

# T7 RNA Polymerase Mediates Fast Promoter-Independent Extension of Unstable Nucleic Acid Complexes<sup>†</sup>

Hani S. Zaher and Peter J. Unrau\*

Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada

Received February 5, 2004; Revised Manuscript Received April 13, 2004

**ABSTRACT:** T7 RNA polymerase is a processive, DNA-dependent RNA polymerase that has a high specificity for its 17 base pair (bp) promoter. In addition to normal transcription, the enzyme can produce anomalous transcripts in the absence of a promoter. We report here the systematic characterization of the transient aspects of this promoter-independent process. Oligonucleotides that are able to form transient unimolecular loop structures closed by as little as one Watson–Crick base pair between the 3′ terminal residue and an internal nucleotide proved to be viable substrates. A single nucleotide triphosphate assay system found that incorporation was encoded by the nucleotide 5′ to the predicted transient base pair. When this coding nucleotide was identical to the internal nucleotide participating in the transient base pair, multiple nucleotide incorporations were observed and could only be explained by a continuous shifting and resetting of the transient base-pairing required for extension. This intermittent extension process can be quite efficient. Short DNA or RNA substrates were good substrates for the enzyme (affinities ranged from 2 to 43  $\mu\text{M}$ ) and were extended rapidly with apparent catalytic rates of up to 240  $\text{min}^{-1}$  being observed, only 2-fold slower than the rate of transcriptional initiation. Our data suggest a possible mechanism for this promoter-independent extension activity and may add to the understanding of viral RNA replicative strategies.

T7 RNA polymerase (T7 RNAP)<sup>1</sup> is a DNA-dependent RNA polymerase consisting of a single polypeptide chain of 883 amino acids (1) that requires no additional transcription factors, making it ideal for the *in vitro* preparation of RNA from synthetic or nonsynthetic DNA templates containing the promoter sequence (2, 3). After binding with nanomolar affinity to its promoter sequence (4, 5), T7 RNAP forms a stable elongation complex capable of sustained, high-speed processive transcription (6–9). The transition from promoter binding to elongation is initiated by the slower synthesis of a short RNA fragment (10–12 nt long) that induces a substantial rearrangement of the N-terminal region of the polymerase (6–9). This process is imperfect, and short abortive and 5′-inhomogeneous transcripts are often created by the enzyme failing or incorrectly transitioning from an initiation to elongation complex (3, 10, 11).

Given the high specificity of the T7 promoter, it is surprising that RNA synthesis can still occur in the absence of a T7 promoter sequence. RNA sequences that can self-replicate (RNA X) in the presence of T7 RNAP have been reported by Konarska and Sharp (12, 13). It has also been

shown that the extension of transcribed RNA can occur by RNA priming using RNA as a template (14). The synthesis of these RNAs, as inferred from their sequence, seems to involve a complicated iterative process of extension and template remodeling (14). This activity appears to be quite unlike the highly processive elongation that occurs during regular transcription and is poorly characterized.

In this paper, we study the transient aspects of T7 RNAP promoter-independent polymerization by limiting the processive ability of the enzyme. By choosing appropriate oligonucleotide sequences and incubating them with individual nucleotide triphosphates, priming regions within an oligonucleotide can be targeted and their ability to sustain nucleotide incorporation strike evaluated. By limiting incorporation to a single nucleotide type, we were able to characterize the minimal substrate requirements for transient nucleotide incorporation and determine the basic kinetics of this process.

## MATERIALS AND METHODS

**T7 RNA Polymerase.** T7 RNA polymerase was purified from *E. coli* strain BL21 carrying the His-tagged plasmid pT7-911Q (15). The enzyme was purified as previously described (16) by immobilized nickel ion affinity chromatography (NTA resin, Qiagen). Protein concentration was determined by optical density at 280 nm with an extinction coefficient of  $1.4 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$  (17). The activity of our enzyme preparation compared favorably to that of commercial T7 RNAP (Roche) in a normal transcription assay, and the enzyme was found to have a specific activity of  $\sim 4.3$

<sup>†</sup> This work was supported by funds from the Michael Smith Foundation for Health Research and the Canadian Institutes of Health Research.

\* Correspondence should be addressed to this author: phone 604-291-3448; fax 604-291-5583; e-mail punrau@sfu.ca.

<sup>1</sup> Abbreviations: T7 RNAP, T7 RNA polymerase; nt, nucleotide; ds, double-stranded; ss, single-stranded; PAGE, polyacrylamide gel electrophoresis; NMP, nucleotide monophosphate; NTP, nucleotide triphosphate; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; PSTV, potato spindle tuber viroid.

Table 1: Kinetic Parameters Determined for Oligonucleotide Extension

substrate	sequence	$K_m$ ( $\mu\text{M}$ )	$k_{\text{app}}$ ( $\text{min}^{-1}$ )	Hill coefficient
promoter complex		$0.016 \pm 0.002$	$500 \pm 22^a$	$1.27 \pm 0.13$
$\text{G}_2\text{CAC}_2\text{C}$	GGCACCC	$43 \pm 4$	$2.0 \pm 0.1$	$0.96 \pm 0.10$
$\text{A}_{10}\text{G}_2\text{A}_2\text{C}_2\text{C}$	AAAAAAAAAAGGAACCC	$18 \pm 5$	$16 \pm 2$	$1.19 \pm 0.26$
$\text{A}_9\text{G}_3\text{A}_2\text{C}_2\text{C}$	AAAAAAAAAGGGAACCC	$41 \pm 6$	$240 \pm 10$	$1.18 \pm 0.15$
$\text{T}_{10}\text{G}_2\text{T}_2\text{C}_2\text{C}$	TTTTTTTTTGGTCC	$1.8 \pm 0.2$	$14 \pm 1$	$0.79 \pm 0.09$
	GGGCC	$6700 \pm 2400^b$	$48 \pm 2$	$1.99 \pm 0.10$

<sup>a</sup> Initiation. <sup>b</sup> Units of ( $\mu\text{m}$ )<sup>2</sup>.

$\times 10^7$  units/ $\mu\text{mol}$  [1 unit of enzyme can incorporate 1 nmol of CTP into acid-precipitable RNA products in 60 min at 37 °C at pH 8.0 (18)].

**Oligonucleotide Synthesis.** RNA oligonucleotides were purchased from Dharmacon and deprotected according to the company's protocol. RNA was dried by use of a SpeedVac and resuspended in water. DNA oligonucleotides were synthesized on an ABI 392 DNA synthesizer by Expedite chemistry.

**5'-Oligonucleotide Labeling.** Oligonucleotides at  $\sim 500$  nM were incubated in T4 polynucleotide kinase buffer (70 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, and 5 mM DTT, pH 7.6) supplemented with 74 kBq/ $\mu\text{L}$  [ $\gamma$ -<sup>32</sup>P]ATP (111 TBq/mmol, NEN) and 0.5 unit/ $\mu\text{L}$  T4 polynucleotide kinase (Invitrogen) at 37 °C for 30 min. The resulting radiolabeled oligonucleotides were purified by 15% preparative PAGE and excised from the gel. Oligonucleotides shorter than 10 nt were eluted in 1 mM EDTA for at least 12 h. The eluate was passed over an equilibrated (50 mM triethylammonium acetate, pH 7.6) C18 SPICE cartridge (Analtech). The cartridge was then washed with the same buffer, and the bound nucleic acids were eluted with 2.5 mL of acetonitrile. The recovered oligonucleotides were dried on a SpeedVac overnight. Oligonucleotides longer than 10 nt were eluted from gel fragments in 300 mM NaCl for at least 12 h and then ethanol-precipitated.

**Extension Reactions.** The standard assay was carried out in 40 mM Tris-HCl (pH 7.9), 2.5 mM spermidine, 26 mM MgCl<sub>2</sub>, 0.01% Triton X-100, 10 mM DTT, 8 mM GTP, 4 mM ATP, 4 mM CTP, and 2 mM UTP. Standard assays with single NTPs were performed at 4 mM, with oligonucleotides at 10 nM unless otherwise stated. Final enzyme concentration was 0.1  $\mu\text{M}$ . Incubations were carried out at 37 °C and stopped by the addition of an equal volume of 2 $\times$  formamide loading dye (95% formamide, 5 mM EDTA, 0.025% xylene cyanol, and 0.025% bromophenol blue). Reactions were resolved on 20% polyacrylamide sequencing gels unless otherwise specified.

**Kinetic Studies of Promoter-Dependent Initiation.** The method used was similar to that of Martin and Coleman (5). The T7 promoter complex was formed by annealing PT7 (5'-CTTTAATACGACTCACTATAGG) and CT7 (5'-AGTCCTATAGTGAGTC GTATTAAAG), each at 4  $\mu\text{M}$ , in TE (pH 7.8) at 94 °C for 2 min before cooling to room temperature.  $K_m$  and  $k_{\text{cat}}$  were determined by varying the T7 promoter complex concentration from 0.5 to 250 nM in the standard extension assay supplemented with 30 kBq/ $\mu\text{L}$  [ $\gamma$ -<sup>32</sup>P]GTP. Samples were incubated for 30 min, within the linear range as determined by a time course, to determine initial velocities. Products were resolved on 23% polyacrylamide sequencing gels and scanned on a Molecular Dynamics Storm 820 phosphorimager. The fraction of total radio-

label incorporation from all products was added and multiplied by the concentration of GTP to give the molarity of the initiated strands at a given time.  $K_m$  and  $k_{\text{app}}$  values were extracted by use of the Michaelis–Menten equation and GraphPad Prism V4.00.

**Kinetic Studies of Promoter-Independent Initiation.** For each oligonucleotide, initial velocities were measured at template concentrations spanning 1–500  $\mu\text{M}$  while enzyme concentration was kept fixed at 0.1  $\mu\text{M}$ . Initial apparent velocities were determined by use of incubation times that were short compared to the time required for trace amounts ( $\sim 10$  nM oligonucleotide concentration) of oligonucleotide to behave in a nonlinear fashion. The total number of CMP incorporations per oligonucleotide was determined by quantification of the resulting gels by use of a phosphorimager. The fractional amount of radioactivity in each extension product was determined and multiplied by the number of extensions required to produce that particular product. These weighted extension values were then summed and multiplied by the oligonucleotide concentration to reflect the total CMP incorporation at a given time and substrate concentration. The resulting initial apparent velocities were then used to extract  $K_m$  and  $k_{\text{app}}$  values by use of the Michaelis–Menten equation.

**dsDNA Inhibition of Extension.** A short template sequence, 5'-AAAGGGAACC (SC), complementary to the 3' end of T<sub>10</sub>G<sub>2</sub>T<sub>2</sub>C<sub>2</sub>C (see Table 1 for sequence), and a fully complementary sequence, 5'-AAAGGGAACCAAAAAAAA-AAA (LC), were titrated from  $\sim 1$  nM to 1  $\mu\text{M}$  in the presence of 10 nM 5'-labeled T<sub>10</sub>G<sub>2</sub>T<sub>2</sub>C<sub>2</sub>C in CTP or UTP extension buffer and annealed as before. Reactions were started by addition of enzyme and stopped with loading dye and a 10-fold excess of unlabeled T<sub>10</sub>G<sub>2</sub>T<sub>2</sub>C<sub>2</sub>C.

## RESULTS

A short RNA oligonucleotide, 5'-r(CUGCCAA), was observed to be specifically extended by GMP upon incubation in the presence of T7 RNAP and GTP (Figure 1). A deoxynucleotide having the sequence 5'-CTGCCAA was also only extended by GMP (extension was  $\sim 1.5$  times faster than the RNA sequence). The extension of these and other oligonucleotides was not dependent on enzyme preparation, as commercial T7 RNAP (Roche) had transcriptional and extension properties nearly identical to those of our enzyme. Both enzyme preparations exhibited a slow exonuclease activity of 0.001–0.01  $\text{min}^{-1}$  oligonucleotide<sup>-1</sup> at 0.1  $\mu\text{M}$  enzyme concentration, resulting from either a contaminating nuclease or intrinsic T7 RNAP activity (Figure 1) (19). The appearance of a faint second band in the GTP lane for both the RNA and DNA sequences (Figure 1) suggested to us the possibility that these constructs form extendable loop

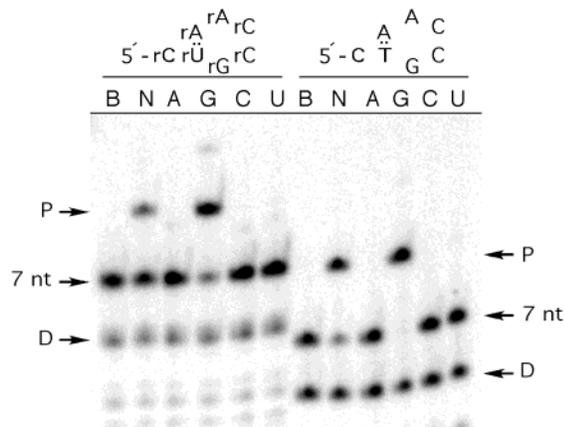


FIGURE 1: Extension reactions of short RNA and DNA oligonucleotides having analogous sequences. Sequence and proposed transient extension structures are indicated. Extension by T7 RNAP in the presence of no NTPs (B), all four NTPs (N), ATP (A), GTP (G), CTP (C), or UTP (U) is shown. Extension is only observed in the presence of GTP (indicated by P). A weak second extension suggests the possible formation of a rG•rU (or rG•T) wobble bp. A weak nuclease activity creates a 6-nt degradation product (indicated by D) after incubation in standard conditions for 45 min.

structures primed by a single base pair between the second and the most terminal residue in the oligonucleotide and using the most 5' residue as a template. This arrangement might, if the template could slip relative to the extension product, allow the incorporation of a second GMP residue by priming from the rG•rU (rG•T) wobble pair created by the first nucleotide incorporation.

*A Single Arbitrary Base Pair Can Initiate Polymerization.*

To test this hypothesis, a series of mutants derived from the DNA sequence 5'-CTGCCAA were synthesized and screened in our single nucleotide triphosphate T7 RNAP extension assay. The DNA sequence 5'-GGCACCC [named G<sub>2</sub>CAC<sub>2</sub>C, with residues in boldface type indicating the hypothetical loop between G(2) and C(7)], was found to specifically incorporate CMP and no other nucleotide. However, in contrast to the sequence 5'-CTGCCAA, numerous CMP incorporations were observed (Figure 2a). Single and double mutations of G<sub>2</sub>CAC<sub>2</sub>C were synthesized that either preserved or destroyed this hypothetical G(2)•C(7) bp. Mutations that preserved base-pairing were found to be active in our single nucleotide primer-extension assay, while mutations that disrupted base-pairing had no detectable extension activity (Figure 2a). Multiple extensions were not observed when the initial base pair was changed to A(2)•T(7), presumably because freshly incorporated rC(8) cannot hybridize with the residue A(2).

*The Incorporated Nucleotide Was Encoded by the Template.* All 7-nt constructs appeared to use the terminal 5' residue of the oligonucleotide to code for nucleotide extension. Constructs that varied the proposed coding nucleotide from G(1) in G<sub>2</sub>CAC<sub>2</sub>C to A, C, or T, while preserving the proposed G(2)•C(7) bp, all showed extension by the correct complementary nucleotide monomer when single nucleotide triphosphates were fed to the enzyme (Figure 2b). Extension in the presence of all four NTPs often took place at a reduced rate or was not observed. The exact dependence of this effect with respect to template sequence was not fully explored.

*Unimolecular and Bimolecular Extension Depends on Template Choice and Concentration.* To investigate the

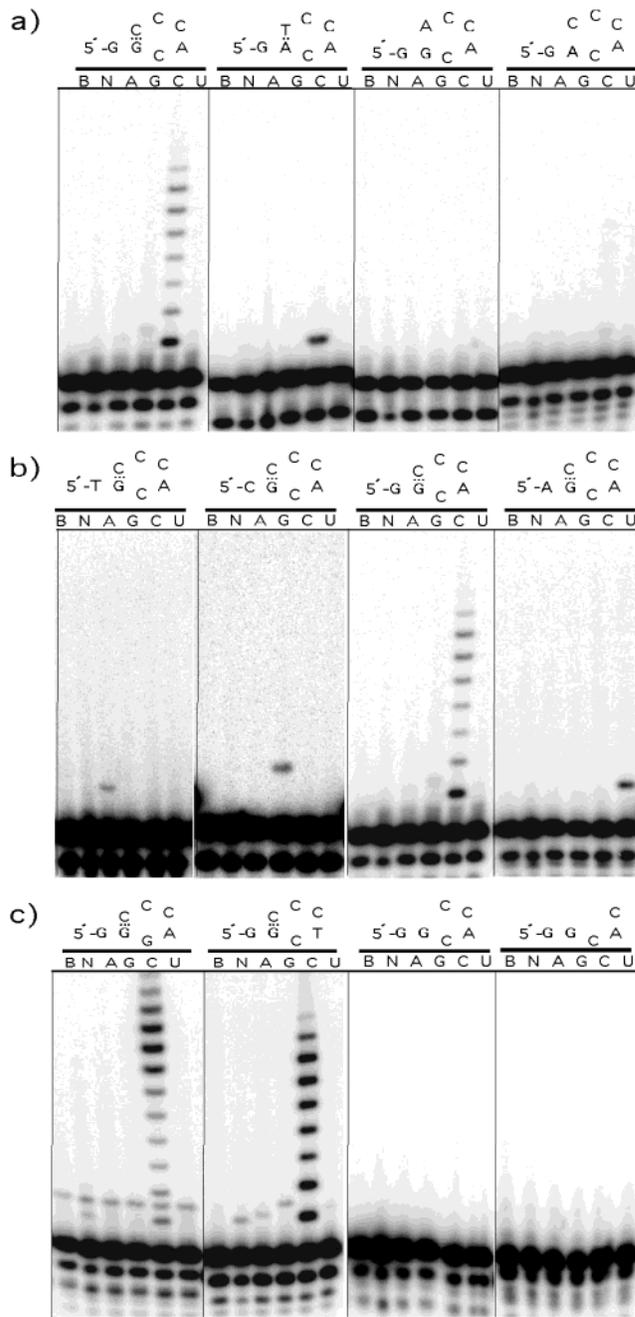


FIGURE 2: Extension of sequence variants of 5'-GGCACCC (G<sub>2</sub>CAC<sub>2</sub>C). Sequences and proposed transient extension structures are shown. (a) Mutations of the transient base pair G(2)•C(7). Extension is only observed when the bp is preserved. (b) Mutations of the coding nucleotide N(1) from left to right: T, C, G, and A. Extension is only observed in the presence of the correct complementary NTP. Multiple extensions are observed only for the 5'-GGCACCC construct in the presence of CTP. (c) Mutations of the loop sequence can result in increased extension rates, while point deletions eliminated extension. All panels show extension by T7 RNAP in the presence of no NTPs (B), all four NTPs (N), ATP (A), GTP (G), CTP (C), or UTP (U). Oligonucleotides were incubated under standard conditions for 45 min.

nature of the primer-template complex further, the kinetics of G<sub>2</sub>CAC<sub>2</sub>C extension were measured. The initial rate of nucleotide incorporation per enzyme climbed rapidly with substrate concentration, plateauing at a  $k_{app}$  of 2.0 min<sup>-1</sup>. A Hill coefficient very close to 1 was observed (Figure 3, coefficient = 0.96 ± 0.10, Table 1). Consistent with the apparent unimolecularity of the G<sub>2</sub>CAC<sub>2</sub>C extension, no

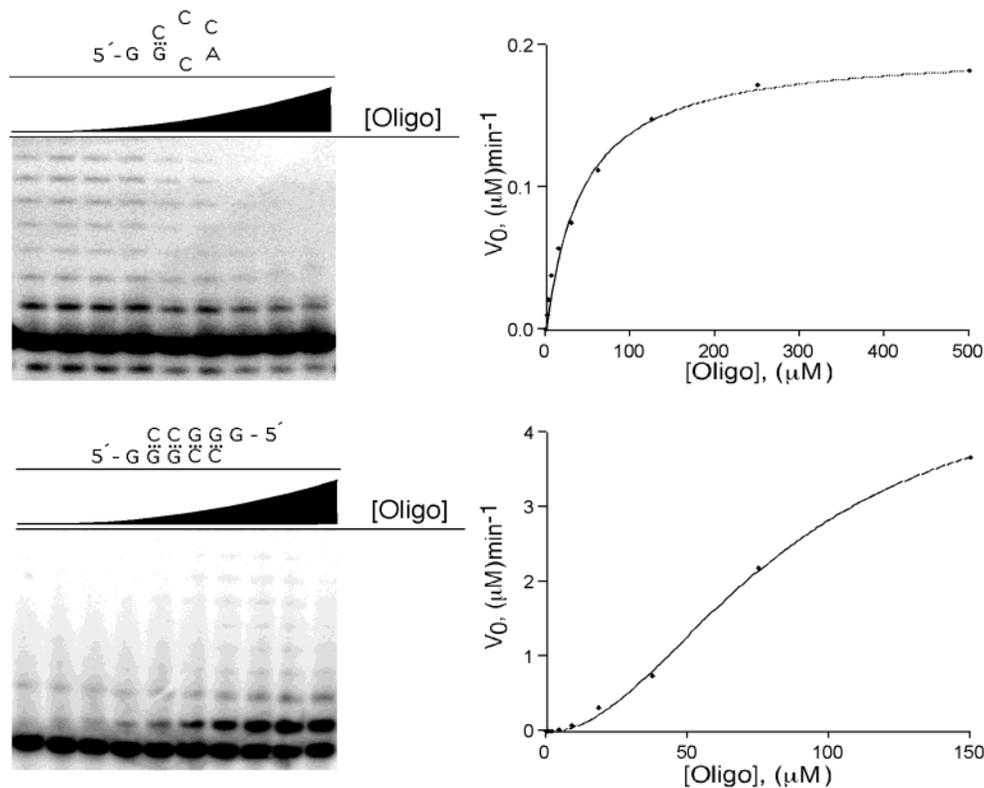


FIGURE 3: Kinetics of CTP extension for unimolecular- and bimolecular-type sequences. (a) The concentration of 5'-GGCACCC was titrated from 2 to 500  $\mu\text{M}$ , while the total amount of radioactivity within each lane was kept fixed. A Hill coefficient of  $0.96 \pm 0.10$  fits well to the data. (b) A self-complementary oligo, 5'-GGGCC, was also titrated from 0.3 to 150  $\mu\text{M}$ . In this case sigmoidal kinetics were observed with a Hill coefficient of  $1.99 \pm 0.10$  fitting the data.

activity was observed for truncated forms of  $\text{G}_2\text{CAC}_2\text{C}$  where the loop closing the proposed  $\text{G}(2)\cdot\text{C}(7)$  bp interaction was decreased from 4 to 3 or 2 nt (Figure 2c). Mutations that preserved loop size resulted in similar or even enhanced activity. In contrast, a 5 nt DNA construct designed to anneal with a copy of itself and encode CMP incorporation (5'-GGGCC) resulted in completely different kinetics. This construct displayed a sigmoidal rate dependence with concentration, resulting in a Hill coefficient very close to 2 and a  $k_{\text{app}}$  of 48  $\text{min}^{-1}$  (Figure 3b, coefficient =  $1.99 \pm 0.10$ , Table 1).

**Extension Kinetics,  $K_m$  and  $k_{\text{app}}$  Determination for Short Oligonucleotide Substrates.** To explore the extension properties of substrates containing coding nucleotides embedded within a longer homogeneous sequence,  $\text{A}_{10}\text{G}_2\text{A}_2\text{C}_2\text{C}$ ,  $\text{A}_9\text{G}_3\text{A}_2\text{C}_2\text{C}$ , and  $\text{T}_{10}\text{G}_2\text{T}_2\text{C}_2\text{C}$  (see Table 1 for sequences) were synthesized and compared to  $\text{G}_2\text{CAC}_2\text{C}$ . The effects of length, base-pairing, and sequence composition on  $K_m$  and  $k_{\text{app}}$  for CMP incorporation were examined. Time courses were carried out for each oligonucleotide by the standard extension assay (Figure 4). All of the oligonucleotides exhibited an initial rapid extension for 8–9 nt, followed by a much slower but sustained extension. Total CMP incorporation (with 10 nM substrate and 0.1  $\mu\text{M}$  enzyme) was linear with respect to time early in the time course. Under these conditions,  $\text{G}_2\text{CAC}_2\text{C}$ ,  $\text{A}_{10}\text{G}_2\text{A}_2\text{C}_2\text{C}$ ,  $\text{A}_9\text{G}_3\text{A}_2\text{C}_2\text{C}$ , and  $\text{T}_{10}\text{G}_2\text{T}_2\text{C}_2\text{C}$  had nucleotide incorporation rates per oligonucleotide of 0.007  $\text{min}^{-1}$ , 0.055  $\text{min}^{-1}$ , 1.5  $\text{min}^{-1}$ , and 1.7  $\text{min}^{-1}$  for up to 4 h, 30 min, 10 min, and 5 min, respectively.

For each oligonucleotide, initial velocities were measured at template concentrations spanning 1–500  $\mu\text{M}$  while enzyme concentration was kept fixed at 0.1  $\mu\text{M}$ . Initial

apparent velocities were determined by use of incubation times that were short compared to the time required for trace amounts of oligonucleotide to behave in a nonlinear fashion ( $\text{G}_2\text{CAC}_2\text{C}$  for 30 min,  $\text{A}_{10}\text{G}_2\text{A}_2\text{C}_2\text{C}$  for 10 min, and both  $\text{A}_9\text{G}_3\text{A}_2\text{C}_2\text{C}$  and  $\text{T}_{10}\text{G}_2\text{T}_2\text{C}_2\text{C}$  for 2 min). The best fit to the Michaelis–Menten equation was obtained for each construct. The  $k_{\text{app}}$  values were found to be in the range of 2–240  $\text{min}^{-1}$ , while the  $K_m$  values were in the 10  $\mu\text{M}$  range (Table 1). A three-parameter fit was used to determine Hill coefficients. In all cases the observed Hill coefficients were close to 1, consistent with a unimolecular extension process.

These values were compared to the rate of initiation of the enzyme in our reaction conditions by an assay developed by Martin and Coleman and co-workers (5, 20). We observed a  $K_m$  of 16 nM and a  $k_{\text{app}}$  of 500  $\text{min}^{-1}$  for a T7 promoter construct capable of runoff transcription (Table 1). While binding affinity was comparable, our catalytic rate of 500  $\text{min}^{-1}$  was 10 times higher than observed by Gardner et al., likely due to our use of saturating NTP conditions and slightly different buffer conditions (20). Extension experiments with  $\text{G}_2\text{CAC}_2\text{C}$  (10 nM concentration) were used to measure the binding affinity of CTP in our single-nucleotide extension assay. CTP was titrated from 0.16 to 20 mM, and a  $K_m(\text{CTP})$  of 0.65 mM was obtained after fitting of the data to Michaelis–Menten kinetics. Interestingly, this binding affinity is only slightly weaker than the equilibrium NTP binding observed for promoter-dependent initiation at +1 and +2 sites (21).

**Sequence and Pairing Ability Weakly Affects the Bounds of Extension.** Having established that constructs  $\text{G}_2\text{CAC}_2\text{C}$ ,  $\text{A}_{10}\text{G}_2\text{A}_2\text{C}_2\text{C}$ ,  $\text{A}_9\text{G}_3\text{A}_2\text{C}_2\text{C}$ , and  $\text{T}_{10}\text{G}_2\text{T}_2\text{C}_2\text{C}$  all had dramatically different extension rates as a function of sequence and

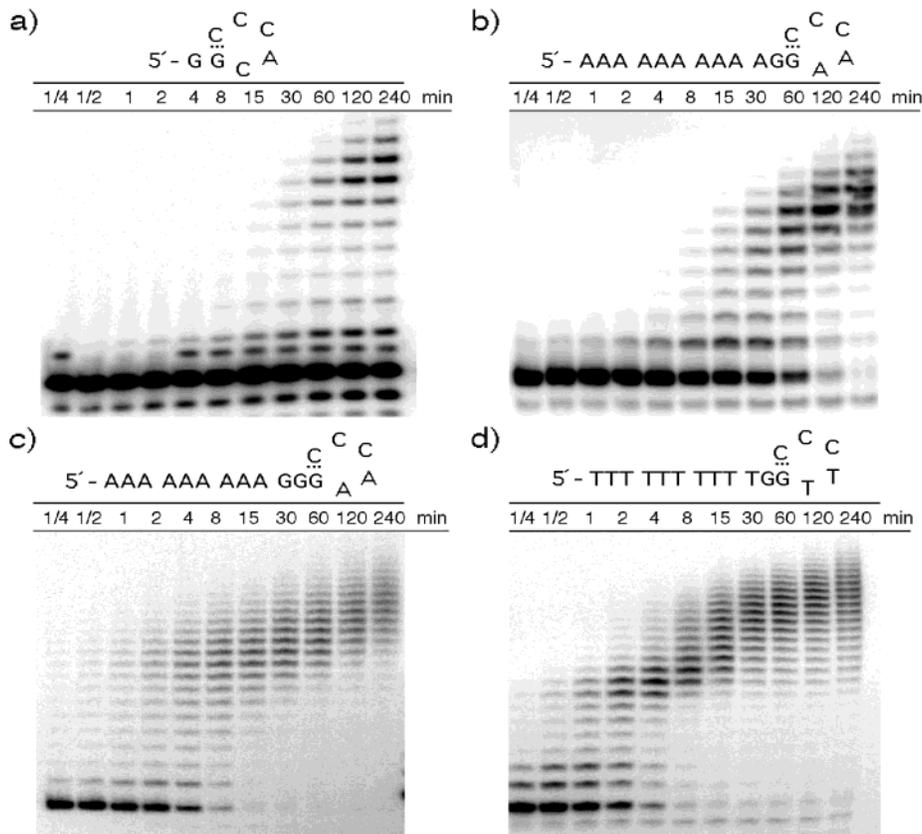


FIGURE 4: Time courses of CTP extension reactions of (a)  $G_2CAC_2C$ , (b)  $A_{10}G_2A_2C_2C$ , (c)  $A_9G_3A_2C_2C$ , and (d)  $T_{10}G_2T_2C_2C$ . These sequences, which have extension rates spanning 2 orders of magnitude (Table 1), all slow their extension rate after incorporation of 8–9 nucleotides. All constructs were incubated under standard conditions.

pairing ability, we explored the decrease in extension rate observed after 8–9 nt of extension in all of these constructs (Figure 4). Seven variants of  $A_{10}G_2A_2C_2C$  ( $AG_2L_l$  series), each having the same length and overall nucleotide composition, were synthesized. The initial loop size of these variants was systematically varied from 2 to 14 nt by moving the GG sequence element in the 5' direction along the poly(A) leader sequence. These constructs were incubated for 45 min in the presence of CTP by our standard assay. An initial loop length of 2 nt was found to be inactive (as would be expected for a unimolecular process). The remaining constructs all showed good initial extension followed by a sharp decrease in extension rate when the total loop size (initial DNA + RNA extension products) reached ~11–12 nt (Figure 5a). The effect of increasing the intrinsic pairing ability from 1 to 2 bp was studied by synthesizing variants of  $A_9G_3A_2C_2C$  ( $AG_3L_l$  series). Even though members of the  $AG_3L_l$  series were dramatically faster than those of the  $AG_2L_l$  series (Table 1, Figure 4), after 45 min of incubation the relative loop sizes of the  $AG_3L_l$  series were again approximately the same with constructs having loops ~14 nt long (Figure 5b).

The only other nucleotide incorporated into the  $A_{10}G_2A_2C_2C$  substrate by T7 RNAP was UMP. The enzyme quickly added more than 30 UMP nucleotides to the  $A_{10}G_2A_2C_2C$  substrate after only 45 min of incubation at a rate approximately 3 times as fast as CMP incorporation. When given all four NTPs,  $A_{10}G_2A_2C_2C$  appeared to be preferentially extended by UMP (as judged by gel mobility patterns) at a rate only slightly slower than that observed with UTP alone.

*Phosphorylation State of Oligonucleotide Substrates Does Not Affect Extension.* Since the measurement of  $K_m$  required the use of large amounts of oligonucleotide substrate and only a trace amount of this substrate was phosphorylated, we were concerned, particularly for small substrates such as  $G_2CAC_2C$ , that the presence or absence of a 5'-phosphate was important for activity. We therefore performed experiments with unlabeled  $G_2CAC_2C$  and  $A_9G_3A_2C_2C$  at 10  $\mu M$  oligonucleotide concentration. Reactions were incubated for 15 and 30 min ( $A_9G_3A_2C_2C$ ) or 1, 2, and 4 h ( $G_2CAC_2C$ ) in our standard CTP extension buffer. Reactions were stopped by heat inactivation of the enzyme at 65  $^{\circ}C$  for 5 min, 1000-fold dilution, and then radiolabeling with polynucleotide kinase and a 6-fold excess of [ $\gamma$ - $^{32}P$ ]ATP. The extension pattern observed was very similar to our standard pre-reaction labeling strategy, indicating that the phosphorylation state of the oligonucleotide was of minor or no importance.

*Competitive Inhibition Found by Use of Inactive Substrates.* The ability of the inactive substrate  $A_{12}G_2C_2C$  (Figure 5a) to compete with the active construct  $A_{10}G_2A_2C_2C$  was determined by adding it at 0, 25, 50, and 100  $\mu M$  concentrations to a series of  $A_{10}G_2A_2C_2C$  extension reactions. For each reaction  $A_{10}G_2A_2C_2C$  was titrated from 8 to 500  $\mu M$  (Figure 6a) in the presence of CTP.  $A_{12}G_2C_2C$  was found capable of inhibiting  $A_{10}G_2A_2C_2C$  extension, and a Lineweaver–Burk plot (Figure 6b) suggested that this inhibition was competitive. Apparent  $K_m$  values were obtained for each inhibitor concentration and plotted against  $A_{12}G_2C_2C$  concentrations (Figure 6c) to yield a  $K_i$  of 45  $\mu M$ . This result is

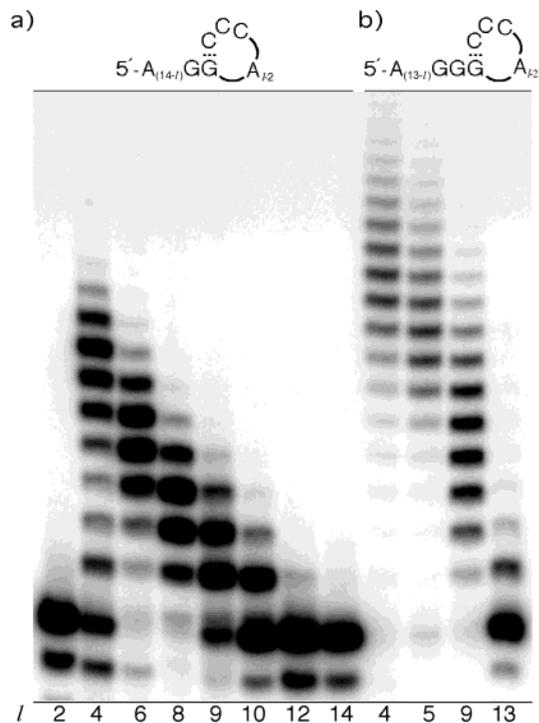


FIGURE 5: Effect of loop size and transient base-pairing ability on the extent of nucleotide incorporation. Standard CTP extension reactions of (a) substrates that can only form one base pair, with initial loop sizes ( $l$ ) ranging from 2 to 14 nt ( $AG_2L_l$  series), and (b) substrates that can potentially form two base pairs, with initial loop sizes ranging from 4 to 13 nt ( $AG_3L_l$  series). As the initial loop size increases, the extension rate slows notably (see also Figure 4) but appears independent of initial loop size. Substrates were incubated for 45 min under standard conditions.

similar to the  $K_m$  values observed for other oligonucleotides (Table 1).

**Stable Primer–Template DNA Complexes Inhibit Extension.** The 17 nt long sequence  $T_{10}G_2T_2C_2C$  efficiently incorporated CMP and AMP.  $T_{10}G_2T_2C_2C$  was also found to poorly incorporate UMP, at a rate at least 50 times slower than the observed rate of CMP incorporation in our single NTP extension assay. This unanticipated incorporation, possibly resulting from coding with a G·U wobble pair (see also Figure 1), was not observed in  $A_{10}G_2A_2C_2C$ , which, however, was ~30-fold slower than  $T_{10}G_2T_2C_2C$ .  $T_{10}G_2T_2C_2C$  was incubated with either a short (SC) or a long (LC) complementary oligonucleotide, each having the potential to code for up to three UMP incorporations. The shorter sequence can form a 7 bp interaction with  $T_{10}G_2T_2C_2C$  with an expected  $T_m$  of 24 °C, while the longer sequence forms 17 bp of interaction and has a theoretical  $T_m$  of 43 °C (Figure 7). In the presence of the shorter oligonucleotide, intramolecular CMP incorporation was not noticeably slowed even at high concentrations of SC, while UMP incorporation was observed to increase only marginally from intrinsic background levels (Figure 7). In contrast, the presence of stoichiometric or higher amounts of LC completely inhibited both CMP and UMP incorporation.

## DISCUSSION

T7 RNAP initiates transcription of RNA from a single nucleotide triphosphate after first tightly binding its DNA promoter sequence. Prior to the formation of an elongation

complex, the enzyme forms a small (~4 bp) bubble of DNA between the bound T7 RNAP promoter and the active site of the enzyme (6). The polymerase synthesizes a short 10–12 nt long RNA fragment in a template-dependent fashion that induces a conformational rearrangement of the protein (8). This rearrangement liberates the T7 RNAP promoter sequence and creates an RNA exit tunnel that stabilizes the larger (~7 bp) replication bubble required for efficient elongation (6). In the absence of a promoter, we have found that short oligonucleotides are surprisingly good substrates for T7 RNAP. The observed extension rates compare well to the rate of initiation in the presence of T7 promoter sequence (Table 1). This, combined with the substrates' respectable binding affinities (2–43  $\mu$ M), indicates that the active site of the enzyme retains considerable catalytic ability for substrates completely lacking a T7 promoter sequence.

A notable feature of promoter-independent extension is that extension products are predictable when suitable incubation conditions are used. The extension of 7 nt long DNA substrates was coded by the most 5' residue in the sequence. Systematically varying this nucleotide resulted in extension by the complementary nucleotide in every case tested (Figure 2b). This extension appeared to rely only on the formation of a transient primer–template complex consisting of a single Watson–Crick base pair (or less efficiently with a wobble pair; see Figures 1 and 7) that can form between the 3' terminus of the oligonucleotide and the nucleotide immediately adjacent and 3' to the coding nucleotide. Consistent with this hypothesis, mutants that disrupted pairing were inactive, while a double mutant that preserved pairing was found to be active (Figure 2a).

This unimolecular model of extension, where active constructs must form unstable loops by folding back their 3' ends so as to make an interaction within their own sequence, is supported by other lines of evidence. The Hill coefficients for constructs predicted to be extended in a unimolecular fashion were always observed to be close to 1 (Table 1, Figure 3a). This effect was observed for a range of substrate lengths and sequence compositions. Further, constructs having Hill coefficients of 1 tolerated the insertion of sequence into their loop region in all cases tested but became inactive whenever the loop region became smaller than 4 nt (Figures 2c and 5a). This loss of activity would be expected, owing to the relatively high thermodynamic instability of such short loop structures (22). In contrast, a self-complementary oligonucleotide (5'-GGGCC), which could plausibly be extended in a bimolecular and not a unimolecular reaction, was observed to have a Hill coefficient of 2, consistent with a bimolecular extension process (Table 1, Figure 3b).

We have isolated the transient extension abilities of T7 RNAP away from its processive polymerization properties. Extension was controlled by the choice of both the construct sequence and the nucleotide triphosphate in the extension reaction. For example, construct  $A_{10}G_2A_2C_2C$ , which in the presence of CTP can use G(11) as a template and G(12)·C(17) to prime a single extension, must be reorganized in order to incorporate a second CMP [after translocation A(10) cannot serve as a template for CMP incorporation] and was therefore used to study the ability of this complex to dissociate from or slip within the enzyme active site. In contrast, substituting UTP for CTP allows processive exten-

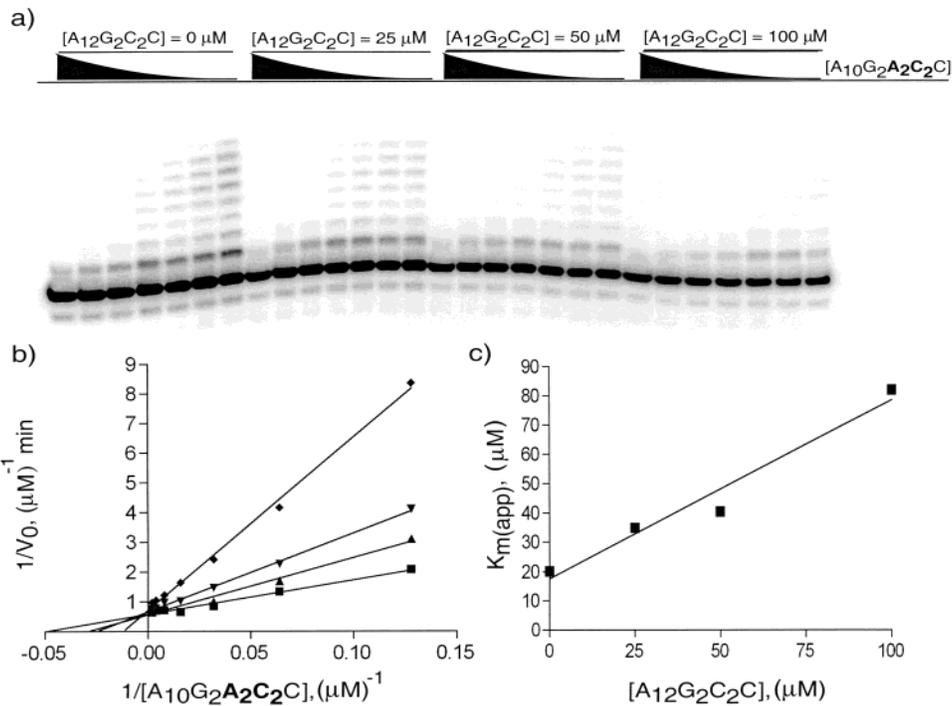


FIGURE 6: Ability of inactive  $A_{12}G_2C_2C$  to inhibit  $A_{10}G_2A_2C_2C$  extension. (a) CTP extension reactions varying  $A_{10}G_2A_2C_2C$  concentrations from 8 to 500  $\mu M$ . The inactive sequence  $A_{12}G_2C_2C$  was titrated from 0 to 100  $\mu M$ . (b) Lineweaver–Burk plot of the inverse initial velocities against inverse  $A_{10}G_2A_2C_2C$  concentrations. (c) Plot of  $K_m(\text{app})$  against  $A_{12}G_2C_2C$  concentration, resulting in a  $K_i$  of 45  $\mu M$ .

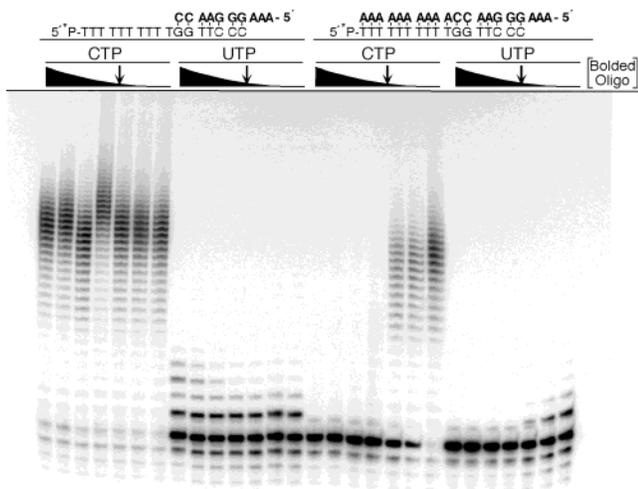


FIGURE 7: dsDNA inhibition of unimolecular extension. 5'-Labeled  $T_{10}G_2T_2C_2C$  (bottom oligonucleotide) was held at 10 nM, while SC or LC (top oligonucleotides in boldface type; left and right panel, respectively) were titrated from 1 to 0  $\mu M$  (1, 1/4, 1/16, 1/64, 1/256, 1/1024, and 0  $\mu M$  respectively) in standard CTP or UTP extension reactions. Arrows indicate the point at which the hybridizing oligonucleotides and  $T_{10}G_2T_2C_2C$  are at approximately equimolar concentrations.

sion as well as transient extension. Indeed, given that over 30 rapid UMP incorporations were observed when  $A_{10}G_2A_2C_2C$  was used as a substrate, a complicated and experimentally unresolvable combination of processive and transient extension steps must have occurred, given that the template sequence contains only 10 sequential dA residues.

Another important feature of promoter-independent extension is that stable double-stranded constructs do not appear to be efficient substrates. The only characterized bimolecular substrate, 5'-GGGCC, had 4 bp of stabilization and was extended efficiently albeit only at high oligonucleotide

concentrations. The construct  $T_{10}G_2T_2C_2C$ , when deliberately complexed with a short template sequence to form a relatively unstable bimolecular complex coding for up to three UMP additions (SC), was a poor substrate for UMP addition (Figure 7). At the same time, the fast unimolecular incorporation of CMP intrinsic to  $T_{10}G_2T_2C_2C$  was not significantly inhibited. The same construct hybridized to a longer, more stable template (LC, 17 bp of hybridization) failed to incorporate either UMP or CMP once stoichiometric or higher amounts of template were added to the reaction. Taken together, these data strongly suggests that the enzyme cannot deal efficiently with long stable dsDNA in the absence of a promoter sequence but must initially recognize and extend substrates that spend a substantial fraction of their time in an unstructured, single-stranded form.

A model consistent with our data is that unimolecular substrates are first bound to the enzyme and then fold transiently into an extendable complex that allows the enzyme to incorporate a nucleotide. If after incorporation the construct is unable to support further addition, either because the end of the template has been reached ( $G_2CAC_2C$ -type constructs) or because the correct cognate NTP was not supplied ( $A_{10}G_2A_2C_2C$ -type constructs), the enzyme–substrate complex can explore two options: either the construct disassociates from the enzyme and rebinds in a later step or it slips within the bound enzyme–substrate complex, allowing the incorporation of a second nucleotide. The ability of the inactive substrate  $A_{12}G_2C_2C$  to act as a competitive inhibitor for  $A_{10}G_2A_2C_2C$ , with a  $K_i$  similar to the binding affinity of  $A_{10}G_2A_2C_2C$ , indicates that substrate release does occur during extension and favors a transient extension mechanism that includes both a dissociative and a slippage component. The interplay between these two effects appears likely to be complicated (observe the huge difference in initial rate between  $A_{10}G_2A_2C_2C$  and  $A_8G_2A_4C_2C$

as implied by Figure 5a) and was not dissected further.

This model of promoter-independent extension is consistent with X-ray structures that capture T7 RNAP complexes during the initiation of polymerization. Cheetham and Steitz have shown that prior to the formation of an elongation complex, the enzyme forms a small four nucleotide bubble of single-stranded DNA between the bound T7 RNAP promoter and the template found in the active site of the enzyme (6). The coding strand of this bubble is bound to an active-site pocket during initiation and appears to accumulate or be "scrunched" during early RNA synthesis (6–8, 23). Constructs lacking the T7 promoter sequence might plausibly bind into this catalytic pocket and be able to loop back their 3' ends and in this way self-prime extension. The finite size of this pocket (6–9 nt), which has been estimated by crystallography (6), may also explain the marked decrease in rate observed for a variety of unimolecular constructs after loop sizes of ~11–14 nt are reached and is consistent with the high initial velocity observed for constructs with small initial loop sizes of 6–8 nt (Figures 4 and 5). Entropically, larger loops are likely to be disfavored within the binding pocket and consequentially slow the rate of extension.

The ability of T7 RNAP to extend short unimolecular substrates by use of only a single base pair of priming may help to explain the evolution of viroid-like nucleic acid sequences capable of self-replication in the presence of the enzyme (12, 13, 24). These short RNA sequences are substantially self-complementary and contain numerous repeats that can form a variety of highly symmetrical secondary structures. Our work, although mainly performed with DNA oligonucleotides, suggests that efficient self-replicating sequences should form neither loop regions nor double-stranded elements near the site of nucleotide synthesis much larger than 14 nt and 10 bp, respectively (see Figures 5 and 7). It is interesting to note that both RNA X and RNA Y appear to respect these constraints (13), as do sequences generated by a completely template-free replicative system (24). The transient nature of priming and the fact that it requires as little as 1 bp to initiate would appear to explain the significant sequence diversity observed among known self-replicating RNA sequences and the sensitivity of their emergence on incubation conditions. The implication of DNA-dependent RNA polymerases in the replication of PSTV viroids and the hepatitis  $\delta$  virus, together with the considerable symmetry evidence in PSTV viroid sequences (25), suggests that the promoter independent properties of T7 RNAP may also prove useful in understanding the emergence of viral RNA sequences.

## ACKNOWLEDGMENT

We are grateful for T7 RNAP preparation by A. Ebhardt and for the careful reading of the manuscript by Drs. M. Leroux, D. Sen, and K. Williams along with members of the Unrau laboratory.

## REFERENCES

- Dunn, J. J., and Studier, F. W. (1983) Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements, *J. Mol. Biol.* 166, 477–535.
- Krupp, G. (1988) RNA synthesis: strategies for the use of bacteriophage RNA polymerases, *Gene* 72, 75–89.
- Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates, *Nucleic Acids Res.* 15, 8783–98.
- Oakley, J. L., Strothkamp, R. E., Sarris, A. H., and Coleman, J. E. (1979) T7 RNA polymerase: promoter structure and polymerase binding, *Biochemistry* 18, 528–37.
- Martin, C. T., and Coleman, J. E. (1987) Kinetic analysis of T7 RNA polymerase–promoter interactions with small synthetic promoters, *Biochemistry* 26, 2690–6.
- Cheetham, G. M., and Steitz, T. A. (1999) Structure of a transcribing T7 RNA polymerase initiation complex, *Science* 286, 2305–9.
- Cheetham, G. M., Jeruzalmi, D., and Steitz, T. A. (1999) Structural basis for initiation of transcription from an RNA polymerase–promoter complex, *Nature* 399, 80–3.
- Yin, Y. W., and Steitz, T. A. (2002) Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase, *Science* 298, 1387–95.
- Cech, T. R., and Golden, B. L. (1999) in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) p 708, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Pleiss, J. A., Derrick, M. L., and Uhlenbeck, O. C. (1998) T7 RNA polymerase produces 5' end heterogeneity during in vitro transcription from certain templates, *RNA* 4, 1313–7.
- Tuschl, T., Sharp, P. A., and Bartel, D. P. (1998) Selection in vitro of novel ribozymes from a partially randomized U2 and U6 snRNA library, *EMBO J.* 17, 2637–2650.
- Konarska, M. M., and Sharp, P. A. (1989) Replication of RNA by the DNA-dependent RNA polymerase of phage T7, *Cell* 57, 423–31.
- Konarska, M. M., and Sharp, P. A. (1990) Structure of RNAs replicated by the DNA-dependent T7 RNA polymerase, *Cell* 63, 609–18.
- Cazenave, C., and Uhlenbeck, O. C. (1994) RNA template-directed RNA synthesis by T7 RNA polymerase, *Proc. Natl. Acad. Sci. U.S.A.* 91, 6972–6.
- Ichetovkin, I. E., Abramochkin, G., and Shrader, T. E. (1997) Substrate recognition by the leucyl/phenylalanyl-tRNA-protein transferase. Conservation within the enzyme family and localization to the trypsin-resistant domain, *J. Biol. Chem.* 272, 33009–14.
- Smith, M. C., Furman, T. C., Ingolia, T. D., and Pidgeon, C. (1988) Chelating peptide-immobilized metal ion affinity chromatography. A new concept in affinity chromatography for recombinant proteins, *J. Biol. Chem.* 263, 7211–5.
- King, G. C., Martin, C. T., Pham, T. T., and Coleman, J. E. (1986) Transcription by T7 RNA polymerase is not zinc-dependent and is abolished on amidomethylation of cysteine-347, *Biochemistry* 25, 36–40.
- Davanloo, P., Rosenberg, A. H., Dunn, J. J., and Studier, F. W. (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase, *Proc. Natl. Acad. Sci. U.S.A.* 81, 2035–9.
- Sastry, S. S., and Ross, B. M. (1997) Nuclease activity of T7 RNA polymerase and the heterogeneity of transcription elongation complexes, *J. Biol. Chem.* 272, 8644–52.
- Gardner, L. P., Mookhtiar, K. A., and Coleman, J. E. (1997) Initiation, elongation, and processivity of carboxyl-terminal mutants of T7 RNA polymerase, *Biochemistry* 36, 2908–18.
- Stano, N. M., Levin, M. K., and Patel, S. S. (2002) The +2 NTP Binding Drives Open Complex Formation in T7 RNA Polymerase, *J. Biol. Chem.* 277, 37292–37300.
- Woese, C. R., Winker, S., and Gutell, R. R. (1990) Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops", *Proc. Natl. Acad. Sci. U.S.A.* 87, 8467–71.
- Kukarin, A., Rong, M. Q., and McAllister, W. T. (2003) Exposure of T7 RNA polymerase to the isolated binding region of the promoter allows transcription from a single-stranded template, *J. Biol. Chem.* 278, 2419–2424.
- Biebricher, C. K., and Luce, R. (1996) Template-free generation of RNA species that replicate with bacteriophage T7 RNA polymerase, *EMBO J.* 15, 3458–65.
- Symons, R. H. (2001) in *RNA* (Soll, D., Nishimura, S., and Moore, P. B., Eds.) Elsevier Science, Oxford, U.K.