

Protein-Nucleic Acid Interactions for RNA Polymerase II Elongation Factors by Molecular Dynamics Simulations

Adan Gallardo, Brandon M. Bogart, Bercem Dutagaci*

Department of Molecular and Cell Biology, University of California Merced, Merced, CA 95340.

* Corresponding authors:

Bercem Dutagaci
5200 North Lake Rd.
Merced, CA 95343
bercemdutagaci@gmail.com

ABSTRACT

RNA polymerase II (Pol II) forms a complex with elongation factors to proceed the elongation stage of the transcription process. In this work, we studied elongation factor SPT5 and explored protein nucleic acid interactions for the isolated systems of KOW1 and KOW4 domains of SPT5 with DNA and RNA, respectively. We performed molecular dynamics (MD) simulations using three commonly used force fields that are CHARMM c36m, AMBER ff14sb and ff19sb. These simulations showed that most of the protein-nucleic acid interactions in the native state were retained with an increased electrostatic binding free energy for all force fields used. RNA was found highly dynamic with all force fields while DNA had relatively more stable conformations with the AMBER force fields compared to CHARMM. Furthermore, we performed MD simulations of the complete elongation complex using CHARMM c36m force field to compare the dynamics and interactions in the isolated systems. Similar strong KOW1 and DNA interactions were observed in the complete elongation complex simulations and DNA was further stabilized by a network of interactions involving SPT5-KOW1, SPT4 and rpb2 of Pol II. Overall, our study showed that the accuracy of force fields and the presence of the entire interaction network are important for elucidating the dynamics of protein-nucleic acid systems.

INTRODUCTION

RNA polymerase II (RNA Pol II) synthesizes RNA in three main steps that are initiation, elongation, and termination¹. Numerous biochemical²⁻⁶ and structural⁷⁻¹¹ studies have added to understanding the mechanisms of these steps, demonstrating that other factors are involved in transcription. During initiation, general transcription factors are associated with RNA Pol II that forms the initiation complex^{9, 11}. This complex dissociates as the RNA Pol II escapes the promoter site and forms the elongation complex with the basal elongation factors^{7-8, 12} to allow transcription with proper processivity. Structure of the transcription initiation complex has been well studied for the last decade^{9, 11} whereas the structure of elongation complex has only recently been revealed^{7-8, 12}. The elucidated structure showed that elongation factors, SPT4 and SPT5, are bound to RNA Pol II to form DNA and RNA clamps at the exit sites suggesting that the interactions of the elongation factors with DNA and RNA are important for the elongation mechanism.

Protein-nucleic acid interactions are not only important for the elongation step of transcription but also in many fundamental biological processes like replication, translation and gene regulation. Thus, research on understanding protein-nucleic acid interactions became an emerging field. X-ray crystallography and cryo-EM provide structural information on protein-nucleic acid systems. In addition to the structural biology techniques, many other experimental techniques including Förster resonance energy transfer (FRET)¹³⁻¹⁴, nuclear magnetic resonance (NMR)¹⁵⁻¹⁶, mass spectrometry¹⁷⁻¹⁸, small angle X-ray scattering (SAXS)¹⁹, electrophoretic mobility shift assay (EMSA)²⁰⁻²², footprinting²⁰⁻²¹, crosslinking^{21, 23} and chromatin immunoprecipitation (ChIP)^{20-21, 24-25} focus on proteins-nucleic acids interactions and their dynamics during the biological processes. In addition to these studies, computational approaches based on molecular dynamics (MD) simulations have provided important insights on the dynamics of protein-nucleic acid interactions at the atomistic details²⁶⁻³². MD simulations are used to obtain mechanistic understanding of the roles that protein-nucleic acid interactions play in biological processes^{27-28, 32-33}. Studies have also shown the importance of force field choice in simulations to reproduce experimental values with high accuracy as they demonstrated that imperfections in the results can occur^{29, 34-35}. The most widely used force fields are CHARMM³⁶⁻³⁹ and AMBER⁴⁰⁻⁴³ since they are largely validated based on various experimental data. Despite the extensive validation, there have been inaccuracies observed with these force fields in DNA^{31, 44-45} and RNA^{34, 46} dynamics, diffusion calculations²⁹ and protein-nucleic acid interactions^{29, 47}. These results suggest that the force fields may generate different results on protein-nucleic acid systems and interpretation of the results should be done with care and consideration of possible simulation artifacts.

In this study, we performed MD simulations of elongation factors with nucleic acids to explore the differences among force fields in generating nucleic acid and protein dynamics and their interactions. We isolated the KOW1 and KOW4 domains of SPT5 and probed their interactions with DNA and RNA, respectively. We used three different force fields, which are CHARMM c36m, AMBER ff14sb and ff19sb. We observed that the interactions between protein and nucleic acids were retained in all force fields. However, force fields effected the flexibility of the DNA chains differently as the CHARMM force field displayed much higher fluctuation

compared to both AMBER force fields. This suggests that AMBER force fields have larger stabilization effects on the intra-molecular interactions of DNA comparing to CHARMM. Furthermore, we used CHARMM c36m to simulate the whole elongation complex that includes RNA Pol II, SPT4, SPT5, upstream and downstream DNA, and nascent RNA. DNA fluctuations were observed in smaller amounts comparing to the isolated KOW1-DNA complex suggesting that DNA is stabilized not only by interactions with KOW1 but by an interaction network of Pol II and elongation factors in the whole complex.

METHODS

MD simulations of KOW1-DNA and KOW4-RNA systems

We used the elongation complex structure with PDB ID 5OIK⁷. For the KOW1-DNA systems we selected the SPT5 residues of 273-399 and DNA residues of 1-12 for the non-template and 32-43 for the template chains that are referred as DNAA and DNAB, respectively, in this paper. For the KOW4-RNA system, SPT5 residues of 536-646 and RNA residues of 31-38 were selected. We isolated those systems from the complete complex and solvated them in a cubic box with a 10 Å cutoff from each edge of the solvation boxes to prevent interactions with the periodic images. We added K⁺ ions to neutralize the systems. System preparation was performed by CHARMM-GUI server⁴⁸⁻⁵⁰ using CHARMM c36m^{36, 39}, AMBER ff14sb⁴⁰ and AMBER ff19sb⁴¹ for proteins. We used BSC1 parameters⁴² for DNA and OL3 parameters⁴³ for RNA with the AMBER protein force fields. We used TIP3P parameters⁵¹ for the explicit water. We minimized the systems for 5000 steps with a tolerance of 100 kJ/mole. Systems were equilibrated for 625 ps with increasing temperature from 100 to 300 K with constraints at the backbone and side chains of proteins with force constant of 400 and 40 kJ/mol/nm², respectively. We, then performed 1 μs of three replicates of MD simulations for each system with a total simulation time of 18 μs. KOW1-DNA system was simulated for four replicates since one of the replicates resulted in partial unfolding for the protein (see RMSD in Fig. S1) that we excluded from the rest of the analysis. Periodic boundary conditions were used with the particle mesh Ewald algorithm for the long-range electrostatic interactions. Lennard Jones interactions were switched between 1.0 to 1.2 Å for the simulations with CHARMM force field and 1.2 Å was used as a cutoff for the simulations with the AMBER force fields. Bonds with H atoms were constraint using the SHAKE algorithm. Simulations were run using Langevin thermostat at 303.15 K and with a friction coefficient of 1 ps⁻¹ Time step was 1 fs for the equilibration and 2 fs for the production runs and trajectories were saved at every 20 ps. Simulations were performed using OpenMM⁵² on GPU machines.

MD simulations of RNA Pol II elongation complex

The complete elongation complex with the PDB ID of 5OIK was used for the simulations. We modelled missing residues of the trigger loop (TL) using MODELLER program⁵³. The system was solvated in a cubic box using a cutoff of 9 Å from each edge with a total box size of 176.5 Å. The system was prepared using MMTSB package⁵⁴ in conjunction with the CHARMM software version 45b2⁵⁵. The system was minimized and equilibrated as described above. We performed 200 ns of production run for each of the three replicates. Simulation details were the same as the small protein-nucleic acid systems described above. We used CHARMM 36m force field with TIP3P water parameters.

Analysis of the simulations

The RMSD, RMSF, distance maps and contact analysis were performed using the MDAnalysis package⁵⁶. For protein, C_α-RMSD was calculated after superimposing the C_α and C_β atoms and for RNA and DNA, heavy atom RMSD was calculated after superimposing the heavy atoms. RMSF values were calculated for C_α atoms of proteins and P atoms for nucleic acids after superimposing the C_α-C_β atoms and heavy atoms to the average structures, respectively. Distance maps were

generated for the initial experimental structure and for the simulations averaged over the trajectories and plotted for distances lower than 10 Å. Contacts between protein and DNA were averaged over the trajectories for each residue pairs that were within 5 Å distance. We calculated the free energy profiles for protein-nucleic acid distances of the simulation trajectories using weighted histogram analysis method⁵⁷⁻⁵⁸. The RMSD analysis was performed for the whole trajectories while all the other analyses were performed for the last 800 ns of the simulations for the KOW1-DNA and KOW4-RNA systems, since the fluctuations of C_α-RMSD decreased after the first 200 ns simulation (Fig S1-2). The analysis for the complete RNA Pol II simulations were performed for the last 150 ns since the RMSD of the KOW1 and KOW4 domains became stable after 50 ns of the simulations (Fig S3).

We performed energy calculations using Molecular Mechanics Generalized Born and Surface Area (MMGB/SA) approach⁵⁹⁻⁶⁰ using the MMTSB package⁵⁴. We calculated total, electrostatic and van der Waals (vdW) binding energies for protein and DNA/RNA pairs without including the entropic contribution. For the energy calculations of the isolated KOW1-DNA and KOW4-RNA systems, we obtained 1,000 conformations extracted at every 2.4 ns from three replicates of simulations for each force fields and applied 500 steps of minimization for each conformation before the energy calculation. For the complete elongation complex, we employed energy calculations onto the total of 900 frames extracted at every 0.5 ns. For the native state, we extracted the KOW1-DNA and KOW4-RNA from the original structure and employed energy calculation after adding hydrogen atoms and applying 5000 steps of minimization. All the minimizations were performed with Adopted Basis Newton-Raphson algorithm. We calculated energies using CHARMM c36 force field.

Hydrogen bond analysis were performed for the Watson-Crick base pairs such that the hydrogen bonds were counted between N1 of adenine and N3 of thymine, N6 of adenine and O4 for thymine, O6 for guanine and N4 of cytosine, N1 of guanine and N3 of cytosine, N2 of guanine and O2 of cytosine. The average number of hydrogen bond counts for the DNA pair was calculated over the trajectories of three replicates. Hydrogen bond analysis was performed using MDAnalysis package⁵⁶.

RESULTS

In this study, we explored the protein-nucleic acid interactions and dynamics for the KOW1 and KOW4 domains of SPT5 elongation factor using AMBER and CHARMM force fields. We isolated the KOW domains and the interacting nucleic acids for the simulations as seen in Fig. 1. We analyzed the simulations in terms of direct interactions between proteins and nucleic acids and the dynamics of both components. We focused on the differences between the simulations with varying force fields to understand the strengths and weaknesses of the force fields for the isolated systems. We, then simulated the complete complex with the CHARMM force field and compared the results with the isolated systems. The results are presented in the following sections.

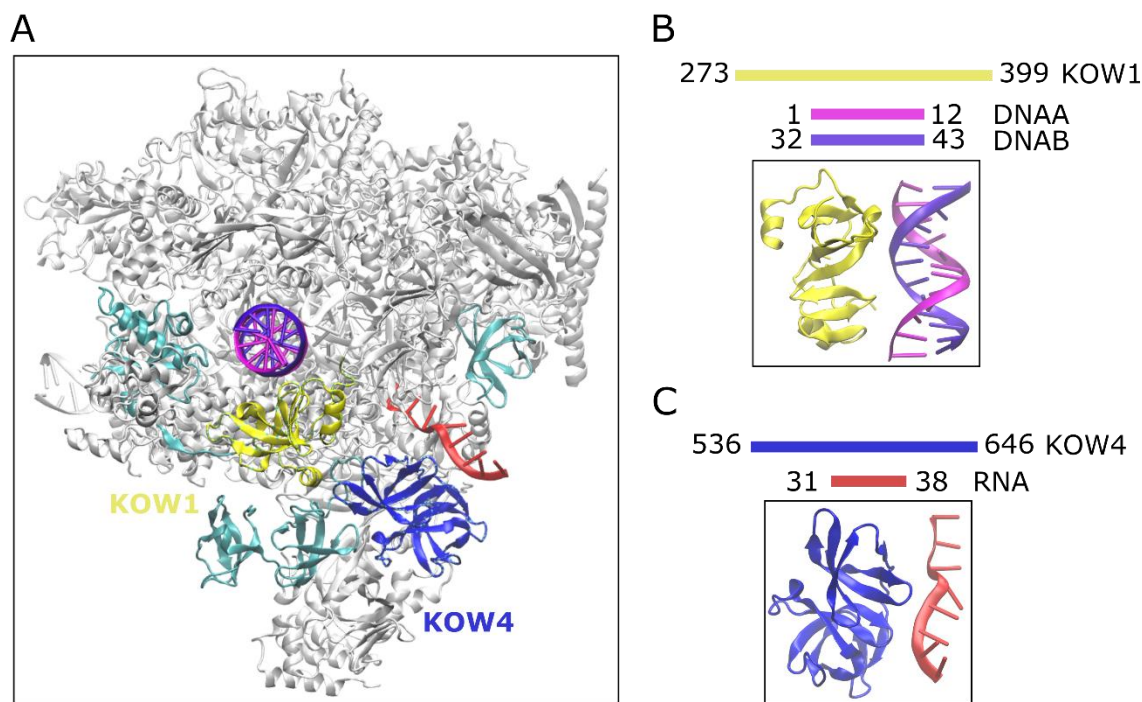


Figure 1. The structures and selections of the simulated systems. (A) the complete elongation complex with the following color codes: KOW1 yellow, KOW4 blue, RNA red, DNAA (non-template) magenta, DNAB (template) violet, other SPT5 domains cyan, and the rest of the complex gray. Selection of the residues and the structures for (B) KOW1-DNA and (C) KOW4-RNA are shown. The figure was generated using the structure with the PDB ID of 5OIK.

KOW1-DNA interactions

The simulations of KOW1 and DNA were performed with three force fields. CHARMM c36m provided a large fluctuation in RMSD values for DNA chains comparing to the simulations with AMBER force fields (Fig. S1). Fig. 2A shows the RMSF profiles for protein (KOW1) and DNA, in which DNA chains show larger RMSF values with CHARMM c36m. For the protein, the RMSD and RMSF profiles are similar for the CHARMM and AMBER force fields although there are larger fluctuations for the residues from 340 to 365 for the AMBER simulations. This is due to the

partial unfolding of the beta sheet region in AMBER simulations, while these beta sheets are retained to some extent in CHARMM c36m simulations.

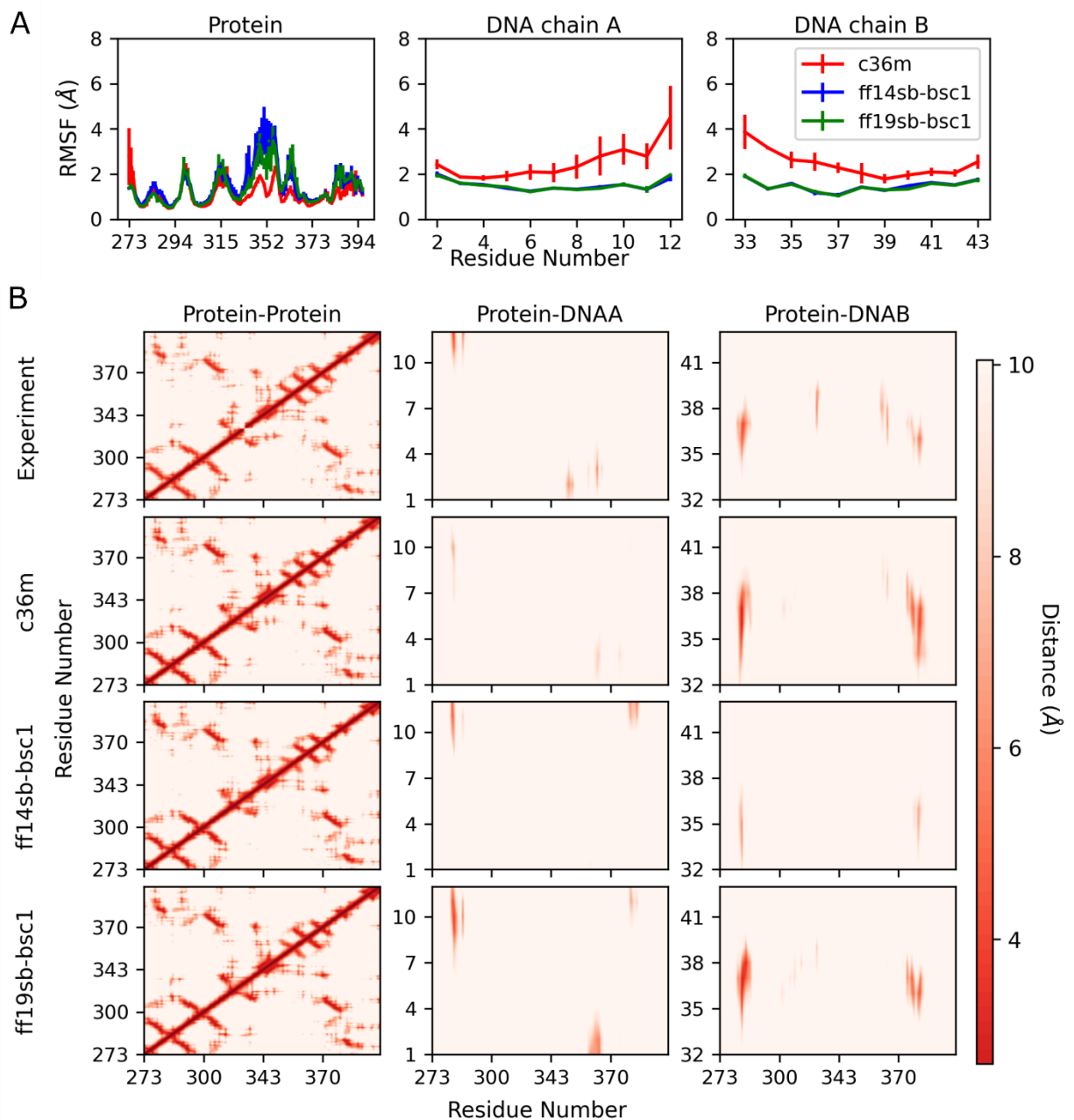


Figure 2. The RMSF and distance maps analysis for the simulations of the KOW1-DNA system with CHARMM c36m, AMBER ff14sb and AMBER ff19sb force fields. (A) RMSF values were calculated for protein, DNAA, and DNAB, (B) distance maps between protein-protein, protein-DNAA and protein-DNAB residues were calculated from the distances for the experimental structure (PDB ID:5oik) and average distances over the trajectories.

In Fig. 2B, we compared the distance maps of native structures for KOW1-DNA systems with the distances averaged over the simulations. Protein-protein interactions are mostly retained during the simulations with CHARMM and AMBER ff19sb force fields, while some of the interactions for the loop regions are lost in the simulations with AMBERff14sb. The two loops with residues, 335-340 and 285-289 have close distances at the native structures, while AMBER ff14sb simulations produced larger distances for these loops. Protein-DNA distances showed larger changes between force fields than the protein-protein distances in the simulations. Protein-DNA interactions were partly retained with three force fields. DNAB shows strong interactions with protein for CHARMM and AMBERff19sb, although some distances were larger especially for the residues K317 and R365. However, most of the distances were increased with AMBERff14sb. This is partly due to one replicate of the AMBER ff14sb simulations, in which DNA and protein interactions were mostly lost comparing to the native state (Fig S4) and this affected the average distances over the replicates. Overall CHARMM and AMBER ff19sb simulations provided more consistent average distance maps with the native structures, while AMBER ff14sb showed increased distances for protein-protein and protein-DNA distance maps.

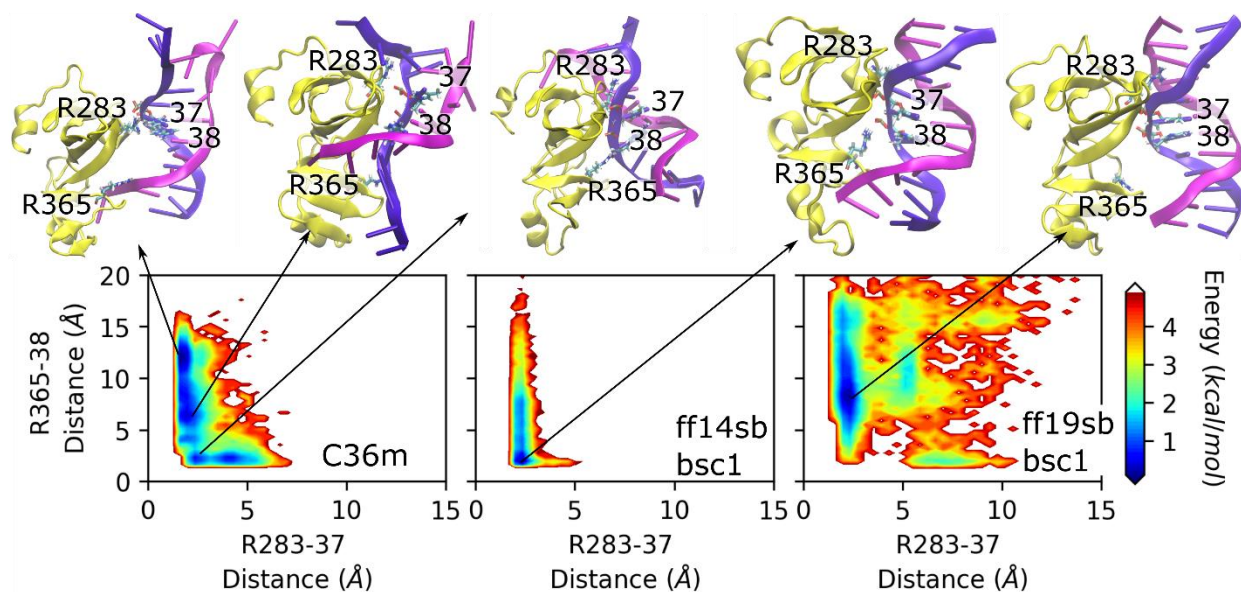


Figure 3. Free energy landscapes for the distances between R283 of protein and residue 37 of DNAB (x-axis) and R365 of protein and residue 38 of DNAB (y-axis). On the top, the snapshots at the minimum energy conformations are shown. Proteins were superimposed to the native state.

To have a closer look on how protein and DNA are interacting, we plotted the free energy landscapes for the selected distances of R283 and R365 with the DNAB residues 37 and 38, respectively. We selected R283 and R365 for this analysis since they have close distances with the DNA residues in the initial structure (Fig. 2B). Fig. 3 shows the free energy plots for the selected distances and the most probable structures. In one of the AMBER ff14sb simulations, protein and DNA lost most of the interactions (see Fig S4) and, therefore, we excluded this trajectory in the analysis shown in Fig 3. CHARMM c36m simulations provided an energy landscape with multiple minimum energy conformations while AMBER force fields provided only one minimum energy

conformations. DNA chains with AMBER force fields were more stable as seen in the RMSF curve and Fig. 3 also shows that the conformations from AMBER simulations retained most of the base-paired interactions. On the other hand, in CHARMM simulations base-pair interactions are lost to some extent. Distance maps between DNA chains in Fig S5 also show that DNAA-DNAB distances at the end of the chains were increased in CHARMM c36m simulations suggesting that base-pair interactions for terminal residues are less stabilized comparing to the AMBER force fields. To quantify the base-pair interactions, we calculated the average number of hydrogen bonds for Watson-Crick pairs and found that CHARMM c36m simulations provided only around the half of the hydrogen bond count obtained by AMBER simulations (Table 1).

Table 1. Average number of hydrogen bonds.

	¹ Number of H-bonds
² Experiment	24
C36m	9.78±0.02
ff14sb-bsc1	19.13±0.02
ff19sb-bsc1	19.25±0.02
C36m (Pol II-EC complex)	14.85±0.02

¹Number of H-bonds between Watson-Crick pairs was calculated for each frame and then averaged over the trajectories. ²Number of H-bonds was calculated from the initial structure that was extracted from the structure with PDB ID:50IK.

To understand electrostatics and vdW energy contributions to the DNA and protein binding energies, we performed MMBG/SA calculations (see details in Methods). All three force fields provided overestimation of both electrostatics and vdW binding energies while in CHARMM c36m simulations, overestimation was far greater compared to the AMBER force fields (Table 2). Overall, AMBER force fields were stabilizing both DNA base-pair and protein-DNA interactions; CHARMM c36m force field stabilized the protein-DNA interactions to a larger extent while it resulted in more fluctuations in DNA chains with decreased base pair interactions.

KOW4-RNA interactions

KOW4 subunit of SPT5 protein has mostly well-defined secondary structure and smaller loop regions comparing to KOW1. This is retained for the simulations with all three force fields as RMSD values converged to less than 2 Å (Fig. S2) and RMSF plots (Fig. 4A) show only small fluctuations. However, for RNA, RMSD values were large and fluctuating for all the force fields and RMSF values were higher than proteins with larger error bars (Fig. 4A). Large fluctuations in RNA structure were expected as it does not have well-defined secondary structure.

Distance maps for the protein (KOW4) were also maintained in large extent during the simulations as protein structures were mostly preserved (Fig. 4B). However, the distance maps between protein and RNA change to some extent with force fields. RNA is in close contact with the residues of K579 and R619 of the protein in the experimental structure. Those interaction sites were also observed in all three force fields with some differences. AMBER force fields showed strong interactions of RNA with protein at the K579 and R619 residues, whereas CHARMM force

field simulations provided longer average distances for the K579 and R619 residues. Fig. 4C shows the free energy plots for the distances of K579 and R619 with RNA residues of 37 and 33, respectively. Simulations with CHARMM force field showed a broader landscape with multiple minimum energies whereas simulations with AMBER force fields provided relatively narrower landscapes especially with ff19sb. Overall, RNA is more flexible in CHARMM simulations and its interaction sites with protein change frequently, while AMBER simulations mostly maintained the interaction sites.

Table 2. Binding free energies calculated using MMGB/SA method.

KOW1-DNA			
	Total binding energy	Electrostatic binding energy	vdW binding energy
¹ Experiment	-21.56	-1364.92	-14.68
C36m	-103.86 ± 0.76	-2134.27 ± 5.25	-60.11 ± 0.48
ff14sb-bsc1	-71.86 ± 0.76	-1765.30 ± 10.36	-45.96 ± 0.46
ff19sb-bsc1	-79.76 ± 0.45	-2031.15 ± 5.93	-46.29 ± 0.32
C36m (Pol II-EC complex)	-55.03 ± 0.52	-1832.52 ± 6.28	-34.36 ± 0.55
KOW4-RNA			
	Total binding energy	Electrostatic binding energy	vdW binding energy
Experiment	-19.34	-844.44	-11.15
C36m	-69.81 ± 0.56	-929.34 ± 2.66	-47.13 ± 0.37
ff14sb-ol3	-96.31 ± 0.69	-1128.82 ± 3.96	-75.34 ± 0.67
ff19sb-ol3	-101.67 ± 0.60	-1114.53 ± 3.16	-74.80 ± 0.36
C36m (Pol II-EC complex)	-64.06 ± 0.57	-985.71 ± 3.14	-43.01 ± 0.35

¹Energies were calculated for the initial structure that was extracted from the structure with PDB ID:50IK after minimization.

Table 2 shows the binding energies for RNA and protein. Like KOW1-DNA interactions, KOW4-RNA interactions were also over-stabilized by the three force fields. In this case, AMBER force fields stabilized the protein-RNA interactions more than CHARMM force field, which is consistent with the distance maps that protein-RNA distances are smaller with AMBER force fields that imply stronger interactions.

Overall, the force fields were able to provide consistent structures for both KOW1 and KOW4. Protein-nucleic acid interactions were more difficult to reproduce by the force-fields, due to high flexibility in the DNA and RNA chains and differences in stabilization of base-pairs and electrostatic interactions in AMBER and CHARMM force fields.

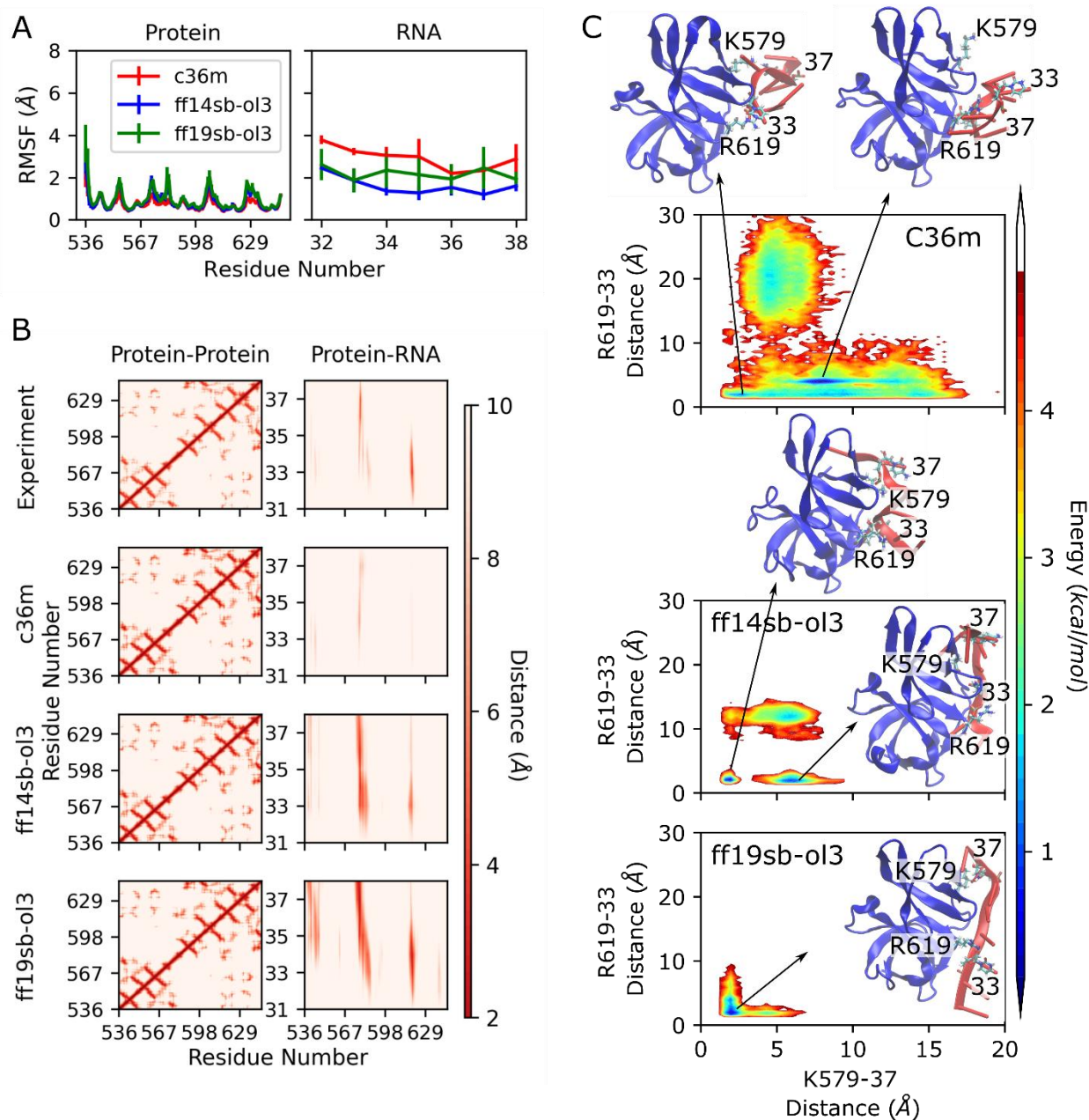


Figure 4. RMSF, distance maps, and free energy landscape analysis for the simulations of KOW4-RNA system with CHARMM c36m, AMBER ff14sb and AMBER ff19sb force fields. (A) RMSF values for protein and RNA were calculated, (B) distance maps calculated from the average distances over the trajectories between protein-protein and protein-RNA, (C) free energy landscapes for the distances between K579 of protein and residue 37 of RNA (x-axis) and R619 of protein and residue 33 of RNA (y-axis) along with the snapshots at the minimum energy conformations are shown.

KOW1-DNA and KOW4-RNA interactions in the complete elongation complex

KOW1 and KOW4 are domains in SPT5 elongation factor that is associated with RNA Pol II during the elongation step. We performed simulations of the complete elongation complex to observe the changes in interactions of KOW1-DNA and KOW4-RNA when they are in complex with RNA Pol II. Fig. S3 showed that DNA chains were converged to lower RMSD values comparing to the isolated system (Fig. S1) using CHARMM force field, which indicates that the DNA chains were stabilized further in the complete complex comparing to the isolated system and gave closer conformations to the experimental structure. Fig. S3 shows that RMSD values for the RNA chain are also lower compared to the isolated KOW4-RNA complex (Fig. S2) while they are much higher than the DNA chains. Similarly, RMSF values were reduced in complete complex for both DNA and RNA chains (Figs 5A and 5C). For the proteins, KOW1 and KOW4, the RMSD and RMSF values are similar for the isolated and full elongation complex systems.

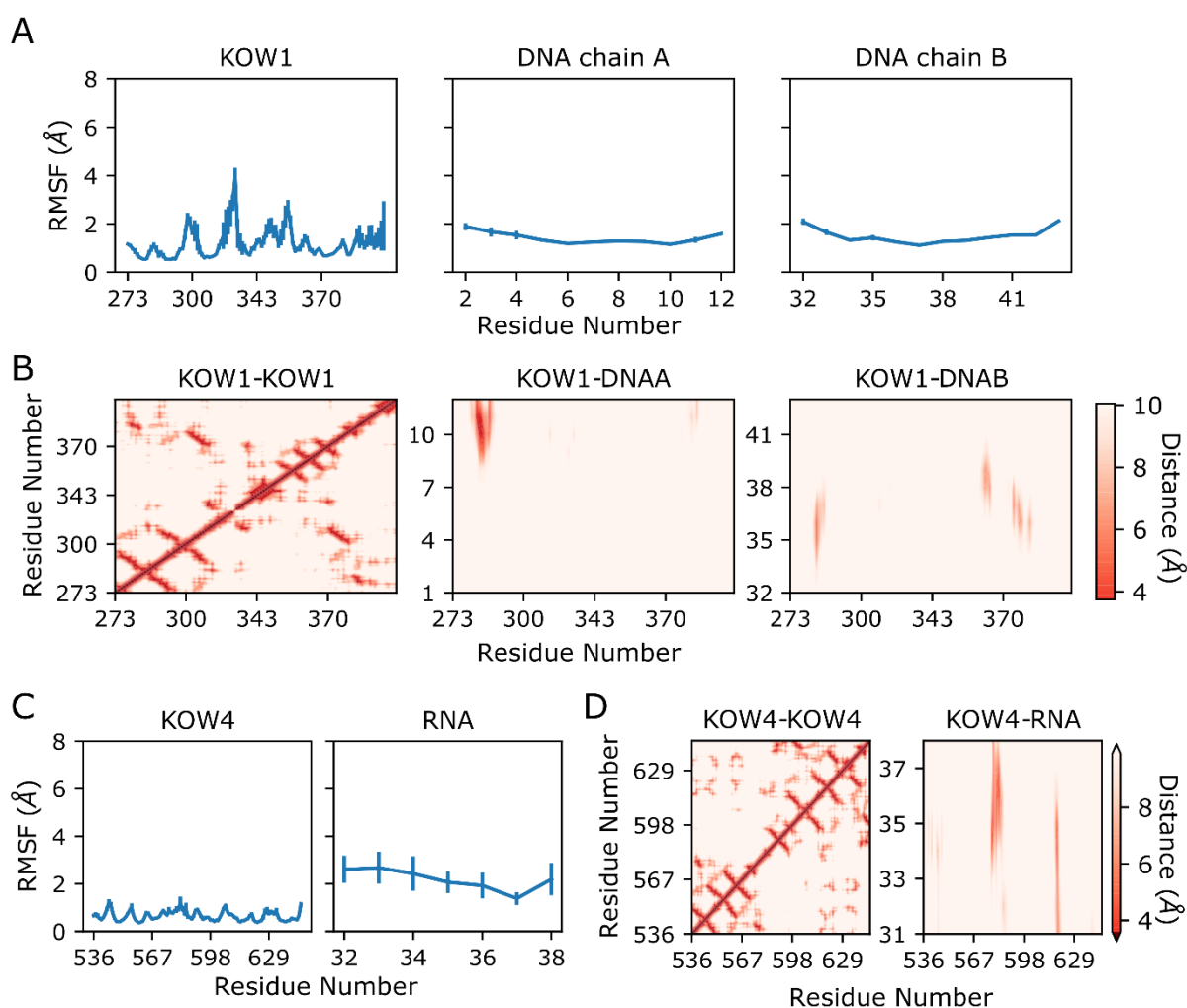


Figure 5. RMSF and distance maps for KOW1-DNA and KOW4-RNA from the simulation of the complete elongation complex with CHARMM c36m force field. (A) RMSF for KOW1 and DNA chains, (B) distance maps for KOW1-KOW1, KOW1-DNAA and KOW1-DNAB, (C) RMSF for KOW4 and RNA chains, (D) distance maps for KOW4-KOW4, KOW4-RNA.

The distance maps (Fig. 5B) showed that DNA and KOW1 were interacting strongly throughout the trajectories and interactions sites were mostly the same sites with the experimental structure (Fig. 2B). KOW4 and RNA distance map (Fig. 5D) was also similar to the map of the experimental structure (Fig. 4B). Table 2 shows that the binding energies for KOW1-DNA are closer to the experimental structure and less over-estimated compared to the isolated system, while the energies are similar to the isolated system for KOW4-RNA.

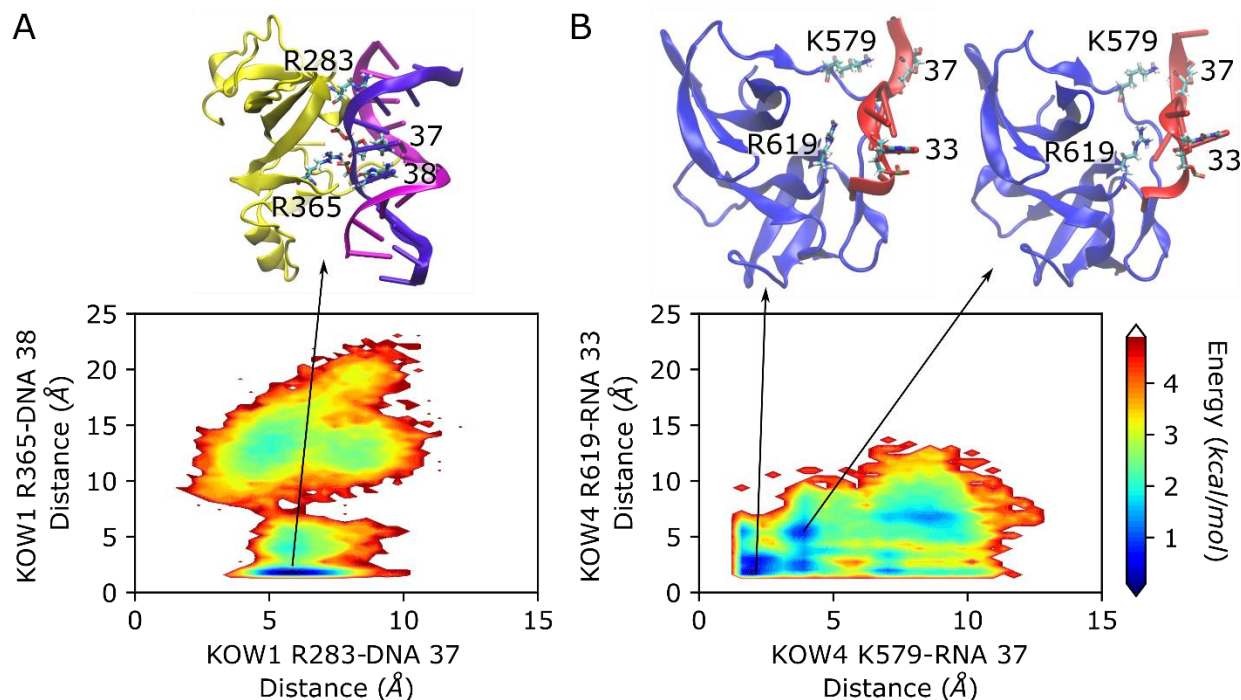


Figure 6. Free energy landscapes for the distances between KOW1 and DNA, and KOW4 and RNA. (A) R283 of KOW1 and residue 37 of DNAB (x-axis) and R365 of KOW1 and residue 38 of DNAB (y-axis); (B) K579 of KOW4 and residue 37 of RNA (x-axis) and R619 of KOW4 and residue 33 of RNA (y-axis); on the top are snapshots at the minimum energy conformations.

Fig. 6 shows the free energy landscape for the selected protein-DNA (Fig. 6A) and protein-RNA distances (Fig. 6B). Protein-DNA distances show a broad distribution like in the isolated CHARMM c36m simulations, however, there is one minimum energy conformation that is mostly populated. The minimum energy conformation shows that DNA chain retained base-pair interactions in larger extent compared to the isolated system. In agreement with this, hydrogen bond count between DNAA and DNAB chains were larger in the complete system (Table 1). Furthermore, we analyzed the number of contacts between the DNA chains and all the proteins in the elongation complex to observe if DNA has interactions with any other protein domain in addition to KOW1 that could contribute to the stabilization of the DNA chains. Table 3 shows that 45.63 % of the contacts of the DNA chains (the selected region in Fig 1) were observed with SPT5-KOW1 domain, while SPT4 domain and Pol II rpb2 subunit also have large number of contacts with DNA. These additional interactions may also contribute to the DNA dynamics such that the

DNA chains are fluctuating less and have more stabilized base pair interactions. This suggests that simulating the isolated KOW1-DNA system may not be sufficient to cover the DNA dynamics that is affected by a network of interactions within the elongation complex as seen in Fig. 7.

Table 3. Number of contacts between DNA and protein in the elongation complex.

Domain	¹ Number of contacts	Percentage of number of contacts
Pol II – rpb1	3.33	3.56
Pol II – rpb2	27.80	29.74
SPT4	13.43	14.36
SPT5 – KOW1	42.66	45.63
SPT5 – NGN	6.27	6.71

¹Number of contacts were calculated as the sum of the average contacts for the pairs for each protein domain. Pairs are defined in contact when they are within 5Å distance.

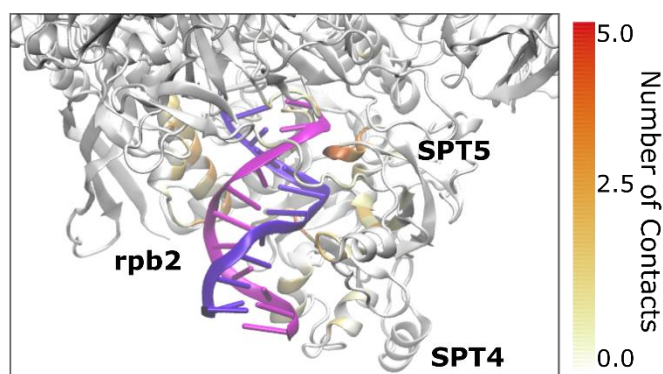


Figure 7. Number of contacts of DNA with the surrounding protein residues colored as depicted in the color bar. The elongation complex structure obtained at the minimum energy state in Fig. 6A was used.

Overall, the interactions of KOW1 and KOW4 with the nucleic acids are strong and stabilized in the full elongation complex similar to the isolated system. On the other hand, DNA fluctuations were reduced in the complete complex with an increased number of hydrogen bonds between the DNA pairs compared to the isolated system. This suggests that simulating DNA-protein systems in isolation may not cover the complete dynamic picture of the nucleic acids and the other interacting proteins should be included in the simulations to obtain higher accuracy in DNA dynamics.

DISCUSSION

Protein-nucleic acid interactions are important for many biological processes and MD simulations are a powerful method to probe these interactions at the atomistic detail in μs time scales. One important part of this approach is the capability of the force field to maintain accurate results in the protein-nucleic acid interactions and in the dynamics of both components. CHARMM and AMBER are the two most used force fields for protein-nucleic acid systems. Although they went through several validations for the proteins^{36, 39-41} and nucleic acids^{37-38, 42-43}, there are still imperfections reported for these force fields for protein-nucleic acid systems^{29, 31, 47}. In our work, we used three recent force fields, AMBER ff14sb and ff19sb with bsc1 for DNA and ol3 for RNA and CHARMM c36m, to investigate the differences between force fields in production of the protein and DNA interactions. We observed that the electrostatic interactions between protein and DNA or RNA are over-estimated with both AMBER and CHARMM force fields. The overestimation of the charge-charge interactions was reported by previous studies as well for both standard AMBER and CHARMM force fields^{28-29, 35, 47}. One remedy for accounting for the unrealistic electrostatic interactions was proposed to modify Lennard Jones parameters to balance pairwise interactions for the charged residues^{29, 35, 47}. Alternatively, there are studies focused on water models, proposing modified water-solute interactions, which could reduce the overestimation of solute-solute interactions⁶¹⁻⁶³.

Another result of our simulations is the difference in the stability of base pair interactions in AMBER and CHARMM force fields, with AMBER having larger stabilization effect on these interactions. For both AMBER and CHARMM force fields, we observed a smaller number of Watson-Crick hydrogen bonds than in the native state. This is consistent with what was found in the literature, that both standard AMBER and CHARMM force fields overestimate the stacking interactions that comes with the underestimation of the base pair interactions^{44-45, 64}, while we found the underestimation is larger in CHARMM c36m simulations. Recent studies on improvement of RNA force fields suggest multiple modifications on vdW parameters and charges on nucleobase to balance stacking interactions with the base pairing strength^{34, 46} although transferability of these modifications into the DNA force fields was not studied. In addition, four sided water models that are better balancing between solute-solute and solute-water interactions were also known to provide improvement in the stacking energies^{34, 64-65}.

The discrepancies in the protein nucleic acid interactions we observed in the isolated systems were reduced when we simulated the complete elongation system. The binding energies between the protein and DNA were still overestimated but closer to the experimental system compared to the isolated KOW1-DNA system. The base pair interactions between DNA chains were also improved in the complete elongation complex, but they are still underestimated comparing to the native state. Overall, the standard CHARMM c36m force field provides more accurate results as the system become more realistic and biologically relevant as a complex, but there are still imperfections in the protein-nucleic acid interactions produced by CHARMM c36m force field.

Our study altogether suggests that the protein nucleic acid interactions are captured well by both standard AMBER and CHARMM force fields, with some imperfections in the electrostatic

interactions and base pair interactions that require attention for obtaining realistic simulations for protein-nucleic acid systems. A future direction could be to apply suggested modifications that includes scaling the water-protein interactions as well as modifying Lennard Jones interactions between the protein and nucleic acids to improve the accuracies of the force fields.

CONCLUSION

Proteins and nucleic acids strongly interacted with each other during the simulations of KOW1 and KOW4 domains of SPT5 elongation factor with DNA and RNA using CHARMM and AMBER force fields. However, the interactions between protein and nucleic acids were over stabilized with both CHARMM and AMBER force fields compared to the native interactions. Simulations of the complete elongation complex, on the other hand, provided a more realistic environment and reduced the discrepancies observed in the isolated system to some extent. In addition, we observed that the DNA was stabilized by a network of interactions with the proteins in the Pol II elongation complex suggesting that isolated systems may not provide accurate dynamics for DNA. Future studies will focus on simulating the complete elongation complex rather than using the isolated systems and applying the modifications suggested in the literature into the standard force fields to explore the interactions of elongation factors with nucleic acids as well as the changes in DNA and RNA dynamics.

AUTHOR INFORMATION

Corresponding Authors

*Bercem Dutagaci

5200 North Lake Rd.

Merced, CA 95343

bercemdutagaci@gmail.com

Note:

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We used the computational resources at the MERCED cluster at University of California Merced and at the National Science Foundation's Extreme Science and Engineering Discovery Environment (XSEDE) facilities under the grant TG-BIO210145.

SUPPORTING INFORMATION:

Figures S1-5 are provided as Supporting Material.

REFERENCES

1. Hantsche, M.; Cramer, P., The Structural Basis of Transcription: 10Years After the Nobel Prize in Chemistry. *Angew Chem Int Edit* **2016**, *55* (52), 15972-15981.
2. Chen, H. T.; Hahn, S., Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: A model for the structure of PIC. *Cell* **2004**, *119* (2), 169-180.
3. Chen, H. T.; Warfield, L.; Hahn, S., The positions of TFIIF and TFIIE in the RNA polymerase II transcription preinitiation complex. *Nat Struct Mol Biol* **2007**, *14* (8), 696-703.
4. Chen, Z. A.; Jawhari, A.; Fischer, L.; Buchen, C.; Tahir, S.; Kamenski, T.; Rasmussen, M.; Lariviere, L.; Bukowski-Wills, J. C.; Nilges, M.; Cramer, P.; Rappsilber, J., Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *Embo J* **2010**, *29* (4), 717-726.
5. 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 Res* **2010**, *38* (12), 4040-4051.
6. Joo, Y. J.; Ficarro, S. B.; Chun, Y. J.; Marto, J. A.; Buratowski, S., In vitro analysis of RNA polymerase II elongation complex dynamics. *Gene Dev* **2019**, *33* (9-10), 578-589.
7. Bernecky, C.; Plitzko, J. M.; Cramer, P., Structure of a transcribing RNA polymerase II-DSIF complex reveals a multidentate DNA-RNA clamp. *Nat Struct Mol Biol* **2017**, *24* (10), 809-815.
8. 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.
9. Nogales, E.; Louder, R. K.; He, Y., Structural Insights into the Eukaryotic Transcription Initiation Machinery. *Annu Rev Biophys* **2017**, *46*, 59-83.
10. Osman, S.; Cramer, P., Structural Biology of RNA Polymerase II Transcription: 20 Years On. *Annu Rev Cell Dev Bi* **2020**, *36*, 1-34.
11. Sainsbury, S.; Bernecky, C.; Cramer, P., Structural basis of transcription initiation by RNA polymerase II. *Nat Rev Mol Cell Bio* **2015**, *16* (3), 129-143.
12. Vos, S. M.; Farnung, L.; Urlaub, H.; Cramer, P., Structure of paused transcription complex Pol II-DSIF-NELF. *Nature* **2018**, *560* (7720), 601-606.
13. Park, H.; Kang, H.; Ko, W.; Lee, W.; Jo, K.; Lee, H. S., FRET-based analysis of protein-nucleic acid interactions by genetically incorporating a fluorescent amino acid. *Amino Acids* **2015**, *47* (4), 729-734.
14. Zhang, S.; Metelev, V.; Tabatadze, D.; Zamecnik, P. C.; Bogdanov, A., Fluorescence resonance energy transfer in near-infrared fluorescent oligonucleotide probes for detecting protein-DNA interactions. *P Natl Acad Sci USA* **2008**, *105* (11), 4156-4161.
15. Kaptein, R., NMR studies on protein-nucleic acid interaction. *J Biomol Nmr* **2013**, *56* (1), 1-2.
16. Lacabanne, D.; Boudet, J.; Malar, A. A.; Wu, P. Z.; Cadalbert, R.; Salmon, L.; Alain, F. H. T.; Meier, B. H.; Wiegand, T., Protein Side-Chain-DNA Contacts Probed by Fast Magic-Angle Spinning NMR. *J Phys Chem B* **2020**, *124* (49), 11089-11097.
17. Cozzolino, F.; Iacobucci, I.; Monaco, V.; Monti, M., Protein-DNA/RNA Interactions: An Overview of Investigation Methods in the -Omics Era. *J Proteome Res* **2021**, *20* (6), 3018-3030.
18. Hanson, C. L.; Robinson, C. V., Protein-nucleic acid interactions and the expanding role of mass spectrometry. *J Biol Chem* **2004**, *279* (24), 24907-24910.
19. Rambo, R. P.; Tainer, J. A., Bridging the solution divide: comprehensive structural analyses of dynamic RNA, DNA, and protein assemblies by small-angle X-ray scattering. *Curr Opin Struct Biol* **2010**, *20* (1), 128-137.
20. Emamjomeh, A.; Choobineh, D.; Hajieghrari, B.; MahdiNezhad, N.; Khodavirdipour, A., DNA-protein interaction: identification, prediction and data analysis. *Mol Biol Rep* **2019**, *46* (3), 3571-3596.

21. Ferraz, R. A. C.; Lopes, A. L. G.; da Silva, J. A. F.; Moreira, D. F. V.; Ferreira, M. J. N.; Coimbra, S. V. D., DNA-protein interaction studies: a historical and comparative analysis. *Plant Methods* **2021**, *17* (1).
22. Fried, M. G., Measurement of Protein-DNA Interaction Parameters by Electrophoresis Mobility Shift Assay. *Electrophoresis* **1989**, *10* (5-6), 366-376.
23. Welsh, J.; Cantor, C. R., Protein - DNA Cross-Linking. *Trends Biochem Sci* **1984**, *9* (12), 505-508.
24. Das, P. M.; Ramachandran, K.; vanWert, J.; Singal, R., Chromatin immunoprecipitation assay. *Biotechniques* **2004**, *37* (6), 961-969.
25. Gilmour, D. S.; Lis, J. T., In vivo Interactions of Rna Polymerase-ii with Genes of Drosophila-Melanogaster. *Mol Cell Biol* **1985**, *5* (8), 2009-2018.
26. Chocholousova, J.; Feig, M., Implicit solvent simulations of DNA and DNA-protein complexes: Agreement with explicit solvent vs experiment. *J Phys Chem B* **2006**, *110* (34), 17240-17251.
27. MacKerell, A. D.; Nilsson, L., Molecular dynamics simulations of nucleic acid-protein complexes. *Curr Opin Struc Biol* **2008**, *18* (2), 194-199.
28. Yoo, J.; Winogradoff, D.; Aksimentiev, A., Molecular dynamics simulations of DNA-DNA and DNA-protein interactions. *Curr Opin Struc Biol* **2020**, *64*, 88-96.
29. You, S.; Lee, H. G.; Kim, K.; Yoo, J., Improved Parameterization of Protein-DNA Interactions for Molecular Dynamics Simulations of PCNA Diffusion on DNA. *J Chem Theory Comput* **2020**, *16* (7), 4006-4013.
30. Gapsys, V.; de Groot, B. L., Alchemical Free Energy Calculations for Nucleotide Mutations in Protein-DNA Complexes. *J Chem Theory Comput* **2017**, *13* (12), 6275-6289.
31. Liao, Q. H.; Luking, M.; Kruger, D. M.; Deindl, S.; Elf, J.; Kasson, P. M.; Kamerlin, S. C. L., Long Time-Scale Atomistic Simulations of the Structure and Dynamics of Transcription Factor-DNA Recognition. *J Phys Chem B* **2019**, *123* (17), 3576-3590.
32. Maffeo, C.; Chou, H. Y.; Aksimentiev, A., Molecular Mechanisms of DNA Replication and Repair Machinery: Insights from Microscopic Simulations. *Adv Theor Simul* **2019**, *2* (5).
33. Wang, B. B.; Feig, M.; Cukier, R. I.; Burton, Z. F., Computational Simulation Strategies for Analysis of Multisubunit RNA Polymerases. *Chem Rev* **2013**, *113* (11), 8546-8566.
34. Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E., RNA force field with accuracy comparable to state-of-the-art protein force fields. *P Natl Acad Sci USA* **2018**, *115* (7), E1346-E1355.
35. Yoo, J.; Aksimentiev, A., New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys Chem Chem Phys* **2018**, *20* (13), 8432-8449.
36. 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. *Nat Methods* **2017**, *14* (1), 71-73.
37. 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. *J Comput Chem* **2011**, *32* (9), 1929-1943.
38. Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D., Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J Chem Theory Comput* **2012**, *8* (1), 348-362.
39. 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. *J Chem Theory Comput* **2012**, *8* (9), 3257-3273.
40. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **2015**, *11* (8), 3696-3713.

41. Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Miguez, A. N.; Bickel, J.; Wang, Y. Z.; Pincay, J.; Wu, Q.; Simmerling, C., ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput* **2020**, *16* (1), 528-552.
42. Ivani, I.; Dans, P. D.; Noy, A.; Perez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goni, R.; Balaceanu, A.; Portella, G.; Battistini, F.; Gelpi, J. L.; Gonzalez, C.; Vendruscolo, M.; Lughton, C. A.; Harris, S. A.; Case, D. A.; Orozco, M., Parmbsc1: a refined force field for DNA simulations. *Nat Methods* **2016**, *13* (1), 55-58.
43. Zgarbova, M.; Otyepka, M.; Sponer, J.; Mladek, A.; Banas, P.; Cheatham, T. E.; Jurecka, P., Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput* **2011**, *7* (9), 2886-2902.
44. Banas, P.; Mladek, A.; Otyepka, M.; Zgarbova, M.; Jurecka, P.; Svozil, D.; Lankas, F.; Sponer, J., Can We Accurately Describe the Structure of Adenine Tracts in B-DNA? Reference Quantum-Chemical Computations Reveal Overstabilization of Stacking by Molecular Mechanics. *J Chem Theory Comput* **2012**, *8* (7), 2448-2460.
45. Morgado, C. A.; Jurecka, P.; Svozil, D.; Hobza, P.; Sponer, J., Balance of Attraction and Repulsion in Nucleic-Acid Base Stacking: CCSD(T)/Complete-Basis-Set-Limit Calculations on Uracil Dimer and a Comparison with the Force-Field Description. *J Chem Theory Comput* **2009**, *5* (6), 1524-1544.
46. Chen, A. A.; Garcia, A. E., High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *P Natl Acad Sci USA* **2013**, *110* (42), 16820-16825.
47. Yoo, J.; Aksimentiev, A., Improved Parameterization of Amine-Carboxylate and Amine-Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *J Chem Theory Comput* **2016**, *12* (1), 430-443.
48. Jo, S.; Kim, T.; Iyer, V. G.; Im, W., Software news and updates - CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem* **2008**, *29* (11), 1859-1865.
49. 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.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W., CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* **2016**, *12* (1), 405-413.
50. Lee, J.; Hitznerberger, M.; Rieger, M.; Kern, N. R.; Zacharias, M.; Im, W., CHARMM-GUI supports the Amber force fields. *J Chem Phys* **2020**, *153* (3), 035103.
51. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J Chem Phys* **1983**, *79* (2), 926-935.
52. 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.; Tye, T.; Houston, M.; Stich, T.; Klein, C.; Shirts, M. R.; Pande, V. S., OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J Chem Theory Comput* **2013**, *9* (1), 461-469.
53. Fiser, A.; Do, R. K. G.; Sali, A., Modeling of loops in protein structures. *Protein Sci* **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. *J Mol Graph Model* **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.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **2009**, *30* (10), 1545-1614.
56. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O., Software News and Updates MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J Comput Chem* **2011**, *32* (10), 2319-2327.

57. Grossfield, A. *Grossfield, Alan, WHAM: the weighted histogram analysis method*”, version 2.0.11, http://membrane.urmc.rochester.edu/?page_id=126/.
58. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J Comput Chem* **1995**, *16* (11), 1339-1350.
59. Thorsteinsdottir, H. B.; Schwede, T.; Zoete, V.; Meuwly, M., How inaccuracies in protein structure models affect estimates of protein-ligand interactions: Computational analysis of HIV-1 protease inhibitor binding. *Proteins* **2006**, *65* (2), 407-423.
60. Wittayanarakul, K.; Hannongbua, S.; Feig, M., Accurate prediction of protonation state as a prerequisite for reliable MM-PB(GB)SA binding free energy calculations of HIV-1 protease inhibitors. *J Comput Chem* **2008**, *29* (5), 673-685.
61. Ashbaugh, H. S.; Collett, N. J.; Hatch, H. W.; Staton, J. A., Assessing the thermodynamic signatures of hydrophobic hydration for several common water models. *J Chem Phys* **2010**, *132* (12), 124504.
62. Best, R. B.; Zheng, W. W.; Mittal, J., Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J Chem Theory Comput* **2014**, *10* (11), 5113-5124.
63. Nerenberg, P. S.; Jo, B.; So, C.; Tripathy, A.; Head-Gordon, T., Optimizing Solute-Water van der Waals Interactions To Reproduce Solvation Free Energies. *J Phys Chem B* **2012**, *116* (15), 4524-4534.
64. Hase, F.; Zacharias, M., Free energy analysis and mechanism of base pair stacking in nicked DNA. *Nucleic Acids Res* **2016**, *44* (15), 7100-7108.
65. Bergonzo, C.; Cheatham, T. E., Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J Chem Theory Comput* **2015**, *11* (9), 3969-3972.