The organization of open complexes between *Escherichia coli* RNA polymerase and DNA fragments carrying promoters either with or without consensus –35 region sequences

Bernard CHAN,*‡ Annick SPASSKY†§ and Stephen BUSBY*||

* School of Biochemistry, University of Birmingham, Birmingham B15 2TT, U.K., and † Département de Biologie Moléculaire, Institut Pasteur, 75724 Paris Cédex 15, France

Transcription initiation at the *Escherichia coli* *galP1* promoter does not depend on specific nucleotide sequences in the –35 region. Footprint analysis of transcriptionally competent complexes between *E. coli* RNA polymerase and DNA fragments carrying *galP1* shows that RNA polymerase protects sequences as far upstream as –55, whereas sequences around the –35 region are exposed. In contrast, with *galP1* derivatives carrying –35 region sequences resembling the consensus, RNA polymerase protects bases as far as –45, and the –35 region is fully protected. Taken together, our data suggest that the overall architecture of RNA polymerase–promoter complexes can vary according to whether or not consensus –35 region sequences are present; in the absence of these sequences, open complex formation requires distortion of the promoter DNA. However, the unwinding of promoter DNA around the transcription start is not affected by the nature of the –35 region sequence. With a *galP1* derivative carrying point mutations in the spacer region that greatly reduce promoter activity, the protection of bases by RNA polymerase around the –10 sequence and transcription start site is reduced. In contrast, protection of the region upstream of –25 is unaffected by the spacer mutations, although sequences from –46 to –54 become hypersensitive to attack by potassium permanganate, indicating severe distortion or kinking of this zone. We suggest that, with this *galP1* derivative, RNA polymerase is blocked in a complex that is an intermediate on the path to open complex formation.

INTRODUCTION

The recognition of nucleotide sequences at promoters by RNA polymerase is crucial to the co-ordination of the transcription of different genes in many organisms. In *Escherichia coli*, the major form of RNA polymerase (holoenzyme containing the sigma-70 subunit) interacts with most promoters by recognizing two hexameric sequence elements, known as the –10 and –35 regions, located around 10 bp and 35 bp upstream of the transcription start site (reviewed by McClure, 1985; Travers, 1987). These elements were defined firstly by mutational studies at a number of promoters, but also by the existence of sequence similarities between different promoters; from these similarities, the consensus hexamer sequences, i.e. 5'-TTGACA-3' and 5'-TATAAT-3' for –35 and –10 regions respectively, have been defined. It is generally assumed that RNA polymerase makes specific contacts with the bases in these sequences, and that the activity of any promoter increases as the actual sequence tends towards consensus. Also, it is thought that although the distance between the two hexamer sequences is important, 17 bp being optimal, the precise sequence of the spacer is not crucial in most cases. However, over the past few years we and others have identified a number of promoter sequences that do not obey these simple rules (Keilty & Rosenberg, 1987; Ponnambalam *et al.*, 1988; Collis *et al.*, 1989). In these cases, RNA polymerase recognizes the –10 sequence together with a combination of sequences both upstream and downstream of the –35 regions, but specific nucleotide sequences in the –35 region are not important. The rationale behind the work described here is that the 'architecture' of complexes formed between RNA polymerase

and any promoter DNA might vary fundamentally according to whether or not that promoter contains a 'conventional' –35 region sequence. Thus, using recombinant DNA methods, we have recently constructed a series of fragments carrying promoters with the same –10 sequence, together with either a consensus –35 region sequence or other upstream sequences that ensure activity (Chan & Busby, 1989). Using DNA footprinting techniques together with a variety of chemical probes, we have now investigated the transcriptionally competent complexes formed between RNA polymerase and these DNA fragments. Our results point to a number of key differences in the organization of these complexes which are dependent on the presence or absence of –35 region sequences.

The starting point of this work was a study of the activation of the *E. coli* *galP1* promoter by the cyclic AMP receptor protein (CRP). Unexpectedly, we found that *galP1* was considerably active in the absence of CRP, both *in vitro* and *in vivo*, despite the lack of any clear –35 region sequence (Bingham *et al.*, 1986). Moreover, this activity was hardly affected by point mutations or deletions in the –35 region: *in vitro*, purified RNA polymerase binds tightly to DNA fragments carrying *P1*, giving a clear footprint, and *in vivo*, *galP1* was active in a Δcrp genetic background (Busby *et al.*, 1987). In our most recent study (Chan & Busby, 1989), we showed that this CRP-independent activity was dependent on nucleotide sequences located in the spacer, between the –10 hexamer and the region where the –35 sequence would normally be located. Additionally, optimal activity required specific contacts of sequences around bp –50 by RNA polymerase. In the absence of these elements we showed that *galP1* activity became dependent on the insertion (by cloning) of

Abbreviation used: CRP, cyclic AMP receptor protein.

‡ Present address: Delta Biotechnology, Castle Court, 59 Castle Boulevard, Nottingham NG7 1FD, U.K.

§ Present address: Institut Curie, 11 Rue Pierre et Marie Curie, 75231 Paris Cédex 05, France

|| To whom correspondence should be addressed.

sequences resembling the consensus, 5'-TTGACA-3', into the -35 region. Here we have exploited the constructions resulting from this work to compare the footprints of RNA polymerase at a series of promoters carrying different sequences upstream of the same *galP1* -10 region.

MATERIALS AND METHODS

Recombinant DNA techniques and promoter cloning

The *lac* expression vector pAA182, described by Jayaraman *et al.* (1987), was used to clone DNA fragments carrying *galP1* (and derivatives) in this work. All fragments were cloned between *EcoRI* and *HindIII* sites (see Fig. 1); recombinants contained the promoterless *lac* operon downstream of the *HindIII* site and the pBR322 gene encoding resistance to ampicillin upstream of the *EcoRI* site. Preparations of plasmid DNA, restriction endonuclease treatments, isolation and labelling of DNA fragments and ligations were performed as described by Sambrook *et al.* (1989). As host we used the Δ *lac* strain of *E. coli*, M182, or its Δ *crp* derivative, described by Busby *et al.* (1983). β -Galactosidase expression from pAA182 derivatives in M182 Δ *crp* cells was measured as described by Busby & Dreyfus (1983) and is taken to reflect the activity of the cloned promoter *in vivo*.

By convention, *gal* promoter sequences were numbered with the *P1* transcription start site as +1, upstream sequences being prefixed with a minus sign. The starting *EcoRI*-*HindIII* fragment carrying just *galP1* (fragment f in this work; Fig. 1) has been described in detail by Gaston *et al.* (1988) and is derived from the sequence published by Busby & Dreyfus (1983). The exploitation of the unique *SfaNI* site to make the derivative fragment g, incorporating unique *BamHI*, *XbaI* and *SphI* sites to facilitate gene manipulation, has been described by Gaston *et al.* (1989) and Chan (1989). The derivatives carrying an altered spacer sequence (fragment h) and containing -35 region sequences resembling the consensus (fragments a-e) were described by Chan & Busby (1989). Assays *in vivo* and *in vitro* have shown that the promoter activity of this set of fragments is due to a unique promoter directing transcription initiation at +1 (Bingham *et al.*, 1986; Busby *et al.*, 1987; Gaston *et al.*, 1988; Chan & Busby, 1989). Recently, using gel retardation assays, we have found that complexes between RNA polymerase and each of these fragments contain one molecule of enzyme per molecule of DNA (S. Minchin & S. Busby, unpublished work).

Promoter sequence i was constructed by changing the hexamer sequence in the -35 region of promoter g from 5'-CACTGC-3' to 5'-TTGACC-3'. This was done by cloning synthetic oligonucleotides between the *SphI* and *XbaI* sites of promoter sequence g (Fig. 1). In the experiment described in Fig. 7(b), promoter fragment h was cleaved just upstream of the -10 sequence by introducing an *NcoI* site, 5'-CCATGG-3', from bp -13 to -8. This was done by restricting construction h with *SphI* and, via synthetic oligonucleotides, linking the resulting *EcoRI*-*SphI* fragment to an *NcoI*-*HindIII* fragment carrying *gal* sequence from -11 to the downstream *HindIII* site. This latter fragment was generated from *gal* promoter DNA carrying the p12 mutation, which creates an *NcoI* site (Bingham *et al.*, 1986).

DNA footprinting techniques

'In vitro' DNA footprinting experiments were performed on *PstI*-*HindIII* fragments purified from pAA182 derivatives. These fragments incorporate the smaller *EcoRI*-*HindIII* fragments carrying promoters a-h, and are illustrated schematically in Fig. 7(a); these longer fragments were used to avoid end-artifacts during footprinting. To examine the lower strand, the *HindIII* ends of these fragments were labelled with [γ -³²P]ATP using T4

polynucleotide kinase, whereas for the upper strand the *HindIII* end was 'filled-in' using Klenow fragment of DNA polymerase and [α -³²P]dATP. Typically, labelled DNA fragments (1-2 nm) were preincubated with or without 50 nm-RNA polymerase holoenzyme (purchased from Pharmacia) in transcription buffer, precisely as described by Chan & Busby (1989), before the addition of DNAase I, copper-orthophenanthroline complexes, dimethyl sulphate or potassium permanganate. The precise protocols for handling each of the footprinting reagents have been presented by Spassky *et al.* (1984) (DNAase I), Sigman *et al.* (1985) (copper-orthophenanthroline), Taniguchi *et al.* (1979) (dimethyl sulphate) and O'Halloran *et al.* (1989) (potassium permanganate). After deproteinization and strand cleavage (where necessary), end-labelled fragments were electrophoresed on standard 10% sequencing gels and autoradiography was used to reveal the pattern of cleavage. Scans of autoradiograms were used to quantify changes in band intensities due to RNA polymerase binding. Each gel was calibrated with Maxam-Gilbert G-specific sequence reactions on the fragments used in each experiment (Sambrook *et al.*, 1989).

RESULTS

Promoter activity of DNA fragments carrying *galP1* and derivatives

A series of DNA fragments with similar sequences carrying derivatives of *galP1* was made and is illustrated in Fig. 1(a) (constructions a-h). Each fragment is bounded by an *EcoRI* linker upstream of the promoter and a *HindIII* linker downstream of the transcription start site. To measure the promoter activity of each derivative, the fragments were cloned between the *EcoRI* and *HindIII* sites of the *lac* expression vector pAA182, such that the *lacZ* gene carried by this vector was regulated by the promoter in the cloned fragment. Fig. 1(b) is a histogram showing β -galactosidase expression by recombinant plasmids carrying the fragments a-h in a Δ *crp* Δ *lac* host. Note that we used a Δ *crp* background to measure the intrinsic *galP1* activity in the absence of any stimulation by CRP.

Fragment f carries the starting *galP1* sequence (Gaston *et al.*, 1988) and fragment g is a derivative containing point mutations that created unique sites for *BamHI* (GGATCC: bp -55 to -50), *XbaI* (TCTAGA: bp -44 to -39) and *SphI* (GCATGC: bp -31 to -26). Although in all experiments *in vitro* we have performed to date the promoters in fragments f and g behave identically (results not shown), the promoter activity of sequence g is unexpectedly always about 50% higher than that of sequence f *in vivo* (Fig. 1).

Promoter sequence h is a derivative of sequence g carrying four point mutations in the spacer region (at -13, -16, -21 and -22); we have previously shown that these mutations reduce run-off transcription from *galP1* *in vitro* and expression from *P1* *in vivo* (Chan & Busby, 1989). Promoter sequence e is a derivative of sequence h in which the hexanucleotide 5'-TAGACA-3', corresponding at five out of six positions to the consensus sequence 5'-TTGACA-3', was cloned into the -35 region, creating a derivative that is 5 times more active than the starting *galP1*. Note that the fragments e-h carry common sequences from the -10 hexamer through the transcription start site to the flanking *HindIII* linker. Promoter activity in fragments f and g is independent of the precise -35 region sequence, whereas the activity of fragment e is dependent on the sequence in this region. This is clearly shown by the activity of the related promoter sequences a-d which carry -35 region sequences that progressively approach the consensus. The data in Fig. 1(b) show that single base pair changes in this region have dramatic effects on promoter activity.

Fig. 1. Promoter constructions and activities *in vivo*

(a) Nucleotide sequences of promoter constructions a–h, between the upstream *Eco*RI and downstream *Hind*III linkers (underlined). For each promoter the –10 hexamer sequence is bracketed by vertical bars, together with the position where –35 hexamer sequences should be located. Positions of bases are numbered above the sequences with respect to the transcription start site (+1). (b) Histogram showing the expression of β -galactosidase in M182 Δ *lacC_{crp}* cells harbouring the pAA182 derivatives carrying the promoter constructions a–h. Each bar on the histogram represents β -galactosidase activity in Miller (1972) units and is the average of at least three independent measurements. The activity values for promoters a–h are 20, 30, 240, 3850, 5050, 1000, 1750 and 260 units respectively.

Footprinting of open complexes at *galP1* and derivatives

Using either the enzyme DNAase I (Fig. 2) or the small inorganic complex between cuprous ions and orthophenanthroline (Fig. 3) as a nucleolytic agent, we have investigated the footprint of RNA polymerase at *galP1* (fragment g) or at derivatives where promoter activity became dependent on specific –35 region sequences (fragments d and e). With fragments d and e, a clear footprint is found with DNAase I (Fig. 2, lanes d and e); the upstream limit of the protection is –45. Comparison of results obtained with fragments a–d confirms that the appearance

of this footprint is dependent on the –35 region sequence: progressive deviation from the consensus hexamer results in a reduction of the amount of protection over the whole range of the footprint. Using copper–orthophenanthroline complexes we obtained similar results with fragments a–e (Fig. 3), although the footprint is less clear as, due to the chemistry, there are less reactive positions and less protection overall. Note, however, that there is clear protection of the –35 region of promoters d and e, and additionally there is enhancement of the reactivity of some bands around –5 as reported by Spassky *et al.* (1988).

Fig. 2. DNAase I footprint analysis of selected promoters

Gel analysis of the pattern of DNAase I attack of the lower strand of promoter constructions a, b, c, d, e, g and h. The left-hand panel compares the footprints of promoters that are dependent on sequences in the –35 region, and the right-hand panel includes the footprints of promoters which function in the absence of a specific –35 sequence. Each column shows a calibration marker (M), the pattern of DNAase I cleavage in the absence of polymerase (–) and the pattern of DNAase I cleavage in the presence of polymerase (+). The numbers to the left of each panel are positions with respect to the transcription start site.

Fig. 3. Footprint analysis using complexes of cuprous ions and orthophenanthroline as nicking agent

The same promoter constructions that were examined by DNAase I footprinting in Fig. 2 were subjected to footprinting with copper-orthophenanthroline ions. The nomenclature for this Figure is identical with that of Fig. 2.

With fragment g, carrying active *galP1* without a consensus -35 region sequence, a clear footprint is seen with both reagents (Figs. 2 and 3). However, comparison of these footprints with those found at, say, derivative e, where the promoter is dependent on the -35 region sequence, shows two striking differences.

Firstly, the upstream limit of protection is -55 at *galP1* (g) whereas it is -45 for promoter e (compare e and g in Figs. 2 and 3). Secondly, with copper-orthophenanthroline as a probe, a massive enhancement of reactivity is seen in the -35 region of sequence g, but not with sequence e (Fig. 3, lanes e and g). A

Fig. 4. Protection of guanine residues by RNA polymerase against methylation with dimethyl sulphate

The autoradiograms show the patterns of piperidine cleavage after methylation of guanine residues by dimethyl sulphate, in either the absence (-) or the presence (+) of RNA polymerase, for promoters e, g and i. Each promoter was probed for protection on the lower or upper strands as described in the Materials and methods section. Each band corresponds to a guanine in the nucleotide sequence, thereby allowing direct calibration.

simple explanation for this would be to suppose that whereas RNA polymerase makes contacts with the -35 region at promoters d and e, it makes little or no contact with this region in promoter g, but instead recognizes bases that are further upstream.

We have also compared the footprints of RNA polymerase at promoter g (*galP1*) with those at the derivative (h) carrying mutations in the spacer that weaken *P1* (see Fig. 1). Results with both reagents show that the protection upstream of -25 , including the enhancement of the reactivity of the -35 region with copper-orthophenanthroline complexes, is unaffected by the mutations, whereas protection downstream of -25 is greatly reduced (compare g and h in Figs. 2 and 3).

Methylation protection by RNA polymerase at *galP1*

Fig. 4 shows the results of an experiment where DNA fragments carrying *galP1* (in sequence g) were exposed to dimethyl sulphate, in either the presence or the absence of RNA polymerase, in order to locate the guanine residues that might be protected from methylation by RNA polymerase. The results with both the lower and upper strands show that there is no strong protection at any position; the most striking effect is the 3-fold enhancement of the modification of G -37 on the lower strand, implying that this region becomes more exposed to the reagent. In contrast, with promoter sequence e, the derivative of *galP1* that is dependent on the -35 sequence, a clear protection was found at G -31 on the lower strand, implying that RNA polymerase makes close contacts with this residue, which corresponds to position 5 of the -35 hexamer sequence 5'-TAGACA-3'. Interestingly, we found no protection of G -33 on the upper strand, which corresponds to position 3 of the hexamer.

In order to check that the differences in methylation protection of the -35 region found between promoters e and g were not due to trivial differences in the flanking sequences, we cloned a near-

consensus -35 hexamer into the -35 region of sequence g. In this construction (i; not shown in Fig. 1), the -35 region of promoter g, i.e. 5'-CACTGC-3', is replaced by 5'-TTGACC-3', which corresponds to the consensus at five out of six positions. With this construct (i in Fig. 4), RNA polymerase protects G -31 on the lower strand and the enhancement at G -37 seen with sequence g is suppressed. Additionally, the lower strand bands at -30 and -13 are enhanced and the upper strand band at -14 is protected. Thus the introduction of a -35 region consensus sequence at *galP1* alters close contacts made by RNA polymerase and allows polymerase to bind tightly in the -35 region.

Unwinding of DNA in open complexes

The results with DNAase I, copper-orthophenanthroline complexes and dimethyl sulphate reveal differences in the organization of open complexes at *galP1* according to whether or not a consensus-like -35 region sequence is present. Fig. 5 shows an experiment designed to determine whether the amount of unwinding of the DNA duplex in open complexes differs according to the nature of the -35 sequence. Complexes between RNA polymerase and promoter sequences e and f were exposed to potassium permanganate, which attacks only bases in unwound or distorted DNA (Borowiec *et al.*, 1987; O'Halloran *et al.*, 1989). The positions at which bases were modified can be deduced from the data shown in Fig. 5; at promoters e and f, modification is dependent on the addition of RNA polymerase. As expected, the major sites for modification are around the -10 region and the transcription start site, and are confined to bases from -11 to $+3$. Significantly, the pattern of modification is independent of the presence or absence of a -35 region consensus-like sequence (compare e and f on both lower and upper strands), strongly suggesting that the extent of unwinding of the duplex in open complexes at *galP1* is independent of the upstream sequences. Note that the fainter bands seen down-

Fig. 5. Detection of promoter unwinding using potassium permanganate

The gels show the pattern of cleavage resulting from potassium permanganate and piperidine treatment of *Pst*I-*Hind*III fragments carrying promoter constructions e, f and h, in the absence (-) or presence (+) of RNA polymerase. Panels (a) and (b) show the attack on the lower and upper strands respectively for each of the promoters. Each experiment also carries a calibration marker (M) made by the Maxam-Gilbert G-specific sequence reaction on one of the promoter constructions. The gels are calibrated with respect to the transcription start site (+1).

Fig. 6. Time course of appearance of permanganate-sensitive bases at *galP1*

The *PstI*–*HindIII* fragments of promoters h and g, labelled on the bottom strand, were incubated with RNA polymerase for different time intervals (0, 1, 2, 5, 10 and 20 min) before treatment with potassium permanganate. The Figure shows autoradiograms of gels on which the resulting fragments were separated; the calibration with a marker (M) is exactly as described in the legend to Fig. 5.

stream of +3 with promoter sequence f and are due to trace amounts of polymerase binding to a separate cryptic promoter (B. Chan, S. Minchin & S. Busby, unpublished work).

Evidence for an intermediate during open complex formation at *galP1*

Fig. 5 also shows the effects of potassium permanganate on complexes between RNA polymerase and promoter sequence h. Recall that sequence h is derived from the starting *galP1* sequence (f and g) but carries four point mutations in the spacer between the –10 and –35 regions, which cause a sharp reduction in *galP1* activity (Fig. 1). Recall also that the footprint study of RNA polymerase binding to this promoter (Figs. 2 and 3) shows that protection upstream of –25 is maintained, whereas downstream protection is weakened. The results in Fig. 5 show that the unwinding of the duplex in the –10 region of promoter h is qualitatively identical to that found with the wild type *galP1* (fragment f) or *galP1* carrying a –35 region sequence that resembles the consensus (fragment e). However, surprisingly, strong bands appear with both the upper and lower strands, corresponding to attack of bases in the region from –45 to –54. As it is unlikely that the duplex is being unwound at this point, our working hypothesis is that the appearance of these bands is due to distortion of the DNA in this region. Since these bands are only faintly visible with the wild-type promoter (sequence f), we propose that this distortion occurs in a transient intermediate during open complex formation, and that with promoter sequence h some of the polymerase–promoter complexes become blocked in this conformation. Some support for this view comes from the experiment shown in Fig. 6, where fragments carrying promoter sequences g or h were incubated with RNA polymerase for different times before exposure to potassium permanganate: the time course of the appearance of bands around –50 is identical with the appearance of bands around –10.

Further evidence that the reactivity around –50 is due to RNA polymerase interacting with *galP1* comes from the experiment shown in Fig. 7, where the *PstI*–*HindIII* fragment

Fig. 7. Potassium permanganate footprints of RNA polymerase binding to *galP1* on fragments of different sizes

(a) Schematic diagram (not to scale) of the *PstI*–*HindIII* fragment of promoter construction h, showing the *HaeIII* site in pBR322 sequence and the *NcoI* site introduced by gene manipulation (see the Materials and methods section). This fragment was cleaved at either of these sites to produce a *HaeIII*–*HindIII* fragment lacking most of the upstream pBR322 DNA, or a *PstI*–*NcoI* fragment that loses the –10 sequence. Note that the diagram is not to scale: the *PstI*–*EcoRI* segment covers co-ordinates 3612–4360 of pBR322 with the *HaeIII* site at 4344, and the *EcoRI*–*HindIII* fragment is 144 bp as described by Busby & Dreyfus (1983). (b) Potassium permanganate footprints of RNA polymerase binding to *PstI*–*HindIII* (i), *HaeIII*–*HindIII* (ii) and *PstI*–*NcoI* (iii) fragments carrying promoter sequence h. The gels are labelled exactly as described in the legend to Fig. 2.

carrying sequence h, used in the experiments in Figs. 5 and 6, was shortened by restriction. Firstly this fragment was cleaved at the *HaeIII* site located in the pBR322 sequence just upstream of the *EcoRI* site (see Fig. 7a). Fig. 7(b) shows that the removal of the flanking pBR322 sequence had no effect on the pattern of permanganate-induced cleavage of open complexes. In contrast, cleavage of the sequence at –12 (at an *NcoI* site that was introduced by gene manipulation: see the Materials and methods section) resulted in total loss of the enhancement of reactivity in the –50 region. We can conclude that the effects of RNA polymerase on the structure of the sequence around –50 in promoter h are dependent on contacts around –10.

DISCUSSION

The starting point of this work was the observation that the recognition of the *E. coli galP1* promoter by RNA polymerase did not require specific contacts with the sequence in the –35 region (Ponnambalam *et al.*, 1986; Chan & Busby, 1989), and the suggestion that a number of *E. coli* promoters may not carry essential –35 region sequences (Keilty & Rosenberg, 1987; Ponnambalam *et al.*, 1988). Although it is clear that promoter activity in many cases is contingent on specific interactions between RNA polymerase sigma factor and –35 region

Fig. 8. Schematic representation of unwinding in open complexes at three different promoters

The nucleotide sequences of promoters e, f and h are shown, with carets above or below these bases that are cleaved by potassium permanganate; double carets indicate strong bands as shown in the gels in Fig. 5. The duplexes are shown separated at the region proposed to be unwound.

sequences resembling the consensus (Gardella *et al.*, 1989; Siegele *et al.*, 1989; Zuber *et al.*, 1989), there are a number of promoters (for example *galP1*) in which the actual -35 region sequence hardly resembles the consensus (Rosenberg & Court, 1979; Harley & Reynolds, 1987). In 1980, Siebenlist *et al.* were able to show that RNA polymerase interacts homologously with two different promoters. Here we suggest that the interactions vary according to whether or not a -35 region consensus-like sequence is present.

DNA manipulation studies with *galP1* have shown that specific nucleotide sequences, both in the spacer region and around -50 , are essential for optimal promoter activity; in the absence of these sequences, expression is dependent on -35 region sequences resembling the consensus (Busby *et al.*, 1987; Chan & Busby, 1989). The results from footprinting studies presented here suggest that although RNA polymerase contacts bases up to -55 at *galP1*, the -35 region is exposed; bases in this region are not protected from dimethyl sulphate and a massive enhancement of reactivity with copper–orthophenanthroline ions is observed. Interestingly, copper–orthophenanthroline ions cleave DNA by attacking the minor groove at particular positions, depending on local geometry or deformation of the groove (Spassky *et al.*, 1988; Sigman & Spassky, 1989). We suggest that, in this case, distortions in the DNA around the -35 region are responsible for the enhancement of cleavage. In sharp contrast, in the cases where the *galP1* promoter carries sequences resembling the consensus, this enhancement is not observed and the -35 region sequences are protected by bound RNA polymerase: clear protection against dimethyl sulphate is seen. Additionally, the upstream limit of the footprint is at -45 , implying that the architecture of the open complex varies according to the presence or absence of -35 region sequences. Interestingly, the unwinding of the duplex from $+1$ to -10 is not influenced by the -35 region sequence. Fig. 8 shows that, according to the pattern of attack by potassium permanganate, the pattern of unwinding is the same for promoters e and f.

Taken together, these results provide evidence to support our proposal (Busby *et al.*, 1987) that at promoters lacking -35 consensus sequences, RNA polymerase, on failing to make a correct contact in the -35 region, ‘gropes’ further upstream to

make compensatory contacts, and that this involves distortion of the DNA. The simplest explanation would be that bending or kinking of the DNA, contingent on the precise sequence of the spacer region, allows the region around -50 to come into contact with RNA polymerase, possibly by wrapping around the protein. Although we have no definite proof that distortion of promoter DNA is involved in this case, two recent reports lend credence to our point of view. Firstly, Bracco *et al.* (1989) reported that synthetic curved DNA sequences can act as transcriptional activators *in vivo*. Secondly, Collis *et al.* (1989), after measuring the promoter activity of a number of synthetic constructs, concluded that “favourable flexibility of curvature of DNA in the close vicinity of potential -35 and -10 promoter regions seems to be a significant factor in determining where *E. coli* RNA polymerase starts RNA chains”. It is worth noting that in both of these cases, because the authors could not demonstrate transcription *in vitro*, it was unclear whether the increased promoter activity was due to the curvature of DNA directly allowing contacts with polymerase or ‘setting up’ the nucleosome proteins in such a way as to favour initiation. In our cases it is clear that sufficient direct polymerase–DNA contacts can form to allow initiation; the presence or absence of -35 consensus sequences determine the nature of the contacts upstream of the -10 sequence, while not affecting contacts around and downstream of -10 . There may therefore be several ways by which RNA polymerase can get to the same start site.

A further strand of evidence arguing in favour of extensive DNA distortion during transcription initiation at *galP1* comes from consideration of the footprint of RNA polymerase at promoter sequence h in Figs. 2, 3 and 5. Recall that this promoter is derived from the synthetic *galP1* promoter sequence g, with the insertion of four point mutations in the spacer region that reduce transcription by 7–8-fold (Fig. 1). In most cases point mutations that lower promoter activity reduce the speed of open complex formation and cause even reduction in the footprint across the polymerase binding site (e.g. Johnston *et al.*, 1987). Very unusually, the effect of the spacer mutations is to reduce the footprint around the transcription start site while hardly affecting protection of upstream sequences, clearly implying that these mutations cause RNA polymerase to be ‘trapped’, possibly at an

intermediate during open complex formation. Interestingly, DNA sequences in this complex on both strands around -50 are hyper-reactive to potassium permanganate (results from Fig. 4 are presented schematically in Fig. 8). Since all residual transcription from this promoter still initiates at +1 (Chan & Busby, 1989), and since it is unlikely that the duplex is unwound at this point, we conclude that the DNA in this region is distorted by RNA polymerase. Circumstantial corroboration for this point of view comes from Borowiec *et al.*, (1987), who reported a study of the binding of *lac* repressor to supercoiled plasmids carrying the *lac* promoter region. These authors showed that a repressor tetramer could bind simultaneously to the primary *lac* operator and to a weaker-binding *lac* pseudo-operator located 93 bp upstream. The formation of this complex was accompanied by the appearance of hypersensitivity to attack by permanganate at a 5 bp sequence located midway between the operators. These results were interpreted in terms of a complex in which the *lac* repressor held the two operators together in a DNA loop which induces a severe distortion or kink in the 5 bp sequence. Remarkably, the nucleotide sequence at this point is 5'-TTTAT-3', which corresponds exactly to the *galP1* sequence from bp -49 to -45 in the zone where we find hyper-reactivity (Fig. 8). The logical conclusion from this coincidence is that this sequence element can behave as a centre for protein-induced distortion.

This work was supported by grants from the U.K. Science and Engineering Research Council (grant GRE 45151) and the Wolfson Foundation.

REFERENCES

- Bingham, A., Ponnambalam, S., Chan, B. & Busby, S. (1986) *Gene* **41**, 67-74.
- Borowiec, J., Zhang, L., Sasse-Dwight, S. & Gralla, J. (1987) *J. Mol. Biol.* **196**, 101-111.
- Bracco, L., Kotlarz, D., Kolb, A., Diekmann, S. & Buc, H. (1989) *EMBO J.* **8**, 4289-4296.
- Busby, S. & Dreyfus, M. (1983) *Gene* **21**, 121-131.
- Busby, S., Kotlarz, D. & Buc, H. (1983) *J. Mol. Biol.* **167**, 259-274.
- Busby, S., Spassky, A. & Chan, B. (1987) *Gene* **53**, 145-152.
- Chan, B. (1989) Ph.D. thesis, University of Birmingham.
- Chan, B. & Busby, S. (1989) *Gene* **84**, 227-236.
- Collis, C., Molloy, P., Both, G. & Drew, H. (1989) *Nucleic Acids Res.* **17**, 9447-9468.
- Gardella, T., Moyle, H. & Susskind, M. (1989) *J. Mol. Biol.* **206**, 579-590.
- Gaston, K., Chan B., Kolb, A., Fox, J. & Busby, S. (1988) *Biochem. J.* **253**, 809-818.
- Gaston, K., Kolb, A. & Busby, S. (1989) *Biochem. J.* **261**, 649-653.
- Harley, C. & Reynolds, R. (1987) *Nucleic Acids Res.* **15**, 2343-2361.
- Jayaraman, P.-S., Peakman, T., Busby, S., Quincey, R. & Cole, J. (1987) *J. Mol. Biol.* **196**, 781-788.
- Johnston, F., Ponnambalam, S. & Busby, S. (1987) *J. Mol. Biol.* **195**, 745-748.
- Keilty, S. & Rosenberg, M. (1987) *J. Biol. Chem.* **262**, 6389-6395.
- McClure, W. (1985) *Annu. Rev. Biochem.* **54**, 171-204.
- Miller, J. (1972) *Experiments in Molecular Genetics*, Cold Spring Harbour Laboratory, Cold Spring Harbor, NY.
- O'Halloran, T., Frantz, B., Shin, M., Ralston, D. & Wright, J. (1989) *Cell* **56**, 119-129.
- Ponnambalam, S., Webster, C., Bingham, A. & Busby, S. (1986) *J. Biol. Chem.* **261**, 16043-16048.
- Ponnambalam, S., Chan, B. & Busby, S. (1988) *Mol. Microbiol.* **2**, 165-172.
- Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genetics* **13**, 319-353.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Siebenlist, U., Simpson, R. & Gilbert, W. (1980) *Cell* **20**, 269-281.
- Siegele, D., Hu, J., Walter, W. & Gross, C. (1989) *J. Mol. Biol.* **206**, 591-603.
- Sigman, D. & Spassky, A. (1989) in *Nucleic Acids and Molecular Biology*, (Eckstein, F. & Lilley, D., eds.), vol. 3, pp. 13-27, Springer-Verlag, Heidelberg.
- Sigman, D., Spassky, A., Rimsky, S. & Buc, H. (1985) *Biopolymers* **24**, 183-197.
- Spassky, A., Busby, S. & Buc, H. (1984) *EMBO J.* **3**, 43-50.
- Spassky, A., Rimsky, S., Buc, H. & Busby, S. (1988) *EMBO J.* **7**, 1871-1879.
- Taniguchi, T., O'Neill, M. & de Crombrughe, B. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5090-5094.
- Travers, A. (1987) *Crit. Rev. Biochem.* **22**, 181-219.
- Zuber, P., Healy, J., Carter, H., Cutting, S., Moran, C. & Losick, R. (1989) *J. Mol. Biol.* **206**, 605-614.

Received 7 March 1990/25 April 1990; accepted 1 May 1990