Structure and Function in Promoter Escape by T7 RNA Polymerase

CRAIG T. MARTIN,*†
EDWARD A. ESPOSITO,*
KARSTEN THEIS,† and
PENG GONG*†

*Department of Chemistry,
University of Massachusetts, Amherst,
Massachusetts 01003
†Department of Biochemistry and
Molecular Biology, University of
Massachusetts, Amherst,
Massachusetts 01003

I. Introduction ................................................................. 324
II. Mapping the Transcription Bubbles at Initiation and
    During Elongation ..................................................... 325
III. Initial Collapse of the Bubble Occurs on Translocation Beyond
    Position +8 .................................................................. 326
IV. Promoter Release Occurs on Translocation Beyond
    Position +8 .................................................................. 329
V. Initial RNA Displacement Occurs on Translocation
    Beyond Position +8 .................................................... 331
VI. Bubble Collapse Contributes to Initial RNA Displacement .......... 332
VII. Retention of Promoter Contacts Interferes with Initial Bubble Collapse.. 337
VIII. Coupling of Promoter Release, Bubble Collapse,
    and RNA Displacement .............................................. 340
IX. A Structural Model for the Transition from Initiation
    to Elongation .................................................................. 342
X. Summary .......................................................................... 343
References ........................................................................... 344

RNA polymerases undergo a complex series of changes in transitioning from a statically and sequence-specifically bound initial complex to a sequence-independent elongation complex clear of the promoter. The single subunit enzyme T7 RNA polymerase has been shown to undergo a large conformational change in its protein structure accompanying a large change in the nature of its interaction with the DNA template and the nascent RNA. As short RNA transcripts are synthesized, the initially melted DNA bubble remains fully open, extending in size as the active site translocates downstream along the DNA, and the initial promoter contacts are retained. Both fluorescence and biochemical assays demonstrate that the upstream edge of the initial
bubble begins to collapse as the active site translocates beyond position +8. At this point, the nascent RNA is long enough to possess inherent stability in its heteroduplex and, indeed, the 5’ end must be coaxed away from the heteroduplex. Collapse of the upstream edge of the bubble provides this coaxing via competitive displacement. Collapse of the bubble is, in turn, allowed by release of the upstream tight binding contacts. Finally, a structural model is proposed, in which growth of the hybrid pushes the N-terminal domain away from the C-terminal domain, but the promoter binding interface is maintained through growth of an 8 base RNA. Translocation past this point pulls the specificity loop away from the N-terminal domain, destroying the promoter binding interface and triggering release of the DNA. Release of the promoter contacts then triggers the series of events previously described, but also allows a 220° rotation of the N-terminal rigid domain, leading to creation of an RNA exit channel in the elongation complex. Thus, initial RNA displacement occurs as the exit channel is created.

I. Introduction

DNA-dependent RNA polymerases face a number of design challenges in carrying out the complex series of events that comprise RNA synthesis. They must initiate with sequence specificity, but they must elongate with no sequence preference. They must not only be able to extend a growing chain (as do the distributive DNA-dependent DNA polymerases), but they must also be able both to initiate dinucleotide synthesis de novo and to extend the transcript from that dinucleotide. During elongation, the RNA strand being extended has the potential to be stabilized by extensive contacts with both the protein and with the template strand of the DNA. However, early in synthesis, the RNA is short and so is expected to make many fewer stabilizing interactions. It is not surprising then, that complexes very early in synthesis (containing RNA products of 2–8 bases) should be less stable than complexes farther out in elongation. It should also not be surprising that nature has design criteria for initiation that are distinct from those for elongation. How is the transition between these two states, broadly described as promoter escape, achieved within the same enzyme complex?

Since the mid-1990s, much has been learned about the transition from an unstable initiation complex to a stable elongation complex in the single subunit T7 RNA polymerase. Both the promoter DNA and the protein can be divided, at least conceptually, into distinct regions, with specific functions. The promoter DNA from positions −17 to −5 appears to be largely involved in the initial tight and specific binding of the upstream DNA to the protein (1–8). In contrast, a primary role of the promoter DNA downstream of position −5 is to melt open and allow positioning of the template strand bases into the
polymerase active site (9–12). The DNA at the transcription start site at positions +1 and +2 (at least) appears to be optimized for the efficient de novo synthesis of the first phosphodiester bond (9–11, 13).

It has been shown that promoter DNA remains bound to the protein during the initial phase of transcription, at least until the polymerase reaches position +6 to +8 (14–18). Retention of these contacts presumably helps to stabilize an otherwise unstable initial complex. In keeping with the need for the complex to polymerize thousands of bases in a sequence-independent manner, early studies showed that these specific promoter contacts appear to be released by the time the polymerase has translocated to position +15 (16). Very little was known about the series of events and structural changes that occur as the polymerase releases its promoter contacts and transitions from an unstable initiation complex to a stable elongation complex. This chapter describes some advancements toward our understanding of this transition.

Valuable insight into the nature of this transition was gained with the determination of crystallographic models describing the structure of an elongating polymerase (19, 20). The structures reveal a large change in the protein accompanying the transition from initiation to elongation. Interestingly, the data reveal that T7 RNA polymerase, although lacking the size and complexity of the multi-subunit polymerases, possesses many of the structural features seen in the more complex systems. The structural change results in a complete disruption of the promoter binding site on the protein, achieving perhaps the same end as release of sigma in the bacterial RNA polymerase. As a result of this disruption, the elongation complex is expected to show little promoter-specific binding. The change in polymerase conformation also generates a putative RNA exit channel, presumably to provide stability to an elongating complex through direct interactions with the single-stranded RNA peeling away from the heteroduplex. Each of these features is expected to contribute to the unique properties of an elongation complex.

While the structural data provide wonderful insight into the initiating and elongating forms of the transcription complex, the data shed little light on how this transition occurs. What interactions drive the structural changes, how are the changes in the structure of the individual components related, and how do the structural changes relate to function? The spectroscopic and biochemical studies that will be presented provide answers to some of these questions.

II. Mapping the Transcription Bubbles at Initiation and During Elongation

Since the mid-1990s, various researchers have used the fluorescent base analog 2-aminopurine as a probe of the local melted state of DNA (21–25). For the purpose of mapping transcription bubbles, it is sufficient to know that
fluorescence of the analog is quenched by interactions with neighboring bases. Stacking interactions are present in single-stranded DNA, but become much stronger as the DNA is annealed to form a duplex. Using this latter property in reverse, for probes placed in a fully duplex environment, local melting of the DNA leads to an increase in the fluorescence of the probe. We have exploited the fluorescence of 2-aminopurine and related base analogs to map the extent of both initially melted transcription bubbles and transcription bubbles in stalled elongation complexes, clear of the promoter (26, 27). The probe is, of course, also useful in assessing dynamics of the transcription process, a topic addressed here only in statically stalled complexes (26, 28–31).

The results summarized in Fig. 1 show that in an initiation complex, bases from positions −4 through at least position −1 are melted open (unpublished results show that the bubble actually extends to position +3 or +4). Thus, the transcription bubble is initially about 7 to 8 bases in length, and in the presence of GTP as the sole substrate allowing transcription to position +3, the bubble size does not increase (unpublished results). This demonstrates that in transcription, the enzyme does not melt the DNA substantially beyond the active site.

In an elongation complex walked to position +15 by withholding one of the nucleoside triphosphates (UTP, in this case), similar fluorescence studies also shown in Fig. 1 demonstrate that the transcription bubble is also about 7 to 8 bases in length. As seen in the initiation complex, DNA downstream of the last incorporated base returns to a DNA:DNA duplex within one base. These results are consistent with crystal structures of elongation complex models (19, 20). In these structures, DNA returns to the duplex just past the active site of the enzyme.

### III. Initial Collapse of the Bubble Occurs on Translocation Beyond Position +8

The previously stated results are deceptive in implying a fixed size for the transcription bubble. In the initial complex, the bubble is about 7 to 8 bases in length, with no RNA product present. In the presence of GTP, allowing synthesis of a 3-base RNA, the bubble is still only 7 to 8 bases in length. However, in order to extend the RNA beyond 3 bases, downstream DNA must be melted. Does the upstream end of the bubble begin collapsing to compensate for this melting and maintain an 8-base bubble or does the bubble initially grow in size beyond 8 bases?

As in the studies presented in Fig. 1, fluorescent probes can be used to follow the progression of the bubble as the enzyme translocates along the DNA. Though not dynamics in the true sense of the word, characterization of
incrementally stalled complexes nevertheless provides snapshots of polymerase translocation, mapping the movement of the bubble as transcription proceeds. In the demonstration of the technique described here, we follow fluorescence from 2-aminopurine placed within the initially melted bubble, just upstream of the start site, to monitor the expected initial collapse of the upstream edge of the bubble as the polymerase active site translocates away from the promoter.

As illustrated in Fig. 2A, in complexes stalled at positions +7 or +8, fluorescence from 2-aminopurine in the template strand at position -2 is high, indicating that the upstream edge of the bubble has not yet begun to collapse.
Fig. 2. Simultaneous monitoring of two ends of the bubble in initially transcribing complexes.
(A) Observation of the collapse of the initial bubble by following the fluorescence quenching of 2-aminopurine placed near the start site, at position –2 of the template strand. Concentrations of enzyme and DNA are 1 μM. Controls representing fully duplex (dsDNA) and fully single-stranded (ssDNA) are shown. The sequences of the DNA templates are such that transcription in the presence of GTP and ATP will walk the complex out to the position shown in the top number of each pair along the y-axis. Subsequent addition of 3′dCTP allows the complex to walk one base pair further, as indicated by the middle number of each group. Finally, addition of UTP should drive translocation still one base farther. High fluorescence, as for complexes stalled at positions +7 and +8, indicates a melted bubble. Low fluorescence, as for complexes stalled at positions +10 and beyond, indicates collapse to a duplex. (B) Simultaneous monitoring of the downstream bubble near the stall site. In the same DNA samples, changing excitation and emission wavelengths allows monitoring of downstream melting, near the stall site (modified from [32]).
At the same time, the fluorescence from pyrrolo-dC in the same construct, shown in Fig. 2B, is also high, confirming that all of the complexes are indeed stalled at the appropriate positions. As the complex is walked beyond position +8, however, the fluorescence from 2-aminopurine at position −2 begins to drop, reaching duplex levels by translocation to position +10. This result shows clearly that the upstream edge of the bubble begins to collapse on translocation beyond position +8.

The simultaneous use of two fluorescent probes is a powerful tool in probing the structures of the stalled complexes. Similarly, the use of different DNA sequences and the sequential addition of 3′-dCTP and UTP provides confirmation that the observations are indeed reasonably general and unlikely to be dependent on the particular sequence around the stall site.

The fluorescence results clearly demonstrate that as the enzyme translocates through the initially transcribed region, the size of the bubble must grow. The initial bubble extends from position −4 upstream (26) to about position +3 or +4 downstream (unpublished results). However, the preceding results show that during initial transcription at position +8, the bubble extends from at least position −2 to position +9, for an overall bubble size of 11 to 13 bases (depending on whether the bubble still extends upstream to its original position at −4). Fluorescence from the probe at position −2 reports that the upstream edge of the bubble begins to collapse when the complex translocates beyond position +8 and is complete by the synthesis of a 10-base RNA.

**IV. Promoter Release Occurs on Translocation Beyond Position +8**

RNA polymerase is initially brought to the transcription start site via fairly classic protein–DNA interactions with upstream promoter bases from position −17 through −5 (2, 4, 5, 7, 8, 33, 34). Early footprinting results indicated that these contacts are retained on translocation through position +6 and are released in complexes stalled at position +15 (14–16) and more recent exonuclease footprinting results suggest that promoter release occurs on translocation to about position +8 (17). We have exploited a “sink challenge” assay to assess functionally when contacts with the upstream promoter elements are lost (18, 35). The cycle shown in Fig. 3A illustrates the basis for the assay. Previous studies have shown that in transcription limited to RNAs six bases in length, the complex remains, for a time, resistant to challenge with an excess of added promoter sink (35). That is, several cycles of transcription continue from the original promoter DNA before the RNA polymerase associates with the sink. The interpretation of these results is that if the original promoter contacts are retained in a stalled complex, release of the RNA will yield a
Fig. 3. Functional measure of promoter release. (A) Susceptibility to sink measures promoter release. Initially transcribing complexes can synthesize and release abortive RNAs without dissociating from the promoter. In contrast, complexes that have escaped the promoter are accessible to competition from sink DNA. (B) Enzyme and DNA were incubated in the presence of (cold) GTP, ATP, and CTP only, providing for stalling at the indicated positions. After 1.0 min, [γ-32P]ATP and a twenty-fold excess of sink DNA were added and transcription continued for an additional 1.0 min. Each column thus represents resistance to the trap promoter in complexes stalled at positions from +6 to +15. Overlaid onto this graph in hatched bars are fluorescence data from Fig 2A, indicating collapse of the initially melted bubble, as probed by changes in the fluorescence from 2-aminopurine placed at position –2 (modified from (32) and (78)).
polymerase–promoter complex fully competent to initiate a new round of transcription and reinitiation will effectively compete with dissociation. In contrast, if promoter contacts are lost, the released RNA polymerase must reencounter promoter DNA. In the presence of an excess of a sink promoter, reinitiation from the original promoter DNA will be at a competitive disadvantage.

We have exploited this assay to determine at what position promoter contacts are lost as the polymerase translocates along the DNA. On templates designed to provide a stall at specific positions, transcription is allowed to proceed for a brief period with the necessary (limiting) NTPs. At time zero, an excess of sink DNA is added, along with an appropriate labeled NTP. Labeled RNA product will only occur from RNA polymerases that have retained their association with the original DNA, complexes that have not lost the upstream promoter contacts. In this case, instead of using an alternate duplex promoter as the sink, we have exploited the fact that partially single-stranded constructs bind more tightly to the enzyme than do fully duplex promoter constructs (1), making the sink that much more effective. We have also used a sink construct that is capable of transcription, as the template strand is extended to position +3. Since transcription from this sink incorporates only G, they will not be seen using [α-32P]ATP. Thus, all observed transcription must originate from complexes remaining associated with the original promoter DNA.

The results shown in Fig. 3B demonstrate that promoter contacts are retained on translocation through position +8, but are lost as the enzyme translocates beyond this position. Clearly, promoter release is occurring on translocation past position +8, the same positioning as seen for collapse of the upstream edge of the initial transcription bubble. This correlation is emphasized in Fig. 3B by the overlay of data from the fluorescence experiments summarized in Fig. 2. This correlation strongly suggests that it is the loss of promoter contacts that allows the collapse of the upstream edge of the bubble.

V. Initial RNA Displacement Occurs on Translocation Beyond Position +8

The preceding results demonstrate that promoter release and collapse of the initial bubble occur on translocation of the complex from position +8 to +10. Retention of the bubble upstream beyond position –1 suggests retention of the RNA in a heteroduplex with the template DNA. This raises the question: when does the 5’ end of the nascent RNA first begin to dissociate from the DNA template? A variation on this fluorescence approach provides the answer.
Fluorescent base analogs show quenched fluorescence within a duplex as a result of increased interactions with neighboring bases. This is true not only in traditional DNA:DNA duplexes, but also in RNA:RNA duplexes and in heteroduplexes (36–38). Consequently, the presence of RNA in heteroduplex with DNA should yield quenching of the fluorescence of a probe in the DNA strand. Replacement of the normal C at positions +1 or +2 of the template strand by pyrrolo-dC should then provide a reporter of the presence of bound RNA. The only drawback to this approach is that if the RNA dissociates, but the nontemplate strand of the DNA immediately re-anneals, we will observe duplex-level quenching throughout the process. This can be overcome by running a parallel experiment in which the corresponding nontemplate strand base is mismatched to the template strand base. In this case, dissociation of the RNA will not lead to reannealing of the template strand at that position and the fluorescence will increase. Indeed, it has been shown that fluorescence of pyrrolo-dC in a singly mismatched duplex is higher than in the single strand (32, 39). Presumably, the base-base interactions are more perturbed in a mismatch than in single-stranded DNA.

By placing a probe at position +1 and then walking the polymerase to positions from +10 to +13, we can follow the fluorescence changes associated with initial dissociation of the RNA. The results presented in Fig. 4A demonstrate that in a construct which correctly pairs G opposite pyrrolo-dC, the fluorescence remains low in each translocational step. The base is paired either with RNA or with the template strand DNA. The data presented in Fig. 4B provide the necessary complement. In the context of an incorrectly paired A in the nontemplate strand opposite pyrrolo-dC, we observe low fluorescence in complexes stalled at position +10, revealing that the RNA is still annealed to the template strand at position +1. As the complex is walked to positions +11 and +12, the fluorescence increases, indicating that the 5’ end of the RNA has now dissociated from the template strand DNA. Similar results with a probe at position +2 show that RNA at this position dissociates as the complex is walked beyond position +11.

VI. Bubble Collapse Contributes to Initial RNA Displacement

It is reasonable to expect that in a complex stalled at position +8, with the bubble extending from at least position −2 (more likely, −4) to +9, the RNA–DNA heteroduplex remains intact throughout its length. The fluorescence studies in which pyrrolo-dC was placed in the template strand at positions +1 and +2, reviewed in the previous section, show that in constructs containing a single mismatch, the 5’ end of the RNA begins to dissociate from the heteroduplex as the complex translocates beyond position +10 (32). Results
summarized in the following text suggest, however, that the use of a mismatched base pair at position $+1$ or $+2$, required for the fluorescence experiment, may weaken reannealing and so artificially delay dissociation of the 5' end of the RNA (18). It is likely that in normal transcription, the 5' end of the RNA begins dissociating slightly sooner, on translocation to position $+9$.

The (near) simultaneous collapse of the upstream edge of the bubble with the initial dissociation of the 5' end of the RNA suggests a functional correlation of these two events. It seems likely that collapse of the DNA base pairs at positions $+1$ and $+2$ serves to competitively displace the 5' end of the RNA.

Early studies showed that the non-template strand of the DNA downstream of position $-5$ is not required for fully functional initiation of transcription (9, 40). Studies done in 2004 have revealed, however, that transcription from such complexes leads to a significant increase in the amount of RNA products 11 to 13 bases in length, longer than traditional abortive RNA products (18). Complexes lacking the non-template strand in the transcribed region effectively lack initial bubble collapse and so might be expected to be deficient in initial displacement of the 5' end of the RNA (41). We have hypothesized that
products 11 to 13 bases in length represent complexes that have not properly resolved the heteroduplex (18). A deficiency in this resolution at the normal position of +9 yields a complex that can be extended 2 to 4 bases, but not farther. The molar ratio of 11–13mer RNA products relative to 11–20mer products represents the fraction of complexes containing 11 base RNAs that are incapable of extension beyond 13 bases. As shown in lane 9 of Fig. 5, in transcription from the control, fully duplex DNA, about 23% of the complexes that have made it through the abortive phase, nevertheless terminate at or before position +13. This suggests an intrinsic barrier in the progression to a stable elongation complex.

In order to test the hypothesis that collapse of the upstream edge of the initial bubble competitively displaces the 5′ end of the nascent RNA, we prepared a series of DNA constructs in which the non-template strand is extended progressively downstream. In a competitive displacement model, collapse of the non-template strand bases at positions −4 to −1 should not be sufficient to assist displacement of the RNA, as they do not compete directly with it. Extension of the DNA duplex downstream of position −1 should, however, begin to provide that competitive displacement.

The results presented in Fig. 5 bear out these expectations. The ratio of 11–13mer RNA products relative to 11–20mer products increases two-fold for complexes lacking the non-template strand from position −5 downstream, as illustrated in lane 1 of Fig. 5. The ratio remains high with incremental extension of the non-template strand (lanes 2–5), but as the non-template strand DNA base at position +1 and, more dramatically, at position +2 is restored, the ratio drops back down to double strand control levels (lanes 6 and 7, respectively). The presence of the bases at positions +1 and +2 is critical, consistent with the idea that collapse of these bases helps the initial displacement of the 5′ end of the RNA.

To further test the model derived in the preceding example, targeted local regions of the DNA were designed to be locally mismatched, as presented in Fig. 6. As expected from the results with partially single-stranded constructs, mismatching the bases at positions +1 through +4 is sufficient to yield the increase in the ratio of 11–13mer to 11–20mer products, as shown in lane 3 of Fig. 6. In contrast, mismatching of bases at positions +5 through +8 (lane 4) has little effect, since these bases are not expected to collapse during the time in which the 5′ end of the RNA dissociates from the heteroduplex.

Indeed, as shown in lanes 5 and 6, mismatching the base pair at either +1 or +2 singly also impairs RNA displacement. This result prompts us to revisit the fluorescence results designed to monitor initial RNA displacement. In those experiments, a single mismatch was introduced into the DNA as a way of distinguishing DNA:DNA from RNA:DNA duplexes. The results of Fig. 6 suggest, however, that the design of that experiment may artificially delay
Fig. 5. Bubble collapse near position +1 competitively displaces the 5’ end of the RNA. Transcription from partially single-stranded constructs in which the non-template strand extends downstream from position −5 to +3 (lanes 1–8) and from the double-stranded control (lane 9) are compared (transcript lengths are indicated above the gel). Relative molar amounts of 11–13mer transcripts are indicated in the bar chart. Of the complexes that have successfully extended past position +10, the ratio of 11–13mer to 11–20mer products represents the fraction that terminate prematurely at positions +11–13 (modified from (18)).
Fig. 6. Collapse of bases +1 and +2 is aided by upstream collapse. (A) Transcription from constructs with windows of four mismatched bases (lanes 2–6) and from constructs with mismatched bases in the proposed critical DNA region (lanes 2–6) are compared to that from a control construct (lane 1). Transcripts are indicated as described in the legend to Fig. 5. (B) Illustration of the effects observed in A (modified from (18)).
initial displacement of the RNA. Thus, we revise the initial conclusion of that study. In mismatched constructs, initial displacement occurs on translocation beyond position $+10$. In fully duplex constructs, the displacement is expected to occur sooner, placing initial RNA displacement approximately coincident with upstream bubble collapse and promoter release.

Since collapse of the bubble at positions $+1$ and $+2$ should be preceded by (or be coincident with) collapse of the upstream bases at positions $-4$ through $-1$, then upstream collapse should facilitate collapse at positions $+1$ and $+2$. This prediction is supported by data presented in Fig. 6. Mismatching the bases at positions $-4$ through $-1$ (but not $+1$ and $+2$) also yields high ratios of 11–13mer to 11–20mer, as shown in lane 2. Indeed, weakening of the duplex in competition with upstream bubble collapse by the mismatching of two bases upstream of the start site, at either positions $-4$ and $-3$ or at positions $-2$ and $-1$, also yields an increase in 11–13mer. Bubble collapse is impaired.

Finally, in order to probe the energetics of this effect more carefully, compensating perturbations were introduced into the DNA composing the upstream bubble. For example, to compensate for the energetic impairment of collapse introduced by mismatches at positions $-4$ and $-3$, the relatively weak (but paired) TA step at positions $-2$ and $-1$ can be replaced by the stronger GC step. As expected, the introduction of the stronger GC pairs rescues the effect of the upstream mismatch, as shown in lane 8 of Fig. 6. Strengthening the bubble in one region compensates for weakening it at another. Finally, although replacement of the TA step at positions $-2$ and $-1$ in the consensus promoter sequence by a GC step is expected to decrease transcription overall (initial promoter melting is impaired), we expect that the resulting enhancement of upstream bubble collapse should yield to more efficient RNA displacement and fewer 11–13mer products. Both predictions are confirmed. As shown in lane 10 of Fig. 6, an overall decrease in products is observed. However, since this substitution strengthens the duplex, it favors collapse of the upstream edge of the bubble in a complex transitioning beyond position $+8$. In this case, the fraction of complexes that lead to 11-13mer RNA products is reduced below the level observed for the consensus duplex.

VII. Retention of Promoter Contacts Interferes with Initial Bubble Collapse

The preceding studies demonstrate that initial collapse of the upstream edge of the transcription bubble and release of promoter contacts both occur on translocation beyond position $+8$. The crystal structure of an initiating T7 RNA polymerase complex suggests that a loop in the protein containing Val 237 helps to maintain open the upstream edge of the bubble, the Val stacking
on the template strand G at position −5 (8, 42, 43). This loop in the protein is an integral part of the N-terminal platform which, together with the specificity loop (residues 738 to 769) and the AT-rich recognition loop (residues 90–100) serves to bind the promoter DNA to the enzyme (34). Thus, binding of the promoter to the protein likely forces the Val loop into the DNA, helping to drive and maintain promoter melting (42, 43).

A logical extension of the preceding results predicts that if promoter contacts are retained, the initial bubble will be impaired in collapse. If collapse is prevented, the 5′ end of the RNA will lose one component driving its dissociation from the template strand (see following text). If the 5′ end of the RNA fails to dissociate, complexes will then experience an increased energetic barrier to translocation beyond position +13.

This prediction is tested in the study summarized in Fig. 7. Noting the presence of an unconserved Ala at position 94 in the polymerase that lies very near the 3′ end of the template strand of the DNA, we engineered an approach to covalently crosslink the RNA polymerase to its promoter DNA (44). Introduction of an alkyl-thiol at the 3′ end of the DNA allows formation of a disulfide crosslink between the enzyme and the DNA, in a location that is far from the active site. As predicted by the preceding model, transcription limited to position +6, where promoter contacts have not yet been lost, is completely normal (44). Promoter contacts are not normally lost at this point in transcription (15, 17, 45) and initiation and initial translocation proceed unimpeded.

Given the evidence already presented that promoter release occurs on translocation past position +8, in the presence of all four NTPs, we expected the complex to halt at about position +8. This is not observed and full-length RNA product is indeed produced. Extensive controls have confirmed that these full-length transcripts are produced from complexes containing bona fide crosslinks, as engineered (44). This suggests that complete release of the promoter is not required for the transition to elongation. Due to the nature of the tethering between the DNA and the protein, it is likely that effective “release” of the promoter contacts remains possible in a covalently crosslinked construct.

Interestingly, of those complexes that translocate beyond position +8, about three-fourths of the complexes produce RNAs 11 to 13 bases in length, and only one-fourth continue on to synthesize full-length runoff transcripts. This indicates that the tethering has indeed created a barrier to promoter escape. As discussed in Section VI, the release of RNAs of this length can be taken as indicative of a failure to resolve the 5′ end of the RNA from the heteroduplex at the appropriate time. This result illustrates that bubble collapse is likely only a contributing factor to resolution of the initial RNA:DNA hybrid (and, possibly, that retention of promoter contacts may not fully
inhibit collapse of the bubble). This is consistent with the results from partially single-stranded DNA constructs. In those experiments, only about 50% of the complexes halt transcription at 11 to 13 bases; the remainder go on to synthesize full-length transcripts. The engineered crosslink not only limits promoter release and bubble collapse, but also likely places topological and steric
restrictions on progression beyond the normal point of promoter release. The 
crosslink does not fully prevent the disruption of some protein–DNA contacts, 
but must surely limit that release.

In summary, the results presented in Fig. 7B show an increase in 
11–13mer RNA only under the oxidizing conditions that favor formation of 
the disulfide bond. Just as predicted, retention of promoter contacts leads to 
an impairment of bubble collapse, which leads to an impairment in resolution 
of the 5’ end of the RNA, which, in turn, limits transcription to 11 to 13 bases.

**VIII. Coupling of Promoter Release, Bubble Collapse, 
and RNA Displacement**

The biochemical and biophysical studies previously described not only 
provide significant details regarding the timing of the individual changes in 
the initially transcribing complex, but also provide insight into the interrelat-
edness of these processes. In particular, we are led to a model in which the loss 
of the initial upstream DNA contacts leads directly to the initial collapse of the 
upstream edge of the DNA bubble. While promoter contacts are maintained, 
the Val loop maintains the upstream edge of the bubble; loss of the contacts, 
and of the Val loop, then allows collapse to proceed spontaneously. Collapse of 
the bubble from the upstream end, in turn, serves to competitively displace the 
5’ end of the RNA, a key event in the progression to a stably transcribing 
elongation complex. Threading of the RNA into the putative exit channel 
completes the progression to a stable elongation complex. These events are 
presented in cartoon form, in the first column of Fig. 8.

Artificial DNA constructs that limit collapse of the nontemplate strand at 
positions +1 and +2 are impaired in displacement of the 5’ end of the RNA, as 
illustrated by the middle column of Fig. 8. The loss of this collapse does not 
completely prevent RNA displacement, however, indicating that collapse is 
only one factor driving displacement of the 5’ end. Steric constraints as well as 
simple thermal breathing of the end of the heteroduplex likely also contribute 
to displacement of the nascent RNA. Finally, failure to release the upstream 
promoter contacts, as would be expected by the introduction of a covalent 
tether between the DNA and protein, prevents collapse of the bubble, leading 
to the same results seen in constructs completely lacking collapse.

The results reviewed here show that the release of upstream promoter 
contacts and the collapse of the upstream edge of the initially melted bubble 
both occur on translocation beyond position +8. We propose that displacement 
of the 5’ end of the nascent transcript also begins on translocation 
beyond position +8, driven, in part, by competition deriving from the collapse
of the DNA bubble. Although our earlier fluorescence results suggested that displacement of the 5' end of the RNA was beginning one or two bases later, more recent results suggest that RNA displacement in these complexes is likely delayed, since a mismatch in the template strand was included in the fluorescence studies (the mismatch was introduced intentionally to distinguish RNA:DNA duplex quenching from DNA:DNA quenching).

Promoter release, collapse of the initial bubble, and resolution of the 5' end of the RNA from the heteroduplex are all key events in transcription. The preceding model demonstrates how one event can lead directly to the next. The convergence of these events in a common mechanistic scheme prompts the question: what triggers promoter release on translocation beyond position +8?

**Fig. 8.** The relationship between promoter release, bubble collapse, and initial RNA displacement. The left column describes these processes during normal transcription from double-stranded DNA templates. The middle column shows how the lack of the nontemplate strand can lead to incorrect displacement of the RNA, which in turn leads to transcripts halted at positions +11 to +13. The right column shows how failure to release the promoter similarly limits bubble collapse, again leading to halted transcripts at positions +11 to +13.
IX. A Structural Model for the Transition from Initiation to Elongation

The structures of initiating and elongating complexes of T7 RNA polymerase reveal a striking change in the conformation of the enzyme (8, 19, 20, 34, 46). In particular, a platform within the N-terminal domain of the protein undergoes a translation and a rotation first described as a 140° clockwise rotation (19, 20). This rearrangement displaces the specificity loop, which contacts the central major groove and provides most of the sequence specificity in binding (2, 3, 5, 7, 33, 47, 48).

A reanalysis of the structures has prompted a revised model for the transition, requiring instead a 220° rotation in the opposite direction (49). In this model, illustrated in Fig. 9, the N-terminal platform initially moves away from the C-terminal domain (driven by growth of the heteroduplex) without substantial rotation, as shown in panels A and B. During this translation of the platform, the specificity loop, which makes direct contact with promoter

![Diagram](https://via.placeholder.com/150)

**Fig. 9.** A model for the structural changes accompanying the transition from an initiation complex to an elongation complex. Panel A is the crystal structure of the initiating complex, with downstream DNA modeled in from the elongation complex. Panel F is the crystal structure of the elongation complex. Panels B–E are models illustrating key features of the proposed transition (modified from (49)). (See Color Insert.)
recognition bases −10 through −8 and which derives from the C-terminal part of the protein, extends to retain its position on the N-terminal platform, thereby allowing retention of the initial promoter binding interface. As has been described, retention of promoter binding maintains the melting of the upstream end of the initial bubble. This is necessary in that the initial heteroduplex is small and so is predicted to be otherwise unstable in competition with a collapsing DNA duplex.

Once the specificity loop is extended beyond its limit and dissociates from the N-terminal platform (panels B and C of Fig. 9), promoter contacts are weakened and released. Biochemical and biophysical data previously reviewed place this at translocation beyond position +8. At this point, the complex contains a full-length (compared to the elongation complex), and therefore maximally stable, heteroduplex. The modeling shown in panel B demonstrates that the specificity loop can readily accommodate translocation to at least position +6 with little or no movement of the C-terminal domain to which it is attached. Photocrosslinking data suggest that the interaction is at least altered on translocation beyond position +5 or +6 (45).

Loss of the protein–promoter contacts with promoter release should remove the Val 237 loop from the upstream edge of the initially melted bubble, allowing initial collapse of the upstream edge of the bubble, as indicated by fluorescence. This collapse, in turn, facilitates displacement of the 5′ end of the nascent RNA, allowing it to thread into the RNA exit channel created by the rotation of the N-terminal platform and the accompanying change in the refolding domain. Translation of the N-terminal domain away from the C-terminal domain also allows the former to clear the latter sufficiently to allow both rotation of the N-terminal domain and the conformational change of the refolding loop, residues 153 to 203. A possible series of events, which may take place with little or no translocation along the DNA, is illustrated in panels C through E. The correlated timing of the rotation and refolding is not known, but the constraints on the motions are. The rotation of the N-terminal domain is not possible until the initial upstream promoter contact with the protein is lost, as the DNA is topologically unable to remain bound during a rotation in the clockwise (220°) direction. Energy supplied by translocational nucleotide hydrolysis (up to position +8) likely provides the energy to disrupt the initial tight binding contacts.

X. Summary

T7 RNA polymerase is an ideal model system in which to study fundamental characteristics of the complex machine that is RNA polymerase. Although structurally unrelated to the more complex multi-subunit RNA polymerases, it
appears that the common functional requirements imposed on all DNA-dependent RNA polymerases is likely to be reflected in many common mechanistic features. Both the small size of the protein and the small size of the promoter DNA allow studies in the single subunit enzymes which are difficult or impossible in the much larger multi-subunit enzymes. Similarly, the ability to prepare large amounts (and concentrations) of highly active T7 RNA polymerase opens doors that remain largely shut to the multi-subunit systems.

The studies presented here add to a substantial body of biochemical literature on structure and function of T7 RNA polymerase (1, 10, 12, 31, 50–57). A self-consistent picture of the critical transition from initiation to elongation is beginning to emerge. The resulting model presented here generates new testable hypotheses that will allow us to further refine our understanding of this complex molecular machine.

ACKNOWLEDGMENTS

This work was supported by grant GM55002 from the National Institutes of Health to C. T. M. E. A. E was supported by National Research Service Award T32 GM08515 from the National Institutes of Health.

REFERENCES


**STRUCTURE AND FUNCTION IN PROMOTER ESCAPE**


