Promoter melting triggered by bacterial RNA polymerase occurs in three steps

Jie Chen[a], Seth A. Darst[b], and D. Thirumalai[a,c,1]

[a]Biophysics Program, Institute for Physical Science and Technology, [c]Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742; and [b]The Rockefeller University, 1230 York Avenue, New York, NY 10065

Edited* by José N. Onuchic, University of California at San Diego, La Jolla, CA, and approved May 28, 2010 (received for review March 18, 2010)

RNA synthesis, carried out by DNA-dependent RNA polymerase (RNAP) in a process called transcription, involves several stages. In bacteria, transcription initiation starts with promoter recognition and binding of RNAP holoenzyme, resulting in the formation of the closed ($R \cdot P_c$) RNAP-promoter DNA complex. Subsequently, a transition to the open $R \cdot P_o$ complex occurs, characterized by separation of the promoter DNA strands in an approximately 12 base-pair region to form the transcription bubble. Using coarse-grained self-organized polymer models of *Thermus aquatics* RNAP holoenzyme and promoter DNA complexes, we performed Brownian dynamics simulations of the $R \cdot P_c \rightarrow R \cdot P_o$ transition. In the fast trajectories, unwinding of the promoter DNA begins by local melting around the −10 element, which is followed by sequential unzipping of DNA till the +2 site. The $R \cdot P_c \rightarrow R \cdot P_o$ transition occurs in three steps. In step I, dsDNA melts and the nontemplate strand makes stable interactions with RNAP. In step II, DNA scrunches into RNA polymerase and the downstream base pairs sequentially open to form the transcription bubble, which results in strain build up. Subsequently, downstream dsDNA bending relieves the strain as $R \cdot P_o$ forms. Entry of the dsDNA into the active-site channel of RNAP requires widening of the channel, which occurs by a swing mechanism involving transient movements of a subdomain of the β subunit caused by steric repulsion with the DNA template strand. If premature local melting away from the −10 element occurs first then the transcription bubble formation is slow involving reformation of the opened base pairs and subsequent sequential unzipping as in the fast trajectories.

DNA scrunching | transcription initiation | self-organized polymer model | molecular simulations | sequential DNA unzipping

The DNA-dependent RNA polymerase (RNAP), whose sequence, structure, and functions are universally conserved from bacteria to man (1, 2), is the key enzyme in the transcription of the genetic information in all organisms (3–5). There are three major stages in the transcription cycle (3), which first requires binding of promoter-specific transcription factors to the catalytically-competent core of RNAP, to form a holoenzyme. They are: (*i*) initiation, which first requires binding of an initiation-specific σ factor to the catalytically-competent core RNAP to form the holoenzyme, followed by recognition of the promoter DNA to form the closed ($R \cdot P_c$) complex and subsequent spontaneous transition to the open ($R \cdot P_o$) complex; (*ii*) elongation of the transcript by nucleotide addition; (*iii*) termination involving cessation of transcription and disassembly of the RNAP elongation complex. Among these highly regulated stages, the most complicated may be the initiation process because it involves promoter recognition, DNA unwinding, and the formation of the transcription bubble inside the RNAP active-site channel, where RNA synthesis occurs.

A simplified transcription initiation pathway is (6, 7)

$$R+P \rightleftharpoons R\cdot P_c \rightleftharpoons R\cdot P_o \rightleftharpoons \mathrm{IC}_{\leq 12} \rightarrow \mathrm{TEC}, \qquad [1]$$

where R is RNAP, P is the promoter DNA (Fig. 1*A*), and $R \cdot P_c$ and $R \cdot P_o$ are the closed and open complexes (Fig. 1*B*), respectively. $\mathrm{IC}_{\leq 12}$ is the abortive initiation complex with transcript size ≤12 nt, and TEC is the transcription elongation complex. Here, we focus only on the dynamics of transcription bubble formation, which occurs during the $R \cdot P_c \rightarrow R \cdot P_o$ transition. The approximately 150 Å long and 110 Å wide core enzyme from the bacterial *Thermus aquaticus* (8, 9) has five subunits, αI, αII, β, β′, and ω (Fig. 1*B*) that are assembled like a crab claw. The two "pincers," formed from the large β and β′ subunits, hold the promoter DNA in the active-site channel between the pincers (Fig. 1). The structural model of RNAP holoenzyme complexes with fork-junction DNA (10) has given insights into the mechanism of the $R \cdot P_c \rightarrow R \cdot P_o$ transition. A variety of structures were pieced together to construct detailed models for the $R \cdot P_c$ and $R \cdot P_o$, which lead to the following mechanism of $R \cdot P_o$ formation (3, 10). Local melting of the promoter −10 element allows binding of the exposed nontemplate strand to conserved region 2.3 of σ ($\sigma_{2.3}$) (Fig. 1*B*), stabilizing the melted region. Melting also renders the promoter DNA flexible, thus facilitating its entry into the active channel.

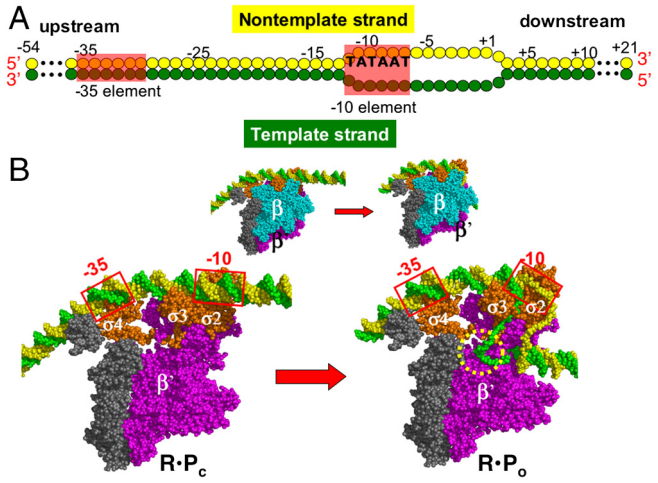

Fig. 1. Structural models for the promoter DNA and RNAP. (*A*) Schematics of the base pairing between the template (*green*) and nontemplate (*yellow*) strands of the promoter. Nucleotide positions are numbered relative to the transcription start site, +1. DNA segments that interact with RNAP, −35 and −10 elements, are shaded *red*. (*B*) Structural models on top correspond to $R \cdot P_c$ (*left*) and $R \cdot P_o$ (*right*) and are color-coded as β *cyan*, β′-*magenta*, σ-*orange*, αI, αII, and ω-*gray*, DNA nontemplate strand-*green*, and template strand-*yellow*. The β subunit is removed to show the interactions of the promoter with the σ subunits (*bottom left*) and the transcription bubble structure on the *bottom right*, while on the *top row* it is shown in full opacity.

Author contributions: D.T. designed research; J.C. performed research; D.T. contributed new reagents/analytic tools; J.C., S.A.D., and D.T. analyzed data; and J.C., S.A.D., and D.T. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: thirum@umd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1003533107/-/DCSupplemental.

The $R \cdot P_o$ structure further suggests that the promoter DNA bends into RNAP active channel to form the transcription bubble (11–13). Although the structural models provide plausible hypothesis for the transcription bubble formation, dynamical studies are required to describe the conformational changes that accompany the $R \cdot P_c \rightarrow R \cdot P_o$ transition.

Here, we use the structural models for $R \cdot P_c$ and $R \cdot P_o$ (updated from (10)) of *T. aquaticus* to address the following questions: (a) What are the steps in transcription bubble formation, and (b) What is the nature of RNAP dynamics that enables the downstream dsDNA entry into the enzyme? To answer these questions, we performed Brownian dynamics simulations of the $R \cdot P_c \rightarrow R \cdot P_o$ transition using a coarse-grained SOP model (14). The simulations of the kinetics of the $R \cdot P_c \rightarrow R \cdot P_o$ transition show that the transcription bubble forms in three distinct steps. In step I, a region near the −10 element on the promoter DNA melts. In step II, the promoter DNA scrunches into the RNAP active channel, forming the transcription bubble and in step III, the downstream DNA bends. DNA bending is a result of downstream DNA relaxation, and occurs only after unwinding of the dsDNA. Widening of the channel needed to accommodate the dsDNA entry into the active channel requires transient expansion of key structural elements in the β subunit of RNAP, which implies that the internal enzyme dynamics plays an important role in the $R \cdot P_c \rightarrow R \cdot P_o$ transition.

Results

A Network of Contacts Trigger the Promoter Melting. The structures of the RNAP-DNA complexes used in this work (10) have 3,122 residues and 150 nucleotides. The holoenzyme has six subunits: αI (Ala6-Glu229), αII (Ala6-Phe225), β (Ala2-Ala1116), β′ (Ala3-Ala1499), ω (Ala2-Ala93), and σ (Ala93-Ala438) (Fig. 1*B*). The promoter DNA has 75 base pairs, −54 to +21, labeled with respect to the transcription start site +1 (Fig. 1*A*). In $R \cdot P_c$, the promoter DNA "sits" on the top of RNAP (Fig. 1*B*) and forms stable interactions between the σ subunit and the −10 and −35 regions. Despite the low-resolution nature of the models, the dynamics reveal key structural changes that occur during the $R \cdot P_c \rightarrow R \cdot P_o$ transition. Several contacts between DNA −10 element and the subunit σ of RNAP rupture, in particular, the contacts involving nucleotides −12 to −8 on the template strand (Fig. S1). Formation of contacts between nucleotides −3 to +5 on the nontemplate strand and the β subunit (Fig. S1), nucleotides −9 to +1 on the template strand and the β subunit (Fig. S1), and nucleotides +1 to 7 on the template strand and the β′ subunit (Fig. S1) stabilize the $R \cdot P_o$ state.

$R \cdot P_c \rightarrow R \cdot P_o$ Transition Trajectories Partition into Fast (Efficient) and Slow (Inefficient) Tracks. The global nature of the $R \cdot P_c \rightarrow R \cdot P_o$ transition is monitored by the time-dependent changes of the root mean square deviation $\Delta_C(t)$ ($\Delta_O(t)$) of the DNA with respect to the closed (open) value. The transition times, τ_ms, calculated using $|\Delta_C(\tau_m) - \Delta_O(\tau_m)| < \epsilon = 0.1$ Å, were used to partition the set of 30 trajectories into fast and slow processes (Fig. S2). In the fast routes, $\Delta_C(t)$ ($\Delta_O(t)$) increases (decreases) very rapidly (Fig. S2), which indicates (see below) that upon opening of the DNA base pairs the transcription bubble forms efficiently. In contrast, in the slow routes, long-lived metastable states, which are indicated by a plateau in $\Delta_C(t)$ ($\Delta_O(t)$) (Fig. S2), are populated. The base pairs open by a complicated pathway resulting in a decreased efficiency in the transcription bubble formation. The conclusions do not depend on the value of the dielectric constant used in the treatment of the electrostatic interactions (see Eq. **S1** in Figs. S2 and S3).

What is the origin of the differences between the fast and slow trajectories? To answer this question, we examined the time-dependent conformational changes of the promoter DNA, which we describe using the distances, $d_i(t)$, between the two complementary nucleotides of each base pair in the promoter DNA. Here, $d_i(t) = |\vec{r}_i^{\,T}(t) - \vec{r}_i^{\,NT}(t)|$ where $\vec{r}_i^{\,T}(t)$ and $\vec{r}_i^{\,NT}(t)$ are the positions of the ith nucleotide on the template and the nontemplate strands, respectively. Since the transcription initiation site is at $i = +1$ and the transcription bubble forms between $i = -12$ and +2 (6), we computed $d_i(t)$ for four representative base pairs, −11, −7, −3, and +2 to describe the transcription bubble formation. In the $R \cdot P_c$ state, $d_i(t) \approx 11$ Å for all of the four base pairs, but in $R \cdot P_o$ state, $d_i(t) \approx 18$ Å, 48 Å, 53 Å, and 28 Å for −11, −7, −3, and +2 base pairs, respectively.

Promoter DNA Unzips Sequentially from −10 Element in the Fast Track. Analysis of $d_i(t)$ in all the fast trajectories shows a consensus sequence of events during the $R \cdot P_c \rightarrow R \cdot P_o$ transition. Base-pair −11 opens first at $t \sim 10$ μs, which is followed by disruption of interactions in −7 at $t \sim 16$ μs. In both cases the equilibrium values corresponding to the structure in $R \cdot P_o$ were reached rapidly (Fig. 2, Fig. S4). At $t \sim 24$ μs, the −3 pair rips and the distance between the nucleotides attains the value in the $R \cdot P_o$ state. The distance $d_{+2}(t)$ fluctuates between 11 Å and 30 Å, and reaches 30 Å shortly after base-pair −3 opens. The opening of the upstream base pairs favor the rupture of the downstream neighbors, which establishes that the base pairs from −12 to +2 rupture abruptly in a sequential manner by an unzipping mechanism. Complete analysis of all the $d_i(t)$s shows that sequential unzipping starting from the −10 site leads to rapid transcription bubble formation.

Initial Base Pair Opening Away from the −10 Element Results in Slow $R \cdot P_c \rightarrow R \cdot P_o$ Transition. Although promoter recognition sequences are localized in the −35 and −10 regions, melting of the base pair can occur stochastically (see also Fig. S5). However, as seen in the fast trajectories for efficient transcription bubble formation, rupture must start from the −10 element. In some trajectories, multiple base pairs melt nearly simultaneously in a non-sequential process, which greatly impedes transcription bubble formation. Fig. 2 shows that in one of the slow trajectories, rupture of base-pairs −3 and +2 compete with the opening of the base-pair −7. At $t \sim 40$ μs, base-pair −7 remains intact while the downstream base-pairs −3 and +2 rip as seen in the increase of $d_{-3}(40\ \mu s)$ and $d_{+2}(40\ \mu s)$ from 11 Å to 40 Å and 11 Å to 50 Å, respectively (Fig. 2). At longer times, the base-pairs −3 and +2 reform as shown by the decrease in d_{-3} and d_{+2} to 11 Å, the value in the $R \cdot P_c$ state, resulting in the "resetting" of the initial state. Subsequently, melting of the base pairs occur sequentially, which is manifested in the increase of $d_i(t)$ to the values in the $R \cdot P_o$ state.

Transcription Bubble Formation in Fast Trajectories Occurs in Three Steps. We used the time-dependent distance changes between the ith (jth) nucleotide on the promoter sequence and jth residue on RNAP, $d_{ij}(t) = |\vec{r}_i(t) - \vec{r}_j(t)|$, $\vec{r}_i$ ($\vec{r}_j$) is the position of the ith nucleotide (residue), to identify three steps in the transcription bubble formation process. The overall dynamics associated with the bubble formation is assessed using $d_{-5}(t)$, which on an average (*black line* in Fig. 3*A*) occurs in about (30–40) μs. Dissection of the events leading to the increase in $d_{-1}(t)$ from about 11 Å at $t = 0$ to 55 Å at $t \approx 35$ μs shows that transcription bubble formation occurs in three major steps.

Step I: The −10 element melts locally. The decreases in $d_{-5',Mg^{2+}}(t)$ (the prime refers to the base-pair number on the NT strand) (Fig. 3*B*) shows that the nontemplate strand moves towards the active site of the enzyme (amino acids around the Mg^{2+} ion). In this stage, $d_{-12,+21}$ also decreases (Fig. 3*B*), but in contrast to $d_{-5',Mg^{2+}}(t)$, the changes in the conformations of +21 base pair relative to the RNAP continues to evolve throughout the transcription bubble formation process (see below). Compared

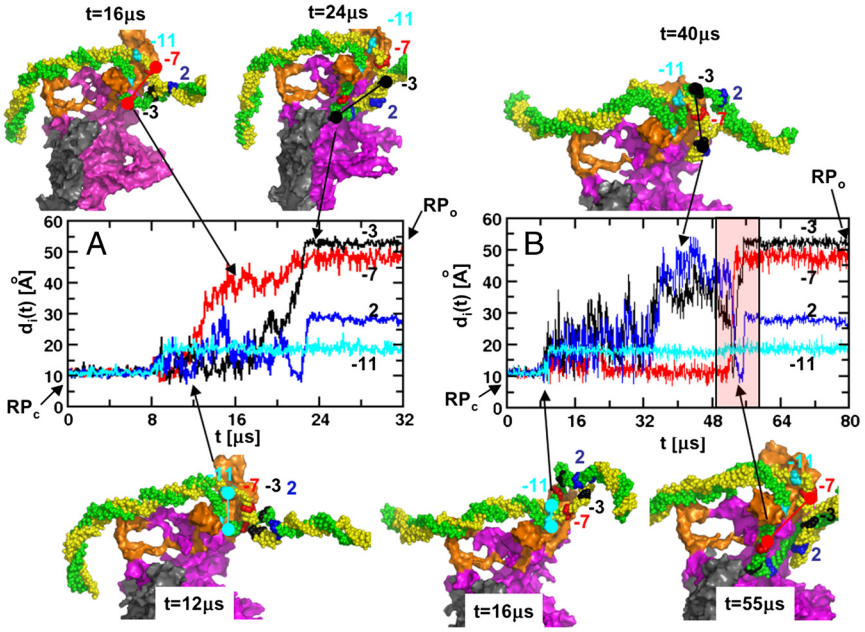

Fig. 2. Dynamics of transcription bubble formation. (*A*) Distance between complementary nucleotides of the base pairs in the bubble region as functions of time for a representative fast trajectory. *Cyan, red, black and blue lines* correspond to $d_i(t)$ for base pairs −11, −7, −3, and +2, respectively. Structures of the RNAP complex at $t = 12$ μs, 16 μs, and 24 μ during the development of the transition bubble are shown. (*B*) Same as (*A*) except the results are for a slow trajectory. The reformation of the prematurely opened base pairs −3 and 2 (*black and blue curves* respectively) is highlighted in the shaded region. The structures sampled during the $R \cdot P_c \rightarrow R \cdot P_o$ transition are highlighted. *Arrows* indicate the starting and ending states.

to the time needed for the bubble opening (Fig. 3*A*), the characteristic time associated with the decrease of $d_{-5',Mg^{2+}}(t)$ ($t \sim 16$ μs) is short, and is the major feature of step I. Although the template strand of the promoter DNA also moves synchronously in this step the dynamics of such a process lasts over the entire duration of bubble formation (at $t \sim 35$ μs). Therefore, we do not consider the movement of the template strand as a major feature of step I. The correlated movement of the nontemplate and the template strands is consistent with the assumption that local melting of DNA in the −10 element renders it flexible. In addition, the observed stabilization of the nontemplate strand in step I agrees with the experimental finding that the conserved aromatic residues of σ are positioned to recognize and stabilize the exposed nontemplate strand (15–17). Loss of base-pair interaction upon melting of the −10 base pairs (TATAAT) results in the formation of favorable interaction of adenine or thymine nucleotides with the aromatic side chains (Phe248, Tyr253, and Trp256) on σ2.3 subunit as well as electrostatic interactions between DNA and the enzyme.

Step II: DNA scrunches into the RNAP active-site channel and forms a bubble. The transcription bubble starts to grow from the −10 to +2 base pair as the promoter sequence unzips. The bubble region quantified in terms of center, $d_{-1}(t)$, increases from 11 Å to 55 Å, the value in the $R \cdot P_o$ state, which implies that in this step promoter DNA unwinds and the strands separate. Meanwhile, the distance between the −10 element and the downstream edge of the promoter DNA (base-pair +21) decreases from 110 Å to 55 Å (see the V-shape $d_{-12,+21}$ in Fig. 3*C*). The observed decrease in Fig. 3*C* during step II is reminiscent of the scrunching

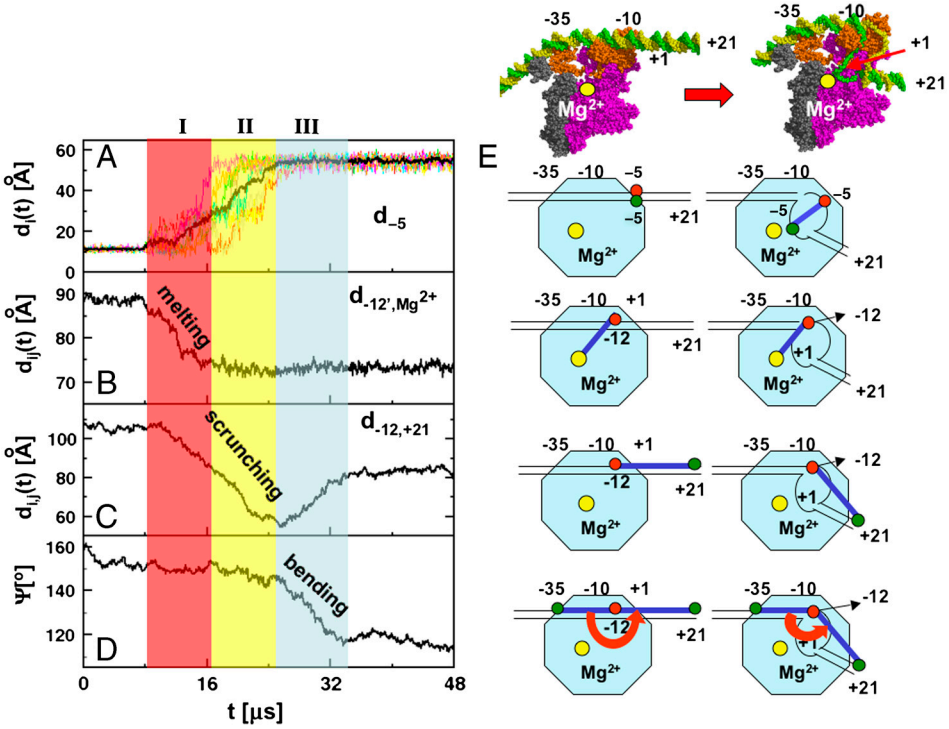

Fig. 3. Three steps (in shaded colors) in the transcription bubble formation in the fast trajectories. (*A*) The time-dependent increase in $d_{-5}(t)$ shows the growth of the transcription bubble. The *black line* shows $d_{-5}(t)$ averaged over multiple trajectories, and the individual traces for a few trajectories are shown for comparison. (*B*) Average changes in the distance between the nucleotide −12 on the nontemplate strand and Mg^{2+} as a function of time. (*C*) Ensemble average changes in $d_{ij}(t)$ for $i = -12$ and $j = +21$ as a function of *t*. (*D*) The time-dependent change in the angle Ψ between the interaction centers at −35, −10, and +21 shows that only after transcription bubble forms (see (*A*), the dsDNA bend (Ψ decreases in step III, which is shaded in l*ight blue color*). (*E*) Schematic representation of dynamical processes in (*A*)–(*D*). The sketches from *top to bottom* correspond to bubble growth ($d_{-5}(t)$ increase), promoter melting (decrease in $d_{-5,Mg^{2+}}(t)$), DNA scrunching (decrease in $d_{-12,+21}(t)$), and dsDNA bending (decrease in $\Psi(t)$). On *top right* are shown the front view of $R \cdot P_c$ (*left*) and $R \cdot P_o$ (*right*) structures without the β subunit and a few relevant nucleotides are labeled for reference.

mechanism proposed for the $R \cdot P_o \rightarrow R \cdot P_{itc}$ based on single molecule FRET experiments (10, 18). According to the scrunching mechanism, in each cycle of the abortive initiation, RNAP pulls the downstream DNA into the active channel and past its active center without substantial change in the enzyme conformation (18, 19). Although the $R \cdot P_{itc}$ transition formation occurs only after the completion of the $R \cdot P_c \rightarrow R \cdot P_o$ transition, the scrunching mechanism appears to also drive the formation of the transcription bubble (see below).

Step III: Downstream DNA bends. In the final stage, the downstream DNA bends to release the strain accumulated during the promoter DNA opening. The strain release results in the increase in $d_{-12,+21}(t)$ after the decrease in step II (Fig. 3*C*). In addition, DNA bends and kinks in the downstream region, which is captured by the time-dependent changes in the angle, $\Psi(t)$ (Fig. 3*D*). The angle $\Psi(t)$ formed between the interacting sites localized at −35, −12, and +21, which is relatively constant during step I and II, decreases in step III from ∼150° to ∼100°, the value in the RP_o state (11–13).

Our finding that DNA bending occurs only after the promoter DNA opening and bubble formation is in apparent conflict with the structural model that DNA bending facilitates entry into the active channel of the enzyme (3). Indeed, it is not realistic for DNA to bend or kink across the entrance of the active-site channel for the following reasons: (*i*) We expect that the flexibility in DNA, which is caused by local melting around −10 region, allows bending to be nearly isotropic restricted only by volume excluded by RNAP; therefore, it is entropically unlikely for DNA to bend in a specific direction and enter the active channel. (*ii*) Because DNA is negatively charged it should be repelled by the negatively charged σ 1.1 (20, 21). The location of σ 1.1 makes it unlikely for DNA to bend into the RNAP active channel prior to unzipping. Since the relaxation of the stressed DNA occurs after the opening of the promoter DNA, the observed DNA bending in the $R \cdot P_o$ state must be associated with the third stage of the bubble formation process.

Scrunching is a universal mechanism in transcription initiation. To illustrate whether the promoter DNA scrunches into RNAP in the $R \cdot P_c \rightarrow R \cdot P_o$ transition, we calculated the time-dependent distance changes, $d_{ij}(t)$, for the same sites that are labeled in the FRET study (18, 22) (Fig. 4). The conclusion that during the $R \cdot P_c \rightarrow R \cdot P_o$, the promoter DNA is scrunched follows the arguments used in the experimental study (18). From the decrease in the distance between the leading edge of RNAP (residue 366 of σ) and the +20 site in downstream DNA (Fig. 4*A*), it follows that RNAP translocates relative to downstream DNA. The lack of distance changes $d_{\sigma 569,-39}(t)$ and $d_{\sigma 569,-20}(t)$ between the upstream edge of RNAP (σ569) and upstream sites in DNA (−39 and −20) during the course of $R \cdot P_c \rightarrow R \cdot P_o$ transition (Fig. 4 *B* and *D*) implies absence of translocation of RNAP relative to upstream promoter sites. Similarly, $d_{\sigma 366,-20}$ is unchanged during the $R \cdot P_c$ to $R \cdot P_o$ transition, which implies that the leading edge of RNAP is stationary with respect to upstream DNA. Direct measurements of distance changes between site −15 and +15 on the promoter DNA shows that $d_{-15,+15}(t)$ decreases by about 5 Å when the $R \cdot P_c \rightarrow R \cdot P_o$ transition is completed. These results show that, to a large extent, RNAP is fixed on the promoter while the distances between downstream edge of DNA decrease.

The "V" shape portion of $d_{\sigma 366,+20}(t)$ and $d_{-15,+15}(t)$ (Fig. 4*A*, *E*) shows that the distance changes associated with leading edge and the specific sites on DNA do not change monotonically in $R \cdot P_c \rightarrow R \cdot P_o$ transition. The distance changes between labeled sites on promoter DNA and changes in $\Psi(t)$ (Fig. 3 *C*,*E*) show that the "V" shape of the distances is largely due to the contraction (or scrunching) of the DNA in the downstream region. The subsequent increase in the variables shown in Fig. 4*A*, *E* is due to the DNA bending.

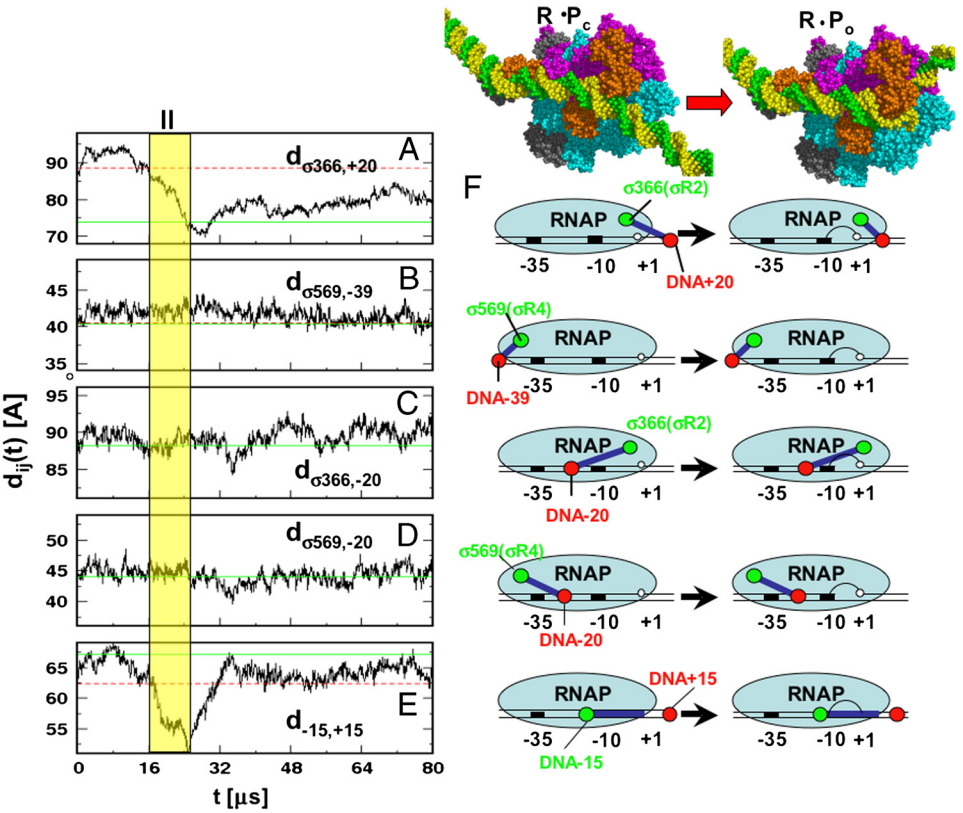

Fig. 4. Dissection of DNA scrunching during Step II. Structures on the *top right*, top view of $R \cdot P_c$ (*left*) and $R \cdot P_o$ (*right*) structures are shown for reference. The distances between nucleotides of the promoter DNA and RNAP as function of time, $d_{ij}(t)$ are on the *left*. In (*A*)–(*D*) the *red dashed lines* (*solid green lines*) are d_{ij} values in $R \cdot P_c$ ($R \cdot P_o$) state. (*A*) Ensemble averaged distance between nucleotide +20 and the leading edge residue σ366 (numbering as in *E. coli*). (*B*) Change in the distance between the trailing head residue σ569 and nucleotide −39 as a function of *t*. (*C*) Time-dependent change in the distance between the leading edge residue σ366 and the upstream nucleotide −20. (*D*) Distance between the trailing edge residue σ569 and nucleotide −20 as a function of *t*. (*E*) Average changes in $d_{-15,+15}(t)$ as a function of *t*. (*F*) The sketches from *top to bottom* capture the distance changes shown on the *left*. Structures on the *top right*, top view of $R \cdot P_c$ (*left*) and $R \cdot P_o$ (*right*) structures are shown for reference.

Active Channel of RNAP Opens by a "Rope-Swing" Mechanism. Despite the enhanced flexibility of the melted region of DNA, it is unclear how the promoter gains access to the binding sites in RNAP. The main DNA binding channel in the RNAP holoenzyme (10–12 Å in diameter for *Thermus thermophilus* and 16–19 Å for *Thermus aquaticus*) is narrow compared to the 22 Å diameter of the dsDNA. (9, 23). Based on static structures alone, two models have been proposed to explain how the RNAP channel expands to accommodate the dsDNA. One possibility is that the repulsive interactions between the negatively charged σ1.1 domain and the dsDNA result in the opening of the channel (3). Alternatively, based on a high resolution structure of the holoenzyme in the closed state, it was suggested that the protruding region of the β-subunit could fit into dsDNA major groove (23). In particular, it was argued that the β-subunit loop serves as a "gate" that allows only single stranded DNA to go through, and the "gate" stabilizes the DNA orientation to favor melting (23). However, in the absence of the crystal structure of the intermediate states of transcription bubble formation or experimental evidence, the mechanism whereby the RNAP channel opens remains elusive. Our simulations of the transcription bubble formation process suggest that channel opening occurs by a "rope-swing" mechanism (see below) and is triggered by transient movements of structural elements of the β subunit.

Analysis of the structures reveals that the Ala132-Ser387 region of the β-subunit makes extensive contacts with the promoter DNA in $R \cdot P_o$ but not in $R \cdot P_c$, which suggests that this region (referred to as the Rope-Swing (*RS*) region) must play an essential role in the channel widening process. To dissect the kinetic role of the *RS* region we partition it into the swing region (*S*), which is the N-terminal globular β2 domain (Ala132-Arg334) (24, 25) and the Rope (*R*) region, a part of the β1 domain corresponding to the C-terminal helix-turn-helix motif (Thr335-Ser387) (Fig. 5*A*). The *R* motif is the analogue of the helix-loop-helix fragment in *T. thermophilus* RNAP (23). The widening of the DNA binding channel is facilitated by the coordinated displacement of the *R* and *S* regions as shown by time-dependent changes in a number of quantities associated with the *R* and *S* regions as well as the β′ subunit. The closest distance, $R(t)$, between the *R* and *S* regions, whose value is 14 Å in $R \cdot P_c$, increases past the value in the final $R \cdot P_o$ state over a "gating" time interval $\Delta\tau_w$ that ranges from 19 μs to about 28 μs. Over the duration $\Delta\tau_w$ the changes in the distance $d_{AB}(t)$ (*A* is the center of mass of the *RS* structural unit and *B* is the center of mass of the residues Ala105–Ala499 of the β′ subunit) as a function of *t* shows (Fig. 5*C*) a total increase of 7 Å during the $R \cdot P_c \rightarrow R \cdot P_o$ transition. The 7 Å increase, which occurs transiently (Fig. 5 *B*, *C*), is sufficient to enable entry of the dsDNA. The change in $d_{AB}(t)$ is also accompanied by an increase of about 8° in the angle, $\Omega(t)$ (Fig 5*D*), between *A*, *O*, and *B* (*O* is the center of mass of residues Glu1264—Lys1426 in the β′ subunit) at $t \approx 24$ μs, which coincides with the increase in $d_{AB}(t)$ (Fig. 5*C*). Both $d_{AB}(t)$ and $\Omega(t)$ decrease after DNA enters RNAP, which results in a tight grip on the dsDNA. Surprisingly, the distance $d_{OB}(t)$ is almost constant (Fig. 5*C*). The totality of the results show that channel expansion occurs predominantly by swinging of the *RS* regions with the β′ subunit being stationary.

Conclusions

Using simulations we have shown that the formation of the transcription-competent $R \cdot P_o$ complex, which is known to occur through a series of intermediates (26, 27), proceeds in several steps and results in the transcription bubble formation. In fully accommodating the DNA, the internal dynamics of RNAP, which is hidden in many conventional experiments, plays a crucial role. Starting only with structural models of $R \cdot P_c$ and $R \cdot P_o$, our work shows that efficient transcription bubble formation requires sequential promoter opening (beginning at around −10 and proceeding downstream), which is supported by experimental data (4, 28, 29). The simulations also make detailed testable quantitative predictions, which can advance our understanding of the dynamics of the promoter melting process.

The proposed three-step formation of the transcription bubble is similar to the time-resolved hydroxyl radical foot printing experiments (26, 30), which characterized the formation of the complex between the T7A1 promoter and the *Escherichia coli* RNA polymerase. From the time-dependent protection against hydroxyl radicals, they surmised that there are three global steps in the formation of $R \cdot P_o$. In the first two stages the protection is attributed to interactions of melted DNA (template and nontemplate strands) with the RNAP, while the slower protection seen in the last stage is due to the entry of the dsDNA (+3 to +21) into RNAP. Our work suggests that the latter process is slow because it involves channel opening that is intimately coupled to the internal dynamics of the RNAP. However, a different study (27) suggested, using analysis of hydroxyl radical and potassium permanganate ($KMnO_4$) footprinting experiments, that the promoter ($\lambda_R P$) does not melt from −11 to +2 until after gaining entry

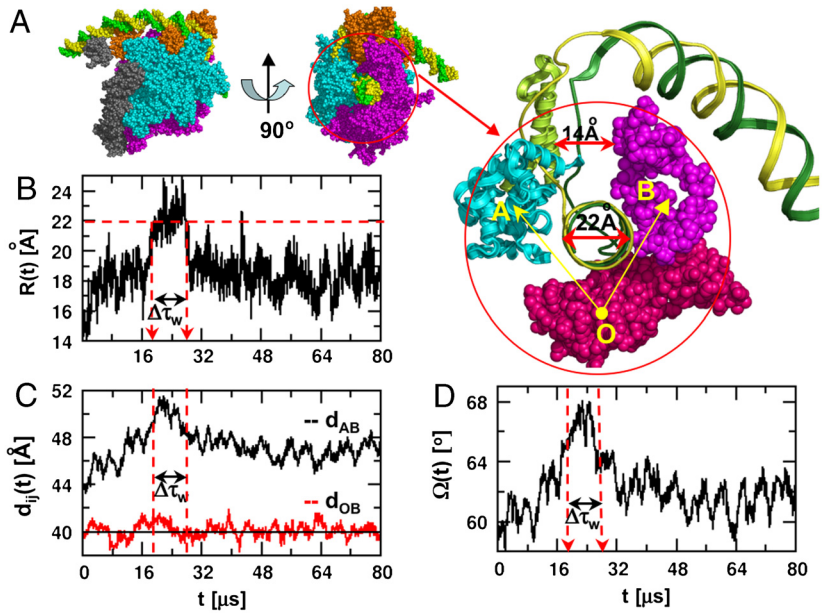

Fig. 5. Channel opening mechanism in fast trajectories. (*A*) View of $R \cdot P_o$. The view on the *right* is obtained by rotating the structure on the left by 90 ° about the vertical axis. The structural units that are close to downstream dsDNA are *circled*. The *RS* region and the promoter DNA are shown as ribbon diagrams. Two domains, Ala105-Ala499 and Glu1264-Lys1462 on the β′ subunit are displayed as *spheres* and colored *magenta* and *pink*, respectively. Residues on *RS* domain that make contact with the promoter DNA. *Red spheres* show that A, B, and O are centers of mass for *RS* domain, and the two selected domains of the β′ subunit. The distance between the closest residues on *RS* region and the β′ subunit (Pro244 and Ala122) is 14 Å, and the width of DNA is 22 Å. (*B*) The distance change between the closest points, residue Pro244 and Ala122 on the *RS* region and the β′ domain as a function of time averaged over ten fast trajectories. (*C*) Averaged distance changes d_{AB} (*black*) and d_{OB} (*red*) as a function of *t*. At $t \approx 19$ μs, $d_{AB}(t)$ increases, while $d_{OB}(t)$ is relatively constant. (*D*) Time-dependent changes in $\Omega(t) = \angle AOB$ (see structure on *top right* in (*A*)) show a transient increase at $t \approx 19$ μs. Results in (*C*) and (*D*) show that the *RS* domain (*R* in *yellowish green* and *S* in *blue ribbons*) swings away from the stationary β′ subunit (*spheres* in the structure on the *right* in (*A*)).

into RNAP. The dynamical structural and energetic changes in our simulations do not support this picture. In particular, the stabilization of melted single-strand DNA around the −10 element by the aromatic residues of σ 2.3 could also prevent $KMnO_4$ from reacting with the intermediate (31). If such a mechanism prevails then the complex formation mechanism would be similar to the present and previous (26) studies. Finally, the predictions linking the internal RNAP dynamics and the bending of the promoter can be tested using single molecule experiments.

Methods

Self-Organized Polymer (SOP) Model for $R \cdot P_c \rightarrow R \cdot P_o$ Transition. The large size of RNAP-DNA complex with 3,122 amino acids and a dsDNA with 150 nucleotides make it necessary to use a coarse-grained (32–35) models. Here we use a SOP model (14, 32, 34) for the enzyme and DNA. In the SOP model (32, 36), the structure of a protein is represented using only the C_α coordinates, $r_i^P (i = 1,2,\ldots N^P)$ with N^P being the number of amino acids). DNA is represented by the centers of mass of the nucleotides, $r_i^{DNA} (i = 1,2,\ldots N^{DNA})$ with N^{DNA} being the number of nucleotides. The state-dependent energy functions for RNAP, DNA, and protein-DNA interaction are given in *SI Text*. The proposed SOP force field yields accurate value of the persistence length (44.3 nm) of the isolated DNA. In addition, the calculated B factors for the core RNAP are in excellent agreement with the measured data based on crystal structures except for a few solvent-exposed residues that are not involved in promoter melting. These results are predictions of the model and not a consequence of adjusting the parameters to obtain agreement with experiments. The accurate description of the properties of the isolated DNA and the core enzyme and previous predictions for forced-unfolding of GFP (37) and rigor to postrigor transition in Myosin V (38) provide ample validation of the coarse-grained model.

Brownian Dynamics Simulation of the $R \cdot P_c \rightarrow R \cdot P_o$ Transition. The simulations of the $R \cdot P_c \rightarrow R \cdot P_o$ are based on the assumption that the local strain that triggers the open complex formation (due to DNA bending), propagates on a time scale that is faster across the structure than the time in which $R \cdot P_c \rightarrow R \cdot P_o$ transition occurs (32, 36, 39). The forces that trigger the $R \cdot P_c \rightarrow R \cdot P_o$ transition are computed from the energy function $H(r_i|R \cdot P_o)$, where the functional for $H(r_i|R \cdot P_o)$ is given by Eq. **S1** in *SI Text*. The trajectories are generated by integrating the Langevin equations of motion for $R \cdot P_c \rightarrow R \cdot P_o$ transition (see *SI Text* for details).

ACKNOWLEDGMENTS. This work was supported in part by a grant from the National Science Foundation (CHE09-14033).

1. Ebright RH (2000) RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol* 304:687–698.
2. Kornberg RD (2007) The molecular basis of eukaryotic transcription. *Proc Nat Acad Sci USA* 104:12955–12961.
3. Murakami KS, Darst SA (2003) Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol* 13:31–39.
4. deHaseth PL, Zupancic ML, Record MT, Jr (1998) RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J Bacteriol* 180:3019–3025.
5. Vassylyev DG (2009) Elongation by RNA polymerase: a race through roadblocks. *Curr Opin Struct Biol* 19:691–700.
6. Record MT, Jr, Reznikoff WS, Craig ML, McQuade KL, Schlax PJ (1996) *Escherichia coli* RNA polymerase ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcription onitiation. (ASM Press, Washington, D.C.), 2nd Ed, pp 792–820.
7. Bai L, Santangelo TJ, Wang MD (2006) Single-molecule analysis of RNA polymerase transcription. *Annu Rev Biophys Biomol Struct* 35:343–360.
8. Zhang GY, et al. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Angstrom resolution. *Cell* 98:811–824.
9. Murakami KS, Masuda S, Darst SA (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Angstrom resolution. *Science* 296:1280–1284.
10. Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296:1285–1290.
11. Rees WA, Keller RW, Vesenka JP, Yang GL, Bustamante C (1993) Evidence of DNA bending in transcription complexes imaged by scanning force microscopy. *Science* 260:1646–1649.
12. Rippe K, Guthold M, vonHippel PH, Bustamante C (1997) Transcriptional activation via DNA-looping: Visualization of intermediates in the activation pathway of *E-coli* RNA polymerase center dot sigma(54) holoenzyme by scanning force microscopy. *J Mol Biol* 270:125–138.
13. Rivetti C, Guthold M, Bustamante C (1999) Wrapping of DNA around the *E.coli* RNA polymerase open promoter complex. *Embo J* 18:4464–4475.
14. Hyeon C, Dima RI, Thirumalai D (2006) Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins. *Structure* 14:1633–1645.
15. Aiyar SE, Juang YL, Helmann JD, Dehaseth PL (1994) Mutations in sigma-factor that affect the temperature-dependence of transcription from a promoter, but not from a mismatch bubble in double-stranded DNA. *Biochemistry* 33:11501–11506.
16. Juang YL, Helmann JD (1994) A promoter melting region in the primary sigma-factor of *Bacillus subtilis*—identification of functionally important aromatic-amino-acids. *J Mol Biol* 235:1470–1488.
17. Rong JC, Helmann JD (1994) Genetic and physiological-studies of *Bacillus subtilis* sigma(a) mutants defective in promoter melting. *J Bacteriol* 176:5218–5224.
18. Kapanidis AN, et al. (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* 314:1144–1147.
19. Revyakin A, Liu CY, Ebright RH, Strick TR (2006) Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314:1139–1143.
20. Camarero JA, et al. (2002) Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc Natl Acad Sci USA* 99:8536–8541.
21. Dombroski AJ, Walter WA, Record MT, Siegele DA, Gross CA (1992) Polypeptides containing highly conserved regions of transcription initiation-factor sigma-70 exhibit specificity of binding to promoter DNA. *Cell* 70:501–512.
22. Tang GQ, Roy R, Ha T, Patel SS (2008) Transcription initiation in a single-subunit RNA polymerase proceeds through DNA scrunching and rotation of the N-terminal subdornains. *Mol Cell* 30:567–577.
23. Vassylyev DG, et al. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Angstrom resolution. *Nature* 417:712–719.
24. Lane WJ, Darst SA (2010) Molecular evolution of multisubunit RNA polymerases: sequence analysis. *J Mol Biol* 395:671–685.
25. Lane WJ, Darst SA (2010) Molecular evolution of multisubunit RNA polymerases: structural analysis. *J Mol Biol* 395:686–704.
26. Sclavi B, et al. (2005) Real-time characterization of intermediates in the pathway to open complex formation by *Escherichia coli* RNA polymerase at the T7A1 promoter. *Proc Natl Acad Sci USA* 102:4706–4711.
27. Davis CA, Bingman CA, Landick R, Record MT, Jr, Saecker RM (2007) Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by *Escherichia coli* RNA polymerase. *Proc Natl Acad Sci USA* 104:7833–7838.
28. Suh WC, Ross W, Record MT (1993) 2 open complexes and a requirement for Mg^{2+} to open the lambda-p(r) transcription start site. *Science* 259:358–361.
29. Chen YF, Helmann JD (1997) DNA-melting at the *Bacillus subtilis* flagellin promoter nucleates near −10 and expands unidirectionally. *J Mol Biol* 267:47–59.
30. Rogozina A, Zaychikov E, Buckle M, Heumann H, Sclavi B (2009) DNA melting by RNA polymerase at the T7A1 promoter precedes the rate-limiting step at 37 °C and results in the accumulation of an off-pathway intermediate. *Nucl Acids Res* 37:5390–5404.
31. Schroeder LA, et al. (2009) Evidence for a tyrosine-adenine stacking interaction and for a short-lived open intermediate subsequent to initial binding of *Escherichia coli* RNA polymerase to promoter DNA. *J Mol Biol* 385:339–349.
32. Hyeon C, Lorimer GH, Thirumalai D (2006) Dynamics of allosteric transitions in GroEL. *Proc Natl Acad Sci USA* 103:18939–18944.
33. Hyeon C, Onuchic JN (2007) Internal strain regulates the nucleotide binding site of the kinesin leading head. *Proc Natl Acad Sci USA* 104:2175–2180.
34. Hyeon C, Thirumalai D (2007) Mechanical unfolding of RNA: from hairpins to structures with internal multiloops. *Biophys J* 92:731–743.
35. Clementi C (2008) Coarse-grained models of protein folding: toy models or predictive tools? *Curr Opin Struct Biol* 18:10–15.
36. Chen J, Dima RI, Thirumalai D (2007) Allosteric communication in dihydrofolate reductase: signaling network and pathways for closed to occluded transition and back. *J Mol Biol* 374:250–266.
37. Mickler M, et al. (2007) Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations. *Proc Nat Acad Sci USA* 105:9604–9609.
38. Tehver R, Thirumalai D (2010) Rigor to post-rigor transition in Myosin V: link between the dynamics and the supporting architecture. *Structure* 18:471–481.
39. Koga N, Takada S (2006) Folding-based molecular simulations reveal mechanisms of the rotary motor F-1-ATPase. *Proc Nat Acad Sci USA* 103:5367–5372.