Molecular Basis for Impacts of DSIF on the Dynamics of RNA Polymerase II Elongation Complex

Weththasinghage D. Amith[a], Brandon M. Bogart[a], and Bercem Dutagaci[a,b*]

[a] Department of Molecular and Cell Biology, University of California, Merced, 5200 North Lake Rd. Merced, CA 95343, USA.

[b] Health Sciences Research Institute, University of California, Merced, 5200 North Lake Rd. Merced, CA 95343, USA.

[*] Corresponding author:
Bercem Dutagaci
5200 North Lake Rd.
Merced, CA 95343, USA
bdutagaci@ucmerced.edu
209-228-3603

ABSTRACT

Transcription elongation is a highly regulated process involving elongation factors associated with RNA polymerase II (Pol II). DRB sensitivity-inducing factor (DSIF) is an elongation factor known to have multiple roles in transcription elongation. Although studies resolved the structures of elongation complexes with DSIF, little is known about the impacts of DSIF on the dynamics of the elongation complex at the molecular level. Here, we used molecular dynamics simulations to elucidate the effects of DSIF on the dynamics and structure of Pol II and upstream nucleic acids, thereby gaining a mechanistic understanding of its role in transcription elongation. We determined three major sites of impact by DSIF, including the upstream nucleic acids, Pol II clamp, and active site, which potentially contribute to its role in transcription processivity. Our results showed that DSIF affects the dynamics of upstream DNA and RNA at the exit sites, preventing the unwinding of DNA and the folding of RNA. In addition, our results suggest that DSIF regulates the motion of the Pol II clamp to potentially maintain a proper size at the central cleft. We also observed a more dynamic active site and increased interactions between active site domains. Based on correlated motion analysis, we proposed that the impacts of DSIF on the active site have an allosteric nature that takes place through the collective motions of the Pol II clamp and nucleic acids.

1. INTRODUCTION

RNA polymerase II (Pol II) is the central enzyme in transcription of all mRNAs and many noncoding RNAs in eukaryotes [1]. Elongation factors accommodate Pol II in transition from the initiation stage of transcription to the stages of promoter-proximal pausing and subsequently elongation [2-4]. Elongation rate and processivity are highly regulated by the collective functions of various elongation factors, including DRB sensitivity-inducing factor (DSIF, also known as Spt4/5), Spt6, polymerase-associated factor 1 complex (PAF), and positive transcription elongation factor (P-TEFb) [5-17]. DSIF is a highly conserved elongation factor in all three domains of life and is formed by the association of Spt4 and Spt5 factors in Eukarya and Archaea and NusG (Spt5 homolog) in Bacteria. DSIF has important roles in promoter-proximal pausing [8,18-21], transcription processivity [6,10,16,22-24], and nucleosomal access [7,25-28], while the molecular mechanisms of the roles of DSIF in transcription elongation remain to be resolved.

Transcription processivity describes the process of transcription elongation in which Pol II remains activated and able to maintain elongation without interruptions [29]. Elongation factors are proposed to maintain transcription processivity by preventing the formation of paused/arrested states of Pol II or releasing Pol II from those states [19,30]. DSIF is one of the elongation factors that was shown to enhance processivity. Earlier studies demonstrated that DSIF stimulates Pol II elongation activity [16,31,32] and reduces or prevents the formation of arrested/paused states [6,33]. There are various hypotheses on the mechanisms of the DSIF function in maintaining transcription processivity. Some studies suggested that DSIF stabilized a closed clamp conformation, which is necessary for maintaining active elongation [24,34]. More

recently, in support of this hypothesis, another study demonstrated the formation of a closer clamp in the presence of Spt4/5 in an archaeal RNA polymerase using the cryo-electron microscopy (cryo-EM) technique [35]. Other studies focused on the interactions between Spt4/5 with DNA and suggested that Spt4/5 increases processivity by encompassing a DNA clamp, which prevents DNA from disengaging or forming secondary structures [6,36]. More recently, structural studies using Cryo-EM techniques showed that DSIF forms DNA and RNA clamps at the exit sites, supporting the importance of DSIF interactions with upstream nucleic acids in transcription processivity [12,22,23,28]. Furthermore, studies suggest an allosteric effect of DSIF on active site dynamics [6,24,31,33,37]. The allosteric function of DSIF was proposed to take place in two potential mechanisms, one is through the impacts of DSIF on the structure and dynamics of DNA and RNA [6,33,37], and the other is through the interactions of DSIF with the Pol II clamp domain [6,24,31]. In both mechanisms, DSIF is potentially impacting the structure and dynamics of the trigger loop (TL) and bridge helix (BH), both located near the active site and may help maintain processivity by preventing pausing and arrest [6] and promoting forward translocation [33,37].

Previous computational studies on transcription elongation predominantly focused on the Pol II active site in the absence of elongation factors [38-47]. These studies used molecular dynamics (MD) simulations, machine learning and Markov state modeling (MSM) and provided important insights into the roles of TL and BH in positioning the incoming NTP, transcription fidelity, translocation and backtracking. After cryo-EM structures of Pol II with elongation factors have been resolved by multiple research groups [12,22,23,25,28,48,49], our lab has started to study the Pol II-DSIF complex [50,51]. We previously showed that the interactions between DSIF and upstream nucleic acids are strongly sustained throughout MD simulations in the complete Pol II-

DSIF complex and isolated Spt5-upstream nucleic acid systems [50], supporting the importance of DSIF's impact on nucleic acid dynamics for its role in transcription elongation. Motivated by our earlier work, here we studied the Pol II elongation complex in the presence and absence of DSIF to investigate the impacts of DSIF on the structure and dynamics of the Pol II elongation complex. Our simulations revealed three important sites of impact of DSIF on the Pol II complex, which are upstream nucleic acids, active site domains, and the Rpb1-clamp domain. DSIF has a strong impact on upstream nucleic acid dynamics to prevent unwinding and secondary structure formation. For the active site domains, we observed increased dynamics, especially for TL. In addition, the interactions between TL and BH increased in the presence of DSIF, and we propose that these interactions may play an important role in the TL opening/closing motion. The impacts of DSIF on the clamp motion were more subtle, probably due to the difficulty in capturing globular motions with the simulation time scale. We observed that the distances of the clamp to the jaw and protrusion domains across the central cleft slightly increased. We propose that this increase reflects DSIF's role in ensuring proper clamp motion rather than closing the clamp further. Overall, based on our results, we suggest that these three impact points may work together to maintain the processivity of transcription.

2. METHODS

2.1. System Preparation.

We prepared three systems of the Pol II complex for this study. Two of the systems were prepared using the published structure with PDB ID: 5OIK [22]. In this structure, Pol II is from *Bos taurus* and DSIF is from *Homo sapiens*. For the first system, we used the complete complex with DSIF, which is referred to as +DSIF. For the second system, we removed DSIF (Spt4/5)

from the complex, which is referred to as -DSIF. In addition, we prepared another system without DSIF using the published Pol II complex structure (PDB ID: 5FLM) [52] and referred it as -DSIF_5flm. In this structure, Pol II is also from *Bos taurus*, and the two structures are similar, without any significant conformational differences (Table S1). We mostly discussed the results for +DSIF and -DSIF systems in the paper, however, we also presented the results of the -DSIF_5flm system and compared it with the -DSIF system. We modeled the missing residues of the TL using the MODELLER program [53]. The histidine at 1108 of Rpb1 was protonated, since it is the equivalent of H1085 in Pol II from *Saccharomyces cerevisiae*, which was suggested to be protonated by an earlier study [40]. Since the initial structure with PDB ID: 5FLM has shorter DNA and RNA at the exit sites, for the -DSIF_5flm system, we modeled upstream DNA and nascent RNA based on the structure with PDB ID: 5OIK by superimposing the two structures nucleic acid constructs. The systems were solvated in a cubic box by using at least a 9 Å cutoff from the edges of the simulation boxes. This resulted in simulation systems with system sizes of 180.2, 173.2, and 174.8 Å and numbers of atoms of 560,945, 497,751, and 514,233 for the +DSIF, -DSIF and -DSIF_5flm, respectively. We neutralize the systems by adding 116, 110 and 114 Na^+ ions in +DSIF, -DSIF and -DSIF_5flm systems, respectively. System preparation was performed by using the MMTSB package [54] together with the CHARMM software version 45b2 [55].

2.2. MD simulations.

The systems were minimized using 5,000 steps of energy minimization. Systems were then equilibrated for 1.6 ns while increasing the temperature from 100 to 300K and using constraints with a force constant of 400 and 40 kJ/mol/nm^2 for the heavy atoms of the backbone and sidechain of proteins, respectively. After equilibration, we performed 500 ns of MD simulations

for five replicates for +DSIF and -DSIF systems, which makes a total of 5 μs simulations. For the -DSIF_5flm system, we performed 250 ns of MD simulations for five replicates. Periodic boundary conditions and Particle Mesh Ewald algorithm [56,57] were applied for the calculation of long-range electrostatic interactions. Lennard-Jones interactions were switched between 10 and 12 Å. Bonds with H atoms were constrained using the SHAKE algorithm. The CHARMM 36m [58] and c36 [59] force fields were applied for proteins and nucleic acids, respectively. The CHARMM modified TIP3P model [60] was used for explicit water molecules. We applied mass repartitioning to the force fields by increasing the masses of H atoms to 3 a.m.u and decreasing the masses of heavy atoms attached to H accordingly, as suggested earlier [61]. This modification allows us to use a 4 fs time step for the production run. We saved the simulation trajectories every 100 ps and obtained 5,000 frames for each replicate for +DSIF and -DSIF systems and 1,500 frames for the -DSIF_5flm system. Simulations were performed using the OpenMM program [62] on GPU machines (single NVIDIA GPU for each simulation) without any bias, restraint, or enhanced sampling using Python scripts derived from the CHARMM-GUI server [63,64].

2.3. Analysis of the Simulations.

Simulations were analyzed for the last 400 ns (last 150 ns for -DSIF_5flm) of the production runs, since root mean square deviations (RMSD) for proteins and nucleic acids were relatively more stable after the first 100 ns of the simulations (Figs. S1-2). We analyzed the root mean square fluctuations (RMSF), contact and distance maps, pairwise distance distributions, radius of gyration (R_g), free energies for distances and conformations, bending angles for the BH domain, Watson-Crick hydrogen bonds, probability distributions of base stacking distance and nucleic acid dihedral angles and dynamic cross-correlation matrix (DCCM) for pairs of protein domains

and nucleic acids. For nucleic acids, we analyzed upstream DNA, residues from -10 to -23 and nascent RNA, the first 12 residues from the 5` end (see Fig. 1). The first 12 residues of RNA were numbered as 1-12 in 5FLM and 31-42 in 5OIK structures.

RMSD values were calculated for C_α and P atoms after superimposing them to the initial structures for proteins and nucleic acids, respectively. RMSF values were calculated for the same atoms with RMSD, but they were superimposed against the average structure throughout the trajectory.

Distance maps and contact maps were calculated over the trajectories and averaged over the replicates for each system. For the contact maps, two residues were assigned to have contact in a specific time frame if the minimum distance between heavy atoms of the residues is within 5 Å.

To obtain free energy for conformations of TL, BH, upstream DNA and nascent RNA, we first extracted the coordinates of C_α and P atoms of each amino acid or nucleotide, respectively, over the trajectory after superimposing them to the initial structure. Then, we applied principal component analysis (PCA) to reduce the dimensionality. We considered the first and second principal components (PCs) for free energy calculation. We calculated free energies by applying the weighted histogram analysis method (WHAM) using the WHAM package [65]. For the free energy based on Pol II clamp distances, we calculated the minimum distances between heavy atoms for each pair over the trajectory and then applied PCA to reduce dimensionality. This analysis was applied for two sets of distances: one is the distances between Rpb1 clamp residues from 280 to 315 and Rpb2 protrusion residues from 405 to 445 and the other is the distances between Rpb1 clamp residues from 118 to 155 and Rpb1 jaw residues from 1229 to 1297 (except the missing residues of the jaw domain in the structure).

We calculated the bending angle of BH as the angle between the principal axes of helices with residue selection of 835-852 and 853-870. For this, we first determined the covariance matrix of the C_α coordinates of the residues within each helix and then calculated eigenvalues and eigenvectors of the covariance matrix. After doing this calculation for the two helices, we calculated the angle between the first principal vectors, the eigenvectors with the highest eigenvalues for each helix, for the two helices from their dot products.

Watson-Crick hydrogen bonds were calculated for the upstream DNA. Hydrogen bonds were calculated between donor and acceptor atoms when they are within 1.2 Å and with a cutoff of 150° for the angle between acceptor-hydrogen-donor atoms. Base-staking distances were calculated between the neighboring bases as their center of mass distances for the heavy atoms. The dihedral angles were calculated for each nucleotide for both non-template and template DNA and distributions were calculated over the trajectory.

DCCMs were calculated for C_α and P atoms for proteins and nucleic acids, respectively, from the normalized covariance $C(i,j) = \langle \Delta\mathbf{r}_i \cdot \Delta\mathbf{r}_j \rangle / (\langle \Delta\mathbf{r}_i^2 \rangle \langle \Delta\mathbf{r}_j^2 \rangle)^{1/2}$, where $\mathbf{r}_{i/j}$ is the displacement vector from the average position and angle brackets denote the average over the trajectory frames. Correlated motion changed between -1 to 1, which means that the motions of two residues are negatively or positively correlated, respectively.

Analyses were performed using the Python programming language together with Scikit-learn [66], NumPy [67], and MDAnalysis [68,69] Python libraries, and the MMTSB package [54]. The VMD program [70] was used to visualize the structures obtained from the simulations.

3. RESULTS

In this study, we investigated the impact of DSIF on the dynamics and structure of the Pol II elongation complex to gain mechanistic insights into the role of DSIF in transcription processivity. We performed MD simulations in the presence and absence of DSIF and probed the effects of DSIF on three impact points (Fig. 1), which are the upstream nucleic acids, Rpb1-clamp, and active sites domains.

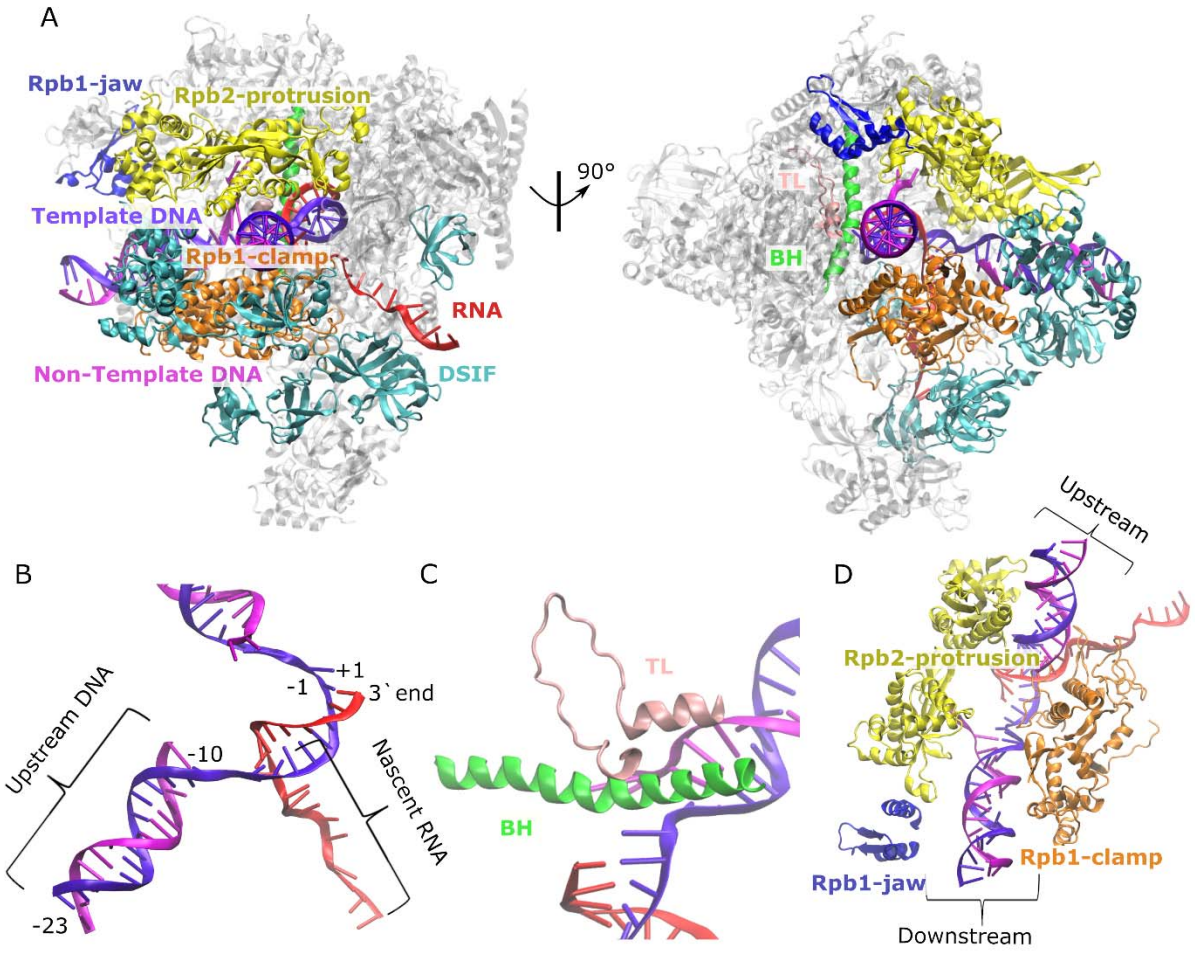

Figure 1. Structure of the initial Pol II – DSIF system. Structure of the complete complex (PDB ID: 5OIK) with DSIF, DNA, RNA, Rpb1 clamp and jaw domains and Rpb2 protrusion domain highlighted (A). The focused regions of the complex were zoomed in, which are upstream DNA and nascent RNA

shown within curly parenthesis (B), TL and BH domains at the active site (C) and Rpb1-clamp, Rpb1-jaw and Rpb2-protrusion domains (D).

3.1. DSIF stabilizes DNA and RNA at exit sites.

Upstream DNA and nascent RNA are known to interact with DSIF, as determined by the cryo-EM studies [12,22,23,28]. At the DNA exit, DSIF forms a "DNA clamp" by Spt5-NGN and KOW1 domains and Spt4 together with Rpb1-wall and Rpb2-protrusion domains [22,23]. Similarly, at the RNA exit, the KOW4 and KOW5 domains form an "RNA clamp" presumably to help the proper exit of RNA [22,23]. We performed analyses to investigate how the interactions with DSIF impact the dynamics of DNA and RNA. RMSF analyses suggested that DNA (template and non-template) has reduced RMSF in the presence of DSIF (Fig. 2A), suggesting a reduced flexibility for the DNA at the exit site. However, differences in the RMSF value of RNA are less significant since there are large error bars, suggesting that fluctuations differ for each replicate simulation. The impact on the RMSF is stronger for DNA, which could be partly due to the overall higher dynamics of the single-stranded nascent RNA compared to base-paired DNA. The increased stability suggested by RMSF values could be the result of the impact of the increased interactions between protein and nucleic acids. The total number of contacts for DNA increased more than twofold for +DSIF, in which DSIF makes 74 % of the contacts with the upstream DNA (Table 1) and those contacts are mostly with Spt4, Spt5-NGN and KOW1 domains (Fig. S3). On the contrary, for RNA, the total number of contacts slightly decreased for +DSIF (Table 1). Spt5 started to have a large percentage of contacts, mostly with the KOW4 domain (Fig. S4), that possibly play a role in stabilizing RNA at the exit site. On the other hand, the contacts with the Rpb2-wall domain were significantly reduced in the presence of DSIF (Fig. S4B).

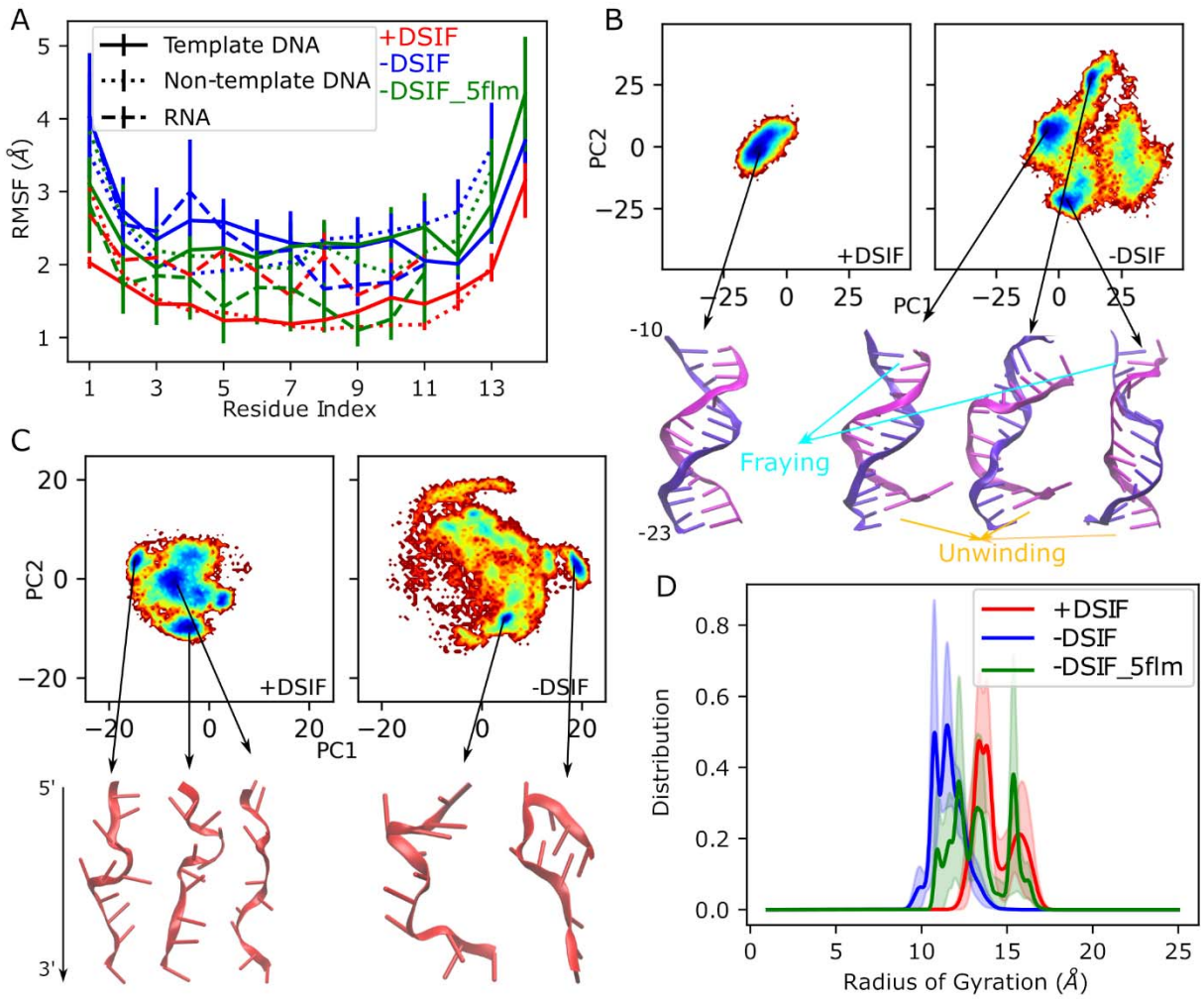

Figure 2. Structure and dynamics of upstream DNA and nascent RNA. RMSF values were reduced for DNA in the presence of DSIF (A). Free energy profiles of DNA (B) and RNA (C) conformations showed more stable structures for DNA and more extended structures for RNA in the presence of DSIF. Free energy profiles were calculated from the first two principles components (PC1 and PC2) from the PCA based on Cartesian coordinates (see Methods for details). The radius of gyration analysis for RNA (D) suggests increased number of compact structures in the absence of DSIF.

Next, we analyzed the structural changes occurring in DNA and RNA in the presence of DSIF. We observed larger RMSD values in the absence of DSIF (Fig. S2), which suggests that nucleic acid structures are deviating from the initial structure in the absence of DSIF. To understand the

nature of structural deviation, we analyzed the conformations of the upstream DNA and nascent RNA. We performed PCA on the Cartesian coordinates of P atoms of DNA and RNA from the superimposed frames along the trajectory. Fig. 2B shows that the DNA conformation is mostly stable in the presence of DSIF. On the other hand, a broader conformational space was observed for DNA when DSIF was removed. The minimum energy structure for the +DSIF system shows a base-paired DNA structure without any significant fraying or unwinding. However, the three minimum energy structures for the -DSIF system show significant deviation from the initial structure and fraying and unwinding were observed. We observed a similar conformational deviation for the DSIF_5flm system, for which two populated structures showed significant unwinding (Fig. S5A). To better understand the structural changes, we analyzed Watson-Crick base-pair hydrogen bonds, base stacking distances, and dihedral angles. The number of base-pair hydrogen bonds slightly decreased for the -DSIF system and significantly reduced for the -DSIF_5flm system (Fig. S6A). For the -DSIF_5flm system, large errors suggest that some of the replicates showed significantly decreased base-pairs than the others, which could be a result of an unstable initial structure due to the modeling of upstream DNA for this system. We also observed a partial increase in base stacking distances, which appears as a small shoulder in the distance distribution plot of -DSIF and -DSIF_5flm systems (Fig. S6B). These increased base stacking distances reflect the fraying we observed in the most populated structures shown in Figs. 2B and S5A. For the dihedral angles, we observed changes in zeta and chi angles, which show slightly broader distributions for the -DSIF and -DSIF_5flm systems (Fig. S7). The changes in these angles further suggest a deviation from the base-pair structures in the absence of DSIF.

Table 1. Number and percentage of contacts of DNA and RNA with protein.

+DSIF				
Protein	Number of DNA contacts	Percentage of DNA contacts	Number of RNA contacts	Percentage of RNA contacts
Total	109.98 ± 5.68	100	112.99 ± 3.99	100
Spt4	18.94 ± 3.60	17.22	0.0 ± 0.0	0
Spt5	62.36 ± 5.32	56.70	35.05 ± 3.79	32.02
Rpb1	5.91 ± 1.93	5.38	28.92 ± 4.77	25.60
Rpb2	22.76 ± 1.37	20.7	49.02 ± 2.69	43.38
-DSIF				
Protein	Number of DNA contacts	Percentage of DNA contacts	Number of RNA contacts	Percentage of RNA contacts
Total	46.90 ± 7.38	100	119.09 ± 6.26	100
Rpb1	12.53 ± 8.14	26.72	50.29 ± 4.71	42.23
Rpb2	34.37 ± 6.05	73.28	68.77 ± 3.95	57.75
-DSIF_5flm				
Protein	Number of DNA contacts	Percentage of DNA contacts	Number of RNA contacts	Percentage of RNA contacts
Total	38.88 ± 6.97	100	123.22 ± 11.79	100
Rpb1	12.09 ± 7.85	31.10	60.78 ± 5.46	49.33
Rpb2	26.78 ± 5.89	68.90	61.34 ± 11.50	49.78

The number of contacts was calculated over the trajectories and averaged over the number of replicate simulations. DNA contacts were calculated for the upstream DNA (from -10 to -23). RNA contacts were calculated for the first 12 residues of nascent RNA where it contacts with DSIF.

RNA conformations were sampled in large spaces for both systems, although for -DSIF, the conformational space was somewhat broader (Fig. 2C). The minimum energy structures for +DSIF are mostly elongated, while for the -DSIF system, the structures are more compact, with one structure forming a hairpin. Similar results were observed for -DSIF_5flm, such that the mostly populated structure was compact and hairpin formation was observed for the second

populated region (Fig. S5B). To understand the impact of DSIF on the compactness of RNA structure, we calculated R_g values over the simulation trajectories. We observed that R_g values are significantly lower for the -DSIF system and lower Rg values were also observed for the -DSIF_5flm system (Fig. 2D), suggesting that DSIF prevents RNA from forming compact structures, including hairpins.

Overall, our results suggest that DSIF forms strong interactions with upstream DNA and nascent RNA, reduces the dynamics of DNA at the exit tunnel, and prevents them from forming unusual structures, like frayed base-pairs for DNA and hairpin-like secondary structures for RNA.

3.2. DSIF alters the dynamics and interactions of TL and BH.

DSIF does not have direct interactions with TL and BH, both are near-active site domains. However, earlier studies suggested that DISF may affect the Pol II active site allosterically [6,24,31,33,37], though the molecular basis of such allosteric impacts is not known. To explore the impacts of DSIF on the active site, we analyzed the structure and dynamics of TL and BH. RMSF plots showed increased flexibility for TL in the presence of DSIF especially when comparing with the -DSIF system, without any significant change for BH (Fig. 3A). The increased flexibility for TL was observed in the loop region between residues 1109-1115, which was shown in color in Fig. 3B. TL is known to play a key role in the nucleotide addition cycle (NAC), and it alternates between “open” and “closed” structures during the NAC [40,44,71-77]. TL in the closed state is proposed to facilitate nucleotide selection and position the incoming nucleotide for catalysis, while the open TL is proposed to play roles in nucleotide binding and forward translocation. In the cryo-EM structure used in our simulations, TL is partially unresolved due to high flexibility in its open state. We modeled the unresolved part and showed

the comparison with the closed state in Fig. 3C. In the presence of DSIF, increased flexibility occurred in this modeled region, which is the tip of the TL that normally moves toward to active site to form the closed state. Based on the increased dynamics on the tip of TL in the presence of DSIF, we hypothesize that DSIF might help facilitate the open-closed TL transition, potentially stabilizing the closed TL.

To gain more insight into the effects of DSIF on TL conformation, we performed PCA on TL C_α coordinates. Fig. 3D shows that TL samples a larger conformational space in the presence of DSIF, which is consistent with higher dynamics observed in the RMSF plots. The free energy plots show multiple minimum energies. We extracted the structures of TL that are closest to the globular minimum in the free energy plots. We observed that for the +DSIF system, the minimum energy structure showed a larger deviation from the initial modeled structure (Fig. 3C), as expected from the larger deviations in the RMSD plots (Fig. S1). BH also sampled a somewhat larger conformational space for the +DSIF system (Fig. 3E), while the lowest energy appears in the same region for both systems, and the corresponding conformations extracted from the trajectories are also structurally similar, with bending angles of 156.2° and 159.7° for +DSIF and -DSIF systems, respectively. In addition, bending angle distributions throughout the trajectories are broad with large errors and, therefore, did not show any significant differences for +DSIF and -DSIF systems (Fig. S8). On the other hand, the -DSIF_5flm system showed sharper distribution for the BH bending angles and the maximum angle (161.2°) appears close to the maximum angle of -DSIF (160.9°) and smaller compared to +DSIF (164.0°). This suggests that in the absence of DSIF, BH tend to bend more, while in the presence of DSIF, BH samples between bent and straight conformations manifested themselves as a broad bending angle distribution.

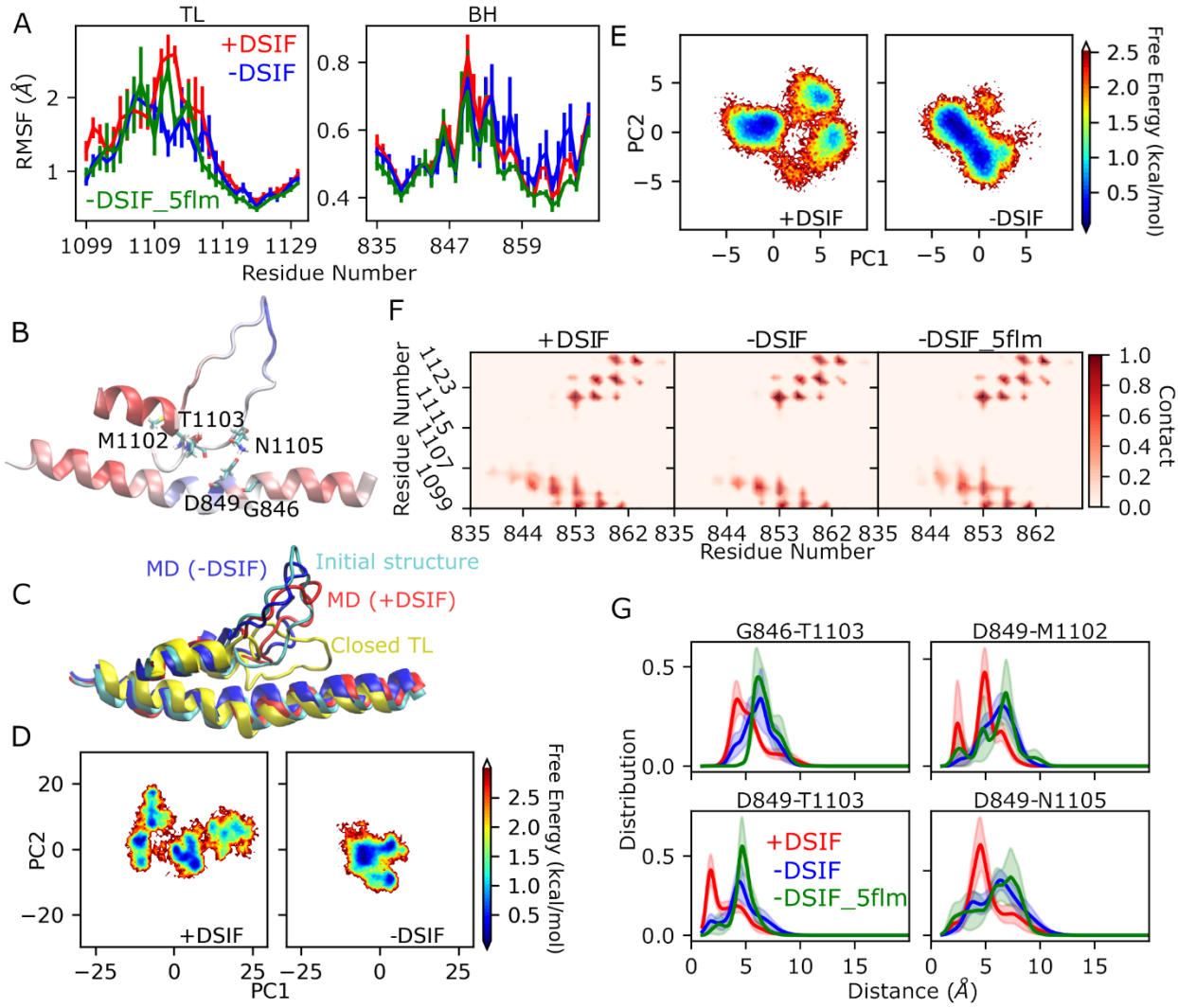

Figure 3. Structure and dynamics of TL and BH. RMSF values for BH and TL showed differences only for TL with increased flexibility in the presence of DSIF (A), and RMSF values are shown in colors from red to blue in increasing order in the structure of TL and BH, the color ranges were RMSF values of 0.40-0.82 for BH and 0.60-2.58 for TL (B). The initial modeled structure of TL, the closed TL structure from PDB ID: 2E2H, and the structures from +DSIF and -DSIF simulations extracted based on PCA are superimposed (C). PCA showed larger conformational spaces for both TL (D) and BH (E), and the structure for TL showed larger deviation from the initial model in the presence of DSIF (C). The number of contacts between BH and TL increased in the presence of DSIF (F), and in consistency with the

number of contacts, the distances between selected TL-BH pairs were reduced (G). These selected pairs are located at the BH-TL interface that have high flexibilities, which are colored by RMSF values from the +DSIF system (B).

Although we did not observe any significant changes in the BH structure and dynamics, we observed changes in its interactions with TL. The number of contacts between the two domains increased for the BH residues between 840-849 and the TL residues between 1102-1105 for +DSIF system (Fig. 3F). To understand which contacts significantly differ for the two systems, we calculated the distance distributions for the pairs that have normalized number of contacts more than 0.7 in any of the two systems and the difference of the contacts between the two systems is larger than 0.1. Fig. 3G shows the distance distributions for the pairs that show the most significant changes for the two systems, as well as for the -DSIF_5flm system, while the distances for the other pairs were only moderately different (Fig. S9). We observed that BH and TL distances are similar for -DSIF and -DSIF_5flm systems and decreased for the +DSIF system, suggesting increased interactions between the two domains for the TL residues M1102, T1103 and N1105 and the BH residues G846 and D849 in the presence of DSIF. The positions of these residues were shown in Fig. 3B. An earlier MD simulation study suggested that TL-BH interactions around that region of the TL play an important role in the TL dynamics [45] supporting our hypothesis that DSIF may play a role in the TL opening/closing motion. Overall, our results suggest that DSIF may increase the dynamics of TL and its interactions with BH and decrease the bending of BH.

3.3. DSIF regulates the clamp conformation at the central cleft.

Earlier studies suggest that DSIF may help maintain the clamp in its closed conformation, thereby enclosing the DNA for transcription elongation [24,34,35,78]. To understand the impact of DSIF on the size of the central cleft, we analyzed the distances of the Rpb1-clamp to Rpb1-jaw and Rbp2-protrusion at downstream and upstream, respectively. We calculated all the pairwise distances for the selected helical regions of the clamp and jaw and the coil-coiled domains of the clamp and protrusion (Fig. 4A, the residue selection is provided in the Methods section). Fig. 4B shows the free energy profiles of the PCs that were obtained by applying PCA on the pairwise distances. Energy profiles span similarly large spaces for +DSIF and -DSIF systems for clamp-jaw distances, while for clamp-protrusion distances, the +DSIF system spans a narrower space, suggesting that direct interactions of DSIF with the clamp stabilize clamp motion at the upstream.

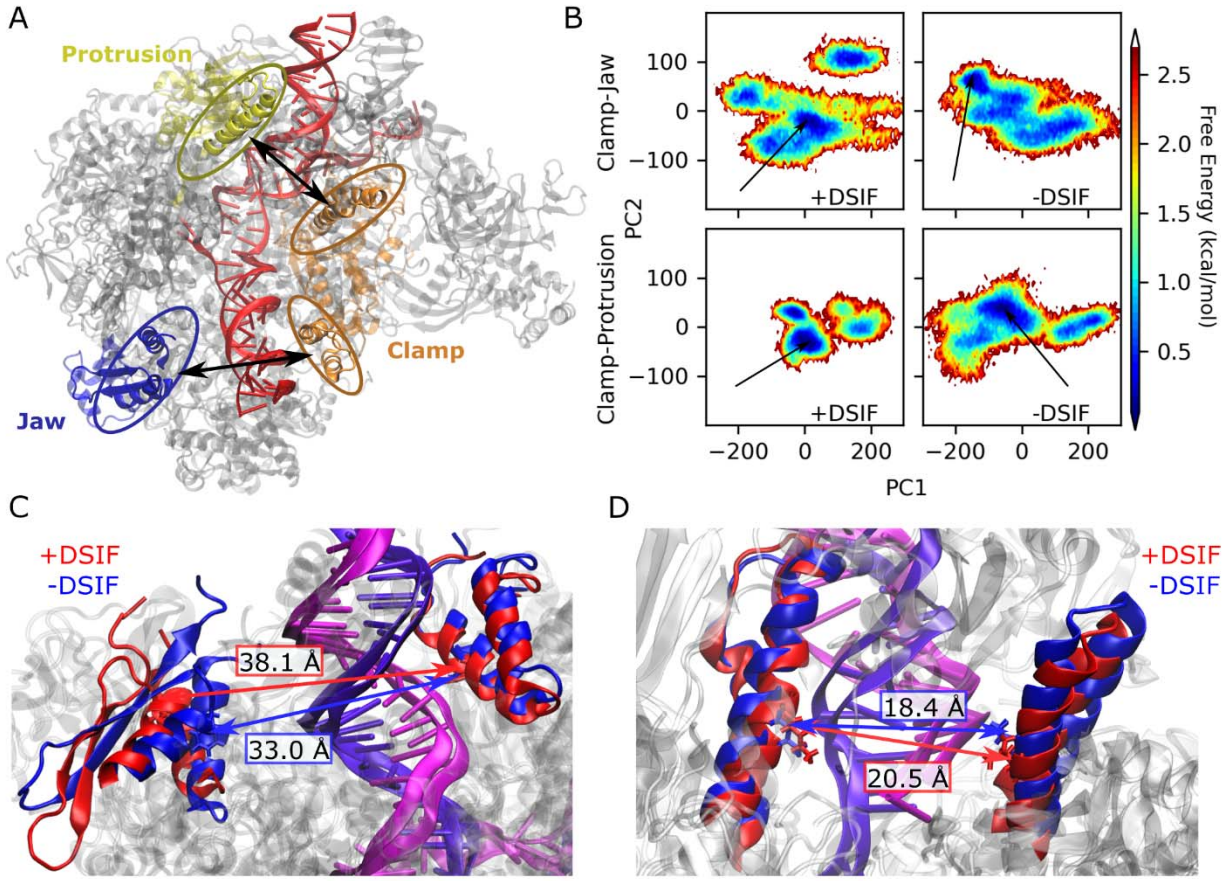

Figure 4. Analysis of clamp motion. The architecture of the central cleft formed by Rpb1-clamp and other domains is demonstrated in a cartoon representation (A). The free energy profiles of PCA based on pairwise distances between Rpb1-clamp and Rpb1-jaw, and between Rpb1-clamp and Rpb2-protrusion (as colored in panel A) are shown (B). The minimum energy structures from the PCA showed that the distances increased in the presence of DSIF, both for clamp-jaw (C) and clamp-protrusion (D).

We then extracted the most populated structures based on the distance PCA free energy profiles and visualized them after superimposing the Rpb1 subunits for +DSIF and -DSIF systems (Fig. 4C-D). Clamp and jaw domains became relatively closer after we removed DSIF (Fig. 4C), suggesting that DSIF may play a role in regulating the distance between the two domains to maintain the proper size in the cleft. Consistently, distance maps also show a decrease in distances between the clamp and jaw domains in the absence of DSIF (Fig. S10). In addition,

distance distribution profiles for the selected residue pairs showed significant differences between the two systems, with the +DSIF system showing overall larger distances (Fig. S11). On the other hand, -DSIF_5flm system showed somewhat increased distances in the distance map (Fig. S10) and distance profiles for selected residues show broader distributions than the other two systems, with a smaller shoulder on the larger distance region (Fig. S11). This may be due to the shorter simulation times for the -DSIF_5flm system, however, overall, the discrepancies between -DSIF and -DSIF_5flm systems may suggest that the globular conformational changes, like clamp motion, may not be fully accessible with nanosecond time scale simulations.

At the upstream, the protrusion domain slightly came closer to the coil-coiled domain of the clamp in the absence of DSIF (Fig. 4D). Consistent with the extracted structures, distance maps also show reduced distances for both -DSIF and -DSIF_5flm (Fig. S10). More prominently, we observed sharper distance profiles (Fig. S12) in the presence of DSIF compared to both -DSIF and -DSIF_5flm systems, suggesting that DSIF stabilizes the coil-coiled domain of the clamp. Since there are direct interactions between DSIF and the coil-coiled domain (Fig. S13), we concluded that direct interactions reduced the dynamics of the coil-coiled domain and consequently caused sharper distance profiles. Overall, our results suggest that DSIF may help maintain the distances of the clamp to the other side of the cleft, while removing DSIF decreases these distances, which may potentially impact transcription processivity. In addition, the initial structures with and without DSIF have similar clamp conformations (Table S1), and the pairwise distances of clamp residues to the jaw and protrusion residues were similar for the two structures (cyan and purple lines in Figs S11-12). In our simulations, we observed overall reduced pairwise distances compared to initial structures with a broader distribution for the systems without DSIF. Therefore, the simulations provided a range of dynamics for clamp motion and suggested that

DSIF partly reduces those dynamics. On the other hand, we observed discrepancies in the distances between the clamp and jaw domains for the simulations of the two systems without DSIF. Therefore, we acknowledge that our simulations may not provide results relevant to long time scale globular motions including clamp motion at the central cleft.

3.4. Motions of active site loops are correlated with clamp and RNA motions.

The impacts of DSIF on the active site have been associated with the allosteric effects of either nucleic acids or Rpb1-clamp. To understand the presence of any allosteric effect, we calculated the correlated motions (see Methods for DCCM calculations) for TL and BH with upstream nucleic acids and Rpb1-clamp. We provided DCCM maps colored in red for both positive and negative correlations in Fig. 5A, while the colors coded by the sign of correlations can be seen in the supplementary figures (see Figs. S14-16). First, we calculated the correlated motion between BH and TL and observed that DSIF increased the correlation between them compared to -DSIF and -DSIF_5flm systems (Figs. 5A and S14), consistent with the increased interactions between these two domains. For the nascent RNA, in the absence of DSIF, for both -DSIF and -DSIF_5flm systems, correlations with TL and BH increased (Figs 5A, S15 and S16), especially for the regions shown with the green arrow in Fig. 5A. Such an increase in the TL/BH-RNA correlated motions could be related to RNA hairpin formation as studies suggested that hairpin formation in the exiting RNA may have an increased effect on the active site structure and dynamics [79-85]. We also observed somewhat increased correlated motions between DNA and TL (Fig. S16), which suggests that both DNA and RNA may have an allosteric impact on TL motion.

For the clamp, we observed changes in three regions in the correlation motions with and without DSIF for TL (Fig. 5A) and similarly for BH (Fig. S14). DSIF decreased the correlations with TL

for the clamp residues 11-92. This region interacted with KOW2-4 domains of the DSIF (Fig. S17), and these interactions potentially reduced its dynamics (Fig. S18). Therefore, we concluded that decreased dynamics reduced the impact of this region on the active site motion. On the contrary, the regions with residues 93-150 and 273-350, which include the coiled-coil domain and close-by residues, showed increased correlations with TL in the presence of DSIF (Figs. 5A and S14). More strikingly, we observed decreased correlations between the clamp and TL loop region (residues 1107-1118) for both -DSIF and -DSIF_5flm systems. This suggests that DSIF impacts the loop region of the TL allosterically by altering its correlated motion with the Rpb1 clamp.

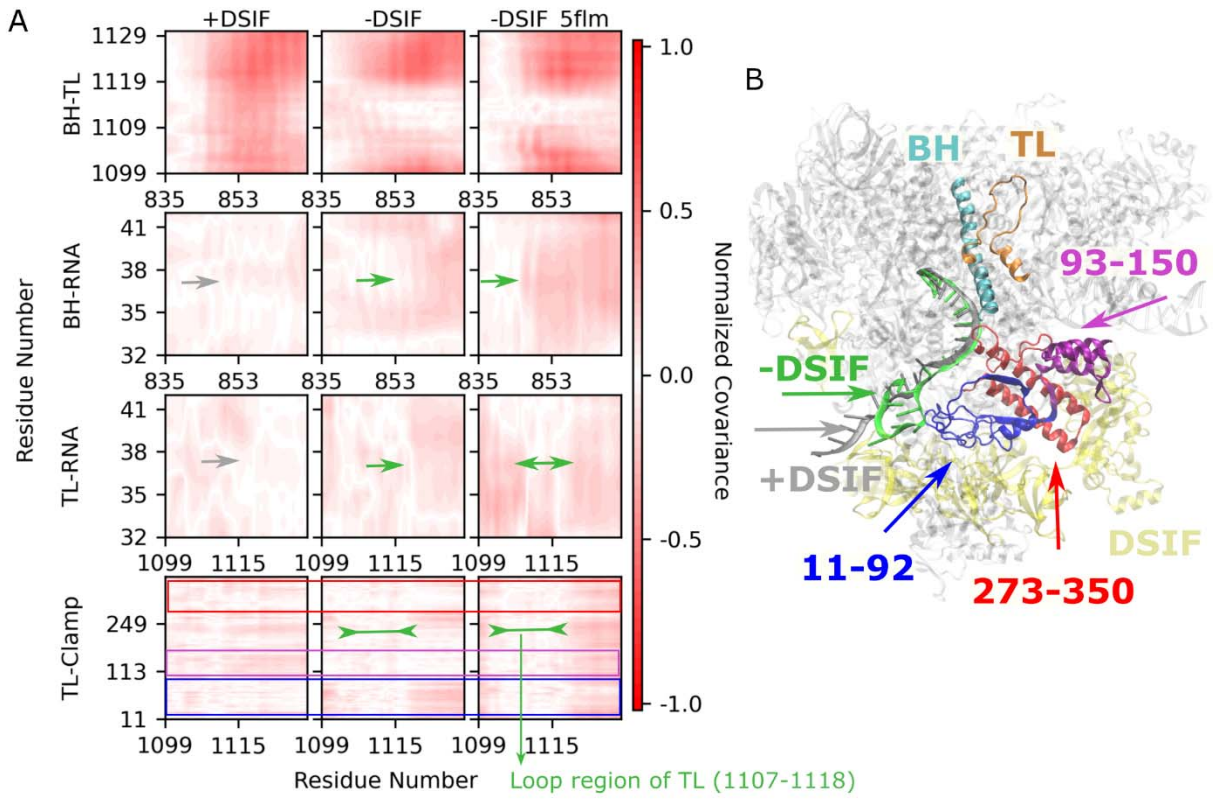

Figure 5. <u>DCCM analysis for protein domains and nucleic acids.</u> Correlated motions decreased between BH and TL and increased between BH and RNA, and TL and RNA in the absence of DSIF (A). In the

presence of DSIF, correlated motion decreased for TL with clamp residues 11-92, increased with clamp residues 93-150 and 273-350, with the correlations of the TL loop region was most impacted (A). Secondary structure formation of RNA and motion of clamp residues correlated with the TL dynamics, suggesting an allosteric impact of DSIF on TL motion (B). The residue ranges 11-92, 93-150, and 273-350 in panel B are highlighted in the same color as cross correlation plot for TL and clamp in panel A to emphasize that these are the regions with changes in cross-correlation maps.

Overall, we suggest that the collective motion of the clamp residues and RNA structural changes impacts motions of TL and BH in the presence of DSIF (Fig. 5B). DSIF affected the correlations of the TL with the clamp residues and, potentially, these correlated motions caused increased flexibility of the TL that we observed in the RMSF plots (Fig. 3A). In addition, DSIF prevents RNA from forming hairpin-like secondary structures and, by this, reduces the impact of RNA on the active site dynamics.

4. DISCUSSION

DSIF is the only universally conserved elongation factor, which is known to have regulatory functions in various steps of transcription. Although it is widely studied, its regulatory mechanism remains unresolved. To obtain mechanistic insights into the DSIF function at the molecular level, we performed MD simulations of the Pol II elongation complex in the presence and absence of DSIF. The most notable impact of DSIF was observed in the nucleic acid structure and dynamics at the exit sites. We observed that DSIF stabilized the upstream DNA, causing it to sample a narrower conformational space, preventing it from forming frayed states and potentially unwinding. DSIF's impact on maintaining upstream DNA structure may be a contributing factor to its role in processivity. Structural studies suggest that DSIF forms a DNA clamp at the upstream together with Rpb1-wall and Rpb2-protrusion domains [22,23]. Forming

the DNA clamp helps stabilize DNA at the exit site, proposed to prevent it from unwinding and potentially rewinding with exiting RNA. Our observation of reduced unwinding of the DNA base pair at the exit site in the presence of DSIF supports this earlier hypothesis on the role of the DNA clamp formed by DSIF. In addition, DSIF demonstrated an even larger impact on the structure of the nascent RNA, preventing it from forming more compact secondary structures, including hairpins. Earlier studies suggest that the hairpin formation of RNA allosterically affects the polymerase active site and stimulates pausing [79-85]. In the light of these earlier studies, our results suggest that DSIF may promote processivity by helping RNA maintain more extended secondary structures at the exit site and consequently preventing RNA from allosterically causing the formation of paused states.

DSIF has also been proposed to impact the clamp motion, potentially stabilizing a closed clamp and, in this way, maintaining processivity [24,34,35,78]. Therefore, we expected to observe the opening of the clamp in our simulations after removing DSIF. On the contrary, we observed a slightly more closed clamp as the distances between the clamp and the domains on the other side of the cleft reduced in the absence of DSIF. Additionally, the clamp, especially the CC domain of the clamp, becomes more dynamic in the absence of DSIF, suggesting that DSIF may play a role in stabilizing the CC domain and, consequently, clamp motion. Based on our results, we propose that, instead of further closing the clamp, DSIF may help maintain the proper size for the central cleft by stabilizing the clamp motion and preventing it from opening or closing further to form unfunctional conformational states. However, we also note that our conclusion is limited to the simulation time scale, which may not be sufficient to completely sample the globular conformational changes in the central cleft. Especially, the discrepancies in the clamp-jaw

distances from the simulations of -DSIF and -DSIF_5flm systems suggest that the clamp motion may not be accurately captured for the nanosecond simulation time scale.

Furthermore, we observed changes in the dynamics of TL and the interactions between TL and BH. TL showed increased flexibility for the residues from Y1109 to K1115 (F1086 to K1092 in yeast), in the presence of DSIF. Previous studies suggest that residues F1086 and V1089 of the TL in yeast Pol II demonstrated loss-of-function mutations [86], and earlier structures [87] suggest an important role for these residues in TL closing dynamics as partial closing of the TL correlated with the loss of the hydrogen bond between the backbones of F1086 and V1089 (Y1109 and V1112 in the mammalian structure). We did not observe any hydrogen bond formation for the backbones of the two residues in either of the simulations. However, the increased flexibility of the TL tip, including the residues Y1109 and V1112 in the presence of DSIF, may suggest DSIF's impact on the TL closing motion. We also observed increased interactions between BH and TL, especially for TL residues M1102, T1103, and N1105 and BH residues G846 and D849. BH and TL interactions are known to be important for the catalytic function [38,76,77,86] and transition between closed and open TL [45]. Also, a decrease in BH-TL interactions was proposed to be one of the mechanisms for the loss-of-function phenotypes of TL mutants [38]. Here, based on our results, we hypothesize that DSIF impacts the TL opening/closing dynamics, which is not feasible to model by standard MD simulations that we applied here. One of our future directions is to perform enhanced sampling simulations to investigate the impacts of DSIF on the energetics of the transition between the open and closed TL, as has been done in an earlier study performed in the absence of DSIF [45].

DSIF was proposed to affect the translocation dynamics during the nucleotide addition cycle [6,25,36,88], Studies suggested that NusG, the bacterial homolog of Spt5, stabilized the post-

translocation state by presumably affecting the active site domains [33,37]. We observed three main impacts of DSIF on active site domains, which are increased loop dynamics of the TL, increased interactions between TL and BH, and broader bending angle distributions of BH. The first two effects on TL suggest that there is a destabilizing effect on TL in its open state, which would be accompanied by increased TL motion toward BH and formation of BH-TL contacts. According to the proposed nucleotide addition cycle, Pol II oscillates between pre- and post-translocation states when the TL is open and the concerted motion of TL and BH moves Pol II to the post-translocation state [39,89]. Our observation of increased TL dynamics and TL/BH interactions may indicate a change in the pre-post translocation equilibrium towards post-translocation and, therefore, may support the earlier hypothesis on stabilization of post-translocation by DSIF. On the other hand, BH bending is proposed to be coupled with TL closing to support post-translocation [37,90]. Our results suggest a broader bending profile compared to the systems without DSIF, with overall less bending. This is not consistent with the proposed role of the BH bending in stabilizing the post-translocation state. However, increased dynamics on the BH bending and straightening may also play a role in affecting the equilibrium between pre- and post-translocation states by the impact of DSIF, as such a role for NusG was suggested by previous experimental studies [33,37].

The role of DSIF in processivity has been associated with its allosteric impact on the structure and dynamics of the active site [6,24,31,33,37], especially through the Rpb1-clamp and upstream nucleic acids. To understand allosteric interactions, we calculated the correlated motions by constructing DCCM for TL and BH with the clamp and nucleic acids. Our results suggest that the clamp has altered correlated motion with TL and BH domains in the presence of DSIF. We especially observed increased correlations for the loop region of TL with the clamp residues.

This is consistent with the earlier studies that focused on the allosteric impact of the clamp and its CC domain [31,34,78]. However, the molecular mechanisms of how the clamp impacts the TL dynamics to promote processivity are not well understood. Hypotheses include altered clamp position by DSIF interactions at the coiled-coil domain affecting active site dynamics [31,91] or stabilized closed clamp conformation by DSIF altering the TL conformation to facilitate catalysis [6,80]. In addition, our results suggest that not only the clamp but also RNA structures impact TL, as in the absence of DSIF, and consequently, upon formation of hairpin-type secondary structures, cross-correlations of TL with RNA increase. Therefore, we conclude that, in the presence of DSIF, the concerted motions of DNA/RNA and Rpb1-clamp impact the active site dynamics and propose that these motions may help prevent interruptions during transcription elongation for maintaining processivity.

5. CONCLUSION

In this study, we examined the impacts of DSIF on the dynamics of the Pol II elongation complex by performing MD simulations in the presence and absence of DSIF. DSIF impacted the structure and dynamics of the upstream nucleic acid substantially, and Rpb1 clamp opening/closing motion moderately. We observed that motions of both nucleic acids and the clamp domain are correlated with the motion of the active site domains, TL and BH, impacting their dynamics and interaction networks. Our results suggest that DSIF may allosterically affect the TL opening/closing motion. The next step to understand DSIF's role in the conformational interplay of TL would be to model the TL opening/closing motion using enhanced sampling techniques. Probing the effects of DSIF on the energetics of the open and closed TL and the transition pathway between the two would provide important insights into the role of DSIF in transcription processivity.

DATA AND SOFTWARE AVAILABILITY

Software used for the simulations and analysis is all open source, which includes OpenMM, CHARMM academic version, MMTSB, and Python libraries (see Methods section for the details of the software). The structures with PDB IDs: 5OIK and 5FLM were used as the initial structures. The Python scripts to run the OpenMM program were generated using the CHARMM-GUI server. The data for ensuring reproducibility of our study is provided in the GitHub repository (https://github.com/bercemd/PolII_DSIF). The data includes coordinates of the initial solvated systems, scripts to prepare the initial systems, run input files, parameter and topology files for the simulations, analysis scripts, and the output data from the analysis. MD simulation trajectories were not included in the repository due to their large sizes; however, they can be made available to researchers upon reasonable requests.

ACKNOWLEDGEMENTS

The authors used computational resources at the Cyberinfrastructure and Research Technologies at the University of California Merced (grant numbers: NSF ACI-2019144, NSF ACI-1429783) and at the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) with ACCESS award number BIO210145.

AUTHOR INFORMATION

Corresponding Author

Bercem Dutagaci

5200 North Lake Rd.

Merced, CA 95343, USA

bdutagaci@ucmerced.edu

209-228-3603

Author Contributions

W. D. A.: Research design and development, analysis, manuscript reviewing and editing; B. M. B.: Research design and development, simulations, analysis; B.D.: Research design, development and management, simulations, analysis, manuscript original writing, reviewing and editing. All authors have given approval to the final version of the manuscript.

Note:

The authors declare no competing financial interest.

NOTES:

This document is the unedited Author's version of a Submitted Work that was subsequently accepted for publication in the *Journal of Chemical Information and Modeling*, copyright © *2025 American Chemical Society* after peer review.

SUPPORTING INFORMATION

Supplementary material associated with this article contains the following analysis results: RMSD for Pol II-Rpb1, TL, BH, RNA and upstream DNA; RMSD of the protein domains for the two initial PDB structures; contact maps of DNA and RNA with the nearby protein domains at the upstream and protein-protein contact maps of Rpb1-clamp with KOW2-4 and CC domain with NGN and KOW2; probability distributions of the number of Watson-Crick base-pair hydrogen bonds, base stacking distances and nucleic acid backbone dihedral angles; probability distributions of BH bending angles; probability distributions of BH-TL distances, distances of Rpb1-clamp to Rpb1-jaw and Rpb2-protrusion; distance maps between Rpb1-clamp and Rpb1-

jaw, and Rpb1-clamp and Rpb2-protrusion; cross-correlation maps of TL and BH with each other, and with Rpb1 clamp, DNA and RNA; RMSF values for Rpb1-clamp.

REFERENCES

[1] Cramer, P., Multisubunit RNA polymerases. *Current Opinion in Structural Biology* **2002**, 12 (1), 89-97.

[2] Cramer, P., Eukaryotic transcription turns 50. *Cell* **2019**, 179 (4), 808-812.

[3] Mohamed, A. A.; Nunez, R. V.; Vos, S. M., Structural advances in transcription elongation. *Current Opinion in Structural Biology* **2022**, 75, 102422.

[4] Saunders, A.; Core, L. J.; Lis, J. T., Breaking barriers to transcription elongation. *Nature Reviews Molecular Cell Biology* **2006**, 7 (8), 557-567.

[5] Chen, F. X.; Woodfin, A. R.; Gardini, A.; Rickels, R. A.; Marshall, S. A.; Smith, E. R.; Shiekhattar, R.; Shilatifard, A., PAF1, a molecular regulator of promoter-proximal pausing by RNA polymerase II. *Cell* **2015**, 162 (5), 1003-1015.

[6] Crickard, J. B.; Fu, J. H.; Reese, J. C., Biochemical analysis of yeast suppressor of Ty 4/5 (Spt4/5) reveals the importance of nucleic acid interactions in the prevention of RNA polymerase II arrest. *Journal of Biological Chemistry* **2016**, 291 (19), 9853-9870.

[7] Crickard, J. B.; Lee, J.; Lee, T. H.; Reese, J. C., The elongation factor Spt4/5 regulates RNA polymerase II transcription through the nucleosome. *Nucleic Acids Research* **2017**, 45 (11), 6362-6374.

[8] Dollinger, R.; Deng, E. B.; Schultz, J.; Wu, S.; Deorio, H. R.; Gilmour, D. S., Assessment of the roles of Spt5-nucleic acid contacts in promoter proximal pausing of RNA polymerase II. *Journal of Biological Chemistry* **2023**, 299 (9), 105106.

[9] Endoh, M.; Zhu, W. Y.; Hasegawa, J.; Watanabe, H.; Kim, D. K.; Aida, M.; Inukai, N.; Narita, T.; Yamada, T.; Furuya, A., et al., Human Spt6 stimulates transcription elongation by RNA polymerase II in vitro. *Molecular and Cellular Biology* **2004**, 24 (8), 3324-3336.

[10] Fitz, J.; Neumann, T.; Pavri, R., Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation. *Embo Journal* **2018**, 37 (8), e97965.

[11] Hou, L. M.; Wang, Y. T.; Liu, Y.; Zhang, N.; Shamovsky, I.; Nudler, E.; Tian, B.; Dynlacht, B. D., Paf1C regulates RNA polymerase II progression by modulating elongation rate. *Proceedings of the National Academy of Sciences of the United States of America* **2019**, 116 (29), 14583-14592.

[12] Vos, S. M.; Farnung, L.; Linden, A.; Urlaub, H.; Cramer, P., Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. *Nature Structural & Molecular Biology* **2020**, 27 (7), 668-677.

[13] Wu, L. P.; Li, L.; Zhou, B.; Qin, Z. H.; Dou, Y. L., H2B ubiquitylation promotes RNA Pol II processivity via PAF1 and pTEFb. *Molecular Cell* **2014**, 54 (6), 920-931.

[14] Yoh, S. M.; Cho, H.; Pickle, L.; Evans, R. M.; Jones, K. A., The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes & Development* **2007**, 21 (2), 160-174.

[15] Yu, M.; Yang, W. J.; Ni, T.; Tang, Z. Y.; Nakadai, T.; Zhu, J.; Roeder, R. G., RNA polymerase II-associated factor 1 regulates the release and phosphorylation of paused RNA polymerase II. *Science* **2015**, 350 (6266), 1383-1386.

[16] Wada, T.; Takagi, T.; Yamaguchi, Y.; Ferdous, A.; Imai, T.; Hirose, S.; Sugimoto, S.; Yano, K.; Hartzog, G. A.; Winston, F., et al., DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes & Development* **1998**, 12 (3), 343-356.

[17] Cui, P.; Jin, H. Y.; Vutukuru, M. R.; Kaplan, C. D., Relationships between RNA polymerase II activity and Spt elongation factors to Spt-phenotype and growth in. *G3-Genes Genomes Genetics* **2016**, 6 (8), 2489-2504.

[18] Aoi, Y.; Takahashi, Y. H.; Shah, A. P.; Iwanaszko, M.; Rendleman, E. J.; Khan, N. H.; Cho, B. K.; Goo, Y. A.; Ganesan, S.; Kelleher, N. L., et al., SPT5 stabilization of promoter-proximal RNA polymerase II. *Molecular Cell* **2021**, 81 (21), 4413-4424.

[19] Core, L.; Adelman, K., Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes & Development* **2019**, 33 (15-16), 960-982.

[20] Decker, T. M., Mechanisms of transcription elongation factor DSIF (Spt4-Spt5). *Journal of Molecular Biology* **2021**, 433 (14), 166657.

[21] Li, J.; Gilmour, D. S., Promoter proximal pausing and the control of gene expression. *Current Opinion in Genetics & Development* **2011**, 21 (2), 231-235.

[22] Bernecky, C.; Plitzko, J. M.; Cramer, P., Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nature Structural & Molecular Biology* **2017**, 24 (10), 809-815.

[23] Ehara, H.; Yokoyama, T.; Shigematsu, H.; Yokoyama, S.; Shirouzu, M.; Sekine, S., Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **2017**, 357 (6354), 921-924.

[24] Martinez-Rucobo, F. W.; Sainsbury, S.; Cheung, A. C. M.; Cramer, P., Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *Embo Journal* **2011**, 30 (7), 1302-1310.

[25] Ehara, H.; Kujirai, T.; Fujino, Y.; Shirouzu, M.; Kurumizaka, H.; Sekine, S., Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. *Science* **2019**, 363 (6428), 744-747.

[26] Farnung, L., Nucleosomes unwrapped: Structural perspectives on transcription through chromatin. *Current Opinion in Structural Biology* **2023**, 82, 102690.

[27] Kujirai, T.; Kurumizaka, H., Transcription through the nucleosome. *Current Opinion in Structural Biology* **2020**, 61, 42-49.

[28] Vos, S. M.; Farnung, L.; Boehning, M.; Wigge, C.; Linden, A.; Urlaub, H.; Cramer, P., Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature* **2018**, 560 (7720), 607-612.

[29] Mason, P. B.; Struhl, K., Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Molecular Cell* **2005**, 17 (6), 831-840.

[30] Hartzog, G. A.; Fu, J. H., The Spt4-Spt5 complex: A multi-faceted regulator of transcription elongation. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms* **2013**, 1829 (1), 105-115.

[31] Hirtreiter, A.; Damsma, G. E.; Cheung, A. C. M.; Klose, D.; Grohmann, D.; Vojnic, E.; Martin, A. C. R.; Cramer, P.; Werner, F., Spt4/5 stimulates transcription elongation through the RNA polymerase clamp coiled-coil motif. *Nucleic Acids Research* **2010**, 38 (12), 4040-4051.

[32] Shetty, A.; Kallgren, S. P.; Demel, C.; Maier, K. C.; Spatt, D.; Alver, B. H.; Cramer, P.; Park, P. J.; Winston, F., Spt5 plays vital roles in the control of sense and antisense transcription elongation. *Molecular Cell* **2017**, 66 (1), 77-88.

[33] Herbert, K. M.; Zhou, J.; Mooney, R. A.; La Porta, A.; Landick, R.; Block, S. M., E. coli NusG inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of RNA polymerase. *Journal of Molecular Biology* **2010**, 399 (1), 17-30.

[34] Li, W. T.; Giles, C.; Li, S. S., Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. *Nucleic Acids Research* **2014**, 42 (11), 7069-7083.

[35] Tarau, D.; Grünberger, F.; Pilsl, M.; Reichelt, R.; Heiss, F.; König, S.; Urlaub, H.; Hausner, W.; Engel, C.; Grohmann, D., Structural basis of archaeal RNA polymerase transcription elongation and Spt4/5 recruitment. *Nucleic Acids Research* **2024**, 52 (10), 6017-6035.

[36] Klein, B. J.; Bose, D.; Baker, K. J.; Yusoff, Z. M.; Zhang, X. D.; Murakami, K. S., RNA polymerase and transcription elongation factor Spt4/5 complex structure. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, 108 (2), 546-550.

[37] Bar-Nahum, G.; Epshtein, V.; Ruckenstein, A. E.; Rafikov, R.; Mustaev, A.; Nudler, E., A ratchet mechanism of transcription elongation and its control. *Cell* **2005**, 120 (2), 183-193.

[38] Dutagaci, B.; Duan, B. B.; Qiu, C. X.; Kaplan, C. D.; Feig, M., Characterization of RNA polymerase II trigger loop mutations using molecular dynamics simulations and machine learning. *Plos Computational Biology* **2023**, 19 (3), e1010999.

[39] Feig, M.; Burton, Z. F., RNA polymerase II with open and closed trigger loops: Active site dynamics and nucleic acid translocation. *Biophysical Journal* **2010**, 99 (8), 2577-2586.

[40] Huang, X. H.; Wang, D.; Weiss, D. R.; Bushnell, D. A.; Kornberg, R. D.; Levitt, M., RNA polymerase II trigger loop residues stabilize and position the incoming nucleotide triphosphate in transcription. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, 107 (36), 15745-15750.

[41] Silva, D. A.; Weiss, D. R.; Avila, F. P.; Da, L. T.; Levitt, M.; Wang, D.; Huang, X. H., Millisecond dynamics of RNA polymerase II translocation at atomic resolution. *Proceedings of the National Academy of Sciences of the United States of America* **2014**, 111 (21), 7665-7670.

[42] Unarta, I. C.; Zhu, L. Z.; Tse, C. K. M.; Cheung, P. P. H.; Yu, J.; Huang, X. H., Molecular mechanisms of RNA polymerase II transcription elongation elucidated by kinetic network models. *Current Opinion in Structural Biology* **2018**, 49, 54-62.

[43] Wang, B. B.; Feig, M.; Cukier, R. I.; Burton, Z. F., Computational simulation strategies for analysis of multisubunit RNA polymerases. *Chemical Reviews* **2013**, 113 (11), 8546-8566.

[44] Wang, B. B.; Opron, K.; Burton, Z. F.; Cukier, R. I.; Feig, M., Five checkpoints maintaining the fidelity of transcription by RNA polymerases in structural and energetic details. *Nucleic Acids Research* **2015**, 43 (2), 1133-1146.

[45] Wang, B. B.; Predeus, A. V.; Burton, Z. F.; Feig, M., Energetic and structural details of the trigger-loop closing transition in RNA polymerase II. *Biophysical Journal* **2013**, 105 (3), 767-775.

[46] Da, L. T.; Pardo-Avila, F.; Xu, L.; Silva, D. A.; Zhang, L.; Gao, X.; Wang, D.; Huang, X. H., Bridge helix bending promotes RNA polymerase II backtracking through a critical and conserved threonine residue. *Nature Communications* **2016**, 7, 11244

[47] Unarta, I. C.; Goonetilleke, E. C.; Wang, D.; Huang, X. H., Nucleotide addition and cleavage by RNA polymerase II: Coordination of two catalytic reactions using a single active site. *Journal of Biological Chemistry* **2023**, 299 (2), 102844.

[48] Farnung, L.; Vos, S. M.; Cramer, P., Structure of transcribing RNA polymerase II-nucleosome complex. *Nature Communications* **2018**, 9, 5432.

[49] Filipovski, M.; Soffers, J. H. M.; Vos, S. M.; Farnung, L., Structural basis of nucleosome retention during transcription elongation. *Science* **2022**, 376 (6599), 1313-1316.

[50] Gallardo, A.; Bogart, B. M.; Dutagaci, B., Protein-nucleic acid interactions for RNA polymerase II elongation factors by molecular dynamics simulations. *Journal of Chemical Information and Modeling* **2022**, 62 (12), 3079–3089.

[51] Gallardo, A.; Dutagaci, B., Binding of small molecule inhibitors to RNA polymerase-Spt5 complex impacts RNA and DNA stability. *Journal of Computer-Aided Molecular Design* **2024**, 38, 1.

[52] Bernecky, C.; Herzog, F.; Baumeister, W.; Plitzko, J. M.; Cramer, P., Structure of transcribing mammalian RNA polymerase II. *Nature* **2016**, 529 (7587), 551-554.

[53] Fiser, A.; Do, R. K. G.; Sali, A., Modeling of loops in protein structures. *Protein Science* **2000**, 9 (9), 1753-1773.

[54] Feig, M.; Karanicolas, J.; Brooks, C. L., MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *Journal of Molecular Graphics & Modelling* **2004**, 22 (5), 377-395.

[55] Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., et al., CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **2009**, 30 (10), 1545-1614.

[56] Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, 98 (12), 10089-10092.

[57] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, 103 (19), 8577-8593.

[58] Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2017**, 14 (1), 71-73.

[59] Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *Journal of Chemical Theory and Computation* **2012**, 8 (9), 3257-3273.

[60] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **1983**, 79 (2), 926-935.

[61] Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-time-step molecular dynamics through hydrogen mass repartitioning. *Journal of Chemical Theory and Computation* **2015**, 11 (4), 1864-1874.

[62] Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L. P.; Shukla, D., et al., OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *Journal of Chemical Theory and Computation* **2013**, 9 (1), 461-469.

[63] Jo, S.; Kim, T.; Iyer, V. G.; Im, W., CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **2008**, 29 (11), 1859-1865.

[64] Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y. F., et al., CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of Chemical Theory and Computation* **2016**, 12 (1), 405-413.

[65] Grossfield, A. WHAM: the weighted histogram analysis method, version 2.0.11, http://membrane.urmc.rochester.edu/?page_id=126/.

[66] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V., et al., Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, 12, 2825-2830.

[67] Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J., et al., Array programming with NumPy. *Nature* **2020**, 585 (7825), 357-362.

[68] Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O., MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **2011**, 32 (10), 2319-2327.

[69] Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D., Impact of 2 '-hydroxyl sampling on the conformational properties of RNA: Update of the CHARMM all-atom additive force field for RNA. *Journal of Computational Chemistry* **2011**, 32 (9), 1929-1943.

[70] Humphrey, W.; Dalke, A.; Schulten, K., VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, 14 (1), 33-38.

[71] Cheung, A. C. M.; Sainsbury, S.; Cramer, P., Structural basis of initial RNA polymerase II transcription. *Embo Journal* **2011**, 30 (23), 4755-4763.

[72] Da, L. T.; Wang, D.; Huang, X. H., Dynamics of pyrophosphate ion release and its coupled trigger loop motion from closed to open state in RNA polymerase II. *Journal of the American Chemical Society* **2012**, 134 (4), 2399-2406.

[73] Kaplan, C. D.; Larsson, K. M.; Kornberg, R. D., The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Molecular Cell* **2008**, 30 (5), 547-556.

[74] Larson, M. H.; Zhou, J.; Kaplan, C. D.; Palangat, M.; Kornberg, R. D.; Landick, R.; Block, S. M., Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, 109 (17), 6555-6560.

[75] Mishanina, T. V.; Palo, M. Z.; Nayak, D.; Mooney, R. A.; Landick, R., Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, 114 (26), E5103-E5112.

[76] Wang, D.; Bushnell, D. A.; Westover, K. D.; Kaplan, C. D.; Kornberg, R. D., Structural basis of transcription: Role of the trigger loop in substrate specificity and catalysis. *Cell* **2006**, 127 (5), 941-954.

[77] Lin, G.; Barnes, C. O.; Weiss, S.; Dutagaci, B.; Qiu, C.; Feig, M.; Lyubimov, A.; Song, J.; Cohen, A. E.; Kaplan, C. D., et al., Structural basis of transcription: RNA Polymerase II

substrate binding and metal coordination at 3.0 Å using a free-electron laser. *Proceedings of the National Academy of Sciences* **2024**, 121 (36), e2318527121.

[78] Sevostyanova, A.; Belogurov, G. A.; Mooney, R. A.; Landick, R.; Artsimovitch, I., The β subunit gate loop is required for RNA polymerase modification by RfaH and NusG. *Molecular Cell* **2011**, 43 (2), 253-262.

[79] Hein, P. P.; Kolb, K. E.; Windgassen, T.; Bellecourt, M. J.; Darst, S. A.; Mooney, R. A.; Landick, R., RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nature Structural & Molecular Biology* **2014**, 21 (9), 794-802.

[80] Nayak, D.; Voss, M.; Windgassen, T.; Mooney, R. A.; Landick, R., Cys-Pair reporters detect a constrained trigger loop in a paused RNA polymerase. *Molecular Cell* **2013**, 50 (6), 882-893.

[81] Sekine, S.; Murayama, Y.; Svetlov, V.; Nudler, E.; Yokoyama, S., The ratcheted and ratchetable structural states of RNA polymerase underlie multiple transcriptional functions. *Molecular Cell* **2015**, 57 (3), 408-421.

[82] Toulokhonov, I.; Artsimovitch, I.; Landick, R., Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* **2001**, 292 (5517), 730-733.

[83] Toulokhonov, I.; Landick, R., The flap domain is required for pause RNA hairpin inhibition of catalysis by RNA polymerase and can modulate intrinsic termination. *Molecular Cell* **2003**, 12 (5), 1125-1136.

[84] Toulokhonov, I.; Zhang, J. W.; Palangat, M.; Landick, R., A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Molecular Cell* **2007**, 27 (3), 406-419.

[85] Yakhnin, A. V.; Yakhnin, H.; Babitzke, P., Function of the transcription elongation factor NusG in hairpin dependent RNA polymerase pausing in the leader. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, 105 (42), 16131-16136.

[86] Qiu, C. X.; Erinne, O. C.; Dave, J. M.; Cui, P.; Jin, H. Y.; Muthukrishnan, N.; Tang, L. K.; Babu, S. G.; Lam, K. C.; Vandeventer, P. J., et al., High-resolution phenotypic landscape of the RNA polymerase II trigger loop. *Plos Genetics* **2016**, 12 (11), e1006321.

[87] Barnes, C. O.; Calero, M.; Malik, I.; Graham, B. W.; Spahr, H.; Lin, G. W.; Cohen, A. E.; Brown, I. S.; Zhang, Q. M.; Pullara, F., et al., Crystal structure of a transcribing RNA polymerase II complex reveals a complete transcription bubble. *Molecular Cell* **2015**, 59 (2), 258-269.

[88] Sanders, T. J.; Lammers, M.; Marshall, C. J.; Walker, J. E.; Lynch, E. R.; Santangelo, T. J., TFS and Spt4/5 accelerate transcription through archaeal histone-based chromatin. *Molecular Microbiology* **2019**, 111 (3), 784-797.

[89] Brueckner, F.; Ortiz, J.; Cramer, P., A movie of the RNA polymerase nucleotide addition cycle. *Current Opinion in Structural Biology* **2009**, 19 (3), 294-299.

[90] Gnatt, A. L.; Cramer, P.; Fu, J. H.; Bushnell, D. A.; Kornberg, R. D., Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 angstrom resolution. *Science* **2001**, 292 (5523), 1876-1882.

[91] Werner, F., A nexus for gene expression-molecular mechanisms of Spt5 and NusG in the three domains of life. *Journal of Molecular Biology* **2012**, 417 (1-2), 13-27.