

RESEARCH ARTICLE

# Structural Model of RNA Polymerase II Elongation Complex with Complete Transcription Bubble Reveals NTP Entry Routes

Lu Zhang<sup>1</sup>, Daniel-Adriano Silva<sup>1‡</sup>, Fátima Pardo-Avila<sup>1</sup>, Dong Wang<sup>2</sup>, Xuhui Huang<sup>1\*</sup>

**1** Department of Chemistry and State Key Laboratory of Molecular Neuroscience, Center for System Biology and Human Health, School of Science and IAS, The Hong Kong University of Science and Technology, Kowloon, Hong Kong, **2** Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California, United States of America

‡ Current Address: Department of Biochemistry, University of Washington, Seattle, Washington, United States of America

\* [xuhuihuang@ust.hk](mailto:xuhuihuang@ust.hk)



OPEN ACCESS

**Citation:** Zhang L, Silva D-A, Pardo-Avila F, Wang D, Huang X (2015) Structural Model of RNA Polymerase II Elongation Complex with Complete Transcription Bubble Reveals NTP Entry Routes. *PLoS Comput Biol* 11(7): e1004354. doi:10.1371/journal.pcbi.1004354

**Editor:** Shi-Jie Chen, University of Missouri, UNITED STATES

**Received:** January 30, 2015

**Accepted:** May 22, 2015

**Published:** July 2, 2015

**Copyright:** © 2015 Zhang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by the Hong Kong Research Grant Council <http://www.ugc.edu.hk/eng/rgc/> [grant numbers 16302214, 609813, AoE/M-09/12, M-HKUST601/13, and T13-607/12R to XH]; National Science Foundation of China <http://www.nsf.gov.cn/publish/portal1/> [grant number 21273188 to XH]; National Institutes of Health <http://grants.nih.gov/grants/oe.htm> [grant number GM102362 to DW]; Hong Kong PhD Fellowship Scheme [www.ugc.edu.hk/eng/rgc/hkphd/](http://www.ugc.edu.hk/eng/rgc/hkphd/) [grant number PF10-17123 to FPA]

## Abstract

The RNA polymerase II (Pol II) is a eukaryotic enzyme that catalyzes the synthesis of the messenger RNA using a DNA template. Despite numerous biochemical and biophysical studies, it remains elusive whether the “secondary channel” is the only route for NTP to reach the active site of the enzyme or if the “main channel” could be an alternative. On this regard, crystallographic structures of Pol II have been extremely useful to understand the structural basis of transcription, however, the conformation of the unpaired non-template DNA part of the full transcription bubble (TB) is still unknown. Since diffusion routes of the nucleoside triphosphate (NTP) substrate through the main channel might overlap with the TB region, gaining structural information of the full TB is critical for a complete understanding of Pol II transcription process. In this study, we have built a structural model of Pol II with a complete transcription bubble based on multiple sources of existing structural data and used Molecular Dynamics (MD) simulations together with structural analysis to shed light on NTP entry pathways. Interestingly, we found that although both channels have enough space to allow NTP loading, the percentage of MD conformations containing enough space for NTP loading through the secondary channel is twice higher than that of the main channel. Further energetic study based on MD simulations with NTP loaded in the channels has revealed that the diffusion of the NTP through the main channel is greatly disfavored by electrostatic repulsion between the NTP and the highly negatively charged backbones of nucleotides in the non-template DNA strand. Taken together, our results suggest that the secondary channel is the major route for NTP entry during Pol II transcription.

and Consejo Nacional de Ciencia y Tecnología Fellowship [www.conacyt.mx](http://www.conacyt.mx) [grant number 215482 to FPA]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

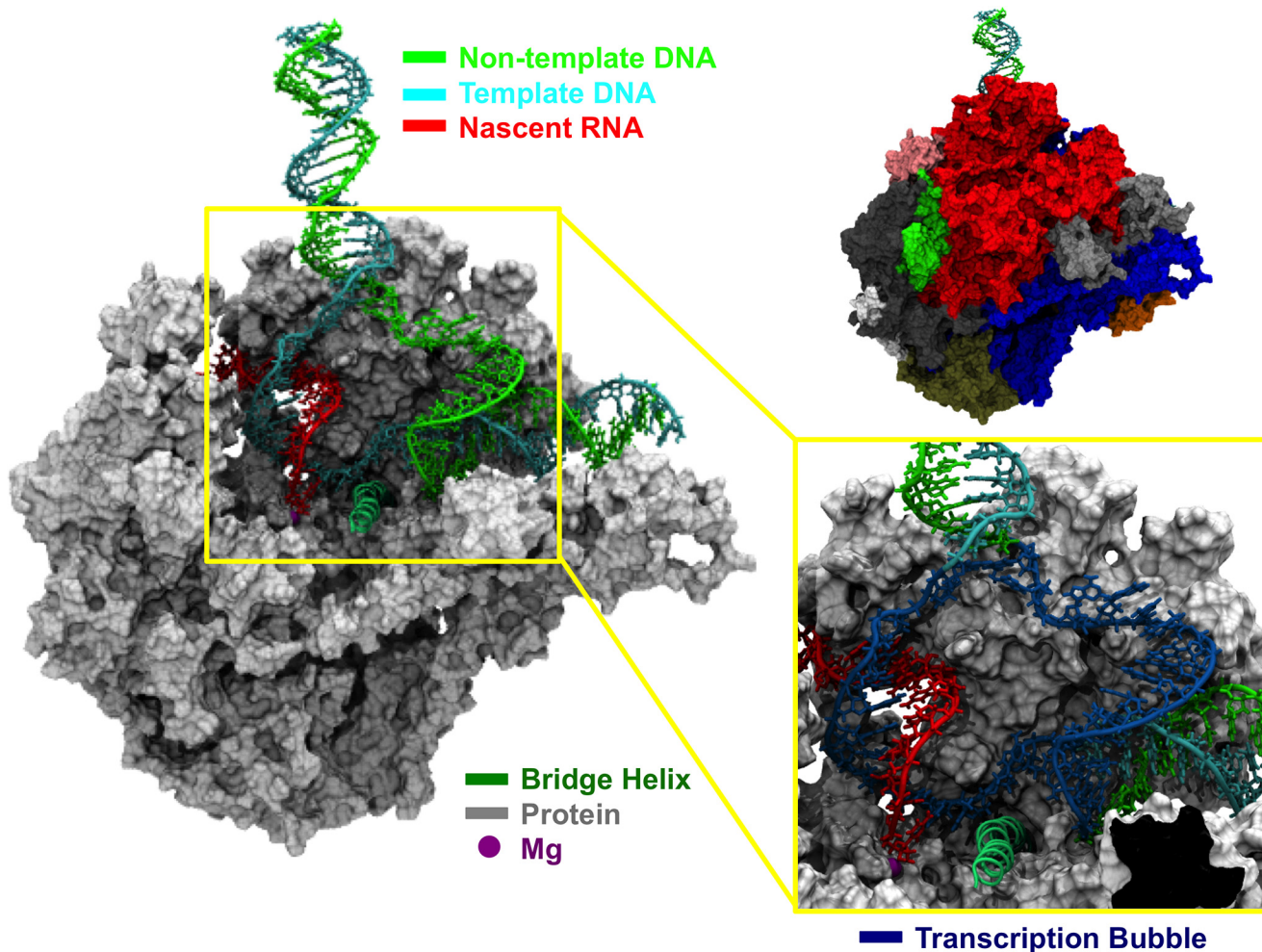
## Author Summary

In eukaryotic cells, the RNA polymerase II (Pol II) is a central enzyme that reads the genetic information encoded in the DNA template to synthesize a messenger RNA. To perform its function, Pol II needs to have the substrate nucleoside triphosphate (NTP) diffuse into its deeply buried active site. Despite numerous efforts, the NTP entry routes remain elusive: NTP could diffuse only through the secondary channel, or also via the main channel. The structural information of the transcription bubble is essential to study this process, however, the unpaired non-template DNA of the transcription bubble is absent in the available X-ray crystal structures. In this regard, we have built a structural model of the Pol II elongation complex with reconstructed transcription bubble using existing experimental data. We then performed Molecular Dynamics (MD) simulations and applied structural analysis to study the routes of NTP diffusion. We found that sterically the probability of NTP loading through the secondary channel is more than twice that of the main channel. Further analysis of the non-bonded energetic contributions to NTP diffusion suggests that NTP diffusion through the main channel is greatly disfavored by the electrostatic repulsion between the substrate and negatively charged backbones of nucleotides in the non-template strand of the transcription bubble. Altogether, our findings suggest that the secondary channel is the more favorable NTP diffusion route for Pol II transcription elongation.

## Introduction

The RNA polymerase II (Pol II) is a processive eukaryotic enzyme that plays a central role in transcription. It catalyzes the synthesis of messenger RNA (mRNA) with high fidelity and efficiency. Double-stranded DNA (dsDNA) enters the enzyme and unwinds around the active site. The dsDNA strand separates and bends at the downstream edge of the transcription bubble, leaving a DNA strand exposed as template (template DNA strand) for mRNA synthesis [1–5]. Nucleoside triphosphates (NTPs) need to diffuse into the active site for the incorporation to the mRNA strand [6–12]. The nascent mRNA chain remains paired to the DNA template in an RNA:DNA hybrid that extends for about 9 nucleotides [1, 2, 6–9, 13, 14]. Then, the template DNA strand separates from the nascent mRNA strand and re-anneals with the non-template DNA strand at the upstream edge of the transcription bubble (Fig 1).

In the past decades, extensive experimental and computational studies have been performed to elucidate the molecular mechanisms of different steps during the transcription process [1–50]. However, how the NTP diffuses from the surface of the enzyme into the active site is still not fully understood. X-ray crystal structures and biochemical studies have proposed two possible entry routes, namely: secondary channel [7–9, 21, 23, 34, 51–55] and main channel [56–62]. Both NTP loading routes are supported by experimental and computational studies. Crystallographic structures have shown that NTPs can bind to the active site as well as an entry site (E-site) in the secondary channel [7–9, 52], which in turn suggests a two-step mechanism of nucleotide selection. That is, NTP may go through the secondary channel to bind at the E site and then rotate its nitrogenous base to reach the active site. This nucleotide rotation may play an important role in discriminating dNTPs from NTPs, and help to ensure the correct base pairing between the incoming NTP and the template DNA [7, 8]. These ideas have also been confirmed by a computational study, which shows that NTP binding to the E-site in the secondary channel could greatly enhance the NTP binding rate at the active center [21]. Alternatively, NTP is suggested to first bind to a non-catalytic (i+2) site near the downstream edge of



**Fig 1. Structural model of Pol II elongation complex with the complete transcription bubble.** The left panel shows a cut-view of our Pol II model with a complete transcription bubble. The protein is shown in grey color (surface representation), the bridge helix is shown in dark green color (cartoon representation) and the magnesium ion is shown by a purple sphere. The DNA template, DNA non-template and mRNA strands are shown in cyan, green and red colors, respectively. The right-top panel shows the whole model without the cut view, with the various subunits of Pol II shown in different colors. The right-bottom panel is a close-up of the transcription bubble shown in dark blue.

doi:10.1371/journal.pcbi.1004354.g001

the transcription bubble, and then shuttles to the active site via the main channel [56–62]. Even though no crystal structure has been captured with NTP bound to the downstream DNA region, a study using millisecond phase kinetics suggested that the NTP could pre-load to the  $i+2$  site in the main channel prior to translocation into the active site, and this two-step mechanism would increase transcriptional fidelity and efficiency [57]. These observations have been further supported by a recent study using transient state kinetics, which demonstrated that the NTP binding to a non-catalytic template-specific site in the main channel could regulate nucleotide incorporation [59]. However, from a structural point of view, it remains unclear if the main channel contains sufficient space to allow loading of the NTP and how is that a pre-loaded NTP at  $i+2$  site reaches the Pol II active site. In particular, the downstream dsDNA occupies the main channel and leaves little space to allow the diffusion of the incoming NTP [6–9, 13–15, 17, 28, 54, 55]. Moreover, the bridge helix and fork regions seem to sterically block the diffusion route of preloaded  $i+2$  NTP to reach active site along with template base crossover above the bridge helix. In addition, the downstream region of the transcription

bubble overlaps with the region proposed to accommodate the incoming NTP, while the transcription bubble is either not included in the crystallographic preparation or its conformation is too flexible to be resolved in the available crystallographic structures of Pol II elongation complex [6–9, 13–15, 17]. Furthermore, the unpaired non-template base in the downstream edge of the transcription bubble might re-anneal with i+2 DNA and over-compete with free NTP. Lacking of full bubble structure makes it difficult to examine the possibility of NTP loading through the main channel based on available elongation structures. A structural model for Pol II elongation complex with the complete transcription bubble is thus crucial for the investigation of the NTP loading routes.

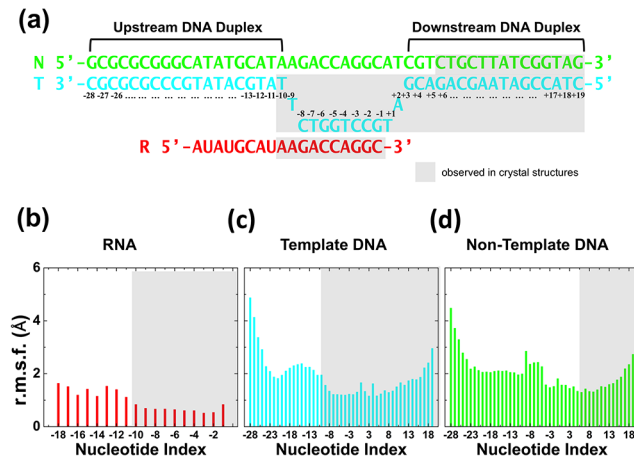
In this work, we have built a structural model of Pol II elongation complex with a complete transcription bubble based on available structures [7–9, 20, 35] (Fig 1). The final model contains 47 nucleotides in both the template and non-template DNA strands and the nascent mRNA strand has been extended to 18 nucleotides in length (Fig 1). Furthermore, we used this model to perform molecular dynamics (MD) simulations to study the effect of the dynamic Pol II complex on the diffusion of NTP. In general, we found that nucleotides that are observed in crystal structures [6–9, 13–15, 17] show less flexibility compared to those for which the electronic density is absent. Based on the stability of base pairing observed in our MD simulations, we identified that the size of the transcription bubble is of 15 nucleotides. To pinpoint possible NTP loading pathways, we analyzed ~1,000 MD snapshots by representing NTP as a sphere. We observed that both channels are possible NTP entry routes, while the number of conformations containing enough space for NTP loading through the secondary channel is more than twice that of the main channel. Based on these results, we constructed a series of structural models with full atomic representation of an NTP molecule located at various positions along the identified diffusion pathways, and performed MD simulations in the presence of explicit solvent. Strikingly, we found that NTPs in the main channel have experienced significantly unfavorable electrostatic interactions compared to those in the secondary channel, mainly due to the strong repulsion from the negatively charged backbones of the nearby nucleotides. Our findings suggest that secondary channel is the major route for the NTP entry during Pol II elongation.

## Results

### Structural model validation by molecular dynamics simulations

We built a structural model of Pol II with complete transcription bubble by extending the DNA (template/non-template) strands and nascent mRNA strand to 47 and 18 nucleotides in length, respectively (see [S1 Text.pdb](#)). This structural model equilibrates after 15ns in the MD simulations ([S1 Fig](#)). The root mean square deviation (RMSD) values of C $\alpha$  atoms reach 3Å in the first 10ns. Afterwards, the RMSD increases only slightly and stays around 3.5Å in the last 5ns of the 20ns simulations.

Nucleotides located in various regions of Pol II show different flexibility in MD simulations ([Fig 2](#)). The nucleotides of the nascent mRNA strand from positions i-1 to i-10 demonstrate small mobility, with the root mean square fluctuation (RMSF) < 1.0Å ([Fig 2B](#)). The mRNA beyond position i-11 is more exposed to the solvent and in consequence exhibits a higher flexibility ([Figs 1](#) and [2B](#)). This is consistent with missing electron density in crystallographic structures [6, 9, 14]. The template DNA strand contains 47 nucleotides and those located in the RNA:DNA hybrid region (from positions i-1 to i-8) show small fluctuations in the MD simulations ([Fig 2C](#)). The nucleotides from positions i+5 to i+16 are stable with an RMSF < 2.0Å as they are paired with the nucleotides of non-template DNA strand under a constrained protein environment. The non-template strand separates from the template DNA strand near the



**Fig 2. Structural validation of the model by MD simulations: Nucleic acids flexibility.** (a) The scheme represents the nucleic acid scaffold used in our simulation (numbers denote nucleotide positions). (b)-(d) The RMSF, per nucleotide position, observed in our MD simulations for: mRNA, template DNA and non-template DNA. Grey background denotes those nucleotides that have been observed in crystallographic structures.

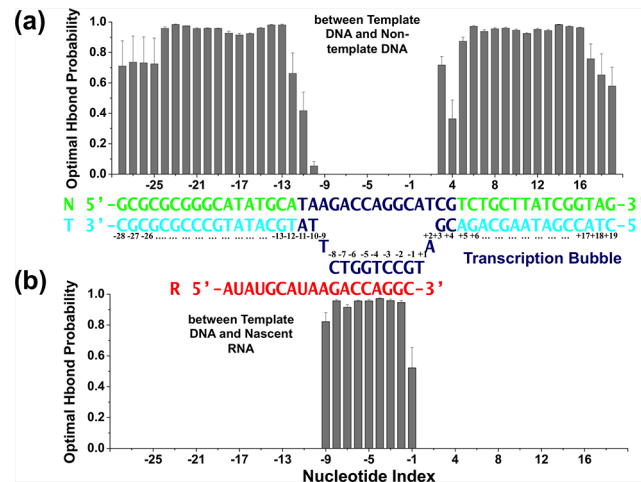
doi:10.1371/journal.pcbi.1004354.g002

bridge helix in the downstream edge, and re-anneals with the template strand in the upstream edge of the transcription bubble. The nucleotides of the non-template strand located near or within the bubble region (from  $i-9$  to  $i+4$ ) do not have stable base pairing partners, thus fluctuate significantly in the simulations (Fig 2D). On the contrary, the non-template nucleotides in the downstream dsDNA show less fluctuation (Fig 2D). Furthermore, terminal nucleotides of both DNA strands present higher fluctuations (Figs 2C and 2D). Especially for the region between  $i-24$  to  $i-28$  of the upstream DNA, the RMSF could be as high as  $4.5\text{\AA}$  (Figs 2C and 2D). Comparison to crystal structures indicates that the structural components observed in the X-ray crystals [6–9, 13–15, 17] are relatively stable in the MD simulations, while those absent or disordered in the crystal structures show more significant fluctuations.

By measuring the base pairing stability in MD simulations we found that the transcription bubble contains 15 nucleotides (Fig 3). The nucleotides from positions  $i-24$  to  $i-12$  region of the upstream DNA strands and from  $i+5$  to  $i+16$  of the downstream duplex show stable base pairing with hydrogen bonds probability  $> 60\%$  (Fig 3A). The dsDNA starts to unwind (probability  $< 60\%$ ) at  $i+4$  and reunites at  $i-12$  site. Furthermore, nucleotides in the RNA:DNA hybrid from  $i-2$  to  $i-9$  are well paired, with hydrogen bonds probability  $> 70\%$  (Fig 3B). The base pairing stability for the 3'-terminus nucleotide of the mRNA ( $i-1$ ) is slightly reduced (with a hydrogen bond probability of  $\sim 50\%$ ) due to its higher flexibility compared to the nucleotides locating from  $i-2$  to  $i-9$  (Figs 2B and 3B). The base pairing between the nucleotides of the hybrid is completely lost (with a probability dropping to zero) starting from  $i-10$ , indicating that the template DNA and nascent mRNA strands separate at position  $i-10$  (Fig 3B).

### Sterically, the secondary channel shows higher probability than the main channel for NTP loading

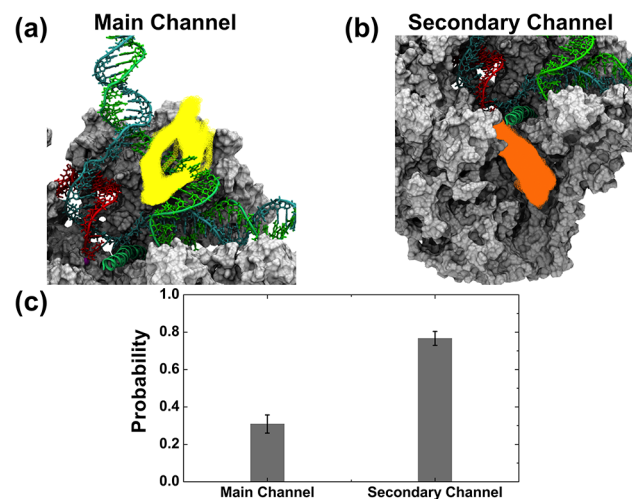
Using the program CAVER [63], we analyzed the MD conformational ensemble to find pathways that could allow the diffusion of a NTP molecule into the active site region of Pol II (see Methods). We discovered potential diffusion pathways through both the main channel and secondary channel. For the main channel, we identified a bifurcated pathway that connects the Pol II surface to the  $i+2$  binding site (Fig 4A). Different from previous models for the NTP entry through the main channel [56–62], neither branch of the pathway is located along the



**Fig 3. Structural validation of the model by MD simulations: Base-pairing stability.** The bar plots show the probability of finding the optimal number of hydrogen bonds for: (a) between the template and non-template DNA strands, and (b) between the DNA-RNA hybrid. In the middle, the nucleic acid scaffold scheme is demonstrated with the region where the DNA duplex opens to form the transcription bubble shown in dark blue.

doi:10.1371/journal.pcbi.1004354.g003

downstream DNA duplex, indicating that the available space in the downstream region is too limited to allow the passage of NTP. However, as shown in Fig 4A, the NTP may diffuse from the solvent to the i+2 binding site through the bifurcated pathway located at both sides of the non-template DNA strand in the transcription bubble region. In the later part of this article, we refer to this branched pathway as the main channel. We also discovered a pathway that can lead directly from the enzyme surface to the i+1 active site (Fig 4B). This pathway goes through



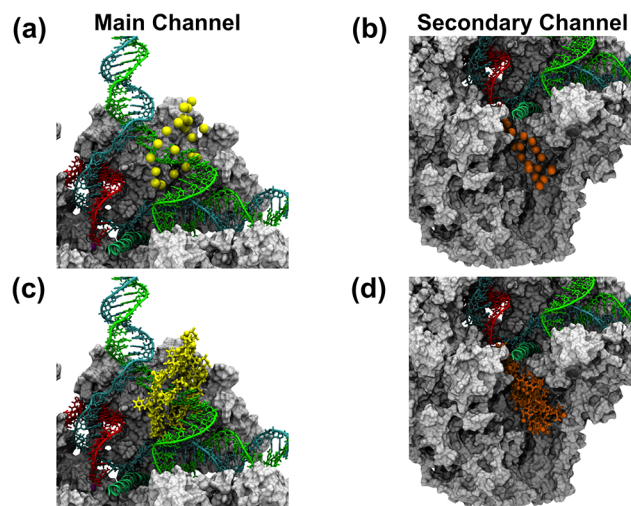
**Fig 4. The secondary channel exhibits higher probability for NTP entry than the main channel.** The cavity analysis was performed using the program CAVER (see Methods) with an ensemble of MD conformations of the elongation complex (without NTP). (a) The most possible pathways with enough space to allow the pre-loading of a NTP to the i+2 site are shown in yellow. Similarly, (b) shows the most possible pathways for a NTP to reach the active site in orange. In (c), the bar plot shows the probability of finding MD conformations with enough space for NTP diffusion through the main channel versus the secondary channel (see Methods). The template DNA (cyan), non-template DNA (green) and RNA (red) strands are shown with tube and licorice representations. The protein components are shown in grey with a cut view with a surface representation.

doi:10.1371/journal.pcbi.1004354.g004

the funnel and pore region, and is consistent with the previously proposed secondary channel. Hence we refer to this pathway as the secondary channel.

The probability of finding pathways for NTP (simulated as a hard sphere with a radius of 3.5 Å) to the *i*+1 active site along the secondary channel is more than twice that of finding the pathways to the *i*+2 binding site along the main channel (Fig 4C). In particular, the main channel pathway was only observed in ~31% of the MD conformations, while almost 77% of the MD conformations contain enough space for NTPs to diffuse to the active site through the secondary channel. Thus, though both channels contain enough space for the loading of NTP, the higher probability of finding pathways to the active site makes the secondary channel the major NTP entry route.

In order to consider the atomic structure of NTP molecules rather than simply treating them as spheres, we further modeled the all-atom NTP conformations at various locations along the above-mentioned loading pathways (Fig 5). In particular, we superimposed the center of mass (c.o.m.) of NTP with the locations of spheres identified by CAVER [63] and performed energy minimization. In this way, the NTPs could fit themselves into the space along the pathways (see Methods). We observed again that the secondary channel was still more favorable than the main channel. We found that the NTP molecules located along the main channel pathway deviated more from their initial positions compared to those along the secondary channel (S2 Fig). In particular, ~50% of the NTP molecules in the main channel move their c.o.m. > 1.0Å away from the initial locations after energy minimization. Three of them even show deviations > 2.0Å, with the highest value being of 3.5Å (S2A Fig). Besides, the shortest atomic distance deviation of NTP molecules in some locations of the main channel is as high as 2.0Å (S2A Fig). In contrast, NTP molecules in the secondary channel do not need to move significantly from their starting positions, where 90% of the conformations show an atomic distance deviation < 1.0Å (S2B Fig). By taking into account the steric effects of the



**Fig 5. MD relaxation of NTP in the main channel and the secondary channel.** (a-b) Initially, the points constituting the entry routes through the main and secondary channels were divided into 20 clusters. The yellow and orange spheres represent the center of each cluster for the main channel and the secondary channel, respectively. (c-d) A NTP molecule (in licorice presentation) accompanied by a bound magnesium atom (sphere representation) was aligned to each of the cluster centers (shown in (a-b)) by its center of mass (see Methods). These conformations were used as starting points for relaxation via MD simulations (see Methods). The template DNA, non-template DNA and mRNA strands are shown in cyan, green and red, respectively, with tube and licorice representations. The cut-view of the protein is shown in grey with a surface representation.

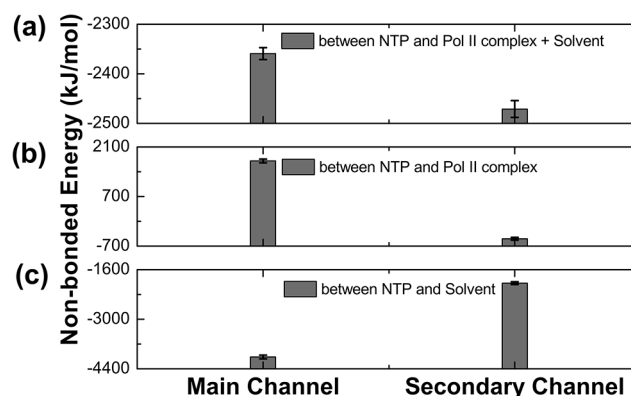
doi:10.1371/journal.pcbi.1004354.g005

NTP molecules along the pathways, the comparison of the NTP molecules' positional deviations along the two channels after energy minimization also supports that the secondary channel is the most favorable route for the NTP entry.

### NTP diffusion through the secondary channel is energetically more favorable than the main channel

To determine the stability of NTP along the pathways under the dynamic protein environment, we performed MD simulations starting from the energy minimized conformations and found that the NTP molecules have more favorable non-bonded interactions with the environment in the secondary channel than in the main channel (Fig 6A). Two components contribute to this energy difference: NTPs interacting with Pol II (protein and nucleotides) and the solvent (water and counter ions) (Figs 6B and 6C). We also find that the non-bonded energy difference is mainly due to the discrepancy of the Coulomb interaction (S3A Fig). Analysis of electrostatic potential shows a negatively charged potential along the main channel, while the secondary channel shows a surface with a nearly even distribution of positive and negative electrostatic potential (S4 Fig), suggesting a more favorable pathway for NTP diffusion through the secondary channel. The van der Waals (vdW) interactions (represented by the LJ interactions) do not show a significant difference between the two channels (S3B Fig).

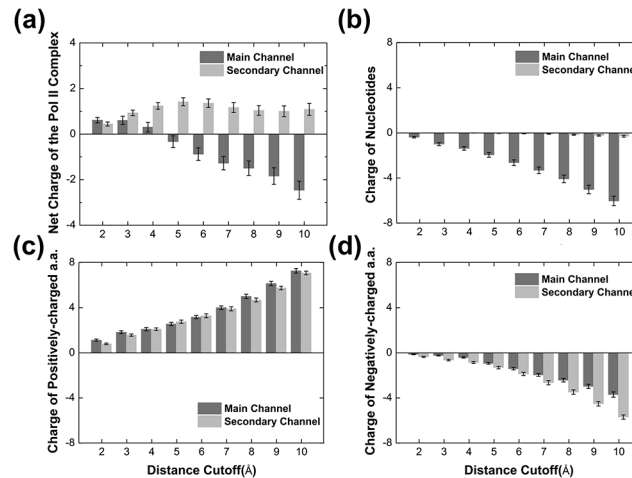
The NTP has more stable non-bonded interactions with the Pol II complex in the secondary channel than along the main channel. As shown in Fig 6B, the interaction between NTP and the Pol II complex in the main channel is quite unstable. This could be explained by the fact that in the main channel the charges surrounding the NTP are mainly negative (Fig 7A), thus destabilizing the also negatively charged NTP. In contrast, the NTP has more stable interactions with the Pol II complex in the secondary channel (Fig 6B), because the overall charge distribution surrounding the NTP is mainly positive, hence a favorable electrostatic environment for the NTP (Fig 7A). Further analysis shows that the charge difference between two channels arises from the nucleotides (Fig 7B) while the charge distributions of the amino acids surrounding NTP are similar along both channels (Figs 7C and 7D). As shown in Fig 4A, the main channel bifurcated pathway locates near to the negatively charged phosphate backbone of the non-template DNA strand, therefore the NTP which also carries negative charge would



**Fig 6. Non-bonded interactions NTP experiences revealed by MD.** The plots compare the energetic contribution of non-bonded interactions (electrostatic + vdW) along the main channel and secondary channel for: (a) NTP and its environment (the Pol II complex and solvent); (b) NTP and the Pol II complex; and (c) NTP and solvent. In these calculations, the Pol II complex includes the protein and nucleotides. The solvent contains water molecules together with the counter-ions. The results shown in (a) are the total contributions from the Pol II complex (shown in (b)) and the solvent (shown in (c)) (see [Methods](#)).

doi:10.1371/journal.pcbi.1004354.g006





**Fig 7. Distribution of charges surrounding the NTP during MD simulations.** (a) The plot shows the net charge of the Pol II (protein and nucleotides, y-axis) within certain radius (x-axis) of the NTP in the diffusion pathways through the main channel (dark grey) or the secondary channel (light grey). (b)-(d) The same as in (a) but for the charges of the nucleotides, positively-charged amino acids (a.a.) and negatively-charged a.a. around NTP, respectively.

doi:10.1371/journal.pcbi.1004354.g007

be destabilized in the main channel by the repulsion from the nearby phosphate backbones. On the contrary, the pathway of the secondary channel is distant from the nucleotide-dense regions, resulting in more stable interaction between NTP the Pol II complex (Figs 4B and 7B).

The analysis of the interaction between NTP and solvent demonstrates an opposite pattern: NTPs in the main channel are more stabilized by the solvent than those in the secondary channel (Fig 6C). This may be related to a higher exposure of the main channel to the solvent and thus the screening effect due to the counter ions and solvent molecules helps to stabilize NTPs (e.g.  $\sim 2 \text{ Na}^+$  within  $10 \text{ \AA}$  of NTP by average). On the contrary, the secondary channel is more deeply buried in the protein, thus NTPs along this pathway are poorly shielded by counter ions and solvent molecules (e.g. around half of MD conformations containing no  $\text{Na}^+$  within  $10 \text{ \AA}$  of NTP), leading to a less stable interaction between NTPs and solvent.

In summary, even though solvent screening effect helps to stabilize NTPs in the main channel, NTPs still prefer the secondary channel if we consider the nucleotide repulsions along the main channel during MD simulations. This also supports that the secondary channel is the major NTP entry route.

## Discussion

Although previous studies have provided great insight into the NTP loading process during Pol II transcription, it is still unclear how the NTP diffuses into the active site from the surface of the enzyme. Two routes have been proposed: a secondary channel directly leading from the enzyme surface to the active site [7–9, 21, 23, 34, 51–55] and a main channel following the downstream DNA duplex binding region into the i+2 binding site [56–62]. Until now, crystal structures have been able to capture NTP binding only to the secondary channel [7–9, 52]. Furthermore, based on the available structural information [6–9, 13–15, 17, 28, 54, 55], it is unclear if the downstream DNA in the main channel would impede the passage of NTP into the i+2 site. However, biochemical studies have proposed that NTPs could diffuse through the main channel and bind to a non-catalytic site to accelerate RNA synthesis [57, 59]. Even so, the role of main channel still remains controversial, as the non-template DNA part of the

transcription bubble overlapping with the main channel is crucial for the study of NTP diffusion routes while its conformation is absent in the available crystal structures of Pol II elongation complex [6–9, 13–15, 17].

In this work, we have built a structural model of Pol II with full transcription bubble (Fig 1). Based on this model, we performed MD simulations and used an ensemble of MD snapshots to study the NTP entry routes. Firstly, the NTP was treated as a sphere and MD conformations were examined to find pathways with enough space to allow NTP diffusion. Consistent with the previously proposed model for NTP loading through the secondary channel, our analysis also identified the same NTP entry pathway [7–9, 21, 23, 34, 51–55] (Fig 4B). In contrast, different from the previously hypothesized main channel pathway along the Pol II downstream DNA duplex [56–62], we found an bifurcated alternative pathway from the enzyme surface to the *i*+2 non-catalytic binding site, locating at two sides of the unpaired non-template DNA strand in the bubble region (Fig 4A). Furthermore, though analysis of the MD conformational ensemble shows that both channels are possible routes for NTP entry by considering only the steric effect, the probability of finding the secondary channel is more than two times higher than that of the main channel (Fig 4C).

Finally, to take into account the atomic details of the NTP molecule, we constructed all-atom models with NTP molecules at various locations along the pathways (Figs 5C and 5D) and further performed MD simulations to examine their stability. Strikingly, we discovered that NTPs in the secondary channel are energetically much more stable than those in the main channel (Fig 6A). In particular, NTP molecules experience significantly more favorable electrostatic interactions with the Pol II complex in the secondary channel than in the main channel (Fig 6B), mainly because the NTP is repelled by the nearby negatively charged phosphate backbone of the non-template DNA in the main channel (Fig 7B). In addition, investigation of the non-bonded interaction between NTP and solvent suggests that the screening effect from the counter-ions and solvent helps to stabilize the NTPs in the main channel (Fig 6C), as the bifurcated pathway is shorter and more exposed to the solvent than the secondary channel. Nevertheless, after taking into account of the overall environment including Pol II and solvent, we found that the NTP is still greatly favored energetically when entering through the secondary channel compared to the main channel. Our findings are consistent with the previous proposal that the secondary channel is the major pathway for NTP entry [7–9, 21, 23, 34, 51–55].

Our current study provides a qualitative comparison of the energetics for NTP to diffuse via the two channels; however, the actual energy difference between the two channels might be not accurate due to simplicity of the energy calculations that we have used. In this regard, free energy calculations may help to investigate an accurate value for the potential of mean force of NTP diffusion through the channels. Nevertheless, due to the complexity of this system, directly performing free energy calculations is challenging and involves enormous computational cost, since it requires sampling all the relevant conformations of the NTP along the diffusion pathways, including not only the conformations of NTP molecule itself, but also its distribution along the pathways and any associated conformational change of Pol II.

Based on our modeling results, we can predict key residues in the NTP diffusion that can be tested by experiment. For example, due to the more limited space for NTP loading through the main channel, we expect that mutations of alanine or glycine along the main channel (for example, A75 in chain B and G178 in chain A) to bulky amino acids or analogs could further reduce the probability for NTP diffusion through the main channel. Since electrostatic interactions play an important role for NTP diffusion, mutations of positively-charged amino acids (for example, K176 in chain A, K422 in chain B and R249 in chain B) to the negatively-charged amino acids could cause stronger repulsion to the NTP in the main channel, thus further impede the NTP entry through the main channel.

In vivo, RNAP may associate with various transcription factors and the presence of these proteins may affect the diffusion of NTP. For example, the elongation factor Spt4/5 [33, 35] locates near the entrance of the main channel pathway. To examine its effect on NTP diffusion, we performed channel analysis and MD simulations of the Spt4/5-Pol II complex (see [S2 Text](#) for modeling details). We found that Spt4/5 could further reduce the probability of NTP diffusion through the main channel (see [S2 Text](#) and [S5 Fig](#)). We are aware that the structure of Spt4/5 that we adopted [35] is not full-length. Therefore, there exist possibilities that the main channel pathway may be further blocked with the presence of full-length Spt4/5. During the proofreading process, the transcription factor TFIIS could bind to the secondary channel of Pol II to cleave the mis-incorporated RNA nucleotides. Its presence will occupy the entire secondary channel [17], and thus exclude the possibility for the NTP to diffuse to the active site through the secondary channel.

In future studies, it will be of interest to consider the coupling of NTP diffusion, DNA:RNA oscillation and translocation. According to the Brownian-ratchet model, the translocation motion is coupled to NTP binding, therefore it exists the possibility that RNAP could oscillate between the pre- and post-translocation states during NTP loading. However, in the current work each individual MD simulation is limited to the length of 20ns, and within this short time-scale it is not possible to observe the oscillation of the DNA/RNA hybrid between pre- and post-translocation states (Previous work suggests the timescale for the oscillation in the absence of NTP is at tens of microseconds [20]). The mechanism of the NTP shuttling from the  $i+2$  site to the active site has been suggested by a few previous studies. For example, the kinetic studies by Kennedy and Erie suggested that the NTP could diffuse through the main channel, bind to the  $i+2$  site and shuttle to the active site along with translocation [59]. Gong *et al.* also found that there is enough space for the NTP shuttling from the  $i+2$  site to the active site based on a minimum scaffold of transcriptional complex [57]. The NTP shuttling to the active site is a plausible model, but there also exists other alternative mechanisms. For example, it is possible that the interaction of the NTP at the  $i+2$  site simply has allosteric regulatory role of translocation, downstream bubble opening or catalysis. Our current work is focused on determining the possibility for NTP to diffuse into Pol II via the main channel (to the  $i+2$  site) or the secondary channel (to the  $i+1$  site, i.e. active site). For this purpose, we chose to use the post-translocation state of Pol II, this is because both sites ( $i+1$  and  $i+2$ ) are available for NTP binding, while in the pre-translocation state only the  $i+2$  binding site is available.

In conclusion, our work has provided insight into the possible routes of NTP entry by performing unbiased pathway searching along both the main channel and secondary channel using an ensemble of MD conformations. Two proposed entry channels are compared sterically and energetically. Our results have shown that the main channel is still a possible route by solely considering the steric effect. However, if we consider the substantial energy difference between the two channels, the secondary channel becomes the favorable route for NTP diffusion. Our work also lays foundation for future studies of the kinetics and thermodynamics of NTP loading in Pol II.

## Methods

### Structural model construction for the complete transcription bubble

The structural model of Pol II with full transcription bubble was built based on available structures [7–9, 20, 35] (see [S1 Text](#)). The post-translocation state model built previously for the translocation study [20] was used as our initial conformation. The downstream DNA was extended by 4 nucleotides after aligning to the crystal structure (PDBID: 2E2H [7]) with the P and O5' atoms of the nucleotides in positions from  $i+1$  to  $i+6$  of the template strand (refer to

the nucleotides positions in Fig 2A). Afterwards, the all-atomic model of eukaryotic RNAP II elongation complex containing full transcription bubble with Spt4/5 bound to the clamp domain [35] (quoted as “Pol II-Spt4/5” later in the text) was used as reference to build the transcription bubble and the upstream DNA. We noticed that there exist other models that may serve as alternative starting structures for our modeling such as the archaeal RNAP elongation complex model with the full transcription bubble [33]. Specifically, we aligned Pol II-Spt4/5 structure to our model with longer downstream DNA by P, O5' and C5' atoms of the nucleotides at  $i+1 \sim i+3$  positions. To model the downstream edge of the full transcription bubble, we used the coordinates of the nucleotides at positions  $i+3$  to  $i+6$  of the aligned Pol II-Spt4/5 to replace the corresponding nucleotides in our model. Furthermore, the fragments of template DNA starting from  $i-10$  to  $i-28$  of the 3'-end and the non-template DNA strand from the 5'-end to  $i+7$  were also extracted from the aligned Pol II-Spt4/5 and inserted into our model. Because the nucleotides were disconnected around the upstream edge of the transcription bubble due to the structural alignment, we applied energy minimization solely for nucleotides located at  $i-8 \sim i-10$  to connect them. One extra nucleotide pair was appended to the downstream dsDNA terminal by aligning the nucleotides' atoms C5 and P from positions  $i+10 \sim i+14$  of the Pol II-Spt4/5 to our model. The nascent mRNA strand was also extended: we aligned Pol II-Spt4/5 to our model by the carbon atoms of nucleotides bases at positions from  $i-6$  to  $i-9$ ; then the fragments of the aligned Pol II-Spt4/5 from  $i-9$  to  $i-18$  were inserted into our model. The final sequences shown in Fig 2A were achieved by point mutations of the nucleotides using the molecular modeling suite Coot [64].

We also manually fixed some structural clashes between the amino acids and the nucleotides. First, we aligned Pol II-Spt4/5 to our model by the C $\alpha$  atoms of Rpb1 residues 200~240 and 270~310. Rpb1 residues 248~260 from the aligned Pol II-Spt4/5 were extracted and used to replace the corresponding amino acids in our model. PHE with residue ID 252 was rotated to avoid its positional overlap with the nearby nucleotide. To correct the clash between the amino acids and the nucleotides in the 5' exit of nascent mRNA strand, we aligned Pol II-Spt4/5 to our model by the phosphate atoms of nucleotides from  $i-10$  to  $i-19$  and then extracted the Rpb1 residues 60~65 from the aligned Pol II-Spt4/5 to replace these six residues in our model. Similar modifications were also made to the Pol II Rpb2. In particular, we aligned Pol II-Spt4/5 to our model by the C $\alpha$  atoms of Rpb2 residues 480~530. Afterwards, the fork loop 4 (Rpb2 residues 501~510) from the aligned Pol II-Spt4/5 was extracted to replace the corresponding residues in our model. The side chain of Rpb2 residue 430 was rotated and residues 436~447 and 918~934 were also manually pulled to avoid their clash with the nucleotides. The final structural model contains template/non-template DNA strands and mRNA strand of 47 and 18 nucleotides in length, respectively (Figs 1 and 2A).

## MD simulations of Pol II elongation complex

The amber99sb force field [65] with modifications on nucleotides [66–71] was used to perform all-atom MD simulations. To maintain the coordination between the zinc and the protein, we have added harmonic restraints with a force constant of  $2261.03 \text{ kJ.mol}^{-1} \cdot \text{\AA}^{-2}$  between zinc ions and their coordinated cysteine residues. The protonation states of the histidines in our model were assigned as previously described [20].

The structural model with the complete transcription bubble was used as our starting conformation. To remove the steric clashes, we first performed a 5,000-steps energy minimization with the steepest descent algorithm by freezing the nucleotides. Furthermore, another 50,000 steps energy minimization was performed for the Pol II system in vacuum to smooth the contact between the amino acids and nucleotides. Next, we solvated the whole system in a water

box of  $160\text{\AA} \times 188\text{\AA} \times 160\text{\AA}$  ( $\alpha = 90^\circ$ ,  $\beta = 90^\circ$ ,  $\gamma = 90^\circ$ ) and neutralized it by adding 141 sodium ions. Our simulation system contains 481,887 atoms in total, including 139,789 TIP3P water molecules [72]. Afterwards, 10,000 steps energy minimization was performed for the whole system. To further relax the Pol II in solvent, position restraints with a force constant of  $10\text{ kJ} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$  was enforced on all the heavy atoms of the Pol II complex and simulation was performed for the whole system for 200ps under an NVT ensemble ( $T = 310\text{K}$ ). The final configuration from the position-restrained simulation was used to initiate 4 independent NVT production ( $T = 310\text{K}$ ) simulations with different initial velocities. The first 500ps were used for temperature annealing from 50K to 310K, followed up by 20ns simulations with the temperature kept at 310K. We stored snapshots every 20ps. The long-range electrostatic interactions beyond the cut-off at  $12\text{\AA}$  were treated with the Particle-Mesh Ewald (PME) method [73]. The Lennard-Jones interactions were smoothly switched off from  $10\text{\AA}$  to  $11\text{\AA}$ . The neighbors list was updated every 10 steps. An integration time step of 2.0ps was used and the LINCS algorithm [74] was applied to constrain all the bonds. All the MD simulations were performed using Gromacs 4.5 [75]. A PDB of the energy-minimized structure is available in the SI as [S1 Text.pdb](#).

## Structural validation

As shown in [S1 Fig](#), the system becomes equilibrated after 15ns. Thus the conformations from the last 5ns of MD simulations were adopted for the structural validations. To investigate the flexibility of the nucleotides, we used the C4 atom of each nucleotide for RMSF calculations for each trajectory. Afterwards, we averaged the data from all the trajectories and the resulting RMSF values for mRNA, template and non-template DNA are shown in [Figs 2B–2D](#). To investigate the base pairing stability, we calculated the probability of optimal hydrogen bonds formation during the MD simulations ([Fig 3](#)). In particular, 2 and 3 hydrogen bonds are defined as the “optimal number of hydrogen bonds” for A-T(U) and C-G pairs, respectively. We applied the default hydrogen bond definition in Gromacs [75] ( $3.5\text{\AA}$  for distance cut-off of donor-acceptor and  $30^\circ$  for angle cut-off of acceptor-donor-hydrogen) to determine the number of hydrogen bonds observed in MD simulations. Then we calculated the percentage of snapshots showing the optimal hydrogen bond number for each trajectory. We applied the bootstrapping algorithm with replacement to get the average hydrogen bond probability (10 iterations with 10 random trajectories per iteration). We also validated the protonation states of titratable residues and discussed the placement of counter ions in our MD simulation model ([S6](#) and [S7 Figs](#) and [S1–S6 Tables](#)).

## Channel analysis based on MD conformations

The last 5ns of MD simulations (1,004 MD snapshots) were used to analyze the channels. The program CAVER [63] was used to search for cavities that might form pathways connecting the protein surface to the binding site ([Figs 4A and 4B](#)). For this purpose, the NTP was simulated as a sphere of radius  $3.5\text{\AA}$  (as suggested previously in [21]). Default settings were used for other input parameters to search the pathway, except that “shell\_radius” and “shell\_depth” were set to be  $30\text{\AA}$  and  $40\text{\AA}$  considering the distance from the Pol II surface to the buried binding site (see [63] for details). For the main channel, the i+2 on-catalytic binding site was used as the initial point for the pathway search. All the pathways that could allow the sphere to go through were identified and divided into groups according to their mutual geometry distances. These groups are then ranked according to the number and the cost of pathways in each group (see [63] for methodology details). During our analysis, we only focused on the first group, which is the most probable or the one of the highest priority ([Fig 4A](#)). We then applied the

bootstrapping algorithm to randomly select 100 MD conformations from our ensemble for 100 times, and calculated fraction of these conformations containing pathways that belong to the most probable group (Fig 4C). A similar procedure was applied to search the pathways and calculate the probability for the secondary channel, but using the *i*+1 active site as the initial point in CAVER (Figs 4B and 4C).

## MD simulations for the NTP loading

We first used the k-centers algorithm in MSMBuild-1.0 [76] to cluster all the points constituting the pathways into 20 microstates according to their positional similarity. We then extracted the central point of each microstate (Figs 5A and 5B) and used these points as the reference positions for placing NTP molecules in the channels. According to the sequence shown in Fig 2A, the matched NTPs for the main channel and secondary channel are UTP and ATP, respectively. Each NTP molecule carries one magnesium ion coordinated with its  $P_{\alpha}$ ,  $P_{\beta}$  and  $P_{\gamma}$  atoms to simulate the Mg B in the crystal structure [7]. The c.o.m. of the NTP with  $Mg^{2+}$  was aligned to the microstate center (Figs 5C and 5D). We then used the aligned NTP and the Pol II MD conformation corresponding to the specific center to build up the structural models with NTP along the channels. Because we used 20 microstate centers for each channel, we have 40 different structural models in total, including 20 models with  $(UTP-Mg)^{2+}$  in the main channel and other 20 models with  $(ATP-Mg)^{2+}$  in the secondary channel.

The parameters of the protein residues, DNA, RNA and the ions were taken from the all-atom amber99sb force field [65] with modifications on nucleotides [66–71]. To simulate NTP, parameter modifications on the polyphosphate [77] of the NTP were also included. The coordination between the  $Mg^{2+}$  and NTP was kept by adding the harmonic restraints with a force constant of  $2261.03 \text{ kJ.mol}^{-1}.\text{\AA}^{-2}$  between the magnesium ion and an oxygen atom attached to the  $P_{\gamma}$  atom. Since the binding of the  $Mg^{2+}$  to the NTP may induce significant charge re-distribution, we have regenerated the partial charges of the  $(NTP-Mg)^{2-}$  group using the restrained electrostatic potential (RESP) [78] fitting to the quantum calculation (HF/6-31G\*). The partial charges of the  $(NTP-Mg)^{2-}$  group ( $(UTP-Mg)^{2-}$  for main channel and  $(ATP-Mg)^{2-}$  for secondary channel) are listed in S7 and S8 Tables. The quantum calculations were performed using Gaussian03 [79].

After building up the structural models, each system was then neutralized by adding 143 sodium ions and solvated in a water box containing 139,784 TIP3P water molecules [72]. Because in the previous channel analysis the  $(NTP-Mg)^{2-}$  group was simulated as a sphere without real geometrical shape, the  $(NTP-Mg)^{2-}$  group in the simulation models may clash with the surrounding residues. Furthermore, to avoid the inserted  $(NTP-Mg)^{2-}$  group from perturbing the Pol II conformation, we first froze the Pol II complex and only performed 1,000-steps energy minimization with the steepest descent algorithm on the  $(NTP-Mg)^{2-}$  group to let it re-orient and fit itself into the channels. To investigate the movement of the  $(NTP-Mg)^{2-}$  group, we measured the positional change for the c.o.m. of  $(NTP-Mg)^{2-}$  group for each energy minimization simulation (S2 Fig). We also calculated the shortest distance between the atoms of  $(NTP-Mg)^{2-}$  group and the initial reference center position (S2 Fig). After energy minimization of only the  $(NTP-Mg)^{2-}$  group in the channels, we performed 10,000-steps energy minimization on the whole system, followed by a 200ps position restrain simulation with a force constant of  $10 \text{ kJ.mol}^{-1}.\text{\AA}^{-2}$  on all the heavy atoms of the Pol II complex together with  $(NTP-Mg)^{2-}$  group under NVT ensemble ( $T = 310\text{K}$ ). Afterwards, the restrain was released and one 10ns simulation was performed with random initial velocity under the NVT ensemble, and the simulated annealing algorithm was applied to elevate the temperature from 50K to 310K in the first 500ps of this simulation. Because we have 40 structural models in total

and we performed one simulation for each model, finally we have 40 MD trajectories. Other simulation parameters are the same as those used for the MD simulations of Pol II elongation complex.

### Calculations of non-bonded interactions that NTP experiences

We collected the last 5ns MD conformations for the calculations. Because we saved the snapshots every 20ps, the ensemble contains 5,020 conformations for each channel. In the calculations, Pol II, (NTP-Mg)<sup>2-</sup> and solvent were treated as three individual energy groups. For each MD conformation, we first calculated the Lennard-Jones ( $E_{LJ}$ ) and short-range Coulomb ( $E_{Coul\_SR}$ ) interaction energies between the (NTP-Mg)<sup>2-</sup> and Pol II. For the long-range Coulomb interaction, we first calculated the total PME energy of (NTP-Mg)<sup>2-</sup> and Pol II ( $E_{PME\_NTP+PoII}$ ) by turning off the partial charges of solvent. Afterwards, we set the partial charges of (NTP-Mg)<sup>2-</sup> and solvent to zero to calculate the PME energy of Pol II ( $E_{PME\_PoII}$ ). Similarly, we turned off the partial charges of Pol II and solvent to calculate the PME energy of (NTP-Mg)<sup>2-</sup> ( $E_{PME\_NTP}$ ). The long-range electrostatic interaction energy ( $E_{Coul\_LR}$ ) between (NTP-Mg)<sup>2-</sup> and Pol II was then derived by subtracting ( $E_{PME\_PoII}$ ) and ( $E_{PME\_NTP}$ ) from ( $E_{PME\_NTP+PoII}$ ). By summing up  $E_{LJ}$ ,  $E_{Coul\_SR}$  and  $E_{Coul\_LR}$ , we could get the non-bonded interaction energy between (NTP-Mg)<sup>2-</sup> and Pol II. Similar procedure was applied to calculate the non-bonded interaction energy between (NTP-Mg)<sup>2-</sup> and solvent. The non-bonded interaction energy between (NTP-Mg)<sup>2-</sup> and environment was then obtained by summing up the interaction energy between (NTP-Mg)<sup>2-</sup> and Pol II/solvent. To demonstrate the non-bonded interaction energy in both channels, we used bootstrapping algorithm to randomly select 100 conformations from the ensemble and repeated it 100 times to obtain the average values (Fig 6). To investigate the non-bonded interaction energies that NTP experiences along the pathways, we divided the conformations from the last 5ns MD simulations into 20 clusters according to the (NTP-Mg)<sup>2-</sup> positional similarity in the channels. We then calculated the non-bonded interaction energy using the 20 MD cluster center conformations for each channel (S3 Fig).

### Calculations of charges around NTP

We calculated the net charge within certain radius of NTP for the Pol II complex (Fig 7). The MD conformations from the last 5ns simulations (5,020 conformations per channel) were used for the calculations. In particular, for one MD conformation, the charge of the Pol II complex was calculated by considering charged amino acids and nucleotides with any atom inside a certain radius of NTP atoms. We repeated the calculation by changing the distance cut-off from 2Å to 10Å. Afterwards, a bootstrapping algorithm was applied to randomly select 100 conformations from the MD conformational ensemble, and this process was repeated 100 times to get the averaged net charges of the Pol II complex (Fig 7A). To study more details about the charges of the Pol II complex, we applied the same calculation method for three separated groups: nucleotides, positive-charged amino acids (LYS, ARG and HIP) and negative-charged amino acids (GLU and ASP) (Figs 7B–7D). In addition, we also calculated the charge of sodium (Na<sup>+</sup>) by using the same method with the distance cut-off 10Å.

### Supporting Information

**S1 Text. Structural model of the Pol II elongation complex with the complete transcription bubble.**  
(PDB)

**S2 Text. Supplementary information text.**  
(DOC)

**S1 Fig. RMSD of the alpha carbon atoms in the four parallel 20ns MD simulations.**  
(TIF)

**S2 Fig. Distance between the NTP and the caver centers after energy minimization.** (a) Distance between NTP and caver center for the main channel. The red circles with solid line show the distance of NTP center of mass to the corresponding caver center, while the black triangles with dashed lines demonstrate the shortest distance of NTP to the caver center. (b) The same as (a) but for the secondary channel.  
(TIF)

**S3 Fig. Non-bonded interactions between NTP and the environment revealed by MD conformations.** (a) Coulomb interactions between NTP and environment along different positions of the main channel (red solid line with triangles) and secondary channel (black dashed line with circles). (b) Similar to (a) but for the Lennard-Jones energies.  
(TIF)

**S4 Fig. Electrostatic potential along the NTP channels.** (a) The surface of the protein and nucleotides within 10Å of the main channel NTP diffusion pathway are shown colored by its electrostatic potential. Amino acids at >10Å from the main channel are shown just in a surface representation and grey color. The template DNA, non-template DNA and RNA strand are shown in licorice representation (cyan, green and red colors, respectively). The right panel is a rotation of 180° respect to the left panel. The empty arrow denotes the NTP pathway. (b) The same as (a) but for the secondary channel.  
(TIF)

**S5 Fig. Spt4/5 elongation factor binding to the upstream DNA locates near the entry of main channel pathway.** In (a)-(c), the main channel pathway discovered is shown in yellow with different views (side view (a), front view (b) and top view (c)). The template DNA (cyan), non-template DNA (green) and RNA (red) strand are shown with tube and licorice representations. The Pol II protein components are shown in grey; (d)-(f) are similar to (a)-(c), but with the elongation factor Spt4 and Spt5 shown in pink and blue, respectively.  
(TIF)

**S6 Fig. Distribution of counter ion Na<sup>+</sup> in the vicinity of Nucleic Acids during MD simulation.** (a) Left and right panels show the distribution of sodium ions (blue sphere) before and after the MD simulation. The two bottom boxes are close-ups of the nucleotides and ions. The template DNA (cyan), non-template DNA (green) and RNA (red) strand are shown in tube and licorice representations. Protein surface is shown in light grey. (b) The plot shows the number of Na<sup>+</sup> ions within 9Å of the nucleotides during the course of 4 independent MD simulations.  
(TIF)

**S7 Fig. Probability of consistent protonation states in our MD setup and the pKa predicted by Propka3.1.** The values were obtained by considering 44 MD conformations. Please refer to [S1–S5 Tables](#) for the residue index.  
(TIF)

**S1 Table. Comparison between protonation states of aspartic acid (ASP) adopted in our MD simulations and those predicted by the Propka software.** <pKa> were obtained by



averaging predictions made by the Propka software using 44 MD conformations.  
(DOC)

**S2 Table. Comparison between protonation states of glutamic acid (GLU) adopted in our MD simulations and those predicted by the Propka software.**  $\langle pK_a \rangle$  were obtained by averaging predictions made by the Propka software using 44 MD conformations. Amino acids whose predicted pKa value (Propka) suggests a protonation state that differs from the one used in our MD simulations are highlighted.  
(DOC)

**S3 Table. Comparison between protonation states of lysine (LYS) adopted in our MD simulations and those predicted by the Propka software.**  $\langle pK_a \rangle$  were obtained by averaging predictions made by the Propka software using 44 MD conformations. Amino acids whose predicted pKa value (Propka) suggests a protonation state that differs from the one used in our MD simulations are highlighted.  
(DOC)

**S4 Table. Comparison between protonation states of arginine (ARG) adopted in our MD simulations and those predicted by the Propka software.**  $\langle pK_a \rangle$  were obtained by averaging predictions made by the Propka software using 44 MD conformations.  
(DOC)

**S5 Table. Comparison between protonation states of histidine (HIS) adopted in our MD simulations and those predicted by the Propka software.**  $\langle pK_a \rangle$  were obtained by averaging predictions made by the Propka software using 44 MD conformations. Amino acids whose predicted pKa value (Propka) suggests a protonation state that differs from the one used in our MD simulations are highlighted.  
(DOC)

**S6 Table. Distances between the amino acids and the main/secondary channel.** The table presents the distance to the NTP diffusion pathways for the six amino acids whose predicted pKa value (Propka) suggests a protonation state that differs from the one used in our MD simulations. The distances are the average minimum distance from the 44 MD conformations used for Propka predictions.  
(DOC)

**S7 Table. Partial charges of (UTP-Mg)<sup>2-</sup> group.**  
(DOC)

**S8 Table. Partial charges of (ATP-Mg)<sup>2-</sup> group.**  
(DOC)

## Acknowledgments

We would like to acknowledge Profs. Roger Kornberg and David Bushnell for helpful discussions. Computing resources were provided by the National Supercomputing Center in Shenzhen. Figures containing the Pol II complex were made with Visual Molecular Dynamics (“VMD”) software (<http://www.ks.uiuc.edu/Research/vmd/>).

## Author Contributions

Conceived and designed the experiments: LZ DAS XH. Performed the experiments: LZ. Analyzed the data: LZ. Wrote the paper: LZ DAS FPA DW XH.

## References

1. Kornberg RD. The molecular basis of eukaryotic transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104(32):12955–61. PMID: [17670940](#)
2. Cheung ACM, Cramer P. A Movie of RNA Polymerase II Transcription. *Cell*. 2012; 149(7):1431–7. doi: [10.1016/J.Cell.2012.06.006](#) PMID: [WOS:000305753800010](#).
3. Wang B, Feig M, Cukier RI, Burton ZF. Computational simulation strategies for analysis of multisubunit RNA polymerases. *Chem Rev*. 2013; 113(11):8546–66. doi: [10.1021/cr400046x](#) PMID: [23987500](#)
4. Pardo-Avila F, Da L-T, Wang Y, Huang X. Theoretical Investigations on Elucidating Fundamental Mechanisms of Catalysis and Dynamics Involved in Transcription by RNA Polymerase. *Journal of Theoretical and Computational Chemistry*. 2013;0(0):1341005-.
5. Martinez-Rucobo FW, Cramer P. Structural basis of transcription elongation. *Bba-Gene Regul Mech*. 2013; 1829(1):9–19.
6. Kettenberger H, Armache KJ, Cramer P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Molecular cell*. 2004; 16:955–65. PMID: [15610738](#)
7. Wang D, Bushnell Da, Westover KD, Kaplan CD, Kornberg RD. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*. 2006; 127(5):941–54. PMID: [17129781](#)
8. Westover KD, Bushnell Da, Kornberg RD. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*. 2004; 119(4):481–9. PMID: [15537538](#)
9. Gnatt AL, Cramer P, Fu JH, Bushnell DA, Kornberg RD. Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 angstrom resolution. *Science*. 2001; 292(5523):1876–82. PMID: [11313499](#)
10. Huang X, Wang D, Weiss DR, Bushnell Da, Kornberg RD, Levitt M. RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(36):15745–50. doi: [10.1073/pnas.1009898107](#) PMID: [20798057](#)
11. Landick R. NTP-entry routes in multi-subunit RNA polymerases. *Trends Biochem Sci*. 2005; 30(12):651–4. PMID: [16243529](#)
12. Kireeva M, Kashlev M, Burton ZF. Translocation by multi-subunit RNA polymerases. *Bba-Gene Regul Mech*. 2010; 1799(5–6):389–401.
13. Westover KD, Bushnell DA, Kornberg RD. Structural basis of transcription: Separation of RNA from DNA by RNA polymerase II. *Science*. 2004; 303(5660):1014–6. PMID: [14963331](#)
14. Sydow JF, Brueckner F, Cheung ACM, Damsma GE, Dengl S, Lehmann E, et al. Structural basis of transcription: mismatch-specific fidelity mechanisms and paused RNA polymerase II with frayed RNA. *Molecular cell*. 2009; 34(6):710–21. doi: [10.1016/j.molcel.2009.06.002](#) PMID: [19560423](#)
15. Cheung ACM, Cramer P. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*. 2011; 471(7337):249–53. doi: [10.1038/nature09785](#) PMID: [21346759](#)
16. Cramer P, Bushnell Da, Kornberg RD. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science (New York, NY)*. 2001; 292(5523):1863–76.
17. Wang D, Bushnell Da, Huang X, Westover KD, Levitt M, Kornberg RD. Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science (New York, NY)*. 2009; 324(5931):1203–6.
18. Da LT, Avila FP, Wang D, Huang XH. A Two-State Model for the Dynamics of the Pyrophosphate Ion Release in Bacterial RNA Polymerase. *Plos Computational Biology*. 2013; 9(4).
19. Da LT, Wang D, Huang X. Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. *Journal of the American Chemical Society*. 2012; 134(4):2399–406. doi: [10.1021/ja210656k](#) PMID: [22206270](#)
20. Silva DA, Weiss DR, Pardo-Avila F, Da LT, Levitt M, Wang D, et al. Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111(21):7665–70. doi: [10.1073/pnas.1315751111](#) PMID: [24753580](#)
21. Batada NN, Westover KD, Bushnell Da, Levitt M, Kornberg RD. Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(50):17361–4. PMID: [15574497](#)
22. Yu J, Da LT, Huang X. Constructing kinetic models to elucidate structural dynamics of a complete RNA polymerase II elongation cycle. *Phys Biol*. 2014; 12(1):016004. doi: [10.1088/1478-3975/12/1/016004](#) PMID: [25475208](#)
23. Cramer P, Bushnell DA, Fu JH, Gnatt AL, Maier-Davis B, Thompson NE, et al. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science*. 2000; 288(5466):640–9. PMID: [10784442](#)

24. Bushnell DA, Cramer P, Kornberg RD. Structural basis of transcription: alpha-Amanitin-RNA polymerase II cocystal at 2.8 Å resolution. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99(3):1218–22. PMID: [11805306](#)
25. Basu RS, Warner BA, Molodtsov V, Pupov D, Esyunina D, Fernandez-Tornero C, et al. Structural Basis of Transcription Initiation by Bacterial RNA Polymerase Holoenzyme. *Journal of Biological Chemistry*. 2014; 289(35):24549–59. doi: [10.1074/jbc.M114.584037](#) PMID: [24973216](#)
26. Jun SH, Hirata A, Kanai T, Santangelo TJ, Imanaka T, Murakami KS. The X-ray crystal structure of the euryarchaeal RNA polymerase in an open-clamp configuration. *Nat Commun*. 2014;5.
27. Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA. Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. *Science*. 2002; 296(5571):1285–90. PMID: [12016307](#)
28. Schwinghammer K, Cheung ACM, Morozov YI, Agaronyan K, Temiakov D, Cramer P. Structure of human mitochondrial RNA polymerase elongation complex. *Nature Structural & Molecular Biology*. 2013; 20(11):1298–U225.
29. Zhang Y, Feng Y, Chatterjee S, Tuske S, Ho MX, Arnold E, et al. Structural Basis of Transcription Initiation. *Science*. 2012; 338(6110):1076–80. doi: [10.1126/science.1227786](#) PMID: [23086998](#)
30. Basu RS, Murakami KS. Watching the Bacteriophage N4 RNA Polymerase Transcription by Time-dependent Soak-trigger-freeze X-ray Crystallography. *Journal of Biological Chemistry*. 2013; 288(5):3305–11. doi: [10.1074/jbc.M112.387712](#) PMID: [23235152](#)
31. Fu JH, Gnatt AL, Bushnell DA, Jensen GJ, Thompson NE, Burgess RR, et al. Yeast RNA polymerase II at 5 Å resolution. *Cell*. 1999; 98(6):799–810. PMID: [10499797](#)
32. Hirata A, Klein BJ, Murakami KS. The X-ray crystal structure of RNA polymerase from Archaea. *Nature*. 2008; 451(7180):851–U12. doi: [10.1038/nature06530](#) PMID: [18235446](#)
33. Klein BJ, Bose D, Baker KJ, Yusoff ZM, Zhang XD, Murakami KS. RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108(2):546–50. doi: [10.1073/pnas.1013828108](#) PMID: [21187417](#)
34. Zhang GY, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell*. 1999; 98(6):811–24. PMID: [10499798](#)
35. Martinez-Rucobo FW, Sainsbury S, Cheung ACM, Cramer P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *Embo J*. 2011; 30(7):1302–10. doi: [10.1038/emboj.2011.64](#) PMID: [21386817](#)
36. Abbondanzieri Ea, Greenleaf WJ, Shaevitz JW, Landick R, Block SM. Direct observation of base-pair stepping by RNA polymerase. *Nature*. 2005; 438(7067):460–5. PMID: [16284617](#)
37. Larson MH, Zhou J, Kaplan CD, Palangat M, Kornberg RD, Landick R, et al. Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(17):6555–60. doi: [10.1073/pnas.1200939109](#) PMID: [22493230](#)
38. Nayak D, Voss M, Windgassen T, Mooney RA, Landick R. Cys-Pair Reporters Detect a Constrained Trigger Loop in a Paused RNA Polymerase. *Molecular Cell*. 2013; 50(6):882–93. doi: [10.1016/j.molcel.2013.05.015](#) PMID: [23769674](#)
39. Mooney RA, Landick R. RNA polymerase unveiled. *Cell*. 1999; 98(6):687–90. PMID: [10499791](#)
40. Mooney RA, Landick R. Building a better stop sign: understanding the signals that terminate transcription. *Nat Methods*. 2013; 10(7):618–9. doi: [10.1038/nmeth.2527](#) PMID: [23807194](#)
41. Weixlbaumer A, Leon K, Landick R, Darst SA. Structural Basis of Transcriptional Pausing in Bacteria. *Cell*. 2013; 152(3):431–41. doi: [10.1016/j.cell.2012.12.020](#) PMID: [23374340](#)
42. Palangat M, Larson MH, Hu X, Gnatt A, Block SM, Landick R. Efficient reconstitution of transcription elongation complexes for single-molecule studies of eukaryotic RNA polymerase II. *Transcription*. 2012; 3(3):146–53. doi: [10.4161/trns.20269](#) PMID: [22771949](#)
43. Palangat M, Grass JA, Langelier MF, Coulombe B, Landick R. The RPB2 Flap Loop of Human RNA Polymerase II Is Dispensable for Transcription Initiation and Elongation. *Mol Cell Biol*. 2011; 31(16):3312–25. doi: [10.1128/MCB.05318-11](#) PMID: [21670157](#)
44. Opalka N, Brown J, Lane WJ, Twist KAF, Landick R, Asturias FJ, et al. Complete Structural Model of *Escherichia coli* RNA Polymerase from a Hybrid Approach. *Plos Biol*. 2010; 8(9).
45. Zhang JW, Palangat M, Landick R. Role of the RNA polymerase trigger loop in catalysis and pausing. *Nature Structural & Molecular Biology*. 2010; 17(1):99–U123.

46. Landick R. Transcriptional pausing without backtracking. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106(22):8797–8. doi: [10.1073/pnas.0904373106](https://doi.org/10.1073/pnas.0904373106) PMID: [19470457](https://pubmed.ncbi.nlm.nih.gov/19470457/)
47. Liu X, Bushnell DA, Silva DA, Huang XH, Kornberg RD. Initiation Complex Structure and Promoter Proofreading. *Science*. 2011; 333(6042):633–7. doi: [10.1126/science.1206629](https://doi.org/10.1126/science.1206629) PMID: [21798951](https://pubmed.ncbi.nlm.nih.gov/21798951/)
48. Liu X, Bushnell DA, Wang D, Calero G, Kornberg RD. Structure of an RNA Polymerase II-TFIIB Complex and the Transcription Initiation Mechanism. *Science*. 2010; 327(5962):206–9. doi: [10.1126/science.1182015](https://doi.org/10.1126/science.1182015) PMID: [19965383](https://pubmed.ncbi.nlm.nih.gov/19965383/)
49. Liu X, Bushnell DA, Kornberg RD. RNA polymerase II transcription: Structure and mechanism. *Bba-Gene Regul Mech*. 2013; 1829(1):2–8.
50. Zuo Y, Steitz Thomas A. Crystal Structures of the E. coli Transcription Initiation Complexes with a Complete Bubble. *Molecular Cell*. (0). doi: <http://dx.doi.org/10.1016/j.molcel.2015.03.010>.
51. Adelman K, Yuzenkova J, La Porta A, Zenkin N, Lee J, Lis JT, et al. Molecular mechanism of transcription inhibition by peptide antibiotic microcin J25. *Molecular Cell*. 2004; 14(6):753–62. PMID: [15200953](https://pubmed.ncbi.nlm.nih.gov/15200953/)
52. Kettenberger H, Armache KJ, Cramer P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Molecular Cell*. 2004; 16(6):955–65. PMID: [15610738](https://pubmed.ncbi.nlm.nih.gov/15610738/)
53. Mukhopadhyay J, Sineva E, Knight J, Levy RM, Ebright RH. Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. *Molecular Cell*. 2004; 14(6):739–51. PMID: [15200952](https://pubmed.ncbi.nlm.nih.gov/15200952/)
54. Vassylyev DG, Vassylyeva MN, Perederina A, Tahirov TH, Artsimovitch I. Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*. 2007; 448(7150):157–U3. PMID: [17581590](https://pubmed.ncbi.nlm.nih.gov/17581590/)
55. Vassylyev DG, Vassylyeva MN, Zhang JW, Palangat M, Artsimovitch I, Landick R. Structural basis for substrate loading in bacterial RNA polymerase. *Nature*. 2007; 448(7150):163–U4. PMID: [17581591](https://pubmed.ncbi.nlm.nih.gov/17581591/)
56. Erie DA, Kennedy SR. Forks, pincers, and triggers: the tools for nucleotide incorporation and translocation in multi-subunit RNA polymerases. *Current Opinion in Structural Biology*. 2009; 19(6):708–14. doi: [10.1016/j.sbi.2009.10.008](https://doi.org/10.1016/j.sbi.2009.10.008) PMID: [19913407](https://pubmed.ncbi.nlm.nih.gov/19913407/)
57. Gong XQ, Zhang CF, Feig M, Burton ZF. Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Molecular Cell*. 2005; 18(4):461–70. PMID: [15893729](https://pubmed.ncbi.nlm.nih.gov/15893729/)
58. Holmes SF, Erie DA. Downstream DNA sequence effects on transcription elongation—Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. *Journal of Biological Chemistry*. 2003; 278(37):35597–608. PMID: [12813036](https://pubmed.ncbi.nlm.nih.gov/12813036/)
59. Kennedy SR, Erie DA. Templated nucleoside triphosphate binding to a noncatalytic site on RNA polymerase regulates transcription. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108:6079–84. doi: [10.1073/pnas.1011274108](https://doi.org/10.1073/pnas.1011274108) PMID: [21447716](https://pubmed.ncbi.nlm.nih.gov/21447716/)
60. Nediaklov YA, Gong XQ, Hovde SL, Yamaguchi Y, Handa H, Geiger JH, et al. NTP-driven translocation by human RNA polymerase II. *Journal of Biological Chemistry*. 2003; 278(20):18303–12. PMID: [12637520](https://pubmed.ncbi.nlm.nih.gov/12637520/)
61. Xiong YL, Burton ZF. A tunable ratchet driving human RNA polymerase II translocation adjusted by accurately templated nucleoside triphosphates loaded at downstream sites and by elongation factors. *Journal of Biological Chemistry*. 2007; 282(50):36582–92. PMID: [17875640](https://pubmed.ncbi.nlm.nih.gov/17875640/)
62. Burton ZF, Feig M, Gong XQ, Zhang CF, Nediaklov YA, Xiong YL. NTP-driven translocation and regulation of downstream template opening by multi-subunit RNA polymerases. *Biochem Cell Biol*. 2005; 83(4):486–96. PMID: [16094452](https://pubmed.ncbi.nlm.nih.gov/16094452/)
63. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, et al. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *Plos Computational Biology*. 2012; 8(10).
64. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D*. 2010; 66:486–501. doi: [10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493) PMID: [20383002](https://pubmed.ncbi.nlm.nih.gov/20383002/)
65. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics*. 2006; 65(3):712–25.
66. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE, Laughton CA, et al. Refinement of the AMBER force field for nucleic acids: Improving the description of alpha/gamma conformers. *Biophys J*. 2007; 92(11):3817–29. PMID: [17351000](https://pubmed.ncbi.nlm.nih.gov/17351000/)
67. Zgarbova M, Otyepka M, Sponer J, Mladek A, Banas P, Cheatham TE, et al. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput*. 2011; 7(9):2886–902. PMID: [21921995](https://pubmed.ncbi.nlm.nih.gov/21921995/)

68. Besseova I, Banas P, Kuhrova P, Kosinova P, Otyepka M, Sponer J. Simulations of A-RNA Duplexes. The Effect of Sequence, Solute Force Field, Water Model, and Salt Concentration. *J Phys Chem B*. 2012; 116(33):9899–916. doi: [10.1021/jp3014817](https://doi.org/10.1021/jp3014817) PMID: [22809319](https://pubmed.ncbi.nlm.nih.gov/22809319/)
69. Krepl M, Zgarbova M, Stadlbauer P, Otyepka M, Banas P, Koca J, et al. Reference Simulations of Non-canonical Nucleic Acids with Different chi Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J Chem Theory Comput*. 2012; 8(7):2506–20. PMID: [23197943](https://pubmed.ncbi.nlm.nih.gov/23197943/)
70. Zgarbova M, Luque FJ, Sponer J, Cheatham TE, Otyepka M, Jurecka P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J Chem Theory Comput*. 2013; 9(5):2339–54. PMID: [24058302](https://pubmed.ncbi.nlm.nih.gov/24058302/)
71. Yildirim I, Kennedy SD, Stern HA, Hart JM, Kierzek R, Turner DH. Revision of AMBER Torsional Parameters for RNA Improves Free Energy Predictions for Tetramer Duplexes with GC and iG/C Base Pairs. *J Chem Theory Comput*. 2012; 8(1):172–81. PMID: [22249447](https://pubmed.ncbi.nlm.nih.gov/22249447/)
72. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys*. 1983; 79(2):926–35.
73. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A Smooth Particle Mesh Ewald Method. *J Chem Phys*. 1995; 103(19):8577–93.
74. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem*. 1997; 18(12):1463–72.
75. Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013; 29(7):845–54. doi: [10.1093/bioinformatics/btt055](https://doi.org/10.1093/bioinformatics/btt055) PMID: [23407358](https://pubmed.ncbi.nlm.nih.gov/23407358/)
76. Bowman GR, Huang XH, Pande VS. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods*. 2009; 49(2):197–201. doi: [10.1016/j.ymeth.2009.04.013](https://doi.org/10.1016/j.ymeth.2009.04.013) PMID: [19410002](https://pubmed.ncbi.nlm.nih.gov/19410002/)
77. Meagher KL, Redman LT, Carlson HA. Development of polyphosphate parameters for use with the AMBER force field. *J Comput Chem*. 2003; 24(9):1016–25. PMID: [12759902](https://pubmed.ncbi.nlm.nih.gov/12759902/)
78. Bayly CI, Cieplak P, Cornell WD, Kollman PA. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges—the Resp Model. *J Phys Chem-U.S.* 1993; 97(40):10269–80.
79. Gaussian03. G03. In: Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al., editors. Wallingford CT: Gaussian, Inc.; 2004.