Nucleic Acids Research, 1994, Vol. 22, No. 22 4667–4672

Start site selection at *lac*UV5 promoter affected by the sequence context around the initiation sites

Woojin Jeong and Changwon Kang*
Department of Life Science, Korea Advanced Institute of Science and Technology, 373-1 Kusong-dong, Yusong-gu, Taejon 305-701, Korea

Received July 14, 1994; Revised September 30 and Accepted October 14, 1994

ABSTRACT

The effects of single base pair substitutions at the initiation sites of *lac*UV5 promoter on the transcription start site selection by *E.coli* RNA polymerase were systematically studied. Transcription start sites were mapped by sizing the cytosine-specifically terminated transcripts produced *in vitro* by using a chain terminator 3′-deoxycytidine 5′-triphosphate (3′-dCTP) in transcription reactions. Transcription of a prototype *lac*UV5 promoter initiated with three purines (−1G, +1A and +2A; +1 representing the predominant start site) located 6–8 bp downstream from the Pribnow box. All the substitutions affected the start site selection, resulting in a change in the number of start sites (from 3 to 2 or 1) and/or a shift of the major start site (to −1 or +2). None of the variants started outside the 3-bp region and at the positions substituted by a pyrimidine. Purine-to-pyrimidine changes suppressed not only initiation at the substituted position but also, in some cases, at the other purine position. Purine-to-purine changes also shifted the major start site or suppressed the initiation at other sites. Changes at −2 and +5 also affected the start site selection. Thus, the sequence context around the initiation sites of *lac*UV5 promoter strongly influences the selection of initiating nucleotides by *E.coli* RNA polymerase.

INTRODUCTION

Transcription initiation is a major process for the regulation of gene expression in prokaryotes and involves many steps: i) binding of RNA polymerase to the promoter, ii) isomerization to an open complex, iii) formation of the first phosphodiester bond, iv) abortive initiation and recycling of the polymerase, v) transition into an elongation complex and promoter clearance (1–3). All these steps can be subjected to regulation. Selection of transcription start site also provides a means for regulation of gene expression. Depending on the intracellular [GTP]/[CTP] pool ratio, transcription of *Salmonella typhimurium pryC* starts at either C or G, which are separated 2 bp apart, leading to different secondary structures and translational attenuation of the transcripts (4).

Compared to other initiation steps, less systematic studies have been performed on the mechanism of start site selection by RNA polymerase. Compilation and comparison of transcription initiation sites of many *E.coli* promoters have revealed some predominant features (5,6). (i) A purine is preferred to a pyrimidine as an initiating nucleotide, and A is slightly more frequently used than G. (ii) The start site is usually located at proper distances from the −10 and −35 regions. (iii) Pyrimidines C and T are found more frequently than the others at the positions −1 and +2 next to the initiation site of the non-coding strand, respectively.

However, these predominant features cannot make general rules for prediction for the start sites. In *E.coli* promoters, transcription mostly (more than 90%) initiates with a purine, but also is known to start with a pyrimidine in some cases (6). The distance between the initiation site and the conserved Pribnow box (−10 region) is mostly 7 bp but also varies from 4 to 12 bp. Many promoters start transcription even at multiple sites. This complexity can probably be resolved by first identifying all the elements involved in the start site selection.

In this initial study the 3 bp from −1 to +2 of the *lac*UV5 promoter, that is one of the most studied *E.coli* promoters, were substituted by other bp, and the start sites of each variant were mapped to the single nucleotide resolution. Sl mapping (7), primer extension (8) and run-off transcription (9) have commonly been used to determine the transcription start sites, but one cannot pinpoint the exact 5′ nucleotides, since the heterogeneous bands can result from a single site initiation. On the other hand, the methods using base-specific RNases (10), nucleotide-specific pausing under limiting nucleotide concentrations (11) and chain terminators 3′-deoxyribonucleoside 5′-triphosphates (12–14) can map the exact start site. Here the start sites of the *lac*UV5 promoter variants were mapped by sizing the transcripts terminated by 3′-dCTP.

MATERIALS AND METHODS

Materials

The restriction endonucleases *Bam*HI, *Eco*RI, *Hin*dIII, *Sma*I and *Xmn*I were purchased from New England Biolab, Inc. and used according to the manufacturer's instructions. The deoxyribo-

*To whom correspondence should be addressed

nucleotides, ribonucleotides and 3'-deoxycytidine 5'-triphosphate (3'-dCTP) were purchased from Boehringer Mannheim. The *Taq* DNA polymerase and the Klenow fragment of *E.coli* DNA polymerase I were purchased from Promega Corporation. The *E.coli* RNA polymerase holoenzyme prepared as previously described (15) was a gift from Prof Akira Ishihama in National Institute of Genetics, Japan.

Construction of *lac*UV5 promoter variants using polymerase chain reaction

The 1.4-kb DNA fragment containing the *lac*UV5 promoter was isolated from digestion of the plasmid pMLUV5(I) with *Xmn*I and *Hin*dIII and was used as a template for polymerase chain reaction. Plasmid pMLUV5(I) has an insert of 53-bp synthetic *lac*UV5 promoter fragment (from −49 to +4) at the *Sma*I site of pUC119 (16). A complementary upstream primer and three sets of mismatched downstream primers (shown in Fig. 1) were synthesized in 0.2-μmole scale by a DNA synthesizer Gene Assembler Plus (Pharmacia LKB Biotechnology) to introduce single bp substitutions at the 3 positions −1, +1 and +2 of the *lac*UV5 promoter.

Polymerase chain reactions (17) were performed using the above mismatched primers in 500-μl microcentrifuge tubes by a thermal cycler (Hook and Tucker Instrument, Ltd.). Samples of 100 μl containing 10 ng of template DNA, 1.0 μM each primer, 200 μM each dNTP and 2 units of *Taq* DNA polymerase in the standard buffer (10 mM Tris−HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$ and 100 μg/ml gelatin) were subjected to 30 cycles of amplification. Each cycle consisted of denaturation at 94°C for 1 min, annealing at 51°C for 1.5 min and extension at 72°C for 1 min. Extension time of the last cycle was 10 min to ensure complete synthesis of the last strands. The completed reactions were extracted twice with an equal volume of phenol and chloroform and the DNA was concentrated by ethanol precipitation.

As shown in Figure 2, the amplified products were digested with *Eco*RI, and the fragments containing promoter variants were inserted into *Eco*RI/*Sma*I site of pUC19. All the expected promoter variants were identified by sequencing 65 clones using the Sanger's dideoxynucleotide method (18). All these plasmids of pLB series contain two adjacent C's at the +16 and +17 positions and these are the first two C's in the transcribed region. In order to utilize sizing of the cytidine-specific terminated transcripts for mapping the initiation sites, the two C's needed to be separated. Therefore each plasmid was digested with *Bam*HI, and its protruding ends were filled by the Klenow fragment followed by self-ligation. The newly constructed plasmids of the pLC series (Fig. 2) have an insert of GATC and a *Cla*I site in place of the *Bam*HI site. All the DNA manipulations were carried out by the methods described by Sambrook *et al.* (19).

Determination of *in vitro* transcription initiation sites

All the plasmids of the pLC series were digested with *Nde*I and *Hin*cII, and the 304-bp fragments containing the promoter region were used as transcription templates. The transcription reaction was performed under the standard single-round reaction conditions as previously described (20) but with some minor modifications. A 10-μl reaction mixture containing 0.5 pmole of the template DNA and 5-fold molar excess of the *E.coli* RNA polymerase holoenzyme in the standard reaction buffer (50 mM Tris−HCl, pH 7.6, 3 mM magnesium acetate, 0.1 mM EDTA, 0.1 mM dithiothreitol, 25 g/ml nuclease-free bovine serum albumin and 50 mM NaCl) was preincubated at 37°C for 10 min. Transcription was initiated by adding 15 μl of prewarmed substrate mixture containing 10 units of RNasin and heparin in the standard reaction buffer. The final concentration of each NTP and heparin was 160 μM and 200 μg/ml, respectively. In the nucleotide-specific pausing experiments 160 μM 3'-dCTP was

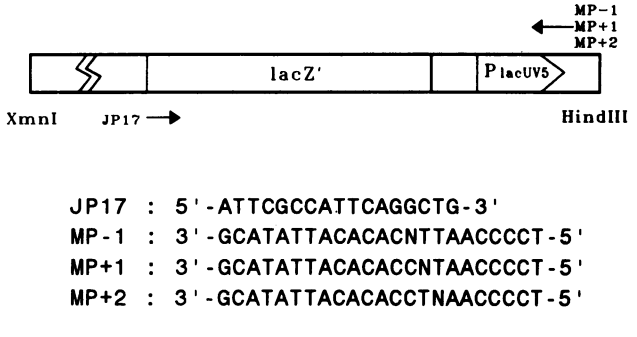

Figure 1. Construction of *lac*UV5 promoter variants by polymerase chain reactions. The upstream, forward primer (JP17) contains a sequence matching with a part of *lacZ'* (located at −220 to −204 from the major start site +1 of the inserted *lac*UV5 promoter) and the 3 different downstream, reverse primers (MP−1, MP+1 and MP+2) corresponding to the sequence from +9 to −14 contain an equal mixture of A, C, G and T in the position N.

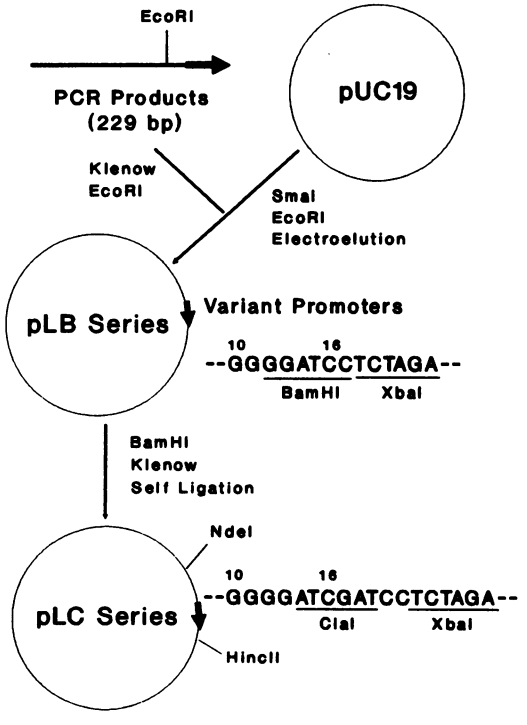

Figure 2. Construction of transcription template plasmids. The PCR amplification products of 229 bp were treated with the Klenow fragment to flush the ends before digested with *Eco*RI. After they were inserted into the *SmaI*/*Eco*RI site of pUC19, the sequence GATC was inserted between +16C and +17C to separate the first two C's in the transcribed region.

used instead of CTP. For radioactive labeling of transcripts, 4 μCi [α-^{32}P]GTP (400 Ci/mmole), 10 μCi [γ-^{32}P]ATP (3,000 Ci/mmole) or 10 μCi [γ-^{32}P]GTP (5,000 Ci/mmole) was added to each reaction. RNA synthesis was allowed to proceed at 37°C for 5 min and stopped by adding 75 μl of the stop solution containing 20 mM EDTA and 150 g/ml *E.coli* tRNA. The radioactive transcripts were sized by denaturing polyacrylamide gel electrophoresis, and each radioactive band was quantified by scanning with a PhosphoImager (Molecular Dynamics).

RESULTS

Transcription initiation sites of the *lac*UV5 promoter

The prototype *lac*UV5 promoter used in this study contains the following sequence from −12 to +20; ...TATAATGTGTGG-AATTGGGGAGGGGATCGATC... (the conventional start site +1 being in bold face). It is identical with the original *lac*UV5 promoter (10) in the sequence from −49 to +5. In order to determine the transcription start sites of our prototype, the cytidine-specific terminated transcripts were precisely sized. Since 3′-dCTP was used instead of CTP, transcription would stop at an underlined position +16. This would result in production of a 16-mer RNA, if the transcription starts at the +1 site. In our *in vitro* transcription reactions with 3′-dCTP, the 15-mer, 16-mer, 17-mer, 19-mer, 20-mer and 21-mer RNAs were produced and labeled with [α-^{32}P]GTP from the prototype promoter as shown in Figure 3, lane 1. Neither longer nor run-off transcripts were produced, suggesting that 3′-dCTP rather efficiently stops transcription.

These results alone may incorrectly suggest that initiation occurs at all the positions from −5 to +2 except −2. If this were the case, when the C-specific terminated transcripts were labeled with [γ-^{32}P]ATP in order to label only the transcripts starting with ATP, only 15-mer and 16-mer should have been labeled. In our experiments, however, in addition to the 15-mer and 16-mer, 19-mer and 20-mer were also labeled as shown in Figure 3, lane 12. Likewise, the reactions with [γ-^{32}P]GTP labeled 17-mer and 21-mer (data not shown) instead of the 17-mer, 18-mer and 20-mer that could have been expected from the initiations at the positions −1G, −2G and −4G. Thus, the 15-mer, 16-mer and 17-mer must have resulted from the transcriptions starting at +2A, +1A and −1G, respectively, and ending at the first C position +16. Likewise, the 19-mer, 20-mer and 21-mer resulted from initiations at +2A, +1A and −1G, respectively, and ending at the second C position +20. The occasional elongations at the first C position might be due to residual presence of CTP in the RNA polymerase preparation and/or DNA preparation, as previously observed (11), or to misincorporation of other nucleotides. The results indicate that transcription starts not only at +1 but also at −1 and +2 from the *lac*UV5 promoter. Quantification of all the C-specific terminated transcripts by radioactivity scanning with a PhosphoImager showed that 55% of the transcripts started at +1A, 23% at −1G and 22% at +2A (Table 1).

Small abortive initiation products approximately upto 9-mer were also labeled as expected (data not shown). The frequency of abortive initiation could vary depending on the initiation site. This can be examined only by sequencing all the abortive initiation products. In this study, however, only the productive transcripts that have escaped from abortive initiation cycling were analyzed. Thus all our results indicate the initiation sites of productive transcriptions. These might not be much different from those of all the initiations including both productive and abortive ones, because the abortive initiation frequency was not much to influence the determination of major initiation sites, except in the C-specific termination experiments with the mutants containing a C in the initiation region (explained below).

Transcription initiation sites of the *lac*UV5 promoter variants

In the prototype *lac*UV5 promoter, transcription starts with only three purines GAA from −1 to +2 (for convenience this region will be called the initiation region). In order to see the effects

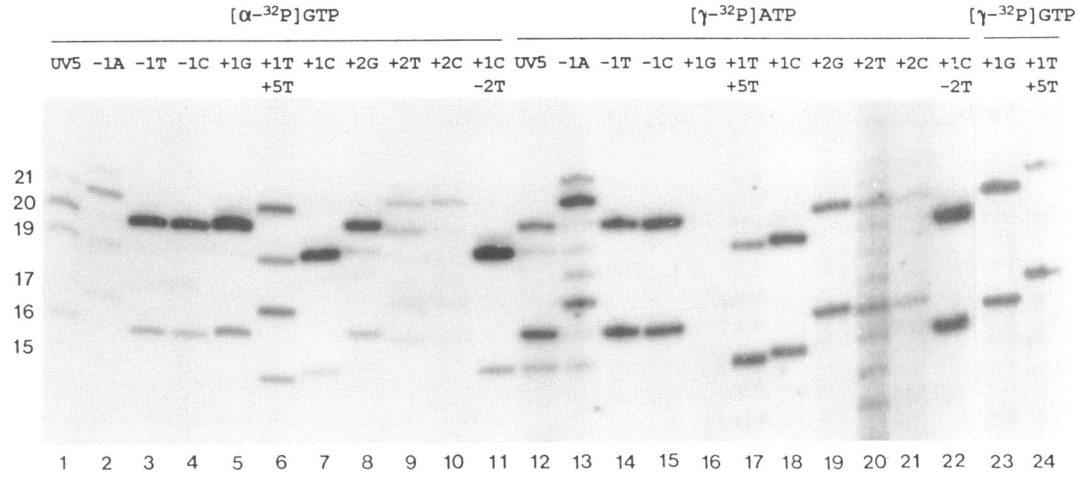

Figure 3. Sizing of the C-specifically terminated transcripts separated by 20% denaturing polyacrylamide gel electrophoresis. In place of CTP, 3′-dCTP was used in single-round *in vitro* transcription reactions of plasmids pLCUV5 (lanes *1* and *12*), pLC−1A (*2* and *13*), pLC−1T (*3* and *14*), pLC−1C (*4* and *15*), pLC+1G (*5*, *16* and *23*), pLC+1T+5T (*6*, *17* and *24*), pLC+1C (*7* and *18*), pLC+2G (*8* and *19*), pLC+2T (*9* and *20*), pLC+2C (*10* and *21*), and pLC+1C−2T (*11* and *22*). The transcripts were labeled with either [α-^{32}P]GTP (lanes *1* to *11*), with [γ-^{32}P]ATP (lanes *12*−*22*), or with [γ-^{32}P]GTP (lanes *23* and *24*). The transcript sizes are shown in the leftmost. Although some artifactual bands produced probably due to contaminating impurities in sample preparations are shown in lane *20*, they disappeared in other experiments and did not affect the determination of initiation sites.

of base pair substitution in the initiation region on the start site selection, all 9 single substitution mutations were introduced in the region using polymerase chain reactions and each variant's start sites were mapped as described in the above. Each of the three purine positions was substituted by the other three bases. The resulting 9 variants can be grouped into 3 T-containing variants, 3 C-containing ones and 3 purine-containing ones.

Variants with a T in the initiation region. The 3 T-containing variants were transcribed *in vitro* in the presence of 3'-dCTP and the C-specific terminated transcript RNAs were sized. When the −1G was changed to T (pLC−1T with TAA in the initiation region), the transcripts labeled with [α-^{32}P]GTP were 16-mer and 20-mer, as shown in Figure 3, lane 3. The same 16-mer and 20-mer were also labeled with [γ-^{32}P]ATP (Fig. 3, lane 14). These results indicated the +1A to be the only start site (Table 1). Thus, transcription of pLC−1T does not initiate with the pyrimidine T at −1 and that it does with A at +1 but not at +2.

When the +2A was changed to T (pLC+2T with GAT), the transcripts labeled with [α-^{32}P]GTP were 16-mer, 17-mer, 20-mer and 21-mer (Fig. 3, lane 9), suggesting the start sites to be −1G and +1A but not +2T. Phosphoimage scanning of the lane 9 revealed the approximately equal chances of initiation at the two purine positions (Table 1). The initiating nucleotides were confirmed by labeling the transcripts with either [γ-^{32}P]ATP (lane 20) or [γ-^{32}P]GTP (data not shown).

The start site of the variant with +1T (pLC+1T with GTA) has previously been reported by us to be only +2A (14). Initiation occurs with neither T at +1 nor G at −1. On the other hand, the start sites of a mutant (pLC+1T+5T) carrying double substitutions at +1 and +5, which was unpurposely obtained from the polymerase chain reactions in this study, were mapped to be not only +2A but also −1G. The transcripts from pLC+1T+5T, labeled with [α-^{32}P]GTP, were 15-mer, 17-mer, 19-mer and 21-mer (Fig. 3, lane 6). The 15-mer and 19-mer were labeled with [γ-^{32}P]ATP (lane 17), and the 17-mer and 21-mer were labeled with [γ-^{32}P]GTP (lane 24). The relative radioactivities of the four bands indicate that the initiation frequencies at the two positions are approximately the same (Table 1). Thus, the second mutation of G to T at +5 (in addition to a mutation of A to T at +1) derepressed the initiation at −1G.

Variants with a C in the initiation region. The 3 promoter variants containing a cytidine within the initiation region (pLC−1C, pLC+1C and pLC+2C) were also assayed for their start sites using a chain terminator 3'-dCTP, as described above. This assay, however, cannot *a priori* detect all the initiations at and before the substituted C positions. Therefore, it was supplemented in each case by sizing the run-off transcripts produced in the absence of 3'-dCTP from the templates that were linearized by digestion with *Nde*I and *Hin*cII. When the transcription starting at the conventional +1 site terminates at the very end made blunt by *Hin*cII, the transcript RNA labeled with [α-^{32}P]GTP should be 30-mer. However, the transcription from a promoter that was shown to have a unique start site, for example pLC−1T with

Table 1. Transcription start sites of the *lac*UV5 promoter variants

Plasmids	Initiation sites[a]	Initiation probabilities (%)
	−2 +5	
pLCUV5	GGAATTG	G_{-1}(23), $\mathbf{A}_{+1}$(55), A_{+2}(22)
pLC+2T	GGAtTTG	G_{-1}(51), A_{+1}(49)
pLC+1T	GGtATTG	$\mathbf{A}_{+2}$(100)
pLC+1T+5T	GGtATTt	G_{-1}(53), A_{+2}(47)
pLC−1T	GtAATTG	$\mathbf{A}_{+1}$(100)
pLC+2C	GGAcTTG	$\mathbf{A}_{+1}$(100)
pLC+1C	GGcATTG	G_{-1}(11), $\mathbf{A}_{+2}$(89)
pLC+1C−2T	tGcATTG	$\mathbf{G}_{-1}$(64), A_{+2}(36)
pLC−1C	GcAATTG	$\mathbf{A}_{+1}$(100)
pLC+2G	GGAgTTG	$\mathbf{A}_{+1}$(83), G_{+2}(17)
pLC+1G	GGgATTG	$\mathbf{G}_{+1}$(100)
pLC−1A	GaAATTG	$\mathbf{A}_{-1}$(70), A_{+1}(8), A_{+2}(22)

[a]The sequence of each variant is shown from −2 to +5. The variation from the prototype *lac*UV5 promoter are shown in lower bold faces. The relative initiation frequency of each start site is indicated by height of the filled bar shown above each position.

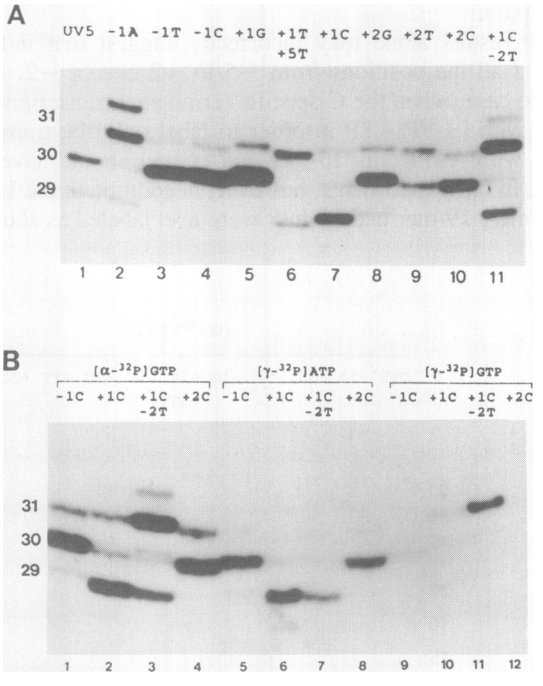

Figure 4. Sizing of the run-off transcripts from the promoter variants separated by 20% denaturing polyacrylamide gel electrophoresis. (**A**) Transcription was performed with plasmids indicated at the top of lanes that were digested with *Nde*I and *Hin*cII, and the transcripts were labeled with [α-^{32}P]GTP. The transcription from +1 to the template end should produce 30-mer RNA, that is shown as a predominant band in lane *3*. (**B**) The promoter variants carrying a C substitution were transcribed with either [α-^{32}P]GTP (lanes *1*−*4*), [γ-^{32}P]ATP (*5*−*8*) or [γ-^{32}P]GTP (*9*−*12*). The plasmids pLC−1C (*1*, *5* and *9*), pLC+1C (*2*, *6* and *10*), pLC+1C−2T (*3*, *7* and *11*), and pLC+2C (*4*, *8* and *12*) were digested by *Nde*I and *Hin*cII. The sizes of transcripts are shown in the leftmost.

+1A start, produced 29-mer and 31-mer bands in addition to the predominant 30-mer band (Fig. 4A, lane 3). The same bands were labeled with [γ-^{32}P]ATP in the same relative intensities, but they disappeared when labeled with [γ-^{32}P]GTP. Thus, the triple band pattern must have been caused by complexity of the termination step (9) and only the predominant band reflects the termination at the very end of the template. This was also confirmed with other promoter variants (Fig. 4A).

The pLC−1C promoter variant (with CAA) produced 16-mer and 20-mer bands when the 3′-dCTP terminated transcripts were labeled with either [α-^{32}P]GTP (Fig. 3, lane 4) or [γ-^{32}P]ATP (lane 15). However, no transcripts were labeled by [γ-^{32}P]GTP. The predominant run-off transcript was 30-mer (Fig. 4B, lane 1) and it was labeled with [γ-^{32}P]ATP (lane 5) but not with [γ-^{32}P]GTP (lane 9). All these results indicate that the start site of pLC−1C is uniquely +1A.

The variant with +1C (pLC+1C, GCA) produced the C-specific 15-mer and 19-mer bands when labeled with [α-^{32}P]GTP (Fig. 3, lane 7). These were labeled also with [γ-^{32}P]ATP (lane 18), but not with [γ-^{32}P]GTP. These results indicated that the +2A is a start site, and show consistency with run-off transcription results, as the 29-mer band was labeled with [γ-^{32}P]ATP (Fig. 4B, lane 6). However, +2A was not the only start site, because the 31-mer band was also labeled with [α-^{32}P]GTP (Fig. 4B, lane 2) and [γ-^{32}P]GTP (lane 10) as a run-off transcription product. Thus, the pLC+1C promoter starts mostly (89%) at +2A but also at −1G (11%) (Table 1). The initiation from −1G could have been artificially aborted at the next position +1C due to the presence of 3′-dCTP.

This abortion was also observed with the double mutation variant pLC+1C−2T that was a side product of polymerase chain reaction. The C-specific terminated transcripts 15-mer and 19-mer (Fig. 3, lane 11) were labeled only with [γ-^{32}P]ATP (lane 22), indicating that the +2A is a start site. It is consistent with the results that a major run-off transcript 29-mer (Fig. 4B, lane 3) was labeled by [γ-^{32}P]ATP (lane 7). However, a more dominant transcript 31-mer was also produced by run-off transcription (lane 3) and labeled by [γ-^{32}P]GTP (lane 11). Thus, this variant starts also at −1G. Compared with the pLC+1C, the major start site shifted from +2A (89% in pLC+1C) to −1G (64% in pLC+1C−2T) (Table 1).

In the case of pLC+2C (with GAC), the results with 3′-dCTP (Fig. 3, lanes 10 and 21) were not taken into account because of the abortions at the +2C position. The run-off transcription products labeled with [α-^{32}P]GTP (Fig. 4B, lane 4) were also labeled with [γ-^{32}P]ATP (lane 8), and the 30-mer was predominant. But they were not labeled with [γ-^{32}P]GTP (lane 10). Thus the start site of pLC+2C appears to be mostly, if not uniquely, +1A.

Variants with a substituted purine in the initiation region. The start sites of the promoter variants, in which an initiating purine was changed to the other purine, were also mapped by transcription reactions with 3′-dCTP. The pLC+1G variant with GGA in the initiation region produced 16-mer and 20-mer (Fig. 3, lane 5) that were labeled by [γ-^{32}P]GTP (lane 23), but not by [γ-^{32}P]ATP (lane 16). These results indicate that transcription of the pLC+1G promoter starts at a single site +1G (Table 1).

When +2A was changed to G (pLC+2G with GAG), transcripts labeled with [α-^{32}P]GTP were 15-mer, 16-mer, 19-mer and 20-mer (Fig. 3, lane 8). The 16-mer and 20-mer RNAs were labeled with [γ-^{32}P]ATP (lane 19), while 15-mer and 19-mer RNAs were labeled with [γ-^{32}P]GTP (data not shown). Thus, transcription from the pLC+2G promoter starts primarily at +1A (83%) and also at +2G (17%) (Table 1).

When −1G was changed to A (pLC−1A with AAA), a series of RNAs from 15-mer to 22-mer were labeled with [α-^{32}P]GTP (Fig. 3, lane 2) and [γ-^{32}P]ATP (lane 13). None of them were labeled with [γ-^{32}P]GTP. Labeling of the 18-mer and 22-mer with [γ-^{32}P]ATP instead of [γ-^{32}P]GTP indicate that those were not produced by initiation at −2G. This can be explained by the possible reiterative copying or stuttering of RNA polymerase during initiation, which was previously observed with *E.coli* RNA polymerase (21, 22), and phage SP6 (23) and T7 RNA polymerases (24). This happens when the initial sequence is AAA or longer oligo(A) (or GGG for T7 polymerase). On the other hand, the sequence AA does not allow slippage (22), which is also shown in cases with our pLC−1T and pLC−1C. Thus, the transcription of this variant appears to start at the three positions from −1A to +2A, like the prototype promoter. Since the sequence of the initiation region of pLC−1A is AAA, the trimer pppApApA transcribed from −1 and the dimer pppApA from +1 can, after slipping one bp back, be annealed to the −1 and +1 positions through 2 A:T base pairs and extended with another A at +2 (21, 22). The initiation frequencies at the 3 positions (Table 1) and the slippage efficiency (approximately 20%) were calculated from radioactivities of the 2 sets of 4 bands. The results suggested that the major start site shifted to −1.

DISCUSSION

Transcription of the *lac*UV5 promoter used in this study initiated *in vitro* with 3 purines −1G, +1A and +2A, when the +1 being the major start site. As an initial study towards understanding the mechanism of start site selection by *E.coli* RNA polymerase, the base pairs at the 3 initiation sites of the *lac*UV5 promoter were substituted individually by all the other base pairs, and the start sites of each variant were mapped to the single nucleotide resolution (Table 1).

All the substitutions, except one, caused changes in the number of start sites, from 3 to 2 or 1, and even in the exception case (pLC−1A) the major start site changed to −1. However, none of the substitution variants started outside the 3-bp region. Another common feature is that the substituted pyrimidine was never chosen as an initiating nucleotide in these promoter variants. It has been previously reported that some *E.coli* promoters (for example, *htpR*-P1) start transcription with pyrimidines, although a purine is available right next to them (25). When any one of the 3 initiating purines of the *lac*UV5 promoter was changed to a pyrimidine, the initiation at the substituted position was entirely suppressed.

Thus, a purine-to-pyrimidine change results in initiation at the other purine positions. For example, the variant carrying +1C started with −1G and +2A, and the variant of +2T with −1G and +1A. However, some of the substitutions suppressed also the initiation at one of the other 2 purine positions, leaving the initiation at only one position (pLC−1T, pLC−1C, pLC+1T and pLC+2C). It is unclear what causes which of the 2 purines is selected. The purine-to-purine changes also resulted in shift of major initiation site (pLC−1A) and suppression of initiation at other sites (pLC+1G and pLC+2G).

All the single mutation results suggested that the sequence context in the initiation region from −1 to +2 of the *lac*UV5

promoter severely affects the start site selection. Also the base pairs at −2 and +5 affected the start site selection as shown in our double mutants (pLC+1T+5T and pLC+1C−2T). The relative frequencies of initiation at the 3 sites −1G, +1A and +2A of our prototype *lac*UV5 promoter (pLCUV5) were in the ratio of 23:55:22 (Table 1). On the other hand, Carpousis *et al.* (10) previously reported that the *lac*UV5 promoter starts at −1G, +1A, +2A and +5G to the ratio of 29:55:9:7. The two promoters are identical in the sequence from −49 to +5 and the difference in the sequence downstream from +6 appears to be responsible for the suppression of initiation at the +5G and the increase in initiation at +2A in our pLCUV5. It was previously reported that deletions up to +9 in the *lac* promoter shifted the start sites (26). All the results indicate that any change in the sequence of a rather broad range around the initiation site can strongly influence the start site selection by *E.coli* RNA polymerase.

Other *E.coli* promoter elements could also have effects on the start site selection. To our knowledge, however, the studies have been very limited. A double mutation including deletion of −37C and change of A to T at −8 of the *lac*UV5 promoter shifted the major start site from +1A to +5G (10). While some changes in sequence and length of the spacer between the −35 and −10 hexamer regions did not shift the start sites of phage λ P_{RM} promoter (9) but heavily affected the start site selection at synthetic promoters containing a pair of consensus hexamers (27, 28). Therefore, we are in the process of examining the effects of all the possible substitutions in the −35 and −10 regions of the *lac*UV5 promoter on the start site selection.

In this study, base pair substitutions in the initiation region of the *lac*UV5 promoter generated various new patterns of initiation sites: three- two- and one-start-site sequences. Especially, single-start-site promoter variants will be useful to synthesize elongation abortive products of defined sequences for studies on transcription elongation mechanism (29). Some single-start-site variants appeared to be more active than the multiple-start-site variants, although quantitative comparisons were not attempted in this study.

Changes in the sequence around initiation sites of some promoters have also shown effects on initiation efficiency. Some few-bp deletions around the start site of a phage SP6 promoter resulted in much reduced initiation with a C at the same position (11, 30). The insertion of a C between the −10 region and the major start site (G) of the *Salmonella typhimurium hisR* promoter resulted in initiation at two C's with the reduced efficiencies that may reflect the diminished capacity of RNA polymerase to initiate with pyrimidines (31). Although our results do not appear to support that the shift of initiation sites is simply due to avoidance of the polymerase from initiating with pyrimidines, it would be necessary to measure the initiation efficiencies of our variants in order to correlate them with the shifts of start sites.

ACKNOWLEDGEMENTS

We are grateful to Dr Akira Ishihama at National Institute of Genetics in Japan for *E.coli* RNA polymerase holoenzyme. This work was supported by grants from Korea Science and Engineering Foundation and Korea Advanced Institute of Science and Technology.

REFERENCES

1. von Hippel,P.H., Bear,D.G., Morgan,W.D. and McSwiggen,J.A. (1984) *Annu. Rev. Biochem.*, 53, 389−446.
2. McClure,W.R. (1985) *Annu. Rev. Biochem.*, 54, 171−204.
3. Rosenberg,M. and Court,D. (1979) *Annu. Rev. Genet.*, 13, 319−353.
4. Sørensen,K.I. and Neuhard,J. (1991) *Mol. Gen. Genet.*, 225, 249−256.
5. Hawley,D.K. and McClure,W.R. (1983) *Nucleic Acids Res.*, 11, 2237−2255.
6. Harley,C.B. and Reynolds,R.P. (1987) *Nucleic Acids Res.*, 15, 2343−2361.
7. Johnson,K., Parker,M.L. and Lory,S. (1986) *J. Biol. Chem.*, 261, 15703−15708.
8. Fisher,R.F., Brierley,H.L., Mulligan,J.T. and Long,S.R. (1987) *J. Biol. Chem.*, 262, 6849−6855.
9. Warne,S.E. and deHaseth,P.L. (1993) *Biochemistry*, 32, 6134−6140.
10. Carpousis,A.J., Stefano,J.E. and Gralla,J.D. (1982) *J. Mol. Biol.*, 157, 619−633.
11. Nam,S.-C. and Kang,C. (1988) *J. Biol. Chem.*, 263, 18123−18127.
12. Axelrod,V.D. and Kramer,F.R. (1985) *Biochemistry*, 24, 5716−5723.
13. Monforte,J.A., Kahn,J.D. and Hearst,J.E. (1990) *Biochemistry*, 29, 7882−7890.
14. Jeong,W. and Kang,C. (1991) *Mol. Cells*, 1, 345−349.
15. Fukuda,R., Iwakura,Y. and Ishihama,A. (1974) *J. Mol. Biol.*, 83, 353−367.
16. Kobayashi,M., Nagata,K. and Ishihama,A. (1990) *Nucleic Acids Res.*, 18, 7367−7372.
17. Innis,M.A., Gelfand,D.H., Sninsky,J.J. and White,T.J. (1990) PCR Protocols: A Guide to Methods and Applications. Academic Press, Inc., California.
18. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl Acad. Sci. USA*, 74, 5463−5467.
19. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular Cloning: A Laboratory Manual. 2nd ed, Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
20. Nomura,T., Fujita,N. and Ishihama,A. (1986) *Nucleic Acids Res.*, 14, 6857−6870.
21. Harley,C.B., Lawrie,J., Boyer,H.W. and Hedgpeth,J. (1990) *Nucleic Acids Res.*, 18, 547−552.
22. Guo,H.-C. and Roberts,J.W. (1990) *Biochemistry*, 29, 10702−10709.
23. Cunningham,P.R., Weitzmann,C.J. and Ofengand,J. (1991) *Nucleic Acids Res.*, 19, 4669−4673.
24. Martin,C.T., Muller,D.K. and Coleman,J.E. (1988) *Biochemistry*, 27, 3966−3974.
25. Landick,R., Vaughn,V., Lau,E., VanBogelen,R.A., Erickson,J.W. and Neidhardt,F.C. (1984) *Cell*, 38, 175−182.
26. Xiong,X., deLaCruz,N. and Reznikoff,W.S. (1991) *J. Bacteriol.*, 173, 4570−4577.
27. Jacquet,M.-A., and Reiss,C. (1990) *Nucleic Acids Res.* 18, 1137−1143.
28. Jacquet,M.-A., Ehrlich,R., and Reiss,C. (1989) *Nucleic Acids Res.* 17, 2933−2945.
29. Krummel,B. and Chamberlin,M.J. (1989) *Biochemistry*, 28, 7829−7842.
30. Kang,C. and Wu,C.-W. (1987) *Nucleic Acids Res.*, 15, 2297−2294.
31. Bossi,L. and Smith,D.M. (1984) *Cell*, 39, 643−652.