Nucleotide addition and cleavage by RNA polymerase II: Coordination of two catalytic reactions using a single active site

Received for publication, July 3, 2022, and in revised form, December 19, 2022 Published, Papers in Press, December 26, 2022,
https://doi.org/10.1016/j.jbc.2022.102844

Ilona Christy Unarta[1], Eshani C. Goonetilleke[1], Dong Wang[2,3,4,*], and Xuhui Huang[1,*]

From the [1]Department of Chemistry, Theoretical Chemistry Institute, University of Wisconsin-Madison, Madison, Wisconsin, USA; [2]Skaggs School of Pharmacy and Pharmaceutical Sciences, [3]Department of Cellular and Molecular Medicine, School of Medicine, and [4]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California, USA

Edited by Patrick Sung

RNA polymerase II (Pol II) incorporates complementary ribonucleotides into the growing RNA chain one at a time *via* the nucleotide addition cycle. The nucleotide addition cycle, however, is prone to misincorporation of noncomplementary nucleotides. Thus, to ensure transcriptional fidelity, Pol II backtracks and then cleaves the misincorporated nucleotides. These two reverse reactions, nucleotide addition and cleavage, are catalyzed in the same active site of Pol II, which is different from DNA polymerases or other endonucleases. Recently, substantial progress has been made to understand how Pol II effectively performs its dual role in the same active site. Our review highlights these recent studies and provides an overall model of the catalytic mechanisms of Pol II. In particular, RNA extension follows the two-metal-ion mechanism, and several Pol II residues play important roles to facilitate the catalysis. In sharp contrast, the cleavage reaction is independent of any Pol II residues. Interestingly, Pol II relies on its residues to recognize the misincorporated nucleotides during the backtracking process, prior to cleavage. In this way, Pol II efficiently compartmentalizes its two distinct catalytic functions using the same active site. Lastly, we also discuss a new perspective on the potential third Mg^{2+} in the nucleotide addition and intrinsic cleavage reactions.

RNA polymerase II (Pol II) is a multisubunit enzyme that synthesizes messenger RNA and regulatory RNAs from a DNA template in eukaryotes. Pol II consists of key motifs that are crucial for the nucleotide addition cycle (NAC), including the trigger loop (TL) and bridge helix (BH) (Fig. 1*A*) (1, 2). Each NAC of Pol II consists of multiple steps (Fig. 1*B*) that involves several conformational changes of TL and BH. Starting from the pretranslocation state, Pol II translocates one base pair forward along the template DNA strand to reach the post-translocation state. Then, nucleotide triphosphate (NTP), alongside an Mg^{2+} (Mg_B) ion, binds to the empty active site. This step is followed by the folding of the TL to stabilize the substrate in the active site. Cognate NTP is then incorporated into the nascent RNA and is facilitated by 2 Mg ions (Mg_A and Mg_B), generating pyrophosphate (PPi) as a by-product. Next, PPi is released alongside Mg_B from the Pol II active site, and the TL unfolds such that the cycle returns to the pre-translocation state. The NAC repeats until Pol II encounters the termination signal. During the elongation process, mismatched or noncognate NTP may be occasionally incorporated into the nascent RNA. If noncognate NTP is incorporated, the forward translocation is compromised, and instead, Pol II backtracks one base pair along the template DNA strand and a short 3′-terminal RNA transcript containing the mis-incorporated nucleotide is removed either by intrinsic cleavage or transcription elongation factor IIS (TFIIS)-assisted cleavage.

Both the nucleotide addition and cleavage reactions have been suggested to follow the universal two-metal-ion reaction mechanism. The two-metal-ion catalytic mechanism was first proposed by T. Steitz and J. Steitz for the 3′-5′ exonuclease activity of bacterial DNA polymerase (3–5). Based on the crystal structure of the cleaving DNA polymerase, two divalent ions ∼3.8 Å apart were observed in the active site, forming a complex with the scissile phosphate. The scissile phosphate bond is positioned in between the two divalent ions. Metal ion A was suggested to help form the nucleophile OH^- to attack the phosphor atom of the scissile phosphate group. Upon the first nucleophilic attack, both metal ions were proposed to stabilize the pentacovalent transition state by coordination with the scissile phosphate oxygen atoms. Then, metal ion B was suggested to help the leaving 3′O^- ion.

Pol II utilizes the same active site to perform both nucleotide addition and intrinsic cleavage. Similar to the exonuclease activity of DNA polymerase (6), two Mg^{2+} ions are required for the chemical reactions, which also form a complex with the scissile phosphate, or the phosphate to be added in intrinsic cleavage or nucleotide addition, respectively. However, the general acid and base that activate the attacking nucleophile and protonate the reaction product are enzyme specific (7).

Alongside X-ray crystallography and biochemical experiments, quantum mechanics (QM) calculation is a powerful

* For correspondence: Dong Wang, dongwang@ucsd.edu; Xuhui Huang, xhuang@chem.wisc.edu.

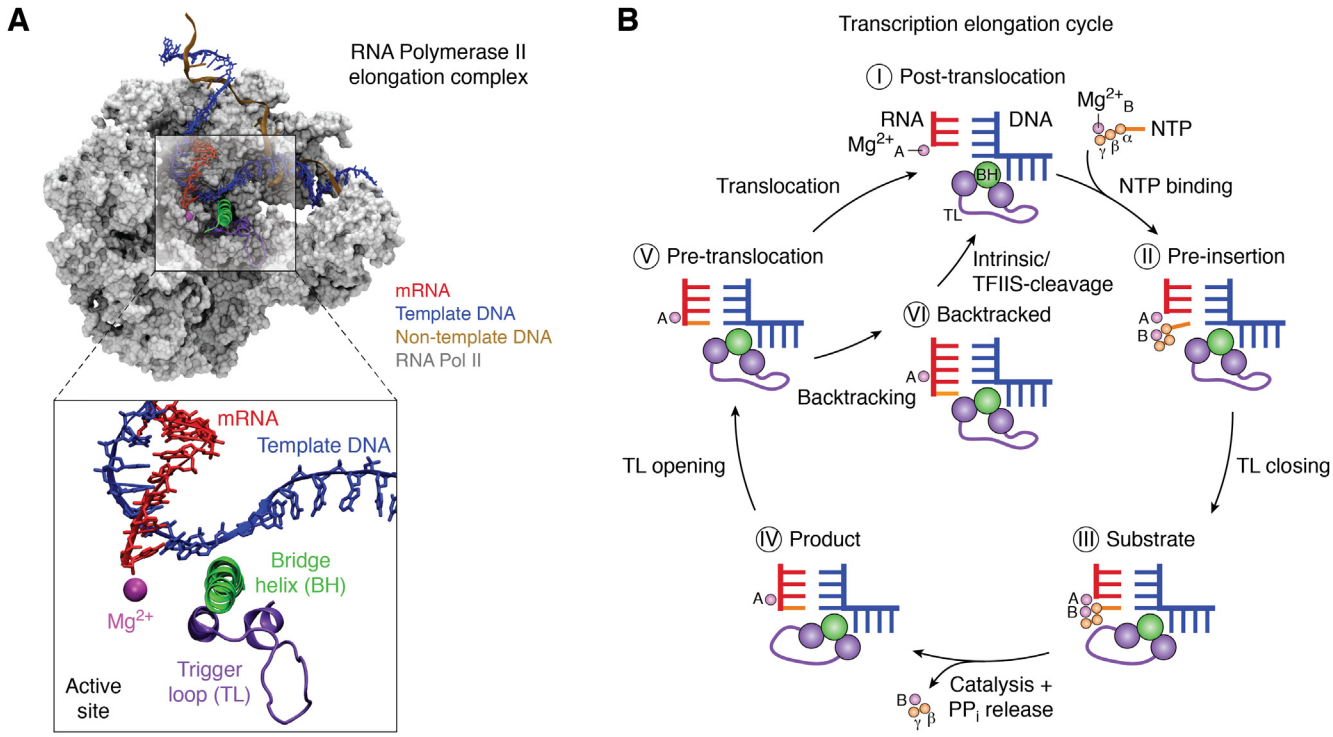

Figure 1. RNA Pol II elongation complex and nucleotide addition cycle. *A*, overview of the Pol II elongation complex. The trigger loop (*purple*) and bridge helix (*green*) are the key motifs involved in the nucleotide addition cycle. Pol II, RNA, and the template DNA are shown in *gray*, *red*, and *blue*, respectively. Mg^{2+} is shown as a *magenta sphere*. *B*, the nucleotide addition cycle and proofreading scheme of Pol II. Panel A and B are reproduced with permission from ref. 2. Copyright 2018 Elsevier. NTP, nucleotide triphosphate; Pol II, RNA polymerase II.

tool that can elucidate the mechanisms of different chemical reactions. QM calculations consider electronic degrees of freedom and can be used to examine the catalysis and cleavage steps of Pol II. However, QM calculations are too computationally demanding when simulating the entire Pol II (around 50,000 atoms). Thus, the QM/molecular mechanics (MM) method, a hybrid of QM and MM, is often employed to study the chemical reactions of large biological complexes. In the QM/MM method, a small part of the system (*e.g.*, the active site) is simulated with QM (8), while the rest of the system is modeled with classical molecular mechanics methods. QM/MM simulations can elucidate the atomistic mechanism and the energy barrier of chemical reactions.

In this article, we review the detailed mechanisms of the nucleotide addition and cleavage reactions that occur at the same active site in Pol II. We particularly highlight the recent works that examine the roles of Pol II residues in both reaction mechanisms. First, we review several computational and experimental studies to elucidate the catalytic mechanisms of the nucleotide addition reaction of Pol II and the mechanism of PPi release. Next, we review the studies that elucidate the molecular dynamics (MD) of Pol II's backtracking upon misincorporation and the catalytic mechanisms of the intrinsic and TFIIS-aided cleavage of Pol II. We also discuss the potential role of the third Mg^{2+} ion during the nucleotide addition and intrinsic cleavage of Pol II. Finally, we discuss the dual role of Pol II in catalyzing two distinct chemical reactions (*i.e.*, nucleotide addition and cleavage) using the same active site.

Catalytic mechanisms of nucleotide addition by pol II

Upon binding of NTP to the active site, TL of Pol II needs to be folded to close the active site and stabilize the bound NTP to proceed to the catalysis step. Specifically, residue L1081 forms hydrophobic interactions with the base of NTP, and H1085 interacts with the phosphate of NTP (9, 10). These interactions are important to position the NTP in the correct geometry for catalysis. To catalyze the reaction, both Mg^{2+} ions coordinate with the phosphate oxygen of NTP, aspartic acid residues of Pol II, and the 3′OH of the terminal RNA (11). It is suggested that Mg_A ion coordinate with the 3′OH of RNA and an oxygen atom of P_α on the NTP to keep the distance between the attacking $3'O^-$ ion and P_α within the range for catalysis (12).

The folding of TL and correct geometry of active site is selective for cognate NTP which ensure high fidelity of the RNA elongation process. MD simulations show that only cognate NTP can stabilize the interactions between TL residues and NTP as well as the correct geometry in the active site (12). On the other hand, when noncognate NTP is in the active site, the interactions are disrupted which may lead to opening of unfolding of TL or expulsion of NTP. The distance between 3′OH of RNA and the P_α increases beyond the range for catalysis when noncognate NTP is in the active site. This observation is in line with single-molecule FRET experiment of another multisubunit RNA polymerase (RNAP), *Escherichia coli* RNAP, that monitors the folding of TL in the presence of either cognate or noncognate NTPs (13). When cognate NTP is added to the elongation complex, TL generally adopts a

folded conformation. Conversely, when only noncognate NTP is added, TL remains in the unfolded conformation.

For the nucleotide addition reaction by Pol II to occur, the 3′ OH of the terminal RNA (OH_{RNA}) must be deprotonated to perform a nucleophilic attack toward the P_α atom on the NTP, resulting in the protonation of PPi. The nucleotide addition reaction mechanism was studied using QM calculations based on density functional theory (14). In this study, the free energy barrier of four proposed nucleotide addition reaction mechanisms was calculated. In the first pathway (HYP1), the phosphate oxygen α of the NTP acts as a general base to deprotonate OH_{RNA}, which is followed by the nucleophilic attack of the P_α and protonation of the PPi by the protonated phosphate oxygen α. The other three pathways differ in the general base that deprotonates the OH_{RNA}, while the general acid that protonates the leaving PPi remains the same, *i.e.*, the protonated H1085 (HIP) of the TL. Histidine is absolutely conserved at the H1085 position (in Pol II) in all cellular RNAPs, which suggests its importance to the chemical reactions in RNAPs. In addition, when TL is folded, H1085 is within the spatial range to act as general acid in the reaction. The general bases for the three pathways are (i) an OH^- ion (HYP2), (ii) an Mg_A-coordinated OH^- (HYP3), and (iii) D485 (HYP4). Based on the density functional theory calculations, HYP2 has the lowest energy barrier of the rate limiting step compared to the other proposed mechanisms (Fig. 2*A*). The low energy barrier of HYP2 is mostly contributed by the deprotonation of OH_{RNA} by external OH^- prior to the nucleophilic attack. This result also suggests that the HIP plays an important role as general acid to protonate the leaving PPi in the nucleotide addition reaction. However, the protonation of the PPi *via* H1085 is not the rate limiting step for any of the proposed mechanisms. Thus, HIP is not exclusively relevant to HYP2 and can act as general acid regardless of the reaction mechanism.

An MD simulation study has also shown that when the TL is folded, L1081 and H1085 interact with the NTP substrate (9). In addition, only the HIP can stabilize the NTP in the active site, which is shown by the small distance between HIP and NTP (Fig. 2, *B* and *C*). This further suggests that H1085 plays a functional role in facilitating nucleotide addition, either by stabilizing the NTP in the catalytically active conformation or directly serving as the general acid during catalysis.

The importance of H1085 in PoI II catalysis is well-established in experimental work. Previous structural studies from the Kornberg lab revealed direct interactions between H1085 and the NTP substrate when the TL is in a closed conformation (10). Furthermore, the Kaplan lab found that several H1085 variants (A/N/D/F) were lethal, mutations to K/R/W/Y caused severe growth defects, and mutations to Q caused a slight growth defect (7, 15–17). Strikingly, H1085Y, a mutant that closely mimics the histidine residue, reduces the nucleotide addition rate by an order of magnitude (7). Similarly, the Thomm lab found that when the homologous histidine (H87) in the TL of *Pyrococcus furiosus* RNAP is mutated into alanine, the nucleotide addition rate reduces by an order of magnitude, and there is a 2-fold decrease in substrate

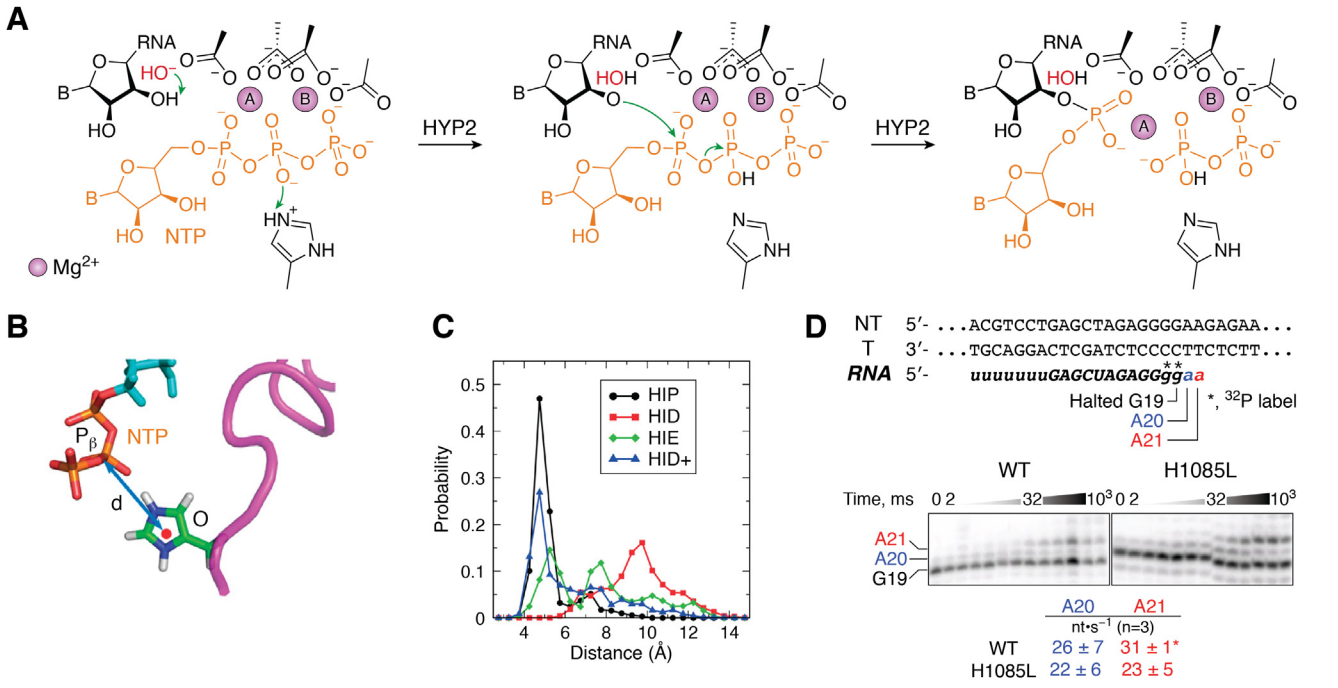

Figure 2. The role of conserved residue, H1085, during the nucleotide addition reaction in Pol II is proposed via experimental and computational methods. *A*, the proposed reaction mechanism of nucleotide addition which has the lowest energy barrier based on quantum mechanics calculations. An external hydroxide ion acts as the nucleophile that deprotonates the 3′ OH of RNA, which is followed by the protonation of the leaving pyrophosphate group. *B*, the distance between H1085 and P_β of NTP was calculated. *C*, Only protonated H1085 (HIP) can stabilize the NTP in the active site, as shown by the small distance between HIP and P_β of NTP. HID and HIE are H1085 with hydrogen on the delta and epsilon nitrogen of the imidazole ring, respectively. HID+ is HID with a +1 net charge. *D*, the rate of the single nucleotide addition assay was measured for both the WT and H1085L Pol II. The mutant H1085L does not significantly affect the nucleotide addition rate. Panel A is reproduced with permission from ref 14. Copyright 2011 American Chemical Society. Panel B and C are reproduced with permission from ref 9. National Academy of Sciences, 2010. Panel D is reproduced with permission from ref 19. Copyright 2021 American Chemical Society. NTP, nucleotide triphosphate; Pol II, RNA polymerase II.

binding (18). On the other hand, among many of the H1085 mutants reported, the H1085L mutant is found to be surprisingly healthy (17). Recently, Landick *et al.* found that this mutant (H1085L) only slightly reduces the single nucleotide addition rate (Fig. 2*D*) (19). It is not fully understood why H1085L only has a modest effect, whereas the other mutants have much more severe defects. It is also unclear why H1085F is lethal, whereas H1085Y is viable. We anticipate that multiple factors may contribute to the complicated TL mutant patterns: (a) As pointed out by Landick *et al.* (19), H1085 may have a structural role to facilitate the formation of the catalytically active conformation and stabilize the cognate substrate at a position poised for reaction. Certain mutants (such as H1085A) may compromise this role; (b) The proton transfer to PPi may require the involvement of the side chain of a TL residue and/or solvent water. While the polar side chains of WT H1085 (imidazole side chain) and viable polar mutants (such as the 4-hydroxyl group of H1085Y and the amide group of H1085Q) may be involved in the proton transfer to PPi, the proton transfer step can also be mediated by solvent water in the hydrophobic H1085L mutant (viable). In contrast, for other nonviable hydrophobic mutants, such as H1085F, the solvent water may be completely excluded by the aromatic bulky side chains and thus compromise the proton transfer step and subsequent PPi release (these findings may explain why H1085Y and H1085L are viable, whereas H1085F is not); (c). It is also possible that the H1085 mutants have distinct effects on the dynamics of TL folding and therefore change its contribution in Pol II catalysis (residue of 1085) or the rate-limiting steps. Nevertheless, further studies are needed to explore this direction.

Furthermore, time-resolved X-ray crystallography has shown the involvement of a third divalent metal in the nucleotide addition reaction of DNA polymerase (20). Specifically, upon the formation of the DNA product and breakage of PPi from the P_α, an Mg^{2+} was shown to bind to newly formed DNA product and the leaving the PPi ion. The third Mg^{2+} was speculated to coordinate with the transition state, between the O_α and O_β, after the nucleophilic attack to the P_α, which will help to overcome the high energy barrier of the product formation (20). A computational study using the QM/MM free energy calculations have shown that the presence of third Mg^{2+} during the catalysis stabilizes and lowering the energy of the product state *via* electrostatic interaction with the leaving PPi (21). Yet, it is still unclear whether the third metal ion is involved in the catalysis reaction or simply stabilizes the product to help the exit of PPi. Considering that the conformation of the active site of DNA polymerase and RNAP are highly conserved, further studies to confirm the involvement of the third metal ion to the nucleotide addition reaction of RNAPs would be worthwhile.

Upon addition of NTP into the nascent RNA, PPi is generated as a by-product. PPi alongside MgB must leave the active site to provide enough space for the next incoming NTP. To study the PPi release pathway, Huang *et al.* employ Markov State Models (MSMs) constructed based on MD simulations (22). MD simulations are a computational method that can estimate the movement of biological molecules in atomic resolution. Although powerful, it is challenging to simulate large biological complexes, such as Pol II, at biologically relevant timescales (μs to ms) due to technological limitations and unfeasible computational times. MSMs can predict the slow timescale dynamics based on short MD simulations, which can be performed in parallel (23–47). MSMs model the dynamics processes estimated from Markovian or Memoryless transitions between metastable states, meaning that the transition to the next state only depends on the current state and is independent of the previous transitions. Importantly, MSMs built from MD simulations can elucidate atomistic mechanisms and the dynamics of functional conformational changes (36, 48–53), such as the release of PPi (22, 54).

MSMs were constructed to study the dynamic process of the $(Ppi\text{-}Mg)^{2-}$ ion leaving the active site to enter the bulk solvent through the secondary channel of Pol II. MSMs identify four metastable states (State S1-S4, Fig. 3) during PPi release, and the transition from State S1 to S4 occurs at ∼1.5μs. $(Ppi\text{-}Mg)^{2-}$ was shown to "hop" between several positively charged residues along the secondary channel. In State S1, $(Ppi\text{-}Mg)^{2-}$ forms a salt bridge interaction with the HIP residue on the TL in the active site (Fig. 3*B*). This interaction is crucial for the release of $(Ppi\text{-}Mg)^{2-}$ from the active site as a simulation with the mutant, H1085F, shows that $(Ppi\text{-}Mg)^{2-}$ is stable in the active site and thus has less of a tendency to be released. This step is followed by a series of salt-bridge interactions between lysine residues in the secondary channel. Importantly, the interaction between $(Ppi\text{-}Mg)^{2-}$ and H1085 during PPi release may increase the flexibility of the TL, leading to the TL unfolding and preparing Pol II for the next NAC.

Molecular mechanisms of pol II's backtracking and cleavage

If a noncognate nucleotide is incorporated into the nascent RNA, Pol II can backtrack (*i.e.*, perform backward

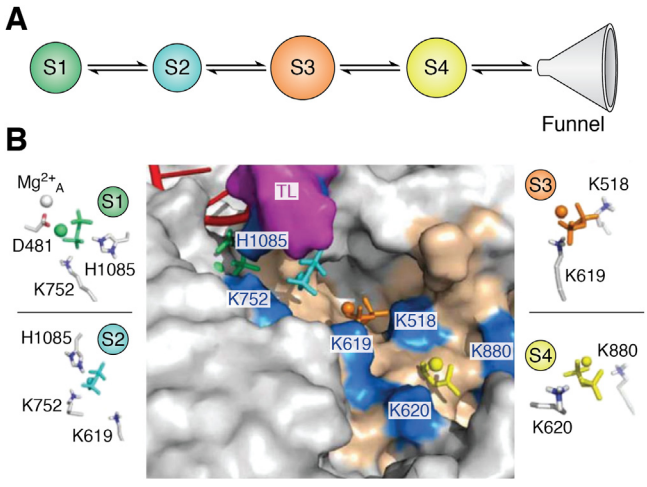

Figure 3. MSM elucidated the molecular mechanism of PPi release from the active site after the nucleotide addition reaction. *A*, MSMs identify four metastable states for the release of $(PPi\text{-}Mg)^{2-}$ from the active site into the bulk solvent (S1-S4). *B*, the representative structures of each metastable state. This figure is reproduced with permission from ref 22. Copyright 2012 American Chemical Society. MSMs, Markov State Models.

translocation on the RNA–DNA hybrid) and then cleave a short piece of the 3′-end of the RNA transcript which contains the misincorporated nucleotide either *via* intrinsic or TFIIS-aided cleavage (55). Upon backtracking, Pol II cleaves the phosphodiester bond between the -1 and +1 position and releases dinucleotides (or a longer piece of RNA) at the 3′ terminal of the nascent RNA. As a result, the Pol II complex is reactivated to generate a new posttranslocation state (step VI, Fig. 1*B*). It has been suggested that the cleavage also follows the two-metal-ions mechanism, requiring two Mg^{2+} (MgA and MgB) to be positioned in the active site to form a complex with the scissile phosphate of the backtracked RNA.

Using MSMs, four metastable states of the backtracking process were identified, showing the stepwise mechanism of the backtracking process (State S-1 to S-4, Fig. 4) (56). In this mechanism, the misincorporated RNA first frays from the active site (S-1 → S-2 → S-3, Fig. 4), after which the backtracking of the template and nontemplate DNA by one base-pair takes place (S-3→S-4, Fig. 4). During the transition from S-1 to S-2, the T831 residue on the BH acts as a sensing probe that examines the stability of the DNA–RNA pair in the active site. Misincorporated RNA forms less stable hydrogen bonds with the template DNA and thus tends to fray and transition to the frayed state (S-2 and S-3, Fig. 4). The role of T831 is validated by a mutagenesis experiment, where a significantly slower cleavage rate is observed for T831A compared to the wildtype Pol II.

Upon backtracking, Pol II removes the downstream terminal RNA nucleotides *via* either intrinsic cleavage or TFIIS-assisted cleavage. Similar to other two-metal-ion reaction mechanisms, the cleavage reaction also requires a general base to activate the attacking nucleophile and a general acid to protonate the by-product. Pol II residues (H1085) or the nucleobase of terminal RNA have been suggested to act as a general base. Yet, a recent biochemical study has shown that the mutation of H1085 to leucine, a nonpolar residue, does not affect the cleavage rate, which indicates that H1085 does not serve as a general base during cleavage (19).

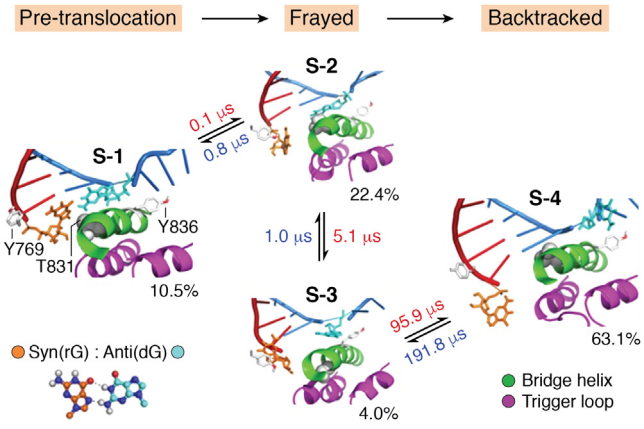

Figure 4. MSM identified four metastable states for the backtracking of misincorporated RNA by one base pair in Pol II (S-1 to S-4). The representative structure for each metastable state is shown. This figure is reproduced with permission from ref 56. Copyright 2016 Springer Nature. MSM, Markov State Model.

QM/MM-MD simulations coupled with biochemical experiments have also been performed to examine the reaction mechanism of intrinsic cleavage (57). Based on the QM/MM calculation, the phosphate oxygen of the backtracked RNA (Rp-O of G12, Fig. 5*A*) acts as a general base to activate the active site water molecule and generate a hydroxide (OH^-) ion which acts as the attacking nucleophile. The OH^- ion attacks the phosphate bond between A10 and U11 of the nascent RNA (Fig. 5*A*), which leads to the cleavage of the terminal dinucleotide of RNA (U11 and G12). The next step is the protonation of the newly formed terminal $O3^-$. QM/MM simulations show that a water molecule in the active site acts as a general acid to protonate the terminal $O3^-$, which is followed by a series of proton transfers to recover the water molecules in the active site.

The role of Rp-O as a general base is validated by two sets of experiments. The first experiment measures the cleavage rate of the elongation complexes with two kinds of RNA, 3′-5′ and 2′-5′ phosphodiester linkages between U11 and G12. The 2′-5′ linkage between the 0 and −1 sites causes the RNA at site 0 to be further away from the active site and discourages its participation in the cleavage reaction. The rate of the 2′-5′ linked RNA indeed was shown to be 14 times slower than the wildtype, 3′-5′ linked RNA. The second experiment measures the cleavage rate when G12 possesses a thiol-substituted phosphate oxygen. The substitution of the oxygen atom with sulfur creates two stereoisomers of the terminal RNA, R_p-S and S_p-S. It was hypothesized that R_p-S would reduce the tendency of the terminal RNA to attack a water molecule. R_p-S was indeed shown to have a 3-fold reduction in cleavage rate, while S_p-S only minimally affected the cleavage rate. This result further confirms the role of R_p-O as the general base in the intrinsic cleavage reaction. To test whether the G12 nucleobase can act as a general base, the cleavage rate of the elongation complex with an RNA scaffold containing a 7-deaza-guanine (substitution at N7 to C) at the G12 position was measured and shown to not affect the cleavage rate. This result indicates that the nucleobase of the backtracked RNA does not act as a general base. Lastly, it has been suggested that a Pol II residue, protonated D485, may act as a general acid during proton transfer. However, based on MD simulations, a protonated D485 cannot stabilize the active site. The role of H1085 residue to the cleavage reactions has also been examined *via* a recent biochemical experiment. The mutation of H1085 to leucine, which is nonpolar and cannot participate in proton transfer, does not affect the rate of the cleavage reaction (19). This result further supports the hypothesis that Pol II residues are not involved in the proton transfer of the cleavage reaction.

A third Mg^{2+} ion has been speculated to facilitate the cleavage of the RNA by Ribonuclease H (58). Another study has also observed that a third Mg^{2+} may lower the activation barrier in the nucleotide addition reaction of DNA Pol η (20, 59) as well as DNA pol β (60). Indeed, structural modeling shows that there is enough space for the third Mg^{2+} in the active site (Fig. 5, *B* and *C*). Experimental results have also shown that an increase in Mg^{2+} concentration increases the

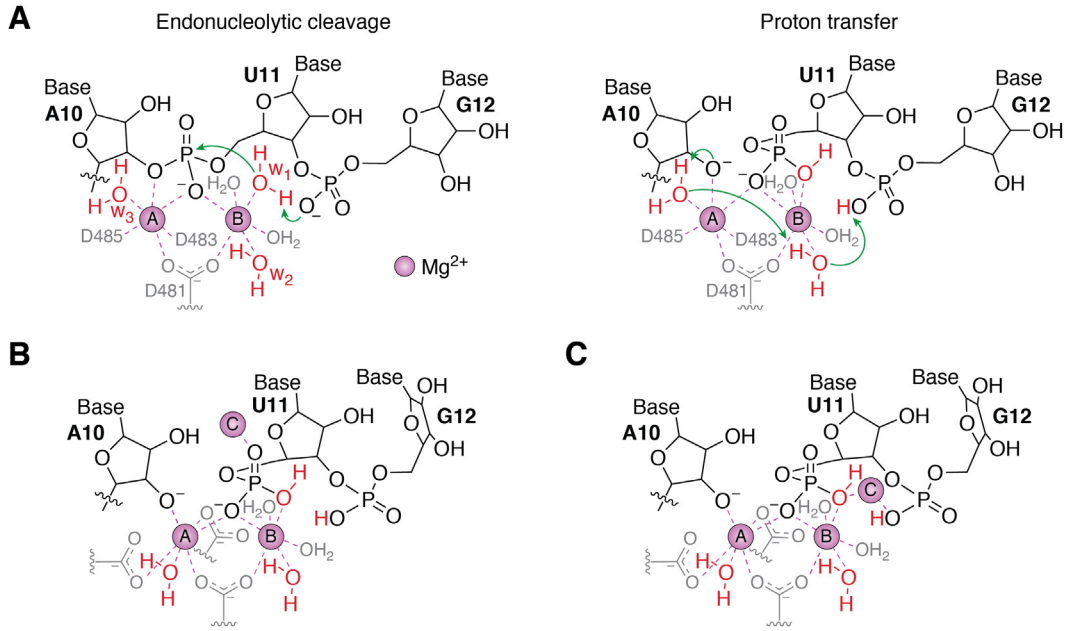

Figure 5. QM/MM-MD simulations showed that the intrinsic cleavage reaction does not involve any Pol II residues. *A*, the proposed reaction mechanism of intrinsic cleavage of Pol II. *B* and *C*, modeled Mg_C ion upon intrinsic cleavage in Pol II, which is positioned on the (*B*) cleaved phosphate and (*C*) near the general base Rp-O of G12. This figure is reproduced from ref 57. Copyright 2019 Springer Nature. Pol II, RNA polymerase II.

k_{cat} of intrinsic cleavage in N7-modified terminal guanine, supporting the participation of the third Mg^{2+} in cleavage. Furthermore, the introduction of the third Mg^{2+} may further facilitate the deprotonation of the attacking nucleophile by increasing the pK_a of the R_p-O of backtracked RNA (Fig. 5*C*). However, the introduction of the third Mg^{2+} only causes minimal rearrangement of the RNA active site and residues and is thus unlikely to alter the previously proposed mechanism.

For longer backtracked RNA, TFIIS is required to reactivate the arrested Pol II state. TFIIS has also been shown to accelerate the cleavage reaction by one order of magnitude compared with intrinsic cleavage (61). Based on the crystal structure of a Pol II elongation complex with TFIIS bound, TFIIS binds through the secondary channel (pore and funnel) of Pol II and extends to the active site of Pol II (62). TFIIS reactivates Pol II by displacing the backtracked RNA in the secondary channel and facilitates the cleavage of the RNA (63). It has been suggested that residues D290-E291 of TFIIS's domain III may coordinate with and stabilize Mg_B. In addition, residue R287 on TFIIS may further stabilize the transition state.

Conclusions

Unlike DNA polymerases or ribonuclease H, Pol II catalyzes both nucleotide addition and intrinsic cleavage (corresponding to the formation and breaking of the phosphodiester bond, respectively) in the same active site. In this article, we review recent progress in elucidating how Pol II performs this dual-function role *via* a single active site. We present an overall model of Pol II's catalytic mechanisms for both reactions (Fig. 6). On one hand, Pol II catalyzes the nucleotide addition following a two-metal-ion mechanism. Several TL residues play important roles in facilitating this reaction (Fig. 6*B*). For example, the HIP may help the protonation of the leaving PPi in the nucleotide addition reaction. In addition, L1081 aligns the incoming NTP base to stabilize the catalytically active conformation. On the other hand, the phosphate Rp-oxygen of RNA serves as the general base in the intrinsic cleavage reaction, and none of the Pol II residues is directly involved in the reaction. (Fig. 6*C*). To recognize misincorporations, Pol II utilizes the conformational change of backtracking, which only occurs when Pol II encounters mis-incorporated nucleotides. In this article, we review recent computational and experimental studies to elucidate catalytic mechanisms of the nucleotide addition reaction of Pol II, the mechanism of PPi release, the MD of Pol II's backtracking upon misincorporation, and the catalytic mechanisms of the intrinsic and TFIIS-aided cleavage of Pol II. As a new perspective, we discuss the potential involvement of a third metal ion to nucleotide addition and cleavage reactions. These studies provide mechanistic insights into how Pol II catalyzes two distinct chemical reactions (*i.e.*, nucleotide addition and cleavage) using a single active site and the universal two-metal-ion mechanism.

Funding and additional information—X. H. acknowledges the support from the start-up funds from University of Wisconsin-Madison and the Hirschfelder Professorship Fund. D. W. acknowledges NIH GM102362. I. C. U. acknowledges the support from Croucher Fellowship for Postdoctoral Researchers from Croucher Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions—I. C. U., D. W., and X. H. conceptualization; I. C. U., E. C. G. writing-original draft; I. C. U., E. C. G., D. W., and X. H. writing-reviewing and editing; X. H. supervision.

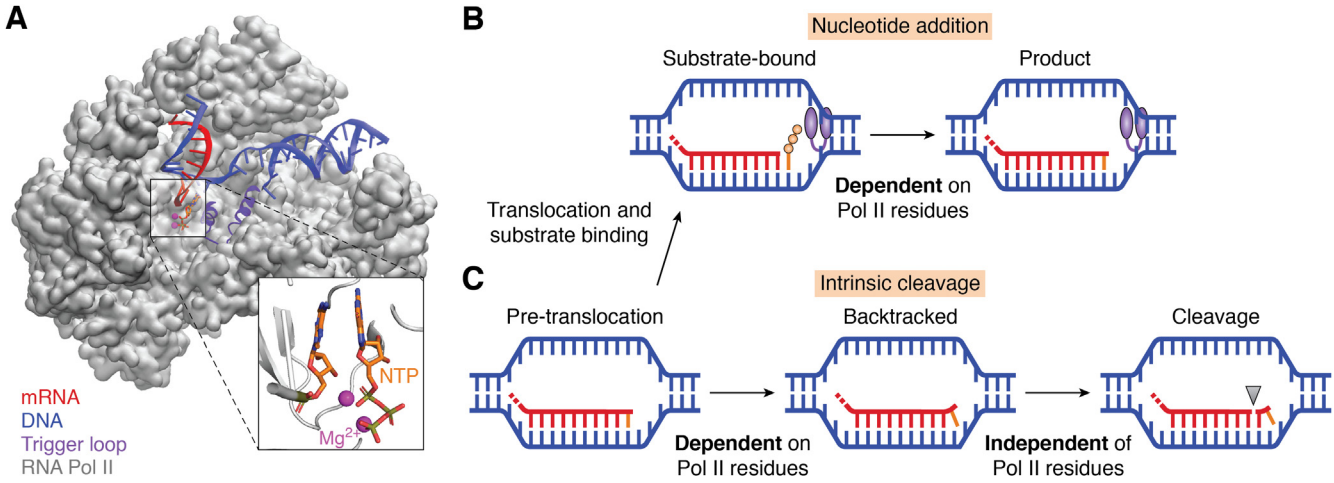

Figure 6. A cartoon scheme illustrating how Pol II catalyzes two opposite reactions using a single active site. *A*, the Pol II elongation complex (PDB ID: 2E2H). The colors represent the following, *gray*: Pol II, *red*: mRNA, *blue*: DNA, *purple*: trigger loop (TL), *orange*: NTP, *magenta*: Mg2+ ions. The inset figure displays the catalytically active conformation of Pol II's active site for the nucleotide addition reaction. *B*, the nucleotide addition reaction adds the substrate NTP to the growing RNA chain. This chemical reaction relies on several Pol II residues on the TL domain (*e.g.*, H1085 and L1081). *C*, if mismatched nucleotides are incorporated, Pol II probes the misincorporation *via* its bridge helix residue (T831) and enables backtracking. The backtracked nucleotides will be subsequently cleaved, and none of the Pol II residues directly involves in the cleavage reaction. NTP, nucleotide triphosphate; Pol II, RNA polymerase II.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: BH, bridge helix; HIP, protonated H1085; MD, molecular dynamics; MM, molecular mechanics; MSM, Markov State Model; NAC, nucleotide addition cycle; NTP, nucleotide triphosphate; OHRNA, 3′ OH of the terminal RNA; PPi, pyrophosphate; Pol II, RNA polymerase II; RNAP, RNA polymerase; TFIIS, transcription elongation factor IIS; TL, trigger loop; QM, quantum mechanics.

References

1. Brueckner, F., Ortiz, J., and Cramer, P. (2009) A movie of the RNA polymerase nucleotide addition cycle. *Curr. Opin. Struct. Biol.* **19**, 294–299
2. Unarta, I. C., Zhu, L., Tse, C. K. M., Cheung, P. P. H., Yu, J., and Huang, X. (2018) Molecular mechanisms of RNA polymerase II transcription elongation elucidated by kinetic network models. *Curr. Opin. Struct. Biol.* **49**, 54–62
3. Beese, L. S., and Steitz, T. A. (1991) Structural basis for the 3′-5′ exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J.* **10**, 25–33
4. Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., and Steitz, T. A. (1988) Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 8924–8928
5. Steitz, T. A., and Steitz, J. A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6498–6502
6. Nudler, E. (2009) RNA polymerase active center: the molecular engine of transcription. *Annu. Rev. Biochem.* **78**, 335–361
7. Kaplan, C. D., Larsson, K. M., and Kornberg, R. D. (2008) The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by α-amanitin. *Mol. Cell* **30**, 547–556
8. Zhou, Y., Wang, S., Li, Y., and Zhang, Y. (2016) Born–oppenheimer ab initio QM/MM molecular dynamics simulations of enzyme reactions. *Methods Enzymol.* **577**, 105–118
9. Huang, X., Wang, D., Weiss, D. R., Bushnell, D. A., Kornberg, R. D., and Levitt, M. (2010) RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15745–15750
10. Wang, D., Bushnell, D. A., Westover, K. D., Kaplan, C. D., and Kornberg, R. D. (2006) Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* **127**, 941–954
11. Svetlov, V., and Nudler, E. (2013) Basic mechanism of transcription by RNA polymerase II. *Biochim. Biophys. Acta* **1829**, 20–28
12. Wang, B., Opron, K., Burton, Z. F., Cukier, R. I., and Feig, M. (2015) Five checkpoints maintaining the fidelity of transcription by RNA polymerases in structural and energetic details. *Nucleic Acids Res.* **43**, 1133–1146
13. Mazumder, A., Lin, M., Kapanidis, A. N., and Ebright, R. H. (2020) Closing and opening of the RNA polymerase trigger loop. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 15642–15649
14. Carvalho, A. T. P., Fernandes, P. A., and Ramos, M. J. (2011) The catalytic mechanism of RNA polymerase II. *J. Chem. Theor. Comput.* **7**, 1177–1188
15. Kaplan, C. D., Jin, H., Zhang, I. L., and Belyanin, A. (2012) Dissection of Pol II Trigger Loop function and Pol II activity-dependent control of start site selection *in vivo*. *PLoS Genet.* **8**, e1002627
16. Braberg, H., Jin, H., Moehle, E. A., Chan, Y. A., Wang, S., Shales, M., *et al.* (2013) From structure to systems: High-resolution, quantitative genetic analysis of RNA polymerase II. *Cell* **154**, 775–788
17. Qiu, C., Erinne, O. C., Dave, J. M., Cui, P., Jin, H., Muthukrishnan, N., *et al.* (2016) High-Resolution phenotypic landscape of the RNA polymerase II trigger loop. *PLoS Genet.* **12**, e1006321
18. Fouqueau, T., Zeller, M. E., Cheung, A. C., Cramer, P., and Thomm, M. (2013) The RNA polymerase trigger loop functions in all three phases of the transcription cycle. *Nucleic Acids Res.* **41**, 7048–7059
19. Palo, M. Z., Zhu, J., Mishanina, T. V., and Landick, R. (2021) Conserved trigger loop histidine of RNA polymerase II functions as a positional catalyst primarily through steric effects. *Biochemistry* **60**, 3323–3336
20. Yang, W., Weng, P. J., and Gao, Y. (2016) A new paradigm of DNA synthesis: Three-metal-ion catalysis. *Cell Biosci.* **6**, 1–7
21. Stevens, D. R., and Hammes-Schiffer, S. (2018) Exploring the role of the third active site metal ion in DNA polymerase η with QM/MM free energy simulations. *J. Am. Chem. Soc.* **140**, 8965–8969
22. Da, L. T., Wang, D., and Huang, X. (2012) 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
23. Chodera, J. D., and Noé, F. (2014) Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144
24. Husic, B. E., and Pande, V. S. (2018) Markov state models: from an art to a science. *J. Am. Chem. Soc.* **140**, 2386–2396
25. Prinz, J. H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., *et al.* (2011) Markov models of molecular kinetics: generation and validation. *J. Chem. Phys.* **134**, 174105

26. Malmstrom, R. D., Lee, C. T., Van Wart, A. T., and Amaro, R. E. (2014) Application of molecular-dynamics based markov state models to functional proteins. *J. Chem. Theor. Comput.* **10**, 2648–2657
27. Bowman, G. R., Pande, V. S., and Noé, F. (2014). In: Bowman, G. R., Pande, V. S., Noé, F., eds. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*, Springer, Netherlands
28. Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A., and Swope, W. C. (2007) Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **126**, 1–17
29. Pan, A. C., and Roux, B. (2008) Building Markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.* **129**, 1–8
30. Morcos, F., Chatterjee, S., Mcclendon, C. L., Brenner, P. R., and Ló Pez-Rendó N, R. (2010) Modeling conformational ensembles of slow functional motions in pin1-WW. *PLoS Comput. Biol.* **6**, 1001015
31. Huang, X., Bowman, G. R., Bacallado, S., and Pande, V. S. (2009) Rapid equilibrium sampling initiated from nonequilibrium data. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19765–19769
32. Buchete, N. V., and Hummer, G. (2008) Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B* **112**, 6057–6069
33. Noé, F., Schütte, C., Vanden-Eijnden, E., Reich, L., and Weikl, T. R. (2009) Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci.* **106**, 19011–19016
34. Bowman, G. R., Voelz, V. A., and Pande, V. S. (2011) Taming the complexity of protein folding. *Curr. Opin. Struct. Biol.* **21**, 4–11
35. Buch, I., Giorgino, T., and De Fabritiis, G. (2011) Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10184–10189
36. Silva, D. A., Bowman, G. R., Sosa-Peinado, A., and Huang, X. (2011) A role for both conformational selection and induced fit in ligand binding by the Lao protein. *PLoS Comput. Biol.* **7**, e1002054
37. Noé, F., Horenko, I., Schütte, C., and Smith, J. C. (2007) Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states. *J. Chem. Phys.* **126**, 155102
38. Bowman, G. R., Ensign, D. L., and Pande, V. S. (2010) Enhanced modeling via network theory: adaptive sampling of markov state models. *J. Chem. Theor. Comput.* **6**, 787–794
39. Sarich, M., Noé, F., and Schütte, C. (2010) On the approximation quality of markov state models. *Multiscale Model. Simul.* **8**, 1154–1177
40. Noé, F., and Nüske, F. (2013) A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Simul.* **11**, 635–655
41. Wu, H., and Noé, F. (2020) Variational approach for learning markov processes from time series data. *J. Nonlinear Sci.* **30**, 23–66
42. Zhang, B. W., Dai, W., Gallicchio, E., He, P., Xia, J., Tan, Z., *et al.* (2016) Simulating replica exchange: markov state models, proposal schemes, and the infinite swapping limit. *J. Phys. Chem. B* **120**, 8289–8301
43. Weng, J., Yang, M., Wang, W., Xu, X., and Tian, Z. (2020) Revealing thermodynamics and kinetics of lipid self-assembly by markov state model analysis. *J. Am. Chem. Soc.* **142**, 21344–21352
44. Zeng, X., Zhu, L., Zheng, X., Cecchini, M., and Huang, X. (2018) Harnessing complexity in molecular self-assembly using computer simulations. *Phys. Chem. Chem. Phys.* **20**, 6767–6776
45. Cao, S., Montoya-Castillo, A., Wang, W., Markland, T. E., and Huang, X. (2020) On the advantages of exploiting memory in Markov state models for biomolecular dynamics. *J. Chem. Phys.* **153**, 1–13
46. Bowman, G. R., Huang, X., and Pande, V. S. (2010) Network models for molecular kinetics and their initial applications to human health. *Cell Res.* **20**, 622–630
47. Yao, Y., Cui, R. Z., Bowman, G. R., Silva, D. A., Sun, J., and Huang, X. (2013) Hierarchical Nyström methods for constructing Markov state models for conformational dynamics. *J. Chem. Phys.* **138**, 174106
48. Konovalov, K. A., Unarta, I. C., Cao, S., Goonetilleke, E. C., and Huang, X. (2021) Markov state models to study the functional dynamics of proteins in the wake of machine learning. *JACS Au* **1**, 1330–1341
49. Wang, W., Cao, S., Zhu, L., and Huang, X. (2018) Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **24**, 1462–1474
50. Wang, X., Unarta, I. C., Cheung, P. P. H., and Huang, X. (2021) Elucidating molecular mechanisms of functional conformational changes of proteins via Markov state models. *Curr. Opin. Struct. Biol.* **67**, 69–77
51. Silva, D.-A., Weiss, D. R., Pardo Avila, F., Da, L.-T., Levitt, M., Wang, D., *et al.* (2014) Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 7665–7670
52. Feng, J., Selvam, B., and Shukla, D. (2021) How do antiporters exchange substrates across the cell membrane? An atomic-level description of the complete exchange cycle in NarK. *Structure* **29**, 922–933.e3
53. Tian, J., Wang, L., and Da, L. T. (2021) Atomic resolution of short-range sliding dynamics of thymine DNA glycosylase along DNA minor-groove for lesion recognition. *Nucleic Acids Res.* **49**, 1278–1293
54. Da, L. T., Pardo Avila, F., Wang, D., and Huang, X. (2013) A two-state model for the dynamics of the pyrophosphate ion release in bacterial RNA polymerase. *PLoS Comput. Biol.* **9**, e1003020
55. Xu, L., Wang, W., Chong, J., Shin, J. H., Xu, J., and Wang, D. (2015) RNA polymerase II transcriptional fidelity control and its functional interplay with DNA modifications. *Crit. Rev. Biochem. Mol. Biol.* **50**, 503–519
56. Da, L.-T., Pardo-Avila, F., Xu, L., Silva, D.-A., Zhang, L., Gao, X., *et al.* (2016) Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nat. Commun.* **7**, 11244
57. Tse, C. K. M., Xu, J., Xu, L., Sheong, F. K., Wang, S., Chow, H. Y., *et al.* (2019) Intrinsic cleavage of RNA polymerase II adopts a nucleobase-independent mechanism assisted by transcript phosphate. *Nat. Catal.* **2**, 228–235
58. Samara, N. L., and Yang, W. (2018) Cation trafficking propels RNA hydrolysis. *Nat. Struct. Mol. Biol.* **25**, 715–721
59. Gao, Y., and Yang, W. (2016) Capture of a third Mg2+ is essential for catalyzing DNA synthesis. *Science* **352**, 1334–1337
60. Freudenthal, B. D., Beard, W. A., Shock, D. D., and Wilson, S. H. (2013) Observing a DNA polymerase choose right from wrong. *Cell* **154**, 157–168
61. Wang, D., Bushnell, D. A., Huang, X., Westover, K. D., Levitt, M., and Kornberg, R. D. (2009) Structural basis of transcription: backtracked RNA polymerase II at 3.4 Ångstrom resolution. *Science* **324**, 1203–1206
62. Kettenberger, H., Armache, K. J., and Cramer, P. (2003) Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell* **114**, 347–357
63. Cheung, A. C. M., and Cramer, P. (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* **471**, 249–253