Volume 17 Number 10 1989 Nucleic Acids Research

# Pseudorevertants of a *lac* promoter mutation reveal overlapping nascent promoters

Russell Karls, Vincent Schulz, Stevan B.Jovanovich[1], Sheila Flynn[+], Alex Pak[§] and William S.Reznikoff*

Department of Biochemistry, College of Agricultural and Life Sciences and [1]McArdle Laboratory for Cancer Research, University of Wisconsin-Madison, Madison, WI 53706, USA

Received December 16, 1988; Revised and Accepted April 21, 1989

ABSTRACT

Four pseudorevertants of a -10 region lacP mutation were isolated. Three of these mutations were found to activate nascent promoters. These mutations were: a -2 G/C → A/T change (-2A) promoting transcription at position +11, a +1 A/T → T/A change (+1T) promoting transcription initiation at position +13, and a +10 C/G → A/T change (+10A) promoting transcription initiation at a complex series of positions. The fourth mutation [a -12 T/A → A/T change (-12A)] promotes transcription initiation at -1. The promoters activated by mutations -12A, -2A and +1T resembled the canonical sigma 70 promoter sequences. The +10A promoter activity is also dependent upon the sigma 70 holoenzyme but can not be readily assigned to a specific promoter sequence.

## INTRODUCTION

The first step in gene expression is the transcription initiation process. This is known to involve the specific interaction of RNA polymerase with DNA sequences termed promoters. For most genes in Escherichia coli the relevant RNA polymerase is the holoenzyme containing the sigma 70 subunit. Examination of over 263 promoters recognized by the sigma 70 holoenzyme has led to the identification of a consensus sequence involving two hexanucleotide pair sequences centered approximately 10 bp and 35 bp upstream of the transcription initiation start point (1,2). These sequences are indicated in Figure 1. Sequence analysis of lactose operon (lac) promoter mutations (also shown in Fig. 1) support the proposed importance of these so-called "canonical" sequences. For instance, mutations such as T743 (-11G) which decreases the similarity to the canonical sequence decreases lac promoter expression as expected (3). It is not known whether individual positions within the canonical sequences are recognized independently (that is the total promoter activity is a summation of the contributions from individual positions) or whether the recognition of these positions occurs through a more complex cooperative process.

Some promoter regions contain more than one overlapping promoter sequence. Examples can be found in the galactose operon regulatory elements (4,5), the

Tn5/IS50 transposition function regulatory elements (6) and the Tn10 tetracycline resistance regulatory elements (7,8). In the case of the lac operon, there exist overlapping nascent promoters (9-13). These are sequences which can act as promoters as a result of single bp changes or which act in vitro but not in vivo. It is not clear whether these nascent promoter sequences exist by chance, whether they are evolutionary remnants or whether they play a physiologically important role.

This paper and an independent report by Rothmel and LeClerc (14) describe the isolation and characterization of second site revertants of lacP mutations (this manuscript studies revertants of the -11G mutation, while the Rothmel and LeClerc report studies revertants of three different mutations including a -11C mutation). Our studies were undertaken in order to study the canonical promoter sequence. We wished to determine whether compensatory mutations could be isolated in the standard lacP sequence which relieved the -11 defect. With the possible exception of one complex case, no such compensatory mutations were found. Alternatively, the canonical promoter sequence could be studied by examining the sequences and properties of any additional nascent promoter sequences which might be revealed by some of the revertants. Three such nascent promoters (one of which has been previously studied, 9,10) appear to qualitatively support the proposed canonical promoter sequence model. However, for the fourth mutation, it was not possible to assign an appropriate promoter sequence. This fourth mutation had been isolated in a previous set of experiments (15), but the previously proposed explanation for its lac activation is not consistent with the data in this and the Rothmel and LeClerc communication (14).

## MATERIALS AND METHODS

(A) Enzymes, media and reagents.

T4 DNA polymerase, T4 DNA ligase, T4 polynucleotide kinase, restriction endonucleases PstI, ScaI, EcoRI, BglII, BamHI, PvuII and AluI were purchased from New England Biolabs. Actinomycin D and glycogen were obtained from Boehringer Mannheim Biochemicals. The indicator dye 5-bromo-4-cloro-3-indolyl-$\beta$-D-galactopyranoside (XG) and the inducer isopropyl-$\beta$-D-thio-galactopyranoside (IPTG) were purchased from Bachem, Inc. AMV reverse transcriptase was purchased from Life Sciences, Inc. Ampicillin, kanamycin, tetracycline, o-nitrophenyl-$\beta$-D-galactopyranoside (ONPG), p-toluenesulfonyl-fluoride hydroxylamine and pancreatic RNase A were purchased from Sigma. The 1212 oligonucleotide primer was obtained from New England Biolabs and from the University of Wisconsin Biotechnology Center. $^{32}$P labeled ribonucleoside tri-

phosphates were from Amersham. All other ribonucleoside triphosphates and all deoxy and dideoxyribonucleoside triphosphates were from Pharmacia.

M9 medium and LB were prepared according to Miller (16). Thiamine (4 μg/ml), glucose (0.1%) or glycerol (0.1%) were added where indicated. TYE agar contains 10 g of tryptone (Difco Laboratories), 10 g of NaCl, 5 g of yeast extract (Difco) and 15 g Bacto-Agar (Difco) per liter. Ampicillin (100 μg/ml), kanamycin (40 μg/ml), tetracycline (15 μg/ml), XG (40 μg/ml), and IPTG (0.1 mM) were added where needed.

(B) Bacterial strains, plasmids and lac mutations.

The bacterial strains used in this study are listed in Table 1a. RZ6520 was construced by P1 transduction according to Miller (16) using a phage lysate kindly provided by the laboratory of Carol Gross. Tetracycline resistance conferred by the rpoD linked Tn10 was used as a selection for the appropriate candidate transductants.

Plasmids used in this study are listed in Table 1b. Plasmids with an EcoRI site located in the lac promoter immediately adjacent to the start site were constructed by the following steps. The -1 to +5 lacP sequence is GAATTG, which can be modified into an EcoRI site by a G→C change at +5. This was accomplished by site directed mutagenesis (17) of two mp8 varients mRZ352 -T743 (mRZ352 -11G) and mRZ361 (18) to generate mRZ352 -11G+5C and mRZ361 +5C. The lacP region of mRZ352 -11G+5C was transferred to pGEM2 (Promega Biotec) by cutting both DNAs with PstI and ScaI and ligating the 352 -11G+5C promoter fragment with the vector DNA. This construct was called pRZ3264. The lacP region of mRZ352 -11G+5C was transferred to pUC119 (19) by a similar approach using PuvII digestion. The resulting plasmid was called pRZ3204. A plasmid lacking lacP sequences upstream of the -1+5 EcoRI site was generated by cloning the EcoRI-BglII fragment from mRZ361+5C into pGEM1 (Promega Biotec) cut with the EcoRI and BamHI to give plasmid pRZ3256.

We use a standard nomenclature in referring to all but three lac mutations. Each mutation is named according to its position relative to the wild type lacP start site and the base change on the non-template strand. Thus T743 is designated -11G. The three exceptions to this nomenclature are: the promoter mutation UV5, the 27 bp insert lacZ multiple cloning site derived from mp8 called 361 (18), and the 18 bp lacZ insert derived from 361 designated 352 (18).

(C) Genetic techniques.

1 - Transfer of mutations from λplac5 to plasmids.

The λplac5 genome is small enough so that λplac5 with one of various pRZ plasmids incorporated into its genome can be packaged. Presumably the inser-

tion of the plasmid into the λplac5 genome occurs through homologous recombination within the shared lac sequences, and thus the excision of the plasmid subsequent to introduction into a host cell has a high probability of exchanging lacP sequences which were originally on λplac5 for those on the original pRZ plasmid. The following technique was used. CSH26 lysogens of the relevant λplac5 derivative were transformed with pRZ3204 or some similar plasmid containing the lacP mutation -11G selecting for ampicillin resistance. The transformants were grown at 30°C in LB to an $OD_{550}$ of approximately 0.4 with

Table 1a Bacterial Strains

| Strain | Genotype | Source |
|---|---|---|
| DH5α | F⁻ endA1, hsdR17, supE44, thi, recA1, gyrA96, relA1, Δ(lac)$_{U169}$, φ80dlacZM15, endA1 | Bethesda Research Laboratories |
| RZ201 | F⁻ Δ(lac-proAB)$_{X111}$, ara, thi, rpsL | Johnson et al. (43) |
| RZ203 | F⁻ Δ(lac-proAB)$_{X111}$, ara, thi, rspL, Δcya | Munson et al. (18) |
| RZ6523 | F' traD36, lacI$^{Q}$Z$_{M15}$, proAB$^{+}$/ Δ(lac-proAB)$_{X111}$, supE, thi, zaf13::Tn10, mutD5 | P1 transduction of Tn10 linked mutD into JM101 |
| RZ6520 | RZ201, rpoD800, zgh3075:Tn10 | P1 transduction of Tn10 linked rpoD800 into RZ203 |
| ER1458 | F⁻ Δ(lac)$_{U169}$, Δ(lon), araD139, rspL, supF, mcrA, mcrB1, hsdR2, trpC22::Tn10, serB28 | New England Biolabs |
| E9003 | F' Δ(lacO-Z)$_{145}$, proAB$^{+}$/ Δ(lac-proAB)$_{X111}$, thi | Munson et al. (18) |
| JV554 | F⁻ Δ(lac)$_{74}$, galK, phoA20, phoR, trp, rspL, relA, nadA::Tn10 | Jovanovich et al. (30) |

Table 1b. Plasmids

| Plasmid | Salient Characteristics | Source |
|---|---|---|
| pGEM1 | pUC derived plasmid containing convergent T7 and SP6 promoters surrounding a polylinker sequence | Promega Biotec |
| pGEM2 | Similar to pGEM1 but with a different polylinker sequence | Promega Biotec |
| pUC119 | pUC19 derived plasmid containing an M13 origin of replication | Vieira and Messing (19) |
| pRZ3204 | pUC119 derivative containing the -11G+5C (352) *lac* sequence | this study |
| pRZ3207 | pUC119 derivative containing the wild type *lac* sequence | this study |
| pRZ3256 | pGEM1 derivative containing a *lac* $\alpha$ complementing sequence without a promoter | this study |
| pRZ3264 | pGEM2 derivative containing the -11G+5C (352) *lac* sequence | this study |
| pRZ6507 | pUC119 derivative containing the -11G (361) *lac* sequence | this study |

vigorous aeration, shifted to 42°C for 15 min, and then incubated an additional 4 h at 37°C. $CHCl_3$ was then added to induce lysis. Aliquots of the supernatant were mixed with CSH26, plated on TYE ampicillin agar and incubated at 30°C overnight. The ampicillin resistant transductants were harvested, grown up overnight at 30°C, plasmid DNA was isolated from the cells and used to transform DH5$\alpha$ cells selecting for ampicillin resistance on TYE-ampicillin-XG-IPTG agar. Blue colonies were picked, purified, grown up and the plasmid DNA isolated for *lacP* sequence analysis. In cases in which the recombinant would be $Lac^-$, multiple individual colonies were picked and the correct recombinants were chosen by sequence analysis.

2 - Transfer of mutations from plasmids to λplac5.

Mutations were transferred from plasmids to the λplac5 chromosome by a procedure analogous to that described above and in Munson et al. (18). A λplac5 -11G lysogen of CSH26 was transformed with the relevant pRZ plasmid, the cells were grown and a phage lysate was formed as described above. Aliquots of this lysate were mixed with CSH26 and ampicillin resistant transductants were isolated. The ampicillin resistant colonies were pooled, grown overnight, subcultured, and a lysate was generated as described above. The lysate was titered using ER1458 host cells on TYE-XG agar, and blue plaques were picked after overnight incubation. The phage from these plaques were grown up and the lacP region was subjected to DNA sequence analysis.

3 - Restriction fragment directed mutagenesis.

In some cases (such as the placement of the +10A mutation in an otherwise wild type lacP sequence) mutations were incorporated into plasmids through a fragment directed mutagenesis procedure. Single stranded pRZ6507 or pRZ3207 DNA was used as a template and a denatured restriction fragment (either a purified PvuII fragment from a plasmid or a mixture of AluI fragments generated from ϕ80plac DNA) was used as the primer. Following the procedure of Munson et al. (18), the annealed primer was extended with T4 DNA polymerase and covalently closed circles were formed with T4 DNA ligase. The DNA molecules were transformed into DH5α cells, selecting for ampicillin resistance by plating the cells on TYE-ampicillin-XG-IPTG agar. The appropriately colored colonies were picked after overnight incubation, and the lacP region of the candidate plasmids was sequenced.

4 - Lysogen formation.

Lysogens were formed by mixing λplac5 derivatives with mid log phase CSH26, RZ203 or E9003 cells at a multiplicity of infection of from 0.1 to 1.0, incubating at 37°C for 15 min, pelleting the cells and resuspending in the same volume of LB and plating out aliquots on TYE-XG-IPTG agar followed by incubation at 30°C overnight. For RZ203 cells, the bacteria were grown for several hours in LB containing 2.5 mg/ml of cAMP prior to mixing with the phage. Blue colonies (extremely light blue colonies for λplac5 -11G) were picked and tested for monolysogeny by the ter excision test (20).

(D) B-galactosidase assays.

B-galactosidase assays were performed as described in Miller (16) except as modified for the coupled transcription-translation assays described later.

(E) Mutagenesis protocols.

The candidate pseudorevertants of -11G were isolated by three different protocols.

pRZ3264 was mutagenized with hydroxylamine. 10 micrograms of plasmid DNA was incubated in 200 microliters of 1 M hydroxylamine for one h at 65°C. The DNA was dialyzed, ethanol precipitated, and digested with EcoRI. The lac promoter fragment was ligated to pRZ3256 which was also cut with EcoRI. This DNA was transformed into DH5α, and plated on TYE-ampicillin-XG-IPTG agar. The lacP region of plasmid DNA from blue colonies was sequenced. About 1% of the colonies were blue.

pRZ3204 was mutagenized by growth in a bacterial mutator strain, RZ6523 (JM101 mutD). pRZ3204 was grown on RZ6523 for 16 h in Luria broth supplemented with 0.5 μg/ml thymidine. Plasmid DNA was extracted, transformed into DH5α and plated on TYE-ampicillin-XG-IPTG agar. Approximately 0.01% of the colonies were blue.

In the final protocol, spontaneous $Lac^+$ "revertants" of -11G were isolated by plating saturated, washed cultures of E9003::λplac5 -11G on lactose minimal agar and incubating for 3 to 4 days at 30°C. The $Lac^+$ colonies were picked and purified. Lysates of the λplac5 -11G revertants were formed and used to lysogenize fresh E9003. These lysogens were tested for their lac phenotype, and all were $Lac^+$. The candidate lacP mutations were transferred to pRZ plasmids as described above and sequenced.

(F) DNA and RNA isolation.

Plasmid DNA was isolated by the alkaline lysis procedure as described in Maniatis et al. (21).

Lambda DNA used for sequence analysis was isolated from lysates prepared as described (21). DNA was extracted from 50 ml of phage lysate as described (22), with the following modifications: The polyethylene glycol (PEG) phage pellet was resuspended in 0.5 ml SM buffer (10 mM Tris, pH 7.9, 10 mM $MgSO_4$) containing 8% PEG, + 0.5 M NaCl, transferred to an eppendorf tube, and precipitated a second time, followed by removal of all traces of supernatant. This pellet was resuspended in 0.5 ml SM buffer and extracted with an equal volume of chloroform. The aqueous phase was removed and treated with RNAse (10 μg/mg final concentration) for 15 min at 37°C. The phage DNA was then extracted with an equal volume hot phenol twice, phenol:chloroform once, and chloroform once. DNA was precipitated by adding 1/2 volume 7.5 M $NH_4OAc$ and 2 volumes cold ethanol, incubating on ice for 15 min, then centrifuging for 15 min at 4°C. The pellet was washed with 70% ethanol, dried, and resuspended in 100 μl TE buffer (10 mM Tris-Cl, 0.1 mM EDTA, pH 7.9).

Lambda DNA used in the coupled transcription-translation system was isolated by the technique of Gardner and Reznikoff (23). DNA was dialyzed versus 10 mM Tris-acetate, 0.1 mM EDTA, precipitated with ethanol in the presence of

2.5 M $NH_4OAc$, washed with 70% ethanol, dried, and resuspended in the above buffer.

Plasmid encoded RNAs were prepared as follows: Cells containing the plasmid of interest were grown to an $OD_{550}$ = 0.4 to 0.8 in 4 ml LB medium. 1 ml culture was transferred to an eppendorf tube. Chloramphenicol was added to 30 μg/ml and allowed to sit for 1 min, followed by centrifuging for 15 sec. The pellet was resuspended in 200 μl STET buffer (21). An equal volume of phenol:chloroform, pH 7.5, was added. Samples were boiled for 40 sec with caps open, then microfuged for 3 min. The aqueous phase was extracted with an equal volume of chloroform and precipitated by adding 1/2 volume 7.5 M $NH_4OAc$ and 2 volumes ethanol. After incubating on ice for 10 min, RNA was pelleted by centrifuging for 10 min at 4°C, washed with absolute ethanol, and resuspended in 50 μl $H_2O$.

(G) In vitro transcription assay.

"Run off" transcription experiments were performed as described in (15) with the following modifications: The template was a BstNI restriction fragment containing the lac promoter region, the RNA polymerase sigma 70 holoenzyme preparation was provided by R. R. Burgess, the $^{32}P$ labeled ribonucleoside triphosphate was rCTP, the RNA polymerase/template molar ratio was 5, the reactions were performed in the presence of 10% glycerol, and the reaction products were analyzed on a 8% denaturing sequencing gel.

(H) RNA primer extension and DNA sequencing.

RNA primer extensions were carried out as described (24) using the 1212 primer, except that the hybridization step was done at 45°C for 15 min, and primer extension incubation was done at 45°C with actinomycin D present at 100 μg/ml (25). DNA sequencing was done as described (24), except that plasmid DNA prepared as described (26) was used as a template; 12.5 μM ddNTP's were used instead of 100 μM ddNTP's; and hybridization and sequencing reactions were carried out at 45°C, not 37°C. Lambda DNA was denatured and sequenced as described above for plasmid sequencing.

(I) Coupled transcription-translation assays.

The S-30 extract was prepared from the E. coli strain JV554 according to a modification of the method of Zubay (27) (28-30). The essential modifications were: The preincubation step was performed at room temperature; 50 μg of p-toluene sulfonyl fluoride per ml was included in the growth medium and in all buffers; and four changes of dialysis buffer were performed at 90 min intervals. The final protein concentration was 20.2 mg/ml as determined by the method of Bradford (31).

The S-30 transcription-translation reactions were performed in a final volume of 50 μl and contained 35 mM TrisOAc, pH 8.0; 120 mM K-glutamate; 27 mM $NH_4OAc$; 2 mM dithiothreitol; 12.2 mM $Mg(OAc)_2$; 0.5 mM of each of the 20 amino acids; 2 mM ATP; 0.5 mM each CTP, UTP, and GTP; 20 mM phosphoenolpyruvate; 1 mg/ml tRNA; 35 mg/ml polyethylene glycol 8000; and 20 μg/ml folinic acid. The appropriate λ DNA in 10 μl, 319 μg of S-30 protein, prepared from JV554, in 15 μl, and 5 μl of monoclonal antibody were added and the mix preincubated on ice for 30 min. The reaction was started by addition of DNA template and incubated at 37°C in a rapidly shaking water bath for 70 min. The reactions were stopped by placing the test tubes on ice.

β-galactosidase assays were performed by adding 0.35 ml of Z-buffer and 0.1 ml of 4 mg/ml o-nitrophenyl-β-D-galactopyranoside to 50 μl of the S-30 reaction. The reaction was incubated at 30°C until a yellow color developed, and the assay stopped by addition of 0.25 ml of 1 M $Na_2CO_3$. After 10 min on ice, the sample was spun for 5 min in a microfuge, and the optical density of the supernatant was measured at 420 nm. Activity units are expressed as described in Miller (16), omitting the correction for light scattering caused by cellular debris.

Purified monoclonal antibodies were prepared as described (30,32) according to the method of Fazekas de St. Goth and Scheidegger (33). Sigma 70 was prepared from an overproducing strain as described (34).

## RESULTS

### Isolation and Sequence Analysis of lacP -11G Revertants.

The lacP -11G reduces lac expression to such a level (see Table 2) that when placed cis to an intact lacZ gene it results in host cells being unable to grow on lactose as the sole carbon source. In addition, when the lacP -11G mutation is carried on pUC type plasmids, lacZ peptide synthesis is so low that DH5α cells containing such plasmids give rise to white colonies on XG containing media. Both of these properties were used to isolate revertants of lacP -11G.

Plasmids which contained candidate revertants of lacP -11G were purified and subjected to dideoxy sequence analysis. Some plasmids contained the wild type lacP sequence (in these cases a precise reversion of -11G had occurred). The other plasmids contained the original -11G mutation plus a second change. The location and type of these second site mutations are displayed in Figure 1. These mutations have been named according to their position in the stan-

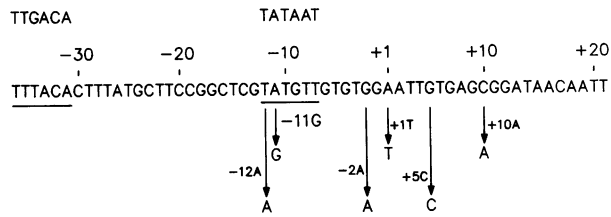


Fig. 1. lac Promoter Sequence. The lac promoter sequence is presented along with the sigma 70 promoter canonical sequences (TTGACA and TATAAT), the lacP version of these sequences (underlined), and the lacP mutations described in this study. The lacP mutation UV5 (not shown) converts the lacP -10 region to the canonical sequence.

dard lac promoter sequence as follows: -12A (a change to A 12 bp upstream of the $lacP^+$ transcription start site), -2A, +1T and +10A. The spontaneous mutations were -2A and +10A, the hydroxylamine induced mutations were -2A (a repeat) and +1T, and the mutD induced mutation was -12A. It should be noted that the different mutation isolation protocols resulted in the mutations being associated with different lac genetic backgrounds. For instance, the spontaneous mutations were isolated in a -11G lac promoter sequence which was otherwise wild type (these mutants are thus designated -11G-2A and -11G+10A). The mutD mutagenesis procedure resulted in a mutation being associated with -11G, +5C and the lacZ insert 352; therefore, this mutant is designated -12A-11G+5C (352). The 2 mutations isolated following hydroxylamine mutagenesis (-2A and +1T) were associated with -11G, +5C and the lacZ insert 361 and are so indicated. Mutations +1T and +10A were previously isolated as mutations lacP115 and lacP111/112 which partially relieved lac expression dependence upon the CAP-cAMP positive regulatory system in an otherwise $lacP^+$ sequence (9,35). Mutations -2A, +1T and +10A were also isolated by Rothmel and LeClerc (14) using a scheme similar to the one presented here.

Analysis of Transcript Start Points

The 5' end of the lac mRNA's programmed by the different mutant promoters in vivo was studied through the use of the primer extension method. Typical results are shown in Figure 2 and are schematically presented in Figure 3. The major primer extension products for the lac promoter mutants are as follows. The lacUV5 promoter primer extension products yields a band corresponding to a start site at +1 (the expected position) and a less prominent band corresponding to +20. This latter product was found for all lac promoters except for +10A and the nul lac promoter mutant -11G (T743). We be-

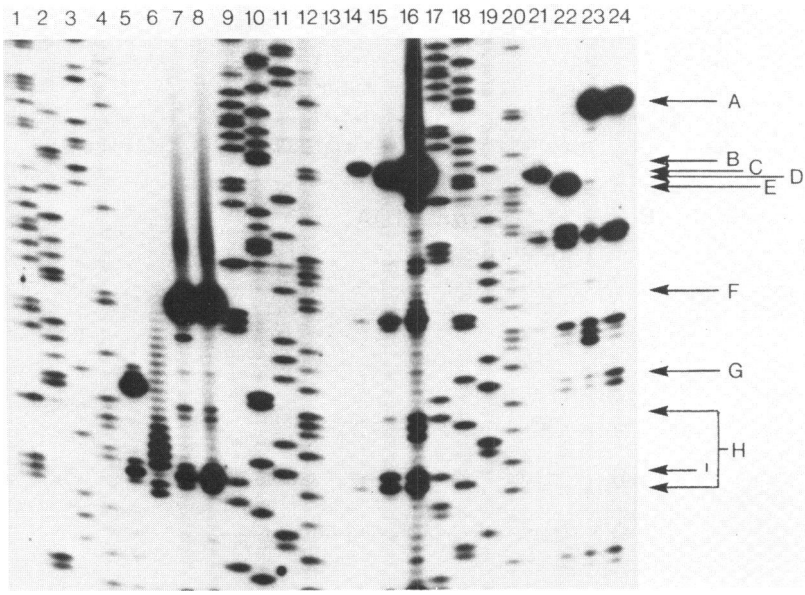


Fig. 2. Primer Extension Analysis of lac Promoter Mutations. The in vivo start points of lac transcripts programmed by lac promoter mutations were examined by primer extension analysis as described in Materials and Methods. Lanes 1,2,3,4 is a sequencing ladder of the lacP region from a wild type lacZ plasmid; lanes 9,10,11,12 is a sequencing ladder of the lacP region from a 352 (18 bp insert) lacZ plasmid. Lanes 17,18,19 and 20 is a sequencing ladder of the lacP region from a 361 (27 bp insert) lacZ plasmid. The primer extension results are as follows:

lane 5: lacP -11G-2A with an apparent start site at +11(G) and a secondary band at +20 (I).
lane 6: lacP -11G+10A with a collection of start sites peaking at +15 to +19, +21, +22 (H).
lane 7: Wild type lacP with a start site at +1 (F) and a secondary band at +20 (I).
lane 8: lacP +5C with a +1 start site (F) and a secondary band at +20 (I).
lane 13: lacP -11G+5C (352) with no apparent start sites.
lane 14: lacP -12A-11G+5C (352) with a -1 start site (B).
lane 15: Wild type lacP (352) with a +1 start site (C).
lane 16: lacPUV5 (352) with a +1 start site (C).
lane 21: lacP -11G-2A+5C (361) with a +11 start site (D).
lane 22: lacP -11G+1T+5C (361) with a +13 start site (E).
lane 23: Wild type lacP (361) with a +1 start site (A).
lane 24: Wild type lacP +5C (361) with a +1 start site.

lieve that the +20 product is an artifact of either mRNA degradation or the primer extension reaction.

The lacP mutants -12A, -2A and +1T yield primer extension products corresponding to starts at -1, +11 and +13, respectively. The lacP mutant +10A

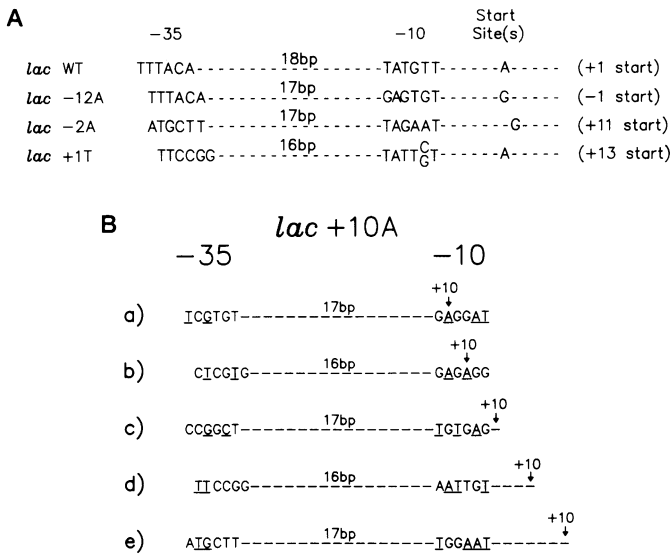


Fig. 3. A. Proposed -12A, -2A and +1T Promoters. The proposed -35 sequence and -10 sequence associated with each promoter activated by the indicated mutation are presented in comparison to the wild type lac promoter.
B. Possible sequence assignments for the +10A promoter. Five different possible +10A promoter sequences are presented with the location of the +10A change indicated in each case. The positions matching the canonical sigma 70 promoter sequence are underlined.

appears to program an unusual distribution of starts with minor bands extending from +2 to +14, bands of increasing intensity from +15 to +19, a break at +20, and bands at +21 and +22. Further analysis of the +10A start points will be presented subsequently.

Physiological Characterization

In order to characterize the expression levels of the mutant promoters in vivo and to test their sensitivity to catabolite repression, λplac5 phage carrying these mutations were used to make lysogens of the lac deletion strains RZ201 and RZ203. Monocopy lysogens of both strains were grown up and assayed for β-galactosidase. The results are complicated by two variables. Different mutations were isolated in different lacZ backgrounds; e.g., the wild type sequence, a sequence with a 27 bp insertion designated 361 (found in pRZ3256), and a sequence with a 18 bp insertion designated 352 (found in pRZ3264). Since these inserts are known to affect the levels of translation initiation, all mutants were compared to $lacP^+$ constructs that had the relevant lacZ

Table 2

| lacP genotype carried on λplac5 | transcript start point | β-galactosidase activity RZ201 | RZ203(Δcya) | +cAMP ΔcAMP ratio |
|---|---|---|---|---|
| -11G[T743] | none | .26 | N.D. | |
| Wild Type | +1 | 2200 | 89 | 25 |
| -11G-2A | +11 | 720 | 120 | 6 |
| -11G+10A | + 15 to +19 + 21, +22 | 430 | 320 | 1.4 |
| Wild Type (352) | +1 | 1320 | 51 | 26 |
| +5C (352) | +1 | 1330 | 61 | 22 |
| -12A-11G+5C (352) | -1 | 310 | 1.5 | 200 |
| +5C (361) | +1 | 510 | 9.4 | 54 |
| -11G-2A+5C (361) | +11 | 62 | 18 | 3.4 |
| -11G+1T+5C (361) | +13 | 450 | 460 | 1.0 |

λplac5 derivatives carrying the indicated lacP genotypes were used to generate monolysogens of RZ201 and RZ203. These lysogens were assayed for β-galactosidase synthesis following growth in M9-glycerol-casamino acid media.

sequences. Secondly, since the different promoters program the synthesis of mRNAs with different 5′ ends, their messages are likely to have different half lives or translation initiation frequencies. In fact, it is known from other studies (N. de la Cruz, personal communication) that the translation initiation inhibition properties of lacZ 361 mRNA are accentuated for transcripts lacking bases at the 5′ end of the +1 message. We have no control for this problem in comparing different mutants, although it should not complicate the analysis of each mutant's sensitivity to CAP-cAMP.

The β-galactosidase assay results are displayed in Table 2. The levels of lacZ expression in the absence of CAP-cAMP (in the RZ203 strain) are presumed to be a measure of RNA polymerase-promoter interactions in the absence of positive effectors. These values have not been corrected for any effects of the differing 5′ ends. Nonetheless, they suggest the following hierarchy in promoter activities: +1T > +10A > -2A > wild type > -12A. The assays also show that the various promoters show different levels of CAP-cAMP stimulation. The

mutant -12A is stimulated over 200-fold, wild type shows 20 to 54-fold stimulation, -2A manifests a 2.7 to 5-fold stimulation, and +1T and +10A are not stimulated.

Sigma 70 Dependence of Promoter Activity

Figure 3A presents a schematic representation for 3 of the lac promoters examined in this report using the location of the proposed mRNA start points and the known canonical sequence requirements for sigma 70 holoenzyme transcription initiation to describe the probable sequences involved. Promoters +1T and -2A program higher levels of β-galactosidase than the wild type lac promoter, yet their sequences are less similar to the proposed canonical sigma 70 holoenzyme promoter sequence (see Mulligan et al. (36) for a quantitative analysis of the +1T (P115) promoter in this regard). The promoter activity associated with the +10A mutation could not be assigned to an obvious sigma 70 dependent promoter sequence (see Figure 3B for 5 different possibilities). Thus, we have examined whether the expression programmed by these promoters is in fact sigma 70 dependent. This was accomplished by both in vivo and in vitro experiments.

The in vivo sigma 70 dependence of the promoter expression was tested by studying whether the RNAs in question were synthesized in a rpoD800 strain under non permissive conditions. The plasmids containing the promoters were transformed into RZ6520, the cells were grown at 30°C to an $OD_{550}$ of approximately 0.5, and then half of each culture was shifted to 42°C. After 1 h of additional incubation, RNAs were extracted from both 30°C and 42°C cultures and were subjected to a lac primer extension analysis as described earlier. The results for mutations +10A, -12A and UV5 are displayed in Figure 4. The promoter specific transcripts were present in all 30°C extracts but absent from all 42°C extracts.

The sigma 70 dependence of the promoters was also tested in a coupled transcription-translation in vitro system. A λplac5 DNA template carrying each promoter mutation was added to two types of S-30 extract. The first extract contained differing amounts of the 3D3 anti sigma 70 monoclonal antibody preparation (the anti-$\sigma^{70}$ monoclonal antibody 2D4 was also used in a separate experiment with similar results [data not shown]). The second extract containing differing amounts of purified sigma 70 (the S-30 extracts are known to be sigma 70 limited, 30). The S-30 extracts were incubated for 70 min and then were assayed for β-galactosidase. The results for lacP +10A (which are qualitatively identical to that for -12A, -2A, +1T and wild type promoters) are shown in Figure 5. The parental lacP -11G template programs

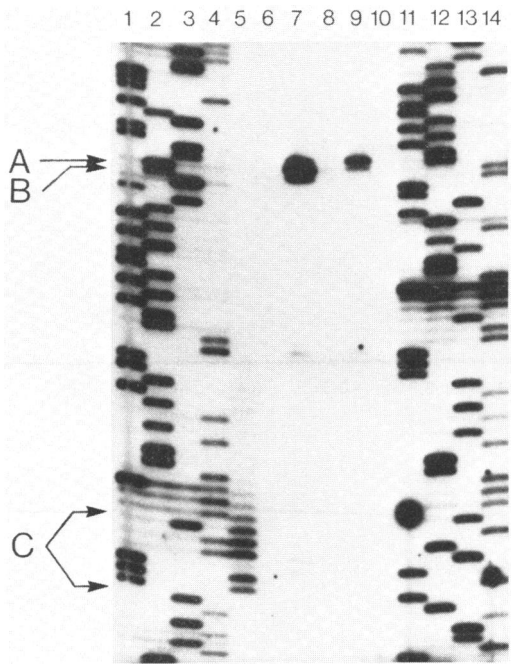


Fig. 4. In vivo rpoD Dependence of Mutant lac Promoter Expression. The primer extension technique was used to analyze mutant lac promoter expression under conditions in which the sigma 70 subunit was inactive. RZ6520 (rpoD800) cells containing lacP plasmids were grown at 30°C (permissive) or 42°C (non-permissive) for 1 h prior to RNA extraction. Lanes 1-4 and 11-14 are sequencing ladders of lacP -11G+10A and lacP -12A-11G+5C (352), respectively. Lanes 5 and 6 analyze lacP -11G+10A expression at 30°C and 42°C (C indicates transcripts). Lanes 7 and 8 analyze lacPUV5 (352) expression at 30°C and 42°C (B indicates transcript). Lanes 9 and 10 analyze lac -12 -11+5 (352) at 30°C and 42°C (A indicates transcript).

the synthesis of ~ 100-fold lower levels of β-galactosidase. Clearly the +10A promoter activity (and that of the other mutant promoters) is sigma 70 dependent.

Reexamination of the lacP111/112 (+10A) Mutation

The +10A mutation is identical to the previously reported lacP111/112 except that the latter is in an otherwise wild type sequence. The RNA start point for lacP111/112, which was only studied in vitro, was thought to be at +1 (37), distinctly different from that found for +10A in vivo. We therefore analyzed the in vivo starts for P111/112 with the results shown in Figure 6. In the presence of CAP-cAMP, P111/112 initiates transcipts primarily at +1,

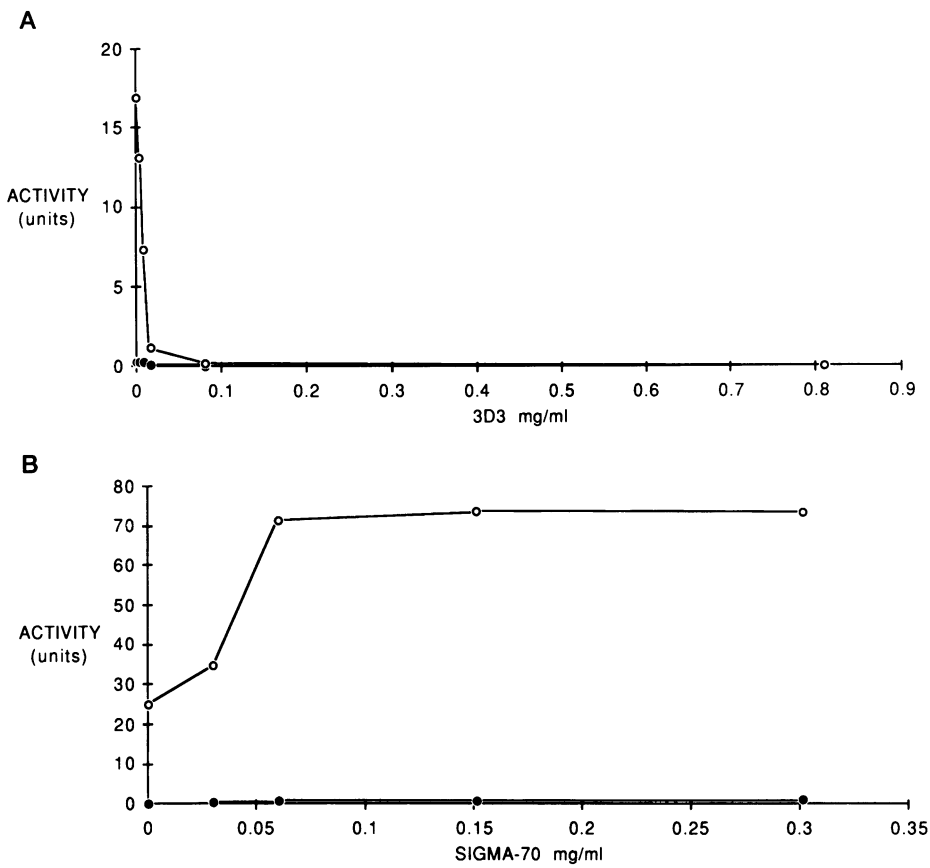


Fig. 5. In vitro Sigma 70 Dependence. The sigma 70 dependence of lacP -11G+10A expression was examined by testing the effect of adding anti sigma 70 antibody (3D3 monoclonal antibody) (Fig. 5A) or purified sigma 70 (Fig. 5B) to a coupled transcription-translation system. The final yield of $\beta$-galactosidase activity for lacP -11G is indicated with O and for lacP -11G+10A with O.

although there are some bands corresponding to those seen for promoter +10A (see Fig. 2). In the absence of CAP-cAMP (presumably in the absence of CAP-cAMP stimulated RNA polymerase binding to the standard lac promoter sequence) the start sites are the same as that found for -11G+10A.

The +10A mutation was reanalyzed in an in vitro transcription experiment using purified sigma 70 holoenzyme. The run off transcripts programmed by BstNI fragments containing either the UV5 mutation, the -11G+10A mutations or the parental -11G mutation are displayed in Figure 7, lanes 1, 2 and 3. The -11G fragment did not program the synthesis of any transcripts; however, use

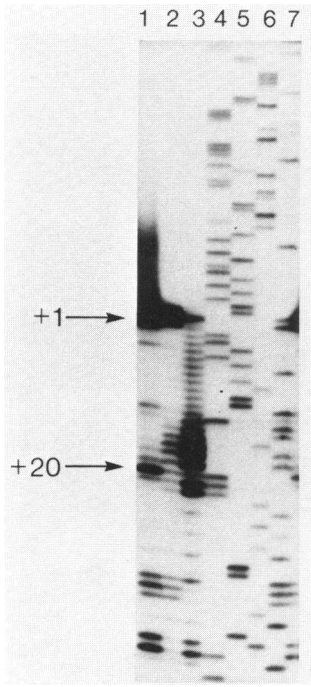


Fig. 6. Response of lacP +10A (P111/P112) to CAP-cAMP. The transcripts encoded by lacP +10A in the presence or absence of CAP-cAMP was studied by a primer extension analysis of lacP transcripts when the relevant plasmids were in RZ201 (+CAP-cAMP) or RZ203 (-CAP-cAMP). Lane 1, lacPUV5 in CSH26; lane 2, lacP +10A in RZ201; lane 3, lacP +10A in RZ203; lanes 4-7, sequencing ladders. The positions corresponding to +1 and +20 are indicated.

of the -11G+10A template resulted in the production of a series of transcripts which closely resemble those found in vivo.

## DISCUSSION

### Failure to Find Compensatory Changes

The promoter sequence on which these studies were based differed from the lac promoter at position -11. This is a strong promoter mutation reducing lac expression (see Table 2), as might be expected by the fact that it alters one of the most highly conserved positions in the canonical sigma 70 promoter sequence. Since modest levels of lacZ expression could be detected by our screening procedures for pseudorevertants (notice the properties of -2A in Table 2), one might have expected to be able to isolate second site com-

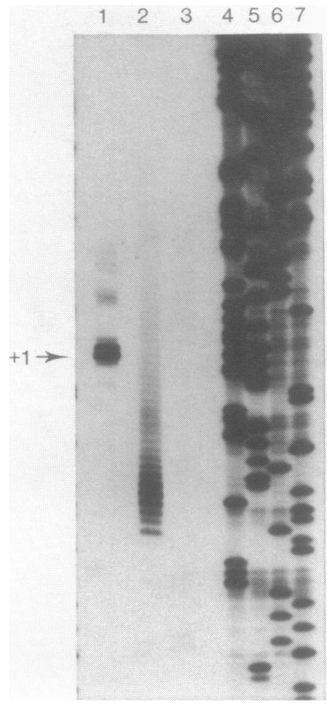


Fig. 7. In vitro Promoter Activity of lacP +10A. Run off transcription assays were performed using three BstNI templates; lacPUV5, lacP -11G+10A and lacP -11G. Lane 1 is the lacPUV5 transcription products diluted 1:10 before loading, lane 2 is the lacP -11G+10A transcription products, lane 3 is the lacP -11G transcription products and lanes 4-7 is a lacP sequence ladder. The location of the +1 position is indicated.

pensatory mutations which would have enhanced expression from position +1. Possible mutations might have occurred by one or a combination of the following changes; -34T to G, 18 bp to 17 bp spacing shift, -9G to A, -8T to A. With the possible exception of the -12A mutation to be discussed later, no such compensatory changes were found either in this study or in the companion study by Rothmel and LeClerc (14; note: the -14G mutation reported in Rothmel and LeClerc activates lacP in the absence of a functional CAP site and is not a second site revertant of a lacP mutation). It is possible that our inability to isolate such changes is a consequence of bad luck or unknown biases in the mutagenesis protocols. Alternatively, the failure might be due to one of the following possibilities. The wild type promoter sequence is activated approximately 50-fold by the CAP-cAMP complex. It could be that none of the proposed double mutant sequences (-11G plus the compensatory change) would

adequately respond to CAP-cAMP and none would yield sufficient lacZ expression in the absence of CAP-cAMP to be detected. It is also possible that the individual positions in the canonical sequence might not act as independent additive recognition sites. There may be a cooperative aspect to the interactions with the -11 position being a crucial site.

A similar attempt to isolate compensatory mutations of phage P22 ant promoter mutations was reported by Grana et al. (38). These investigations were successful in the isolation of such mutations, and these mutations appear to follow the predictions of the canonical sequence model. However, an examination of the properties of a series of such double mutations indicated that the phenotypic consequence of each single mutation varies with the precise sequence context with which it was associated (39). Thus this study also suggests that the individual base pair positions in a promoter do not act as independent recognition sites.

Activation of Nascent Promoters

The mutations which were isolated in this study were second site revertants that activated different nascent promoters. Two were reoccurrences of previously isolated mutations (+1T and +10A) although the mRNA start sites for +10A were not identified previously. Two previously unidentified nascent promoters were also detected. Each of these mutations and the relevant nascent promoter is analyzed below.

The +1T Mutation

This change is precisely the same as the previously described lacP115 mutation (9,10) and the 23-1 mutation described in Rothmel and LeClerc (14). All of the results in our analysis of this mutation are entirely consistant with previous conclusions. +1T activates a sigma 70 holoenzyme recognized promoter as a consequence of generating a new -10 region type sequence (AATTGT→TATTGT, yielding mRNA starts at +13). The only new observation is that this promoter is also active in the presence of a +5G→C change (yielding a TATTCT -10 region; see Table 2 and Figure 2). It should be noted, however, that this promoter does not precisely fit the canonical model since its level of activity is greater than that predicted from its "homology score" (36). This anomaly may be explained by the observation that sequences upstream of the proposed -35 region affect the level of this promoter's activity (40, Rothmel and LeClerc, personal communication).

The -2A Mutation

This pseudorevertant has not been previously described. It is clear that it too activates a sigma 70 dependent nascent promoter by generating a new -10 canonical sequence (TGGAAT→TAGAAT) resulting in an mRNA start at +11. As with

the +1T promoter, the overall similarity between the proposed -35 and -10 regions for the -2A promoter and the canonical sigma 70 promoter sequence seems too weak to explain its level of activity.

The -2A mutation was studied in two slightly different sequence contexts (TAGAATTG and TAGAATTC) with no obvious difference in its promoter properties. Thus a G→C change 6 bp prior to the new start point has no dramatic effect on its activity.

An important question unresolved by these studies is whether one should consider promoters as closely aligned as that generated by the -2A and +1T mutations the same or different promoters. There are many promoters which manifest multiple start points equivalent to a combination of that programmed by -2A and +1T, and if these two are to be considered distinct promoters (as proposed in this communication), perhaps one should reevaluate whether some previously described promoters with multiple start points might not consist of more than one tightly linked overlapping promoter. Also left unresolved by these studies is what, if any, physiological role the nascent promoter sequences play in wild type lac operon expression.

The -12A Mutation

This mutation is unique to this study. It is questionable whether this mutation actually reveals an additional overlapping nascent promoter. An alternative explanation is that this promoter activity is a slight modification of the wild type promoter, generating a -10 region displaced one bp upstream but using the already existing -35 region.

The -12A promoter activity has 3 interesting properties:

(1) The -12A promoter's -10 region sequence is an even poorer match to the canonical sequence than that of the wild type promoter. This might explain the extremely low level of CAP-cAMP independent expression programmed by this promoter.

(2) The -12A promoter's proposed -35 and -10 regions have a preferred spacing relative to that found for the wild type promoter.

(3) The -12A promoter manifests enhanced stimulation by the CAP-cAMP complex. We have no explanation for this property.

The +10A Mutation

This mutation is identical to the previously described mutations lacP111 and P112 which were isolated as partially CAP-cAMP independent lacP mutations (15,35). The data in this communication indicates that the previous analysis of this mutation (that it enhanced expression at the +1 start site) did not correctly identify the major cause for its promoter phenotype, probably be-

cause of the gel technology used in these earlier experiments. It is clear from the data displayed in Figures 2, 4, 6 and 7 that +10A activates a complex series of apparent starts, especially at positions +15 to +19 and +21 and +22.

One alternative explanation for this data (that the +10A mutation results in vivo in enhanced 5' mRNA end degradation of transcripts which start at +1 to give the apparent start sites) can be ruled out by the experiments shown in Figure 6. When the in vivo transcripts of the template containing the +10A mutation but lacking the -11G mutation are examined, it can be seen that in the presence of CAP-cAMP the major start site is at +1 (proving that 5' end degradation is not occurring because of the +10A mutation), while in the absence of CAP-cAMP the apparent starts are distributed between +15 and +22. (The results of this experiment also indicate that competition is occurring in vivo between holoenzyme complexes starting at the two locations). Furthermore, the in vitro transcripts programmed by -11G+10A show a pattern of transcripts very similar to that found in vivo (Figure 7), suggesting that 5' end degradation is not responsible for the observed distribution of bands. Other possible reasons for the unusual collection of apparent starts have not been examined. For instance, Patient (41) and Parker (42) have found colE1 plasmid promoters which appear to program "primed" transcription initiation events at least in vitro (e.g., the 5' end nucleotide of the transcript is not homologous to the template). One or more of the +10A promoter starts may involve such "primed" events.

The appropriate -35 and -10 sequence arrangment for the +10A promoter is not obvious. In part this is because of the presence of multiple apparent start points. In addition, various candidate sequences have only weak similarities to the canonical sigma 70 promoter sequence. These sequences are displayed in Figure 3B. Sequences (a) and (b) are interesting because they directly involve the +10A change in the proposed -10 regions. Sequences (d) and (e) are the sequences from which the +1T and -2A promoters were derived.

An alternative explanation for the unusual +10A promoter sequence might be that it is recognized by an alternate form of RNA polymerase. However, our experiments clearly indicate that the sigma 70 holoenzyme is responsible for its activity. We are led to propose two explanations for the +10A promoter activity.

(1) There is an alternative sequence which can be recognized for sigma 70 holoenzyme programmed transcription initiation.

(2) The +10A mutation generates an altered DNA structure which compensates for the lack of a canonical sequence. If this latter were true, it should yield

insights as to the functional significance of the -10 region.
It is surprising that other promoters with properties similar to +10A have not been discovered previously.

## CONCLUSION

The studies presented in this communication allow us to make two conclusions.
(1) The lac promoter elements contain multiple overlapping nascent promoter sequences. It is not clear what, if any, role these sequences play in lac operon expression as programmed by the "normal" lac promoter.
(2) The sigma 70 holoenzyme can initiate transcription at a sequence which bears little resemblance to the so-called canonical sequence. It is not clear whether this potential is utilized at any "normal" promoter although, if so, this certainly must be a minor species.

## Acknowledgements

We thank J. E. LeClerc and R. K. Rothmel for sharing their unpublished results with us and for critically evaluating our manuscript. We thank John Makris, William Marshall and Kathleen Thornton for technical help. This work was supported by Public Health Service grant GM19670 from the National Institutes of Health.

*To whom correspondence should be addressed

Present addresses: +University of Minnesota Medical School, Rochester, MN and §University of Wisconsin Medical School, Madison, WI, USA

## REFERENCES

1. Hawley,D.L. and McClure,W.R. (1983) Nucleic Acids Res. **11**, 2237-2255.
2. Harley,C.B. and Reynolds,R.P. (1987) Nucleic Acids Res. **15**, 2343-2361.
3. Reznikoff,W.S. (1984) In Kelley,D.P. and Carr,N.G. (eds), The Microbe 1984: Part II Prokaryotes and Eukaryotes, Cambridge University Press, pp.195-218.
4. Musso,R.E., DiLauro,R., Adhya,S. and deCrombrugghe,B. (1977) Cell **12**, 847-854.
5. Aiba,H., Adhya,S. and deCrombrugghe,B. (1981) J. Biol. Chem. **256**, 11905-11910.
6. Krebs,M.P. and Reznikoff,W.S. (1986) J. Mol. Biol. **192**, 781-791.
7. Bertrand,K.P., Postle,K., Wray,L.V., Jr. and Reznikoff,W.S. (1983) Gene **23**, 149-156.
8. Hillen,W., Scholmeier,K. and Gatz,C. (1984) J. Mol. Biol. **172**, 185-201.
9. Maquat,L.E. and Reznikoff,W.S. (1980) J. Mol. Biol. **139**, 551-556.
10. Peterson,M.L. and Reznikoff,W.S. (1985a). J. Mol. Biol. **185**, 525-533.
11. Peterson,M.L. and Reznikoff,W.S. (1985b) J. Mol. Biol. **185**, 535-543.
12. Reznikoff,W.S., Maquat,L.E., Munson,L.M., Johnson,R.C. and Mandecki,W.

(1982) In Rodriguez,R.L. and Chamberlin,M.J. (eds), Promoters: Structure and Function, Praeger Press, New York, pp.80-95.
13. McClure,W.R., Hawley,D.K. and Malan,T.P. (1982) In Rodriguez,R.L. and Chamberlin,M.J. (eds), Promoters: Structure and Function, pp.111-120.
14. Rothmel,R.K. and LeClerk,J.E. (1989) Nucleic Acids Res. In press.
15. Maquat,L.E. and Reznikoff,W.S. (1978) J. Mol. Biol. **125**, 467-490.
16. Miller,J.H. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
17. Kunkel,T.A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 488-491.
18. Munson,L.M., Stormo,G.D., Niece,R.L. and Reznikoff,W.S. (1984) J. Mol. Biol. **177**, 663-683.
19. Vieira,J. and Messing,J. (1987) Methods in Enzymology **153**, 3-11.
20. Mousset,S. and Thomas,R. (1969) Nature **221**, 242-244.
21. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
22. Kaslow,D. (1986) Nucleic Acids Res. **14**, 6767
23. Gardner,J.F. and Reznikoff,W.S. (1978) J. Mol. Biol. **126**, 241-258.
24. Inoue,T. and Cech,T.R. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 648-652.
25. Roth,M.J., Tanese,N. and Goff,S.P. (1985) J. Biol. Chem. **260**, 9326-9335.
26. Chen,E.Y. and Seeburg,P.H. (1985) DNA **4**, 165-170.
27. Zubay,G. (973) Ann. Rev. Genet. **7**, 267-287.
28. Artz,S.W. and Broach,J.R. (1975) Proc. Natl. Acad. Sci. U.S.A. **72**, 3453-3457.
29. Jovanovich,S.B. (1983) Ph.D. Thesis, University of California-Davis.
30. Jovanovich,S.B., Lesley,S.A. and Burgess,R.R. (1988) J. Biol. Chem., in press.
31. Bradford,M.M. (1976) Anal. Biochem. **72**, 248-254.
32. Strickland,M., Thompson,N.E. and Burgess,R.R. (1988) Biochemistry **27**, 5755-5762.
33. Fazekas de St. Goth,S. and Sheidegger,D. (1980) J. Immunol. Methods **35**, 1-21.
34. Gribskov,M. and Burgess,R.R. (1983) Gene **26**, 109-118.
35. Maquat,L.E., Thornton,K. and Reznikoff,W.S. (1980) J. Mol. Biol. **139**, 537-549.
36. Mulligan,M.E., Hawley,D.K., Entriken,R. and McClure,W.R. (1984) Nucleic Acids Res. **12**, 789-800.
37. Maquat,L.E. (1979) Ph.D. Thesis, University of Wisconsin-Madison.
38. Grana,D., Youderian,P. and Susskind,M.M. (1985) Genetics **110**, 1-16.
39. Grana,D., Gardella,T. and Susskind,M.M. (1988) Genetics **120**, 319-327.
40. Yu,X.-M. and Reznikoff,W.S. (1986) J. Mol. Biol. **188**, 545-553.
41. Patient,R.K. (1979) Nucleic Acids Res. **6**, 2647-2665.
42. Parker,R.C. (1983) Gene **26**, 127-136.
43. Johnson,R.C., Yin,J.C.-P. and Reznikoff,W.S. (1982) Cell **31**, 873-882.