Journal of Theoretical and Computational Chemistry
Vol. 12, No. 8 (2013) 1341005 (16 pages)

DOI: 10.1142/S0219633613410058

THEORETICAL INVESTIGATIONS ON ELUCIDATING FUNDAMENTAL MECHANISMS OF CATALYSIS AND DYNAMICS INVOLVED IN TRANSCRIPTION BY RNA POLYMERASE

FÁTIMA PARDO-AVILA*, LIN-TAI DA*, YING WANG*
and XUHUI HUANG*,†,‡

*Department of Chemistry
Center of Systems Biology and Human Health
Institute for Advance Study and School of Science
Hong Kong University of Science and Technology
Clear Water Bay Road, Kowloon, Hong Kong

†Division of Biomedical Engineering
Center of Systems Biology and Human Health
Institute for Advance Study and School of Science
Hong Kong University of Science and Technology
Clear Water Bay Road, Kowloon, Hong Kong
‡xuhuihuang@ust.hk

Received 27 March 2013
Accepted 26 May 2013
Published 30 July 2013

RNA polymerase is the enzyme that synthesizes RNA during the transcription process. To understand its mechanism, structural studies have provided us pictures of the series of steps necessary to add a new nucleotide to the nascent RNA chain, the steps altogether known as the nucleotide addition cycle (NAC). However, these static snapshots do not provide dynamic information of these processes involved in NAC, such as the conformational changes of the protein and the atomistic details of the catalysis. Computational studies have made efforts to fill these knowledge gaps. In this review, we provide examples of different computational approaches that have improved our understanding of the transcription elongation process for RNA polymerase, such as normal mode analysis, molecular dynamic (MD) simulations, Markov state models (MSMs). We also point out some unsolved questions that could be addressed using computational tools in the future.

Keywords: Molecular dynamics; Markov state models; molecular motors.

1. Introduction

In both eukaryotic and prokaryotic cells, gene transcription is carried out by multi-subunit RNA polymerases that share a homologous catalytic core (see Fig. 1).[1,2] During the RNA synthesis, the transcription elongation complex (TEC) repeatedly

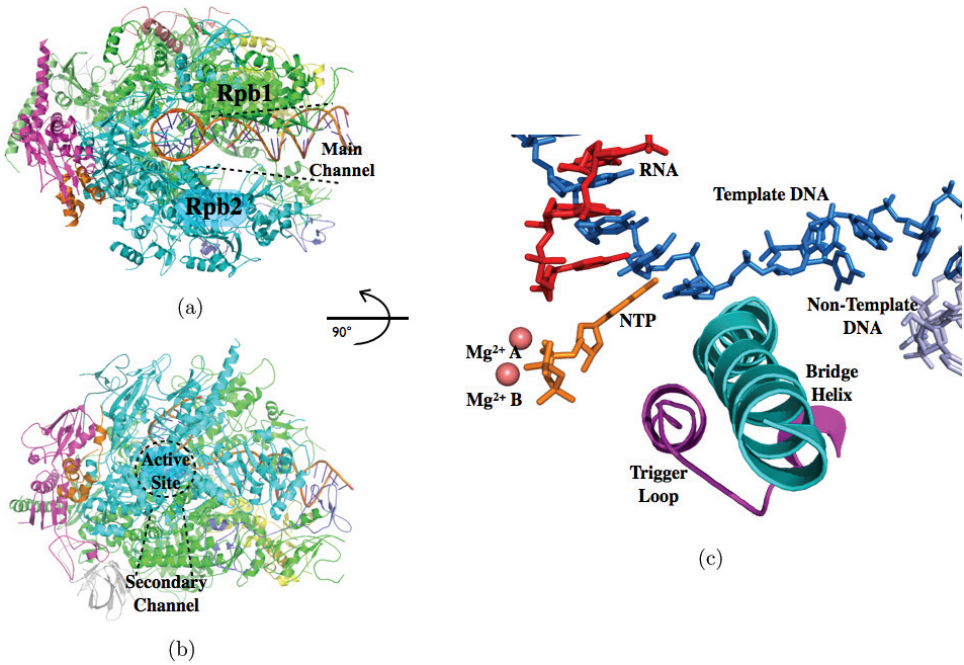

Fig. 1. Structural features of RNA Pol II. (a) Top view of RNA Pol II, PDB entry 2E2H. Each chain is shown in a different color, Rpb1 in green and Rpb2 in cyan. The main channel is signaled by dashed black lines and with template DNA inside. (b) Side view of Pol II. A rotation of the structure in panel A by 90° produces the side view of RNA polymerase. Dashed lines signal the secondary channel, and a circle marks the location of the active site. (c) Active site of the Pol II. The main components of the active site are shown. The template DNA in marine, nascent RNA chain in red, and incoming NTP is shown in orange. Mg^{2+} A and Mg^{2+} B are shown as salmon spheres. Two important and conserved domains are also shown: BH in cyan and closed TL in magenta.

reads through the genetic information in each nucleotide addition cycle (NAC) with the assistance of two critical domains underneath the polymerase active site, termed as trigger loop (TL) and bridge helix (BH). In general, each NAC includes nucleoside triphosphate (NTP) entry, followed by the catalytic reaction, the pyrophosphate ion (PP_i) release and the translocation of the DNA template (see Fig. 2).

X-ray crystallographic studies have greatly enhanced our understandings of the atomic structures of the RNA polymerases at different transcription stages. During the NTP entry, two binding sites: "E" and "A" sites were identified in the eukaryotic RNA polymerase (Pol II). The entry (E) site allows binding of all four different NTPs, while the addition (A) site only allows the binding of the correctly matched NTP.[3,4] Then the TL can form a close interaction network with the incoming NTP in the A-site to facilitate the catalysis and assure the fidelity of transcription.[4] The catalysis involves a two-magnesium catalyzed reaction resulting in formation of the PP_i group, which can be released from the active site. Finally, the TEC must translocate to complete the cycle and create an empty A-site in order to re-initiate

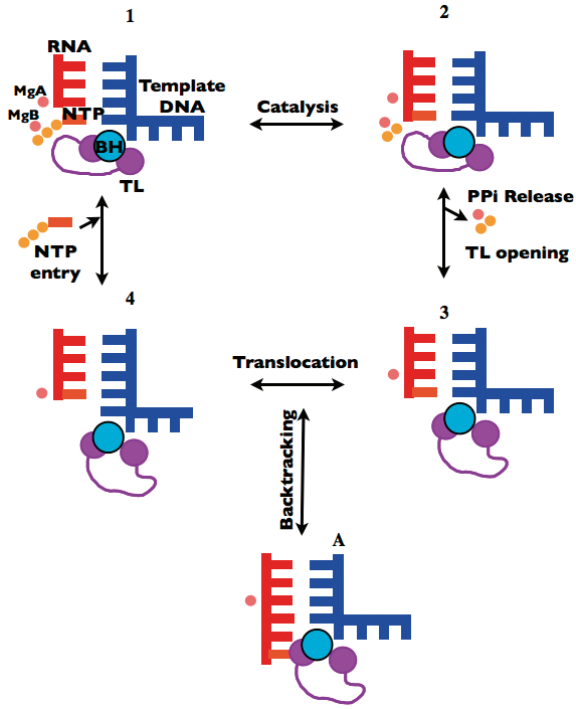

Fig. 2. NAC for RNA polymerase II. (1) NTP entry: The NTP is in the active site and the TL is closed, all the elements ready for catalysis. (2) After the addition of the NTP to the nascent RNA chain, the PP_i remains in the active site and TL is still closed. (3) Pre-translocation. After the PP_i release and the opening of the TL, the system is in the pre-translocation state. (4) Post-translocation. The nucleotides move by exactly one register to go from pre- to post-translocation. After translocation, the active site is empty and another cycle can begin. (A) Backtracked state. In the case of mismatch or DNA damage, the Polymerase can go to a backtracked state in which the RNA can be cleaved.

another NAC. To date, people have already obtained most of the X-ray structures of RNA polymerase participating in the different steps of NAC, providing us a static picture of how the RNA polymerases perform their functions.[1,3–10]

In addition, mutagenesis experiments have revealed many residues that may play critical roles during the transcription elongation process, such as the active site residue N479, TL residues H1085, Q1078, L1081, and the YFI motif in the BH (*E. coli* β' residues 772-YFI-774).[11–13] Moreover, combinations of single-molecule experiments, fluorescence and kinetic studies were employed to directly determine the rates of the different steps in the NAC, as well as characterizing other important states such as pausing and backtracking.[14–17] Nevertheless, crystal structures can only provide the static snapshots, but important dynamic information is still missing, such as how the enzyme conducts the catalytic reaction step by step, how the PP_i group is released from the active site and whether it is tightly coupled with the TL opening motion, and how the TL performs its function dynamically. Yet currently it is still difficult to address the above questions using only experimental methods.

Instead, theoretical studies can provide detailed dynamic information of extensive biological systems at an atomic description. Molecular dynamics (MD) simulations have been widely used as a theoretical tool to study the conformational changes of many biomolecules, and have been successfully applied to study the Pol II transcription elongation complex.[12,18–20] However, one major challenge for simulations of big systems — like Pol II, with nearly 500,000 atoms in a solvated simulation box — is to reach its biologically relevant timescales (∼ millisecond or longer). Even with the modern hardware like Anton — a supercomputer specially designed for MD simulations of biological system[21] — it is still challenging to directly simulate Pol II at millisecond timescale.

Coarse-grained (CG) models provide a powerful tool to greatly accelerate the MD simulation of biological systems compared with the all-atom (AA) MD simulations.[22–24] Current CG models allow to perform MD simulations of systems of ∼10 million atoms by using only ∼1 million particles to represent them. Thus, CG models have a great potential to bridge the gap between the MD simulations and biologically relevant events. In the past years, several CG force fields (FF), such as the MARTINI FF[22] and Shinoda-Devane-Kevin (SDK) FF[24] have been developed and successfully used to study biological systems like bio-membranes. However, in some cases it is important to have atomistic detail for specific regions of the systems under study (such as the active site of the enzymes), in those cases it is necessary to have an AA representation to describe the solute. Thus, a hybrid model could be used to embed the AA biological system into a CG solvent.[25] Although these CG models have not been applied to study the RNA polymerase, they have promising potential to capture the conformational changes that cannot be accessible by AA MD simulations.

Another approach to overcome the timescale gap is to construct Markov state models (MSMs) from extensive short MD simulations. The MSM can not only predict the long timescale dynamic behaviors of the systems we are interested, but also can provide the kinetic properties that are inaccessible through conventional MD simulations.[26–32]

Apart from this, in order to understand the catalysis of the biological enzymes, quantum mechanics (QM) and the hybrid quantum mechanics/molecular mechanics (QM/MM) have been widely applied to understand how the enzymes perform their functions step by step.[33] In both DNA polymerase and RNA polymerases, people have suggested a general two-metal catalyzed reaction mechanism during the DNA/RNA synthesis in which two-hydrogen transfer processes were proposed to occur.[34,35] Currently, several groups have successively applied the QM/MM methodology on both DNA and RNA polymerases to understand their catalysis.[36]

This review summarizes the recent studies about the transcription elongation process with emphasis on theoretical studies. The contents are described in the following order: NTP loading, TL motion and its regulation of the transcription, catalytic reaction, PP_i release mechanism, translocation and off-pathway states. We also introduced the current understandings of the each step involved in the transcription elongation process as well as some open questions that still remain

unsolved. At the end, we wrapped all the discussions up in order to give an overall picture of the complete elongation process.

2. NTP Loading

The first step of the NAC is the entry of the NTP into the enzyme, followed by the recognition and positioning of the correctly matched NTP in the active site. This process raises many interesting questions. The first set of questions arise from the fact that the active site is buried more than 30 Å away from the surface inside the enzyme, and the pathway that the NTPs follow to reach the active site remains unclear.

From inspection of the X-ray crystal structures, it is possible to see that the secondary channel provides a route of entry from the solvent directly to the active site (see Fig. 1(b)). For this reason the secondary channel has been proposed as a possible entry of the nucleotides.[3,37] In contrast, other studies proposed that the nucleotides can enter through the main channel, the cavity through which the template DNA enters the enzyme (see Fig. 1(a)).[38] These studies show that the presence of NTPs that match the downstream template DNA can increase the fidelity of RNAP, taken together with single molecule experiments that proposed that the NTP binding can occur in the pre-translocation state, leading to the hypothesis that the incoming NTPs can bind the template DNA in the main channel before they reach the active site.[17,38] However, since the X-ray crystal structures do not show the complete transcription bubble, it may be possible that the non-template DNA could interfere with the binding of the NTPs in the main channel. The entry path of the NTP is an interesting and challenging question that could be addressed by computational methods.

Another important question concerning this process is the rate at which the nucleotides reach the active site, which in turn could determine the rate of the elongation process. We also need to consider that some of the nucleotides that can reach the active site will not be complimentary to the template DNA, so the non-matching nucleotide must leave the active site to let another nucleotide enter, until the correctly matched NTP is found. Batada *et al.* calculated the diffusion rate of the NTPs to the active site through the secondary channel.[39] They used the topography of the secondary channel from X-ray crystal (PDB id: 1R9T) and calculated the trajectory of the diffusion of the NTPs using Langevin stochastic differential equations. They found that at concentrations near *in vivo* conditions, the rate at which the NTPs reached the active site was ~ 200/s, and the rate of successful binding was less than 20/s. Since the addition rate measured for RNA Pol II is around 10 nucleotides per second, the calculated diffusion rate seems proper to achieve the synthesis rate.

After the NTP reaches the active site, the NTP needs to be correctly positioned before the catalytic reaction. However, the mechanism by which the NTP is correctly positioned in the active site is still elusive. It has been proposed that this loading could be aided by the binding of NTPs to another site near the active site, which has

been captured in crystal structures and termed as E-site (entry site).[4] In the study by Batada *et al.*, they found that adding the E-site in their calculations increased the binding rate to the A-site. Thus, they proposed that as a result of the ionic nature of the E-site, the E-site can chelate the Mg ions and can provide a binding site with less steric clashes than the A-site, increasing the binding affinity to the E-site.[39] Nonetheless, the role of the E-site is not clear, so as how the incoming NTP can rotate from its position in the E-site to its final catalytically active position in the A-site. These problems still need to be further addressed.

3. Trigger Loop (TL) Motion and Its Regulation of the Transcription

TL domain is located below the BH, near the active site. It can adopt two conformations: closed and open. In its closed conformation it can directly interact with the loaded NTP (see Fig. 3). This suggests that the TL could play a role in catalysis and the regulation of the reaction.[4,17,37] Computational approaches have been used to identify the role of the residues that directly interact with the NTP. These studies complement experiments since they can provide a molecular explanation to the phenotypes observed in mutagenesis studies.

In the NTP-Pol II binding complex, the TL residue L1081 forms hydrophobic interactions with the incoming NTP, which can help stabilize and position the base

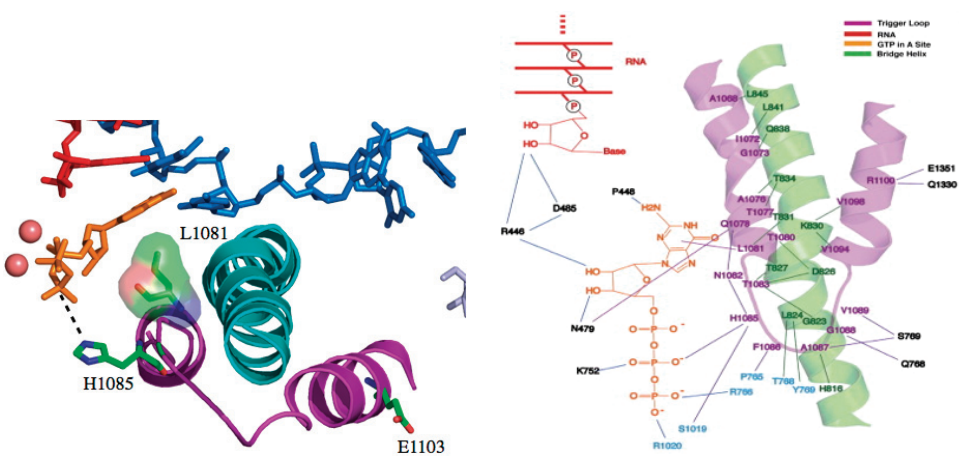

Fig. 3. Interaction of the residues on the TL with the incoming NTP. (a) The role of L1081 and H1085 for stabilizing the incoming NTP. L1081 is shown in surface representation to show the hydrophobic interaction with the base of the nucleotide. H1085 is shown in sticks representation and a dashed line shows the possible interaction of Nε with the P$_\beta$ of the NTP. E1103 is shown in sticks representations, to highlight its location far from the active site. The color code used is the same as the one used in Fig. 1. (b) Network of interactions between the residues of the protein in the active site and the incoming nucleotide, reproduced from (4).

of the incoming NTP. In order to test this, Huang *et al.* performed MD simulations and *in silico* mutations to understand the roles of L1081. They found that this residue helps to position the incoming NTP, by promoting base stacking, base pairing and stabilizing the NTP.[12]

The residue H1085 interacts with the incoming NTP, and has also been regarded as a participant in the catalytic reaction.[4] Huang *et al.* observed that the *in silico* mutant to phenylalanine was unstable and diminished the interaction between the residue and the NTP. The H1085Y mutant was more stable than the H1085F mutant and could keep some interactions with the NTP. These results are consistent with the experimental evidence that the H1085F mutant is lethal but the H1085Y mutant only decreases the elongation rate. Furthermore, H1085 can potentially adopt several protonation states (HIE — when the hydrogen is connect to Nε, HID — hydrogen connected to Nδ, HIP — hydrogen connected both to Nε and Nδ). Huang *et al.* found that the stability of the active site is dependent of the H1085 protonation state, and the most stable species is the protonated form (HIP) of the H1085 due to its ability to form both hydrogen bonds and salt bridges interactions with the phosphate group of the NTP.[12]

In addition, the TL has also been shown to participate in the discrimination between RNA and DNA.[4,40] Furthermore, many mutational studies have revealed striking mutants, like E1003G that despite being far from the active site (see Fig. 3(a)), can increase the elongation rate of Pol II, and also rescue some lethal mutants.[11,18] These findings suggest that the TL plays important roles in the regulation of the elongation process, which could be related to its motility, thus understanding the motion of the TL can greatly enhance our understanding of the NTP loading.

4. Catalytic Reaction

After the NTP loading in the active site, the nucleic addition reaction occurs through a proposed two-metal catalyzed mechanism, resulting in the formation of a phosphodiester bond between the incoming nucleotide (on register +1) and the preceding nucleotides. Binding of the metal ions to the nucleotides has been found to play an important role on the regulation of the biological functions of the nucleotides including catalysis.[41,42] Recently, Carvalho *et al.* performed an ONIOM study to reveal the detailed catalytic reaction mechanism in Pol II.[36] They built the NTP-bound Pol II complex based on a recent crystal structure (PDB id: 2E2H), and the final solvated complete model was pre-equilibrated by performing MD simulations using amber99 force field. The resulting snapshot was used as the input structure for the QM/MM study with two-layers ONIOM methodology. The model was optimized at B3LYP/6-31G(d):PM3MM level and the energies were obtained at B3LYP/6-3111++G(2d,2p) level. In addition, several potential bases were considered to activate the 3′-terminal RNA nucleotide in the study, and different protonation states for the leaving PP_i group were also included.

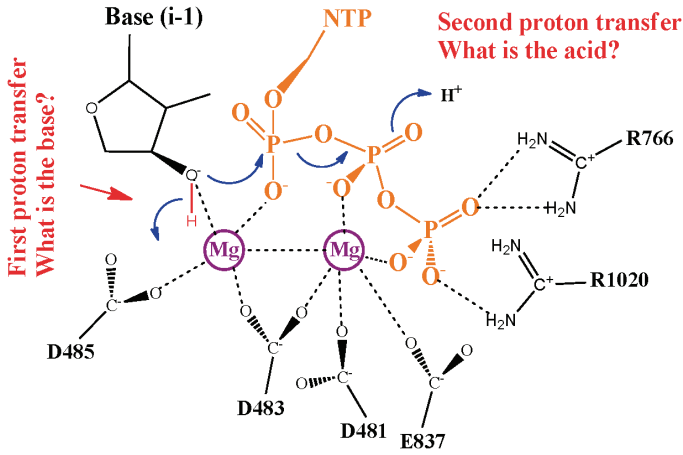

Fig. 4. Catalytic mechanism of RNA polymerase. A two metal ion mechanism has been proposed for the reaction. The two main questions regarding the mechanism are highlighted in red.

Their ONIOM calculations indicate that the 3′-OH of the terminal RNA nucleotide can first deprotonate the hydrogen atom to hydroxide ion in the solvent. Then, the formed nucleophile attacks the alpha-phosphate atom of the incoming NTP (see Fig. 4). Finally, one PP_i group forms and accepts a proton from the TL histidine residue H1085. This two-proton transfer mechanism is consistent with experimental observations.[35] However, our modeling studies suggest that the protonated H1085 does not necessarily donate the proton to the PP_i group after the catalytic reaction, because both PP_i group and PPH_i group have similar tendency to release from the active site.[43] In addition, mutagenesis data from bacterial RNA polymerase also indicate that the single mutation of the TL residue H1242A decreases the nucleic addition rate only approximately three-fold, which is similar to the effect of mutation R1239A, suggesting that H1242 is likely to act as a transition-state stabilizer, instead of being a proton donor.[44]

5. PP_i Release Mechanism

PP_i release from the active site after the catalytic reaction is a necessary step during the transcription process and is suggested as a pre-requisite step for the translocation by both experimental and theoretical studies.[14] PP_i release in T7 RNA polymerase was found to be able to directly promote the translocation process.[45] However, in Pol II, the coupling between the PP_i release, TL opening motion and translocation are still inconclusive. Therefore, in order to reveal the interplay between the above processes, Da *et al.* created MSMs based on extensive MD simulations to capture the dynamics of the PP_i release in both Pol II and bacterial RNA polymerase (RNAP).[43]

The starting PP_i-bound Pol II complex was built from a crystal structure (PDB id: 2E2H). The final complex model was subject to several pre-equilibrium MD

simulations. Then, five representative snapshots were chosen for the subsequent steered MD (SMD) simulations to pull the PP_i group out of the active site. Finally, in order to eliminate the biases introduced from the SMD, they performed 122 conventional MD simulations starting form geometrically different conformations from the SMD simulations. Based on these MD simulations, a MSM was built to identify the metastable states and obtain the dynamic properties of the PP_i release process.

This study suggested that the TL tip domain becomes more flexible after formation of the PP_i group, which is likely due to the loss of the linkage between the PP_i group and the newly formed RNA chain. The MSM indicates that PP_i release in Pol II adopt a hopping model in which the PP_i group can transfer between three metastable states (see Fig. 5). In specific, the fluctuation of the TL tip domain can help the PP_i group transfer from the active site to the first metastable state (S1), facilitated by the residues K752 and H1085. Further single mutant MD simulations proved that the substitutions to alanine of either K752 or H1085 in the active site could delay the PP_i transfer to the S1 state. Next, two other positively charged residues along the secondary channel, K518 and K619, can transfer the PP_i group from S1 to S2 state, and mutation to alanine of the residue K619 at S1 state can prevent this transition. Finally, the PP_i group hops to the last metastable state (S3) in the pore region of the channel. The release of the PP_i group from S3 state results in the diffusion of the PP_i group into the funnel region and finally out of the secondary channel. The MFPT for the PP_i release along the pore region of the secondary channel was calculated to be $\sim 1.5\,\mu$s.

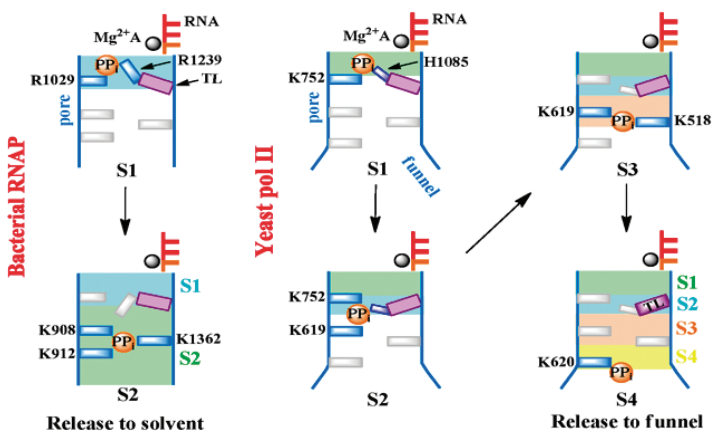

Fig. 5. PP_i release in prokaryotic and eukaryotic RNA polymerases. (a) Bacterial RNAP follows a two state model. The residues involved in the process are signaled. (b) PP_i release in RNA pol II follows a hopping model composed of four states. The cartoon shows the coupling between the motion of the tip of the TL and the PP_i release. In both panels, the background color represents the location of the different states along the channel.

The MSM suggested that the PP_i release is only coupled with the TL tip motion but due to the fast dynamics of the PP_i release, the process is uncoupled with the TL fully opening. The TL tip motion can facilitate the PP_i group transfer from active site to S1 state through the interactions between the TL residue H1085 and the PP_i group. The further release of the PP_i group from the S1 state can conversely flip down the TL tip domain.

For the PP_i release in bacteria RNA RNAP, we built the starting PP_i-bound RNAP complex from a crystal structure (PDB id: 2O5J). To obtain the unbiased MD simulations used to build a MSM, Da *et al.* adopted a strategy similar to the one described for PP_i release in Pol II. The MSM revealed that PP_i release in these two RNA polymerases follow different mechanisms (see Fig. 5).[46] The PP_i release in RNAP follows a two-states model that contrasts with the hopping model formed by four states observed in Pol II. In addition, the PP_i release in RNAP is three-fold faster compared to Pol II, which is consistent with the faster elongation rates observed for RNAP. Surprisingly, in RNAP the coupling between the PP_i release and the unfolding of the TL was not observed, although there exists correlation between PP_i release and the side-chain rotation of TL residue R1239. Furthermore, the PP_i release in both species is unlikely to induce the translocation process because no translocation was observed in our studies.

6. Translocation

The RNA polymerase reads processively through the DNA template, and synthesizes the RNA chain at one nucleotide at a time. In each step the nucleotides (DNA and RNA) must move exactly by one register in the process named translocation, in order to form an empty active site and be able to continue the NAC (see Fig. 2).

Two main theories have been proposed to explain the translocation mechanism. The power stroke model proposes that the translocation is driven by the energy derived from the NTP addition.[45] On the other hand, the two-pawl Brownian Ratchet model proposes that the translocation is driven by thermal fluctuation, and that the vibration of the BH promotes the motion of the nucleotides, while the incoming nucleotide acts like a pawl and can stall the system in the post-translocated state.[47] The second model has been proved experimentally in the multi-subunit RNA polymerases, while a combined model is suggested for T7 RNA polymerase.[48,49]

Due to the complexity of the transcription process, it is difficult to separate the contribution of the different component steps of the NAC to the elongation rates measured experimentally. Recently, a method to directly measure translocation rates was described. Using fluorescent probes located in strategic points in the enzyme and in the template DNA, Malinen *et al.* could measure the translocation rate and can determine the contribution of translocation to the elongation rate. However, the molecular mechanism of the process is still unknown.[14]

A crystal structure obtained in Cramer's group identified a possible intermediate of Pol II during the translocation process, captured by binding with the inhibitor

α-amanitin. In this structure the central part of the BH is bent towards the DNA: RNA hybrid, stabilized by a wedged TL through the interaction of L1081 with V829 in the BH, these features can stabilize the nucleic acid from the template DNA in an intermediate position between pre and post-translocation.[5] This still represents only a static structure and there are still open questions regarding the mechanism and dynamics of the translocation process.

Computational efforts have been made to improve the understanding of translocation. Normal mode analysis by Feig and Burton provided some insights into the motions involved in the translocation process and suggested that the translocation would require an open state of TL.[19] Another study by the same authors analyzed 100 ns MD simulations of RNA polymerase with open and closed conformations of the TL.[20] Their data suggest that the downstream DNA and the upstream DNA−RNA hybrid do not translocate simultaneously. They also observed that the bent BH conformation is more favorable when the TL is open, and that the TL open favored more significant forward translocation associated movements in their simulations. However, these simulations are not long enough to observe the whole translocation events that could occur at tens of microseconds or even longer.

Recently, Kireeva *et al.* performed short (10 ns) MD simulations for the bacterial RNA polymerase, with open and closed TL.[18] They confirmed their previous results that the TL needs to open before the translocation. Their simulations and experiments also help identify hinge residues — located in the N-terminal of the BH and at the base of the TL — that help regulate the BH bending. However, further studies are still needed to address the detailed dynamic process of the translocation process.

7. Pausing and Backtracking

As described before, the RNA polymerase can restart a new cycle after the translocation. However, if RNA polymerases detect special potential points of regulation or errors — such as intrinsic pausing sites, splicing sites, DNA damage, and nucleotide wrong incorporation — they can enter into a paused state.[50−52] The paused state can constitute some element of regulation in the transcription process. If the pausing period is too long it can lead to the termination of elongation of the RNA polymerase, or in the cause of encountering errors, it can move backwards into a backtracked state to enable proofreading. In the backtracked state, the RNA transcript is allowed to extrude, making it available for cleavage. X-ray crystallography has revealed the structures of the bacterial RNAP paused state and of the Pol II backtracked states.[10,53,54] On the other hand, single molecule experiments have provided information about how often the polymerase pauses along the elongation and footprinting experiments allow determine when the polymerase are trapped in the backtracked state.[13,15,55,56]

In a recent study, the structure of the *Thermus thermophilus* RNA polymerase in the paused state was described.[53] One of the key features of the paused state is a relaxed open clamp (the domain formed by Rpb1 and Rpb2, that forms the main

channel and holds the template DNA), which is a conformation that had also been proposed by fluorescence experiments.[57] The paused structure also possesses a kinked BH that blocks the RNAP active site, which explains the inhibition of the RNAP catalytic activity.

The backtracked RNA polymerases have been captured by X-ray crystallography. One of these structures observed few nucleotides backtracked that can be removed through the intrinsic hydrolysis mechanism.[10,54] However, when two or more nucleotides are backtracked, the enzyme stalls and an extra transcription factor (like TFIIS in eukaryotic cell) is needed to reactivate the protein. Crystal structures of TFIIS interacting with nucleotides in the active site of the enzyme are available. In the case of backtracking, it would be interesting to understand how the enzyme moves backwards, how the enzyme performs the intrinsic hydrolysis, how TFIIS enters the polymerase and how TFIIS performs the hydrolysis.[58,59]

Probing these mechanisms at atomic resolution is also difficult for other experimental techniques such as biochemical and genetic experiments. However, elucidation of such mechanisms is crucial for understanding the fundamental basis of transcription elongation. Computer simulations can be an ideal complementary tool to capture the dynamics of the backtracking process in the RNA polymerase system.

8. The Complete Picture: RNA Polymerase as a Molecular Motor

RNA polymerase moves processively along DNA during the transcription elongation process, driven by the free energy liberated after the NTP incorporation. Due to this chemical-to-mechanical conversion, the RNA polymerase can be considered as a molecular motor, such like kinesins, dyneins or myosins.[60] Many single molecule experiments have been performed to identify the coupling of the hydrolysis and the movement of the RNA polymerase, as well as to explain the possible limiting steps.[15,61] These experiments have shown that RNA polymerase stall force is at least twice the measured for cytoskeletal motors, and it also exceeds the calculated force for DNA supercoiling. Moreover, RNA polymerase is more complex than other molecular motors, due to the reactions that it performs, and also because it is the center of regulation of a basic process like transcription.[62]

So far we have reviewed the studies on individual steps involved in the transcription elongation process for the RNA polymerase. However, if we really want to understand how RNA polymerase works we need to look at all the steps as a whole picture. An example of such efforts is the study of T7 RNA polymerase, a single subunit enzyme. The authors built a kinetic model of the NAC, they constructed their pathway scheme using high-resolution structural information and they calibrated the kinetics of the transitions with single molecule force measurements.[49] Although the paper provides a good approximation for understanding T7 RNA polymerase, many of the kinetic parameters used were obtained empirically. A more reliable model can be obtained if we can include more reliable parameters, many of which could come from computational studies.

In conclusion, this review point out what we have learnt about RNA polymerase, as well as the challenges we still confront and their potential solutions. We have also provided an overview of the computational tools that could be used to study other big and complex systems, and we have shown that now we can study systems that before were unreachable.

Acknowledgment

X.H. acknowledges National Basic Research Program of China (973 program, 2013CB834703), National Science Foundation of China (No. 21273188), and Hong Kong Research Grants Council (661011, SBI12SC01, HKUST6/CRF/10, and T13-607/12R). F.P. acknowledges support from Hong Kong PhD Fellowship Scheme (2011/2012) and CONACYT Fellowship 215482.

References

1. Cramer P, Bushnell Da, Kornberg RD, Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution, *Science* **292**:1863−1876, 2001.
2. Kornberg RD, The molecular basis of eukaryotic transcription, *Proc Natl Acad Sci USA* **104**:12955−12961, 2007.
3. Westover KD, Bushnell DA, Kornberg RD, Structural basis of transcription: Nucleotide selection by rotation in the RNA polymerase II active center, *Cell* **119**:481−489, 2004.
4. Wang D, Bushnell DA, Westover KD, Kaplan CD, Kornberg RD, Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis, *Cell* **127**:941−954, 2006.
5. Brueckner F, Cramer P, Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation, *Nature Struct Mol Biol* **15**:811−818, 2008.
6. Brueckner F, Ortiz J, Cramer P, A movie of the RNA polymerase nucleotide addition cycle, *Curr Opin Struct Biol* **19**:294−299, 2009.
7. Bushnell DA, Cramer P, Kornberg RD, Structural basis of transcription: Alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution, *Proc Natl Acad Sci USA* **99**:1218−1222, 2002.
8. Cramer P, Architecture of RNA polymerase II and implications for the transcription mechanism, *Science* **288**:640−649, 2000.
9. Liu X, Bushnell DA, Silva D-A, Huang X, Kornberg RD, Initiation complex structure and promoter proofreading, *Science* **333**:633−637, 2011.
10. Wang D *et al.*, Structural basis of transcription: Backtracked RNA polymerase II at 3.4 angstrom resolution, *Science* **324**:1203−1206, 2009.
11. Kaplan CD, Jin H, Zhang IL, Belyanin A, Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo, *PLoS Genetics* **8**:e1002627, 2012.
12. Huang X *et al.*, RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription, *Proc Natl Acad Sci USA* **107**:15745−15750, 2010.
13. Nedialkov YA *et al.*, The RNA polymerase bridge helix YFI motif in catalysis, fidelity and translocation, *Biochim Biophys Acta* **1829**:187−198, 2012.
14. Malinen AM *et al.*, Active site opening and closure control translocation of multisubunit RNA polymerase, *Nucleic Acids Res* **40**:7442−7451, 2012.

15. Galburt EA, Grill SW, Bustamante C, Single molecule transcription elongation, *Methods* **48**:323−332, 2009.
16. Bai L, Fulbright R, Wang M, Mechanochemical kinetics of transcription elongation, *Phys Rev Lett* **98**:1−4, 2007.
17. Larson MH *et al.*, Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II, *Proc Natl Acad Sci USA* **109**:6555−6560, 2012.
18. Kireeva ML *et al.*, Molecular dynamics and mutational analysis of the catalytic and translocation cycle of RNA polymerase, *BMC Biophys* **5**:11, 2012.
19. Feig M, Burton ZF, RNA polymerase II flexibility during translocation from normal mode analysis, *Proteins* **78**:434−446, 2010.
20. Feig M, Burton ZF, RNA polymerase II with open and closed trigger loops: Active site dynamics and nucleic acid translocation, *Biophys J* **99**:2577−2586, 2010.
21. Shaw D, Dror R, Salmon J, Millisecond-scale molecular dynamics simulations on Anton, *Proc Conf High Performance Computing, Networking, Storage and Analysis (SC09)*, pp. 1−11, 2009, available at <dl.acm.org/citation.cfm?id=1654099> (accessed 13 July 2013).
22. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH, The MARTINI force field: Coarse grained model for biomolecular simulations, *The J Phys Chem B* **111**:7812−7824, 2007.
23. Klein ML, Shinoda W, Large-scale molecular dynamics simulations of self-assembling systems, *Science* **321**:798−800, 2008.
24. Shinoda W, DeVane R, Klein ML, Multi-property fitting and parameterization of a coarse grained model for aqueous surfactants, *Mol Simul* **33**:27−36, 2007.
25. Wassenaar Ta, Ingólfsson HI, Priess M, Marrink SJ, Schäfer LV, Mixing MARTINI: Electrostatic coupling in hybrid atomistic-coarse-grained biomolecular simulations, *The J Phys Chem B* **117**:3516−3530, 2013.
26. Bowman GR, Beauchamp KA, Boxer G, Pande VS, Progress and challenges in the automated construction of Markov state models for full protein systems, *J Chem Phys* **131**:124101, 2009.
27. Bowman GR, Voelz VA, Pande VS, Atomistic folding simulations of the five-helix bundle protein λ(6−85), *J Am Chem Soc* **133**:664−667, 2011.
28. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC, Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics, *J Chem Phys* **126**:155101, 2007.
29. Morcos F *et al.*, Modeling conformational ensembles of slow functional motions in Pin1-WW, *PLoS Comput Biol* **6**:e1001015, 2010.
30. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR, Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations, *Proc Natl Acad Sci USA* **106**:19011−19016, 2009.
31. Silva D-A, Bowman GR, Sosa-Peinado A, Huang X, A role for both conformational selection and induced fit in ligand binding by the LAO protein, *PLoS Comput Biol* **7**:e1002054, 2011.
32. Voelz VA, Bowman GR, Beauchamp K, Pande VS, Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39), *J Am Chem Soc* **132**:1526−1528, 2010.
33. Das M, Electron transfer through non-hydrogen and hydrogen bonded intermolecular tunnel junctions: A computational study, *J Theoret Comput Chem* **11**:997−1004, 2012.
34. Liu J, Tsai M-D, DNA polymerase β: Pre-steady-state kinetic analyses of dATPαS stereoselectivity and alteration of the stereoselectivity by various metal ions and by site-directed mutagenesis, *Biochemistry* **40**:9014−9022, 2001.

35. Castro C *et al.*, Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases, *Proc Natl Acad Sci USA* **104**:4267−4272, 2007.
36. Carvalho A, The catalytic mechanism of RNA polymerase II, *J Chem Theor Comput* **7**:1177−1188, 2011.
37. Vassylyev DG *et al.*, Structural basis for substrate loading in bacterial RNA polymerase, *Nature* **448**:163−168, 2007.
38. Nedialkov YA *et al.*, NTP-driven translocation by human RNA polymerase II, *J Biol Chem* **278**:18303−18312, 2003.
39. Batada NN, Westover KD, Bushnell DA, Levitt M, Kornberg RD, Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center, *Proc Natl Acad Sci USA* **101**:17361−17364, 2004.
40. Zhang J, Palangat M, Landick R, Role of the RNA polymerase trigger loop in catalysis and pausing, *Nature Struct Mol Biol* **17**:99−104, 2010.
41. QIN P-H *et al.*, Interactions of the nucleotides with the metal ions Mg 2+, Ca 2+, Mn 2+, Na +, and K +, *J Theoret Comput Chem* **11**:1183−1199, 2012.
42. Tehrani ZA, Javan MJ, Fattahi A, Hashemi MM, Effect of cation radical formation on reactivity and acidity enhancement of cytosine nucleobase: Natural bond orbital and atom in molecule analysis, *J Theoret Comput Chem* **11**:313−327, 2012.
43. Da L, Wang D, Huang X, Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II, *J Am Chem Soc* **134**:2399−2406, 2012.
44. Yuzenkova Y *et al.*, Stepwise mechanism for transcription fidelity, *BMC Biol* **8**:54, 2010.
45. Yin YW, Steitz TA, The structural mechanism of translocation and helicase activity in T7 RNA polymerase, *Cell* **116**:393−404, 2004.
46. Da L-T, Pardo-Avila F, Wang D, Huang X, A Two-state model for the dynamics of the pyrophosphate ion release in bacterial RNA polymerase, *PLOS Comput Biolo* **9**: e1003020, 2013.
47. Bar-Nahum G *et al.*, A ratchet mechanism of transcription elongation and its control, *Cell* **120**:183−193, 2005.
48. Guo Q, Sousa R, Translocation by T7 RNA polymerase: A sensitively poised Brownian ratchet, *J Mol Biol* **358**:241−254, 2006.
49. Yu J, Oster G, A small post-translocation energy bias aids nucleotide selection in T7 RNA polymerase transcription, *Biophys J* **102**:532−541, 2012.
50. Kwak H, Fuda NJ, Core LJ, Lis JT, Precise maps of RNA polymerase reveal how promoters direct initiation and pausing, *Science* **339**:950−953, 2013.
51. Landick R, Transcriptional pausing without backtracking, *Proc Natl Acad Sci USA* **106**:8797−8798, 2009.
52. Kireeva ML, Kashlev M, Mechanism of sequence-specific pausing of bacterial RNA polymerase, *Proc Natl Acad Sci USA* **106**:8900−8905, 2009.
53. Weixlbaumer A, Leon K, Landick R, Darst SA, Structural basis of transcriptional pausing in bacteria, *Cell* **152**:431−441, 2013.
54. Cheung ACM, Cramer P, Structural basis of RNA polymerase II backtracking, arrest and reactivation, *Nature* **471**:249−253, 2011.
55. Maoiléidigh DO, Tadigotla VR, Nudler E, Ruckenstein AE, A unified model of transcription elongation: What have we learned from single-molecule experiments? *Biophys J* **100**:1157−1166, 2011.
56. Nedialkov YA, Nudler E, Burton ZF, RNA polymerase stalls in a post-translocated register and can hyper-translocate, *Transcription* **3**:260−269, 2012.

57. Chakraborty A *et al.*, Opening and closing of the bacterial RNA polymerase clamp, *Science* **337**:591−595, 2012.
58. Kettenberger H, Armache K-J, Cramer P, Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage, *Cell* **114**:347−357, 2003.
59. Kettenberger H, Armache K, Cramer P, Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS, *Molecular Cell* **16**:955−965, 2004.
60. Chowdhury D, Stochastic mechano-chemical kinetics of molecular motors: A multidisciplinary enterprise from a physicist's perspective, pp. 1−370, arXiv:1207.6070 [physics.bio-ph], 2012.
61. Abbondanzieri EA, Greenleaf WJ, Shaevitz JW, Landick R, Block SM, Direct observation of base-pair stepping by RNA polymerase, *Nature* **438**:460−465, 2005.
62. Gelles J, Landick R, RNA polymerase as a molecular motor, *Cell* **93**:13−16, 1998.