Continuous *in Vitro* Evolution of Bacteriophage RNA Polymerase Promoters†

Ronald R. Breaker, Amrita Banerji, and Gerald F. Joyce*

Departments of Chemistry and Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, California 92037

Received May 23, 1994; Revised Manuscript Received July 21, 1994®

ABSTRACT: Rapid *in vitro* evolution of bacteriophage T7, T3, and SP6 RNA polymerase promoters was achieved by a method that allows continuous enrichment of DNAs that contain functional promoter elements. This method exploits the ability of a special class of nucleic acid molecules to replicate continuously in the presence of both a reverse transcriptase and a DNA-dependent RNA polymerase. Replication involves the synthesis of both RNA and cDNA intermediates. The cDNA strand contains an embedded promoter sequence, which becomes converted to a functional double-stranded promoter element, leading to the production of RNA transcripts. Synthetic cDNAs, including those that contain randomized promoter sequences, can be used to initiate the amplification cycle. However, only those cDNAs that contain functional promoter sequences are able to produce RNA transcripts. Furthermore, each RNA transcript encodes the RNA polymerase promoter sequence that was responsible for initiation of its own transcription. Thus, the population of amplifying molecules quickly becomes enriched for those templates that encode functional promoters. Optimal promoter sequences for phage T7, T3, and SP6 RNA polymerase were identified after a 2-h amplification reaction, initiated in each case with a pool of synthetic cDNAs encoding greater than 10^{10} promoter sequence variants.

The method of 3SR[1] (self-sustained sequence replication) allows the rapid amplification of specific RNA molecules under isothermal conditions (Guatelli et al., 1990; Joyce, 1992). RNA amplification is the result of the combined activities of a reverse transcriptase, which produces cDNA from the RNA, and a bacteriophage RNA polymerase (RNAP), which produces multiple copies of RNA from the newly-synthesized cDNA. Each new RNA can enter another replicative cycle, resulting in the exponential amplification of the target molecule. Because of the continuous nature of the amplification process and the potential for mutation, this procedure and a related procedure based on Qβ RNA replicase (Miele et al., 1983; Wu et al., 1992) are inherently prone to generate selfish RNAs (Joyce, 1992). Several classes of selfish RNA molecules, each employing a different amplification scheme, are known to exist as "molecular parasites" of *in vitro* RNA amplification reactions [for a review, see Breaker and Joyce (1993)].

One example of a selfish RNA of the 3SR amplification procedure is "RNA Z", whose mechanism of replication is known in detail (Breaker & Joyce, 1994). Amplification of RNA Z requires the activity of both reverse transcriptase and RNAP and involves the synthesis of a cDNA replication intermediate (Figure 1). The cDNA contains the (–) strand of the RNAP promoter element, corresponding to the nontemplate strand of double-stranded DNA. The promoter (–) domain is located immediately adjacent to a DNA hairpin structure. This hairpin serves as a primer for reverse transcriptase-catalyzed second-strand synthesis, resulting in the formation of a double-stranded promoter element. RNAP, which requires a double-stranded DNA promoter (Milligan et al., 1987; Maslak & Martin, 1993) of defined sequence (Diaz et al., 1993), then operates on the elongated DNA hairpin to produce new RNA Z molecules.

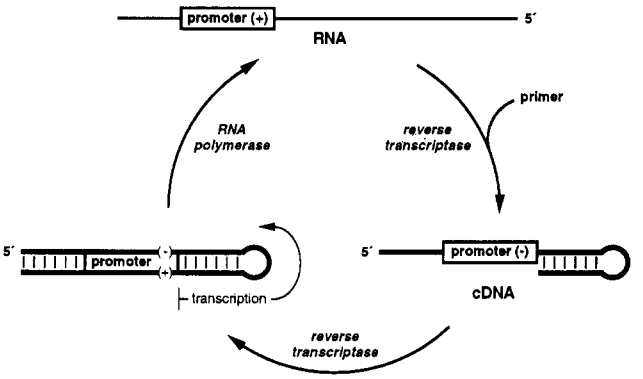

FIGURE 1: Amplification mechanism of RNA Z (Breaker & Joyce, 1994). A DNA primer hybridizes to the 3′ end of RNA Z and is extended by reverse transcriptase to yield a complementary DNA. This cDNA contains a bacteriophage RNAP promoter (–) sequence, corresponding to the sequence normally present in the nontemplate strand of double-stranded DNA. The cDNA rearranges to form a hairpin structure that is further extended by reverse transcriptase, thereby creating a functional, double-stranded promoter element. RNA polymerase then initiates transcription, continuing around the hairpin loop, to generate a complete copy of the original RNA. Both the input RNA and the RNA amplification products contain a promoter (+) sequence, corresponding to the template strand of double-stranded DNA.

In their natural context, phage RNAPs operate on a duplex DNA comprised of separable "template" and "nontemplate" strands. In contrast, the template and nontemplate strands

† Supported by Research Grant NAGW-3118 from the National Aeronautics and Space Administration.

* To whom correspondence should be addressed at the Department of Molecular Biology, The Scripps Research Institute.

® Abstract published in *Advance ACS Abstracts,* September 1, 1994.

[1] Abbreviations: HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; 3SR, self-sustained sequence replication; Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid; AMV RT, avian myeloblastosis virus reverse transcriptase; M-MuLV RT, Moloney murine leukemia virus reverse transcriptase; RNAP, RNA polymerase; Taq polymerase, *Thermus aquaticus* DNA polymerase; PAGE, polyacrylamide gel electrophoresis.

0006-2960/94/0433-11980$04.50/0

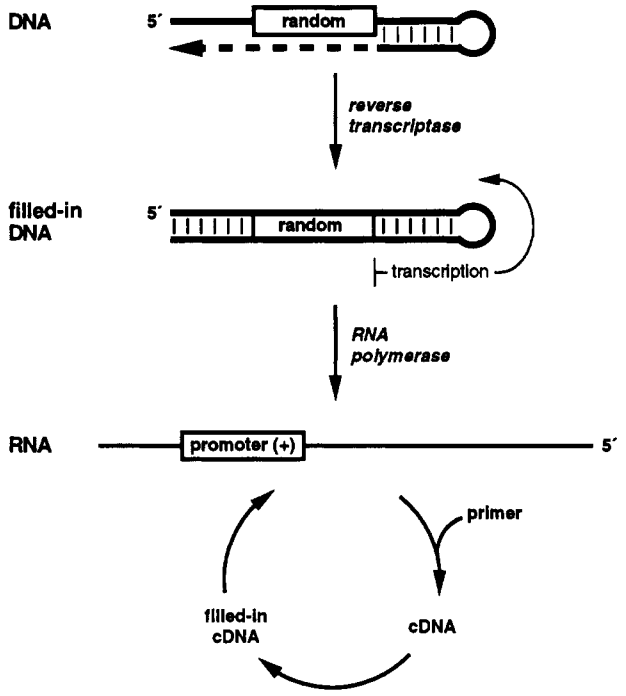

FIGURE 2: Continuous *in vitro* evolution of bacteriophage RNAP promoters. Synthetic DNA molecules that contain random sequences in place of the consensus RNAP promoter sequence are extended by reverse transcriptase to generate a complete stem–loop structure. Only those DNAs that carry a functional promoter will produce RNA products upon addition of a particular RNA polymerase. The resulting RNAs then enter an isothermal amplification reaction, which enriches the population with those molecules that encode the most efficient RNAP promoter. Alternatively, promoter evolution reactions can be initiated directly with randomized synthetic DNAs that correspond to the cDNA replication intermediate.

of the cDNA replication intermediate of RNA Z are joined through the hairpin structure. During transcription, this arrangement causes the polymerase to read around the hairpin loop, resulting in the synthesis of an RNA product that contains an embedded promoter sequence. This sequence corresponds to the promoter element at the DNA level that was responsible for directing the synthesis of the RNA. The RNA also contains structural features that are necessary for its conversion to a new extended DNA hairpin that contains the same promoter sequence. Because the promoter sequence of RNA Z is heritable, the replication mechanism of this selfish RNA can be adapted for the *in vitro* evolution of RNAP promoters.

We have developed a continuous *in vitro* evolution procedure, based on the amplification mechanism of RNA Z, that allows rapid optimization of RNAP promoters (Figure 2). We prepared synthetic DNA oligonucleotides, corresponding to the cDNA amplification intermediate of RNA Z, but containing degenerate or completely random sequences in place of the promoter (–) sequence. A pool of these synthetic cDNAs was first treated with reverse transcriptase to convert the randomized promoter (–) domain to a double-stranded element. Subsequent addition of either T7, T3, or SP6 RNAP led to amplification and progressive enrichment of molecules that contain promoter sequences that most efficiently initiate transcription. For the most part, the resulting "optimal" promoter sequences closely resemble those of the naturally-occurring promoter consensus sequences.

EXPERIMENTAL PROCEDURES

Materials. AMV and M-MuLV reverse transcriptases were obtained from Life Sciences and U. S. Biochemical, respectively. T7 RNAP was prepared as previously described (Davanloo et al., 1984) and was purified according to a procedure originally developed for SP6 RNAP (Butler & Chamberlin, 1982). SP6 RNAP was purchased from Pharmacia, and T3 RNAP was purchased from both Promega and Stratagene. Taq polymerase was obtained from Boehringer-Mannheim and *Sca*I from New England BioLabs. NTPs, dNTPs, and ddNTPs (2',3'-dideoxynucleoside triphosphates) were purchased from Pharmacia. Primer 1 (5'-GTGCCAAG-CTTGTGATAGGTAGATCTTTACAAATTTTCCC) and the synthetic DNA constructs I, II, and III were obtained from Operon Technologies and purified by PAGE and subsequent size-exclusion chromatography on Sephadex G-25.

Nucleic Acid Amplification. Standard amplification reaction mixtures contained 0.5 μM primer 1, 50 mM HEPES (pH 7.5, 23 °C), 50 mM NaCl, 15 mM $MgCl_2$, 5 mM DTT, 2 mM spermidine, 2 mM each NTP, 0.2 mM each dNTP, 0.25 μCi/μL [α-^{32}P]ATP, 5 units/μL T7 RNAP, and 5 units/μL M-MuLV RT, and were incubated at 37 °C for 1 h. Reactions were initiated by the addition of RNA Z molecules and were terminated by the addition of EDTA (30 mM final concentration). Serial transfer amplification reactions were conducted in a total volume of 100 μL; each reaction was initiated by introducing 1/10 000 of the previous reaction mixture.

Promoter Evolution. Evolution *in vitro* of phage T7, T3, and SP6 RNAP promoter sequences, employing DNA construct I, utilized 1 nmol of gel-purified synthetic DNA, which was converted to a double-stranded DNA hairpin by reverse transcriptase. Each 1-mL fill-in reaction contained 50 mM HEPES (pH 8.5, 23 °C), 40 mM KCl, 1 mM DTT, 2.5 mM $MgCl_2$, 0.5 mM each dNTP, and 5 units/μL M-MuLV RT, which was incubated at 37 °C for 1 h. The reaction products were extracted with phenol and with chloroform/isoamyl alcohol (24/1, v/v) and were ethanol-precipitated. The resulting DNA was transcribed in a 200-μL reaction mixture containing 50 mM Tris-HCl (pH 7.5, 23 °C), 15 mM $MgCl_2$, 2 mM spermidine, 5 mM DTT, 2 mM each NTP, and 15 units/μL of either T7, T3, or SP6 RNAP, which was incubated at 37 °C for 2 h. Subsequent amplification reactions were initiated by addition of 1 μL of the transcription mixture to a 100-μL volume containing primer 1, the appropriate RNAP, and the other components used in the standard amplification mixture (see above). Each round of promoter evolution was carried out at 37 °C for 2 h; serial transfer of 1/50 000 of the completed reaction mixture was performed to initiate the next round of amplification.

Amplification reactions with DNA construct II were initiated directly by addition of 7 pmol of synthetic DNA to a reaction mixture containing 50 mM HEPES (pH 7.5, 23 °C), 50 mM NaCl, 15 mM $MgCl_2$, 5 mM DTT, 2 mM spermidine, 2 mM each NTP, 0.2 mM each dNTP, and 500 units of M-MuLV RT, which was incubated at 37 °C for 15 min, followed by addition of primer 1 (0.5 μM) and either T7, T3, or SP6 RNAP (5 units/μL) in a final volume of 100 μL, which was incubated at 37 °C for 2 h. Amplification reactions with DNA construct III were carried out as described for construct II, except that 14 pmol of synthetic DNA was used. Amplification reactions were terminated by the addition of 6 μL of 0.5 M EDTA. RNA products were purified by ethanol precipitation and denaturing PAGE.

Sequence Analysis. The RNA products of amplification were sequenced by the dideoxy chain-termination method (Sanger et al., 1977) employing AMV reverse transcriptase. (5'-^{32}P)-labeled DNA primer 1 (~0.1 pmol, 3.5 μCi/pmol)

and gel-purified RNA template (~0.3 pmol) were annealed in a 9-μL volume containing 40 mM Tris-HCl (pH 8.0, 23 °C), 50 mM NaCl, and 5 mM DTT by heating to 65 °C for 5 min, slow-cooling to room temperature, and placing on ice. $MgCl_2$ (4 mM final concentration) and 20 units of AMV RT were added to give a total volume of 15 μL, from which five individual extension reaction mixtures were prepared. Extension reactions of 12 μL each contained either ddGTP, ddATP, ddTTP, and ddCTP (0.1 mM), or no ddNTP in addition to the four natural dNTPs (0.1 mM each), 15 mM Tris-HCl (pH 8.0, 23 °C), 20 mM NaCl, 2.5 mM $MgCl_2$, 2 mM DTT, and 0.3 unit of AMV RT, and were incubated at 37 °C for various times. The reactions were terminated by addition of a 6-μL stop solution [96% formamide, 7.5 mM EDTA, and 0.05% (w/v) each of bromophenyl blue and xylene cyanol] and analyzed by denaturing 6% PAGE, followed by autoradiography. Positions of sequence heterogeneity within the promoter domain, noted upon examination of autoradiograms, were confirmed by analysis with a phosphorimager (Molecular Dynamics).

Promoter Assays. The efficiency of an SP6 promoter sequence variant relative to the natural consensus SP6 promoter was determined by *in vitro* transcription of a DNA construct containing both the variant and natural consensus promoter elements. A double-stranded DNA template was created by annealing the synthetic oligomers "SP6-wt" (5′-CTGCAGAATTC*ATTAGGTGACACTATA*GAAATTATTATGGTGGAAGTTTAGGAGTTAGGAGTACTCCCGGCCGCACCGACC) and "SP6-mt1" (5′-CTGCAAAGCTT*ATTTATGAGACACTATA*GAAATTGGTCGGTGCGGCCGGGAGTACT) and extending their 3′ termini with Taq polymerase (nucleotides in italics denote the promoter region; underlined nucleotides are complementary between the two oligomers). Annealing reactions contained 50 pmol each of SP6-wt and SP6-mt1, 10 mM Tris-HCl (pH 8.3, 23 °C), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin (v/v), and 0.2 mM each dNTP in 100 μL, which were heated in a thermal cycler (Ericomp) to 92 °C for 4 min and then to 85 °C for 4 min, during which time 0.5 μL of Taq polymerase (5 units/μL) was added. The resulting mixture was incubated at 65 °C for 1 min and at 72 °C for 10 min and then chilled on ice. The reaction products were digested with *Sca*I (cleaves the sequence 5′-AGT↓ACT, located at one end of the region of complementarity), and were transcribed in a 50-μL reaction mixture containing 2.5 pmol of each DNA fragment, 50 mM Tris-HCl (pH 7.5, 23 °C), 15 mM $MgCl_2$, 5 mM dithiothreitol, 2 mM spermidine, 2 mM each NTP, 10 μCi of [α-^{32}P]GTP, and 5 units/μL of SP6 RNAP, which was incubated at 37 °C for 1 h. The reaction products were separated by 10% denaturing PAGE, and the relative amounts of 35mer and 25mer RNA, produced from the wild-type and mutant promoter, respectively, were determined by phosphorimager analysis.

RESULTS

Optimization of RNA Z. RNA Z amplification reactions that were initiated with synthetic DNA constructs of our own design usually resulted in inefficient amplification of the desired construct, and soon led to the emergence of novel sequence variants with improved amplification efficiency. Typically, these variants contained a different hairpin sequence, but were unchanged throughout the remainder of the molecule (Breaker & Joyce, 1994). We allowed RNA Z to optimize its own sequence by carrying out serial transfer amplification reactions (see Experimental Procedures). Beginning with a synthetic DNA construct, we carried out 10 serial transfer reactions, providing an overall amplification factor of about 10^{40}. The resulting RNA was isolated by PAGE and sequenced by the dideoxy-mediated chain-termination method. The consensus sequence of the ensemble (Figure 3A) differed from the input sequence only in the hairpin domain. The optimized sequence remains stable through subsequent rounds of amplification. Furthermore, a progression of competitive serial transfer reactions, initiated with 0.5 pmol each of the optimized form and a previously-characterized form of RNA Z (Breaker & Joyce, 1994), demonstrated that the optimized form could replicate more efficiently, driving the competitor to extinction within two successive amplification reactions (data not shown). Accordingly, the sequence of the optimized RNA Z molecule was adopted as the starting point for *in vitro* evolution of bacteriophage RNAP promoters.

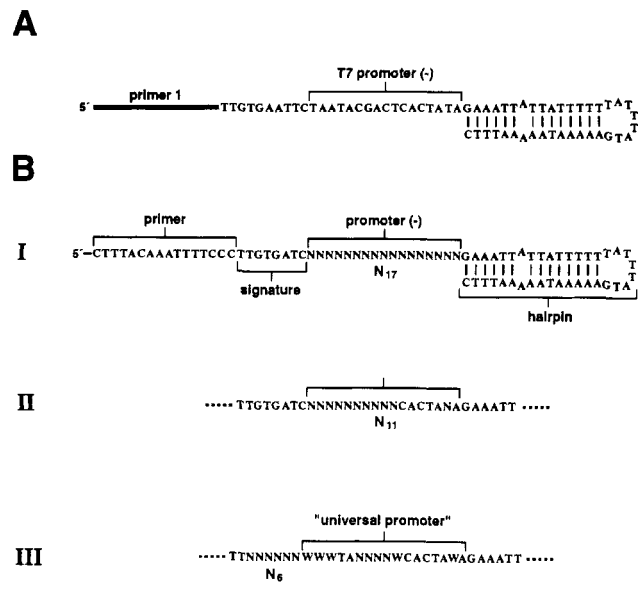

FIGURE 3: Constructs for *in vitro* evolution of promoters. (A) Optimized cDNA form of RNA Z; (B) randomized constructs used to initiate *in vitro* promoter evolution. Construct I contains 17 random nucleotides in place of an RNAP promoter (–) sequence (N = G, A, T, or C). Construct II contains 11 random nucleotides and 6 nucleotides of fixed identity in the promoter region. Construct III contains a mixture of non-, partially-, or fully-randomized positions (W = A or T) spanning both the 17-nucleotide promoter region and the 6 nucleotides that lie immediately upstream from the promoter. The "universal promoter" sequence fixes those nucleotides that are common to the promoter for T7, T3, and SP6 RNAP. Sequences not depicted for constructs II and III are identical to those given for construct I.

Design of Synthetic DNA Constructs. Three synthetic DNA molecules (constructs I, II, and III) were designed in accordance with the structure of the cDNA form of RNA Z, and were used to probe the sequence requirements of RNAP promoters (Figure 3B). Each cDNA construct is comprised of four domains: "primer", "signature", "hairpin", and "promoter". The primer domain (5′-terminal 16 nucleotides) is identical to nucleotides 25–40 of primer 1 (used to initiate cDNA synthesis; see Experimental Procedures). The signature domain, which lies between the primer and promoter domains, is not critical for efficient amplification, but provides an identifying tag sequence to distinguish these constructs from others that have been produced in our laboratory. In this way, we can distinguish the molecules being tested from those that might appear as a result of cross-contamination. The sequence of the hairpin domain is that which resulted from our optimization experiments (see above). The promoter

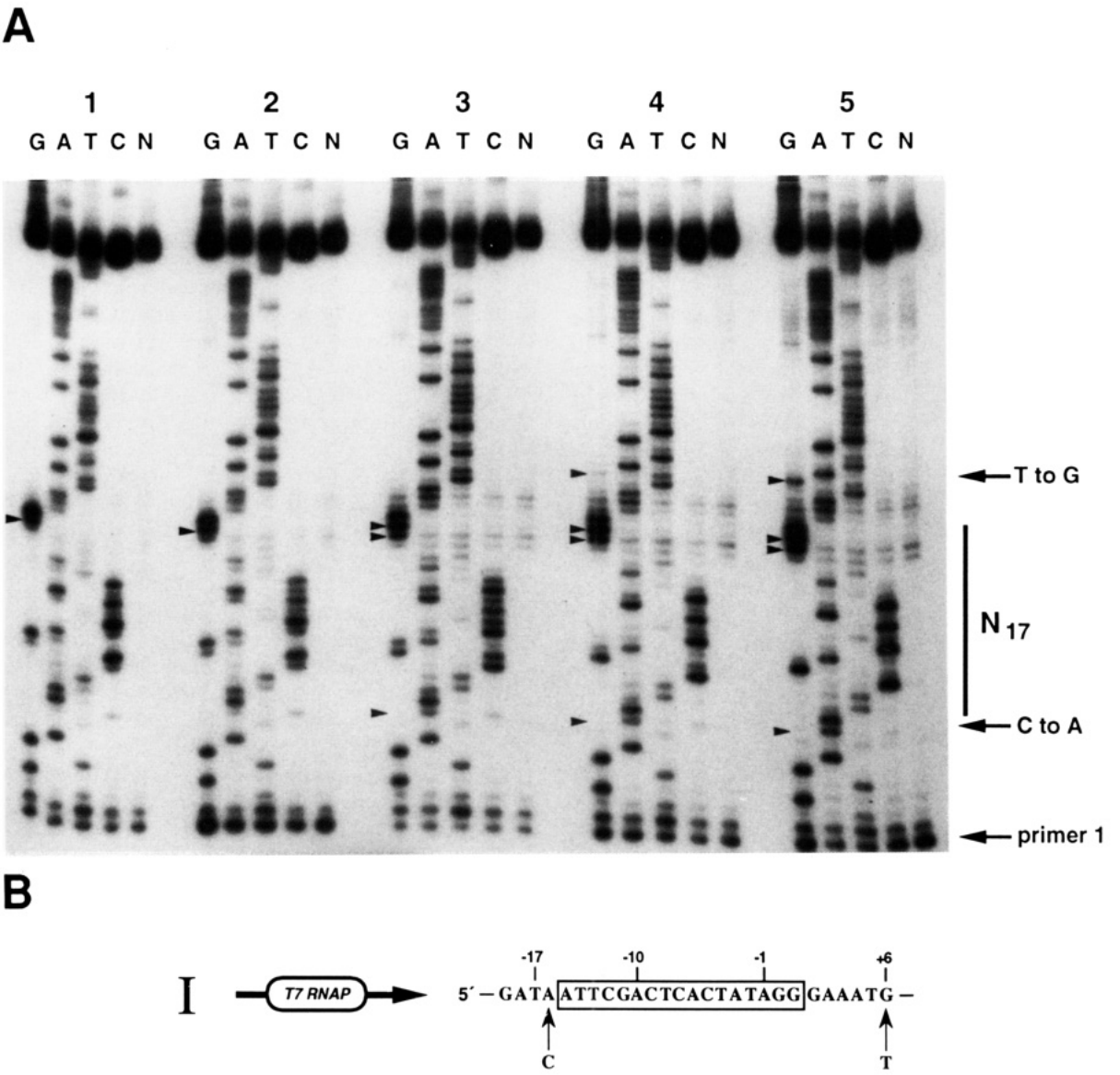

FIGURE 4: *In vitro* evolution of DNA construct I with T7 RNAP. (A) Sequence analysis of the RNA products over five successive amplification reactions. The vertical bar denotes the region corresponding to the 17-nucleotide promoter (–) sequence that was randomized in the original DNA construct. Arrowheads indicate unexpected mutations that arose over the course of evolution. (B) Sequence of the promoter (–) region of the cDNAs present after the fifth serial transfer reaction. Boxed nucleotides are the 17 positions that were initially randomized in construct I. Arrows indicate the two mutations that arose at nonrandomized positions.

domain consists either of random nucleotides (construct I) or of a mixture of random, degenerate, and fixed nucleotides (constructs II and III).

Promoter Evolution in Vitro. Construct I is a 79mer synthetic DNA that contains 17 random nucleotides in place of the T7 RNAP promoter (–) sequence (Figure 3B). A 1 nmol pool of construct I, containing roughly 35 000 copies of each of the 4^{17} ($\sim 10^{10}$) possible sequences, was extended by reverse transcriptase to generate a double-stranded promoter region. The resulting hairpin DNAs were transcribed in a reaction mixture containing either T7, T3, or SP6 RNAP. Serial transfer amplification reactions were then carried out, initiated with 1 μL of the 200-μL transcription reaction mixture, and perpetuated with 1/50 000 of the previous reaction mixture.

RNAs produced by transcription of construct I with T7 RNAP were amplified over five serial transfer reactions, providing an opportunity for enrichment of the most efficacious promoter sequences. The RNAs were isolated from each amplification mixture by denaturing 5% PAGE and were sequenced, as an ensemble, by the dideoxy-mediated chain-termination method using (5′-^{32}P)-labeled primer 1 (Figure 4A). The RNA population isolated after the first amplification reaction contains several sequence variants that, over the course of the five serial transfers, were culled to yield a single dominant sequence. This dominant sequence contains a promoter domain (positions –17 through –1) that is nearly identical to the natural promoter consensus sequence for T7 RNAP, with the exception of a single A to T change at position –13 (Figure 4B).

The dominant sequence obtained after five serial transfers contains three changes outside of the randomized region. First, the promoter domain of the evolved cDNA is followed by 3 G residues at positions +1 through +3 (Figure 4), while the original DNA construct I contains 17 random nucleotides in the promoter domain followed by only a single G residue (Figure 3B). Second, the signature domain of the evolved cDNA is 5′-TTGTGATA, compared to 5′-TTGTGATC in the original construct. Third, a single nucleotide in the hairpin domain has mutated from T to G (Figure 4A, transfer rounds

4 and 5). This mutation, which disrupts a potential T·A base pair in the hairpin domain, is not likely to be important for promoter function, but may confer a selective advantage compared to the original sequence at some other stage in the amplification cycle.

Acquisition of the two new G residues at the transcription start site occurs sequentially over the course of the serial transfer reactions (Figure 4A). It is highly unlikely that the gain of G residues at the transcription initiation site is the result of an insertion event, but rather is due to a 2-nucleotide shift in the position of the promoter domain relative to the original 17-nucleotide random region of construct I. This is consistent with the fact that there is no corresponding shift in the dideoxy-sequencing ladder of the signature and hairpin domains, which would result from an insertion event. Because of the shift in promoter position, the last two initially-random nucleotides are now located at positions +1 and +2, and have become fixed as G residues through the selective amplification process.

The shift in promoter position causes the promoter domain to begin prematurely, encroaching into the original signature domain. The last two nucleotides of the signature domain (TC) now comprise the first two nucleotides of the promoter (–17T, –16A; Figure 4B). The progressive C to A change at position –16 enables the shifted promoter domain to conform to the natural T7 RNAP promoter consensus sequence. This acquired mutation lies outside the region that was initially randomized in construct I, reflecting the ongoing possibility of mutation by the enzymes that are responsible for amplification.

Parallel amplification reactions were conducted employing construct I and either T3 or SP6 RNAP. Ensemble sequencing of the RNA populations isolated after three serial transfers revealed a high degree of similarity between the promoter sequences obtained by *in vitro* evolution and the natural promoter consensus sequence for each polymerase (data not shown). As was the case with T7 RNAP, when the pool was amplified with T3 RNAP, the resulting cDNAs exhibited two extra G residues at the transcription start site and a concomitant 2-nucleotide shift in promoter position. The cDNAs resulting from amplification with SP6 RNAP did not exhibit extra G residues at the initiation site or a shift in promoter position.

Refinement of Promoter Sequence Requirements. In order to fix the location of the promoter domain, we designed construct II, which contains 11 random nucleotides and 6 defined nucleotides within the promoter (–) region (Figure 3B). The six defined nucleotides, located at positions –7 through –3 (CACTA) and –1 (A), are strictly conserved in nearly all known phage RNAP promoter sequences (Diaz et al., 1993). Moreover, these specific nucleotides became fixed over the course of *in vitro* evolution experiments involving construct I and T7 RNAP (Figure 4) as well as T3 and SP6 RNAP (data not shown).

Amplification reactions involving construct II and either T7, T3, or SP6 RNAP were initiated with 7 pmol of synthetic DNA. This pool contains approximately 10^6 copies of each of the 4^{11} ($\sim 10^6$) possible sequence variants. Due to the high copy number of each sequence, the amplification reactions were initiated directly with synthetic DNA. The RNA products were isolated by PAGE after a single 2-h incubation, and were sequenced as an ensemble, as described above (Figure 5A).

The consensus sequence obtained for each RNAP as a result of *in vitro* evolution of construct II is identical to the natural consensus sequence for that polymerase. However, several nucleotide positions exhibit significant sequence heterogeneity (Figure 5B). Further serial transfer experiments did not resolve this heterogeneity, suggesting that the deviations from consensus do not greatly alter promoter efficiency. Interestingly, constructs amplified with T3, but not with T7 or SP6 RNAP, acquire a mutation in the nonrandomized region (A to G change at position +2), which allows transcription to initiate with two G residues (Figure 5). This same mutation arose in several replicate experiments.

The products obtained with construct II and SP6 RNAP contain two nucleotides within the promoter domain that are clearly heterogeneous in sequence, at positions –12 (G or T) and –10 (T, G, or A). The double mutant, with a G to T change at position –12 and a T to A change at position –10, was prepared and assayed for transcription efficiency in direct comparison with the natural promoter consensus sequence (see Experimental Procedures). This promoter variant was less than 50% as efficient compared to the natural promoter under our assay conditions, suggesting that the two single mutants may be more abundant than the double mutant in the population of amplifying molecules.

In vitro promoter evolution also was used to address the importance of the nucleotide sequence that lies immediately upstream from the 17-nucleotide promoter domain. Although the promoter is usually considered to begin at nucleotide position –17, chemical and enzymatic footprinting experiments have shown that T7 RNAP protects the promoter region as far upstream as position –21 (Ikeda & Richardson, 1986; Basu & Umadas, 1986; Gunderson et al., 1987). Moreover, the most efficient "class III" T7 RNAP promoters conform to the consensus sequence GAAAT at positions –22 through –18 (Dunn & Studier, 1983).

In vitro evolution of the region immediately upstream of the traditional promoter domain was conducted using construct III (Figure 3B), in which nucleotides –23 through –18 were completely randomized. The promoter domain of construct III was designed to contain a minimum amount of degeneracy while accommodating the natural consensus promoters for T7, T3, and SP6 RNAP, so that all three promoters could be evolved from a common starting pool of synthetic DNA. The amplification reactions were initiated with 14 pmol of DNA, representing roughly 10^5 copies of each of the 10^7 possible sequence variants. Ensemble sequencing of the RNA products after a single amplification reaction with either T7, T3, or SP6 RNAP revealed no significant preference in the nucleotide composition of the –23 through –18 region (data not shown). Subsequent serial transfer reactions with T7 RNAP failed to distinguish any preferred sequence, with the ensemble of amplifying molecules continuing to exhibit a roughly equimolar mixture of the four nucleotides at each of these positions.

DISCUSSION

Optimal Bacteriophage Promoters. The promoter sequences obtained in this study for the RNA polymerases of T7, T3, and SP6 bacteriophage are very similar to their respective natural consensus sequences. *In vitro* evolution of the T7 RNAP promoter, beginning with construct I, gave rise to transient replicating molecules whose sequences differ from the natural promoter consensus sequence by an A to T change at position –13 and an A to C change at position –16 (transfer rounds 1–3, Figure 4a). A T7 RNAP promoter that contains a nonconsensus T at position –13 is known to function with ~65% efficiency compared to the consensus promoter (Diaz et al., 1993). The A to C change at position –16 can be

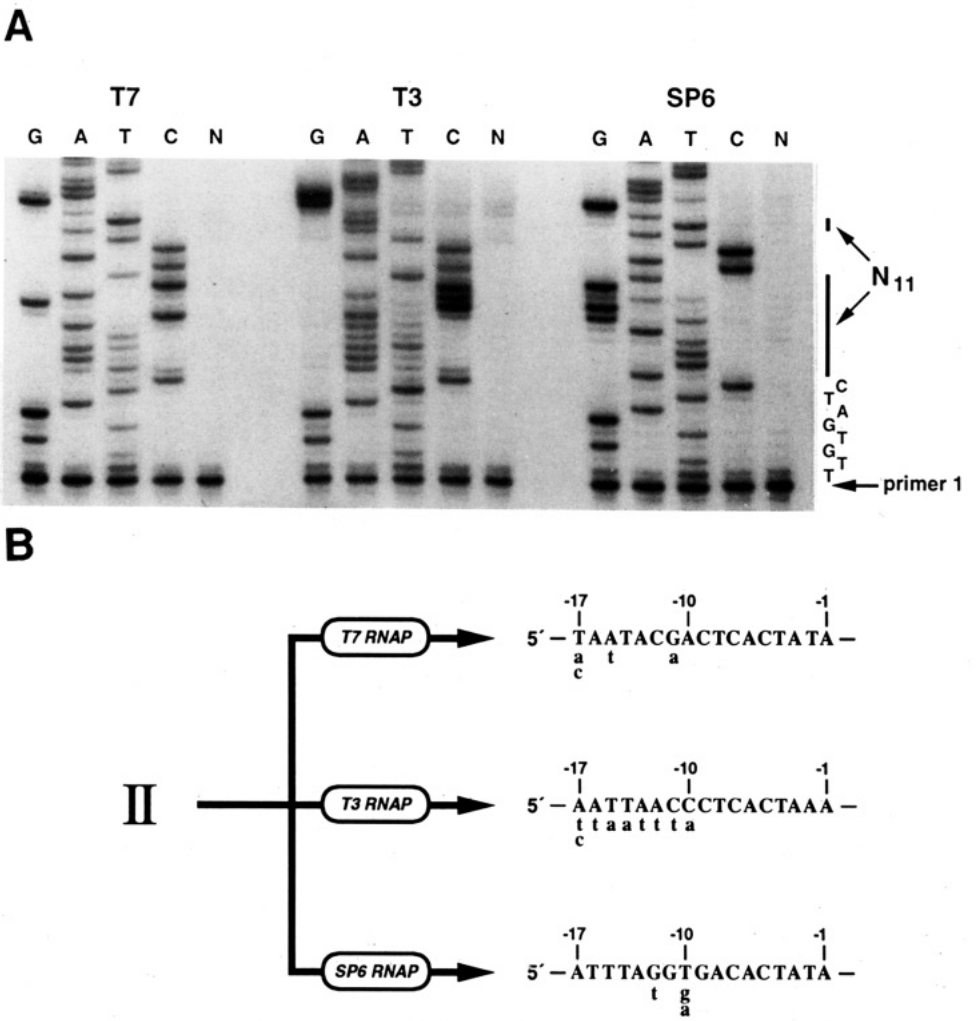

FIGURE 5: Polymerase-specific differentiation of DNA construct II. (A) Sequence analysis of the population of RNAs obtained through promoter evolution with either T7, T3, or SP6 RNAP. Vertical bars denote the 11 nucleotides that were randomized in the original DNA construct. (B) Sequence of the promoter (–) region of cDNAs present after the amplification reaction for each of the three RNAPs. Uppercase letters denote the consensus sequence for each RNAP promoter; lowercase letters indicate other frequent nucleotides at positions of heterogeneity.

attributed to the shift in promoter position, which caused two upstream nucleotides to become part of the promoter. In our case, T7 RNAP transiently tolerates a nonconsensus nucleotide at position –16 in order to shift the promoter by two nucleotides and accommodate G residues at the last two random positions. Previous studies have shown that T7 RNAP operates most efficiently when there are consecutive G residues at positions +1 through +3 (Milligan et al., 1987). This is consistent with the observation that in natural sequences G residues are highly favored in the +1 through +3 positions for initiation of transcription by either T7 or T3 RNAP (Diaz et al., 1993). Eventually, the inherent mutation rate of the amplification process generates a variant that restores the consensus nucleotide at position –16, which then grows to dominate the population (transfer rounds 3–5, Figure 4a). Such events reflect the evolutionary nature of our method for *in vitro* optimization of promoters.

Amplification of construct I by T3 RNAP resulted in an ensemble of promoter sequences that differ from the natural consensus sequence by several nucleotides (data not shown). These differences are also most likely due to the shift in promoter position relative to the initially-randomized promoter domain. Amplification of construct I with SP6 RNAP, on the other hand, did not result in a shift in promoter location. The natural consensus sequence for SP6 RNAP is GAA at positions +1 through +3, matching the sequence of DNA construct I.

The promoter consensus sequences of T7, T3, and SP6 RNAP, whether natural or derived through *in vitro* evolution with construct I, contain common nucleotides at positions –7 through –3 (CACTA) and –1 (A). This observation was used to rationalize the design of construct II, which fixes these nucleotides while randomizing the remaining 11 nucleotides of the promoter, thereby preventing the promoter from shifting and forcing the polymerases to accept a single G at the site of initiation. Indeed, amplification of construct II with either T7, T3, or SP6 RNAP did not result in a shift in promoter position (Figure 5). Interestingly, however, reactions with T3 RNAP invariably gave rise to variants that had added a second G residue at the initiation site, suggesting either a greater propensity for mutation during amplification with T3 RNAP or an enhanced selective advantage of the second G residue in the context of T3 compared to T7 or SP6 RNAP.

The proportion of each nucleotide at positions of sequence heterogeneity cannot be determined precisely by the method of ensemble sequencing. However, significant amounts of heterogeneity can be recognized by autoradiography and phosphorimager analysis. The results we obtained with construct II and T7 RNAP can be compared to a previous systematic analysis of the effects on promoter efficiency of single base mutations at nucleotide positions –15 through –6 (Diaz et al., 1993). *In vitro* evolution yields a T7 RNAP promoter consensus that contains three positions of obvious sequence heterogeneity: –17 (T, A, or C), –15 (A or T), and

−11 (G or A) (Figure 5B). These findings are consistant with the observation that only T can substitute for the natural consensus A at position −15, resulting in ~65% trancription efficiency compared to the natural sequence (Diaz et al., 1993). Substituting either G or C at this position results in only ~3% relative transcription efficiency. Likewise, only A can substitute for the natural consensus G at position −11, resulting in ~70% relative efficiency, whereas substituting either C or T gives <3% relative efficiency.

In our experimental system, the sequence that lies immediately upstream from the promoter region is not important for efficient transcription by any of the RNAPs that were tested. At the outset of this study, we wondered whether any variant of the natural consensus sequence, containing mutations either within or upstream of the promoter domain, would prove to be more efficient than the natural sequence in an *in vitro* context. This is clearly not the case for T7, T3, and SP6 RNAP. The natural consensus sequence is already optimal, or nearly optimal, for eliciting maximal transcription efficiency by the polymerase. This does not exclude the possibility that a mutant form of the polymerase, in combination with a variant form of the promoter, might allow a higher level of transcription. The catalytic properties and promoter sequence requirements of various mutant forms of T7 RNAP have been characterized (Raskin et al., 1993; Ikeda et al., 1993). Continuous *in vitro* promoter evolution could be used to define the optimal promoter sequence for these and other mutant RNAPs.

Continuous in Vitro Evolution. The first continuous *in vitro* evolution experiments involving nucleic acids were carried out by Speigelman and co-workers (Mills et al., 1967), who used Qβ replicase, the RNA-dependent RNA polymerase from bacteriophage Qβ. This enzyme selectively amplifies the genomic RNA of bacteriophage Qβ, which is identifiable to the enzyme because of its characteristic structural features (Nishihara et al., 1983). Extensive serial transfer amplification of Qβ genomic RNA with Qβ replicase resulted in the evolution of small variant RNAs that amplified more efficiently than the input RNA (Mills et al., 1967). Related experiments were conducted to evolve RNA templates that are efficiently amplified by Qβ replicase under conditions of limiting substrate availability or in the presence of the intercalating agent ethidium bromide (Spiegelman, 1971). These *in vitro* evolution studies did not focus on a specific attribute of the RNA, but rather on its overall fitness in the context of the Qβ amplification system.

In contrast, the continuous *in vitro* evolution scheme outlined in Figure 2 specifically selects those RNA and corresponding DNA molecules that encode a functional RNAP promoter element. The population of molecules becomes enriched, through differential amplification, with those variants that carry the most efficient promoters. Other sequence and structural features of RNA Z-like molecules may have some influence on amplification efficiency. However, by judicious design of the synthetic DNA constructs that were used to initiate the amplification process, we have focused evolutionary change primarily on the promoter region.

A number of protocols have been developed for the *in vivo* selection of both *Escherichia coli* RNA polymerase promoters (Horwitz & Loeb, 1986, 1988; Oliphant & Struhl, 1988) and T7 RNAP promoters (Ikeda et al., 1992a,b). A pool of plasmids, each carrying a variant RNAP promoter located upstream from a reporter gene, is used to transform bacterial cells. Functional promoters then are identified on the basis of the phenotype of individual bacterial colonies, which express the gene in proportion to the efficiency of the adjacent promoter. *In vivo* selection allows evaluation of promoter efficacy in a native context. However, the number of sequence variants that can be surveyed by these methods is limited by the efficiency and scale of bacterial transformation.

The replication cycle of RNA Z can be exploited for the continuous *in vitro* evolution of bacteriophage RNAP control elements, resulting in the rapid determination of optimal promoter sequences. Experiments of this type may also be utilized to investigate the promoter and enhancer sequence requirements of other RNA polymerases and to complement studies employing systematic mutational analysis.

REFERENCES

Basu, S., & Umadas, M. (1986) *J. Mol. Biol. 190*, 425–437.

Breaker, R. R., & Joyce, G. F. (1994) *Proceedings of the NATO Conference: Self-reproduction of Supramolecular Structures*, Maratea, Italy, September 12–16, 1993 (in press).

Breaker, R. R., & Joyce, G. F. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 6093–6097.

Butler, E. T., & Chamberlin, M. J. (1982) *J. Biol. Chem. 257*, 5772–5778.

Davanloo, P., Rosenberg, A. H., Dunn, J. J., & Studier, F. W. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2035–2039.

Diaz, G. A., Raskin, C. A., & McAllister, W. T. (1993) *J. Mol. Biol. 229*, 805–811.

Dunn, J. J., & Studier, F. W. (1983) *J. Mol. Biol. 166*, 477–535.

Guatelli, J. C., Whitfield, K. M., Kwoh, D. Y., Barringer, K. J., Richman, D. D., & Gingeras, T. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1874–1878.

Gunderson, S. I., Chapman, K. A., & Burgess, R. R. (1987) *Biochemistry 26*, 1539–1546.

Horwitz, M. S. Z., & Loeb, L. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7405–7409.

Horwitz, M. S. Z., & Loeb, L. A. (1988) *J. Biol. Chem. 263*, 14724–14731.

Ikeda, R. A., & Richardson, C. C. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3614–3618.

Ikeda, R. A., Ligman, C. M., & Warshamana, S. (1992a) *Nucleic Acids Res. 20*, 2517–2524.

Ikeda, R. A., Warshamana, G. S., & Chang, L. L. (1992b) *Biochemistry 31*, 9073–9080.

Ikeda, R. A., Chang, L. L., & Warshamana, G. S. (1993) *Biochemistry 32*, 9115–9124.

Joyce, G. F. (1992) in *Antisense RNA and DNA* (Murray, J. A. H., Ed.) pp 353–372, Wiley-Liss, Inc., New York.

Maslak, M., & Martin, C. T. (1993) *Biochemistry 32*, 4281–4285.

Miele, E. A., Mills, D. R., & Kramer, F. R. (1983) *J. Mol. Biol. 171*, 281–295.

Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987) *Nucleic Acids Res. 15*, 8783–8798.

Mills, D. R., Peterson, R. L., & Spiegelman, S. (1967) *Proc. Natl. Acad. Sci. U.S.A. 58*, 217–224.

Nishihara, T., Mills, D. R., & Kramer, F. R. (1983) *J. Biochem. 93*, 669–674.

Oliphant, A. R., & Struhl, K. (1988) *Nucleic Acids Res. 16*, 7673–7683.

Raskin, C. A., Diaz, G. A., & McAllister, W. T. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3147–3151.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Spiegelman, S. (1971) *Q. Rev. Biophys. 4*, 213–253.

Wu, Y., Zhang, D. Y., & Kramer, F. R. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 11769–11773.