Essential Structure of *E. coli*. Promoter

Takashi AOYAMA*

A synthetic promoter having the perfect consensus sequences (TTGACA and TATAAT) at −35 and −10 regions was constructed, and by altering sequences of the promoter systematically, the correlation of the sequences and promoter function was investigated in an *in vitro* transcription system. It was demonstrated that not only the sequences but also the spacing of the two consensus sequences are the important determinant of promoter strength. It was also shown that the RNA start point is principally fixed by the distance from the −10 consensus sequence, and that the sequences around the start point exert influence not the strength but the salt and temperature dependencies of promoter. In the next step, alternating CG sequences were inserted in the upstream of the promoter, and influence of negative helical turns on promoter function was analysed. It was found that supercoiling induced conformational change in the CG sequence strongly blocks function of the close-neighboring promoter, while such inhibitory effect was not observed in the relaxed state. These results provided evidence that promoter function is determined by not only the primary sequence but also the three dimensional conformation, and that at least the −35 region is required to retain the B-form helix for the binding of RNA polymerase to promoter. On the basis of the above observations, the structure essential for promoter function was discussed.

KEY WORDS: Synthetic promoter/ Transcription/ Promoter strength/ DNA supercoiling/ Z-DNA/

INTRODUCTION

A grate deal of information on the sequences of *E. coli* promoters have been accumulated.[1-5] The sequence comparison has revealed that promoter contains two regions of homology about 10 and 35 base pairs prior to the start point of transcription.[2,3] The most conserved base pairs at the −35 region are TTGACA and those at the −10 region are TATAAT (consensus seuqences). It has also been shown that RNA polymerase contacts with promoter at these two regions.[2] Nevertheless, the sequences of naturally occurring promoters reveal considerable diversity, and none of them contain both the consensus sequences in the corresponding regions.[3] The distance between the two consensus sequences also varies by several base pairs.[2,3] At present, information is still not enough to predict the site and strength of promoter from a given sequence. More systematical analysis is required on the correlation between the structure and function of promoter. For this purpose, I have constructed a promoter having the perfect consensus sequences at the −35 and −10 regions starting from the *tet* gene promoter of pBR322, and altered its structure systematically, and the activities of modified promoters were compared in an *in vitro* transcription system. The structures altered are mainly the spacing be-

---

* 青山卓史 : Laboratory of Molecular Biology, Institute for Chemical Research, Kyoto University, Uji, Kyoto-fu 611.

tween the two consensus sequences,[6] and the sequence around the RNA start point.[7] In the next, alternating CG sequences which can form the Z DNA structure in the supercoiling state were placed in the region upstream of the -35 consensus sequence,[8] and influence of negative helical turns was analysed in an *in vitro* system. Based on the results of these experiments, I deduced the structures which determine the function and strength of *E. coli* promoter.

MATERIALS AND METHODS

Preparation of plasmids and restriction fragments

E. coli C600, HB101 and GM33 (*dam⁻*) were used as recipients. For preparation of plasmid DNA, the sodium-SDS method[6] or the alkaline-SDS method[10] were used. After RNase-treatment the plasmid fraction was prepared by passage through a short agarose column (Bio Rad A5m). The preparation was then fractionated by CsCl-ethidium bromide density-gradient centrifugation[11] for isolation of covalently closed circular DNA(cccDNA). When used cccDNA as a template in the *in vitro* transcription experiments, the cccDNA fraction was purified by two cycles of the density gradient centrifugation. Restriction fragments were prepared as follows. Plasmid DNAs digested by restriction enzymes were resolved by 5% polyacrylamide gel electrophoresis, and band regions identified by ethidium bromide staining were cut out, from which DNA fragments were extracted, followed by n-butanol treatment.

Introduction of negative linking differences

Negative linking differences were introduced into plasmids by treatment of plasmids with calf thymus topoioosmerase I in the presence of different amounts of ethidium bromide.[12] The resulting topoisomers were resolved by agarose gel electrophoresis in the presence of chloroquine, and the superhelical densities were determined by the band-counting method of Keller.[13]

In vitro synthesis of RNA

Unless otherwise noted, standard reaction mixtures used for RNA synthesis contained 50 mM tris-HCl (pH 7.9), 8 mM $MgCl_2$, 150 mM KCl, 0.1 mM dithiothreitol, 0.1 mM (α-^{32}P)UTP and 0.4 mM each three other NTPs. The total volume of reaction mixture were 40 to 100 μl, and 0.2 to 1 pmol of template DNA and 2 to 10 pmol RNA polymerase were used depending on the scale of experiments. RNA polymerase used was the product of New England Biolabs. The standard reaction conditions used for measurement of the rate of RNA synthesis were as follows. Two sets of solutions, one containing template DNA and the other containing RNA polymerase were prepared. After pre-incubation for 5 min at the reaction temperature, generally 37°C, the solutions were mixed and incubated for indicated periods. Reaction was stopped by phenol treatment. Two volumes of ethanol was added to the aqueous layer and after holding for 30 min at -70°C, precipitates were collected by centrifugation.

Preparation of in vivo transcripts

E. coli HB101 cells harboring plasmids were grown to $A_{550}=0.3$ in 5 ml of L-

broth. The cells were harvested and suspended in 0.3 ml of 20 mM sodium acetate (pH 5.5)–0.5% SDS-1 mM EDTA. An equal volume of phenol equilibrated with 20 mM sodium acetate (pH 5.5) was added, and the mixture was incubated at 60°C for 5 min with gentle shaking. The aqueous layer was once re-extracted with phenol. Three volumes of cold ethanol was added, and after holding for 30 min at −70°C, precipitates were collected by centrifugation. The precipitates were dissolved in 0.3 ml of the same acetate/SDS buffer. The ethanol precipitation was repeated two more times. The final products were dissolved in 0.1 ml of distilled water, and 10 μl aliquots were used per assay.

S1 mapping

Analysis was essentially performed by the method of Berk and Sharp.[14] Synthesized RNA were mixed with probe DNA (about 1.0×10^4 cpm) in 20 μl of the hybridization buffer containing 80% formamide, 20 mM Hepes (pH 6.5) and 0.4 M NaCl. After incubation for 10 min at 75°C, the solution was slowly cooled to 37°C, and held for 2 hrs. 20 μl of ×10 S1 buffer containing 0.4 M sodium acetate (pH 4.5), 1 M NaCl and 40 mM ZnCl$_2$, and 160 μl of distilled water were added, and the solution was incubated for 20 min at 20°C with 300 units of S1 nuclease. Reaction was terminated by adding ethanol, and the resulting precipitates were dissolved in 80% formamide, containing 10 mM EDTA and marker dyes, and electrophoresed on 8% polyacrylamide sequencing gel.

Detection and quantitative analysis of transcripts

Labeled run-off transcripts of relevant promoters were detected by using 8% polyacrylamide sequencing gels, and their radioactivities were determined by Cerenkov counting of the respective bands on gels.

Measurement of the amount of RNA synthesized on plasmid DNA *in vitro* and RNA transcribed in *E. coli* cells were done as follows. *In vitro* or *in vivo* transcripts were mixed with an excess of a terminal labeled DNA fragment containing the promoter, and after hybridization, the amounts of hybrids protected against S1 nuclease were measured. The conditions for hybridization and S1 nuclease-digestion were essentially identical to those of S1-mapping.[14] The *Hap*II fragment indicated on the plasmid circle in Fig. 9 contains two promoters in the counter direction; one is the synthetic promoter and the other the *bla* gene promoter. Thus, both 5′ termini of this fragment of each plasmid were labeled with (γ-^{32}P)ATP, and specific activities were adjusted to 1.0×10^6 cpm/pmol by adding the non-labeled fragment. The *in vitro* products obtained in the indicated reaction conditions or *in vivo* products obtained from 0.5 ml cultures were hybridized with 0.8 pmol of the labeled DNA fragment, and after digested with S1 nuclease, the products were analysed by sequencing gels. Respective bands corresponding to the transcription from relevant promoters were cut out, and radioactivities were measured by Cerenkov counting.

Other methods

DNA sequencing was carried out by the Maxam-Gilbert method[15] or the dideoxy method.[16-18] Conditions for T4 DNA polymerase reaction, S1 nuclease digestion and ligation have been described elsewhere.[6,7]

RESULTS

Construction of a promoter having the perfect consensus sequences, and its derivatives having altered spacing

The *tet* promoter of pBR322 contains TTGACA at the −35 region, TTTAAT at the −10 region and a *Hind*III site just before the −10 region.[19]   It is therefore possible to replace the −10 region with the consensus sequence by using the *Hind*III site.   However, a counter-directed promoter which affects analysis of promoter function is located just after the start point of transcription.[20]   Thus deletion was first introduced from the *Hind*III site to the position of the counter-directed promoter (about 25 base-pairs from the *Hind*III site) by stepwise exonucleolysis of T4 DNA polymerase followed by S1 nuclease-treatment.   The *Hind*III site was reconstructed by using a *Hind*III linker (dCCAAGCTTGG), and the upstream region was preserved.   The plasmid constructed (pDEL5) carries only the −35 region of the *tet* promoter (TTGACA) and confers ampicillin(Ap)-resistance, but not the tetracycline (Tc)-resistance, on *E. coli* cells by transformation.   In order to place the consensus sequence in the −10 region, the duplex of synthesized pentadecamers containing TATAAT was inserted into the *Hind*III site of pDEL5.   After transformation, plasmids were prepared from Tc^r-Ap^r transformants, and the one containing the consensus sequences both in the −35 and −10 regions was named pSPI57 (Fig. 1).

pSPI57 contains *Cla*I and *Hind*III sites between the two consensus sequences, so that the spacing was independently adjusted at these sites.   pSPI57 was linearized
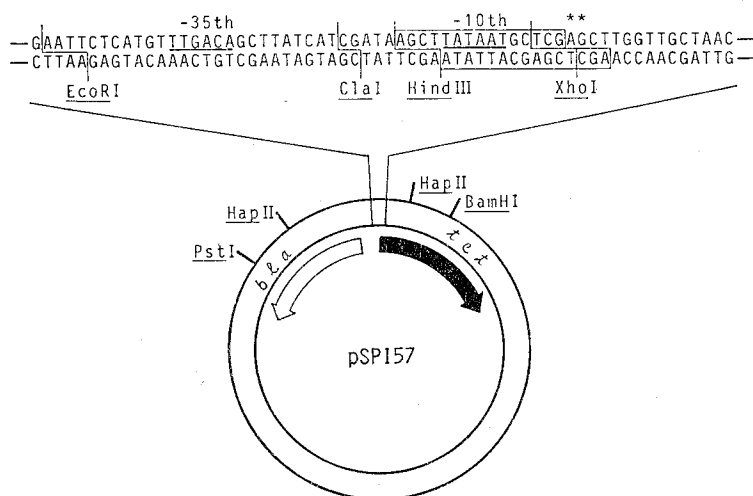

Fig. 1.   The sequence and restriction sites around the constructed promoter of
pSPI57.   The inserted pentadecamer-duplex was boxed, and consensus
sequences were underlined.   Asterisks indicate the RNA start points.
Location and direction of the *bla* promoter and constructed promoter
are shown on the circular map of pSPI57 together with the restriction
sites used for generation of the fragments which were used as templates
for *in vitro* transcription.

at either the *Cla*I or *Hind*III site, and the positions of 3′-ends were adjusted by exonucleolysis of T4 DNA polymerase. The ends were then converted to blunt end by S1 digestion, and circularized by ligation, and transformation was carried out. Plasmids were prepared from Ap^r transformants randomly selected, and the sizes of fragment generated by *Eco*RI and *Xho*I were determined by sequencing gels. As a result, two series of promoter with different spacer lengths were obtained. When assayed promoter function by Tc-resistance of transformants, the *tet* gene expression was observed with promoters having the spacer length of 16 to 18. The sequences of the constructed promoters with the spacer lengths of 15 to 19 were determined by the Maxam-Gilbert method[15] (Fig. 2).

Activity of the constructed promoters having different spacings

The constructed promoter was cut out together with the *bla* gene promoter by *Hap*II as a fragment of about 610 base-pairs (see Fig. 1), and transcription was carried out for 1 hr at 37°C with 1 pmol of the fragment and 10 pmol RNA polymerase in 0.1 ml of the standard transcription mixture. Products formed are shown in Fig. 3, in which RNA–1 corresponds to the transcript from the constructed promoter and RNA–2 to the transcript from the *bla* promoter, respectively. It is clear that the fragments from pPSH329 and pPSC7 both having the 17 bp-spacing yield very intense bands at the RNA–1 position (lane 3,8 in Fig. 3). As judged from the band intensity, the promoter activity appears to be much stronger than that of the *bla* gene promoter as an internal reference. In contrast, those having the spacer length of 16 bp or 18 bp gave only weak bands, and at the spacer length of 15 bp or 19 bp, no bands were observed. The promoter strength was compared with those of other promoter showed about the same strength as that of fd gene X promoter,[21] suggesting that it belongs to "high efficiency" promoters. The original *tet* promoter gave a band weaker than that of the constructed promoter with the 16 or 18 bp-spacing, even though it has the 17 bp-spacing. Those results reveal that not only the sequences in the −35 and −10 region but also their spacing are important factors for promoter strength.

RNA start points of promoters having different spacings

The start point of the RNA from the constructed promoter of pPSC7 was de-

Plasmid	Spacer length	−35th	−10th **
pPSH2	19 bp	GT[TGACA]GCTTATCATCGATAAGGCT	[TATAAT]GCTCGAGCTT
pPSC1	19	GT[TGACA]GCTTATCATCCGATAAGCT	[TATAAT]GCTCGAGCTT
pSPI57	18	TGT[TGACA]GCTTATCATCGATAAGCT	[TATAAT]GCTCGAGCTT
pPSH329	17	ATGT[TGACA]GCTTATCATCGATAGCT	[TATAAT]GCTCGAGCTT
pPSC7	17	ATGT[TGACA]GCTTATCACGATAAGCT	[TATAAT]GCTCGAGCTT
pPSH561	16	CATGT[TGACA]GCTTATCATCGATAAT	[TATAAT]GCTCGAGCTT
pPSC8	16	CATGT[TGACA]GCTTATCATATAAGCT	[TATAAT]GCTCGAGCTT
pPSH509	15	TCATGT[TGACA]GCTTATCATCGATAA	[TATAAT]GCTCGAGCTT
pBR322	17	ATGT[TGACA]GCTTATCATCGATAAGC	[TTTAAT]GCGGTAGTTT

Fig. 2　The sequences of constructed promoters with different spacer lengths. The consensus sequences are shown in boxes, and asterisks indicate initiation points of transcription. The sequence and RNA start point of the original *tet* promoter are shown for comparison.
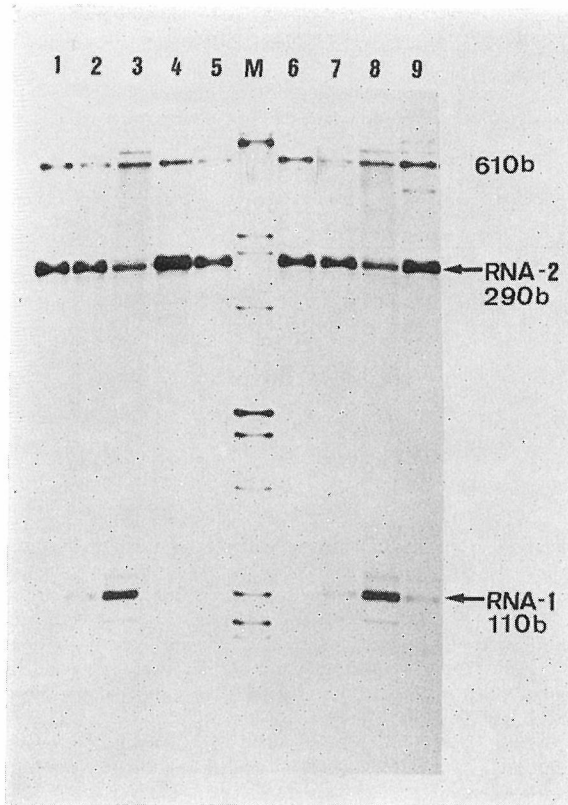
Fig. 3 Analysis of *in vitro* transcripts from promoters with different spacer lengths. The constructed promoter was generated together with the *bla* promoter by *Hap* II from respective plasmids listed in Fig. 2, and transcripts formed *in vitro* were analysed. Products from pPSH2, pSPI57, pPSH329, pPSH561, pPSH509, pPSC1, pSPI57, pPSC7 and pPSC8 are shown in lane 1 to 9, respectively. Lane M in the middle provides size markers. RNA-1 and RNA-2 indicate the positions of run-off transcripts from the constructed and *bla* promoters, respectively. Bands of about 610 bases were assigned to full-length transcripts from termini. Thin bands near the full-length transcripts are assumed to be hybrids of transcripts.

termined by S1 mapping in which RNA transcribed from the *Hae*III fragment (see Fig. 1) of pPSC7 was hybridized to the terminal labeled *Hap*II fragment (see Fig. 1). It was found that the promoter initiated transcription at either adenine or guanine residues, 6 or 7 bases downstream from the -10 consensus sequence respectively (data not shown). As the next step, the run-off transcripts from the *Hap*II fragments containing the promoters with different spacer lengths were electrophoresed on a sequencing gel for a period long enough to resolve one base difference. The obtained band patterns of transcripts from promoters with the 16 to 18 bp-spacing are the same as that of pPSC7, although their intensities were different (data not shown). The result indicates that the starting points of transcription were not affected by the position of the consensus sequence at the -35 region.

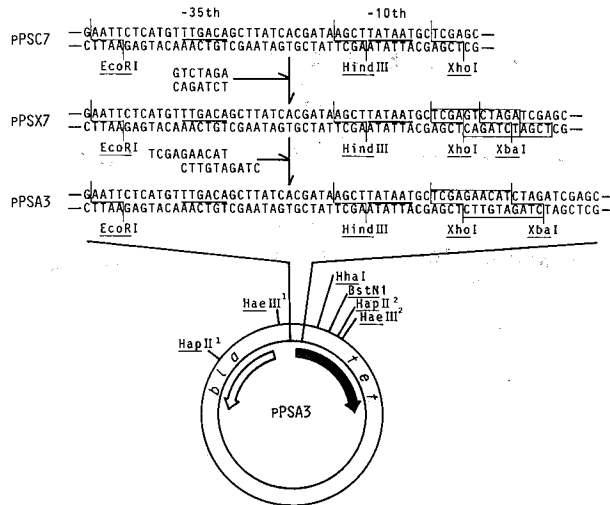*Construction of promoters having different sequences in RNA start points*

Fig. 4 Construction of promoters having different sequences in the downstream from the −10 consensus sequence. The promoter sequences, the insertion sites in the plasmid, and the steps of sequence modification are schematically shown. The −35 and −10 consensus sequences were indicated by underlines, and inserted sequences in derivatives were boxed. When constructed pPSG series, the indicated decamer-duplex was replaced by the other decamer-duplex (see text). Restriction sites indicated on the circular map of pPSA3 are those used for generation of fragments. The HapII[1]-BstN1 fragment and HaeIII[1]-BstN1 fragment were used as template for in vitro transcription, and the HhaI-HapII[2] fragment was used as primer for dideoxy-supercoil-sequencing.

To construct promoters having different sequences in RNA start points, the promoter carried by pPSC7 which has the two consensus sequences at an optimal distance (17 base-pairs) was altered according to the scheme shown in Fig. 4. pPSC7 was linearized by XhoI, and after repairing by T4 DNA polymerase, the synthetic duplex, dGTCTAGA:dTCTAGAC, was introduced by blunt-end ligation. To the resulting plasmid, pPSX7, having a XbaI site at the outside of the XhoI site, two kinds of synthetic duplexes, dTCGAGGGCAT:dCTAGATGCCC and dTCGA-

Plasmids	Sequences	Relative Activities
pPSA303	-AAGCT TATAAT GCTCGCGAGAACATCT-	10.5
pPSA3	-AAGCT TATAAT GCTCGAGAACATCTAG-	12.2
pPSA323	-AAGCT TATAAT GCTGAGAACATCTAGA-	11.6
pPSA213	-AAGCT TATAAT GCGAGAACATCTAGAT-	14.2
pPSA302	-AAGCT TATAAT CGAGAACATCTAGATC-	9.5
pPSA403	-AAGCT TATAAT GCGAACATCTAGATCG-	13.2
pPSG10	-AAGCT TATAAT GCTCGCGAGGGCATCT-	11.1
pPSG3	-AAGCT TATAAT GCTCGAGGGCATCTAG-	10.0
pPSG105	-AAGCT TATAAT GCTCAGGGCATCTAGA-	7.2

Fig. 5 Sequences, initiation points and relative activities of constructed promoters. The −10 consensus sequences were boxed, and start points of transcription were indicated by asterisks. Relative activities of promoters to the bla promoter are given in the right column.

GAACAT:dCTAGATGTTC, were inserted between the *Xho*I and *Xba*I sites and the plasmids obtained were named pPSG3 and pPSA3 respectively. The two promoters differ in GC-contents of the sequences around the RNA start points.

To change the sequences around the RNA start points further, insertion or deletion was introduced at the *Xho*I sites of pPSG3 and pPSA3 by the same procedure that used for alternating of spacing. As the result, a series of promoters having different sequences in the RNA start points were obtained (Fig. 5). These promoter have the CAT sequence, which frequently occurs in the RNA start points of natural promoters, at different distances from the −10 consensus sequences.

Activity and RNA start points of the constructed promoters with different sequences around the RNA start points

The constructed promoter was cut out together with the *bla* gene promoter by *Hap*II and *Bst*N1 as a fragment of about 600 base pairs, and the promoter activity
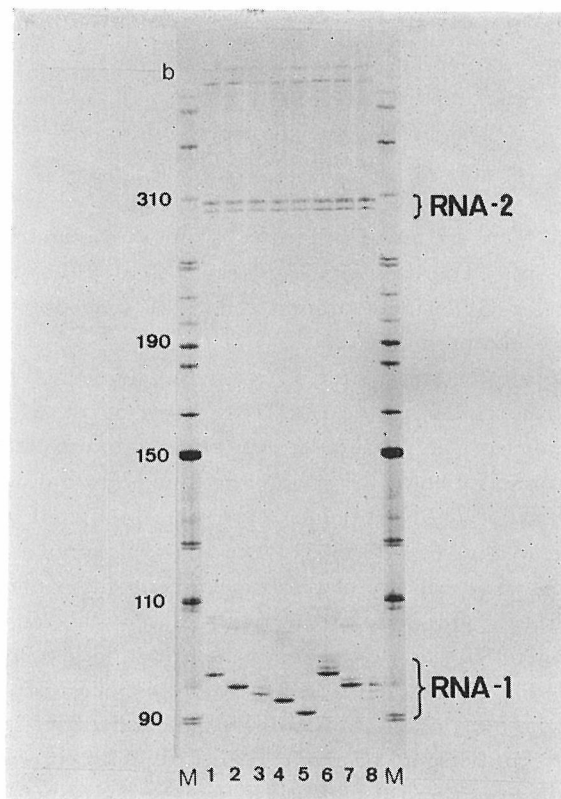

Fig. 6 Transcripts formed on *Hap*II¹-*Bst*N1 fragments from constructed plasmids. Products from pPSA303, pPSA3, pPSA323, pPSA213, pPSA302, pPSG10, pPSG3 and pPSG105 are shown in lanes 1 to 8 (see Fig. 5 for plasmid names). Marker fragments are shown in lane M. The band regions for RNA-1 (transcripts from constructed promoters) and RNA-2 (transcripts from the *bla* promoter) were indicated by the side of lanes. RNA-1 and RNA-2 are transcribed in different directions. A few bands seen above RNA-2 are read-back transcripts.

was determined in the *in vitro* transcription system. Transcription was carried out for 15 min at 37°C with 1 pmol of the fragment and 10 pmol RNA polymerase in 0.1 *ml* of the standard transcription mixture. Product synthesized are shown in Fig. 6, in which RNA–1 corresponds to the transcript from the constructed promoter and RNA–2 to the transcripts from the *bla* promoter, respectively. Analysis was performed by using 0.3 mm-thin 8% polyacrylamide sequencing gel, so that resolution was high enough to determine the product sizes by one-base length, and the relative positions of the RNA start points could be assigned on the template from the product sizes. Three bands having slightly different mobilities are seen at the RNA–2 position in ecah lane of Fig. 6. Analysis of truncated products indicated that all of them originated from a single site at the *bla* gene promoter, and appearance of three bands was attributed to the variation of the RNA stop position at the *Hap*II terminus. However, this was not the case for RNA–1, which was transcribed in the reverse direction. This is not surprising, for RNA stop positions are markedly influenced by stability of the helical structure at termini.[22]

Radioactivities of RNA–1 and RNA–2 in each lane were measured by Cerenkov counting of the bands, and the values of RNA–1 corrected for the U contents were compared by molar ratios relative to those of RNA-2. The result is given in the right column in Fig. 5. The levels of the constructed promoters were not significantly different, except for the one on pPSG105. This promoter differed from the pPSG3 promoter only by one base, but it gave a low value beyond the measurement errors. As discussed later, we assume that this is due to the presence of a G cluster in the RNA start point. The data in Fig. 5 reveal that neither the position of the CAT sequence nor the GC-content around the RNA start point bears significant relation to the level of the promoter.

In order to allocate the start point of transcription precisely, S1 mapping was carried out to the transcript from the *Hae*III fragment of pPSA303 (see Fig. 4) by using the terminal labeled *Hap*II fragment as a probe. It was determined that this promoter initiated transcription at the guanine residue, 7 base downstream from the −10 consensus sequence. The start points of the other promoters were then assigned by positions relative to that of pPSA303. The starting points assigned are shown by asterisk in Fig. 5. The validity of these assignment was confirmed by other experiments, in which only products starting with A were labeled by using $(\gamma\text{-}^{32}P)$ ATP (data not shown). In 7 out of 9 promoters, initiation occurred by Pu(purine) at the 7th position from TATAAT, regardless of the sequences around the initiation sites. Initiation of the remaining two from pPSA302 and pPSG105 occurred by Pu at the 8th and 6th positions, respectively. The pPSA302 promoter contains Py (pyrimidine) at the 7th position, so that it is likely that RNA starts at the neighboring Pu, if no Pu is present at the 7th position. Unexplained is the case of the pPSG105 promoter containing Pu at the 7th position. As mentioned in the previous section, activity of this promoter was also low. I assume that the abnormality is due to the occurrence of a G-cluster around the RNA start point.

In order to determine by which of the consensus sequences the starting point of transcription was determined, pPSA3 and pPSA303 were recombined with pSPI57

Plasmids	Sequences
pPSA303-I57	-GT TTGACA GCTTATCATCGATAAGCT TATAAT GCTCGCĠAGAA-
pPSA303	-TGT TTGACA GCTTATCACGATAAGCT TATAAT GCTCGCĠAGAA-
pPSA303-C8	-ATGT TTGACA GCTTATCATATAAGCT TATAAT GCTCGCĠAGAA-
pPSA3-I57	-GT TTGACA GCTTATCATCGATAAGCT TATAAT GCTCGAĠAACA-
pPSA3	-TGT TTGACA GCTTATCACGATAAGCT TATAAT GCTCGAĠAACA-
pPSA3-C8	-ATGT TTGACA GCTTATCATATAAGCT TATAAT GCTCGAĠAACA-

Fig. 7 Sequences and RNA start points of spacer-mutant promoters. The −35 and −10 consensus sequences were boxed, and the start points of transcription were indicated by asterisks.

and pPSC8 at the *Hin*dIII site between −35 and −10 regions, and two sets of spacer mutant promoters (Fig. 7) were constructed. Transcripts from these mutant promoter were analysed as in the experiment in Fig. 6. As indicated by asterisks in Fig. 7, RNA synthesis was initiated in the constant position from the −10 consensus sequence.

Effect of the sequences around the RNA start points on properties of promoter

Dependency of promoter function on temperature and ionic strength were compared between the promoter on pPSA3 and pPSG3 or between those on pPSA3 and pPSG10. The promoters in these plasmids initiate RNA synthesis at the 7th position with about the same efficiencies but the GC-contents of the sequences around the start points are significantly different each other (see Fig. 5). The *Hae*III-*Bst*N1 fragments carrying only the constructed promoter were generated from these plasmids and used as templates for RNA synthesis. The RNA synthesis was done in the 40 μl of the standard mixture containing 0.25 pmol of the DNA fragment and 5 pmol of RNA polymerase. For analysis of the salt effect, KCl concentrations
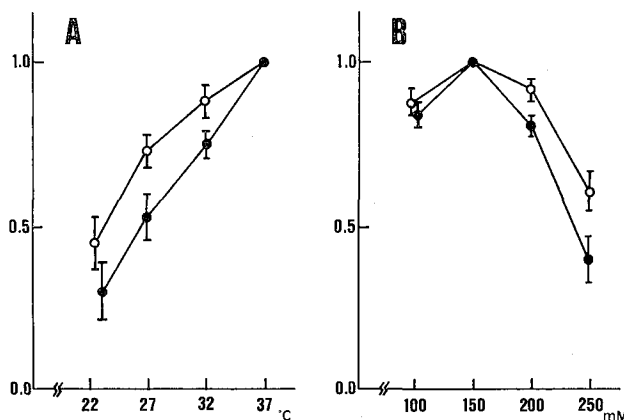
Fig. 8 Effects of temperature and salt-concentration on activities of promoters from pPSA3 (○) and pPSG3 (●). *Hae*III[1]-*Bst*N1 fragments were generated from these plasmids, and RNA synthesis was carried out either at indicated temperatures (A) or at indicated KCl concentrations (B) under the conditions described in the text. The data were compared by normalizing to the values at 37°C(A) or to the value at 0.15 M KCl (B). Vertical bars represent standard deviations in the experiments.

were altered. The reaction mixture was inucbated at 37°C (for the salt effect) or at different temperature (for the temperature effect) for 1 min, heparin was added at 0.1 mg/ml to prevent re-initiation, and RNA synthesis was allowed to continue at 37°C for another 15 min. The major band formed on the template was RNA–1 of about 95 bases long. Its relative amounts were estimated by Cernekov counting of respective bands on gels. In Fig. 8, the amounts of RNA formed for promoters on pPSA3 and pPSG3 at different temperatures (Fig. 8A) or at different salts (Fig. 8B) are shown by normalizing the values to those at 37°C or those at 0.15 M KCl, respectively. Compared with the promoter on pPSA3, that on pPSG3 seemed to be more sensitive to temperature and salts. A similar correlation was observed between the promoters on pPSA3 and pPSG10, and that of pPSG10 was more intensely depressed at low temperature and at high salts (data not shown).

Construction of promoters containing an alternating CG block upstream the −35 consensus sequence

To investigate the correlation between promoter strength and DNA conformation, alternating CG sequences, which can form Z-DNA structure under negative superhelical stress,[12,23] were inserted in the upstream region of the −35 consensus sequence. The promoter, PSA403 on pPSA403 (see the previous section) was used as the starting promoter. Since this promoter has a *Eco*RI site at the −44 position, the synthetic CG block was inserted by using this site. The alternating CG block of 18 nucleotides was annealed to form heterogeneous duplexes, and repaired in the T4 DNA polymerase reaction, and inserted to the *Eco*RI site of pPSA403 by repair
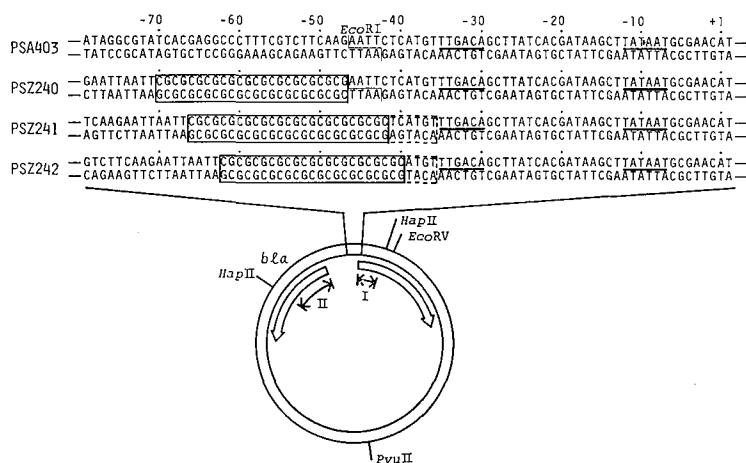

Fig. 9 Sequences of the synthesized promoters. Nucleotides were numbered from the RNA start point, and those in the upstream region were denoted by (−) sign. The −35 and −10 consensus sequences were underlined, and *Eco*RI sites used for CG insertion were indicated. The CG blocks and adjacent Pu-Py tracts are shown in solid and dotted boxes, respectively. The circle in the lower part illustrates plasmids, on which the transcripts form *bla* and synthetic promoters are shown by leftward and rightward arrows. Restriction sites indicated are those used in this study. Bands I and II are the regions of transcripts quantitatively analysed.

ligation. Among the resulting plasmids carrying different length of CG blocks, the one having $d(CG)_{12}$ was selected. Since this plasmid creates *Eco*RI sites at both sides of the CG block, the one at the distal end was destroyed, and by using the proximal one, the distance between the CG block and the -35 consensus sequence was adjusted by S1 nuclease-digestion. The sequences of resulting three promoters, PSZ240, PSZ241 and PSZ242 are shown in Fig. 9.

Evidence has been reported that short CG blocks in plasmids form the Z-form structure in pysiological conditions of ionic strength and superhelical density.[12,23] When the cccDNA fractions of plasmids carrying these promoters, pPSZ240, pPSZ241 and pPSZ242, were digested by restriction enzyme *Bss*HII, which recognizes d(GC-GCGC), these plasmids were hardly attacked. After conversion to the relaxed state by weak DNase I digestion or to the linear form by restriction enzyme digestion,

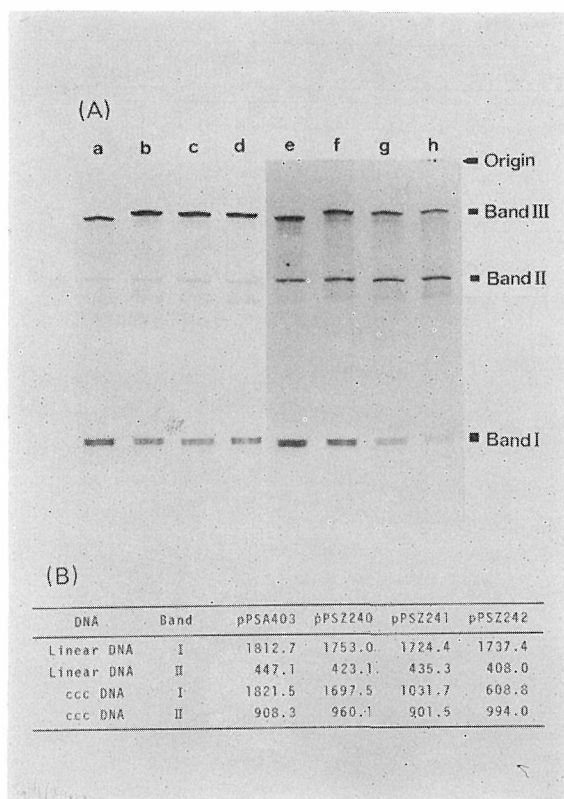| DNA	Band	pPSA403	pPSZ240	pPSZ241	pPSZ242
Linear DNA	I	1812.7	1753.0	1724.4	1737.4
Linear DNA	II	447.1	423.1	435.3	408.0
ccc DNA	I	1821.5	1697.5	1031.7	608.8
ccc DNA	II	908.3	960.1	901.5	994.0

Fig. 10 The amounts of transcripts *in vitro* formed on natively supercoiled and linear plasmid DNAs. RNA synthesis was carried out on the cccDNA fractions purified from cells and those opened at the unique *Pvu*II site, and the amounts of transcripts from the *bla* and synthetic promoters were determined by S1 mapping procedure. (A) Gel electrophoretic patterns obtained for linear (left four lanes) and supercoiled (right four lanes) DNAs (lanes a, e: pPSA403, lanes b, f: pPSZ240, lanes c, g: pPSZ241, lanes d, h: pPSZ242). Bands I and II correspond to probe DNA bands protected by transcripts from synthetic and *bla* promoters, respectively (see Fig. 9). (B) The radioactivities (c.p.m.) of bands, determined by Cerenkov counting method. Each value is the average of 2 independent experiments.

these DNAs were easily digested by *Bss*HII. The insensitivity was thought to be due to the conformational change of DNA from B-form to Z-form, as reported in other papers.[24]

Influence of the CG block on promoter activity in vitro

Transcription experiments *in vitro* were first carried out by using the natively supercoiled plasmids (the cccDNA fraction) and those linearized at the unique *Pvu*II site as control. The *in vitro* transcription was done for 3 min at 37°C in 50 μl of standard reaction mixture containing 0.2 pmol of plasmid DNA and 2 pmol of RNA polymerase, and the amount of RNA from the synthetic promoter were determined as in the method section. Upon resolution of the terminal-labeled probe DNA protected by hybrid formation with transcripts, three major bands were obtained (Fig.
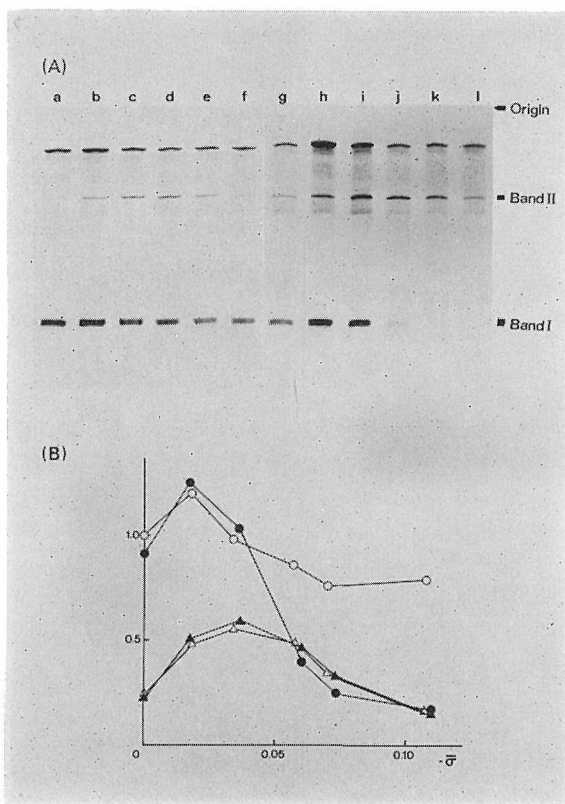

Fig. 11 Effect of supercoiling on the rate of transcription *in vitro* from the synthetic promoters. Different degrees of negative linking differences were introduced into pPSA403 and pPSZ242, and the amounts of transcripts from *bla* and synthetic promoters were determined as in Fig. 10. (A) Gel electrophoretic patterns obtained for pPSA403 series (a to f) and pPSZ242 series (g to l). The mean superhelical densities ($-\bar{\sigma}$) determined by the method of Keller[13] are follows. a: 0, b: 0.018, c: 0.034, d: 0.057, e: 0.070, f: 0.108, g: 0, h: 0.017, i: 0.036, j: 0.060, k: 0.073, l: 0.110. (B) The radioactivities (c.p.m.) in bands I and II were compared by the values relative to that obtained for the relaxed pPSA403, and plotted against the $-\bar{\sigma}$ values (◯, △: pPSA403, ●, ▲: pPSZ242, ◯, ●: band I, ▲, △: band II).

10A). Band I (120 to 121 bases long) and band II (about 290 bases long) wree assigned to probe DNA fragmnets protected by transcripts from the synthetic and *bla* promoters, and band III (about 610 bases long) to the original *Hap*II fragment protected by read-through transcripts. As shown in Fig. 10B, all the synthetic promoters in the linear form showed almost the same activities. In the natively supercoiling state, however, the activities of PSZ241 and PSZ242 were markedly reduced. The amounts of transcripts from the *bla* promoter which provides the internal control were not influenced by the CG block insertion, although the relative amounts formed on the supercoiled DNA were higher than those on the linear DNA. The result strongly suggested that the Z-DNA formation in the close vicinity of the −35 position results in strong inhibition of promoter activity. In support this view, the depressed activity of the PSZ242 promoter was partially recovered by increasing salts to 0.2 M (data not shown), in which the B-DNA structure is stabilized.[25]

In the next step, different negative helical densities were introduced into the

(A)

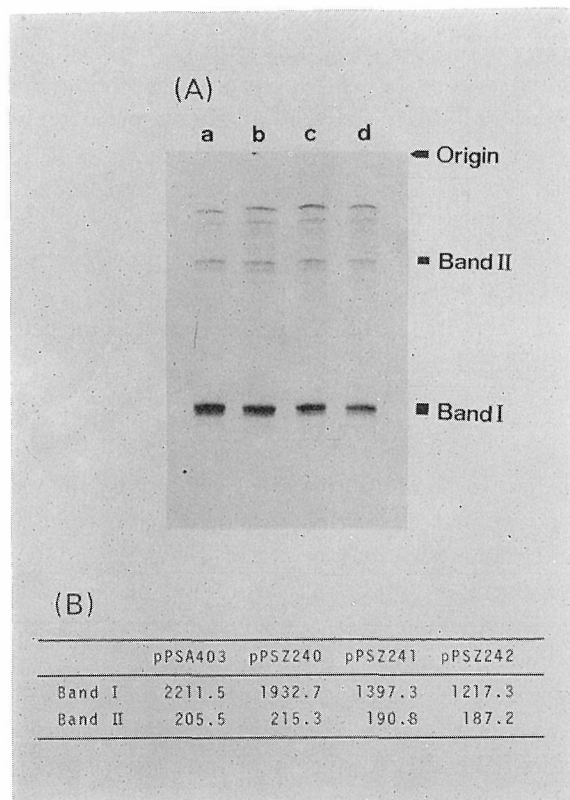|	pPSA403	pPSZ240	pPSZ241	pPSZ242
Band I	2211.5	1932.7	1397.3	1217.3
Band II	205.5	215.3	190.8	187.2

Fig. 12 The amounts of transcripts *in vivo* from the synthetic promoters. The RNA fractions were prepared from cells harboring plasmids, and the amounts of transcripts from synthetic and *bla* promoters were determined by the S1-mapping procedure. (A) Gel electrophoretic patterns obtained for the cells harboring pPSA403(a), pPSZ240(b), pPSZ241(c) and pSPZ242(d). (B) The radioactivities (c.p.m.) in bands, as average values of three independent experiments.

plasmid, pPSZ242, which showed the strongest effect of the CG-insertion, and pPSA-403 as the control, and the correlation of the helical densities and promoter activities were examined in the *in vitro* system. The result of experiments is shown in Fig. 11. Without the CG block, the activity gave a maximum value at the negative helical density $(-\bar{\sigma})$ of 0.02 and then slightly decreased by increasing negative helical densities. By placing the CG block, however, the activity was drastically reduced at the $-\bar{\sigma}$ value larger than about 0.04. Below this critical value, alternating CG sequences of similar sizes is known to form the Z-DNA structure.[26] It is therefore conclusive that the reduction of the PSZ242 promoter activity in the supercoiling state is due to the inhibition of the interaction of RNA polymerase and promoter by the Z-DNA structure formation in the adjacent region.

We have also compared the supercoiling effect on the *bla* promoter, and found that the activity was significantly raised by transition from the relaxing to supercoiling state (Fig. 11B), like most of *E. coli* promoters.[27-31]

Influence of CG blocks on promoter activity in vivo

In order to examine whether the CG block insertion affects promoter activity *in vivo*, the RNA fraction was prepared from cells harboring plasmids, and the total amounts of RNA initiated from the synthetic promoter were determined. The values were also compared with those initiated from the *bla* promoter which provides the internal reference.

As shown in Fig. 12, the transcripts from PSZ241 and PSZ242 were significantly less than that from PSA403. No significant changes were observed in the quantities of the *bla* transcripts from the four plasmids, so that the copy number of these plasmids are assume to be about the same in *E. coli* cells. It is therefore likely that the apparent reduction of the activities of PSZ241 and PSZ242 in the cells is due to the conformational change induced by the CG block insertion.

DISCUSSION

Importance of the two homologous regions deduced by the sequence comparison in promoter function is strongly supported by genetic evidence.[2,3] Most of promoter mutations occur within the hexa-nucleotides in the -10 and -35 regions, and in general, up-mutation that increase the initiation frequency increase the homology and down-mutations decrease the homology to the consensus sequences.[3,4] In this study, a synthetic promoter with the consensus sequences in the -10 and -35 regions was constructed, and it was demonstrated that such an ideal promoter showed strong activity, but only when the spacer length between the conserved regions was adjusted to 17 base pairs. The space adjustment was made at two different restriction sites, and the same result was obtained. Evidence has also been presented with other promoters that the 17 bp-spacing yields higher activity.[32,33] It is therefore concluded that the promoter strength is determined by both the sequences in the consensus sequences and the spacing between the two regions. It is likely that the variation of the activities in natively occurring promoters reflects the deviation of the sequences from the perfect consensus sequences and that from the optimal spaicng

between the two consensus sequences.

Comparison of the promoter sequences indicates that weakly conserved base-pairs are present at the outside of the conserved regions.[2,3] It is especially noted that a sequence CAT frequently appears in the RNA start points.[2,3] I thus constructed a series of promoters having the CAT sequence in the different positions from the −10 region, as well as those having different sequences in the regions between the two consensus sequences and in the region upstream of the −35 consensus sequences. However, neither the promoter strength nor the RNA start point was significantly influenced by other seuqences including the CAT sequence.

Our data rather demonstrate that the −10 consensus sequence is the determinant of the RNA start point, and that initiation predominantly occurs by Pu base at the 7th position from the −10 consensus sequence, TATAAT. Based on these observations the start points of known promoters for *E. coli* RNA polymerase were compared according to the data compiled by Hawley and McClure.[3] The data are summarizde in Table 1. Among 85 promoters of which the initiation sites had been assigned, occurrences of Pu and Py base at the 7th position are 42 and 43, respectively. Most of the promoters having Pu at the 7th position initiate by Pu. But some initiate by purine at adjacent positions. Although little is known about the mechanism of such exceptions, we noted that initiation by Pu at other positions often occur if a stretch of the same Pu base encompasses the 7th position. This was also the case of the constructed promoter PSG105, which initiates RNA synthesis by G at the 6th position within the G cluster. On the other hand, initiation of promoters having Py base at the 7th position mostly occurs by Pu at adjacent positions, especially at the 8th position. The constructed promoter in pPSA302 initiating at 8th position belongs to this category. Only a few promoters initiate by Py bases, but none of them contain Pu at the 7th and 8th positions. Combining with the result obtained in this paper, we deduced a general conclusion that initiation principally occurs by Pu base at the 7th position from the −10 consensus sequence and that if

Table 1 Start points of transcription in 85 known promoters having different bases (Pu or Py) at the 7th position from TATAAT. Occurrence of Pu and Py at indicated positions (Number from TATAAT) and number of promoters initiating by boxed bases were indicated. Data were taken from ref. 4.

Bases at 7th position (Total Number)	6th	7th	8th	9th	Number of Initiation
			(Positions from TATAAT)		
Pu (42)	·	[Pu]	·	·	31
	[Pu]	Pu	·	·	3
	·	Pu	[Pu]	·	4
	·	Pu	·	[Pu]	4
Py (43)	·	Py	[Pu]	·	26
	[Pu]	Py	·	·	5
	Py	Py	Py	[Pu]	6
	Py	[Py]	Py	·	3
	[Py]	Py	Py	·	1
	Pu	Py	[Py]	·	2

Pu base is not present at the 7th position, RNA synthesis is initiated by Pu base at adjacent position.

Since promoter strength was not significantly influenced by the sequences around the RNA start points, the sequence effect on other properties of promoters were examined, and it was found that promoters having relatively GC-rich sequences in the RNA start points are more sensitive to temperature and salts. It has generally been thought that partial unwinding of DNA helix is a prerequisite for successful initiation of transcription.[34,35] In fact local melting of the promoter region from positions −9 to +3 by binding of RNA polymerase has been demonstrated from the sensitivity to modifiaction agents and S1 nuclease.[2,36] It is therefore reasonable that dependency of promoter on temperature and salt concentrations could be influenced by the GC-contents in the region around the RNA start point, as GC-ocntents and ionic strength relate to thermal stability of DNA helix.[37]

I have demonstrated that in the negatievly supercoiling state, the placement of the CG blocks in the close vicinity of the −35 region caused a strong inhibitory effect on promoter activity. Such inhibition was observed at the negative helical densities higher than about $-\bar{\sigma}=0.04$, including those of naturally supercoiling plasmids. Evidence has been presented that short CG blocks in plasmids form the Z-DNA structure in physiological conditions of ionic strength and superhelical density.[12,23] Consistent is our observation that the cccDNAs of the plasmids carrying the CG blocks were not digested by BssHII specific for the GCGCGC sequence. It is therefore likely that the CG regions in supercoiled plasmids form the Z-DNA structure, and the resulting conformational change exerts influence on the activity of the close-neighboring promoter, by blocking the interaction of RNA polymerase. This means that at least the −35 region is strictly required to form the B-DNA structure for the interaction of RNA polymerase. The exact boundaries and the size of the conformational transition at B-DNA to Z-DNA junction are not known, as the boundaries are not always identical to the ends of CG stretch.[38] In the case of PSZ241 and PS1Z242 promoters, the CG blocks are followed by a short stretch of Pu-Py tracts in frame, so that the junction could be expanded to the left end of the −35 consensus sequence (see Fig. 9). However, the inhibitory effect becomes significant only when the CG block was placed in the closed vicinity of the −35 region, and the CG block at 11 bp upstream from the −35 region has little effect. The result implied that the effect of the Z-DNA formation is local and dose not exert influence beyond the range of 11 bp.

It is generally accepted that RNA polymerase approaches to promoter along one side of the helix so as to closely conatct at the two consensus sequences,[2] and subsequently forms a stable complex or the "open" complex, by unwinding the −9 to +3 region.[2,36] This view is consistent with the observations that the activities of most E. coli promoetrs are raised by the transition of DNA from the relaxing state to the negatively supercoiling state which facilitates unwinding of the helical structure.[39] However, some other promoters show different response to supercoiling.[40] It is also shown for the lac promoter that conversion of the sequence in the −10 region to the perfect consensus sequence by substitution of two base makes the pro-

moter (*lac* UV–5) insensitive to supercoiling.[30] It is now demonstrated that the synthetic promoter having the perfect consensus sequences shows only slightly dependency to the supercoiling, while the *bla* promoter is highly sensitive.

The important role of the two consensus sequences in the promoter recognition of RNA polymerase is unquestionable, but the sequences in naturally occurring promoters reveal considerable diversity in the consensus sequence region, and none of them contain both the ocnsensus sequences.[3] The distance between two consensus sequences also varies by several bp, although an optimal distance has been determined to be 17 pb. It is generally assumed that such structural variations in promoter play a major role in gene regulation by affecting the efficiency of RNA polymerase-binding. A view emerged from the different response of promoters to supercoiling is that supercoiling dose not merely enhance the melting of DNA, but rather alters the strength of promoters, by modulating the helical face toward the complete contactness with which RNA polymerase interacts. Insensitivity to supercoiling of the *lac* UV–5 as well as the synthetic promoter in this study may be attributable to the perfect consensus sequences which allow the complete fitness of RNA polymerase without any additional torsion.

ACKNOWLEDGMENT

REFERENCES

(1)   Rosenberg, M. and Court, D. C., *Ann. Rev. Genet.,* **13**, 319 (1979).
(2)   Siebenlist, U., Simpson, R. B. and Gilbert, W., *Cell,* **20**, 269 (1980).
(3)   Hawley, D. K. and McClure, W. R., *Nucleic Acids Res.,* **11**, 2237 (1983).
(4)   Youderian, P., Bouvier, S. and Susskind, M. M., *Cell,* **30**, 843 (1982)
(5)   von Hippel, P. H., Bear, D. G., Morgan, W. D. and McSwiggen, J. A.. *Ann. Rev. Genet.,* **53**, 389 (1984).
(6)   Aoyama, T., Takanami, M., Ohtsuka, E., Taniyama, Y., Marumoto, R., Sato, H. and Ikehara, M., *Nucleic Acids Res.,* **11**, 5855 (1983).
(7)   Aoyama, T. and Takanami, M., *Nucleic Acids Res.,* **13**, 4085 (1985).
(8)   Aoyama, T. and Takanami, M. To be published.
(9)   Ohtsuka, E., Taniyama, Y., Marumoto, R., Sato, H., Ikehara, M., Kanazawa, S. and Takanami, M., *Nucleic Acids Res.,* **10**, 191 (1983).
(10)   Maniatis, T., Fritsch, E. F. and Sambrook, J. Ed., "Molecular Cloning," Cold Spring Harbor Laboratory, New York, 1983, p. 86.
(11)   Radloff, R., Bauer, W. and Vinograd, J., *Proc. Natl. Acad. Sci. U.S.A.,* **57**, 1514 (1967).
(12)   Peck, L., J., Nordheim, A., Rich, A. and Wang, J. C., *Proc. Natl. Acad. Sci. U.S.A.,* **79**, 4560 (1982).
(13)   Keller, W., *Proc. Natl. Acad. Sci. U.S.A.,* **72**, 4876 (1975).
(14)   Berk, A. and Sharp, P. A., *Cell,* 12, 721 (1977).
(15)   Maxam, A. M. and Gilbert, W., *Proc. Natl. Acad. Sci. U.S.A.,* **74**, 560 (1977).
(16)   Sanger, F., Nicklen, S. and Coulson, A. R., *Proc. Natl. Acad. Sci. U.S.A.,* **74**, 5463 (1977).
(17)   Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. and Petersen, G. B., *J. Mol. Biol.,* **162**, 729 (1982).
(18)   Wallace, R. B., Johnson, M. J., Suggs, S. V., Miyoshi, K., Bhatt, R. and Itakura, K., *Gene,* **16**, 21, (1981).

(19) Sutcliffe, J. G., *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 77 (1979).

(20) Stüber, D. and Bujard, H., *Proc. Natl. Acad. Sci. U.S.A.*, **78**, 167 (1981).

(21) Seeburg, P. H., Nüsslein, C. and Schaller, H., *Eur. J. Biochem.*, **74**, 107 (1977).

(22) Sugimoto, K., Sugisaki, H., Okamoto, T. and Takanami, M., *J. Mol. Biol.*, **111**, 487 (1977).

(23) Singleton, C. K., Klysik, J., Stirdvant, S. M. and Wells, R. D., *Nature*, **229**, 312 (1982).

(24) Zacharias, W., Larson, J. E., Kilpatrick, M. W. and Wells, R. D., *Nucleic Acids Res.*, **12**, 7677 (1984).

(25) Azorin, F., Nordheim, A. and Rich, A., *EMBO J.*, **2**, 649 (1983).

(26) Peck, L. and Wang, J. C., *Proc. Natl. Acad. Sci. U.S.A.*, **80**, 6206 (1983).

(27) Botchan, P., Wang, J. C. and Echols, H., *Proc. Natl. Acad. Sci. U.S.A.*, **70**, 3077 (1973).

(28) Richardson, J. P., *J. Mol. Biol.*, **91**, 477 (1975).

(29) Botchan, P., *J. Mol. Biol.*, **105**, 161 (1976).

(30) Wei-Jue, C., Yang, H.-L., Zubay, G. and Gellert, M., "Promoters" Rodriguez, R. and Chamberlin, M. J. Ed., Praeger, New York, 1982, p. 242.

(31) McClure, W. R., Hawley, D. K. and Malan, T. P., "Promoters" Rodriguez, R. and Chamberlin, M. J. Ed., Praeger, New York, 1982, p. 111.

(32) Russel, D. R. and Bennett, G. N., *Gene*, **20**, 231 (1982).

(33) Stefano, J. E. and Gralla, J. D., *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 1069 (1982).

(34) Wang, J. C., Jacobsen, J. H. and Saucier, J.-M., *Nucleic Acids Res.*, **4**, 1225 (1977).

(35) Melnikova, A., Beabealashvilli, R. and Mirzabekov, A. D., *Eur. J. Biochem.*, **84**, 301 (1978).

(36) Siebenlist, U., *Nature*, **279**, 651 (1979).

(37) Marmur, J. and Doty, P., *J. Mol. Biol.*, **5**, 109 (1962).

(38) Johnston, B. H. and Rich, A., *Cell*, **42**, 713 (1985).

(39) Yang, H.-L., Heller, K., Gellert, M. and Zubay, G., *Proc. Natl. Acad. Sci. U.S.A.*, **76**, 3304 (1979).

(40) Wang, J. C., "Genetic Rearrangement: Biological Consequences of DNA Structure and Genome Arrangement" Chater, K. F., Cullis, C. A., Hopwood, D. A., Johnston, A. W. B. and Woolhouse, H. W. Ed., Croom helm, Ltd., London, 1983, p. 1.