Volume 10 Number 16 1982 Nucleic Acids Research

Transcription initiation sites within an IS2 insertion in a Gal-constitutive mutant of *Escherichia coli*

Deborah M.Hinton and Richard E.Musso

Cancer Biology Program, Frederick Cancer Research Center, Frederick, MD 21701, USA

Received 28 October 1981; Revised and Accepted 17 July 1982

ABSTRACT

Insertion of the insertion sequence IS2(I) directly before the galE gene of the galactose operon results in a Gal minus phenotype (1,2). The Gal-constitutive allele gal^c200 (and its deletion derivative gal^c200 Δ31) arise from such a Gal minus mutant by the insertion of IS2(II) DNA within the IS2(I) sequence (3). We have transcribed *in vitro* a DNA template representing the IS2-galE region of gal^c200 Δ31. Gal-directed transcription initiates at two sites within the IS2(I) sequence, 51 and 52 bp from the IS2-galE junction. The promoter for these transcripts, Pgal200Δ31, is composed of a novel joint between a -10 region from the IS2(I) DNA and a -35 region contributed by the IS2(II) insertion. No promoters intrinsic to the 121 bp of the IS2(II) sequence also present on the template were detected. The relevance of Pgal200Δ31 to the Gal^c phenotype of gal^c200 and to general mechanisms for the constitutive expression of genes adjacent to IS2 is discussed.

INTRODUCTION

Bacterial insertion sequences (IS elements) are among the simplest known transposable elements, being less than 1500 bp in size and lacking known genetic markers (for reviews see 4-6). One of these elements, IS2, occurs normally within the chromosome of *Escherichia coli* (7) and has also been isolated as a rare mutational insertion at various sites in bacterial and phage operons (1,2,8-12). Genetic analyses of these mutations have shown that IS2 affects the expression of genes adjacent to its insertion. Insertion of IS2 in one orientation (I) near the operator/promoter region of an operon blocks transcription from the promoter, thus preventing the expression of downstream genes (1,2,9,12). This polar effect of IS2(I) arises from the presence of a rho-dependent termination site within the insertion element in this orientation (13). On the other hand, the disruption of an operon by IS2 in orientation II can lead to the constitutive expression of genes distal to its insertion (8,11,14). This observation has led to the speculation that, in orientation II, the IS2 DNA contains promoter(s) that direct transcription

© IRL Press Limited, Oxford, England.

into adjacent DNA (5,8,11). The DNA sequence of IS2, however, does not appear to contain sequences unambiguously related to other known bacterial promoters (15).

In addition to the insertion of IS2 in orientation II, genes downstream from the insertion of altered IS2(I) sequences may also be expressed constitutively (3,16-19). Several such IS2(I) derivatives have been obtained by selection of spontaneous Gal$^+$ pseudo-revertants from the Gal$^-$ strains, gal3::IS2(I) (9) and galOP-308::IS2(I) (1,2). These independently isolated Gal$^-$ mutants both have a polar IS2(I) insertion in the galactose operon immediately preceding the start of the galE message (3,15) (Figure 1). DNA sequence analyses of the IS2(I) derivatives of gal3 and galOP-308 have revealed that they all involve DNA insertions or sequence changes located 50 to 100 bp upstream from the IS2(I)-galE junction (3,19-21). The mechanism whereby these localized alterations of the polar IS2(I) sequence convert a Gal$^-$ phenotype into Gal-constitutive expression is unclear. Although one might speculate that the IS2(I) sequence alterations generate new gal-directed promoters, the DNA sequence analyses have not generally demonstrated obvious homologies to known promoter sequences (3,20,21).

To help elucidate how the insertion of IS2(II) or altered IS2(I) DNA results in the constitutive expression of adjacent genes, we have investigated the alleles gal^{c}200 and gal^{c}200 Δ31 (Figure 1). The allele gal^{c}200 is a Gal-constitutive derivative of gal3 which expresses the gal genes at a level twice that of the fully induced gal operon (3). Our previous analyses of gal^{c}200 have shown that it contains a complete IS2 in orientation II nested within the IS2(I) DNA (3). This IS2(II) insertion occurs between positions -65 and -66 of the IS2(I) sequence. Thus, in gal^{c}200, only 65 bp of IS2(I) DNA are adjacent to galE; the remaining IS2(I) DNA (-66 to -1327) lies upstream from the IS2(II) insert. Such a structure for promoting constitutive gal expression is apparently not unique. Besemer et al. (19) have reported an identical structure for an independently isolated Gal$^+$ revertant of galOP-308, designated gal-308^c-1-0. Interestingly, the Gal-constitutive phenotype of gal^{c}200 is retained even when most of the IS2 DNA has been removed by a deletion called Δ31 (22). This deletion leaves only 121 bp of the IS2(II) sequence (-1207 to -1327) and fuses the remaining IS2(II) segment to bacterial DNA that is normally several kilobases upstream from the gal genes. Thus, gal^{c}200 and gal^{c}200 Δ31 contain both an IS2(II) and an altered IS2(I) sequence (at the IS2(II) insertion site), providing an opportunity to examine how these IS2 sequences result in the constitutive expression

of adjacent genes. *In vitro* transcription of DNA from these alleles should indicate whether the IS*2*(II) or the altered IS*2*(I) DNA contain functional promoter sequences. In this paper, we report the results of the *in vitro* transcription from one such DNA template, *gal*^{c}*200 Δ31*.

MATERIALS AND METHODS

a) Chemicals and enzymes. Restriction enzymes, T_4 DNA Ligase and T_4 DNA polymerase were purchased from Bethesda Research Laboratories, Inc. Ribonucleases (RNases) T_1,T_2, and U_2 were obtained from Calbiochem. Pancreatic ribonuclease was obtained from Worthington and ribonuclease P_1, from P-L Biochemicals, Inc. RNA polymerase holoenzyme was prepared according to the procedure of Berg *et al*. through step five (23).

The [α-^{32}P] ribonucleoside triphosphates (>400 Ci/mmol) were purchased from New England Nuclear Corp. All other nucleotides were from P-L Biochemicals, Inc. The plasmid pBRHE*gal*^{c}*200 Δ31* (3) was the generous gift of A. Ahmed. Native DNA was extracted and separated strands were prepared as previously described (24) from the lambda phages λ*c*I857*r*32*S*7 (λr32) (2) and λp*gal*8*c*I857*S*7 (λp*gal*8) (35). [α-^{32}P]ATP-labeled 6S RNA (25) was generated by *in vitro* transcription of a *Hae*III restriction fragment of λ DNA. Cellogel 250 strips were purchased from Kalex Scientific Co. and polyethyleneimine (PEI) thin layer chromatograms (Cel 300) were obtained from Brinkmann Instruments, Inc. DEAE cellulose plates (250 micron, cellulose: DEAE:: 9:1) were obtained from Analtech, Inc. DE 81 paper was purchased from Whatman, Inc. The mobilities of nucleotides and oligonucleotides on this paper differed somewhat from earlier published values (26, 27) due to a change in the paper by the manufacturer.

b) Isolation of HincII DNA fragment. A 550 bp *Hin*cII fragment for *in vitro* transcription was isolated from a total *Hin*cII digest of the plasmid pBRHE*gal*^{c}*200 Δ31* by electrophoretic separation on a 1.5% agarose gel. The band was cut from the gel and eluted electrophoretically. After concentration and removal of ethidium bromide by extractions with isobutanol, the DNA was loaded on a DEAE cellulose column (1.0 x 1.0 cm). The column was washed with buffer A (50 mM Tris(hydroxymethyl)aminomethane (Tris)•HCl, pH 8.0; 1 mM ethylenediamine tetraacetic acid (EDTA)) and 150 mM NaCl in buffer A. The DNA was eluted by a solution of 1 M NaCl in buffer A, precipitated with ethanol, dried, and stored in water at -20°C.

c) In vitro transcription reactions. *In vitro* transcriptions were performed in reaction mixtures (25 μl) containing 20 mM Tris.HCl, pH 8.0;

75 mM KCl; 5 mM $MgCl_2$; 0.1 mM EDTA; 0.5 mM dithiothreitol; 50 μg/ml bovine serum albumin; 0.05 μg DNA; 15 μg/ml RNA polymerase holoenzyme; 100 μg/ml heparin; and all four 5'-ribonucleoside triphosphates (NTPs) (27). After incubation of the first seven components at 37°C for 10 min, heparin was added and the resulting mixture kept at 37°C for an additional 2 min. In vitro transcription was begun by adding a solution of the 5'-ribonucleoside triphosphates, giving the following nucleotide concentrations: for either labeled UTP or CTP precursor: 10 μM [α-^{32}P]UTP (or CTP) (300 Ci/mmol), 50 μM CTP (or UTP), 200 μM ATP, 200 μM GTP; for either labeled ATP or GTP precursor: 100 μM [α-^{32}P]ATP (or GTP) (40 Ci/mmol), 200 μM GTP (or ATP), 50 μM CTP, 10μM UTP. After 20 min, a solution (155 μl) containing 65 mM Tris•HCl, pH 7.5; 44 mM $MgCl_2$; 0.3% sodium dodecyl sulfate; and 87 μg/ml tRNA was added, and the reaction stopped by phenol extraction and ethanol precipitation. The RNA was resuspended in a sample buffer containing 89 mM Tris•borate, pH 8.3, 2.5 mM EDTA, 7 M urea, 0.1% bromophenol blue, and 0.1% xylene cyanol FF. The sample was heated at 70°C for 3 min and electrophoresed on a 5% polyacrylamide, 7 M urea denaturing gel. Labeled RNA species were detected by autoradiography, cut from the gel, and eluted electrophoretically. Carrier tRNA (100 μg) and sodium acetate (to give a final concentration of 1%) were added and the RNA was precipitated in ethanol.

d) Hybridization of RNA to single-stranded DNA. The major RNA species after in vitro transcription was hybridized to the r strand of λr32 DNA in a solution (1 ml) containing 4.7 μg/ml r strand λr32 DNA, ^{32}P-labeled RNA isolated as described above, 3% phenol, 3 mM NaOH, and 2 x SSC (1 x SSC: 0.15 M NaCl, 15 mM sodium citrate, pH 7.0) (28). After incubation at 66°C for 4 hr, the RNA-DNA hybrid was collected on a nitrocellulose filter by dropwise filtration. Single-stranded RNA was eliminated by incubating the filter at 22°C for 30 min in a solution (2 ml) of 5 U/ml ribonuclease T_1 in 2 x SSC. Following the ribonuclease treatment, the nuclease T_1 was inactivated by incubating the filter at 55°C for 1 hr in a solution of 0.17 M iodoacetic acid, 0.17 M NaOH, 0.10 M NaCl, and 50 mM sodium acetate, pH 5.6. The RNA was released from the filter by heating in water (90°C) for 3 min and 100 μg carrier tRNA added. After ethanol precipitation, the RNA was dried and analyzed as described below.

e) Enzymatic digestion of RNA. Two-dimensional fingerprints of the ^{32}P-labeled RNA were generated by standard techniques (26): RNA prepared with each of the four [α-^{32}P]NTPs (plus 100 μg carrier tRNA) was dissolved in a

solution (6 µl) containing 1 x RNase buffer (20 mM Tris•HCl, pH 7.5; 1 mM EDTA) and 15 U T_1 ribonuclease and incubated at 37°C for 45 min. The T_1 products were then fractionated in a two-dimensional system (first dimension: electrophoresis on Cellogel strips at pH 3.5; second dimension: chromatography on DEAE cellulose plates developed with a 1:1 ratio of 45 min and 60 min hydrolyzed Homomixture B.) The separated T_1 products were eluted from the DEAE cellulose with 30% triethylammonium bicarbonate, pH 9.0, and dried. Secondary digestion of these oligomers was then performed as described (26) by treatment with pancreatic ribonuclease (2.5 µg in 5 µl of 1 x RNase buffer) or ribonuclease U_2 (0.1 U in 5 µl containing 20 mM sodium acetate, pH 4.8; 1 mM EDTA). The secondary products were separated on DE 81 paper by electrophoresis at pH 3.5.

To determine the 5' end of the transcript, ^{32}P-labeled RNA (plus 100 µg carrier tRNA) was digested with nuclease P_1 (10 µg in 10 µl sodium acetate pH 5.6) (29,30). The P_1 products were separated by two-dimensional chromatography on PEI thin layer plates. (First dimension: step by step separation in 0.5, 2.0 and 4.0 M sodium formate, pH 3.4; second dimension: 0.75 M KH_2PO_4 pH 3.4) (31). To determine the penultimate 5'-residue, ^{32}P-labeled RNA (plus 100 µg carrier tRNA) was digested with 2 U nuclease T_2, 20 U ribonuclease T_1 and 0.5 µg ribonuclease A in 10 µl 50 mM ammonium acetate, pH 4.5 (26). The T_2 products were separated by electrophoresis on DE 81 paper at pH 1.7.

f) Plasmid Constructions. To construct pAST202, we deleted DNA between the EcoRI and HpaI sites of pBRHEgal^{c}200 Δ31 as follows. DNA of the latter plasmid was digested with EcoRI. The staggered EcoRI ends were filled with T_4 DNA polymerase as described by Wartell and Reznikoff (36). The DNA was then digested with HpaI and finally circularized by ligation with T_4 DNA ligase (0.7Weiss units for 16 hr at 16° and 2 µg/ml DNA concentration). The bacterial strain SA1171 [E. coli K12 Δ(galETK) galR$^-$] (obtained from Dr. S. Adhya) was made competent for transformation (37) and transformants were selected as ampicillin resistant colonies. Plasmid DNA was prepared (38) from several isolates and characterized by size and restriction sites. Most isolates had the plasmid structure expected for pAST202 as indicated in Fig. 5.

The plasmid pAST100 was constructed from pBR322 and an EcoRI-HindIII fragment containing the gal promoter using λpgal8 DNA as a source for the latter. DNA from λpgal8 and pBR322 were each digested to completion by HindIII and EcoRI, mixed together at 15 µg/ml of the former and 5 µg/ml of the latter and ligated with T_4 DNA ligase (0.05U) for 16 hr at 16°. Competent cells of SA1171 were transformed and Ap^r cells selected and then screened

for growth on L-agar plates containing tetracycline at 0, 10, or 20 μg/ml. From 50 Ap^r colonies eight were found to grow well at 10 μg/ml tetracycline and each was found to have the desired plasmid, pAST100, with the 1.1 Kb EcoRI-HindIII fragment containing the gal promoter.

g) Determination of promoter strength in vivo. In the plasmid pBR322 part of the promoter for the tet gene lies between the EcoRI and HindIII sites (39). Thus a deletion of this region will result in transformants which are tetracycline-sensitive unless other DNA is inserted which provides a new promoter for transcribing the tet gene. Previous studies on a plasmid pBdCl, which is identical to pASTI, have shown that tet is expressed from the gal promoter (40) and we have used this plasmid as a reference to estimate the relative promoter activity of DNA segments derived from gal^c200 Δ31. This is based on determining the level of tetracycline resistance conferred by specific plasmids as follows. Single colonies of SA1171 transformants containing the desired plasmids were grown overnight in L broth containing 100 μg/ml ampicillin (L-amp) and then diluted 100X into L-amp and grown to mid-log phase. Each culture was then diluted and tested for plating efficiency on L-agar plates containing ampicillin (40 μg/ml) and different concentrations of tetracycline (0,5,10,20, or 50 μg/ml). The results for the plate containing only ampicillin was taken as an efficiency of plating (EOP) = 1.0.

RESULTS

To characterize the transcription products from gal^c200 Δ31, we isolated a 550 bp HincII fragment which contains both the IS2(II) and the IS2(I) segments, ~ 320 bp of galE and 39 bp of bacterial DNA to the left of Δ31 (Figure 1). This template was transcribed in vitro under standard conditions in the presence of [α-^{32}P] ribonucleoside triphosphates. Separation of the radioactive RNA products on a denaturing gel revealed one major RNA species migrating with an apparent size of ~ 400 bp (Figure 2).

The isolated RNA from the HincII template was digested with T_1 ribonuclease and the resulting oligonucleotides were separated by standard two-dimensional fingerprinting techniques (Figure 3A). Although a complex fingerprint was generated, the overall pattern was similar to that of galE maps obtained in our laboratory by in vitro transcription from the gal promoter (data not shown). This suggested that most of the T_1 oligomers were derived from the transcription of the galE gene. To determine whether

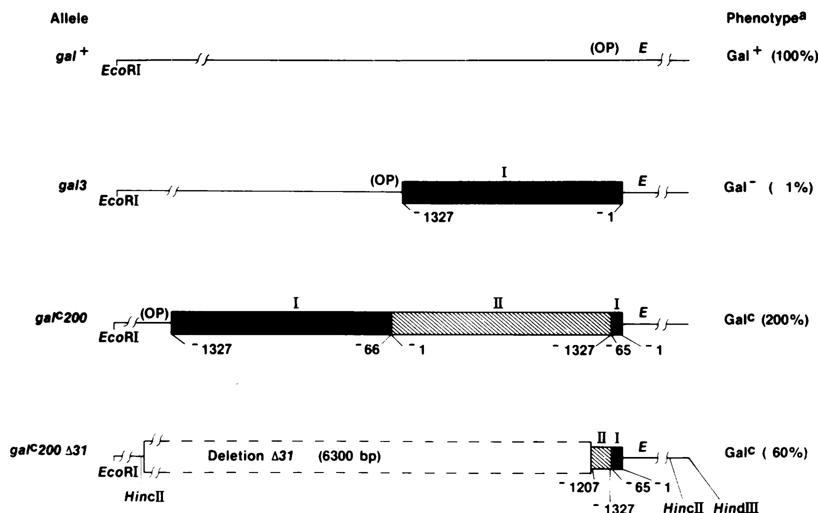

Figure 1. Structural and functional relationships among the alleles gal+, gal3, galc200, galc200 Δ31. (OP) and E designate the operator/ promoter region and the epimerase gene of the gal operon, respectively. The orientation of the IS2 DNA is designated as follows I: ▬ ;II: ▨ . In gal3 an IS2 in orientation I precedes the start of the galE message; position -1 represents the end of the IS2(I) DNA at the IS2(I)-galE junction while position -1327 represents the end at (OP)-IS2(I). In galc200 an IS2(II) is nested within the IS2 sequence. In galc200 Δ31 a deletion, indicated by dotted lines, has removed all of the IS2 DNA upstream from position -1207 of IS2(II). DNA sequence analyses of galc200 Δ31 (3) indicated that the IS2(II) has inserted between positions -65 and -66 of the IS2(I) sequence. The same sequence is presumed to occur in galc200 since galc200 Δ31 is derived from this allele. The positions of HindIII, HincII and EcoRI restriction sites referred to in the text are shown. [a]Numbers in parentheses indicate the % of gal expression relative to a fully induced Gal+.

the RNA also contained sequences from the transcription of the IS2(I) DNA upstream from the galE gene, we hybridized the isolated RNA to the r strand of λr32. (This λ DNA contains a copy of an IS2 sequence but does not contain the gal operon; the r strand would complement any gal-directed RNA from the 65 bp of IS2(I) adjacent to galE.) Any sequences that did not hybridize to the λr32 DNA, including the gal-specific sequences, were then eliminated by trimming the RNA-DNA hybrid with T_1 ribonuclease. A simplified fingerprint, containing a portion of the T_1 oligomers from the complete map, was obtained (Figure 3B). This analysis demonstrates that the RNA initiated upstream from the galE sequence.

The T_1 oligonucleotide products from both the complete and IS2-specific fingerprints were analyzed by treatment with pancreatic ribonuclease and/or

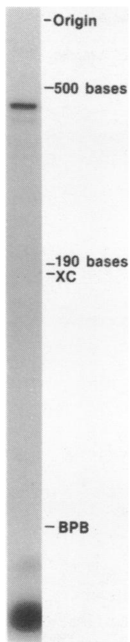

Figure 2. In vitro transcription products of the HincII DNA template from gal^{c}200 Δ31 after separation on a 5% polyacrylamide, 7 M urea gel. The in vitro transcription was performed as described in Methods. BPB and XC refer to the positions of the marker dyes bromophenol blue and xylene cyanol FF, respectively. The sizes designated on the gel indicate the mobilities of in vitro transcripts of known lengths. Minor bands observed on the gel account for less than 10% of the radioactivity incorporated into the major band. T_1 fingerprint analyses of these minor bands indicated that they represent shorter transcripts derived from the major band.

ribonuclease U_2. As expected, all the T_1 products present on the complete map, but missing from the IS2-specific map, correlated with transcription of the galE sequence present on the template (data not shown). The T_1 sequences observed in the IS2-specific fingerprint represented the products expected from the transcription of the IS2 DNA upstream from the IS2-galE junction. As shown in Figure 3C, the T_1 product T3, representing a portion of the expected T_1 product from positions -54 to -38 of IS2(I), was the farthest upstream T_1 oligomer observed. The next upstream potential product, a heptamer, did not appear on the IS2-specific map. Thus, T3 contains the 5'-end of the RNA and transcription starts within the IS2(I) DNA sequence, ACTATCACTTATTTAAG.

Several lines of evidence were used to determine the exact nucleotide start of the RNA. Secondary analyses of T3 revealed the presence of a polyphosphated moiety (Table 1). In addition, this characterization indicated that while the pancreatic RNase product ACp was obtained, the ribonuclease U_2 products Ap and CUAp were not observed. These results showed that the RNA must initiate within the sequence CTATC (positions -53 to -49 of the IS2(I) sequence). To identify which of these nucleotides represents a 5'-start, we digested the isolated RNA with nuclease P_1, an enzyme that yields

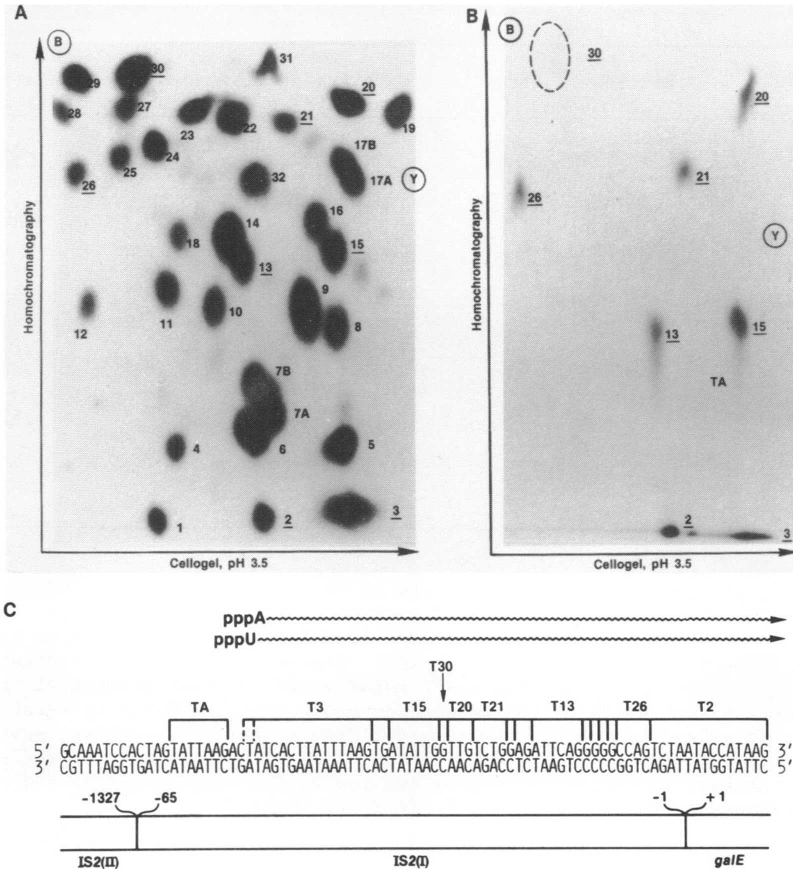

Figure 3. Characterization of the transcript from galc200 Δ31. (A and B) Two-dimensional fingerprints of ribonuclease T_1 oligonucleotide products derived from the isolated [α-^{32}P]UTP-labeled RNA shown in Figure 2 before (panel A) and after (panel B) hybridization to the r strand of λr32. The fingerprints were generated as described in the text. B denotes the position of the marker dye, xylene cyanol FF;Y denotes orange G. The underlined T_1 products represent those common to both maps. (C) The DNA sequence at the IS2-galE juncture of gal^{c}200 Δ31 is shown with the matching T_1 oligonucleotides common to both fingerprints (panels A and B) denoted above. These T_1 products were analyzed as described in the text and assigned as follows: T2: UCUAAUACCAUAAGp; T26: CCAGp[U]; T13: AUUCAGp; T21: UCUGp; T20: UUGp; T30: Gp[U]; T15: AUAUUGp; T3 :pppUAUCACUUAUUUAAGp + pppAUCACUUAUUUAAGp. (Brackets denote nearest neighbor.) TA denotes the predicted position of the T_1 product UAUUAAGp, representing positions -55 to -61 of the IS2(I) DNA, which is not observed on the map in Panel B. The arrows above the sequence indicate the two initiation sites (at positions -51 and -52) and the direction of transcription. The schematic diagram below the sequence shows the location of the sequence relative to the gal^{c}200 Δ31 DNA; the numbers refer to the IS2 sequence positions as given in Figure 1.

TABLE-1 Secondary Digestion Products from the T_1 Product T3

$[\alpha-^{32}P]$NTP Precursor	Pancreatic RNase Products	Relative Mobility[a]	RNase U_2 Products	Relative Mobility[a]
UTP	Up	2.2	Gp[U]	1.38
	ACp[U]	1.0	CUUAp	.22
	AUp	.89	UUUAp	.14
	AAGp[U]	.24	$ppp(Xp)_n$[b,c]	.03
	$ppp(Xp)_n$[b,c]	.07		
ATP	Up[A]	2.2	UCAp	.56
	Cp[A]	1.4	CUUAp	.24
	AAGp	.23	UUUAp	.15
	$ppp(Xp)_n$[b,c]	.06	$ppp(Xp)_n$[b,c]	.02
CTP	ACp	1.0	UCAp	.57
	AUp[C][c]	.93		
	$ppp(Xp)_n$[b,c]	.10		
GTP	AAGp	.24		

The T_1 oligonucleotide T3 (a mixture of the oligomers pppUAUCACUUAUUUAAGp and pppAUCACUUAUUUAAGp) was digested with pancreatic ribonuclease or ribonuclease U_2 as described in Methods and the products fractionated by electrophoresis at pH 3.5 on DE 81 paper. Nearest neighbors are indicated in brackets. Underscoring denotes a ^{32}P molar yield for that product >1 (as determined by intensity of the autoradiographic spot). [a]Mobility relative to that of xylene cyanol FF. These mobilities differ from earlier, published values (26,27) because of a change in the DE 81 paper. [b]Characterized as a 5'-triphosphate (oligo)nucleotide by its low R_b but exact sequence could not be determined by these analyses. [c] ^{32}P molar yield <1.

the nucleoside triphosphate pppN from the 5'terminal residue. As shown in Figure 4, analysis of the transcript by P_1 digestion demonstrated that transcription with either $[\alpha-^{32}P]$ATP or $[\alpha-^{32}P]$UTP labels the 5'-terminus but transcription with $[\alpha-^{32}P]$CTP does not. This analysis then eliminated the possible C starts at positions -49 and -53 and indicated that the A residue (-51) and one or both of the U residues (-52, -50) represent 5'-starts. To establish the 5'-penultimate residue, we digested the isolated RNA with ribonuclease T_2, an enzyme that yields pppNp from the 5'-end of the RNA. This nucleoside tetraphosphate is labeled by ^{32}P if either the first or the second residue of a transcript is derived from an $[\alpha-^{32}P]$NTP precursor. As seen in Figure 4D, no pppNp product was obtained after the T_2 digestion of $[\alpha-^{32}P]$CTP-labeled RNA. Two identical T_2 products, however, were derived from the RNA after transcription with either $[\alpha-^{32}P]$UTP or $[\alpha-^{32}P]$ATP. These products, identified as pppAp and pppUp, demonstrate

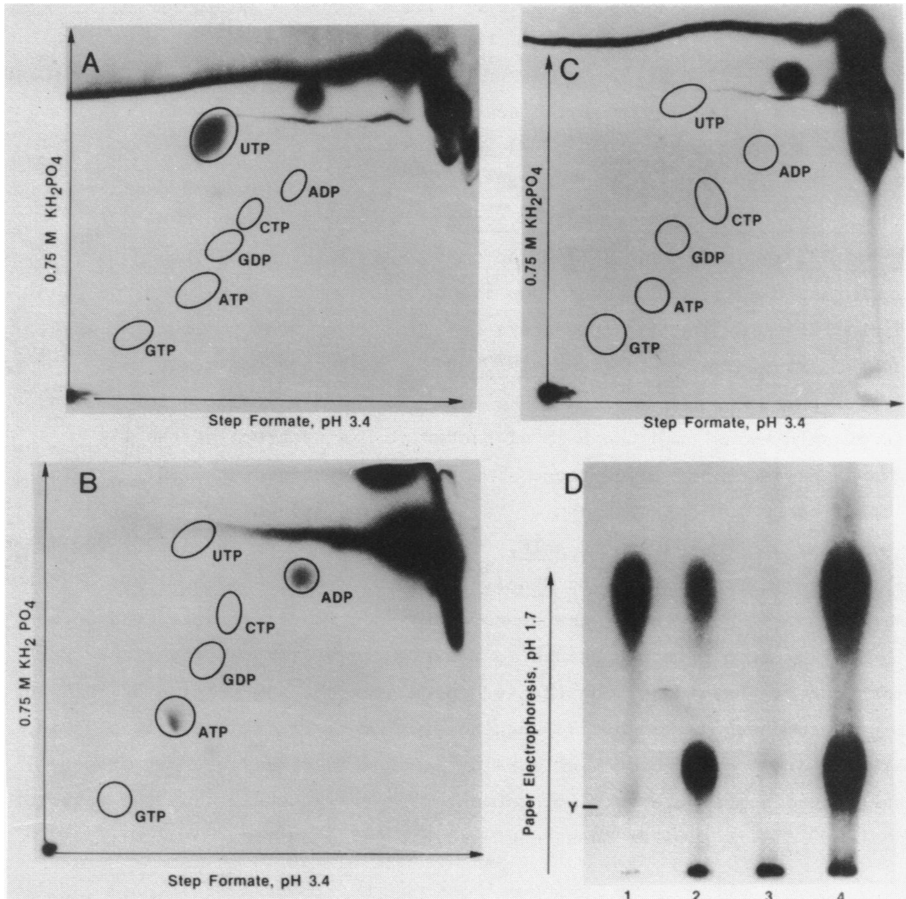

Figure 4. Digestion of ^{32}P-labeled transcript from galc200 Δ31 with nuclease P_1 and ribonuclease T_2. A–C) Autoradiographs of two-dimensional separations after P_1 digestion of the transcript labeled with [α-^{32}P]UTP (Panel A), [α-^{32}P]ATP (Panel B), and [α-^{32}P]CTP (Panel C). P_1 digestions and product fractionations were performed as described in Methods. The direction of development for the two different solvents is shown by the arrows; the position of marker nucleotides is indicated by circles. D) Autoradiograph of the products generated by T_2 digestion of the transcript labeled with [α-^{32}P]ATP (lane 2), [α-^{32}P]CTP (lane 3), or [α-^{32}P]UTP (lane 4) after electrophoresis on DE 81 paper at pH 1.7. The procedure for T_2 digestion is given in Methods. Lane 1 shows the position of [α-^{32}P]pppAp generated by the T_2 digestion of [α-^{32}P]ATP-labeled 6S RNA. Y denotes the position of the marker dye, orange G.

that the RNA is composed of two transcripts. One starts at position -51 (pppAUp...); the other, at position -52 (pppUAp...) (Figure 3C). Since the

intensities of the radioactive pppUp and pppAp moieties generated by T_2 digestion were similar (Figure 4D), there is not a strong preference for either start.

To ascertain whether these IS2 segments can also promote transcription *in vivo* we examined the level of tetracycline resistance conferred by pBRHE*gal*^{c}200 Δ31 and certain related plasmids (Figure 5). The plasmid pBRHE*gal*^{c}200 Δ31 contains a 6.6 kb *Eco*RI - *Hin*dIII fragment from *gal*^{c}200 Δ31 inserted into pBR322 and confers ampicillin-resistance to transformed bacteria. Although the resulting loss of DNA between the *Eco*RI and *Hin*dIII sites in pBR322 destroys the normal promoter for the *tet* gene, bacteria transformed with pBRHE*gal*^{c}200 Δ31 also become resistant to moderate levels of tetracycline. To localize the new promoter for *tet* in this plasmid we deleted the 6 kb of bacterial DNA upstream of the IS2 segments as described in Methods. The resulting plasmid, pAST202, retains 39 bp of bacterial DNA upstream of the 121 bp IS2(II) and 65 bp IS2(I) segments as well as 430 bp of *galE*. Thus the DNA insert in pAST202 is nearly identical to the 550 bp *Hin*cII fragment transcribed *in vitro*. This DNA region must contain the new promoter for *tet* since bacteria transformed with pAST202 also gain resistance to moderate levels of tetracycline. For comparison, a plasmid was constructed containing the same region of *galE* but the *gal* promoter instead of IS2 sequences. Transformation of this plasmid, pAST100, into a *galR*$^-$ bacterial strain (which derepresses the *gal* promoter) results in a comparable level of tetracycline resistance. Thus the promoter of *gal*^{c}200 Δ31 is similar in strength to the *gal* promoter.

DISCUSSION

In vitro transcription of a DNA fragment from *gal*^{c}200 Δ31 yields *gal*-directed RNA, initiating at two adjacent sites within the IS2 DNA, 51 and 52 bp upstream from the IS2-*galE* junction. These transcripts are observed using a purified *in vitro* transcription system, and the RNAs initiate specifically well within the interior of the DNA template. We have not observed any significant amounts of nonspecific initiation or transcription from the ends of the template fragment, two of the most common artefacts encountered with this system (41). Thus, we conclude that this transcription defines a bacterial promoter, designated P*gal*200Δ31, located within the ~ 40 bp upstream of the transcription initiation sites. As seen in Figure 6, the DNA sequence of P*gal*200Δ31 resembles the consensus sequence of other characterized bacterial promoters (32). The hexamer TAGTAT located at the -10 region of P*gal*200Δ31 is similar to the conserved sequence TATAAT.

Plasmid	EOP in the Presence of Tetracycline (μg/ml) 0	5	10	20	50
a) pBRHEgal^{c}200 Δ31	1.0	0.8	0.9	0.7	0[a]
b) pAST202 (ΔEcoRI-HpaI)	1.0	1.2	0.9	1.1	0[a]
c) pAST100	1.0	1.0	1.1	0.8	0[a]

EcoRI HpaI HpaI HindIII EcoRI
IS2 galE tet bla
HindIII
galE tet bla
EcoRI EcoRI
galopE tet bla
0 2 4 6 8 10 kb

Figure-5. Expression of the tet gene promoted from bacterial DNA. Linear structures of the plasmids tested are shown with the thin lines representing pBR322 vector DNA, thick lines for the bacterial DNA inserts, and the open box indicating the IS2 segments of gal^{c}200 Δ31. The plating efficiency (EOP) in the presence of different concentrations of tetracycline was determined for SA1171 transformants as described in Methods. The bacteria SA1171 without plasmid fails to grow in the presence of 5μg/ml tetracycline.

Similarly, 30 bp upstream from the initiation sites, the sequence TGGAAA is designated as the -35 region based on its homology with the expected sequence TTGACA and its distance (17 bp) from the -10 region. Neither the -10 nor -35 region of Pgal200Δ31 matches the consensus sequence exactly, however, and other portions of the gal^{c}200 Δ31 DNA template might be designated promoters based on their homologies with the consensus sequence.

-1327 -65
IS2(II) IS2(I) galE

ATGTCTGGAAATATAGGGGCAAATCCACTAGTATTAAGACTATCACTTAT• • •
-40 -30 -20 -10 -1 +1

TTGACA TATAAT

Figure-6. The sequence of the promoter, Pgal200Δ31. The DNA sequence surrounding the start of the transcript from gal^{c}200 Δ31 is shown. The location of the sequence relative to the gal^{c}200 Δ31 DNA is indicated by the schematic diagram showing the positions of the IS2(II) and IS2(I) DNA. The numbers above the diagram refer to the IS2 sequence positions as given in Figure 1. The numbers below the sequence correspond to the pppUp start of the transcript (+1) and the region of the promotor Pgal200Δ31 (-1 to -40). The DNA sequence of the conserved -10 (Pribnow box) and -35 regions are shown.

Thus, the DNA sequence of the gal^{c}200 Δ31 template is not sufficient to predict the transcription from Pgal200Δ31 observed in vitro. In particular, the transcription initiation site at position -52, representing a UTP start, might not be expected since few examples of transcription starts at pyrimidine residues are known (32). Thus, these in vitro transcription studies have been useful in defining the existence and position of a gal-directed promoter located within the IS2 sequence of gal^{c}200 Δ31.

The detection of Pgal200Δ31 in vitro suggests that this promoter might be responsible for the Gal^c phenotype of gal^{c}200Δ31. Although we have not proven this proposal by direct analysis of the in vivo RNA, our data is generally supportive. First, tetracycline resistance conferred by the plasmid pAST202 indicates that the IS2 segments composing Pgal200Δ31 can promote transcription of the tet gene in vivo. In addition, the level of expression from Pgal200Δ31 is comparable to that obtained from an induced gal promoter as judged by in vitro transcription (data not shown) or level of tetracycline resistance in vivo. This correlates well with the relative level of gal expression observed with the original strain with gal^{c}200Δ31 in the chromosome (22). Furthermore, the position of Pgal200Δ31 is also consistent with the view that this promoter directs gal transcription in gal^{c}200 Δ31. The allele, gal^{c}200, differs from its Gal$^-$ parent, gal3, by the insertion of an IS2 in orientation II, 65 bp from the IS2(I)-galE junction (Figure 1) (3). Since the DNA region defining Pgal200Δ31 lies upstream from the transcript starts at -51 and -52, this promoter overlaps the IS2 sequence change between gal^{c}200 and gal3. Thus, the creation of Pgal200Δ31 results from the structural difference between gal^{c}200 and its Gal$^-$parent.

The position of Pgal200Δ31 also defines this promoter as one derived from an altered IS2(I) sequence rather than a promoter intrinsic to the IS2(II) DNA. As seen in Figure 6, 15 bp of Pgal200Δ31, including the -10 (Pribnow box) region, are contributed by IS2(I) DNA. The remaining upstream sequence, containing the -35 region, stems from the IS2(II) insert. Based on this type of structure, we can make two predictions. First, given the phenotypes of gal^{c}200 and gal3, we would expect that in the gal3 template, the -10 region is not coupled with a -35 region adequate for promoter function. In fact, our in vitro transcription of DNA templates representing the IS2(I)-galE region of gal3 do not result in any gal-directed transcription within the IS2(I) sequence (33). Second, we would expect that other alterations besides the sequence change exhibited by gal^{c}200 could result in a -35 region needed for a functional promoter. We have shown this to be true for another Gal-constit-

utive revertant of gal3, called gal^c331 (33). In this revertant, 108 bp of IS2 derived DNA have been inserted at the same site as the IS2(II) insert of gal^c200 (3). Like gal^c200 Δ31, in vitro transcription of gal^c331 yields a gal-directed transcript initiating at positions -51 and -52 of the IS2(I) DNA. This RNA, which is identical to the gal-directed transcript of gal^c200 Δ31, occurs even though the gal^c331 allele provides a -35 region distinct from that of gal^c200. Thus, these results indicate that the IS2(I) present in gal3 contains the -10 portion of a functional promoter. Other Gal^c revertants might also arise by sequence changes that supply the needed DNA upstream from this region.

By transcribing a DNA template from the Gal^c revertant, gal^c200 Δ31, we have been able to investigate the transcription potential of the altered IS2(I) sequence as well as part of the IS2(II) DNA. Both of these sequences have been associated with the constitutive expression of genes adjacent to their insertions. This association has led to the speculation that functional promoters are contained within the IS2(II) or altered IS2(I) sequences. Our results demonstrate that, in gal^c200 Δ31, a promoter is created by the alteration of the IS2(I) DNA from the insertion of the IS2(II) sequence. We have not detected a promoter within the portion of the IS2(II) sequence present on the gal^c200Δ31 template. From these results, one might speculate that other cases of the constitutive expression of genes adjacent to IS2(II) (8,11,14) arise by the creation of a novel joint promoter at the site of the IS2(II) insertion. This proposal is supported by the fact that not every IS2(II) insertion has resulted in a constitutive phenotype. For example, IS2(II) insertions within the arg operon are known to be polar (12,34). However, since gal^c200 Δ31 contains only 121 bp of the IS2(II) DNA, it is clear that another promoter could exist in the portion of the IS2(II) sequence (or the upstream bacterial DNA) deleted by Δ31. Such a promoter, if it exists, should be present in gal^c200 but not gal^c200 Δ31. Certainly the higher level of gal expression in gal^c200 (200% Gal^+ level) as compared to gal^c200 Δ31 (60%) suggests that gal^c200 may indeed have a gal-directed promoter not present in gal^c200 Δ31. The existence of an active IS2(II) promoter is in fact supported by our in vitro transcription of other templates containing IS2 elements (33). These studies have demonstrated the presence of a weak IS2 promoter which transcribes in orientation II and is located in the region of the IS2(II) DNA deleted from gal^c200 by Δ31. Thus, either this promoter, or another as yet undetected promoter, might also contribute to the constitutive gal expression of gal^c200.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the generous gift of the plasmid pBRHEgal^{c}200 Δ31 from A. Ahmed and the technical assistance of K. Bidwell and J. Bear. We also thank C. Benyajati, M. Pearson, and R. Yuan for reviewing the manuscript and B. Traynor for her diligence in typing the manuscript. This work was supported by The National Cancer Institute (under contract No-C01-75380 with Litton Bionetics, Inc.). D.M. Hinton is an American Cancer Society Postdoctoral Fellow.

REFERENCES

1. Hirsch, H.-J., Starlinger, P. and Brachet, P. (1972) Molec. gen. Genet. 119, 191-206.
2. Fiandt, M., Szybalski, W. and Malamy, M.H. (1972) Molec. gen. Genet. 119, 223-231.
3. Ahmed, A., Bidwell, K. and Musso, R. (1980) in Cold Spring Harbor Symposia on Quantitative Biology, Vol. XLV, pp. 141-151, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
4. DNA Insertion Elements, Plasmids, and Episomes, Bukhari, A.I., Shapiro J.A., and Adhya, S.L. Eds., pp. 25-139, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
5. Starlinger, P. (1980) Plasmid 3, 241-259.
6. Calos, M.P. and Miller, J.H. (1980) Cell 20, 579-595.
7. Saedler, H. and Heiss, B. (1973) Molec. gen. Genet. 122, 267-277.
8. Saedler, H., Reif, H.J., Hu, S. and Davidson, N. (1974) Molec. gen. Genet. 132, 265-289.
9. Ahmed, A. and Scraba, D. (1975) Molec. gen. Genet. 136, 233-242.
10. Mosharrafa, E., Pilacinski, W., Zissler, J., Fiandt, M. and Szybalski, W. (1976) Molec. gen. Genet. 147, 103-109.
11. Pilacinski, W., Mosharrafa, E., Edmundson, R., Zissler, J., Fiandt, M. and Szybalski, W. (1977) Gene 2, 61-74.
12. Charlier, D., Crabeel, M., Palchaudhuri, S. Cunin, R., Boyer, A. and Glansdorff, N. (1978) Molec. gen. Genet. 161, 175-184.
13. deCrombrugghe, B., Adhya, S., Gottesman, M. and Pastan, I. (1973) Nature (New Biol.) 241, 260-264.
14. Walz, A., Ratzkin, B., and Carbon, J. (1978) Proc. Natl. Acad. Sci. USA 75, 6172-6176.
15. Ghosal, D., Sommer, H. and Saedler, H. (1979) Nucleic Acids Res. 6, 1111-1122.
16. Ghosal, D. and Saedler, H. (1977) Molec. gen. Genet. 158, 123-128.
17. Peterson, P.A., Ghosal, D., Sommer, H. and Saedler, H. (1979) Molec. gen. Genet. 173, 15-21.
18. Delius, H., Charlier, D. and Besemer, J. (1980) Molec. gen. Genet. 179, 391-397.
19. Besemer, J., Gortz, G. and Charlier, D. (1980) Nucleic Acids Res. 8, 5825-5833.
20. Ghosal, D, and Saedler, H. (1978) Nature 275, 611-617. 21.
21. Sommer, H., Cullum, J. and Saedler, H. (1979) Molec. gen. Genet. 175, 53-56.
22. Ahmed, A. and Johansen, E. (1975) Molec. gen. Genet. 142, 263-275.
23. Berg, D., Barrett, K. and Chamberlin, M. (1971) in Methods in Enzymology, Grossman, L. and Moldare, K. Eds., Vol 21, pp 506-519,

Academic Press, New York.
24. Nissley, S.P., Anderson, W.B., Gottesman, M.E., Perlman, R.L. and Pastan, I. (1971) J. Biol. Chem. 246, 4671-4678.
25. Lebowitz, P., Weissman, S.M. and Radding, C.M. (1971) J. Biol. Chem. 246, 5120-5139.
26. Brownlee, G.G. (1972) Determination of Sequences in RNA, American Elsevier Publishing Co., Inc., New York.
27. Musso, R.E., DiLauro, R. Adhya, S. and deCrombrugge, B. (1977) Cell 12, 847-854.
28. Bovre, K., Lozeron, H.A. and Szybalski, W. (1971) Methods Virol. 5, 271-292.
29. Fujimoto, M., Kuninaka, A. and Yoshino, H. (1974) Agric. Biol. Chem. 38, 1555-1561.
30. Johnson, L.D. and Lazzarini, R.A. (1977) Virology 77, 863-866.
31. Cashel, M., Lazzarini, R.A. and Kalbacher, B. (1969) J. Chromatogr. 40, 103-109.
32. Rosenberg, M. and Court, D. (1979) Annu. Rev. Genet. 13, 319-353.
33. Hinton, D. and Musso, R. manuscript in preparation.
34. Boyen, A., Charlier, D., Crabeel, M., Cunin, R., Palchaudhuri, S. and Glansdorff, N. (1978) Molec. gen. Genet. 161, 185-196.
35. Feiss, M., Adhya, S. and Court, D. (1972) Genetics 71, 189-206.
36. Wartell, R.M. and Reznikoff, W.S. (1980) Gene 9, 307-319.
37. Dagert, M. and Ehlich, S.D. (1979) Gene 6, 23-28.
38. Klein, R.D., Selsing, E. and Wells, R.D. (1980) Plasmid 3, 88-91.
39. Rodriguez, R.L., West, R.W. and Heyneker, H.L. (1979) Nucleic Acids Res. 6, 3267-3288.
40. deCrombrugghe, B., Mudryj, M., DiLauro, R. and Gottesman, M. (1979) Cell 18, 1145-1151.
41. Küpper, H., Contreras, R., Khorana, H.G. and Landy, A. (1976). In RNA Polymerase (Losick, R. and Chamberlin, M., eds.) Cold Spring Harbor, New York, pp. 473-484.