Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 3203–3207, June 1983
Biochemistry

Biological expression of an *Escherichia coli* consensus sequence promoter and some mutant derivatives

(synthetic DNA/mutagenesis/heparin resistance)

JOHN J. ROSSI, XAVIER SOBERON*, YASUMASA MARUMOTO, JEANNETTE MCMAHON, AND KEIICHI ITAKURA

Department of Molecular Genetics, City of Hope Research Institute, 1450 E. Duarte Road, Duarte, California 91010; and *Instituto de Investigaciones Biomedicas, Departmento de Biologia Molecular, Ciudad Universitaria 20, Districto Federal, Mexico

Communicated by Mark Ptashne, February 16, 1983

ABSTRACT **A prokaryotic consensus sequence promoter has been chemically synthesized and cloned in bacterial plasmid vectors. This designed sequence is biologically active and promotes efficient expression of the genes to which it is fused. It is an unusually strong promoter *in vitro*, capable of specifying multiple rounds of transcription even when there is a large molar excess of heparin present prior to the addition of RNA polymerase. These properties make this a useful sequence for the *in vitro* production of RNAs. A 2-base-pair spacer mutant and a −35 region transversion mutant have been created *in vitro* in the synthetic promoter by synthetic-DNA-mediated, site-specific mutagenesis. The spacer mutant has a marginal *in vivo* effect on promoter strength but virtually abolishes the *in vitro* heparin resistance. The −35 region transversion changes a highly conserved nucleotide into the statistically least preferred base. This mutation has no marked effect on *in vivo* or *in vitro* promoter strength.**

By using a limited amount of sequence information, it was recognized several years ago that there exist two regions of general homology in prokaryotic promoter sequences. These are the Pribnow box, or −10 region, and the −35 region (1–3) centered about 10 and 35 base pairs, respectively, upstream from the transcriptional initiation site. More recently, statistical analyses of promoter sequence composition have resulted in the generation of model promoter sequences containing the most frequently found bases at given positions throughout the RNA polymerase-recognized region (4–6).

We have taken advantage of these statistically derived sequences and have chemically synthesized and cloned a consensus sequence promoter. Our model promoter sequence design is similar to the consensus sequence derived by Rosenberg and Court (5), which is based on 46 compiled prokaryotic promoter sequences. We show by *in vitro* and *in vivo* criteria that the 42-base-pair consensus sequence contains all the information necessary for efficient and accurate transcription by *Escherichia coli* RNA polymerase. These studies conclusively demonstrate that a totally designed consensus regulatory sequence, predicted from genetical and biochemical data, is biologically functional. This consensus promoter is an unusually strong recognition sequence for RNA polymerase, making it useful for the large-scale preparation of transcripts *in vitro*.

MATERIALS AND METHODS

Chemical Synthesis of DNA. The synthesis of deoxyribonucleotides, by the phosphotriester approach with block coupling and a solid-phase support, was carried out according to published procedures (7, 8).

DNA Polymerase Reactions. These were carried out as described (9).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Site-Specific Mutagenesis. Single-stranded DNA from M13:pc2 (see Fig. 3) containing the synthetic promoter was used as the template for *in vitro* mutagenesis under conditions similar to those described by Zoeller and Smith (10).

Recombinant single-stranded phage from 100 individual plaques were spotted onto nitrocellulose and hybridized with the ^{32}P-labeled synthetic oligonucleotides under conditions that discriminated between complete and mismatch hybridization (11). After a final plaque purification, replicative form DNA was prepared from the mutant viruses and was used as the source of restriction fragments for subcloning. The DNA sequences were determined after subcloning in both cases.

***In Vitro* Transcriptions.** RNAs were synthesized in 50-μl reaction mixtures of 18 mM Tris (pH 7.8), 12 mM $MgCl_2$, 70 mM KCl, 1 mM dithiothreitol, nonradioactive nucleoside triphosphates at 500 μM each, and a 25 μM mixture of the nucleoside [α-^{32}P]triphosphate, and an unlabeled carrier. RNA polymerase (New England BioLabs) was used at 7.5–15 pmol and the purified restriction fragment template was at 1–2 pmol per reaction. When restriction-digested whole plasmid DNAs were used, 0.4 pmol of DNA template and 0.8 pmol of RNA polymerase were added to the mixture. Heparin was used at either 50 or 100 μg/ml and was added before, after, or simultaneously with RNA polymerase as indicated.

Other Methods. Methods used for DNA sequence analysis, plasmid preparation, and bacterial transformation have been described (9).

Determinations of the *in vitro* transcriptional initiation sites were done by abortive initiation (12). The *in vivo* start sites were determined by nuclease S1 mapping (13).

RESULTS

Sequence Design and Synthesis. The −35 region sequence T-T-G-A-C-A and the −10 region sequence T-A-T-A-A-T-G contain the most statistically favored base at each position (Fig. 1). Nucleotides between the −35 and −10 regions of the natural promoters are variable in length and sequence composition (4–6). In our promoter design, we utilized the average spacing, which is 17 base pairs, separating the last base in the −35 region sequence from the first base in the −10 region sequence. A synthetic *Hin*dIII linker was ligated to the 3′ end of the promoter, with the resultant C-C-C-A-A-G-C-T-T sequence having either a potential C or A start. Transcription initiates *in vitro* and *in vivo* at the position shown in Fig. 1.

The scheme for the final synthesis of the 42-base-pair sequence is depicted in Fig. 2. The fidelity of the *in vitro* DNA polymerase I-catalyzed reaction was determined by DNA sequence analyses both prior to and subsequent to cloning (data not presented).

Cloning and Assessment of Biological Function of the Synthetic Promoter. The fragment containing the *Eco*RI–*Hin*dIII flanked promoter was ligated with a pBR327 (14) derivative in

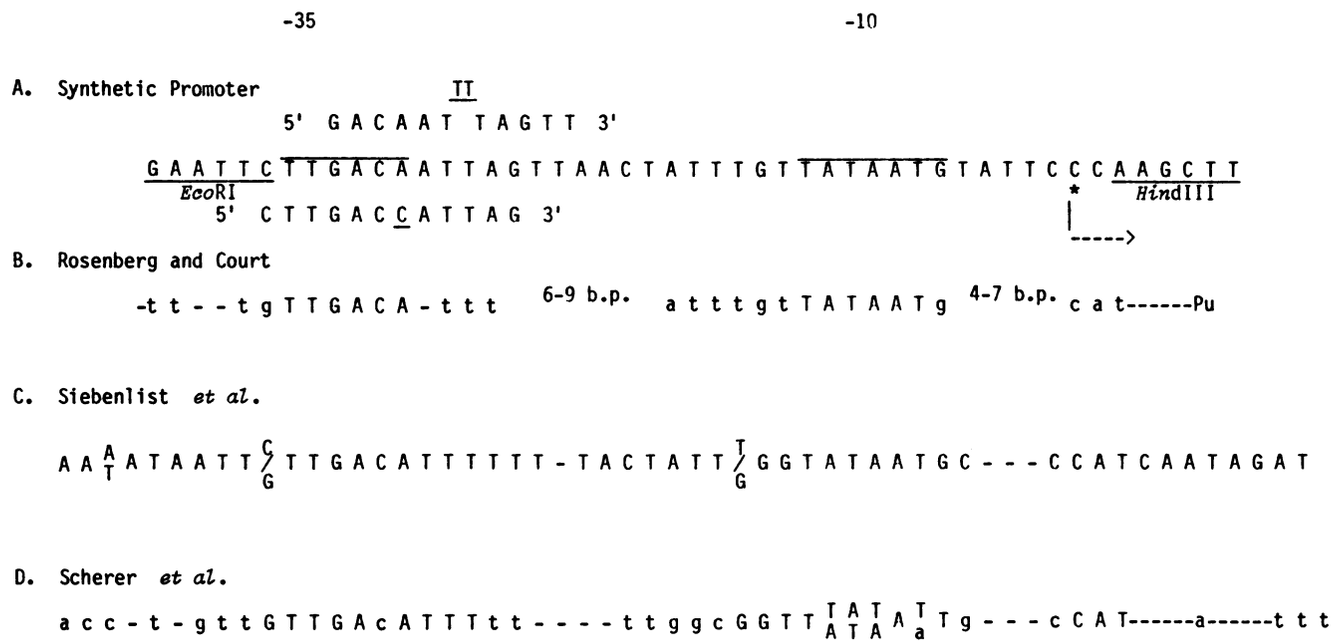

FIG. 1. Synthetic and consensus or model promoter sequences. The nontemplate strand is depicted in each case. In *B* and *D*, the uppercase letters correspond to highly conserved bases and the lowercase letters correspond to less-conserved bases. The Rosenberg and Court sequence (5) is based on statistical analyses of 46 prokaryotic promoters; the Siebenlist *et al.* sequence (6) is derived statistically from comparisons of 54 prokaryotic promoters; the Scherer *et al.* sequence (4) is derived from a computer-based analysis of 17 prokaryotic promoters. For the synthetic promoter and the spacer mutant, the experimentally determined (data not presented) transcriptional start is identified by !*----->. The oligonucleotide sequences above and below the synthetic promoter represent the mutators used to generate the 2-base-pair spacer mutation (upper) and −35 region transversion mutation (lower). b.p., Base pair(s).

which the *tet* promoter had been deleted by nuclease BAL 31 treatment (data not presented). Ligation of the synthetic promoter into the *Eco*RI and *Hin*dIII sites of this plasmid restored tetracycline resistance to bacteria transformed with this DNA (data not presented). This plasmid was used for the constructions depicted in Fig. 3. Plasmid pXJ002 contains a 780-base-pair chloramphenicol acetyltransferase segment (15) which is joined to the synthetic promoter. pXJ003 was derived from pXJ002 and pMLB1034 as depicted. This created a fusion protein encoding sequence consisting of the first 52 $^2/_3$ amino acid codons of chloramphenicol acetyltransferase fused to a defective *lazZ* (β-galactosidase) at the 8 $^1/_3$ codon. Both constructions result in functional expression of the proteins to which they are fused (Table 1).

We also have constructed, by nuclease BAL 31 mutagenesis, a promoter deletion derivative of pXJ002 in which 49 base pairs in the synthetic promoter region were deleted but the *Hin*dIII insert containing the entire chloramphenicol acetyltransferase encoding sequence was left intact. This plasmid, designated pYJ026 (Table 1), did not promote expression of any detectable chloramphenicol acetyltransferase activity. These results verify that *in vivo* expression of the gene for this enzyme is dependent on transcription from the synthetic promoter.

Construction of Mutations in the Synthetic Promoter. The oligonucleotide primers used for the *in vitro* mutagenesis are

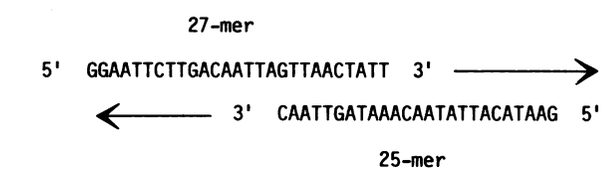

FIG. 2. Nucleotide sequences of the chemically synthesized single-stranded fragments and the duplex structures they form as substrates for DNA polymerase I. The arrows indicate the direction of DNA polymerase I-mediated repair synthesis.

illustrated in Fig. 1A. The mutant alterations were verified by DNA sequence analysis after subcloning into plasmids pXJ002 and pXJ003 (data not presented). Neither the spacer nor the −35 region mutants showed altered phenotypes on selective or indicator media.

Assays of β-galactosidase and chloramphenicol acetyltransferase activities were carried out with bacteria harboring both the mutant and nonmutant promoter-containing plasmids (Table 1). The spacer mutation resulted in a reproducibly small (20–40%) reduction in β-galactosidase or chloramphenicol acetyltransferase expression. The A- to C-transversion in the −35 region resulted in slightly increased levels of chloramphenicol acetyltransferase expression.

The Synthetic Promoter Directs Strong *In Vitro* Transcription. The consensus sequence is an unusually strong *in vitro* promoter. Runoff transcription experiments using DNA restriction fragments were carried out as described in the legend to Fig. 4. We observed strong *in vitro* transcription from the synthetic promoter even when RNA polymerase was used to initiate transcription in a preincubation mixture consisting of template DNA, NTPs, and heparin at 100 μg/ml (Fig. 4*C*). No transcription from the spacer mutant was observed under these conditions. Initiation of transcription with a mixture of heparin (50 μg/ml) and nucleoside triphosphates resulted in strong runoff transcriptions from both the consensus sequence and spacer mutant promoters (Fig. 4*B*). Under these same conditions, weaker transcription was observed from the β-lactamase promoter.

We utilized an assay developed by Stefano and Gralla (21) for measuring open complex stability. Briefly, RNA polymerase, DNA template, and heparin are preincubated together for varying periods of time followed by a 10-min transcriptional runoff initiated with the addition of nucleoside triphosphates at the indicated times from the start of preincubation.

Even after 3 hr of preincubation of the consensus sequence promoter with RNA polymerase in the presence of heparin,

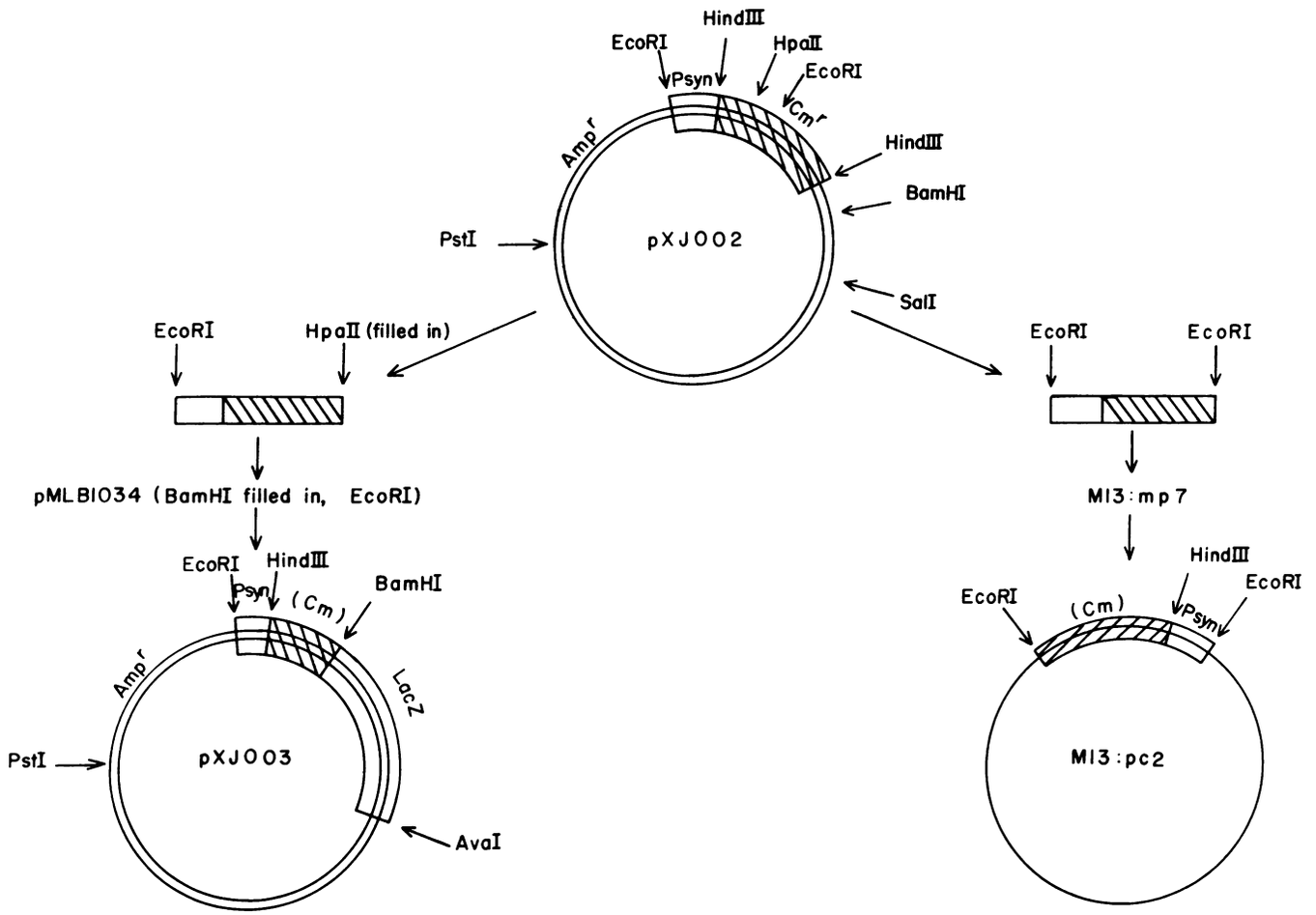

FIG. 3. Synthetic promoter expression vectors and M13 derivative for *in vitro* mutagenesis. The M13:pc2 virus has the promoter insert oriented so as to be complementary to the oligonucleotide mutators depicted in Fig. 1 when the (+)-strand viral template was used for *in vitro* mutagenesis.

there was little or no detectable reduction in runoff transcriptional efficiency (Fig. 5). In contrast, transcription from the spacer mutant decreased markedly over the same time course. In this same experiment, the promoter deletion plasmid pYJ026 (see above) did not promote any chloramphenicol fragment runoff transcription, although a small amount of transcription from the β-lactamase promoter was observed. When DNA from the −35 region mutant (pXJ005) was examined in an identical assay, the open complex stability was similar to that observed for pXJ002 (data not presented).

As was determined for the consensus sequence promoter (see above), the spacer mutant transcription initiated with CTP at the position illustrated in Fig. 1.

Table 1. Chloramphenicol acetyltransferase (CAT) and β-galactosidase activities

Plasmid promoter	Enzyme, units
Chloramphenicol acetyltransferase*	
Experiment 1	
pXJ002, synthetic	0.840
pXJ004, synthetic, T-T insertion	0.530
pYJ026, deleted synthetic	ND
Experiment 2	
pXJ002, synthetic	1.3
pXJ005, synthetic, −35 region transversion	2.1
β-Galactosidase†	
pXJ003, synthetic	1,516 ± 324
pXJ003-4, synthetic, T-T insertion	1,227 ± 395

*CAT measurements were carried out at ambient temperature by using the assay described by Shaw (16). Extracts were prepared from single-colony innocula grown overnight in L broth plus ampicillin. The data presented are from single determinations, but the relative differences observed are fully reproducible. Units are expressed as μmol/min per mg of protein. Proteins were measured by the method of Lowry *et al.* (17) or Bradford (18). ND, none detected.

†β-Galactosidase assays and units were as described by Miller (19). The means ± SD presented were determined from eight independent measurements for each construction.

DISCUSSION

Prior to our studies, Dobrynin *et al.* (22) chemically synthesized a prokaryotic promoter largely homologous to the statistically derived sequence of Scherer *et al.* (4) (see Fig. 1). The biological function of their promoter was implied by experiments which partially replaced the endogenous pBR322 *tet* promoter with the synthetic sequence. This construction resulted in maintaining tetracycline-resistance levels equivalent to those obtained with the natural *tet* promoter. No further characterizations were reported.

The nucleotide sequence of our synthetic promoter conforms closely to the statistically derived consensus sequence reported by Rosenberg and Court (5) (Fig. 1), with the exception of the sequences in the region of transcription initiation. The synthetic consensus sequence has been shown by a number of criteria to be biologically functional (Table 1; Figs. 4 and 5). This promoter has several interesting and important properties.

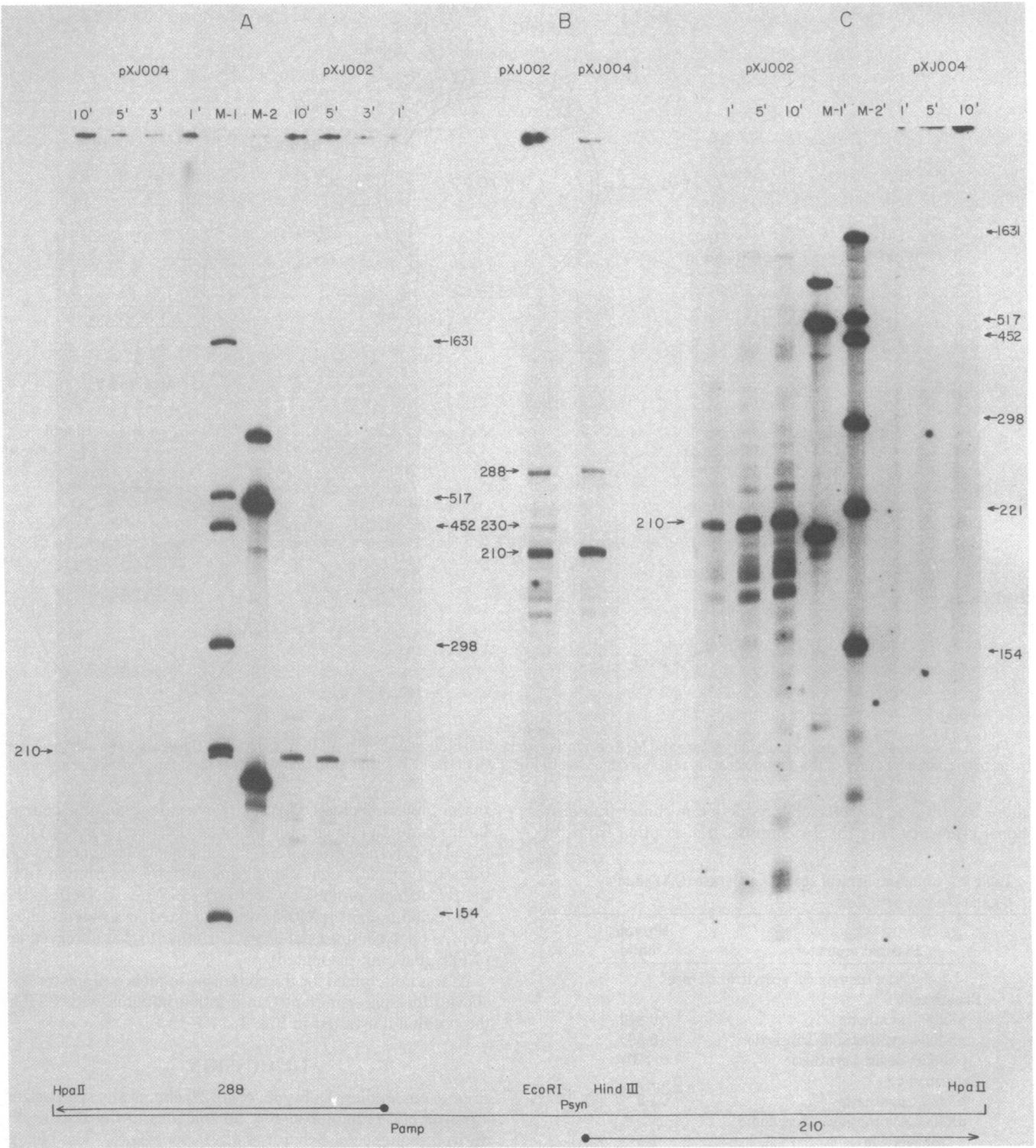

FIG. 4. Heparin challenge and *in vitro* runoff transcription from the consensus sequence (pXJ002) and spacer mutant (pXJ004) promoters. The 706-base-pair restriction fragment used as the template for transcription is depicted below. This fragment contains the "leftward" transcribing β-lactamase promoter (20) and the "rightward" transcribing synthetic promoter. (*A* and *C*) Heparin (50 and 100 μg/ml, respectively) was added prior to the addition of RNA polymerase. (*B*) Heparin (100 μg/ml) was added along with NTPs, subsequent to the addition of RNA polymerase. The numbers above the lanes in *A* and *C* indicate the times (min) of transcriptional runoff; transcription runoffs were for 10 min in *B*. Molecular weight markers were 3'-^{32}P-labeled *Hin*fI-digested pBR327 (M1), and 3'-^{32}P-labeled *Hin*dIII/*Hpa* II-digested transcription template (M2). The migrations of the RNA transcripts are somewhat slower than those of the DNA molecular weight markers in this gel system.

Runoff transcription studies suggest this to be an unusually strong *in vitro* promoter. Because some of these experiments involved relatively low RNA polymerase/DNA ratios, we infer that the affinity of RNA polymerase for the promoter is high. This high affinity is apparent under conditions where RNA polymerase is used to initiate transcriptions in a preincubation mix of DNA and heparin, plus nucleoside triphosphates (Fig. 4). We know of no other promoter that competes as efficiently for RNA polymerase binding in the presence of such huge molar excesses of heparin. We have tested a *trp–lac* UV-5 fusion promoter con-

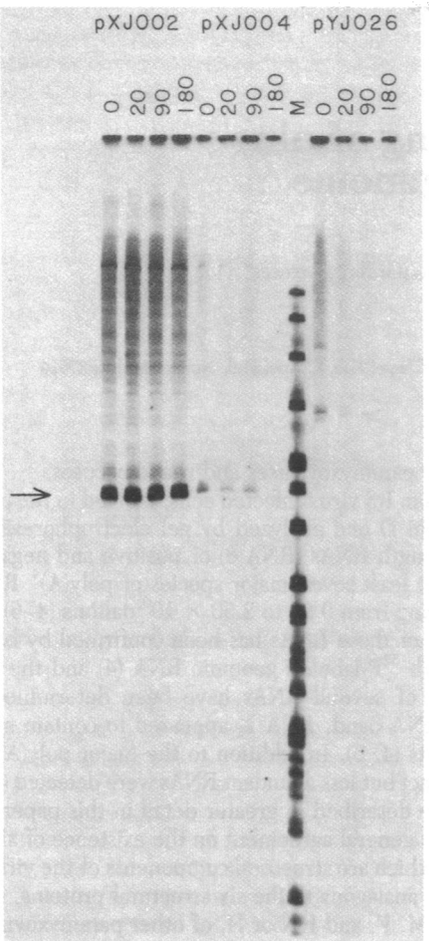

FIG. 5. Open complex stabilities of the consensus sequence (pXJ002) and spacer mutant (pXJ004) promoters. Conditions for this assay were essentially as described (21). The numbers above the lanes refer to the time (min) of preincubation of template DNAs, RNA polymerase, and heparin (100 μg/ml). In each case, the templates were *Hpa* II-digested whole plasmid DNAs. Arrow, Expected runoff transcripts from the consensus and spacer mutant promoters.

taining fragment (supplied by J. Brosius) in a similar reaction. This promoter contains consensus −35 and −10 sequences but has a 16-base-pair spacing between these regions. Poor transcription was observed *in vitro* (data not presented), even though this is a strong promoter *in vivo* (J. Brosius, personal communication).

This unusual property of heparin resistance makes the consensus sequence promoter useful for large-scale *in vitro* production of runoff transcripts by using linearized whole plasmid DNA as template. When heparin is added prior to RNA polymerase, transcriptional initiations from other, weaker promoters are greatly decreased or completely eliminated.

The strong *in vitro* competence of the consensus sequence promoter also enables efficient *in vitro* transcription through A^+T-rich sequences such as those found in noncoding regions of many eukaryotic genes. Thus, this promoter can be utilized for *in vitro* transcribing of eukaryotic RNAs that could be used as substrates for studies of RNA processing or modification.

The importance of spacing between the −10 and −35 regions has been demonstrated for the *lac* promoter (21, 23). We have extended these observations to the synthetic promoter. The 2-base-pair insertion (Fig. 1*A*) increased the −35 to −10 spacing to 19 base pairs, a spacing not found among natural nonmutant promoters (5, 6). Although the spacer mutant significantly alters *in vitro* RNA polymerase–promoter interaction, reduction of *in vivo* expression is less than 50%. Our data demonstrate that the effects of this spacer mutation on *in vitro* promoter function appear to reside in both initial binding of RNA polymerase and maintenance of a stable, open promoter complex (Figs. 4 and 5).

It is interesting to note that our spacer mutant does not alter the *in vitro* or *in vivo* transcriptional start site. We can conclude from this, and other data, that the positioning of the transcriptional initiation site in the synthetic promoter is determined by the −10 region sequence (unpublished data).

In choosing to alter the −35 region sequence T-T-G-A-C-A to T-T-G-A-C-C, we altered a highly conserved nucleotide for which there are no known naturally occurring mutants. The transversion mutation generated changed the dA, which is the most highly conserved base at this position, to dC, which is the least conserved (5, 6). Our results from *in vivo* and *in vitro* analyses indicate that this change has no marked consequence on promoter strength. The strong conservation of this nucleotide among natural promoters thus remains an enigma.

The authors thank Dr. R. Rodriguez for the chloramphenicol acetyltransferase plasmid and Dr. M. Berman for the pMLB1034 plasmid. They also acknowledge the expert technical assistance of P. R. Reeve and M. J. Rossi. This work was supported by National Science Foundation Grant PCM8108280 (to J.R.), and Public Health Service Grant GM30395 (to K.I.). K.I. is a member of the Cancer Research Center (CA16434) at the City of Hope Research Institute. Y.M. was supported by a training grant from Daiichi Seiyaku Co., Ltd., of Japan.

1. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 784–788.
2. Schaller, H., Gray, C. & Herrman, K. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 737–741.
3. Gilbert, W. (1976) in *RNA Polymerase*, eds. Losick, R. & Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 193–205.
4. Scherer, G. E. F., Walkinshaw, M. D. & Arnott, S. (1978) *Nucleic Acids Res.* **5**, 3759–3773.
5. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genetics* **13**, 319–353.
6. Siebenlist, U., Simpson, R. & Gilbert, W. (1980) *Cell* **20**, 269–281.
7. Dembek, P., Miyoshi, K. & Itakura, K. (1981) *J. Am. Chem. Soc.* **103**, 706–708.
8. Miyoshi, K., Huang, T. & Itakura, K. (1980) *Nucleic Acids Res.* **8**, 5491–5505.
9. Rossi, J. J., Kierzek, R., Huang, T., Walker, P. A. & Itakura, K. (1982) *J. Biol. Chem.* **257**, 9226–9229.
10. Zoeller, M. J. & Smith, M. (1982) *Nucleic Acids Res.* **10**, 6487–6499.
11. Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T. & Itakura, K. (1979) *Nucleic Acids Res.* **6**, 3543–3557.
12. Cech, C. L., Lichy, J. & McClure, W. R. (1980) *J. Biol. Chem.* **255**, 1763–1766.
13. Brosius, J., Cate, R. L. & Perlmutter, A. P. (1982) *J. Biol. Chem.* **257**, 9205–9210.
14. Soberon, X., Covarrubias, L. & Bolivar, F. (1980) *Gene* **9**, 287–305.
15. Close, T. & Rodriquez, R. (1982) *Gene* **20**, 305–316.
16. Shaw, W. (1975) in *Methods Enzymol.* **43**, 737–755.
17. Lowry, O., Rosebrough, N., Farr, A. & Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275.
18. Bradford, M. (1976) *Anal. Biochem.* **72**, 248–257.
19. Miller, J. (1972) *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Russell, D. & Bennett, G. N. (1981) *Nucleic Acids Res.* **9**, 2517–2533.
21. Stefano, J. & Gralla, J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 1062–1072.
22. Dobrynin, V., Korobko, V., Severtsova, I., Bystrov, N., Chuvpilo, S. & Koslov, M. (1980) *Nucleic Acids Res. Symp. Series* **7**, 365–376.
23. Mandecki, W. & Reznikoff, W. S. (1982) *Nucleic Acids Res.* **10**, 903–912.