CHAPTER 7

Substrate Loading, Nucleotide Addition, and Translocation by RNA Polymerase

JINWEI ZHANG[a] AND ROBERT LANDICK[b]

[a] Departments of Biomolecular Chemistry, University of Wisconsin-Madison, 1550 Linden Drive, Madison, WI 53706, USA; [b] Department of Biochemistry and Department of Bacteriology, University of Wisconsin-Madison, 1550 Linden Drive, Madison, WI 53706, USA

7.1 Basic Mechanisms of Transcript Elongation by RNA Polymerase

After the RNA polymerase (RNAP) holoenzyme commences RNA synthesis, the initiation complex isomerizes into the elongation complex (EC) in which the RNAP translocates the template DNA and synthesizes RNA transcripts of up to $\sim 10^4$ (bacterial) to $\sim 10^6$ (eukaryotic) nucleotides. At the ends of transcription units, the EC can be dissociated by either intrinsic or factor-dependent (*e.g.*, rho) termination processes.[1–5] An elongating RNAP maintains a 12–14 base-pair transcription bubble and an 8–9 base-pair RNA-DNA hybrid, places the RNA 3′ end in an active site near the center of the enzyme, and catalyzes nucleotide (nt) addition at 15–30 nt s^{-1} (for eukaryotic RNA Polymerase II, or Pol II) or 50–100 nt s^{-1} (for bacterial RNAPs).[6–8] Each round of nucleotide addition requires several steps collectively known as the nucleotide addition cycle.[9–12] Occasionally the nucleotide addition cycle can be interrupted by pause

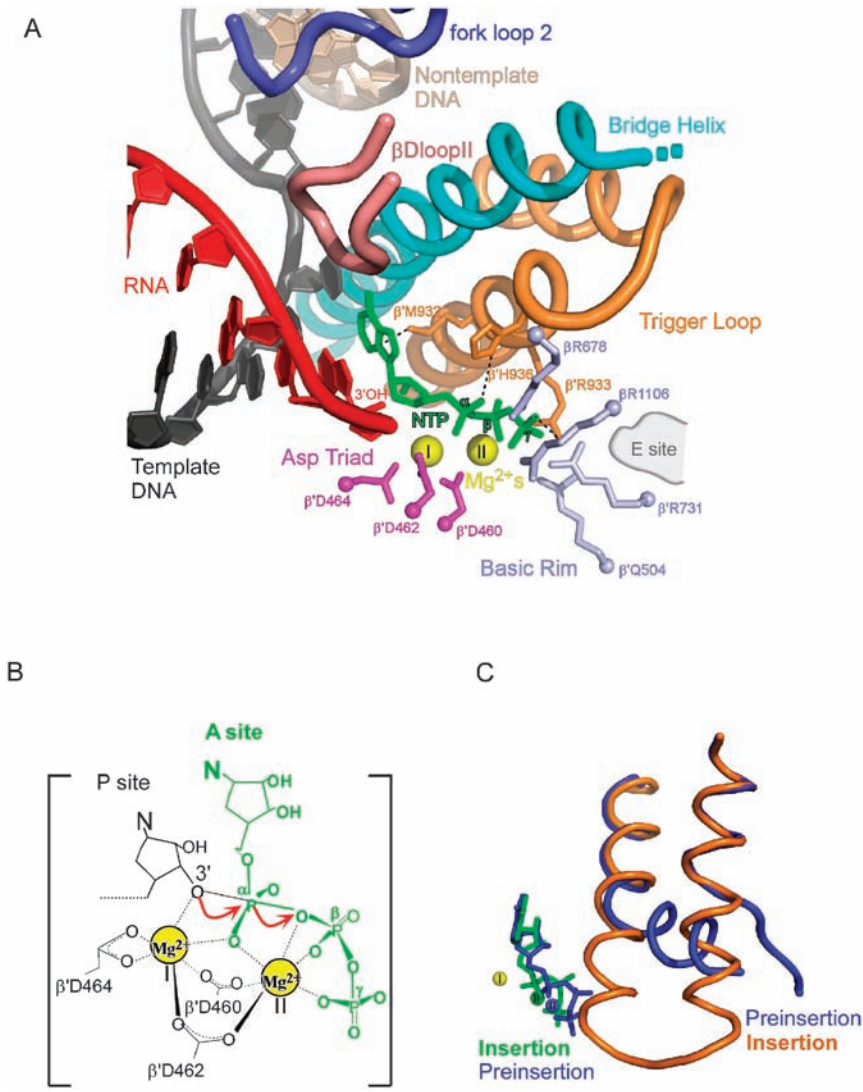
RSC Biomolecular Sciences
RNA Polymerases as Molecular Motors
Edited by Henri Buc and Terence Strick

Published by the Royal Society of Chemistry, www.rsc.org

signals encoded in the DNA and RNA that play important regulatory roles such as ensuring synchronization of transcription and translation in bacteria.[1,13,14]

7.1.1 Active-site Features of an Elongation Complex

Recent crystallographic studies of bacterial and eukaryotic RNAPs and transcription complexes have produced a largely consistent picture of the RNAP active site (Figure 7.1).[15–28] Together with extensive biochemical and RNAP-nucleic-acid crosslinking data, these crystal structures suggest that multisubunit



RNAPs from bacteria and eukaryotes evolved from an common ancestral core structure composed of two double psi-β barrels (DPBB) by accretion of both conserved and variable (*i.e.*, lineage-specific) sequences.[29] The active site of RNAP resides in the center of the structural core, buried 30–40 Å deep from the enzyme surface. Within the RNAP active site, an invariant aspartic acid triad (β′ D460, β′ D462, β′ D464 in *E. coli* numbering, which will be used throughout this chapter) coordinates two catalytic Mg^{2+} ions (Figure 7.1A). Mg^{2+} I is usually bound in the RNAP active site due to its higher affinity ($K_d = \sim 100$ μM), whereas Mg^{2+} II is only loosely coordinated by an EC ($K_d > 10$ mM) and thus may only bind in complex with substrate nucleotide triphosphate (NTP) to participate in the two-metal-mediated catalysis of nucleotide addition.[30] Alteration of any of the three Asp residues dramatically reduces all known catalytic activities of the RNAP active site and strongly decreases Mg^{2+} binding to RNAP, which is consistent with the structural analysis of the active site.[31]

7.1.2 Nucleotide Addition Cycle

First it is important to define the concept of register of the RNAP, which can be thought of as the position of the protein's catalytic site relative to the 3′ end of the RNA:DNA hybrid (Figure 7.2A). We begin in the pre-translocated register: here a new nucleotide has been added to the nascent RNA and still fills the

Figure 7.1 RNA polymerase active site and mechanism of catalysis. (A) A structural model of multi-subunit RNA polymerase active site and adjacent structures in the catalytically competent insertion state. Template DNA (black), nontemplate DNA (light orange), RNA (red), bridge helix (cyan), trigger loop (orange); βDloopII (salmon), fork loop 2 (blue), Asp triad (magenta), Mg^{2+} ions (yellow), incoming NTP substrate (green), basic rim (light blue) and E site (gray). Residues in the trigger loop that directly contact the incoming NTP, β′ Met 932, β′ Arg 933 and β′ His 936, are indicated and the interactions are shown as dashed lines. β′ Q504 of the basic rim in *Escherichia coli* corresponds to and is shown as β′ Arg 783 in *Thermus thermophilus*. The structural model is built based on Vassylyev *et al.*, 2007 PDB ID 2O5J,[15] and Wang *et al.*, 2006 PDB ID 2E2H.[17] All molecular graphics were prepared using Pymol (DeLano Scientific, Palo Alto, CA). (B) Mechanism for two-Mg^{2+}-mediated catalysis of nucleotide addition. Two subsites of the active site, P site and A site, are shown. NTP is shown in green. The red arrows illustrate the S_N2-type nucleophilic attack of the RNA 3′ OH on the NTP α-phosphorus atom and the reaction involves stabilization of a trigonal-bipyramidyl transition state by two Mg^{2+} ions.[30,33] (C) Comparison of the pre-insertion state (denoted PS) and the insertion state (A, also referred to sometimes as IS) during nucleotide addition. In the pre-insertion state, the trigger loop (blue) is partially unfolded away from the active site and the NTP is also shown in blue. In the insertion state, folding of the trigger loop (orange) shifts the NTP phosphates (green) towards the active site and facilitates catalysis. Note the differences in positioning of Mg^{2+} II in the pre-insertion state (blue sphere) and insertion state (green sphere).

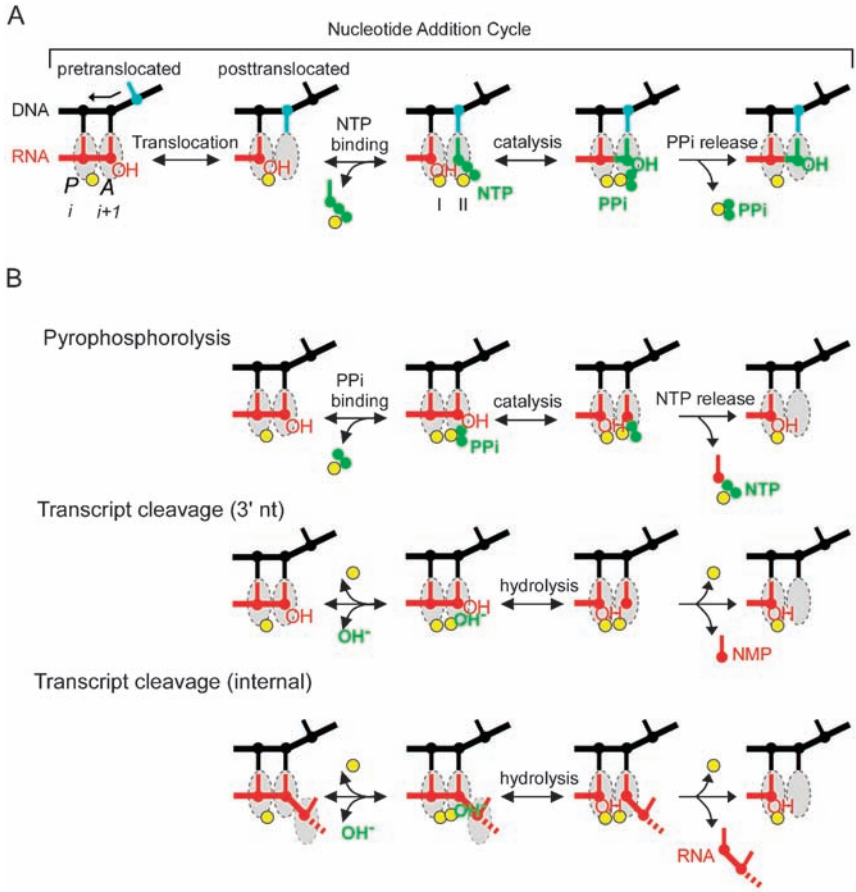


Figure 7.2 Nucleotide addition cycle and alternative reactions catalyzed by RNAP. (A) Steps of nucleotide addition cycle include translocation, NTP-binding, catalysis and pyrophosphate (PPi) release. Template DNA (black), RNA (red) Mg^{2+} (yellow sphere) and NTP (green). (B) Alternative reactions catalyzed by RNAP include pyrophosphorolysis (top panel) and transcript cleavage at the 3′ nt (middle panel) or internal transcript cleavage (bottom panel). Pyrophosphorolysis occurs in the presence of high concentrations of PPi and low concentrations of NTP. RNAP enters pretranslocated register, binds PPi at the A site, then the non-bridging oxygen of PPi initiates nucleophilic attack on the scissile phosphorus atom, forming NTP, which then exits from the active site. Transcript cleavage of the 3′ nt or internal transcript cleavage occurs by nucleophilic attack by the OH^- group on the scissile phosphate, hydrolysis of the phosphodiester bond and removal of NMP or polynucleotide RNA.

nucleotide insertion site (A site, also referred to as IS, or $i+1$). To add a new nucleotide to the RNA, RNAP must first move the RNA 3′ nt from the insertion site to the product site (P site, also referred to as i site) to make room for the incoming nucleotide, a process termed translocation. Then in the

post-translocated register the incoming nucleotide triphosphate (NTP) can enter the A site and align with the RNA 3′ OH. Subsequently, catalysis of nucleotide addition occurs by S_N2-type nucleophilic attack of the RNA 3′ OH on the NTP α-phosphorus atom and involves stabilization of a trigonal-bipyramidyl transition state by two Mg^{2+} ions.[11,32] This two-Mg^{2+}-mediated catalysis appears to be universal for all polynucleotide polymerases (Figure 7.1B).[33] The catalysis step covalently attaches one NMP to the RNA 3′ end *via* a new phosphodiester bond, extending the RNA by one nucleotide and generating pyrophosphate (PPi) and a new RNA 3′ end (PPi). Release of pyrophosphate from the active site completes the cycle and sets the stage for the next round of nucleotide addition (Figure 7.2A). A mechanistic understanding of each step in this cycle is essential to produce a complete molecular picture of the nucleotide addition process and to shed light on the complex regulatory mechanisms by which intrinsic signals in RNA and DNA and extrinsic factors regulators modify and reprogram the transcribing complex.

7.1.3 Pyrophosphorolysis and Transcript Cleavage

The RNAP active site can catalyze two types of reverse reactions that shorten the RNA chain: pyrophosphorolysis and transcript cleavage.

7.1.3.1 Pyrophosphorolysis

When RNAP is in the pretranslocated register the nucleotide addition reaction can be reversed in the presence of pyrophosphate (PPi), removing the 3′ nt in the form of NTP (Figure 7.2B, top panel). This chemical reversal of nucleotide addition is termed pyrophosphorolysis and can progressively shorten the RNA transcript at rates greater than $\sim 1\ min^{-1}$ in the presence of high concentrations of PPi and low concentrations of NTP.[34] Thus pyrophosphorolysis is one mechanism for removal of misincorporated nucleotides to ensure transcription fidelity.[12] However, pyrophosphorolysis cannot be responsible for most error correction in RNA *in vivo* for the following reasons. In the *E. coli* cell, PPi levels remain almost constant at $\sim 1\ \mu M$ due to the action of pyrophosphatase. However, the apparent K_d for pyrophosphorolysis is ~ 1–$3\ mM$. In contrast, NTPs levels are at ~ 1–$3\ mM$ in cells whereas their apparent K_d for RNAP active site are ~ 50–$200\ \mu M$.[12,34,35] Thus, the equilibrium constant for nucleotide addition *versus* pyrophosphorolysis is estimated to favor nucleotide addition over pyrophosphorolysis by more than 100-fold at cellular NTP and PPi concentrations.[12,36] Therefore, the pyrophosphorolysis reaction presumably does not occur to any appreciable extent in most cellular conditions.[36] Also as noted by Greive *et al.*, using pyrophosphorolysis to correct errors in RNA does not meet the ''Hopfield Criterion,'' which posits that error corrections are only effective when using a different mechanism than the synthesis mechanism.[37] As the chemical reversal of nucleotide addition, pyrophosphorolysis requires the

same active-site geometry as nucleotide addition, and thus also suffers from inefficient catalysis due to misincorporated, misaligned RNA 3′ nt. A different mechanism, namely transcript cleavage, is likely utilized by the cell to remove misincorporated nucleotides.

7.1.3.2 Transcript Cleavage

The transcript cleavage reaction hydrolytically removes one or more nucleotides from the RNA 3′ end, rescues arrested complex due to misincorporation, and contributes to transcription fidelity.[38,39] Transcript cleavage is mechanistically distinct from pyrophosphorolysis because in transcript cleavage the nucleophile that initiates the nucleophilic attack on the α-phosphorus is the OH^- group instead of pyrophosphate. Nevertheless, all three types of reactions, nucleotide addition, pyrophosphorolysis and transcript cleavage, share the requirement of two catalytic Mg^{2+} ions to stabilize a trigonal-bipyramidyl transition state during catalysis (Figure 7.1B).

Transcript cleavage is central to the rescue of arrested transcription complexes. During transcript elongation, incorporation of incorrect nucleotides (termed "misincorporation") into the RNA 3′ nt occurs at probabilities ranging from 10^{-3} to 10^{-5}, leading to formation of arrested transcription complexes.[12,40] In these arrested complexes, the incorrect RNA 3′ nt is unable to base pair with the template DNA base. Lack of base pairing at the 3′ end of the RNA:DNA hybrid can in turn cause the RNAP to slide back (upstream) relative to the DNA-RNA scaffold, placing the RNA 3′ proximal region containing the incorrect nucleotide(s) in the secondary channel, a process termed backtracking.[41] Backtracked transcription complexes are unable to incorporate nucleotides but can be rescued by hydrolytic removal of backtracked RNA *via* nucleophilic attack of OH^- on the scissile phosphate, placing a newly generated 3′ nt in the P site, ready to resume nucleotide addition (Figure 7.2B).[42] Such a reaction also requires Mg^{2+} II, which binds RNAP only weakly in the absence of NTP. Therefore, higher concentrations of Mg^{2+} or elevated pH (providing more OH^- as nucleophiles) accelerate transcript cleavage.[30,38,42]

In the presence of transcription cleavage factors such as GreA and GreB in bacteria and TFIIS in eukaryotes, the transcript cleavage reaction can be accelerated by more than 3000-fold.[24,39, 43–45] These cleavage factors bind RNAP and insert a reactive domain into the RNAP secondary channel, delivering several acidic residues to close proximity of the active site.[39,46,47] These factors can then induce transcript cleavage presumably by stabilizing Mg^{2+} II binding.[30,39] In addition, the identity of the RNA 3′ nt affects its ability to stabilize Mg^{2+} II and to promote transcript cleavage. A nucleotide misincorporated at the 3′ end of the RNA appears to stimulate transcript cleavage at the penultimate position, thus providing an intrinsic proof-reading mechanism (Chapter 8).[48] In summary, transcript cleavage removes misincorporated nucleotides and converts arrested complexes back into active ECs.

7.1.4 Regulation of Transcript Elongation by Pauses

Occasionally the nucleotide addition cycle can be interrupted by pause signals encoded in the DNA and RNA that play important regulatory roles.[1,13,14] These pause signals embedded in the DNA and RNA sequence, coupled with the effect of extrinsic regulators, can transiently interrupt nucleotide addition in a fraction of elongating RNAPs and cause transcriptional pauses.[13,14,49,50] Transcriptional pauses ensure coupling of transcription and translation in bacteria, allow time for proper folding of RNA secondary and tertiary structures, facilitate properly-timed loading of elongation regulators, and precede intrinsic and regulator (*e.g.*, Rho) dependent termination.[51] Although no general consensus sequence for pausing has emerged, transcriptional pauses are strongly sequence-dependent.[14,50,52,53]

Different types of pauses appear to result from a common initially paused state sometimes called the elemental pause and are categorized based on different ways to extend the elemental pause into stabilized pauses: hairpin-stabilized pauses (*e.g.*, the *his* pause from the histidine biosynthesis operon leader region), backtracking pauses (*e.g.*, the *ops* pause), and regulator-dependent pauses (*e.g.*, the promoter-proximal σ pause).[14,50] Structurally, the elemental pause results from a rearrangement of the protein and nucleic acid components in the active site, leading to inhibition of nucleotide addition. In the elemental pause the system is likely in the pretranslocated register.[13]

At the best understood hairpin-stabilized *his* pause site, interactions between RNAP and nucleic acid sequences (in the downstream DNA, in the active site, in the RNA:DNA hybrid, in the pause RNA hairpin, and in the 2–3 nt spacer between the hybrid and the hairpin) collectively contribute to the active site rearrangement in the elemental pause. These components plus the pause RNA hairpin subsequently stabilize the paused state, resulting in a long-lived pause.

Kinetically, pauses represent local free energy minima that precede relatively high free energy barriers to nucleotide addition. Kinetic models that seek to describe transcript elongation quantitatively and to predict pause sites based on sequence context are discussed in Chapter 9.

7.2 Structural Basis of NTP Loading and Nucleotide Addition

A single RNAP active site, deeply buried more than 30 Å from the RNAP surface, is responsible for all known activities of RNAP, including nucleotide addition, pyrophosphorolysis and transcript cleavage.[25,31] Key conformational changes in RNAP, involving protein secondary structure motifs, DNA, and RNA in the active site, mediate nucleotide loading and addition (Figure 7.1). Interference of these key movements by regulatory signals, extrinsic factors, and small molecules such as streptolydigin can dramatically impact or even completely shut off catalysis of nucleotide addition.[13,15,19,20] Therefore, it is crucial to understand the structural basis of nucleotide loading and addition

before one can understand regulation of transcript elongation. However, detection of such small movements on the order of 1–10 Å buried 30–40 Å deep from RNAP surface at 10^{-2}–10^{-3} s temporal resolution is no easy task. Nonetheless, recent high-resolution crystallographic and extensive biochemical studies have yielded considerable insight and have begun to allow visualization of the nucleotide addition process in atomic detail.

An important question raised by these crystal structures is how substrate NTP initially gains access to the active site. The only evident route that connects the active site to the RNAP surface based on available RNAP crystal structures is a funnel-shaped narrow pore termed the secondary channel.[27,28,54] Recent crystallographic, biochemical and simulation studies have identified an NTP "entry" site (E site) located at the inner pore of the secondary channel, lending support to the secondary-channel model of NTP entry.[22,54,55] However, based on biochemical studies of human Pol II, Burton and co-workers proposed an alternative NTP-entry model in which NTPs initially base-pair to downstream DNA template and load through the main channel.[23,56,57] These two models are further discussed in Chapter 8.

Within a ~15 Å radius of the active site, the following conserved RNAP structural features have been identified as important for RNAP catalytic activities and regulation: a long α-helical segment termed the bridge helix (β′780-815) that spans the main channel of RNAP and a flexible trigger loop (β′928-942 and β′1131-1145) just downstream of the active site. The bridge helix and the trigger loop have been observed to move toward and away from the active site or to fold and unfold near the active site, respectively, in structures containing or lacking nucleic acids. Additional structures identified as important for RNAP function and regulation include a basic rim (βR678, βR1106, β′Q504, β′R731, which contact substrate phosphates), a fork loop 2 (β533–541, termed βDloopI or streptolydigin-binding loop, proposed to bind NTP and control nucleotide addition) and a βDloopII (β558-575, which contacts the RNA:DNA hybrid) (Figure 7.1A).

These structures have led to several models for the crucial movements in the RNAP active site, which we will describe in the following sections.

7.2.1 Bridge-helix-centric Models of Nucleotide Addition and Translocation

Several structural models of nucleotide addition have postulated that a putative bridge helix oscillation between a uniform α-helical conformation (cyan, Figure 7.3A) and a locally unfolded conformation (magenta, Figure 7.3A) sterically drives translocation of DNA through RNAP and thereby controls the nucleotide addition cycle.[25,27,58,59] In these models, movements of the trigger loop located adjacent to the bridge helix are proposed to facilitate bridge helix oscillation and thus indirectly affect nucleotide addition *via* effects on bridge helix conformation.[58,59] These bridge-helix-centric models are based on the crystallographic observation of different conformations of the bridge helix in

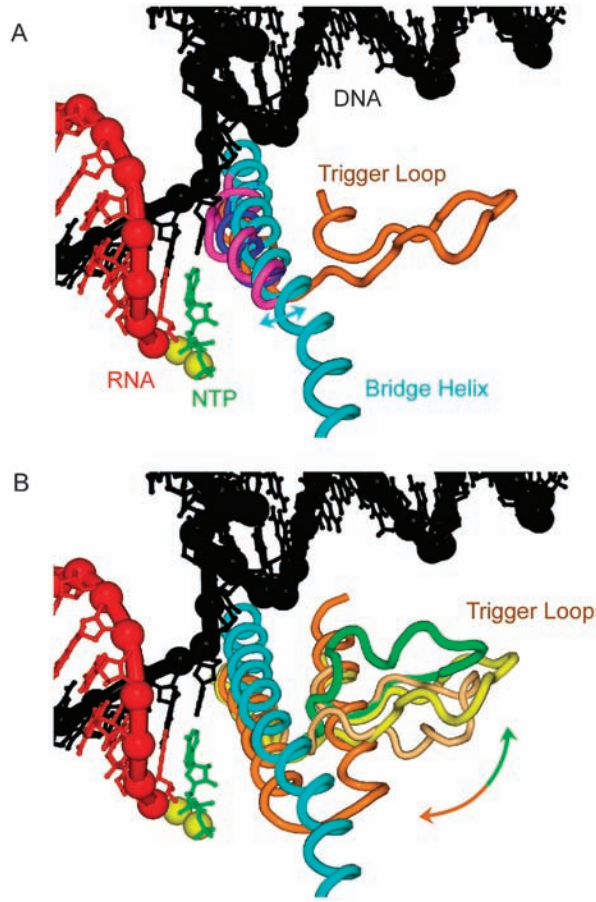


Figure 7.3 Proposed movements of the bridge helix and of the trigger loop during nucleotide addition. (A) Bridge-helix movements. The bridge helix is proposed to oscillate between an α-helical conformation (“straight,” in cyan, modeled as in Gnatt *et al.*,[27] PDB ID 1I6H) and an unfolded conformation (“bent” or “flipped out,” in magenta, modeled as in Vassylyev *et al.*,[25] PDB ID 1IW7) driving DNA translocation through RNAP. Alternatively, the bridge helix can assume a shifted conformation (in blue, modeled as in Wang *et al.*,[17] PDB ID 2E2H) and play roles in stabilizing trigger loop folding. Note that the unfolded bridge helix conformation (magenta) is sterically incompatible with the template DNA base in the *i* + *1* site (black). The frequency of the bridge-helix oscillation (cyan arrow) therefore could determine the rate of nucleotide addition and control transcript elongation.[58] (B) Trigger-loop movements. The trigger loop is proposed to oscillate between a folded α-helical hairpin conformation [orange, modeled as in Vassylyev *et al.*[15] (PDB ID 2PPB) and Wang *et al.*[17] (PDB ID 2E2H)] and multiple unfolded loop conformations [green conformation is modeled as in Vassylyev *et al.*[25] (PDB ID 1IW7); yellow conformation is modeled as in Tuske *et al.*[19] (PDB ID 1ZYR); light orange conformation is modeled as in Kettenberger *et al.*[22] (PDB ID 1Y1V)]. Trigger loop folding brings its residues β′ 933–936 into direct contact with the substrate NTP (Figure 7.1A), closing the active site, and aligning the RNA 3′ OH, NTP phosphates, and two Mg^{2+}s to facilitate catalysis.[15] Therefore, folding or movement of the trigger loop (orange–green arrow) controls nucleotide addition.

bacterial RNAP core/holoenzymes and in yeast Pol II. In the yeast Pol II structures, the bridge helix is uniformly alpha-helical[17,27,55,60,61] (Figure 7.3A, cyan), whereas in *Thermus aquaticus* core enzyme[62] and *Thermus thermophilus* holoenzyme[25] the bridge helix is locally unfolded or flipped out in the middle (Figure 7.3A, magenta). In this unfolded conformation the domain would sterically clash with the $i+1$ template DNA base based on these models and structures of the EC.[25]

These structural observations led to proposals that the bridge helix unfolding pushes the $i+1$ template DNA to the i position, thus sterically driving DNA translocation through RNAP. Epshtein *et al*. reported that the RNA 3′ nt (substituted with crosslinkable nucleotide analog 2′-deoxy-3′-isothiocyanate) crosslinks to β′ R789 of the bridge helix and β′ M932 of the trigger loop and postulated that an unfolded (''bent'') bridge helix would place β′ R789 in the A site, evicting the pretranslocated RNA 3′ nt from the A site and preventing substrate NTP binding in the A site.[59] Therefore, it was suggested that the bridge helix oscillation involving the residues β′ R789, β′ T790, β′ A791 supported by a coordinated trigger loop movement involving β′ M932 not only could drive DNA/RNA translocation but also could define a ''swinging gate'' mechanism that controlled NTP entry.[59]

Further work by Bar-Nahum *et al*. isolated two dominant lethal substitutions in the trigger loop, G1136S and I1134V, which increase or decrease RNAP elongation rate, respectively. Exonuclease III footprinting with these mutants suggest that the fast G1136S RNAP is more likely to be in the post-translocated state than wild-type RNAP, whereas the slow I1134V RNAP is more likely to be in the pretranslocated state than wild type.[58] Therefore, these substitutions in the trigger loop were suggested to affect the folding/unfolding dynamics of the bridge helix and thus indirectly alter the translocation register and elongation rate of RNAP by affecting bridge-helix-controlled translocation. The RNAP inhibitor streptolydigin has been proposed to inhibit nucleotide addition by binding to and interfering with this conformational cycling of the bridge helix.[19]

The evidence supporting the bridge-helix-centric models faces certain objections. First, in recent yeast EC structures trapped in the post-translocated register, the observed displacement of the bridge helix towards the RNA:DNA hybrid (2.0 to 2.7 Å) takes place without unfolding.[17,55] Bacterial EC structures complexed with AMPcPP or AMPcPP and streptolydigin did not reveal the unfolding (bending) of the bridge helix either.[15] To date, the unfolded bridge helix has been only observed in bacterial core or holoenzymes with no nucleic acid or substrate NTPs and has never been observed in yeast Pol II structures trapped in a number of states.[17,22,24,27,60] These structural findings cast doubt on the mechanistic relevance of an unfolded bridge helix.

Second, biochemical evidence supporting the bridge helix oscillation relies on the usage of artificial crosslinking groups such as 2′-deoxy-3′-isothiocyanate nucleotide analog, which lacks both 2′ OH and 3′ OH and is such a poor substrate for RNAP that Mn^{2+} rather than Mg^{2+} has to be used to boost incorporation efficiency.[58,59] Usage of Mn^{2+} is known to increase

misincorporation and thus may cause distortions of the active-site structures and lead to artifactual crosslinking results.[63–65]

Third, it remains unclear how substitutions in the trigger loop isolated by Bar-Nahum *et al.*, which are located in the bridge-helix-distal side of the trigger loop, can dramatically alter the bridge-helix conformation. Indeed, to examine the possible level of involvement of bridge helix unfolding in nucleotide addition, Toulokhonov *et al.*[13] characterized a mutant RNAP carrying the A791G substitution in the center of the unfolded segment of the bridge helix. Predictions based on folding studies with isolated peptides suggested that such an alteration from helix stabilizer Ala to helix destabilizer Gly should destabilize the alpha-helical conformation by ~20-fold.[66] However, the A791G mutation only slows nucleotide addition at a defined pause site by ~30%[13] and supports *E. coli* growth.[19] Furthermore, a more drastic disruption of the bridge helix obtained by replacing the "oscillating" segment of five amino acids with a 3-glycine linker [$\Delta\beta'$K789-S793Ω(Gly)$_3$] gave a ~90-fold reduction of the nucleotide addition rate at a pause site whereas deleting the trigger loop had a ~460-fold effect.[13] Disrupting the bridge helix in an RNAP lacking a trigger-loop only reduced the rate of NTP incorporation four-fold. These results are at odds with the proposed pivotal role of bridge helix oscillations in nucleotide addition and pausing.[13,19,25,27,58]

7.2.2 Central Role of the Trigger Loop in Nucleotide Addition and Pausing

In contrast to the bridge-helix-centric structural models, recent biochemical and structural studies of bacterial and yeast ECs revealed that the trigger loop, located close to the bridge helix at the inner pore of the secondary channel, plays the key role in NTP loading, nucleotide addition and regulation of elongation (Figures 7.1A and 7.3B).[13,15,17] Movement of the trigger loop is proposed to be associated with the rate-limiting active-site rearrangement that aligns the substrate NTP and thus facilitates catalysis. It also may orchestrate the creation of paused transcription complexes when its movement is restricted.[13,15–17]

In earlier RNAP crystal structures, the conserved trigger loop is disordered, which is indicative of high degree of mobility. Among the structures in which the trigger loop is modeled, different conformations have been observed depending on the particular enzyme, conditions of crystallization, and presence or absence of transcription factors and small molecules such as streptolydigin (Figure 7.3B).[15–17,19,20,22] Recent crystal structures of bacterial EC complexed with AMPcPP and of yeast Pol II EC with 3′deoxy RNA complexed with GTP both reveal a novel trigger loop conformation, a folded α-helical hairpin (Figures 7.1A and 7.3B) that is reminiscent of the O, O′ helices in DNA polymerases.[15,17] The folded trigger loop makes extensive interactions with the nearby bridge helix: together they form a three-helical bundle. Trigger loop folding brings its conserved residues β' Met932-His936 into direct contact with

the incoming NTP (contacts occur both at the base and at the phosphates) and properly aligns the RNA 3′ nt OH, NTP triphosphate, and two catalytic Mg^{2+} ions for catalysis (Figures 7.1A and 7.3B).[13,15,17]

RNA–protein crosslinking experiments provided biochemical evidence that trigger loop folding occurs at least in some ECs. In non-backtracked, active transcription complexes halted at a nonpause site, the crosslinkable nucleotide analogue 4thioUMP incorporated at the RNA 3′ nt crosslinks to the A site.[67] This crosslink to the A site was shifted to the trigger loop when the halted EC gradually went into arrest, indicative of an active-site reconfiguration. More recent work showed that, at the non-backtracked *his* pause site, 4thioUMP incorporated at the RNA 3′ nt in the A site primarily crosslinks to the β′ Arg933-His936 segment of the trigger loop.[13] These results provide strong support for the existence of a folded trigger loop conformation and its close proximity to the RNA 3′ nt. The crosslinkable NTP analogue 4thioUMP used in these studies is superior to 2′-deoxy-3′-isothiocyanate used by other studies[58,59] in that it is a much better substrate for RNAP, does not require Mn^{2+} for efficient incorporation, and only moderately affects RNAP catalytic activity.[13]

Various alterations of the trigger loop establish its central role in nucleotide addition. Single amino-acid substitutions on the trigger loop profoundly alter elongation rate, transcriptional pausing, transcript cleavage, pyrophosphorolysis and termination.[13,15,58,59,68] Substitutions within the trigger loop that do not involve the NTP-interacting residues but that should specifically destabilize the folded conformation drastically reduce rates of nucleotide addition by as much as ~10 000-fold with minimal effects on apparent NTP affinity.[13,15] Deletion of the trigger loop either in *E. coli* or in *T. aquaticus* also reduces nucleotide addition by a factor of ~10 000 with minimal effects on apparent NTP affinity.[13,20] These biochemical results suggest that trigger loop folding is required for catalysis but not for NTP binding in the active site.

The trigger loop also plays important roles in transcriptional pausing. Various alterations of trigger loop residues greatly alter pausing. Importantly, deletion of the trigger loop from *E. coli* RNAP significantly compromised its ability to recognize a pause signal, and largely abolished the ~100-fold difference in nucleotide addition rates at pause sites and non-pause sites.[13] These observations are consistent with the view that restriction of the trigger loop's movement contributes to formation of the paused state. Thus, the trigger loop not only serves as a facilitator of nucleotide addition, but also serves as a controller of the active site in response to regulatory signals such as pause signals embedded in the DNA and RNA.

7.2.3 A Trigger-loop Centric Mechanism for Substrate Loading and Catalysis

The picture that emerges from the aforementioned structural and biochemical studies supports the following trigger-loop model of nucleotide addition. (Figures 7.3B and 7.4).[13,15,17] First, while the trigger loop is unfolded, NTP

initially lands at an entry site (E site[55]) located at the junction of the main channel with the secondary channel. This initial binding can even occur in the absence of the trigger loop. After initial NTP binding, the trigger loop partially folds and NTP enters the A site. However, this first conformation of NTP in the A site is not aligned for catalysis and is termed a "pre-insertion" state. Here the basic rim plays an important role: composed of multiple arginine residues that directly contact the NTP's phosphate groups (specifically, residues βR678, βR1106, β′Q504, β′R731) it is thought to interact with and prevent the phosphate groups from aligning with the RNA 3′ OH and Mg^{2+} cations and undergoing premature catalysis. Thus the basic rim could play a role in proofreading the incoming NTP.

After binding of a correct NTP in the A site, the trigger loop completely folds into an α-helical hairpin, closing the active site, and aligning the incoming NTP triphosphates with the RNA 3′ OH and Mg^{2+}s to facilitate catalysis (Figure 7.1A). This key conformational change of the trigger loop appears to be associated with the rate-limiting step in nucleotide addition. S_N2-type nucleophilic attack of the RNA 3′ OH on the NTP α-phosphorus atom then ensues and the two properly-aligned Mg^{2+} ions stabilize a trigonal-bipyramidyl transition state (Figure 7.1B). Lastly, PPi exits from the active site, the trigger loop reverts to its unfolded conformation, and the enzyme translocates DNA by one bp to await the next round of nucleotide addition. In this model, the bridge helix plays a secondary role by facilitating trigger loop folding *via* formation of a three-helical bundle.[13,15]

In conclusion, extensive structural and biochemical studies have revealed that trigger loop folding or movement, rather than the unfolding of the bridge helix, controls nucleotide loading and addition. A trigger-loop-centric model incorporating a two-step substrate loading mechanism appears to explain most experimental observations and is consistent with conservation of the fundamental mechanism of nucleotide addition throughout bacterial and eukaryotic RNAPs.[13,15]

The trigger loop has also been proposed to participate in ensuring transcription fidelity and preventing misincorporation.[17] It remains unclear what exact role the trigger loop plays in transcription fidelity, as RNAP with no trigger loop can still distinguish among different NTPs.[13] Quantitative analysis of transcription fidelity in mutant RNAPs with various trigger-loop alterations are necessary to examine the trigger loop's involvement in fidelity control. Specific mechanisms of fidelity control are discussed in more detail Chapter 8.

7.3 Models of Translocation: Power-stroke *versus* Brownian Ratchet

RNAP is a powerful and processive mechanoenzyme that converts chemical energy stored in NTP into the mechanical work of threading DNA through RNAP. RNAP can generate forces as high as 30 pN[69] and can synthesize RNAs

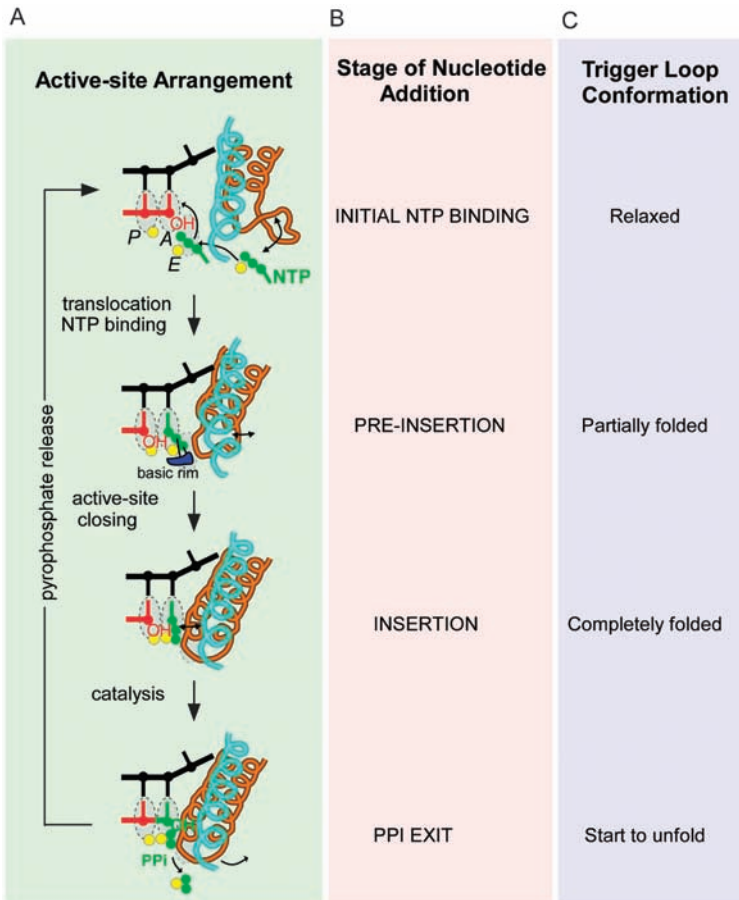


Figure 7.4 A trigger-loop-centric model of nucleotide loading and addition. The left-hand column depicts cartoon views of active-site arrangement at each stage of nucleotide addition cycle (middle column) and corresponding conformation of the trigger loop (right-hand column). Concomitant with or following translocation, incoming NTP initially binds at the E site when the trigger loop is in a relaxed loop conformation. NTP then moves into the A site, possibly inducing partial folding of the trigger loop, together constituting an open "pre-insertion" state, where NTP base-pairing with DNA template can be checked whereas catalysis is inhibited. The basic rim (blue) interacts extensively with the NTP phosphates at this stage (black lines) and likely prevents formation of catalytic competent insertion state. Incorrect NTPs can then be checked and rejected from the active site before a correct NTP induces complete folding of the trigger loop, closing the active site and aligning the RNA 3′ OH, NTP phosphates, and Mg^{2+}s, facilitating catalysis of nucleotide addition at the insertion site. Following catalysis, PPi release from the active site removes the bridging interactions between the active site and the trigger loop. The trigger loop thus likely spontaneously unfolds in the absence of PPi and the active site is now ready to undergo another round of nucleotide addition.

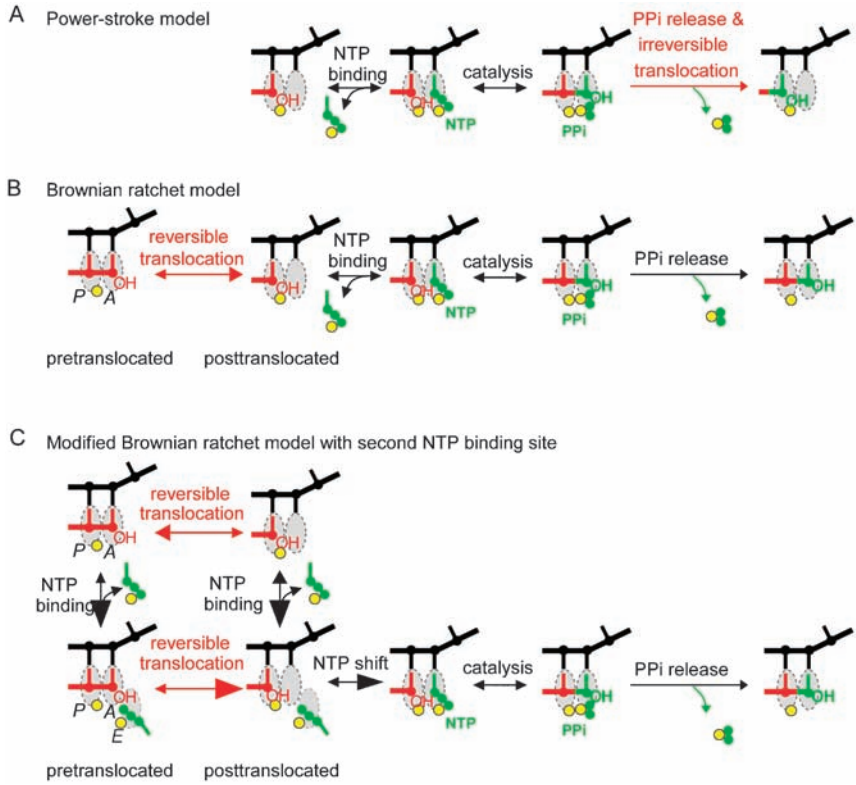


Figure 7.5 Power-stroke and Brownian ratchet models of RNAP translocation. Comparison of different kinetic models for RNAP translocation. (A) A power-stroke model where translocation is tightly coupled to irreversible PPi release. (B) A Brownian ratchet model where reversible interconversion between the pre- and post-translocated states occurs faster than nucleotide addition. NTP binding favors the post-translocated state and bias the EC towards forward translocation. (C) A modified Brownian ratchet model where translocation and NTP binding can occur in either order. This model requires the existence of a secondary NTP binding site to accommodate the possibility of nucleotide binding when the EC is in the pre-translocated state. Differences in the size of arrow heads illustrate the directionality of change in equilibrium.

as long as 2000 kb.[70] Each nucleotide addition cycle involves a translocation step in which the DNA moves relative to RNAP by $\sim$3.4 Å, the distance of a single base pair. In the active site, the RNA 3′ nt moves from the A site to the P site, one base pair is unwound in the downstream DNA duplex and another base pair is rewound in the upstream DNA. As already mentioned in Chapter 4, two types of models offer mechanistic explanations for the translocation process: "power-stroke" models and "Brownian ratchet" models (Figure 7.5).[34,58,71–75]

7.3.1 Key Distinctions between Power-stroke and Brownian Ratchet Models

Power-stroke models posit that translocation occurs by directly coupling the chemical energy of triphosphate hydrolysis to irreversible mechanical translocation (Figure 7.5A). In contrast, Brownian ratchet models postulate that the pretranslocated and post-translocated states interconvert more rapidly than nucleotide is added and that NTP binding in the A site biases RNAP towards forward translocation (Figures 7.5B and 7.5C). The key distinction between these two types of models lies in whether the pretranslocated and post-translocated states interconvert fast enough with respect to nucleotide incorporation. The rate of interconversion is determined by the height of the energy barrier between these two states, or translocation registers. This energy barrier has not been directly measured, although there is evidence that PPi biases the translocation register from post-translocated state to pretranslocated state and that NTP has the opposite effect.[34,58] Thus, definitive evidence that would unequivocally distinguish these models does not yet exist. This being said, biochemical, single-molecule and modeling efforts have been undertaken to address the question and have yielded considerable insights into the mechanism of NTP-driven translocation.

7.3.2 Power-stroke Models

A power-stroke model for RNAP translocation was proposed for T7 RNAP based on crystallographic snapshots of transcribing T7 RNAP ECs complexed with either incoming NTP or PPi.[71] The observation of a closed form of the O helix and fingers domain in the presence of PPi and of a more open form in the presence of NTP led to the proposal that irreversible PPi release destabilizes the closed conformation causing opening of the O helix, which then drives translocation by physically pushing the RNA:DNA hybrid from the A site to the P site (Chapter 4). The essence of this model is that irreversible PPi release is tightly coupled with translocation *via* the defined motion of the O-helix. This would ensure that the power-stroke operates unidirectionally, converting chemical energy into mechanical work of translocation (Figure 7.5A). Similar power-stroke models have been proposed for other molecular mechanical systems such as myosin-actin contraction and mitochondria F1-ATP synthase rotation. In such systems, the release of phosphates promotes certain well-defined conformational changes that are coupled to mechanical movement.[76,77]

7.3.3 Brownian Ratchet Models

Although the power-stroke model agrees well with the T7 RNAP EC structures, it has received little support from solution biochemistry or single-molecule studies. Instead, models analogous to Brownian ratchets have received general support from experimental and theoretical studies of both T7 and bacterial RNAPs.[34,58,72–75] Sousa and others lent biochemical support for

Brownian ratchet models by showing that (1) halted ECs can slide on DNA, indicating that translocation is reversible; (2) roadblocks preventing RNAP movement reduce the apparent NTP affinity for RNAP, presumably by decreasing the population of post-translocated ECs (recall that it is in the post-translocated register that the unoccupied A site can bind NTPs); and (3) rapidly elongating and halted ECs are equally sensitive to PPi, suggesting that the occupancy of the pre- and post-translocated states are not significantly different in the two cases and, therefore, that interconversion can be uncoupled from nucleotide addition.[34,72,73]

Furthermore, by using single-molecule force-clamp transcription assays where the resolution reaches the level of a single base pair, Abbondanzieri *et al.* characterized the force–velocity behavior of *E. coli* RNAP. They reported results in favor of Brownian ratchet models over power-stroke models in which PPi release is tightly coupled with translocation.[78] Specifically, statistical analysis of multiple single-molecule transcription traces at various assisting or hindering forces suggested that higher hindering forces are required to slow elongation at increasing concentrations of NTP. This suggests that NTP binding is coupled to translocation, which is consistent with Brownian ratchet models. Power-stroke models, in which translocation is driven by PPi release but not by NTP binding, do not predict significant effects of NTP on the force dependence of translocation. In addition, the observed insensitivity of elongation rate to [PPi] also argues against a power-stroke model in which translocation is coupled to PPi release.[69] In fact, the measured force–velocity relationship is well-described by a modified Brownian ratchet model in which either translocation or NTP binding can occur first prior to NTP alignment in the A site and catalysis (Figure 7.5C). This modified model requires the existence of a secondary NTP binding site that can accommodate NTP binding in pretranslocated ECs, which fits well with the observation of initial NTP binding in the E site observed in yeast Pol II.[54,55] An allosteric site proposed by Erie and co-workers (Section 7.4.1) would also be consistent with this observation, although there is currently no structural evidence for an allosteric NTP-binding site (for further discussion of all these points see Chapter 9). Similar single-molecule, force-clamp studies conducted with T7 RNAP also supported the general Brownian ratchet model.[79] Altogether these results suggest that the fundamental Brownian ratchet mechanisms of RNAP translocation may be conserved between single-subunit T7 RNAP and multisubunit RNAPs.

7.3.4 Technical Outlook in Detecting the Precise Translocation Register

A mechanistic understanding of the translocation process requires the ability to precisely and non-invasively measure the translocation register of RNAP. A major concern in interpreting experiments that try to probe the translocation register is that most techniques can perturb the translocation register one is trying to measure. For example, exonuclease III footprinting of ECs can be

biased by exonuclease III interacting with RNAP and altering its original register. $KMnO_4$ and other chemical probing methods can suffer from insufficient spatial or temporal resolution since direct observation of single-base-pair translocation during rapid elongation would require $\sim 10^{-10}$ m spatial resolution and $\sim 10^{-3}$ s temporal resolution or better and also may chemically damage RNAP and alter translocation register. Guo *et al*. have reported sub-Å resolution in measurement of the average position of halted T7 ECs using an RNAP-tethered chemical nuclease (Fe^{2+}-BABE) to generate detectable cleavage products of radiolabeled RNA/DNA strands.[34] This approach is rapid enough, precise (sub-Å spatial resolution) and, more importantly, does not perturb the translocation register. Future adaptation of this technique for multi-subunit RNAPs would be highly beneficial. The key technical challenge of utilizing such an assay is to guarantee that a single chemical nuclease is specifically tethered to the same amino acid (a reactive cysteine) for each and every RNAP molecule. To date, there have been no documented attempts to examine how many of the cysteines in multi-subunit RNAPs are sufficiently exposed for conjugation to Fe^{2+}-BABE.

Fluorescence-based assays offer another promising technique to measure rapid translocation in real time. Successful applications of fluorescent nucleotide analogues have enabled detection of opening or closing of a single base pair.[80] Extension of these studies could lead to direct observation of fluorescence changes during rapid translocation. Alternatively, FRET-based distance measurements in rapidly elongating RNAPs could allow for time-resolved measurements of translocation (Chapter 5).

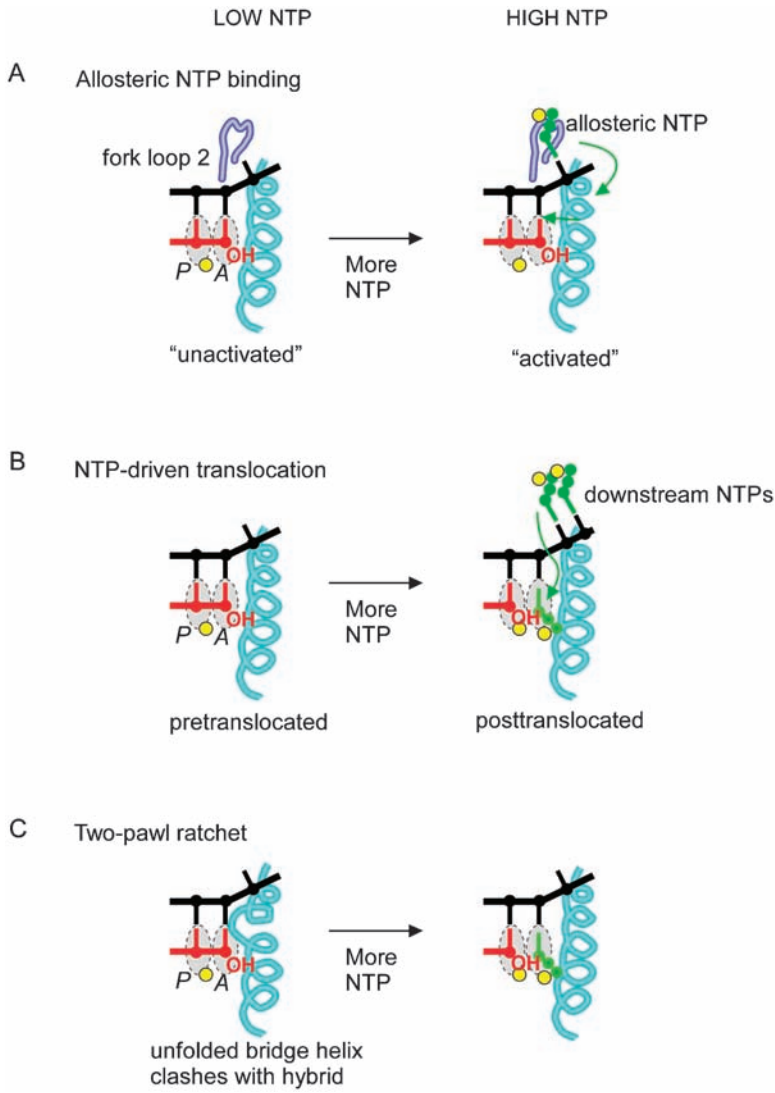
7.4 Kinetic Models of Nucleotide Addition

Landmark studies of rapid nucleotide addition by the Erie and Burton groups have demonstrated that the rate of nucleotide addition by bacterial and human RNAPs exhibited a greater-than-hyperbolic dependency on substrate NTP concentration.[23,35,56,81] Furthermore, the rate of incorporation after addition of a single-nucleotide displays a ''biphasic'' nature at intermediate NTP concentrations. This suggests the co-existence of two molecular species – a faster one and a slower one – that do not rapidly interconvert.[81] To explain these results, Erie and colleagues proposed a kinetic model for NTP addition termed the allosteric NTP binding model, whereas Burton and colleagues have proposed an alternative model termed NTP-driven translocation.

7.4.1 Allosteric NTP Binding Model

Erie and colleagues proposed that two NTPs can bind RNAP simultaneously at two distinct sites characterized by different affinities: one incoming NTP can bind the A site to be incorporated next, while a second NTP can bind an ''allosteric'' site on RNAP and switch the EC from an ''unactivated'' state

where nucleotide addition is slow to an "activated" state, where nucleotide addition is fast (Figure 7.6A). This allosteric model can explain the empirical rates measured at various NTP concentrations and this mechanism is proposed to adjust the overall rate of nucleotide addition based on the availability of NTPs.[35,81] It also has been observed that the identity of downstream template base at the $i+2$ position significantly affects nucleotide addition immediately upstream at the $i+1$ position.[81] It is suggested that a small conserved loop in the β subunit called fork loop 2 (β533–541, also termed βDloopI or streptolydigin-binding loop) is the allosteric NTP binding site. Fork loop 2 is located



only ~5–6 Å away from the $i+2$ downstream template base, close enough to directly contact an NTP bound there. The fork loop 2 sequence resembles that of well-characterized "P-loops" found in various nucleotide binding proteins, suggesting it may be a good candidate for an NTP binding site. Additionally, fork loop 2 assumes different conformations in *T. aquaticus* and *T. thermophilus* RNAPs.[25,62] Further work led to the proposal that an allosteric NTP bound at fork loop 2 can drive translocation *via* a "ratchet motion." Specifically, upon NTP binding at this allosteric site, the fork loop 2 would change its conformation and alter its direct contact with the bridge helix or the rifampicin-binding pocket residues (Figure 7.6A), both of which extensively contact the RNA:DNA hybrid and could directly affect the translocation process.[81]

This allosteric NTP binding model can explain the biphasic nature of the kinetics of nucleotide addition. It also suggests the existence of a regulatory pathway for relating the rate of transcript elongation to cellular NTP levels. Interestingly, recent yeast EC structures complexed with NTP have observed a 90° rotation of fork loop 2 in between downstream DNA duplex strands, suggesting a role in unwinding downstream DNA,[17] supporting the proposal that repositioning of fork loop 2 is coupled to translocation. However, there is yet to be a direct demonstration that NTPs can bind to the proposed allosteric site, or that the conformation of fork loop 2 directly affects translocation. It seems at least possible that the fast and slow species of ECs simply reflect

Figure 7.6 Kinetic models of nucleotide addition. Three kinetic models, (A)–(C), describe nucleotide addition at low (sub-saturating) and high (near-saturating or saturating) concentrations of NTPs, each providing mechanistic explanations for the empirical greater-than-hyperbolic NTP dependency of nucleotide addition. (A) Allosteric NTP binding model. In this model, two NTPs can bind simultaneously to two distinct sites on RNAP: one at the A site and the other at the allosteric site (fork loop 2, in blue). At low NTP, only the A site is populated by NTP, the EC is in an "unactivated" state, where nucleotide addition is slow; at high NTP, NTP binding at the allosteric site switches the EC to an "activated" state where nucleotide addition is fast, possibly by conformational changes in the fork loop 2 and consequent movements in the rifampicin-binding pocket (not shown) or the bridge helix (in cyan, green arrows).[35,81] (B) NTP-driven translocation model. In this model, at low NTPs, ECs are predominantly in the pretranslocated state. Therefore, nucleotide addition is slow. At high NTPs, increased NTP occupancy of the A site displaces the equilibrium towards the post-translocated state and thus accelerates nucleotide addition by alleviating the inhibition due to pretranslocated states at low NTPs. This model distinguishes itself from the model in (A) by postulating that the NTP bound at downstream DNA is not "allosteric" or "non-essential," but rather is *en route* to the A site for incorporation.[23,56,57] (C) Two-pawl ratchet model. In this model, bridge helix oscillations occur at every nucleotide addition cycle. At low NTPs, an unfolded bridge helix can clash with the RNA 3′ nt occupying the A site, leading to inhibition of nucleotide addition and even backtracking. At high NTPs, due to increased NTP occupancy of the A site and thus decreased RNA 3′ nt occupancy of the same site, the probability for the unfolded bridge clashing with RNA 3′ nt is reduced, leading to faster nucleotide addition.[58]

different semi-stable translocation registers of RNAP that interconvert more slowly than NTP addition but whose interconversion might be affected by NTP binding. Mutational analysis of the residues constituting the proposed allosteric site and structural demonstration of NTP binding to the site will be crucial to test this model.

7.4.2 NTP-driven Translocation Model

A similar model termed "NTP-driven translocation" was proposed by Burton and co-workers. It is based on presteady-state kinetic studies of human Pol II. As opposed to the allosteric model, this model posits that the two kinetically distinct species identified in nucleotide addition simply correspond to the pre-translocated (slow species) and post-translocated (fast species) complexes. In this model, NTPs are proposed to line up base-paired to the downstream DNA template at positions $i+2$, $i+3$, and even $i+4$, favoring forward translocation of the enzyme by facilitating delivery of the NTP to the A site (Figure 7.6B).[23,56,57] This model is analogous to the aforementioned "allosteric NTP binding" model in that an NTP can base pair to the downstream $i+2$ position and accelerate the NTP incorporation at the A site.[57] However, it suggests that the NTP bound at downstream $i+2$ position is not "allosteric" but rather a substrate NTP *en route* to the A site (Figure 7.6B; see also Chapter 8).

7.4.3 Two-pawl Ratchet Model

Bar-Nahum *et al.* have proposed an alternative model to explain the greater-than-hyperbolic NTP dependency of nucleotide addition without invoking an additional NTP binding site.[58] In this model, nucleotide addition and unidirectional translocation are mediated by two pawls: NTP as a "stationary pawl," which binds to the A site and prevents reverse translocation of the RNA 3′ end, and the oscillating bridge helix as a "reciprocating pawl," which periodically unfolds in the middle, sterically pushing the 3′ end of the hybrid from the A site to the P site to drive translocation. Importantly, unfolding of the bridge helix can fail to translocate the hybrid, but instead peel the RNA 3′ nt away from the DNA template base, destabilizing the hybrid and allowing backtracking (Figure 7.6C).[58] The frequency of this occurring is affected by NTP occupancy of the A site: at low NTP the RNA 3′ nt has a higher probability of occupying the A site. In this case nucleotide addition is thus limited by the oscillation frequency of the bridge helix (Figure 7.6C).

The same model could also explain the greater-than-hyperbolic NTP dependence of incorporation kinetics. NTP can bind the A site and diffuse back out faster than bridge helix oscillations, leaving the active site temporarily in an NTP-primed state (the post-translocated state), which can now incorporate a nucleotide more rapidly than an unprimed active site. As the NTP concentration increases, translocation is made easier and incorporation of NTP is in turn

facilitated, a mechanism leading to apparent substrate cooperativity. In addition, the state in which the unfolded bridge-helix clashes with and peels off the pre-translocated hybrid could represent a "trapped" or "paused" state, which could be stabilized by an RNA hairpin or engage in further backtracking or even termination.[58] The peeling of a DNA base off of the 3′ end of the hybrid is similar to the previously proposed fraying of RNA 3′ nt away from the template DNA at a pause.[1,13]

The "two-pawl ratchet" model is subject to several concerns. First, the model centers on the idea that bridge helix oscillations sterically control translocation and therefore nucleotide addition. For such a model to work, translocation has to be associated with the rate-limiting step of nucleotide addition. However, structural and biochemical evidence suggests that trigger loop folding, rather than translocation, is likely the key conformational change that is rate-limiting for nucleotide addition.[13,15,17] Second, the model requires that oscillations of the bridge helix occur in the context of transcribing ECs and must occur on the same timescale as nucleotide addition; these timescales remain unexplored. Finally, as noted earlier, several structural and genetic data argues against a central role of the bridge helix in nucleotide addition (Section 7.2.1).

7.4.4 Biophysical Models for Transcript Elongation

Pioneering work from von Hippel and colleagues has offered a quantitative thermodynamic approach to understanding transcript elongation. These models consider the transcribing EC at any given template position as facing three competing energy barriers to three potential reaction pathways, including elongation, backtracking ("editing") and termination.[36,82,83]

By extending the work of von Hippel, as well as that of Sousa and colleagues,[36,73,83] Bai *et al.* proposed a simplified sequence-dependent kinetic model for transcript elongation and used it to predict backtrack pauses along different DNA templates.[74] A more recent kinetic model by Tadigotla *et al.* sought to improve upon the Bai *et al.* model by incorporating sequence-specific contributions of RNA secondary structures on suppression of backtrack pauses and accounting for thermal fluctuations of the transcription bubble.[75] The respective values of these two models are discussed in Chapter 9.

These thermodynamic and kinetic modeling efforts have allowed for a more quantitative understanding of the elongation process. However, they also reveal that our current knowledge remains largely insufficient to produce an accurate picture of the mechanical behavior of elongating RNAP. These models can be further refined as we gain more mechanistic knowledge of the elongating process, for instance by taking into account energy involved in forming and breaking RNA tertiary structures or free energy associated with trigger loop/bridge helix movements. Conceivably, these models can be refined by varying the assisting or hindering load on RNAP in single-molecule transcription assays and examining how the thermodynamic model respond to external perturbations. A further step in the modeling efforts would be to

consider the chromosomal topology, density of RNAPs on DNA, cellular spatial constraints, and numerous DNA modifications RNAP encounter in living cells.

7.5 Technological Advances in Studies of Transcript Elongation

The fundamental catalytic mechanisms and regulatory functions of transcript elongation have been extensively studied since the discovery of RNAPs, primarily using classical biochemistry and genetics as tools. The first high-resolution X-ray crystal structure of a multi-subunit RNAP in 1999[62] sets the stage for the elucidation of a series of RNAP crystal structures ranging from core RNAP to transcription elongation complexes with bound NTP. These structures have produced a high-resolution picture of transcript elongation largely consistent with results obtained from classical biochemical and genetic approaches. Only in the past few years have several novel biophysical techniques such as single-molecule force clamps,[50,69,78] and the renewed interest in fluorescence-based spectroscopy and microscopy such as fluorescence resonance energy transfer experiments (FRET),[84] and chromatin immunoprecipitation followed by whole-genome microarrays(ChIP-chip)[85,86] been successfully applied to investigate the mechanisms and regulation of transcript elongation. These powerful and increasingly accessible tools, backed up by classical biochemical and genetic approaches, have greatly advanced the field and will allow investigators to begin unraveling the complex structural mechanisms of transcript elongation and its regulation. More importantly, these new technologies may allow us to seek answers to several fundamental questions inaccessible to the classical experimental approaches. The characterization of force–velocity relationships revealed by single-molecule force-clamp assays[78] and the observation of DNA scrunching in initiating complex measured by FRET assays[84] and single-molecule DNA nanomanipulation[87] are excellent examples of such applications.

7.6 Concluding Remarks

In this chapter we have provided an overview of our current understanding of the mechanisms underlying transcript elongation and its control. Specifically, we have addressed research questions involving how NTP substrates enter the RNAP active site, of how key conformational changes near the active site involving the trigger loop and the bridge helix control nucleotide addition, of how catalysis of nucleotide addition, pyrophosphorolysis, and transcript cleavage occur, and of how RNAP threads DNA and RNA through the main channel *via* repeated translocation events.

Nonetheless, several fundamental questions concerning the molecular mechanisms of transcript elongation remain important challenges that await

investigators. It remains a technical challenge to obtain direct evidence that distinguishes the secondary-channel and main-channel models of NTP entry, to directly observe and measure key movements of the trigger loop/bridge helix during nucleotide addition, and to accurately and non-invasively monitor translocation in real-time. We are hopeful that the burgeoning applications of novel biophysical and biochemical approaches in the transcription field can help provide answers to these important questions in the near future.

References

1. I. Toulokhonov and R. Landick, The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination., *Mol. Cell*, 2003, **12**, 1125–36.
2. T. Santangelo, R. Mooney, R. Landick and J. W. Roberts, RNA polymerase mutations that impair reconfiguration to a termination resistant complex by Q antiterminator proteins., *Genes Dev.*, 2003, **17**, 1281–1292.
3. J. P. Richardson, Loading Rho to terminate transcription., *Cell*, 2003, **114**, 157–9.
4. P. H. von Hippel and Z. Pasman, Reaction pathways in transcript elongation., *Biophys. Chem.*, 2002, **101–102**, 401–23.
5. I. Gusarov and E. Nudler, Control of intrinsic transcription termination by N and NusA: the basic mechanisms., *Cell*, 2001, **107**, 437–49.
6. U. Vogel and K. F. Jensen, The RNA chain elongation rate in *Escherichia coli* depends on the growth rate., *J. Bacteriol.*, 1994, **176**, 2807–13.
7. B. R. Bochner and B. N. Ames, Complete analysis of cellular nucleotides by two-dimensional thin layer chromatography., *J. Biol. Chem.*, 1982, **257**, 9759–69.
8. A. M. Edwards, C. M. Kane, R. A. Young and R. D. Kornberg, Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro., *J. Biol. Chem.*, 1991, **266**, 71–5.
9. G. Rhodes and M. J. Chamberlin, Ribonucleic acid chain elongation by *Escherichia coli* ribonucleic acid polymerase: Isolation of ternary complexes and the kinetics of elongation., *J. Biol. Chem.*, 1974, **249** 6675–6683.
10. S. A. Kumar and J. S. Krakow, Studies on the product binding site of the *Azotobacter vinelandii* ribonucleic acid polymerase., *J. Biol. Chem.*, 1975, **250**, 2878–2884.
11. V. W. Armstrong, D. Yee and F. Eckstein, Mechanistic studies on deoxyribonucleic acid dependent ribonucleic acid polymerase from *Escherichia coli* using phosphorothioate analogues. 2. The elongation reaction., *Biochemistry*, 1979, **18**, 4120–4123.
12. D. A. Erie, T. D. Yager and P. H. von Hippel, The single-nucleotide addition cycle in transcription: a biophysical and biochemical perspective., *Annu. Rev. Biophys. Biomol. Struct.*, 1992, **21**, 379–415.

13. I. Toulokhonov, J. Zhang, M. Palangat and R. Landick, A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing., *Mol. Cell*, 2007, **27**, 406–19.
14. I. Artsimovitch and R. Landick, Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals., *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 7090–5.
15. D. G. Vassylyev, M. N. Vassylyeva, J. Zhang, M. Palangat, I. Artsimovitch and R. Landick, Structural basis for substrate loading in bacterial RNA polymerase., *Nature*, 2007, **448**, 163–8.
16. D. G. Vassylyev, M. N. Vassylyeva, A. Perederina, T. H. Tahirov and I. Artsimovitch, Structural basis for transcription elongation by bacterial RNA polymerase., *Nature*, 2007, **448**, 157–62.
17. D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan and R. D. Kornberg, Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis., *Cell*, 2006, **127**, 941–54.
18. H. Kettenberger, A. Eisenfuhr, F. Brueckner, M. Theis, M. Famulok and P. Cramer, Structure of an RNA polymerase II-RNA inhibitor complex elucidates transcription regulation by noncoding RNAs., *Nat Struct. Mol. Biol.*, 2006, **13**, 44–8.
19. S. Tuske, S. G. Sarafianos, X. Wang, B. Hudson, E. Sineva, J. Mukhopadhyay, J. J. Birktoft, O. Leroy, S. Ismail, A. D. Clark Jr, C. Dharia, A. Napoli, O. Laptenko, J. Lee, S. Borukhov, R. H. Ebright and E. Arnold, Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation., *Cell*, 2005, **122**, 541–52.
20. D. Temiakov, N. Zenkin, M. N. Vassylyeva, A. Perederina, T. H. Tahirov, E. Kashkina, M. Savkina, S. Zorov, V. Nikiforov, N. Igarashi, N. Matsugaki, S. Wakatsuki, K. Severinov and D. G. Vassylyev, Structural basis of transcription inhibition by antibiotic streptolydigin., *Mol Cell*, 2005, **19**, 655–66.
21. K. D. Westover, D. A. Bushnell and R. D. Kornberg, Structural basis of transcription: separation of RNA from DNA by RNA polymerase II., *Science*, 2004, **303**, 1014–6.
22. H. Kettenberger, K. J. Armache and P. Cramer, Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS., *Mol Cell*, 2004, **16**, 955–65.
23. Y. A. Nedialkov, X. Q. Gong, S. L. Hovde, Y. Yamaguchi, H. Handa, J. H. Geiger, H. Yan and Z. F. Burton, NTP-driven translocation by human RNA polymerase II., *J. Biol. Chem.*, 2003, **278**, 18303–12.
24. H. Kettenberger, K. J. Armache and P. Cramer, Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage., *Cell*, 2003, **114**, 347–57.
25. D. G. Vassylyev, S. Sekine, O. Laptenko, J. Lee, M. N. Vassylyeva, S. Borukhov and S. Yokoyama, Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution., *Nature*, 2002, **417**, 712–9.

26. S. A. Darst, N. Opalka, P. Chacon, A. Polyakov, C. Richter, G. Zhang and W. Wriggers, Conformational flexibility of bacterial RNA polymerase., *Proc. Natl. Acad. Sci. USA.*, 2002, **99**, 4296–301.
27. A. Gnatt, P. Cramer, J. Fu, D. Bushnell and R. D. Kornberg, Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution., *Science*, 2001, **292**, 1876–82.
28. P. Cramer, D. Bushnell, J. Fu, A. Gnatt, B. Maier-Davis, N. Thompson, R. Burgess, A. Edwards, P. David and R. Kornberg, Architecture of RNA polymerase II and implications for the transcription mechanism., *Science*, 2000, **288**, 640–649.
29. L. M. Iyer, E. V. Koonin and L. Aravind, Evolution of bacterial RNA polymerase: implications for large-scale bacterial phylogeny, domain accretion, and horizontal gene transfer., *Gene*, 2004, **335**, 73–88.
30. V. Sosunov, E. Sosunova, A. Mustaev, I. Bass, V. Nikiforov and A. Goldfarb, Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase., *EMBO J*, 2003, **22**, 2234–44.
31. V. Sosunov, S. Zorov, E. Sosunova, A. Nikolaev, I. Zakeyeva, I. Bass, A. Goldfarb, V. Nikiforov, K. Severinov and A. Mustaev, The involvement of the aspartate triad of the active center in all catalytic activities of multisubunit RNA polymerase., *Nucleic Acids Res.*, 2005, **33**, 4202–11.
32. D. Yee, V. W. Armstrong and F. Eckstein, Mechanistic studies on deoxyribonucleic acid dependent ribonucleic acid polymerase from *Escherichia coli* using phosphorothioate analogues. 1. Initiation and pyrophosphate exchange reactions., *Biochemistry*, 1979, **18**, 4116–4120.
33. T. A. Steitz, A mechanism for all polymerases., *Nature*, 1998, **391**, 231–2.
34. Q. Guo and R. Sousa, Translocation by T7 RNA polymerase: a sensitively poised Brownian ratchet., *J Mol Biol*, 2006, **358**, 241–54.
35. J. E. Foster, S. F. Holmes and D. A. Erie, Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation., *Cell*, 2001, **106**, 243–52.
36. S. J. Greive and P. H. von Hippel, Thinking quantitatively about transcriptional regulation., *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 221–32.
37. J. J. Hopfield, Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity., *Proc. Natl. Acad. Sci. USA*, 1974, **71**, 4135–9.
38. M. Orlova, J. Newlands, A. Das, A. Goldfarb and S. Borukhov, Intrinsic transcript cleavage activity of RNA polymerase., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 4596–4600.
39. E. Sosunova, V. Sosunov, M. Kozlov, V. Nikiforov, A. Goldfarb and A. Mustaev, Donation of catalytic residues to RNA polymerase active center by transcription factor Gre., *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 15469–74.
40. D. A. Erie, O. Hajiseyedjavadi, M. C. Young and P. H. von Hippel, Multiple RNA polymerase conformations and GreA: control of fidelity of transcription., *Science*, 1993, **262**, 867–873.

41. N. Komissarova and M. Kashlev, Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3′ end of the RNA intact and extruded., *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 1755–1760.
42. C. K. Surratt, S. C. Milan and M. J. Chamberlin, Spontaneous cleavage of RNA in ternary complexes of *Escherichia coli* RNA polymerase and its significance for the mechanism of transcription., *Proc. Natl. Acad. Sci. USA*., 1991, **88**, 7983–7987.
43. S. Borukhov, V. Sagitov and A. Goldfarb, Transcript cleavage factors from *E. coli*., *Cell*, 1993, **72**, 459–466.
44. S. Borukhov, A. Polyakov, V. Nikiforov and A. Goldfarb, GreA protein: a transcription elongation factor from *Escherichia coli*., *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 8899–8902.
45. M. Izban and D. Luse, The RNA polymerase II ternary complex cleaves the nascent transcript in a 3′->5′ direction in the presence of elongation factor SII., *Genes Dev*., 1992, **6**, 1342–1356.
46. N. Opalka, M. Chlenov, P. Chacon, W. J. Rice, W. Wriggers and S. A. Darst, Structure and function of the transcription elongation factor GreB Bound to bacterial RNA polymerase., *Cell*, 2003, **114**, 335–45.
47. O. Laptenko, J. Lee, I. Lomakin and S. Borukhov, Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase., *EMBO J*., 2003, **22**, 6322–34.
48. N. Zenkin, Y. Yuzenkova and K. Severinov, Transcript-assisted transcriptional proofreading., *Science*, 2006, **313**, 518–20.
49. S. Kyzer, K. S. Ha, R. Landick and M. Palangat, Direct versus limited-step reconstitution reveals key features of an RNA hairpin-stabilized paused transcription complex., *J. Biol. Chem*., 2007, **282**, 19020–8.
50. K. M. Herbert, A. La Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick and S. M. Block, Sequence-resolved detection of pausing by single RNA polymerase molecules., *Cell*, 2006, **125**, 1083–94.
51. R. Landick, The regulatory roles and mechanism of transcriptional pausing., *Biochem. Soc. Trans*., 2006, **34**, 1062–6.
52. M. Palangat, C. T. Hittinger and R. Landick, Downstream DNA selectively affects a paused conformation of human RNA polymerase II., *J. Mol. Biol*., 2004, **341**, 429–442.
53. C. Chan, D. Wang and R. Landick, Multiple interactions stabilize a single paused transcription intermediate in which hairpin to 3′ end spacing distinguishes pause and termination pathways., *J. Mol. Biol*., 1997, **268**, 54–68.
54. N. N. Batada, K. D. Westover, D. A. Bushnell, M. Levitt and R. D. Kornberg, Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center., *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 17361–4.
55. K. D. Westover, D. A. Bushnell and R. D. Kornberg, Structural basis of transcription; nucleotide selection by rotation in the RNA polymerase II active center., *Cell*, 2004, **119**, 481–9.

56. X. Q. Gong, C. Zhang, M. Feig and Z. F. Burton, Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II., *Mol. Cell*, 2005, **18**, 461–70.
57. Z. F. Burton, M. Feig, X. Q. Gong, C. Zhang, Y. A. Nedialkov and Y. Xiong, NTP-driven translocation and regulation of downstream template opening by multi-subunit RNA polymerases., *Biochem. Cell Biol.*, 2005, **83**, 486–96.
58. G. Bar-Nahum, V. Epshtein, A. Ruckenstein, R. Rafikov, A. Mustaev and E. Nudler, A ratchet mechanism of transcription elongation and its control., *Cell*, 2005, **120**, 183–193.
59. V. Epshtein, A. Mustaev, V. Markovtsov, O. Bereshchenko, V. Nikiforov and A. Goldfarb, Swing-gate model of nucleotide entry into the RNA polymerase active center., *Mol. Cell*, 2002, **10**, 623–634.
60. P. Cramer, D. Bushnell and R. Kornberg, Structural basis of transcription: RNA polymerase II at 2.8 Å resolution., *Science*, 2001, **292**, 1863–76.
61. A. Gnatt, J. Fu and R. D. Kornberg, Formation and crystallization of yeast RNA polymerase II elongation complexes., *J. Biol. Chem.*, 1997, **272**, 30799–805.
62. G. Zhang, E. A. Campbell, L. Minakhin, C. Richter, K. Severinov and S. A. Darst, Crystal structure of *Thermus aquaticus core* RNA polymerase at 3.3 Å resolution., *Cell*, 1999, **98**, 811–24.
63. Y. Huang, A. Beaudry, J. McSwiggen and R. Sousa, Determinants of ribose specificity in RNA polymerization: effects of Mn2+ and deoxynucleoside monophosphate incorporation into transcripts., *Biochemistry*, 1997, **36**, 13718–28.
64. J. J. Arnold, D. W. Gohara and C. E. Cameron, Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mn2+., *Biochemistry*, 2004, **43**, 5138–48.
65. D. Pinto, M. T. Sarocchi-Landousy and W. Guschlbauer, 2′-Deoxy-2′-fluorouridine-5′-triphosphates: a possible substrate for *E. coli* RNA polymerase., *Nucleic. Acids Res.*, 1979, **6**, 1041–8.
66. A. Chakrabartty, T. Kortemme and R. L. Baldwin, Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions., *Protein Sci.*, 1994, **193**, 843–52.
67. V. Markovtsov, A. Mustaev and A. Goldfarb, Protein-RNA interactions in the active center of the transcription elongation complex., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 3221–3226.
68. R. Weilbaecher, C. Hebron, G. Feng and R. Landick, Termination-altering amino acid substitutions in the β′ subunit of *Escherichia coli* RNA polymerase identify regions involved in RNA chain elongation., *Genes Dev.*, 1994, **8**, 2913–2917.
69. M. Wang, M. Schnitzer, H. Yin, R. Landick, J. Gelles and S. Block, Force and velocity measured for single molecules of RNA polymerase., *Science*, 1998, **282**, 902–907.

70. M. Burmeister, A. P. Monaco, E. F. Gillard, G. J. van Ommen, N. A. Affara, M. A. Ferguson-Smith, L. M. Kunkel and H. Lehrach, A 10-megabase physical map of human Xp21, including the Duchenne muscular dystrophy gene., *Genomics*, 1988, **2**, 189–202.
71. Y. W. Yin and T. A. Steitz, The structural mechanism of translocation and helicase activity in T7 RNA polymerase., *Cell*, 2004, **116**, 393–404.
72. R. Guajardo, P. Lopez, M. Dreyfus and R. Sousa, NTP concentration effects on initial transcription by T7 RNAP indicate that translocation occurs through passive sliding and reveal that divergent promoters have distinct NTP concentration requirements for productive initiation., *J. Mol. Biol.*, 1998, **281**, 777–92.
73. R. Guajardo and R. Sousa, A model for the mechanism of polymerase translocation., *J. Mol. Biol.*, 1997, **265**, 8–19.
74. L. Bai, A. Shundrovsky and M. D. Wang, Sequence-dependent kinetic model for transcription elongation by RNA polymerase., *J. Mol. Biol.*, 2004, **344**, 335–49.
75. V. R. Tadigotla, O. M. D, A. M. Sengupta, V. Epshtein, R. H. Ebright, E. Nudler and A. E. Ruckenstein, Thermodynamic and kinetic modeling of transcriptional pausing., *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 4439–44.
76. G. Oster and H. Wang, Rotary protein motors., *Trends. Cell Biol.*, 2003, **13**, 114–21.
77. R. D. Vale and R. A. Milligan, The way things move: looking under the hood of molecular motor proteins., *Science*, 2000, **288**, 88–95.
78. E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick and S. M. Block, Direct observation of base-pair stepping by RNA polymerase., *Nature*, 2005, **438**, 460–5.
79. P. Thomen, P. J. Lopez and F. Heslot, Unravelling the mechanism of RNA-polymerase forward motion by using mechanical force., *Phys. Rev. Lett.*, 2005, **94**, 128102.
80. E. Kashkina, M. Anikin, F. Brueckner, E. Lehmann, S. N. Kochetkov, W. T. McAllister, P. Cramer and D. Temiakov, Multisubunit RNA polymerases melt only a single DNA base pair downstream of the active site., *J. Biol. Chem.*, 2007, **282**, 21578–82.
81. S. F. Holmes and D. A. Erie, Downstream DNA sequence effects on transcription elongation. Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion., *J. Biol. Chem.*, 2003, **278**, 35597–608.
82. Z. Pasman and P. H. von Hippel, Active *Escherichia coli* transcription elongation complexes are functionally homogeneous., *J. Mol. Biol.*, 2002, **322**, 505–19.
83. T. D. Yager and P. H. von Hippel, A thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli*., *Biochemistry*, 1991, **30**, 1097–1118.
84. A. N. Kapanidis, E. Margeat, S. O. Ho, E. Kortkhonjia, S. Weiss and R. H. Ebright, Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism., *Science*, 2006, **314**, 1144–7.

85. N. B. Reppas, J. T. Wade, G. M. Church and K. Struhl, The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting., *Mol. Cell*, 2006, **24**, 747–57.
86. C. D. Herring, M. Raffaelle, T. E. Allen, E. I. Kanin, R. Landick, A. Z. Ansari and B. O. Palsson, Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays., *J. Bacteriol.*, 2005, **187**, 6166–74.
87. A. Revyakin, C. Liu, R. H. Ebright and T. R. Strick, Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching., *Science*, 2006, **314**, 1139–43.