Sciencexpress Research Article

Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution

Averell L. Gnatt,* Patrick Cramer,† Jianhua Fu,‡ David A. Bushnell, Roger D. Kornberg§

Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305–5126, USA.

*Present address: Department of Pharmacology and Experimental Therapy, University of Maryland, 655 West Baltimore Street, HH403, Baltimore, MD 21201, USA.

†Present address: Institute of Biochemistry, Gene Center, University of Munich, 81377 Munich, Germany.

‡Present address: Department of Molecular Biology and Genetics, Cornell University, 223 Biotechnology Building, Ithaca, NY 14853, USA.

§To whom correspondence should be addressed. E-mail: kornberg@stanford.edu

The crystal structure of RNA polymerase II in the act of transcription has been determined at 3.3 Å resolution. Duplex DNA is seen entering the main cleft of the enzyme and unwound before the active site. Nine base pairs of DNA-RNA hybrid extend from the active center at nearly right angles to the entering DNA, with the 3'-end of the RNA in the nucleotide addition site. The 3'-end is positioned above a pore, through which nucleotides may enter, and through which RNA may be extruded during backtracking. The 5'-most residue of the RNA is close to the point of entry to an exit groove. Changes in protein structure between the transcribing complex and free enzyme include closure of a clamp over the DNA and RNA, and ordering of a series of "switches" at the base of the clamp to create a binding site complementary to the DNA-RNA hybrid. Protein-nucleic acid contacts help explain DNA and RNA strand separation, the specificity of RNA synthesis, "abortive cycling" during transcription initiation, and RNA and DNA translocation during transcription elongation.

The recent structure determination of yeast RNA polymerase II at 2.8 Å resolution and of bacterial RNA polymerase at 3.3 Å resolution has led to proposals for polymerase-DNA and –RNA interactions (1, 2, accompanying paper by Cramer *et al.*). A DNA duplex was suggested to enter a positively charged cleft between the two largest subunits and to make a right angle bend at the active center, where the DNA strands are separated, and from which a DNA-RNA hybrid emerges. Avenues for entry of substrate nucleoside triphosphates and for exit of RNA could also be surmised. While consistent with results of cross-linking experiments (3-7), these general proposals for polymerase-nucleic acid interaction have not been proven, and they do not address key questions about the transcription process: How is an unwound "bubble" of DNA established and maintained in the active center? Why does the enzyme initiate repeatedly, generating many short transcripts, before a transition is made to a stable elongating complex? What is the nature of the presumptive DNA-RNA hybrid duplex? How are DNA and RNA translocated across the surface of the enzyme, forward and backward, during RNA synthesis and backtracking? Here we report the crystal structure determination of yeast RNA polymerase II in the form of an actively transcribing complex, from which answers to these questions and additional insights into the transcription mechanism are derived.

The main technical challenge of this work was the isolation and crystallization of a transcribing complex. Initiation at an RNA polymerase II promoter requires a complex set of general transcription factors, and is poorly efficient in reconstituted systems (8, 9). Moreover, most preparations contain many inactive polymerases, and the transcribing complexes obtained would have to be purified, by mild methods to preserve their integrity (10). The initiation problem was overcome with the use of a DNA duplex bearing a single stranded "tail" at one 3'-end (Fig. 1A) (11, 12). Pol II starts transcription in the tail, 2-3 nucleotides from the junction with duplex DNA, with no requirement for general transcription factors. All active polymerase molecules are converted to transcribing complexes, which pause at a specific site when one of the four nucleoside triphosphates is withheld. The problem of contamination by inactive polymerases was solved by passage through a heparin column (12); inactive molecules were adsorbed, whereas transcribing complexes flowed through, presumably because heparin binds in the positively charged cleft of the enzyme, which is occupied by DNA and RNA in transcribing complexes. The purified complexes formed crystals diffracting anisotropically to 3.1 Å resolution (13).

Structure of a pol II transcribing complex. Diffraction data complete to 3.3 Å resolution were used for structure determination by molecular replacement with the 2.8 Å pol II structure (14). A native zinc anomalous difference Fourier map showed peaks coinciding with five of the eight zinc ions of the pol II structure, confirming the molecular replacement solution (15). The remaining three zinc ions were located in the clamp, a region shown previously to undergo a large conformational change between different pol II crystal forms (accompanying paper by Cramer *et al.*). The locations of the three zinc ions served as a guide for manual repositioning of the clamp in the transcribing complex structure. An initial electron density map revealed nucleic acids in the vicinity of the active center. After adjustment of the protein model, the

nucleic acid density improved and nine base pairs of DNA-RNA hybrid could be built (16). Additional density along the DNA template strand allowed another three nucleotides downstream and one nucleotide upstream to be built. Modeling of the nucleic acids assumed the 3'-end of the RNA at the biochemically defined pause site (Fig. 1A), since the nucleic acid sequences could not be inferred from the crystallographic data (17).

The final model contains 3521 amino acid residues, 22 nucleotides, eight Zn^{2+} ions, and a Mg^{2+} ion, and has a free R-factor of 29.8% (R-factor 25.0%, 40-3.3 Å) (Fig. 2). A simulated-annealing omit map computed from a model of the protein alone revealed the phosphate groups and most bases in the DNA-RNA hybrid region, confirming the modeling of the nucleic acids (Fig. 2A). Density for DNA in the downstream region was very weak and discontinuous, but revealed the major groove, allowing a canonical B-DNA duplex to be approximately placed (not included in the model (18)). Numbering of nucleotides in the DNA begins with +1 immediately downstream and –1 upstream of the Mg^{2+} ion (Fig. 1A).

Closure of the clamp. The structures of free and transcribing pol II differ mainly in the position of the clamp (Fig. 2B). As previously suggested (1), and now demonstrated, the clamp swings over the cleft during formation of the transcribing complex, trapping the template and transcript. The clamp rotates by about 30°, with a maximum displacement of over 30 Å at external sites (at the Rpb1 "zipper"). While most of the clamp moves as a rigid body, five "switch" regions undergo conformational changes and folding transitions (Table 1). Switches 1, 2, 4, and 5 form the base of the clamp (Fig. 3). Switches 1 and 2 are poorly ordered and switch 3 is disordered in free pol II; all three switches become well ordered in the transcribing complex. Ordering is likely induced by binding of the switches to DNA downstream and within the DNA-RNA hybrid (see below). Binding to the hybrid may help couple clamp closure to the presence of RNA. The conformational changes of the switch regions may be concerted, since the switches interact with one another. The conformational changes are accompanied by changes in a network of salt linkages to the "bridge" helix across the cleft (Rpb1 residues R839, R840 and K843).

Downstream DNA mobility. Downstream DNA lies in the cleft between the clamp and Rpb2 (Figs. 1B, 2B, 2C), consistent with results from electron crystallography of the transcribing complex (19) and results of DNA-protein cross-linking (3-7). The DNA contacts the Rpb5 "jaw" domain at a loop containing proline residue P118, as previously suggested (1), and then passes between the Rpb2 "lobe" region and the Rpb1 "clamp head". The sequence of the Rpb2 lobe is divergent between yeast and bacteria, but the fold is conserved, whereas the clamp head is not conserved.

Details of downstream DNA-pol II interaction are lacking because the electron density is weak, indicative of mobility of the DNA. Further, downstream DNAs from neighboring transcribing complexes in the crystal interact end-to-end, stacking on one another, so the precise location of the DNA may be determined by crystal packing forces. This could be the reason why there is no apparent contact between downstream DNA and the upper jaw. In addition, the length of DNA used here is possibly too short for passage all the way through the jaws.

Transcription bubble. The downstream edge of the transcription bubble lies between the poorly ordered downstream duplex DNA and the first ordered nucleotide of the template strand at position +4, three nucleotides before the beginning of the RNA-DNA hybrid (Fig. 3B). The nucleotide at position +4 in the nontemplate strand and the remainder of this strand are disordered. The template strand follows a path along the bottom of the clamp and over the "bridge" helix. Template nucleotides +4, +3, and +2 are stacked in the manner of right-handed B-DNA. The base of nucleotide +1 is flipped with respect to that of nucleotide +2 by a left-handed twist of 90°. The base at +1 therefore points downwards into the floor of the cleft for readout at the active site, whereas the base at +2 is directed upwards into the opening of the cleft. This unusual conformation of the DNA results from binding to switches 1 and 2, as well as to the bridge helix (Figs. 1C, 1D, 3). Invariant bridge helix residues A832 and T831 position the coding nucleotide through van der Waals interactions, while Y836 binds nucleotide +2, and may correspond to a tyrosine in the "O-helix" of some single subunit DNA polymerases (20, 21).

Maintenance of the downstream edge of the transcription bubble may be attributed not only to the binding of nucleotides +2, +3, and +4, but also to Rpb2 "fork loop" 2 (Figs. 1D, 4). While this loop includes several disordered residues (22), it would likely clash with the nontemplate strand at position +3 if the nontemplate strand was still base paired with the template strand. A corresponding loop in the bacterial enzyme ("βD loop I"), four residues longer than that in yeast, was previously suggested to play such a role (4). Rpb2 fork loop 1 may help maintain the transcription bubble further upstream (Figs. 1D, 4). This loop is absent from the bacterial enzyme, perhaps reflecting a difference in promoter melting between eukaryotes, which require general transcription factors for the process, and bacteria, which do not. Both fork loops, although exposed, are highly conserved between yeast and human polymerases.

DNA-RNA hybrid. The base in the template strand at position +1 forms the first of nine base pairs of DNA-RNA hybrid, located between the bridge helix and Rpb2 "wall" (Figs. 1D, 4). The length of the hybrid corroborates the value of 8-9 base pairs determined biochemically (23). The hybrid heteroduplex adopts a nonstandard conformation, intermediate between those of standard A- and B-DNA (Fig. 5), and is underwound (24), in comparison with the crystal structure of a free DNA-RNA hybrid, which is closely related to the A-form (25).

The electron density for the hybrid is strongest in the downstream region around the active center, indicative of a high degree of order, important for the high fidelity of transcription. The electron density remains strong for the DNA template strand further upstream, but the density for the RNA strand becomes weaker (Fig. 2A). This gradual loss of density reflects a diminution in the number of RNA-protein contacts. The template DNA strand is bound by protein over the entire length of the hybrid, whereas RNA contacts are limited to the downstream region (Fig. 1C). The five upstream ribonucleotides are held mainly through base pairing with the template DNA.

Contacts to the downstream and upstream parts of the hybrid are made by Rpb1 and Rpb2, respectively (Fig. 1C). Fifteen protein regions are involved, with a significant portion of the contacts arising from the ordering of Rpb1 switches 1, 2, and 3 upon nucleic acid binding. The entire set of protein contacts forms an extended, highly complementary binding surface. A surface area of 3400 $Å^2$ is buried in the protein-nucleic acid interface, comparable to values for transcription factors bound specifically to DNA sites of

similar size. Biochemical studies have shown the binding interaction contributes significantly to the stability of a transcribing complex, and thus to the high processivity of transcription (26, 27).

While a strong pol II-nucleic acid interaction is important for the ordering of nucleic acids in the active center region and for the stability of a transcribing complex, the interaction must not interfere with the translocation of nucleic acids during transcription. Indeed, the nucleic acids in the transcribing complex are mobile, as shown by the partial order of the downstream DNA (see above), and by a high overall crystallographic temperature factor of the hybrid, which appears to reflect mobility rather than static disorder (28). The conflicting requirements of tight binding and mobility may be reconciled in at least three ways. First, almost all protein contacts are to the sugar-phosphate backbones of the DNA and RNA. There are no contacts with the edges of the bases, so there is no base specificity. A large open space between pol II and the major groove of the hybrid is a prominent feature of the structure. Second, several side chains interact with two phosphate groups along the backbone simultaneously (Fig. 1C), which may reduce the activation barrier for translocation. Finally, about twenty positively charged side chains form a "second shell" around the hybrid at a distance of 4-8 Å, which may attract the hybrid without restraining its movement across the enzyme surface (29).

RNA synthesis. The active site metal ion in the transcribing complex structure corresponds to one of two metal ions in the 2.8 Å pol II structure, referred to as metal A (accompanying paper by Cramer *et al.*). The location of this metal in the transcribing complex is appropriate for binding the phosphate group between the nucleotide at the 3'-end of the RNA and the adjacent nucleotide, designated +1 and –1, respectively (Fig. 1C). In the two-metal-ion mechanism proposed for single subunit polymerases, metal A contacts the α-phosphate of the incoming nucleoside triphosphate and metal B binds all three phosphates (30-35). Metal B may be absent from the transcribing complex structure because it has left with the pyrophosphate following nucleotide addition. On this basis, position +1 in the transcribing complex would be that of a nucleotide just added to the growing RNA, before translocation to bring the next template base into position opposite an empty nucleotide-binding site at the end of the RNA (36) (Fig. 6).

The ribonucleotide in position +1 lies in the entrance to the previously noted "pore 1," which extends from the floor of the cleft through to the backside of the enzyme. This location and orientation of the 3'-end of the RNA lend strong support to the previous proposal that nucleoside triphosphates enter through the pore during RNA synthesis, and that RNA is extruded through the pore during backtracking (1). The close fit of the DNA-RNA hybrid to the surrounding protein leaves no alternative to the pore for access of nucleotides to the active site. (Major conformational changes creating access are unlikely, since they would disrupt protein-nucleic acid contacts important for the fidelity and processivity of transcription.)

Specificity for ribo- rather than deoxyribonucleotides may be attributed to recognition of both the ribose sugar and the DNA-RNA hybrid helix. The 2'-hydroxyl group of a ribonucleotide in the substrate binding site (position +1) is 5 Å from the side chain of the highly conserved Rpb1 residue N479. While this distance is too great for specific interaction, a slightly different positioning of an incoming nucleoside triphosphate might permit hydrogen bonding and discrimination of the ribose sugar. Different positioning of the nucleoside triphosphate could result from chelation by metal B, bound at a site in the structure of free pol II (accompanying paper by Cramer *et al.*). RNA 2'-hydroxyl groups at positions -1, -3, and -5 are at hydrogen bonding distance from the side chains of Rpb1 residue R446 and Rpb2 residues H1097 and Q481. The nucleic acid binding site is, furthermore, highly complementary to the nonstandard conformation of the hybrid helix and not to the standard conformation of a DNA double helix. Such indirect discrimination was previously suggested to contribute to the specificity of T7 RNA polymerase transcription (37). Recognition of RNA in the transcribing complex from positions –1 to –5, by both hydrogen bonding and indirect discrimination, can contribute to the specificity of RNA synthesis through proofreading. The presence of a deoxyribonucleotide or of an incorrect base anywhere in this region of the RNA will be destabilizing. A backtracked complex, with previously correctly synthesized RNA in the hybrid region, and with the RNA containing the misincorporated nucleotide extruded at the 3'-end, will be favored. The extruded RNA can be removed by cleavage at the active site, through the action of transcription factor TFIIS.

Key nonspecific (van der Waals) contacts to the nucleotide base at the end of the hybrid region, in position +1, are made by residues T831 and A832 from the Rpb1 bridge helix, as mentioned above. Although highly conserved, the bridge helix is essentially straight in the pol II structures so far determined but bent in the bacterial enzyme structure in the vicinity of the residues corresponding to T831 and A832 (1, 2). The bend would produce a movement of this region of the bridge helix by 3-4 Å, resulting in a clash with the nucleotide at position +1 (Fig. 6). Modeling of a bacterial transcribing complex resulted in such a clash (4). We speculate that the bridge helix oscillates between straight and bent states, and that this movement accompanies the translocation of nucleic acids during transcription: addition of a nucleotide at position +1 would occur in the straight state; translocation to position –1 and movement of nucleic acids through the distance between base pairs, about 3.2 Å, would be accompanied by a conformational change to the bent state; reversion to the straight state without movement of nucleic acids would create an empty site at position +1 for entry of the next nucleotide, completing a cycle of

Protenucleotide addition during RNA synthesis (Fig. 6).in-RNA contacts are of special significance at the very beginning of transcription. Nucleoside triphosphates must be held in positions +1 and –1 for the synthesis of the first phosphodiester bond. Following translocation to positions –1 and –2, the dinucleotide product must still be held by protein-RNA contacts, as the energy of base-pairing alone is insufficient for retention in the complex. Indeed, RNA is deeply buried in the transcribing complex as far as position –3 (Fig. 1C). Di- and trinucleotides are nevertheless occasionally released, and transcription must restart, resulting in "abortive cycling" (38). RNA is exposed at position –4 and beyond, with no direct protein contacts except for the hydrogen bond at position –5 mentioned above. Coincident with exposure of the RNA, biochemical studies reveal a transition in stability at a transcript length of four residues, beyond which the RNA is generally retained (39). While the direct protein-RNA contacts observed up to this point may be largely responsible for retention, long range interactions also play a role. For example, a highly conserved arginine makes

long range electrostatic interactions with the RNA around position –4 (R497 in Rpb2, R529 in *E. coli* β), and mutation of this residue results in the overproduction of abortive transcripts (40).

RNA exit. Abortive cycling yields an abundance of 2-3 residue transcripts, as well as transcripts of up to 10 residues (41). An initiating complex evidently undergoes a second transition when the transcript reaches ten residues in length. At this point the newly synthesized RNA must separate from the DNA-RNA hybrid and enter an exit channel on the surface of the enzyme, where it remains protected from nuclease attack for about six more residues (42). Three loops extending from the clamp, termed "rudder," "lid," and "zipper," have been suggested to play roles in hybrid dissociation, RNA exit, and maintenance of the upstream end of the transcription bubble (2, and accompanying paper by Cramer *et al*., Fig. 4). Modeling of the DNA-RNA hybrid beyond the nine base pairs seen in the transcribing complex structure would produce a clash with the rudder. Extension of the RNA from the last hybrid base pair leads beneath the rudder to the previously proposed "exit groove 1." Continuation of this RNA path also leads beneath the lid, whose role may be to maintain the separation of RNA and template DNA strands. The zipper may play a similar role in separating template and nontemplate DNA strands. The lid and a small portion of the rudder are disordered in the transcribing complex structure, but are ordered in the free pol II structure. The lid and rudder may become ordered in the transcribing complex in conjunction with the second transition (43) and with the establishment of a stable, elongating complex.

Conclusions and prospects. The atomic structure of RNA polymerase II in the act of transcription reveals the protein-DNA and -RNA interactions underlying the process. The structure shows a right angle bend of the DNA path at the active center. This feature is understandable in retrospect. The bend orients the DNA-RNA hybrid optimally for transcription, which occurs along the direction of the hybrid axis. Nucleotides enter through the funnel and pore, add to the RNA at the end of the RNA-DNA hybrid, translocate through the hybrid-binding region and exit beneath the rudder and lid.

Answers to many longstanding questions about the transcription mechanism may be found in the structure of the clamp. This mobile, multifunctional element does more than close over the nucleic acids in the active center to enhance the processivity of transcription. First, switch regions at the base of the clamp couple its closure to the presence of DNA-RNA hybrid in the active center. This coupling satisfies the dual requirement for retention of nucleic acids during transcript elongation and their release following termination. Second, through the rudder, lid, and zipper, the clamp plays a key role in the events of hybrid melting and template reannealing at the upstream end of the transcription bubble.
Experiments to test the proposed roles for these structural elements by site-directed mutagenesis are among the many that can now be designed on the basis of the structure. In addition, polymerase may be cocrystallized with synthetic transcription bubbles (44) and other forms of RNA and DNA.

References and Notes

1. P. Cramer, *et al.*, *Science* **288**, 640 (2000).
2. G. Zhang, *et al.*, *Cell* **98**, 811 (1999).
3. E. Nudler, *J. Mol. Biol.* **288**, 1 (1999).
4. N. Korzheva, *et al.*, *Science* **289**, 619 (2000).
5. N. Naryshkin, A. Revyakin, Y. Kim, V. Mekler, R. H. Ebright, *Cell* **101**, 601 (2000).
6. C. I. Wooddell, R. R. Burgess, *Biochemistry* **39**, 13405 (2000).
7. M. S. Bartlett, M. Thomm, E. P. Geiduschek, *Nat. Struct. Biol.* **7**, 782 (2000).
8. M. H. Sayre, H. Tschochner, R. D. Kornberg, *J. Biol. Chem.* **267**, 23376 (1992).
9. R. Conaway, J. Conaway, *Prog. Nucleic Acid Res. Mol. Biol.* **56**, 327 (1997).
10. A. M. Edwards, C. M. Kane, R. A. Young, R. D. Kornberg**,** *J. Biol. Chem.* **266**, 71 (1991).
11. T. R. Kadesch, M. J. Chamberlin, *J. Biol. Chem.* **257**, 5286 (1982).
12. A. Gnatt, J. Fu, R. D. Kornberg, *J. Biol. Chem.* **272**, 30799 (1997).
13. Plate-like monoclinic crystals of space group C2 with unit cell dimensions a=157.3 Å, b=220.7 Å, c=191.3 Å, β=97.5° were grown by the sitting drop vapor diffusion method under the conditions previously developed for free pol II [J. Fu, *et al*., *Cell* **98**, 799-810 (1999)]. Crystals were transferred slowly to freezing buffer (1) and flash frozen in liquid nitrogen. Diffraction data were collected at a wavelength of 0.998 Å at beamline 9.2 at the Stanford Synchrotron Radiation Laboratory. While diffraction to 3.1 Å resolution could be observed in two directions, anisotropy limited the useable data to 3.3 Å resolution.
14. Data processing with DENZO and SCALEPACK [Z. Otwinowski, W. Minor, *Meth. Enzym.* **276**, 307 (1996)] showed that the data collected at 0.998 Å were 100% complete in the resolution range 40-3.3 Å. A total of 96,867 unique reflections were measured. At a redundancy of 4.4, the R_{sym} was 11.1% (31.7% at 3.4-3.3 Å). The structure was solved by molecular replacement using AMORE [J. Navaza, *Acta Cryst.* **A50**, 157-163 (1994)]. A modified atomic pol II structure lacking the mobile clamp was used as search model. A single strong peak was obtained after rotation and translation searches (Correlation coefficient=59, R-factor=43%, 15-6.0 Å resolution).
15. Diffraction data were recollected at the zinc anomalous peak wavelength (1.283 Å) from the crystal used in structure determination. Initial phases were calculated from the pol II search model after rigid body refinement in CNS (45).
16. Model building was carried out with the program O [T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Cryst.* **A47**, 110 (1991)] and refinement was carried out with CNS (45). For cross-validation, 10% of the data were excluded from refinement. The four mobile modules defined for free pol II (accompanying paper by Cramer *et al*.) were used for rigid body refinement, followed by bulk solvent correction and anisotropic scaling. After positional and restrained B-factor refinement, a free R-factor of 35% was obtained using all data. The resulting sigmaa-weighted electron density maps allowed building of switch 3 and rebuilding of the other switch regions. Loops that were present in free pol II but disordered in the transcribing complex were removed. The final protein electron density was generally of good quality and most side chains were visible. Some flexible regions, including the jaws, parts of Rpb8, and the upper portions of the wall and clamp, showed only main chain density. In these regions, the refined pol II structure was not rebuilt. A few rounds of model building and refinement of the protein

lowered the free R-factor to 31.0%. At this stage, difference density with a helical shape was observed for the nucleic acids in the hybrid region and phosphates and bases were revealed. The density originating at the active site metal was assigned to the RNA strand and the opposite continuous density was assigned to the DNA template strand. A total of 22 nucleotides were placed individually, resulting in a 0.7% drop in the free R-factor after refinement.
17. The 3.3 Å electron density map did not allow distinction of purine from pyrimidine bases. Placement of the particular sequences thus assumed complete RNA synthesis until the pause site and no backtracking. Modeling resulted in a length of the downstream DNA that agrees with end-to-end packing of DNAs from neighboring complexes. The ambiguity in the assignment of nucleic acid sequences does not affect our conclusions since there are no base-specific protein contacts. The density map included a few weak, disconnected peaks in pore 1 that may arise from backtracked RNA in a subpopulation of complexes or from incoming nucleoside triphosphates.
18. At the standard contour level of 1.0σ, only a few disconnected peaks are observed for the downstream DNA. At a contour level of 0.8σ, extended density features are observed, which identify the approximate helix axis and major groove of the downstream DNA, with only a few disconnected noise peaks in the surrounding solvent region. Inclusion of the DNA duplex placed in this way in the refinement led to an increase in the free R-factor.
19. C. L. Poglitsch, *et al.*, *Cell* **98**, 791 (1999).
20. J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **391**, 304 (1998).
21. A. C. Rodriguez, H. W. Park, C. Mao, L. S. Beese, *J. Mol. Biol.* **299**, 447 (2000).
22. Rpb2 residues 468-476 and 503-508 are disordered in fork loop 1 and 2, respectively.
23. E. Nudler, A. Mustaev, E. Lukhtanov, A. Goldfarb, *Cell* **89**, 38 (1997). M. L. Kireeva, N. Komissarova, D. S. Waugh, M. J. Kashlev, *J. Biol. Chem.* **275**, 6530 (2000).
24. The nucleic acid model was obtained by placing nucleotides manually into unbiased electron density peaks. At 3.3 Å resolution, the location of phosphate groups and the approximate axes through base pairs were revealed. After refinement, the positions of the nucleotides changed only slightly, showing that the final nucleic acid model reflects the experimental data and that the model is not primarily a result of the geometrical constraints applied during refinement. While the available data define the overall hybrid conformation, stereochemical details are not revealed and the parameters of the hybrid helix must be viewed as approximate. The hybrid shows an average rise per residue of 3.2 Å [program CURVES, R. Lavery, H. Sklenar, *J. Biomol. Struct. Dyn.* **6**, 63 (1988)], compared with 2.8 and 3.4 Å for A- and B-DNA, respectively. The average minor groove width is 10.4 Å (CURVES), compared with 11 and 7.4 Å for A- and B-DNA, respectively. The r.m.s. deviation in phosphorus atom positions between the hybrid and canonical A- and B-DNA is 3.1 and 5.5 Å, respectively. The helical twist is 12.6 residues/turn [program NEWHELIX, K. Grzeskowiak, D. S. Goodsell, M. Kaczor-Grzeskowiak, D. Cascio, R. E. Dickerson, *Biochemistry* **32**, 8923 (1993)]. The phosphorus atom positions show an r.m.s. deviation of 2.7 Å from the structure of a free hybrid (25).
25. N. C. Horton, B. C. Finzel, *J. Mol. Biol.* **264**, 521 (1996).
26. I. Sidorenkov, N. Komissarova, M. Kashlev, *Mol. Cell* **2**, 55 (1998).
27. M. L. Kireeva, N. Komissarova, D. S. Waugh, M. Kashlev, *J. Biol. Chem.* **275**, 6530 (2000).
28. The average atomic B-factor is 97 $Å^2$ for the hybrid, as compared to 63 $Å^2$ for the entire structure. The bases and backbone groups show similar B-factors. This likely indicates mobility since static disorder, arising from the presence of complexes at different register, would be expected to result in low B-factors for the backbone and higher B-factors for the bases. We believe that refinement of atomic B-factors is justified at the given resolution, and that the resulting B-factors are meaningful, since refinement of all protein atoms, starting from a constant value of 30 $Å^2$, results in an overall B-factor that is very close to that obtained for the free pol II structure at 2.8 Å resolution. Moreover, the general distribution of B-factors is similar to that for the structure of free pol II.
29. These residues include arginines 320, 326, 839 and 840 and lysines 317, 323, 330, 343 and 830 of Rpb1, and arginines 476, 497, 766, 1020, 1096 and 1124 and lysines 210, 458, 507, 775, 865, 965 and 1102 of Rpb2.
30. L. S. Beese, T. A. Steitz, *EMBO J.* **10**, 25 (1991).
31. S. Doublie, S. Tabor, A. M. Long, C. C. Richardson, T. Ellenberger, *Nature* **391**, 251 (1998).
32. J. R. Kiefer, C. Mao, J. C. Braman, L. S. Beese, *Nature* **391**, 304 (1998).
33. M. R. Sawaya, R. Prasad, S. H. Wilson, J. Kraut, H. Pelletier, *Biochemistry* **36**, 11205 (1997).
34. T. A. Steitz, *Curr. Op. Struct. Biol.* **3**, 31 (1993).
35. T. A. Steitz, *Nature* **391**, 231 (1998).
36. Although the 3'-most residue of the RNA is in the position of a nucleotide just added to the chain, it must have undergone translocation and then returned to this position prior to crystallization. Translocation is necessary to create a site for the next nucleotide, whose absence from the reaction results in a paused complex.
37. G. M. Cheetham, T. A. Steitz, *Science* **286**, 2305 (1999).
38. Y. Tintut, J. T. Wang, J. D. Gralla, *J. Biol. Chem.* **270**, 24392 (1995).
39. L. Kinsella, C. Y. Hsu, W. Schulz, D. Dennis, *Biochemistry* **21**, 2719 (1982).
40. D. J. Jin, C. L. Turnbough, *J. Mol. Biol.* **236**, 72 (1994).
41. J. R. Levin, B. Krummel, M. J. Chamberlin, *J. Mol. Biol.* 196, 85 (1987). B. Krummel, M. J. Chamberlin, *Biochemistry* 28, 7829 (1989). F. C. P. Holstege, U. Fiedler, H. T. M. Timmers, *EMBO J.* **16**, 7468 (1997).
42. W. Gu, M. Wind, D. Reines, *Proc. Natl. Acad. Sci. USA* **93**, 6935 (1996).
43. Ordering of the rudder and lid may not be observed because of structural heterogeneity of the transcribing complexes in this region. Heterogeneity might be expected as a consequence of inefficient displacement of RNA from DNA-RNA hybrid during transcription of tailed templates.
44. S. S. Daube, P. H. von Hippel, *Science* **258**, 1320 (1992).
45. A. T. Brünger, *et al.*, *Acta Cryst.* **D54**, 905 (1998).
46. For assistance at beamlines 1-5, 7-1, 9-1, and 9-2 of the Stanford Synchrotron Radiation Laboratory we thank H. Bellamy, A. Cohen, P. Ellis, P. Kuhn, T. McPhillips, M. Soltis, and the other members of the SSRL user support staff. This research is based in part on work done at SSRL which is funded by the U.S. Department of Energy Office

of Basic Energy Sciences. The structural biology program is supported by the NIH National Center for Research Resources Biomedical Technology Program and the DOE Office of Biological and Environmental Research. We thank COMPAQ for providing a Unix workstation. We thank N. Thompson and R. Burgess for generously providing antibody for protein purification. We thank J. Puglisi and members of the Kornberg laboratory for comments on the manuscript. The contribution of A.L.G. was sponsored by USAMRMC Breast Cancer Initiative, DAMD17-97-7099, and does not necessarily reflect the policy of the government. P.C. was supported by a postdoctoral fellowship of the Deutsche Forschungsgemeinschaft (DFG). D.A.B. was supported by postdoctoral fellowship #PF-00-014-01-GMC from the American Cancer Society. This research was supported by NIH grant GM49985 to R.D.K. Coordinates have been deposited at the protein data bank (accession code 1I6H).

1 February 2001; accepted 28 March 2001
Published online 19 April 2001; 10.1126/science.1059495
Include this information when citing this paper.

Fig. 1. Nucleic acids in the transcribing complex and their interactions with pol II. (**A**) DNA ("tailed template") and RNA sequences. DNA template and non-template strands are in blue and green, respectively, and RNA is in red. This color scheme is used throughout. (**B**) Ordering of nucleic acids in the transcribing complex structure. Nucleotides in the solid box are well ordered. Nucleotides in the dashed box are partially ordered, while those outside the boxes are disordered. Three protein regions that abut the downstream DNA are indicated. (**C**) Protein contacts to the ordered nucleotides boxed in (B). Amino acid residues within 4 Å of the DNA are indicated, colored according to the scheme for domain or domain-like regions of Rpb1 or Rpb2 (see accompanying paper by Cramer *et al.*). Ribose sugars are shown as pentagons, phosphates as dots, and bases as single letters. Amino acid residues listed beside phosphates contact only this nucleotide. Amino acid residues listed beside riboses contact this nucleotide and its 3'-neighbor. (**D**) Schematic representation of protein features participating in the detailed interactions shown in (C). Same notation as in (C), except bases are shown as thick bars.

Fig. 2. Crystal structure of the pol II transcribing complex. (**A**) Electron density for the nucleic acids. On the left, the final sigmaa-weighted $2mF_{obs}$-DF_{calc} electron density for the downstream DNA duplex (dashed box in Fig. 1B) is contoured at 0.8σ (green). At this contour level, the surrounding solvent region shows only scattered noise peaks. A canonical 16-base pair B-DNA duplex was placed into the density. On the right, the final model of the DNA-RNA hybrid and flanking nucleotides (boxed in Fig. 1B) is superimposed on a simulated-annealing F_{obs}-F_{calc} omit map, calculated from the protein model alone using CNS (45) (green, contoured at 2.6σ). The location of the active site metal A is indicated. (**B**) Comparison of structures of free pol II (top) and the pol II transcribing complex (bottom). The clamp (yellow) closes on DNA and RNA, which are bound in the cleft above the active center. The remainder of the protein is in gray. (**C**) Structure of the pol II transcribing complex. Portions of Rpb2 that form one side of the cleft are omitted to reveal the nucleic acids. Bases of ordered nucleotides (boxed in Fig. 1B) are depicted as cylinders protruding from the backbone ribbons. The Rpb1 bridge helix traversing the cleft is highlighted in green. The active site metal A is shown as a pink sphere.

Fig. 3. Switches, clamp loops, and the hybrid binding site. (**A**) Stereoview of the clamp core (1, yellow) and the DNA and RNA backbones. The view is as in Fig. 2C. The five switches are shown in pink and are numbered. Three loops, which extend from the clamp and may be involved in transactions at the upstream end of the transcription bubble, are in violet. Major portions of the protein are omitted for clarity. (**B**) Stereoview of nucleic acids bound in the active center.

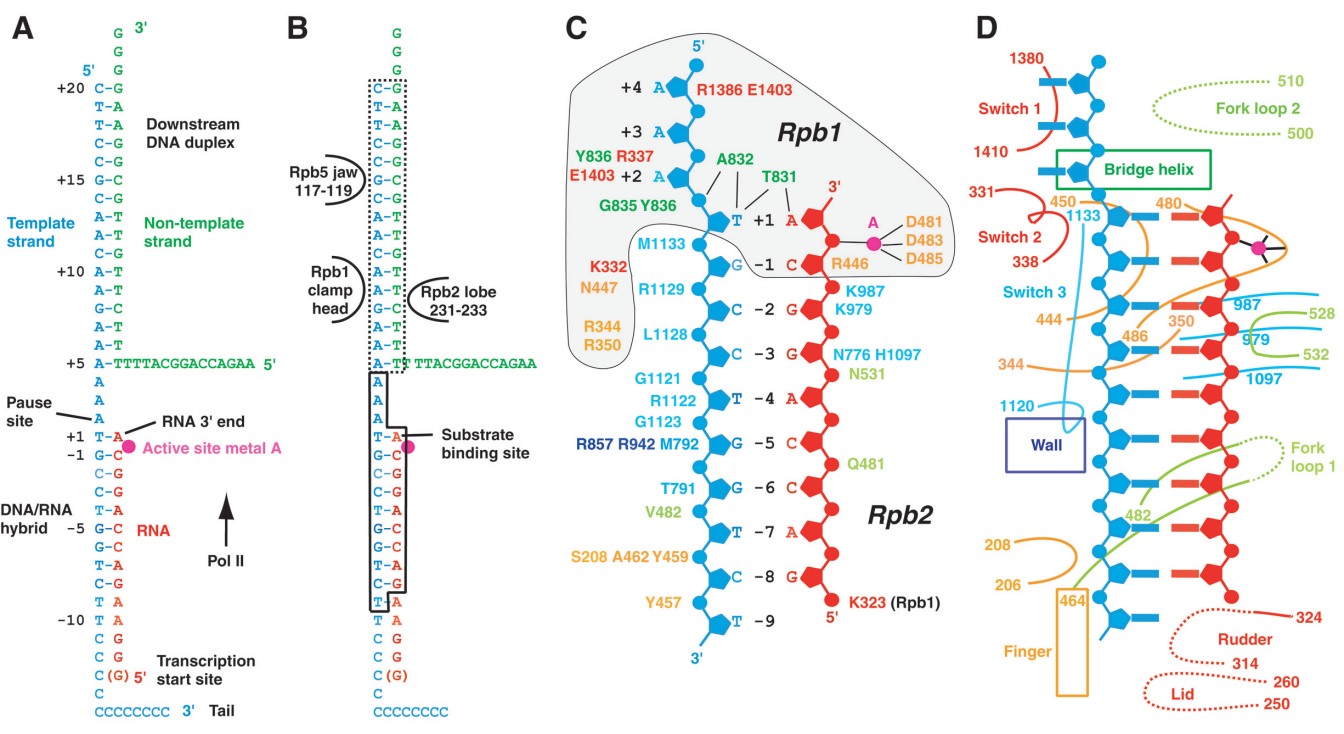
Fig. 4. Maintenance of the transcription bubble. (**A**) Schematic representation of nucleic acids in the transcribing complex. Solid ribbons represent nucleic acid backbones from the crystal structure. Dashed lines indicate possible paths of nucleic acids not present in the structure. (**B**) Protein elements proposed to be involved in maintaining the transcription bubble. Protein elements from Rpb1 and Rpb2 are shown in silver and gold, respectively.

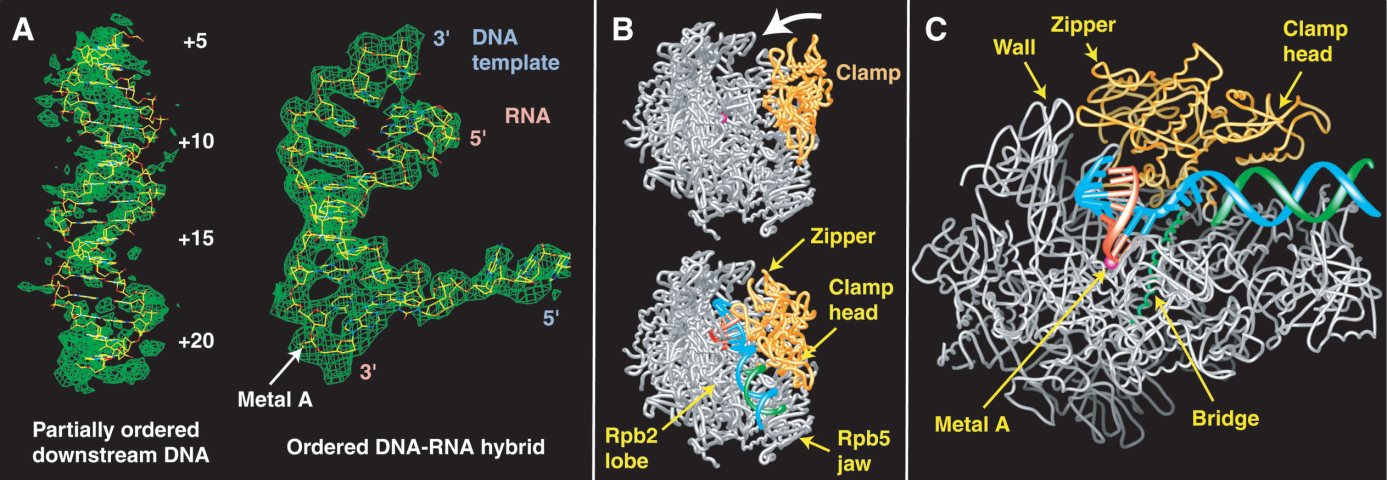
Fig. 5. DNA-RNA hybrid conformation. The view is similar to that in Fig. 2C. The conformation of the DNA-RNA hybrid is intermediary between canonical A- and B-DNA. DNA, blue; RNA, red.

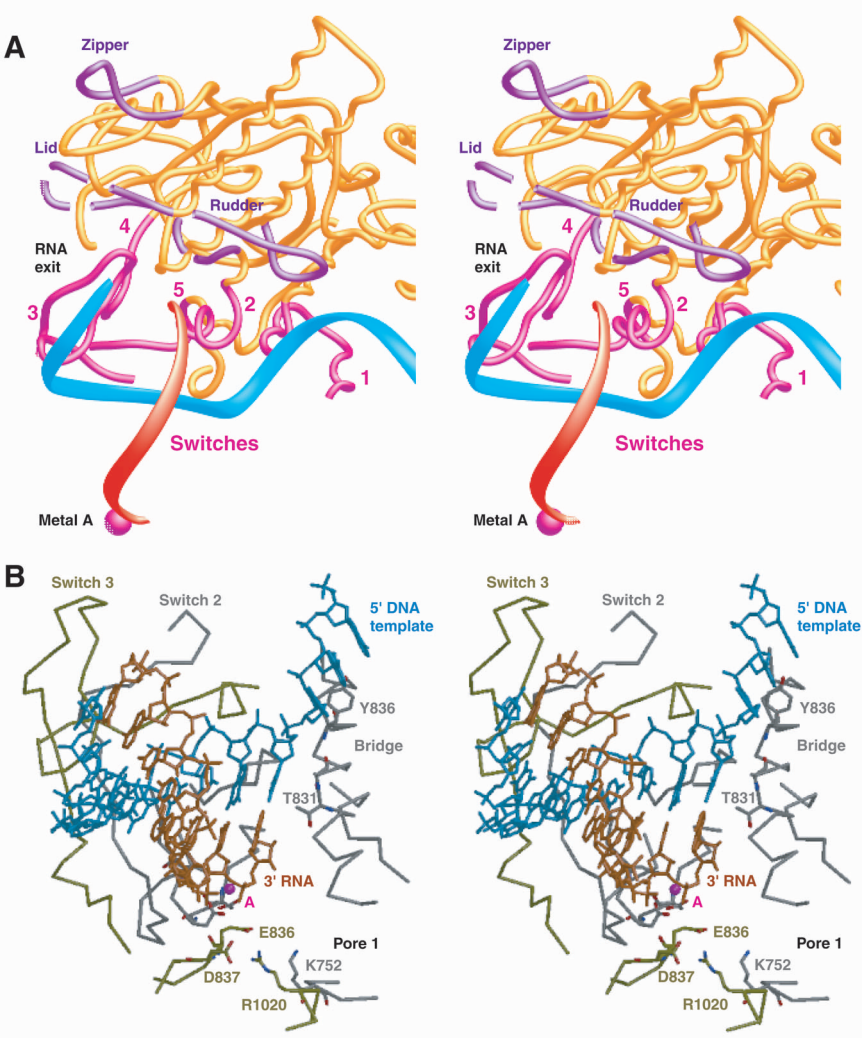
Fig. 6. Proposed transcription cycle and translocation mechanism. (**A**) Schematic representation of the nucleotide addition cycle. The nucleotide triphosphate (NTP) fills the open substrate site (top) and forms a phosphodiester bond at the active site ("Synthesis"). This results in the state of the transcribing complex seen in the crystal structure (middle). We speculate that “Translocation” of the nucleic acids with respect to the active site (marked by a pink dot for metal A) involves a change of the bridge helix from a straight (silver circle) to a bent conformation (violet circle, bottom). Relaxation of the bridge helix back to a straight conformation without movement of the nucleic acids would result in an open substrate site one nucleotide downstream and would complete the cycle. (**B**) Different conformations of the bridge helix in pol II and bacterial RNA polymerase structures. The view is the same as in Fig. 2C. The bacterial RNA polymerase structure (2) was superimposed on the pol II transcribing complex by fitting residues around the active site. The resulting fit of the bridge helices of pol II (silver) and the bacterial polymerase (violet) is shown. The bend in the bridge helix in the bacterial polymerase structure causes a clash of amino acid side chains (extending from the backbone shown here) with the hybrid base pair at position +1.

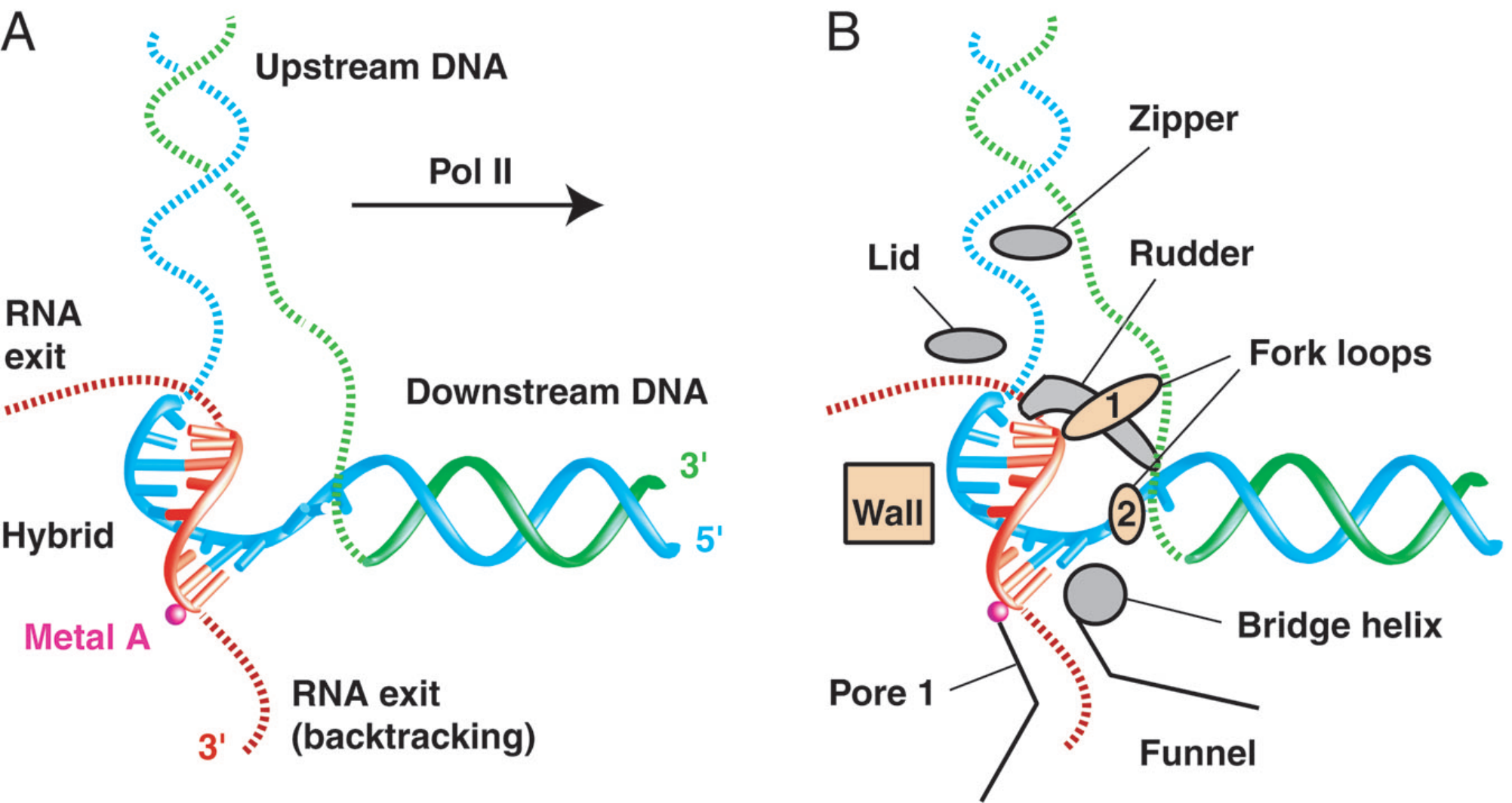
Table 1. Switch regions.

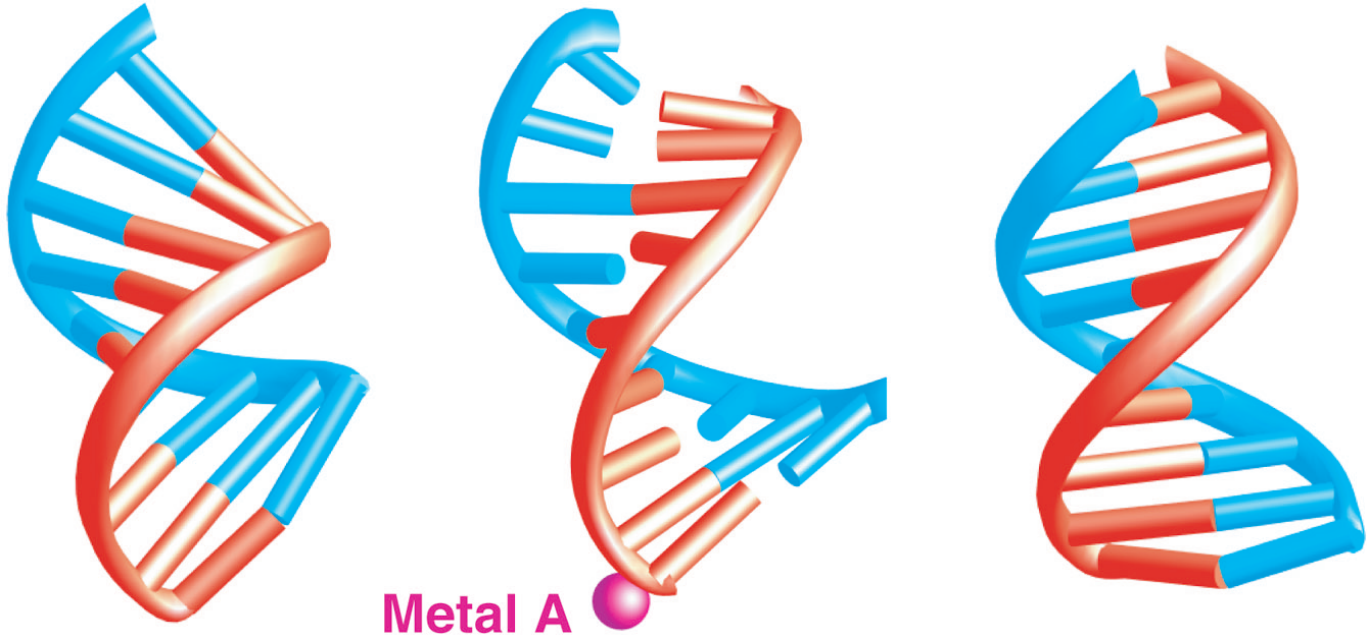
Switch	Subunit	Domain	Residues	DNA contact	Structural changes upon clamp closure
1	Rpb1	Cleft-Clamp core	1384-1406	+1 to +4	two short helices formed (α47a, α47b)
2	Rpb1	Clamp core	328-346	-2, -1, +2	helical turn flipped out
3	Rpb2	Hybrid binding-Anchor	1107-1129	-5 to -1	loop becomes ordered
4	Rpb2	Clamp	1152-1159	-	one turn added to helix α32 in the anchor region
5	Rpb1	Clamp core	1431-1433	-	hinge-like bending

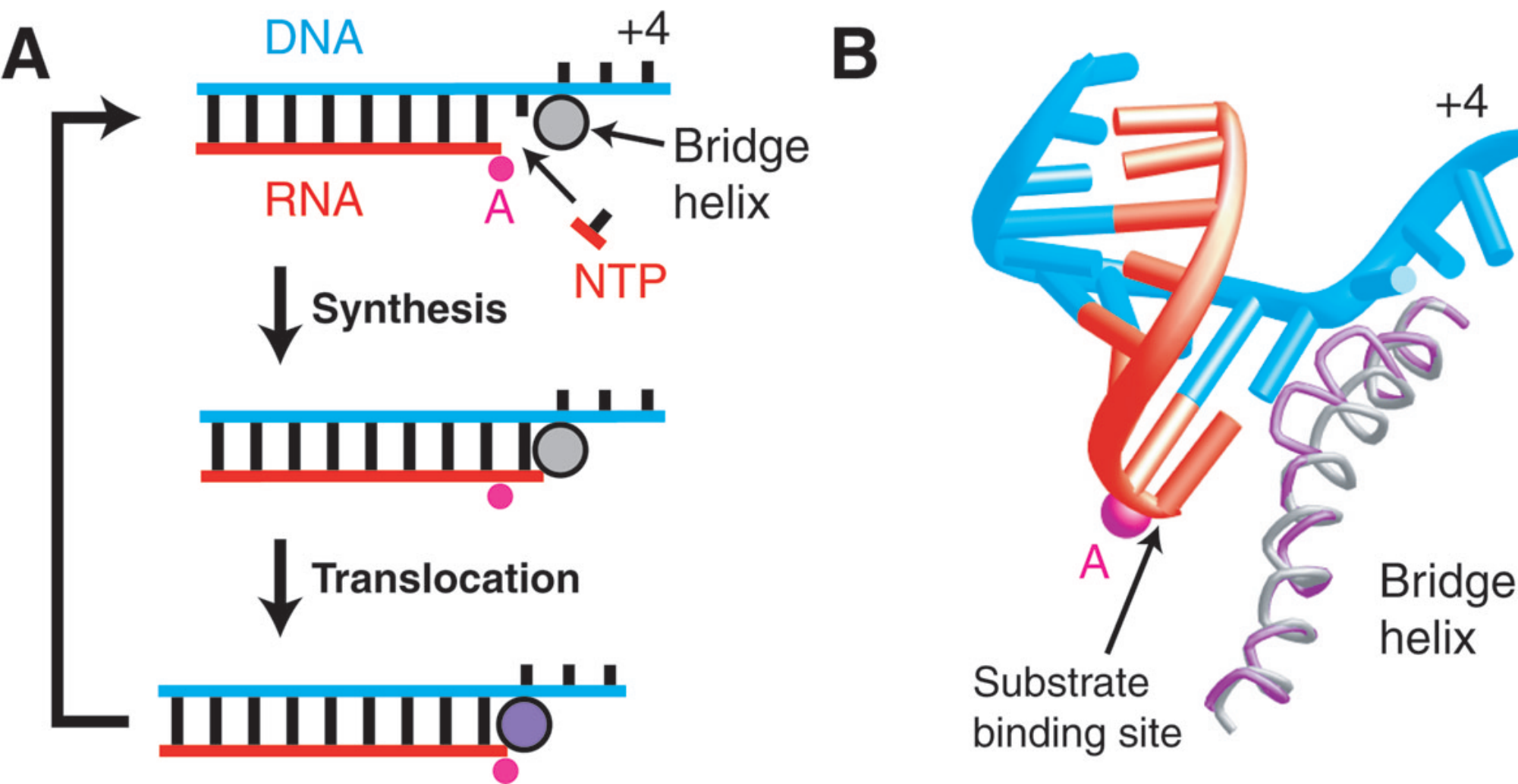

A
B
C
D
Downstream DNA duplex
Template strand
Non-template strand
Pause site
RNA 3' end
Active site metal A
DNA/RNA hybrid
RNA
Pol II
Transcription start site
Tail
Rpb5 jaw 117-119
Rpb1 clamp head
Rpb2 lobe 231-233
Substrate binding site
Rpb1
Rpb2
R1386 E1403
Y836 R337 E1403
A832
T831
G835 Y836
M1133
K332 N447
R1129
R344 R350
L1128
G1121
R1122
G1123
R857 R942 M792
T791
V482
S208 A462 Y459
Y457
D481 D483 D485
R446
K987 K979
N776 H1097 N531
Q481
K323 (Rpb1)
Switch 1
Fork loop 2
Bridge helix
Switch 2
Switch 3
Wall
Fork loop 1
Finger
Rudder
Lid


A
+5
+10
+15
+20
3'
DNA
template
RNA
5'
5'
3'
Metal A
Partially ordered
downstream DNA
Ordered DNA-RNA hybrid
B
Clamp
Zipper
Clamp
head
Rpb2
lobe
Rpb5
jaw
C
Wall
Zipper
Clamp
head
Metal A
Bridge


A
Zipper
Lid
Rudder
RNA exit
4
5
2
3
1
Switches
Metal A
Zipper
Lid
Rudder
RNA exit
4
5
2
3
1
Switches
Metal A
B
Switch 3
Switch 2
5' DNA template
Y836
Bridge
T831
3' RNA
A
E836
Pore 1
D837
K752
R1020
Switch 3
Switch 2
5' DNA template
Y836
Bridge
T831
3' RNA
A
E836
Pore 1
D837
K752
R1020


A
Upstream DNA
Pol II
RNA exit
Downstream DNA
3'
5'
Hybrid
Metal A
3'
RNA exit (backtracking)
B
Zipper
Lid
Rudder
Fork loops
1
2
Wall
Bridge helix
Pore 1
Funnel


Metal A
A-DNA
DNA-RNA Hybrid
B-DNA


A
DNA
+4
RNA
A
Bridge
helix
NTP
Synthesis
Translocation
B
+4
A
Bridge
helix
Substrate
binding site