

DNA opening during transcription initiation by RNA polymerase II in atomic detail

Jeremy Lapierre¹ and Jochen S. Hub^{1,*}

¹Theoretical Physics and Center for Biophysics, Saarland University, Saarbrücken, Germany

ABSTRACT RNA polymerase II (RNAP II) synthesizes RNA by reading the DNA code. During transcription initiation, RNAP II opens the double-stranded DNA to expose the DNA template to the active site. The molecular interactions driving and controlling DNA opening are not well understood. We used all-atom steered molecular dynamics simulations to derive a continuous pathway of DNA opening in human RNAP II, involving a 55 Å DNA strand displacement and a nearly 360° DNA helix rotation. To drive such large-scale transitions, we used a combination of RMSD-based collective variables, a newly designed rotational coordinate, and a path collective variable. The simulations reveal extensive interactions of the DNA with three conserved protein loops near the active site, namely with the rudder, fork loop 1, and fork loop 2. According to the simulations, DNA–protein interactions support DNA opening by a twofold mechanism; they catalyze DNA opening by attacking Watson-Crick hydrogen bonds, and they stabilize the open DNA bubble by the formation of a wide set of DNA–protein salt bridges.

SIGNIFICANCE Transcription, the synthesis of RNA from a DNA template, is catalyzed by a macromolecular complex denoted RNA polymerase II. Transcription is initiated by the assembly of the polymerase with transcription factors and with double-stranded DNA. Henceforth, the DNA template strand is loaded into the active site, involving the rupture of ~13 DNA basepairs as well as large-scale rotations and translations of the DNA strands. To obtain insight into the protein–DNA interactions involved in initiation, we derived a continuous pathway of DNA opening using all-atom steered MD simulations. The simulations revealed extensive interactions between DNA and protein loops, which enable the disruption of Watson-Crick hydrogen bonds and the stabilization of the open DNA conformation.

INTRODUCTION

Transcription of DNA to RNA is catalyzed by RNA polymerases (RNAPs), a cornerstone of the central dogma of molecular biology (1). In eukaryotes, RNAP II carries out the synthesis of coding RNAs and of many noncoding RNAs. Transcription involves three main steps: initiation, elongation, and termination. To trigger initiation, the 12-subunit RNAP II first assembles with general transcription factors to form the preinitiation complex (PIC) (2). Within the 12 RNAP II subunits, RNA polymerase subunits 1 and 2 (RPB1 and RPB2, respectively, Fig. 1 A) form the cleft and the active site. Several loops protrude from the two large subunits (Fig. 1 A), which are well conserved among eukaryotes, including the rudder (in RPB1), fork loop 1 (FL1, in RPB2), and fork loop 2 (FL2, in RPB2)

(3,4). During initiation, these loops are in proximity with the DNA as the transcription bubble forms. The architecture of RNAP II and the mechanism of transcription initiation have been described in several excellent reviews (5,6).

Structural studies provided snapshots of the two end states of the PIC during transcription initiation in eukaryotes (2,7–15): snapshots of the closed complex (CC), in which DNA is double stranded and located on top of the RNAP II cleft, and of the open complex (OC), in which the transcription bubble has formed and is loaded into the active site (Fig. 1, A and B). While a cryo-EM structural study of the bacterial RNAP also revealed intermediate states of DNA opening (16), atomic details of the DNA opening pathway during transcription initiation in eukaryotes are missing. Consequently, the roles of conserved amino acid motifs of the rudder and of FL1 and FL2 during transcription initiation are largely unclear.

Previous molecular dynamics (MD) simulations focused on the elongation step of transcription (17–25) and on the clamp dynamics during initiation in bacterial RNAP (26). A recent coarse-grained MD study addressed DNA melting

Submitted May 13, 2022, and accepted for publication October 7, 2022.

*Correspondence: jochen.hub@uni-saarland.de

Editor: Karissa Sanbonmatsu.

<https://doi.org/10.1016/j.bpj.2022.10.012>

© 2022 Biophysical Society.



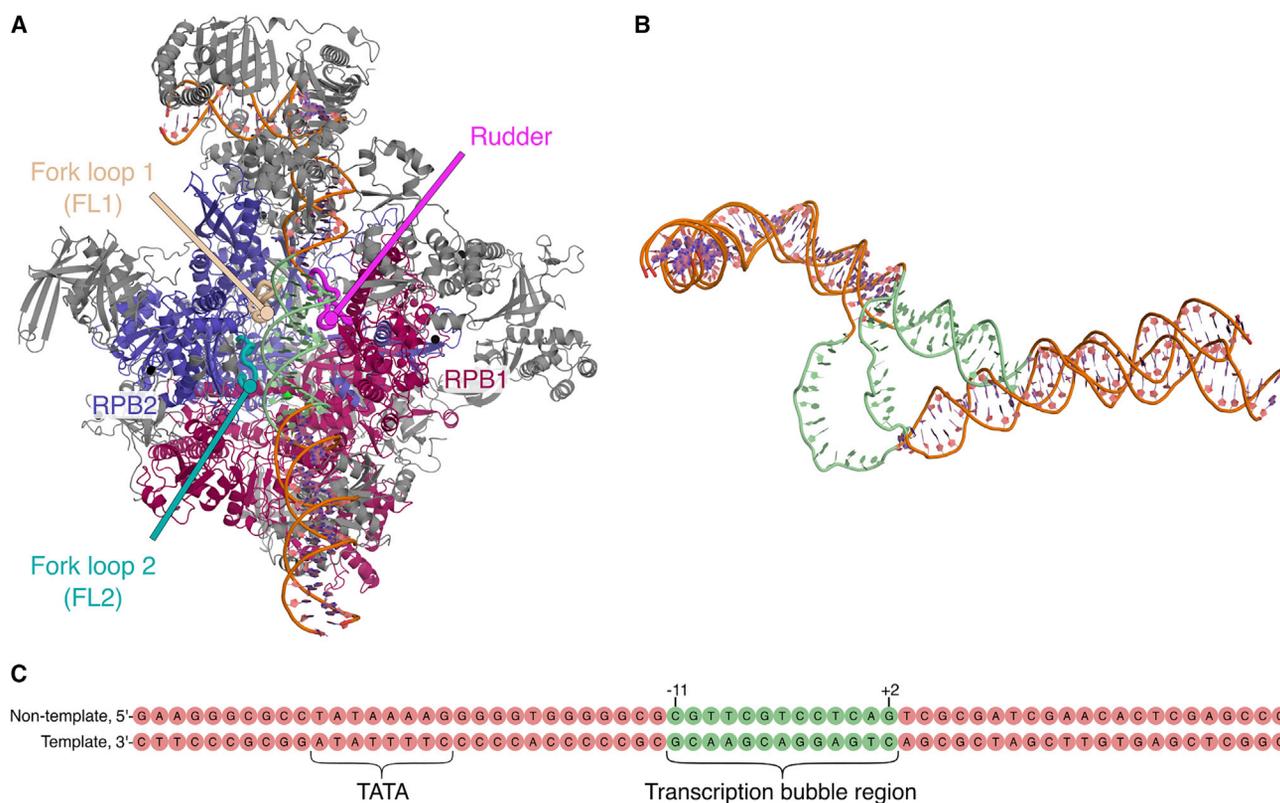


FIGURE 1 PIC complex in CC and overlap of DNA in CC and OC. (A) Cryo-EM structure of the CC without TFIIF and TFIIS (PDB: 5IY6 (7)). Zinc ions shown as black spheres. (B) Overlay of the DNA in CC and OC, taken from structures 5IY6 and 5IYB, respectively (7). The DNA region involved in the DNA bubble formation is highlighted in green. (C) DNA sequence simulated in this work, corresponding to the DNA sequence found in 5IYB. DNA numbering according to (7), where +1 refers to the transcription start site in the OC structure. To see this figure in color, go online.

by inserting DNA base mismatches (27). However, DNA opening has not been simulated with atomistic models or without DNA base mismatches.

In this work, we used MD simulations to obtain a continuous opening transition from the CC to the OC in atomic detail. Because the CC-to-OC transition involves conformational rearrangements on the scale of several nanometers, obtaining such transition by brute-force MD simulations is computationally prohibitive. Therefore, we used steered MD simulations (28,29) along a set of collective variables (CVs) to drive DNA opening and to enhance the sampling along the DNA opening pathway. Our CC-to-OC simulation provides insight into the spatial rearrangements of the DNA and of the protein loops during initiation, and they reveal extensive polar interactions of the DNA with the rudder, FL1, and FL2. These observed interactions suggest roles of the protein loops in supporting DNA strand separation and in stabilizing the transcription bubble in the OC.

MATERIALS AND METHODS

Simulation setup

MD simulations were carried out with Gromacs (30) version 2020.2 patched with Plumed (31,32) version 2.6.1, and with Gromacs version

2021 patched with Plumed version 2.7.0. The initial atomic coordinates for the CC were obtained from the Protein Data Bank (PDB: 5IY6 (7)) from which we removed TFIIF and TFIIS. We used YASARA version 20.8.23 to add acetyl and N-methyl amide capping groups at the ends of the missing protein regions and at the C- and N-termini. We also used YASARA to add missing atoms (33). The system was solvated with TIP3P water molecules and Na/Cl counterions were added to neutralize the system with a salt concentration of 100 mM (34). In total, the system contained 832,078 atoms. The OL15 force field was used for the DNA (35). The ff14sb force field was used for the protein (36) except for the zinc(II)-coordinating Cys and His residues, for which the improved parameters by Macchiagodena et al. were used (37).

Electrostatic interactions were computed with the particle-mesh Ewald method (38) using a real-space cutoff at 1 nm and a Fourier spacing of 0.16 nm. Dispersion interactions and short-range repulsion were described with a Lennard-Jones potential with a cutoff at 1 nm. Bonds and angles of water were constrained with the SETTLE algorithm (39) and bonds involving other hydrogen atoms were constrained with LINCS (40). To remove atomic clashes, the system was energy minimized with the steepest-descent algorithm. We next equilibrated the system under *NVT* conditions for 100 ps at 300 K using the velocity-rescale thermostat with one heat bath for the coordinated ions, DNA and protein and another heat bath for water and counterions (41). Then, we equilibrated the system at 1 bar for 10 ns under *NPT* conditions using Parrinello-Rahman pressure coupling and using the same thermostat as in the *NVT* equilibration (42). During both equilibration steps, all heavy atoms were position restrained with a force constant of $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$.

To enable the use of a 4 fs time step for further pulling simulations we used hydrogen mass repartitioning (HMR) (43). Accordingly, to increase the oscillation period of the bond angles involving hydrogen atoms, the

hydrogen masses were scaled up by a factor of f_H and the heavy atom masses connected to hydrogens were scaled down, while keeping the overall masses of chemical moieties constant. To choose a scaling factor f_H that yields stable simulations at a 4 fs time step, we tested scaling factors from 2 to 3 in steps of 0.2, where three simulations were carried out for each scaling factor. Each simulation was carried out for 20 ns in *NPT* conditions. None of the simulations with a scaling factor of 2.8 or 3 were stable, whereas all other simulations were stable. For production simulations, we decided to use $f_H = 2.5$.

To exclude that HMR leads to excessive energy drift, we carried out three *NVE* simulations with $f_H = 2.5$ using 4 fs time steps for 500 ps and, for reference, three *NVE* simulation without HMR using a 2 fs time step. On average, we obtained an energy drift of 0.06% ns^{-1} with HMR, which was even smaller than the average value of -0.13% ns^{-1} without HMR (Table S1). Hence, integrating Newton's equations of motion with HMR models was numerically stable and exhibited only a marginal energy drift.

Simulation of initial DNA opening pathway

We generated an initial path from the CC to the OC with a steered MD simulation of 175 ns using a combination of one rotational CV and two root mean-square distance RMSD-based CVs.

As the first CV, we used a rotational CV defined as $\xi_1 = 1/4 \sum_{i=1}^4 \gamma_i(X; X_0)$, where each dihedral angle γ_i was defined as:

$$\gamma_i(X; X_0) = \text{dih}(\mathbf{v}_i(X_0), \mathbf{u}_{\text{axis}}, \mathbf{v}_i(X)) \quad (1)$$

Here, \mathbf{u}_{axis} denotes the helix axis of the DNA in region +3 to +23 (Fig. S1 A). X_0 is the configuration at $t = 0$ ns, and X is the configuration at a later simulation time t . Further, $\mathbf{v}_i(X_0)$ denotes the vector connecting the two centers of mass (COMs) $\mathbf{c}_{i,1}(X_0)$ and $\mathbf{c}_{i,2}(X_0)$ at $t = 0$ ns (Fig. S1, B and D–F), and $\mathbf{v}_i(X)$ is the instantaneous vector connecting the same COMs at simulation time t (Fig. S1, C and H). The two groups of atoms used to define $\mathbf{c}_{i,1}$ and $\mathbf{c}_{i,2}$, respectively, were constructed by splitting the helix DNA region +3 to +23 along the axis, as illustrated in Fig. S1, B and D–F. The four $\mathbf{v}_i(X_0)$ defining ξ_1 are depicted in Fig. S1 G.

As a second CV, we used $\xi_2 = \Delta(X, X_{\text{OC1}})$. Here, X_{OC1} denotes the DNA backbone atoms of the OC in the region -17 to -5 , taken from the 5IYB structure (7), and $\Delta(X, X_{\text{OC1}})$ denotes the RMSD of the instantaneous structure X relative to the reference structure X_{OC1} . As a third CV, we used $\xi_3 = \Delta(X, X_{\text{OC2}})$. Here, X_{OC2} denotes the backbone atoms of the OC in the region -17 to $+2$, again taken from the 5IYB structure.

During the 175 ns of steered MD simulation, we applied different forces on the three CVs described above:

1. The rotational CV (ξ_1) was pulled from 6.27 rad (close to 2π) to 0.01 rad over the first 100 ns using a force constant of $7000 \text{ kJ mol}^{-1} \text{ rad}^{-2}$. Over the next 4 ns, the force applied on ξ_1 was turned off by linearly decreasing the force constant from 7000 to $0 \text{ kJ mol}^{-1} \text{ rad}^{-2}$.
2. The RMSD relative to X_{OC1} (ξ_2) was pulled from 2.35 to 0.4 nm over the first 50 ns using a force constant of $10,000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Over the next 50 ns, ξ_2 was pulled from 0.4 to 0 nm using a force constant decreasing linearly from 10,000 to $0 \text{ kJ mol}^{-1} \text{ nm}^{-2}$.
3. The RMSD relative to X_{OC2} (ξ_3) was pulled from 1.99 to 0 nm between simulation times of 50 and 100 ns, using a linearly increasing force constant between 20,000 and $30,000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Over the next 75 ns, ξ_3 was restrained at 0 nm using a force constant of $30,000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Notably, the relatively large force constants acting on the RMSD-based CV propagate to only moderate forces on the 451 individual atoms that were used for defining the RMSD.

To exclude that independent steered MD simulations would lead to qualitatively different opening pathways, 5 independent 175 ns simulations were carried out and analyzed in terms of number of hydrogen bonds (H-bonds), contacts, and interaction energies (Fig. S4). Whereas indepen-

dent simulations exhibit different fluctuations in the downstream DNA helix, the trends of the interactions and conformations of the DNA bubble are largely preserved.

Relaxation of the initial DNA opening pathway

To relax and sample intermediate states along the opening pathway, we first used constant-velocity pulling from the CC to the OC with a path collective variable (PCV) (44). The two components of a PCV are S_{path} and Z_{path} , and are defined as:

$$S_{\text{path}}(X) = \frac{\sum_{i=1}^N i e^{-\lambda M_i(X)}}{\sum_{i=1}^N e^{-\lambda M_i(X)}} \quad (2)$$

$$Z_{\text{path}}(X) = -\frac{1}{\lambda} \ln \sum_{i=1}^N e^{-\lambda M_i(X)} \quad (3)$$

where the unitless S_{path} describes the progression along the path and Z_{path} describes the deviation from the path. N denotes to the number of reference configurations defining the DNA opening path. The distance metric $M_i(X) = \Delta^2(X, X_i)$ is the mean-squared deviation (MSD) of the instantaneous configuration X relative to the reference configuration X_i . The choice of the N reference configurations was optimized to obtain similar MSDs between neighboring configurations and a good flatness of the surface spanned by the $N \times N$ MSD matrix (44). The symbol λ is the smoothing parameter, proportional to the inverse of the MSD between adjacent reference configurations.

To relax the initial path along Z_{path} , we performed two rounds of constant-velocity pulling along S_{path} , while applying a wall potential on Z_{path} acting above $Z_{\text{path}}(X) = 0.035 \text{ nm}^{-2}$. For the first relaxation round, we took $N = 72$ reference configurations from the initial path, set $\lambda = 21.7 \text{ nm}^{-2}$, and carried out 100 ns of constant-velocity pulling along S_{path} from 1.1 to 71.3, corresponding to configurations close to the first or last reference configuration. We used a force constant of 5000 kJ mol^{-1} for S_{path} and a force constant of $2.8 \times 10^6 \text{ kJ mol}^{-1}$ for Z_{path} . To build a new path for the following relaxation round satisfying the two criteria for reference selection mentioned above without discarding configurations in the OC, another 20 ns simulation was required with a harmonic restraint centered on $S_{\text{path}}(X) = 71.3$, with a force constant of 5000 kJ mol^{-1} and keeping the wall potential acting above $Z_{\text{path}}(X) = 0.035 \text{ nm}^2$ with an offset of 0.005 nm^2 and a force constant of $2.8 \times 10^6 \text{ kJ mol}^{-1} \text{ nm}^{-4}$. For the second relaxation round, we used $N = 91$ reference configurations from the first relaxation round, set $\lambda = 45.43 \text{ nm}^{-2}$, and carried out 100 ns of constant-velocity pulling along S_{path} from 1.15 to 90.85 using a force constant of 4000 kJ mol^{-1} for S_{path} and a force constant of $4 \times 10^6 \text{ kJ mol}^{-1}$ for Z_{path} . For the same reasons as for the first relaxation round, we extended the second relaxation simulation for another 20 ns with a harmonic restraint centered on $S_{\text{path}}(X) = 90.85$, with a force constant of 4000 kJ mol^{-1} and keeping the wall potential acting above $Z_{\text{path}}(X) = 0.035 \text{ nm}^2$ with an offset of 0.005 nm^2 and a force constant of $4 \times 10^6 \text{ kJ mol}^{-1} \text{ nm}^{-4}$.

Sampling the final DNA opening path

From the second relaxation round of constant-velocity pulling described above, we selected $N = 63$ reference configurations to define our final PCV. This final PCV was set up with $\lambda = 36 \text{ nm}^{-2}$. To sample DNA and loop conformations along the DNA opening path, we extracted 90 configurations from the second round of constant-velocity pulling with S_{path} ranging from 1.250 to 62.9. Each of those 90 configurations was used to start a simulation of 50 ns restrained on a particular value of S_{path} with a harmonic potential.

The reference positions and the force constants of the harmonic potentials of these 90 simulations are shown in Fig. S2. These 90 simulations were used to characterize the opening path in terms of H-bonds, potential energies, and atomic contacts (Figs. 4 A and 5).

Simulation analysis

H-bonds were defined with a cutoff distance of 0.35 nm between the hydrogen atom and the H-bond acceptor, and with a cutoff angle hydrogen-donor-acceptor of 30°. The basepairs +2 and +3 exhibited disrupted H-bonds but were mismatched with other bases in the downstream DNA fork; hence, during the analysis, we did not consider these bases as part of the open DNA bubble. Contacts were defined with a cutoff distance of 0.3 nm. The potential energies were computed as the average of the sum of Lennard-Jones and short-range Coulomb interactions with a cutoff at 1 nm. Standard errors (SE) were all computed with the gmx analyze tool and confidence intervals were defined as one SE. Simulation trajectories were visualized with PyMOL (45) and VMD(46) Images were generated with PyMOL.

Conformational stability of the OC

To test the stability of the OC obtained from the second relaxation round, we performed a free simulation of the OC, i.e., without any biasing potential. To this end, a 200 ns simulation was started from the final configuration of the second relaxation round. We quantified the stability of the OC by monitoring the distances between the 12 disrupted basepairs of the transcription bubble (Fig. S3). The 12 distances were computed with the center of geometries of the heavy backbone atoms of the two complementary nucleotides. The 12 distances were then averaged in each time frame. For reference, the same protocol was used to compute the average distances between the 13 dissociated basepairs in the reference OC (5IYB (7)).

Multiple sequence alignments

RPB1 and RPB2 protein sequences were taken from the UniProtKB/Swiss-Prot database (47). The organisms were chosen to cover different kingdoms of the eukaryotic domain. The alignments were carried out with CLUSTAL W version 2.1 (48).

RESULTS

Steering a 55 Å conformational transition with a combination of CVs

Upon forming the transcription bubble, DNA carries out a transition involving a rotation of the DNA double strand by $\sim 370^\circ$ as well as a translation of the DNA strands by up to 55 Å relative to the protein (7). Simulating such large-scale, nonlinear conformational transitions in atomic detail imposes considerable challenges. One possible strategy for favoring these large-scale motions is to introduce base mismatches between the two DNA strands, as used for obtaining the OC cryo-EM structure by He et al. (7) or used for favoring DNA melting in coarse-grained MD simulations (27). In contrast to these previous studies, we simulated DNA opening without base mismatches according to the biologically relevant state of the CC (Fig. 1 C). We obtained a relaxed pathway of DNA opening with a combination of two methods. First, we obtained an initial pathway using steered MD simulations

along a combination of three CVs; second, the initial pathway was relaxed using the PCV method (44).

To guide the opening pathway, steered MD simulations were carried out along a combination of the following three CVs: 1) a rotational CV applied to the downstream DNA helix, thereby driving the melting of the DNA strand (Fig. 2 A, ξ_1); 2) two CVs given by the RMSD of the sugar-phosphate backbone of the DNA relative to the conformation in the OC, taken from the 5IYB structure (ξ_2 and ξ_3) (7). Fig. 2 A illustrates the evolution of the rotational CV ξ_1 and of the RMSD-based CV ξ_3 . By pulling along these CVs we obtained an initial path of DNA opening (Videos S1 and S2).

A PCV for steering and relaxing the DNA opening path

Because our steered MD simulations were carried out on much shorter timescales compared with experimental timescales, it is reasonable to believe that the initial path is still biased by nonequilibrium effects. To relax the conformations along the opening pathway and, thereby, to mitigate such nonequilibrium effects, we applied the PCV method (44). Generally, PCVs are defined using two CVs: the position S_{path} along the initial path and the distance Z_{path} from the path, where the path is defined along N intermediate conformations (see materials and methods for details). In this study, the initial PCV was defined with 72 intermediate conformations taken from the steered MD simulation. Then, we carried out two rounds of constant-velocity pulling along S_{path} . Within each round, the path was allowed to relax, providing us with an updated set of increasingly relaxed intermediate conformation and, thereby, an updated PCV. The final PCV along 63 relaxed intermediate conformations allows convenient opening simulations by pulling along the single S_{path} , instead of pulling along the three CVs used for obtaining the initial path (see above). In addition, projection onto the final S_{path} provides a convenient measure for the progress of the opening pathway, as used below in our figures and analysis.

Atomistic transition from the closed to a stable open DNA

By pulling along the aforementioned S_{path} , we obtained all-atom continuous trajectories of DNA opening from the CC to the OC (Fig. 2, A and C and Videos S1 and S2). To test whether we have reached a state with a stable open DNA bubble, we simulated the final state without any biasing potential for 200 ns. In this simulation, the distances between the disrupted basepairs were reasonably stable (Fig. S3, A and E), demonstrating that the strands did not reanneal, as expected for a stable OC. Three additional free simulations of 300 ns each corroborated the stability of the OC (Fig. S3, B–D and F). The open DNA bubble exhibited a length of 12 basepairs

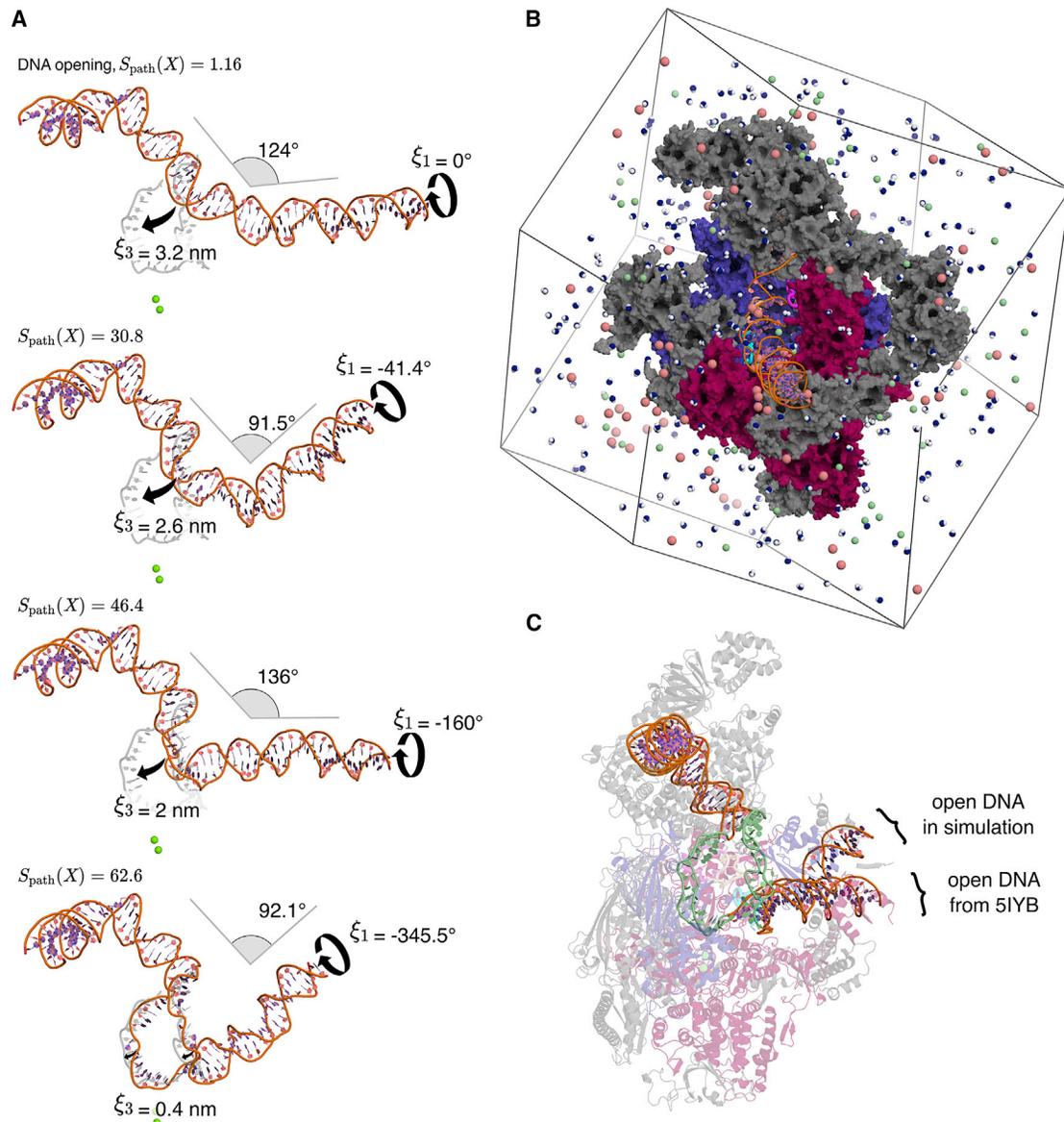


FIGURE 2 Transition from closed to open DNA in atomic detail. (A) Opening transition snapshots with corresponding S_{path} value (progression along the DNA opening), rotation of the downstream DNA helix (ξ_1), RMSD relative to the open bubble (ξ_3), and DNA helix bending angle. The target open bubble conformation is depicted in light gray. For reference, two catalytic magnesium ions are shown as green spheres. (B) Simulation box. Colors for the PIC in CC are consistent with Fig. 1. Water molecules, sodium ions, and chloride ions are colored in blue and white, pale pink, and pale green, respectively. Most water molecules and ions have been removed for clarity. (C) Section of the PIC in OC obtained from steered MD simulations. Open DNA from the 5IYB structure is shown for reference. The transcription bubble is colored in pale green. To see this figure in color, go online.

(bp), in reasonable agreement with the length of 13 bp in the reference structure by He et al. (7). We quantified the spatial extension of the bubble with the average distance d_{bb} between the 12 disrupted basepairs. In the unbiased 200 ns simulation of the OC, we obtained $d_{\text{bb}} = 2.36 \text{ nm}$ (Fig. S3), in reasonable agreement with the value of 2.67 nm in the reference structure. Minor structural differences relative to the reference OC are expected because 1) the DNA bubble is flexible and 2) RNAP II in the OC accommodates various DNA bubble lengths and widths during initiation and, more generally, during the entire transcription process (49). Overall, the stability

of the DNA transcription bubble in our free simulation implies that the previous pulling simulation represents a complete DNA opening pathway.

Fig. 2 A and Videos S1 and S2 provide an atomic view on the large DNA rearrangements. First, due to the clockwise rotational motion carried out by the downstream DNA, the DNA became underwound in the transcription bubble region. Second, due to the translational motions induced on the transcription bubble toward the active site, the DNA bent at the transcription bubble region. These two topological changes of DNA led ultimately to the disruption of

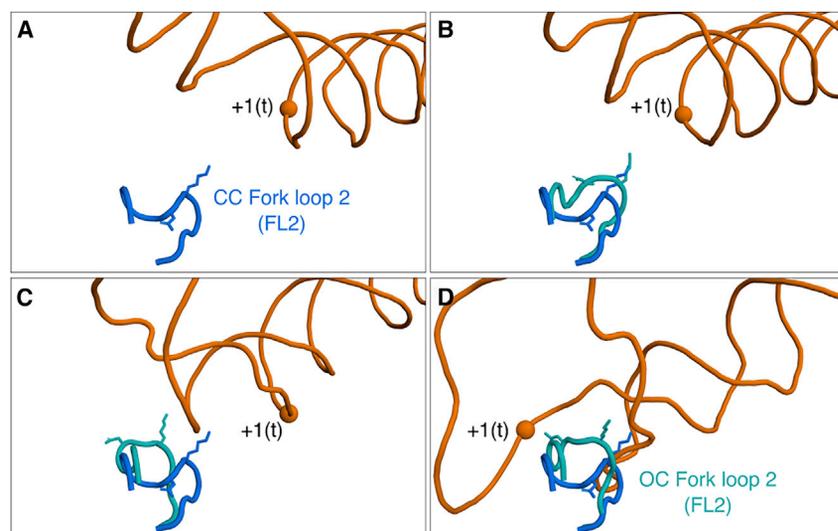


FIGURE 3 Tilting of sensor fork loop 2 (FL2) into the transcription bubble during DNA opening. (A–D) Snapshots of FL2 (cyan) during DNA opening corresponding to $S_{\text{path}} = 1.16, 31.8, 51.5,$ and 62.6 , respectively. The final tilted state of FL2 is depicted in (D). For reference, the starting position of FL2 (marine blue), Asp-492 and Lys-494 are shown. To see this figure in color, go online.

12 bp. DNA rotational angles, bending angles, and RMSD relative to the reference OC are depicted in Fig. 2 A for four snapshots of our DNA opening trajectory. This interplay between negative supercoiling (clockwise rotational motion of DNA), DNA bending, and basepair disruptions have been reported previously in DNA minicircles (50–53). In addition, negative supercoiling has been shown to promote DNA opening during transcription with minimal transcription factors (54), further corroborating that our simulations reflect experimentally relevant conditions. Together, we obtained a simulation protocol that provides continuous atomistic transitions from the CC to a stable OC with an open DNA transcription bubble within computationally accessible simulation times. The protein–DNA contacts and interaction energies obtained from the simulations are discussed in the following sections.

Fork loop 2 tilts during DNA opening

Because DNA opening occurs inside the PIC, the DNA extensively interacts with protein domains and, in particular, with the protein loops. While DNA was loaded into the active site in the simulations, FL2 tilted into the transcription bubble, between the two DNA strands (Fig. 3, A–D and Video S3). Whereas solvent-exposed protein loops are often flexible, the FL2 conformation pointing into the open bubble was remarkably stable, locked by electrostatic protein–DNA interactions, as observed in the free 200 ns simulation following the opening transition described above. The FL2 tilting in our simulations is compatible with a hypothesized role of FL2 as a sensor for the open transcription bubble (7). A recent study revealed a similar conformational change of FL2 during the transition from the CC to the OC, further supporting that FL2 is acting as a sensor for DNA opening (15).

Fork loops 1 and 2 support DNA opening by H-bond attack on Watson-Crick pairs

The simulations revealed how DNA opening is supported by the rearrangement of H-bonds between DNA, protein, and water, as shown in Fig. 4. Namely, the loss of 31 Watson-Crick (WC) DNA–DNA H-bonds (Fig. 4 A, orange curve) was predominantly compensated by the formation of approximately 40 DNA–water H-bonds (Fig. 4 A, blue curve). The open DNA was further stabilized by the formation of approximately 12 DNA–protein H-bonds (Fig. 4 A, green curve), among which $\sim 50\%$ formed with the WC edge (Fig. 4 A, red curve), and the other $\sim 50\%$ formed with other edges or with the DNA backbone.

The progression of WC H-bonds (Fig. 4 A, orange curve), together with visual inspection of the MD trajectories, revealed three key protein residues involved in destabilizing the double-stranded DNA by attacking the WC H-bonds. These events are reflected by marked decreases of the number of WC H-bonds at $S_{\text{path}} = 45.9, 49.6,$ and 55.4 (Fig. 4 A, vertical lines) and are visualized in the molecular representations of Fig. 4, B–D. First, the side chain of Gln-461 of FL1 interacted with the basepairs at -6 , thereby competing with the WC basepairing (Fig. 4 B, $S_{\text{path}} = 45.9$). Second, the side chain of Lys-458 of FL1 interacted via H-bonds with basepairs at position -9 and -10 (Fig. 4 C, $S_{\text{path}} = 49.6$). Third, at a later stage of the opening process and after FL2 tilted into the open bubble, Arg-491 of FL2 destabilizes WC H-bonds, thereby promoting the unzipping of the double-stranded DNA (Fig. 4 D, $S_{\text{path}} = 55.4$). Hence, DNA–protein interactions do not merely serve as a compensation for the loss of DNA–DNA interactions to energetically stabilize the final OC state, but they might also catalyze the rupture of the WC basepairing.

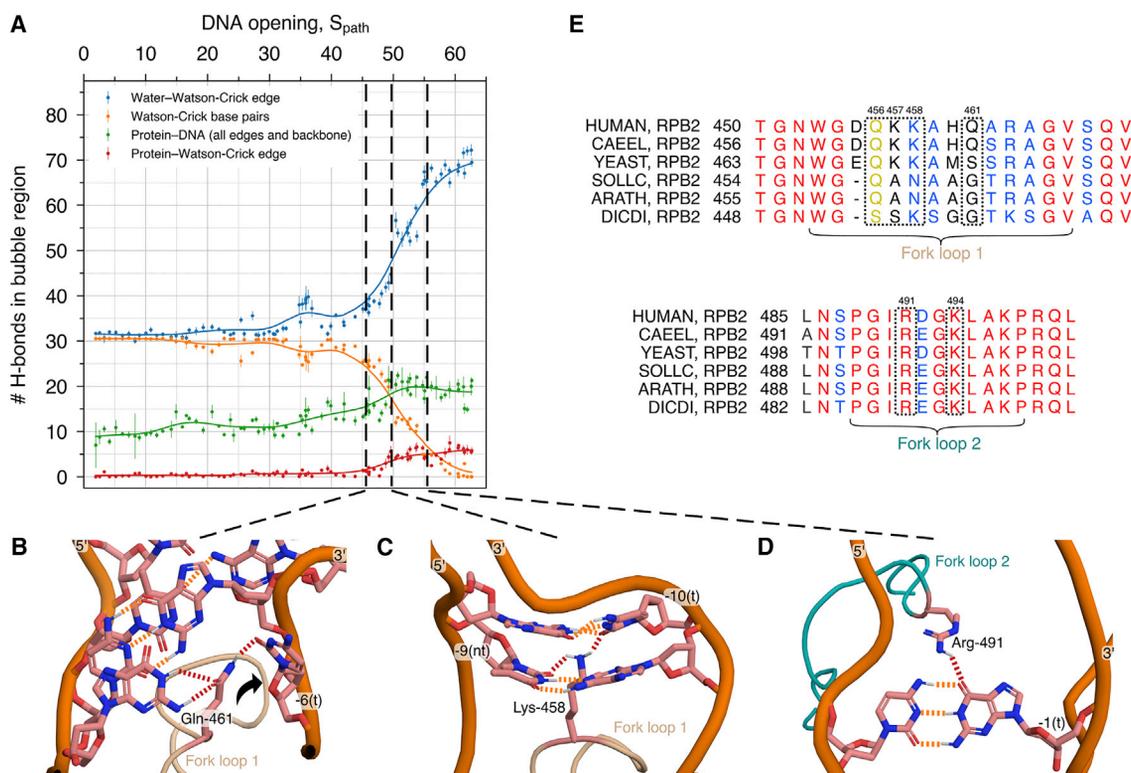


FIGURE 4 Rupture of Watson-Crick H-bonds in the transcription bubble and formation of DNA–protein and DNA–water H-bonds during the DNA opening process. (A) Development of the H-bonds of the transcription bubble region: number of H-bonds between WC edge and water (blue), between basepairs (orange), between DNA and protein (green), and between DNA WC edge and protein (red). Smooth lines are shown to guide the eye. (B) WC H-bond disruption driven by Gln-461 and base flipping (black arrow) of DNA residue -6 in the template strand. (C and D) Attack of WC H-bond by fork loops 1 and 2, respectively. (E) Sequence alignment (CLUSTAL W (48)) of fork loop 1 and 2 from six different eukaryotes: *Homo sapiens*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Solanum lycopersicum*, *Arabidopsis thaliana*, and *Dictyostelium discoideum*. The residue color highlights the degree of conservation: invariant residues (red), residues with similar properties (blue), or with weakly similar properties (yellow). Residues in dotted boxes are discussed in the text. To see this figure in color, go online.

The DNA dynamics in RNAP II driven by FL1 and FL2 are not unique but instead resemble dynamics observed in other DNA-interacting enzymes. For instance, base flipping has been suggested as an early mechanistic stage for DNA opening in a bacterial promoter (55). Likewise, H-bond attack to WC basepairs has been proposed for the cytosine 5-methyltransferase, where the enzyme infiltrates the DNA helix by forming H-bonds with nucleic bases, consequently destabilizing WC H-bonds and inducing base flipping (56,57). Similarly, a base flipping event at position -6 of the template strand occurred during DNA opening in our simulation. Here, base flipping was promoted by the aforementioned Gln-461, via disruption of the WC H-bonds during the DNA opening (Fig. 4 B, black arrow).

To get additional insights into the role of DNA–protein interactions during DNA opening, and to identify selection pressure on the three key residues mentioned above, we analyzed the residue conservation of FL1 and FL2 among six eukaryotic organisms by means of multiple sequence alignments. Overall, FL1 and FL2 are strongly conserved

among eukaryotes, demonstrating their critical biological roles (Fig. 4 E). However, Gln-461 is not conserved among eukaryotes, suggesting that the DNA–Gln-461 interactions observed in our simulations are either not critical for DNA opening or may be replaced with other interactions. In contrast, Lys-458 is well conserved among the analyzed eukaryotes (Fig. 4 E); we hypothesize that the substitutions with Asn in *Solanum lycopersicum* and *Arabidopsis thaliana* may interact with DNA similar to Lys, thus supporting the role of residue 458 in destabilizing the double-stranded DNA. Arg-491 is invariant among all the eukaryotic organisms chosen here (Fig. 4 E), underlining its biological relevance in DNA strand separation. Taken together, our data suggest that residues of FL1 and FL2 catalyze DNA opening by attacking WC H-bonds between double-stranded DNA, providing a rationale for the marked sequence conservation of FL1 and FL2. According to the simulations and the sequence alignment, the conserved residues Lys-458 and Arg-491 are involved in H-bond attack; however, we cannot exclude the possibility that other conserved residues play similar roles.

Fork loop–DNA and rudder–DNA electrostatic interactions stabilize the open DNA conformation

To rationalize the energetic driving forces for DNA opening, we monitored the potential energy from DNA–DNA, DNA–protein, and DNA–water interactions (Fig. 5 A). Here, potential energies were taken as the sum of Lennard-Jones and short-range Coulomb interactions, averaged over 50 ns of simulation and normalized relative to the state of the CC. The loss of interactions between the DNA strands

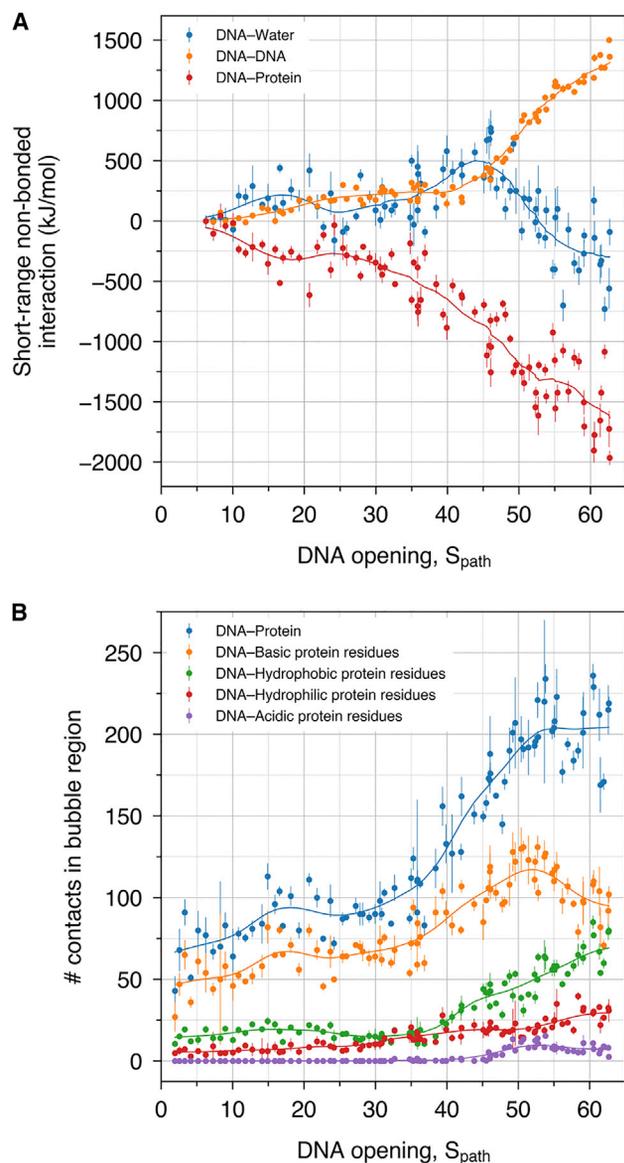


FIGURE 5 Electrostatic interactions support DNA opening. (A) Coulomb and Lennard-Jones short-ranged interactions between DNA and water (blue), between DNA and DNA (orange), and between DNA and protein (red). (B) Number of contacts within the DNA bubble region with all protein residue types (blue), with basic protein residues (orange), with hydrophobic protein residues (green), with hydrophilic protein residues (red), and with acidic protein residues (purple). Smooth lines are shown to guide the eye. To see this figure in color, go online.

is primarily compensated by a large gain of DNA–protein interactions, as evident from the large negative DNA–protein potential energies (Fig. 5 A, orange and red). Although DNA opening leads to an increase of DNA–water interactions, as expected from the formation of H-bonds between water and the WC edge (Fig. 4 A, blue), DNA–water interactions (Fig. 5 A, blue) play a much smaller role compared with DNA–protein interactions (Fig. 5 A, red).

To quantify which type of DNA–protein interactions drive DNA opening, we further analyzed the number of contacts of the DNA bubble region with different groups of amino acids of common physicochemical properties (Fig. 5 B). Evidently, the DNA forms ~50 new contacts with basic protein residues, far more compared with contacts with polar or acidic residues. This finding reflects that the RNAP II cleft is highly positively charged, which helps to attract the negatively charged DNA backbone deeper into the cleft and, in particular, into the active site. This finding demonstrates, not surprisingly, that electrostatic interactions between the DNA and RNAP II are the key energetic driver for transcription bubble formation.

Visual inspections of the simulations revealed reoccurring salt bridges and H-bonds between the protein and the open DNA. Gln-456 and Lys-457 (in FL1) form H-bonds with the template strand of DNA (Fig. 6 A), suggesting that FL1 stabilizes the open bubble by compensating the loss of H-bonds between the two DNA strands and by imposing a steric obstacle against strands reannealing. In close proximity to FL1, Arg-327 (in the rudder) forms a salt bridge with the template strand (Fig. 6 B). Likewise, Lys-494 (in FL2) and Arg-222 (in RPB2) interact with the nontemplate strand via salt bridges and H-bonds (Fig. 6 C). Arg-222, Lys-413, and Arg-416 are examples of residues outside the fork loops or the rudder that form electrostatic interactions with the nontemplate strand of the open DNA conformation (Fig. 6 C–D).

To corroborate the relevance of the DNA–protein interactions observed in our simulations for stabilizing the transcription bubble, we inspected the conservation of the residues mentioned above with sequence alignments. Accordingly, Lys-494 (Fig. 4 E, FL2) and Arg-222 (Fig. 6 E, RPB2) are invariant, while Lys-413 (Fig. 6 E, RPB2) is well conserved, supporting their biological relevance. Gln-456 (Fig. 4 E, FL1) and Arg-416 (Fig. 6 E, RPB2) are largely conserved except in *Dictyostelium discoideum* and in *Saccharomyces cerevisiae*, respectively, underlining their putative role in stabilizing the transcription bubble. In contrast, Arg-327 (Fig. 6 E, rudder) is not conserved but may be replaced with Thr or Gln; however, all those residues are capable of forming H-bonds with the DNA backbone, and thereby may all stabilize the open bubble. Finally, Lys-457 (Fig. 4 E, FL1) is not conserved, suggesting that this residue is less critical for stabilizing the transcription bubble. Together, our simulations show extensive interactions between the PIC and DNA that stabilize the

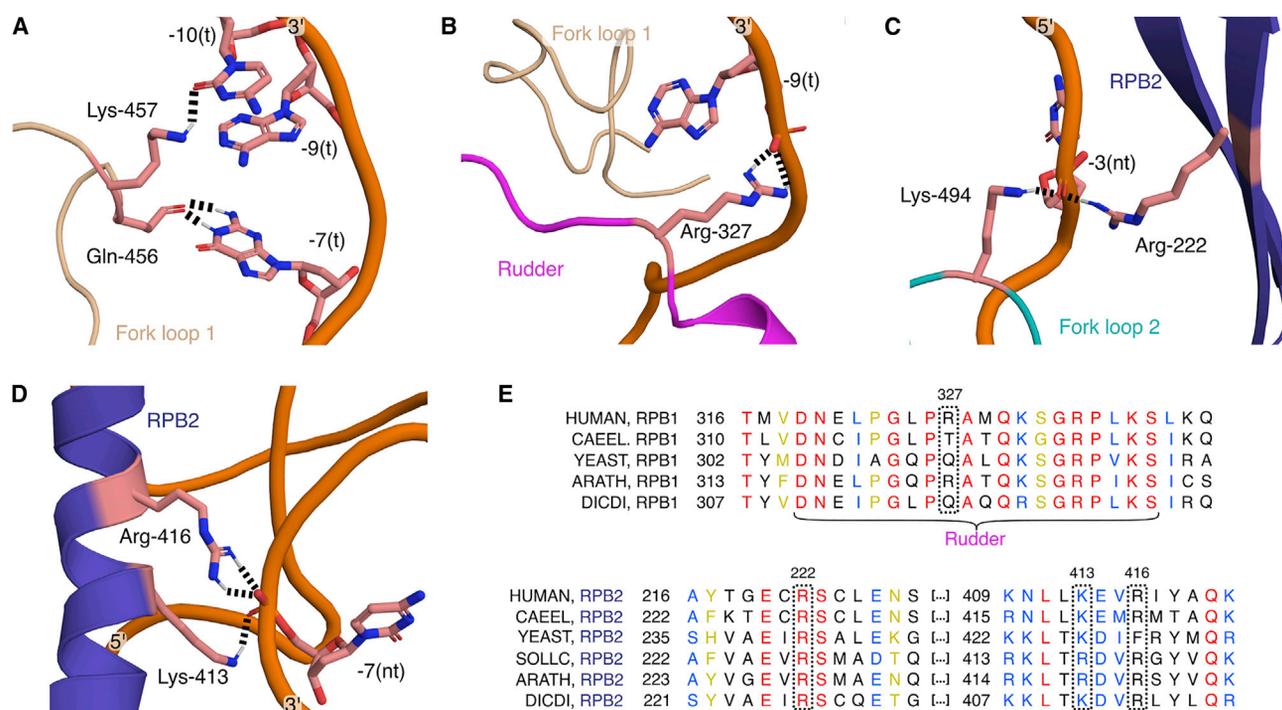


FIGURE 6 Electrostatic interactions between DNA and PIC stabilize open DNA in the OC. (A–D) H-bonds between protein and DNA. (B–D) Salt bridges between cationic residues with the anionic DNA backbone. (E) Sequence alignment (CLUSTAL W) of the rudder from different eukaryote organisms. Red residues are invariant, blue are from groups of strongly similar property, and yellow from groups of weakly similar properties. To see this figure in color, go online.

DNA bubble. In the light of the sequence alignments, many of these interactions are critical, whereas some may be replaced by other interaction.

DISCUSSION

We have presented the first all-atom simulation of a continuous DNA opening transition within human RNAP II. The simulations revealed extensive electrostatic and polar interactions of the DNA with the protein, predominantly with the two fork loops and with the rudder. Closer inspection of these interactions suggested that the rudder and the two fork loops are involved in 1) the separation of the two DNA strands by means of H-bond attack to WC basepairs and 2) in maintaining the open DNA conformation by a combination of steric hindrance and electrostatic interactions. The biological relevance of the protein residues involved in the observed interactions was further scrutinized by analyzing their conservation among eukaryotic amino acid sequences. Finally, we observed a base flipping event as well as the flipping of FL2 into the transcription bubble, in line with previous experiments (7,15,55–57).

Mutagenesis experiments targeting the fork loops and the rudder have been carried out in archaeal RNAP II and have revealed that the rudder helps in stabilizing the melted DNA in the OC (58). In addition, these experiments suggested that FL2 and, in particular Arg-451—the archaeal equivalent of

Arg-491 mentioned in this work—plays a role in unwinding downstream DNA during elongation. However, mutagenesis of FL1 did not impact DNA opening in archaeal RNAP II. Whereas archaeal and human RNAP II exhibit high sequence conservation, the physiological temperature at which DNA opening occurs may strongly differ, which might influence DNA melting. Indeed, a temperature of 70°C was used in the permanganate footprinting experiment by Naji et al. compared with a temperature of 37°C expected for human physiological conditions (58). The same kind of mutagenesis experiments in eukaryotic RNAP II system, ideally for human RNAP II, would be of high interest to confirm that FL1, FL2, and the rudder are essential for DNA opening.

Simulating and relaxing such a large-scale roto-translational conformational transition with atomic MD force fields is computationally challenging. A recent coarse-grained MD study introduced basepair mismatches in the transcription bubble region to favor DNA opening (27). However, if DNA melting occurs without simultaneous rotation of the downstream DNA, high DNA strains in the upstream or downstream DNA region emerge, which is incompatible with the open DNA conformation from experimental structures (7,9,59). Therefore, in this work, we used steered MD simulations along a combination of three CVs to guide the large-scale displacement of DNA by up to 55 Å simultaneously with DNA rotation by ~346°. Finding

a suitable set of CVs for obtaining a stable OC without undesired DNA melting outside the transcription bubble required extensive optimization and human supervision; hence, future studies may aim toward more automated protocols for finding suitable sets of CVs. Having obtained an initial DNA opening simulation from steered MD, we used the PCV framework to relax the initial opening pathway. Notably, we tried to compute the potential of mean force along the PCV with umbrella sampling (60) or metadynamics (61), with the aim to obtain an estimate for the free energy of DNA opening; however, we observed considerable hysteresis problems, suggesting that it is difficult to sample all the degrees of freedom orthogonal to the PCV, such as all alternative protein–DNA interaction motifs. This observation further implies that our simulations provided a plausible pathway for DNA opening, but not necessarily the minimum free energy pathway. For instance, alternative pathways may involve sets of protein–DNA pair interactions in addition to the interactions presented in Figs. 4 and 5. To enable exhaustive conformational sampling and free energy calculations, future simulations may investigate the use of additional enhanced sampling techniques, such as bias-exchange umbrella sampling (62) or the use of extensive compute power (63).

In eukaryotes, the transcription factor TFIID catalyzes both DNA translocation and DNA opening by using the energy released by ATP hydrolysis (64–67). DNA opening can also be triggered by torsional stresses generated by negative supercoiling produced mostly by remote transcription processes; therefore, ATP is also the indirect energy source for DNA opening under negative supercoiling conditions (54,68–73). However, it has been suggested that translocase activity is not necessary for RNA transcription and, thus, that ATP-independent DNA opening is achievable by RNAP II (12) with the use of binding energy generated from PIC assembly (74). In this study, we modeled DNA opening in the absence of TFIID with the use of rotational and translational CVs. However, since TFIID also produces torsional stress to downstream DNA, our current protocol for DNA opening will be useful to study DNA opening in the presence of TFIID. Simulations with TFIID will be particularly relevant to understand the role of its XBP subunit, the TFIID subunit containing the translocase activity and the motor for DNA unwinding (66).

The transcription factor TFIIB contains the B-reader and B-linker elements, which also help DNA opening (10,75). Because refined atomic models of the B-reader and B-linker were not resolved in the CC structure by He et al. (7), we simulated the CC-to-OC transition in the absence of these TFIIB elements. Further atomistic simulations will be needed to investigate whether the B-reader and B-linker support DNA opening using similar interaction motifs as observed here for FL1, FL2, and for the rudder.

To conclude, we obtained an atomic model for a continuous DNA opening event in the human PIC. The simula-

tions revealed extensive interactions of the DNA with the protein, in particular with loops protruding into the polymerase cleft: FL1, FL2, and the rudder. According to the simulations, the loops play multiple roles for DNA opening: 1) by attacking WC H-bonds, they may catalyze the melting of the DNA; 2) extensive polar interactions via salt bridges with the DNA backbone and, to a lower degree, via H-bonds with the DNA backbone and bases stabilize the open DNA conformation; 3) FL2 tilted into the DNA bubble during opening, a conformational transition that is compatible with a function of FL2 as a sensor for an open transcription bubble.

SUPPORTING MATERIAL

Supporting material can be found online at <https://doi.org/10.1016/j.bpj.2022.10.012>.

AUTHOR CONTRIBUTIONS

J.S.H. and J.L. designed the research. J.L. carried out the simulations. J.L. analyzed the data with inputs from J.S.H. J.L. and J.S.H. wrote the article.

ACKNOWLEDGMENTS

This study was supported by the Deutsche Forschungsgemeinschaft via SFB 860/A16. We thank Christian Dienemann, Sandra Schilbach, Shintaro Aibara, and Patrick Cramer for insightful discussions.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Crick, F. 1970. Central dogma of molecular biology. *Nature*. 227: 561–563.
2. He, Y., J. Fang, ..., E. Nogales. 2013. Structural visualization of key steps in human transcription initiation. *Nature*. 495:481–486.
3. Cramer, P., D. A. Bushnell, and R. D. Kornberg. 2001. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science*. 292:1863–1876.
4. Gnatt, A. L., P. Cramer, ..., R. D. Kornberg. 2001. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science*. 292:1876–1882.
5. Nogales, E., R. K. Louder, and Y. He. 2017. Structural insights into the eukaryotic transcription initiation machinery. *Annu. Rev. Biophys.* 46:59–83.
6. Osman, S., and P. Cramer. 2020. Structural biology of RNA polymerase II transcription: 20 Years on. *Annu. Rev. Cell Dev. Biol.* 36:1–34.
7. He, Y., C. Yan, ..., E. Nogales. 2016. Near-atomic resolution visualization of human transcription promoter opening. *Nature*. 533:359–365.
8. Murakami, K., K. L. Tsai, ..., R. D. Kornberg. 2015. Structure of an RNA polymerase II preinitiation complex. *Proc. Natl. Acad. Sci. USA*. 112:13543–13548.
9. Plaschka, C., L. Larivière, ..., P. Cramer. 2015. Architecture of the RNA polymerase II-Mediator core initiation complex. *Nature*. 518:376–380.

10. Plaschka, C., M. Hantsche, ..., P. Cramer. 2016. Transcription initiation complex structures elucidate DNA opening. *Nature*. 533:353–358.
11. Schilbach, S., M. Hantsche, ..., P. Cramer. 2017. Structures of transcription pre-initiation complex with TFIIF and Mediator. *Nature*. 551:204–209.
12. Dienemann, C., B. Schwalb, ..., P. Cramer. 2019. Promoter distortion and opening in the RNA polymerase II cleft. *Mol. Cell*. 73:97–106.e4.
13. Yan, C., T. Dodd, ..., I. Ivanov. 2019. Transcription preinitiation complex structure and dynamics provide insight into genetic diseases. *Nat. Struct. Mol. Biol.* 26:397–406.
14. Schilbach, S., S. Aibara, ..., P. Cramer. 2021. Structure of RNA polymerase II pre-initiation complex at 2.9 Å defines initial DNA opening. *Cell*. 184:4064–4072.e28.
15. Aibara, S., S. Schilbach, and P. Cramer. 2021. Structures of mammalian RNA polymerase II pre-initiation complexes. *Nature*. 594:124–128.
16. Chen, J., C. Chiu, ..., S. A. Darst. 2020. Stepwise promoter melting by bacterial RNA polymerase. *Mol. Cell*. 78:275–288.e6.
17. Huang, X., D. Wang, ..., M. Levitt. 2010. RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription. *Proc. Natl. Acad. Sci. USA*. 107:15745–15750.
18. Feig, M., and Z. F. Burton. 2010. RNA polymerase II with open and closed trigger loops: active site dynamics and nucleic acid translocation. *Biophys. J.* 99:2577–2586.
19. Da, L.-T., D. Wang, and X. Huang. 2012. Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. *J. Am. Chem. Soc.* 134:2399–2406.
20. Wang, B., A. V. Predeus, ..., M. Feig. 2013. Energetic and structural details of the trigger-loop closing transition in RNA polymerase II. *Biophys. J.* 105:767–775.
21. Yu, J., L. T. Da, ..., X. Huang. 2014. Constructing kinetic models to elucidate structural dynamics of a complete RNA polymerase II elongation cycle. *Phys. Biol.* 12, 016004.
22. Silva, D. A., D. R. Weiss, ..., X. Huang. 2014. Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proc. Natl. Acad. Sci. USA*. 111:7665–7670.
23. Zhang, L., D. A. Silva, ..., X. Huang. 2015. Structural model of RNA polymerase II elongation complex with complete transcription bubble reveals NTP entry routes. *PLoS Comput. Biol.* 11:e1004354.
24. Unarta, I. C., L. Zhu, ..., X. Huang. 2018. Molecular mechanisms of RNA polymerase II transcription elongation elucidated by kinetic network models. *Curr. Opin. Struct. Biol.* 49:54–62.
25. Ka Man Tse, C., ..., X. Huang. 2019. Intrinsic cleavage of RNA polymerase II adopts a nucleobase-independent mechanism assisted by transcript phosphate. *Nat. Energy*. 2:228–235.
26. Unarta, I. C., S. Cao, ..., X. Huang. 2021. Role of bacterial RNA polymerase gate opening dynamics in DNA loading and antibiotics inhibition elucidated by quasi-Markov state model. *Proc. Natl. Acad. Sci. USA*. <https://doi.org/10.1073/pnas.2024324118>.
27. Shino, G., and S. Takada. 2021. Modeling DNA opening in the eukaryotic transcription initiation complexes via coarse-grained models. *Front. Mol. Biosci.* 8:817343–817412.
28. Grubmüller, H., B. Heymann, and P. Tavan. 1996. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*. 271:997–999.
29. Leech, J., J. F. Prins, and J. Hermans. 1996. SMD: visual steering of molecular dynamics for protein design. *IEEE Comput. Sci. Eng.* 3:38–45.
30. Abraham, M. J., T. Murtola, ..., E. Lindahl. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 1-2:19–25.
31. Tribello, G. A., M. Bonomi, ..., G. Bussi. 2014. Plumed 2: new feathers for an old bird. *Comput. Phys. Commun.* 185:604–613.
32. PLUMED consortium, Bussi, G., ..., A. White. 2019. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods*. 16:670–673.
33. Krieger, E., and G. Vriend. 2014. YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics*. 30:2981–2982.
34. Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
35. Zgarbová, M., J. Šponer, ..., P. Jurečka. 2015. Refinement of the sugar-phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J. Chem. Theory Comput.* 11:5723–5736.
36. Maier, J. A., C. Martinez, ..., C. Simmerling. 2015. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* 11:3696–3713.
37. Macchiagodena, M., M. Pagliai, ..., P. Procacci. 2019. Upgrading and validation of the AMBER force field for histidine and cysteine zinc(II)-Binding residues in sites with four protein ligands. *J. Chem. Inf. Model.* 59:3803–3816.
38. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
39. Miyamoto, S., and P. A. Kollman. 1992. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13:952–962.
40. Hess, B., H. Bekker, ..., J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.
41. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101.
42. Parrinello, M., and A. Rahman. 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52:7182–7190.
43. Feenstra, K. A., B. Hess, and H. J. C. Berendsen. 1999. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.* 20:786–798.
44. Branduardi, D., F. L. Gervasio, and M. Parrinello. 2007. From A to B in free energy space. *J. Chem. Phys.* 126, 054103.
45. Schrödinger, L., and W. DeLano. The PyMOL Molecular Graphics System, Version 2.3.0. <http://www.pymol.org/pymol>.
46. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD – visual molecular dynamics. *J. Mol. Graph.* 14:33–38.
47. UniProt Consortium, Martin, M. J., ..., J. Zhang. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49:D480–D489.
48. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.
49. Pal, M., A. S. Ponticelli, and D. S. Luse. 2005. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell*. 19:101–110.
50. Lankaš, F., R. Lavery, and J. H. Maddocks. 2006. Kinking occurs during molecular dynamics simulations of small DNA minicircles. *Structure*. 14:1527–1534.
51. Randall, G. L., L. Zechiedrich, and B. M. Pettitt. 2009. In the absence of writhe, DNA relieves torsional stress with localized, sequence-dependent structural failure to preserve B-form. *Nucleic Acids Res.* 37:5568–5577.
52. Mitchell, J. S., C. A. Laughton, and S. A. Harris. 2011. Atomistic simulations reveal bubbles, kinks and wrinkles in supercoiled DNA. *Nucleic Acids Res.* 39:3928–3938.
53. Irobalieva, R. N., J. M. Fogg, ..., L. Zechiedrich. 2015. Structural diversity of supercoiled DNA. *Nat. Commun.* 6:8851.
54. Parvin, J. D., and P. A. Sharp. 1993. DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell*. 73:533–540.

55. Feklistov, A., and S. A. Darst. 2011. Structural basis for promoter -10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell*. 147:1257–1269.
56. Klimasauskas, S., S. Kumar, ..., X. Cheng. 1994. HhaI methyltransferase flips its target base out of the DNA helix. *Cell*. 76:357–369.
57. Huang, N., N. K. Banavali, and A. D. MacKerell. 2003. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. *Proc. Natl. Acad. Sci. USA*. 100:68–73.
58. Najji, S., M. G. Bertero, ..., M. Thomm. 2008. Structure-function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement. *Nucleic Acids Res.* 36:676–687.
59. Barnes, C. O., M. Calero, ..., G. Calero. 2015. Crystal structure of a transcribing RNA polymerase II complex reveals a complete transcription bubble. *Mol. Cell*. 59:258–269.
60. Torrie, G. M., and J. P. Valleau. 1974. Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard-Jones fluid. *Chem. Phys. Lett.* 28:578–581.
61. Laio, A., and M. Parrinello. 2002. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA*. <https://doi.org/10.1073/pnas.202427399>.
62. Fukunishi, H., O. Watanabe, and S. Takada. 2002. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure prediction. *J. Chem. Phys.* 116:9058–9067.
63. Shaw, D. E., P. J. Adams, ..., K. A. Yuh. 2021. Anton 3: twenty microseconds of molecular dynamics simulation before lunch. In International Conference for High Performance Computing, Networking, Storage and Analysis.
64. Holstege, F. C., U. Fiedler, and H. T. Timmers. 1997. Three transitions in the RNA polymerase II transcription complex during initiation. *EMBO J.* 16:7468–7480.
65. Kim, T. K., R. H. Ebricht, and D. Reinberg. 2000. Mechanism of ATP-dependent promoter melting by transcription factor IIIH. *Science*. 288:1418–1422.
66. Grünberg, S., L. Warfield, and S. Hahn. 2012. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nat. Struct. Mol. Biol.* 19:788–796.
67. Fishburn, J., E. Tomko, ..., S. Hahn. 2015. Double-stranded DNA translocase activity of transcription factor TFIID and the mechanism of RNA polymerase II open complex formation. *Proc. Natl. Acad. Sci. USA*. 112:3961–3966.
68. Liu, L. F., and J. C. Wang. 1987. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci. USA*. 84:7024–7027.
69. Kouzine, F., S. Sanford, ..., D. Levens. 2008. The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat. Struct. Mol. Biol.* 15:146–154.
70. Naughton, C., N. Avlonitis, ..., N. Gilbert. 2013. Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat. Struct. Mol. Biol.* 20:387–395.
71. Kouzine, F., A. Gupta, ..., D. Levens. 2013. Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat. Struct. Mol. Biol.* 20:396–403.
72. Teves, S. S., and S. Henikoff. 2014. Transcription-generated torsional stress destabilizes nucleosomes. *Nat. Struct. Mol. Biol.* 21:88–94.
73. Corless, S., and N. Gilbert. 2016. Effects of DNA supercoiling on chromatin architecture. *Biophys. Rev.* 8:51–64.
74. Alekseev, S., Z. Nagy, ..., F. Coin. 2017. Transcription without XPB establishes a unified helicase-independent mechanism of promoter opening in eukaryotic gene expression. *Mol. Cell*. 65:504–514.e4.
75. Kostrewa, D., M. E. Zeller, ..., P. Cramer. 2009. RNA polymerase II-TFIIB structure and mechanism of transcription initiation. *Nature*. 462:323–330.