Major Groove Recognition Elements in the Middle of the T7 RNA Polymerase Promoter†

Tong Li,‡ Hoi Hung Ho,‡ Maribeth Maslak,‡ Charlie Schick,§ and Craig T. Martin*,‡

Department of Chemistry and Program in Molecular and Cellular Biology, University of Massachusetts, Amherst, Massachusetts 01003-4510

Received October 12, 1995; Revised Manuscript Received January 18, 1996⊗

ABSTRACT: T7 RNA polymerase recognizes a relatively small promoter extending only 17 base pairs upstream from the start site for transcription. A model for this recognition suggests that the enzyme interacts with the major groove of duplex DNA in the region centered at position −9 [Muller, D. K., et al. (1989) *Biochemistry 28*, 3306−3313], and recent kinetic analyses of promoters containing base analogs at positions −10 and −11 have provided support for this model [Schick, C., & Martin, C. T. (1993) *Biochemistry 32*, 4275−4280; Schick, C., & Martin, C. T. (1995) *Biochemistry 34*, 666−672]. In the current work, we extend this analysis across the proposed major groove, identifying specific base functional group contacts at positions −9 through −5. Specifically, the 6-carbonyl of guanine at positions −9 and −7, the 6-amino group of adenine at position −8, the 5-methyl group of thymine at position −6, and the 2-amino group of guanine at position −5 are identified as primary contacts. The results strongly support the model for duplex recognition in this region of the promoter and suggest that recognition continues along one face of the helix beyond the major groove and into the adjoining minor groove at position −5, where helix melting begins.

The T7 family of DNA-dependent RNA polymerases presents an ideal model system in which to understand detailed structure−function relationships in transcription. The consensus promoter sequences extend only 17 base pairs upstream from the start site for transcription (Oakley & Coleman, 1977), yet the enzymes are highly specific for their respective promoters. In the past two decades, numerous studies have been carried out *in vitro* and *in vivo* to determine important contacts in promoter recognition. Very early studies used base analogs within the DNA in an attempt to identify functional groups which might contact the enzyme (Stahl & Chamberlin, 1976, 1978), but most studies have focused on transcription from promoters containing single or multiple base pair substitutions (Chapman & Burgess, 1987; Chapman et al., 1988; Schneider & Stormo, 1989; Klement et al., 1990; Jorgensen et al., 1991; Ikeda, 1992; Ikeda et al., 1992a,b; Raskin et al., 1992; Diaz et al., 1993). The incorporation of single base pair substitutions can identify specific positions in the promoter critical to recognition but provides little information on DNA contacts at a structural level. Chemical modification interference studies have identified potential phosphate and guanine N7 base contacts (Jorgensen et al., 1991). Finally, footprinting of the static enzyme−promoter complex with Fe(II)EDTA/H_2O_2 has provided a broad overview of potential contacts and has prompted a general model for promoter recognition (Muller et al., 1989).

The development of an oligonucleotide-based kinetic assay for the initiation of transcription allowed a new approach to the elucidation of promoter recognition in this system (Martin & Coleman, 1987). Via the incorporation of base analogs at single sites within the promoter, the relative contributions to promoter recognition of individual chemical groups in the DNA can be determined. The controlled nature of this assay also allows measurement of effects which might be missed in other approaches (for example, a 10-fold weakening of promoter binding might not be observed in a single measurement of transcription at high enzyme and DNA concentrations). Recent studies using the oligonucleotide-based approach have identified a number of specific functional group contacts within the promoter (Maslak et al., 1993; Schick & Martin, 1993, 1995). Near positions −10 and −11, this approach has confirmed a detailed model for a specific protein−DNA contact in the major groove (Raskin et al., 1992).

These data, combined with the determination of the crystal structure of the uncomplexed enzyme (Sousa et al., 1993), have supported a model for promoter recognition in which the enzyme recognizes one face of a closed DNA duplex, spanning two major grooves and the adjoining minor groove (Muller et al., 1989). However, results from transcription using partially single-stranded synthetic promoters strongly indicate that, near the start site for transcription, energetically important contacts occur only with the template strand of the DNA, presumably in a melted form of the promoter (Milligan et al., 1987; Maslak & Martin, 1993). This result has prompted a return to a two-domain model for promoter function (Chapman & Burgess, 1987), in which the upstream contacts contribute simple binding to the major groove of duplex DNA while the downstream sequence facilitates helix melting and the direction of the initiating DNA template bases to the enzyme active site.

The question remains as to which bases are duplex, and which are single-stranded, in the energetically important recognition complex (or complexes) involved in the initiation

† Supported by Grant MCB-9308670 from the National Science Foundation.

‡ Department of Chemistry.

§ Program in Molecular and Cellular Biology.

⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1996.

0006-2960/96/0435-3722$12.00/0

of transcription. In other words, how far downstream does the upstream duplex recognition region extend? To address this question and to further test the model, the current studies target bases at positions −9 to −5 of the T7 (and T3) RNA polymerase promoters.

EXPERIMENTAL PROCEDURES

RNA Polymerase. T7 and T3 RNA polymerases were prepared from *Escherichia coli* strain BL21 carrying the overproducing plasmid pAR1219 (kindly supplied by F. W. Studier) or plasmid pCM56 (kindly supplied by W. T. McAllister), respectively. In this system, the RNA polymerase is expressed under inducible control of the *lac* UV5 promoter (Davanloo et al., 1984; Morris et al., 1986). 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, Prime Synthesis, and Milligen/Biosearch. Detritylation was monitored throughout each synthesis to verify the efficiency of coupling. Single strands from a 1 μmol scale synthesis were purified trityl-on using an Amberchrome reverse phase resin as described previously (Schick & Martin, 1993).

Modified bases were incorporated using standard coupling procedures on the synthesizer. To conserve reagents, the base analogs were coupled off-line, but on-column, using a procedure communicated to us by Hugh Mackey of Glen Research (Schick & Martin, 1995).

Kinetics Assays. Assays of transcription initiation (Martin & Coleman, 1987) were carried out in a total volume of 20 μL 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 previously (Maslak et al., 1993). For each template, reaction velocities were measured at various enzyme and DNA concentrations. The error in each velocity was approximated as the higher of 0.1 μM/min or the *t* distribution 80% confidence interval of the fitted slope for the three time points. Velocity data were then fit as previously described (Martin & Coleman, 1987) to the exact solution of the steady state equation, using a weighted nonlinear least-squares minimization algorithm based on the Gauss−Newton method (Johnson et al., 1981). Ranges in the values represent a 67% joint confidence interval of the fitted parameters. As a result of the nonlinear nature of the velocity equation and the potential interdependence of the fit parameters, increases in K_m typically have more confidence than do decreases (Johnson, 1983). In this study, the best fit k_{cat} measured for the native promoter varied somewhat (27−35 min^{-1}) between different preparations of the enzyme. Changes for mutant promoters were only deemed significant relative to a native promoter control using the same enzyme preparation. In any case, small changes in k_{cat} (less than a factor of 2) should be interpreted with caution.

[1] Abbreviations: EDTA, ethylenediaminetetraacetic acid; DTT, dithiothreitol; PAGE, polyacrylamide gel electrophoresis; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; SDS, sodium dodecyl sulfate.

Table 1: Comparison of the Kinetic Parameters for Modified Promoters

	K_m (nM)	range[a]	k_{cat} (min^{-1})	range[a]
native				
T7[b]	2.0	0.8−4.2	28	26−30
T3	1.0	0.4−2.4	40	39−42
position −9				
dCdG → dTdG	0.2	<0.1−2	42	39−44
→ dCdP	−		low	
position −9 (T3)				
→ dCdP	−		low	
→ dTdA	−		low	
→ d^{5me}CdG	1.0	0.3−3.0	33	31−35
→ dCdI	3.0	1.0−6.0	35	33−36
position −8				
dTdA → dUdA	2.9	1.6−5.0	39	37−41
→ dTdP	42	25−69	22	18−25
→ dTdI	244	143−461	23	15−31
→ dTdG	−		low	
position −7				
dCdG → d^{5me}CdG	2.5	0.8−6.8	44	40−47
→ dCdA	66	20−246	1.0	0.6−1.6
→ dTdA	40	18−93	8.7	6.5−11
→ dTdG	10.4	5.1−21	28	25−31
→ dCdI	5.0	1.8−13	15	13−17
position −6				
dAdT → dAdU[b]	23	13−40	23	20−25
→ dPdT	5.4	2.8−10	24	22−2
→ dGdT	6.0	3.5−10.5	34	31−37
→ dAdC	16	10−23	18	16−20
→ dGdC	27	20−37	18	16−20
→ dAd^{5me}C	3.9	0.4−19	11.4	9.5−12.9
→ dGdI	−		low	
position −5				
dCdG → d^{5me}CdG	4.3	1.9−9.4	35	31−39
→ dTdG	12	7.2−21	23	20−25
→ dCdI	163	112−246	13	10−16
→ dTdA	40	18−93	8.7	6.5−10.9
→ dCdA	−		low	
→ dCd^{2A}P	−		low	

[a] Indicated ranges represent a 67% joint confidence interval for the best fit parameters. [b] From Maslak et al. (1993). For native promoter constructs, best fit values for K_m typically fall within the range 0.8-4.2 nM, while values for k_{cat} depend somewhat on the enzyme preparation, varying from 27 to 35 min^{-1}.

RESULTS

The synthesis of oligonucleotides containing base analogs at unique positions in the promoter DNA provides a powerful tool for the elucidation of critical structural contacts in a protein−DNA complex. The application of the steady state kinetic assay for the initiation of transcription to assess the effects of the perturbations provides a functional measure of the relative contributions of individual groups to mechanistically important step(s) in the initiation process. In a previous study, we have presented evidence that the steady state parameter K_m provides at least a relative measure of changes in promoter binding (Maslak & Martin, 1994). The parameter k_{cat} appears to reflect changes in the rate-limiting step or steps in initiation, but melting of the DNA helix near the start site is only a part of this rate (removing completely the barrier to melting results in only a doubling of k_{cat}). While some of the mutations presented below do not significantly perturb the initiation kinetics, others result in activity so low that valid kinetic parameters could not be obtained (indicated in Table 1 as low). The kinetic data are presented in Table 1 and summarized pictorially in Figure 1, which shows for each base pair, regions identified as involved or not involved in promoter recognition.

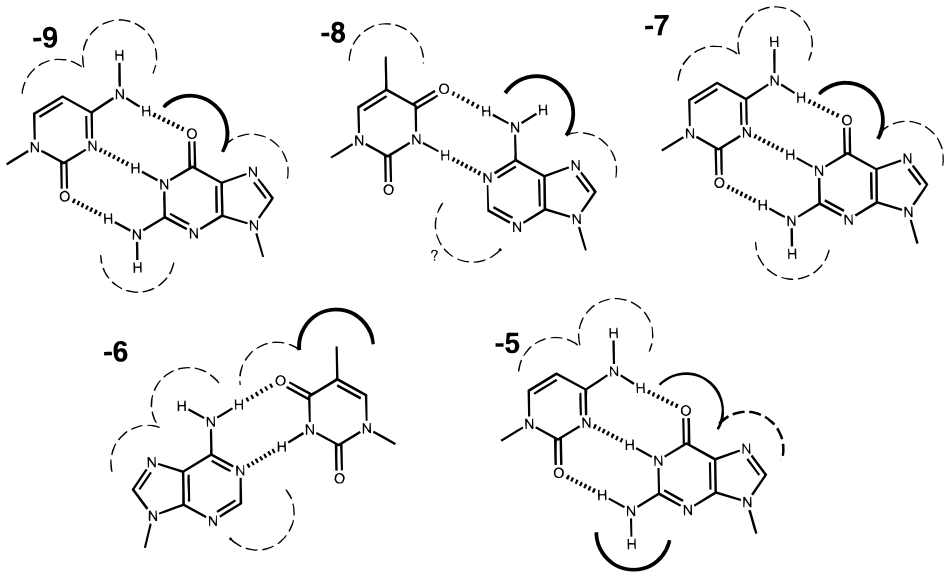


FIGURE 1: Base contact profiles. The five base pairs examined in this study are shown above. Functional groups identified as probable contact sites are outlined by a bold curve, while those identified as unlikely to be involved in promoter recognition are indicated by dashed curves.

Position −9. T7 and T3 RNA polymerase both possess a CG base pair at position −9 of their consensus promoters. The results in Table 1 show that in the T7 system the substitution dCdG → dTdG, which introduces a base mismatch and alters substantially the base in the nontemplate strand, has little effect on the steady state kinetics of transcription initiation, suggesting that recognition does not occur with cytidine. In contrast, the substitution dCdG → dCdP (P = purine), which removes from guanine both the 6-carbonyl group in the major groove and the 2-amino group in the minor groove, results in transcription too low to obtain reliable kinetic parameters, implicating recognition of the template strand guanine. In related studies in the T3 system (these results were not repeated in the T7 system since both recognize the same base at this position), the same very large effect on transcription is seen not only for the dCdG → dCdP substitution but also for the dCdG → dTdA substitution, as expected for recognition of guanine. Finally, the simple substitution dCdG → d^{5me}CdG results in no change in kinetic parameters, consistent with a lack of recognition of the nontemplate base C. Together, these results very clearly identify the template strand as the location of the recognized functional group(s). This leaves the guanine 6-carbonyl and/or 2-amino groups as recognition contacts. To resolve this ambiguity, the simple substitution dCdG → dCdI removes the 2-amino group from the minor groove side (while retaining the 6-carbonyl) and results in no change in kinetic parameters. This result, combined with the large disruptive effect of the dCdG → dCdP substitution, points very clearly to the 6-carbonyl of guanine as the primary recognition determinant at position −9.

Position −8. In a previous study in the T7 system, we demonstrated that the thymine methyl group of the TA pair at position −8 is not involved in recognition (Maslak et al., 1993), implicating the template strand at this position. By manipulation of the template strand, the substitution dTdA → dTdP removes the adenine 6-amino group and results in a substantial increase in K_m, with a small decrease in k_{cat}. Replacement of the hydrogen bond donor amino group by an acceptor carbonyl (dTdA → dTdI) further weakens binding as reflected in K_m, while the mismatch pair introduced by the substitution dTdA → dTdG weakens initiation so severely that kinetic parameters are not readily determined. As for position −9, these results again demonstrate recognition of the template strand base and point strongly to the adenine 6-amino group as the primary recognition element at position −8.

Position −7. The simple addition of a methyl group at the 5-position of cytidine (dCdG → d^{5me}CdG) at position −7 results in no significant change in the kinetic parameters, as predicted by the model. The substitution dCdG → dCdA at this position introduces a mismatch and replaces the guanine base in the template strand by adenine. The kinetic data from this construct show both a large increase in K_m and a large decrease in k_{cat}, indicating a major disruption of promoter recognition. That this does not result simply from the introduction of a mismatch is supported by similar results for the substitution dCdG → dTdA. Indeed, incorporation of the substitution dCdG → dTdG, which introduces a mismatch while preserving the nature of the template strand base, results in kinetic parameters much closer to those of wild type. Finally, the very simple substitution dCdG → dCdI, which removes the guanine 2-amino group, results in only a modest reduction in k_{cat}, suggesting a lack of minor groove contacts. These results demonstrate clearly the critical nature of the base pair at position −7 and point strongly to recognition of the guanine 6-carbonyl on the template strand of the DNA.

Position −6. Previous results from incorporation of deoxyuridine (dAdT → dAdU) at position −6 pointed to the thymine methyl group on the template strand as a recognition determinant (Maslak et al., 1993). That the template strand is the only contact at this position is supported by the result that the substitutions dAdT → dPdT and dAdT → dGdT have very little effect on the kinetics. In fact, the latter substitution introduces a base mismatch, yet near wild type kinetic parameters are retained. In contrast, the constructs dAdT → dAdC and dAdT → dGdC result in significant increases in K_m and decreases in k_{cat}. The primary importance of the pyrimidine 5-methyl group is clearly evident from the results of the substitution dAdT → $dAd^{5me}C$, in which K_m returns almost to wild type levels (although k_{cat} is still slightly

depressed). Interestingly, none of these substitutions reduces transcription as dramatically as do some substitutions at positions −7 and −8. To introduce an even larger perturbation at this site, we incorporated the bulky purine−purine mismatch dAdT → dGdI, which replaces thymine by the purine base inosine. As expected for this very large structural perturbation, the initiation of transcription for this construct is too low to obtain valid kinetic parameters. Although our current results point clearly to the thymine 5-methyl group as a recognition determinant, other factors may also be important at this position.

Position −5. The above results support the model for promoter recognition derived from the footprinting data, showing contacts across the full width of the major groove. The model further predicts that at position −5 template strand contacts will continue to dominate the interaction but that recognition contacts may be expected to cross from the major to the minor groove as the protein−DNA contacts continue along one face of the DNA helix. The prediction that nontemplate strand contacts are not involved in this recognition is confirmed by results from the substitutions dCdG → d^{5me}CdG and dCdG → dTdG, which perturb the base in the nontemplate strand yet show very little change in the kinetic parameters. In contrast, the much simpler substitution dCdG → dCdI, which simply removes the 2-amino group from the minor groove side of the template strand base, results in an increase in K_m of more than 1 order of magnitude and a large decrease in k_{cat}. The substitutions dCdG → dTdA and dCdG → dCdA, which introduce mismatches but also result in the removal of the guanine 2-amino group, similarly have large effects on the kinetics. Finally, the substitution dCdG → $dCd^{2A}P$ removes the guanine 6-carbonyl, while the 2-amino group is retained, yet still results in very low kinetics, suggesting that contact with the 6-carbonyl is also significant. These results clearly demonstrate recognition of the guanine base at position −5, as predicted by a model in which the DNA is recognized primarily as a duplex at position −5. They further suggest that recognition has crossed over to the minor groove, as predicted by the model of duplex recognition in this region of the promoter.

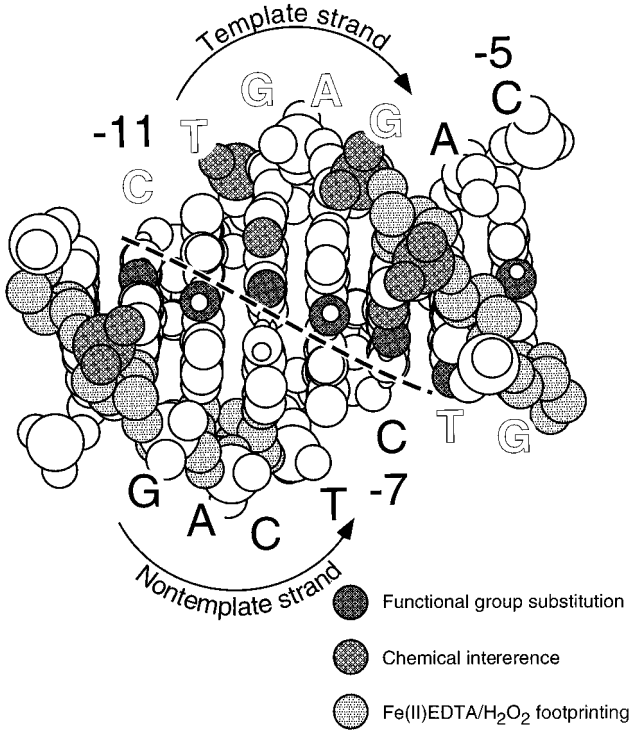
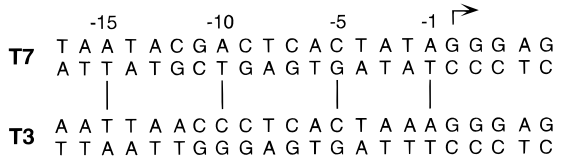


FIGURE 2: Summary of promoter recognition contacts. The T7 RNA polymerase promoter sequence modeled as B-form DNA, showing identified recognition elements. The sugars protected in footprinting with Fe(II)EDTA/H_2O_2 are shown in light gray (Muller et al., 1989). Guanine N7 and phosphate groups identified via chemical modification interference studies are shown in medium gray (Jorgensen et al., 1991). Base functional groups identified as contacts in current and previous kinetic analyses (Maslak et al., 1993; Schick & Martin, 1993, 1995) are shown in dark gray. The dashed line separates bases derived from the template and nontemplate strands of the DNA. Note that this pattern of contacts strongly supports a model in which the enzyme recognizes one face of double-stranded DNA in this region of the promoter.

DISCUSSION

The consensus promoter sequences for the T7 family of RNA polymerases extend 17 base pairs upstream from the start site for transcription. For the T7 and T3 enzymes, the consensus promoters are identical from positions −3 to −9 (Oakley & Coleman, 1977; Bailey et al., 1983; Dunn & Studier, 1983). Among the known promoter sequences in the T7, T3, and SP6 genomes, this is also the most highly conserved region of the promoter sequences. Studies incorporating base pair substitutions both *in vitro* and *in vivo* have generally confirmed the critical nature of this region (Chapman & Burgess, 1987; Chapman et al., 1988; Schneider & Stormo, 1989; Klement et al., 1990; Jorgensen et al., 1991; Ikeda, 1992; Ikeda et al., 1992a,b; Raskin et al., 1992; Diaz et al., 1993).

The introduction of base analogs into specific sites within the DNA allows a detailed probing of individual base contacts in promoter recognition. Using this approach, we have recently identified specific functional group contacts at positions −10 and −11 of the T3 and T7 promoters (Schick & Martin, 1993, 1995).



In this region, the T7 and T3 enzymes show species-specific recognition of their respective consensus promoter sequences; however, the results show that the overall approach of the two enzymes to the DNA is the same. Specifically, the data confirm an earlier model in which Asn748 in T7 and Asp749 in T3 make bidentate contacts with major groove functional groups at positions −10 and −11 (Raskin et al., 1992). The illustration in Figure 2 highlights the guanine 6-carbonyl and adenine 6-amino groups identified as contacts at positions −11 and −10, respectively, of the T7 promoter. These results, combined with data from hydroxyl radical footprinting (Muller et al., 1989), predict that the enzyme makes contacts across the major groove, starting on the nontemplate strand at the upstream end (positions −11 and −10) and progressing to the template strand at the downstream end (positions −8, −7, and −6).

Major Groove Contacts are Recognized from Positions −9 to −6. The current results confirm recognition of major

groove functional groups at positions −9 through −6. As predicted by the model, these contacts cross the major groove, moving from the nontemplate strand side at position −10 to the template strand side at position −9. In particular, the results show clearly that the guanine 6-carbonyl group is the major contact at position −9 and that contacts do not occur on the minor groove side of this base pair. The involvement of the guanine 6-carbonyl group is consistent with previous results, which showed that methylation of the 7-imino group of guanine at position −9 interferes with static promoter binding (Jorgensen et al., 1991). Although such interference could occur via direct disruption of enzyme contacts with the guanine N7 group, the enzyme's inability to recognize adenine at this position suggests that methylation of the guanine N7 group causes an indirect disruption of interactions at the neighboring 6-carbonyl group, which is positioned in line with other identified contacts.

At position −8, conversion of adenine to either purine, which removes the adenine N6 amino group, or guanine, which replaces it with a carbonyl, weakens initiation of transcription by much more than 1 order of magnitude (to an extent which does not allow the accurate determination of the steady state kinetic parameters). The large decrease in initiation accompanying the substitution of adenine by purine, in particular, argues strongly for recognition of the adenine N6 amino group, since the rest of the base pair remains unchanged in structure. Similarly, the current substitutions at position −7 point clearly toward recognition of the guanine 6-carbonyl and against recognition of either the minor groove or the nontemplate side of the major groove. The previous methylation interference results implicate recognition of the template strand side of the major groove at position −7, and possibly direct recognition of the guanine N7 group (Jorgensen et al., 1991), in addition to the guanine 6-carbonyl. Previous studies employing single base substitutions have confirmed the sensitive nature of the promoter sequence between positions −10 and −7. A study using genetic selection of functional promoters concluded that the consensus bases confer the maximum recognition potential at each position (Schneider & Stormo, 1989), consistent with the placement of position −9 at the center of the major groove recognition (Muller et al., 1989).

Previous results have implicated in promoter recognition the thymine 5-methyl group at position −6 (Maslak et al., 1993). In the current study, the substitution dAdT → dGdC leads to an increase in K_m of more than 1 order of magnitude, while the substitution dAdT → dGd^{5me}C, which differs only in the presence of the thymine 5-methyl group, results in near native kinetic parameters. This result suggests that the thymine methyl is the principle, and perhaps the only, base recognition determinant at this position. In the current model, the 5-methyl group is expected to lie close to the backbone along the downstream edge of the major groove in duplex DNA. In contrast, minor groove functional groups at position −6 do not lie along the proposed recognition face. Further evidence that this region is recognized as double-stranded DNA comes from comparison of the full base pair substitution dAdT → dGdC, which should not disrupt the helical nature of the DNA and results in only about a factor of 10 increase in K_m, with the substitution dAdT → dGdI, which introduces a purine−purine mismatch and results in a much larger disruption in initiation kinetics.

Recognition of One Face Predicts Major to Minor Groove Transition. As illustrated in Figure 2, the duplex binding model derived from the footprinting results predicts that the major groove contacts should move closer to the phosphate backbone at position −6 and cross over (at least partially) to the minor groove side of the base pairs at position −5 (Muller et al., 1989). This crossing over of contacts is supported by ethylation interference studies, in which ethylation of phosphates at positions −5 through −7 on the template strand interferes with static binding (Jorgensen et al., 1991). It is also of interest to note that, unlike the results at positions −9 through −7, all full base pair substitutions studied at positions −6 and −5 allow measurable kinetic parameters. At the backbone crossover point, the base functional groups are presumably less directly accessible to the protein interface, such that base specificity is less stringent. This was also observed in previous studies examining full base pair substitutions *in vitro* and *in vivo* (Chapman & Burgess, 1987; Chapman et al., 1988; Ikeda et al., 1992a; Diaz et al., 1993). In fact, in a study with selection for functional mutant promoters in a plasmid-based expression system, all possible base pairs were found at position −6 but only the consensus sequence was observed at positions −9 to −7 (Schneider & Stormo, 1989). These results are consistent with the model in which position −6 begins the crossover between major and minor groove recognition, such that direct contacts with base functional groups are more hindered than at the center of the recognized major groove.

At position −5, recognition is expected to cross, at least partially, to the minor groove as the contacts continue downstream along one face of the helix. The current results demonstrate that at position −5 recognition has in fact crossed over to the minor groove. The simple substitution dCdG → dCdI at position −5 results in a large increase in K_m and a significant decrease in k_{cat}, strongly implicating in specific recognition the guanine 2-amino group on the minor groove side of the base. This represents the first identifiable minor groove contact in the region from position −12 to −5. The magnitude of this perturbation suggests that this single modification may explain an early observation in which incorporation of deoxyinosine (hypoxanthine) throughout the nontemplate strand of a T7 phage promoter resulted in complete loss of activity (Stahl & Chamberlin, 1978).

Methylation interference results have suggested a possible major groove contact at position −5 as well (Jorgensen et al., 1991). In this case, methylation of the adjacent N7 position of adenine is reported to *enhance* binding. As at position −6, the duplex binding model predicts that recognition at position −5 should straddle the phosphate backbone and could involve contributions from both the major and minor grooves. The results from the other modifications of guanine are consistent with the additional recognition of a major groove determinant, most likely the guanine 6-carbonyl.

Transition to Single-Stranded Recognition Downstream. Recent studies targeting bases closer to the start site for transcription (Maslak & Martin, 1993; H. H. Ho, T. Li, M. Maslak, and C. T. Martin, manuscript in preparation) indicate that recognition four to five base pairs upstream of the initiation site occurs in a melted form of the DNA, interactions occurring exclusively with the template strand (in this context, distinctions between major and minor groove are

irrelevant, although the nomenclature remains useful to describe the side of each base on which specific functional groups lie). The results presented here suggest that this melted region must be confined to positions −1 to −4. Within the resolution possible in this type of analysis, melting is expected to begin between positions −4 and −5 of the promoter. This refinement of the model has important implications for attempts at docking DNA models onto the crystal structure for the enzyme (Sousa et al., 1993).

The recognition of base elements straddling the major and minor grooves of DNA is an unusual motif and raises the question of whether this recognition plays a specific role in the initiation of transcription. An early promoter analysis *in vitro* showed that, at position −6, the substitution dAdT → dTdA decreases activity to 6% of wild type on linear plasmid DNA but to only 84% of wild type on supercoiled plasmid DNA (Chapman & Burgess, 1987). Subsequent measurements *in vivo* with the promoter on (supercoiled) plasmid DNA showed a relatively small decrease in activity for this same mutation (Chapman et al., 1988; Ikeda et al., 1992a). Indeed, an *in vivo* (plasmid-based) promoter selection study concluded that this position is not recognized at all, despite its complete conservation in the (linear) bacteriophage genome (Schneider & Stormo, 1989). These results suggest a role for this downstream end of the duplex recognition region. The contacts in the major and minor grooves at positions −6 and −5 may play a role in facilitation of the melting of the region downstream of position −5. In assays on negatively supercoiled DNA, the energetic barrier to melting should be lower, such that in those studies contacts at positions −5 and −6 are less critical to initiation (note that the T7 genome is linear *in vivo*). This proposal can now be tested *in vitro* on partially single-stranded templates which are effectively premelted near the start site for transcription (Maslak & Martin, 1993).

REFERENCES

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

Chapman, K. A., & Burgess, R. R. (1987) *Nucleic Acids Res. 15*, 5413−5432.

Chapman, K. A., Gunderson, S. I., Anello, M., Wells, R. D., & Burgess, R. R. (1988) *Nucleic Acids Res. 16*, 4511−4524.

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

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

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

Ikeda, R. A. (1992) *J. Biol. Chem. 267*, 11322−11328.

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

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

Johnson, M. L. (1983) *Biophys. J. 44*, 101−106.

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

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.

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.

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

Morris, C. E., Klement, J. F., & McAllister, W. T. (1986) *Gene 41*, 193−200.

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.

Schick, C., & Martin, C. T. (1993) *Biochemistry 32*, 4275−4280.

Schick, C., & Martin, C. T. (1995) *Biochemistry 34*, 666−672.

Schneider, T. D., & Stormo, G. D. (1989) *Nucleic Acids Res. 17*, 659−674.

Sousa, R., Chung, Y. J., Rose, J. P., & Wang, B. C. (1993) *Nature 364*, 593−599.

Stahl, S. J., & Chamberlin, M. J. (1976) in *RNA polymerase* (Losick, R., & Chamberlin, M., Eds.) pp 429−440, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Stahl, S. J., & Chamberlin, M. J. (1978) *J. Biol. Chem. 253*, 4951−4959.

BI9524373