1* by promoting this structure (Lavigne et al., 1992a and references therein).

We are grateful to Richard Hayward and Annie Kolb for communicating results prior to publication. This work was funded by S.E.R.C. with grant no. GRE 45151. S.B. is an EPA Cephalosporin Fund Research Fellow.

REFERENCES

Bingham, A., Ponnambalam, S., Chan, B. and Busby, S. (1986) Gene **41**, 67–74
Busby, S. and Dreyfus, M. (1983) Gene **21**, 121–131
Busby, S. and Buc, H. (1987) Microbiol. Sci. **4**, 371–375
Casadaban, M. and Cohen, S. (1980) J. Mol. Biol. **138**, 179–207
Chan, B. and Busby, S. (1989) Gene **84**, 227–236
Chan, B., Spassky, A. and Busby, S. (1990) Biochem. J. **270**, 141–148
Gardella, T., Moyle, H. and Susskind, M. (1989) J. Mol. Biol. **206**, 579–590
Gaston, K., Kolb, A. and Busby, S. (1989) Biochem. J. **261**, 649–653
Grimes, E., Busby, S. and Minchin, S. (1991) Nucleic Acids Res. **19**, 6113–6118
Hawley, D. and McClure, W. (1983) Nucleic Acids Res. **11**, 2237–2254
Keilty, S. and Rosenberg, M. (1987) J. Biol. Chem. **262**, 6389–6395
Kobayashi, M., Nagata, K. and Ishihama, A. (1990) Nucleic Acids Res. **18**, 767–772
Kumar, A., Malloch, R., Fujita, N., Smillie, D., Ishihama, A. and Hayward, R. (1992) J. Cell. Biochem. (**suppl. 16E**), 144
Lavigne, M., Herbert, M., Kolb, A. and Buc, H. (1992a) J. Mol. Biol. **224**, 293–306
Lavigne, M., Kolb, A. and Buc, H. (1992b) Biochemistry **31**, 9647–9656
Lodge, J., Fear, J., Busby, S., Gunasekaran, P. and Kamini, N. R. (1992) FEMS Lett. **95**, 271–276
Maxam, A. and Gilbert, W. (1980) Methods Enzymol. **65**, 499–560
McClure, W. (1985) Annu. Rev. Biochem. **54**, 171–204
Miller, J. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
Peakman, T., Busby, S. and Cole, J. (1990) Eur. J. Biochem. **191**, 325–331
Ponnambalam, S., Webster, C., Bingham, A. and Busby, S. (1986) J. Biol. Chem. **261**, 16043–16048
Ponnambalam, S., Chan, B. and Busby, S. (1988) Mol. Microbiol. **2**, 165–172
Siebenlist, U., Simpson, R. and Gilbert, W. (1980) Cell **20**, 269–281
Siegele, D., Hu, J., Walter, W. and Gross, C. (1989) J. Mol. Biol. **206**, 591–603
Steitz, T. (1990) Q. Rev. Biophys. **23**, 205–280
Waldburger, C., Gardella, T., Wong, R. and Susskind, M. (1990) J. Mol. Biol. **215**, 267–276
Zinkel, S. and Crothers, D. (1991) J. Mol. Biol. **219**, 201–215

Received 13 July 1992/19 August 1992; accepted 21 August 1992