Volume 17 Number 10 1989 Nucleic Acids Research

Mutational analysis of the *lac* regulatory region: second-site changes that activate mutant promoters

Randi Kubrick Rothmel+ and J.Eugene LeClerc*

Department of Biochemistry, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642, USA

Received December 19, 1988; Revised and Accepted April 21, 1989

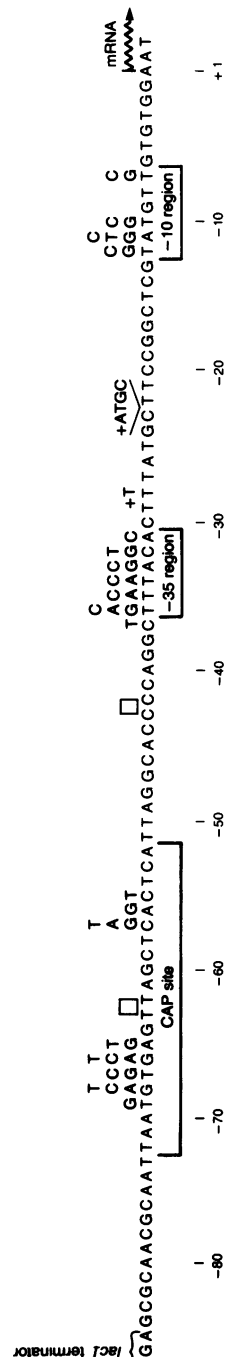
ABSTRACT

Second-site mutations that restored activity to severe *lacP1* down-promoter mutants were isolated. This was accomplished by using a bacteriophage f1 vector containing a fusion of the mutant *E. coli lac* promoters with the structural gene for chloramphenicol acetyltransferase (CAT), so that a system was provided for selecting phage revertants (or pseudorevertants) that conferred resistance of phage-infected cells to chloramphenicol. Among the second-site changes that relieved defects in mutant *lac* promoters, the only one that restored *lacP1* activity was a T→G substitution at position -14, a weakly conserved site in *E. coli* promoters. Three other sequence changes, G→A at -2, A→T at +1, and C→A at +10, activated nascent promoters in the *lac* regulatory region. The nascent promoters conformed to the consensus rule, that activty is gained by sequence changes toward homology with consensus sequences at the -35 and -10 regions of the promoter. However, the relative activities of some promoters cannot be explained solely by consideration of their conserved sequence elements.

INTRODUCTION

Studies on the promoters for *Escherichia coli* genes have identified several features of the promoter sequence that are critical to the initiation of transcription by the RNA polymerase holoenzyme. Similarity to the consensus sequences TTGACA and TATAAT, located around 35 base-pairs and 10 base-pairs upstream from the transcription start site, and a spacing of 16-18 nucleotides between those sequences, are the key features deduced from comparisons of natural promoters (1,2). However, a promoter with both derived consensus sequences has not been identified in nature; rather, the existence of a hierarchy of promoter strengths suggests that each promoter is optimized for its *in vivo* function and that promoters of equal strength can be encoded in diverse sequences (3,4). For given promoters, furthermore, mutagenesis to the consensus sequence(s) does not necessarily enhance transcription (5,6). It is not yet understood what contributions to promoter activity are made by individual bases in the promoter-RNA polymerase interaction.

A compilation of promoter mutations that enhance or decrease transcription largely confirms that the bases in the -35 and -10 regions, and the spacing between


lacI terminator
TGAGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAAT
T T
CCCT
GAGAG
T
A
GGT
C
ACCCT
TGAAGGC
+T
+ATGC
C
CTC
GGG
C
G
mRNA
CAP site
−35 region
−10 region
−80
−70
−60
−50
−40
−30
−20
−10
+1

the regions, are the main determinants for promoter activity (1). Figure 1 shows the "down" mutations in the promoter for the lactose operon, collected from the hybrid *lac* phage M13mp2. The sites of base changes that cause defects in transcription from this catabolite-activated promoter nicely conform to the binding site for the catabolite activator protein (CAP) and the conserved elements of the -35 and -10 regions; down mutations at the non-consensus G-T (positions -11,-12) in the -10 region have not been found, nor has mutation to consensus G at position -34. Since the mutational map is probably saturated for base substitutions by the variety of mutagenic treatments used, it is likely that other changes have only minor effects on transcription activity. The *lac* regulatory region is exquisitely sensitive to spacing mutations, between the symmetrical recognition sites for CAP (7) and between the CAP site and -35 region. The only mutations that have been observed between the -35 and -10 regions are addition mutations, which increase the spacing from the 18 bases in the *lac* promoter that is the maximal inter-region distance for most wild-type promoters (2).

In this report, we describe the isolation and characterization of *lac* promoter sequences that activate transcription from several of the mutant promoters shown in Fig. 1. We considered that several possible outcomes might result from such a route to examining promoter activity: 1) identifying the minimum consensus bases in each promotor hexamer required for *lac* promoter actvity; 2) determining if changes in one region of the promoter may compensate for defects in another region; and 3) revealing bases outside conserved regions that affect promoter activity. With only one exception, the changes that we found rather created new promoter regions, in a manner analogous to the CAP-independent promoter mutants isolated and analyzed by Reznikoff and co-workers (8,9,10,11). A similar outcome is reported in an accompanying paper from the Reznikoff laboratory (12).

MATERIALS AND METHODS

Media, enzymes and reagents

Media for growth of bacteria and phage were YT and 2xYT, respectively, and were prepared according to Miller (13). Antibiotic plates used for selecting clones contained either 50 μg ampicillin (Sigma) per ml or 10 μg chloramphenicol (Calbiochem) per ml.

Figure 1. Down mutations in the *lac* promoter of M13mp2 *lac* hybrid phage. Single-base substitutions are shown above the wild-type *lacP1* sequence, open boxes indicate single-nucleotide deletions, and + indicates addition mutations. Data are from LeClerc and Istock (23), Kunkel and Alexander (43), and unpublished results.

Reagent grade chemicals were used. Non-radioactive nucleotides were from Pharmacia. [^{35}S] dATP (500 Ci/mmol) and [^{32}P]ATP (>3000 Ci/mmol) were from NEN.

Enzymes and proteins were from the following suppliers: proteinase K, Beckman; DNA polymerase (Klenow fragment), BRL or Pharmacia; AMV reverse transcriptase and RNaseIn, Boehringer-Mannhein; T4 polynucleotide kinase, U.S. Biochemicals; and T4 DNA ligase, N.E. Biolabs.

Bacterial strains

E. coli JM101 (14) was used for preparation of phage stocks and for DNA transfection. Revertants of *lacP*-CAT phages were isolated after UV mutagenesis using SMH51, an F$^+$ derivative of a Δ(*pro-lac*) strain, CSH50 (13), or by growth in an F$^+$ derivative of the *mutD5* strain LE30 (15). Strain SMH51 was used for chloramphenicol inactivation studies and the determination of CAT activity in phage-infected cells.

DNA isolation and sequencing

RF DNA was isolated from phage-infected JM101 cells according to the procedure of Godson and Vapnek (16), except that the cleared lysate was used directly on cesium chloride-ethidium bromide gradients. The DNA was then dialyzed, phenol extracted, and precipitated with ethanol.

Single-stranded DNA for sequence analysis was purified from phage preparations as previously described (17), except that the ether extraction was omitted. Sequencing was according to the procedures of Sanger *et al.* (18) and Biggin *et al.* (19). An oligonucleotide primer, 5' AGCTTCCTTAGCTCC, which anneals to a site in the CAT structural gene (20), was synthesized using an Applied Biosystems 380 synthesizer.

Construction of RL-CAT vector

RL-CAT, a bacteriophage f1 vector containing a fusion of the *lac* promoter and a chloramphenicol acetyltransferase (CAT) structural gene, was constructed from portions of three vectors: RK112, an f1 phage derived from the f1-pBR322 chimera R208 and the f1 cloning vector R199 (21); RL1, an M13*lac* phage derivative of M13mp2 and M13mp8 (22) containing the *lac* promoter in a *Hin*dIII-*Sal*I fragment; and pCM1 (20), containing the CAT structural gene in a *Sal*I fragment.

RK112 The ColE1 replication origin of R208 was removed so that only the f1 phage origin would be utilized. This was accomplished by restriction enzyme cutting at a unique *Pvu*II site and digestion with Bal31 nuclease, the latter reaction monitored by loss of a *Taq*I site present in the plasmid origin. The orientation of the pBR-derived DNA, encoding resistance to ampicillin (Ap) and tetracycline, was

switched by removing the insert as an *Eco*RI fragment and cloning it into the *Eco*RI site of the f1 cloning vector R199. The resulting f1 vector, RK112, encodes resistance of phage-infected cells to both antibiotics and contains unique *Sal*I and *Hin*dIII sites in the *tet* gene.

RL1 This vector was constucted for a source of *lac* promoter DNA available as a convenient restriction fragment. The *lac* promoter of M13mp8 was first removed by *Eco*RI and *Ava*I digestion, the ends filled in using DNA polymerase (Klenow), and the vector was ligated. The *lac* promoter of M13mp7 was removed by *Pvu*I and *Eco*RI digestion and, after filling ends, was cloned into a blunt-ended *Pst*I site of the promoterless M13mp8 vector, utilizing the blue color reaction of phage-infected JM101 cells on Xgal/IPTG indicator plates as the cloning detection system. The construct, RL1, provided the *lac* promoter as a *Hin*dIII-*Sal*I fragment.

RL-CAT The *lac* promoter, isolated as a *Hin*dIII-*Sal*I fragment from RL1, and the CAT structural gene, isolated as a *Sal*I fragment from pCM1, were cloned into RK112 at *Hin*dIII and *Sal*I sites, thereby disrupting the *tet* gene of RK112. The resulting f1 vector, RL-CAT, has a *lacP*-CAT fusion gene, with chloramphenicol (Cm) resistance driven by *lacP1* and Ap resistance encoded in pBR-derived DNA. Phage-infected cells are resistant to both antibiotics and produce high titers of phage (~ 10^{12}/ml). For subsequent genetic manipulations with *lacP* mutations, phage infected cells were either treated as colonies on agar plates containing Cm and/or Ap, or phage plaques were picked from a lawn of cells and tested for cell growth on antibiotic plates. In the experiments reported here, the host cells used were deleted for chromosomal *lac* genes in order to avoid *lacI*-mediated repression (which only slows the growth of phage-infected cells on Cm) and to ensure against recombination with the host *lac* promoter.

Transfer of *lacP* mutations to RL-CAT

The *lacP* mutations used in this study were from a collection of spontaneous and ultraviolet (UV)-induced mutants of M13mp2 phage, isolated as colorless plaques of phage-infected cells on Xgal indicator plates (17,23). They were numbered 31, -T at position -62 of the CAP site; -35 region mutants 45, T→G at -36, and 174, T→G at -35; -10 region mutants 23 and 204, A→C and A→T, respectively, at -11; and 37, a +ATGC spacing mutation between the -35 and -10 regions (cf. Fig.2). In order to transfer the mutations from M13mp2 to the *lac* promoter of the RL-CAT vector, an *Eco*RI-*Pvu*II fragment containing *lacP* was isolated from the RF form of each mutant phage DNA and used for oligomer-directed mutagenesis on single-stranded RL-CAT DNA (24,25). After transfection of $CaCl_2$-treated JM101 cells, candidate plaques were tested for Ap resistance and

Cm sensitivity. The desired sequence change in each mutant clone, now named RL31, RL45, *etc.*, was verified by sequence analysis.

Isolation of *lacP* revertants

Two mutagenesis procedures were used to induce phage mutants that confer Cm resistance in phage-infected cells, presumably by direct reversion of the *lacP* mutations or by second-site changes that activate the mutant promoters. The mutagenic treatments, ultraviolet irradiation and growth of hybrid phages in a mutator strain of *E. coli*, were chosen because they are relatively non-specific with respect to nucleotide specificity.

UV mutagenesis. The procedure was carried out as described in LeClerc *et al.* (17), except that UV-irradiated (50 Jm^{-2}) RL phages were used to infect UV-irradiated (15 Jm^{-2}) SMH51 cells (moi ~ 0.25) and the infected cells were plated on agar containing Cm (10 μg/ml). After overnight growth, phage from individual Cm^R colonies were grown on a lawn of JM101 cells in order to isolate single phage plaques for retesting on antibiotic plates. The *lacP* region of candidate clones was then sequenced to determine nucleotide changes. By the UV mutagenesis procedure, the number of Cm^R revertants was enhanced from twofold to 30-fold over the spontaneous level for different RL phages.

mutD mutagenesis. Each RL phage was grown overnight in a culture of a *mutD* mutator strain. Phage stocks from this growth were regrown for five hours in a liquid culture of SMH51 cells in the presence of Cm (10 μg/ml) in order to enrich mutants that rendered cells Cm^R. Infected cells were then grown on Cm plates to isolate single colonies and the procedure was continued as described above for UV-induced revertants.

Chloramphenicol inactivation and CAT assay

Phage-infected SMH51 cells were plated on YT agar containing varying levels of Cm. After overnight growth, plates were scored for inactivation of colony-forming ability.

Cell lysates for the determination of CAT activity in phage-infected cells were prepared as follows. SMH51 cells were grown at 37°C to $1x10^8$/ml, infected with RL phage at moi 10-20, and the growth was continued to $5x10^8$/ml (about two hours). The cells were harvested and extract was prepared according to the method of Wickner et al. (26). CAT activity was determined by the spectrophotometric assay of Shaw (27), using 5 μl to 100 μl of extract. Total protein was determined by the method of Bradford (28) using Bio-Rad reagents.

mRNA isolation and mapping

mRNA from phage-infected cells was isolated according to the guanidinum/hot phenol procedure of Feramisco *et al.* (29), with the following modifications.

Five ml of SMH51 cells at $1x10^8$/ml were infected with phage (moi 10-20) and grown for two hours at 37°C. After RNA extraction and ethanol precipitation, RNA was dissolved in a buffer of 0.1M Tris-HCl (pH 7.4), 50 mM EDTA, 0.2% SDS, 2 mM $CaCl_2$, and 10 mM $MgCl_2$. Proteinase K was added to a final concentration of 200 µg/ml, DNase to 3.3 µg/ml and RNasin to 10 units per reaction, followed by incubation at 37°C for 2 hrs. The reaction mixture was extracted twice with phenol/ chloroform, and the mRNA was precipitated in two volumes of 95% ethanol and stored at -20°C in 70% ethanol.

mRNA start sites were identified by primer extension using a 5' ^{32}P-end labeled oligonucleotide, pACGATGCCATTGGGA, and AMV reverse transcriptase (30,31). Five ng primer was annealed to 15 µg RNA at 42°C for 60 min in 50 mM Tris-HCl (pH 8.3), 100 mM KCl, 10 mM $MgCl_2$, 1.0 mM DTT, and 1.0 mg/ml BSA. Four dNTPs were added to a final concentration of 0.5 mM each. The reaction was started by adding 5 units of AMV reverse transcriptase and incubated at 37°C for 20 min. The reaction was stopped by adding formamide/dye and heating at 90°C. Five µg RNA was loaded on an 8% polyacrylamide–7M urea gel for electrophoresis. Sequence standards were prepared using the same primer and RL-CAT template DNA.

RESULTS

Isolation of *lacP* revertants

The f1 RL-CAT vector contains a fusion gene of the *lac* promoter and the structural gene for chloramphenicol acetyltransferase (CAT). The phage vector was designed for isolating revertants of *lacP1* mutations available from previous studies on mutagenesis of the *lac* region of M13mp2 phage (17,23). Since the Cm resistance of phage-infected cells is driven by the *lac* promoter in RL-CAT, cells infected with phage containing *lacP1* down mutations remained Cm-sensitive and a strong selection system was provided for isolating phage mutants that rendered cells Cm^R. Among the phages that promote colony growth on Cm plates should be those that contain direct reversion mutations back to the wild-type *lacP1* sequence and those of interest containing second-site changes that activate mutant *lac* promoters. The RL-CAT vector was based in the single-stranded DNA bacteriophage f1, for easy genetic manipulation of *lacP* mutations, and it carries a second antibiotic resistance gene (Ap) for selection of phage-infected cells (see Materials and Methods).

The *lacP* mutations chosen for reversion analysis are shown in Figure 2; in the M13mp2 background, they caused a complete defect in transcription activity as assessed by the lack of α-complementation for ß-galactosidase activity in phage-

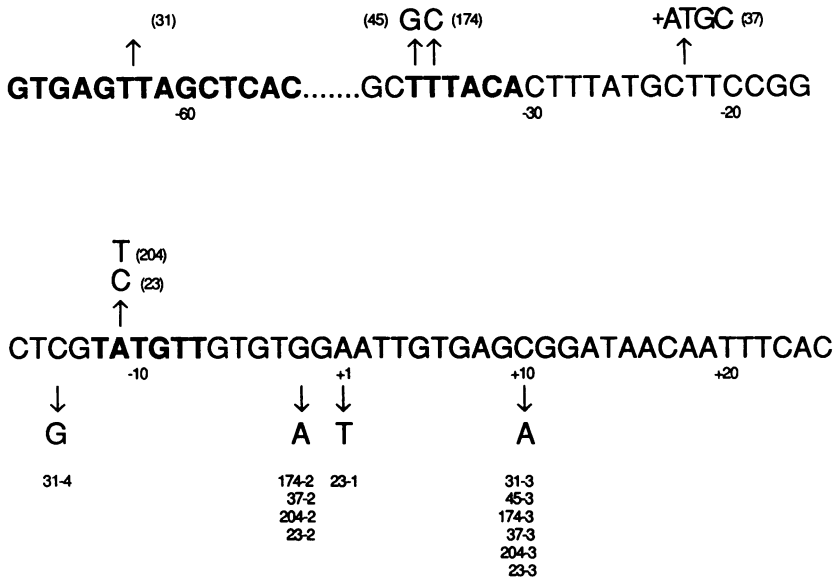


Figure 2. Second-site changes that activate mutant *lac* promoters. The down-promoter mutations used in this study are indicated above the *lacP1* sequence. Second-site changes that activate each mutant promoter are shown below the sequence; they are numbered for the corresponding mutant background indicated above the sequence.

infected cells (17). When the mutations were transferred to the RL-CAT vector, phage-infected cells showed no growth on plates with a low level of Cm (5 μg/ml). The mutations represent defects in each of the regions critical for *lac* promoter activity: the CAP site (RL31); the -35 region (RL45 and RL174); the -10 region (RL23 and RL204); and the spacing between the -35 and -10 regions (RL37). Since the roles of the -35 and -10 region bases in the RNA polymerase-promoter interaction are not known, and hence the effects of specific substitutions unknown, we chose two base changes in each region for the reversion analysis.

UV- and *mutD*-induced revertants of phage that promoted colony growth in the presence of Cm (10 μg/ml) were collected. Sequence analysis of the *lacP* region of 142 isolates showed that ten phages contained the wild-type (*lacP1*) sequence and thus were direct revertants; 40 isolates contained a second-site change within the *lac* regulatory region and are described below; and 92 phages showed only the *lacP* sequence change of the starting RL phage and were not analyzed further.

Second-site sequence changes and transcription start sites

The second-site sequence changes found in the *lac* regulatory region of

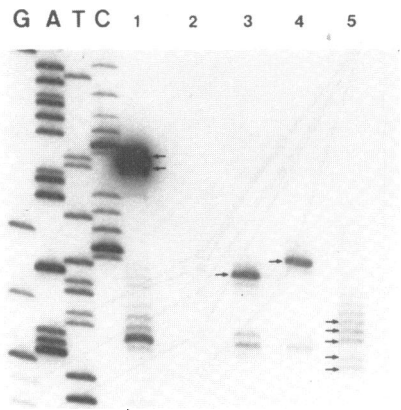


Figure 3. Primer extension products on mRNA from RL phage-infected cells. Transcription start sites (indicated by arrows) were mapped on *in vivo* mRNA from RL-CAT (*lacP1*), lane 1; RL23, lane 2; RL23-1 (+1T), lane 3; RL23-2 (-2A), lane 4; and RL23-3 (+10A), lane 5. Sequencing ladders indicated by G,A,T,C were from RL-CAT DNA.

revertant phages are summarized in Figure 2 (downward-pointing arrows). Out of 40 revertants analyzed, several showed identical changes and four unique sites were identified: G at position -14 (relative to the *lacP1* sequence); A at -2; T at +1; and A at +10. The +10A change was found in all six mutant backgrounds tested and the -2A change was found in four mutant backgrounds. With respect to the origin of the mutants, the -14G and +1T changes were unique isolates from the ***mutD*** and UV mutagenesis procedures, respectively, and multiple isolates of the -2A and +10A changes arose from both protocols.

Second-site changes in the revertant phages may activate mutant promoters by enhancing transcription from *lacP1* (*i.e.* use of the +1 start site) or by creating new promoters within the *lac* regulatory region. In order to distinguish these possibilities, start sites were mapped by primer extension on *in vivo* mRNA from phage-infected cells. Figure 3 shows an example of the results for the original RL23 mutant and for revertants RL23-1, RL23-2 and RL23-3. Of the four second-site changes found, only the -14G revertant (RL31-4) used the *lacP1* +1 start site. Revertants containing -2A and +1T showed unique start sites at +11 and +13, respectively, relative to the *lacP1* sequence. Primer extension using the +10A revertants gave termination bands that, in different experiments, were most pronounced at positions +17, +18, +19, +21 and +22 (cf. Fig. 2). Multiple primer extension products in this region were observed in transcript maps for all active *lac* promoters; they appear to be artifacts of the primer extension method. In the +10A

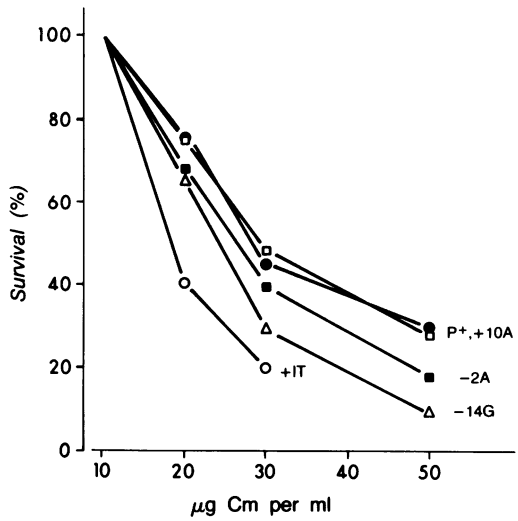


Figure 4. Chloramphenicol inhibition of colony-forming ability. Survival at the indicated concentrations of Cm is shown for RL phage-infected cells: RL-CAT (*lacP1*), ● ; RL174-3 (+10A), □ ; RL174-2 (-2A), ■ ; RL31-4 (-14G), △ ; and RL23-1 (+1T), ○ .

revertant, the absence of such a product at +20 leads us to believe that the other products represent a family of start sites, but this conclusion is not secure.

Cm resistance and CAT activity

The relative activities of the revertant promoters were compared by assessing the colony-forming ability of phage-infected cells on agar plates containing Cm. As shown in Figure 4 for revertants representing each second-site change, RL174-3 of the +10A promoter group conferred about the same resistance to Cm as RL-CAT with a wild-type promoter. The decreasing order of resistance was then RL174-2 (-2A) > RL31-4 (-14G) > RL23-1 (+1T). Some differences in the inactivation curves for revertants with different mutant backgrounds were observed; in particular, RL37-2 of the -2A group conferred the same resistance as RL-CAT (data not shown). When the revertants were tested using SMH92 cells containing a *crp* mutation that inactivates the catabolite activator protein, results similar to those with SMH51 *crp*+ were obtained, indicating that the revertants were not dependent on CAP-cAMP for promoter activity (data not shown). It should be noted that the RL31-4 (-14G) promoter, which activates the *lacP1* +1 start site, is not susceptible to CAP-cAMP stimulation because of the CAP site mutation in RL31.

A more quantitative measure of gene expression from the mutant promoters

Table 1. CAT activity in extracts of RL phage infected cells.

Promoter	Vector	CAT activity[a]	Relative activity
lacP1	RL-CAT	412 ± 25	(1.0)
—	RL31	7.3 ± 2.8	—
—	RL45	11.8 ± 4.5	—
—	RL174	6.2 ± 0.9	—
—	RL37	3.9 ± 1.4	—
—	RL204	4.6 ± 0.6	—
—	RL23	2.3 ± 0.5	—
-14G	RL31-4	135 ± 21	0.33
-2A	RL174-2	191 ± 14	0.46
	RL37-2	299 ± 25	0.73
	RL204-2	112 ± 14	0.27
	RL23-2	107 ± 24	0.26
+1T	RL23-1	87 ± 5	0.21
+10A	RL31-3	244 ± 18	0.59
	RL45-3	235 ± 18	0.57
	RL174-3	225 ± 15	0.55
	RL37-3	221 ± 34	0.54
	RL204-3	147 ± 15	0.36
	RL23-3	121 ± 20	0.29

[a] Activity (±S.D.) given in units/mg according to Shaw (27), determined at 6-8 time points using two or more amounts of extract. The activity measured in extracts of uninfected cells was approximately 1.2 units/mg.

was obtained by measuring CAT activity in extracts of phage-infected cells, given in Table 1. The parent mutants showed only 1-2% of the activity of RL-CAT infected cells, and in contrast to the results from the Cm resistance assay, none of the second-site changes restored to the mutants the high activity of the wild-type promoter. It is also apparent that the different mutant backgrounds affect the activities of the -2A promoters and the +10A promoters. As assessed by CAT activity, the -2A promoter of RL37-2 is the most active of all the revertants. Four of the +10A promoters form a homogeneous group, but promoter activity is diminished in the parent -10 region mutants RL23 and RL204.

DISCUSSION

The aim of this study was to investigate the sequence requirements for an active promoter in *E. coli*, by identifying compensatory mutations that relieve defects in mutant *lac* promoters. Analysis of the second-site reversion mutations in

Table 2. Promoter regions activated by second-site reversion mutations.

Promoter	-35 region	Spacing	-10 region[a]	Start site(s)	Conserved bases[b] -35	-10
Consensus	TTGACA	16-18	t·tggTATAATg	a/g	(6)	(6)
lacP1	TTTACA	18	t·gTATGTTg	a (+1)	5	4
-14G	TTTACA	18	tggTATGTTg	a (+1)	5	4
-2A	ATGCTT	17	ttgtg**TA**G**A**AT	g (+11)	2	5
+1T	TTCCGG	16	tgg**T**ATTG**T**g	a (+13)	2	4
+10A						
RL23-3	TCGTCT	17	t·tg·G**A**GG**A**T	caa·tt (+17,18,19,21,22)	3	3
RL204-3	TCGTTT	"	"	"	2	3
others	TCGTAT	"	"	"	2	3

[a] Second-site changes that activate mutant promoters are shown in bold print. Weakly conserved positions of the extended -10 region are given in lower case print.
[b] Only the conserved positions of the -35 and -10 region hexamers (shown underlined) are considered here.

six different *lacP* mutants devoid of promoter activity revealed only one such compensatory mutation, to a weakly conserved G 5' to the -10 region of a *lac* promoter inactivated by a CAP-site mutation. As assessed by the altered mRNA start sites of the three other revertant promoters we found, their second-site changes created new promoters within the *lac* regulatory region. In this respect, our results are similar to those from an earlier search for CAP-cAMP independent *lac* promoters conducted by Reznikoff and colleagues (8) and the results reported in an accompanying paper by Karls *et al.* (12) on revertants of *lacP* -11G mutants. Indeed, all three second-site changes that activate nascent promoters have been isolated independently in the Reznikoff laboratory: the +1T and +10A changes in the CAP-cAMP independent mutants P115 and P111/112, respectively (8), and the -2A, +1T and +10A changes in the *lacP* -11G background (12). Nevertheless, two lines of evidence indicate that our search for second-site revertants was not exhaustive. First, Karls *et al.* (12) identified another change, -12A, which activates transcription at a -1 start site; such a promoter would likely be active in the RL204

background, since the resulting -10 region (GATTGT) would be closer to consensus than in the -11G background (GAGTGT). Second, we have introduced the +1T change found in RL23-1 into the RL37 background by oligonucleotide-directed mutagenesis, and the resulting mutant conferred Cm resistance to phage-infected cells. Hence, other changes, including compensatory mutations, may have been missed by our protocols.

In Table 2, we have assigned the sequences that most likely serve as -35 and -10 regions in the revertant promoters. These assignments were based upon the positions of the determined mRNA start sites and similarity to the consensus sequences and spacing from compiled *E. coli* promoters (1,2). It can be seen that the promoter hexamers have at least two of the most highly conserved bases in the -35 region and three or more consensus -10 bases, so that the assigned sequences fit within the "rules" for *E. coli* promoters recognized by *E. coli* RNA polymerase-sigma 70 holoenzyme (4). We have evidence that the sequence assignments in Table 2 are correct for the +1T and -2A promoters. First, we have shown by oligonucleotide-directed mutagenesis that changes at a conserved site in the putative -35 regions of the RL23-1 (TTCCGG → TCCCGG) and RL37-2 (ATGCTT → ACGCTT) promoters inactivated CAT expression, completely in the case of RL23-1 and six- to sevenfold in RL37-2 (manuscript submitted). These results are consistent with the effects of promoter mutations at this position of the -35 region (1). Second, in a search for pseudorevertants of *lacP1* mutations in M13mp2 phage, we identified a -26T mutant that used the +11 start site, as in -2A promoters (unpublished results). In this case, moving to consensus homology in a nascent -35 region (ATGCTT → TTGCTT), rather than creating a -10 region (TGGAAT→ TAGAAT), programmed the same start site.

In addition to the most conserved positions shown in Table 2, we have given the sequence of an extended -10 region containing weakly conserved bases around the -10 hexamer. The frequency of the indicated base at these positions is only 33-37% in the 263 promoters analyzed by Harley and Reynolds (2), compared to 50-90% for the most conserved positions. Consensus bases are found in the extended -10 regions of the revertant promoters and, as demonstrated by the -14G promoter, they may be essential for promoter activity in some sequence contexts. In particular, extensive literature exists on the significance of T-G (positions -15, -14) 5' to the -10 region. Promoter-down mutations away from G have been identified in λPRE (32), λPRM (33), *argC* (34), *galP1* (35) and *galP2* (36), and away from T in *tyrT* (37) and *galP1* (38). More analogous to our study, Graña *et al.* (39) found that an A→G change at position -14 activated phage P22 Pant promoters containing severe -35 or -10 region down mutations. In RL31-4, the -14G reversion activated

lacP1 that contained a severe CAP-site mutation; it could be considered that the -14G change restored promoter activity by relieving *lacP1* dependence on cAMP-CAP for activation. That -14G relieved the effects of CAP-site (*lacP1*) and -10 region (Pant) mutations, in addition to -35 region mutations, argues against a unique function for the TG sequence that has been proposed to operate in promoters with little -35 region homology (40,41), since both *lacP1* and Pant have highly conserved -35 regions.

In contrast to the "consensus-is-best" aspects of the preceding Discussion, the relative activities of the nascent promoters are not explained by homology to the consensus promoter regions. In particular, gene expression from the -2A promoters in different mutant backgrounds is significantly different, as assessed by the Cm resistance of phage-infected cells and CAT activity, despite identical consensus homology. More striking is that the +10A promoters, with the lowest consensus homology of all the revertants, are among the most active promoters. The basis for the enhanced activity of the RL37-2 promoter involves sequence upstream to its -35 region, including the four base-pair duplication of the RL37 background that is not present in the other -2A revertants (cf. Fig.2). Using site-specific mutagenesis methodology, we have shown that the sequence adjacent to the -35 regions of both the RL37-2 and RL23-1 promoters enhances, or in the case of RL23-1 is required for, the activity of the nascent promoters (manuscript submitted). Yu and Reznikoff (11) have previously shown by deletion analysis that the sequence is required for the full activity of the +1T promoter of *lac*P115. We also believe that the basis for the enhanced activity of the -2A promoter in RL174-2, compared to RL204-2 and RL23-2, is the silent *lacP1* -35 region of the RL174 mutant background. Inactivation of the -35 region by site-specific mutation significantly increased the activity of nascent promoters in other mutant backgrounds, possibly by removing a non-productive binding site that competes with the nascent promoter for RNA polymerase (unpublished results).

We applied the computer program of Mulligan *et al.* (42) to the sequences of our mutant promoters, in order to predict the most probable promoter sites and rank them by homology score. All promoter sites listed in Table 2 were selected, with the exception of the +10A promoters. Specifically, the silent promoter site of the -2A parent mutants (*i.e.*, without the -2A change) outranked all +10A promoters, and the site of the +1T parent mutant outranked +10A promoters other than RL23-3 (the RL23 mutation increases homology at its putative -35 region). In other words, homology to consensus sequences does not explain activation of nascent promoters by the +10A change, since homology considerations predict that the +11 or +13 start sites would more probably be used. This is clearly not the case

seen in the primer extension assay (cf. Fig. 3), but rather staggered termination bands at either side of +20 represent the only start sites that we find to account for active transcription from +10A promoters. Another conundrum is that CAT expression from the +10A promoter with the best consensus homology, RL23-3, is the lowest of the six +10A promoters. Karls *et al.* (12), who have obtained experimental results consistent with those reported here, have suggested possible -10 and -35 regions for the +10 promoter that are moved slightly upstream from those proposed in Table 2, and that have even lower consensus homology: CTCGTA--16bp--GAGAGG. The data of Table 1 provide some support for that assignment, since the RL204 and RL23 mutant backgrounds would further decrease -35 region consensus homology (to CTCGTT and CTCGTC, respectively; cf. Fig. 2) and CAT expression from the +10A promoters in those backgrounds is reduced by twofold. An alternative explanation is that the other mutant backgrounds enhance promoter activity in a manner analogous to RL174-2. In either case, further analysis of the +10A promoter is required to understand its utilization by RNA polymerase and to provide a molecular basis for the multiple start sites that it programs.

The *lac* regulatory region appears to be rich in sequences that serve as elements for active promoters; at least four sets of natural sequences function as -35 regions if paired with nascent -10 regions created by mutation. It may not be surprising that several of these elements exert an effect on the function of the nascent promoter. The variable activities of homologous promoters in different *lacP* mutant backgrounds also demonstrate that sequences outside the normally conserved regions have specific effects on promoter activity, which may not be surmised by comparisons of compiled promoters. An optimistic view is that an orderly analysis of these sequence effects in overlapping promoter regions will provide some additional rules to describe the function of natural promoters.

ACKNOWLEDGEMENTS

We gratefully acknowledge W.S. Reznikoff for useful discussions and providing information before publication, P. Model for providing phage vectors, and Daniel Lang for preparation of the manuscript. This work was supported by NIH grant GM27817.

*To whom correspondence should be addressed

+Present address: Department of Microbiology and Immunology, University of Illinois at Chicago, College of Medicine, Chicago, IL 60680, USA

REFERENCES

1. Hawley,D. and McClure,W.R. (1983) Nucl. Acids Res. **11**, 2237-2255.
2. Harley,C.B. and Reynolds,R.P. (1987) Nucl. Acids Res. **15**, 2343-2361.
3. Bujard,H. (1980) Trends Biochem. Sci. **5**, 274-278.
4. McClure,W.R. (1985) Ann. Rev. Biochem. **54**, 171-204.
5. Inouye,S. and Inouye,M. (1985) Nucl. Acids Res. **13**, 3101-3110.
6. Mandecki,W., Goldman,R.A., Powell,B.S. and Caruthers,M.H. (1985) J. Bacteriol. **164**, 1353-1355.
7. Ebright,R.H., Cossart,P., Gicquel-Sanzey,B. and Beckwith,J. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 7274-7278.
8. Maquat,L.E., Thornton,K. and Reznikoff,W.S. (1980) J. Mol. Biol. **139**, 537-549.
9. Maquat,L.E. and Reznikoff,W.S. (1980) J. Mol. Biol. **139**, 551-556.
10. Peterson,M.L. and Reznikoff,W.S. (1985) J. Mol. Biol. **185**, 525-533.
11. Yu,X.E. and Reznikoff,W.S. (1986) J. Mol. Biol. **188**, 545-553.
12. Karls,R., Flynn,S., Pak,A., Schulz,V. and Reznikoff,W.S. (1989) Nucl. Acids Res. This volume.
13. Miller,J. (1972) Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
14. Messing,J. (1979) Recombinant DNA Technical Bulletin **2**, 43-48.
15. Silhavy,T.J., Berman,M.L. and Enquist,L.W. (1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
16. Godson,G.N. and Vapnek,D. (1973) Biochem. Biophys. Acta. **299**, 516-520.
17. LeClerc,J.E., Istock,N.L., Saran,B.R. and Allen Jr.,R.A. (1984) J. Mol. Biol. **180**, 217-237.
18. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463-5467.
19. Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) Proc. Natl. Acad. Sci. U.S.A. **80**, 3963-3965.
20. Close,T. and Rodriguez,R. (1982) Gene **20**, 305-316.
21. Boeke,J.D., Vovis,G.F. and Zinder,N.D. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 2699-2702.
22. Messing,J. (1983) Methods Enzymol. **101**, 20-78.
23. LeClerc,J.E. and Istock,N.L. (1982) Nature **297**, 596-598.
24. Zoller,M.J. and Smith,M. (1982) Nucl. Acids Res. **10**, 6487-6500.
25. Kunkel,T.A. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 488-492.
26. Wickner,W., Brutlag,D., Sheckman,R. and Kornberg,A. (1972) Proc. Natl. Acad. Sci. U.S.A. **69**, 965-969.
27. Shaw,W.V. (1975) Methods Enzymol. **43**, 737-755.
28. Bradford,M. (1976) Analytical Biochem. **72**, 248-261.
29. Feramisco,J.R., Smart,J.E., Burridge,K., Helfman,D.M. and Thomas,G.P. (1982) J. Biol. Chem. **257**, 11024-11031.
30. Treisman,R., Proudfoot,N.J., Shader,M. and Maniatis,T. (1982) Cell **29**, 903-911.
31. Bina-Stein,M., Thoren,M., Salzman,N. and Thompson,J.A. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 731-735.

32. Rosenberg,M., Court,D., Shimitake,H., Brady,C. and Wulff,D.L. (1978) Nature **272**, 414-423.
33. Rosen,E.D., Hartley,J.L., Matz,K., Nickols,B.P., Young,K.M., Donelson,J.E. and Gussin,G.N. (1980) Gene **11**, 197-205.
34. Piette,J., Cunin,R., Boyen,A., Charlier,D., Crabeel,M., van Vliet,F., Glansdorff,N., Squires,C. and Squires,C.L. (1982) Nucl. Acids Res. **10**, 8031-8048.
35. Busby,S., Truelle,N., Spassky,A., Dreyfus,M. and Buc,H. (1984) Gene **28**, 201-209.
36. Bingham,A.H.A., Ponnambalam,S., Chan,B. and Busby,S. (1986) Gene **41**, 67-74.
37. Berman,M.L. and Landy,A. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4303-4307.
38. Busby,S.J.W., Aiba,H. and deCrombrugghe,B. (1982) J. Mol. Biol. **154**, 211-227.
39. Graña,D., Youderian,P. and Susskind,M.M. (1985) Genetics **110**, 1-16.
40. Ponnambalam,S., Webster,C., Bingham,A. and Busby,S. (1986) J. Biol. Chem. **261**, 16043-16048.
41. Keilty,S. and Rosenberg,M. (1987) J. Biol. Chem. **262**, 6389-6395.
42. Mulligan,M.E. and McClure,W.R. (1986) Nucl. Acids Res. **14**, 109-126.
43. Kunkel,T.A. and Alexander,P.S. (1986) J. Biol. Chem. **261**, 160-166.