JMB

Positioning of the Start Site in the Initiation of Transcription by Bacteriophage T7 RNA Polymerase

Benjamin F. Weston, Iaroslav Kuzmine and Craig T. Martin*

Department of Chemistry
University of Massachusetts
Amherst, MA 01003-4510, USA

The determination of various polymerase structures has sparked interest in understanding how the polynucleotide template interacts with the active site. In the primer-independent initiation of transcription, an additional question arises as to how the complex directs the first two bases of the template uniquely into the active site. Recent studies in the model RNA polymerase from bacteriophage T7 demonstrate that upstream duplex contacts provide at least some of the binding specificity and suggest that the enzyme interacts with the template strand in a melted context near the start site for transcription. The current work probes the role of the template strand in positioning of the first two templating bases during initiation. The results suggest that such positioning is not rate-limiting in steady-state turnover, and that the insertion of a very large and flexible linker three or four bases upstream of the start site has no significant effect on the fidelity of start site selection. The insertion of linkers immediately adjacent to the start site, however, does significantly decrease the fidelity of start site selection (as evidenced by a large increase in misinitiation at position $+2$, with little change in the observed rate of correct initiation), suggesting that some of the non-transcribed template DNA does help to position the first two templating bases into the active site of the RNA polymerase. Finally, incorporation of an abasic site at position -1 yields a similar decrease in initiation fidelity, suggesting a role for stacking of the bases at positions -1 and $+1$.

© 1997 Academic Press Limited

*Keywords:* RNA polymerase; transcription; nucleoside analogs; initiation; abasic

*Corresponding author

Introduction

In determining a unique site along the DNA for the initiation of transcription, an RNA polymerase, like many sequence-specific DNA-binding proteins, must interact strongly with specific functional groups along the promoter DNA. In addition, an RNA polymerase must direct the initial templating bases ($+1$ and $+2$) of the melted DNA precisely into the polymerase active site. However, after the initiation of transcription, an RNA polymerase should behave as a sequence-independent polymerase, accommodating any templating bases into the active site. In the simplest model to accommodate these requirements, the polymerase (phospho-transfer) active site would not contain direct specificity elements, but rather would achieve specificity through adjacent upstream (non-transcribed) contacts with the promoter. In fact, previous studies in the bacteriophage T7 system indicate that its promoter can be loosely divided into an upstream duplex recognition domain and a

downstream melted domain comprising the site of transcription initiation (Chapman & Burgess, 1987; Li *et al.*, 1996; Maslak & Martin, 1993).

T7 RNA polymerase presents an ideal model system in which to study fundamental mechanistic questions in transcription. The single subunit enzyme, which requires no additional factors, is highly specific for a relatively short promoter sequence, and initiation begins with good fidelity at a unique site within the promoter. An emerging model for promoter recognition in this system, summarized in Figure 1, suggests that the recognition of promoter elements upstream of about position -5 occurs *via* simple major groove contacts in duplex DNA (Chapman & Burgess, 1987; Li *et al.*, 1996; Schick & Martin, 1995). Downstream of position -5, specificity determinants shift exclusively to the template strand, in a melted context (Maslak & Martin, 1993; unpublished results). It seems obvious that one function of the sequence "TATA" (positions -4 to -1 of the T7 promoter) is to provide a low barrier to the melting of the DNA
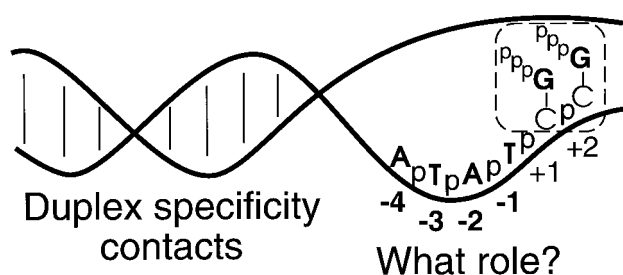
Figure 1. A model for promoter recognition by T7 RNA polymerase, emphasizing upstream duplex recognition elements and downstream template strand contacts. Two nucleoside triphosphate bases are shown bound at the active site.

at the adjoining transcription start site (Breslauer *et al.*, 1986; Jia *et al.*, 1996; Újvári & Martin, 1996). Indeed, *Escherichia coli* RNA polymerase can initiate primer-independent transcription from non-promoter sequences containing an artificially created bubble in the DNA, although with significant loss in start site selection fidelity (Aiyar *et al.*, 1994). Finally, a narrow region of sequence-derived melting is not sufficient to uniquely define the start site, since T7 RNA polymerase initiates with native fidelity from constructs that are single-stranded downstream from position −1 (Maslak & Martin, 1993; Milligan *et al.*, 1987; and see below).

One might envision different mechanisms to achieve the initiation of transcription at a unique site in the promoter. In one extreme, specificity contacts on the template strand at positions +1 and +2 would serve to direct those bases into the active site. In another extreme, mentioned above, all specificity would come from upstream contacts, several bases distant from the start site, with simple linear spacing of the template strand from those contacts directing the positioning of the initial templating bases. Non-specific interactions with the phosphate backbone might facilitate the direction of the templating bases to the active site.

The introduction of nucleoside analogs into synthetic oligonucleotides allows simple tests of structural constraints in DNA function. A specific analog that is useful in maintaining backbone connectivity, while removing all base and sugar constraints, is the simple propyl spacer shown in Figure 2. This spacer and longer polyethylene glycol spacers have been used previously to cross-link the ends of duplexes (Altmann *et al.*, 1995) or as components of hairpin structures (Durand *et al.*, 1990; Williams & Hall, 1996). Non-nucleosidic spacers have also been exploited to link two domains of DNA or RNA, to test the role of specific regions of the DNA backbone in helicase function (Amaratunga & Lohman, 1993) or of the RNA backbone in RNA splice site selection (Pasman & Garcia-Blanco, 1996).

The promoter DNA construct used in the current studies consists of the T7 RNA polymerase consen-

sus sequence from positions −17 through −1, followed by DNA encoding the 5 base runoff transcript GGGAA.

```
 -17        -10       -5      +1    +5
 T A A T A C G A C T C A C T A T A G G G A A   nontemplate
 A T T A T G C T G A G T G A T A T C C C T T   template
                                  ⌐→
```

We have shown previously that for full promoter function, only the template strand is required in the region (positions −4 to −1) linking the upstream recognition element of the T7 RNA polymerase promoter with the start site at position +1 (Maslak & Martin, 1993). In order to characterize the role of this DNA in positioning the start site, we have incorporated into the promoter at unique positions, a simple non-nucleoside spacer that mimics the normal DNA backbone in its overall connectivity. Although the backbone in this construct can adopt the path taken by the backbone in the native promoter complex, it is not constrained to do so by interactions with, or conformational constraints of, the sugar and/or base. We have extended this analysis to include the substitution of individual nucleosides by longer alkyl spacers, to disrupt the spacing and to effectively lengthen the connection between the upstream specificity element and the downstream start site for transcription. The effects of these substitutions on both initiation kinetics and start site fidelity are examined, in order to provide insight into the mechanism of template strand positioning in the initiation of transcription.

Results

In transcription, the template strand of the DNA acts as a cofactor, binding to the enzyme and forming a part of the active site. The templating bases in the DNA then direct the corresponding ribonucleoside triphosphate bases to the active site, in order to allow specific phosphotransfer chemistry. In initiation, the template strand presumably serves to direct the placement of the first two nucleoside triphosphates. Thus positions +1 and +2 of the DNA template (the templating bases) must be placed into well-defined positions within the active site prior to initiation. Previous studies have defined upstream promoter elements involved in the specific binding of T7 RNA polymerase to its promoter (Li *et al.*, 1996; Maslak *et al.*, 1993; Schick & Martin, 1995). Studies have also shown that constructs containing the (single-stranded) template DNA coupled to an upstream duplex portion of the promoter are sufficient to allow initiation of transcription (Maslak & Martin, 1993; Milligan *et al.*, 1987). Indeed, the steady-state rate of initiation increases twofold on such constructs (Maslak & Martin, 1993). What then are the roles of template strand nucleotides immediately adjacent to the templating bases? Distance and conformational constraints imposed by the linkage to upstream duplex recognition elements might ac-
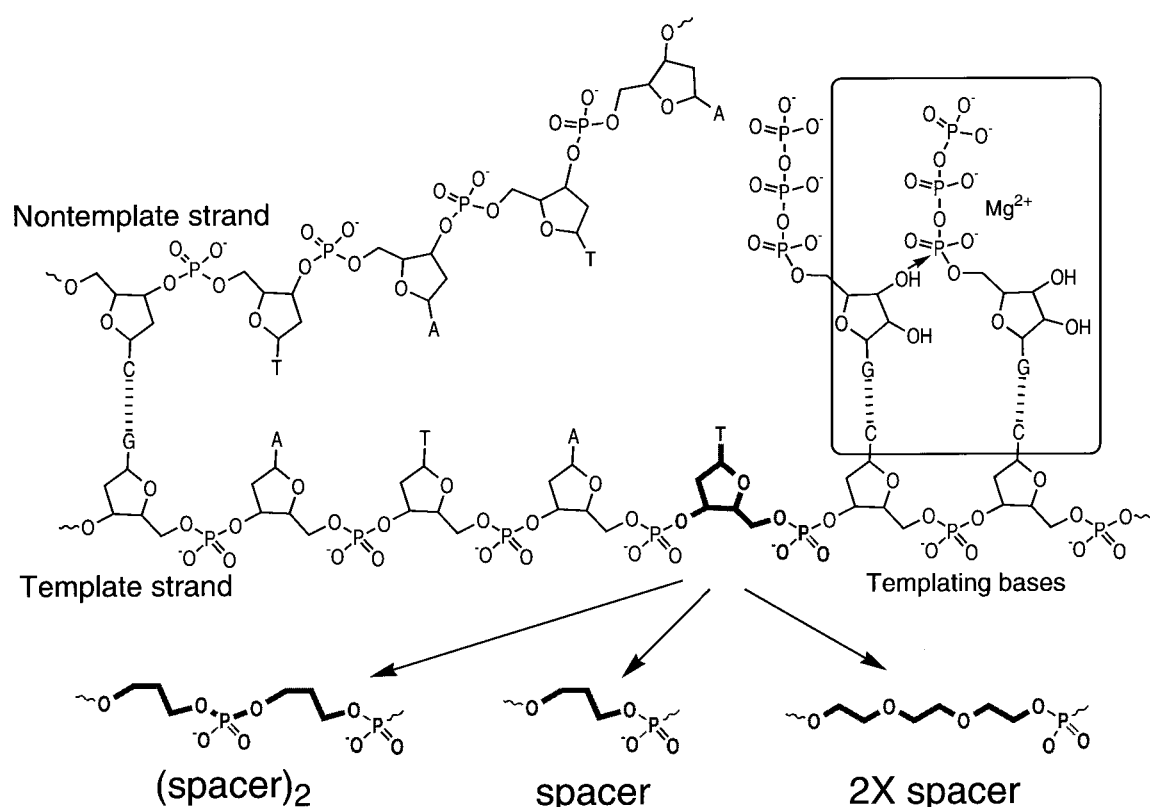
Figure 2. Description of the non-nucleosidic linkers and their incorporation at position −1, immediately adjacent to the start site for transcription. Note that (spacer)$_2$ and 2X-spacer have comparable extended lengths, but very different linear charge densities.

tively direct the templating bases into the active site. Other factors, such as stacking of the +1 and −1 bases, could further stabilize the correct structure.

Spacer substitutions at position −1

Replacement of the normal nucleotide linkage at position −1 by a simple alkyl linkage allows the correct placement of the downstream DNA into the active site, but removes steric constraints imposed by the sugar and removes any potential interactions of the base at position −1 with the templating base at position +1. The results summarized in Table 1 show that in multiple turnover transcription from a double-stranded construct encoding a five base runoff transcript, substitution of the nucleoside at position −1 by a simple spacer leads to a twofold increase in the apparent activity. That transcription is not dramatically reduced by this substitution suggests that the base at position −1 does not play an essential role in the rate-limiting step(s) governing the initiation of transcription.

Under the conditions used for this simple activity assay, the enzyme should be about 75 to 80% saturated by promoter, assuming a native K_d of 5 nM (Újvári & Martin, 1996). Therefore a twofold increase in activity cannot arise solely from an increase in binding affinity. More detailed steady-state kinetic assays yield the kinetic parameters

$K_m = 6$ nM and $k_{cat} = 80$ minute^{-1} for the promoter construct containing a simple spacer in the template strand at position −1 (compared with $K_m = 6$ nM and $k_{cat} = 30$ minute^{-1} for the native promoter construct). A two- to threefold increase in apparent activity is therefore predicted from the increase in k_{cat}. Promoter binding (as reflected in the parameter K_m) appears not to be affected by this substitution.

The substitution of a spacer that mimics the length of a normal nucleoside results in a relatively small change in the rate of initiation. To what extent can more dramatic structural perturbations be incorporated and still allow effective initiation? To address this question, the single nucleotide linkage at position −1 was replaced by two non-nucleosidic spacers, a construct labeled (spacer)$_2$ in Figure 2. This construct formally allows more flexibility, but also introduces steric bulk and added charge density (an extra phosphate group). The results summarized in Table 1 show that in this case, the apparent activity still does not change significantly relative to the wild-type promoter. To achieve a similar linkage, but without an added phosphate group, a construct was prepared containing a glycol spacer of approximately the same length, labeled (2X spacer) in Figure 2. The results again show a relative activity the same as that of the (spacer)$_2$ construct.

**Table 1.** Relative activities of double stranded constructs containing non-nucleoside spacers at position −1

Substitution	Relative activity
Native	1.0
Spacer	1.7
(spacer)$_2$	1.2
2X spacer	1.1
(2X spacer)$_2$	1.1
(2X spacer)$_3$	0.5

Relative amount of runoff synthesis in a ten minute incubation at 37°C, containing 0.08 μM T7 RNA polymerase and 0.20 μM promoter encoding the transcript GGGAA, and using [α-^{32}P]GTP as a label, as described in Materials and Methods.

With DNA bound at the active site, it appears that the enzyme can accommodate substantial added bulk and charge at position −1. In addition, the similar activities of the promoters containing (2X spacer) or (spacer)$_2$, which differ in the presence of a large negatively charged phosphate group in the latter, suggest that the DNA backbone in this region is not substantially constrained in the enzyme-DNA complex. To further test these proposals, constructs (2X spacer)$_2$ and (2X spacer)$_3$, with still longer linkages, were prepared and tested. The data presented in Table 1 indicate that for both constructs, the relative activity is comparable to that of the smaller (2X spacer) construct, with a small reduction for the largest construct (2X spacer)$_3$. Steady-state kinetic assays on the promoter containing (2X spacer)$_3$ at position −1 yield values of $K_m = 31$ (21 to 46) nM and $k_{cat} = 13.6$ (12.4 to 14.8) minute^{-1}, suggesting at most a five-fold decrease in binding affinity and a twofold decrease in initiation, relative to wild-type. In previous studies, the removal of a single methyl group, or hydrogen bond donor or acceptor, in the upstream region was sometimes found to result in much larger perturbations to the kinetics (Li *et al.*, 1996; Maslak *et al.*, 1993; Schick & Martin, 1995). Preliminary fluorescence measurements of binding confirm that K_d is not increased substantially in the (2X spacer)$_3$ construct (unpublished results). It is clear that substantial added charge and steric bulk are reasonably accommodated very near the active site, with only mild perturbations in promoter binding and in the apparent steady-state kinetics.

Fidelity of initiation site selection

In the above experiments, the relative activity presented represents the total amount of labeled ATP incorporated into RNA product in a defined period of time (for the template used, encoding the transcript GGGAA, only the four and five base transcripts will be detected). Although the results clearly show that transcription continues at a reasonable rate in the presence of substantial spacer substitutions, they do not indicate whether the full-length (five base) transcript, corresponding to initiation at position +1, is the major product. In order to assess the fidelity of initiation, RNA pro-
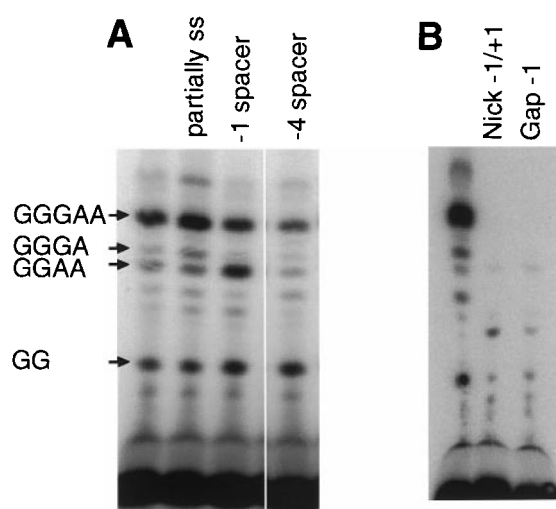
Figure 3. Representative product profiles from transcription reactions, analyzed by denaturing gel electrophoresis (see Materials and Methods). In all cases, only the nucleoside triphosphates GTP and ATP were present in solution. (a) Representative reactions on a template encoding the five base runoff transcript GGGAA, and incorporating [γ-^{32}P]GTP as the label, so that each transcript is labeled only once. Partially ss denotes a promoter construct that is single-stranded from positions +1 through +5; all other constructs are fully duplex. Lanes denoted spacer substitute the simple spacer for the template strand base at the indicated position in the promoter. (b) Reactions on a longer template, encoding a 12 base runoff transcript (although transcription halts at the same five base transcript due to a lack of UTP in the solution). In these reactions, [α-^{32}P]GTP was used as the label in order to emphasize longer transcripts.

ducts from a similar transcription reaction were characterized by denaturing gel electrophoresis, examples of which are shown in Figure 3. To provide a simple analysis of transcript sequence, parallel reactions were carried out in the presence of either [γ-^{32}P]GTP, [α-^{32}P]GTP or [α-^{32}P]ATP (the template encodes only G and A). The nucleotide [γ-^{32}P]GTP results in the labeling of each transcript only once (at the 5′ end), while [α-^{32}P]GTP and [α-^{32}P]ATP label each transcript proportional to the number of G or A residues, respectively. For each band on the gel, the approximate length derived from its gel mobility (note that for short RNA transcripts there will be some sequence-dependence of mobility), combined with the measured composition of G and A, provide a very good assignment of transcript identity. Examples of the compositional information used in assignment of transcripts are shown in Table 2. Note that misinitiation at position −1, followed by premature termination at position +4 would yield a transcript with the sequence AGGGA, with the same length and formal composition as the correct transcript GGGAA. However, in analyses of the type shown in Table 2, misinitiation at position −1 would not incorporate any [γ-^{32}P]GTP, leading to artificially

Table 2. Assignment of transcription products

	Native promoter		Partially ss/spacer at −1	
RNA seq	$\dfrac{[\alpha\text{-}^{32}\text{P}]\text{GTP}}{[\gamma\text{-}^{32}\text{P}]\text{GTP}}$	$\dfrac{[\alpha\text{-}^{32}\text{P}]\text{ATP}}{[\gamma\text{-}^{32}\text{P}]\text{GTP}}$	$\dfrac{[\alpha\text{-}^{32}\text{P}]\text{GTP}}{[\gamma\text{-}^{32}\text{P}]\text{GTP}}$	$\dfrac{[\alpha\text{-}^{32}\text{P}]\text{ATP}}{[\gamma\text{-}^{32}\text{P}]\text{GTP}}$
GGGAA	2.7	2.1	2.7	2.2
GGGA	2.9	1.2	3.2	1.1
GGAA	1.9	2.0	1.9	2.3
GGG	2.9		3.4	
GGA	2.1	1.3	2.1	1.3
GG	2.2		2.2	

Ratios represent the relative incorporation of each nucleotide, in a ten minute reaction, under the conditions (0.2 μM each of enzyme and promoter DNA) described in Materials and Methods.

Promoter constructs were either: fully duplex or duplex from position −17 to −1, but single-stranded (template strand only) from positions +1 to +5.

high values for the ratios presented. This is not observed, consistent with the conclusion that the enzyme does not misinitiate at position −1 to any significant extent.

Analysis of transcripts produced from the native promoter, containing the normal thymidine at position −1 of the template strand, summarized in Table 3, shows that the 5mer GGGAA is the predominant product. In addition, very small amounts of the transcript, GGAA, corresponding to misinitiation at position +2, and some abortive transcript of sequence GG (which could arise from initiation at either position +1 or +2) are observed. Substitution of the simple spacer at position −1 leads to a fivefold increase (13 *versus* 2.6 μM) in the amount of the incorrect product GGAA, with only a slight (14 *versus* 17 μM) decrease in the amount of the correct five base transcript. Thus the increase in apparent activity associated with the substitution of the simple spacer arises almost completely from a fivefold increase in incorrect initiation at position +2. In contrast, the twofold increase in activity associated with the removal of the non-template strand in this region, denoted (native)/ss in Table 3, does not arise from incorrect initiation; fidelity is not decreased in the partially single-stranded construct containing the consensus thymidine at position −1.

For longer spacer substitutions, as for the simple spacer substitution, the amount of misinitiation at position +2 remains comparable to the amount of

correct initiation at position +1. This is true even for the very dramatic substitution of the (2X spacer)$_3$ at position −1. This construct, however, may lead to a small increase in the synthesis of the product (GGGA) that is correctly initiated, but which is truncated prematurely at the 3′ end. The substantial added steric bulk and/or charge may facilitate abortive termination prior to successful runoff transcription.

The above results suggest a unique role for the nucleotide at position −1 in determining the fidelity of start site selection. The replacement of the thymidine nucleotide at position −1 by a simple spacer alters potential constraints imposed by the backbone, and removes potential interactions with the thymine base. To elucidate the nature of the interactions at this position, we incorporated an abasic site at position −1. The results summarized in Table 4 indicate that this construct has the same increased misinitiation at position +2 as does the spacer substitution, suggesting that base stacking interactions may be important in determining the positioning of the start site.

The result that partially single-stranded constructs (which lack the non-template strand from position +1 to +5) initiate transcription well with either the consensus thymidine or a spacer at position −1 clearly demonstrates that the non-template strand is not necessary to direct the transcribed template strand into the active site. To confirm that the direct connection of the template strand to up-

Table 3. Product distributions for constructs containing non-nucleoside spacers at position −1

Substitution	[RNA product] (μM)[a]						
	GGGAA	GGGA	**GGAA**	GGG	GGA	**GG**	**Total**
(Native)	**17**	0.9	**2.6**	0.5	0.4	**6.7**	**28**
Spacer	**14**	<0.5	**13**	0.3	1.0	**14**	**42**
(Native)/ss[b]	**38**	3.2	**4.3**	1.0	0.9	**9.4**	**57**
Spacer/ss[b]	**23**	1.3	**15**	0.8	2.2	**16**	**58**
(Spacer)$_2$	**5.8**	0.9	**11**	0.2	1.6	**17**	**36**
2X spacer	**3.7**	0.5	**4.0**	0.4	0.7	**10**	**36**
(2X spacer)$_3$	**3.3**	2.2	**1.1**	0.3	0.7	**7.3**	**15**
abasic site	**11**	3.2	**8.6**	2.4	1.1	**6.3**	**33**

[a] Final concentration of each RNA in a ten minute reaction, as described for Table 2.
[b] Partially single stranded construct, as described in Table 2.

Table 4. Product distribution for a construct containing an abasic site at position −1

| Substitution | [RNA product] (μM)[a] | | | | | | |
	GGGAA	GGGA	**GGAA**	GGG	GGA	**GG**	**Total**
(Native)	**17**	0.9	**2.6**	0.5	0.4	**6.7**	**28**
Abasic site	**11**	3.2	**8.6**	2.4	1.1	**6.3**	**33**

[a] Final concentration of each RNA in a ten minute reaction, as described for Table 2.

stream promoter elements is required for initiation, we prepared two constructs that completely break the connection between position +1 and the upstream template DNA. In one construct, three separate pieces of DNA were annealed so as to present a nick between positions −1 and +1 of the template strand (flanked by 3′ and 5′ hydroxyl groups). In the second construct, a gap was introduced at position −1 (deleting the template strand thymidine base completely). In both cases, in order that a stable duplex downstream would be maintained, the downstream sequence was extended to position +12, as shown below.

```
 -17        -10      -5       +1    +5          +12
  T A A T A C G A C T C A C T A T A G G G A A T T C G A G C
  A T T A T G C T G A G T G A T A T C C C T T A A G C T C G
                                 ⌐→
```

The sequence of this construct is such that in the presence of only GTP and ATP, the same full-length RNA sequence (GGGAA) is expected. As shown in Figure 3, while the fully connected construct produces the expected RNA products in good yield, no detectable transcription is observed from either the nicked or the gapped construct. A covalent tethering of the downstream template strand to the upstream promoter element is required.

Spacer substitutions at other sites

The above results demonstrate that substitutions at position −1, immediately adjacent to the start site for initiation, do influence positioning of the templating bases. In order to determine the extent to which interactions more distant (upstream) from the start site influence positioning, substitutions were incorporated along the template strand at positions −2, −3 and −4 within the TATA sequence. As seen in Table 5, the introduction of a simple spacer at position −2 yields a threefold increase in apparent activity, similar to the effect of the substitution at position −1, while the same substitution at position −4 yields near-native activity. In addition, the results presented in Table 6 show clearly that the fidelity of initiation from the promoter containing a spacer at position −2 is not as good as that of wild-type, but is better than that of the construct containing a spacer at position −1. However, placement of a simple spacer further upstream at position −4 or a double length spacer at position −3 yields native levels of correctly and misinitiated products, indicating that template

DNA downstream of position −3 is sufficient to accurately position the templating bases.

Substitution of a simple spacer for the templating base at position +1, now within the normally transcribed region, must significantly disrupt the initiation of transcription, forcing the enzyme to initiate at position +2. The analysis of apparent activity presented in Table 5 confirms that initiation is substantially inhibited, while analysis of product distributions, shown in Table 6, demonstrates that the complex can initiate transcription at position +2, at a rate that is an order of magnitude reduced from the initiation at +1 by the native complex (approximately one-fourth the rate of misinitiation at +2 by the native complex).

Discussion

The promoter sequence in DNA directs an RNA polymerase to a specific base for the (primer-independent) initiation of transcription. In normal initiation of transcription, the RNA polymerase must minimally bind to the specific sequence of DNA and facilitate melting of the DNA duplex, providing some site selectivity. In fact, previous studies have shown that both T7 and *E. coli* RNA polymerases can bind to and elongate from pre-formed DNA-RNA transcription bubbles, in a primer-dependent process (Daube *et al.*, 1994; Daube & von Hippel, 1992, 1994). More recent studies have shown that T7 RNA polymerase can transcribe (elongate) through a large gap in the template strand, requiring the direction of a disconnected template strand into the active site, presumably without regard to sequence (Liu & Doetsch, 1996; Zhou *et al.*, 1995). Elongation then continues using the newly positioned template strand. Finally, it has been shown that *E. coli* RNA polymerase can initiate (primer-independent) RNA synthesis from

Table 5. Relative activities of constructs containing non-nucleoside spacers at other positions

Position	Substitution	Relative activity[a]
	native	1.0
−2	Spacer	2.9
−2	2X spacer	0.5
−3	Spacer	2.7
−3	2X spacer	0.7
−4	Spacer	1.2
−4[b]	Spacer	2.2
+1	Spacer	<0.1

[a] Relative activity as described for Table 1.
[b] Partially single-stranded as described for Table 2.

Table 6. Product distributions for constructs containing non-nucleoside spacers at other positions

Substitution	Pos	[RNA product] (μM)[a]						
		GGGAA	GGGA	**GGAA**	GGG	GGA	**GG**	**Total**
(Native)		**17**	0.9	**2.6**	0.5	0.4	**6.7**	**28**
Spacer	−2	**45**	4.9	**11**	1.3	1.9	**23**	**87**
2X spacer	−2	**4.9**	0.8	**2.4**	0.6	2.5	**6.4**	**18**
2X spacer	−3	**8.2**	0.9	**1.3**	0.5	1.3	**19**	**31**
Spacer	−4	**6.4**	0.3	**1.0**	0.8	0.3	**13**	**21**
Spacer/ss	−4	**24**	2.3	**2.7**	1.9	1.0	**20**	**52**
Spacer	+1			**0.7**		0.3[b]	**2.3**	**4**

[a] Final concentration of each RNA in a ten minute reaction, as described for Table 2.
[b] For the construct containing a spacer at position +1, a product assigned as GAA (0.5 μM) was also observed.

artificially constructed transcription bubbles that contain no promoter sequence (Aiyar *et al.*, 1994; Tripatara & deHaseth, 1993).

In the T7 RNA polymerase system, previous studies have shown that promoter recognition is achieved, at least in part, by specific (duplex) major groove contacts 5 to 11 bases upstream of the start site (Li *et al.*, 1996; Maslak *et al.*, 1993; Schick & Martin, 1995) and that melting represents only a small barrier to the steady-state rate of initiation (Maslak & Martin, 1994). Upstream binding interactions would seem sufficient to direct the transcription start site at position +1 to the vicinity of the enzyme active site, but not to precisely direct initiation to a unique site on the DNA template (i.e. placement of the two templating bases into the active site of the enzyme). The latter should require specific interactions with functional groups on the template strand, restricted conformations of the template strand, and/or a precise spacing of the DNA from the upstream recognition elements.

A direct linkage to upstream elements is required for initiation

Previous studies in which a gap was introduced into the template DNA some distance downstream of the promoter indicate that during the processive elongation phase of transcription, T7 RNA polymerase can bypass a gap in the DNA, bringing a disconnected template strand into the active site (Liu & Doetsch, 1996; Zhou *et al.*, 1995). In this case, it was proposed that the non-template strand serves as an indirect linkage between the two pieces of template DNA. This mechanism for template strand positioning clearly does not apply to the initiation of transcription from spacer-containing constructs, since the absence of the non-template strand in this region has little effect on initiation. Furthermore, the introduction of a nick or gap in the template strand, as used in the previous studies of processive elongation, completely abolishes the primer-independent initiation of transcription.

Positioning of the template strand is not rate-limiting

The incorporation of a simple linker at position −1 results in a twofold increase in the apparent

activity, consistent with a previous study which showed that removal of the barrier to melting in this region (in this case, the loss of an AT base-pair) leads to a twofold increase in the steady-state parameter k_{cat} (Maslak & Martin, 1993). The incorporation of a variety of more perturbing linkers at positions −1 through −4, including a linker, $(2X spacer)_3$, which is potentially six times the linkage length of the single nucleotide it replaces, results in very minor changes in the apparent activity in this system. Although statistical mechanical arguments predict that the average separation between the adjoining nucleotides will be smaller than the maximum extended length (Williams & Hall, 1996), it is clear that these substitutions provide very different (and fewer) restrictions on the path of the DNA backbone connecting upstream elements with the start site bases. Moreover, the larger substitutions require the accommodation of significantly more steric bulk and negative charge, suggesting that the backbone linkage at position −1 of the template strand enjoys ready access to the aqueous phase, while the immediately adjacent templating bases (+1 and +2) are positioned correctly at the active site.

These results strongly support the assignment of a step beyond template positioning as rate-limiting. This is consistent with recent results, which indicate that promoter binding and helix melting occur on a much faster time-scale than does initiation (Jia *et al.*, 1996; Újvári & Martin, 1996).

Base stacking interactions at position −1 are uniquely required for maximal fidelity

Although the overall rate of initiation is not decreased substantially by the above substitutions, simple assays of activity do not provide information on the precise fidelity of start site selection. It is well known that promoter-dependent initiation by T7 RNA polymerase begins with good fidelity at the +1 site in the promoter, and the results presented here for the native promoter demonstrate this. Ignoring the dinucleotide product (which could, in principle, initiate from either position +1 or position +2), at most 14% of transcripts misinitiate at position +2 (this is an upper limit, since some of the GGAA or GGA product could arise from correct initiation, followed by forward

slippage of the dinucleotide), No detectable product corresponding to misinitiation at position −1 is observed. This fidelity occurs despite the fact that initiation at position −1 would incorporate ATP as the first base in the transcript, and some T7 phage promoters do begin with adenine as the first coded base (Dunn & Studier, 1983). Finally, it is clear from the current studies that the non-template strand in this region is not required for native fidelity. What is the physical basis for this specificity?

The incorporation at position −1 of a flexible linker that mimics the backbone connectivity of a normal nucleoside results in normal initiation at position +1, but leads to a fivefold increase in the amount of products misinitiated at position +2 (in this case, misinitiation one base upstream is not expected, since there is no base at position −1). This suggests that torsional constraints imposed by the sugar and/or stacking interactions between the template strand bases at positions −1 and +1 are important in directing position +1 into the active site of the enzyme. That the incorporation of an abasic site at position −1 yields similar results points more strongly to stacking interactions, presumably between the bases at positions −1 and +1, as the essential determinants (although the sugar in an abasic site may adopt alternative conformations, the backbone in this construct is certainly more constrained than in constructs containing spacers at position −1). These observations are qualitatively the same on a construct lacking the non-template strand in this region, consistent with the proposal that the non-template strand is not involved in the precise placement of the template strand into the active site. As the length and charge of the spacer at position −1 are increased, the relative amounts of the correctly initiated and misinitiated transcripts remain comparable.

The nature of the upstream linkage is not critical

The above results suggest that the precise nature of the connectivity between the start site and upstream elements is important for ensuring the fidelity of initiation. However, spacer substitutions at positions −2 through −4 result in ratios of correctly initiated to misinitiated transcripts that are more like that of the native system. Although incorporation of the ''2X linker'' at position −2 leads to only a twofold preference for correct to incorrect initiation, the other substitutions all lead to ratios greater than four, comparable to the ratio observed for the native promoter. Finally, these results are consistent with the proposal that stacking between the bases at positions −1 and +1 plays a role in start site positioning.

The observation that significant disruptions in the template strand at positions −3 and −4 are tolerated with no measurable effect on the initiation of transcription casts doubt on the simplest model for initiation in which upstream specificity elements (duplex major groove contacts) direct downstream non-specific interactions *via* a simple spacing of the phosphate backbone. In fact, interactions very close to the start site are involved in the precise positioning of the first two templating bases into the enzyme active site. The critical nature of the template strand in this region is underscored by the observation that the incorporation of a simple spacer at position +1 leads to a significant decrease in initiation at position +2, as measured by the synthesis of both the expected runoff transcript and the aborted dinucleotide.

Materials and Methods

RNA polymerase

T7 RNA polymerase was prepared from *E. coli* strain BL21 carrying the overproducing plasmid pAR1219 (kindly supplied by F. W. Studier), in which RNA polymerase is expressed under inducible control of the *lac* UV5 promoter (Davanloo *et al.*, 1984; Morris *et al.*, 1986). The enzyme was purified and the concentration determined ($\varepsilon_{280} = 1.4 \times 10^5$ M^{-1} cm^{-1}) as described previously (King *et al.*, 1986). Purity of the enzyme (>95%) was verified by denaturing polyacrylamide gel electrophoresis.

Oligonucleotides

Oligonucleotides were synthesized by the phosphoramidite method on a Milligen/Biosearch Cyclone Plus DNA synthesizer. The standard phosphoramidites (CPG and Prime Synthesis) were coupled normally. The spacer analogs (Prime Synthesis) 3-*O*-dimethoxytrityl-propyl-1-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite (spacer) and 9-*O*-dimethoxytrityl-triethyleneglycol,1-[(2-cyanoethyl)-(*N*,*N*-diisopropyl)]-phosphoramidite (2X spacer) were coupled off-machine as described (Schick & Martin, 1993). 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 Amberchrom CG-161cd reverse phase resin as described (Schick & Martin, 1993). Purity of the oligonucleotides was confirmed by denaturing (urea) gel electrophoresis of 5′ end-labeled single strands.

An abasic site was incorporated into DNA by first synthesizing the appropriate single-stranded construct containing deoxyuracil at the target site, followed by enzymatic removal of the base using the enzyme uracil-*N*-glycosylase. Briefly, single-stranded oligonucleotides containing deoxyuracil were incubated for one hour at 37°C in 20 mM Tris-HCl (pH 8.0), 50 mM KCl, 2.5 mM MgCl$_2$, 0.1 mg/ml bovine serum albumin, and one unit of uracil-*N*-glycosylase (Life Technologies). Cleavage of the glycosidic bond was verified by heating a small amount of 5′-labeled sample for one hour at 95°C (resulting in strand cleavage under these conditions) followed by denaturing gel electrophoresis (Schick & Martin, 1993).

Kinetics assays

Steady-state assays of transcription initiation (Martin & Coleman, 1987) were carried out in a total volume of 20 μl. Reaction conditions as originally reported (Maslak *et al.*, 1993) and revised (Újvári & Martin, 1996) were

30 mM Hepes (pH 7.8), 15 mM magnesium acetate, 25 mM potassium glutamate, 0.25 mM EDTA, 0.05% (v/v) TWEEN-20 (Calbiochem, protein grade), 0.8 mM GTP and 0.4 mM ATP, using [α-^{32}P]ATP as a label. It is important to note that in these assays, full-length product is generally assumed, since the velocities used in the determination of k_{cat} and K_m derive from total radioactivity incorporated divided by the predicted amount of label per transcript (two adenosine bases per GGGAA transcript). This assumption is good for the native promoter, but is less so for some of the spacer-containing promoters. In the current study, the transcripts GGGAA and GGAA represent the dominant products, allowing the assumption above. The dinucleotide GG also appears in electrophoretic product analyses (see below), but does not separate well in the paper chromatography used in the steady-state assays.

For the steady-state kinetic analyses of each template, reaction velocities were measured at different enzyme (0.04, 0.08 μM) and DNA (0.02 to 0.30 μM) concentrations. The error in each velocity was approximated as the higher of 0.1 μM/minute or the *t*-distribution 80% confidence interval of the fitted slope for the three time-points. Velocity data were then fit as described (Martin & Coleman, 1987) to the exact solution of the steady-state equation, using a weighted non-linear 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 non-linear 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), and we have recently shown (Újvári & Martin, 1996) that K_m may not accurately reflect increases in binding affinity (decreases in K_d). Similarly, small changes in k_{cat} (less than a factor of 2) should be interpreted with caution.

Analysis of transcription products

Relative activities of the various promoter constructs were determined under the conditions described above, using [α-^{32}P]GTP as a label. Reactions were incubated for ten minutes at 37°C with 0.08 μM enzyme and 0.2 μM duplex DNA, and were quenched by spotting onto chromatography paper pre-spotted with 200 mM EDTA. Activities were proportional to the total amount of radioactivity that did not migrate significantly from the origin in ascending paper chromatography in 60% saturated ammonium sulfate (Mulligan *et al.*, 1985). For constructs producing more than one major transcript, relative activity is not necessarily proportional to the frequency of initiation, since different transcripts will contain different amounts of the radioactive label. Also note that the chromatographic separation does not typically isolate products shorter than three or four bases.

To determine the fidelity of initiation, reactions were carried out as for the determination of relative activities, except that reactions contained 0.2 μM enzyme and 0.2 μM duplex DNA and ten minute time-points were quenched into a 95% formamide mixture and loaded onto 7 M urea/18% polyacrylamide, sequencing gels. The upper reservoir contained 0.5× TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA), while the lower reservoir contained 1.0 M sodium acetate and 0.67× TBE buffer (60 mM Tris, 60 mM boric acid, 1.3 mM EDTA). Gels were electrophoresed at 2000 V (50 W) until the bromophenol blue dye had migrated approximately 70% the length of the gel (approximately three hours). Following electrophoresis, gels were dried and quantified using a Molecular Dynamics Storm840 Phosphorimager. This allowed quantification of individual transcription products for comparison between the constructs. Typically, parallel experiments were run containing [α-^{32}P]ATP, [α-^{32}P]GTP or [γ-^{32}P]GTP. The ratios of the extent of incorporation of [α-^{32}P]ATP or [α-^{32}P]GTP to that of [γ-^{32}P]GTP provide unambiguous identification of almost all RNA products (note that the templates encode RNA products containing only G or A, and that CTP and UTP are left out of the reactions).

References

Aiyar, S. E., Helmann, J. D. & deHaseth, P. L. (1994). A mismatch bubble in double-stranded DNA suffices to direct precise transcription initiation by *Escherichia coli* RNA polymerase. *J. Biol. Chem.* **269(18)**, 13179–84.

Altmann, S., Labhardt, A. M., Bur, D., Lehmann, C., Bannwarth, W., Billeter, M., Wüthrich, K. & Leupin, W. (1995). NMR studies of DNA duplexes singly cross-linked by different synthetic linkers. *Nucl. Acids Res*, **23(23)**, 4827–4835.

Amaratunga, M. & Lohman, T. M. (1993). *Escherichia coli* rep helicase unwinds DNA by an active mechanism. *Biochemistry*, **32**, 6815–6820.

Breslauer, K. J., Frank, R., Blocker, H. & Marky, L. A. (1986). Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA,* **83(11)**, 3746–3750.

Chapman, K. A. & Burgess, R. R. (1987). Construction of bacteriophage T7 late promoters with point mutations and characterization by *in vitro* transcription properties. *Nucl. Acids Res*, **15(13)**, 5413–32.

Daube, S. S. & von Hippel, P. H. (1992). Functional transcription elongation complexes from synthetic RNA-DNA bubble duplexes. *Science*, **258(5086)**, 1320–1324.

Daube, S. S. & von Hippel, P. H. (1994). RNA displacement pathways during transcription from synthetic RNA-DNA bubble duplexes. *Biochemistry*, **33(1)**, 340–347.

Daube, S. S., Hart, C. R. & von Hippel, P. H. (1994). Coupling of RNA displacement and intrinsic termination in transcription from synthetic RNA.DNA bubble duplex constructs. *Proc. Natl Acad. Sci. USA,* **91(20)**, 9539–9543.

Davanloo, P., Rosenberg, A. H., Dunn, J. J. & Studier, F. W. (1984). Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proc. Natl Acad. Sci. USA*, **81(7)**, 2035–9.

Dunn, J. J. & Studier, F. W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* **166(4)**, 477–535.

Durand, M., Chevrie, K., Chassignol, M., Thuong, N. T. & Maurizot, J. C. (1990). Circular dichroism studies of an oligodeoxyribonucleotide containing a hairpin loop made of a hexaethylene glycol chain: confor-

mation and stability. *Nucl. Acids Res,* **18(21)**, 6353–6359.

Jia, Y., Kumar, A. & Patel, S. (1996). Equilibrium and stopped-flow kinetic studies of interaction between T7 RNA polymerase and its promoters measured by protein and 2-aminopurine fluorescence changes. *J. Biol. Chem.* **271(48)**, 30451–30458.

Johnson, M. L. (1983). Evaluation and propagation of confidence intervals in nonlinear, asymmetrical variance spaces. Analysis of ligand-binding data. *Biophys. J.* **44(1)**, 101–106.

Johnson, M. L., Correia, J. J., Yphantis, D. A. & Halvorson, H. R. (1981). Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys. J.* **36(3)**, 575–588.

King, G. C., Martin, C. T., Pham, T. T. & Coleman, J. E. (1986). Transcription by T7 RNA polymerase is not zinc-dependent and is abolished on amidomethylation of cysteine-347. *Biochemistry,* **25(1)**, 36–40.

Li, T., Ho, H., Maslak, M., Schick, C. & Martin, C. (1996). Major groove recognition elements in the middle of the T7 RNA polymerase promoter. *Biochemistry,* **35(12)**, 3722–3727.

Liu, J. & Doetsch, P. (1996). Template strand gap bypass is a general property of prokaryotic RNA polymerases: implications for elongation mechanisms. *Biochemistry,* **35(47)**, 14999–15008.

Martin, C. T. & Coleman, J. E. (1987). Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry,* **26(10)**, 2690–2696.

Maslak, M. & Martin, C. T. (1993). Kinetic analysis of T7 RNA polymerase transcription initiation from promoters containing single-stranded regions. *Biochemistry,* **32(16)**, 4281–4285.

Maslak, M. & Martin, C. T. (1994). Effects of solution conditions on the steady-state kinetics of initiation of transcription by T7 RNA polymerase. *Biochemistry,* **33(22)**, 6918–6924.

Maslak, M., Jaworski, M. D. & Martin, C. T. (1993). Tests of a model for promoter recognition by T7

RNA polymerase: thymine methyl group contacts. *Biochemistry,* **32(16)**, 4270–4274.

Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucl. Acids Res,* **15(21)**, 8783–8798.

Morris, C. E., Klement, J. F. & McAllister, W. T. (1986). Cloning and expression of the bacteriophage T3 RNA polymerase gene. *Gene,* **41(2-3)**, 193–200.

Mulligan, M. E., Brosius, J. & McClure, W. R. (1985). Characterization *in vitro* of the effect of spacer length on the activity of *Escherichia coli* RNA polymerase at the TAC promoter. *J. Biol. Chem.* **260(6)**, 3529–3538.

Pasman, Z. & Garcia-Blanco, M. A. (1996). The 5′ and 3′ splice sites come together *via* a three dimensional diffusion mechanism. *Nucl. Acids Res.* **24(9)**, 1638–1645.

Schick, C. & Martin, C. T. (1993). Identification of specific contacts in T3 RNA polymerase-promoter interactions: kinetic analysis using small synthetic promoters. *Biochemistry,* **32(16)**, 4275–4280.

Schick, C. & Martin, C. T. (1995). Tests of a model of specific contacts in T7 RNA polymerase-promoter interactions. *Biochemistry,* **34(2)**, 666–672.

Tripatara, A. & deHaseth, P. L. (1993). A new start site for *Escherichia coli* RNA polymerase at an engineered short region of non-complementarity in double-stranded DNA. *J. Mol. Biol.* **233(3)**, 349–358.

Újvári, A. & Martin, C. T. (1996). Thermodynamic and kinetic measurements of promoter binding by T7 RNA polymerase. *Biochemistry,* **35(46)**, 14574–14582.

Williams, D. J. & Hall, K. B. (1996). RNA hairpins with non-nucleotide spacers bind efficiently to the human U1A protein. *J. Mol. Biol.* **257(2)**, 265–275.

Zhou, W., Reines, D. & Doetsch, P. W. (1995). T7 RNA polymerase bypass of large gaps on the template strand reveals a critical role of the nontemplate strand in elongation. *Cell,* **82(4)**, 577–585.