



Modeling DNA Opening in the Eukaryotic Transcription Initiation Complexes via Coarse-Grained Models

Genki Shino and Shoji Takada*

Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto, Japan

OPEN ACCESS

Edited by:

Maciej Maciejczyk,
University of Warmia and Mazury in
Olsztyn, Poland

Reviewed by:

Cezary Czaplowski,
University of Gdansk, Poland
Wei Chen,
Independent Researcher, Austin, TX,
United States

*Correspondence:

Shoji Takada
takada@biophys.kyoto-u.ac.jp

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 08 September 2021

Accepted: 07 October 2021

Published: 15 November 2021

Citation:

Shino G and Takada S (2021) Modeling
DNA Opening in the Eukaryotic
Transcription Initiation Complexes via
Coarse-Grained Models.
Front. Mol. Biosci. 8:772486.
doi: 10.3389/fmolb.2021.772486

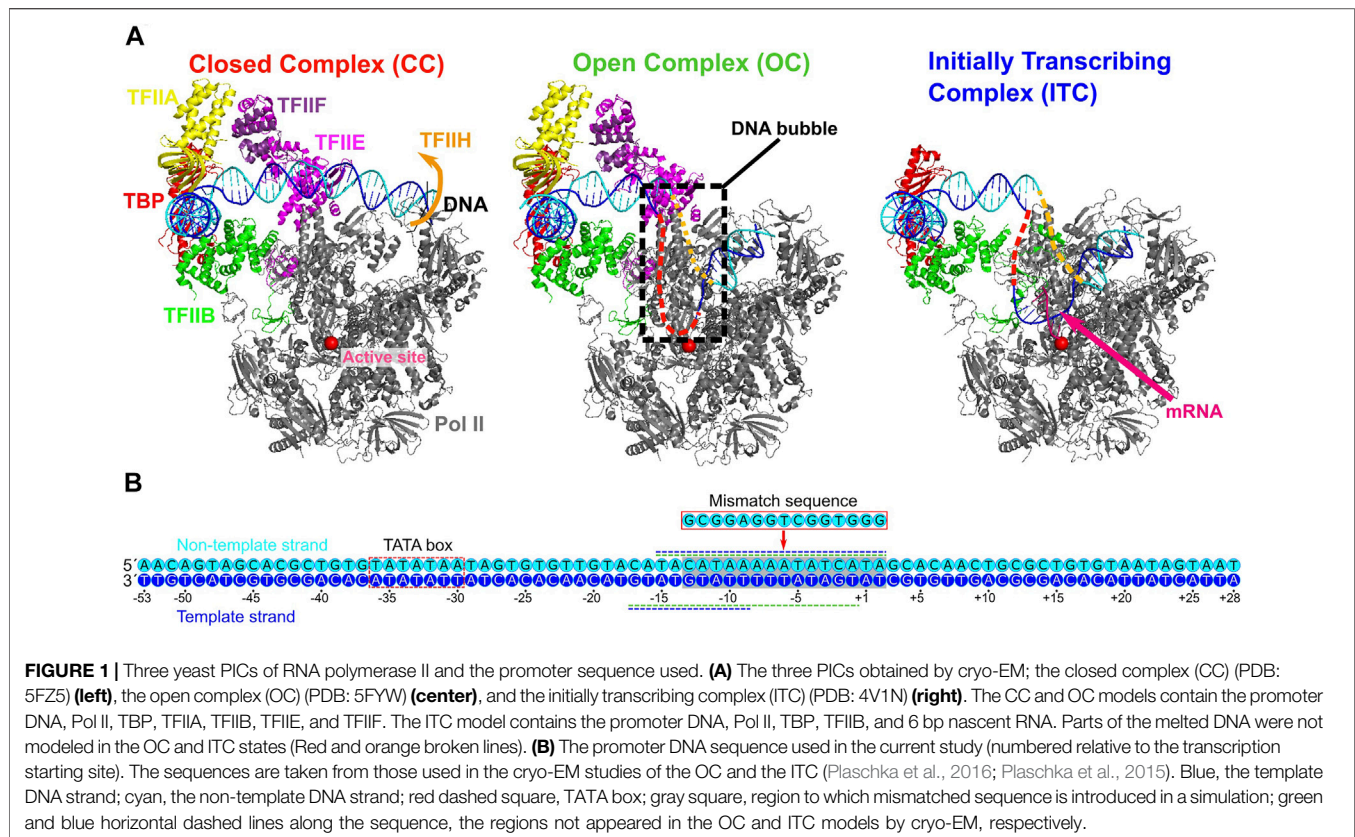
Recently, the molecular mechanisms of transcription initiation have been intensively studied. Especially, the cryo-electron microscopy revealed atomic structure details in key states in the eukaryotic transcription initiation. Yet, the dynamic processes of the promoter DNA opening in the pre-initiation complex remain obscured. In this study, based on the three cryo-electron microscopic yeast structures for the closed, open, and initially transcribing complexes, we performed multiscale molecular dynamics (MD) simulations to model structures and dynamic processes of DNA opening. Combining coarse-grained and all-atom MD simulations, we first obtained the atomic model for the DNA bubble in the open complexes. Then, in the MD simulation from the open to the initially transcribing complexes, we found a previously unidentified intermediate state which is formed by the bottleneck in the fork loop 1 of Pol II: The loop opening triggered the escape from the intermediate, serving as a gatekeeper of the promoter DNA opening. In the initially transcribing complex, the non-template DNA strand passes a groove made of the protrusion, the lobe, and the fork of Rpb2 subunit of Pol II, in which several positively charged and highly conserved residues exhibit key interactions to the non-template DNA strand. The back-mapped all-atom models provided further insights on atomistic interactions such as hydrogen bonding and can be used for future simulations.

Keywords: transcription, eukaryotes, protein-DNA complex, DNA opening, molecular dynamics simulation

INTRODUCTION

Transcription is fundamental to virtually all area of biology. In eukaryotic cells, RNA polymerase II (Pol II) transcribes all messenger RNAs, making it of central importance. The Pol II transcription initiation requires progressive assembly of several general transcription factors (TFs) and Pol II on the promoter DNA sequence, forming the pre-initiation complex (PIC). After the initial transcription of short RNAs, the transcription machinery escapes the promoter region converting its architecture for the transcription elongation. Much of the transcriptional regulations are related to these early stages of transcription and thus it is of utmost importance to understand the molecular mechanisms of the transcription initiation, which we focus in this study.

Overall processes in the Pol II transcription initiation have been characterized via decades of studies. The PIC consists of Pol II and six general TFs, TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, and TFIIF (Buratowski et al., 1989; Roeder, 1996; Grünberg and Hahn, 2013; Sainsbury et al., 2015). In addition, coactivators such as Mediator are involved in its regulation (Schilbach et al., 2017; Nozawa et al.,



2017). The initiation process begins with the recognition of the promoter DNA sequence by TFIID. For the promoter sequences that contain the classic TATA box, the TATA-binding protein (TBP) in TFIID binds the TATA box DNA sequence, leading to ~90-degree bend of the DNA. Then, TFIIA, TFIIB, Pol II-TFIIF complex assemble in this bent site. Further, TFIIE and TFIIH are recruited in order, to form the PIC with the bent duplex DNA (termed the closed complex, CC). In particular, PIC without TFIIH is called as core PIC (cPIC). Next, the promoter DNA melts into the template and non-template DNA strands, driven by the ATP-dependent translocase activity of TFIIH (termed the open complex, OC). The template DNA strand moves toward the active site of Pol II. The melted DNA region is called “DNA bubble”, of which size is experimentally characterized as ~6 bp in the OC state (Tomko et al., 2017). Subsequently, the DNA bubble expands to ~13 bp (Tomko et al., 2017), which allows the template DNA strand reaching to the active site to begin the messenger RNA synthesis. The complex in which the initial transcription begins is called the initially transcribing complex (ITC). Notably, while the ATP-dependent translocase activity of TFIIH facilitates the promoter DNA opening (Compe and Egly, 2012; Fishburn et al., 2015), some promoter DNAs can open spontaneously without TFIIH (Plaschka et al., 2016).

Recently, the cryo-electron microscopy (cryo-EM) revealed near-atomic structures in key stages of the Pol II transcription initiation (Schilbach et al., 2017; Nozawa et al., 2017; Plaschka et al., 2016), which provided the model of DNA opening process in the PICs (**Figure 1A**) (Plaschka et al., 2016). The model is

based on the yeast CC, OC (Plaschka et al., 2016), and ITC structures (Plaschka et al., 2015), and the highly conserved human CC structure (He et al., 2013). However, the state transitions from CC to OC, and to ITC were not directly observed. Moreover, the modeled structures of OC and ITC by cryo-EM do not contain parts of DNA strands because of high flexibility in the DNA bubble. Thus, how the template and non-template DNA strands behave inside Pol II has not been fully understood. Complementarily, the DNA bubble size in OC and ITC states has been detected via optical and magnetic tweezer experiments (Tomko et al., 2017; Fazal et al., 2015). However, the structural details in the DNA bubble is currently missing.

Given such situations, molecular dynamics (MD) simulations can offer another complementary approach to address the structural dynamics of the Pol II transcription initiation since MD simulations can provide high-resolution spatiotemporal information (Chen et al., 2010; Hyeon and Thirumalai, 2011; Feig and Burton, 2010; Huang et al., 2010; Silva et al., 2014). However, since the DNA opening process involves rather large-scale and slow movements of DNA within large complexes, conventional MD simulations with fully-atomic resolution (designated as the all-atom (AA) MD hereafter) cannot easily sample these structural dynamics. To circumvent this difficulty, one can alternatively use coarse-grained (CG) MD simulations, which can speed up the simulation by orders of magnitude at the cost of accuracy (Liwo et al., 2014; Takada et al., 2015; Kmiecik et al., 2016; Pak and Voth, 2018). Once comprehensively sampled by CG-MD, one can back-map the sampled CG structure models

into AA models, followed by AA-MD simulations (Shimizu and Takada, 2018). A recent study employed such a protocol to gain comprehensive and high-resolution energy landscape in a bacterial RNA polymerase (Unarta et al., 2021).

In this study, using the cryo-EM yeast structures for the CC, OC, and ITC, we performed multiscale MD simulations to model structures and dynamic processes of DNA opening.

Combining CG- and AA- MD simulations, we first obtained the atomic model for the DNA bubble in the OC. Then, in the CG-MD simulation from the OC to the ITC, we found a previously unidentified intermediate state which is formed by the bottleneck in the fork loop 1 of Pol II: The loop opening triggered the escape from the intermediate, serving as a gatekeeper of the promoter DNA opening. In the ITC, the non-template DNA strand passes a groove made of the protrusion, the lobe, and the fork of Rpb2 subunit of Pol II, in which several positively charged and highly conserved residues exhibit key interaction to the non-template DNA strand.

MATERIALS AND METHODS

Preparation of the Simulation System

We modeled the three yeast structures, CC, OC, and ITC, based on the cryo-EM structure models, 5FZ5 (Plaschka et al., 2016) and 6GYL (Dienemann et al., 2018) for CC, 5FYW (Plaschka et al., 2016) for OC, and 4VIN (Plaschka et al., 2015) for ITC. Missing residues in the original models were modeled by the software MODELLER (Webb and Sali, 2016; Marti-Renom et al., 2000; Sali and Blundell, 1993; Fiser et al., 2000).

We used the DNA sequence identical to that used in cryo-EM studies (Plaschka et al., 2016; Plaschka et al., 2015). The sequence was derived from the promoter sequence of HIS4 gene locus, from which 28 bp were deleted at the downstream of the TATA box.

Coarse-Grained MD Simulations

In this study, we applied the coarse-grained (CG) simulation model that has been developed previously and extensively applied to protein-DNA complex systems (Levy et al., 2007; Terakawa et al., 2012; Freeman et al., 2014; Zhang et al., 2016; Shimizu et al., 2016; Lequieu et al., 2017; Niina et al., 2017; Brandani et al., 2018; Tan and Takada, 2020). We used AICG2+ model for proteins, 3SPN.2 model for DNA (Li et al., 2014; Hinckley et al., 2013). Briefly, in AICG2+, each amino acid in proteins are represented by one CG particle placed at its C α position and the structure-based contact potential biases its energy landscape towards the reference structure. In 3SPN.2 model, each nucleotide is modeled by three CG particles corresponding to the phosphate, the sugar, and the base. Orientation-dependent potentials for base-base interactions and others are designed to reproduce basic experimentally-characterized properties of duplex and, to some extent, single strands. Between proteins and DNA, we applied the structure-based contact potential for representing the specific interactions, as well as a general excluded volume term and the electrostatic interaction via the Debye-Huckel approximation (the monovalent salt concentration was set to 200 mM

throughout this study). For the electrostatic interaction, we employed partial charges on the surface residues of proteins, which were optimized to reproduce the electrostatic potential around the protein obtained by the all-atom model via the RESPAC method (Terakawa and Takada, 2014). For time propagation, including the solvent effect implicitly, we employed a simple Langevin dynamics at the temperature 300 K. For all the CG-MD simulations, we used the software CafeMol 3.2 (Kenzaki et al., 2011).

The specific protein-DNA interaction, i.e., the structure-based contact potential is, as usual, expressed as

$$\sum_{ij \in \text{contact}} \epsilon_{go} \left[5 \left(\frac{r_{ij0}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{ij0}}{r_{ij}} \right)^{10} \right]$$

where r_{ij} is the distance between CG particles i and j , r_{ij0} is the corresponding distance at its reference structure, and ϵ_{go} is a parameter that modulates the strength of the interaction, of which value was calibrated to be 1.2 kcal/mol, to maintain experimentally characterized contacts in the three states, CC, OC, and ITC (Plaschka et al., 2016; Plaschka et al., 2015). Specifically, we set the following four conditions to be satisfied:

1. In the CC state, the contact between DNA and the E-wing of TFIIE is maintained.
2. In the OC state, the template DNA strand can maintain the native contacts with a region close to the active site of Pol II.
3. In the OC and ITC states, an upstream side of DNA maintains its contact with N50, K51, and T52 of TFIIE.
4. In the OC and ITC states, the triple mutations N50E, K51E, and T52E lead to loss of the DNA-TFIIE contacts (Plaschka et al., 2016).

For the specific protein-DNA interaction, we collected protein-DNA contacts in the three complexes structures and used its union set for the structure-based contact potential. This union set includes the particular contacts satisfying above four conditions.

Trajectory Analysis

The state-to-state transitions were characterized by protein-DNA contacts that depend on the state. There are 92, 22, and 118 contacts between proteins and DNA in CC, OC, and ITC states, respectively. The contacts in CC are all specific to the CC state and are not shared with the other two states. Thus, all the 92 contacts are used to characterize the CC state. The contacts in OC are mostly a subset of the contacts in the ITC state (19 out of 22 included in the ITC contacts). We used all the 22 contacts to characterize the OC state. Of the 118 contacts in ITC states, 99 are unique to the ITC state and thus are used to characterize the ITC state. Once the sets of contacts are defined, we quantify the state-to-state transition by the fraction of contacts formed in each snapshot.

The size of the DNA bubble was estimated as the sum of the broken base pairs, which are defined by the distance between the CG particles of the base pairs larger than 6.2 Å: In preliminary CG-MD simulations of duplex DNA at the same solvent

condition, the probability that the base-base distance is larger than 6.2 Å was 0.3%. In apparently melted DNA configurations, their base-base distances were almost surely larger than this threshold distance.

Back-Mapping to All-Atom Model and All-Atom MD Simulations

Following the previously developed protocol (Shimizu and Takada, 2018), we performed the back-mapping from our CG models to all-atom models. While we used the cryo-EM-based CC protein structures as the reference structures of all the three states in the dynamical modeling, we moved them back to the respective cryo-EM protein structures aiming at more accurate modeling of all-atom structures. For the intermediate state I₂, we used the OC structure as the reference. For each state, we began with the CG-MD simulation at 300 K for 10⁵ MD steps. Then, to reduce local fluctuations, we performed a quick annealing simulation, quenching the temperature from 300 to 1 K, followed by a 10⁵ MD step simulation at 1 K. The final structure was put into the back-mapping toolset. For DNA, we applied the CG to AA reconstruction tool (Shimizu and Takada, 2018), whereas for proteins, we employed the PD2 ca2main (Moore et al., 2013) for backbone and SCWRL4 (Krivov et al., 2009) for sidechain reconstruction.

Once the all-atom model for the PIC were obtained, we performed all-atom MD simulations using the software GROMACS 2020.2 (Abraham et al., 2015) with the protein, DNA, and water force fields, ff14SB (Maier et al., 2015), and parmbsc1 (Ivani et al., 2016), and TIP3P (Jorgensen et al., 1983), respectively. We used the standard protocol: We set the box size of 182.2 × 232.1 × 186.4 Å³ solvating with water molecules and 171 Na⁺ ions to neutralize the system. After the energy minimization, we equilibrated the local system with NVT and then NPT ensembles (T = 300 [K], the pressure 1 bar), followed by 10 ns MD simulations. We used the cutoff distance of 1 nm for the Coulomb interaction with the particle-mesh-Ewald for long range treatment.

RESULTS

Multiscale Modeling of Pre-Initiation Complexes

Our multiscale modeling begins with CG-MD simulations that connect the three states of the PIC; the CC, OC, and ITC. The constructed CG models were then back-mapped to AA models, which is followed by short-time MD simulations with the AA models.

We employ the CG model that has been extensively used to protein-DNA complexes (Levy et al., 2007; Terakawa et al., 2012; Freeman et al., 2014; Zhang et al., 2016; Shimizu et al., 2016; Lequieu et al., 2017; Niina et al., 2017; Brandani et al., 2018; Tan and Takada, 2020). In the CG model, each amino acid in proteins is represented as one particle located at the Ca atom position and each nucleotide in DNA is modeled by three particles each representing the phosphate, the sugar, and the base. The

protein energy function AICG2+ contains the contact potentials that stabilizes the predefined reference (native) structure, i.e., the structure-based model (Li et al., 2012). The DNA energy function 3SPN.2 is empirically tuned to reproduce several experimental data such as the sequence-dependent melting temperature and bending modulus (Hinckley et al., 2013). The protein-DNA energy function consists of the generic terms; the short-range repulsion and the electrostatic interaction, and the specific interactions; structure-based contact potentials (See *Materials and Methods* for more details).

The simulation system consists of an 81-bp promoter DNA (the sequence shown in **Figure 1B**) and the protein complex that contains Pol II, TBP, TFIIA, TFIIB, TFIIE, and TFIIF (that is, this study deals with cPIC). TFIIF is not included because the DNA bubble can form without TFIIF (Plaschka et al., 2016; Alekseev et al., 2017) and because the structural information on ATP-dependent conformational change in TFIIF is incomplete albeit some structures previously reported (Schilbach et al., 2017; Dienemann et al., 2018; Osman and Cramer, 2020).

The root-mean-square-differences (RMSD) of protein complexes were 0.85 Å between CC and OC, and 3.5 Å between OC and ITC, which are smaller than the resolutions reported in the cryo-EM analysis (8.8, 4.4, and 7.8 Å, for CC, OC, and ITC models, respectively). We note that we excluded DNA in the calculations of these RMSDs. In the CG-MD simulations, these modest-sized structure changes in proteins should appear via the interaction to DNA (and a short RNA in the case of ITC). Since CC contains the weakest protein-DNA interaction among the three complexes structure models, we took the protein structure of CC as a reference structure of the protein complex in the CG model throughout this study.

Modeling the DNA Bubble in the Open Complex

First, to obtain the OC model with the open DNA, we performed 40 independent CG-MD simulations of 5 × 10⁶ MD steps, starting from the CC structure (**Supplementary Figure S1**). In any of the simulations, the promoter DNA did not melt spontaneously and most of the OC specific protein-DNA contact did not appear although the particular region of the promoter (−18~ +7 relative to the transcription start site (TSS)) was distorted toward the cleft of Pol II (**Supplementary Figure S1**). This distortion in DNA was observed in a previous cryo-EM study, which indicates the pre-stage of DNA opening (Dienemann et al., 2018). A previous study shows that the DNA opening in the absence of TFIIF takes a very long time; the real-time observation of the formation of the DNA bubble shows that it takes a few seconds (Fazal et al., 2015). Therefore, it is reasonable that we did not observe spontaneous DNA opening in our CG-MD simulations that cannot cover second time scales.

Then, to enforce the prompt DNA opening, we modified the non-template DNA sequence to introduce the DNA mismatch of 15 bases into the promoter (**Figure 1B**). The introduced mismatch is identical to that used for the cryo-EM structures of OC and ITC (Plaschka et al., 2016; Plaschka et al., 2015). Under this condition, we performed 40-independent CG-MD

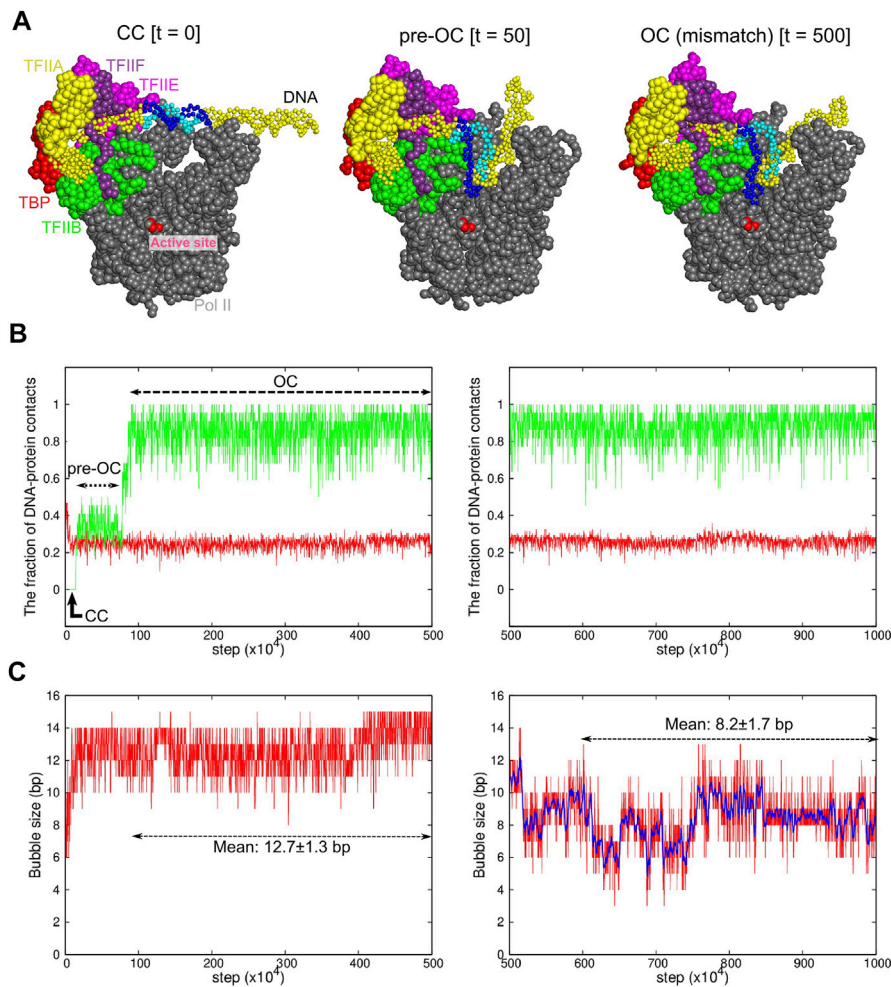
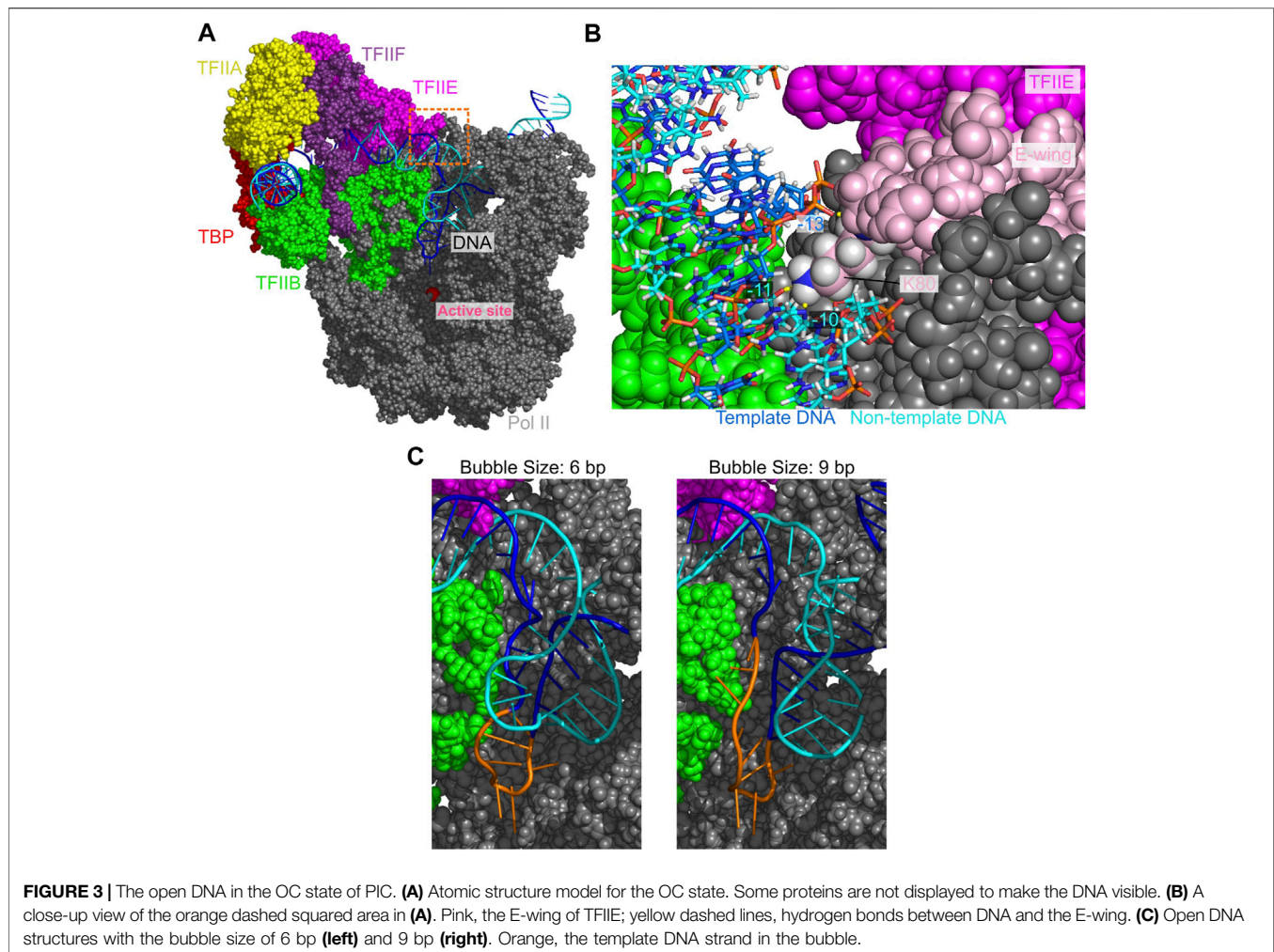


FIGURE 2 | Coarse-grained MD simulation for the transition from the CC to OC states. Results of a representative trajectory are shown. **(A)** Snapshots at 0 MD step (left, the CC state), at 50×10^4 MD steps (center, the pre-OC state), and at 500×10^4 MD steps (right, the OC state with the mismatch). Some proteins are not displayed to make DNA visible. The same colors are used as **Figure 1A**. Blue and cyan region in DNA indicates the 15-bp mismatch region, forming the DNA bubble. **(B)** The time course of the fractions of protein-DNA contacts specific to CC (red) and OC (green). **(C)** The time course of the DNA bubble size. In **(B,C)**, the left/right panels are from the first/second halves of MD simulations with/without the DNA mismatch. The blue curve in the right panel in **(C)** shows a moving average over 11 points.

simulations of 2×10^7 MD steps (a representative trajectory is shown in **Figure 2**). **Figure 2B**, left panel depicts a representative time course of the fraction of protein-DNA contacts specific to CC (red) and to OC (green). In this trajectory, in the very initial phase, $\sim 80\%$ of the CC-specific contacts were lost, whereas $\sim 40\%$ of the OC-specific contacts were formed to reach an intermediate state, which we call “pre-OC” state (**Figure 2A**, center). In the pre-OC state, most of the mismatched DNA region melted to form a bubble of ~ 13 bp (**Figure 2C**). This caused the +2 site of the template DNA strand to form new contacts with Pol II (**Supplementary Movie S1**). All the 40 trajectories paused at this pre-OC state (40/40 cases). In the trajectory in **Figure 2B**, left panel, the template DNA strand jumped further down to the active site at $\sim 0.9 \times 10^6$ MD steps, reaching to the OC-like state with the mismatch (**Figure 2A**, right, **Figure 2B**) (22/40 cases). In the 22 cases, we observed that the complex further moved

towards the ITC state in five cases). The transition was driven by the new contact formation of the +1 site of the template DNA strand with Pol II (**Supplementary Movie S1**).

To obtain the OC structure model without the DNA mismatch, for obtained OC-like structures with the mismatch, we changed the DNA sequence back to the original sequence without mismatch, followed by 5×10^6 MD steps CG-MD simulations. A representative trajectory is depicted in the right panels of **Figures 2B,C**. While the overall positioning of DNA did not change (**Figure 2B**, right), part of the melted DNA regained the base pairing during the trajectory (**Figure 2C**, right). The observed bubble size fluctuated in time in the range of 6–10 bp, with the mean and the standard deviation 8.2 ± 1.7 bp. In the 22 cases, we did not see significant difference in the fraction of DNA-protein contacts and the DNA bubble size (Another trajectory shown in **Supplementary Figure S2**). Notably, we observed that



the bubble size depends on the promoter sequence to some extent; with the promoter sequence used in the single-molecule magnetic tweezer experiment, our simulation resulted in the bubble size of 5.5 ± 1.4 bp, which is fairly compared with the experimental estimate, 6.1 ± 0.3 bp (Tomko et al., 2017).

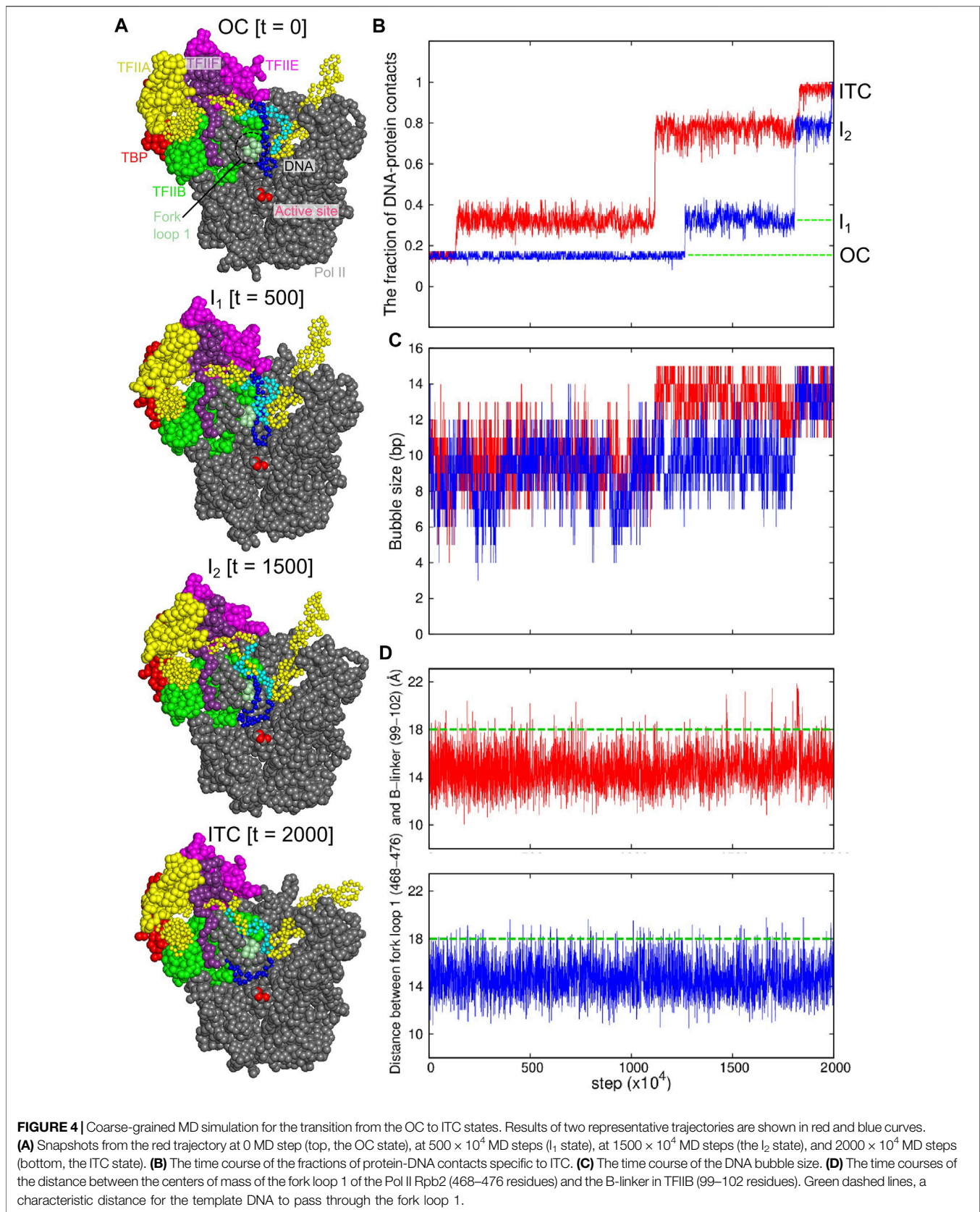
To detect protein-DNA interactions in the OC state at atomic level, we modeled all-atom structures via back-mapping from the snapshots of a CG-MD trajectory with the DNA bubble sizes of 6 and 9 bp. The obtained all-atom models were further relaxed/refined by 10 ns MD simulations with explicit water solvent (**Figure 3**). In the upstream of the DNA bubble, we find that the hydrogen bonds of the TFIIE E-wing residues K80 with the non-template DNA at -11 to -10 sites, and with the template DNA at -13 site, which are present in the CC state, are maintained (**Figure 3B**). These interactions are suggested to facilitate the promoter opening and contribute to the efficiency of transcription initiation (Forget et al., 2004). Comparing the structures with 6 and 9 bp DNA bubbles, we find that the template DNA strand is rather similar each other while the non-template DNA strand is more mobile. The 6 bp in the downstream side (from -4 to $+2$ sites) were melted in both

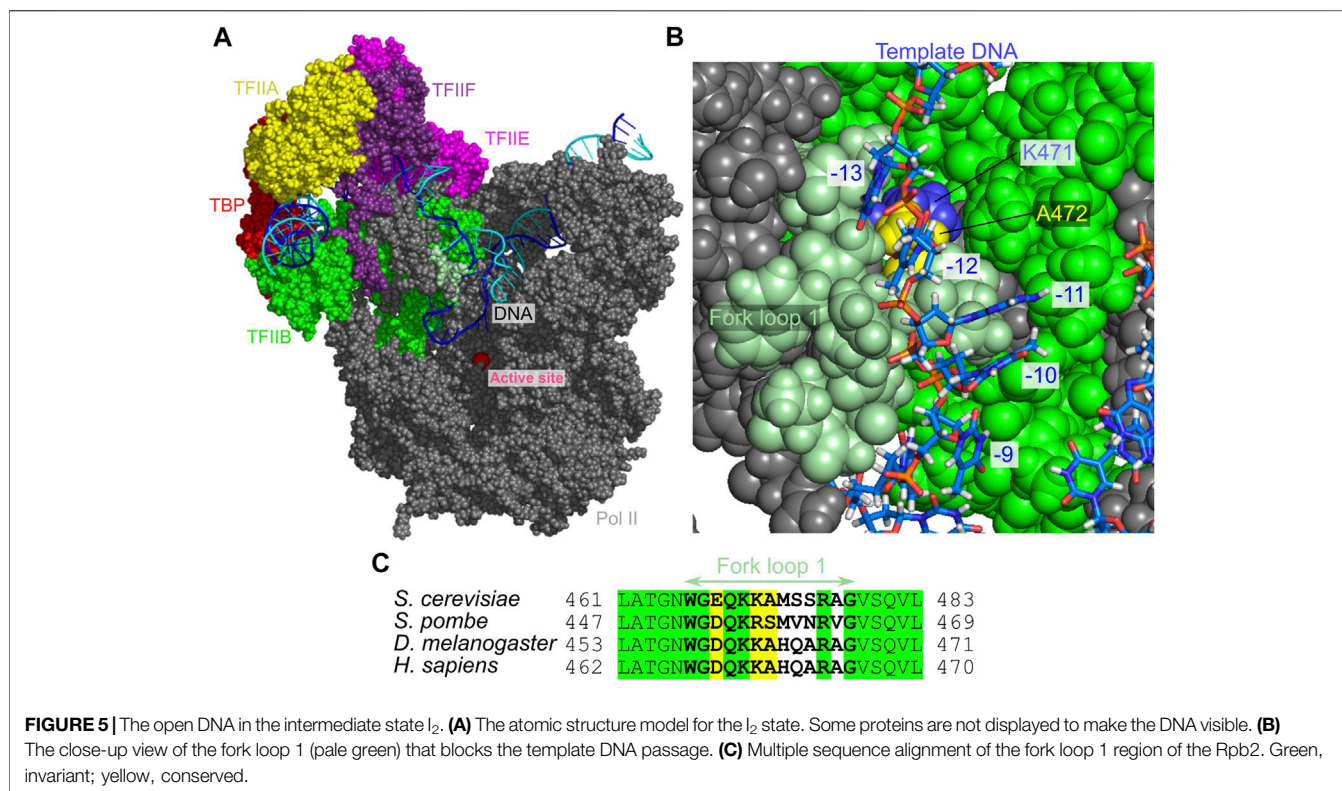
structures, while the 3 bp in the upstream side (from -7 to -5 sites) were formed/melted in the 6 bp/9 bp DNA bubbles.

Dynamical Modeling of the Transition from the Open Complex to the Initially Transcribing Complex

Next, we addressed dynamic process of the transition from the OC state to the ITC state. Starting from a final snapshot of the previous simulation that paused at the OC state for the promoter DNA without the DNA mismatch, we performed 140 independent CG-MD simulations of 2×10^7 MD steps (**Figure 4**). In most trajectories (130/140 cases), the contact between the promoter DNA (from -16 to -9 sites) and the E-wing of TFIIE persisted for the whole simulation time, which clearly precluded the template DNA strand from accessing the active site. Only in 10 cases, we observed the disruption of this contact, which directly triggered the template DNA strand to move down towards the active site (**Figure 4**). 9 out of these 10 trajectories reached the ITC state.

In these successful trajectories, we found two intermediate states (**Figure 4A** the second and the third models) before





reaching the ITC state (Figure 4A, bottom). In a representative trajectory (red in Figures 4B–D), the first transition occurred at $\sim 1 \times 10^6$ MD steps, after which about 35% of the ITC specific contacts were formed. In this intermediate state I_1 , ~ 4 bp of the template DNA strand, ($-2 \sim +2$ sites, relative to the TSS) approached the active site, while the contact between the DNA and the TFIIE E-wing is maintained (Figure 4A the second structure). After a long waiting time, the DNA was detached from the TFIIE E-wing region (at 1.1×10^7 MD steps in the red trajectory), followed by the motion of the entire DNA bubble towards the active site (Figure 4A the third structure). However, the template strand DNA in the upstream side of the DNA bubble, $-13 \sim -9$ sites, collides with the fork loop 1 of Rpb2 subunit of Pol II (Figure 4A, the third structure, and Supplementary Figure S4, left). This forms a metastable intermediate state I_2 . After some duration time at this intermediate state, the complex made the final transition to the ITC state (at $\sim 1.7 \times 10^7$ MD steps in the red trajectory). The other successful trajectories followed similar pathways.

To increase the samples of transitions to the ITC state, we performed 160 extra-simulations of 5×10^6 MD steps in which the contact between DNA and the TFIIE Ewing was weakened intentionally (see Materials and Methods; Supplementary Figure S3). In this setup, we observed the successful transition to the ITC state for 102 out of 160 cases with the transition pathway unchanged. The rest of trajectories stayed at the intermediate state I_2 until the end of trajectories (58/160 cases).

In the ITC state, the DNA bubble size was, on average, 13.4 ± 1.1 bp (Figure 4C; Supplementary Figure S3B), which perfectly agrees with the previous estimate (13.4 bp) (Tomko et al., 2017).

Fork Loop 1 Serve as a Gatekeeper

The intermediate state I_2 appears because of the blockage by the fork loop 1, which led us to hypothesize that the fork loop 1 may serve as a gatekeeper. To monitor large-scale motions of the fork loop 1, we plotted in Figure 4D the time courses of the distance between the fork loop 1 and the B-linker of TFIIB, finding that the fork loop 1 exhibits intermittent large-scale fluctuation to open the gate (green dashed lines in Figure 4D). In the representative time course (red), the time of the transition from I_2 to ITC states in Figure 4B coincides with a large-scale opening. Looking into structure changes at the time, we found that the template DNA strand passed the fork loop 1 upon the loop opening, and moved toward the active site (Figure 4A; Supplementary Figure S4, right). Notably, in any trajectory, the non-template DNA strand never passed the fork loop 1. Instead, the non-template DNA strand approached to the wall of Pol II. Therefore, after the passage of the template strand, the fork loop 1 is located inside the DNA bubble. This support the hypothesis that the fork loop 1 serve as a gatekeeper; it is only passed by the template, but not the non-template DNA strand. This role is supported by previous studies (Plaschka et al., 2016; Meyer et al., 2006; Meyer et al., 2009).

The fork loop 1 sequence is fairly well conserved from yeast to human (Figure 5). In the human Pol II, it has been reported that a mutant that deletes two residues in the fork loop 1 (K458, A459 in human Pol II, which align with K471, A472 in yeast Pol II) abolishes the transcription *in vitro* (Jeronimo et al., 2004). This supports the crucial role of the fork loop 1. The mutation may

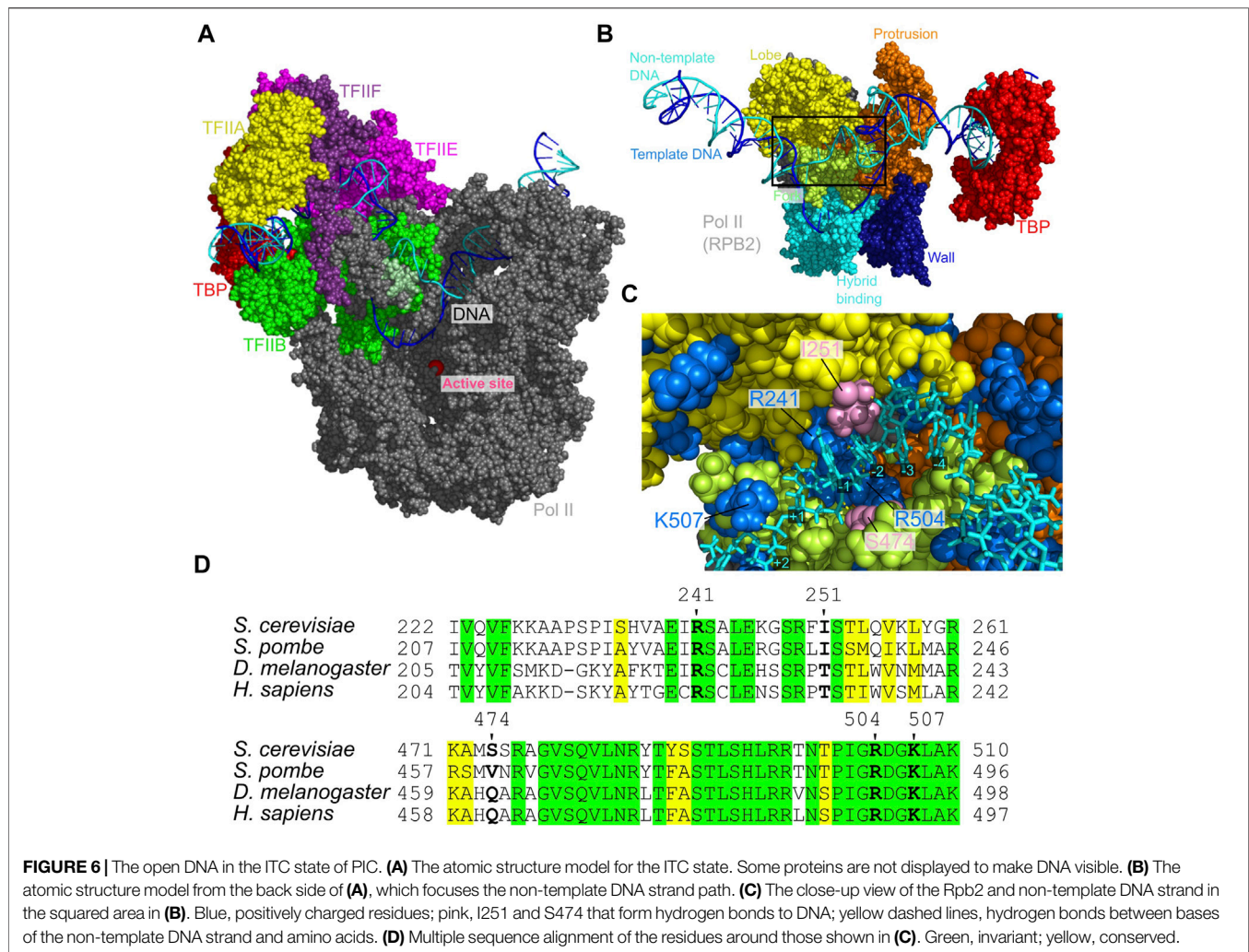


FIGURE 6 | The open DNA in the ITC state of PIC. **(A)** The atomic structure model for the ITC state. Some proteins are not displayed to make DNA visible. **(B)** The atomic structure model from the back side of **(A)**, which focuses the non-template DNA strand path. **(C)** The close-up view of the Rpb2 and non-template DNA strand in the squared area in **(B)**. Blue, positively charged residues; pink, I251 and S474 that form hydrogen bonds to DNA; yellow dashed lines, hydrogen bonds between bases of the non-template DNA strand and amino acids. **(D)** Multiple sequence alignment of the residues around those shown in **(C)**. Green, invariant; yellow, conserved.

alter the loop opening dynamics, which led to the malfunction of Pol II.

The Template and Non-Template DNA Strands in the Initially Transcribing Complex

To predict the placement of the template and non-template DNA strands and probe protein-DNA interactions in the ITC state, we constructed all-atom structure model via the back-mapping from snapshots in a CG-MD trajectory (Figure 6A), which is followed by 10 ns AA-MD simulations. We note that the template DNA strand was anticipated to be in the wall from the cryo-EM study even though the cryo-EM structure model for the ITC state does not contain the segment of the template and non-template DNA strands (Plaschka et al., 2016). The constructed model in this work supports this prediction; the template DNA strand is indeed placed in the wall of Pol II (Supplementary Figure S4, right). More specifically, our model suggests that the non-template DNA strand is localized at the protrusion, the lobe, and the fork of RPB2 subunit of Pol II (Figures 6B,C). This placement of the non-template DNA strand is fairly close

to that found in the yeast elongation complex structures solved by X-ray diffraction (Supplementary Figure S5) (Barnes et al., 2015). These regions form the groove with many basic amino acids (Figure 6C). Along the groove the three basic residues, R241, R504, and K507 made specific interactions to the DNA at -1, +1, and +2 sites (Figure 6C). These three residues are strictly conserved across broad range of eukaryotes (Figure 6D).

DISCUSSION

Our computational modeling revealed that, after passing the OC state, the PIC passes two intermediate states before reaching the ITC, through which a small DNA bubble in the OC is expanded to complete the DNA bubble ready for RNA synthesis. One key gating state is I_2 , where the upstream part of the template DNA strand (-9 to -13 sites) interacts with the fork loop 1. The fluctuation of fork loop 1 was obligatory to escape from I_2 that leads to engaging the template DNA into the active center at ITC (Figures 4, 5). The non-template DNA strand did not pass

the fork loop 1, suggesting that the fork loop 1 serves as the gatekeeper for the DNA bubble.

Previous studies implicated two critical roles of fork loop 1. Based on the structural change in the fork loop 1 between the nucleic-acid free state and in the transcribing state (Meyer et al., 2006; Meyer et al., 2009) as well as the mutation assays (cite), the fork loop 1 was considered to play an important role in the transcription initiation. Alternatively, since the fork loop 1 is located at around the terminus of the DNA/RNA hybrid in the elongation complex, it may have important roles in separation of the product RNA from the template DNA strand. Our current simulation clearly supports the former functional role. The fork loop 1 forms a gate together with the rudder of RPB1 subunit in Pol II and serves as a gatekeeper for the engagement of the template DNA strand, but not the non-template DNA strand. The structural model obtained can be used to guess key residues as the gatekeeper, which can be examined by mutagenesis experiments. Furthermore, the shapes of the DNA transcription bubble of the intermediates I_1 and I_2 are different from those of OC and ITC. Especially, it will be interesting to investigate the structure change of the template strand (e.g., the opening of the DNA transcription bubble upstream) experimentally, for example, by the FRET technology.

Moreover, the current study found that non-template DNA strand in ITC is localized in the groove, formed by the protrusion, the lobe, and the fork of RPB2 subunit (Figure 6). This pass is close to the non-template DNA strand pass in the yeast elongation complex (Barnes et al., 2015), suggesting its ubiquitous importance. However, to our knowledge, these interaction sites were not investigated by mutagenesis. Systematic mutation assays in these sites would clarify the roles of stabilizing the non-template DNA strand in the transcription process.

In this study, we only mentioned the formation of DNA transcription bubble and did not discuss the initial transcription proceed by RNA polymerase. The ITC modeled in this study is a state in which transcription has not yet occurred, followed by the scanning of the transcription-start-site, early RNA transcription, and the promoter escape. It has been proposed that the initial transcription proceeds in prokaryotes via the “scrunching” model (Kapanidis et al., 2006; Winkelman and Gourse, 2017), but it is unclear whether the same is true for eukaryotic transcription.

Obviously missing in the current work is the kinetic and energetic arguments on the very initial process of the DNA opening in the transition from the CC to the OC states. In this study, even with the use of CG-MD simulations, the DNA opening was too slow to be simulated directly in MD simulations for the native promoter sequence. Instead, we needed to introduce a mismatch sequence in the promoter region. This is clearly a limitation. To study kinetic and energetic aspects in

this initial DNA opening without the mismatch sequence, we need some advanced sampling methods, such as the umbrella sampling, the Markov-state modeling (Husic and Pande, 2018), and the string method (Weinan et al., 2002). Alternatively, since the ATP-driven motor activity of TFIIF helicase is expected to accelerate the DNA opening, including this effect either explicitly or implicitly may enable to simulate the dynamic process of the bubble formation more directly. These developments are left for future studies.

Related to this, it has been known that Pol I and Pol III systems do not have TFIIF-like helicases (Vannini and Cramer, 2012; Paule and White, 2000; Han et al., 2017; Gouge et al., 2017), yet initiating the transcription efficiently via similar three states (Engel et al., 2017; Abascal-Palacios et al., 2018; Vorlander et al., 2018). Comparison of the transcription initiation in the three RNA polymerase systems can put forward comprehensive understanding of transcription initiation in Eukaryote.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

GS and ST contributed to conception and design of the study. GS performed simulations and analyzed data. GS and ST wrote the draft of the manuscript, contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the Ministry of Education, Culture, Sport, Science, and Technology (MEXT) grant JPMXP1020200101 as “Program for Promoting Researches on the Supercomputer Fugaku” (ST), by the Japan Science and Technology Agency (JST) grant (JPMJCR1762) (ST), and by Japan Society for the Promotion of Science (Award number(s): 20H05934, 21H02441).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.772486/full#supplementary-material>

REFERENCES

- Abascal-Palacios, G., Ramsay, E. P., Beuron, F., Morris, E., and Vannini, A. (2018). Structural Basis of RNA Polymerase III Transcription Initiation. *Nature* 553, 301–306. doi:10.1038/nature25441
- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., and Hess, B. (2015). Gromacs: High Performance Molecular Simulations through Multi-level Parallelism from Laptops to Supercomputers. *SoftwareX* 1–2, 19–25. doi:10.1016/j.softx.2015.06.001
- Alekseev, S., Nagy, Z., Sandoz, J., Weiss, A., Egly, J.-M., Le May, N., et al. (2017). Transcription without XPB Establishes a Unified Helicase-independent

- Mechanism of Promoter Opening in Eukaryotic Gene Expression. *Mol. Cell* 65, 504–514. doi:10.1016/j.molcel.2017.01.012
- Barnes, C. O., Calero, M., Malik, I., Graham, B. W., Spahr, H., Lin, G., et al. (2015). Crystal Structure of a Transcribing RNA Polymerase II Complex Reveals a Complete Transcription Bubble. *Mol. Cell* 59, 258–269. doi:10.1016/j.molcel.2015.06.034
- Brandani, G. B., Niina, T., Tan, C., and Takada, S. (2018). DNA Sliding in Nucleosomes via Twist Defect Propagation Revealed by Molecular Simulations. *Nucleic Acids Reserach* 46 (6), 2788–2801. doi:10.1093/nar/gky158
- Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989). Five Intermediate Complexes in Transcription Initiation by RNA Polymerase II. *Cell* 56 (4), 549–561. doi:10.1016/0092-8674(89)90578-3
- Chen, J., Darst, S. A., and Thirumalai, D. (2010). Promoter Melting Triggered by Bacterial RNA Polymerase Occurs in Three Steps. *Proc. Natl. Acad. Sci. USA* 107 (28), 12523–12528. doi:10.1073/pnas.1003533107
- Compe, E., and Egly, J.-M. (2012). TFIIF: when Transcription Met DNA Repair. *Nat. Rev. Mol. Cell Biol* 13 (6), 343–354. doi:10.1038/nrm3350
- Dienemann, C., Schwalb, B., Schilbach, S., and Cramer, P. (2018). Promoter Distortion and Opening in the RNA Polymerase II Cleft. *Mol. Cell* 73, 97–e4. doi:10.1016/j.molcel.2018.10.014
- Engel, C., Gubbe, T., Neyer, S., Sainsbury, S., Oberthuer, C., Baejen, C., et al. (2017). Structural Basis of RNA Polymerase I Transcription Initiation. *Cell* 169 (1), 120–131. doi:10.1016/j.cell.2017.03.003
- Fazal, F. M., Meng, C. A., Murakami, K., Kornberg, R. D., and Block, S. M. (2015). Real-time Observation of the Initiation of RNA Polymerase II Transcription. *Nature* 525 (10), 274–277. doi:10.1038/nature14882
- Feig, M., and Burton, Z. F. (2010). RNA Polymerase II with Open and Closed Trigger Loops: Active Site Dynamics and Nucleic Acid Translocation. *Biophys. J.* 9 (8), 577–2586. doi:10.1016/j.bpj.2010.08.010
- Fiser, A., Do, R. K. G., and Šali, A. (2000). Modeling of Loops in Protein Structures. *Protein Sci.* 9, 1753–1773. doi:10.1110/ps.9.9.1753
- Fishburn, J., Tomko, E., Galburt, E., and Hahn, S. (2015). Double-stranded DNA Translocase Activity of Transcription Factor TFIIF and the Mechanism of RNA Polymerase II Open Complex Formation. *Proc. Natl. Acad. Sci. USA* 112 (13), 3961–3966. doi:10.1073/pnas.1417709112
- Forget, D., Langelier, M.-F., The'rien, C., Trinh, V., and Coulombe, B. (2004). Photo-Cross-Linking of a Purified Preinitiation Complex Reveals Central Roles for the RNA Polymerase II Mobile Clamp and TFIIE in Initiation Mechanisms. *Mol. Cell Biol* 24 (3), 1122–1131. doi:10.1128/mcb.24.3.1122-1131.2004
- Freeman, G. S., Lequieu, J. P., Hinckley, D. M., Whitmer, J. K., and de Pablo, J. J. (2014). DNA Shape Dominates Sequence Affinity in Nucleosome Formation. *Phys. Rev. Lett.* 113, 168101. doi:10.1103/physrevlett.113.168101
- Gouge, J., Guthertz, N., Kramm, K., Dergai, O., Abascal-Palacios, G., Satia, K., et al. (2017). Molecular Mechanisms of Bdp1 in TFIIF Assembly and RNA Polymerase III Transcription Initiation. *Nat. Commun.* 8, 130. doi:10.1038/s41467-017-00126-1
- Grünberg, S., and Hahn, S. (2013). Structural Insights into Transcription Initiation by RNA Polymerase II. *Trends Biochem. Sci.* 38 (12), 603–611. doi:10.1016/j.tibs.2013.09.002
- Han, Y., Yan, C., Nguyen, T. H. D., Jackobel, A. J., Ivanov, I., Knutson, B. A., et al. (2017). Structural Mechanism of ATP-independent Transcription Initiation by RNA Polymerase I. *eLife* 6, e27414. doi:10.7554/eLife.27414
- He, Y., Fang, J., Taatjes, D. J., and Nogales, E. (2013). Structural Visualization of Key Steps in Human Transcription Initiation. *Nature* 495, 481–486. doi:10.1038/nature11991
- Hinckley, D. M., Freeman, G. S., Whitmer, J. K., and de Pablo, J. J. (2013). An Experimentally-Informed Coarse-Grained 3-Site-Per-Nucleotide Model of DNA: Structure, Thermodynamics, and Dynamics of Hybridization. *J. Chem. Phys.* 139, 144903. doi:10.1063/1.4822042
- Huang, X., Wang, D., Weiss, D. R., Bushnell, D. A., Kornberg, R. D., and Levitt, M. (2010). RNA Polymerase II Trigger Loop Residues Stabilize and Position the Incoming Nucleotide Triphosphate in Transcription. *Proc. Natl. Acad. Sci.* 107 (36), 15745–15750. doi:10.1073/pnas.1009898107
- Husic, B. E., and Pande, V. S. (2018). Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* 140 (7), 2386–2396. doi:10.1021/jacs.7b12191
- Hyeon, C., and Thirumalai, D. (2011). Capturing the Essence of Folding and Functions of Biomolecules Using Coarse-Grained Models. *Nat. Commun.* 2, 487. doi:10.1038/ncomms1481
- Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., et al. (2016). Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* 13, 55–58. doi:10.1038/nmeth.3658
- Jeronimo, C., Langelier, M.-F., Zeghouf, M., Cojocar, M., Bergeron, D., Baali, D., et al. (2004). RPA1, a Novel Human RNA Polymerase II-Associated Protein Affinity Purified with Recombinant Wild-type and Mutated Polymerase Subunits. *Mol. Cell Biol* 24 (16), 7043–7058. doi:10.1128/mcb.24.16.7043-7058.2004
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79, 926–935. doi:10.1063/1.445869
- Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S., and Ebricht, R. H. (2006). Initial Transcription by RNA Polymerase Proceeds through a DNA-Scrunching Mechanism. *Science* 314 (5802), 1144–1147. doi:10.1126/science.1131399
- Kenzaki, H., Koga, N., Hori, N., Kanada, R., Li, W., Okazaki, K., et al. (2011). CafeMol: A Coarse-Grained Biomolecular Simulator for Simulating Proteins at Work. *J. Chem. Theor. Comput.* 7 (6), 1979–1989. doi:10.1021/ct2001045
- Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016). Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* 116 (14), 7898–7936. doi:10.1021/acs.chemrev.6b00163
- Krivov, G. G., Shapovalov, M. V., and Dunbrack, R. L., Jr. (2009). Improved Prediction of Protein Side-Chain Conformations with SCWRL4. *Proteins* 77 (4), 778–795. doi:10.1002/prot.22488
- Lequieu, J., Schwartz, D. C., and de Pablo, J. J. (2017). In Silico evidence for Sequence-dependent Nucleosome Sliding. *Proc. Natl. Acad. Sci. USA* 114 (44), E9197–E9205. doi:10.1073/pnas.1705685114
- Levy, Y., Onuchic, J. N., and Wolynes, P. G. (2007). Fly-Casting in Protein–DNA Binding: Frustration between Protein Folding and Electrostatics Facilitates Target Recognition. *J. Am. Chem. Soc.* 129 (4), 738–739. doi:10.1021/ja065531n
- Li, W., Wang, W., and Takada, S. (2014). Energy Landscape Views for Interplays Among Folding, Binding, and Allostery of Calmodulin Domains. *Proc. Natl. Acad. Sci. U S A* 111 (9), 10550–10555. doi:10.1073/pnas.1402768111
- Li, W., Terakawa, T., Wang, W., and Takada, S. (2012). Energy Landscape and Multiroute Folding of Topologically Complex Proteins Adenylate Kinase and 2ouf-Knot. *Proc. Natl. Acad. Sci.* 109 (44), 17789–17794. doi:10.1073/pnas.1201807109
- Liwo, A., Baranowski, M., Czaplowski, C., Golaś, E., He, Y., Jagieła, D., et al. (2014). A Unified Coarse-Grained Model of Biological Macromolecules Based on Mean-Field Multipole-Multipole Interactions. *J. Mol. Model.* 20, 2306. doi:10.1007/s00894-014-2306-5
- Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theor. Comput.* 11, 3696–3713. doi:10.1021/acs.jctc.5b00255
- Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Šali, A. (2000). Comparative Protein Structure Modeling of Genes and Genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291–325. doi:10.1146/annurev.biophys.29.1.291
- Meyer, P. A., Ye, P., Suh, M.-H., Zhang, M., and Fu, J. (2009). Structure of the 12-Subunit RNA Polymerase II Refined with the Aid of Anomalous Diffraction Data. *J. Biol. Chem.* 284 (19), 12933–12939. doi:10.1074/jbc.m809199200
- Meyer, P. A., Ye, P., Zhang, M., Suh, M.-H., and Fu, J. (2006). Phasing RNA Polymerase II Using Intrinsically Bound Zn Atoms: An Updated Structural Model. *Structure* 14, 973–982. doi:10.1016/j.str.2006.04.003
- Moore, B. L., Kelley, L. A., Barber, J., Murray, J. W., and MacDonald, J. T. (2013). High-quality Protein Backbone Reconstruction from Alpha Carbons Using Gaussian Mixture Models. *J. Comput. Chem.* 34 (22), 1881–1889. doi:10.1002/jcc.23330
- Niina, T., Brandani, G. B., Tan, C., and Takada, S. (2017). Sequence-dependent Nucleosome Sliding in Rotation-Coupled and Uncoupled Modes Revealed by Molecular Simulations. *Plos Comput. Biol.* 13 (12), e1005880. doi:10.1371/journal.pcbi.1005880
- Nozawa, K., Schneider, T. R., and Cramer, P. (2017). Core Mediator Structure at 3.4 Å Extends Model of Transcription Initiation Complex. *Nature* 545, 248–251. doi:10.1038/nature22328

- Osman, S., and Cramer, P. (2020). Structural Biology of RNA Polymerase II Transcription: 20 Years on. *Annu. Rev. Cel Dev. Biol.* 36, 1–34. doi:10.1146/annurev-cellbio-042020-021954
- Pak, A. J., and Voth, G. A. (2018). Advances in Coarse-Grained Modeling of Macromolecular Complexes. *Curr. Opin. Struct. Biol.* 52, 119–126. doi:10.1016/j.sbi.2018.11.005
- Paule, M. R., and White, R. J. (2000). Survey and Summary Transcription by RNA Polymerases I and III. *Nucleic Acids Res.* 28 (6), 1283–1298. doi:10.1093/nar/28.6.1283
- Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C., Plitzko, J., and Cramer, P. (2016). Transcription Initiation Complex Structures Elucidate DNA Opening. *Nature* 533, 353–358. doi:10.1038/nature17990
- Plaschka, C., Larivière, L., Wenzek, L., Seizl, M., Hemann, M., Tegunov, D., et al. (2015). Architecture of the RNA Polymerase II-Mediator Core Initiation Complex. *Nature* 518, 376–380. doi:10.1038/nature14229
- Roeder, R. (1996). The Role of General Initiation Factors in Transcription by RNA Polymerase II. *Trends Biochem. Sci.* 21 (9), 327–335. doi:10.1016/s0968-0004(96)10050-5
- Sainsbury, S., Bernecky, C., and Cramer, P. (2015). Structural Basis of Transcription Initiation by RNA Polymerase II. *Nat. Rev. Mol. Cel Biol* 16, 129–143. doi:10.1038/nrm3952
- Sali, A., and Blundell, T. L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* 234, 779–815. doi:10.1006/jmbi.1993.1626
- Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., et al. (2017). Structures of Transcription Pre-initiation Complex with TFIIF and Mediator. *Nature* 551, 204–209. doi:10.1038/nature24282
- Shimizu, M., Noguchi, Y., Sakiyama, Y., Kawakami, H., Katayama, T., and Takada, S. (2016). Near-atomic Structural Model for Bacterial DNA Replication Initiation Complex and its Functional Insights. *Proc. Natl. Acad. Sci. USA* 113 (50), E8021–E8030. doi:10.1073/pnas.1609649113
- Shimizu, M., and Takada, S. (2018). Reconstruction of Atomistic Structures from Coarse-Grained Models for Protein-DNA Complexes. *J. Chem. Theor. Comput.* 14 (3), 1682–1694. doi:10.1021/acs.jctc.7b00954
- Silva, D.-A., Weiss, D. R., Pardo Avila, F., Da, L.-T., Levitt, M., Wang, D., et al. (2014). Millisecond Dynamics of RNA Polymerase II Translocation at Atomic Resolution. *Proc. Natl. Acad. Sci.* 111 (21), 7665–7670. doi:10.1073/pnas.1315751111
- Takada, S., Kanada, R., Tan, C., Terakawa, T., Li, W., and Kenzaki, H. (2015). Modeling Structural Dynamics of Biomolecular Complexes by Coarse-Grained Molecular Simulations. *Acc. Chem. Res.* 48 (12), 3026–3035. doi:10.1021/acs.accounts.5b00338
- Tan, C., and Takada, S. (2020). Nucleosome Allosterism in pioneer Transcription Factor Binding. *Proc. Natl. Acad. Sci. USA* 117 (34), 20586–20596. doi:10.1073/pnas.2005500117
- Terakawa, T., Kenzaki, H., and Takada, S. (2012). p53 Searches on DNA by Rotation-Uncoupled Sliding at C-Terminal Tails and Restricted Hopping of Core Domains. *J. Am. Chem. Soc.* 134 (35), 14555–14562. doi:10.1021/ja305369u
- Terakawa, T., and Takada, S. (2014). RESPAC: Method to Determine Partial Charges in Coarse-Grained Protein Model and its Application to DNA-Binding Proteins. *J. Chem. Theor. Comput.* 10, 711–721. doi:10.1021/ct4007162
- Tomko, E. J., Fishburn, J., Hahn, S., and Galburt, E. A. (2017). TFIIF Generates a Six-Base-Pair Open Complex during RNAP II Transcription Initiation and Start-Site Scanning. *Nat. Struct. Mol. Biol.* 24 (12), 1139–1145. doi:10.1038/nsmb.3500
- Unarta, I. C., Cao, S., Kubo, S., Wang, W., Cheung, P. P. H., Gao, X., et al. (2021). Role of Bacterial RNA Polymerase Gate Opening Dynamics in DNA Loading and Antibiotics Inhibition Elucidated by Quasi-Markov State Model. *Proc. Nat. Acad. Sci. U.S.A.* 118, e2024324118. doi:10.1073/pnas.2024324118
- Vannini, A., and Cramer, P. (2012). Conservation between the RNA Polymerase I, II, and III Transcription Initiation Machineries. *Mol. Cel* 45 (4), 439–446. doi:10.1016/j.molcel.2012.01.023
- Vorländer, M. K., Khatter, H., Wetzel, R., Hagen, W. J. H., and Müller, C. W. (2018). Molecular Mechanism of Promoter Opening by RNA Polymerase III. *Nature* 553, 295–300. doi:10.1038/nature25440
- Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinformatics* 54, 5.6.1–5.6.37. doi:10.1002/cpbi.3
- Weinan, E., Ren, W., and Vanden-Eijnden, E. (2002). String Method for the Study of Rare Events. *Phys. Rev. B* 66, 052301. doi:10.1103/physrevb.66.052301
- Winkelman, J. T., and Gourse, R. L. (2017). Open Complex DNA Scrunching: A Key to Transcription Start Site Selection and Promoter Escape. *Bioessays* 39 (2), 1600193. doi:10.1002/bies.201600193
- Zhang, B., Zheng, W., Papoian, G. A., and Wolynes, P. G. (2016). Exploring the Free Energy Landscape of Nucleosomes. *J. Am. Chem. Soc.* 138 (26), 8126–8133. doi:10.1021/jacs.6b02893

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Shino and Takada. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.