Kinetic Analysis of T7 RNA Polymerase–Promoter Interactions with Small Synthetic Promoters†

Craig T. Martin‡ and Joseph E. Coleman*
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06510
Received July 16, 1986; Revised Manuscript Received November 20, 1986

ABSTRACT: Specific interactions between T7 RNA polymerase and its promoter have been studied by a simple steady-state kinetic assay using synthetic oligonucleotide promoters that produce a short five-base message. A series of promoters with upstream lengths extending to promoter positions −19, −17, −14, and −12 show that promoters extending to −19 and −17 produce very specific transcripts with initiation rate constant k_{cat} = 50 min^{-1} and a Michaelis constant K_m = 0.02 μM, indicating that the consensus sequence to position −17 is sufficient for maximum promoter usage. Shortening the upstream region of the promoter to −14 substantially increases K_m (0.3 μM) but does not significantly reduce the maximum velocity (k_{cat} = 30 min^{-1}). Finally, truncation of the promoter at position −12 results in extremely low levels of specific transcription. The coding and noncoding strands appear to make different contributions to promoter recognition. Although the double-stranded promoter of upstream length −12 is very poor as a transcription template, extension of only the noncoding strand to −17 very significantly improves both k_{cat} and K_m. In contrast, extension of only the coding strand results in no significant improvement. Substitution of an AT base pair at position −10 by CG (as found in T3 RNA polymerase promoters) produces a 10-fold increase in K_m, with little effect on k_{cat}. Comparison of two promoters containing a base pair mismatch at this site (AG or CT) demonstrates that promoter recognition is very sensitive to the nature of the base on the noncoding strand and is only slightly affected by the presence of a mismatch created by a wrong base in the coding strand. These results lead to a model for transcription initiation by T7 RNA polymerase in which promoter recognition occurs in the melted form of the DNA and in which upstream determinants on the noncoding strand provide significant specific interactions with the protein.

T7 RNA polymerase is one of several apparently related bacteriophage RNA polymerases (Chamberlin & Ryan, 1982; McAllister et al., 1983). These DNA-dependent RNA polymerases are highly promoter specific, single-subunit enzymes with M_r 100 000. They are much less complex, structurally and functionally, than bacterial and eukaryotic RNA polymerases and so present ideal systems in which to study the molecular details of transcription. The complete primary sequences of the enzymes from T7 and T3 indicate that they share substantial structural homology (Stahl & Zinn, 1981; Moffatt et al., 1984; McGraw et al., 1985). The DNA sequences for all 17 T7 promoters are known, and from these data a consensus sequence for the promoter from positions −17 to +5 has been established (Oakley & Coleman, 1977; Dunn & Studier, 1983). Eleven promoter sequences from the T3 enzyme (Bailey et al., 1983; Basu et al., 1984), and four from the SP6 enzyme (Brown et al., 1986), have been determined. Their consensus sequences share regions of homology with that of T7 but also contain significant differences, consistent with the observation of low cross-recognition of their promoters (Bailey et al., 1983).

Of the three enzymes, T7 RNA polymerase is the best understood. The consensus promoter sequence from positions −17 to +5

```
-15          -1
 |            |
TAATACGACTCACTATA GGGAGA
```

is strictly conserved in all class III promoters. Early studies of promoter recognition predicted a minimum upstream promoter length of between 14 and 22 base pairs (Oakley et al., 1979; Osterman & Coleman, 1981). Although truncation of the promoter at position −14 destroys transcription, religation of random sequences upstream of this position restores activity. Digestion of the enzyme–promoter complex by a purified single-strand endonuclease has revealed that the promoter region from positions +2 to −6 is open and that the noncoding strand (only) in this region is susceptible to digestion (Osterman & Coleman, 1981). Recent DNase I footprinting studies have confirmed this enhanced susceptibility of the noncoding strand to cleavage (Basu & Maitra, 1986). Complementary footprinting experiments with the double-strand intercalator methidiumpropyl-EDTA-Fe(II)[1] have shown that the promoter bases of both strands from positions −2 to −21 are protected by the enzyme (Ikeda & Richardson, 1986). Taken together, the results available to date give a rough picture of the positioning of the polymerase bound to promoter, but the specific interactions required for promoter recognition remain unclear.

In the hope of better understanding the requirements for promoter recognition and initiation of transcription in T7 RNA polymerase, we have developed a transcription system consisting of synthetic oligonucleotides of defined length and sequence as templates, all of which produce a simple five-base mRNA. This system allows the direct measurement of transcription rates under steady-state conditions and allows for controlled manipulation of promoter structure. We have examined promoters with upstream lengths of 19, 17, 14, and 12 bases, as well as hybrids constructed with unequal lengths of the coding and noncoding strands. We have also initiated

† This work was supported by NIH Grant GM21919-11.
‡ C.T.M. is the recipient of National Institutes of Health Postdoctoral Fellowship GM10902.

[1] Abbreviations: BSA, bovine serum albumin; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; HPLC, high-performance liquid chromatography; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/87/0426-2690$01.50/0

a study of the effects of base substitution at specific sites in the promoter. These data suggest a model for promoter recognition by the T7-related RNA polymerases in which different regions of both strands of the promoter DNA are involved in specific DNA–protein interactions.

Materials and Methods

T7 RNA polymerase was prepared from *Escherichia coli* strain BL 21 containing plasmid pAR1219 (kindly supplied by William Studier and John Dunn), with T7 gene 1 (RNA polymerase) cloned under inducible control of the *lac*UV5 promoter (Davanloo et al., 1984). Enzyme was purified as previously described (King et al., 1986) by fractionation with Polymin P (less than 1.25%) and ammonium sulfate, followed by chromatography on SP-Trisacryl (LKB), TSK CM-Fractogel (EM Science), and TSK DEAE-Fractogel (EM Science). A molar extinction coefficient of $\epsilon_{280} = 1.4 \times 10^5\ M^{-1}$ was used to determine enzyme concentrations (King et al., 1986). Enzyme activity was assayed under standard reaction conditions as previously described (Oakley et al., 1975; King et al., 1986) and routinely showed a specific activity on T7 DNA of 300 000–400 000 units/mg. One unit of specific activity is defined as nanomoles of AMP incorporated per hour at 37 °C in the standard assay mix (Chamberlin et al., 1970; Chamberlin & Ring, 1973).

Oligonucleotides were synthesized by the phosphoramidate method on an Applied Biosystems Model 380B synthesizer. Single strands were purified by HPLC[1] on a Nucleogen DEAE 60-7 column in 20 mM sodium acetate and 20% acetonitrile, eluting with a gradient from 0.0 to 1.0 M KCl. Purity of single strands end labeled with ^{32}P was determined by electrophoresis on 20% acrylamide and 7 M urea gels. Concentrations of single-strand oligonucleotides were estimated by assuming an average molar extinction coefficient per base of $\epsilon_{260} = 8.4 \times 10^3\ M^{-1}$. Double-strand templates were prepared by heating a 1:1 mixture of complementary single strands in TE buffer (10 mM Tris, pH 7.8, 1 mM EDTA) to 70 °C for 5 min. The solutions were then allowed to cool slowly to room temperature and stored at 4 °C until use. All templates used in the kinetic assays possessed 3′- and 5′-hydroxyl groups.

Kinetic assays of transcription on oligonucleotide templates were carried out in 20 μL of 40 mM Tris-HCl, pH 7.8, 20 mM $MgCl_2$, 10 mM NaCl, 0.25 mM EDTA, 1 mM DTT, 0.05 mg/mL BSA (Boehringer Mannheim), 0.8 mM GTP, and 0.4 mM each UTP, CTP, and ATP. Unless otherwise indicated, [α-^{32}P]UTP was added to approximately 100 μCi/mL. A 10.0-μL aliquot of a 2X mix (55 mM Tris, 40 mM $MgCl_2$, 2 mM DTT, 0.1 mg/mL BSA, 1.6 mM GTP, [α-^{32}P]UTP, and 0.8 mM each UTP, CTP, ATP) on ice was premixed with 5.0 μL of 4X oligonucleotide in TE buffer and incubated in 0.5-mL plastic tubes for 3 min at 37 °C. To start the reaction, 5.0 μL of 4X enzyme (freshly diluted into 40 mM Tris-HCl, pH 7.8, and 40 mM NaCl) on ice was added and the sample returned to 37 °C. Aliquots of 5.0 μL were withdrawn at 3, 6, and 9 min and spotted onto individual lanes of Whatman 3MM filter paper, prespotted with 10 μL of 200 mM EDTA. The samples were then chromatographed in 60% saturated ammonium sulfate, pH 8, to separate products from unincorporated nucleotide (Mulligan et al., 1985). For each sample, a distinct band of radioactivity migrating away from the origin, but well separated from free nucleotide, was visualized by exposure to X-ray film, then cut out, and quantitated with OptiFluor (United Technologies Packard) liquid scintillant.

Analysis of transcription products was performed with typical kinetic assay samples, but the reactions were quenched by addition into twice the volume of 8 M urea, 20 mM EDTA, and 20 mM Tris, pH 7.8. These samples were heated to 90 °C for 5 min, then loaded directly to 20% acrylamide and 7 M urea transcript gels, and electrophoresed under standard conditions (Maniatis et al., 1982).

Results

Characterization of Transcription Products. The synthetic promoters used in this study are derived from the sequence

```
          -10          +1
           |           |
5′-TAATACGACTCACTATA GGACT-3′ (noncoding)
3′-ATTATGCTGAGTGATAT CCTGA-5′ (coding)
```

This oligonucleotide consists of the known promoter consensus sequence from positions −17 to −1, plus the code for a five-base message utilizing all four nucleotides. All templates should produce the same five-base runoff message GGACU. If the proper five-base message is the only product synthesized, direct quantitation of completed initiation can be determined as the amount of [α-^{32}P]UMP incorporated into product.

To characterize the messages transcribed from the various oligonucleotide templates used in this study, transcript gels were prepared in which [α-^{32}P]UMP and [γ-^{32}P]GTP were incorporated into mRNA in parallel reactions. Enzyme and template concentrations were chosen for each reaction so as to produce approximately 5% incorporation of [α-^{32}P]UMP in a 9-min assay (determined from kinetic data below). The specific activities of [α-^{32}P]UTP and [γ-^{32}P]GTP were approximately equal. The transcript gel in Figure 1 shows the RNA products from some of the templates used in this study. The templates showing the highest activity (such as −17ds, see below) produce one major product. The reactions incorporating [γ-^{32}P]GTP show clearly that initiation is occurring at a G start site, as expected. Similarly, incorporation of [α-^{32}P]UMP confirms that the major product results from transcription to the end of the template. Although there is minor formation of a smaller transcript in the [γ-^{32}P]GTP experiments, the kinetic assays below incorporate [α-^{32}P]UMP and so measure only completed transcripts. Gels of reaction mixtures containing [γ-^{32}P]GTP and incomplete subsets of the remaining three ribonucleotides (not shown) yield a ladder pattern of shortened transcripts as expected from initiation at the unique +1 G promoter position.

In order to achieve detectable ribonucleotide incorporation on poor templates such as −12ds and $-12_N/-17_C$, enzyme and template concentrations were sufficiently high so that nonspecific transcription produced multiple products. The lack of incorporation of [γ-^{32}P]GTP shows that the predominant products in these cases arise from initiation at other than the correct (+1 G) start site. The large products are most likely runoff transcripts due to initiation at the ends of the DNA template.

In the kinetic analyses that follow, transcription products were separated by ascending paper chromatography as described under Materials and Methods. It is generally assumed for each template that only the correct five-base transcript is synthesized; however, in the case of some of the poor templates, this assumption is not entirely correct, as can be seen in the transcript gel shown in Figure 1. In order to avoid artifacts in the paper chromatograms, the chromatogram origins (which may contain transcripts larger than five bases) were cut out, quantitated separately, and not used in calculating reaction velocities. Since U occurs only in the last base of the correct five-base transcript, prematurely terminated transcripts will not be counted; however, random-start short transcripts may

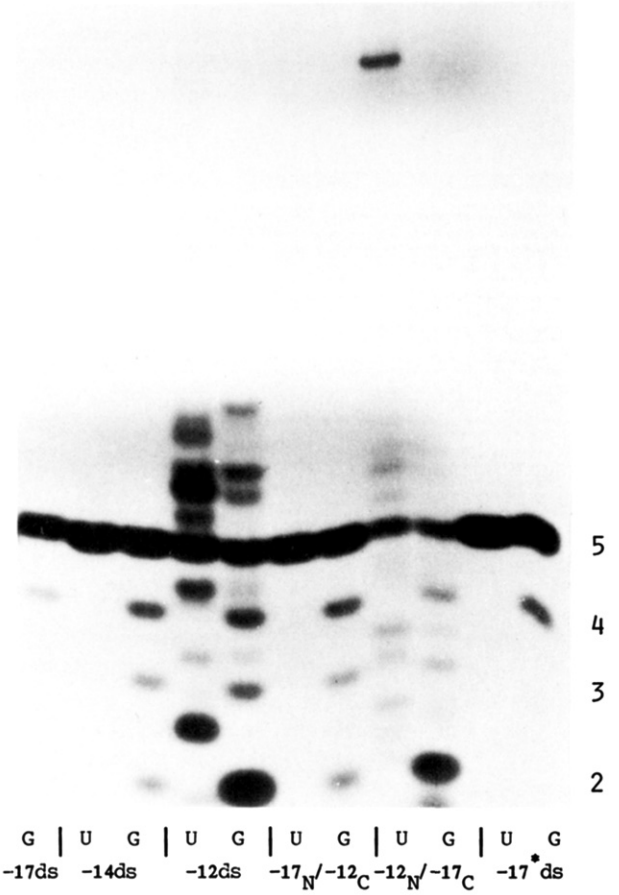

FIGURE 1: Transcript gel comparing RNA messages produced from some of the promoters used in this study. Parallel reactions incorporating either [α-^{32}P]UMP or [γ-^{32}P]GTP were incubated for 9 min at 37 °C as described in the text. For each template, enzyme and promoter were present in a 1:1 ratio at the following concentrations: −17ds, 0.1 μM; −14ds, 0.3 μM; −12ds, 4.0 μM; $-17_N/-12_C$, 0.6 μM; $-12_N/-17_C$, 4.0 μM; −17*ds, 0.2 μM. Bands corresponding to correctly initiated transcripts (i.e., those incorporating [γ-^{32}P]GTP) of lengths 2, 3, 4, and 5 are indicated on the gel. Displaced bands incorporating [α-^{32}P]UMP are dimers, trimers, and tetramers of other than the correct message sequence.

contain varying amounts of UMP. Consequently, the rate constants (see below) determined for those templates that produce incorrect transcripts will be upper limits, with similar effects on the predicted K_m.

Assays for Determination of Kinetic Constants. Traditional RNA polymerase assays quantitate the incorporation of radiolabeled nucleotide into acid-insoluble product within a given period of time, generally 10 min (Chamberlin et al., 1970; Chamberlin & Ring, 1973). The activities obtained are not readily interpretable in terms of specific steps in transcription. The simple kinetic assay outlined here should allow accurate measurement of initiation kinetics (see Discussion). A steady-state kinetic mechanism has been assumed for each template:

$$\text{enzyme} + \text{DNA} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{enzyme–DNA} \xrightarrow{k_{cat}} \text{enzyme} + \text{DNA} + \text{message} \quad (1)$$

The concentrations of nucleotides remained well above their K_m's (0.1 mM; Oakley et al., 1979) throughout the reactions, so k_{cat} can be considered a simple first-order rate constant. Since in most cases neither enzyme nor DNA was in excess, the data were fit to the exact steady-state solution of eq 1

$$V = \tfrac{1}{2}k_{cat}\{D_T + E_T + K_m - [(D_T + E_T + K_m)^2 - 4E_TD_T]^{1/2}\} \quad (2)$$

Table I: Comparison of Kinetic Data for Promoters of Varying Upstream Lengths[a]

template	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m
−19ds	54 (44–67)	0.020 (0.015–0.037)	2700
−17ds	52 (46–58)	0.015 (0.008–0.029)	3600
−14ds	19 (17–20)	0.27 (0.22–0.33)	70
$-17_N/-14_C$	60 (36–89)	0.32 (0.16–0.75)	190
$-14_N/-17_C$	18 (15–21)	0.32 (0.22–0.48)	60
−12ds	1.4 (0.9–2.0)	2.3 (1.0–5.8)	0.6
$-17_N/-12_C$	8.8 (6.4–11)	0.4 (0.1–1.1)	24
$-12_N/-17_C$	0.7 (0.5–1.1)	1.6 (0.6–5.0)	0.5

[a] Kinetic parameters k_{cat} and K_m were obtained from fits of the data in Figures 2–4 to eq 1, as described in the text. Ranges of values reflect the 65% joint confidence interval of the fitted parameters.

Table II: Comparison of Kinetic Data for Consensus Sequence Promoters Containing a Base Substitution at Position −10[a]

template	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m
−17ds	52 (46–58)	0.015 (0.008–0.029)	3600
−17*ds	30 (27–34)	0.14 (0.10–0.20)	210
$-17_N/-17_C$*	41 (30–52)	0.08 (0.04–0.17)	490
-17_N*$/-17_C$	24 (19–28)	0.35 (0.22–0.57)	80

[a] Strands denoted 17_N* and 17_C* contain C and G, respectively, at position −10 of their sequence. Parameters were fit to the kinetic data presented in Figure 4 and reported as in Table I.

by using a nonlinear least-squares minimization algorithm based on the Gauss–Newton method (Johnson et al., 1981). In this analysis, the velocity of UMP incorporation (V, the dependent variable) is fit to the independent variables total enzyme and DNA concentration, E_T and D_T, respectively, and the fit parameters K_m and k_{cat}.

Rather than estimating velocities of nucleotide incorporation from a single time point, we have determined them from a linear least-squares fit to plots of UMP incorporation vs. time (3, 6, and 9 min). The error in each velocity was approximated as the t distribution 80% confidence interval of the fitted slope. The linear least-squares fits were generally as expected for steady-state kinetics. At high concentrations of enzyme and template, some drop in the velocity was observed with time, possibly due to the presence of secondary low-affinity binding sites.

For each template, reaction velocities were measured for several DNA concentrations and at least two enzyme concentrations. Each velocity was weighted according to its calculated error. A single best fit to eq 2 was then obtained for all data from that template. In the figures that follow, predicted velocity curves for each enzyme concentration are presented with the data to illustrate the fit. Note, however, that all curves for a given template reflect the same best fit values of k_{cat} and K_m. The fit values k_{cat} and K_m are presented in the tables, along with their 65% joint confidence intervals (Turner et al., 1981).

As noted above, transcription initiation at other than the promoter G at position +1 will cause both k_{cat} and K_m to be overestimated. Similarly, as K_m becomes increasingly large, competition by nonspecific binding to DNA will complicate the assumptions of eq 1. Therefore, discretion in the interpretation of data from the very poor promoter templates (−12ds and $-12_N/-17_C$) is advised.

Dependence of Transcription on Upstream Length of the Template. The velocities of incorporation observed for transcription from the various oligonucleotide promoters are plotted vs. DNA concentration in Figures 2–4. The kinetic data derived from these plots, summarized in Tables I and II, show quite clear trends in the catalytic properties of the various promoter templates.

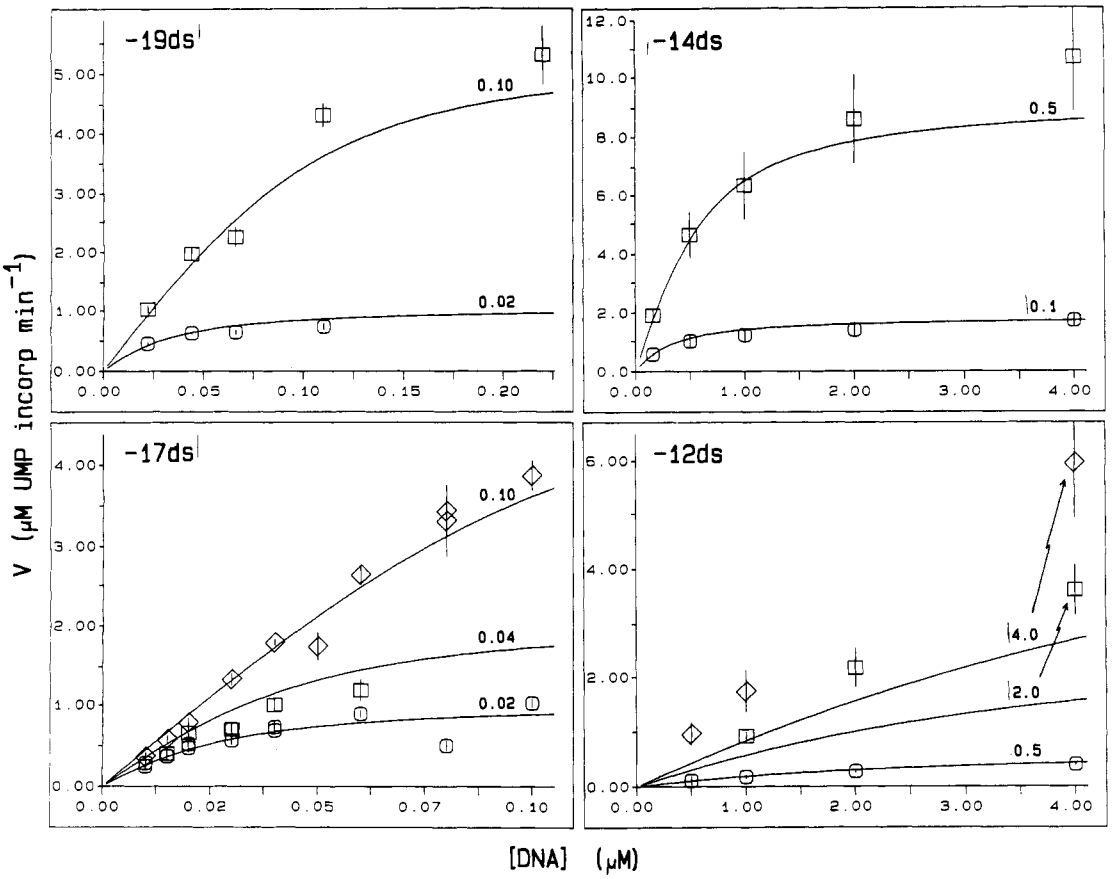

FIGURE 2: Comparison of transcription kinetics for double-stranded promoters with varying upstream lengths of −19, −17, −14, and −12. Plots for each template show velocities of UMP incorporation vs. oligonucleotide concentration for at least two enzyme concentrations (0.02–4.0 μM, as by the numbers on the plots). The predicted set of curves in each panel reflects the single best fit values of k_{cat} and K_m for each template, as presented in Table I.

The full-length consensus template (−17ds) incorporates UMP at a maximum velocity of 50 min^{-1} with K_m = 0.015 μM. To determine whether the consensus sequence is sufficient to allow maximal transcription, we prepared a synthetic promoter (−19ds) containing an additional two bases (AT) upstream. The (message-strand) sequence of this double-strand template is

AT TAATACGACTCACTATA GGACT

The kinetic data for this template (k_{cat} = 50 min^{-1}; K_m = 0.020 μM) demonstrate that the presence of the two additional bases upstream of position −17 provides no significant benefit to recognition or the rate of transcription initiation.

To better determine the effects of template length on transcription initiation, we prepared oligonucleotide templates with shortened upstream promoter lengths extending only to positions −14 and −12. The −14ds promoter incorporates UMP maximally at about half the rate of the full-length template (k_{cat} = 20 min^{-1}). However, significantly higher concentrations of enzyme and DNA are required to achieve this rate, since the estimated K_m of 0.3 μM for the −14ds template is at least an order of magnitude greater than that for the −17ds template. Removing two more bases from the upstream end of the promoter (−12ds) almost completely destroys both binding and catalytic activity (K_m > 1 μM; k_{cat} < 2 min^{-1}). The transcript gel in Figure 1 shows that the majority of the observed nucleotide incorporation from −12ds at high enzyme and DNA concentrations is not promoter specific.

Effects on Transcription of Single-Stranded Regions in the Promoter. The results presented above suggest that the presence of both strands beyond position −14 is required for optimal promoter recognition. To study the relative contributions of the coding and noncoding strands to binding and initiation, we prepared hybrid templates, each containing complementary DNA strands of unequal upstream lengths. We will hereafter refer to a hybrid containing noncoding and coding strands extending to (upstream) positions $-i$ and $-j$, respectively, as the $-i_N/-j_C$ template. The data summarized in Table I for the −14ds template show that extension of either strand to full length (−17) without its complement does not significantly improve K_m. The catalytic rate constant, however, does show some dependence on the nature of the upstream region. Although a single-stranded extension of only the coding strand to −17 does not enhance k_{cat}, extension of only the noncoding strand to full length restores k_{cat} to its maximal value.

Further evidence for preferential recognition of the noncoding strand is found with the hybrid templates formed from strands with upstream lengths from −12 to −17. As noted previously, the short promoter, −12ds, is almost completely inactive as a template for T7 RNA polymerase. Extension of only the coding strand to its full length of −17 ($-12_N/-17_C$) does not significantly improve promoter utilization (k_{cat} = 0.7 min^{-1}; K_m = 2 μM). Extension of only the *noncoding* strand to position −17, however, does significantly improve promoter recognition (k_{cat} = 9 min^{-1}; K_m = 0.4 μM). This effect is seen more dramatically in Figure 1, where the correct product is formed with much greater fidelity by the $-17_N/-12_C$ template than by either the $-12_N/-17_C$ or the −12ds templates.

Effect of a Single Base Pair Substitution and of Base Mismatch on Promoter Recognition. The promoter sequence features that determine differential recognition between the T7, T3, and SP6 transcription systems cannot be stated categorically from simple inspection of the consensus sequences

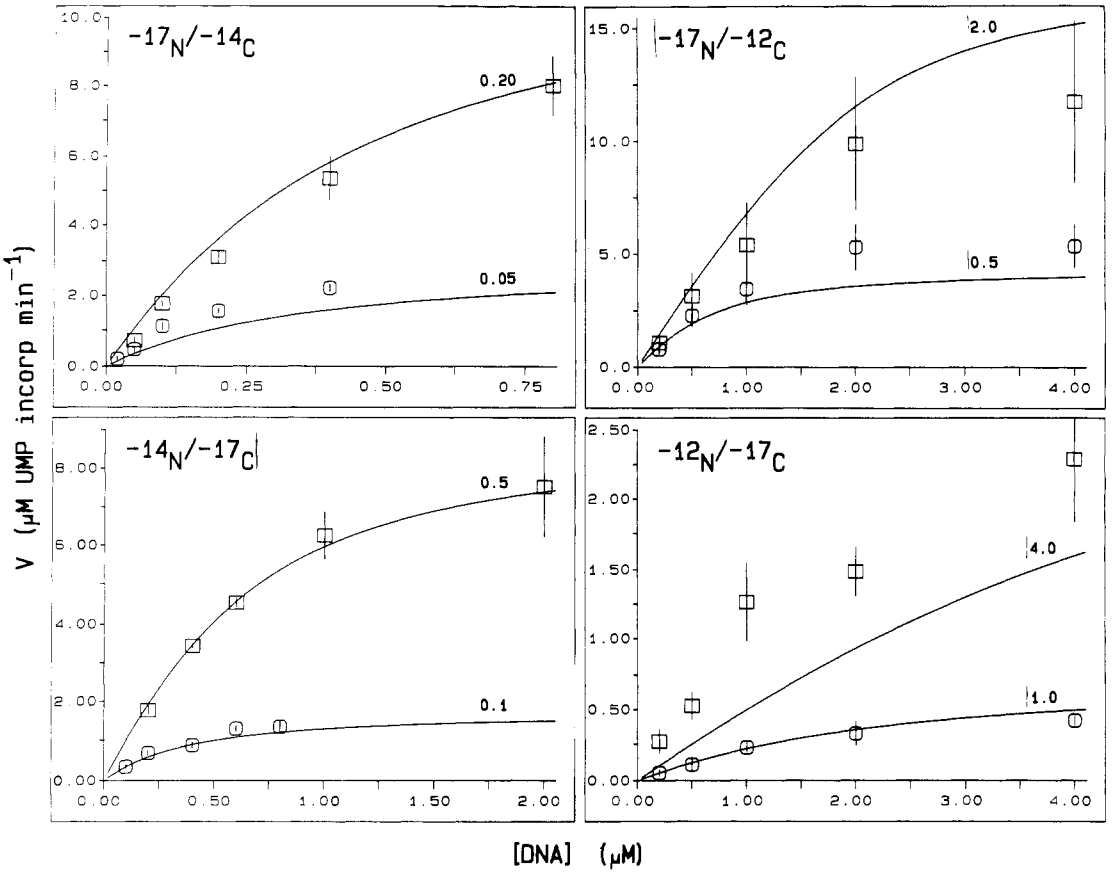

FIGURE 3: Comparison of transcription kinetics for promoters containing single-stranded ends at the upstream limit of the promoter. For each template, the upstream lengths of the coding (C) and noncoding (N) strands are indicated according to the nomenclature presented in the text. Data and predicted curves are presented as outlined in Figure 2.

of T7 and T3 promoters. Position −10 is a rigorously conserved A in the T7 promoters and is C in all T3 promoters sequenced to date. To investigate the importance of this base to promoter recognition in T7, we constructed a double-stranded consensus promoter −17*ds, in which the AT base pair at position −10

```
          -10          +1
           |           |
5'-TAATACGCCTCACTATA GGACT-3' (noncoding)
3'-ATTATGCGGAGTGATAT CCTGA-5' (coding)
```

is converted to a CG base pair. The results summarized in Table II show that recognition and proper message formation (see Figure 1) do occur at a maximum rate (k_{cat} = 34 min^{-1}) only slightly reduced from that of the consensus sequence. Specific promoter affinity, however, has decreased dramatically for this altered promoter (K_m = 0.14 μM).

To investigate whether this reduction in apparent binding affinity is more sensitive to changes in one strand than in the other, we prepared hybrids of this promoter in which only one of the two strands contains an errant base; i.e., there is a base mismatch at position −10. The kinetic data for these species (Table II) show the same trend as was observed in the studies of template length dependence. If the template contains a T3-type base in the noncoding strand ($-17^*_N/-17_C$), the binding and catalysis (K_m = 0.35 μM; k_{cat} = 24 min^{-1}) are comparable to or worse than the doubly mutated promoter. On the other hand, if the T3-like base is contained in the coding strand ($17_N/17^*_C$), specific transcription by T7 RNA polymerase is significantly retained (k_{cat} = 30 μM; K_m = 0.04 μM).

Discussion

The abortive initiation assay developed for *E. coli* RNA polymerase has provided a powerful tool for studying initiation in the bacterial enzyme (McClure, 1980). The kinetic assay system presented here provides the first opportunity to study isolated transcription initiation by the much simpler RNA polymerase from T7. The small size of the T7 promoters allows development of an assay involving transcription of a short runoff transcript from a synthetic oligonucleotide template. To date, activity assays for T7 RNA polymerase have relied on transcription from large templates, such as whole or fragmented T7 DNA (Chamberlin & Ring, 1973; Chamberlin & Ryan, 1982; McAllister et al., 1981), or from T7 promoters cloned into plasmids (Smeekens & Romano, 1986). These assays are complicated in that most do not produce a single short transcript of defined length and thus cannot be adapted to yield absolute quantitation of initiation.

The maximum rate of synthesis of the five-base message in the present assay system is 0.8 s^{-1} (50 min^{-1}). The average rate of nucleotide incorporation into RNA by T7 RNA polymerase has been estimated to be ~230 s^{-1} (Golomb & Chamberlin, 1974), a rate confirmed for the cloned enzyme by measuring the time intervals between partial and complete transcript formation during transcription from whole T7 DNA (Muller and Coleman, unpublished experiments). Hence, for messages less than 30 bases long the elongation phase of transcription should contribute less than 10% to the observed rate of transcript synthesis. In the present assay, the rate of synthesis of the 5-mer should therefore reflect the rate of initiation. Thus this system provides a well-defined assay for T7 RNA polymerase with which to determine the requirement for specific bases in promoter recognition and initiation.

Interpretation of Kinetic Data. The data obtained in this study fit well to Michaelis–Menten kinetics, as described in the previous section. This does not preclude the existence of

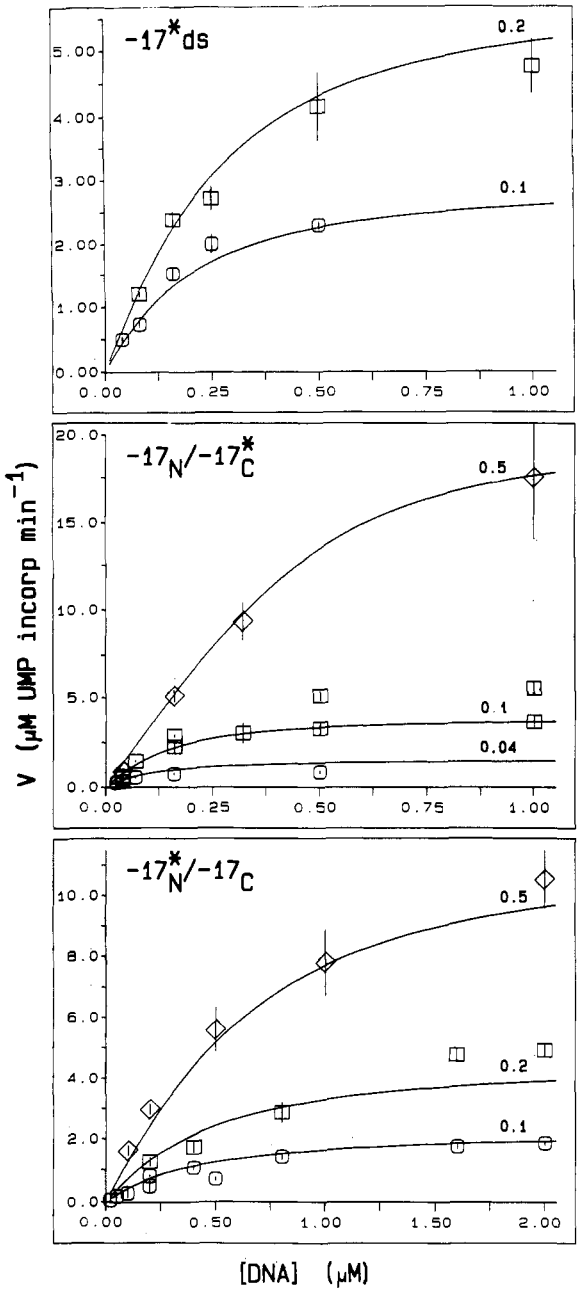

FIGURE 4: Comparison of transcription kinetics for promoters containing substitution of one or both bases at position −10. Promoter −17ds* is equivalent to −17ds but contains a CG base pair substituted for AT at position −10. The promoters containing a base mismatch possess either an AG pair ($17_N/17_C$*) or a CT pair (17_N*/17_C) at position −10. Data and predicted curves are presented as outlined in Figure 2.

a multistep mechanistic pathway in transcription initiation. The reversible binding of promoter, for example, may involve multiple discrete steps. Similarly, k_{cat} reflects only the slowest step or steps in the transition to a functioning elongation complex that can involve multiple intermediates as well.

One interpretation of the kinetic data that must be considered is that the rate-limiting step in the present short transcript assay is not initiation but release of enzyme from the template after synthesis of the five-base message (runoff). Although runoff has never been proposed to be slow in transcription of long messages, it is conceivable that residual interaction with the promoter remains after synthesis of a five-base message, thereby slowing enzyme–template dissociation. In this case, however, one would expect nonconsensus promoters which interact more weakly to yield larger k_{cat} values at saturation, a phenomenon not observed in any of our deviant promoters (Table I).

The current assay also provides the opportunity to determine the percentage of active enzyme in a given preparation. We can include in eq 2 a factor (α) indicating the fraction of total enzyme that is active. This factor may be fit in the nonlinear least-squares analysis along with K_m and k_{cat}. The resulting best fit to the data for the −17ds template (not shown) is essentially identical with that obtained assuming full enzymatic activity; *i.e.*, the enzyme preparation employed here is completely active ($\alpha = 1.0 \pm 0.2$).

Promoter Length Requirements: Upstream Length of Double-Stranded Promoter. Comparison of the kinetic data for the double-stranded promoters −19ds and −17ds shows that the consensus sequence up to position −17 is sufficient for maximum promoter recognition and transcription initiation. The substantially higher K_m measured for the −14ds template, however, verifies that a double-stranded promoter to position −14 is not sufficient for *maximum* usage, although specific recognition and initiation do occur. The major effect of the loss of three base pairs upstream from position −14 appears to be a decrease in binding affinity of the enzyme for the promoter template. Finally, the very poor transcription observed from the −12ds template suggests that critical determinants for promoter recognition lie upstream of position −12.

Contributions of Individual Coding and Noncoding Strands to Transcription Efficiency. Synthetic oligonucleotides also provide the ability to assess the relative contributions of the individual DNA strands to transcription initiation. Elements of promoter recognition that require specific features of a closed double-stranded DNA helix should be sensitive to the deletion of either DNA strand. On the other hand, if recognition occurs primarily through interactions with one of the individual strands after melting to form the open complex, then such interactions may be sensitive only to the alteration of one or the other of the individual DNA strands. The data presented in Table I for the $-17_N/-14_C$ and $-14_N/-17_C$ templates suggest only a small difference between the truncation of the coding strand and that of the noncoding strand at position −14. Hence the changes caused by truncation at position −14 are not due to specific interactions on one strand. It may be that recognition requires the promoter to be double-stranded from positions −14 to −17 or that a number of interactions occur separately with each of the complementary single strands.

A much more dramatic effect is observed when the truncation is moved to position −12. As shown in Table I, truncation of both strands beyond position −12 results in very poor promoter recognition and utilization. Extension of the coding strand to full length ($-12_N/-17_C$) results in very little improvement—specific recognition is almost insignificant. Extension of the *noncoding* strand, however, substantially improves both K_m and k_{cat}, suggesting a unique role for the upstream region of the noncoding strand in promoter recognition. Clearly, recognition determinants on the noncoding strand are effective in the absence of their complementary bases. By comparison of the −12/−17 pairs and the −14/−17 pairs it can be concluded that bases −13 and −14 of the noncoding strand are responsible for some important interactions with the protein.

Recognition by T7 RNA Polymerase of the Base(s) at Position −10 of the Promoter. The change of an AT base pair at position −10 to a CG base pair (as found in T3 RNA polymerase promoters) produces a 10-fold decrease in promoter recognition by the T7 enzyme as measured by K_m. Comparison of hybrids containing a mutated base on only *one*

of the two strands, i.e., containing a base mismatch at that site in the DNA, reveals that the decreased binding is mostly due to the presence of the wrong base in the *noncoding* strand. With the proper base in the noncoding strand ($-17_N/-17^*_C$), even the resulting base mismatch does not affect binding to the extent that an error in the noncoding strand does.

Implications for Promoter Binding to T7 RNA Polymerase. With all the synthetic promoters examined, decreased promoter recognition accompanying changes in promoter length or base sequence is significantly worse when the change occurs in the *noncoding* strand. It appears that the promoter recognition determinants in the upstream region (positions −12 to −17) lie mostly on the noncoding strand. Furthermore, recognition of these determinants does *not* require the presence of a double-stranded DNA helix. Although there is no *direct* evidence, it seems likely that the coding strand in the region near the first base of the message must interact specifically with the protein. The sequence from positions −2 to −9 is absolutely conserved in all T7, T3, and SP6 promoters found to date, yet the endonuclease digestion data indicate that the noncoding strand from positions −6 to +2 is, at least transiently, available for digestion. While positions −2 to −21 are protected from footprinting with the intercalator methidium-propyl-EDTA-Fe(II) (Ikeda & Richardson, 1986), the intercalator would not be expected to interact with the single-stranded regions of the promoter.

These considerations lead to a model for promoter recognition in which the open coding strand near the initiating base (approximately positions +2 to −6) and the noncoding strand farther upstream (positions −10 to −17) provide essential components of the DNA–protein interactions involved in specific promoter recognition and initiation of transcription. It is possible that the promoter is melted beyond position −6 (the position indicated by endonuclease digestion), perhaps as far as position −17. The initiation of transcription on the promoter $-17_N/-12_C$ shows that bases −13 to −17 in the noncoding strand need not be hydrogen bonded to complementary DNA bases for specific recognition (Table I). It is true, however, that the coding sequences of promoters present as single strands do not initiate transcription (Osterman & Coleman, 1981), a characteristic of these synthetic promoters as well (Martin and Coleman, unpublished results). Since the present data show that specific interactions between the noncoding strand and the protein are essential for promoter recognition, this fact alone would account for the finding that single-stranded promoter sequences will not initiate transcription. The question of how many base pairs of the double-stranded promoter are melted in the open complex or whether any specific bases must remain hydrogen bonded requires further investigation. Considering that a DNA duplex of 16 bases would have a minimum length of ~70 Å, a DNA binding groove that accommodates a melted region of this size would be expected to occupy a substantial region of the protein surface.

The kinetic assay system described here will allow further detailed study of promoter recognition by T7 RNA polymerase. In particular, the strictly conserved bases from positions −2 to −9 most certainly provide significant additional interactions with the enzyme. This system will also allow direct comparisons of promoter recognition between the enzymes from the T7, T3, and SP6 phages to better determine if there are specific inter- and intraclass promoter recognition domains.

References

Bailey, J. N., Klement, J. F., & McAllister, W. T. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 2814–2818.

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

Basu, S., Sarkar, P., Adhya, S., & Maitra, U. (1984) *J. Biol. Chem. 259*, 1993–1998.

Brown, J. E., Klement, J. F., & McAllister, W. T. (1986) *Nucleic Acids Res. 14*, 3521–3526.

Chamberlin, M., & Ring, J. (1973) *J. Biol. Chem. 248*, 2235–2244.

Chamberlin, M., & Ryan, T. (1982) *Enzymes (3rd Ed.) 15*, 87–108.

Chamberlin, M., McGrath, J., & Waskell, L. (1970) *Nature (London) 228*, 227–231.

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

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

Golomb, M., & Chamberlin, M. (1974) *J. Biol. Chem. 249*, 2858–2863.

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

Johnson, M. L., Correia, J. J., Yphantis, D. A., & Halvorson, H. R. (1981) *Biophys. J. 36*, 575–588.

King, G. C., Martin, C. T., Pham, T. T., & Coleman, J. E. (1986) *Biochemistry 25*, 36–40.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

McAllister, W. T., Morris, C., Rosenberg, A. H., & Studier, F. W. (1981) *J. Mol. Biol. 153*, 527–544.

McAllister, W. T., Horn, N. J., Bailey, J. N., MacWright, R. S., Jolliffe, L., Gocke, C., Klement, J. F., Dembinski, D. R., & Cleaves, G. R. (1983) in *Gene Expression* (Hamer, D., & Rosenberg, M., Eds.) pp 33–41, Liss, New York.

McClure, W. R. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5634–5638.

McGraw, N. J., Bailey, J. N., Cleaves, G. R., Dembinski, D. R., Gocke, C. R., Joliffe, L. K., MacWright, R. S., & McAllister, W. T. (1985) *Nucleic Acids Res. 13*, 6753–6766.

Moffatt, B. A., Dunn, J. J., & Studier, F. W. (1984) *J. Mol. Biol. 173*, 265–269.

Mulligan, M. E., Brosius, J., & McClure, W. R. (1985) *J. Biol. Chem. 260*, 3529–3538.

Oakley, J. L., & Coleman, J. E. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4266–4270.

Oakley, J. L., Pascale, J. A., & Coleman, J. E. (1975) *Biochemistry 14*, 4684–4691.

Oakley, J. L., Strothkamp, R. E., Sarris, A. H., & Coleman, J. E. (1979) *Biochemistry 18*, 528–537.

Osterman, H. L., & Coleman, J. E. (1981) *Biochemistry 20*, 4884–4892.

Smeekens, S. P., & Romano, L. J. (1986) *Nucleic Acids Res. 14*, 2811–2827.

Stahl, S. J., & Zinn, K. (1981) *J. Mol. Biol. 148*, 481–485.

Turner, B. W., Pettigrew, D. W., & Ackers, G. K. (1981) *Methods Enzymol. 76*, 596–628.