Effects of saturation mutagenesis of the phage SP6 promoter on transcription activity, presented by activity logos

Inkyung Shin*, Jinsuk Kim*, Charles R. Cantor†‡, and Changwon Kang*§

*Department of Biological Sciences, Korea Advanced Institute of Science and Technology, 373-1 Kusong-dong, Yusong-gu, Taejon 305–701, Korea; †Department of Biomedical Engineering, Boston University, Boston, MA 02215; and ‡Sequenom, Inc., San Diego, CA 92121

Contributed by Charles R. Cantor, December 23, 1999

A full set of SP6 promoter variants with all possible single substitutions at positions −17 to +5 was constructed. Transcription activities of these variants were individually measured *in vivo* and *in vitro* to determine the contribution of each base pair to the promoter activity. The *in vivo* activity was measured indirectly by transcriptional interference of the replication of promoter-bearing plasmids. This activity depends most highly on residues −11, −9, −8, −7, and +1 (initiation site). All substitutions at −11, −9, −8, and −7 abolished formation of closed complexes, except for A−8C. These residues are involved in base-specific interactions with the polymerase, and the substitutions exhibit the same strong inhibition *in vitro*. In contrast, the *in vitro* activities of some other variants, measured on linearized templates, were different from those *in vivo*. Some variants at −13, −4, and −2, among others, showed exceptionally higher activities *in vivo* than *in vitro*, supporting the possibility that these residues are involved in post-binding steps, including template melting and bending. The A−3T variant showed much lower activity *in vivo* than *in vitro*, but it bound to the polymerase 2-fold more than the consensus sequence and is possibly involved in polymerase binding. A quantitative hierarchy of all the base pairs is graphically displayed by activity logos, revealing the energetic contribution of each base pair to the activity.

The bacteriophage SP6 genome encodes a single-subunit RNA polymerase that is very similar to the T7, T3, and K11 RNA polymerases. SP6 promoters consist of a highly conserved 20-bp sequence that extends from positions −17 to +3 and exhibits a strong homology to the T7, T3, and K11 promoter sequences (1). Despite their homology, each polymerase shows highly stringent specificity for its own promoter sequence. Mutational studies with phage T7 promoters have yielded detailed information about the relationship of structure to function and the recognition of them by RNA polymerase. The promoter consists of two domains: an initiation domain downstream of −4 and a binding domain upstream of −5 (2, 3). Single base changes in the initiation domain have little effect on promoter binding but reduce the rate of transcription initiation. In contrast, changes in the binding domain reduce the efficiency of promoter binding but have little effect on the initiation of transcription. Base pairs at positions −11 through −7 of the T7 promoter are essential to the binding (4–7).

The interactions of phage RNA polymerases with their promoters have also been studied by a variety of biochemical methods. Footprinting studies with methidiumpropyl-EDTA-Fe(II) indicate that the T7 polymerase protects the region from −17 to −4 (8, 9). Methylation and ethylation interference studies indicate that the major groove of the promoter between −5 and −12 is important for polymerase binding (10). More recent studies with base analog substitutions reveal that, at positions −11 and −10 of the T3 and T7 promoters, their RNA polymerases recognize functional groups along the nontemplate strand wall of the major groove (11, 12). Li *et al.* (13) identified specific functional groups of residues −9 to −5 as primary contacts with T7 RNA polymerase.

The SP6 promoter has not been studied as thoroughly as the T7 promoter. Although the two promoters share a high degree of homology, there are some differences, especially in the promoter–RNA polymerase interaction region. Mutational studies with the SP6 promoter suggest that the SP6 polymerase binding domain could extend approximately to position −3 and that the polymerase-DNA contacts could be distributed in a broader region than in the T7 promoter (1, 14). *In vitro* transcription results with −9 and −8 mutants at various salt conditions suggest that SP6 RNA polymerase uses more nonionic forces for promoter interaction than T7 polymerase does (1).

In this study, we have constructed a library of SP6 promoter variants with all possible single base pair substitutions at positions −17 to +5. The strengths of these promoter variants were measured *in vivo* and *in vitro* to determine the contribution of each base pair to the promoter activity. Such a full saturation mutagenesis has not been applied even to the T7 promoter. A quantitative view of the base pair contributions is displayed by a method called activity logos, a term analogous to sequence logos (15).

Materials and Methods

Saturation Mutagenesis of the Phage SP6 Promoter. A full set of single residue substitutions from −17 to +5 of the SP6 promoter was constructed by chemical synthesis. Oligonucleotides MutAC (5′-CTGGATCCaTTTaGGTGacacTaTaGaaGaAGTGATCA-GTCTAGATGCG-3′) and MutGT (5′-CTGGATCCAtttAggt-gACACtAtAgAAgAAGTGATCAGTCTAGATGCG-3′) were synthesized on a 0.2-μmol scale with a Gene Assembler Plus DNA synthesizer (Amersham Pharmacia). They carried mixtures of bases in the positions designated by lower case letters; in these positions, the indicated base was 91%, and the other three bases were 3% each. The sequences are of the upper, nontemplate strands from −25 to +24. A 16-nt primer was synthesized, annealed to the 3′ ends of MutAC and MutGT, and extended by the Klenow fragment of *Escherichia coli* DNA polymerase I to make double-stranded DNA. The resulting 49-bp duplexes were then cleaved with *Xba*I and inserted into the *Xba*I–*Sma*I site of pSV2. The promoter assay vector pSV2 was derived from pGEM4Z (Promega). The 2.5-kilobase *Nde*I–*Eco*RI fragment of pGEM4Z was self-ligated after the ends were filled in. The 1-kilobase *Bam*HI–*Bgl*II fragment of pGEMEX-2 (Promega) was inserted at the *Bam*HI site of the resulting promoterless plasmid, resulting in pSV2. Individual promoter variants cloned in pSV2 (pSV-no.) were sequenced, and 40 of the 66 possible mutants were isolated from these random pools. The

§To whom reprint requests should be addressed. E-mail: ckang@sorak.kaist.ac.kr.

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.

remaining 26 point mutants were obtained from the other nine oligonucleotides in the same way.

Determination of Plasmid Copy Numbers for *in Vivo* Promoter Strength Assays. When *E. coli* cells harboring each pSV-no. were transformed with pACSP6R containing the SP6 RNA polymerase gene (16), each pSV copy number depended on the strength of its promoter (17). The copy number of pACSP6R was invariantly 220 per cell in the presence of strong or weak promoters, measured against the chromosome copy number as previously described by Lin-Chao and Bremer (18). Thus, the variable copy number of each pSV was determined by comparison to pACSP6R. JM109 cells grown in LB medium containing 100 μg/ml ampicillin and 25 μg/ml tetracycline were harvested, and plasmids were prepared from 3-ml aliquots by the alkaline lysis method (19). After electrophoresis in 0.8% agarose, gels were stained with ethidium bromide and destained in water. DNA bands were photographed by using Polaroid film type 667, and negative scan images were subjected to density analysis by IMAGEQUANT version 3.3 (Molecular Dynamics). The copy number of each pSV, N_i, was calculated as follows:

$$N_i = \frac{\text{band intensity of each pSV-no. per bp}}{\text{band intensity of pACSP6R per bp}} \times 220.$$

Measurement of Promoter Strength *in Vitro*. The strengths of SP6 promoter variants were individually determined *in vitro* by measuring the production of RNA from a linearized plasmid containing each variant. Transcription reactions were carried out in 20 μl of mixture containing 40 mM Tris·HCl (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM DTT, 0.5 mM each ribonucleotide, 0.2 units/μl RNasin, 1 μg of linear plasmid DNA, 0.5 units/μl of SP6 RNA polymerase (Promega), and 0.3 μM [α-^{32}P]CTP (400 Ci/mmol; 1 Ci = 37 GBq). The reactions were terminated after a 30-min incubation at 37°C by adding 20 μl of the gel loading buffer containing 80% (vol/vol) deionized formamide, 10 mM EDTA (pH 8.0), 0.025% xylene cyanole, and 0.025% bromophenol blue. After 8 M urea/6% polyacrylamide gel electrophoresis, the gels were exposed to Agfa x-ray film Curix at −70°C. By using the autoradiogram as a guide, the radioactive bands were excised from the gel, and the radioactivity was measured by liquid scintillation counting. Alternatively, it was measured by PhosphorImager analysis with a Storm 860 scanner (Molecular Dynamics).

Results

Relative Promoter Strength *in Vivo*. To determine the contribution of each residue of the SP6 promoter to the transcription initiation efficiency, all single residue substitutions were introduced in every position from −17 to +5. (The initiation site is numbered +1.) Transcription activity of each cloned variant was measured *in vivo* by measuring the copy number of an SP6 promoter-bearing plasmid, rather than by measuring a reporter gene expression. When SP6 RNA polymerase is present in *E. coli* JM109 cells, the stability of SP6 promoter-bearing plasmids depends on the orientation of the promoter relative to the replication origin and on the sequence of the origin (17). When the SP6 promoter is directed toward the replication origin of pGEM4Z, the copy number is quantitatively correlated with the efficiency of phage transcription (17). The promoter variants were cloned in pSV2, a derivative of pGEM4Z, and the resulting recombinants, pSV-nos., were introduced into JM109/pACSP6R cells producing SP6 RNA polymerase under the control of *lac* promoter, as shown in Fig. 1. When a promoter variant is active, an extended form of RNA I is synthesized that acts as an inhibitor of the replication primer, RNA II, resulting in reduction in the pSV copy number.

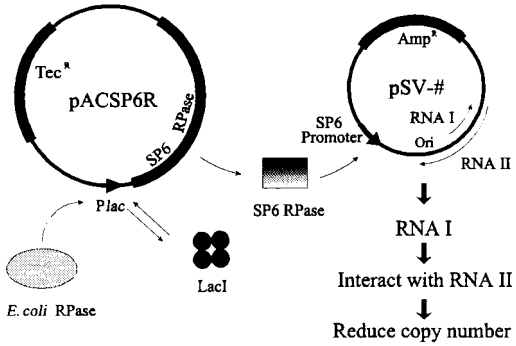

Fig. 1. *In vivo* transcription assay based on interference of the replication of pSV-no. plasmids by SP6 transcription. SP6 RNA polymerase is produced from the gene under the control of the *lac* promoter in plasmid pACSP6R. It recognizes an SP6 promoter variant on a pSV-no. plasmid and produces a transcript that contains a 3′ RNA I sequence. This RNA interacts with RNA II and reduces the pSV-no. copy number. Thus, the copy number depends on strength of the promoter variant.

A kinetic model for replication control of ColE1-type plasmids was previously developed by Brendel and Perelson (20). The plasmid copy number depends most strongly on the rate of RNA I synthesis. It also increases linearly with RNA II synthesis rate and with cell doubling time, but in our studies, these two parameters were constant. Minor modifications were made in our studies to relate the *in vivo* promoter activity (rate of RNA I synthesis) to the copy number of pSV (Fig. 2). In our system, Rom protein is not involved, and the interaction between RNA I and RNA II is weakened, resulting in the high copy number of pGEM4Z (19). All the kinetic parameters were adopted from the original model, except for those involving the RNA I–RNA II complex (k_1, k_{-1}, and k_2). The parameter k_{-2} was not changed, because its value was too small compared with k_2. Values of the three parameters were adjusted to set the copy number (sum of all the forms in Fig. 2) at 400 per cell (for pGEM4Z derivatives).

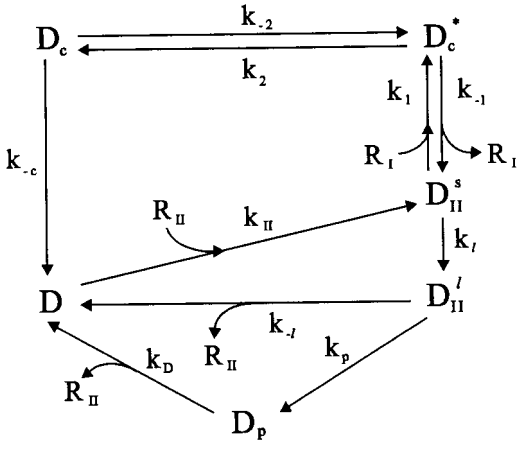

Fig. 2. Kinetic scheme of the ColE1 replication control mechanism slightly modified from the model of Brendel and Perelson (20). The parent plasmid of pSV-no., pGEM4Z, does not produce the Rom protein. The values of parameters k_1, k_{-1}, and k_2 were modified to 4.2×10^7 M^{-1}·min^{-1}, 202 min^{-1}, and 16.5 min^{-1}, respectively, to fix the copy number of pGEM4Z at 400 per cell. Other parameters are the same as described (20). Plasmid DNA occurs either free (D) or in association with RNA II transcripts (D_{II}^s for short transcripts or D_{II}^l for longer transcripts), with RNA II primer (D_p), or with bound complex (D_c* and D_c). Short-length plasmid-bound RNA II (D_{II}^s) forms an unstable complex with RNA I (D_c*). Replication converts primed DNA (D_p) to free DNA (D). The conversions between the different states occur at the rates indicated.

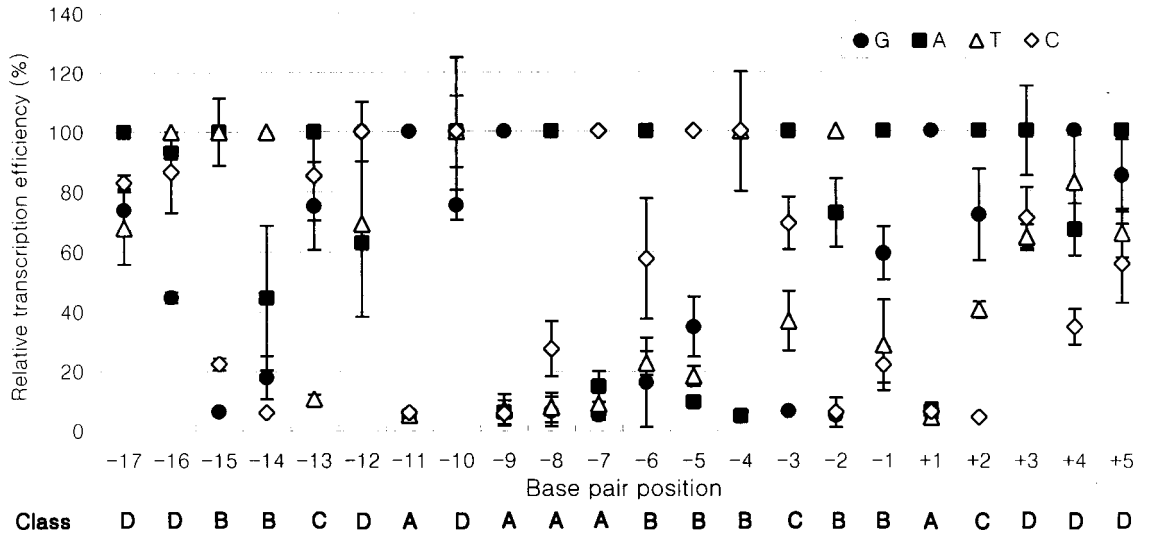

Fig. 3. Relative *in vivo* transcription efficiencies of SP6 promoter variants compared with the consensus SP6 promoter. Each position was classified as described in the text, and the classifications (classes A to D) are shown at the bottom. An average of more than three measurements was taken for each variant.

The copy number of each pSV-no. variant will depend on its rate of RNA I synthesis, k_I, that is controlled by an SP6 promoter variant. The k_I for each promoter variant, k_I^i, was calculated from the measured plasmid copy number, N_i, of that variant.

$$k_I^i = 2{,}330 \times \left(\frac{1}{N_i} + 8.4 \times 10^{-5}\right)$$

The relative activity of each promoter variant, a_i, compared with the consensus promoter, was calculated as follows:

$$a_i = \frac{k_I^i - k_I^{no}}{k_I^{con} - k_I^{no}},$$

where k_I^{con} is for the plasmid containing the consensus SP6 promoter and k_I^{no} is for pSV2 that does not contain an SP6 promoter sequence.

The relative *in vivo* activities of the 66 SP6 promoter variants determined are shown in Fig. 3. None showed significantly higher activity than the consensus promoter. The effects of the substitutions varied greatly: 6 substitutions did not affect the activity at all, and a third (n = 22) of the 66 substitutions reduced the activity to less than 10% of that of the consensus promoter.

The 22 positions from −17 to +5 were divided into four classes depending on the substitution effects. All three substitutions reduced the activity to below one-third the consensus level in positions −11, −9, −8, −7, and +1 (class A). Two of the three substitutions showed such reduction in positions −15, −14, −6, −5, −4, −2, and −1 (class B), whereas only one substitution did so in positions −13, −3, and +2 (class C). In the remaining seven positions, no substitutions showed this much reduction (class D). This classification qualitatively reflects the relative importance of the position to the promoter activity (see below), class A being the highest.

Relative Promoter Strength *in Vitro*. Transcription activities of SP6 promoter variants on linearized plasmid templates were measured individually under standard conditions (at 6 mM $MgCl_2$) and compared with the consensus promoter activity (Fig. 4). Again, the effects of the substitutions greatly varied. The activity range was greater *in vitro* (from 0.01 to 120% of the consensus promoter activity) than *in vivo* (from 5 to 100%). Interestingly, T-to-A substitution at the −10 position (designated T−10A here) increased the activity by 20%, and T−10G, A−3T, and G+4T substitutions did not affect it (98–105 ± 13%) within experimental error. About a half of the variants (n = 35) had less than one-third of the consensus promoter activity. When the 22 positions were divided into four classes based on *in vitro* data, 6 positions changed class compared with the *in vivo* results. Three more positions, −5, −4, and −2, joined the class A *in vitro*. Position −13 joined class B, and positions −1 and +3 joined class C. All these changes represent an increase in importance *in vitro*, except for −1.

The differences between the *in vivo* and *in vitro* activities of each variant are shown in Fig. 5. Some are significantly outside the maximum standard deviation for the difference values (about ± 20%). The most dramatic difference was shown by the T−4C variant: 100 ± 10% *in vivo* but 2 ± 1% *in vitro*. Several other variants showed much higher activities *in vivo* than *in vitro*, including A−13C. Variant A−3T was the only one that showed the opposite effect.

Some of the variants were assayed for polymerase binding or formation of closed complex by gel retardation (data not shown), as previously described (21). All substitutions at −11, −9, −8, and −7 abolished polymerase binding, except for the variant A−8C that retained a third of the binding affinity of the consensus promoter. Thus, loss or reduction of transcription activity by these substitutions is attributed by loss or reduction of polymerase binding. The binding affinity of T−10A, the only variant that was significantly more active *in vitro* than the consensus promoter, was similar to that of the consensus, suggesting that the mutation enhanced a kinetic step after polymerase binding. Variant A−3T bound to the polymerase over 2-fold more than the consensus, suggesting that a subsequent step or steps were inhibited both *in vitro* and *in vivo*.

Abortive initiation products from variants carrying substitutions at −1 through +5, except for +1, were closely examined. The largest products of abortive initiation cycling were 6 nt long as previously reported (22). Only the G+4T variant produced greatly increased amounts of 6-mers, suggesting that this residue plays an important role in abortive initiation cycling.

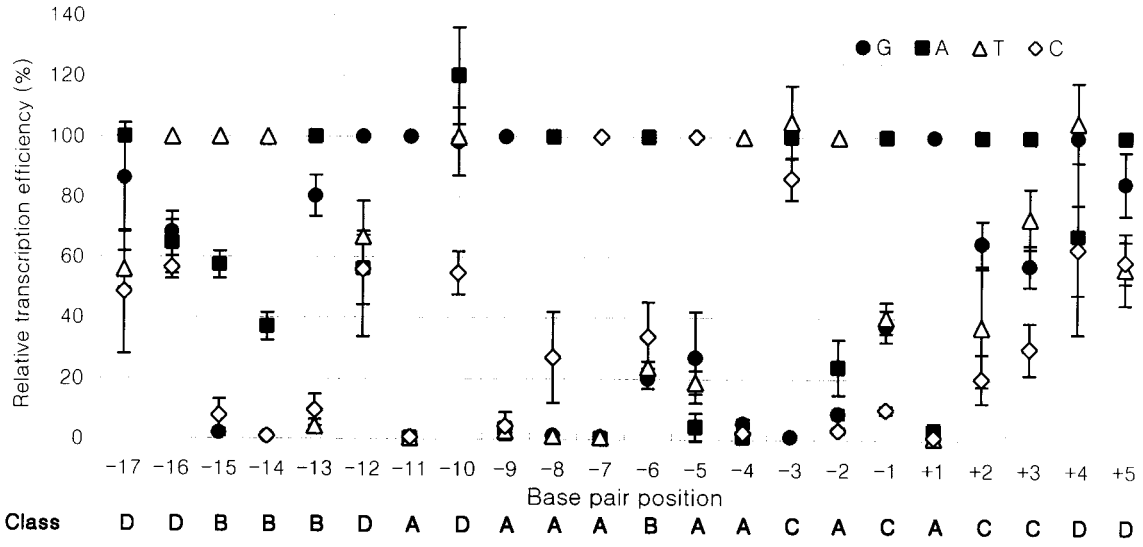

Fig. 4. Relative *in vitro* transcription efficiencies of SP6 promoter variants compared with consensus SP6 promoter. Each position was classified as described in the text, and the classifications (classes A to D) are shown at the bottom. An average of more than three measurements was taken for each variant.

Activity Logos: A Method for the Display of Saturation Mutagenesis Data. All possible base pairs at each position were individually tested in this study for their effects on promoter activity in *in vivo* and *in vitro* assays. A graphic method to display all these data is desirable. Sequence logos (15) display consensus sequence information from a set of comparative sequence data and provide a quantitative view of the amount of information at each point in the functional site, based on information theory. In this study, a method displaying any saturation mutagenesis data was developed based on energetics; the resulting displays are called activity logos, a term analogous to sequence logos.

Activity is not simply a sum of individual residue contributions, but as a first approximation, it can be treated this way (23). In a binding assay, the binding constant, K, can be considered to reflect the free energies, ΔG_i, contributed by each residue in the binding site.

$$K = e^{-\Sigma_i \Delta G_i/RT}$$

If a pure rate is measured, the rate constant k_2 reflects an activation energy, with contribution from each residue, E_i:

$$k_2 = e^{-\Sigma_i E_i/RT}.$$

When the assay depends on the product of the rate and a preequilibrium binding, the activity, V, will depend on:

$$V = k_2 K = e^{-\Sigma_i(\Delta G_i + E_i)/RT}.$$

For example, transcription assays in this study depend on the product of the equilibrium of polymerase-promoter binding and the rates of isomerization and promoter clearance.

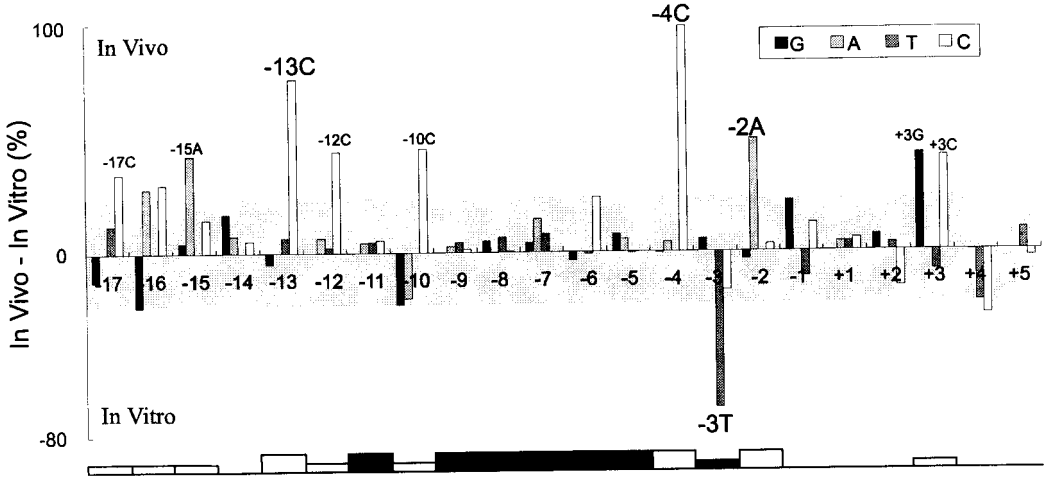

Fig. 5. Difference between *in vivo* and *in vitro* activities of the promoter variants. Each bar results from subtracting the *in vitro* activity from the *in vivo* activity, separately expressed as the percentage of the consensus promoter activity. Differences shown in the shaded area are within experimental error. Residues possibly involved in polymerase binding (filled boxes) and those displaying high *in vivo*:*in vitro* activity ratios (open boxes) are highlighted at the bottom.

Because all of these equations have the same form, we can describe the activity a, for any assay as:

$$RT \ln a = -\sum_{i} F_i,$$

where F_i indicates the relative energetic contribution of the ith position. If we change the identity of the base at position i from the normal base to base b, an altered activity a_{bi} may result. The energy change at that position will be F_{bi}, if we treat the positions as independent.

$$\frac{F_{bi}}{RT} = \ln a - \ln a_{bi} = \ln\frac{a}{a_{bi}}$$

The relative energetic importance of the ith position, H_i^o can be estimated by adding the effects of each possible base on activity.

$$H_i^o = \frac{F_i}{RT} = \sum_{b=1}^{4} \ln\frac{a}{a_{bi}}$$

The normal base should refer to the base yielding the maximum activity, which is not necessarily a consensus or wild-type base, such that $\ln(a/a_{bi})$ is always positive. This formalism weights equally any depression in energetics that affects activity. Thus, a position with relative activities $A = 1$, $C = G = T = 0.1$ and a position with $A = C = G = 1$, $T = 0.001$ would contribute equally.

At each position, an activity logo, the equivalent of a sequence logo, can be constructed by weighting H_i^o by the relative activities seen with each of the four bases.

$$H_{bi}^o = \frac{a_{bi}}{\sum_{b=1}^{4} a_{bi}} H_i^o$$

A correction must be made for the limited dynamic range and signal to noise of activity measurements. To prevent a_{bi} from ever assuming an unrealistically small value, all observed activities can be incremented by a small constant ε, say 0.01 a. Thus, the final form of the activity logos becomes

$$H_{bi} = \frac{(a_{bi}+\varepsilon)}{\sum_{b=1}^{4}(a_{bi}+\varepsilon)} H_i = \frac{(a_{bi}+\varepsilon)}{\sum_{b=1}^{4}(a_{bi}+\varepsilon)} \sum_{b=1}^{4} \ln\frac{a+\varepsilon}{a_{bi}+\varepsilon}.$$

H_i is the height of the logo for position i, and H_{bi} is the portion for base b.

Our extensive set of *in vivo* and *in vitro* activity data on all the point mutations of the SP6 promoter is displayed by activity logos in Fig. 6. The qualitative classification of base pair positions, described earlier, is fairly consistent with this quantitative view of the energetic contribution of every position, except for a few positions in classes B and C. The activity logos in Fig. 6 are also compared with sequence logos constructed from 11 SP6 promoter sequences (1). Sequence logos appear to be high-contrast, low-resolution versions of the activity logos. They may resemble *in vivo* logos more than *in vitro* logos, as reflected by the results for position −3.

Discussion

Although extensive information on phage T7 promoters has accumulated through many independent approaches, none of them covered the entire length of the promoter. In this study, all possible substitutions were introduced into every position of the phage SP6 promoter, and their effects on the promoter activity were measured *in vivo* and *in vitro*, presenting the first example of full saturation promoter mutagenesis. Analysis of 11 known SP6 promoter sequences showed that only positions −17 to +3 contain significant sequence information (1), as shown in the sequence logos of Fig. 6. An *in vitro* evolution study showed no sequence requirements at positions −22 to −18 (24). Escape from abortive initiation cycling occurs after formation of 6-nt RNA in SP6 transcription (22). Thus, our coverage from −17 to +5 includes all the residues of the SP6 promoter.

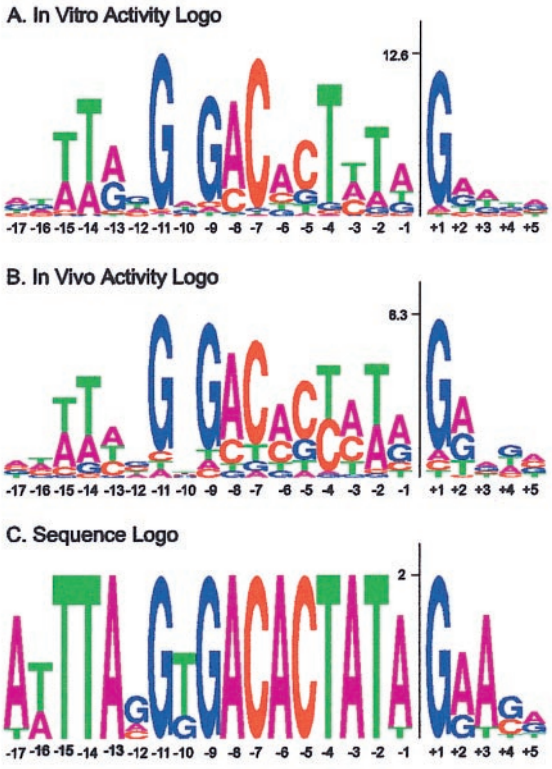

Fig. 6. Activity and sequence logos of the phage SP6 promoter. The two activity logos were constructed based on energetic contribution of each base pair in every position to the promoter activity as described in the text. Maximum heights of the activity logos simply reflect lower limits of measurements. The maximum information content in each residue is 2 in sequence logos. The sequence logos were generated from 11 known SP6 promoter sequences (1) by http://www.bio.cam.ac.uk/seqlogo/logo.cgi (15).

The activity logos shown in Fig. 6 reveal quantitative hierarchies of all the possible base pairs *in vivo* and *in vitro.* Requirement for GTP as an initiating residue seems to be much more stringent in SP6 transcription than in T7 (5). Phage promoters have been thought to consist of two parts, largely based on studies on the T7 promoter. The region from −17 to −5 is for polymerase binding, and that from −4 to +3 is for melting of template duplex and initiation (2, 3). Within the presumably polymerase-binding region, the residues at −17, −16, −12, and −10 of the SP6 promoter show no evidence of base-specific interactions. The most important positions are −11, −9, −8, and −7, and these are involved in polymerase binding. We previously reported that the two base pairs at positions −9 and −8 distinguish between the SP6 and T7 promoters (1). The two consensus promoters are identical at −11 and −7. Although they differ also at −17, −16, −15, −12, and −10 (1), these positions do not contribute to the activity, except for −15. Substitution of

T7-specific A for the SP6 T at −15 resulted in no activity loss at all *in vivo* and only half activity *in vitro* (Figs. 3 and 4).

Jorgensen *et al.* (10) reported that methylation of the G residues at positions −12, −11, −9, −7, and −5 interfered with binding of the polymerase and suggested that the major groove between −5 and −12 is a contact region. These results are consistent with ours, except for the role of the −12 residue. Any mutations in this position retained 56–100% activity, and a substitution of T (having a methyl group) resulted in 67–69% activity both *in vivo* and *in vitro* (Figs. 3 and 4).

The effects of mutation at the four most important positions on transcription activity did not show any significant differences *in vivo* and *in vitro.* These results probably reflect that the base-specific interactions with the polymerase are not significantly different *in vivo* and *in vitro*.

On the other hand, positions −17, −16, −15, −13, −12, −10, −4, −3, −2, and +3 showed great differences *in vivo* and *in vitro* (Fig. 5). Among these, −13C, −4C, −3T, −2A, and +3C showed especially meaningful differences (higher than two-thirds activity under one condition but lower than one-third activity under the other), whereas the others exhibited always higher than 50% activity. These differences in activity may be due to the difference in DNA template structure. The templates presumably formed negatively supercoiled structures *in vivo* but were linear *in vitro*. In fact, all the variants showing significant differential effects favored *in vivo* conditions, except for the A−3T (Fig. 5). The T7 RNA polymerase indirectly recognizes the A+T-rich sequence in the minor groove at −17 to −13 through its inherent flexibility, and these distortions cause a slight bend (25). If the SP6 polymerase recognizes the A+T-rich region in the same way, our results suggest that supercoiled structure might help the structural modulation.

The T7 promoter residues −4 to +3 are involved in postbinding steps, including melting of duplex in the complex (2, 3). The SP6 variant A−3T that showed higher activity *in vitro* than *in vivo* was found to bind the polymerase much more than the consensus promoter, according to the gel shift assay. Thus, it is possible that the SP6 −3 residue is involved in polymerase binding in a relatively nonspecific manner (14).

The relative *in vivo* activity of each promoter variant was measured in this study by the interference in plasmid replication by transcription. This indirect assay for promoter strength has an advantage over a direct assay of RNA measurement (like in our *in vitro* assay) or an indirect assay for activity of a reporter protein. In this *in vivo* assay, strong mutational effects result in high, rather than low, copy numbers of the plasmid, which can be measured very accurately. This advantage is significant especially in construction of activity logos, because strong mutational effects are weighted logarithmically more than weak ones.

The exploding availability of DNA sequence data creates a challenge to devise ways of displaying those data in a manner that conveys comprehensible clues about function. The kind of data most frequently available is comparative sequence information such as which sequences function naturally as promoters for a particular RNA polymerase or which mutations are known to inactivate a promoter. Molecular biologists have most often dealt with such data by constructing a consensus sequence. However, such sequences convey only part of the information available from a set of comparative sequence data. Sequence logos, soundly based on information theory, provide a quantitative view of the amount of information at each point in the functional site. They do not concern the different activities of individual sequences. Activity logos, based on energetics, provide a quantitative view of the energetic contribution of each residue to the function, from measurements of individual activities.

We thank Dr. John Rush and Prof. Charles C. Richardson at Harvard Medical School for sharing unpublished sequences of phage SP6 genomic promoters and Prof. Peter von Hippel at University of Oregon for very helpful comments on the manuscript. This work was partially supported by grants from the Ministry of Education Genetic Engineering Program and from the Korea Advanced Institute of Science and Technology.

1. Lee, S. S. & Kang, C. (1992) *J. Biol. Chem.* **268,** 19299−19304.
2. Chapman, K. A. & Burgess, R. R. (1987) *Nucleic Acids Res.* **15,** 5413−5432.
3. Chapman, K. A., Gunderson, S. I., Anello, M., Wells, R. D. & Burgess, R. R. (1988) *Nucleic Acids Res.* **16,** 4511−4524.
4. Schneider, T. D. & Stormo, G. D. (1989) *Nucleic Acids Res.* **17,** 659−674.
5. Ikeda, R. A., Ligman, C. M. & Warshamana, S. (1992) *Nucleic Acids Res.* **20,** 2517−2524.
6. Ikeda, R. A., Warshamana, G. S. & Chang, L. L. (1992) *Biochemistry* **31,** 9073−9080.
7. Diaz, G. A., Raskin, C. A. & McAllister, W. T. (1993) *J. Mol. Biol.* **229,** 805−811.
8. Ikeda, R. A. & Richardson, C. C. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3614–3618.
9. Gunderson, S. I., Chapman, K. A. & Burgess, R. R. (1987) *Biochemistry* **26,** 1539−1546.
10. Jorgensen, E. D., Durbin, R. K., Risman, S. S. & McAllister, W. T. (1991) *J. Biol. Chem.* **266,** 645−651.
11. Schick, C. & Martin, C. T. (1993) *Biochemistry* **32,** 4275−4280.
12. Schick, C. & Martin, C. T. (1995) *Biochemistry* **34,** 666−672.
13. Li, T., Ho, H. H., Maslak, M., Schick, C. & Martin, C. T. (1996) *Biochemistry* **35,** 3722−3727.
14. Kim, S., Hong, W. J. & Kang, C. (1993) *Biochem. Mol. Biol. Int.* **31,** 153−159.
15. Schneider, T. D. & Stephens, R. M. (1990) *Nucleic Acids Res.* **18,** 6097−6100.
16. Jeong, W. J. & Kang, C. (1997) *Biochem. Mol. Biol. Int.* **42,** 711−716.
17. Kwon, Y., Kim, J. & Kang, C. (1998) *Genet. Anal. Biomol. Eng.* **14,** 133−139.
18. Lin-Chao, S. & Bremer, H. (1986) *Mol. Gen. Genet.* **203,** 143−149.
19. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
20. Brendel, V. & Perelson, A. S. (1993) *J. Mol. Biol.* **229,** 860−872.
21. Muller, D. K., Martin, C. T. & Coleman, J. E. (1988) *Biochemistry* **27,** 5763–5771.
22. Nam, S. C. & Kang, C. (1988) *J. Biol. Chem.* **263,** 18123−18127.
23. Berg, O. G. & von Hippel, P. H. (1988) *Trends Biochem. Sci.* **13,** 207–211.
24. Breakder, R. R., Banerji, A. & Joyce, G. F. (1994) *Biochemistry* **33,** 11980−11986.
25. Cheetham, G. M., Jeruzalmi, D. & Steitz, T. (1999) *Nature (London)* **399,** 80–83.