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.

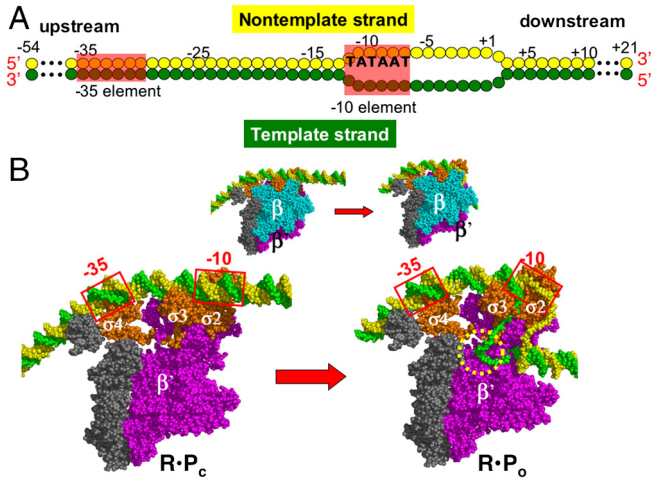

Fig. 1. Structural models for the promoter DNA and RNAP. (*A*) Schematics of the base pairing between the template (*green*) and nontemplate (*yellow*) strands of the promoter. Nucleotide positions are numbered relative to the transcription start site, +1. DNA segments that interact with RNAP, −35 and −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 $R \cdot P_c \rightarrow R \cdot P_o$

to the time needed for the bubble opening (Fig. 3*A*), the characteristic time associated with the decrease of $d_{-5'}$

mechanism proposed for the $R \cdot P_o \rightarrow R \cdot P_{\text{itc}}$ based on single

Active Channel of RNAP Opens by a "Rope-Swing" Mechanism. Despite the enhanced flexibility of the melted region of DNA,

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.