

Tests of a Model of Specific Contacts in T7 RNA Polymerase–Promoter Interactions[†]

Charlie Schick[‡] and Craig T. Martin^{*,§}

Program in Molecular and Cellular Biology and Department of Chemistry, University of Massachusetts, Amherst, Massachusetts 01003

Received August 9, 1994; Revised Manuscript Received November 1, 1994[⊗]

ABSTRACT: The T7, T3, and SP6 RNA polymerases represent a highly homologous family of enzymes that recognize similarly homologous promoter DNA sequences. Despite these similarities, the enzymes are highly specific for their respective promoters. Studies of mutant RNA polymerases have linked a specific amino acid residue in the protein to recognition of bases at positions –11 and –10 in the promoter [Raskin, C. A., *et al.* (1992) *J. Mol. Biol.* 228, 506–515]. In kinetic analyses of transcription from synthetic promoters containing base-analog substitutions, we have recently shown that at positions –11 and –10 of the T3 promoter, T3 RNA polymerase recognizes functional groups along the nontemplate strand wall of the major groove [Schick, C., & Martin, C. T. (1993) *Biochemistry* 32, 4275–4780]. We now extend these studies to the homologous region of the T7 promoter. The results confirm extrapolations from the T3 system and show that T7 RNA polymerase recognizes corresponding functional groups at positions –11 and –10 of the T7 promoter. The results are consistent with a direct readout model for recognition of these bases [Raskin, C. A., *et al.* (1992) *J. Mol. Biol.*, 228, 506–515], in which the 6-carbonyl and 7-imino groups of the nontemplate guanine at position –11 and the 6-amino group of the nontemplate adenine at position –10 of the T7 promoter are directly involved in binding. The results further support an overall model for promoter recognition in which the enzyme binds to one face of the duplex DNA in this upstream region of the promoter.

The homologous family of DNA-dependent RNA polymerases from bacteriophages T7, T3, and SP6 are single-subunit enzymes, with a molecular mass of ~100 kDa. The T7 and T3 enzymes are 82% exactly conserved in protein sequence (Stahl & Zinn, 1981; Moffatt *et al.*, 1984; McGraw *et al.*, 1985) yet are very specific for their respective promoters (Bailey *et al.*, 1983; McAllister *et al.*, 1983). The consensus DNA sequences for each of the T7 and T3 promoters are highly conserved over a 17 base pair region from positions –17 to –1 relative to the transcription start site (Oakley & Coleman, 1977; Dunn & Studier, 1983):

```

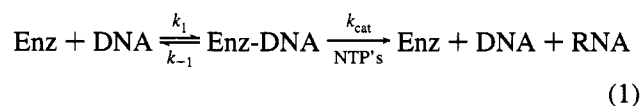
      -15      -10      -5      -1  →
T3.  A A T T A A C C C T C A C T A A A G G G A G
      T T A A T T G G G A G T G A T T T C C C T C

      -15      -10      -5      -1  →
T7.  T A A T A C G A C T C A C T A T A G G G A G
      A T T A T G C T G A G T G A T A T C C C T C
  
```

Comparison of the T7 and T3 consensus sequences reveals that the promoters differ primarily at base pairs –10, –11, and –12, and it has been proposed that these three base pairs are involved in the species-specific differential recognition of the promoters (Bailey *et al.*, 1983; Klement *et al.*, 1990).

The high degree of homology in protein and promoter sequences and the ease of manipulation of these simple enzymes and their promoters allow a detailed analysis of

the molecular interactions involved in differential promoter recognition. Using an *in vitro* steady-state kinetic assay of initiation (Martin & Coleman, 1987), we have determined steady-state kinetic parameters associated with transcription initiation from short oligonucleotide promoter templates. Equation 1 shows the simplest steady-state equation that fully describes the observed kinetics.



$$K_m = \frac{k_{-1} + k_{\text{cat}}}{k_1}$$

In these assays, the velocity of RNA synthesis is measured as a function of both total DNA and total enzyme concentration. A nonlinear least-squares fit of the velocity data as a function of the concentrations of both enzyme and DNA provides the single best fit set of the parameters k_{cat} and K_m , with associated joint confidence intervals. Although the mechanistic details underlying K_m and k_{cat} are not yet known, various lines of evidence indicate that K_m reflects primarily promoter binding and k_{cat} reflects the rate-limiting step(s) in the initiation of transcription (Maslak & Martin, 1993, 1994). In any case, relatively small changes in promoter function, which result from simple and well-controlled changes in the DNA or enzyme, lead to measurable changes in K_m and/or k_{cat} . Comparisons of K_m and k_{cat} values resulting from a series of systematic modifications of the enzyme–DNA system allow one to establish functional group interactions involved in the site-specific initiation of transcription.

In a recent study, we identified specific DNA contacts within the differential specificity region of the T3 promoter

[†] Supported by Grants NP-722 from the American Cancer Society and MCB-9308670 from the National Science Foundation.

[‡] Program in Molecular and Cellular Biology.

[§] Department of Chemistry.

[⊗] Abstract published in *Advance ACS Abstracts*, December 15, 1994.

(Schick & Martin, 1993). In order to test detailed models for promoter recognition in this family of RNA polymerases, we now extend the analysis to the corresponding region of the T7 promoter. Utilizing different nucleotide modifications, we probe recognition of the -10 and -11 base pairs in order to directly test the model for promoter recognition in this region of the promoter.

EXPERIMENTAL PROCEDURES

RNA Polymerase. T7 RNA polymerase was prepared from *Escherichia coli* strain BL21 carrying the overproducing plasmid pAR1219 (kindly supplied by F. W. Studier) cloned under inducible control of the *lac* UV5 promoter (Davanloo *et al.*, 1984). The enzyme was purified as described in King *et al.* (1986). Purity of the enzyme was verified by SDS–PAGE.¹

Oligonucleotides. Oligonucleotides were synthesized by the phosphoramidite method on a Milligen/Biosearch Cyclone Plus DNA synthesizer. Reagents were from Glen Research, Cruachem, BioGenex, and Milligen/Biosearch. Detritylation was monitored throughout each synthesis to verify the efficiency of coupling. Single strands from a 1 μ mol scale synthesis were purified trityl-on using an Amberchrome reverse-phase resin as described in Schick and Martin (1993). DNA purity was verified by denaturing gel electrophoresis.

Modified bases were incorporated using standard coupling procedures on the synthesizer. For practical reasons, the deoxyinosine base was coupled off-line, but on-column, using a procedure communicated to us by Hugh Mackey of Glen Research: the oligonucleotide was synthesized on-machine through oxidation of the base preceding the base to be coupled off-line, and the trityl group was removed on-machine. A 1-mL syringe, containing 10 μ mol of the phosphoramidite dissolved in 100 μ L of activator, was attached to the column, and the syringe contents were pushed through to an empty syringe on the other side of the column. The solution was alternately pushed back and forth through the column for 1 min. The column was then washed well with dry acetonitrile. Equal volumes (200 μ L) of Cap A and Cap B were loaded into a syringe and pushed through the column, reacting for 30 s. The column was then washed again with dry acetonitrile. Finally, 600 μ L of oxidizer solution was pushed through the column, reacting for 30 s, and the column was washed a final time with dry acetonitrile. At this point, the column was placed back on the machine and the synthesis of the remaining sequence was completed, trityl-on.

Since 7-deaza-dG is not readily available as a phosphoramidite derivative, the nucleoside triphosphate analog was incorporated into DNA enzymatically. To facilitate the construction of a mutant promoter containing a dGdC \rightarrow 7-deaza-dGdC substitution at position -11 , a larger fragment of DNA was used as the final transcription template. We have previously shown (Maslak & Martin, 1993) that, in the presence of only the nucleotides GTP and ATP, transcription from a 51 base pair synthetic oligodeoxynucleotide promoter, extending $-32\rightarrow+19$, gives the same kinetic parameters as

obtained with the 22 base pair ($-17\rightarrow+5$) promoter (the resultant RNA product is GGGAA in both cases). The sequence of the 7-deaza-dG construct (and its controls) is identical to that of this 51-mer.

51T: 5' CGCCTGCAGCCTGGTTCCTATAGTGTGATGATTAATTCGAAATCCGCC 3'

19P: 5' GGCGGAATTCGAAATTAAT 3'

28P: 5' GGCGGAATTCGAAATTAATACGACTCAC 3'

In the enzymatic synthesis of DNA, complementary single strands (51T and 19P, or 51T and 28P) were annealed and polymerization reactions were carried out in 0.5 nmol of 19P or 28P, 0.5 nmol of 51T, 750 μ M dATP, 500 μ M each of dCTP and dTTP, 50 mM NaCl, 10 mM Tris-HCl, pH 8.0, 10 mM MgCl₂, 1 mM DTT, and 3 units of T4 DNA polymerase (New England Biolabs). For the 7-deaza-dG construction, a final concentration of 500 μ M 7-deaza-dGTP (Boehringer Mannheim) was included in the polymerization reaction (in the corresponding controls, 500 μ M dGTP was used). After a 60-min incubation at 37 °C, the DNA polymerase was heat inactivated for 30 min at 80 °C. Nucleotides and salt from the polymerization reaction mix were removed by passage through a G-25 Sephadex spun column (Maniatis *et al.*, 1982). Resultant DNA solutions were diluted in 40 mM HEPES, pH 7.8, and 1 mM EDTA to bring the double-stranded DNA to a final concentration of 10 μ M. Purity of the constructs was verified by denaturing gel electrophoresis and restriction digest analysis (using the *Eco*RI site at position -27 and the *Hin*FI site at position -10).

Kinetics Assays. Assays of transcription initiation (Martin & Coleman, 1987) were carried out in a total volume of 20 μ L containing 30 mM HEPES pH 7.8, 15 mM magnesium acetate, 100 mM potassium glutamate, 0.25 mM EDTA, 1 mM DTT, 0.1 mg/mL *N,N*-dimethylated casein (Sigma), 0.05% TWEEN-20 (Calbiochem, protein grade), 0.8 mM GTP, and 0.4 mM ATP, as described in Maslak *et al.* (1993).

RESULTS

Members of the T7 family of RNA polymerases recognize a promoter approximately two helix turns in length. It has been proposed that promoter recognition at positions -11 and -10 occurs via direct contacts with base functional groups along the nontemplate wall of the major groove (Raskin *et al.*, 1992). In the T3 system, it has been shown that T3 RNA polymerase makes specific contacts in the major groove with the nontemplate cytidines at positions -11 and -10 , interacting with the exocyclic 4-amino groups on the cytidines (Schick & Martin, 1993). The current model predicts that the T7 enzyme must take the same general approach to its promoter and implicates involvement of the 6-carbonyl or the 7-imino group of guanine at position -11 and the 6-amino group of adenine at position -10 .

The synthetic promoters used in this study, excluding the construct containing 7-deaza-dG, are derived from the sequence shown above, consisting of the known T7 consensus sequence from positions -17 to -1 plus the DNA coding for the five-base message GGGAA. The 7-deaza-dG construct contains additional flanking sequences $-32\rightarrow-18$ upstream and $+6\rightarrow+19$ downstream. The message in this construct is GGGAAC..., such that the presence of only GTP and ATP in the reaction mixture will also lead to the

¹ Abbreviations: EDTA, ethylenediaminetetraacetic acid; DTT, dithiothreitol; PAGE, polyacrylamide gel electrophoresis; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid.

Table 1: Comparison of the Kinetic Parameters for T7 Native and Modified Promoters^a

DNA	K_m (nM)	k_{cat} (min ⁻¹)
Native ^b	2.0 (0.8-4.2)	28 (27-30)
-11dGdC→dIdC	2.7 (1.4-4.8)	30 (28-32)
-11dGdC→dG5medC	6.1 (3.8-9.8)	33 (31-35)
-11dGdC→dGdT	3.0 (1.5-5.8)	23 (22-24)
-11dGdC→dAdC	16 (9.5-27)	22 (20-24)
-11dGdC→dAdT	10 (6.5-15)	26 (24-29)
-11dGdC→dPdC	8.2 (5.2-13)	27 (25-29)
-11dGdC→d7-deaza-dGdC	35 (18-72)	21 (17-26)
-10dAdT→dAdU ^b	3.1 (1.6-6.1)	26 (24-27)
-10dAdT→dA5medC	6.9 (3.5-13)	38 (28-33)
-10dAdT→dGdT	32 (20-51)	25 (22-29)
-10dAdT→dG5medC	60 (34-112)	33 (27-41)
-10dAdT→dPdT	5.0 (2.6-10)	28 (26-30)
-10dAdT→dP5medC	14 (8.1-25)	35 (32-38)
-10dAdT→dCdG	30 (17-56)	20 (17-24)

^a Indicated ranges represent a 67% joint confidence interval for the best fit parameters. ^b From Maslak *et al.* (1993).

synthesis of the abortive initiation product GGGAA (Maslak & Martin, 1993). Steady-state kinetic parameters for the initiation of T7 RNA polymerase from its consensus promoter are $K_m = 2$ nM and $k_{cat} = 28$ min⁻¹ (Maslak *et al.*, 1993). These kinetic constants are almost identical to the corresponding values for the T3 enzyme under the same conditions (Schick & Martin, 1993), consistent with the high degree of homology between the enzymes.

Incorporation of nucleotide analogs at positions -11 and -10 allows direct tests of the predictions of the model for differential recognition. Modifications of the DNA begin with simple changes to the nucleotide base functional groups. The intent is to alter a local region of the major or minor groove, while minimally perturbing other potential contacts. Changes in K_m and k_{cat} resulting from the various modifications to the T7 promoters are compared in Table 1.

Position -11. Previous results from the T3 system strongly indicate the lack of any contacts in the minor groove at position -11 (Schick & Martin, 1993). In the T7 system, the presence of minor groove contacts can similarly be tested by replacing the nontemplate guanine with an inosine (dGdC→dIdC). This modification effectively removes the 2-amino group of guanine from the minor groove while leaving other functional groups unchanged. As predicted by extrapolation from the T3 system, this minor groove modification of the T7 promoter does not alter K_m or k_{cat} (3 nM and 28 min⁻¹, respectively).

On the opposite side of the helix at this position, direct contacts along the template strand wall in the major groove may be probed by the introduction of a methyl group into the major groove via the substitution dC→5-methyl-dC. To first order, the minor groove profile and the Watson-Crick base pairing are not disrupted by this change. As predicted from the model for the T3 system, in T7 the addition of a methyl group to the 5 substituent of cytosine at position -11 (dGdC→dG5medC) has minimal, if any, effect on the kinetic parameters ($K_m = 6$ nM; $k_{cat} = 33$ min⁻¹). From this platform, a more substantive alteration of the template strand base can be constructed by replacement of the template 5-methylcytosine with a thymine. In addition to replacing the 4-amino group of cytosine by a carbonyl, this modification disrupts conventional Watson-Crick base pairing at this position and leaves a mismatch. Nevertheless, this very large

substitution (dG5medC→dGdT) does not seem to affect binding or catalysis significantly ($K_m = 3$ nM; $k_{cat} = 23$ min⁻¹). This result suggests that the template strand base at this position is not important for binding and agrees with previous studies implicating recognition of the nontemplate strand in this region of the promoter (Jorgensen *et al.*, 1991; Ikeda *et al.*, 1992; Schick & Martin, 1993).

These results direct attention to the remaining functional groups in the major groove at this position: the 6-carbonyl and 7-imino groups of guanine on the nontemplate strand. Beginning with the dIdC base pair at position -11, which has native-like parameters, the substitution dIdC→dAdC introduces a mismatch and replaces the inosine 6-carbonyl by an amine while preserving the template base. This substitution does have a significant weakening effect on the kinetics ($K_m = 16$ nM). Since the introduction of a mismatch above had little effect on the kinetics, the disruptive effect of the dIdC→dAdC substitution is preliminarily assigned to the functional group substitution. Further support for this assignment is provided by the substitution dIdC→dAdT ($K_m = 10$ nM), which introduces the same modification of the 6-carbonyl, but in the context of a conventional base pair. These results strongly identify the replacement of the guanine 6-carbonyl by an amino group as disruptive to promoter binding.

Finally, in order to test whether the disruptive effect of these substitutions is due primarily to the removal of the native 6-carbonyl or to the addition of the non-native 6-amino group, the guanine base can be replaced by purine. The substitution dIdC→dPdC simply replaces the inosine 6-carbonyl by a hydrogen and introduces a mismatch. This substitution at position -11 weakens binding ($K_m = 8$ nM), with no effect on k_{cat} .

The above results strongly implicate the guanine 6-carbonyl in recognition of position -11. However, the data do not rule out contacts with the 7-imino group of guanine along the same wall of the major groove. Incorporation of the 7-deaza-dG analog into DNA effectively replaces a hydrogen bond acceptor by a hydrogen bond donor at the 7 position of guanine. Since 7-deaza-dG is difficult to incorporate into DNA chemically, we purchased the analog as a nucleoside triphosphate and constructed the mutant promoter containing dGdC→7-deaza-dGdC at position -11 using a larger fragment of DNA as the final transcription template.

To prepare promoter constructs containing 7-deaza-dG in the nontemplate DNA strand, the full-length template strand (51T) and a primer corresponding to the upstream nontemplate strand (19P, encompassing positions -32 to -14) were synthesized chemically. The remainder of the nontemplate strand was then synthesized using T4 DNA polymerase, substituting 7-deaza-dGTP in place of dGTP. In addition to the guanine at position -11, this synthesis places 7-deaza-dG in the nontemplate strand at positions +1, +2, +3, +9, +10, +13, +16, +17, and +19. To control for these additions, DNA was synthesized in which the -11 position contains guanine as part of the primer (28P). In this case, all guanines downstream of position -11 in the nontemplate strand contain 7-deaza-dG. Finally, as a check on the synthesis process, parallel constructs were synthesized with regular dGTP in place of 7-deaza-dGTP.

The two promoter constructs which contained dGTP throughout yielded native kinetic values in the initiation

assay, both demonstrating the fidelity of the enzymatic DNA synthesis and verifying that transcription from the longer transcription templates yields the same kinetics as transcription from shorter oligonucleotide constructs (Martin & Coleman, 1987). The kinetics of transcription on the promoter construct containing 7-deaza-dG downstream of position -5 show that the presence of 7-deaza-dG at positions +1, +2, and +3 in the nontemplate strand message region does not affect the kinetic parameters ($K_m = 2$ nM; $k_{cat} = 41$ min⁻¹). Finally, the construct additionally containing the dGdC→7-deaza-dGdC substitution at position -11 shows a significant increase in K_m (35 nM) and a small decrease in k_{cat} (21 min⁻¹). This result unequivocally demonstrates the importance of the 7-imino group of guanine in the major groove of the nontemplate strand at position -11.

Position -10. It has been shown previously that recognition at position -10, like that at position -11, occurs on the nontemplate side of the major groove (Jorgensen *et al.*, 1991; Schick & Martin, 1993). Additionally, it has been shown that in the T7 system, removal of the methyl group of thymine on the template strand (T→dU) at position -10 has no effect on binding (Maslak & Martin, 1993). To further test the proposal that the template strand is not critical for recognition at this position, the template strand thymine can be replaced by 5-methylcytosine. The more disruptive substitution dAdT→dA5medC introduces a mismatch and replaces the 4-carbonyl of thymine with an amine. This change results in a relatively small effect on the measured kinetic parameters ($k_{cat} = 31$ min⁻¹; $K_m = 7$ nM). This small increase in K_m , if significant, may be a result of the mismatch created, as has been seen previously for the same position in the T3 system (Schick & Martin, 1993). These results suggest that the template strand thymine is not directly involved in binding at this position.

Returning to potential recognition elements on the nontemplate side of the major groove at position -10, the 7-imino group and the 6-amino group of adenine are the most likely contacts. The substitution dAdT→dGdT replaces the 6-amino group of adenine with a carbonyl in the major groove and adds an amino group to the minor groove. This change weakens binding significantly ($K_m = 32$ nM) while leaving k_{cat} unchanged (25 min⁻¹). Although this result is consistent with recognition of the 6-amino group of adenine, the disruption of base pairing forced by the dGdT substitution may lead to displacement of other critical functional groups. Incorporation of dG5-methyl-dC restores the base pairing, but the large disruption in the kinetics ($K_m = 60$ nM) remains. These results support a model for recognition in which the 6-amino group of adenine at this position is involved in binding.

To test more simply for recognition of the 6-amino group of adenine, a dAdT→dPdT substitution can be introduced at position -10, to first order replacing the 6-amino group of adenine by a hydrogen and leaving the 7-imino group unchanged. Contrary to simple predictions, this removal of the 6-amino group (dAdT→dPdT) has no significant effect on binding or catalysis ($K_m = 5$ nM; $k_{cat} = 28$ min⁻¹). However, in the substitution of adenine by purine (dAdT→dPdT), the 6-amino group is replaced by a hydrogen, leaving a functional group vacancy. It is likely that in this construct a water molecule would fill this vacancy and could position as shown in Figure 1. In adopting this position, the water

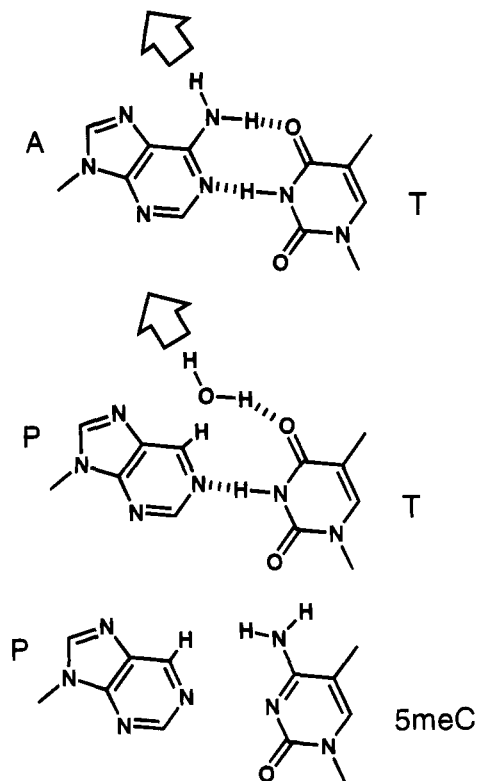


FIGURE 1: Model of water-mediated contact at position -10 in T7. Replacement of the 6-amino group by a hydrogen may allow a water molecule to position as an alternate hydrogen bond donor. Incorporation of an amino group in place of the stabilizing 4-carbonyl group from the opposing template base disrupts this coordination of water.

molecule could establish a stabilizing contact with the 4-carbonyl of thymine and possibly with the 6-hydrogen of purine, while leaving a proper hydrogen bond donor in a position similar to that of the native adenine amino group. To test this proposal, the opposing thymine on the template strand can be replaced by 5-methyl-dC, replacing the thymine 4-carbonyl by an amino group, which cannot stabilize the proposed orientation of water. Indeed, the substitution dPdT→dP5medC has a substantial weakening effect on initiation ($K_m = 14$ nM; $k_{cat} = 35$ min⁻¹), even though the same substitutions, taken separately (dAdT→dA5medC or dAdT→dPdT), have no significant effect on the kinetic parameters. This result suggests that in the dPdT base pair a water molecule can be positioned to make stabilizing contacts in the major groove, replacing the original contact from the 6-amino group of the nontemplate adenine.

DISCUSSION

Studies of artificially constructed hybrid polymerases have located a region in the T7 and T3 proteins, from amino acid residues 674–754, that is involved in the differential specificity of the T7 and T3 enzymes for their promoters (Joho *et al.*, 1990). More recently, Raskin *et al.* (1992) have identified a specific amino acid within this region which is involved in differential recognition of the two polymerases. A change to one residue in each of the T7 and T3 RNA polymerases (T7 Asn748→Asp; T3 Asp749→Asn) is sufficient to switch the local specificities at positions -11 and -10 of the polymerases to the bases of the heterologous sequence. From these studies, they have proposed a model of differential recognition involving the base pair at position

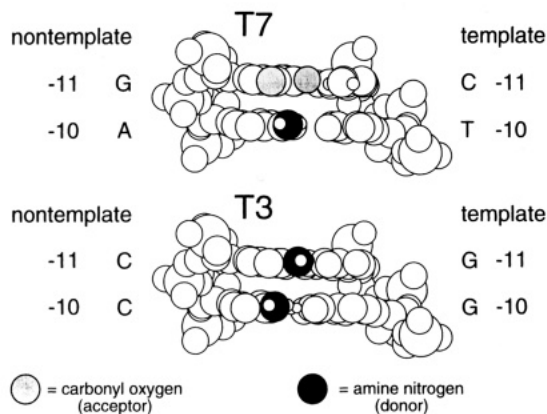


FIGURE 2: Model of the dinucleotide step for positions -11 and -10 . A view into the major groove of B-form DNA shows shaded atoms indicating the base contacts implicated for T7 (this study) and the corresponding contacts in T3 (Schick & Martin, 1993; Raskin *et al.*, 1992).

-11 , with possible contributions from the base pair at position -10 . Specifically, the model proposes that the carboxyl group of Asp749 in T3 makes a bidentate contact with the 4-amino groups of the nontemplate strand cytosines at positions -11 and -10 . In T7, Asn748 is proposed to make a similar bidentate contact with the 6-amino group of adenine at position -10 and the 7-imino group of guanine at position -11 . Crystallographic studies of T7 RNA polymerase suggest that Asn748 lies in a putative DNA binding cleft at the appropriate distance from the proposed active site to form contacts with positions -10 and -11 of the promoter (Sousa *et al.*, 1993).

Specific Protein–DNA Interactions at Positions -10 and -11 . Through functional group mutagenesis of the DNA, specific base functional groups which contribute to promoter recognition at positions -10 and -11 in T3 RNA polymerase have recently been identified (Schick & Martin, 1993). The results clearly indicate a recognition of only major groove functional groups and strongly suggest a local and specific interaction with the cytosine 4-amino group at position -10 in the T3 promoter. The results also suggest that at position -11 it is the 4-amino group that is involved in recognition, consistent with the model proposed by Raskin *et al.* (1992).

The current recognition model, supported by results in the T3 system, leads to specific predictions regarding the mechanism of promoter recognition in the T7 system. As in T3, recognition in T7 is predicted to be mostly through direct contacts of the major groove functional groups of the nontemplate bases at positions -10 and -11 . Functional groups predicted to be available for recognition are the 6-carbonyl or the 7-imino group of the nontemplate strand guanine at position -11 and the 6-amino group of the nontemplate strand adenine at position -10 . Using kinetic analysis of transcription from synthetic promoters specifically modified at positions -10 and -11 , we have directly tested these predictions.

Position -10 Contacts. The results at position -10 are consistent with a model in which the nature of the template strand base is not important. Changes to the template strand have at most a slight effect on the kinetic parameters. The data suggest that contact at this position is through the base on the nontemplate strand, specifically with the 6-amino group of adenine as shown in Figure 2. These results are consistent with extrapolations from the T3 system (Schick

& Martin, 1993) and agree with the proposal from Raskin *et al.* (1993). The data suggest that the replacement of the 6-amino group by a hydrogen (A→P) allows a water molecule to coordinate across the major groove, supplying a hydrogen donor functionally equivalent to the native 6-amino group of adenine. In this case, the nature of the base on the opposing strand is significant in providing buttressing of the water molecule. This interpretation supports the model for recognition in which the 6-amino group of adenine is the dominant contact at this position in T7.

Position -11 Contacts. The changes at position -11 are consistent with a model in which the polymerase makes contacts with the nontemplate strand base in the major groove, in particular, with the 6-carbonyl and 7-imino groups of the nontemplate guanine, as illustrated in Figure 2. As predicted from results in the T3 system, replacement of the template strand base at this position has no effect on binding, while modifications of the nontemplate strand base increase K_m . Modification of the 6-carbonyl group, by changing the guanine either to a purine or to an adenine, has similar weakening effects. Evidence for weaker interactions with the 6-carbonyl of guanine and the 4-amino group of cytosine at position -11 in T7 and T3, respectively, relative to the interactions observed at position -10 , suggests that other amino acids may contribute to the specificity at position -11 , and that contacts of the 7-imino group may be influenced weakly by the nature of the functional group at the 6-position of guanine. Finally, this result suggests the possibility that position -10 , rather than position -11 , is the dominant contact in this region. In any case, these results agree with the proposal (Raskin *et al.*, 1992) that T7 RNA polymerase makes a hydrogen bond contact with an acceptor on the nontemplate strand guanine at position -11 .

Evidence for a Direct Readout of Duplex DNA. Footprinting studies in the T7 system have suggested that the polymerase binds primarily to one face of a fully or partially duplex form of the promoter (Muller *et al.*, 1989), and more recent studies support this model in the region of the promoter upstream from about position -6 (Maslak *et al.*, 1993; Maslak & Martin, 1993; Schick & Martin, 1993). Specifically, the protection pattern suggests that the bound protein spans the major groove of duplex DNA in the region of differential specificity (positions -12 to -8), directly reading the major groove base functional groups.

As summarized in Figure 3, ethylation and methylation interference data on T7 and T3 promoters (Jorgensen *et al.*, 1989, 1991), along with current and previous kinetic data from T7 and T3 promoters (Martin & Coleman, 1987; Maslak *et al.*, 1993; Schick & Martin, 1993), clearly demonstrate that from position -12 to position -10 the polymerase binds the major groove, interacting preferentially with the bases on the nontemplate strand upstream of, and including, position -10 . Contacts shift to the template strand wall of the major groove at position -9 in T3 and T7; changes in the major groove of the *template* strand base at position -9 disrupt binding, while changes in the nontemplate base have no effect (data not shown).

All data presented to date are consistent with a simple direct readout of functional groups in the major groove in this region of the promoter. Results from the introduction of mismatches and whole base changes argue against recognition based on sequence-dependent changes in the

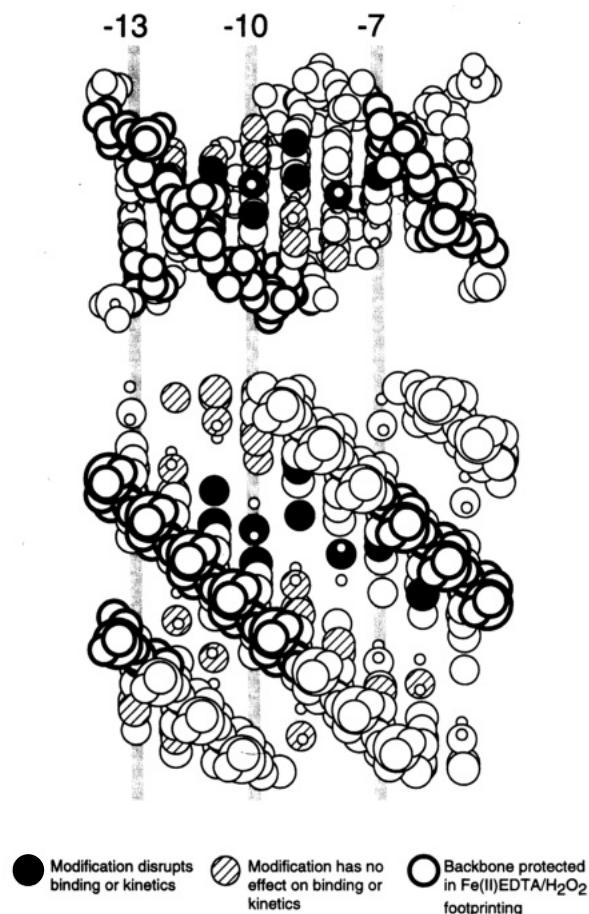


FIGURE 3: Summary of the recognition pattern identified in the current work [see also Schick (1994)] and in previous footprinting (Muller *et al.*, 1989) and interference studies (Jorgensen *et al.*, 1989, 1991).

structure of the DNA backbone. In many cases, base pair mismatches are tolerated, while in other cases, very simple functional group substitutions are not. The clear differences between predicted major and minor groove effects observed in these studies provide further support for a direct readout of major groove functional groups in this region of the promoter. For example, the lack of any perturbation in the kinetics accompanying removal of the guanine 2-amino groups at positions -11 and -10 in T3 (Schick & Martin, 1993) and position -11 in T7 strongly suggests that no minor groove contacts are occurring in this region.

These results highlight the utility of base functional group modifications in the elucidation of promoter recognition contacts. Modifications to the DNA described in this work range from simple (G \rightarrow I, C \rightarrow 5meC) to more substantial (base pair mismatches). In principle in this type of study, the effect of a given substitution might result from the simple disruption of direct contacts or from more indirect influences on the structure of the DNA. However, the results presented here, combined with previous results, present a compelling picture of simple direct readout through major groove contacts. Furthermore, the combined results suggest that a major component of the differential specificity displayed by the T7 and T3 enzymes arises from simple complementary changes in a single amino acid and one or two bases in the DNA (Raskin *et al.*, 1992).

Similar analyses in the SP6 system are underway to provide a still more rigorous test of this model. Alignment

of the amino acid sequences from the T7, T3, and SP6 RNA polymerases implicates an arginine in the recognition by the SP6 enzyme of a dGdC base pair at position -11 and both dTdA and dGdC at position -10 in the consensus SP6 promoter. The current model (Raskin *et al.*, 1992) predicts that recognition of this base pair step will occur via nontemplate strand functional groups in the major groove. Specifically, arginine has the proper geometry to form bidentate contacts with the hydrogen bond acceptors of guanine at position -11 and the hydrogen bond acceptors of thymine or guanine at position -10 . Finally, the addition to these studies of an array of T7, T3, and SP6 enzymes containing directed mutations in the proposed recognition region of the protein (Raskin *et al.*, 1992, 1993) will further refine the developing model of this central region of the promoter.

ACKNOWLEDGMENT

We are grateful to Maribeth Maslak for valuable discussions and to William T. McAllister for numerous discussions and for communicating results prior to publication.

SUPPLEMENTARY MATERIAL AVAILABLE

Steady-state kinetic fits for constructs listed in Table 1 (3 pages). Ordering information is given on any current masthead page.

REFERENCES

- Bailey, J. N., Klement, J. F., & McAllister, W. T. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2814–2818.
- 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.
- Ikeda, R. A., Warshamana, G. S., & Chang, L. L. (1992) *Biochemistry* **31**, 9073–9080.
- Joho, K. E., Gross, L. B., McGraw, N. J., Raskin, C., & McAllister, W. T. (1990) *J. Mol. Biol.* **215**, 31–39.
- Jorgensen, E. D., Joho, K., Risman, S., Moorefield, M. B., & McAllister, W. T. (1989) *DNA Protein Interactions in Transcription*, pp 79–88, Alan R. Liss, Inc., New York.
- Jorgensen, E. D., Durbin, R. K., Risman, S. S., & McAllister, W. T. (1991) *J. Biol. Chem.* **266**, 645–651.
- King, G. C., Martin, C. T., Pham, T. T., & Coleman, J. E. (1986) *Biochemistry* **25**, 36–40.
- Klement, J. F., Moorefield, M. B., Jorgensen, E., Brown, J. E., Risman, S., & McAllister, W. T. (1990) *J. Mol. Biol.* **215**, 21–29.
- Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Martin, C. T., & Coleman, J. E. (1987) *Biochemistry* **26**, 2690–2696.
- Maslak, M., & Martin, C. T. (1993) *Biochemistry* **32**, 4281–4285.
- Maslak, M., & Martin, C. T. (1994) *Biochemistry* **33**, 6918–6924.
- Maslak, M., Jaworski, M. D., & Martin, C. T. (1993) *Biochemistry* **32**, 4270–4274.
- McAllister, W. T., Horn, N. J., Bailey, J. N., MacWright, R. S., Jolliffe, L., Gocke, C., Klement, J. F., Dembrinski, D. R., & Cleaves, G. R. (1983) *Gene Expression* (Hamer, D. H., & Rosenberg, M. J., Eds.) pp 33–41, Alan R. Liss, Inc., New York.

- McGraw, N. J., Bailey, J. N., Cleaves, G. R., Dembrinski, D. R., Gocke, C. R., Jolliffe, 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.
- Muller, D. K., Martin, C. T., & Coleman, J. E. (1989) *Biochemistry* 28, 3306–3313.
- Oakley, J. L., & Coleman, J. E. (1977) *Proc. Natl. Acad. Sci. U.S.A.* 74, 4266–4270.
- Raskin, C. A., Diaz, G., Joho, K., & McAllister, W. T. (1992) *J. Mol. Biol.* 228, 506–515.
- Raskin, C. A., Diaz, G., & McAllister, W. T. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90, 3147–3151.
- Schick, C. (1994) Ph.D. Thesis, University of Massachusetts at Amherst.
- Schick, C., & Martin, C. T. (1993) *Biochemistry* 32, 4275–4280.
- Sousa, R. J., Chung, Y. J., Rose, J. P., & Wang, B. C. (1993) *Nature* 364, 593–599.
- Stahl, S. J., & Zinn, K. (1981) *J. Mol. Biol.* 148, 481–485.

BI941826E