Published in final edited form as:

Nat Energy. 2019 March ; 2(3): 228–235. doi:10.1038/s41929-019-0227-5.

Intrinsic Cleavage of RNA Polymerase II Adopts a Nucleobase-independent Mechanism Assisted by Transcript Phosphate

Carmen Ka Man Tse^{#1}, Jun Xu^{#2}, Liang Xu^{2,4}, Fu Kit Sheong¹, Shenglong Wang³, Hoi Yee Chow⁵, Xin Gao⁷, Xuechen Li⁵, Peter Pak-Hang Cheung¹, Dong Wang², Yingkai Zhang^{3,6}, and Xuhui Huang¹

¹Department of Chemistry, Centre of Systems Biology and Human Health, State Key Laboratory of Molecular Neuroscience, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

²Department of Cellular and Molecular Medicine, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093, United States

³Department of Chemistry, New York University, New York, New York 10003 United States

⁴Department of Chemistry, Sun Yat-Sen University, Guangzhou, China

⁵Department of Chemistry, State Key Lab of Synthetic Chemistry, The University of Hong Kong, Hong Kong

⁶NYU-ECNU Centre for Computational Chemistry at NYU Shanghai, Shanghai 200062, China

⁷Computational Bioscience Research Centre (CBRC), CEMSE Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

[#] These authors contributed equally to this work.

Abstract

RNA polymerase II (Pol II) utilises the same active site for polymerization and intrinsic cleavage. Pol II proofreads the nascent transcript by its intrinsic nuclease activity to maintain high transcriptional fidelity critical for cell growth and viability. The detailed catalytic mechanism of intrinsic cleavage remains unknown. Here, we combined *ab initio* quantum mechanics/molecular mechanics studies and biochemical cleavage assays to show that Pol II utilises downstream

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Author contributions

J.X. and X.L. prepared the proteins and performed the biochemical analyses. C.K.M.T., X.G., and S.W. performed aiQM/MM-MD simulations. H.Y.C. and X.L. performed Reverse phase-FPLC purification of thio-substituted oligonucleotides. C.K.M.T., J.X., P.P.-H.C., F.K.S., D.W., Y.Z., and X.H. analysed the data. C.K.M.T., J.X., P.P.-H.C., D.W., Y.Z., and X.H. wrote the manuscript with inputs from all authors. D.W., Y.Z. and X.H. directed and supervised the research.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

Custom computer code used to analyse simulation data is available from the corresponding author on request.

Competing interests

The authors declare no competing interests

phosphate oxygen to activate the attacking nucleophile in hydrolysis, while the newly formed 3'-end is protonated through active-site water without a defined general acid. Experimentally, alteration of downstream phosphate oxygen either by 2'-5' sugar linkage or stereo-specific thio-substitution of phosphate oxygen drastically reduced cleavage rate. We showed by N7-modification that guanine nucleobase does not directly involve as acid-base catalyst. Our proposed mechanism provides important insights into the understanding of intrinsic transcriptional cleavage reaction, an essential step of transcriptional fidelity control.

Introduction

Inaccurate gene expression would impose adverse effect on proteostasis and cellular lifespan¹. To maintain high transcriptional fidelity, RNA polymerases (RNAPs) are capable to proofread the nascent RNA transcript by its intrinsic nuclease activity, a hydrolytic cleavage reaction that can be further stimulated by other transcription factors such as GreA/GreB in bacteria^{2,3}, TFIIS in eukarya⁴ and TFS in archaea⁵. It has been shown in *Saccharomyces cerevisiae* that TFIIS-stimulated cleavage is not essential, as the factor-independent, intrinsic cleavage activity by RNA polymerase II (Pol II) is sufficient for cell viability⁶.

RNAP utilises the same active site for both nucleotide polymerization and intrinsic cleavage reaction. Upon single nucleotide misincorporation, RNAP adopts a stable one-nucleotide-backtracked state^{7,8} that allows intrinsic cleavage to remove dinucleotide at 3'-terminus by hydrolysing the penultimate phosphodiester bond^{7,9,10} without the participation of external transcription factors. It was suggested that intrinsic cleavage follows the two-metal mechanism¹¹ that is observed in various enzymes, including human DNA polymerase η ^{12,13}, β ¹⁴⁻¹⁷ and μ ¹⁸, N4 RNAP¹⁹, archaeon *Desulfurococcus mobilis* homing endonuclease I-Dmol²⁰, *Thermus thermophilus* Argonaute protein²¹, and *Bacillus Halodurans* Ribonuclease H1²². The central question remains, however, on the identities of the general base and acid involved in this cleavage reaction.

The reaction mechanism of intrinsic cleavage by Pol II remains elusive. Although time-resolved X-ray studies have emerged to be a powerful technique for real-time monitoring the enzymatic reaction progress^{12-20,22}, it remains difficult to fully reveal the reaction mechanism using this technique for Pol II due to the resolution limit of current available crystal structures. The currently available crystal structures of one-residue-backtracked RNAP^{7,8} are unable to reveal a general base for activating the attacking nucleophile, as well as a general acid for protonating the oxyanion leaving group. The hydrogen atoms are difficult to be located accurately and precisely due to resolution limit. Previous biochemical studies suggested either the nucleobase of the terminal nucleotide^{9,10} or a particular enzyme residue (invariant His in the trigger loop (TL) motif)²³⁻²⁵ could function as the general base. The possible candidate serving as general acid is not clear. Furthermore, the characterization of the transition state is presumably more challenging due to its short lifetime (~fs). Considering the current experimental limitations, computational studies can serve as a powerful and complementary approach to elucidate reaction mechanisms with atomistic details²⁶⁻²⁹.

In the present work, we applied the state-of-the-art Born-Oppenheimer *ab initio* quantum mechanics/molecular mechanics-molecular dynamics (aiQM/MM-MD) simulations to elucidate the intrinsic cleavage mechanism in backtracked Pol II. Our results suggest that phosphate oxygen on the terminal misincorporated nucleotide unexpectedly serves as general base and assists in deprotonating the attacking water. The exposed 3'-oxyanion is protonated through water-mediated proton transfer without a general acid. Our theoretical predictions are validated by biochemical experiments. Taken together, our proposed cleavage mechanism provides important insights into understanding the intrinsic cleavage reaction by Pol II.

Results

A concerted reaction mechanism for Pol II's intrinsic cleavage.

In order to investigate the molecular mechanism of Pol II intrinsic cleavage reaction, we performed large-scale aiQM/MM-MD simulations on the backtracked Pol II system consisting of 103 QM atoms and 372,091 MM atoms (Fig. 1(a) and Supplementary Note 1). In typical aiQM/MM-MD calculations, spherical boundary scheme is applied where only the core region is enclosed in a truncated sphere. Our recently developed QChem/Amber³⁰ interface applies periodic boundary condition, allowing a more accurate description on long range electrostatic interaction. We first carefully validated our initial structure by showing that important structural features for catalysis are well maintained, including the distance between the two Mg²⁺ ions and the coordination of each of them (Supplementary Note 1–7 and Supplementary Fig. 1–2, 19).

We performed aiQM/MM calculations along the reaction coordinates to explore the 1-Dimensional (1D) and subsequently 2-Dimensional (2D) potential energy profiles after obtaining an active reaction complex structure from MD simulations, followed by extensive aiQM/MM-MD simulations along the minimum potential energy pathway on the 2D potential energy surface to explore the 2D free energy surface via umbrella sampling^{31,32} (Methods and supplementary Note 4). In particular, we collected 7 ps production aiQM/MM-MD simulations from 115 windows to calculate the 2D potential of mean force (PMF) or free energy profiles using weighted histogram analysis method (WHAM)³³. We further demonstrate that our reported free energy profiles are converged by doubling the length of our production aiQM/MM-MD simulations at each window (Supplementary Fig. 3) and the choices of reaction coordinates can effectively distinguish conformations with different covalent bonding (Supplementary Fig. 4).

We show that Pol II follows a concerted mechanism to cleave the phosphodiester bond of the backtracked mismatched nucleotide. We systematically surveyed atoms near the attracting nucleophile that can potentially serve as the general base to activate the nucleophile. Interestingly, we found that rather than a Pol II residue, the downstream phosphate oxygen (G12 pro-Rp) serves as the general base to activate the Mg_B-binding attacking nucleophile (the activation barrier is ~29 kcal/mol, see Fig. 1(b) and 2). The resulting overall free energies of Reactant State (RS), 1st Transition State (TS1), Intermediate State (IS), 2nd Transition State (TS2) and Product State (PS) are 0, 29.18±0.58, 20.43±0.99, 28.66±0.31 and 0.77±0.91 kcal/mol, respectively.

The aiQM/MM-MD simulations (Fig. 3) suggest that the concerted mechanism for the intrinsic cleavage is associated with a transition state (TS1) containing partial bonding to both the attacking nucleophile and leaving group. This results in a metastable intermediate state (IS), which contains an exposed A10 3'-end that is further protonated by the Mg_A-coordinating water molecule (Supplementary Note 6). The last step of the reaction is marked by the completion of proton transfer that recovers neutral water molecules in the active site. To validate the TS1 structure, we have relaxed the TS1 towards the two connecting minima (Supplementary Note 11–12). The relaxation pathway is fully consistent with our identified mechanism: the attacking water is deprotonated by the terminal phosphate oxygen (Supplementary Fig. 5). We have also examined the stability of the identified IS containing a protonated terminal phosphate and a cleaved phosphodiester bond, which further supports the ability of the 3'-nucleotide phosphate oxygen to accept a proton in serving as the general base (Supplementary Fig. 6). Our suggestion of G12 pro-Rp oxygen serving as the general base in intrinsic cleavage is supported by the resolved one-nucleotide-backtracked structures of *Saccharomyces cerevisiae* Pol II⁷ and *Thermus thermophilus* RNAP⁸, where both structures contain a twisted backbone between the backtracked and upstream nucleotides that places the pro-Rp atom of the former in close proximity to the active site (Supplementary Fig. 7).

Phosphate oxygen of 3'-RNA nucleotide as general base.

In our proposed mechanism (Fig. 3), pro-Rp oxygen of G12 serves as a general base that activates the attacking water for cleavage. The distance between pro-Rp oxygen of G12 and attacking water is important for activation. To directly test this proposed mechanism, we employed two different strategies to investigate the roles of pro-Rp oxygen in intrinsic cleavage.

In the first approach, we increased the distance between pro-Rp oxygen of G12 and attacking water by replacing 3'–5' phosphodiester linkage with a specific 2'–5' phosphodiester linkage between terminal G12 and U11. As the result, the distance between G12 non-bridging phosphate oxygen and the attacking water is greatly increased, thereby disabling it from acting as a general base. To demonstrate this, we performed MD simulations using 2'–5' RNA transcript and showed that the average distance between the G12 pro-Rp oxygen and the hydrogen on the attacking water is 5.47 ± 0.29 Å, larger than the value of 3.44 ± 0.29 Å obtained from 3'–5' linkage; hence, significantly compromising the catalysed cleavage activity (Fig. 4(e)). Apart from the distance between the G12 pro-Rp oxygen and the attacking water, we further show that the active-site architecture remains nearly intact with the change from the 3'–5' to 2'–5' linkage (Supplementary Fig. 10 and Supplementary Information Note 9 and 10).

To further validate the above theoretical observations, we performed an intrinsic cleavage assay using the scaffold described in Fig. 4(a), in which the G12 of the RNA transcript was either in a 3'–5' or 2'–5' linked orientation. The experimental assay of Pol II intrinsic cleavage clearly shows that the alteration of the linkage between the U11 and G12 nucleotide greatly reduces the overall cleavage rate of backtracked transcript (Supplementary Fig. 4(b, c)).

To quantify the cleavage rate, we found that double phase method fits the experimental data much better than the single-phase fitting method (Supplementary Fig. 11). Strikingly, the k_{fast} of 2'-5' linkage is more than 14-fold slower than that of 3'-5' linkage ($0.14 \pm 0.02 \text{ min}^{-1}$ for 3'-5' linkage vs. $0.0091 \pm 0.0022 \text{ min}^{-1}$ for 2'-5' linkage). In addition, the 3'-5' linked scaffold has higher percentage of fast phase than the 2'-5' linkage ($69.5 \pm 6.0 \%$ for 3'-5' linkage vs. $32.1 \pm 8.6 \%$ for 2'-5' linkage). The slow rate constants of these two scaffolds are similar ($0.00088 \pm 0.00031 \text{ min}^{-1}$ for 3'-5' linkage vs. $0.00081 \pm 0.00010 \text{ min}^{-1}$ for 2'-5' linkage).

These experimental results reveal two significant observations related to the role of G12 during backtracked intrinsic cleavage. First, the two-phase fittings suggest that two different intrinsic cleavage mechanisms may co-exist in the backtracked elongation complex. Second, we showed that the cleavage rates (k_{fast}) are highly affected by the backtracked nucleotide phosphodiester backbone linkage (3'-5' vs 2'-5' linkage). This result explicitly indicates that the linkage alteration of the backtracked nucleotide can strongly influence the intrinsic cleavage activity. This change in orientation directly shifts the position of the phosphate group as described in Fig. 4(f). Hence, the 14-fold reduction in cleavage activity, as well as the lower percentage of fast phase (Supplementary Fig. 11) for the altered 2'-5' linkage compared with the natural 3'-5' linkage supports the facilitative role of the phosphate group during intrinsic cleavage as proposed by our theoretical studies. Also, the reduced population of the fast phase in the 2'-5' linkage could be related to the enlarged distance between the G12 pro-Rp oxygen and the hydrogen on the attacking water, which further validates the importance of the phosphate group.

High pH stimulates phosphodiester hydrolysis through deprotonating the attacking water. The profiles and shapes of pH-dependent cleavage curves provide important insights into the reaction mechanisms. To test whether the cleavage mechanism of 2'-5' and 3'-5' linked scaffolds are the same, we also compared the pH profile of the 2'-5' and 3'-5' linked scaffolds. The pH profiles of cleavage reactions of 2'-5' and 3'-5' linked scaffolds were quite similar (Fig. 4(d)), suggesting that the same set of atoms are involved in the cleavage reactions. This result is consistent with our hypothesis.

To further confirm the involvement of pro-Rp oxygen of G12, we took a second approach using thio-substitution of this critical phosphate oxygen atom without significantly altering the atomic configuration of the active site (Fig. 5(a) and Methods). The production of oligonucleotides containing thio-substitution yields two stereoisomers, namely Rp-S and Sp-S (Fig. 5(a)). We performed reversed-phase high-performance liquid chromatography to separate Rp-S and Sp-S oligonucleotides from the mixture to determine the cleavage rates for the separated components (Methods). Among the two peaks we obtained (Thio-Peak 1 and Thio-Peak 2), the cleavage rate of Thio-Peak 2 is significantly reduced, while the cleavage rate of Thio-Peak 1 has little change (Fig. 5(b, c)). This result is consistent with our proposed reaction mechanism. On one hand, the stereo-specific Rp-S-substitution of terminal phosphate oxygen (Left panel of Fig. 5(a)) would reduce the ability to deprotonate water (Right panel of Fig. 5(a)) and hence a reduced cleavage rate, so we suggest Rp-S-substitution to correspond to Thio-Peak 2 (Fig. 5(c)). On the other hand, in the Sp-S substitution, oxygen stays as the proton acceptor and the cleavage rate would be minimally

affected, so we anticipate that the Sp-S-substitution correspond to Thio-Peak 1. Consistent with our results, a previous study also showed that the thio-substitution of terminal adenine phosphate oxygen resulted in a rate reduction in cleavage reaction rate by Pol II¹⁰.

Limited role of 3'-RNA nucleobase as alternative general base.

Previous studies suggested that the nucleobase of the backtracked nucleotide may directly participate as general acid/base in intrinsic cleavage, although divergence was observed between bacteria and eukaryotes^{9,10}. In particular, it was demonstrated that substituting N7 atom of adenine nucleobase with a methyl group has little effect on intrinsic cleavage in *Thermus aquaticus* RNAP⁹, but reduced the reaction rate by 3-fold in *Saccharomyces cerevisiae* Pol II¹⁰. Methyl-substitution of N7 atom of guanine nucleotide was demonstrated to yield 5-fold increase in reaction rate in *Thermus aquaticus* RNAP⁹, but its effect in *Saccharomyces cerevisiae* Pol II was still unknown. In order to investigate whether the guanine nucleobase participates as a general acid/base in intrinsic cleavage, we compared the intrinsic cleavage rates with the scaffold containing a terminal 7-deaza-guanine (7-deaza-G with substitution of the N7 atom) or a guanine (G) (Supplementary Fig. 12(a) and Supplementary Note 13). The observed cleavage rates of G and 7-deaza-G are at similar levels in the presence of 5 mM Mg²⁺ (0.021 ± 0.005 vs 0.024 ± 0.004 min⁻¹), suggesting that both the N7 atom and the nucleobase have limited contribution to intrinsic cleavage in Pol II under these conditions. Consistent with these experiments, our QM/MM calculations of the activation energy barrier considering the N7 atom of G12 as the general base lead to a significant increase by 20 kcal/mol, rendering the reaction inactive (Supplementary Fig. 9(a)). Interestingly, we found about 2.5-fold difference in cleavage rates between the G and 7-deaza-G template in the presence of 50 mM Mg²⁺ (0.022 ± 0.003 vs 0.054 ± 0.016 min⁻¹) (Supplementary Fig. 12(b)). This suggests that the N7 atom may involve in coordination of a transient third metal at high Mg²⁺ concentrations (see Supplementary Fig. 13 for two possible positions of Mg_C in our catalytic site after QM/MM energy optimization). A similar role of N7 of guanine for chelating Mg²⁺ was also proposed for bacterial RNAP using 7-deaza-G⁹.

To test if other nucleobase atoms such as O6 on the backtracked nucleotide base G12 could play a role in intrinsic cleavage serving as general base, we also performed QM/MM calculations for G12 O6 serving as possible alternative general base and our results show that it increases the activation energy by ~20 kcal/mol (Supplementary Fig. 9(c)). Taken together, we concluded that the contribution of the nucleobase of G12 to Pol II intrinsic cleavage is minor, if there is any.

Pol II residues are not likely to serve as general acid.

Given that D485 is the closest protein residue to the newly generated A10 3'-end oxygen, we performed MD simulations with protonated D485 to test whether D485 could serve as general acid and whether a protonated Asp is stable in the active site of backtracked Pol II (Supplementary Note 8). In the three MD trajectories produced, one of them shows a distorted active site configuration with U11 not coordinating with Mg_A, while the other two are stable with the proton on D485 pointing away from the active site (Supplementary Fig. 14).

To further examine whether the proton can be successfully accommodated when the nearest water molecule is removed, we repeated the MD simulation with five-coordinated Mg_A . The results show that this configuration is not associated with increased stability as the proton on D485 points away from the active site during position-restrained equilibration in all the three trials. To eliminate the possibility that the instability originates from force-field artefact, MD simulation was repeated with the proton on D485 with its position restrained. In all three trials, the Mg_A - Mg_B distance increases to near 5 Å during equilibration; therefore, we conclude that protonated D485 is not a stable configuration of the active site of backtracked Pol II complex and is not able to serve as the general acid. The unavailability of the general acid in the active site is consistent with our view that protein residues in Pol II are not directly involved in proton transfer during the cleavage reaction.

Discussion and Conclusion

In our proposed mechanism, intrinsic cleavage is assisted by the non-bridging phosphate oxygen on the terminal nucleotide, which is only available in the backtracked state. The equilibrium of pol II translocation states is driven by the 3'-misincorporated base and moves toward the backtracked state. Since the downstream pro-Rp oxygen atoms in the crystal structures occupy a similar position regardless of the identity and orientation of the base of the misincorporated nucleotide (Supplementary Fig. 7), we anticipate that the mechanism we elucidated for the G:G mismatch is general to all types of mismatch. Therefore, the observed sequence dependence of proofreading rates is likely to reflect the difference in tendency of backtracking instead of the distinct chemical reaction mechanisms. The identified important role of backtracking motion in transcription proofreading is supported by our previous study showing that a threonine residue on the Pol II bridge helix can probe the mismatch of the terminal base pair at i+1 site and facilitate backtracking³⁴. In addition to the proposed mechanism described in this work, it is also possible that Trigger Loop plays additional role in proofreading as a positional instead of acid-base catalyst²⁵. Our proposed catalytic role of phosphate is also consistent with a recent study showing a stimulating effect of inorganic phosphate on RNAP exonuclease cleavage³⁵.

The observed rate of Pol II's intrinsic cleavage is relatively slow in comparison with that of the TFIIS-assisted cleavage (slower by ~400 times)⁷. This slow rate is reflected by a relatively high free energy barrier (~29 kcal/mol) from our calculations for the intrinsic cleavage reaction. Interestingly, recent structural work suggested the potential involvement of additional transiently bound metal ions in DNA polymerase reaction by DNA polymerase η ¹³ and DNA polymerase β ¹⁴, as well as RNA cleavage conducted by RNase H²². To date, direct structural evidences in supporting the involvement of additional transiently-bound metal in transcriptional intrinsic cleavage by Pol II are not available. Nevertheless, we tested the potential role of a third Mg^{2+} (Mg_C) by modelling its possible locations based on previously published Mg_C conformations^{14,22} (Supplementary Fig. 13). In our modelled structures, the introduction of Mg_C only induces minor rearrangement of the active site and does not affect the alignment of the attacking nucleophile, which indicates that Mg_C will unlikely alter the reaction mechanisms, especially the identity of the general base. Subsequent QM/MM energy optimizations show that Mg_C could stabilize the energy change from reactant to product state by 3.4 kcal/mol in one structural model, suggesting its

potential role in lowering the free energy barrier (Supplementary Fig. 13(b)). It is worthy to further investigate the potential involvement of Mg_C in the future studies.

In this study, by employing state-of-the-art simulations accompanied by biochemical studies, we have elucidated key features of the mechanism of intrinsic cleavage catalysed by RNA polymerase II. The non-bridging pro-Rp oxygen of the terminal nucleotide G12 in the backtracked complex positions and deprotonates the attacking water molecule. The exposed 3'-oxyanion is protonated through proton transfer through metal-coordinating water molecules without the participation of a general acid. Our QM/MM calculation results indicate that all these three alternative bases (N7, O6, and U11 pro-Rp) are substantially less favourable in potential energy for endonucleolytic cleavage than our proposed general base bearing the G12 pro-Rp oxygen (Supplementary Note 13 and Supplementary Fig. 9). The proposed mechanism is supported by biochemical experiments with 3'-5'-to-2'-5' modification of the phosphodiester linkage between terminal G12 and U11, thio-substitution of the non-bridging pro-Rp oxygen of the terminal nucleotide G12, and N7 modification of nucleobase of G12. We consider this non-bridging phosphate-based reaction mechanism to be more reliable than a protein-based mechanism because the non-bridging pro-Rp oxygen is activated in response to the shift in translocation equilibrium upon misincorporation. Although other intrinsic cleavage mechanisms may co-exist, they appear as much less efficient and secondary. The dominating nucleobase-independent reaction mechanism suggests that backtracking instead of chemical reaction is the underlying means for observed sequence-dependent proofreading. Furthermore, unlike other specific endonucleases, a unique feature of Pol II is that the same active site can catalyse both polymerization (nucleotide addition) and cleavage reactions, which is distinct from Ribonuclease H²² and Flap endonucleases³⁶. Moreover, the enzyme active site as well as its RNA substrate are very different for Pol II and Ribonuclease H. The molecular mechanism for RNA cleavage reaction catalysed by Pol II is poorly understood. It is not clear whether Pol II has a similar mechanism like endonucleases or not. Our proposed mechanism, in which the phosphate Rp-oxygen serves as the general base rather than any Pol II residues, suggests that Pol II compartmentalises its backtracking (with conformational change highly reliant on Pol II residues to recognise NTP misincorporation) and cleavage mechanisms (Pol II independent reaction mechanism) in order to coordinate these two functions. This study therefore provides important insights into understanding how RNA is cleaved by Pol II. In fact, the similarity between the substrate-assisted reaction mechanism identified in this work and the other reported RNA cleaving mechanisms by a specific endonuclease is quite intriguing. It possibly provides a link on how dual functions of RNAP (polymerization and cleavage activities from the same active site) have evolved.

METHODS

System Preparation

The 3.4 Å x-ray crystal structure of *Saccharomyces cerevisiae* Pol II with G-dG mismatch (PDBID: 3GTG)⁷ was used in this study. Middle missing loops in Rpb 1, Rpb 2 and Rpb 8 were modelled using SwissPDB Viewer³⁷. Specifically, the long missing loop from residue 133 to 162 in Rpb 2 was modelled with flexible GAG sequence. Based on the pKa

calculations performed with H⁺⁺³⁸ and PROPKA3.1³⁹, all residues were assigned as their standard protonation state. The active Rpb1-H1085 was specifically assigned to be doubly protonated, because this protonation state was suggested to allow better interaction with the triphosphate group of incoming NTP⁴⁰. For details on MD calculations and for selecting the initial conformation from MD simulation were given in Supplementary Note 1–4 and Supplementary Fig. 14–17.

QM/MM and aiQM/MM-MD Calculations

QChem/Amber program was used to perform QM/MM calculations. The largest advantage of QChem/Amber is allowing full description of the whole system under periodic boundary condition. The usual approach of truncating the system into a small sphere would remove a large portion of the protein and nucleic acids, as well as the neutralizing Na⁺ ions which distribute widely in the solvent. QChem/Amber thus provides a more accurate description on long range electrostatic interactions.

The QM region consists of 103 atoms, including the complete G12 and U11 nucleotides, the ribose and the base of A10, the two Mg²⁺ ions, and three coordinating water molecules that involve in bond breaking and formation. The C5' atom of A10 was described using pseudopotential. The electrostatic sphere with radius of 25 Å was updated together with the position of centre U11 pro-Sp oxygen atom. The remaining 372,091 atoms were all included in the free MM region. B3LYP^{41–43} functional and 6–31G*^{44,45} basis set were used for all the QM atoms. MM atoms were described using the same force field parameters as stated above.

2D potential energy surface was first identified using reaction coordinate driving method. The reaction pathway was divided into two steps: (1) Endonucleolytic cleavage and (2) Proton transfer. In step 1, the attacking nucleophile W1 is deprotonated by the Rp-oxygen on the backtracked nucleotide, followed by cleavage of P-O bond between A10 and U11. The four variables O_{A10}-P_{U11}, P_{U11}-O_{W1}, O_{W1}-H_{W1} and H_{W1}-O_{G12} are denoted as r1, r2, r3 and r4, respectively. The two reaction coordinates were chosen as r1-r2 and r3-r4. In step2, the protonation of exposed 3'-end by water-mediated proton transfer was described by six variables: H_{W1}-O_{G12} (r4), O_{W1}-H_{W3} (r5), H_{W2}-O_{W1} (r6), O_{W3}-H_{W2} (r7), H_{W3}-O_{W3} (r8) and O_{A10}-H_{W3} (r9). The reaction coordinates were r4-r5+r6-r7 and r8-r9. Totally, 1810 and 2862 points were used to construct the 2D potential energy surface of step 1 and step 2, respectively. We further show that these choices of reaction coordinates can effectively distinguish conformations with different covalent bonding. In particular, a monotonic change is observed when variations of actual bond distances were projected on to our reaction coordinate for both step 1 and step 2 (Supplementary Fig. 4).

Along the minimum energy pathway on the 2D potential energy surface, we selected initial structures (i.e. windows) for free energy calculation using umbrella sampling method^{31,32}. For each window, 1 ns equilibration with the QM region kept frozen was performed to ensure adequate relaxation of the MM part such as the hydrogen-bonding network in the active site. The final frames served as the initial structures of aiQM/MM-MD simulations. At each window of the umbrella sampling, force constants in the range of 10 to 300 kcal/

mol·Å² were imposed on the reaction coordinates. The 2D PMF profile was calculated using weighted histogram analysis method (WHAM)³³.

Data in the first 4ps was considered as equilibration and thus was not included for analysis. As a result, the data within the 4–11 ps in the 115 windows was used for analysis. The minimum free energy pathway was mapped out using the Dijkstra's algorithm⁴⁶ from the 2D PMF of Step 1 (RC=r1-r2, r3-r4) and Step 2 (RC=r4-r5+r6-r7,r8-r9) respectively to elucidate the reaction mechanism as shown in Fig. 2(a). The resulting overall free energies of RS, TS1, IS, TS2 and PS are 0, 29.18±0.58, 20.43±0.99, 28.66±0.31 and 0.77±0.91 kcal/mol, respectively. The error bars were reported as standard deviations obtained from PMFs obtained from 1-ps segments of aiQM/MM-MD simulations. To examine the convergence of free energy profiles, we have further extended aiQM/MM-MD simulations of each window to 18ps, and the results indicate that our reported PMFs are converged (Supplementary Fig. 3). We note that our aiQM/MM-MD free energy simulations are limited by tens of picoseconds. If there exist protein conformational changes occurring at substantially longer timescales that can affect the reaction mechanisms, other approaches such as semi-empirical methods⁴⁷ may need to be adopted to enhance the conformational sampling.

Intrinsic Transcript Cleavage Assay

Cleavage reactions were performed by pre-incubating purified Pol II with the RNA:DNA scaffolds containing either a 3'–5' or 2'–5' linkage between U11 and G12 in RNA as described in Fig. 4. The elongation complex was assembled in a 20 mM Tris-HCl (pH = 9) buffer with 40 mM KCl.

Intrinsic cleavage was initiated upon addition of Mg²⁺. Final concentrations for intrinsic cleavage were 20 mM Tris-HCl (pH = 9), 100 nM Pol II, 25 nM scaffold, 40 mM KCl, and 50 mM MgCl₂. The reaction was quenched in 0.5 mM EDTA at various time points and analysed by denatured PAGE. Time points in Fig. 4(b) were 1 min, 5 min, 20 min, 1 hr, 3 hr, 8 hr, 24 hr, 48hr, and 96hr. In the pH profile of intrinsic cleavage assay, the buffers and pH values were: MES-6.2, HEPES-7.0, HEPES-7.9, CHES-8.8, CAPS-9.9; with a buffer concentration of 50mM. To fit the cleavage data, we tested both the single phase and the double phase dissociation equations. Intriguingly, the single-phase fitting was in an extremely poor quality, whereas the double phase fit the experimental data very well (Supplementary Fig. 11). The two-phase equation we used was described as follows: $Y = \text{Plateau} + A_1 \cdot \exp(-k_{\text{Fast}} \cdot X) + A_2 \cdot \exp(-k_{\text{Slow}} \cdot X)$ (1), where, A₁ and A₂ represent the population of the fast and slow phase, respectively.

For the thio-substitution cleavage assay, the pH value of the buffer was 7.5 and MgCl₂ concentration was 5mM. The time points in Figure. 5b were 1 min, 4 min, 15 min, 1 hr, 3 hr, 8 hr, 24 hr. The time dependence of cleavage product formation was fit to a one-phase association Equation to determine the observed rate (k_{obs}). To separate the two different RNA thio-substitution stereoisomers, oligonucleotides were purified by reversed-phase high-performance liquid chromatography on an Agilent 1200 series HPLC system equipped with a quaternary pump (Waters 1525), a multiple wavelength detector and fraction collector involving a mobile phase of 0.1M TEAA in water (Solvent A) and acetonitrile (Solvent B)

with flow rate 1mL/min, using Agilent ZORBAX XDB C18 column (4.6 mm x 250 mm, 5 μ m, and 50°C).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dr. Zhenyang Lin for helpful discussions. This work was supported by the Hong Kong Research Grant Council (HKUST C6009–15G and AoE/P-705/16 to X.H. and X.L.; 16302214 and T31–605/18-W to X.H.), King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) (OSR-2016-CRG5–3007 to X.H. and X.G.), Shenzhen Science and Technology Innovation Committee (JCYJ20170413173837121 to X.H.), and Innovation and Technology Commission (ITC-CNERC14SC01 to X.H.), the National Institutes of Health (Grant R35-GM127040 to Y.Z.; Grant GM102362 to D.W.). X.H. is the Padma Harilela Associate Professor of Science. This research made use of the computing resources of the Supercomputing Laboratory at King Abdullah University of Science and Technology.

References

1. Vermulst M et al. Transcription errors induce proteotoxic stress and shorten cellular lifespan. *Nat. Commun.* 6, 8065 (2015). [PubMed: 26304740]
2. Borukhov S, Polyakov A, Nikiforov V & Goldfarb A GreA protein: A transcription elongation factor from *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 89, 8899–902 (1992). [PubMed: 1384037]
3. Borukhov S, Sagitov V & Goldfarb A Transcript cleavage factors from *E. coli*. *Cell* 72, 459–466 (1993). [PubMed: 8431948]
4. Reines D Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II. *J. Biol. Chem.* 267, 3795–3800 (1992). [PubMed: 1371280]
5. Lange U & Hausner W Transcriptional fidelity and proofreading in Archaea and implications for the mechanism of TFS-induced RNA cleavage. *Mol. Microbiol.* 52, 1133–1143 (2004). [PubMed: 15130130]
6. Sigurdsson S, Dirac-Svejstrup AB & Svestrup JQ Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol. Cell* 38, 202–210 (2010). [PubMed: 20417599]
7. Wang D et al. Structural basis of transcription: Backtracked RNA polymerase II at 3.4 Å resolution. *Science* 324, 1203–1206 (2009). [PubMed: 19478184]
8. Sekine S, Murayama Y, Svetlov V, Nudler E & Yokoyama S The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Mol. Cell* 57, 408–421 (2015). [PubMed: 25601758]
9. Zenkin N, Yuzenkova Y & Severinov K Transcript-assisted transcriptional proofreading. *Science* 313, 518–520 (2006). [PubMed: 16873663]
10. Nielsen S & Zenkin N Transcript assisted phosphodiester bond hydrolysis by eukaryotic RNA polymerase II. *Transcription* 4, 209–212 (2013). [PubMed: 24270513]
11. Steitz TA & Steitz JA A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U. S. A.* 90, 6498–6502 (1993). [PubMed: 8341661]
12. Nakamura T, Zhao Y, Yamagata Y, Hua Y & Yang W Watching DNA polymerase η make a phosphodiester bond. *Nature* 487, 196–201 (2012). [PubMed: 22785315]
13. Gao Y & Yang W Capture of a third Mg^{2+} is essential for catalyzing DNA synthesis. *Science* 352, 1334–1337 (2016). [PubMed: 27284197]
14. Freudenthal BD, Beard WA, Shock DD & Wilson SH Observing a DNA polymerase choose right from wrong. *Cell* 154, 157–168 (2013). [PubMed: 23827680]
15. Freudenthal BD et al. Uncovering the polymerase-induced cytotoxicity of an oxidized nucleotide. *Nature* 517, 635–639 (2015). [PubMed: 25409153]

16. Vyas R, Reed AJ, Tokarsky EJ & Suo Z Viewing human DNA polymerase β faithfully and unfaithfully bypass an oxidative lesion by time-dependent crystallography. *J. Am. Chem. Soc.* 137, 5225–5230 (2015). [PubMed: 25825995]
17. Reed AJ & Suo Z Time-dependent extension from an 8-oxoguanine lesion by human DNA polymerase beta. *J. Am. Chem. Soc.* 139, 9684–9690 (2017). [PubMed: 28682600]
18. Jamsen JA et al. Time-lapse crystallography snapshots of a double-strand break repair polymerase in action. *Nat. Commun.* 8, 253 (2017). [PubMed: 28811466]
19. Basu RS & Murakami KS Watching the bacteriophage N4 RNA polymerase transcription by time-dependent soak-trigger-freeze x-ray crystallography. *J. Biol. Chem.* 288, 3305–3311 (2013). [PubMed: 23235152]
20. Molina R et al. Visualizing phosphodiester-bond hydrolysis by an endonuclease. *Nat. Struct. Mol. Biol.* 22, 65–72 (2014). [PubMed: 25486305]
21. Sheng G et al. Structure-based cleavage mechanism of *Thermus thermophilus* Argonaute DNA guide strand-mediated DNA target cleavage. *Proc. Natl. Acad. Sci. U. S. A.* 111, 652–657 (2014). [PubMed: 24374628]
22. Samara NL & Yang W Cation trafficking propels RNA hydrolysis. *Nat. Struct. Mol. Biol.* 25, 715–721 (2018). [PubMed: 30076410]
23. Yuzenkova Y & Zenkin N Central role of the RNA polymerase trigger loop in intrinsic RNA hydrolysis. *Proc. Natl. Acad. Sci. U. S. A.* 107, 10878–10883 (2010). [PubMed: 20534498]
24. Zhang J, Palangat M & Landick R Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat. Struct. Mol. Biol.* 17, 99–104 (2010). [PubMed: 19966797]
25. Mishanina TV, Palo MZ, Nayak D, Mooney RA & Landick R Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading. *Proc. Natl. Acad. Sci. U. S. A.* 114, E5103–E5112 (2017). [PubMed: 28607053]
26. Warshel A & Levitt M Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103, 227–249 (1976). [PubMed: 985660]
27. Warshel A Computer simulations of enzyme catalysis: Methods, progress, and insights. *Annu. Rev. Biophys. Biomol. Struct.* 32, 425–443 (2003). [PubMed: 12574064]
28. Riccardi D et al. Development of effective quantum mechanical/molecular mechanical (QM/MM) methods for complex biological processes. *J. Phys. Chem. B* 110, 6458–6469 (2006). [PubMed: 16570942]
29. Lu X et al. QM/MM free energy simulations: Recent progress and challenges. *Mol. Simul.* 42, 1056–1078 (2016). [PubMed: 27563170]
30. Zhou Y, Wang S, Li Y & Zhang Y Born-Oppenheimer ab initio QM/MM molecular dynamics simulations of enzyme reactions. *Methods Enzymol.* 577, 105–118 (2016). [PubMed: 27498636]
31. Hu P, Wang S & Zhang Y How do SET-domain protein lysine methyltransferases achieve the methylation state specificity? Revisited by Ab initio QM/MM molecular dynamics simulations. *J. Am. Chem. Soc.* 130, 3806–3813 (2008). [PubMed: 18311969]
32. Hu P, Wang S & Zhang Y Highly dissociative and concerted mechanism for the nicotinamide cleavage reaction in Sir2Tm enzyme suggested by ab initio QM/MM molecular dynamics simulations. *J. Am. Chem. Soc.* 130, 16721–16728 (2008). [PubMed: 19049465]
33. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH & Kollman PA The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13, 1011–1021 (1992).
34. Da L-T et al. Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nat. Commun.* 7, 11244 (2016). [PubMed: 27091704]
35. Gottesman ME & Mustaev A Inorganic phosphate, arsenate, and vanadate enhance exonuclease transcript cleavage by RNA polymerase by 2000-fold. *Proc. Natl. Acad. Sci. U. S. A.* 115, 2746–2751 (2018). [PubMed: 29483274]
36. Patel N et al. Flap endonucleases pass 5'-flaps through a flexible arch using a disorder-thread-order mechanism to confer specificity for free 5'-ends. *Nucleic Acids Res.* 40, 4507–4519 (2012). [PubMed: 22319208]

37. Guex N & Peitsch MC SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18, 2714–2723 (1997). [PubMed: 9504803]
38. Anandakrishnan R, Aguilar B & Onufriev AV H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* 40, W537–W541 (2012). [PubMed: 22570416]
39. Søndergaard CR, Olsson MHM, Rostkowski M & Jensen JH Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* 7, 2284–2295 (2011). [PubMed: 26606496]
40. Huang X et al. 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 (2010). [PubMed: 20798057]
41. Becke AD Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* 38, 3098–3100 (1988).
42. Lee Yang & Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B. Condens. Matter* 37, 785–789 (1988). [PubMed: 9944570]
43. Becke AD Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98, 5648–5652 (1993).
44. Petersson GA et al. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *J. Chem. Phys.* 89, 2193–2218 (1988).
45. Petersson GA & Al-Laham MA A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *J. Chem. Phys.* 94, 6081–6090 (1991).
46. Dijkstra EW A note on two problems in connexion with graphs. *Numer. Math.* 1, 269–271 (1959).
47. Cui Q & Elstner M Density functional tight binding: values of semi-empirical methods in an ab initio era. *Phys. Chem. Chem. Phys.* 16, 14368–14377 (2014). [PubMed: 24850383]

Figure 1. Elucidating Pol II intrinsic cleavage mechanism using QM/MM-MD simulations.

(a) Full description of the system with the use of QChem/AMBER simulation package. The system consists of 372,091 MM atoms and 103 QM atoms. Active site of backtracked RNA Polymerase II (PDBID: 3GTG) was modelled with MD following the procedure as described in Methods. Key atoms being included in QM partition are shown in coloured ball and stick representations. **(b)** Proposed reaction mechanism shown in chemical representation. Step 1: Endonucleolytic cleavage using water as the nucleophile and terminal phosphate oxygen as the general base (left). Step 2: Water-mediated proton transfer to the exposed 3'-end oxygen (right).

Figure 2. Identifying reaction pathway using free energy profiles.

(a) The overall 1D free energy profile. (b) Reaction coordinates and 2D free energy surfaces of step 1: endonucleolytic cleavage (left) and step 2: proton transfer (right). Please refer to the Methods section for our choices of reaction coordinates and Supplementary Note 11 for validation of the energy maxima of TS1 and TS2 and Supplementary Note 12 for validation of the energy minima of RS, IS and PS.

Figure 3. Structural representation along the reaction pathway.

After the formation of reactant state RS, deprotonation of the attacking water by the terminal G12 pro-Rp oxygen and SN2 cleavage of phosphodiester bond occur in an associative manner, leading to a trigonal bipyramidal TS1 (see also Supplementary Note 5). After the inversion of configuration of the cleaved phosphate is completed, a metastable intermediate IS with exposed 3'-oxanion is formed. The oxanion is protonated by a nearby metal-coordinating water molecule, leading to TS2. Subsequent proton transfer process occurs, resulting in the final product state PS with cleaved phosphate and newly formed 3'-end.

Figure 4. Alteration of the RNA 3'-terminal linkage reduces the intrinsic cleavage rate.

(a) Sequence of scaffold used in the intrinsic cleavage assay. (b-c). Results from the intrinsic cleavage assay show that 2'-5' modification significantly decreases the intrinsic cleavage rate compared to the natural 3'-5' linkage. The time point for the cleavage assay was: 0, 1 min, 5 min, 20 min, 1 h, 3 h, 8 h, 24 h, 48 h, 96 h. (d) Comparison of pH profiles of intrinsic cleavage (P2) among 2'-5' and 3'-5' linked scaffolds. For (c) and (d), the error bars are reported as standard deviations of three independent experiments. (e) The average distance between G12 pro-Rp oxygen and the attacking water. The error bars are reported as sample standard deviations obtained from distances computed from an ensemble of MD conformations as described in Supplementary Note 9–10. (f) Three representative reactant-like structures with 2'-5' linked scaffolds chosen from MD simulations, displaying the distances between G12 pro-Rp oxygen and nucleophilic water.

Figure 5. The effect of thio-substitution of non-bridging phosphate oxygen on intrinsic cleavage. (a) Chemical structure of the active-site atoms distinguishing the two different RNA thio-substitution stereoisomers: *Rp*-S substitution and *Sp*-S substitution. (b, c) The effects of thio-substitution stereoisomers on the intrinsic cleavage rate. The stereoisomer from Thio-Peak 2 decreases the cleavage rate by about 3-fold, whereas the stereoisomer from Thio-Peak 1 has minor effect. The time point for the cleavage assay was: 1 min, 4 min, 15 min, 1h, 3h, 8h, 24h. The error bars are reported as standard deviations of three independent experiments.