

Mechanism of Dynamic Binding of Replication Protein A to ssDNA

Anupam Mondal and Arnab Bhattacharjee

School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi, India.

* Arnab Bhattacharjee.

Email: arnab@jnu.ac.in

<https://orcid.org/0000-0002-7714-2619>

Classification

Biophysics and Computational Biology

Keywords

ssDNA-protein interactions, Replication protein A, dynamic binding

Author Contributions

A.B. designed the research; A.M., and A.B. performed the research, analysed the data and wrote the paper.

This PDF file includes:

Main Text

Figures 1 to 6

Tables 1 to 2

Abstract

Replication protein A (RPA) serves as hub protein inside eukaryotic cells, where it coordinates crucial DNA metabolic processes and activates the DNA-damage response system. A characteristic feature of its action is to associate with ssDNA intermediates before handing over them to downstream proteins. The length of ssDNA intermediates differs for different pathways. This means RPA must have mechanisms for selective processing of ssDNA intermediates based on their length, the knowledge of which is fundamental to elucidate when and how DNA repair and replication processes are symphonized. By employing extensive molecular simulations, we investigated the mechanism of binding of RPA to ssDNA of different lengths. We show that the binding involves dynamic equilibrium with a stable intermediate, the population of which increases with the length of ssDNA. The vital underlying factors are decoded through collective variable principal component analysis. It suggests a differently orchestrated set of interactions that define the action of RPA based on the sizes of ssDNA intermediates. We further estimated the association kinetics and probed the diffusion mechanism of RPA to ssDNA. RPA diffuses on short ssDNA through progressive 'bulge' formation. With long ssDNA, we observed a conformational change in ssDNA coupled with its binding to RPA in a cooperative fashion. Our analysis explains how the 'short-lived,' long ssDNA intermediates are processed quickly in vivo. The study thus reveals the molecular basis of several recent experimental observations related to RPA binding to ssDNA and provides novel insights into the RPA functioning in DNA repair and replication.

Significance Statement

Despite ssDNA be the common intermediate to all pathways involving RPA, how does the latter function differently in the DNA processing events such as DNA repair, replication, and recombination just based on the length of ssDNA intermediates remains unknown. The major hindrance is the difficulty in capturing the transient interactions between the molecules. Even attempts to crystallize RPA complexes with 32nt and 62nt ssDNA have yielded a resolved structure of only 25nt ssDNA wrapped with RPA. Here, we used a state-of-the-art coarse-grained protein-ssDNA model to unravel the detailed mechanism of binding of RPA to ssDNA. Our study illustrates the molecular origin of variations in RPA action during various DNA processing events depending on the length of ssDNA intermediates.

Main Text

Introduction

Replication protein A (RPA) in eukaryotes plays a crucial role in DNA metabolic processes such as DNA replication, recombination, and repair (1–3). A common feature in all its pathways is that RPA interacts with ssDNA intermediates. For example, RPA rapidly coats the ssDNA regions during the initiation phase of DNA replication to provide stability and prevent the formation of secondary structures in ssDNA (4, 5). During the elongation phase, RPA coordinates the lagging strand polymerase switching and processing of Okazaki fragments (6). RPA also participates in DNA nucleotide excision repair (NER) process through its interactions with XPA to stabilize the open complex after locating the lesion and help in positioning the nuclease for the dual incision (7). Similarly, RPA sequesters the ssDNA during repair of DNA double-strand breaks (DSBs) and promotes the loading of RAD51 recombinase to form RAD51– ssDNA nucleoprotein filament (8).

Despite the myriad roles of RPA, the mechanism by which it can direct different ssDNA intermediates in distinct pathways to coordinate the various DNA metabolic processes remains obscure. The functions of RPA have been found to depend on its binding affinity to ssDNA and its ability to interact with other proteins(9) physically. Interestingly, RPA is a multi-domain protein, and its affinity for ssDNA is determined by a combined contributions from its six DNA binding domains (DBDs A-F). The DBDs are made of structurally-related, oligonucleotide binding (OB) folds, which are further grouped into three subunits, RPA70, RPA32, and RPA14 (see Fig 1). RPA70 is composed of four DBDs, namely F, A, B, and C (flexible F domain which remains disconnected from the rest of the complex is not shown in Fig 1A), whereas RPA32 and RPA14 comprise DBD-D with a winged-helix domain and DBD-E respectively. Among six DBDs, DBD-A and -B have the highest affinity ($k_a \sim 5 \times 10^5 \text{ M}^{-1}$ and $5 \times 10^4 \text{ M}^{-1}$ respectively) for ssDNA (10, 11), whereas DBD-C and DBD-D exhibit weaker binding affinities (11–13). The latter two DBDs, along with DBD-E, form a trimerization core (14). The high affinity of DBD-A and DBD-B is primarily due to their interactions with ssDNA through a combination of polar and aromatic amino acids (15). This is evident from mutational analysis that suggests a significant reduction in affinity of RPA for ssDNA caused by mutations of polar amino acids. It is, however, interesting to note that the same is not valid for mutations at aromatic sites (11, 16). Mutations of aromatic residues have minimal impact on RPA affinity, but they are linked with a loss of DNA repair activity of RPA without any significant effects on its role in DNA replication(16). This indicates that while the affinity of RPA for ssDNA is essential for their assembly, the function of RPA in different pathways differ based on the details of intermolecular interactions. A difference also exists in the length of ssDNA intermediates involved in various DNA metabolic pathways (4, 11, 17, 18). For example, long (100-200 nt) but short-lived ssDNA intermediates are observed in DNA replication, whereas much shorter ssDNA intermediates (<30 nt) form a stable RPA-ssDNA complex for processing and repairing the DNA damage via the NER pathway (19). Does this mean the length of ssDNA intermediates regulate the action of RPA in various DNA metabolic pathways? If yes, what is the underlying molecular mechanism and what are the molecular determinants that differentiate the association of RPA to ssDNA based on the length of the latter? While answers to the questions are essential to know when and how specific DNA repair and recombination processes are orchestrated, very little has been explored so far.

In this study, we have computationally explored the RPA-ssDNA binding energy landscape using a state-of-the-art coarse-grained (CG) protein-DNA model to investigate their association mechanism at molecular details. The model parameters were precisely tuned by employing the model in successfully predicting the binding of different ssDNA sequences to their respective protein partners. Our results suggest the presence of two distinct RPA-ssDNA binding modes that are connected via a dynamic equilibrium. The relative population of these states is a function of the length of ssDNA intermediates and their stability is governed by a set of intermolecular interactions. By applying a collective variable principal component analysis (PCA), we analysed the weights of each component of intermolecular interactions. The finding, together with the mechanistic details of ssDNA propagation on RPA, unravels an unanticipated mechanism of their association and provides crucial insights into the RPA functioning in DNA repair and replication.

Results

Binding of RPA to ssDNA varies with the length of ssDNA

To begin with, we first test if our model has achieved the correct balance between the weighted strengths of various protein-nucleic acid interactions and the conformational flexibility of the ssDNA tract to capture the RPA-ssDNA binding dynamics. For this, we performed binding simulations of a (dT)₂₅ ssDNA with a RPA molecule starting from a completely unbound state. To monitor the progress of RPA association to ssDNA, we select the specific interfacial contacts between them as a reaction coordinate. An interfacial contact is recognised if two non-hydrogenous atoms, each from an amino acid of RPA and a nucleotide of ssDNA are within 3.7 Å in the crystal structure of RPA-(dT)₂₅ (PDB ID: 4gop.pdb) complex. We found 57 such contacts (17 charged, 13 aromatic, 15 polar and 12 hydrophobic residues). The analysis is presented in Fig 1B, which closely resembles the RPA-ssDNA contacts probed by Fan et. al. (20). The heterogeneity in the type of interfacial residues suggests an orchestrated set of interactions behind the formation of the bound state complex. Domain wise, these contacts are spread over all four DBDs (A-D), among which DBD-C forms the maximum number of specific interfacial contacts with ssDNA despite its lowest affinity towards ssDNA. Towards this end, we note that the DBDs of RPA are interconnected by flexible linkers that allow them to interact independently with the ssDNA and assume an ensemble of conformations. Therefore, to understand if the specific interfacial contacts are formed sequentially along the length of ssDNA, we divided the length of the ssDNA into approximately two halves. The strand containing consecutive 13 nucleotides from the 3'-end is referred as *arm-1* and feature ~27 specific contacts ($N_{specific}^{Arm1}$, 8 charged, 7 aromatic, 8 polar and 4 hydrophobic residues), whereas 12 nucleotides from the 5'-end is termed as *arm-2* that contains ~30 specific contacts ($N_{specific}^{Arm2}$, 9 charged, 6 aromatic, 7 polar and 8 hydrophobic residues). We estimated the arm-wise specific contact formation probabilities from trajectories of RPA-(dT)₂₅ binding simulations. The result presented in Fig 1C shows the most probable region corresponds to a high number of specific contacts in both *arm-1* and *arm-2* of ssDNA. This essentially indicates capturing the correct bound state of RPA-(dT)₂₅ in our simulations, which is also supported by an excellent structural overlap (see Fig 1D) obtained from the superimposition of our simulation generated bound state of RPA-(dT)₂₅ on its crystal structure. Fig 1C also suggests that the route to the bound state of RPA-(dT)₂₅ goes through a simultaneous formation of the specific interfacial contacts (evident from the diagonal data set) along both the arms of ssDNA. This means binding of different DBDs of RPA to ssDNA is not sequential, rather coherent.

Having seen that our model precisely captures the heterogeneity in RPA-ssDNA binding, we turned to investigate how RPA interacts with ssDNA intermediates of different lengths. It is important to note that previously complexes of RPA with a short (32 nt) and a long (62 nt) ssDNA molecule were crystallised (4gop.pdb and 4gnx.pdb) with 3.1 Å and 2.8 Å resolutions respectively. Interestingly, only 25 nt long identical ssDNA structures were resolved in both the complexes. While this suggests that ~25 nt ssDNA stretch is sufficient to cover all four DBDs (A-D) of RPA, it fails to capture all the dynamic and transient contacts between nucleobases of ssDNA and amino acid residues of RPA that might define the differences in functioning of RPA during DNA replication and repair. Therefore, we probed the dynamics of binding of RPA separately with (dT)₂₅ (short ssDNA), (dT)₃₂ (dT)₄₀ (dT)₅₀ and (dT)₆₂ (long ssDNA). The free energy profiles corresponding to short ((dT)₂₅) and long ((dT)₆₂) ssDNA with respect to the number of arm-wise contacts between nucleobases of ssDNA and amino acid residues ($N_{contact}^{Arm1}$ and $N_{contact}^{Arm2}$) of RPA are presented in Fig 2A-B. The same for the other lengths of ssDNA can be found in Fig S3 in Supplementary text. Note that unlike $N_{specific}^{Arm1}$ and $N_{specific}^{Arm2}$, here we consider the formation of contacts whenever a nucleobase of ssDNA comes within 6 Å to an amino acid residue of RPA in our CG model. This allows us to monitor the progress of RPA binding to ssDNA even when they are non-specifically interacting. Our results show distinctly different free energy surfaces for (dT)₂₅ and (dT)₆₂. A single basin (denoted by H₂₅) is observed in the free energy surface of RPA-(dT)₂₅. The cluster of snapshots in Fig 2C corresponding to the basin H₂₅ shows a horse-shoe shaped binding mode similar to its crystal structure (4gop.pdb). In

contrast, the free energy surface of RPA-(dT)₆₂ shows two basins (L₆₂ and H₆₂). L₆₂ represents a stable intermediate state in RPA-(dT)₆₂ binding where $N_{contact}^{Arm1}$ is significantly high (~28 contacts) but $N_{contact}^{Arm2}$ is limited to very few contacts (~5 – 10 contacts). A cluster of snapshots (Fig 2D) corresponding to this state (state L₆₂), show almost a linear conformation adopted by ssDNA on the RPA surface. In these conformations, the *arm-1* of ssDNA is associated with DBD-D and DBD-C of RPA, but *arm-2* remains flanking on RPA surface. The second basin (H₆₂) on the free energy surface of RPA-(dT)₆₂ corresponds to high $N_{contact}^{Arm1}$ and $N_{contact}^{Arm2}$ (~28 contacts on each arm). The cluster of snapshots corresponding to H₆₂ presented in Fig 2D elucidates a conformational transition from linear ssDNA of L₆₂ to a ‘horse-shoe’ shaped bent conformation such that it can bind to DBD-B and DBD-A of RPA. We estimated the associated change in free energy, which suggests state-H₆₂ is slightly more stable ($\Delta G \sim 0.7 \pm 0.001$ kcal/mol, see Fig S4) compared to L₆₂. The small free energy gap signifies a dynamic interconversion equilibrium exists between states L₆₂ and H₆₂. An analogous free energy gap (~1.2 k_BT) was reported for protein diffusion around the major grooves of dsDNA during its ‘sliding dynamics’ (21). The variation in the dynamic binding of RPA to ssDNA depending on the length of the latter is further manifested from our binding studies of RPA with an intermediate length of ssDNA molecules ((dT)₃₂, (dT)₄₀, (dT)₅₀). Fig S5 shows as the length of ssDNA intermediate increases, the population of state-L_n (n is length of DNA) rises, indicating a shift in binding mechanism of RPA with the change in length of ssDNA intermediates. To this end, we also compared the stability of bound states of RPA with (dT)₂₅ (state H₂₅) and (dT)₆₂ (state H₆₂). We find the former is more stable ($\Delta G \sim 0.46 \pm 0.001$ kcal/mol), which is in agreement with the observation that RPA forms a more stable complex with short ssDNA intermediate during DNA repair compared to the long ssDNA tract in DNA replication (19).

Molecular determinants of dynamic binding of RPA to ssDNA

What could be the physical basis of differences in RPA binding mechanism to ssDNA of different lengths? To understand this, we employed a collective variable (CV) principal component analyses (PCA) (22) that can pull apart the hidden cross-correlations among the collective variables. Thus, the technique is advantageous over the free energy analysis based on projection solely onto a single input variable. Here, as input basis variables, we considered different types of interactions between the amino acid residues of RPA and nucleotides of ssDNA that work in symphony to lead to the formation of RPA-ssDNA bound complex. This includes the following collective variables (CV1 to CV4): (1) electrostatic energy ($E_{elec}(RPA:ssDNA)$) between phosphate residues of ssDNA and charged amino acids of RPA, (2) aromatic interaction energy (E_{aro}), (3) hydrophobic interaction energy (E_{hp}) (4) polar interaction energy (E_{polar}) between the nucleobases of ssDNA and the relevant classes of amino acids of RPA. Our aim is to re-analyse the binding free energy profiles of RPA with (dT)₂₅ and (dT)₆₂.

Fig 3 and 4 present the free-energy surfaces of RPA with (dT)₂₅ and (dT)₆₂ respectively plotted against various estimated principal components (PCs). There are four principal components available (PC1 to PC4) that represent a new basis set of the transformed coordinates. The basis sets are obtained from linear combinations of the collective basis variables that we have taken as input to PCA. The coefficients used in the linear combinations and the variance are listed in Table 1 and 2 for (dT)₂₅ and (dT)₆₂, respectively. For moderate lengths of ssDNA, analysis were done and presented in supplementary text (Fig S6, S7 and Table S12 and S13). Our analysis suggests that $E_{elec}(RPA:ssDNA)$ is the major driving force (common to all PCs) for forming the bound state complex. We, therefore, consider $E_{elec}(RPA:ssDNA)$ as a common coordinate and presented the free energy profiles of the systems individually against each principal component. For example, the first panel of Fig 3 and 4 represent the free energy surfaces projected onto the first principal component (PC1) that account for ~43% and ~42% of the total variance. In addition, PC2 and PC3 are also important components that together with PC1 present 95.3% and 96% of the total variances in RPA-(dT)₂₅ and RPA-(dT)₆₂ assembly respectively. The variation of PC4 is less than 5%, indicating its minimal impact in regulating the RPA-ssDNA assembly. A comparison of the associated free energy surfaces of the two systems confirms the presence of a single basin for RPA binding to short ssDNA but two distinct basins for long ssDNA. The result suggests that the RPA-(dT)₆₂ complex undergoes a dynamic interconversion equilibrium between the binding

modes corresponding to the two basins. For RPA-(dT)₂₅ the basin appears at $E_{elec}(RPA:ssDNA) = -25$ kcal/mol, whereas the basins for RPA-(dT)₆₂ complex appear at $E_{elec}(RPA:ssDNA) = -35$ kcal/mol and -55 kcal/mol respectively. The increasing electrostatic energy with the increasing length of ssDNA is due to the involvement of more number of phosphate residues of ssDNA interacting with the positively charged amino acids of RPA. Another important insight emerges from the analysis of the combinations of collective basis variables corresponding to the significant principal components. In RPA-(dT)₂₅ binding, PC1, PC2 and PC3 are made up of approximately 27%, 61% and 12% E_{polar} . Together these account for ~39% polar interactions between the nucleobases and the polar amino acids of RPA that stabilize the bound RPA-(dT)₂₅ complex. The analysis explains why mutations of the polar residues in RPA lower its affinity for short ssDNA (11, 16). Similar estimations for E_{aro} and E_{hp} indicate that the contributions of interactions of nucleobases with aromatic and hydrophobic amino acids towards the precise binding of RPA to (dT)₂₅ are 9.3% and 23.1% respectively. In contrast, the total contributions from polar, hydrophobic and aromatic interactions in the first three PCs for the RPA-(dT)₆₂ binding are 13.2%, 34.13% and 27.6% respectively. The comparable weights of all three types of interactions suggest that more heterogeneous intermolecular interactions drive the binding of RPA to long ssDNA. The interactions work in symphony to regulate the binding process and reduction of one component due to point mutations of a particular type of amino acids can be easily compensated by the other interactions. This is supported by the experimental observation that mutations of aromatic residues in RPA only minimally impact its affinity for ssDNA (16). Apart from polar and aromatic interactions, we found hydrophobic amino acids also play a significant role. We note that mutations of hydrophobic residues such as F238A, F269A, and F386A have reduced the affinity of RPA to ssDNA (16). However, it is difficult to assess to what extent the impact is due to a reduction in hydrophobic interactions as the mutations cause a simultaneous loss in aromatic interactions as well. The differences in the interaction pattern of RPA with varying length of ssDNA, thus define its differences in the mechanism of dynamic binding to ssDNA during DNA repair and replication processes.

Association mechanism of RPA differ for short and long ssDNA intermediates

Here, we investigate the molecular mechanism of RPA association to ssDNA, which requires probing the ssDNA propagation on RPA until it wraps the four DBDs (A-D) completely to form the bound state complex. Intuitively this is a long time-scale problem. Previous studies have rather investigated the translocation of ssDNA on protein in bound states only(23, 24). The analyses based on crystal structures of ssDNA bound E. coli SSB and RPA proteins demonstrate that short ssDNA segments(1-7nt) erect over the protein surface and form bulges due to defects in interactions between nucleobases of ssDNA and interfacial protein residues. With the dissolution of bulges, released nucleotides slide on the protein surface. The mechanism, referred as 'reptation dynamics', has been captured experimentally(25) although the formation of bulges is not. The picture is reminiscent of sliding mechanism of dsDNA on nucleosome via twist defect propagation(26).

To illustrate the RPA-ssDNA association mechanism, we start monitoring the movement of each nucleotide on RPA surface starting from when they established the first nonspecific contact with RPA. We noticed the formation of bulges on ssDNA as its characteristic feature. The bulges on RPA surface are identified if the distance (r_{ij}) between *i*-th and *j*-th phosphate atoms of ssDNA is less than at least by 60% to its Kuhn length in a stretched conformation and the first phosphate residue participating in a bulge is positioned within 10 Å from the RPA surface. Fig 5A and B represent characterization of such bulges and suggest the most preferred size of the bulges are 3-4 nt long and they are transient (lifetime decays exponentially, see Fig 5B). Too small or large bulges are energetically unstable. The result is in agreement with the previous studies done on ssDNA and RPA protein(23, 24). To probe if the formation of bulges on ssDNA helps it to associate with RPA, we studied the time evolution of all phosphate residues that appear first when a bulge forms. Green dots present indices of those phosphate residues in Fig 5C, D, and E representing the movement of bulges in (dT)₂₅ and arm-1 and arm-2 of (dT)₆₂ respectively. The results reveal three intriguing aspects of bulge propagation in the context of the association mechanism of RPA to ssDNA: (i) bulges form throughout the length of (dT)₂₅ tract. Some of them

dissolute quickly and do not propagate with time. These are ‘static bulges’ observed previously for ssDNA sliding on *E. coli* SSB protein. The net movement of ssDNA due to the dissolution of such bulges is modest. Alternatively, some bulges are dynamic that aid in transporting nucleotides of ssDNA on RPA surface. (ii) The dynamic bulges move from 3’ to 5’ direction in (dT)₂₅ ssDNA. Snapshots in Fig 5F taken at different time steps clearly portray the propagation of the bulges along the DBDs of RPA, leading to formation of the final RPA-(dT)₂₅ association complex. The molecular driving force for the directed motion of the bulges is possibly the long-ranged electrostatic interactions. (iii) A similar directed movement of bulges can also be noticed in arm-1 of (dT)₆₂. However, unlike arm-1 of (dT)₆₂, a significant length of *arm-2* does not show any bulge formation (a black region in Fig 5E) up to MD steps $\sim 12 \times 10^6$ MD. This corresponds to state L₆₂ (see Fig 2), where ssDNA assumes a linear configuration with approximately one-third of its length flanking on the RPA surface. Transition from this state to the final bound state (H₆₂) may be accompanied by formation and dissolution of bulges (‘reptation dynamics’) as it is in RPA-(dT)₂₅. However, our analysis of bulges in *arm-2* indicates they are mostly static. Therefore, we hypothesize that alternative association mechanism of RPA might exist that can annotate how the long ‘short-lived’ ssDNA intermediates (19) are rapidly processed by RPA during DNA replication.

Association kinetics and diffusion efficiency of RPA vary with the length of interacting ssDNA molecule

In order to test our hypothesis, we first estimated the association kinetics of RPA to ssDNA intermediates of different length. The rate of association was measured at DBD-A and DBD-D of RPA from the time required to form at least 80% of the specific RPA-ssDNA contacts. The choice of the DBDs is to ensure complete wrap of RPA by ssDNA as it does in the final bound complex. Our results presented in Fig 6A exhibits association rates of ssDNA at DBD-A and DBD-D respectively as a function of the length of ssDNA intermediates. We find there exist two different rates of ssDNA association to DBD-A for (dT)₂₅, whereas a single for DBD-D. The observation is consistent with the experimental result based on interactions between surface tethered wild type RPA and Cy3-labeled (dT)₃₅ that has measured two equilibrium constants (19) corresponding to a fast ($K_a = 1.47 \pm 0.27 \times 10^9 \text{ M}^{-1}$) and a ten times slow binding ($K_a = 1.66 \pm 0.59 \times 10^{10} \text{ M}^{-1}$) mode of RPA to ssDNA. A recent study (9) has proceeded further to assess the individual contributions of different DBDs by site-specific labelling technique with MB543 and measured the association rate of (dT)₂₅ at DBD-D given by $k \sim 36.2 \pm 2.3 \text{ s}^{-1}$. In contrast, the association of ssDNA at DBD-A features a fast ($k \sim 30.6 \pm 9.8 \text{ s}^{-1}$) and a slow kinetics ($k \sim 10.3 \pm 9.8 \text{ s}^{-1}$). The measurement of these rates from our simulations shows $k_{fast-association}^{DBD-A} \sim 27.54 \pm 2.8 \text{ s}^{-1}$ and $k_{slow-association}^{DBD-A} \sim 12.9 \pm 0.6 \text{ s}^{-1}$, indicating an excellent match between simulations and experiment. However, our measured rate at DBD-D ($k_{association}^{DBD-D} \sim 108.43 \pm 19.7 \text{ s}^{-1}$) is ~ 3 times higher than the experimental value. This could be because of the presence of winged helix associated with DBD-D, which we have not considered in the present simulations. Moving further, our analysis demonstrates that $k_{association}^{DBD-D}$ is roughly independent of the length of ssDNA intermediate. In comparison, $k_{fast-association}^{DBD-A}$ increases linearly (up to 3 times) with the increasing length of ssDNA intermediates. $k_{slow-association}^{DBD-A}$ does not vary much with the length of ssDNA and even this binding mode does not exist for ssDNA length more than 40 nt. To explain the observations, we estimated the contact-time of specific interfacial RPA residues from the time they spent interacting (distance from nucleobase less than 6 Å) with at least one nucleotide. The analysis for both slow and fast binding events presented in Fig 6B clearly indicates a major role played by DBD-C and DBD-D of the trimerization core. The ssDNA spends significantly long time close to the interfacial residues of the trimerization core, resulting in a slow propagation of the ssDNA tract to DBD-A. This is counterintuitive as the DBDs of trimerization core has much lower affinity for ssDNA compared to DBD-B ($k_a \sim 5 \times 10^4 \text{ M}^{-1}$) and DBD-A ($k_a \sim 5 \times 10^5 \text{ M}^{-1}$). This is due to the high propensity of trimerization core to form more contacts with ssDNA that outcompetes the high affinity DBD-B and DBD-A. In some cases however, the conformational flexibility of ssDNA facilitates its rapid propagation without being stuck long at the trimerization core, resulting in a fast association kinetics at DBD-A. The impact of conformational flexibility is more prominent for longer ssDNA intermediates and it fully offsets the ability of trimerization core to retard the ssDNA association kinetics at DBD-A for ssDNA length greater than 40 nts. The slow binding

mode thus does not exist for longer ssDNA. The same conformational feature (high conformational entropy) of long ssDNA also supports its extended conformation in state- L_n (n denotes the length of ssDNA, see Fig S5). The transition from state- L_n to the final bound complex of RPA-long ssDNA requires a conformational transition of ssDNA and its association at DBD-B and DBD-A. We note that both the processes happens cooperatively (see Fig S8 and Fig S9 in Supplementary text). Longer the ssDNA, greater is the cooperativity (see Fig S10 where cooperativity η_H increases linearly with ssDNA length). The molecular driving force is the long ranged electrostatic interactions between ssDNA and DBD-A and DBD-B. With the increasing flanking length of the flexible ssDNA, its capture radius increases that promotes faster nonspecific association of ssDNA at DBD-B and DBD-A (see Fig 6A, where $k_{fast-association}^{DBD-A}$ increases with ssDNA length). After initial nonspecific association, short ranged aromatic, polar and hydrophobic interactions act concertedly to form the final bound complex. The mechanism is analogous to the 'fly-casting' technique, where disordered proteins because of their high capture radius can associate faster to their binding partners (27–29).

In addition to the association kinetics, we also investigated the diffusion efficiency of RPA on ssDNA in its bound state. Despite their high-affinity binding, RPA needs to be displaced or repositioned quickly for unhindered action of the downstream proteins involved in various metabolic pathways. This is common to many ssDNA binding proteins (30, 31). We measured the diffusion efficiency of RPA from its one-dimensional diffusion coefficient (D_1) at 291 K from the slope of mean-square-displacement (MSD) of nucleotide bases at DBD-A. Our estimated diffusion coefficient D_1 for $(dT)_{25}$ is ~ 0.00033 nt^2/MD time step. This corresponds to $D_1 \sim 1650 \pm 426$ $\text{nt}^2 \text{ s}^{-1}$ (see Fig S11 and Supplementary text for D_1 calculation) considering each MD step in our simulations of 2 ps. The value is comparable to experimentally reported D_1 is 5000 ± 400 $\text{nt}^2 \text{ s}^{-1}$ for hRPA on ssDNA at 310 K (4). We further note that D_1 increases linearly by 8 fold (D_1 for $(dT)_{62} \sim 12859 \pm 2727$ $\text{nt}^2 \text{ s}^{-1}$) as length of ssDNA changes from 25 nt to 62 nt, indicating a significant role of length of ssDNA intermediates in determining the fate of RPA-ssDNA bound complex during its post-processing. Our analysis suggests the molecular origin of faster diffusion of RPA on longer ssDNA is the formation of greater number of 'dynamic bulges' (see Fig S12 in Supplementary text) compared to that on shorter ssDNA due to the higher heterogeneity in intermolecular interactions in the former.

Discussion

In this paper, we have computationally explored the full binding energy landscape of RPA with ssDNA of different lengths using a state-of-the-art coarse-grained protein-DNA model. The applicability of the model has been tested extensively by employing it in capturing the binding of different ssDNA sequences to their respective protein partners. Our motivation is to understand how the same RPA molecule results in different outcomes in DNA metabolic pathways such as DNA repair, recombination, and replication, depending on its interactions with ssDNA intermediates of varying lengths. Our study reveals the underlying molecular mechanism that provides crucial new insights into how and when DNA repair/replication processes are orchestrated. For example, we observed that RPA binds very differently to short ($(dT)_{25}$) and long ($(dT)_{62}$) ssDNA tracts. The short ssDNA transiently associates with RPA and then through 'reptation dynamics', where dynamic bulges on ssDNA form and dissolve continuously, ssDNA tract slowly wraps the four DBDs (A-D) of RPA and forms the RPA-ssDNA bound complex as shown by a single basin in the respective free energy landscape. The formation and movement of bulges from 3' to 5' end are triggered by mainly electrostatic interactions and defects in polar interactions respectively. The overall binding is coherent and dynamic, where the flexibly interconnected DBDs of RPA interact individually with $(dT)_{25}$ through innumerable ways to stabilize the association of ssDNA finally into a horse-shoe shaped bent conformation on the RPA surface. The coherency further suggests that the ssDNA associates with the two extreme DBDs of RPA (A and D) roughly with comparable association kinetics, as was confirmed by the experimentally observed kinetic rates at these DBDs as well. While this is direct evidence that ssDNA binding to RPA is not sequential (ssDNA does not propagate along the sequence of DBDs in RPA), some association events were also reported that indicate ~ 3 times slower association

kinetics of ssDNA at DBD-A. Our analysis elucidates a major role played by the trimerization core of RPA in this slow kinetics. The high propensity of trimerization core to form the maximum number of contacts with (dT)₂₅ outcompetes the high-affinity DBD-B and DBD-A in halting the ssDNA longer close to the trimerization core, resulting into slower association kinetics.

In comparison, RPA association to longer ssDNA ((dT)₆₂) is quite different, although the initial association between (dT)₆₂ and RPA occurs through the same 'reptation dynamics'. The major difference between the free energy profiles of RPA-(dT)₂₅ and RPA-(dT)₆₂ is that the latter shows two basins, indicating the formation of final bound state via a stable intermediate state. The structural characterization of the intermediate state shows a linear conformation adopted by (dT)₆₂, where roughly half of (dT)₆₂ is bound to DBD-D and DBD-C but the remaining half at 5' end flanks on the RPA surface. Our analysis suggests the high conformational flexibility (entropy) favours the extended conformation of (dT)₆₂, the direct evidence of which comes from linear increment of the population of the stable intermediate state with the increasing length of ssDNA molecules. The high conformational entropy of longer ssDNA is also beneficial to offset the impact of trimerization core in retarding the association kinetics of ssDNA at DBD-A. In fact, the slow association kinetics of ssDNA at DBD-A diminishes to non-existence for ssDNA length longer than 40 nt. The transition from here to the final bound RPA-(dT)₆₂ complex is unique for longer ssDNA only. To this end our results reveal two interesting aspects. (i) Firstly, RPA-(dT)₂₅ complex is more stable ($\Delta G \sim 0.46 \pm 0.001$ kcal/mol) than RPA-(dT)₆₂ complex, which validates the experimental observation that RPA associates more strongly with short ssDNA intermediates during DNA repair compared to the long ssDNA tract in DNA replication (19). The molecular basis is the higher heterogeneity observed in the interactions between RPA and (dT)₆₂. We show the association of RPA to long ssDNA is stabilized by a weighted combination of electrostatic, aromatic, polar and hydrophobic interactions. Reduction of one component via point mutations of a particular type of amino acids can be easily compensated by the other interactions. The prediction is supported by the experimental observation that mutations of aromatic residues in RPA only minimally impact its affinity for ssDNA. In contrast, electrostatic and polar interactions define the free energy profile of RPA to short ssDNA ((dT)₂₅). This explains why mutations of the polar residues in RPA lower its affinity for short ssDNA (11, 16). (ii) Secondly, the intermediate state, exclusive to longer ssDNA, is only marginally less stable ($\Delta G \sim 0.7 \pm 0.001$ kcal/mol) compared to the RPA-(dT)₆₂ complex. The low energy barrier indicates a dynamic equilibrium between the states, unlike a single bound state for RPA with short ssDNA. During the transition from the intermediate state to the RPA-(dT)₆₂ bound complex, the ssDNA undergoes a conformational change and binds to DBD-B and DBD-A of RPA. We find the conformational switch in ssDNA is facilitated by a cooperative binding at DBD-B and DBD-A of RPA. The cooperativity originates from a 'fly-casting' mechanism, where the negatively charged phosphates on the flanking ssDNA segment senses the positively charged residues of RPA from far to trigger their association. With the increasing flanking length of ssDNA, its capture radius increases that promote faster association. The mechanism illustrates how short-lived, long ssDNA intermediates rapidly associate with RPA during DNA replication.

To summarize, our study unravels the mechanism of dynamic binding of RPA to ssDNA. It explains the molecular origin of variations in RPA action during various DNA processing events depending on the length of ssDNA intermediates. Furthermore, the advantage of longer ssDNA is observed in faster diffusion of RPA on the ssDNA tract as well, indicating an ssDNA length-dependent fate of RPA-ssDNA bound complexes during their post-processing. The study also lays the foundation to analyse the impacts of mutation in RPA on its function and investigate how other RPA interacting proteins (RIPs) such as RAD52 that has lower a affinity for ssDNA displaces RPA from its complex with ssDNA during DNA metabolism.

Materials and Methods

We developed a coarse-grained model to study ssDNA-protein interactions. The model resolution is similar to a previously developed model (32, 33) but varies significantly in terms of energetics.

We adopted a coarse-grained representation of protein, where each amino acid is represented by a single bead placed at the respective C_{α} position (34). The energetics of the protein is described by a structure-based potential (35) (for details, see the Supporting text), which represents the funnel-like energy landscape for protein folding (35) and has been used extensively in studying protein-protein (36) and protein-nucleic acid interactions (37–43).

The resolution of the ssDNA molecule is three beads per nucleotide (phosphate, sugar and nitrogenous base), placed at their respective geometric centers. The energetics of the ssDNA molecule follows that of 3SPN.2 model of DNA by Hinckley et. al. (44). The model has successfully captured the structural features of ssDNA including precise predict the persistence length of ssDNA in agreement to the experimental measurements.

We have considered two components to describe the ssDNA-protein interactions namely, (i) long-ranged and (ii) short-ranged interactions. The long-ranged interaction includes electrostatic interactions between negatively charged phosphate beads of ssDNA and charged amino acids (R, K, D, E) residues of protein. For pH is <7, we have considered His residue to be positively charged (see Table S11). We assigned a unit positive charge to Lys, Arg and His and a unit negative charge on Glu and Asp. The phosphate bead was assigned a negative charge of 0.6 to take counterion condensation into account. We modelled the electrostatic interactions by Debye-Hückel potential that considers salt effect. It is noteworthy that previously Debye-Hückel potential has been applied successfully to understand crucial aspects of protein-nucleic acid interactions (37–43, 45, 46) despite its limited applicability for dilute solutions only. The effective strength of protein-DNA interaction is scaled by a factor of 1.67 to bring the local charge of phosphate beads back to -1, as described in their previous works (47).

The short-ranged interactions include aromatic stacking interactions between aromatic amino acids (F, W, H, Y) and the nucleobases of ssDNA, hydrogen bond interactions between non-aromatic amino acids (both polar and hydrophobic) and ssDNA bases and repulsive excluded volume interactions between protein and ssDNA residues. The aromatic stacking interactions are modelled by a Lennard-Jones potential, where reweighted sequence-dependent pairwise interaction strength parameters ϵ_{ij}^{B-AA} were chosen (48)(see Table S3). The reweighting factor is 0.6 to fit into our CG model in order to achieve the correct binding modes for various protein-ssDNA systems. The hydrogen bond interaction, similar to aromatic stacking, is also modelled by the Lennard-Jones potential, with the interaction strength of $\epsilon_{ij}^{Hydrogen}$. The non-aromatic amino acid residues (both polar and hydrophobic) have a strong tendency to form direct hydrogen bonds with nucleotides (20). We analysed the number of direct hydrogen bonds between non-aromatic amino acids and nucleobases from a set of 51 protein-ssDNA complexes (see Table S4). The profile shows a large variation in the interaction pattern between different protein residues with four different nucleotide sequences. We tuned the interaction strength values of $\epsilon_{ij}^{Hydrogen}$ according to the hydrogen bond formation pattern analysed in Figure S1 and a complete list of sequence dependent pairwise base-non aromatic amino acid interaction strength values of $\epsilon_{ij}^{Hydrogen}$ provided in Table S5. These values successfully capture the correct binding modes of various protein-ssDNA complexes (see Figure S2) irrespective of amino acid sequences of protein and nucleotide sequences of ssDNA molecule. The variety in sequence and structure of the studied complex strongly supports the generality and appropriateness of our model to study binding dynamics of ssDNA with proteins. Finally, a purely repulsive interaction is applied to consider the excluded volume effect. For further details, see the supporting information. To this end, it is important to note that our coarse-grained potential energy functions for the structure and energetics of the protein-ssDNA binding are fully transferable and does not include any bias towards the final specific complex.

The unbiased binding simulations were performed using Langevin dynamics with a friction coefficient $\gamma = 0.1$ kg/s. The integration time step used in the simulation is 0.05, which corresponds to a real time scale of 2 ps (49) (see Supporting material for time step conversion). The implicit solvent effect is considered by using the dielectric constant of water (78). Simulations were performed at temperatures as mentioned in the corresponding protein-ssDNA PDB complex

(see Table S11), otherwise a temperature of 291K was used. The salt concentration is taken 10 mM to ensure strong protein-ssDNA interactions. To explore RPA-ssDNA binding energy surface, we performed 50 independent simulations for each ssDNA length ((dT)₂₅, (dT)₃₂, (dT)₄₀, (dT)₅₀, and (dT)₆₂). The simulations were 2 x 10⁸ MD steps long to ensure enough sampling of the conformational space.

Acknowledgments

We gratefully acknowledge the financial support from DST India (DST PURSE, DST/INSPIRE/04/2013/000100, DST SERB CRG/2019/001001), and DBT CoE research grant. A.M. acknowledges financial support from CSIR India in the form of a Senior Research Fellow.

References

1. M. S. Wold, Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66**, 61–92 (1997).
2. G. G. Oakley, Replication protein A: directing traffic at the intersection of replication and repair. *Front. Biosci.* **15**, 883 (2010).
3. R. Chen, M. S. Wold, Replication protein A: Single-stranded DNA's first responder. *BioEssays* **36**, 1156–1161 (2014).
4. B. Nguyen, *et al.*, Diffusion of Human Replication Protein A along Single-Stranded DNA. *J. Mol. Biol.* **426**, 3246–3261 (2014).
5. K. Treuner, U. Ramsperger, R. Knippers, Replication Protein A Induces the Unwinding of Long Double-stranded DNA Regions. *J. Mol. Biol.* **259**, 104–112 (1996).
6. S.-H. Bae, K.-H. Bae, J.-A. Kim, Y.-S. Seo, RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes. *Nature* **412**, 456–461 (2001).
7. O. D. Scharer, Nucleotide Excision Repair in Eukaryotes. *Cold Spring Harb. Perspect. Biol.* **5**, a012609–a012609 (2013).
8. L. S. Symington, End Resection at Double-Strand Breaks: Mechanism and Regulation. *Cold Spring Harb. Perspect. Biol.* **6**, a016436–a016436 (2014).
9. N. Pokhrel, *et al.*, Dynamics and selective remodeling of the DNA-binding domains of RPA. *Nat. Struct. Mol. Biol.* **26**, 129–136 (2019).
10. A. I. Arunkumar, M. E. Stauffer, E. Bochkareva, A. Bochkarev, W. J. Chazin, Independent and Coordinated Functions of Replication Protein A Tandem High Affinity Single-stranded DNA Binding Domains. *J. Biol. Chem.* **278**, 41077–41082 (2003).
11. I. M. Wyka, K. Dhar, S. K. Binz, M. S. Wold, Replication Protein A Interactions with DNA: Differential Binding of the Core Domains and Analysis of the DNA Interaction Surface †. *Biochemistry* **42**, 12909–12918 (2003).
12. E. Bochkareva, S. Korolev, S. P. Lees-Miller, A. Bochkarev, Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J.* **21**, 1855–1863 (2002).
13. A. Bochkarev, E. Bochkareva, L. Frappier, A. M. Edwards, The crystal structure of the complex of replication protein A subunits RPA32 and RPA14 reveals a mechanism for single-stranded DNA binding. *EMBO J.* **18**, 4498–4504 (1999).
14. L. A. Yates, *et al.*, A structural and dynamic model for the assembly of Replication Protein A on single-stranded DNA. *Nat. Commun.* **9**, 5447 (2018).
15. A. Bochkarev, R. A. Pfuetzner, A. M. Edwards, L. Frappier, Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* **385**, 176–181 (1997).
16. C. S. Hass, K. Lam, M. S. Wold, Repair-specific Functions of Replication Protein A. *J. Biol. Chem.* **287**, 3908–3918 (2012).
17. C. Kim, B. F. Paulus, M. S. Wold, Interactions of human replication protein A with oligonucleotides. *Biochemistry* **33**, 14197–14206 (1994).
18. C. A. Brosey, *et al.*, A new structural framework for integrating replication protein A into

- DNA processing machinery. *Nucleic Acids Res.* **41**, 2313–2327 (2013).
19. R. Chen, S. Subramanyam, A. H. Elcock, M. Spies, M. S. Wold, Dynamic binding of replication protein a is required for DNA repair. *Nucleic Acids Res.* **44**, 5758–5772 (2016).
 20. J. Fan, N. P. Pavletich, Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. *Genes Dev.* **26**, 2337–2347 (2012).
 21. P. C. Blainey, *et al.*, Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.* **16**, 1224–1229 (2009).
 22. M.-Y. Tsai, W. Zheng, M. Chen, P. G. Wolynes, Multiple Binding Configurations of Fis Protein Pairs on DNA: Facilitated Dissociation versus Cooperative Dissociation. *J. Am. Chem. Soc.* **141**, 18113–18126 (2019).
 23. C. Maffeo, A. Aksimentiev, Molecular mechanism of DNA association with single-stranded DNA binding protein. *Nucleic Acids Res.* **45**, 12125–12139 (2017).
 24. G. Mishra, L. S. Bigman, Y. Levy, ssDNA diffuses along replication protein A via a reptation mechanism. *Nucleic Acids Res.* **48**, 1701–1714 (2020).
 25. R. Zhou, *et al.*, SSB Functions as a Sliding Platform that Migrates on DNA via Reptation. *Cell* **146**, 222–232 (2011).
 26. G. B. Brandani, T. Niina, C. Tan, S. Takada, DNA sliding in nucleosomes via twist defect propagation revealed by molecular simulations. *Nucleic Acids Res.* **46**, 2788–2801 (2018).
 27. Y. Levy, J. N. Onuchic, P. G. Wolynes, Fly-casting in protein-DNA binding: Frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* **129**, 738–739 (2007).
 28. E. Trizac, Y. Levy, P. G. Wolynes, Capillarity theory for the fly-casting mechanism. *Proc. Natl. Acad. Sci.* **107**, 2746–2750 (2010).
 29. A. Bhattacharjee, S. Wallin, Coupled folding-binding in a hydrophobic/polar protein model: Impact of synergistic folding and disordered flanks. *Biophys. J.* (2012) <https://doi.org/10.1016/j.bpj.2011.12.008>.
 30. R. Roy, A. G. Kozlov, T. M. Lohman, T. Ha, SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. *Nature* **461**, 1092–1097 (2009).
 31. J. Zhang, R. Zhou, J. Inoue, T. Mikawa, T. Ha, Single molecule analysis of *Thermus thermophilus* SSB protein dynamics on single-stranded DNA. *Nucleic Acids Res.* **42**, 3821–3832 (2014).
 32. G. Mishra, Y. Levy, Molecular determinants of the interactions between proteins and ssDNA. *Proc. Natl. Acad. Sci.* **112**, 5033–5038 (2015).
 33. C. Maffeo, T. T. M. Ngo, T. Ha, A. Aksimentiev, A Coarse-Grained Model of Unstructured Single-Stranded DNA Derived from Atomistic Simulation and Single-Molecule Experiment. *J. Chem. Theory Comput.* **10**, 2891–2896 (2014).
 34. A. Bhattacharjee, D. Krepel, Y. Levy, Coarse-grained models for studying protein diffusion along DNA. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **6**, 515–531 (2016).
 35. C. Clementi, H. Nymeyer, J. N. Onuchic, Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–53 (2000).
 36. W. Zheng, N. P. Schafer, A. Davtyan, G. A. Papoian, P. G. Wolynes, Predictive energy landscapes for protein-protein association. *Proc. Natl. Acad. Sci.* **109**, 19244–19249 (2012).
 37. A. Bhattacharjee, Y. Levy, Search by proteins for their DNA target site: 1. The effect of DNA conformation on protein sliding. *Nucleic Acids Res.* **42**, 12404–12414 (2014).
 38. A. Bhattacharjee, Y. Levy, Search by proteins for their DNA target site: 2. The effect of DNA conformation on the dynamics of multidomain proteins. *Nucleic Acids Res.* **42**, 12415–12424 (2014).
 39. A. Mondal, A. Bhattacharjee, Searching target sites on DNA by proteins: Role of DNA dynamics under confinement. *Nucleic Acids Res.* **43**, 9176–9186 (2015).
 40. P. Dey, A. Bhattacharjee, Structural Basis of Enhanced Facilitated Diffusion of DNA-Binding Protein in Crowded Cellular Milieu. *Biophys. J.* **118**, 505–517 (2020).
 41. P. Dey, A. Bhattacharjee, Mechanism of Facilitated Diffusion of DNA Repair Proteins in Crowded Environment: Case Study with Human Uracil DNA Glycosylase. *J. Phys. Chem.*

- B* **123**, 10354–10364 (2019).
42. P. Dey, A. Bhattacharjee, Role of Macromolecular Crowding on the Intracellular Diffusion of DNA Binding Proteins. *Sci. Rep.* **8**, 844 (2018).
 43. P. Dey, A. Bhattacharjee, Disparity in anomalous diffusion of proteins searching for their target DNA sites in a crowded medium is controlled by the size, shape and mobility of macromolecular crowders. *Soft Matter* **15**, 1960–1969 (2019).
 44. D. M. Hinckley, G. S. Freeman, J. K. Whitmer, J. J. De Pablo, An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization. *J. Chem. Phys.* **139** (2013).
 45. M.-Y. Tsai, B. Zhang, W. Zheng, P. G. Wolynes, Molecular Mechanism of Facilitated Dissociation of Fis Protein from DNA. *J. Am. Chem. Soc.* **138**, 13497–13500 (2016).
 46. B. Zhang, W. Zheng, G. A. Papoian, P. G. Wolynes, Exploring the Free Energy Landscape of Nucleosomes. *J. Am. Chem. Soc.* **138**, 8126–8133 (2016).
 47. J. Lequieu, A. Córdoba, D. C. Schwartz, J. J. de Pablo, Tension-Dependent Free Energies of Nucleosome Unwrapping. *ACS Cent. Sci.* **2**, 660–666 (2016).
 48. L. R. Rutledge, L. S. Campbell-Verduyn, S. D. Wetmore, Characterization of the stacking interactions between DNA or RNA nucleobases and the aromatic amino acids. *Chem. Phys. Lett.* **444**, 167–175 (2007).
 49. T. Veitshans, D. Klimov, D. Thirumalai, Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold. Des.* **2**, 1–22 (1997).

Figures and Tables

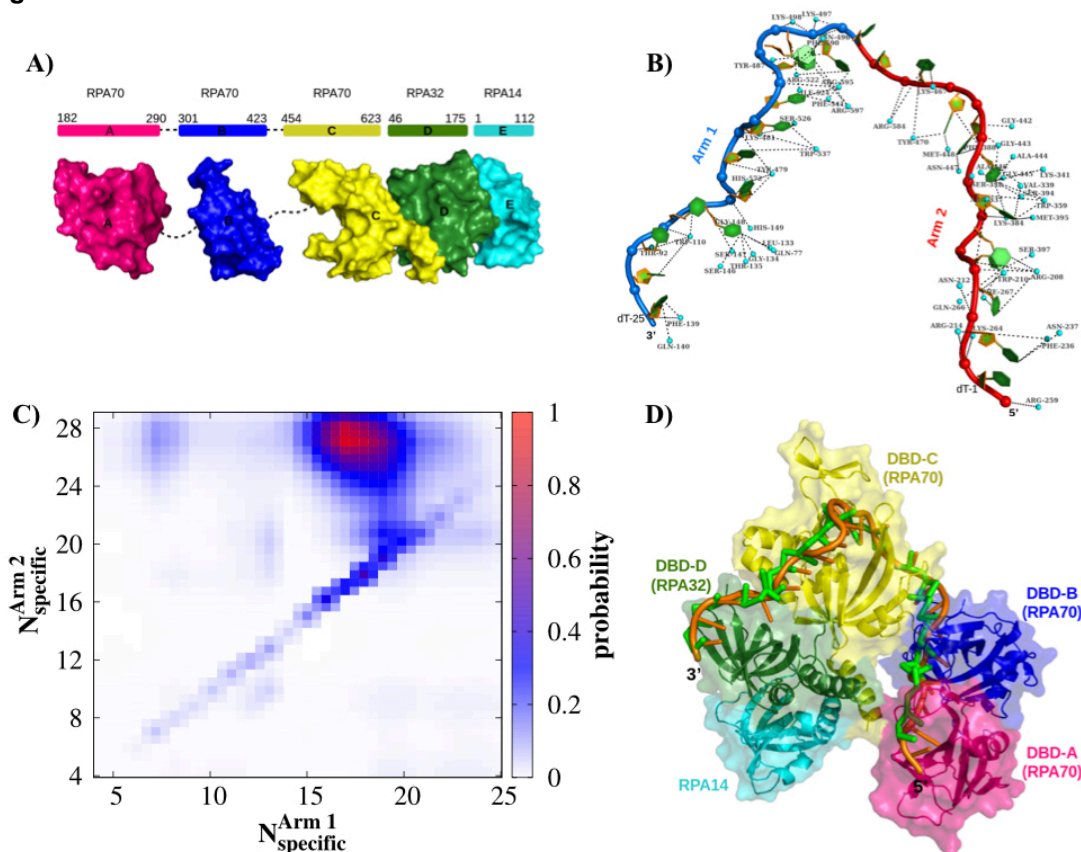


Figure 1. Structure and binding of RPA with ssDNA. (A) Schematic representation of the three RPA subunits RPA70, RPA32 and RPA14 with their residue numbers. RPA70 is composed of three DNA binding domains (DBD-A-C), whereas each RPA32 and RPA14 has DBD-D and DBD-E respectively. (B) The specific interfacial contacts of RPA residues (small cyan sphere) with ssDNA phosphate (blue sphere), sugar (orange colour) and base atoms (green ring). The ssDNA length is divided into approximately two equal halves and referred to as *arm-1* (13-25 nucleotides,

blue colour) and *arm-2* (1-12 nucleotides, red colour). (C) Arm-wise specific interfacial contact formation probabilities are presented. The red regime indicates the bound state. (D) Crystal structure of the DBDs of *U. maydis* RPA bound to (dT)₂₅ ssDNA (PDB ID 4GNX). The simulated ssDNA (green colour) is superimposed with the crystal one (orange color) after achieving the proper bound state conformation.

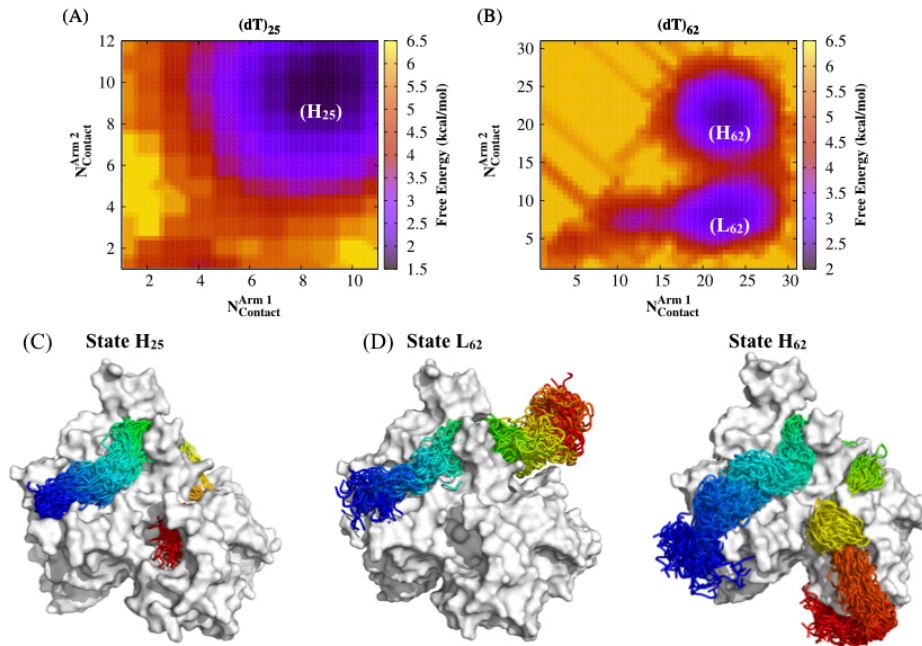


Figure 2. Length dependency in RPA-ssDNA binding. (A-B) Two dimensional free energy profiles for short ((dT)₂₅) and long ((dT)₆₂) length of ssDNA as a function of arm-wise contacts ($N_{Contact}^{Arm1}$ and $N_{Contact}^{Arm2}$) between nucleobases and the closest amino acid residues (within a distance cutoff of 6 Å). The basin H_{25} in (dT)₂₅ indicates horse-shoe shaped bound state, whereas the two minima L_{62} and H_{62} in (dT)₆₂ corresponds to a linear and a horse-shoe shaped conformation respectively. (C-D) Superposition of the cluster of snapshots corresponding to H_{25} , L_{62} and H_{62} states are presented. RPA is shown in grey surface and ssDNA in rainbow colour.

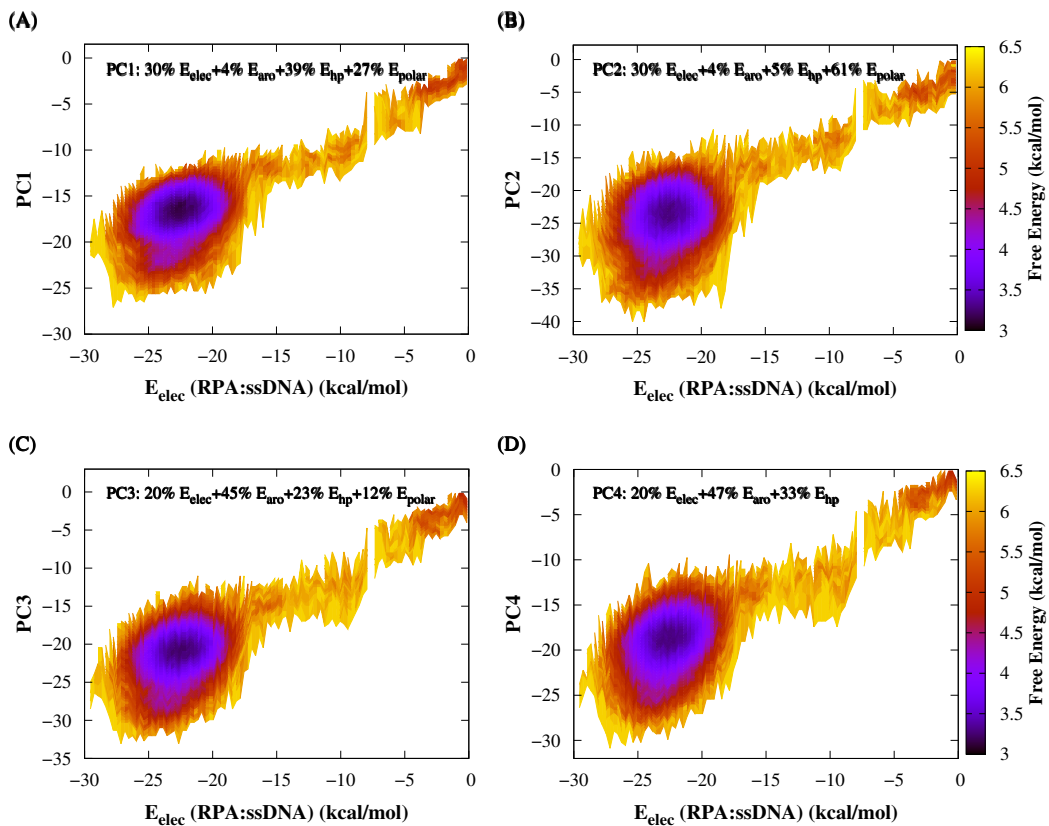


Figure 3. Two dimensional free energy surfaces for binding of RPA with (dT)₂₅ as a function of various principal components (PCs). Each panel (A-D) shows the free energy is projected on each individual PC (PC1 to PC4) along with a common coordinate E_{elec} (RPA:ssDNA). Each PC is a linear combination of the four collective basis variables in which the coefficients (number percentage) represents the fraction of these basis variables (see Table 1 for details).

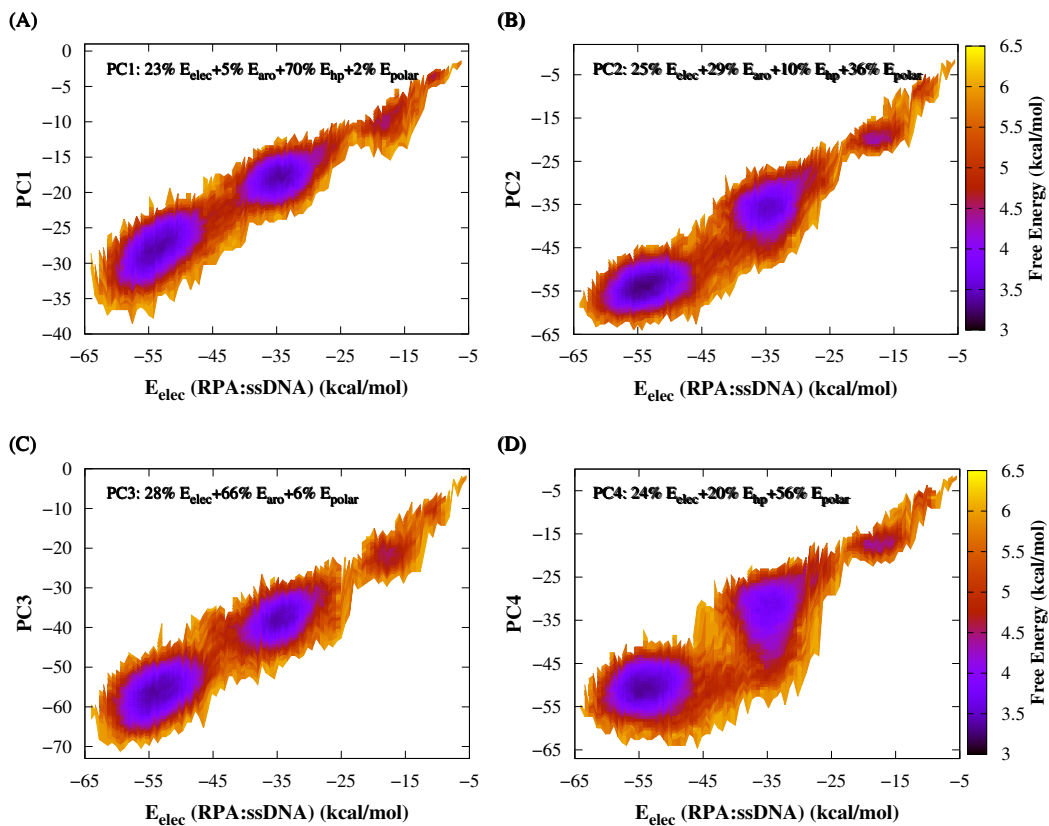


Figure 4. Two dimensional free energy surfaces for binding of RPA with $(dT)_{62}$ as a function of various principal components (PCs). Each panel (A-D) shows the binding free energy is projected on a particular PC (PC1 to PC4) along with a common coordinate $E_{elec}(RPA:ssDNA)$. The number percentage in each PC represents the fraction of the collective basis variables (see Table 2 for details).

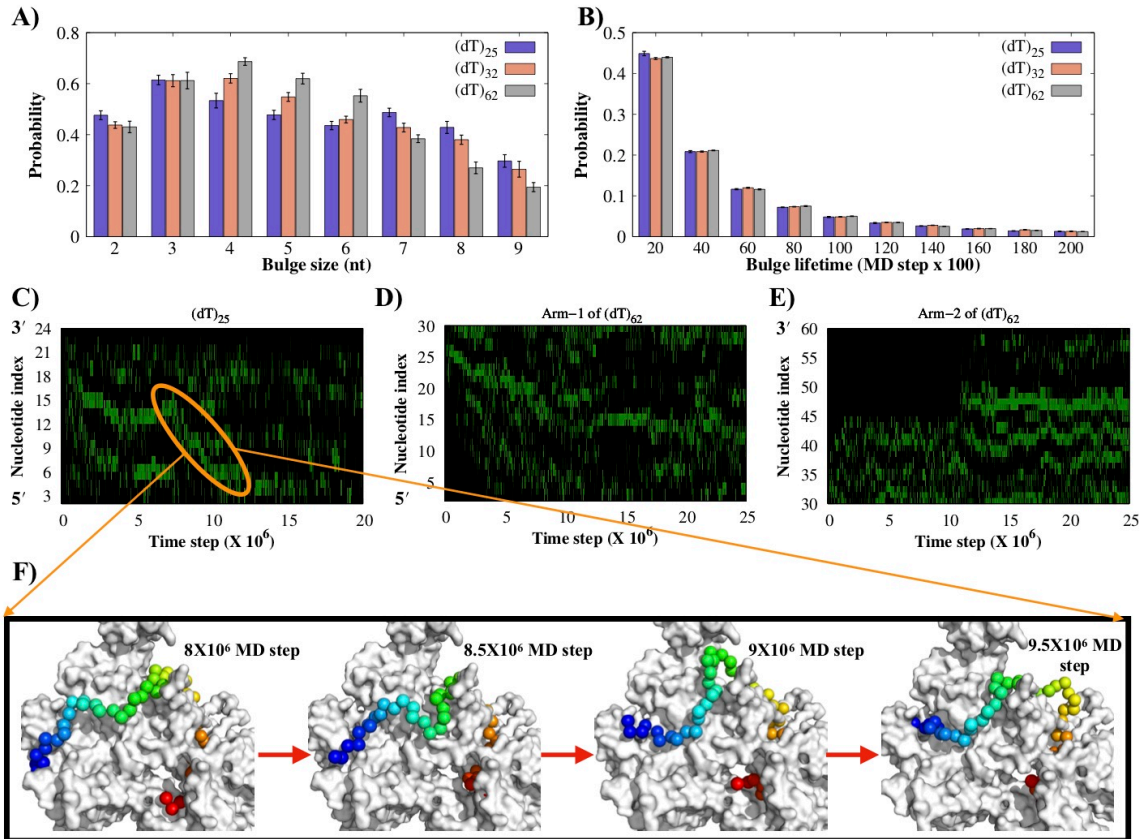


Figure 5. Characterization of ssDNA bulges on the RPA surface. (A-B) Probability distribution of different bulge sizes and their lifetime for (dT)₂₅, (dT)₃₂ and (dT)₆₂ length of ssDNA. Bulges were quantified by examining the distance between those phosphate atoms which was less than at least 60% to its Kuhn length in a stretched conformation. (C-E) Time evolution of the nucleotide bulge index which was positioned on the RPA surface are shown for (dT)₂₅ and *arm-1* and *arm-2* of (dT)₆₂. 3' and 5' indicates the direction of the bulge movement. (F) Four snapshots, which were sampled at different MD time steps (highlighted region), represent the displacement of a bulge on the RPA surface. RPA is shown in grey surface and ssDNA in rainbow colour.

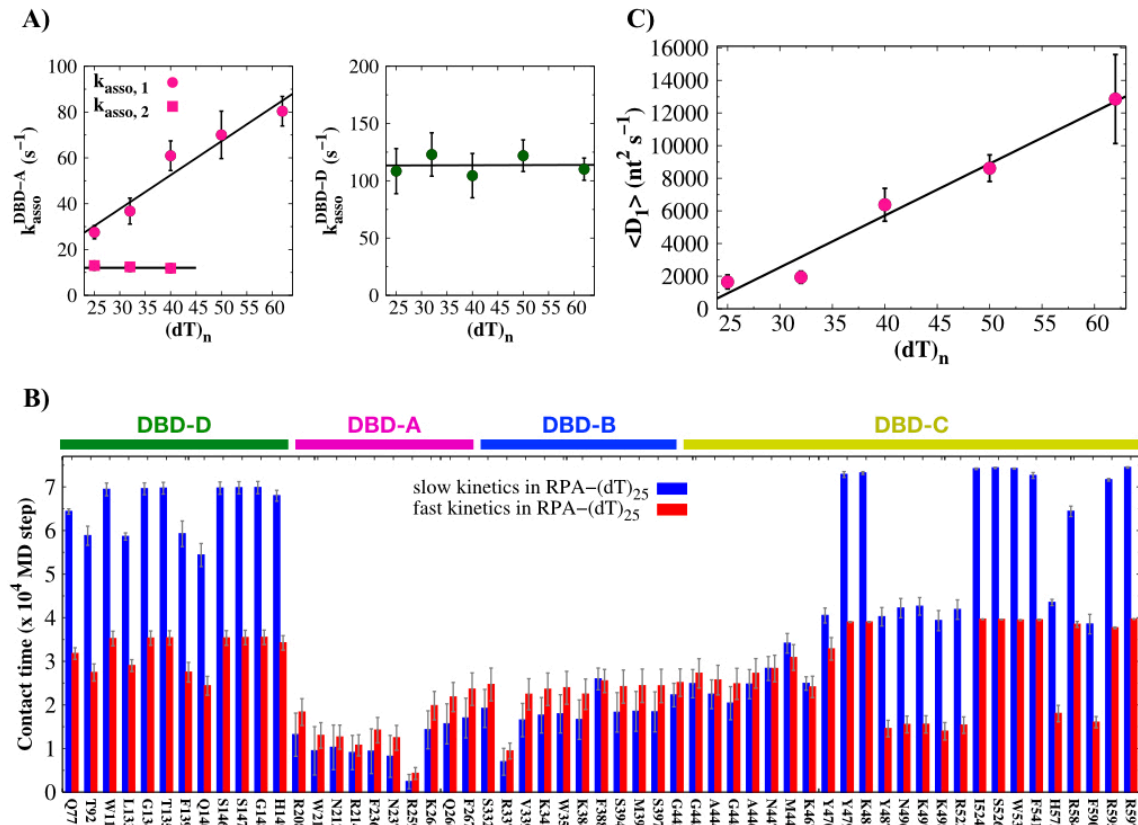


Figure 6. Kinetics of RPA binding with varying length of ssDNA. (A) The association rate of ssDNA to DBD-A and DBD-D of RPA with respect to ssDNA length $(dT)_n$, where 'n' denotes the length of the ssDNA. The rates for $k_{\text{association}}^{\text{DBD-A}}$ are best fit to two association kinetics, whereas the rate for $k_{\text{association}}^{\text{DBD-D}}$ fit to one association kinetics. (B) The contact time of nucleotide bases in $(dT)_{25}$ with interfacial RPA residues of different DBDs for slow and fast binding events. The range of each individual DBD is mentioned by a thick line at the top. (C) Variation in the 1D diffusion coefficient D_1 at DBD-A of RPA as a function of $(dT)_n$.

Table 1. Coefficients for Principal Components (PCs) for (dT)₂₅ binding mode

	E_{elec}	E_{aro}	E_{hp}	E_{polar}
PC1 (43.5%)	0.295	0.037	0.394	0.273
PC2 (38.9%)	0.304	0.036	0.053	0.607
PC3 (12.9%)	0.204	0.456	0.221	0.120
PC4 (4.7%)	0.198	0.470	0.332	0

Table 2. Coefficients for Principal Components (PCs) for (dT)₆₂ binding mode

	E_{elec}	E_{aro}	E_{hp}	E_{polar}
PC1 (42.6%)	0.229	0.048	0.70	0.022
PC2 (28.3%)	0.255	0.282	0.103	0.360
PC3 (25.1%)	0.277	0.658	0.001	0.064
PC4 (4.0%)	0.240	0	0.194	0.555