cades is their size: A three-input sorter implemented in next-generation (CMOS 9S) technology requires an area of 53 μm^2, whereas our cascade implementation uses only 200 nm^2, a factor of 260,000 difference. Even if CMOS density were to continue to double every 2.5 years, it would still take 45 years to shrink to the size of these cascades.

Conclusions. Molecule cascades provide new ways to study and exploit the motion of individual molecules in nanometer-scale structures. We have shown that below 6 K, the motion of CO molecules in our cascades is due to quantum tunneling. Our results at higher temperatures suggest that thermally assisted tunneling can play an important role in chemical kinetics when Arrhenius behavior exhibits anomalously low prefactors. The extreme sensitivity of the molecule hopping rate to the height and width of the energy barrier provides opportunities for probing the interactions between an adsorbate and its surroundings. Through exactly tailored model systems, detailed comparisons with theoretical calculations should be possible. The ability to engineer the direction and rate of molecular motion has enabled us to implement extraordinarily small (albeit exceedingly slow) logic circuits.

References and Notes

1. B. G. Briner, M. Doering, H.-P. Rust, A. M. Bradshaw, *Science* **278**, 257 (1997).
2. R. Baer, Y. Zeiri, R. Kosloff, *Surf. Sci.* **411**, L783 (1998).
3. L. J. Lauhon, W. Ho, *Phys. Rev. Lett.* **85**, 4566 (2000).
4. J. Kua, L. J. Lauhon, W. Ho, W. A. Goddard III, *J. Chem. Phys.* **115**, 5620 (2001).
5. J. D. Meindl, Q. Chen, J. A. Davis, *Science* **293**, 2044 (2001).
6. D. Goldhaber-Gordon, M. S. Montemerlo, J. C. Love, G. J. Opiteck, J. C. Ellenbogen, *Proc. IEEE* **85**, 521 (1997).
7. G. Y. Tseng, J. C. Ellenbogen, *Science* **294**, 1293 (2001).
8. S. J. Wind, J. Appenzeller, R. Martel, V. Derycke, Ph. Avouris, *Appl. Phys. Lett.* **80**, 3817 (2002).
9. Y. Huang *et al.*, *Science* **294**, 1313 (2001).
10. D. M. Eigler, C. P. Lutz, W. E. Rudge, *Nature* **352**, 600 (1991).
11. C. S. Lent, P. D. Tougaw, *Proc. IEEE* **85**, 541 (1997).
12. I. Amlani *et al.*, *Science* **284**, 289 (1999).
13. R. P. Cowburn, M. E. Welland, *Science* **287**, 1466 (2000).
14. R. S. Braich, N. Chelyapov, C. Johnson, P. W. K. Rothemund, L. Adleman, *Science* **296**, 499 (2002); published online 14 March 2002 (10.1126/science.1069528).
15. K. E. Drexler, *Nanosystems* (Wiley, New York, 1992).
16. D. D. Swade, *Sci. Am.* **268**, 86 (February 1993).
17. L. Bartels, G. Meyer, K.-H. Rieder, *Surf. Sci.* **432**, L621 (1999).
18. Mass spectroscopic analysis of the CO isotope distribution showed an upper bound of 2% for the total amount of ^{17}O and ^{18}O.
19. D. M. Eigler, E. K. Schweizer, *Nature* **344**, 524 (1990).
20. M. F. Crommie, C. P. Lutz, D. M. Eigler, *Science* **262**, 218 (1993).
21. G. Witte, *Surf. Sci.* **502–503**, 405 (2002).
22. L. Bartels, G. Meyer, K.-H. Rieder, *Appl. Phys. Lett.* **71**, 213 (1997).
23. A single CO molecule apparently decreases the local electronic density of states (LDOS) at the adsorption site at the Fermi level, which causes the feedback mechanism to move the STM tip closer to the surface in order to keep a constant tunnel current. In contrast, the dimer shows an increase in the LDOS between the two CO molecules.
24. The hopping rates were influenced by the presence of the tip (either slowed or sped up, depending on the tip position) even at low tunnel current ($I \approx 50$ pA) and bias voltage between $V = 5$ mV and $V = 50$ mV. The size of the influence depends primarily on the tip height (rather than on I or V) when V is kept well below the 35-meV vibrational excitation.
25. N. Knorr *et al.*, *Phys. Rev. B* **65**, 115420 (2002).
26. B. C. Stipe, M. A. Rezaei, W. Ho, *Science* **280**, 1732 (1998).
27. J. I. Pascual *et al.*, *Surf. Sci.* **502–503**, 1 (2002).
28. L. J. Lauhon, W. Ho, *Phys. Rev. B* **60**, R8525 (1999).
29. R. Raval *et al.*, *Surf. Sci.* **203**, 353 (1988).
30. J. Braun *et al.*, *J. Chem. Phys.* **105**, 3258 (1996).
31. The step in conductance is ~20% at $|V| = 35$ mV and ~15% at $|V| = 4$ mV for CO arranged in the $\sqrt{3}$ by $\sqrt{3}$ lattice, but only half as large for isolated CO molecules. We did not observe the 41-mV CO to Cu external stretch mode.
32. The exponent in the WKB approximation for the tunneling rate can be evaluated for one isotope by using the measured rate and assuming an attempt rate for tunneling on the order of the external vibrational frequencies ($A_{QT} = 5 \times 10^{12}\ s^{-1}$). The tunneling mass ratio is then determined with the rate for the other isotope.
33. Tunneling and classical over-the-barrier processes both depend on the stiffness of the full many-dimensional potential energy surface in the coordinates perpendicular to the reaction path. This makes it difficult to determine the value, and even the sign, of isotope shifts; for example, modes that are stiffer in the barrier than in the initial state favor passage of heavier isotopes.
34. R. Baer, Y. Zeiri, R. Kosloff, *Phys. Rev. B* **54**, R5287 (1996).
35. S. Ovesson *et al.*, *Phys. Rev. B* **64**, 125423 (2001).
36. J. V. Barth, H. Brune, B. Fischer, J. Weckesser, K. Kern, *Phys. Rev. Lett.* **84**, 1732 (2000).
37. J. V. Barth, *Surf. Sci. Rep.* **40** 75 (2000).
38. G. X. Cao, E. Nabighian, X. D. Zhu, *Phys. Rev. Lett.* **79**, 3696 (1997).
39. P. J. Price, *Am. J. Phys.* **66**, 1119 (1998).
40. F. Wittl *et al.*, *J. Chem. Phys.* **98**, 9554 (1993).
41. C. L. Seitz, in *Introduction to VLSI Systems*, C. A. Mead, L. A. Conway, Eds. (Addison-Wesley, Reading, MA, 1980), p. 218–262.
42. R. Landauer, *Physica A* **168**, 75 (1990).
43. Chain polymerization initiated with an STM tip has been demonstrated at room temperature (*44*).
44. Y. Okawa, M. Aono, *J. Chem. Phys.* **115**, 2317 (2001).

45. We thank J. P. Sethna for useful discussions and preliminary analysis of these results and G. Northrop for his design of a three-input sorter implemented in CMOS 9S technology.

30 July 2002; accepted 4 October 2002
Published online 24 October 2002;
10.1126/science.1076768
Include this information when citing this paper.

Structural Basis for the Transition from Initiation to Elongation Transcription in T7 RNA Polymerase

Y. Whitney Yin[1] and Thomas A. Steitz[1,2,3]*

To make messenger RNA transcripts, bacteriophage T7 RNA polymerase (T7 RNAP) undergoes a transition from an initiation phase, which only makes short RNA fragments, to a stable elongation phase. We have determined at 2.1 angstrom resolution the crystal structure of a T7 RNAP elongation complex with 30 base pairs of duplex DNA containing a "transcription bubble" interacting with a 17-nucleotide RNA transcript. The transition from an initiation to an elongation complex is accompanied by a major refolding of the amino-terminal 300 residues. This results in loss of the promoter binding site, facilitating promoter clearance, and creates a tunnel that surrounds the RNA transcript after it peels off a seven–base pair heteroduplex. Formation of the exit tunnel explains the enhanced processivity of the elongation complex. Downstream duplex DNA binds to the fingers domain, and its orientation relative to upstream DNA in the initiation complex implies an unwinding that could facilitate formation of the open promoter complex.

Despite structural differences, the 99-kD single-subunit RNA polymerase from the bacteriophage T7 (T7 RNAP) and the multisubunit cellular RNAPs share numerous functional characteristics. Both families of RNAPs have initiation and elongation phases of transcription (*1*, *2*). During the initiation phase, RNAP binds to a specific DNA promoter, opens the duplex at the transcription start site, and initiates RNA synthesis de novo. Transcription during this phase is unstable and characterized by repeated abortive initiation events that produce short RNA fragments [2 to 6 nucleotides (nt)] (*3*, *4*). After synthesis of 10- to 12-nt-long RNA, the polymerase enters the elongation phase and completes transcription of the mRNA processively without dissociating until termination. There are significant biochemical differences between the initiation and elongation states. Footprinting assays show differences in DNA protection (*5–7*). The T7 RNAP-DNA complex is substantially more stable in the elongation phase (*4*, *8*, *9*),

[1]Department of Molecular Biophysics and Biochemistry, [2]Department of Chemistry, [3]Howard Hughes Medical Institute, Yale University, 266 Whitney Avenue, New Haven, CT 06520–8114, USA.

*To whom correspondence should be addressed. E-mail: eatherton@csb.yale.edu

and T7 lysozyme, a natural inhibitor of T7 RNAP, inhibits the transition from the initiation to the elongation state but has little effect on the activity of the transcribing elongation complex (*10*, *11*). Here, we describe the structural basis for promoter opening, the transition from the abortive initiation to processive elongation phases, promoter clearance, the regulation by T7 lysozyme, and the unwinding of downstream DNA.

The structure of T7 RNAP was largely unchanged whether complexed either with the transcription inhibitor T7 lysozyme (*12*), a 17–base pair (bp) open promoter DNA (*13*), or with a 17-bp promoter containing a 5′ template extension of 5 nt and a 3-nt RNA transcript (*14*); the COOH-terminal two-thirds of T7 RNAP is homologous to the polymerase domain of the Pol I family DNA polymerase (*15*–*18*), whereas a novel NH_2-terminal domain (residues 1 to 325) is unique to the RNAP. The structures show that one antiparallel β loop, named the specificity loop (residues 740 to 770), makes sequence-specific contacts with the promoter, whereas another, the intercalating hairpin (residues 230 to 240), opens the upstream end of the transcription bubble (*13*, *14*). The structures of the two promoter-containing complexes also provide support for the "scrunching" model of transcription initiation, in which RNA synthesis leads to an accumulation of the DNA template within the active site before the promoter is released (*14*).

Nevertheless, the structure of the transcribing initiation complex could not explain many aspects of the elongation phase. Extensive proteolysis that results in loss of the NH_2-terminal 180-residue fragment abolishes the elongation phase; although the polymerase can initiate transcription from a promoter, it makes only abortive transcripts (*19*). A proteolytic cleavage of T7 RNAP after residue 173 or 180 results in somewhat decreased efficiency of elongation and decreased single-strand RNA binding, which suggests that the integrity of the region between residues 173 to 180 plays a role in elongation (*19*, *20*). However, because the proteolytically nicked region is located at least 40 Å away from the 5′ end of the RNA transcript in the initiation complex structure, it was not clear how this remote site could affect elongation transcription. Similarly, a single mutation from Glu^{148} to Ala^{148} (E148A), which abolished synthesis of any transcript longer than 5 nt (*21*), is located at least 35 Å away from the 5′ end of the mRNA in the initiation complex structures.

The most puzzling paradox, however, arose from the apparent incompatibility of the biochemical and structural evidence for the maximum length of the DNA-RNA heteroduplex during transcription. The structure of the initiation complex contained only 3 bp of DNA-RNA heteroduplex, and, indeed, any extension of the heteroduplex beyond 3 bp was deemed to be sterically excluded by the protein structure (*14*). In contrast, numerous biochemical studies, including a recent cross-linking of the RNA and DNA strands, led to the conclusion that the length of the DNA-RNA heteroduplex during elongation is about 8 bp (*7*, *22*–*24*). Various attempts (*24*) to accommodate these data with the use of existing T7 RNAP structures seemed implausible. We show here that these apparent conflicts arose because an important piece of the puzzle in understanding the transition from transcription initiation to elongation by T7 RNAP was missing.

The crystal structure of a T7 RNAP complex trapped in a functional elongation mode with a transcription bubble of DNA and heteroduplex RNA 7-nt long shows that the NH_2-terminal domain changes conformation substantially, as compared to the structure of the initiation complex. As a consequence of this change, the promoter binding site is destroyed, and a channel that accommodates the heteroduplex in the active site and an exit tunnel through which RNA can pass are created. These structural features account for promoter clearance and processivity in the elongation phase.

Structure of an elongation complex. The T7 RNAP was cocrystallized with DNA containing a transcription bubble and mRNA, a complex that mimics the elongation phase, and its structure was determined at 2.1 Å resolution. The 30-bp duplex DNA contains a central region of 11 noncomplementary bases and a 17-nt RNA that is complementary to the template for 10 nt at its 3' end (Fig. 1A). The RNA of this substrate can be extended by the polymerase in a template-directed and processive manner in the absence of promoter (*25*), and it possesses other features of a normal promoter-initiated elongation complex, as seen in the earlier work of von Hippel and Daube (*26*, *27*). The T7 RNAP elongation complex was assembled by mixing the polymerase with the substrate after the three oligonucleotide strands of template DNA, nontemplate DNA, and RNA had been annealed (*28*). The structure was determined by single-wavelength anomalous diffraction using selenomethionine-substituted T7 RNAP and by molecular replacement using the polymerase domain of the T7 RNAP initiation complex as a search model. The phases derived from these two sources were weighted and combined. Density modification was applied to the initial electron density map calculated with combined phases to further reduce the phase errors and improve the map (Fig. 1B) (*29*). The final refined model has an R_{factor} of 24.1% (R_{free} = 27.3%). The data collection and refinement statistics are provided in Table 1.

The T7 RNAP protein is seen bound to the duplex DNA and its RNA transcript annealed to a central, opened section of DNA in the active site (Fig. 1B). The active site is located in an enlarged channel bounded by the polymerase's fingers, thumb, and palm domains of the COOH-terminus. Although the annealed construct contained 10 nt of complementary heteroduplex RNA-DNA by design, the elongation complex active site contains only 7 bp of heteroduplex DNA-RNA. After 7 bp, the 5′ end of the mRNA is separated from the template by an α helix of the thumb domain and enters a positively charged tunnel

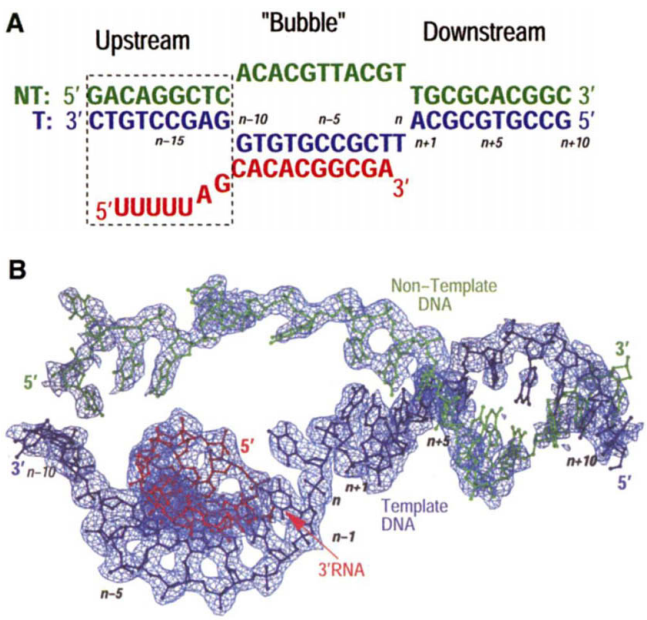

Fig. 1. Substrates in the T7 RNAP elongation complex. (**A**) The substrate construct co-crystallized with T7 RNAP consisted of 30 nt each of template DNA (blue) and nontemplate DNA (green) that are complementary (except for a central 11 nt) and a 17-nt RNA (red) whose 3′ 10 nt are complementary to the template DNA. The nucleotide that templates the nascent NTP is numbered *n* and nt upstream are given numbers $n - i$ and downstream given numbers $n + i$ ($i \geq 1$). The portions of RNA and DNA that are not visible in the map are outlined by dashed lines. (**B**) A portion of the composite omit electron density map corresponding to the substrate is contoured at 1.1 σ. Figs 1 and 4 were made with the program SPOCK (*58*).

in the protein, while the template strand remains bound to the thumb domain. The single-stranded, nontemplate DNA in the bubble is separate from the template DNA-RNA heteroduplex and makes sequence-independent contacts with the protein. The template and nontemplate strands merge to form duplex DNA at both the upstream and downstream ends of the bubble (Fig. 1B). Only 3 nt of DNA upstream of the DNA-RNA heteroduplex are visible in the electron density map. At the downstream end of the bubble, the nontemplate DNA strand base-pairs with the template strand at a position that is one nucleotide beyond the incoming nucleotide (n + 1), and all 10 downstream base pairs are clearly visible in a positively charged cleft formed by the fingers domain. Overall, the 3′-terminal 10 nt of RNA as well as 21 nt each of template and nontemplate DNA are visible in the electron density maps, whereas 9 bp of upstream DNA duplex and the 5′-terminal 7 nt of single-stranded RNA are not (Fig. 1A). The 7-bp RNA-DNA heteroduplex seen in this structure is in good agreement with the ~8-bp length derived by biochemical studies (*7*, *22*–*24*). The total of 21 nt of DNA seen in the structure also agrees with the 21- to 24-nt length of DNA protected by the polymerase in footprinting experiments of an elongation complex (*5*, *7*, *20*).

This elongation complex is analogous to the binary complexes of primer-template DNA with DNA polymerases; the primer terminus is located at the posttranslocation position ready to accept an incoming nucleotide. In this complex, as with the corresponding DNA polymerase complexes (*30*), the base that is to form the template for the incoming nucleotide lies in a pocket in the fingers domain, rather than in an orientation that would allow it to pair with the incoming nucleoside triphosphate (NTP). The position of the primer terminus relative to the palm domain also is identical to its position in the transcription initiation complex lacking the NTP (*14*). We have additionally determined the structure of an elongation complex after insertion of the nucleotide at the n position, yielding an 8-bp heteroduplex and insights into the mechanism of translocation (*31*).

T7 RNAP structural transition from initiation to elongation. A comparison of the polymerase structure in the initiation complex with its structure in the elongation complex shows that portions of the enzyme, most notably the NH_2-terminal domain, have undergone major conformational changes that alter its shape and tertiary structure (Fig. 2). The structural changes seen in the NH_2-terminal domain involve three different regions, each undergoing a different kind of conformational alteration: (i) Rigid body motion. Six α helices, D, E, F, G, I, and J, together with the intercalating hairpin, rotate by 140° and translate by 30 Å as a rigid body from their location in the initiation complex (Fig. 2F). The ends of helices F and G pack against the third bp of heteroduplex in the initiation complex and must move to accommodate a longer heteroduplex. The six helices are repositioned into the region that is occupied by the promoter DNA in the initiation complex,

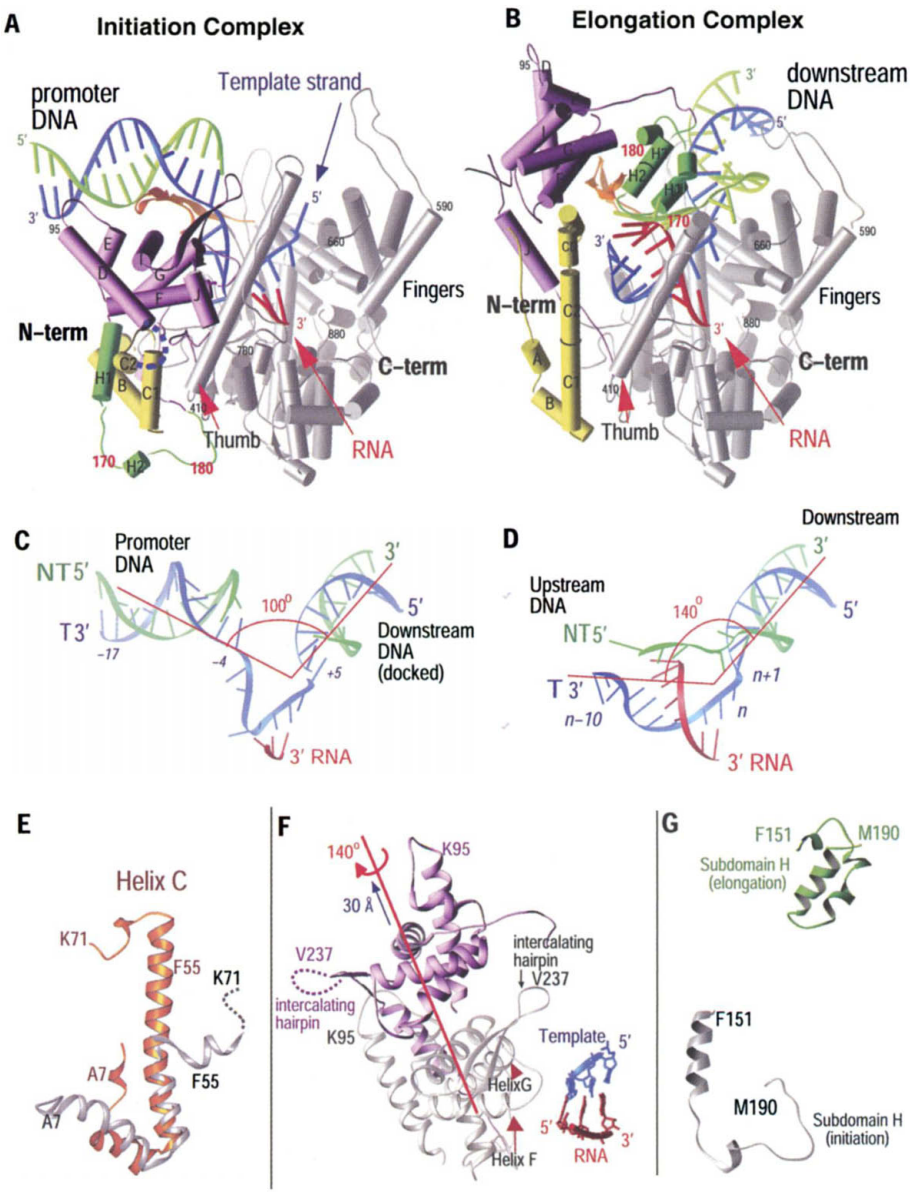

Fig. 2. Comparison of the structures of the T7 RNAP initiation and elongation complexes. The initiation complex (**A**) and elongation complex (**B**) have been orientated equivalently by superimposing their palm domains. Helices are represented by cylinders and β strands by arrows. The corresponding residues in the NH_2-terminal domains of the two complexes that undergo major refolding are colored in yellow, green, and purple, and the COOH-terminal domain (residues 300 to 883) is colored in gray. The template DNA (blue), nontemplate DNA (green), and RNA (red) are represented with ribbon backbones. The proteolysis-susceptible region (residues 170 to 180) is a part of subdomain H (green) in the elongation complex and has moved more than 70 Å from its location in the initiation complex. The specificity loop (brown) recognizes the promoter during initiation and contacts the 5′ end of RNA during elongation, whereas the intercalating hairpin (purple) opens the upstream end of the bubble in the initiation phase and is not involved in elongation. The large conformational change in the NH_2-terminal region of T7 RNAP facilitates promoter clearance. (**C**) The DNA and trinucleotides of RNA seen in the structure of the initiation complex docked to the downstream DNA of the elongation complex shows a 100° bend between upstream and downstream segments. (**D**) The DNA and 7 nt of RNA observed in the structure of the elongation complex show a decreased angle of bending between the base-paired upstream and downstream segments. Panels (**E**), (**F**), and (**G**) show the three different conformational changes—helix formation, rigid body movement, and refolding—undergone in the transition between initiation and elongation. This figure and Figs. 3, 5, and 6 were made with the program Ribbons (*60*).

thereby abolishing the interaction between T7 RNAP and the promoter and thus explaining promoter clearance. The intercalating hairpin, which opens the promoter in the initiation complex, moves and becomes disordered in the elongation complex, accounting for mutations in this loop that reduce the efficiency of initiation but not elongation (*32*, *33*). (ii) Extension of an α helix. In a conformational change that is reminiscent of one undergone by the influenza virus hemagglutinin protein upon a pH change (*34*), helix C1 becomes significantly elongated from 22 to 50 Å by the stacking of helix C2 on top of C1 and the refolding of the disordered loop between C2 and helix D to further elongate the C1-C2 helix (Fig. 2E). The elongated C helix now protrudes into the region formerly occupied by the six-helices assembly in the initiation complex, implying that the two conformational changes are likely to be coordinated. (iii) Formation of antiparallel α helices. Perhaps the most unprecedented conformational change involves residues 160 to 190, which not only extensively refold, but move about 70 Å from one side of the polymerase to the other (Fig. 2G). This region refolds from a short helix and an extended loop into a pair of antiparallel helices (H1 and H2/3) (Fig. 2; A, B, and G). The newly formed compact structure, named subdomain H, forms part of the RNA-transcript exit tunnel and contacts the 5′ end of the RNA transcript on one surface and the nontemplate DNA on the opposite surface.

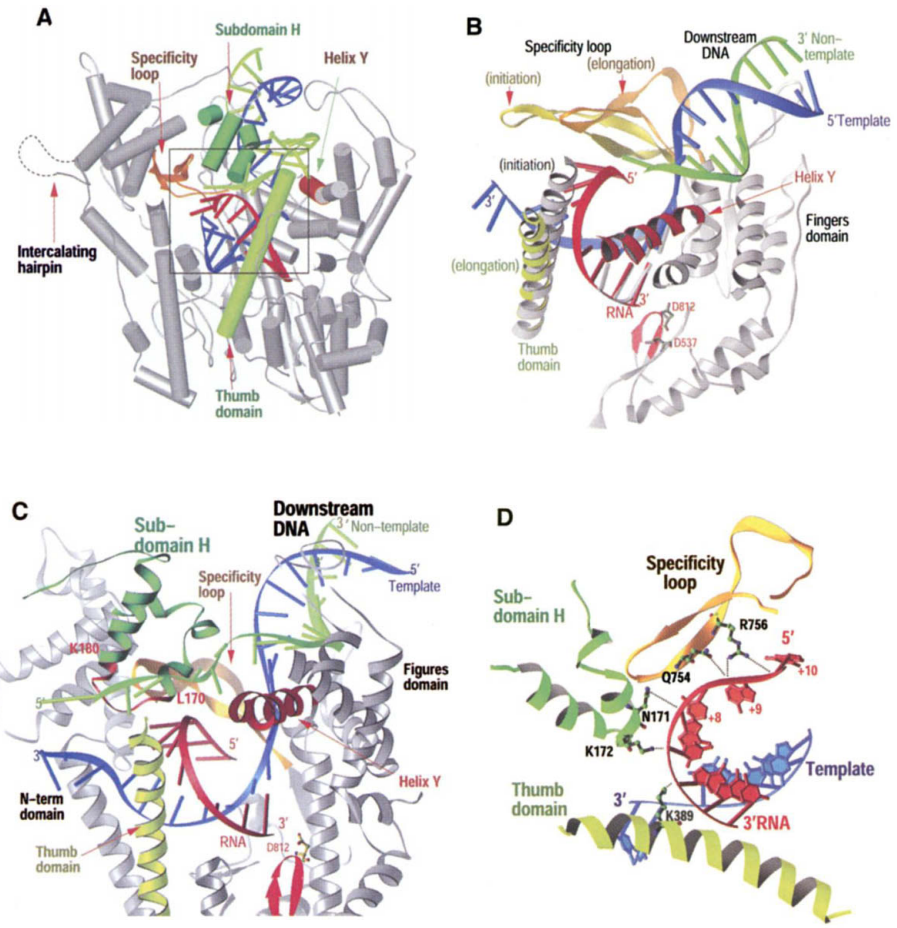

Fig. 3. Views of the transcription bubble. (**A**) Global view of the elongation complex with a box outlining the active site region that is enlarged in (B), (C), and (D) with the thumb (yellow green), subdomain H (green), specificity loop (yellow), and helix Y (red). (**B**) Conformational changes of the thumb and the specificity loop. The thumb domain as observed in the initiation complex (gray) has rotated about 15° in the elongation complex (green) and assists in the separation of the RNA transcript from the template DNA. The position of the specificity loop in the initiation complex (yellow) blocks the exit of RNA and has moved in the elongation complex (brown) to open the exit tunnel and interact with the exiting RNA. The 3 bp of heteroduplex in the initiation complex (gray) superimposes on that of the elongation complex. (**C**) Interactions of the transcription bubble and heteroduplex in the elongation complex with domain H (green and red) and specificity loop (brown). Proteolytic cuts within the red loop in subdomain H reduce elongation synthesis (*19*, *20*). Thumb σ helix (yellow) and σ-helix Y (orange) are analogously involved in strand separation. (**D**) Side chains from subdomain H (green), the specificity loop (brown), and the thumb that interact with the single-stranded 5′ end of the RNA transcript and facilitate its separation from the template.

In contrast with the NH_2-terminal domain, the COOH-terminal domain undergoes fewer structural changes (Fig. 3B). The thumb domain rotates by about 15° from its orientation in the initiation complex and, together with the subdomain H, creates a binding cleft for the nontemplate strand DNA (Fig. 3, C and D). In the initiation complex, the specificity loop crosses the active site to make sequence-specific interactions in the major groove of the promoter DNA (*13*, *14*), and it lies in a position that blocks the path through which RNA exits in the elongation complex. In the elongation complex, the specificity loop moves sideways to open the exit tunnel and to become a part of it. The tip of the specificity loop, which contacts promoter DNA in the initiation complex, then contacts the 5′ end of the mRNA in its passage through the tunnel (Figs. 3D and 4), consistent with the observation that the RNA transcript can be photocross-linked to the specificity loop (*24*). This conformational change of the specificity loop may also be associated with promoter release.

The mutation E148A, which lies remote from substrates in the initiation complex, abolishes synthesis of transcripts longer than 5 nt (*21*). This may be due to the inability of the mutant Ala^{148} to make interactions necessary to the structural transition of the specificity loop in forming an elongation complex. Glu^{148} stacks directly against Met^{750} and interacts indirectly with Asn^{748} at the tip of the specificity loop through Arg^{155} to bend the specificity loop toward the 5′ end of the RNA. The mutant Ala^{148} cannot make the interactions to secure the bending configuration of the specificity loop and therefore would affect the integrity of the exit tunnel. Consequently, both structural and biochemical studies agree on these dual functions of the specificity loop.

This massive structural reorganization of the NH_2-terminal domain upon formation of the elongation complex creates a tunnel through which the RNA can exit and a binding site for the single-stranded nontemplate DNA of the transcription bubble from $n - 7$ to n. The newly formed exit tunnel, whose interior is positively charged, measures about 8 Å in diameter and 20 Å in length and is formed by the thumb domain, the specificity loop, and subdomain H (Fig. 3, C and D, and Fig. 4). After 7 bp of heteroduplex, 3 nt of RNA are separated from the DNA by the rim of the exit tunnel (residues 170 to 180) and the thumb domain (Fig. 3C). The single-stranded 5′ end of the RNA transcript is seen entering the exit tunnel (Fig. 4, B and D). Model-building suggests that the tunnel may accommodate five extended nucleotides, implying that an RNA transcript longer than 12 nt would emerge from the side of the tunnel opposite the active site. The exit tunnel contacts the RNA strictly through interactions with the phosphates of the sugar-phosphate backbone.

Residues Arg^{756} and Gln^{754} from the specificity loop, as well as Asn^{171} and Lys^{172} on the subdomain H, are all within hydrogen-bonding distance of phosphates at the 5' end of the single-stranded mRNA (Fig. 3D).

The processivity of the elongation complex, in contrast to the initiation complex, could be explained by the appearance of the mRNA exit tunnel, which topologically surrounds transcripts longer than 8 to 12 nt. Its effect on processivity is entirely analogous to that of the sliding clamp on DNA replication (*35*). Further, the stability of the elongation complex, as compared to the abortive phase complex, is enhanced by the extensive interaction between the 7 bp of heteroduplex and its binding site. That proteolytic cleavage of residues between 173 and 180 results in reduced elongation efficiency (*19*, *20*) may be a consequence of a reduced integrity of subdomain H and a reduced stability of the tunnel in the elongation phase. Complete proteolytic digestion of the NH_2-terminal 180 residues results in a T7 RNAP capable of abortive synthesis and incapable of elongating transcripts beyond 8 nt (*18*, *19*). Because this enzyme would be missing helices D to G and subdomain H, it may be incapable of destroying the promoter binding site, which is required for clearance, and may be unable to form the RNA exit tunnel required for processive synthesis.

During the transition from initiation to elongation, T7 RNAP relinquishes its sequence-specific grasp of the promoter and begins translocation along DNA, a process often referred to as promoter clearance, which is achieved by the destruction of the promoter binding site and movement of the six helices by 30 Å into the position formerly occupied by the promoter DNA (Fig. 2). Upstream DNA now binds in a sequence-independent manner to a newly created cleft that is formed in part by the thumb domain and helix C2 (Figs. 2B and 4B). These two upstream DNA binding sites are separated by at least 40 Å and cannot be occupied simultaneously by upstream DNA because formation of one binding site dismantles the other. The DNA that is upstream of the transcription bubble and visible in our complex is not base-paired due to the noncomplementarity of the designed sequence. Presumably, the upstream DNA of complementary sequence would form a duplex that would lie in the upstream channel.

What causes the T7 RNAP to undergo such conformational change and what stabilizes the elongation phase structure? Because the apoenzyme has essentially the same structure as the initiation complex structure (*13*, *14*), it seems likely that the formation of the longer RNA:DNA heteroduplex is playing a critical role. We previously noted that it was not possible to elongate the heteroduplex beyond 3 bp in the initiation complex because of a steric clash with helices F and G (*14*). Here, we observe that the 3-bp heteroduplex in the initiation complex and the 7-bp heteroduplex in the elongation complex are bound and oriented identically on the polymerase active site (Fig. 3B). When the C_α backbone atoms of the palm domains of the initiation and elongation complexes are superimposed, the first 2 bp of the initiation and elongation complexes superimpose with an root means squared (rms) deviation of 0.34 Å (Fig. 3B). Thus, it seems likely that incorporation of a fourth nt into the transcript would result in a steric clash and a destabilization of the initiation complex structure. Indeed, our attempts to elongate the 3-nt transcript by 1 nt destroyed transcribing crystals of the initiation complex. However, DNA protection experiments suggest that T7 RNAP is still bound to the promoter in the presence of a 6-nt transcript (*5*), which is inconsistent with the complete conversion of T7 RNAP to the elongation complex structure seen here. Taken together, the structural

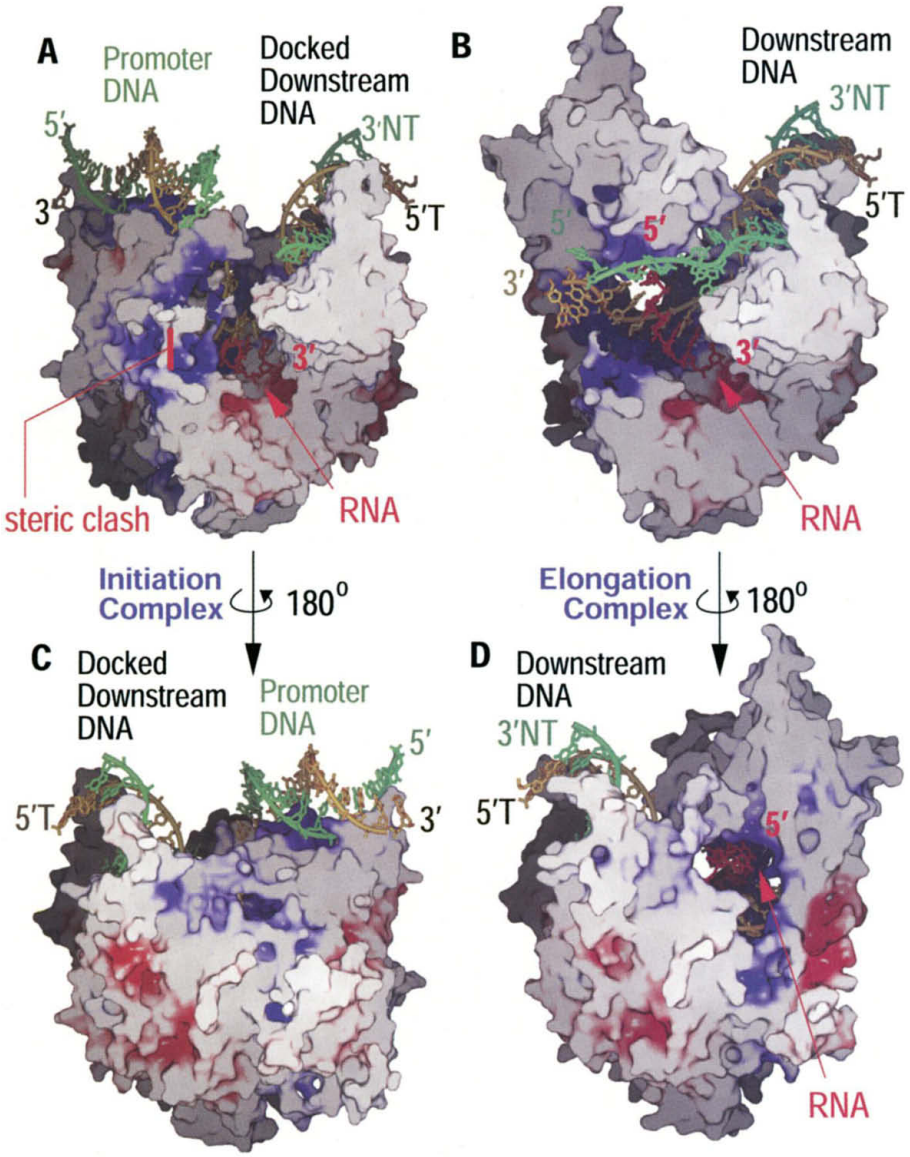

Fig. 4. Formation of the RNA exit tunnel and upstream DNA binding site in the elongation complex. The solvent contact surface for the initiation complex conformation of T7 RNAP (**A**) with the observed upstream promoter DNA and heteroduplex along with the downstream DNA modeled from the elongation complex. The thumb domain has been removed from both (A) and (B) to allow a view into the heteroduplex binding site. The elongation complex (**B**) shows the disappearance of the promoter DNA binding site, the formation of a new channel that binds to heteroduplex and upstream DNA nonspecifically and a tunnel through which the transcript (red) exits. (**C**) and (**D**) are rotated by 180° about a vertical axis. The appearance of the tunnel that contains the 5′ end of an RNA transcript (red) in the elongation complex (D). The positive electrostatic potential is blue and the negative is red. The two complexes have been oriented identically by superposition of their palm domains.

and footprinting data imply either that an additional T7 RNAP conformation exists that allows the formation of a longer heteroduplex than can be accommodated by the initiation complex while still making promoter-specific interactions, or, alternatively, that the RNA peels off the template after 3 nt, as suggested earlier (*14*), until it becomes long enough to form the 7-bp heteroduplex seen in the elongation complex. It is not obvious what kind of intermediate conformational change would move helices F and G from blocking an elongating heteroduplex without destroying the promoter binding site, because helices F and G move in concert with helices D to G in the observed transition (Fig. 2). In any case, the longer heteroduplex should destabilize the initiation complex conformation of T7 RNAP and make interactions that stabilize the elongation conformation of the enzyme. Indeed, the total surface area of contact between T7 RNAP and the DNA/RNA substrates is the same in the initiation and elongation structures, but only after the heteroduplex reaches 7 bp in length.

One might ask why the abortive synthesis of short oligonucleotides exists and why the enzyme might not be "designed" to carry out the stable RNA synthesis that occurs in the elongation phase right from the start. The initiation of RNA synthesis at a particular site that is required for specific gene expression and regulation as well as the need for de novo, unprimed synthesis necessitates binding of the polymerase at a specific DNA location, the promoter. Furthermore, the binding of T7 RNAP to both the promoter and the downstream DNA appears to be essential for opening the bubble. Because short transcripts (2 to 4 nt) cannot form stable heteroduplexes, polymerase leaving the promoter prematurely would presumably lead to bubble closure and transcript displacement by the nontemplate strand. An enzyme locked in the elongation mode conformation seems unlikely to be capable of specific initiation and bubble opening.

T7 RNAP opening of the transcription bubble. T7 RNAP appears to facilitate the formation of a transcription bubble by untwisting and bending duplex DNA. To derive the degree of promoter untwisting and bending upon binding to T7 RNAP, we constructed a complete open-promoter DNA by superimposing the palm domains of the open promoter and the elongation complexes and combining the upstream DNA from the former with the downstream DNA of the latter. This complete open promoter contained 13 bp of upstream promoter duplex, 10 bp of downstream duplex and 6 nt of template between −4 and +2, for a total of 29 nt of template DNA. After superimposing the 13-bp promoter on one end of a 29-bp B-form duplex, the other end of the straight DNA has to be bent by 80° and untwisted by 146° to superimpose on the downstream DNA of the complete open promoter. The bubble is 6 nt long and includes nucleotides −4 to +2 (promoter numbering). The template strand of the bubble, which is visible in the complexes, bends sharply (~90°) at position −4 and descends into the active site and likewise bends about 80° after +2 to reemerge from the deeply buried active site and to rejoin the downstream duplex at +3.

We suggest that the energy required to melt the 6 bp of the duplex to form the bubble (about 9 to 16 kcal/mol) may arise from the reduction of the DNA twist by about 146° and from changing the relative orientations of the upstream and downstream DNA axes by 80°. The underwinding of DNA produced by promoter binding could destabilize the duplex by up to 24 kcal/mol (*36*), and the bending may destabilize it by as much as 25 kcal/mol (*36*), either one of which is sufficient to melt the duplex. It is presumably the extensive interaction between the enzyme and a bent, unwound, open-promoter DNA that produces the free energy required to distort the DNA, thereby destabilizing and opening the duplex. The total surface area of the initial open-promoter DNA that interacts with T7 RNAP is 2700 Å^2. With the use of the conversion factor of 25 cal/Å^2 of buried surface area, which is often used to evaluate the energetic contributions of hydrophobic interactions to binding (*37*), we can calculate that as much as 68 kcal of intrinsic interaction energy may be available for DNA distortion, entropy of immobilization reducing conformational entropy, and a 10^{-9} M dissociation constant (K_d) (*38–40*). In this regard, it may be interesting to note that the several insertions in the fingers domain of T7 RNAP as compared with Klenow fragment serve to greatly increase the interaction surface with downstream DNA in the RNAPs.

Nontemplate DNA. The single-stranded nontemplate DNA is well separated from the heteroduplex in the transcription bubble (Figs. 1B and 3C), which is held open by extensive interactions between the polymerase and both the template DNA-RNA heteroduplex and the single-stranded nontemplate strand. The nontemplate DNA is immobilized at two points along the transcription bubble. At the upstream fork, it interacts with the thumb domain and the outer surface of the RNA exit tunnel that is formed by subdomain H. Bases at positions $n - 2$ and $n - 3$ are flipped out of the helical axis and stack with Arg^{173} of subdomain H and Tyr^{385} of the thumb domain, respectively. At the downstream fork, it interacts with the fingers domain. The nontemplate DNA interacts with one side of subdomain H, while the mRNA interacts with the other (Fig. 3C).

Crystal structures of the initiation complex and the open promoter complex did not show nontemplate DNA downstream of −5 (promoter numbering). In both complexes, the DNA duplex from −1 to −4 was melted with the template strand firmly bound and plunging into the active site, whereas the nontemplate strand was disordered (*13*, *14*). If the nontemplate strand is modeled into

Table 1. Summary of crystallographic analysis.

Data collection and SAD analysis	Native1	Native2	SeMet(SAD)
Wavelength (Å)	0.98	0.98	0.979
Resolution (Å)	2.1	2.2	2.9
Completeness (overall/last shell, %)	85.0 (35.6)	89.2 (55.4)	100 (100)
Space group	$P2_1$ (2 copies/a.u.)	$C222_1$ (1 copy/a.u.)	$P2_1$ (2 copies/a.u.)
Unit cell (Å)	100.7 144.8 101.2 90.0 90.6 90.0	142.9 145.5 145.6 90.0 90.0 90.0	100.8 144.2 102.0 90.0 91.1 90.0
No. of sites			50
R_{linear}	6.5	8.2	7.2
Phasing power (acentric reflections)			1.2
Solvent-flipping density modification			
SAD FOM			0.62
Structure refinement			
Resolution (Å)	40–2.09 ($F_p > 2\sigma\|F\|$)		
R_{factor}/R_{free}	24.1/27.3		
No. protein residues, nucleic acid, water	862 amino acids (missing residues: 1, 233–240, 363–374), 47 nucleotides, 190 water molecules.		

$R_{linear} = |I - \langle I \rangle|/I$, where I is the observed intensity and $\langle I \rangle$ is the average intensity for multiple measurements of reflections.
Phasing power = rms($|Fph|/E$), where E is the residual lack of closure.
FOM, figure of merit.
R_{free} and R_{factor} calculated with the use of the test data set that was excluded from the refinement.
$R = (|F_p|-|F_c|)/|F_p|$, where F_p and F_c are the observed and calculated structure factors.
a.u., asymmetric unit; σ, standard deviation.

these initiation complexes at the position it occupies in the elongation complex, there are no plausible interactions apparent, because subdomain H lies on the opposite side of the molecule in the initiation complex and the position of the thumb is also altered. This observation agrees with biochemical studies showing that the presence of nontemplate DNA in the bubble region stabilizes the elongation complex but has little effect on stabilizing the initiation complex (*40*).

Strand separation of downstream DNA. Because DNA-dependent RNAPs transcribe double-stranded DNA, they must displace the nontemplate strand from the downstream duplex as it enters the bubble to generate the single-stranded template, thereby functioning as a helicase in addition to a polymerase. Two components of the elongation complex structure, subdomain H and helix Y (residues 644 to 661), appear to be involved in this process. Helix Y is wedged in the fork where the template and nontemplate strands separate from the downstream duplex, whereas subdomain H stabilizes the nontemplate strand of DNA. A bulky amino acid residue, Phe^{644}, at the end of the Y helix extends outward and stacks on the template base at position $n + 1$, the first base pair at the downstream end of the transcription bubble (Fig. 5). Helix Y serves to divert the direction of the nontemplate strand, promoting its separation from the template; this is analogous to the role of the thumb helix in diverting the direction of the 5′ end of the RNA transcript as it separates from template.

Ever since the first structural studies of the *Escherischia coli* Klenow fragment (KF) of Pol I (*16*, *42*) and continuing through those of substrate complexes with the pol I family of enzymes (*17*, *30*, *43*, *44*), it has remained obscure how DNA polymerase I is able not only to fill single-stranded gaps but also to displace the RNA primers of Okazaki fragments and to synthesize DNA, leaving only a nick in the DNA duplex (*45*). Comparison of the structure of T7 RNAP bound to downstream duplex DNA with that of the Pol I family of DNA polymerases provides structural insights into this process. Superposition of the C_α backbones of the palm domains of T7 RNAP and *E. coli* KF aligns the homologous portions of the respective fingers domains, and the downstream duplex DNA bound to T7 RNAP fits well onto the fingers domain of KF (Fig. 5). Helix Y of T7 RNAP aligns precisely on helix O1 (residues 770 to 778) of KF, which lies between the template and nontemplate strands. Furthermore, Phe^{644} of T7 RNAP that stacks on the last template base of the downstream duplex is identically positioned as Phe^{771} of KF. The Phe at this position is conserved in all Pol I family polymerases as a large hydrophobic residue, implying that the Pol I family polymerases all use a similar mechanism for binding downstream duplex and for separating the two strands. A similar structure is not present in other DNA polymerases, such as the B family of replication polymerases that do not exhibit strand displacement ability.

In the model of Pol I constructed with downstream DNA, the nontemplate strand departs the Pol I downstream duplex in the direction of the 5′ nuclease domain (*46*), which is responsible for cleaving the Okazaki RNA, whereas the template strand enters the polymerase active site in the same way as the template strand in Pol I DNA polymerase binary complexes (*30*). In the DNA Pol I binary complexes as in T7 RNAP, the template base that will pair with the incoming dNTP is the *n* position and lies in a pocket until the incoming nt arrives to form the ternary complex of enzyme, primer-template, and NTP. After nt insertion in DNA Pol I, the template nt *n* now becomes base-paired, creating a continuous duplex DNA with a nick in the nontemplate strand between the *n* base pair and $n + 1$ base pair, which is the first in the downstream duplex.

Inhibition of T7 RNAP transcription by T7 lysozyme. Biochemical data implied that T7 lysozyme may inhibit transcription by preventing T7 RNAP from undergoing the transition from the initiation to the elongation phase (*47*, *48*), although previous structural data implied that lysozyme binding may additionally alter the site of catalysis by repositioning the COOH terminus (*13*). T7 lysozyme negatively regulates T7 RNAP by binding to it either during or before the initiation phase of transcription, in which case only short abortive transcripts are made, but T7 RNAP is largely unaffected by T7 lysozyme once it has entered the elongation phase, except that there is reduced synthesis past pause sites containing an RNA helical hairpin (*7*, *10*, *11*). The cocrystal structure of T7 RNAP and lysozyme shows the polymerase largely in its initiation phase conformation (except for the extreme COOH-terminus) with lysozyme bound to the COOH-terminal domain at some distance from the promoter and catalytic sites (*12*, *18*). The T7 lysozyme structure was modeled onto the elongation conformation of T7 RNAP after superposition of the palm domain of the polymerase in the lysozome complex and elongation complex. The lysozyme fits well over most of its interaction surface, except the specificity loop moves 5 Å closer to the lysozyme in its elongation complex conformation. A comparison of the structure of the lysozyme binding sites on T7 RNAP in the initiation and elongation conformations reveals no striking differences. Lysozyme bound to the elongation conformation, however, would be immediately adjacent to the specificity loop and not far from the tunnel exit. Perhaps the 5′ end of the message and/or the specificity loop make a new interaction that prevents elongation of the transcript beyond 15 nt as is observed biochemically (*10*, *11*). We must conclude that our understanding of the structural basis for T7 RNAP inhibition by lysozyme is at present incomplete.

Comparison of T7 RNAP with the multisubunit RNAP. Comparisons of the structures of T7 RNAP and its various substrate complexes with those of the multi-subunit DNA-dependent RNAP that have been recently determined (*49*, *50*) show several similarities as well as a few differences. The T7 RNAP elongation complex has a similar angle (~90°) between the axes of the down-

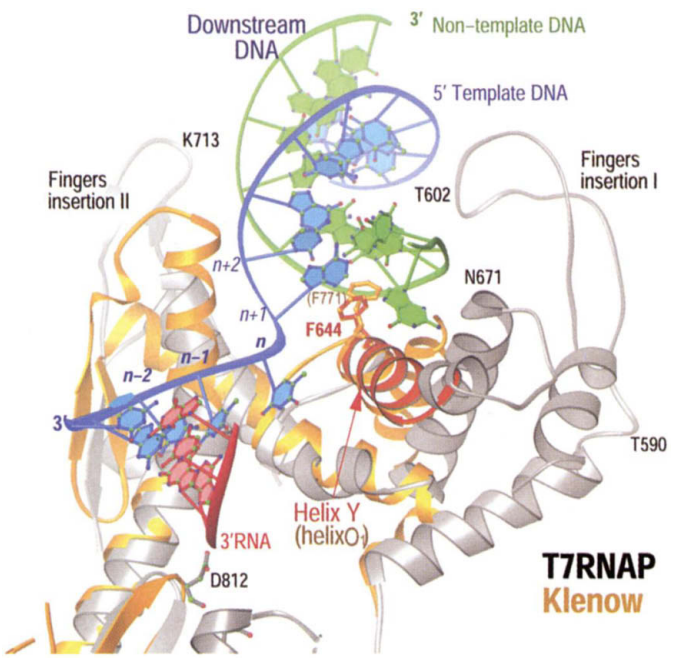

Fig. 5. The T7 RNAP fingers domain (gray) bound to the downstream DNA superimposed on the corresponding part of the Klenow fragment fingers domain (yellow). The template DNA (blue) is redirected and separated from the nontemplate DNA (green) by the α helix Y and Phe^{644} in T7 RNAP. A corresponding α helix O1 and Phe are found in the KF, as are other portions of the downstream DNA binding site. Part of the RNA transcript is shown in red.

stream DNA and the heteroduplex as that observed in the yeast Pol II elongation complex (*39*) (although the dihedral angles relative to the primer terminus differ), and the length of the heteroduplex is similar (8 bp versus 9 bp when comparing postinsertion states). Although there are three unpaired template bases between the last nucleotide of the heteroduplex and the first observed base pair of downstream duplex of Pol II, they are in the B-form DNA conformation and could be base-paired in a true transcription bubble, in which case the terminal bp of the heteroduplex and the first bp of the downstream duplex would be adjacent as in T7 RNAP. The functional reasons for these similarities in the structures of substrates bound to two nonhomologous RNAPs are unclear, but they may be related to common mechanisms of translocation, duplex opening, or access to and/or correct selection of incoming NTP. T7 RNAP has a tunnel-like opening, as do the DNA polymerases, most notably the B-family polymerases (*51*) and the multisubunit RNAPs (referred to as the "funnel" in yeast Pol II) (*49*), which provides access for the incoming NTP or dNTP. The large angle between the heteroduplex and downstream DNA allows unhindered access of the incoming nucleotide to the primer terminus. Also analogous in the two polymerases, the binding of downstream DNA to yeast Pol II results in the rotation of a domain, called the flap, that sequesters the DNA, a conformational change that may be functionally similar to the closing of loops in the fingers domain of T7 RNAP around the downstream DNA subsequent to its binding. Such sequestering of downstream DNA upon its binding to either polymerase family seems unlikely to be responsible for processive elongation synthesis (*49*, *50*) because such changes presumably also occur upon formation of the initiation complex.

After superimposing the structure of Taq RNAP complex with promoter DNA (*52*) on that of yeast pol II complexed with downstream DNA (*50*), we measured the bend angle between the upstream DNA on the former and the downstream DNA on the latter, and it is again similar between the multisubunit polymerase (110°) and T7 RNAP (100°). The size of the upstream and downstream duplex binding sites are about 2 and 1.5 times larger, respectively, in the multisubunit polymerase than the corresponding interaction sites in T7 RNAP. This difference may be related to the larger energy required to open the 12-nt bubble in the former compared with the 6-nt bubble in the latter. Further, like T7 RNAP, both the bacterial and the yeast RNAPs have a presumed RNA exit tunnel that lies near the 5′ terminus of the RNA in the heteroduplex, which exists also in the apoenzymes. However, in the bacterial holoenzyme, which also contains the σ subunit that is responsible for promoter-specific initiation, the tunnel is blocked by a domain of σ (*49*). It has been suggested that the transition from the initiation phase to the elongation phase in bacteria is triggered by the formation of an RNA transcript that is sufficiently long to displace σ from the tunnel, thereby facilitating σ's dissociation from the complex and promoter release (*52–54*), a hypothesis yet to be verified. Only when a transcript is long enough to displace σ and pass through the tunnel would processive synthesis commence. Presumably, processive RNA synthesis in the elongation phase results from the RNA transcript being surrounded by protein in both polymerase families. A major difference, then, between the multisubunit and T7 RNAPs is the massive conformational change exhibited by the latter to form an exit tunnel that already exists in the former but is blocked in the initiation phase by σ. The unprecedented conformational dexterity exhibited by T7 RNAP may be a consequence of the limited genome space of the T7 phage, which may impose the requirement for this dual functionality of promoter recognition and tunnel formation by the NH_2-terminal domain.

Conclusion. The crystal structure of a T7 RNAP elongation complex shows that the NH_2-terminal domain rearranges from its structure in the initiation complex, which destroys the promoter binding site and creates a channel that accommodates a 7-bp heteroduplex as well as a tunnel through which the transcript passes after peeling off the heteroduplex. These features account for the enzyme's processivity in the elongation phase as well as the phenomenon of promoter clearance. The fingers domain forms a binding site for 10 bp of downstream DNA, whose orientation relative to the upstream promoter as seen in previous complexes suggests that the enzyme uses the interaction with upstream and downstream duplex to bend and unwind DNA at the transcription start site and thus facilitate promoter opening. The template and nontemplate strands entering the active site from the downstream DNA are separated by an α helix that is also present in the homologous DNA polymerase I and may explain how DNA polymerase I is able to displace the 5′ end of the nontemplate strand. The comparison of the structural differences between the transcribing RNAP complexes and the homologous DNA polymerase I explains how the additional functional properties exhibited by the RNAP are acquired.

References and Notes

1. P. H. von Hippel, D. G. Bear, W. D. Morgan, J. A. McSwiggen, *Annu. Rev. Biochem.* **53**, 389 (1984).
2. P. H. von Hippel, *Science* **281**, 660 (1998).
3. A. J. Carpousis, J. D. Gralla, *Biochemistry* **19**, 3245 (1980).
4. C. T. Martin, D. K. Muller, J. E. Coleman, *Biochemistry* **27**, 3966 (1988).
5. R. A. Ikeda, C. C. Richardson, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 3614 (1986).
6. S. I. Gunderson, K. A. Chapman, R. R. Burgess, *Biochemistry* **26**, 1539 (1987).
7. J. Huang, R. Sousa, *J. Mol. Biol.* **303**, 347 (2000).
8. Y. Jia, S. S. Patel, *Biochemistry* **36**, 4223 (1997).
9. M. L. Ling, S. S. Risman, J. F. Klement, N. McGraw, W. T. McAllister, *Nucleic Acids Res.* **17**, 1605 (1989).
10. B. A. Moffatt, F. W. Studier, *Cell* **49**, 221 (1987).
11. X. Zhang, F. W. Studier, *J. Mol. Biol.* **269**, 10 (1997).
12. D. Jeruzalmi, T. A. Steitz, *EMBO J.* **17**, 4101 (1998).
13. G. M. Cheetham, D. Jeruzalmi, T. A. Steitz, *Nature* **399**, 80 (1999).
14. ———, T. A. Steitz, *Science* **286**, 2305 (1999).
15. P. Davanloo, A. H. Rosenberg, J. J. Dunn, F. W. Studier, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2035 (1984).
16. D. L. Ollis, P. Brick, R. Hamlin, N. G. Xuong, T. A. Steitz, *Nature* **313**, 762 (1985).
17. C. A. Brautigam, T. A. Steitz, *Curr. Opin. Struct. Biol.* **8**, 54 (1998).
18. G. M. Cheetham, T. A. Steitz, *Curr. Opin. Struct. Biol.* **10**, 117 (2000).
19. D. K. Muller, C. T. Martin, J. E. Coleman, *Biochemistry* **27**, 5763 (1998).
20. R. A. Ikeda, C. C. Richardson, *J. Biol. Chem.* **262**, 3790 (1987).
21. B. He, M. Rong, R. K. Durbin, W. T. McAllister, *J. Mol. Biol.* **265**, 275 (1997).
22. E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, *Cell* **89**, 33 (1997).
23. N. Korzheva, A. Mustaev, E. Nudler, V. Nikiforov, A. Goldfarb, *Cold Spring Harbor Symp. Quant. Biol.* **63**, 337 (1998).
24. D. Temiakov *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14109 (2000).
25. Template-dependent transcription by T7 RNAP was assayed by incorporation of P32-labeled nucleotides into the RNA transcript using the designed DNA substrate. Binding affinity of the designed bubble construct (Fig. 1A) to T7 RNAP was determined by gel shift assay and florescence polarization measurement. The apparent K_d was estimated to be 5 μM under the crystallization conditions.
26. S. S. Daube, P. H. von Hippel, *Science* **258**, 1320 (1992).
27. ———, *Biochemistry* **33**, 340 (1994).
28. The transcription bubble substrate was assembled by mixing three oligonucleotides at a molar ratio of template/nontemplate/RNA of 1.0:1.0:1.1. The annealing solution contained 50 mM of pH 7.0 tris-HCl, 10 mM Mg acetate, and 20 mM $NaCl_2$ and was heated to 75° for 5 min. The mixture was then slowly cooled to 20°C over a period of 2 hours. The transcription elongation complex was then assembled by mixing the annealed substrate with equimolar T7 RNAP.
29. Crystals of the T7 RNAP transcription elongation complex were obtained by vapor diffusion; 250 μM of elongation complex solution was mixed with an equal volume of a well solution containing 50 mM tris-HCl, pH 7.0, 200 mM Li_2SO_4, 6% polyethylene glycol 8000, 10 mM Mg acetate and 5 mM dithiothreitol. The crystals were flash-frozen in liquid propane before data collection. Selenomethionine- substituted T7 RNAP elongation complex was also prepared for multiwavelength anomalous dispersion experiments (*55*). Native data were collected at Brookhaven National Laboratory (BNL), beamline X25, and single-wavelength anomalous dispersive (SAD) data were collected at Advance Photon Source (APS), beamline ID19. Data were reduced with DENZO and Scalepack (*56*). The derivative and one native data set were in monoclinic form (p21), whereas the other native data set was in orthorhombic form (C221). Phases from molecular replacement using the COOH-terminal domain (residues 300 to 883) of the T7 RNAP initiation complex were used to locate 50 Se sites. The phases from molecular replacement and Se SAD were then combined with program AMORE (*57*) and CNS (*58*). Experimental phases obtained from monoclinic data were used to refine against the orthorhombic data to yield the final atomic model, since this data set revealed more ordered downstream DNA.
30. J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **391**, 304 (1998).
31. Y. W. Yin, T. A. Steitz, in preparation.

32. N. M. Stano, S. S. Patel, *J. Mol. Biol.* **315**, 1009 (2002).
33. L. G. Brieba, R. Sousa, *Biochemistry* **40**, 3882 (2001).
34. P. A. Bullough, F. M. Hughson, J. J. Skehel, D. C. Wiley, *Nature* **371**, 37 (1994).
35. J. Kuriyan, M. O'Donnell, *J. Mol. Biol.* **234**, 915 (1993).
36. V. A. Bloomfield, D. M. Crothers, I. Tinoco Jr., *Nucleic Acids: Structures, Properties, and Functions* (University Science Books, Sausalito, CA, 2000).
37. F. M. Richards, *Annu. Rev. Biophys. Bioeng.* **6**, 151 (1977).
38. G. A. Diaz, M. Rong, W. T. McAllister, R. K. Durbin, *Biochemistry* **35**, 10837 (1996).
39. Y. Jia, A. Kumar, S. S. Patel, *J. Biol. Chem.* **271**, 30451 (1996).
40. A. Ujvari, C. T. Martin, *J. Mol. Biol.* **273**, 775 (1997).
41. V. Gopal, L. G. Brieba, R. Guajardo, W. T. McAllister, R. Sousa, *J. Mol. Biol.* **290**, 411 (1999).
42. P. S. Freemont, J. M. Friedman, L. S. Beese, M. R. Sanderson, T. A. Steitz, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 8924 (1988).
43. S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **391**, 251 (1998).
44. _____, T. Ellenberger, *Curr. Opin. Struct. Biol.* **8**, 704 (1998).
45. A. Kornberg, T. Baker, *DNA Replication* (Freeman, New York, ed. 2, 1991).
46. U. K. Urs, R. Murali, H. M. Krishna Murthy, *Acta Crystallogr.* **D 55**, 1971 (1999).
47. J. Huang, J. Villemain, R. Padilla, R. Sousa, *J. Mol. Biol.* **293**, 457 (1999).
48. A. Kumar, S. S. Patel, *Biochemistry* **36**, 13954 (1997).
49. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1876 (2001).
50. P. Cramer, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1863 (2001).
51. M. C. Franklin, J. Wang, T. A. Steitz, *Cell* **105**, 657 (2001).
52. K. S. Murakami, S. Masuda, E. A. Campbell, O. Muzzin, S. A. Darst, *Science* **296**, 1285 (2002).
53. D. G. Vassylyev *et al.*, *Nature* **417**, 712 (2002).
54. S. S. Daube, P. H. von Hippel, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8390 (1999).
55. S. Doublie, in *Methods in Enzymology*, C. W. Carter and R. M. Sweet, Eds. (Academic Press, San Diego, CA, 1997), vol. 277, pp. 523–530.
56. Z. Otwinowski, W. Minor, in *Methods in Enzymology* C. W. Carter and R. M. Sweet, Eds. (Academic Press, San Diego, CA 1997), vol. 277, pp. 307–326.
57. M. Navarro, G. A. Cross, E. Wirtz, *EMBO J.* **18**, 2265 (1999).
58. A. T. Brünger *et al.*, *Acta Crystallogr.* **D54**, 905 (1998).
59. J. A. Christopher, SPOCK: The Structural Properties Observation and Calculation Kit (Program Manual) (The Center for Macromolecular Design, Texas A&M University, College Station, TX, 1998).
60. M. Carson, in *Methods in Enzymology*, C. W. Carter and R. M. Sweet, Eds. (Academic Press, San Diego, CA, 1997), vol. 277, pp. 493–505.

61. Data were collected at two synchrotron sources for this work, ID-19 at Argonne National Labs (APS) and X25 at National Synchrotron Light Source (NSLS). Use of the Argonne National Laboratory Structural Biology Center beamlines at the Advanced Photon Source was supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research, under Contract No. W-31-109-ENG-38. Research carried out in part at the NSLS, BNL, was supported by the DOE, Division of Materials Sciences and Division of Chemical Sciences, under Contract No. DE-AC02-98CH10886. We thank M. Becker (X25) and A. Joachimiak (ID-19) for their support during beamline data collections; W. Kennedy, C. Joyce, and S. Kamtekar for many helpful discussions and critical reading of the manuscript; and D. Crothers for help with the thermodynamic calculations. The coordinates for the T7 RNAP elongation complex have been deposited in the PDB under accession code 1msw. Supported by NIH grant GM57510 to T.A.S.

16 August 2002; accepted 11 September 2002
Published online 19 September 2002;
10.1126/science.1077464
Include this information when citing this paper.

Projection of an Immunological Self Shadow Within the Thymus by the Aire Protein

Mark S. Anderson,[1] Emily S. Venanzi,[1] Ludger Klein,[2] Zhibin Chen,[1] Stuart P. Berzins,[1] Shannon J. Turley,[1] Harald von Boehmer,[2] Roderick Bronson,[3] Andrée Dierich,[4] Christophe Benoist,[1]* Diane Mathis[1]*

Humans expressing a defective form of the transcription factor AIRE (autoimmune regulator) develop multiorgan autoimmune disease. We used *aire*-deficient mice to test the hypothesis that this transcription factor regulates autoimmunity by promoting the ectopic expression of peripheral tissue–restricted antigens in medullary epithelial cells of the thymus. This hypothesis proved correct. The mutant animals exhibited a defined profile of autoimmune diseases that depended on the absence of aire in stromal cells of the thymus. Aire-deficient thymic medullary epithelial cells showed a specific reduction in ectopic transcription of genes encoding peripheral antigens. These findings highlight the importance of thymically imposed "central" tolerance in controlling autoimmunity.

A problem that has intrigued immunologists for decades is how animals achieve immunological tolerance to autoantigens (*1*). For example, T cells are generated in the thymus, and because their antigen-specific receptors are encoded by genes assembled through random somatic DNA rearrangement, the emergent repertoire of receptors inevitably includes specificities capable of reacting to self constituents. To avoid the potentially pathological state of autoimmunity, it is necessary to purge these self-reactive cells from the repertoire, either by removal or silencing. Some are removed in the thymus soon after generation, but this raises the question of how thymocytes that are reactive to proteins expressed only in nonthymic parenchymal tissues can be identified and dealt with. A commonly held notion involves a dichotomy of "central" and "peripheral" mechanisms: Tolerance to ubiquitously expressed or blood-borne antigens is achieved in the thymus, whereas tolerance to tissue-restricted antigens is secured by means of diverse extrathymic processes.

Surprisingly, RNA transcripts encoding a multiplicity of proteins previously considered to be synthesized only in particular peripheral tissues can be detected in the thymus (*2, 3*), specifically in very rare epithelial cells in the medulla (*4, 5*). Examples include transcripts encoding transcription factors, structural proteins, membrane proteins, hormones, and secreted proteins. Thymic medullary epithelial cells (MECs) have been increasingly implicated in the clonal deletion or inactivation of semimature self-reactive thymocytes (*6, 7*), fueling interest in the precise function of these ectopically expressed transcripts. Several transgenic (*4, 8–17*) and nontransgenic (*18, 19*) mouse systems have revealed a direct link between ectopic synthesis of a designated protein in MECs and the absence of peripheral lymphocyte reactivity to that protein. Many of the ectopically expressed antigens (insulin, thyroglobulin, myelin basic protein, and retinal S-antigen) are associated with organ-specific autoimmune diseases (type 1 diabetes, thyroiditis, multiple sclerosis, and uveitis, respectively). In addition, there are some very suggestive correlations between antigen expression levels in the thymus and disease susceptibility in humans (*20, 21*) and rodents (*22*). We sought to determine what drives ectopic synthesis of peripheral tissue–restricted proteins in MECs, and what impact this expression has on an animal's state of immunological self tolerance.

Following clues from the human system, we anticipated that mice lacking the *aire* gene would prove key to addressing these issues. Autoimmune polyendocrinopathy–candidia-

[1]Section on Immunology and Immunogenetics, Joslin Diabetes Center; Department of Medicine, Brigham and Women's Hospital; Harvard Medical School, 1 Joslin Place, Boston, MA 02215, USA. [2]Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA. [3]Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA. [4]Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, 1 rue Laurent Fries, 67404 Strasbourg, France.

*To whom correspondence should be addressed. E-mail: cbdm@joslin.harvard.edu