Binding of Small Molecule Inhibitors to RNA Polymerase-Spt5 Complex Impacts RNA and DNA Stability

Adan Gallardo[1], Bercem Dutagaci[1*]

[1]Department of Molecular and Cell Biology, University of California Merced, Merced, CA 95343, USA.

* Corresponding authors:
Bercem Dutagaci
5200 North Lake Rd.
Merced, CA 95343
bdutagaci@ucmerced.edu
orcid.org/0000-0003-0333-5757

ABSTRACT

Spt5 is an elongation factor that associates with RNA polymerase II (Pol II) during transcription and has important functions in promoter-proximal pausing and elongation processivity. Spt5 was also recognized for its roles in transcription of expanded-repeat genes that are related to neurogenerative diseases. Recently, a set of Spt5-Pol II small molecule inhibitors (SPIs) were reported, which selectively inhibit mutant huntingtin (Htt) gene transcription. Inhibition mechanisms as well as interaction sites of these SPIs with Pol II and Spt5 are not entirely known. In this study, we predicted the binding sites of three selected SPIs at the Pol II-Spt5 interface by docking and molecular dynamics (MD) simulations. Two molecules out of three demonstrated strong binding with Spt5 and Pol II, while the other molecule was more loosely bound and wandered around multiple binding sites. Strongly bound SPIs indirectly affected RNA and DNA dynamics at the exit site as DNA became more flexible while RNA was stabilized by increased interactions with Spt5. Our results suggest that the transcription inhibition mechanism induced by SPIs can be related to Spt5-nucleic acid interactions, which were altered to some extent with strong binding of SPIs.

Keywords

RNA polymerase II; Transcription elongation; Small molecule inhibitors; DRB sensitivity-inducing factor; Molecular dynamics simulations

INTRODUCTION

RNA polymerase II (Pol II) is a vital protein known for its fundamental role in RNA synthesis, which takes place in three main stages that are initiation, elongation, and termination [1,2]. The transcription elongation stage starts with promoter-proximal pausing that causes Pol II to escape from the promoter site to maintain RNA elongation. During the elongation stage, several transcription elongation factors associate with Pol II and collectively regulate the transcription process [3]. Spt5 is one of the elongation factors that associates with Spt4 to form DRB sensitivity-inducing factor (DSIF) [4]. DSIF is known to have many roles in transcription including in promoter-pausing [5,6], processivity of transcription [7,8], nucleosomal access [9-11], and is recently recognized for its role in selective regulation of transcription of repetitive genes [12-15]. Although several studies in structural biology [6,7,10,16-22], biochemistry, and genetics [9,23-30] provided substantial data and valuable insights into DSIF function, understanding the mechanisms, in which DSIF plays a role in these transcription events is still limited. Structural studies showed that Spt5 interacts with Pol II upstream and stabilizes DNA and RNA at the exit tunnels [7,16]. Our previous study also shows that KOW1 and KOW4 domains stayed strongly in contact with the upstream DNA and nascent RNA, respectively, during MD simulations [31]. These studies suggest strong interactions of DSIF with DNA and RNA, and these interactions could be directly related to the functional roles of DSIF in transcription elongation.

The role of DSIF in transcription of genes with multiple repeats was associated with several neurodegenerative diseases [12-15]. Studies showed that reducing the expression of DSIF orthologs selectively inhibits transcription of repetitive genes [12-15], which suggests a crucial role of DSIF in transcription of such multi-repeat genes. Therefore, interference with DSIF function may have therapeutic consequences by decreasing pathological outcomes of neurodegenerative diseases such as Huntington's disease [32], spinocerebellar ataxia type 36 [33], amyotrophic lateral sclerosis (ALS) and frontotemporal dementia [34,35], which are all known to be related to expanded gene mutations. Following these studies on the effects of DSIF in multi-repeat gene transcription, a recent study suggested a novel therapeutic approach for combating Huntington's disease by introducing Spt5-Pol II small molecule inhibitors (SPIs), which selectively inhibit transcription of the huntingtin (Htt) mutant gene, which has extended CAG repeats, by reproducing the DSIF knockdown effects [36]. They observed inhibitory effects of SPIs on the transcription of Htt mutant gene as well as inflammatory genes, while the mechanisms of how these molecules interfere with DSIF function and how they interact with the Pol II complex are open questions. Fluorescence intensity measurements on selected two SPIs suggest different binding patterns as one of the SPIs interacts with mostly Pol II Rpb1 subunit, while the other one seems to have interactions with both Rpb1 and Spt5. Although the earlier study provides some insight into their binding, the structural details of the binding sites are not fully resolved, yet crucial to understand the molecular mechanisms of actions for these SPIs in transcription inhibition.

In this study, we predicted the binding sites of SPIs at the Pol II-DSIF complex and further elucidated interactions between the SPIs and the residues of Pol II and Spt5. We selected three SPIs that are known to selectively inhibit Htt mutant gene transcription and performed docking calculations to determine their initial binding poses. Then, we performed molecular dynamics

(MD) simulations to obtain binding strengths, interacting residues, and the impact of these interactions on the dynamics of the Pol II-DSIF complex. We observed that two SPIs are strongly bound to the initial binding sites predicted by docking at the Pol II-Spt5 interface, whereas one SPI demonstrated multiple binding sites. Notably, we observed that the strong binding of the formers altered Spt5-nucleic acid interactions that further impacted the stability of DNA and RNA at the exit site while weakly binding SPI had less impact on the interactions and dynamics of the nucleic acids.

METHODS

Initial structures of SPIs

We selected three molecules that inhibit the mutant Htt gene from the previously published list of SPIs.[36] The SPIs I, II, and III correspond to 18, 21, and 86 in the previous paper [36] and the structures were shown in Fig. 1. The initial coordinates were generated using the ChemDraw program online version (19.0.0-CDJS-19.0.x.9+da9bec968) and the Avogadro program version 1.2.0. Their conformations were minimized using Gaussian09 [37] with MP2 geometry optimization method and 6-31G(d) basis.

Docking of SPIs on the Pol II-DSIF complex

We used a recently published cryo-EM structure of the Pol II elongation complex with the PDB ID of 5OIK [16]. We performed docking using the AutoDock program version 4.2.6 [38]. Kollman [39] and Gasteiger [40] charges were used for protein and ligand partial charges, respectively. Only polar hydrogens were added to the protein. For docking, a grid was generated to cover the entire protein complex to reduce any biases in finding the binding sites. Grids were generated around the center of mass of the complete Pol II elongation complex with 1 Å grid spacing. 1000 runs were performed for each ligand-protein combination. The best binding conformation with a minimum total binding energy among all the docking conformations was used for each ligand for MD simulations.

MD simulations

We performed MD simulations of protein-ligand complexes for three SPIs as well as for the apoprotein (apo). All the systems were prepared using the MMTSB toolset [41] and CHARMM software version 45b2 [42]. The protein complexes were solvated in a cubic box with a 10 Å cutoff from each side of the box. We added Na^+ ions to neutralize the systems. The systems have 177 Å box sixes that consist of around 560,000 atoms. We used the CHARMM c36 and c36m force fields [43-45] for nucleic acids and proteins, respectively, and CGenFF [46] for SPIs by generating topology and parameter files using the CHARMM-GUI server [47,48]. TIP3P water model was used for water molecules [49]. Force fields were modified to repartition the mass of the heavy atoms that are attached to H atoms to increase the mass of H to 3 a.m.u. as suggested earlier [50]. This modification increases the sampling of the simulations by allowing the use of a 4 fs time step without disturbing the system [50]. We minimized the systems using 5000 steps with an energy tolerance of 100 kJ/mol. Systems were equilibrated for around 1.6 ns while the temperature was increased from 100 to 300 K and the time step was increased from 1 to 4 fs. During the

equilibration, backbone and side chain heavy atoms of proteins were constrained using force constants of 400 and 40 kJ/mol/nm^2, respectively. After the equilibration, we performed MD simulations for 200 ns. Equilibration and MD simulations were repeated for three replicates for each system. For the long-range interactions, periodic boundary conditions were used with the particle mesh Ewald algorithm. Lennard-Jones interactions were switched between 1.0 to 1.2 nm. Langevin thermostat was used at 303.15 K and with 1 ps^{-1} friction coefficient. All the production runs were performed using 4 fs time step and trajectories were saved at every 40 ps. Simulations were performed using the OpenMM program [51] on GPU machines.

Analysis of the simulations

Simulations were analyzed for RMSD, RMSF, distance, and contact maps, the relative center of mass of ligands, and ligand densities using the MDAnalysis package [52] and MMTSB toolset [41]. RMSD and RMSF of the proteins and nucleic acids were calculated for C_α and P atoms, respectively, after aligning C_α and P atoms of the frames onto a reference structure. For the RMSD calculations, the reference structure was the initial structure, while for the RMSF calculations, frames were aligned to the average structure over the trajectories. RMSD values of ligands were calculated for heavy atoms after superimposing the frames to the initial structure. Distance maps were calculated as the average minimum distance for each residue pair along the simulations. Contact maps were calculated for the minimum distance of ligand to each residue of Spt5 and Rpb1, and 5 Å used as the cutoff for the contact.

To evaluate the location of the ligands along the trajectories we calculated the relative center of mass of the ligands and ligand densities. The relative center of mass was calculated as the center of mass of the ligand with respect to the center of mass of the protein complex throughout the trajectory. Then, a principal component analysis (PCA) was applied to the relative center of mass of the ligand to obtain the corresponding positions of the ligand in a two-dimensional plot. After that, we performed the weighted histogram analysis method (WHAM) [53] to obtain the most probable positions of ligands along the protein complex. Ligand densities were calculated by applying a three-dimensional grid with 1 Å resolution on the trajectories that are aligned to the initial structures.

Binding energies from MD simulations were calculated by MMGBSA calculations using the CHARMM software and MMTSB toolset. For the calculations, we extracted frames at every 1 ns with a total number of 600 frames over three replicate trajectories for each system. After a short minimization for each frame (50 steps of steepest descent and 50 steps of adopted basis Newton-Raphson), we calculated the total, electrostatic and van der Waals binding energies between the ligands and Pol II-DSIF complex.

RESULTS

SPIs were reported to interfere with the DSIF function in Htt mutant gene transcription. The previous study [36] predicted that SPI I interacts with the coiled-coil (CC) domain of the Rpb1 subunit and Spt5 elongation factor, while II showed interactions only with Rpb1. However, their exact binding sites were not predicted in the earlier study, therefore, in this work, we studied the binding of the three SPIs (Fig. 1) to the Pol II-DSIF complex. Below, we presented binding sites

obtained by docking and MD simulations, and the impacts of the binding to the protein complex and nucleic acids.

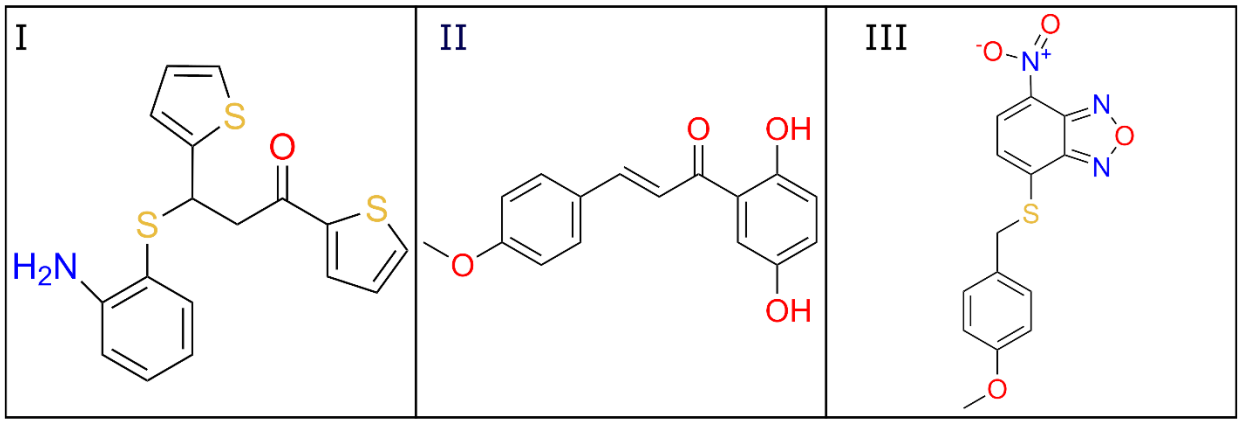

Fig. 1 The chemical structures of SPIs used in this study.

Docking predicted binding of the SPIs to the Pol II-Spt5 interface.

We applied docking of the three SPIs to the complete Pol II-DSIF complex. The most favorable docking poses with the minimum energies for all the molecules were similar, which were located at the interface between Pol II and Spt5 (Fig. S1). The docking positions are close to the CC domain of the Rpb1 subunit of Pol II for all the SPIs in agreement with the previous experimental study [36]. Fig. 2 shows docking positions for the SPIs with highlighted interaction sites within 5 Å distance. SPI-I showed interactions with polar and charged residues (R291, R292, N287, N288) as well as with the nonpolar (I88) and aromatic (Y43) residues of the Rpb1 subunit of Pol II. The only nearby residue from Spt5 was E384, which may form H-bonds with the amine group of the SPI-I. II has interaction sites similar to I, but it is located closer to K42 and Q295, and lacks the interaction site with Y43 compared to I. III is supported by hydrophobic residues of Y43 and I88, and polar residue R292, while it is located relatively far from E384 of Spt5 compared to I and II. In addition, docking energies are provided in Table 1. II and III have lower energies compared to I, suggesting that they are more strongly bound to the docking site. However, our MD simulations suggest that II is more loosely bound compared to I and III, which we discussed below in more detail.

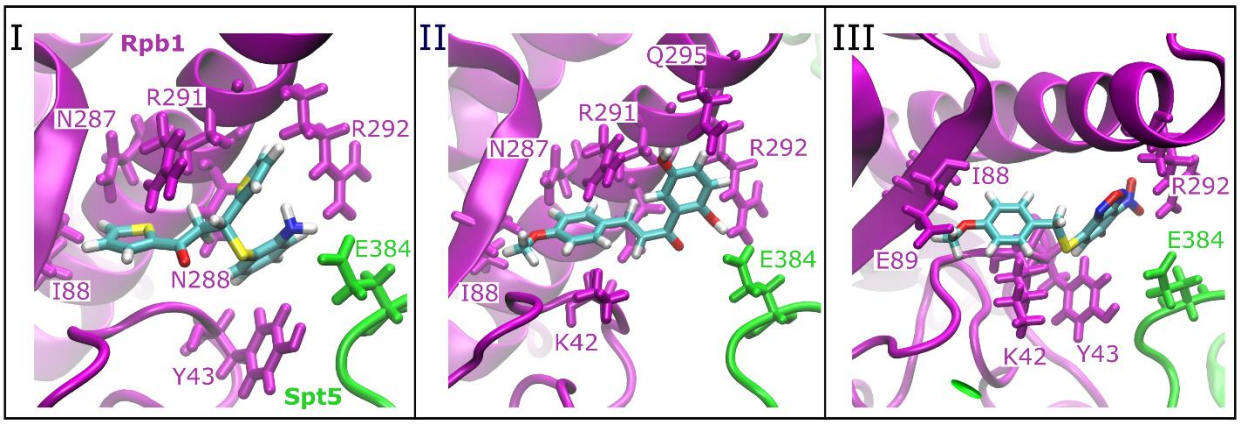

Fig. 2 Docking positions for SPIs I, II and III at the Pol II-DSIF complex and the nearby residues of Rpb1 and Spt5 that are within 5 Å of the SPIs. The color code for the Pol II-DSIF complex is as follows: Rpb1 is in magenta, and Spt5 is in green; SPIs are shown in colors coded by atom name: C is cyan, H is white, O is red, N is blue, and S is yellow.

Table 1. Binding energies from docking and MMGBSA calculations.

SPI	Docking	MMGBSA		
	Energy	Total	Electrostatic	Van der Waals
I	-5.84	-26.41±0.21	-19.39±0.98	-32.41±0.25
II	-6.95	-16.93±0.54	-17.78±1.17	-19.08±0.54
III	-7.13	-30.53±0.27	-23.16±1.05	-34.63±0.18

Docking energies of the minimum energy poses are shown in kcal/mol. MMGBSA energies were calculated from the MD simulations trajectories and reported in kcal/mol.

Two SPIs manifested strong binding to the Pol II-DSIF complex.

We performed MD simulations of Pol II-DSIF complexes with SPIs at the minimum energy docking poses as the initial structures. Fig. S2 shows the RMSD plots of the SPIs during the three replicates of the simulations. SPI-I shows higher flexibility due to the rotation of the aromatic rings around the single bonds. SPI-II and III have fewer possibilities of rotation due to the higher number of resonances in their structure and relatively less flexibility. The docking poses for the three SPIs were at the interface between the CC domain of Rpb1 and Spt5 (Fig. S1). Fig. 3a shows the densities of SPIs along the MD simulation trajectories, which show that for all three SPIs the initial docking site is one of the major binding sites. I and III strongly bound to the CC-Spt5 interface, while II wandered between multiple binding sites. Fig. 3b-c shows that the major binding sites of SPI I and III are very similar to each other. These binding sites consist of the CC domain of Rpb1 and spatially nearby residues, and mostly NGN, KOW1 and KOW2-3 domains of Spt5. For II, the contacts are more scattered, nevertheless around similar regions, which are the CC domain of Rpb1, and NGN, KOW1, and KOW2-3 domains of Spt5. In addition, binding energies calculated by the MMGB-SA method (Table 1) are lower for I and III compared to II consistent with the strong binding of I and III observed in Fig. 3. For SPIs I and III, van der Waals binding energies

are lower than the electrostatic binding energies, whereas, for SPI-II, van der Waals and electrostatic binding energies are close to each other and contribute almost equally to the total binding.

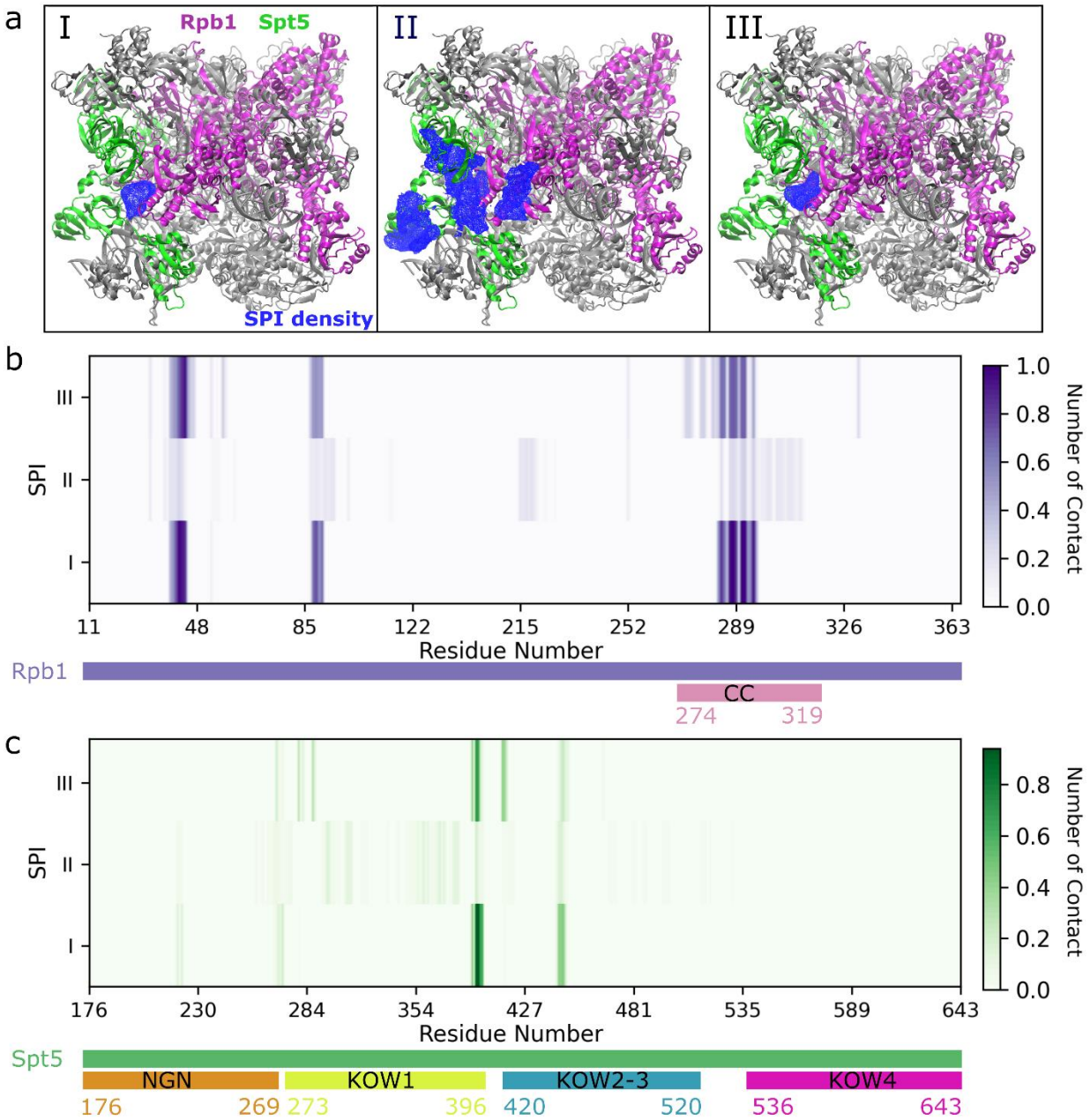

Fig. 3 (a) The densities of SPIs were calculated for the trajectories of three replicates for I, II and III. Rpb1 is in magenta, Spt5 is in green, and the rest of the complex is in silver. SPI densities were shown in blue with 2 % iso-value occupancy. (b, c) Normalized number of contacts of SPIs with Rpb1 (b) and Spt5 (c). Rpb1 contacts are observed only for residues between 11-363. The CC domain of Rpb1 and NGN, KOW1, KOW2-3 and KOW4 domains of Spt5 were shown below the plots in different colors.

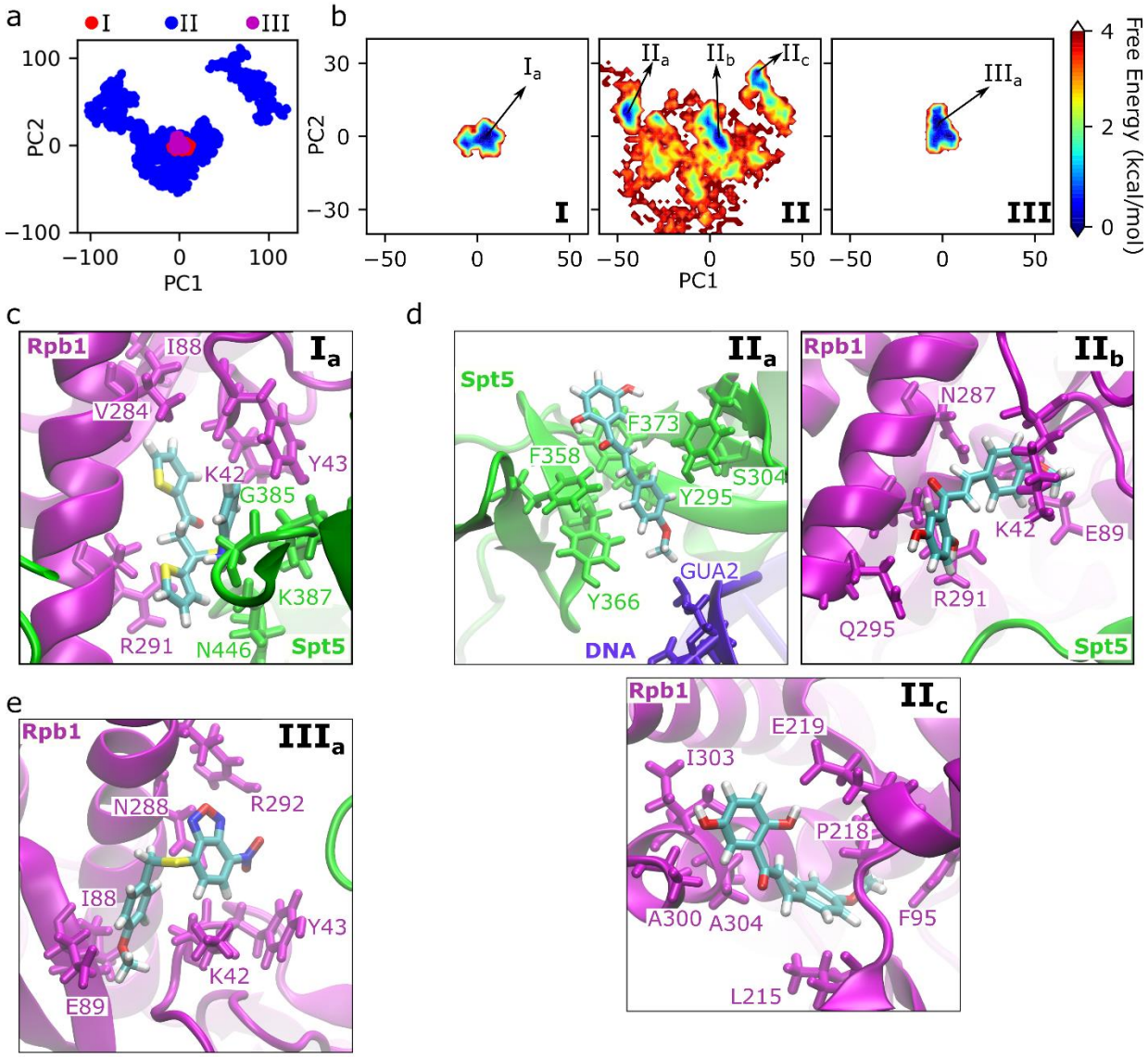

Fig. 4 Binding positions of SPIs obtained from the PCA of the relative center of masses of SPIs with respect to the center of mass of the complete Pol II-DSIF complex. (a) PCA for three SPIs together; (b) potential mean force for the PC1 and PC2 separately calculated for three SPIs by applying WHAM analysis; (c-e) lowest energy positions obtained from the free energy plots for SPIs (c) I, (d) II, and (e) III.

In order to obtain mostly populated binding sites and interaction networks, we applied PCA on the relative position of SPIs with respect to the center of mass of the Pol II-DSIF complex. Fig. 4a shows that I and III are located around the same region in the PCA plot as their binding sites are similar, while II spans a larger area in the PCA plot consistent with the densities observed in Fig. 3a. Fig. 4b shows the free energy plot obtained from the PCA, where we extracted mostly populated (minimum energy) binding sites for each SPI shown in Fig. 4c, d, and e for I, II and III, respectively. We extracted one major binding position for I and III (Fig. 4c and e), and three positions for II (Fig. 4d). I_a has interactions with both Rpb1 and Spt5 (Fig. 4c), while III_a has

mostly interactions with Rpb1 residues rather than Spt5 (Fig. 4e). For I_a, the aromatic rings are mostly stabilized by the hydrophobic pocket formed by V284, Y43 and hydrophobic chains of K42 and R291. On the other hand, III_a has more hydrophilic interactions; forms H-bonds with E89 and N188, and has electrostatic interactions with R292.

For II, we extracted three positions, which are shown as II_a, II_b and II_c in Fig. 4d. The II_a position was at the interface between the KOW1 domain of Spt5 and the upstream DNA without demonstrating any interactions with Pol II. It was located at the hydrophobic pocket formed by aromatic residues, which are F358, F373, Y295 and Y366 of Spt5. The II_b position is around the initial docking pose, where the molecule is located between the CC domain of Rpb1 and Spt5. It was surrounded by polar and charged residues which are stabilizing the SPI via H-bonds. The II_c position was also at the CC domain but facing the opposite side from the Spt5. It was stabilized by hydrophobic interaction with the residues I303, A300, A304, F95 and L215, and it also forms H-bond with E219. Overall, II was located around the CC domain and Spt5, but presented multiple binding positions in contrast to single binding sites of I and III.

SPIs impact DNA and RNA stability.

In order to understand the effects of SPIs on the dynamics of Pol II, Spt5 and DNA/RNA at the exit tunnels, we analyzed RMSF and RMSD values of proteins and nucleic acids along the simulations. Fig. 5a shows the RMSF plots for the sections of the template and non-template DNA that are interacting with the KOW1 domain of Spt5 and the nascent RNA that is interacting with the KOW4 domain of Spt5. We observed increased fluctuations in DNA for the systems with SPIs I and III, while fluctuations were relatively smaller for the system with II compared to apo. As SPI-I and III were more strongly bound to the Pol II complex, this result suggests that the binding of SPIs to the Rpb1-Spt5 interface may destabilize the upstream DNA, which is known to be stabilized by interactions with Spt5 [7,16]. For the nascent RNA, we observed an opposite effect with reduced fluctuations in the RMSF plots for the systems with SPIs I, II and III. In addition to this, we observed high RMSD values for RNA as expected since it does not have a globular shape, while the RMSD values were relatively smaller and converged more quickly for the systems with I and III (Fig. 5b). On the other hand, RMSD values of DNA (Fig. S3) are mostly similar for the complexes with SPIs compared to the apo-complex, except that one replicate for each SPI-I and III showed larger RMSD values for DNA that supports a decreased stability of DNA observed in the RMSF plots.

To further understand the stability change of RNA, we focused on the interactions of RNA and the KOW4 domain of Spt5. Since the SPIs did not demonstrate any direct interactions with RNA, we assumed that SPIs impact the stability of RNA through the interactions of Spt5. Fig. 5c shows the distance maps of RNA with the KOW4 domain of Spt5. Apo showed strong interactions for KOW4 residues 578-585 and 617-620 and weaker interactions for 539-541. All these interaction sites were observed in the complexes with SPIs with reduced average distances suggesting the formation of stronger interactions compared to the apo-protein. Furthermore, additional interaction sites (563-564 and 639-642) were observed for systems with I and III. The analyses of the distance maps and RMSF/RMSD of RNA suggested that the SPIs increased the interactions between Spt5 and RNA and, as a result, stabilized RNA at the exit tunnel. For DNA,

Figs. S4 and S5 show the distance maps between the KOW1 domain of Spt5 and non-template and template DNA at the exit tunnel. There are not any significant changes in the distance maps in the presence of SPIs, while interactions with the template chain somewhat weakened especially for the residues between 361-366.

For the proteins, we calculated RMSF values of the CC domain of Rpb1, and NGN, KOW1 and KOW4 domains of Spt5 (Fig. S6). For the CC, we observed some small changes in the flexible parts of the domain (residues 290-302 and 320-326). For Spt5, there are not any significant changes in the RMSF values, except for the increased fluctuations at the KOW1 domain (residues 344-357) with SPI III. RMSD values were slightly higher for the apo-complex and the complex with II for Spt5 (Fig. S7), while RMSD values for Rpb1 did not show any significant changes (Fig. S8).

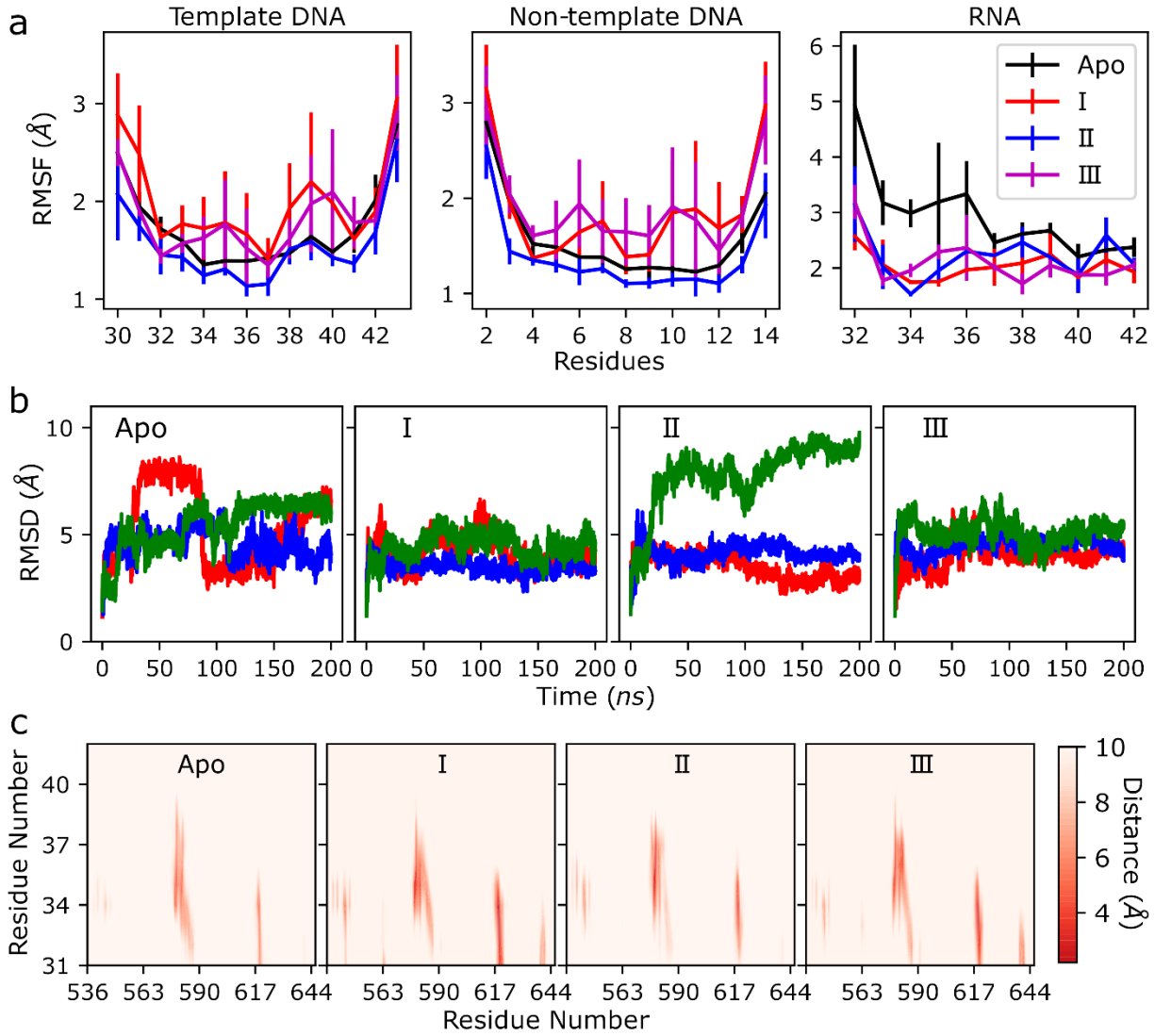

Fig. 5 Analysis of the dynamics and interactions of upstream nucleic acids. (a) RMSF of template and non-template DNA and nascent RNA; (b) RMSD of the nascent RNA, red, blue

and green lines represent the values of each replicate simulation; (C) distance maps between the KOW4 domain of Spt5 and the nascent RNA.

Overall, RMSF and RMSD plots of the nascent RNA suggest that the SPIs, especially I and III, cause stabilization of RNA at the exit tunnel; and distance maps suggest that SPIs I and III indirectly impact RNA dynamics by increasing interactions between the Spt5-KOW4 domain and RNA. On the other hand, upstream DNA was destabilized in the presence of I and III, and interactions between Spt5-KOW1 and DNA were slightly weakened suggesting an indirect impact of SPIs on DNA dynamics as well, but in the opposite direction.

DISCUSSION

In this study, we predicted binding sites of potential drug molecules that are known to inhibit Htt mutant gene transcription. We studied three SPIs, I, II and III, that were reported earlier by Bahat et al with the numbers 18, 21 and 86. We choose I and II since there are experimental measurements for these two SPIs that suggest direct interactions between them and Rpb1 and/or Spt5, therefore, we can compare the experimental results with our binding predictions. Experiments suggested that both SPI-I and II have direct interaction with the Rpb1 CC domain. Intriguingly, our docking and MD simulation results also predicted interactions of both SPIs with the CC domain. On the other hand, experiments showed that SPI-I affected Spt5 while II did not show any effect on the fluorescence spectrum of Spt5 suggesting that only SPI-I has direct interactions with Spt5. Consistent with experiments, we observed that SPI-I was located at the interface of Rpb1 and Spt5 and had interactions with both the CC-domain and Spt5. However, we also observed that SPI-II interacts with Spt5, which was not predicted by experiments. SPI-II manifested multiple binding sites that include both the CC domain of Rpb1 and Spt5. Additionally, SPI-III also shared a similar binding site with SPI-I, located at the interface of Rpb1 and Spt5. Binding to the Pol II-DSIF complex was stronger for I and III, compared to II and this may be related to their effects on the inhibition of Htt mutant gene transcription. We hypothesized that SPI-II has multiple binding sites that may present multiple mechanisms of inhibiting different genes including inflammatory genes and Htt gene. SPI I and II both selectively inhibited mutant Htt gene transcription, but SPI-II had significant effects on inflammatory gene transcription at both basal and induced conditions while SPI-I had only effects at induced conditions [36]. SPI-III, which has inhibition patterns similar to II, but it shows stronger inhibition of Htt mutant gene transcription as it inhibited it at very small concentrations; reported 0.1 μM compared to 30 μM of I and II [36]. The inhibition by SPI-II for transcription of a wider range of genes could be related to its multiple binding sites, which may cause multiple mechanisms for inhibitions. However, this hypothesis needs to be further tested by experimental and computational studies.

We observed that SPIs affect the dynamics of RNA and DNA at the exit tunnel. The effects are larger for SPIs I and III, as expected since they demonstrated a stronger binding than II. We showed that Spt5-RNA interactions became stronger in the presence of SPIs suggesting that SPIs indirectly affect RNA dynamics through interactions with Spt5. Similarly, Spt5-DNA interactions became weaker, and this potentially caused the increased flexibility of the DNA chains, however the impact of SPIs on Spt5-DNA interactions was weaker compared to Spt5-RNA interactions. Overall, our observations suggest that SPIs stabilize RNA and destabilize DNA. However, in this

study, we directly used the DNA template sequence and the sequence of the corresponding nascent RNA that were present in the transcription bubble of the initial cryo-EM structure. To obtain a mechanistic understanding of the inhibition of Htt mutant gene transcription, we need to repeat these simulations with Htt mutant gene in the transcription bubble. We hypothesize that RNA stabilization at the exit site could be larger with mutant Htt gene that may induce pausing, or alternatively largely destabilized upstream DNA may cause transcription defects. Our future plans include testing these hypotheses by performing simulations of the Pol II-DSIF complex in the presence of WT and mutant Htt gene. A structure of the Pol II elongation complex with the Htt gene currently does not exist, however, it can be computationally modeled using the existing DNA/RNA hybrid in the Pol II elongation complex structure as a reference and incorporating the available Htt gene structures [54-56] followed by extensive minimization and equilibration steps.

CONCLUSION

In this study, we elucidated the binding of recently discovered SPIs to the Pol II-DSIF complex. We observed that two out of three SPIs were strongly bound to the Spt5 and Pol II interface as suggested by experiments, while the remaining SPI bound to multiple binding sites. SPIs indirectly affected dynamics of the nascent RNA and upstream DNA via interactions with Spt5. We concluded that the transcription inhibition mechanism could be related to Spt5-nucleic acid interactions. Future work will focus on computational modeling and the simulation of the Pol II elongation complex in the presence of WT and mutant Htt genes to investigate further the role of Spt5 in the inhibition mechanisms of multi-repeat genes.

Acknowledgments

We used the computational resources at the MERCED and Pinnacles clusters at the University of California Merced and the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) facilities funded by the National Science Foundation (ACCESS award number is TG-BIO210145). We thank Brandon Michael Bogart for setting up the initial systems for the apo elongation complex simulations.

Author contributions

Adan Gallardo: Conceptualization, Formal Analysis and Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing.

Bercem Dutagaci: Conceptualization, Formal Analysis and Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing, Project Administration, Resources, Supervision.

Supporting Information

Figures S1-S8 are provided as Supporting Information.

Data availability statement

The initial structure of the Pol II elongation complex was obtained from the Protein Data Bank with the PDB ID:5OIK. Details of the generation of initial SPI structures and their complexes with Pol II elongation complex were provided in the Methods section. We used open software

for the simulations (OpenMM) and analysis (MDAnalysis and MMTSB). We provided the extracted frames from the PCA plots on our GitHub page (https://github.com/bercemd/SPIs_binding). Simulation trajectories can be available upon request.

Statements and Declarations

Competing Interests: The authors declare no competing interests.

Funding: No funding was received for conducting this study.

REFERENCES

1. Hantsche, M. and Cramer, P. (2016) The structural basis of transcription: 10 years after the Nobel Prize in chemistry. *Angew Chem Int Edit*, **55**, 15972-15981.
2. Osman, S. and Cramer, P. (2020) Structural biology of RNA polymerase II transcription: 20 years on. *Annu Rev Cell Dev Bi*, **36**, 1-34.
3. Kwak, H. and Lis, J.T. (2013) Control of Transcriptional Elongation. *Annu Rev Genet*, **47**, 483-508.
4. Decker, T.M. (2021) Mechanisms of Transcription Elongation Factor DSIF (Spt4-Spt5). *J Mol Biol*, **433**.
5. Li, J. and Gilmour, D.S. (2011) Promoter proximal pausing and the control of gene expression. *Curr Opin Genet Dev*, **21**, 231-235.
6. Vos, S.M., Farnung, L., Urlaub, H. and Cramer, P. (2018) Structure of paused transcription complex Pol II-DSIF-NELF. *Nature*, **560**, 601-606.
7. Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M. and Sekine, S. (2017) Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science*, **357**, 921-924.
8. Fitz, J., Neumann, T. and Pavri, R. (2018) Regulation of RNA polymerase II processivity by Spt5 is restricted to a narrow window during elongation. *Embo J*, **37**, e97965.
9. Crickard, J.B., Lee, J., Lee, T.H. and Reese, J.C. (2017) The elongation factor Spt4/5 regulates RNA polymerase II transcription through the nucleosome. *Nucleic Acids Res*, **45**, 6362-6374.
10. Ehara, H., Kujirai, T., Fujino, Y., Shirouzu, M., Kurumizaka, H. and Sekine, S. (2019) Structural insight into nucleosome transcription by RNA polymerase II with elongation factors. *Science*, **363**, 744-747.
11. Kujirai, T. and Kurumizaka, H. (2020) Transcription through the nucleosome. *Curr Opin Struc Biol*, **61**, 42-49.
12. Cheng, H.M., Chern, Y., Chen, I.H., Liu, C.R., Li, S.H., Chun, S.J., Rigo, F., Bennett, C.F., Deng, N., Feng, Y.A. *et al.* (2015) Effects on murine behavior and lifespan of selectively decreasing expression of mutant huntingtin allele by Supt4h knockdown. *Plos Genet*, **11**, e1005043.
13. Furuta, N., Tsukagoshi, S., Hirayanagi, K. and Ikeda, Y. (2019) Suppression of the yeast elongation factor Spt4 ortholog reduces expanded SCA36 GGCCUG repeat aggregation and cytotoxicity. *Brain Res*, **1711**, 29-40.
14. Kramer, N.J., Carlomagno, Y., Zhang, Y.J., Almeida, S., Cook, C.N., Gendron, T.F., Prudencio, M., Van Blitterswijk, M., Belzil, V., Couthouis, J. *et al.* (2016) Spt4 selectively regulates the expression of C9orf72 sense and antisense mutant transcripts. *Science*, **353**, 708-712.
15. Liu, C.R., Chang, C.R., Chern, Y., Wang, T.H., Hsieh, W.C., Shen, W.C., Chang, C.Y., Chu, I.C., Deng, N., Cohen, S.N. *et al.* (2012) Spt4 is selectively required for transcription of extended trinucleotide repeats. *Cell*, **148**, 690-701.
16. Bernecky, C., Plitzko, J.M. and Cramer, P. (2017) Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat Struct Mol Biol*, **24**, 809-815.

17. Ehara, H., Kujirai, T., Shirouzu, M., Kurumizaka, H. and Sekine, S.I. (2022) Structural basis of nucleosome disassembly and reassembly by RNAPII elongation complex with FACT. *Science*, **377**, eabp9466.
18. Farnung, L., Ochmann, M., Garg, G., Vos, S.M. and Cramer, P. (2022) Structure of a backtracked hexasomal intermediate of nucleosome transcription. *Mol Cell*, **82**, 1-9.
19. Farnung, L., Vos, S.M. and Cramer, P. (2018) Structure of transcribing RNA polymerase II-nucleosome complex. *Nat Commun*, **9**, 5432.
20. Filipovski, M., Soffers, J.H.M., Vos, S.M. and Farnung, L. (2022) Structural basis of nucleosome retention during transcription elongation. *Science*, **376**, 1313-1316.
21. Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H. and Cramer, P. (2018) Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. *Nature*, **560**, 607-612.
22. Vos, S.M., Farnung, L., Linden, A., Urlaub, H. and Cramer, P. (2020) Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. *Nat Struct Mol Biol*, **27**, 668-677.
23. 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.* (2021) SPT5 stabilization of promoter-proximal RNA polymerase II. *Mol Cell*, **81**, 4413-4424.
24. Baejen, C., Andreani, J., Torkler, P., Battaglia, S., Schwalb, B., Lidschreiber, M., Maier, K.C., Boltendahl, A., Rus, P., Esslinger, S. *et al.* (2017) Genome-wide Analysis of RNA Polymerase II Termination at Protein-Coding Genes. *Mol Cell*, **66**, 38-49.
25. Crickard, J.B., Fu, J.H. and Reese, J.C. (2016) 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. *J Biol Chem*, **291**, 9853-9870.
26. Evrin, C., Serra-Cardona, A., Duan, S.F., Mukherjee, P.P., Zhang, Z.G. and Labib, K.P.M. (2022) Spt5 histone binding activity preserves chromatin during transcription by RNA polymerase II. *Embo J*, **41**, e109783.
27. Fitz, J., Neumann, T., Steininger, M., Wiedemann, E.M., Garcia, A.C., Athanasiadis, A., Schoeberl, U.E. and Pavri, R. (2020) Spt5-mediated enhancer transcription directly couples enhancer activation with physical promoter interaction. *Nat Genet*, **52**, 505-515.
28. Hu, S.B., Peng, L.N., Xu, C.L., Wang, Z.N., Song, A. and Chen, F.X. (2021) SPT5 stabilizes RNA polymerase II , orchestrates transcription cycles, and maintains the enhancer landscape. *Mol Cell*, **81**, 4425-4439.
29. Li, W.T., Giles, C. and Li, S.S. (2014) Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. *Nucleic Acids Res*, **42**, 7069-7083.
30. Shetty, A., Kallgren, S.P., Demel, C., Maier, K.C., Spatt, D., Alver, B.H., Cramer, P., Park, P.J. and Winston, F. (2017) Spt5 Plays Vital Roles in the Control of Sense and Antisense Transcription Elongation. *Mol Cell*, **66**, 77-88.
31. Gallardo, A., Bogart, B.M. and Dutagaci, B. (2022) Protein-Nucleic Acid Interactions for RNA Polymerase II Elongation Factors by Molecular Dynamics Simulations. *J Chem Inf Model*, **62**, 3079–3089.
32. Ashley, C.T. and Warren, S.T. (1995) Trinucleotide repeat expansion and human disease. *Annu Rev Genet*, **29**, 703-728.
33. Kobayashi, H., Abe, K., Matsuura, T., Ikeda, Y., Hitomi, T., Akechi, Y., Habu, T., Liu, W.Y., Okuda, H. and Koizumi, A. (2011) Expansion of Intronic GGCCTG

Hexanucleotide Repeat in NOP56 Causes SCA36, a Type of Spinocerebellar Ataxia Accompanied by Motor Neuron Involvement. *Am J Hum Genet*, **89**, 121-130.
34. DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J. *et al.* (2011) Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron*, **72**, 245-256.
35. Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L. *et al.* (2011) A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron*, **72**, 257-268.
36. Bahat, A., Lahav, O., Plotnikov, A., Leshkowitz, D. and Dikstein, R. (2019) Targeting Spt5-Pol II by small-molecule inhibitors uncouples distinct activities and reveals additional regulatory roles. *Mol Cell*, **76**, 617-631.
37. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H. *et al.* (2016) Gaussian 09, Revision D.01.
38. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J. (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem*, **30**, 2785-2791.
39. Singh, U.C. and Kollman, P.A. (1984) An Approach to Computing Electrostatic Charges for Molecules. *J Comput Chem*, **5**, 129-145.
40. Gasteiger, J. and Marsili, M. (1980) Iterative Partial Equalization of Orbital Electronegativity - a Rapid Access to Atomic Charges. *Tetrahedron*, **36**, 3219-3228.
41. Feig, M., Karanicolas, J. and Brooks, C.L. (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J Mol Graph Model*, **22**, 377-395.
42. 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.* (2009) CHARMM: The biomolecular simulation program. *J Comput Chem*, **30**, 1545-1614.
43. Best, R.B., Zhu, X., Shim, J., Lopes, P.E.M., Mittal, J., Feig, M. and MacKerell, A.D. (2012) 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. *J Chem Theory Comput*, **8**, 3257-3273.
44. Foloppe, N. and MacKerell, A.D. (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem*, **21**, 86-104.
45. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmuller, H. and MacKerell, A.D. (2017) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods*, **14**, 71-73.
46. Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I. *et al.* (2010) CHARMM General Force Field: A Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J Comput Chem*, **31**, 671-690.
47. Jo, S., Kim, T., Iyer, V.G. and Im, W. (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem*, **29**, 1859-1865.

48. Kim, S., Lee, J., Jo, S., Brooks, C.L., Lee, H.S. and Im, W. (2017) CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J Comput Chem*, **38**, 1879-1886.
49. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys*, **79**, 926-935.
50. Hopkins, C.W., Le Grand, S., Walker, R.C. and Roitberg, A.E. (2015) Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput*, **11**, 1864-1874.
51. 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.* (2013) OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *J Chem Theory Comput*, **9**, 461-469.
52. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B. and Beckstein, O. (2011) Software news and updates MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*, **32**, 2319-2327.
53. Grossfield, A. WHAM: the weighted histogram analysis method, version 2.0.11, http://membrane.urmc.rochester.edu/?page_id=126/.
54. Chen, J.L., VanEtten, D.M., Fountain, M.A., Yildirim, I. and Disney, M.D. (2017) Structure and Dynamics of RNA Repeat Expansions That Cause Huntington's Disease and Myotonic Dystrophy Type 1. *Biochemistry*, **56**, 3463-3474.
55. Kiliszek, A., Kierzek, R., Krzyzosiak, W.J. and Rypniewski, W. (2010) Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases. *Nucleic Acids Res*, **38**, 8370-8376.
56. Tawani, A. and Kumar, A. (2015) Structural Insights Reveal the Dynamics of the Repeating r(CAG) Transcript Found in Huntington's Disease (HD) and Spinocerebellar Ataxias (SCAs). *Plos One*, **10**, e0131788.