The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance

Dmitry Temiakov*, Pamela E. Mentesana*†, Kaiyu Ma*, Arkady Mustaev‡, Sergei Borukhov*, and William T. McAllister*§

*Morse Institute of Molecular Genetics, Department of Microbiology and Immunology, State University of New York Health Science Center at Brooklyn, 450 Clarkson Avenue, Box 44, Brooklyn, NY 11203-2098; and ‡Public Health Research Institute, New York, NY 10016

Communicated by F. William Studier, Brookhaven National Laboratory, Upton, NY, October 5, 2000 (received for review May 25, 2000)

During the early stages of transcription, T7 RNA polymerase forms an unstable initiation complex that synthesizes and releases transcripts 2–8 nt in length before disengaging from the promoter and isomerizing to a stable elongation complex. In this study, we used RNA·protein and RNA·DNA crosslinking methods to probe the location of newly synthesized RNA in halted elongation complexes. The results indicate that the RNA in an elongation complex remains in an RNA·DNA hybrid for about 8 nt from the site of nucleotide addition and emerges to the surface of the enzyme about 12 nt from the addition site. Strikingly, as the transcript leaves its hybrid with the template, the crosslinks it forms with the RNA polymerase involve a portion of a hairpin loop (the specificity loop) that makes specific contacts with the binding region of the promoter during initiation. This observation suggests that the specificity loop may have a dual role in transcription, binding first to the promoter and subsequently interacting with the RNA product. It seems likely that association of the nascent RNA with the specificity loop facilitates disengagement from the promoter and is an important part of the process that leads to a stable elongation complex.

Despite a lack of obvious sequence or structural homology between the single subunit T7 RNA polymerase (RNAP) and the multisubunit RNAPs, the basic features of the transcription process are highly conserved among the two groups of enzyme (reviewed in ref. 1). As is the situation with other RNAPs, T7 RNAP forms an unstable initiation complex (IC) that synthesizes and releases transcripts 2–8 nt in length (abortive initiation products) before disengaging from the promoter and isomerizing to a stable elongation complex (EC). Whereas considerable biochemical and structural data are available concerning T7 RNAP initiation complexes, little is known about the transition to an EC or about the properties of the stable complex. Recognition of the promoter involves a specificity loop in the RNAP (amino acid residues 739–770) that projects into the DNA binding cleft and interacts with the binding region of the promoter, which lies 7 to 11 bp upstream from the active site (i.e., positions −7 to −11) (2, 3). The transition from duplex DNA in the binding region to open or melted DNA in the initiation region commences between −5 and −4, and involves an intercalating β hairpin loop; the template strand is then led down into the active site by additional contacts with the surface of the enzyme (3). During abortive initiation, the contacts with the binding region of the promoter are maintained while the leading edge of the initiation complex moves downstream, resulting in a more extended footprint of the complex on the template (4, 5). Packing of the DNA into the complex is accomplished by "scrunching" of the intervening portion of the template strand into a hydrophobic binding pocket (6).

The transformation to a stable EC commences when the nascent RNA has achieved a length of ≈9 nt and is accompanied by release of the upstream promoter contacts (4, 7). However, a transcription complex with all of the properties of a fully processive EC does not appear to be formed until after 12–14 nt of RNA have been synthesized (8). It has been proposed that the progression to a stable EC is triggered by filling of the template strand binding pocket (6) and/or by association of the transcript with an RNA product binding site in the N-terminal domain of the RNAP (9). The work shown here indicates that interactions between the nascent RNA and the specificity loop are an important element in this transition.

Materials and Methods

RNA Polymerase and Templates. Mutant RNAPs were constructed and purified as described (10, 11). All RNAPs described here have an N-terminal His_6 leader and exhibited normal activity. Templates that allow the incorporation of UTP analogs at defined positions in the transcript and the use of immobilized RNAPs to extend the transcript by successive cycles of limited elongation have been previously described (8, 11, 12).

RNA·DNA and RNA·Protein Crosslinking. To prepare halted elongation complexes, 20 pmol of T7 RNAP was incubated with an equimolar concentration of template, 0.3 mM GTP, 0.1 mM ATP, and 50 μM UTP analog (see below) in 20 μl of transcription buffer (40 mM Tris-acetate, pH 7.9/8 mM magnesium acetate/5 mM β-mercaptoethanol/0.1 mM EDTA) for 5 min at 37°C. The startup complexes were immobilized on Ni^{2+}-agarose beads (Qiagen, Chatsworth, CA) and extended in the presence of limiting mixtures of nucleoside triphosphates at a final concentration of 10 μM each (11). Samples were chilled on ice, and crosslinking was activated as described below.

Crosslinking with the UTP analog U● was carried out as described in (13). This analog has two reactive groups that are activated in the presence of $NaBH_4$, an aldehyde that may form a Schiff's base with primary amines in the protein, and an aromatic bis(2-iodoethyl)amino group linked to the fifth position of pyrimidine, which forms specific crosslinks to the base to which the analog is paired in an RNA·DNA hybrid (13).

Crosslinking with 4-thio-UTP (sUTP; Amersham Pharmacia)

Abbreviations: RNAP, RNA polymerase; IC, initiation complex; EC, elongation complex; HA, hydroxylamine; NTCB, 2-nitro-5-thiocyanobenzoic acid; WT, wild type; NCS, *N*-chlorosuccinimide.

†Present address: Department of Microbiology and Molecular Genetics, Life Sciences, Room 166, SUNY Stony Brook, Stony Brook, NY, 11794.

§To whom reprint requests should be addressed. E-mail: pogo51@aol.com.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.250473197. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.250473197

was activated by exposure to a UV lamp (Cole–Palmer, 6 W) at 365 nm for 10 min (11, 12). Stop buffer was added, and the samples were resolved by electrophoresis in 6% polyacrylamide gels in the presence of 0.1% SDS/PAGE and visualized by autoradiography (14). The crosslinked RNAP was isolated by electroelution, precipitated twice with acetone, dried *in vacuo*, and taken up in 20 μl of distilled water. Crosslinking was specific and required the presence of sUTP, UV irradiation, and active transcription (12).

Peptide Mapping. *Hydroxylamine (HA) cleavage.* A 2-μl aliquot of crosslinked RNAP was mixed with 30 μl of HA (Sigma) in 6 M urea, 4.5 M lithium hydroxide (pH 10), and incubated at 45°C for 2–4 h. The sample was precipitated with 10% trichloroacetic acid (TCA) at 0°C, washed in 5% TCA, taken up in 10 μl of loading buffer, and resolved by PAGE in 10% gels.

Cleavage with 2-nitro-5-thiocyano-benzoic acid (NTCB) and CNBr. Samples prepared as described above were digested with NTCB and CNBr as in ref. 15. Products of cleavage were resolved either by PAGE in 10% gels or in a 4–12% NuPage Bis-Tris gradient gel using an MES buffer system and Seeblue size markers (Invitrogen), as noted in the figure legends.

N-chlorosuccinimide (NCS) cleavage. A 2-μl aliquot of crosslinked material was taken up in 3 μl of 150 mM HCl and mixed with 5 μl of NCS (Sigma; 10 mg/ml in water). After 15 min incubation at room temperature, a fresh portion (5 μl) of NCS was added, and the samples were incubated for an additional 5 min. The reaction was stopped by addition of 5 μl of loading buffer, and the samples were analyzed by electrophoresis as described above.

RESULTS

Characterization of the Elongation Complex. To determine the disposition of the RNA in T7 RNAP elongation complexes, we incorporated into the RNA analogs of UTP that may be crosslinked either to DNA or to protein, and subsequently identified the locations of the crosslinks. To place the analogs at defined positions in the transcript, we took advantage of a modified form of T7 RNAP having a His_6 leader at the N terminus (10). Use of the modified RNAP allowed us to immobilize transcription complexes on Ni^{+2} agarose beads and to "walk" the complexes along the template by successive cycles of washing and incubation with limited mixtures of substrates (Fig. 1). The halted complexes studied here have all made transcripts of at least 15 nt and appear to be true elongation complexes; they are highly stable (half-life over 1 h) and are nearly quantitatively extended during each subsequent cycle, indicating either that they represent true intermediates in the reaction pathway or are readily able to reenter the pathway.

To determine the range over which the transcript remains in association with the DNA template, we used a UTP analog (U●) that has been shown to form crosslinks exclusively with the adenine to which it is paired in an RNA·DNA hybrid (13). In addition to its ability to form specific crosslinks with DNA, this analog is also able to form crosslinks with the protein via a reactive aldehyde group (13). Whereas crosslinking to the RNAP was observed when the analog was placed at any position between −1 and −17, efficient crosslinking to the DNA was observed only when the analog was positioned 1 to 8 nt upstream from the 3′ end of the transcript (i.e., positions −1 to −8). (In this work, we identify positions in the transcript relative to the elongating 3′ end of the RNA at −1; see ref. 15.) These results demonstrate that the RNA in an EC remains in close proximity to the DNA (presumably in an RNA·DNA hybrid) from −1 to −8.

To determine at which point the RNA becomes accessible to the solvent, we labeled the 3′ end of a 17-nt transcript in a stable EC and then treated the complexes with RNase T1 (Fig. 2).

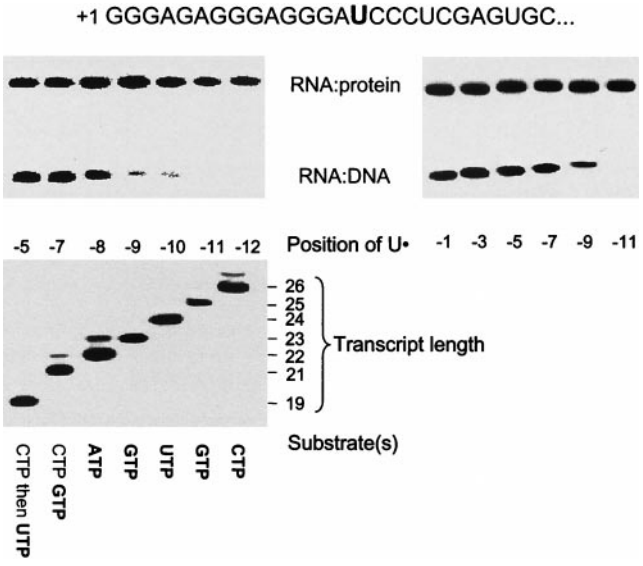

Fig. 1. Crosslinking of RNA to DNA and RNAP. A 120-bp template that directs synthesis of a transcript with the sequence indicated (*Top*) was constructed by PCR amplification of pPK10 (8). Start up complexes that extended to +15 were formed by incubation of His_6-T7 RNAP in the presence of GTP, ATP, and U● (bold). The complexes were immobilized on Ni^{2+}-agarose beads, and the transcripts were incrementally extended by sequential cycles of washing and incubation with the substrates indicated (11). Transcripts were labeled during each cycle by inclusion of the [α-^{32}P]NTP indicated in bold, and each sample was divided into two portions. One portion was examined directly by electrophoresis in 20% gels to verify appropriate extension of the transcript (*Lower*; the position of the U analog in the transcript is expressed relative to the 3′ end of the RNA at −1). Crosslinking of the transcripts in the other portion was activated by exposure to $NaBH_4$ (13), and the samples were analyzed by PAGE in a 12% gel in the presence of 6 M urea (*Upper*). Whereas crosslinking to the RNAP was observed at all positions from −5 to −12, efficient crosslinking to the DNA was observed only from −5 to −8. Similar results were obtained using other templates (PK10, PK12, PK13, PK14, D2, and DT3; see ref. 11) that allow U● to be positioned from −1 to −11 (*Right*). RNAP·RNA complexes were identified by their sensitivity to proteinase K and retention on Ni^{2+} agarose beads; RNA·DNA complexes had a mobility that corresponds to the template strand crosslinked to the expected transcript.

Twelve nucleotides of nascent RNA were protected in the intact EC, indicating that the transcript does not emerge from the interior of the RNAP until this point. A similar conclusion was reached in previous studies by using templates that direct the synthesis of a self-cleaving hammerhead structure in the RNA (16). Here, it was observed that 13 nucleotides past the cleavage point must be synthesized before the transcript can fold and self-cleave, suggesting that the RNA is not free of steric constraints until this point.

To probe contacts between the nascent transcript and the RNAP, we incorporated 4-thio-UTP (sUTP) into the RNA. This analog reacts only with protein residues that are in close proximity to the base (17). As shown in Fig. 3, transcripts containing this analog at all positions from −1 to −17 formed crosslinks with the RNAP. However, the most efficient crosslinking was observed when the analog was positioned at −1 and −9. It had previously been shown that T7 RNAP can bind single-stranded RNA (ssRNA) in a nonspecific fashion, and that exogenous oligomers of ssRNA or ssDNA are effective competitors for this binding (9). Based on the RNase T1 protection experiments described above, it is likely that crosslinking of transcripts beyond −12 involves a surface binding site(s). Consistent with this notion, we have found that ssDNA inhibits crosslinking of transcripts having sUTP positioned at −12 and beyond (not shown).

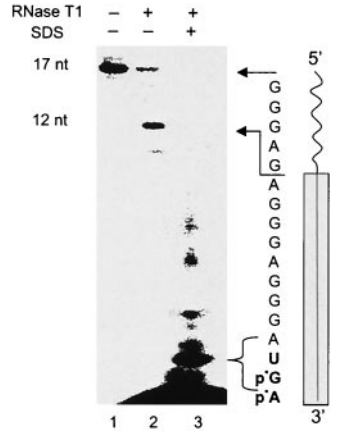

Fig. 2. Twelve nucleotides of nascent RNA are protected within the EC. A startup complex that extends to +14 was formed on a template derived from pPK12 (11), and the transcripts were extended by incubation with UTP, [α-^{32}P]ATP, and [α-^{32}P]GTP (bold). The sample was divided into three portions that were exposed to 0.1% SDS and/or RNase T1 (0.1 unit, 15 min at 25°C) as indicated, and analyzed by PAGE in a 20% gel in the presence of 6 M urea. An intact transcript of the expected length (17 nt) was observed in the untreated sample. This transcript was reduced to a limit size of 12 nt (shaded box) in undissociated complexes (−SDS) but was degraded in complexes that had been disrupted by exposure to SDS. The size of the limit digest was determined from an overexposed film in which a complete ladder of transcripts extending to +17 was observed (data not shown).

To determine whether the crosslinked RNA could be further extended, complexes formed in the presence of unlabeled substrates were crosslinked and subsequently incubated with [α-^{32}P]NTPs (Fig. 3). Whereas transcripts crosslinked from −1 to −7 were extended only poorly, transcripts crosslinked at −9 and −11 were readily extended (Fig. 3). The former observation is consistent with the notion that the RNA nucleotides from −1 to −7 are close to the active site or are involved in an RNA·DNA hybrid, and that crosslinking would alter the structure and/or movement of the hybrid. Transcripts crosslinked at −9 and −11 could be extended by at least 10 nt (Fig. 3), suggesting either that the element in the RNAP to which the transcript is crosslinked is flexible, and/or that the RNA is extruded between the site of its displacement from the template and the site on the protein to which it is crosslinked. Interestingly, transcripts crosslinked from −13 to −17 were extended only poorly, perhaps because the region of the RNAP affected by these crosslinks is crucial for processive elongation (see *Discussion*).

Mapping of the Crosslink at −9. For a number or reasons, the crosslink at −9 was of particular interest. First, as noted above, it seems likely that the base at this position is near the point at which the transcript is displaced from the template. More importantly, it corresponds to the length of the nascent transcript that is present when the transition from an unstable IC to a stable EC commences (4, 7, 8). To map this contact in the EC, we used a combination of conventional protein mapping methods together with site-directed mutagenesis.

HA cleaves between asparagine (N) and glycine (G) residues (11). There are two occurrences of the NG pair in T7 RNAP, at positions 289 and 588, and HA cleavage is therefore expected to generate ≈30-kDa fragments from both the N- and C-terminal regions as well as two partial digestion products of ≈60 kDa. As shown in Fig. 4, peptide fragments of these sizes were labeled by a transcript crosslinked at −9. To discriminate whether it was the C-terminal or the N-terminal fragment that was being illuminated, we constructed a mutant T7 RNAP in which the N-terminal cleavage site was eliminated by substitution of Asn-289 with Asp (N289D). Digestion of the crosslinked mutant protein resulted in a labeled 30-kDa fragment but no label in the 60-kDa fragment (Fig. 4). Because the only cleavage site in the mutant protein is the NG at 589, the crosslink with RNA at −9 must involve amino acid residues in the 30-kDa fragment between 589 and 883.

NTCB modifies and cleaves proteins at cysteine residues (17), of which there are twelve in T7 RNAP. Cleavage with this agent again resulted in a labeling pattern that was consistent with a C-terminal crosslink (Fig. 4). To clarify the interpretation of this pattern and to verify the position of the crosslink, we used RNAPs in which C723 and C839 had been mutated to eliminate cleavage at these sites. The results indicate that the crosslink is between residues 723 and 839. Thus, the two smallest labeled bands from wild-type polymerase are both eliminated by the

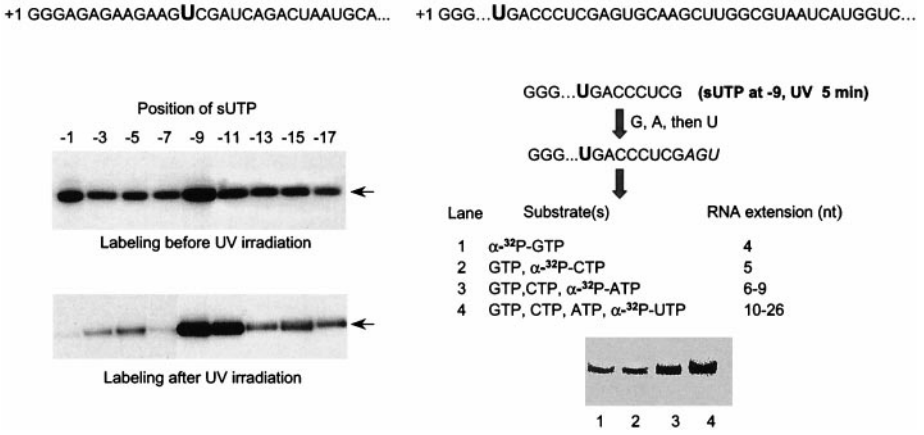

Fig. 3. Crosslinking of the nascent transcript to the RNAP (*Left*). Startup complexes that incorporate sUTP at +14 were formed by using a synthetic template that directs transcription with the sequence indicated, and the transcripts were extended incrementally as described in Fig. 1. The complexes were exposed to UV light and analyzed by PAGE in a 10% gel (11). The arrow indicates the position of the RNA·RNAP complex; this species is observed only when sUTP is incorporated into the transcript, is sensitive to proteinase K, and is retained on Ni^{2+} beads (11, 12). Transcripts were labeled either before crosslinking by incorporation of [α-^{32}P]ATP during formation of the startup complex, or after crosslinking by incubation with the next [α-^{32}P]NTP to be incorporated (*Right*). Complexes having sUTP at −9 were formed as in the *Left* panel, UV-irradiated, and extended 3 nt further by incubation with GTP, ATP, and then UTP (italics). The sample was divided into four portions and incubated with the substrates indicated; the minimal length by which the transcript must be extended after crosslinking to incorporate the labeled substrate is given.

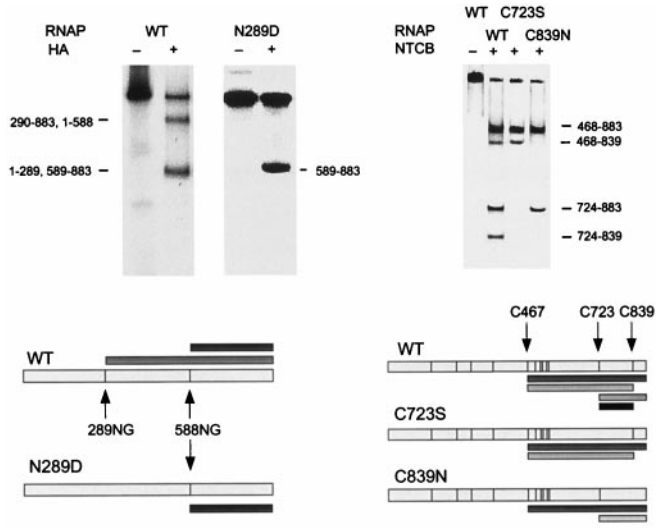

Fig. 4. Coarse mapping of the crosslink at −9. Complexes having a transcript crosslinked at −9 were formed using WT or mutant RNAPs, as indicated. Samples were digested with HA (*Left*) (11) or NTCB (*Right*) (17) and analyzed as in Fig. 3. The bars at the bottom provide a schematic representation of the fit of the observed labeled peptides to the predicted cleavage sites in the RNAP. Cleavage within the cluster of six cysteine residues from 467 to 540 is represented as a single cleavage event at C467.

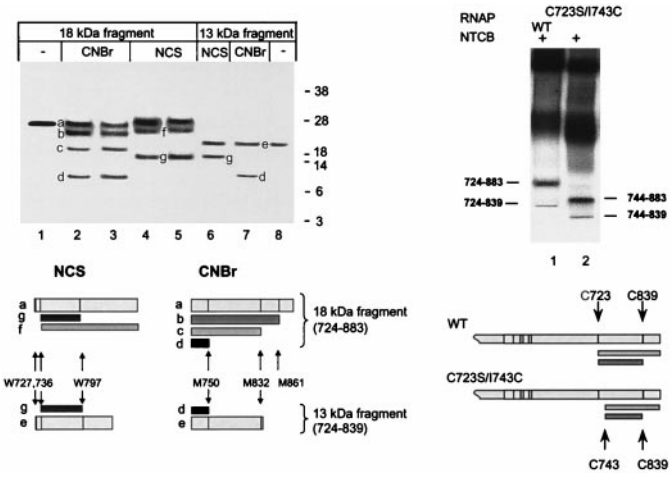

Fig. 5. Further localization of the crosslink at −9. (*Left*) The 13-kDa and 18-kDa NTCB cleavage fragments that correspond to the intervals from 724–839 and 724–883 in the WT enzyme (Fig. 4) were excised from the gel and digested with NCS or CNBr, which cleave after tryptophan (W) or methionine (M) residues, respectively (17, 31). Samples were run in a 4–12% NuPage Bis-Tris gradient gel using an MES buffer system and Seeblue size markers (Invitrogen); the positions of size markers are given to the right. Individual peptides are identified by letter; the fit of these fragments to the predicted cleavage sites is presented at the bottom. (*Right*) Complexes having a crosslink at −9 were formed with WT RNAP or the double mutant C723S, I743C, and treated with NTCB.

C723S mutation, consistent with their representing the 724–839 fragment (13 kDa) and a partial digestion product from 724 to 883 (18 kDa). The smallest labeled band and a considerably larger one are both eliminated by the C839N mutation, consistent with their representing the 724–839 fragment and a 468–839 partial digestion product, but the labeled 724–883 and 468–883 fragments remain.

To localize the −9 crosslink further, the 13-kDa and 18-kDa NTCB fragments from the wild-type (WT) enzyme were purified by gel electrophoresis and subsequently cleaved with NCS or CNBr. As shown in Fig. 5, each treatment produced a mixture of labeled products resulting from partial digestion. The only consistent interpretation of these data is that the smallest labeled band apparent after NCS digestion represents the interval W737–W797, and the smallest labeled band after CNBr treatment represents the interval from A724 to M750. (In these experiments, the labeled bands migrate somewhat more slowly than protein markers of equivalent size because of the crosslinked RNA.) Thus, the −9 crosslink must lie between residues 737 and 750. This assignment was corroborated by cleavage of the 30-kDa HA fragment with NTCB, CNBr, and NCS (not shown).

Further localization of the crosslink within the interval between 737 and 750 involved the construction of a double mutant of T7 RNAP in which the cysteine at 723 was eliminated and a new cysteine was inserted at position 743 (C723S, I743C). The labeling pattern of the NTCB-cleaved mutant enzyme is similar to that of the WT enzyme except that the sizes of the smallest crosslinked peptides are decreased by 2.5 kDa (Fig. 5). This change in size is consistent with a shift in the position of the cysteine residue that defines the N terminus of the cleavage fragment from 723 to 743, indicating that the crosslink lies between residues 744 and 750.

Discussion

The specificity loop of T7 RNAP (which comprises amino acid residues 739–770) projects into the DNA binding cleft of the RNAP and makes specific contacts with the upstream region of the promoter during binding and initiation. The 7-aa interval between Q744 and M750 to which the crosslink at −9 of the RNA in an EC is made corresponds to one arm of this loop and encompasses residues that are directly involved in promoter recognition. These residues include N748, which interacts with the base pairs at −10 and −11 of the promoter DNA, and R746, which interacts with the base pair at −7 (2, 3). The observation that the transition from IC to EC commences when the RNA achieves a length of ≈9 (4, 7, 8), together with the observation that the nucleotide at −9 in the RNA in an EC forms crosslinks with the promoter-recognition region of the polymerase, suggests that an interaction between the nascent transcript and the specificity loop plays an important role in promoter clearance and/or stabilizing the EC. Perhaps the growing RNA chain disrupts the interaction between the specificity loop and the promoter, or stabilizes a conformation of the loop that inhibits its re-association with the DNA.

Four separate crystal structures have been solved for T7 RNAP: the apoenzyme, the enzyme complexed with T7 lysozyme (an inhibitor of T7 RNAP), a binary complex of the RNAP bound to its promoter, and an initiation complex in which the first three nucleotides of RNA have been synthesized (3, 6, 18, 19). As yet, there is no information with regard to the structure of an elongation complex such as we have characterized here.

In earlier studies, Jeruzalmi and Steitz (19) modeled a putative RNA·DNA hybrid into the structure of a T7 RNAP·lysozyme complex by homology with the *Taq*I DNA polymerase primer/template complex, and found that the binding cleft could neatly accommodate 6–8 bp of hybrid with little steric clash. However, in a more recently solved structure of an initiation complex formed in the presence of GTP (which allows incorporation only of the first three G residues), Cheetham and Steitz (6) observed a different trajectory for the 3-bp RNA·DNA hybrid than predicted in the earlier study, and noted that further extension of the hybrid would result in a steric clash with the N-terminal domain. Furthermore, they noted that the base at the 5′ end of the transcript appeared to be "peeling off" from the template

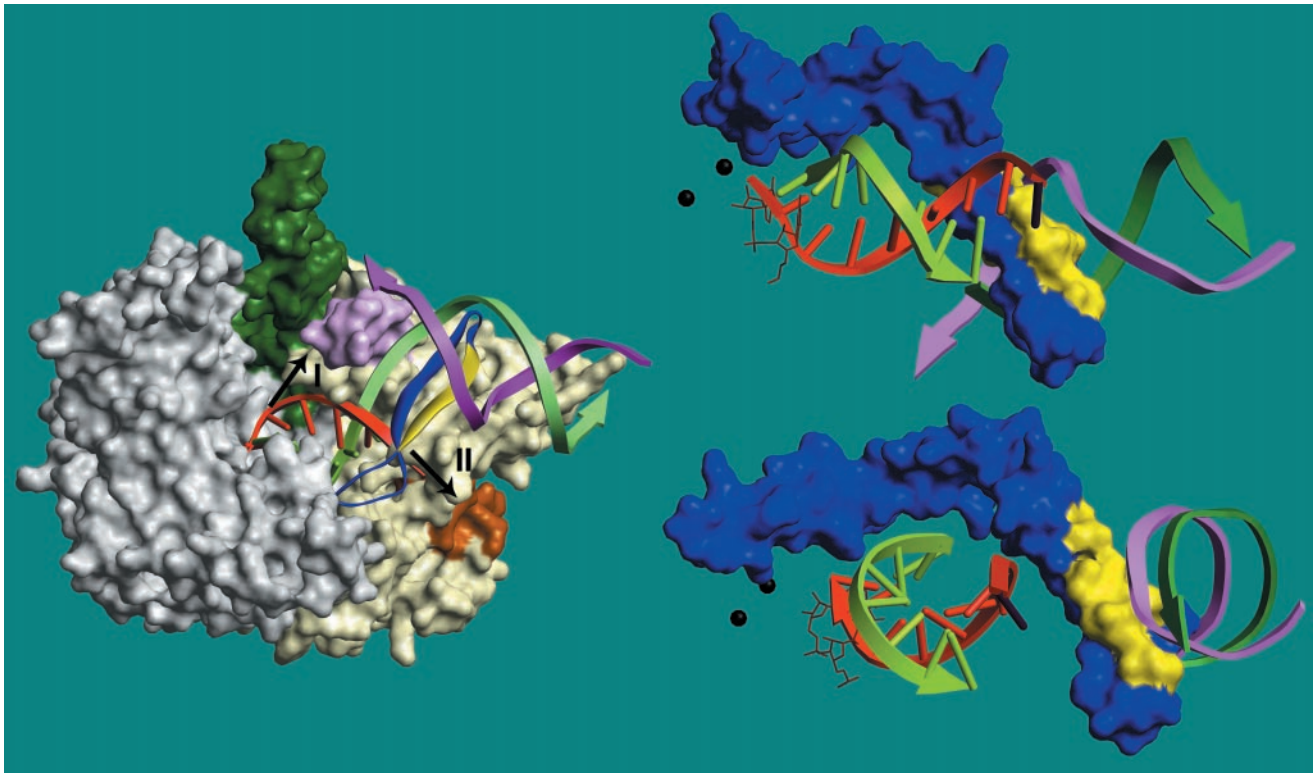

Fig. 6. Model of a T7 RNAP transcription complex. The structures shown are based on the experimentally determined structure of the T7 RNAP initiation complex (6). In the polymerase, the N-terminal domain is tan, the thumb domain is green, the intercalating hairpin that directs the template strand into the active site (residues 230–250) is pink, the N-terminal surface-exposed binding site for exogenous RNA (21) is brown, the specificity loop (739–770) is dark blue, and the 7-aa interval to which the −9 crosslink is made (744–750) is yellow. A putative 8-bp RNA·DNA hybrid has been modeled into the DNA binding cleft by homology modeling to the *Taq*I DNA polymerase primer/template complex (20) by superimposing D537, D812, and Y639 in T7 RNAP (PDB ID QLN), with the corresponding residues in *Taq*I DNAP (PDB ID TAU) using WEBLAB VIEWERPRO 3.5 (Molecular Simulations, Waltham, MA). The template strand is green, the non-template strand cyan, and the newly made RNA red. In this view, the promoter is to the right, the active site is at the left end of the RNA·DNA hybrid, and the front of the cleft obscures the active site and much of the template strand of the RNA·DNA hybrid. Arrows show the suggested exit pathways for the nascent transcript as proposed by Cheetham and Steitz (6) or in this work (arrows I and II, respectively). To the right of the overall complex are shown close-up views of the specificity loop, the RNA·DNA hybrid, and the binding region of the promoter from two different perspectives. The complex has been rotated such that the view is now under the specificity loop and the template strand of the RNA·DNA hybrid is now visible. The first two nucleotides of the nascent RNA are shown in wireframe, and the alpha carbons of the two Asp residues that define the active site (D537, D812) are shown as black spheres. The RNA has been extended by 1 nt (black) to show the proximity of the modeled base at −9 to the specificity loop. In this model, 2–3 nt of unpaired template strand would be required to transit the specificity loop and reestablish duplex DNA at the upstream border of the transcription bubble.

(i.e., was not involved in true Watson-Crick base pairing). Based on these observations, the authors suggested that the RNA·DNA hybrid could not exceed 3 bp in the IC. The exit pathway for the displaced RNA proposed in the latter study is not consistent with the results obtained here (Fig. 6), but may correspond to the path for poly(G) products that are synthesized by transcript slippage at promoters that initiate with +1 GGG (7) or for short products that are released during abortive initiation.

In considering the possible position of the RNA·DNA hybrid in an elongation complex, we repeated the earlier strategy of Jeruzalmi and Steitz (19) by homology modeling the *Taq*I DNA polymerase primer/template complex (20) into the T7 RNAP initiation complex, superimposing the highly conserved residues D537, D812, and Y639 in the active site of T7 RNAP with the corresponding residues in *Taq*I DNAP (Fig. 6). As in the earlier studies (19), we observed few steric clashes between the RNA·DNA hybrid and residues in the binding cleft. Although the position (and shape) of the hybrid should be considered tentative, the model predicts that the transcript nucleotide at −9 would be near the region in the specificity loop to which it forms a crosslink. Furthermore, the trajectory of the displaced RNA would direct it toward a previously identified surface binding site in the N-terminal domain (21).

As noted above, the transition to a stable EC commences when the RNA has achieved a length of ≈9 nt, but the transition is not completed until after 12–14 nt have been synthesized (8). The later stages in this process may involve binding of the emerging transcript to the surface binding site. Preliminary experiments have shown that transcripts crosslinked at −14 are attached to a region of the RNAP that lies near the HA cleavage site at position 289, which is consistent with this expectation (unpublished observations). A number of mutations that affect processivity and termination map to this region of the RNAP (22–24), which may explain why crosslinking of transcripts beyond −13 prevents further elongation.

The model shown in Fig. 6 suggests the possibility that the specificity loop may continue to be involved in displacement of the transcript and resolution of the transcription bubble after the polymerase has cleared the promoter, and might also monitor the DNA and/or RNA for sequences that are involved in termination or pausing (25). Clearly, substantial rearrangements would need to occur during the transition from an IC to an EC in order to accommodate the model proposed. The notion that significant structural alterations occur during isomerization is consistent with a variety of experimental data in the T7 system, as well as results with multisubunit RNAPs (1, 15, 26–29). In

further support of the proposed position of the RNA·DNA hybrid in the T7 EC, we note that many of the structures that are involved in interactions with the primer/template in the pol I family of DNAPs are conserved in T7 RNAP (19), suggesting that these regions may be involved in similar functions in both RNA and DNA polymerization.

The overall organization of the T7 elongation complex described here bears a remarkable similarity to the organization of ternary complexes formed by the multisubunit RNAPs (15). Thus, for both types of RNAP, the RNA·DNA hybrid is proposed to be 8–9 bp in length, and the RNA does not emerge to the surface of the enzyme until 12–14 nt have been synthesized. In this work, we have proposed a role for the specificity loop of T7 RNAP in binding the nascent transcript and resolving the trailing edge of the transcription bubble. A similar role has been proposed for the ''rudder'' element in *Escherichia coli* RNAP (15, 30). (We prefer the more nautically correct term of marlinespike, a tool that is used to separate the strands of a rope during splicing.) Interestingly, we note the presence of highly conserved basic residues (arginine and lysine) on the surface of the rudder that faces the putative RNA·DNA hybrid in *E. coli* RNAP, and the presence of a conserved arginine residue (R746) in the region of the specificity loop of T7 RNAP that interacts with the nascent transcript. Whereas the rudder is not directly involved in promoter interactions, a coiled-coil structure that projects from this element, and the flexible ''flap'' under which the nascent RNA emerges after its displacement, are both thought to interact with the sigma subunit (the transcription specificity factor that is involved in promoter recognition) (15, 29). Thus, as in the case of T7 RNAP, interactions of the nascent RNA with the rudder or the flap may trigger release of promoter interactions and/or stabilize the EC.

The observation that transcripts crosslinked at −9 may be further extended, perhaps by looping out of the RNA, suggests that the organization of the EC may be somewhat flexible. This finding has potential implications for models of pausing and termination, as conditions that slow the transit of the emerging transcript over the surface of the RNAP while maintaining a high rate of polymerization might also lead to extrusion of the RNA, providing an opportunity for the transcript to contact additional surfaces of the enzyme.

We thank Ray Castagna, Manli Jiang, and Dr. Michael Anikin for technical advice and assistance, and Dr. Vadim Nikoforov for critical comments on the manuscript. This work was supported by National Institutes of Health Grants GM381477 to W.T.M., GM54098 to S.B., and GM30717 and GM49242 to A.M.

1. McAllister, W. T. (1997) in *Transcription by T7 RNA Polymerase*, Nucleic Acids and Molecular Biology, eds. Eckstein, F. & Lilley, D. M. J. (Springer, Berlin), Vol. 11, pp. 15–25.
2. Raskin, C. A., Diaz, G. A. & McAllister, W. T. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 3147–3151.
3. Cheetham, G., Jeruzalmi, D. & Steitz, T. A. (1999) *Nature (London)* **399,** 80–83.
4. Ikeda, R. A. & Richardson, C. C. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3614–3618.
5. Place, C., Oddos, J., Buc, H., McAllister, W. T. & Buckle, M. (1999) *Biochemistry* **38,** 4948–4957.
6. Cheetham, G. & Steitz, T. A. (1999) *Science* **286,** 2305–2309.
7. Martin, C. T., Muller, D. K. & Coleman, J. E. (1988) *Biochemistry* **27,** 3966–3974.
8. Mentesana, P. E., Chin-Bow, S. T., Sousa, R. & McAllister, W. T. (2000) *J. Mol. Biol*., in press.
9. Muller, D. K., Martin, C. T. & Coleman, J. E. (1988) *Biochemistry* **27,** 5763–5771.
10. He, B., Rong, M., Lyakhov, D. L., Gartenstein, H., Diaz, G. A., Castagna, R. C., McAllister, W. T. & Durbin, R. K. (1997) *Protein Expression Purif*. **9,** 142–151.
11. Temiakov, D., Mentesana, P. E. & McAllister, W. T. (2000) in *Characterization of T7 RNA Polymerase Protein:DNA Interactions During the Initiation and Elongation Phases,* Protein:DNA Interactions: A Practical Approach, eds. Travers, A. A., Buckle, M. & Hames, B. D. (Oxford Univ. Press, Oxford, U.K.), pp. 351–364.
12. Karasavas, P. E. (1998) Dissertation (State University of New York, Brooklyn, NY).
13. Nudler, E., Mustaev, A., Lukhtanov, E. A. & Goldfarb, A. (1997) *Cell* **89,** 33–41.
14. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
15. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. & Darst, S. A. (2000) *Science* **289,** 619–624.
16. Tyagarajan, K., Monforte, J. A. & Hearst, J. E. (1991) *Biochemistry* **30,** 10920–10924.
17. Markovtsov, V., Mustaev, A. & Goldfarb, A. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 3221–3226.
18. Sousa, R., Chung, Y. J., Rose, J. P. & Wang, B. C. (1993) *Nature (London)* **364,** 593–599.
19. Jeruzalmi, D. & Steitz, T. A. (1998) *EMBO J.* **17,** 4101–4113.
20. Eom, S. H., Wang, J. & Steitz, T. A. (1996) *Nature (London)* **382,** 278–281.
21. Sastry, S. S. & Ross, B. M. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 9111–9116.
22. He, B., Rong, M., Durbin, R. K. & McAllister, W. T. (1997) *J. Mol. Biol.* **265,** 275–288.
23. Lyakhov, D. L., He, B., Zhang, X., Studier, F. W., Dunn, J. J. & McAllister, W. T. (1997) *J. Mol. Biol*. **269,** 28–40.
24. Macdonald, L. E., Durbin, R. K., Dunn, J. J. & McAllister, W. T. (1994) *J. Mol. Biol.* **238,** 145–158.
25. He, B., Kukarin, A., Temiakov, D., Chin-Bow, S. T., Lyakhov, D. L., Rong, M., Durbin, R. K. & McAllister, W. T. (1998) *J. Biol. Chem.* **273,** 18802–18811.
26. Bonner, G., Lafer, E. M. & Sousa, R. (1994) *J. Biol. Chem.* **269,** 25129–25136.
27. Sousa, R. (1997) in *Fundamental Aspects of T7 RNA Polymerase Structure and Mechanism*, Nucleic Acids and Molecular Biology, eds. Eckstein, F. & Lilley, D. M. J. (Springer, Berlin), Vol. 11, pp. 1–14.
28. Sousa, R., Patra, D. & Lafer, E. M. (1992) *J. Mol. Biol.* **224,** 319–334.
29. Mooney, R. A. & Landick, R. (1999) *Cell* **98,** 687–690.
30. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999) *Cell* **98,** 811–824.
31. Lischwe, M. A. & Sung, M. T. (1977) *J. Biol. Chem.* **252,** 4976–4980.