ARTICLES

Structural basis for substrate loading in bacterial RNA polymerase

Dmitry G. Vassylyev[1], Marina N. Vassylyeva[1], Jinwei Zhang[2], Murali Palangat[3], Irina Artsimovitch[5] & Robert Landick[3,4]

The mechanism of substrate loading in multisubunit RNA polymerase is crucial for understanding the general principles of transcription yet remains hotly debated. Here we report the 3.0-Å resolution structures of the *Thermus thermophilus* elongation complex (EC) with a non-hydrolysable substrate analogue, adenosine-5′-[(α,β)-methyleno]-triphosphate (AMPcPP), and with AMPcPP plus the inhibitor streptolydigin. In the EC/AMPcPP structure, the substrate binds to the active ('insertion') site closed through refolding of the trigger loop (TL) into two α-helices. In contrast, the EC/AMPcPP/streptolydigin structure reveals an inactive ('preinsertion') substrate configuration stabilized by streptolydigin-induced displacement of the TL. Our structural and biochemical data suggest that refolding of the TL is vital for catalysis and have three main implications. First, despite differences in the details, the two-step preinsertion/insertion mechanism of substrate loading may be universal for all RNA polymerases. Second, freezing of the preinsertion state is an attractive target for the design of novel antibiotics. Last, the TL emerges as a prominent target whose refolding can be modulated by regulatory factors.

The mechanism of substrate selection and loading into the active site is of central importance in the studies of the multisubunit RNA polymerases (RNAPs), which mediate gene expression in all cellular organisms and in which substrate loading is tightly coupled to the overall processivity, fidelity and regulation of transcription. Not only will understanding of this mechanism illuminate the basic principles of transcription, but it will also have important medical applications by facilitating the design of novel antibiotics targeting the nucleotide addition cycle (NAC) of bacterial RNAP.

Structural snapshots of the substrate-bound transcription intermediates obtained by X-ray crystallography have provided a clear view of the NAC in single-subunit RNAPs, suggesting a two-step substrate loading mechanism[1,2]. In the T7 EC, the incoming substrate NTP first binds to the inactive ('preinsertion') conformation. In the preinsertion state the substrate is base-paired with the acceptor template base while the catalytic O/O′ helices anchoring the NTP phosphates are in the open configuration, thereby positioning the substrate several ångströms from the catalytic centre. Second, the EC isomerizes into the catalytically active ('insertion') state by closure of the active site after rotation of the catalytic O/O′ helices and the concerted movement of the substrate into the catalytic centre.

In contrast, the pathway of substrate loading into the active site of multisubunit RNAPs remains a matter of debate[3]. Although an analogous two-step mechanism was proposed for bacterial enzymes[1,4], one group working on a structural analysis of eukaryotic RNAPs reported findings of a templated NTP in the preinsertion site[5], whereas another emphasized a template-independent site of entry[6] (overlapping, but distinct from, a previously proposed E-site whose existence was inferred from an analysis of RNAP active-site exonuclease activity, in the location of the nucleotide base[7]) while ruling out the existence of the preinsertion state. Moreover, 'multisubstrate' models based on kinetic analyses posit the existence of at least two other sites to which substrate can bind simultaneously with the insertion site[8,9]. These apparently irreconcilable models of substrate selection and loading underscore the need for additional investigation including high-resolution structural analysis.

We have determined the crystal structures of the *T. thermophilus* EC (ttEC[10],) with the non-hydrolysable substrate analogue AMPcPP, and with AMPcPP and antibiotic streptolydigin (Stl) at 3.0 Å resolution (Fig. 1, Supplementary Fig. 1 and Supplementary Table 1). Our results argue that NTP loading occurs in two steps: first, substrate binding to an open, inactive ('preinsertion') state, and second, refolding of the TL into two trigger helices (TH) that stabilize the closed, catalytically active ('insertion') intermediate (Fig. 1). Stl induces displacement of the TH, thereby freezing the EC in the inactive state (Fig. 1).

Insertion site: closing by folding of the TL

In the ttEC/AMPcPP complex (Fig. 1a), formation of a closed substrate-bound intermediate occurs on folding of the TL (β′1236–1254) into two α-helical segments that extend the α-helices flanking the TL to form two long anti-parallel helices (TH, β′1221–1265); together with the adjacent bridge helix (BH, β′1066–1103), this creates a three-helical bundle (Fig. 2a and Supplementary Fig. 2a). TL refolding reduces the dimensions of the secondary channel[11] at the entrance to the active site from ~15 Å × 22 Å in the substrate-free EC to ~11 Å × 11 Å in the substrate-bound EC (Supplementary Fig. 2b, c). This constriction probably hinders substrate dissociation and competition but would still permit the entry of regulators that are thought to directly remodel the RNAP catalytic centre on binding through the secondary channel[12]. Closure of the active site, although not accompanied by the TL→TH transition, was also proposed to occur in eukaryotic EC[13].

In the substrate-free EC, where TL is disordered[10], certain structural elements may adopt a conformation that is unfavourable for TH formation. Superposition of the ttEC and ttEC/AMPcPP structures (root mean squared deviation 0.37 Å over the 500 C_α atoms) revealed local alterations near the active site apparently triggered by substrate binding. First, the rearrangement of the downstream DNA (dwDNA) is accompanied by a repositioning of fork loop 2 (fork 2, β413–451)

[1]Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Schools of Medicine and Dentistry, 402B Kaul Genetics Building, 720 20th Street South, Birmingham, Alabama 35294, USA. [2]Department of Biomolecular Chemistry, [3]Department of Biochemistry and [4]Department of Bacteriology, University of Wisconsin, Madison, Wisconsin 53706, USA. [5]Department of Microbiology, The Ohio State University, 484 West 12th Avenue, Columbus, Ohio 43210, USA.

(Supplementary Fig. 3b) that is likely functionally significant: the fork 2 amino-terminal hairpin stacks on the $(i+2)$ base pair, whereas its carboxy-terminal portion contributes to the formation of the closed active site, thereby bridging the dwDNA and substrate in the EC[10]. Second, a ~1.5-Å shift of the central portion of the BH (β′1077–1095) positions the BH between the TH, so that the BH side chains fit into the grooves of these helices. This BH displacement (which differs from the BH bending/unfolding proposed to control translocation and NTP loading[14–16]) seems crucial for the TL→TH transition: it supports an optimal network of contacts within the three-helical bundle while eliminating steric clash that would occur between the TH and BH positioned as in the substrate-free EC (Supplementary Fig. 3b, c; Phe β′1241 from TH and Thr β′1084 from BH would produce the major clash; an alignment of all the residues discussed in this paper with *Escherichia coli* and yeast RNAPs is presented in Supplementary Table 2).

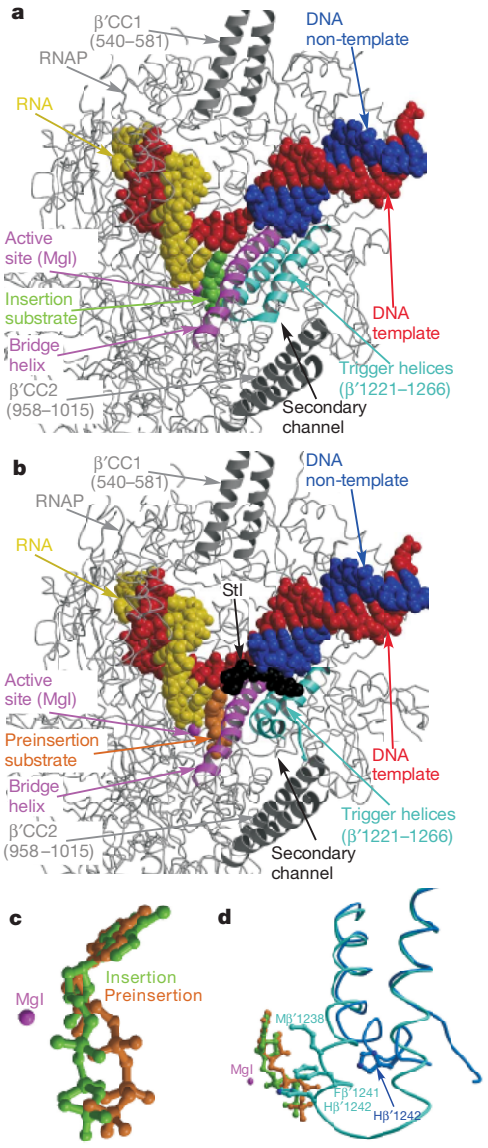


Figure 1 | Structures of the substrate complexes. The same colour scheme is used in all figures. The DNA template, non-template and RNA strands are in red, blue and yellow, respectively. The BH, the TH and the rest of the RNAP molecule are in magenta, cyan and grey, respectively. The insertion and preinsertion NTP analogues and Stl are designated by green, orange and black, respectively. The catalytic Mg^{2+} ions (MgI and MgII) are shown as magenta spheres. **a**, **b**, Overall views of the ttEC/AMPcPP (**a**) and EC/AMPcPP/Stl (**b**) complexes. CC, coiled coil. **c**, **d**, Superposition of the NTPs (**c**, **d**) and TH (**d**) in the insertion (green, NTP; cyan, TH) and preinsertion (orange, NTP; blue, TH) complexes.

Insertion site: substrate binding

In the ttEC/AMPcPP complex, AMPcPP forms a Watson–Crick base pair with the acceptor template sandwiched between the 3′ end of the RNA/DNA hybrid on one side and the protein residues from the TH and BH on the other. Met β′1238 (TH) and Thr β′1088 (BH), which stack directly on the substrate and template bases, respectively, seem essential for the positioning and selection of the substrate (Fig. 2); the corresponding RNAPII residues Leu 1081 and Thr 831 seem to have similar functions in the yeast EC[13]. In agreement with modelling and biochemical studies[1,4,15], Asn β′737 forms hydrogen bonds with both O3′ and O2′ of the substrate ribose, thereby permitting discrimination against the non-cognate dNTPs lacking either or both of these atoms (Fig. 2b, c). This is in contrast with the eukaryotic EC structures in which the Asn β′737 counterpart interacts only with the O3′ atom[13]. However, in agreement with the eukaryotic structures, Arg β′704 forms hydrogen bonds bridging the O2′ atom of the substrate and the 3′ RNA nucleotide (Fig. 2c and Supplementary Fig. 4), which is consistent with its proposed role in discriminating between rNTP and dNTP.

The substrate phosphates reside in the active site and adopt a folded conformation that permits the concerted coordination of both the high-affinity (MgI) and low-affinity (MgII) catalytic Mg^{2+} ions, which are located ~3.9 Å apart (as was also observed in the high-resolution 'insertion' structures of T7 RNAP and DNA polymerase[2,17]). All three phosphates coordinate MgII, and the α-phosphate additionally binds MgI. As expected, the catalytic Asp triad (β′739, β′741 and β′743) provides three ligands to MgI (Fig. 2d and Supplementary Figs 4a and 5). However, in contradiction of the expectations that MgII would be directly coordinated by two β′ triad Asp residues and by the invariant Asp β686 side chain[6,18], only one triad residue, Asp β′739, bridges both Mg^{2+} ions (Fig. 2d and Supplementary Figs 4a and 5). A similar coordination pattern was observed in DNA polymerases[2,17].

The electron density (ED) reveals a water molecule near MgII (Supplementary Fig. 6) whose position is fixed by the substrate phosphates

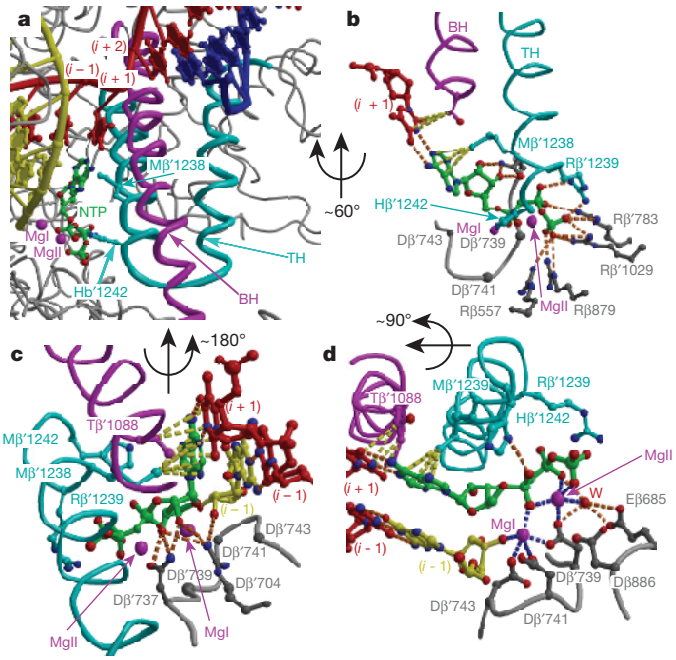


Figure 2 | The insertion complex. **a**, Close-up view through the main channel of the folded TH closing the RNAP active site. **b**–**d**, Three views showing a three-dimensional pattern of protein direct (**b**) and Mg^{2+}-mediated (**d**) interactions with the AMPcPP phosphates, discriminating contacts with the ribose (**c**) and stacking interactions with the base (**b**–**d**) of AMPcPP. The rotations necessary to obtain the views in **b** and **c** relative to **a** and of **b** relative to **d** are indicated. Hydrogen bonds, van der Waals interactions and Mg^{2+} coordination bonds are shown by orange, yellow and blue dashed lines, respectively.

and several protein residues including invariant Glu β685 (W in Fig. 2d and Supplementary Figs 4a and 5). In the final structure, two ligands for the metal ions are missing (one for each Mg^{2+} ion); however, their positions can be inferred because all other bonds agree well with the octahedral Mg^{2+} coordination. Modelling shows that both missing bonds may be donated by a single water molecule located within the interacting distance of Asp β686 (pW in Supplementary Fig. 7), which may have a catalytic function. The direct and indirect (water-mediated) coordination of the catalytic metal ions by the β′ triad and the β acidic pair, respectively, is consistent with the observations that replacements of any of the β′ Asp residues[19–21] confer a much more deleterious effect than those of the β residues[7,22–25].

The conformation of the substrate phosphates is stabilized through 11 hydrogen bonds with six basic residues, among which four were predicted from the yeast EC structure[5] and modelling of the bacterial substrate complex[1]; the other two, Arg β′1239 and His β′1242, belong to the TH (Fig. 2b and Supplementary Fig. 4a). Together with two additional TH residues, Met β′1238 and Phe β′1241, these residues form a complex network of interactions with the substrate that may determine the optimal shape of the closed active-site cavity and favour the proper loading of cognate NTPs (see below) while excluding non-cognate NTPs.

Taken together, the highly specific mode of substrate binding, the arrangement and coordination of the catalytic Mg^{2+} ions, the stable closure of the active site by the TH, and the apparent similarity to functionally homologous systems[2,17] strongly argue that the ttEC/AMPcPP structure corresponds to the catalytically active insertion transcription intermediate.

Preinsertion state: Stl-induced displacement of the TH

Our structural and functional analyses of the mechanism of Stl[26], which inhibits bacterial RNAP during both initiation and elongation, led us to conclude that Stl does not compete with the substrate but rather stabilizes the inactive, substrate-bound preinsertion state, presumably by interfering with closure of the active site (but see ref. 27). The ttEC/AMPcPP/Stl structure confirms this prediction: whereas both Stl and AMPcPP were bound to the ttEC, the TH are only partly folded and are displaced from their position in the ttEC/AMPcPP complex by as much as 15 Å, leaving the active site open (Fig. 3a).

Stl binds in the dwDNA channel along the BH, ~20 Å and ~10 Å from the β′ Asp triad and substrate base, respectively (Fig. 3a and Supplementary Fig. 8). The Stl contacts to the ttEC are different from those in the RNAP/Stl complex[26] by a few polar and multiple van der Waals interactions that Stl forms along its entire length with the dwDNA (Supplementary Fig. 8b, c). The effect of Stl on TH displacement is probably indirect and consists of two major components. First, binding of the streptolol ring in the pocket formed by the BH fragment (β′1082–1086) and fork 2 and its van der Waals contacts with the acceptor base would hinder their movements after substrate binding in the insertion complex; in particular, the BH in the Stl-induced conformation would probably clash with the closed TH, as described above for a substrate-free EC. Second, on the opposite side of the Stl molecule, the acetamide moiety stacks on Ile β′1260 and Pro β′1257, slightly displacing the latter (Supplementary Fig. 8b, c); these interactions might disturb the network on the interhelical TH contacts required for the proper folding of this TL segment. As a result, the TL C terminus (β′1251–1255) is in an extended rather than an α-helical conformation, whereas the TH tip (β′1244–1250) seems disordered (Figs 1d and 3a). The N-terminal half of the TL retains the α-helical configuration; however, the helix is sharply bent, allowing it to pack in the dwDNA channel near Stl (and far from the substrate), and makes extensive hydrophobic interactions with both the BH and Stl.

Preinsertion state: substrate binding

Apart from TL displacement from the active position, Stl binding does not affect any structural elements essential for substrate binding in the active site. Consistently, our structural data show that substrate may still bind with high affinity and specificity in the open conformation. However, the open intermediate seems catalytically inactive as a result of the displacement of the substrate phosphates (a ~35° rotation occurring approximately about the C4′–C5′ bond resulting in a ~2.5 Å overall shift; Fig. 1c). Although still in the folded configuration that is capable of anchoring the MgII, the phosphates reside too far from the MgI (~5.5 Å to the α-phosphate) to permit the catalytic reaction (Fig. 2b–d and Supplementary Figs 4b and 9). Given that the base-specific and ribose-specific interactions are nearly identical to those in the insertion complex (see below), the ttEC/AMPcPP/Stl structure probably represents a structural snapshot of a template-dependent inactive substrate complex. This complex may be functionally similar to the T7 preinsertion complex, even though the difference in preinsertion and insertion substrate positions is less marked than was observed in the T7 EC.

Whereas the absence of the folded TH probably accounts for the inactive orientation of the triphosphate, its binding is mediated by 15 hydrogen bonds with four basic protein side chains (Fig. 3b) and should be at least as stable as in the insertion complex. Among these, Arg β557 donates only one hydrogen bond to the substrate in the insertion complex vs eight hydrogen bonds bridging the α- and γ-phosphates in the preinsertion complex, thereby probably compensating for the lack of the polar interactions with the TH in the preinsertion state (Fig. 3b and Supplementary Fig. 4). Four basic residues form a rim around the NTP phosphates in both the preinsertion and insertion states (Arg β557, Arg β879, Arg β′783 and Arg β′1029; Figs 2b and 3b, and Supplementary Fig. 4)[1,5]. The rim is stabilized through stacking interactions between the Arg side chains that largely determine and restrain their orientations. Replacement of one of these residues may increase their mobility and induce rearrangement of the remaining basic residues in the rim that would allow them to maintain the same number of contacts with NTP as in the wild-type enzyme. Indeed, the substitution of Arg β557 (β R678 in *E. coli*) for Cys does not have large effects on

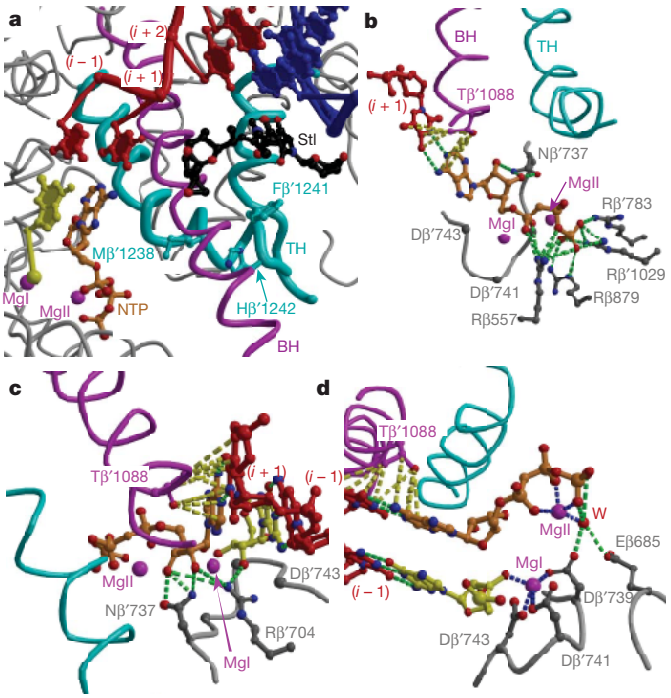
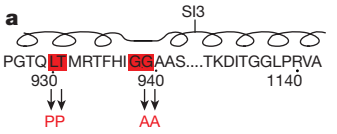
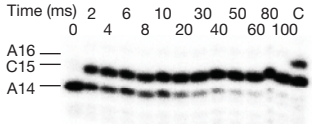


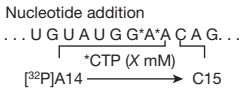
Figure 3 | The preinsertion complex. a, The TH in the Stl-bound EC. **b–d**, Three views (in similar orientations to those in Fig. 2b–d) showing a three-dimensional pattern of protein direct (**b**) and Mg^{2+}-mediated (**d**) interactions with the AMPcPP phosphates, discriminating contacts with the ribose (**b**) and stacking interactions with the base (**b–d**) of AMPcPP. Hydrogen bonds, van der Waals interactions and Mg^{2+} coordination bonds are shown by green, yellow and blue dashed lines, respectively.

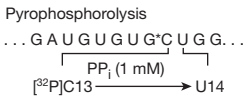
the substrate affinity or the catalytic rate[25]. Whereas in the insertion state all three NTP phosphates have protein ligands, in the preinsertion state the β-phosphate lacks interactions with the protein; this suggests that the triphosphate adopts a somewhat relaxed conformation that resists loading into the catalytic position. The basic rim may therefore serve as a 'gate' that prevents catalysis in the open EC conformation. Tyr 639, which sterically blocks the active site in the preinsertion state, may have a similar function in T7 RNAP[1]. Isomerization into the insertion state would require the gate to 'open', probably in response to TH folding that might alter the NTP interactions to facilitate its release from the preinsertion contacts and at the same time stabilize the insertion configuration. His β′1242 might be important for this transition, because in the TH configuration it clashes most severely with the preinsertion NTP phosphates; however, in the model the clash also occurs with the TH backbone around position β′1239, suggesting that TH folding may be the key for the preinsertion/insertion transition.

Despite the lack of contacts with the TH in the open active site, the base and sugar substrate moieties occupy positions similar to those in the insertion state (root mean squared deviation ~0.6 Å); thus, both Watson–Crick base pairing with the template and ribose-discriminating contacts with Asn β′737 (Fig. 3b–d and Supplementary Fig. 4b) are already established in the preinsertion state. Interestingly, Thr β′1088 (BH), which in the insertion complex stacks only on the acceptor template base (Fig. 2), forms additional stacking interactions with the NTP base in the preinsertion state that seem to compensate for the absence of Met β′1238 (TH). The BH and, in particular, Thr β′1088 may therefore serve as an important fidelity determinants in both the preinsertion and insertion states. Consistently, substitution of Thr β′1088 for Val, which would enhance van der Waals interactions with the NTP/DNA template base pair, also enhanced selectivity for rNTP over dNTP[4]. The effect of Asp β554 mutations on transcription fidelity[25] may be also explained by the repositioning of the BH resulting from these substitutions: in both complexes Asp β554 resides far from the NTP but in close proximity to the BH.









d

RNAP	$V_{25\ \mu m}$ (s^{-1})	$CTP_{0.5}$ (μm)	V_{max}^{app} (s^{-1})	V_{PP_i}(min^{-1})
Wild type*	130 ± 30	~140†	~800†	n.d.
Wild type	230 ± 80	~50†	~500†	0.9 ± 0.1
ΔSI3	110 ± 30	70 ± 40	440 ± 120	2 ± 1
LT → PP	0.008 ± 0.001	130 ± 20	0.04 ± 0.004	<0.01
GG → AA	0.4 ± 0.1	25 ± 10	1.0 ± 0.1	0.1 ± 0.01

*Ref.8
†Not accurately measurable with 2 ms reactions.

Figure 4 | Effect of TL alterations on nucleotide addition rate. a, Location of TL alterations in the *E. coli* RNAP. The diagram above the sequence indicates the position of the TH. SI3 denotes the location of SI3 in wild-type *E. coli* RNAP. *E. coli* LT930,1 and GG938,9 correspond to *T. thermophilus* LT1236,7 and GG1244,5, respectively. **b**, Sequence of the transcription scaffolds surrounding the sites used to assay nucleotide addition and pyrophosphorolysis (see Methods). C15 corresponds to C25 in ref. 8. 'X mM CTP' represents the range of CTP concentrations used in the assay. **c**, Electrophoretic separation of reaction products from a quench-flow assay of wild-type RNAP at 25 μM CTP (see Methods). **d**, Kinetic parameters measured for wild-type, ΔSI3, PP and AA RNAPs. n.d., not determined. Values estimated from Fig. 1C in ref. 8 are shown for comparison. $CTP_{0.5}$, the concentration of CTP giving half-maximal velocity, is reported rather than K_{NTP} because equilibrium binding may not be a valid assumption. The rate of pyrophosphorolysis (V_{PP_i}) was determined at 1 mM PP_i. Errors are s.e.m. and were calculated as described in Methods.

TL folding is essential for catalysis

The TH are apparently less flexible than the unfolded TL, because their conformation is additionally stabilized through intrahelical and interhelical interactions, as well as through contacts with the adjacent protein segments. However, the TH may be unstable in the absence of the latter contacts, suggesting that the substrate not only anchors the TL, as implied by the eukaryotic TL model[13], but may also induce changes that create a network of the protein–protein contacts along the entire TH length to stabilize the flexible portion of the TL. Indeed, even subtle alterations would disrupt these crucial interactions and could alter folding/positioning of the TH even in the presence of the substrate (for example, as observed in the ttEC/AMPcPP/Stl structure).

To test the hypothesis that folding of the TL, rather than its movement in the unfolded state[13], is a crucial component of the substrate-loading mechanism in bacterial RNAPs, we created substitutions within the TH that do not involve the substrate-interacting residues and should not affect seriously the movement and conformation of the unfolded TL, but would instead have a specific effect on the TL→TH transition (Leu-Thr→Pro-Pro and Gly-Gly→Ala-Ala; Fig. 4a). Replacement of Leu-Thr with Pro-Pro (LT→PP) should completely block TL folding, whereas replacement of Gly-Gly with Ala-Ala (GG→AA) should restrict flexibility of the turn between the folded TH and perturb their packing and folding. We purified mutant *E. coli* RNAP enzymes containing these substitutions but lacking SI3, a 188-residue insertion that is absent from the *T. thermophilus* TL. Deletion of SI3 has only modest effects on nucleotide addition[28]. We tested the effect of the PP and AA substitutions on the rate of CMP addition in TECs directly reconstituted on a DNA/RNA scaffold (see Methods), using a sequence previously found to support rapid nucleotide addition (Fig. 4b)[8]. With wild-type *E. coli* RNAP, we observed rates at saturating, half-maximal and 25 μM CTP that were comparable to those reported previously[8] but too fast for kinetic analysis (Fig. 4c and Supplementary Fig. 10). Deletion of SI3 slowed nucleotide addition to a measurable but still rapid range. However, the PP and AA substitutions slowed CMP addition markedly, with the PP mutant ~10^4 slower than ΔSI3 RNAP, even at saturating CTP, and the AA mutant ~400 times slower (Fig. 4d). However, little difference was seen in the concentration of CTP required for half-maximal velocity, suggesting the principal defect was in catalysis rather than in substrate binding. Large defects in rates of pyrophosphorolysis (Fig. 4d) are consistent with this interpretation. The large effects on catalytic rate with minimal effects on substrate affinity of these substitutions argue strongly for the existence of the preinsertion state deduced from our structural analysis.

Substrate-induced opening of the dwDNA claws

Superposition of the ttEC and ttEC/AMPcPP structures by the dwDNA atoms (root mean squared deviation ~0.2 Å) revealed a substantial opening of the dwDNA claws that results in the ~4 Å displacement of the β-pincer (whose orientation in the substrate complex corresponds well to that observed in the 'open' holoenzyme structure[29]), with a concomitant loss of several polar and van der Waals interactions with the dwDNA in the substrate-bound complex (Supplementary Fig. 11). Because no such opening was observed in the ttEC/AMPcPP/Stl complex, we conclude that these alterations may arise from changes near the active site that occur on isomerization to the closed, insertion state. Indeed, in both the substrate-free and substrate-bound complexes, the N-terminal segment of fork 2 forms an extensive hydrophobic interface with the β-pincer, suggesting that its repositioning is a consequence of substrate-dependent fork 2 movement. Thus, along with TH formation, opening of the dwDNA claws seems to be a hallmark of the closed insertion complex. The insertion complex remains stable as a result of substrate binding despite the increased dwDNA mobility; however, it may become prone to translocation after addition of the nucleotide and formation of the pre-translocation complex (see below).

Nucleotide addition cycle

We propose the following pathway for the NAC in the multisubunit enzymes (Fig. 5). Before NTP loading, the EC exists in equilibrium between the pre-translocated and post-translocated states, whose interconversion is driven by thermal motions. In the pre-translocated EC, where the 3′ RNA nucleotide occupies the position taken by the substrate in the post-translocated EC, the TL→TH transition might have a function in the exo-nucleolytic activity of RNAP[7] (by promoting closure of the active site) or in facilitating translocation (by promoting the opening of the dwDNA claws, which probably accompanies TH formation). Substrate loading occurs in two steps. First, the NTP binds to the open EC to form an inactive, preinsertion intermediate, working as a 'ratchet' that stabilizes the post-translocated EC[30] and providing the first sieve for substrate selection. Second, on folding of the TL into the TH, the NTP is repositioned to form the catalytically competent insertion complex, in which the active site is closed. The inherent instability of the TH, the opening of the dwDNA claws, and the potential competition between the preinsertion and insertion contacts for the substrate triphosphate suggest that the insertion complex may be a high-energy state that is more sensitive to the shape or chemical properties of the incoming NTP than the preinsertion EC, thereby providing the second, finer sieve for substrate selection. The NAC culminates with the catalytic reaction, resulting in transcript extension and pyrophosphate release. The loss of the phosphate moieties then destabilizes the insertion complex and triggers relaxation to the ground state, which may allow transition from the closed, pre-translocated state to an open, presumably post-translocated, state[2] characterized by the unfolded TL and closed β-pincer, as observed in the substrate-free eukaryotic[5,13] and bacterial ECs[10]. However, this is unlikely to create a 'power-stroke'-like mechanism[2] because the structural alterations do not seem to form a 'ratchet' that would stabilize the post-translocated state irreversibly[10], and physical measurements[30] also suggest that translocation is best explained by an NTP-biased, brownian-ratchet model.

Comparison with T7 and eukaryotic RNAPs

In spite of their drastically different architectures, multisubunit bacterial and single-subunit phage ECs use structurally analogous strategies to handle the nucleic acids[10]. The substrate-bound ttEC

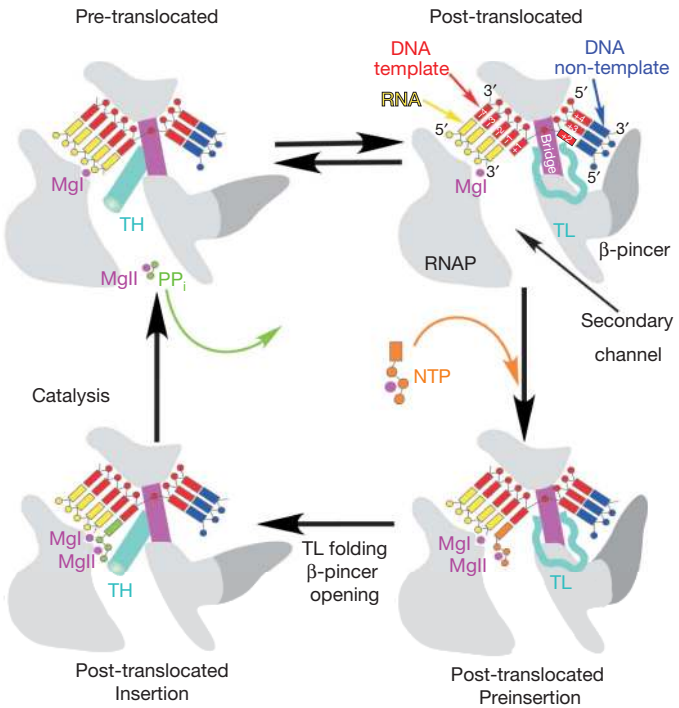


Figure 5 | Nucleotide addition cycle. The substrate loading pathway in bacterial RNAP.

structures revealed additional similarities between these systems. First, in both enzyme classes substrate loading occurs in two steps, in which the inactive, preinsertion state probably provides quality control for the initial substrate before its delivery to the insertion state. Second, both systems use preinsertion 'gates' to prevent premature substrate alignment in the catalytic centre. Third, both types of insertion complex might represent high-energy states used for the final selection of the cognate substrate. Last, the structural elements that control the transition from open to closed configurations seem strikingly similar in both conformation (two anti-parallel α-helices called TH and O/O′-helices in the ttEC and T7 EC, respectively) and mode of substrate binding (key hydrophobic and basic residues located in the N-terminal helices bind to the substrate aligned roughly parallel to the helical axes; Supplementary Fig. 12). However, these catalytic helices have different functions in substrate loading. In the T7 EC, the folded O/O′ helices anchor the substrate in the preinsertion state, positioning the substrate and acceptor DNA template far from the active site. In contrast, in the preinsertion bacterial EC the TH are probably unfolded and bind the substrate only after their folding in the insertion complex. Moreover, whereas TH function is apparently limited to the proper positioning of the NTP in the active site, the O/O′ helices also contain the 'gate' residue Tyr 639 and an important fidelity determinant, Phe 644 (ref. 31), which is involved in the separation of the dwDNA strand.

Several medium-resolution yeast EC structures bound to the NTP substrates and their analogues have been reported[5,6,13]. As expected on the basis of the significant structural and functional similarity among all multisubunit RNAPs, our analysis of the bacterial EC has revealed numerous similarities with the yeast structures, including that with a proposed preinsertion mechanism of substrate loading[5]. Nonetheless, several key aspects of the bacterial substrate complexes differ from the eukaryotic closed EC/NTP complex[13]. The most significant differences include the dwDNA configuration and contacts, the TL conformations in the closed complexes, and the binding modes/sites for the NTP substrate. We think that these discrepancies (analysed in the Supplementary Discussion) are due largely to the differences in the crystallization conditions and nucleic-acid scaffolds, to difficulties in the interpretation of the lower-resolution structural data available for the eukaryotic complexes, or both.

A detailed comparison of the relevant bacterial and eukaryotic structures (Supplementary Discussion and Supplementary Figs 13–15) argues for the existence of a catalytically inactive preinsertion state that may have key functions in all RNAPs. First, the preinsertion substrate may act as a ratchet to stabilize the post-translocated state. Second, the substrate is able to be initially surveyed in (and rejected from) the open RNAP configuration. Third, the preinsertion state prevents premature catalysis in the open state. Last, stabilization of NTP in the preinsertion state might be used as a response to the lesions present in the dwDNA (see Supplementary Discussion). The alternative template-independent substrate entry site that was observed experimentally[6] cannot have any of these functions. However, the entry site may still function not as an alternative but rather as an additional, third substrate intermediate en route to the preinsertion state[26].

Taken together, the structural analysis and the very high conservation of the active site residues in all kingdoms of life imply that the eukaryotic and bacterial RNAPs use the same mechanism of substrate loading and catalysis. The rigorous, detailed comparison between the molecular mechanisms of bacterial and eukaryotic RNAPs awaits the structural explanation of ECs assembled on identical, more natural scaffolds for both systems.

Concluding remarks

Our conclusion that the isolated TH conformation may require stabilization through interactions with both the substrates and numerous adjacent RNAP structural elements has several implications. First, all reactions catalysed by RNAP (nucleotide addition, exonucleolytic and endonucleolytic cleavages, and pyrophosphorolysis)

probably require the TL→TH transition if they are to achieve the closed, catalytically active state. This implies that the extrinsic factors that markedly increase the RNAP intrinsic exonucleolytic (non-complementary NTP) and endonucleolytic (Gre factors) activities not only donate the missing ligands to the catalytic Mg^{2+} ions[7,32,33] but may also interact with the TH (directly or indirectly) to stabilize the closed intermediate. Indeed, the eukaryotic counterpart of the Gre-factors, TFIIS, interacts with the TL[5], and the TL→TH conversion may be required for the proper positioning of its 'catalytic' acidic residues in the RNAP catalytic centre. Second, changes in the geometry of the RNAP active site that are thought to accompany the formation of paused ECs probably involve an alternative TL/TH configuration stabilized in part through interactions of the frayed 3′ RNA nucleotide with the residues in the E-site, which may be formed in part by TL/TH residues[34]. Third, various auxiliary factors might activate long-distance allosteric signals modulating the TL↔TH switch, thereby conferring a range of regulatory effects. Thus, along with the catalytic residues and Mg^{2+} ions, the dynamic TL/TH structure emerges as a central regulatory element in the transcription complexes. Our observation that the antibiotic Stl blocks the closure of the active site and freezes the EC in the inactive state without even making direct polar interactions with the TL opens a range of opportunities for the design of novel antibiotics targeting the TL→TH transition in bacterial RNAPs.

METHODS SUMMARY

Crystallization of both complexes was performed under essentially the same conditions as for the ligand-free EC[10]. The structures were refined to a final *R*-factor/R_{free} ratio of 0.225/0.257 and 0.234/0.266 for the ttEC/AMPcPP and ttEC/AMPcPP/Stl structures, respectively, at 3.0 Å resolution (Supplementary Table 1). Both nucleotide addition and pyrophosphorolysis reactions were performed with ECs reconstituted with *E. coli* RNAP and nucleic acid scaffolds as described in Supplementary Information. For nucleotide addition, kinetic parameters (V_{max} and $NTP_{0.5}$) were determined from reactions over a wide range of substrate concentrations without assumption of any particular reaction mechanism.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 28 January; accepted 11 May 2007.
Published online 20 June 2007.

1. Temiakov, D. *et al.* Structural basis for substrate selection by t7 RNA polymerase. *Cell* **116,** 381–391 (2004).
2. Yin, Y. W. & Steitz, T. A. The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* **116,** 393–404 (2004).
3. Landick, R. NTP-entry routes in multi-subunit RNA polymerases. *Trends Biochem. Sci.* **30,** 651–654 (2005).
4. Svetlov, V., Vassylyev, D. G. & Artsimovitch, I. Discrimination against deoxyribonucleotide substrates by bacterial RNA polymerase. *J. Biol. Chem.* **279,** 38087–38090 (2004).
5. Kettenberger, H., Armache, K. J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16,** 955–965 (2004).
6. Westover, K. D., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **119,** 481–489 (2004).
7. Sosunov, V. *et al.* Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *EMBO J.* **22,** 2234–2244 (2003).
8. Foster, J. E., Holmes, S. F. & Erie, D. A. Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell* **106,** 243–252 (2001).
9. Gong, X. Q., Zhang, C., Feig, M. & Burton, Z. F. Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol. Cell* **18,** 461–470 (2005).
10. Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. H. & Artsimovitch, I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* doi:10.1038/nature05932 (this issue)..
11. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98,** 811–824 (1999).
12. Nickels, B. E. & Hochschild, A. Regulation of RNA polymerase through the secondary channel. *Cell* **118,** 281–284 (2004).
13. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D. & Kornberg, R. D. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127,** 941–954 (2006).
14. Epshtein, V. *et al.* Swing-gate model of nucleotide entry into the RNA polymerase active center. *Mol. Cell* **10,** 623–634 (2002).
15. Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* **292,** 1876–1882 (2001).
16. Bar-Nahum, G. *et al.* A ratchet mechanism of transcription elongation and its control. *Cell* **120,** 183–193 (2005).
17. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391,** 251–258 (1998).
18. Artsimovitch, I. *et al.* Structural basis for transcription regulation by alarmone ppGpp. *Cell* **117,** 299–310 (2004).
19. Dieci, G. *et al.* A universally conserved region of the largest subunit participates in the active site of RNA polymerase III. *EMBO J.* **14,** 3766–3776 (1995).
20. Sosunov, V. *et al.* The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase. *Nucleic Acids Res.* **33,** 4202–4211 (2005).
21. Zaychikov, E. *et al.* Mapping of catalytic residues in the RNA polymerase active center. *Science* **273,** 107–109 (1996).
22. Langelier, M. F. *et al.* The highly conserved glutamic acid 791 of Rpb2 is involved in the binding of NTP and Mg(B) in the active center of human RNA polymerase II. *Nucleic Acids Res.* **33,** 2629–2639 (2005).
23. Lee, J., Kashlev, M., Borukhov, S. & Goldfarb, A. A β subunit mutation disrupting the catalytic function of *Escherichia coli* RNA polymerase. *Proc. Natl Acad. Sci. USA* **88,** 6018–6022 (1991).
24. Werner, F. & Weinzierl, R. O. A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol. Cell* **10,** 635–646 (2002).
25. Holmes, S. F., Santangelo, T. J., Cunningham, C. K., Roberts, J. W. & Erie, D. A. Kinetic investigation of *Escherichia coli* RNA polymerase mutants that influence nucleotide discrimination and transcription fidelity. *J. Biol. Chem.* **281,** 18677–18683 (2006).
26. Temiakov, D. *et al.* Structural basis of transcription inhibition by antibiotic streptolydigin. *Mol. Cell* **19,** 655–666 (2005).
27. Tuske, S. *et al.* Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation. *Cell* **122,** 541–552 (2005).
28. Artsimovitch, I., Svetlov, V., Murakami, K. S. & Landick, R. Co-overexpression of *Escherichia coli* RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J. Biol. Chem.* **278,** 12344–12355 (2003).
29. Vassylyev, D. G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417,** 712–719 (2002).
30. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438,** 460–465 (2005).
31. Huang, J., Brieba, L. G. & Sousa, R. Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* **39,** 11571–11580 (2000).
32. Laptenko, O., Lee, J., Lomakin, I. & Borukhov, S. Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* **22,** 6322–6334 (2003).
33. Sosunova, E. *et al.* Donation of catalytic residues to RNA polymerase active center by transcription factor Gre. *Proc. Natl Acad. Sci. USA* **100,** 15469–15474 (2003).
34. Toulokhonov, I., Zhang, J., Palangat, M. & Landick, R. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol. Cell* (in the press)..

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Svetlov for help with sequence alignments and critical reading of the manuscript. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Energy Research. This work was supported by grants from the NIH to D.G.V., I.A. and R.L.

Author Contributions D.G.V. determined and analysed the structure and supervised the project. M.N.V. performed crystallization and data collection. J.Z. and M.P. performed biochemical analysis of the trigger loop mutants under the guidance of R.L. I.A. performed biochemical analysis of the *T. thermophilus* TECs. D.G.V., I.A. and R.L. jointly wrote the manuscript.

Author Information The atomic coordinates and structure factors of the ttEC/AMPcPP and ttEC/AMPcPP/Stl complexes are deposited in the Protein Data Bank under accession numbers 2O5J and 2PPB, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.G.V. (dmitry@uab.edu).

METHODS

Crystallization and data collection. To obtain the crystals of the ttEC/AMPcPP and ttEC/AMPcPP/Stl complexes, the ttEC prepared in the same way as described previously[10] was co-crystallized with AMPcPP (3 mM) or with AMPcPP (3 mM) and Stl (3 mM), respectively, under essentially the same crystallization conditions (the well solution contained 30% 2-methyl-2,4-pentandiol, 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium acetate) as the ligand-free ttEC[35]. The crystallization trials yielded the crystals of the same size and shape as for the ligand-free ttEC that belong to the same $P4_12_12$ space group. Whereas the crystals of the ttEC/AMPcPP/Stl complex had nearly the same unit cell dimensions ($a = b = 155.38$ Å, $c = 496.99$ Å) as those of the ligand-free ttEC, the crystals of the ttEC/AMPcPP complex were characterized by a large change in the unit cell parameters ($a = b = 152.38$ Å, $c = 524.57$ Å) (Supplementary Table S1). The crystals of both complexes were improved and selected for data collection with the same strategy as for the ligand-free complex[10]. This allowed us finally to select a few well-diffracting crystals, from which the high-resolution data were collected at the SERCAT beam line (APS, Argonne, USA; Supplementary Table S1). The structure determination and refinement were performed with the same approaches and procedures as for the ligand-free ttEC[10] (Supplementary Table S1).

Determination of the structures. The ttEC/AMPcPP/Stl crystals were characterized by essentially the same unit cell parameters as those of the ligand-free ttEC. In contrast, the ttEC/AMPcPP crystals consistently demonstrated a marked increase (~25 Å) in the dimensions of the *c* axis (Supplementary Table S1). Given that all the crystals were obtained under the same conditions, this result suggests significant alterations of the enzyme structure induced by substrate binding, probably blocked in the presence of Stl, and incompatible with the previous crystal form. Indeed, the application of zonal scaling to the ttEC/AMPcPP/Stl data in the $P4_12_12$ space group yielded an initial R-factor/R_{free} ratio of ~40% that decreased to 38.5% at 3.0 Å resolution on domain rigid-body refinement, a value very similar to that of the ttEC at the same stage of refinement[10]. In contrast, the rigid-body refinement did not converge for the EC/AMPcPP structure, whereas molecular replacement with the program AMORE[36] revealed a ~7.5° rotation of the entire molecule, giving an initial R-factor/R_{free} of ~42%. The domain rigid-body refinement[36] further improved the R-factor by ~4% (38.2%), indicating more extensive domain rearrangement in the AMPcPP-bound EC than in the ttEC/AMPcPP/Stl structure. The twinning option of the CNS program[37] allowed us to improve the ED and the R-factor values further (Supplementary Table S1).

The focal point in both complexes is the active site area, in which local but functionally significant alterations may occur on substrate binding. To detect these putative alterations and to minimize the model bias, the first base pair of the RNA/DNA hybrid (position −1) and the acceptor template nucleotide (nt, register $i + 1$) were omitted from the model at the initial stages of refinement. The first base pair of the dwDNA (register $i + 2$) was also omitted: if the 'multi-substrate' hypothesis[9] were correct, the substrate would disrupt the $(i + 2)$ base pair because the scaffold contained Ts in the template strand at both $(i + 1)$ and $(i + 2)$. The resulting electron density (ED) map showed clear ED and allowed us to model[38] all the missing portions of the structure, including the $(i + 2)$ base pair[10] arguing against the 'multisubstrate' model. The refinement converged to the final R-factor/R_{free} values of 0.225/0.257 and 0.234/0.266 for the ttEC/AMPcPP and ttEC/AMPcPP/Stl structures, respectively (Fig. 1 and Supplementary Table S1). All structural figures were prepared with the programs Molscript, Bobscript and Raster3D[39–41].

Isolation and assay of mutant *E. coli* RNAPs. Core wild-type and ΔSI3 RNAPs were purified after overexpression from plasmids as described previously[28]. PP and AA mutant RNAPs were obtained by oligonucleotide mutagenesis of plasmid pRL662, transfer of a *Bsm*I–*Xho*I fragment to pRL4455 (ref. 34), and purification after overexpression by polyethyleneimine precipitation, Ni^{2+}-nitrilotriacetate chromatography and heparin-agarose chromatography. TECs were reconstituted as described[42], using for A14 TEC (nucleotide addition assay) oligonucleotides 5′-CTCTGAATCTCTTCCAGCACACATTAGGACGTACTG-ACC-3′ (template DNA), 5′-biotin-GGTCAGTACGTCCATTCGATCTCCC-GAAGAGATTCAGAG-3′ (non-template DNA) and 5′-AGGCGAUGUGUG-3′ (RNA). The reconstituted TECs were 3′-labelled by incubation with 2 μM [α-^{32}P]ATP (A14) or UTPs (each at 10 μM) and [α-^{32}P]CTP. For elongation assays, A14 TECs were diluted in buffer A (25 mM HEPES-KOH pH 8.0, 130 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol, 0.15 mM EDTA, 25 μg ml^{-1} acetylated BSA) to 25–50 nM TEC, and assayed at various concentrations of CTP at 37 °C with a KinTek rapid-quench flow apparatus. The concentration of $MgCl_2$ was adjusted to maintain Mg^{2+} in 5 mM excess of the CTP concentration. Samples were quenched by addition of 1 vol. of 2 M HCl, immediately neutralized to pH 8.0 by addition of the same volume of 3 M unbuffered Tris, extracted with phenol, precipitated with ethanol, and suspended in formamide loading dye. For pyrophosphorolysis assays, U14 TECs were immobilized on streptavidin–agarose beads (Sigma), washed extensively with buffer B (10 mM Tris-HCl pH 7.9, 20 mM NaCl, 10 mM $MgCl_2$, 14 mM 2-mercaptoethanol, 0.1 mM EDTA), incubated with PP_i (1 mM) in buffer B for 10 min at 37 °C, and dissociated by heating at 94 °C for 3 min after the addition of an equal volume of saturated urea in 90 mM Tris-borate pH 8.3, 50 mM EDTA to halt the reactions. Reaction products were separated electrophoretically on denaturing polyacrylamide gels (7 M urea, 17.5% acrylamide:bisacrylamide, 19:1), and RNA quantities were determined from Phosphorimager scans of the gels.

Reaction rates (addition of CMP to A14 ECs and pyrophosphorolysis of U14 ECs; Fig. 4d and Supplementary Fig. 10) were determined as described[8] by non-linear regression of RNA concentrations against time with errors calculated from the standard error of regression and Student's t distribution at the 95% confidence level. Apparent V_{max} and $CTP_{0.5}$ values (Fig. 4d) were calculated by non-linear regression of the average rate weighted by their respective error estimates against CTP concentration assuming a hyperbolic dependence, with error calculated as described for the rate measurements.

35. Kashkina, E. *et al.* Elongation complexes of *Thermus thermophilus* RNA polymerase that possess distinct translocation conformations. *Nucleic Acids Res.* **34,** 4036–4045 (2006).
36. Navaza, J. Implementation of molecular replacement in AMoRe. *Acta Crystallogr. D Biol. Crystallogr.* **57,** 1367–1372 (2001).
37. Brunger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54,** 905–921 (1998).
38. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
39. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24,** 946–950 (1991).
40. Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.* **55,** 938–940 (1999).
41. Merrit, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277,** 505–524 (1997).
42. Sidorenkov, I., Komissarova, N. & Kashlev, M. Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol. Cell* **2,** 55–64 (1998).