MUTATIONAL ANALYSIS OF PHAGE SP6 TRANSCRIPTION INITIATION AND A NEW TRANSCRIPTION VECTOR

Changwon Kang*, Sang-Chul Nam and In-Woo Lee
Department of Biological Science and Engineering
Korea Advanced Institute of Science and Technology

Efficient *in vitro* RNA synthesis can be easily accomplished from cloned DNA using bacteriophage SP6, T7 or T3 RNA polymerase. Despite its popularity as *in vitro* transcription system, molecular mechanisms of bacteriophage transcription has not been studied, although physical and catalytic properties of several phage RNA polymerases have well been documented (1). Only recently the T7 promoter has been physically mapped by footprinting of the T7 RNA polymerase (2, 3). These simple phage systems, however, could be useful for detailed molecular studies of transcription.

The 3′ halves (from -7 to +1) of the phage SP6, T7 and T3 promoters share almost identical sequence 5′CACTATAG3′, which appears to be important for transcription initiation. Another useful tool for determination of the DNA base pairs crucial for an activity besides consensus sequence and footprinting is genetic analysis combined with recombinant DNA techniques.

For molecular genetic analysis of the phage SP6 promoter and transcription initiation, a new SP6 transcription plasmid vector pCKSP6 was constructed. Since it contains a BamHI cleavage site precisely at the initiation site, this vector would be useful for obtaining RNAs with desired 5′ end sequence.

A series of the plasmid pCKSP6 mutants containing deletions around the initiation site was generated using mung bean nuclease. Activity analysis of these SP6 promoter/initiation sequence mutants revealed the correlation between the sequence around the initiation site and the initiation activity. It was also found that there is a hierachy in the importance of each base pair. Furthermore, the transcription start sites of these mutants were determined by a novel method using sequence-dependent RNA polymerase pausings. The abortive initiation process was found to continue up to 6 nucleotide long RNA chains.

1. New transcription vector

The new plasmid vector pCKSP6 containing a SP6 RNA polymerase promoter is unique in that a gene can be inserted precisely at the transcription start site through the BamHI site. So the resulting transcripts would contain no plasmid derived extraneous sequence at the 5′ end except the initiating base G (4).

A *Xenopus borealis* somatic (Xbs) 5 S RNA gene has been inserted at the initiation site and the termination signal for the *Xenopus* RNA polymerase III in the cloned gene was not recognized at all by the SP6 RNA polymerase. In order to obtain run-off termination at the 3′ end of the structural gene the restriction enzyme Dra I site was established there by oligonucleotide-directed site-specific mutagenesis (5). Thus, we were able to obtain Xbs 5 S RNA with authentic 5′ and 3′ ends by *in vitro* run-off transcription. The transcripts will be used for studies of RNA folding during transcription.

2. Mutations using mung bean nuclease

Taking advantage of the "nibbling" capability of mung bean nuclease to penetrate into frayed DNA double-stranded edges, various deletion mutants of pCKSP6 were obtained, that lack a small number of base pairs around the transcription initiation site (4). The BamHI site of the pCKSP6 was opened and the ends were treated by

* To whom correspondence should be addressed.
The work was partially supported by a research grant from the Korea Science and Engineering Foundation.

mung bean nuclease and religated. The resulting circular plasmids are pCKSP6d3, pCKSP6d4, pCKSP6d5, pCKSP6d6, and pCKSP6d7, the number at the end defining the size of deletion in base pairs. The pertinant partial sequences of all the mutants' noncoding strands are listed in Table I.

Interestingly the penetration of mung bean nuclease is asymmetric (Table I). No base pairs were removed at the downstream end of BamHI cleaved pCKSP6, but up to three base pairs were deleted at the upstream end. This is probably because the downstream end of 8 G:C base pairs is more stable than the A:T rich upstream end.

3. Transcriptional activities of mutants

The relative efficacies of the mutants as templates in the *in vitro* SP6 transcription reactions compared to that of the pCKSP6 are listed in Table I. The template activities of the plasmids pCKSP6, pCK+Xf and pCK+Xf are very similar. Since the three plasmids share the identical sequence down to position +6 and their further downstream sequences are different, the measured activities indicate that the sequences downstream from +7 do not measurably affect the promoter function of initiation.

The activities of the 7 plasmids listed below, sharing the identical sequence upstream from -1, vary greatly. This can only be explained if promoter efficacy depends on the six base pairs from +1 to +6, as well as the previously recognized consensus sequence upstream of the initiation site (5-8).

Plasmids	+1 2 3 4 5 6	*rel. act.*
pCKSP6	GGATCC	100%
pCKSP6d4	GCCCGG	79%
pCKSP6d3	GCCCCG	65%
pCK-X	GCGGCT	42%
pCK+X	GCGCCT	29%
pCKSP6d5	CCCGGG	19%
pCK+Xd	CGCCTA	4%

This comparison made it possible to define the relative importance of individual base pairs. Two pairs of plasmids differ only at one base pair. When the altered base is at +4 (pCK-X and pCK

Table 1. Transcription activities of the phage SP6 initiation sequence mutants

Plasmid	Non-coding strand nucleotide sequence[a]	Relative activity[b]
	-10 +1 +10	
pSP64	TGACACTATAGAATACAAGCTTGGG	
pCKSP6	TGACACTATAGGATCCCCGGGCGAG	100%
pCKSP6d3	TGACACTATAGCCCCGGGCGAGCTC	65%
pCKSP6d4	TGACACTATAGCCCCGGGCGAGCTCG	79%
pCKSP6d5	TGACACTATACCCGGGCGAGCTCGA	19%
pCKSP6d6	TGACACTATACCCGGGCGAGCTCGAA	18%
pCKSP6d7	TGACACTATACCCGGGCGAGCTCGAAT	3%
	1 2 3 4 5 6 7 8 9	
pCK+Xf	TGACACTATAGGATCCGCCTACGGC	87%
pCK+X	TGACACTATAGCGCCTACGGCCATA	29%
pCK+Xd	TGACACTATACGCCTACGGCCATAC	4%
	1 2 3 4 5 6 7 8 9	
pCK-Xf	TGACACTATAGGATCCGGCTGTTCG	102%
pCK-X	TGACACTATAGCGGCTGTTCGGGGA	42%
pCK-Xa	TGACACTATAAGACACGACTTATCG	

[a] The position +1 is defined as the transcription initiation site of the wild type promoter (pSP64).
[b] The relative transcription activities are compared to pCKSP6 (100%).

+X) or at +5 (pCKSP6d3 and pCKSP6d4), there is a small difference in the activity. All the plasmids with G at +1 have higher activities than those with C at the position. From these observations, it appears that the ability of a change to affect transcriptional efficiency decreases as one moves away from +1.

Using this assumption, similar comparisons for the two base pairs at +1 and +2 would give the order of relative activities, GG>GC>CC>CG; and for the first three base pairs (+1 to +3), GGA>GCC>GCG>CCC>CGC. In no case does a downstream mutation override an additional mutation occurring further upstream. For example, pCKSP6d3 differs from pCK-X at +3, +4 and +6. The change from G to C at +4, in the absence of any other change (pCK-X→pCK+X), results in a loss of transcriptional activity. The transversion from G to C at +3, however, leads to increased activity (GCG→GCC). The overall effect of this triple mutation is increased activity, consistent with the net effect being dominated by the alteration occurring closest to +1. Thus, it is entirely consistent with the above assumption.

Moreover, most changes are transversions between G/C and C/G base pairs. So the observed activity differences are probably not due to intrinsic instability of the DNA duplexes. Free energy changes of duplex melting (from -1 to +6), calculated from parameters recently determined by Breslauer et al. (9), are not correlated with the order of relative transcriptional activities observed with the corresponding mutants. This indicates that the phage SP6 RNA polymerase does not directly bind to an open structure, but first forms a closed complex with the promoter, like the *E. coli* RNA polymerase.

The phage SP6 consensus promoter contains the characteristic "TATA" sequence from -4 to -1. Surprisingly, when the A/T base pair at -1 is altered to C/G (pCKSP6d5→pCKSP6d6), while keeping the two base pairs (CC) at +1 and +2 constant, there is little change in activities (19%→18%). This indicates that the base pairs at +1 and +2 are more important than at -1 for transcription initiation.

4. SP6 transcription promoter

The above results are in good agreement with the recent T7 promoter footprinting data by Ikeda & Richardson (2). From their results one can assume that when only the first ribonucleotide is bound, the phage T7 RNA polymerase protects the DNA from -21 to +6. Our genetic analysis with the phage SP6 RNA polymerase suggests that only the sequence upstream from +6 is involved in initiation. Furthermore, our results suggest that there is a hierachy of importance of each base pair near the initiation site, in the order of +1>+2>+3>+4, +5, +6, -1.

In the absence of RNA synthesis the T7 RNA polymerase protects the DNA from -21 to only -3 as forming the closed complex of initiation (2). Taking into consideration of the similarity of the T7 and the SP6 promoters, this may explain why the mutation at -1 is rather silent. The base pair at -1 might not be directly involved in the initial binding of the RNA polymerase.

During transcription initiation, as our results suggest, a closed complex is first formed and only then the two strands of DNA template in the downstream part of promoter are separated to form an open complex. Since A:T base pairs can melt more easily than G:C, one would predict that A:T to G:C mutants should disfavor the initiation. Our results, however, do not support this hypothesis. The RNA polymerase must play an active role in the melting and the sequence-dependent DNA structural perturbations may affect the unwinding efficiency.

5. Initiation site selection

How dose an RN polymerase choose the initiation site? One hypothetical answer is that it chooses only purine site within a certain range. An *E. coli lac* promoter mutant leads to multiple purine site initiation (10). No one has yet, however, studied the effect of initiation sequence mutations on the start site selection. The set of mutants described here presents good subjects for this study.

The initiation site of each mutant was deter-

mined by a novel method, using sequence-dependent pausings of the phage SP6 RNA polymerase (11). The *in vitro* transcription by the SP6 RNA polymerase was carried out under the conditions where the limiting concentration of a ribonucleotide causes the polymerase to pause long enough and terminate at the positions of the nucleotide. Thus, it was possible to obtain sequencing ladders in high-resolution polyacrylamide-urea gel electrophoresis, as shown in Fig. 1, for example. Precise sizing of these abortive elongation transcripts determined the initiation site of each mutant at the nucleotide level.

The results are shown in Fig. 2. The wild-type (pSP64) starts at +1 G. When the sequence upstream from +1 G is kept constant and only the

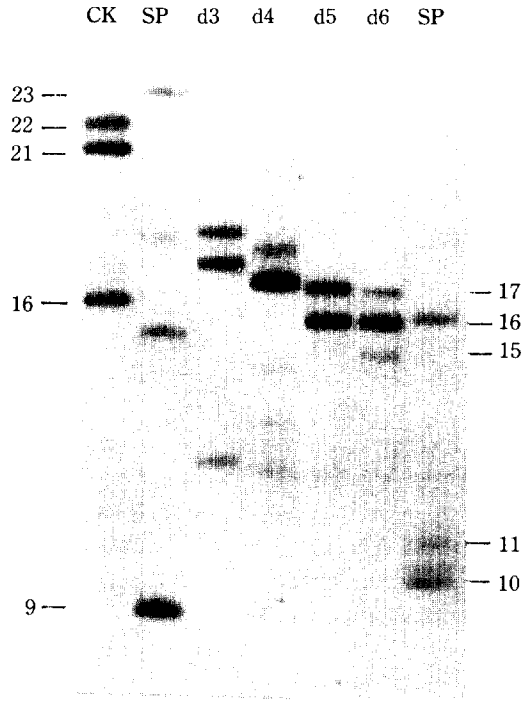

Fig. 1. Abortive elongation transcripts from the reactions with limiting nucleotide concentrations. The limiting nucletide was UTP, except in the 2nd lane from the left where CTP was limiting. The plasmid samples are, from the left, pCK SP6 (CK), pSP64 (SP), pCKSP6d3 (d3), pCK SP6d4 (d4), pCKSP6d5 (d5), pCKSP6d6 (d6), and pSP64 (SP). [α-^{32}P]GTP was added to 2-5 μCi, depending on the transcription efficiency shown in Table I.

PLASMID	+123456
pSP64	TATAGAATAC
pCKSP6	TATAGGATCC
pCKSP6D4	TATAGCCCGG
pCKSP6D3	TATAGCCCCG
pCK-X	TATAGCGGCT
pCK+X	TATAGCGCCT
pCK-XA	TATAAGACAC
pCKSP6D5	TATACCCGGG
pCKSP6D6	TATCCCGGGC

Fig. 2. Transcription initiation sites of the phage SP6 initiation sequence mutants. All mutants starts transcription only at the +1 position, except the pCKSP6d6 which starts at both −1 and +1 positions with equal frequency.

downstream sequence changes, the transcription still starts only at the +1 G, even if it is surrounded by purines at −1 and +2. When the +1 G is changed to either A (pCK-Xa) or C (pCKSP6d5), it still starts only at the site. These results argue against the hypothesis that the phage RNA polymerase starts transcription only with GTP (12), or only with purines. They suggest, however, that the physical position is important for the initiation site selection. The polymerase may start to read the coding strand at a certain distance from a direct contact point on the template DNA.

This hypothesis is further supported by the results with the plasmid pCKSP6d6, where the sequence downstream from the −1 is different from the wild-type (Fig. 2). It starts both at −1 C and +1 C with equal frequency (11). This could happen if the helical structure of the region is altered, so that the start position gets located between the −1 and +1 residues. This suggests that the DNA sequence-dependent structural perturbations would shift the start site.

6. Abortive initiation cycling

The *E. coli* RNA polymerase was found to get

into the stable elongation mode only after escaping from the reiterative synthesis of RNA oligomers up to about 10 bases long (13, 14). According to our results the SP6 RNA polymerase escapes from this abortive initiation cycling after synthesizing 6 base long RNA (11). This may explain why transcription initiation efficiency depends on the sequence from +1 to +6 as well as the upstream promoter sequence (4).

This abortive initiation process is more prevalent at lower concentration of nucleotides, suggesting that abortive release of a transcript competes with productive formation of elongation complex.

REFERENCES

1. Chamberlin, M. & Ryan, T. (1982) The Enzymes, **15**, 87-108.
2. Ikeda, R.A. & Richardson, C.C. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3614-3618.
3. Basu, S. & Maitra, U. (1986) *J. Mol. Biol.* **190,** 425-437.
4. Kang, C. & Wu, C.-W. (1987) *Nucl. Acids Res.* **15,** 2279-2294.
5. Lee, I.-W. (1988) M.S. Thesis, KAIST.
6. Brown, J.E., Klement, J.F. & McAlister, W.R. (1986) *Nucl. Acids Res.* **14,** 3521-3526.
7. Dunn, J.J & Studier, F.W. (1983) *J. Mol. Biol.* **166,** 477-535, and erratum (1984) *J. Mol. Biol.* **175,** 112.
8. McGraw, N.J., Bailey, J.N., Cleaves, G.R., Dembinski, D.R., Gocke, C.R., Joliffe, L.K., MacWright, R.S. & McAllister, W.T. (1985) *Nucl. Acids Res.* **13,** 6753-6766.
9. Breslauer, K.J., Frank, R., Bloecker, H. & Marky, L A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3746-3750.
10. Carpousis, A.J., Stefano, J.E., & Gralla, J.D. (1982) *J. Mol. Biol.* **157,** 619-633.
11. Nam, S.-C. (1988) M.S. Thesis, KAIST.
12. Basu, S. & Maitra, U. (1986) *J. Mol. Biol.* **190,** 425-437.
13. Carpousis, A.J. & Gralla, J.D. (1980) *J. Biol. Chem.* **259,** 8664-8671.
14. Munson, L.M. & Reznikoff, W.S. (1981) *Biochem.* **20,** 2081-2085.