Published online 10 January 2020

Nucleic Acids Research, 2020, Vol. 48, No. 4 ***1701–1714***
doi: 10.1093/nar/gkz1202

ssDNA diffuses along replication protein A via a reptation mechanism

Garima Mishra[1,2,*], Lavi S. Bigman[1] and Yaakov Levy[1,*]

[1]Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel and [2]Department of Physics, Indian Institute of Technology Kanpur, Kanpur, 208016, India

Received July 15, 2019; Revised November 23, 2019; Editorial Decision December 14, 2019; Accepted January 08, 2020

ABSTRACT

Replication protein A (RPA) plays a critical role in all eukaryotic DNA processing involving single-stranded DNA (ssDNA). Contrary to the notion that RPA provides solely inert protection to transiently formed ssDNA, the RPA–ssDNA complex acts as a dynamic DNA processing unit. Here, we studied the diffusion of RPA along 60 nt ssDNA using a coarse-grained model in which the ssDNA–RPA interface was modeled by both aromatic and electrostatic interactions. Our study provides direct evidence of bulge formation during the diffusion of ssDNA along RPA. Bulges can form at a few sites along the interface and store 1–7 nt of ssDNA whose release, upon bulge dissolution, leads to propagation of ssDNA diffusion. These findings thus support the reptation mechanism, which involves bulge formation linked to the aromatic interactions, whose short range nature reduces cooperativity in ssDNA diffusion. Greater cooperativity and a larger diffusion coefficient for ssDNA diffusion along RPA are observed for RPA variants with weaker aromatic interactions and for interfaces homogenously stabilized by electrostatic interactions. ssDNA propagation in the latter instance is characterized by lower probabilities of bulge formation; thus, it may fit the sliding-without-bulge model better than the reptation model. Thus, the reptation mechanism allows ssDNA mobility despite the extensive and high affinity interface of RPA with ssDNA. The short-range aromatic interactions support bulge formation while the long-range electrostatic interactions support the release of the stored excess ssDNA in the bulge and thus the overall diffusion.

INTRODUCTION

DNA processing in the cell is regulated by proteins that interact with DNA, in either its double-stranded (ds) or single-stranded (ss) forms. Proteins that interact with dsDNA have different biophysical and biochemical characteristics compared with those that interact with ssDNA, nevertheless, both families show the ability to diffuse linearly along DNA. While the biophysical properties of protein diffusion along dsDNA are quite well understood (1–4), the corresponding understanding for diffusion along ssDNA is limited.

Single-stranded DNA binding (SSB) proteins are the first responders to the transiently formed ssDNA, which is an intermediate during DNA replication, recombination, and repair processes (5). SSB proteins bind to ssDNA with high affinity to form very stable SSB–ssDNA complexes whose primary purpose is to secure the information stored in ssDNA. Subsequently, SSBs recycle (i.e. they dissociate from and re-associate with ssDNA) and reposition themselves within the complexes they form with ssDNA to facilitate metabolic processes. Experimental studies shed some light on the dynamic activity of different types of SSBs (6–8), and link between the dynamics of SSB–ssDNA complexes and biological function. In addition to diffusion of SSBs along ssDNA, their dynamic activities also include their ability to undergo direct transfer (9) which has been shown to be important in DNA replication both *in vitro* and *in vivo* (10).

The experimentally measured one-dimensional (1D) diffusion coefficient (D_1) of SSBs along ssDNA (6,8) is ∼3–4 orders of magnitude lower than that of proteins diffusing along dsDNA (11–15). The large differences between their 1D diffusion coefficients may arise from the different functionalities of these two classes of protein. dsDNA binding proteins must locate and bind a specific site, which requires a rapid search and fast diffusion. By contrast, protecting the ssDNA does not require that SSBs bind a specific site, however nor should they hinder the functions of other proteins that need to bind the SSBs—consequently SSBs are required to stably bind the ssDNA without remaining static at any single site. The binding affinities of SSB proteins are much higher than those of the dsDNA binding proteins in their non-specific DNA binding modes. Several features contribute to the complexity of diffusion of ssDNA. The greater flexibility of ssDNA, the extensive interface it forms

*To whom correspondence should be addressed. Tel: +972 8 9344587; Email: koby.levy@weizmann.ac.il
Correspondence may also be addressed to Garima Mishra. Email: garima.phy@gmail.com

with SSBs, and especially the high affinity of the complex have led to its biophysics being less well understood than that of dsDNA.

Escherichia coli SSB, a homotetrameric protein, wraps fully around ssDNA (16) and forms a compact structure with an extensive interface ($(SSB)_{65}$ at high salt concentrations) (17,18). Several mechanisms have been proposed to explain the migration of ssDNA along *E. coli* SSB, among them rolling and sliding dynamics (9,19). In the rolling mechanism, a segment at one end of the ssDNA is partially released by an SSB tetramer unit that immediately wraps around another portion of the ssDNA strand in its place (6). However, a single molecule study of the diffusion of SSB along a stretch of ssDNA ruled out the possibility of rolling being the dominant progression mechanism (7,20,21). In the sliding mechanism, breaking of all the interfacial contacts may result in a net motion of the entire ssDNA as it moves together along the SSB interface. Alternatively, a reptation mechanism was proposed. This mechanism is similar to the sliding mechanism with the exception that it does not require a concerted migration of the entire ssDNA but only of smaller stretches, which are stored in bulges. These bulges might be viewed as defects at the interface that, because of thermal fluctuations, accumulate excess ssDNA following breakage of the interfacial contacts at these sites. Following the reptation mechanism (also called 'sliding-with-bulge'), the linear diffusion of the ssDNA is achieved by local formation of transient bulges that assist in fragmenting the large interface and enable sliding in a stepwise manner as the ssDNA progressively moves through the SSB interface. Force-dependent ssDNA diffusion along SSB provides experimental support for the reptation model (20).

Escherichia coli SSB diffuses along short ssDNA with a 1D diffusion coefficient of 270 nt^2/s at 37°C, however, a 600-fold increase in the SSB diffusion coefficient is observed on long ssDNA (22). The large differences in diffusion coefficient may suggest that SSBs adopt different diffusion mechanisms under different circumstances. Unlike the sliding and reptation mechanisms that are utilized to diffuse along a short stretch of ssDNA, it has been hypothesized that SSB movement would occur through intersegment transfer on a single long ssDNA strand (22).

Recently, the diffusion of another SSB protein, the heterotrimeric replication protein A (RPA) (23,24), was probed and very fast diffusion along a short stretch of ssDNA was observed (8). RPA and *E. coli* SSB are very different structurally as the former is a heterotrimer and the latter is a homotetramer, yet they both contain several oligonucleotide/oligosaccharide binding domains. RPA is composed of three tightly associated subunits of ∼70, 32 and 14 kDa. These subunits include six oligonucleotide/oligosaccharide binding domains. Four of these binding domains interact with ssDNA (dT25), which adopts an extended conformation (Figure 1A). The number of nucleotides involved in the RPA–ssDNA interface is known for *U. maydis* RPA with ssDNA (PDB 4GNX), whose crystal structure contains 25 nucleotides (25). The size of the interface ssDNA forms with RPA was estimated from ensemble fluorescence experiments for human and S. *cerevisiae* RPA. Similarly to *E. coli* SSB, these two RPA bind to ssDNA in multiple DNA binding modes which is influenced by salt concentration. The number of nucleotides interacting with human [8] and *S. cerevisiae* RPA [26] are $\sim 20 \pm 2$ nt for [NaCl] < 50 mM and $\sim 28 \pm 2$ nt for [NaCl] > 1 M. The extensive RPA–ssDNA interface defines a high affinity complex with a K_D in the sub-nM range (8). The estimated 1D diffusion coefficient of RPA on ssDNA is $D_1 \sim 5000$ nt^2/s, which is about 10 times faster than for ssDNA diffusion on *E. coli* SSB. The high mobility of ssDNA notwithstanding maintenance of its interface with RPA might be linked to its ability to exchange, while bound to RPA, with a free RPA (23,26,27).

In the current study, a coarse-grained model for protein–ssDNA interaction (28) is applied to investigate the molecular mechanism for the linear diffusion of ssDNA along RPA and to explore how the rearrangement of the interface during the rapid diffusion of the ssDNA is consistent with the high affinity of the complex. While ssDNA exhibits high mobility when bound to two structurally different SSB proteins, *E. coli* SSB and RPA, it is unclear whether they share a similar mechanism of ssDNA diffusion. Furthermore, no direct evidence has been provided yet for the reptation mechanism. Bulge formation has not been observed experimentally, and although a bulge was detected computationally, it was static (29). Here, we provide a comprehensive analysis of the driving forces of the reptation diffusion of ssDNA along RPA via bulges.

MATERIALS AND METHODS

The diffusion of ssDNA along RPA was studied using a coarse-grained model based on a model that was developed to predict the interactions between proteins and ssDNA (28,30). The model used here is identical to the model presented in (28) with the exception that electrostatic forces are also included intramolecularly within the ssDNA. This model refers to Model I. Five additional models were studied here, as described below, to decipher the molecular determinants that govern the diffusion of ssDNA along RPA. In the following, we summarize all the models used in this study.

Coarse-grained model for RPA

We used a coarse-grained model to explore the binding of ssDNA to RPA at the molecular level. The model represents each protein residue by two beads placed at the Cα and Cβ positions. Beads representing charged amino acids (Lys, Arg, Asp, Glu and His) are charged in the model by placing the charge at the Cβ position of the residue. We have selected only His residues whose predicted pK_a is <6 to be positively charged. The protein is simulated by a native topology-based model that include nonspecific electrostatic (i.e., nonnative) interactions and uses the Lennard–Jones (L–J) potential to represent native contact interactions. Overall, we followed an approach to protein modeling similar to that described in previous studies (1,31) and will be briefly described here. The internal energy of the protein is designated by $E_{\mathrm{Prot}}(\Gamma, \Gamma_0)$,

$$E_{\mathrm{Prot}}(\Gamma,\Gamma_0) = E_{\mathrm{Prot}}^{\mathrm{Bond}} + E_{\mathrm{Prot}}^{\mathrm{Angle}} + E_{\mathrm{Prot}}^{\mathrm{Dihedral}} + E_{\mathrm{Prot}}^{\mathrm{Native-}\mathit{contacts}} + E_{\mathrm{Prot}}^{\mathrm{Electrostatics}} + E_{\mathrm{Prot}}^{\mathit{Repulsion}}$$

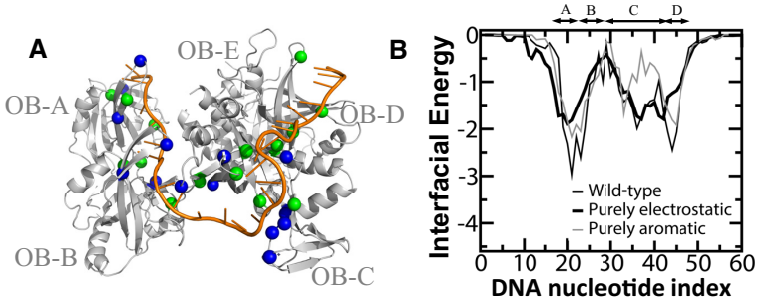

Figure 1. Structure and energetics of RPA–ssDNA interactions. (**A**) Distribution of the 15 aromatic residues (green spheres) and 14 positively charged residues (blue spheres) at the interface of ssDNA with *U. maydis* RPA (PDB 4GNX). The five OB domains in this crystal structure are indicated by A-E. The ssDNA interacts with OB domains A–D. The surface of RPA includes many more aromatic and charged residues (including negatively charged residues that are not shown in this structure). (**B**) Distribution of the interfacial energy along the ssDNA nucleotide from the coarse-grained model for wild-type RPA (thin line) and for chemically homogeneous mutants of RPA in which all the interfacial residues are positively charged (thick line) or aromatic (gray line). The OB domains (A–D) with which the ssDNA sequence interact with are indicated in the top.

where Γ denotes a particular conformation and Γ_0 denotes the native conformation along the MD simulation trajectory. The electrostatic interactions were modeled by the Debye–Hückel potential and we followed the parameters used in previous studies (1,32). We point out that the simulations were ran at relatively low temperature at which the protein is folded and fluctuates around the native state. In addition to the inherent flexibility of the proteins in the coarse-grained model, which is dictated by the density of the native contacts, we incorporated enhanced flexibility for regions characterized by high B-factors (i.e. higher than the mean B-factor) or residues with no electron density.

Coarse-grained model for ssDNA

The ssDNA was modeled by three beads per nucleotide (representing the phosphate (P), sugar (S) and base (B)) that were positioned at the geometric center of each represented group. The phosphate bead in the model bears a negative charge. The model potential for ssDNA used in our study, which follows other models (33–35) is given by:

$$E_{\mathrm{ssDNA}} = E^{\mathrm{Bond}}_{\mathrm{ssDNA}} + E^{\mathrm{Angle}}_{\mathrm{ssDNA}} + E^{\mathrm{Dihedral}}_{\mathrm{ssDNA}} + E^{Base-pairing}_{\mathrm{ssDNA}} + \mathrm{E}^{\mathrm{Stacking}}_{\mathrm{ssDNA}} + \mathrm{E}^{\mathrm{Repulsion}}_{\mathrm{ssDNA}}$$

The first three terms dictate the flexibility of the ssDNA backbone while the last three terms govern the global structure of the ssDNA. The first three terms have identical forms to the corresponding terms in E_{Prot}. The term $E^{\mathrm{Bond}}_{\mathrm{ssDNA}}$ represents the contribution from the covalently linked [(P-S)$_n$] backbone and the base, which is covalently linked to the sugar. The terms $E^{\mathrm{Angle}}_{\mathrm{ssDNA}}$ is the potential for bond angles which is applied between all the following three neighboring beads: (P_i–S_i–B_i), (B_i–S_i–P_{i+1}), (P_i–S_i–P_{i+1}), (S_i–P_{i+1}–S_{i+1}). $E^{\mathrm{Dihedral}}_{\mathrm{ssDNA}}$ is the potential of the dihedral angles, formed by the four beads B_i, S*i*, S_{i+1} and B_{i+1}. The values of the native bond lengths and angles were obtained from the PDB representation of the helical structure that ssDNA adopts in the duplex form.

Two major interactions involved in the stability of ssDNA are base pairing and stacking interactions. Usually, a homopolymeric ssDNA (poly T or poly C) is used in experimental studies of SSB–ssDNA complexes, which prohibits the formation of base-paring ($E^{\mathrm{Base-pairing}}_{\mathrm{ssDNA}} = 0$), as it requires complementary nucleotides in ssDNA. However, the formation of stacks still remains possible, depending on the type of nucleotides comprising the homopolymeric ssDNA. Therefore, to consider the effect of base stacking in ssDNA, we added a short-range attraction between consecutive bases of ssDNA in the form of the L-J potential. Similarly to the protein, a repulsion term (i.e. excluded volume) is applied between all non-bonded beads in ssDNA. As in the original model (28), the persistence length is governed in the model mostly by the dihedral angle between four consecutive phosphate beads. In this study, the strength of this angle is 0.7 which yields a persistence length of 32 Å, in agreement with other computational models (33,35,36) and experimental estimates of the persistence length of ssDNA (37–39). To examine the effect of the ssDNA persistence length, we changed the strength of this angle to either 0.2 or 2.0, which corresponds respectively to persistence length of 26 and 43 Å.

Coarse-grained model for ssDNA–RPA interactions

The interaction potential between a protein amino acid (AA) and ssDNA nucleotides arises from three contributions. (i) The electrostatic interaction between the Cβ bead representing a charged AA (K, R, H, D, E) and the negatively charged phosphate (P) of ssDNA. (ii) The aromatic stacking interaction between the Cβ bead representing an aromatic AA (W, F, Y, H) and an ssDNA base (B). (iii) The repulsive interactions between other beads of the protein and ssDNA. Thus: $E_{\mathrm{prot-ssDNA}} = E^{\mathrm{Elec}}_{\mathrm{prot-ssDNA}} + E^{\mathrm{Aromatic}}_{\mathrm{prot-ssDNA}} + E^{\mathrm{Repulsion}}_{\mathrm{prot-ssDNA}}$. The repulsion is applied between all beads of the protein with all beads of the ssDNA when setting the radii of DNA beads to 2 Å. The electrostatic interactions acting between all the charged beads in the system are modeled by the Debye–Hückel potential (1). These interactions are nonspecific and the phosphate groups of the ssDNA can interact with any charged residue of the RPA. The aromatic interaction, similar to base stacking, is modeled by the L–J potential, with a base–aromatic amino acid interaction strength of $\varepsilon_{\mathrm{B-AA}}$. The parameter $\varepsilon_{\mathrm{B-AA}} = 0$ for all the non-aromatic AAs and $\varepsilon_{\mathrm{B-AA}}$

> 0 only for aromatic amino acids (W, F, Y, H). Another important point is that W has a large surface area compared with other aromatic residues (F, Y and H) and hence $\varepsilon_{T\text{-}W}$ may make a larger contribution than $\varepsilon_{T\text{-}F}$, $\varepsilon_{T\text{-}Y}$ and $\varepsilon_{T\text{-}H}$ when the aromatic AAs form stacks with the bases of ssDNA. Given the coarse-grained nature of the model, we calibrated the relative strengths of the electrostatic and aromatic contributions and selected the following values: $\varepsilon_{T\text{-}F} = \varepsilon_{T\text{-}Y} = \varepsilon_{T\text{-}H} = 2$ and $\varepsilon_{T\text{-}W} = 3$. These values, which are consistent with other studies that quantified the strength of these interactions (40,41), successfully captured the structures of various protein–ssDNA complexes (see ref. (28)). We note that in the current study, the ssDNA was modeled as poly-T so no further parameters were needed.

In the simulations, the initial structure of ssDNA and RPA corresponded to the crystal structure of RPA–ssDNA (pdb ID 4GNX). This crystal structure includes OB folds A, B, C, D and E but lacks domain F, which is therefore not included in our simulations. Using the total potential energy ($E_{prot} + E_{ssDNA} + E_{prot-ssDNA}$) of the RPA–ssDNA complex, the dynamics of the protein and ssDNA was simulated using Langevin dynamics.

To study the diffusion of ssDNA along RPA, we used Pymol to extend the ssDNA from both its 3′ and 5′ ends in order to allow continuous diffusion of ssDNA along the RPA interface. For the Pymol extension, we considered a double stranded (ds) DNA of 20 base-pair length. We removed one of the dsDNA strands and, utilizing it as a newly generated stretch of 20 nt polyT, applied it at each end of the existing 25 bp ssDNA 4GNX structure while maintaining its 3′ or 5′ orientation. In this way, a 60 nt stretch (18+25+17 nt) of ssDNA was formed, which is long enough to allow continuous diffusion of ssDNA along the RPA interface. To verify the behavior of the newly generated 60 nt polyT, we simulated it in the absence of RPA and calculated its persistence length as 32 Å. This matches well with the persistence length of ssDNA reported in earlier studies. Furthermore, the designed RPA–ssDNA structure with the extended DNA was simulated for a short time of 10^5 time-steps to allow it to reach equilibrium. During this timeframe, the conformation of these flanking ssDNA stretches was changed based on the force-field of the ssDNA. Given the high flexibility of the flanking ssDNA, this short equilibration simulation is sufficient to allow the structure to adopt representative conformations, which are governed by E_{ssDNA} (see Supporting Information). The flanking ssDNA regions are unlikely to interact with the RPA until after diffusion takes place because of the electrostatic repulsive forces acting between the ssDNA nucleotides.

We performed the simulations in solution where the dielectric constant was that of water and using a salt concentration of 0.01 M. The simplicity of Debye–Hückle representation of charge-charge interactions does not capture electrostatic effects due to cations binding to ssDNA or cation and anion binding to RPA. Simulating RPA-ssDNA using the coarse-grained model with [NaCl] = 0.01 M guarantees an extensive ssDNA–RPA interface with 25 nucleotides interacting with charged or aromatic residues, as was concluded experimentally at [NaCl] ∼ 0.5 M. Accordingly, the used salt concentration in the simulations represents an effectively higher salt concentration. The simulations were performed at a temperature of 0.40 (arbitrary units), at which the protein does not undergo unfolding. To estimate the temperature of the coarse-grained simulations, fluctuations of internal distances in RPA from atomistic simulations at several temperatures were compared to those from the coarse-grained model indicating that temperature of 0.40 corresponds to 310 K (Supplementary Figure S1). The presented results for each model were obtained by averaging 50 simulations of 10^7 time steps each. The mean square displacements (MSD) were calculated based on the time evolution of the ssDNA indices relative to reference points defined along the ssDNA–RPA interface. Additional details of the MSD calculations can be found in previous publications (1,4).

Studied models of RPA–ssDNA interactions

To decipher the diffusion mechanism of ssDNA along RPA, it was studied for five additional models in which the RPA–ssDNA interface was manipulated. Models I–III were designed by reducing the strength of the aromatic interactions. Model I refers to the original model as described above. In models II and III, a single value was used for the aromatic interactions (i.e. without discriminating between the different aromatic side-chains $\varepsilon_{T\text{-}F} = \varepsilon_{T\text{-}Y} = \varepsilon_{T\text{-}H} = \varepsilon_{T\text{-}W}$) and its value was 2.0 or 1.5. These parameters produced a gradual decrease in the overall stability of the RPA–ssDNA interface. A distance criterion of 9 Å was applied between any ssDNA phosphate or any base groups in the crystal structure and the Cβ bead of the protein residues (being either positively charged or aromatic amino acids). By defining all residues within this distance criterion as interfacial, we identified 15 aromatic and 14 positively charge residues at the binding interface of RPA with ssDNA (Figure 1A). For RPA with only a positively charged interface (model IV), we changed all the 15 identified aromatic interfacial residues to positively charged residues by placing a (+1) charge at the Cβ-bead of these residues while eliminating aromatic interactions for these residues. Similarly, for an RPA interface made up of aromatic residues only, we neutralized the charge on the 14 positively charged residues from the RPA binding interface and considered them as aromatic residues with an interaction strength of 2. We identified the nucleotide interacting with an RPA residue in any given snapshot by monitoring the nucleotide for which the interaction energy was lowest. Model V, which represents the RPA–ssDNA interface as purely electrostatic but having short-range characteristics, was designed by using a modified form of the Debye–Hückel potential. When designing this model, it was important to shorten the long-range interactions while maintaining identical overall interface stability. This goal cannot be achieved by simply increasing salt concentrations, so we also changed the dielectric constants. We employed two different dielectric constants: one that controls the Debye screening length (i.e. controls the range of interaction) and another that controls the amplitude of interaction. The parameters of the Debye-Hückel for the long range potential are $\varepsilon_1 = \varepsilon_2 = 70$ and salt concentration of 0.01 M. The shorter-range Debye–Hückel was achieved by $\varepsilon_1 = 70$, $\varepsilon_2 = 25$, and salt concentration of 0.06 M. Model VI, in which the RPA–ssDNA interface is gov-

erned solely by electrostatic interactions, differs from Models IV and V with respect to the identity of the negatively charged beads on the ssDNA. Whereas, in models IV and V, the negative charges were placed on the phosphate groups, in model VI they were placed alternately on a phosphate and then on a base. This was achieved by neutralizing the charge on every other phosphate group and placing the negative charge on the base group of the same nucleotide.

To better understand the molecular factors that govern the diffusion of ssDNA along RPA, some mutations were introduced at the ssDNA–RPA interface. The diffusion coefficient of ssDNA diffusion along these mutants was measured as well as for variants with different flexibility of the ssDNA and of RPA. More methodological details can be found in the supporting information.

RESULTS AND DISCUSSION

Stability of the ssDNA–RPA interface

ssDNA often adopts an extended conformation and therefore the interface in many ssDNA–SSB complexes is elongated and narrow. However, the high plasticity of the ssDNA means that the interface can be very curved and non-trivial to predict. Furthermore, these quasi 1D interfaces are chemically quite heterogeneous, as they are stabilized by aromatic and electrostatic interactions. Aromatic interactions can form between the bases of ssDNA and the aromatic side-chains of the SSB, whereas electrostatic interactions can form between positively charged residues (Lys, Arg and His) and the ssDNA phosphate groups. The stability of the ssDNA–SSB interface may be governed by the different strengths of the electrostatic and aromatic interactions and by their spatial organization, because the former have a much longer range than the latter.

Figure 1A shows the 29 residues of RPA that interact directly with ssDNA. Of these, 14 are positively charged residues (Lys, Arg or His) and 15 are aromatic residues (Trp, Phe, Tyr or His). The distribution of aromatic and electrostatic residues describes a heterogeneous interface. To quantify the heterogeneity of the interface of the RPA–ssDNA complex and study its energetics, we applied a coarse-grained model that was shown to successfully predict the complexes formed between various SSBs and ssDNA (28). Figure 1B shows the interaction energy between each nucleotide and the interfacial residues, illustrating that the energy landscape is indeed heterogeneous because of the different chemical environments of each of the ssDNA nucleotides.

To examine whether the energetic heterogeneity originates from the geometry of the interface or from the identity of the interfacial residues, we estimated two chemically homogenous variants of RPA in which the interface was either purely electrostatic or purely aromatic (formed by mutating the interfacial positively charged residues to aromatic, or vice versa). The RPA–ssDNA interface was marginally stable and deviated from the X-ray conformation when it included only aromatic interactions (even though the energetic strengths of electrostatic and aromatic pairwise interactions are comparable). The energetic profile of the purely aromatic interface was more rugged than for wild-type RPA. The stability of the variant with the purely electrostatic interface illustrates the importance of long-range electrostatic interactions for the integrity of the RPA–ssDNA interface. The distribution of interfacial energy along the ssDNA nucleotide shows irregular behavior with several local minima for wild-type RPA. When the interfacial aromatic residues (F, W, Y, and H) are replaced by positively charged residues (K and R), the energy profile for the interface, which is then stabilized solely by electrostatic interactions, is more regular with fewer numbers of minima and lower energy barriers in comparison with wild-type RPA (Figure 1B).

The diffusion of ssDNA along RPA involves bulge formation

The texture of the energy surface controls diffusion dynamics, and may suggest faster dynamics for ssDNA moving along the protein interface when all the RPA interfacial residues are positively charged in comparison with the chemically heterogeneous wild-type RPA protein (Figure 1B). To explore the mechanism of ssDNA diffusion along RPA, we simulated the complexes formed between polyT ssDNA and RPA, either wild-type or mutants, using a coarse-grained model that was earlier used to capture the structure of various ssDNA–ssDBP complexes (28,30) (see Materials and Methods). To probe the dynamics of ssDNA while it moves along RPA, we selected six residues on RPA (designated as sites 1–6) distributed along the elongated interface it forms with ssDNA (Figure 2A). The dynamics can be probed by following the evolution of the ssDNA nucleotide index interacting with each of the six selected RPA sites over time, with sliding dynamics identified as a change in the ssDNA nucleotide index with which the RPA interacts. When the same change in ssDNA nucleotide index is identified for adjacent RPA sites (i.e. when an ssDNA index increase at RPA site i is accompanied by an ssDNA index increase at RPA site i+1) then 'correlated' or 'cooperative' movement is observed across these two sites.

Figure 2B and C shows trajectories for the interaction of ssDNA with wild-type and mutant RPA, respectively, and illustrate that the nucleotide index of the ssDNA that interacts with the six sites changes over time. While the overall movement of the ssDNA along the interface with RPA is correlated across the six sites, there are some periods during which the correlation is weaker. These periods (marked by short black lines in Figure 2B and C) may correspond to the formation of bulges of variable lifetime. An examination of the dynamics of ssDNA indicates slower diffusion of ssDNA along wild-type RPA compared with its diffusion along a wholly charged RPA interface in which the entire interface is stabilized by charge–charge interactions and the ssDNA diffuses back and forth along the homogenous RPA interface.

To understand the mechanism at work in the sliding of ssDNA along the RPA surface, we zoomed in on a short period of ssDNA diffusion in the two models and looked at snapshots of the RPA–ssDNA complexes at different time steps (Figure 3A, Supplementary Figures S3 and S4). We colored three ssDNA segments (each of four nucleotides in length, for representative purposes) pink, yellow, and green to follow their location over time. The movement of the ss-

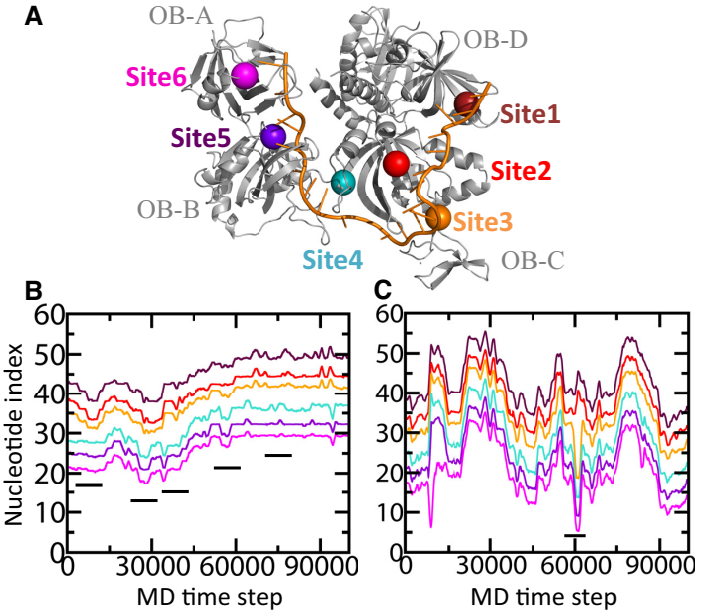

Figure 2. Diffusion of ssDNA along its interface with RPA. (**A**) Six selected sites (numbered 1–6) on the RPA–ssDNA interface are indicated by maroon, red, orange, cyan, purple, and magenta spheres. These sites, which were selected from the 29 sites shown in Figure 1a, are located at the Cβ positions of residues W71, W358, Y308, Y291, F217 and F57, respectively. (B and C) Time evolution of the DNA nucleotide index that interacts with each of sites 1–6 on RPA: (**B**) for wild-type RPA and (**C**) for a mutant RPA in which all aromatic interfacial residues were rendered positively charged. The short black horizontal lines indicate molecular dynamic (MD) timestep periods during which changes in the index of the interacting nucleotides occur differently at different sites, such that DNA progression proceeds in a less consistent manner across the six sites, so suggesting bulge formation.

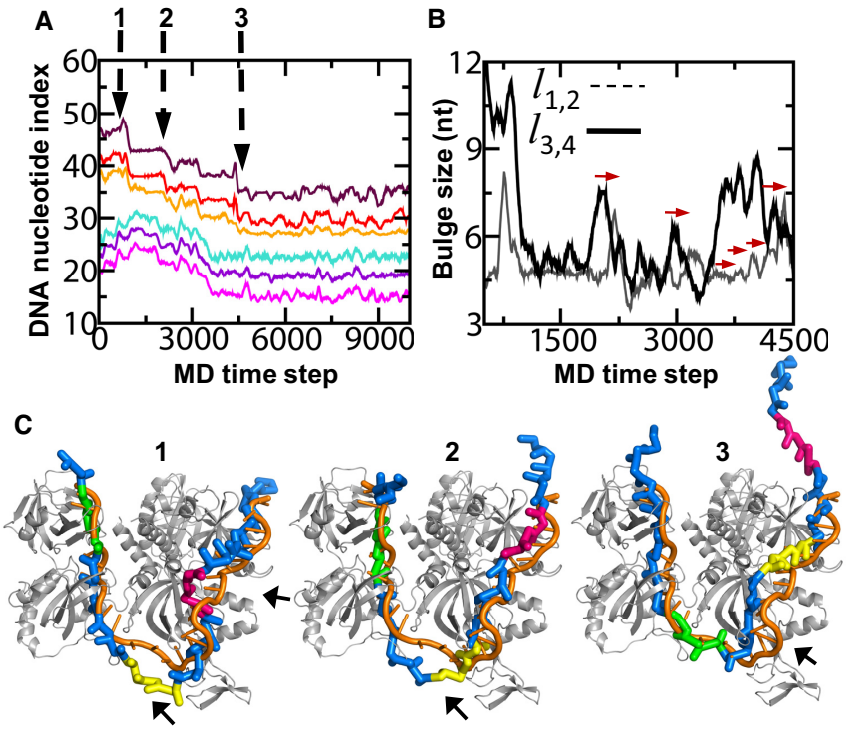

Figure 3. Snapshots of 1D diffusion of ssDNA along RPA. (**A**) Time evolution of the ssDNA nucleotide index interacting with the six selected RPA sites (shown in Figure 2A) for wild-type RPA. (**B**) The time evolution of the size of the bulge formed between sites 1 and 2 (indicated by $l_{1,2}$; dashed grey line) and between sites 3 and 4 (indicated by $l_{3,4}$; thick black line). The short red arrows indicate major events in the migration of the bulge formed between sites 3 and 4 as it diffuses to lie between sites 1 and 2. (**C**) Three snapshots, which were sampled at different times (1, 2 and 3) as indicated by the arrows in the trajectories, illustrate the displacement of the ssDNA. In these snapshots, the sampled ssDNA conformations are shown in blue and, for comparison, their conformation in the crystal structure is shown in orange. To highlight the sliding dynamics, three stretches of the ssDNA were colored in green, yellow, and pink. The small arrows indicate bulges in the ssDNA structure.

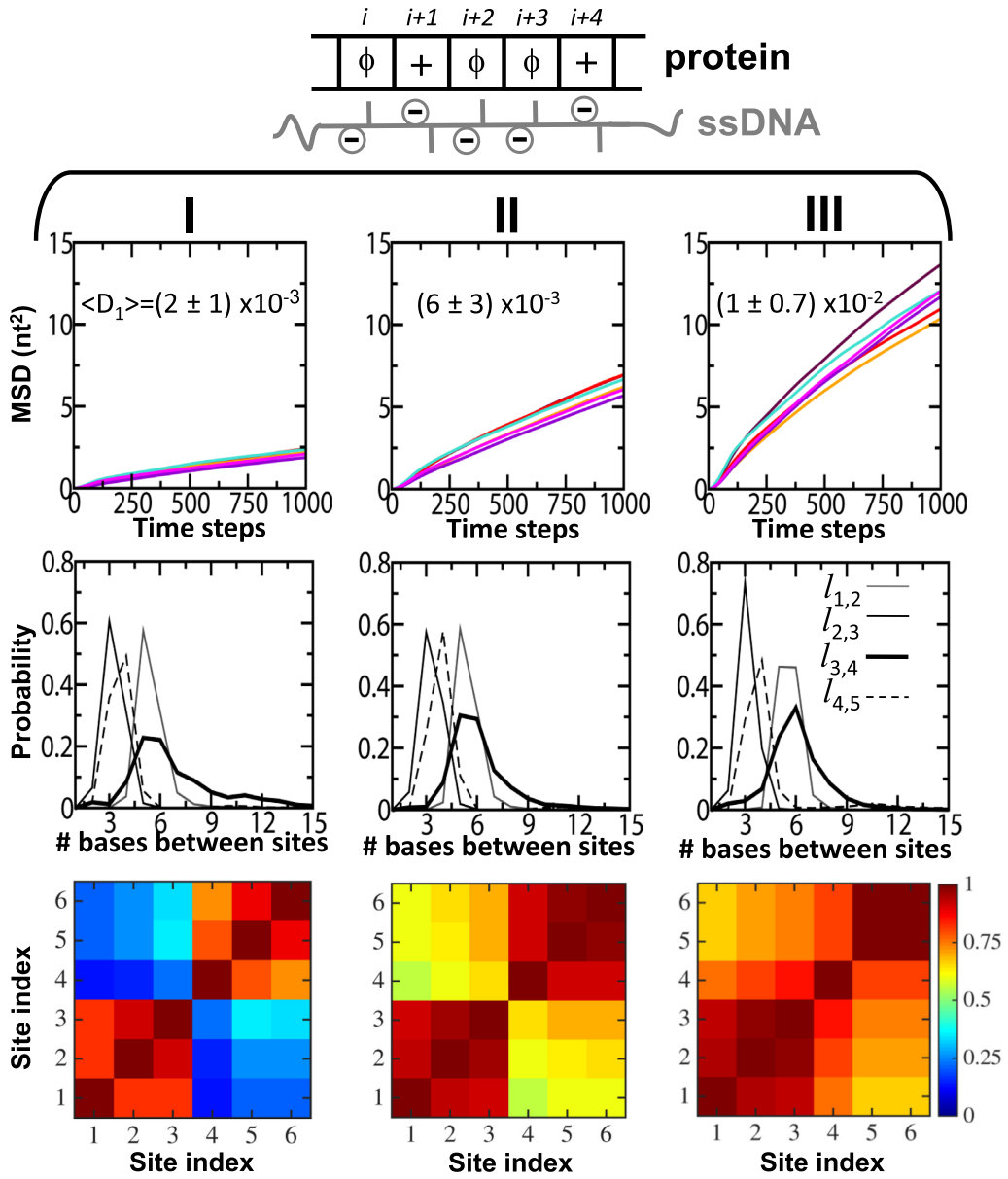

Figure 4. The effect of the stability of the ssDNA–RPA interface on ssDNA diffusion. The scheme (top center) illustrates the interface between ssDNA (gray) and protein (black). The interface is stabilized by the interactions of bases (sticks) with aromatic residues (ϕ) and of ssDNA phosphate groups (circles with minus sign) with positively charges residues (plus sign). Protein sites within the interface are identified by *i*. The diffusion of ssDNA along RPA is studied for three models (I–III), characterized by different ssDNA–RPA interface stabilities that were modulated by changing the strength of the interactions between the aromatic residues and the ssDNA bases: strongest wild-type stability (I), medium interface stability (II), weakest interface stability (III). Diffusion is quantified by the mean square displacement (MSD) as a function of time at sites 1–6 (see Figure 2A) (top row). The number shown on each model is the average value of the diffusion coefficient, $<D_1>$, and its corresponding standard deviation in nt^2/time-step. The distributions of the ssDNA length between sites *i* and *j*, $l_{i,j}$ (i.e. the bulge size; middle row) are given for: $l_{1,2}$ (dashed line), $l_{2,3}$ (solid line), $l_{3,4}$ (thick line) and $l_{4,5}$ (gray line). The width of the distribution at half maximum is narrower by 5% and 11% for models II and III, respectively, in comparison to model I. The matrices (bottom row) show the degree of correlation between ssDNA nucleotide indexes interacting with any pair of RPA sites 1–6 throughout the simulation (red and blue refer to high and low correlation coefficients, respectively). The analysis for each model was obtained by averaging data from 50 simulations each of 10^7 time steps.

DNA nucleotide was not cooperative at all the six sites, in that there were some pauses and jumps in the nucleotide index across different sites. This suggests that the sliding of ssDNA along RPA does not require a simultaneous breaking of contact across the entire interface. Instead, diffusion propagates within smaller segments only, whose boundaries are defined by the points at which the ssDNA dissociates from the RPA. These bulging segments are clearly visible in the snapshots, as demarcated by the arrows (Figure 3C). There are several sites at which bulges form and dissolve, with bulge lifetime and size potentially varying even with respect to a given site.

To examine the effect of interface stability on the accumulation of bulges, we monitored the length of ssDNA ac-

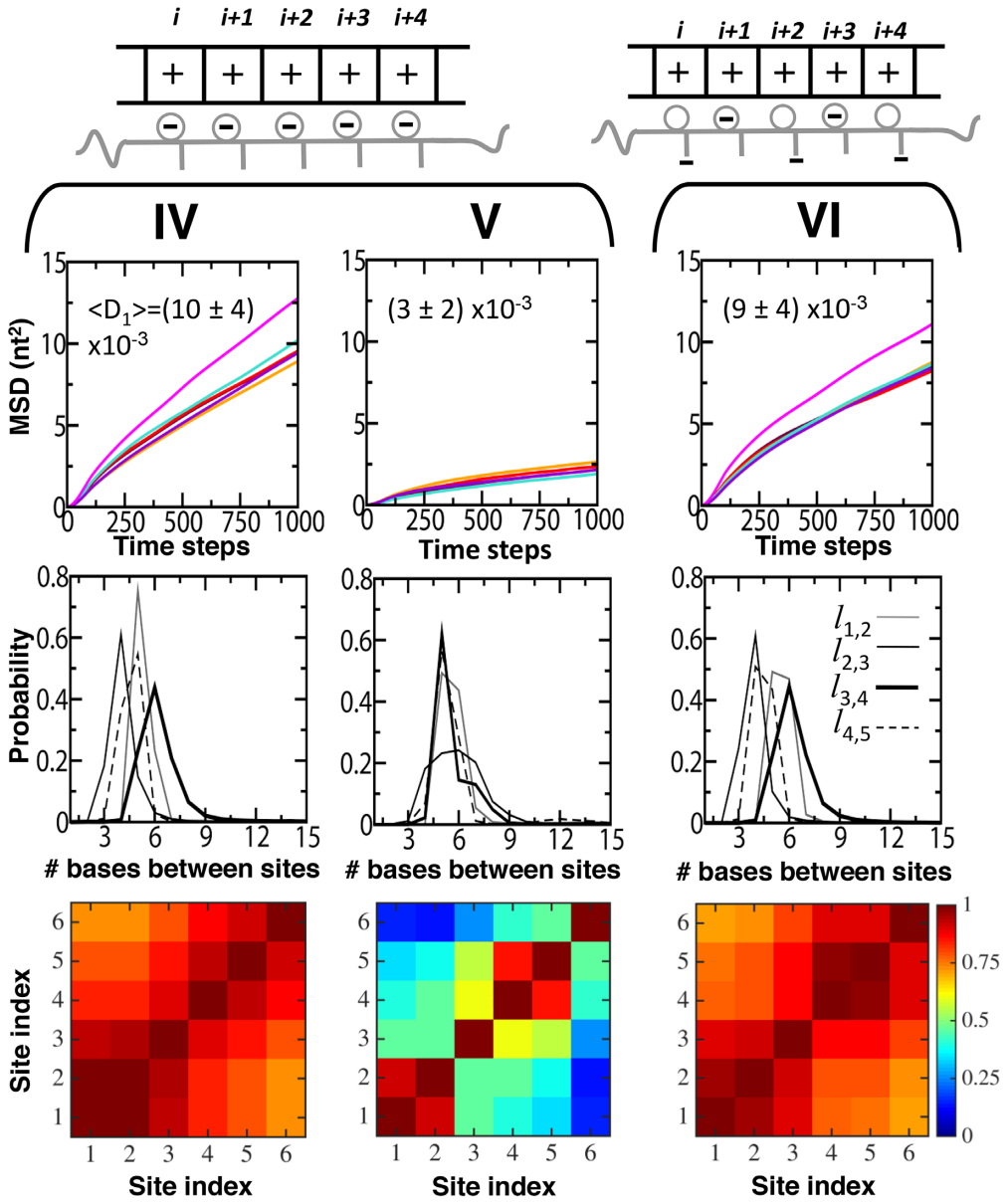

Figure 5. The effect of electrostatic interactions on ssDNA diffusion along RPA. The schemes at the top illustrate the nature of the purely electrostatic ssDNA–RPA interface when the aromatic residues of RPA are mutated to positively charged residues and the negative charges of the ssDNA are placed on its phosphate groups (left) or alternately on its phosphates and bases (right). The linear diffusion was quantified for three different ssDNA–RPA interfaces that are stabilized by purely electrostatic interactions having different characteristics (models IV–VI): long range electrostatic interactions (IV), short-range electrostatic interactions (V), and a system with the negative ssDNA charges placed alternately on bases and phosphates (VI). The linear diffusion of ssDNA along the RPA was quantified by the linear diffusion coefficient (D_1) at sites 1–6 (see Figure 2A) (upper row), the distribution of ssDNA length between selected sites (middle row), and the correlation of the dynamics between sites 1–6 (bottom raw). The number shown on each model is the average value of the diffusion coefficient, $<D_1>$, and its corresponding standard deviation in nt^2/time-step. The width of the distribution at half maximum is narrower by 24%, 58%, and 27%, for models IV, V and VI, respectively, in comparison to model I. This analysis is similar to that shown in Figure 4.

cumulated between some of the six selected sites. If diffusion takes place without bulges, the length of ssDNA through the simulation will be invariant. However, variation in the length of the ssDNA during diffusion may suggest the formation of bulges. To further understand ssDNA diffusion via bulges, we probed the evolution of bulge formation and dissolution over time. The reptation mechanism suggests that the ssDNA diffuses via diffusion of the bulge. Accordingly, release of the excess ssDNA stored in a bulge must be accompanied by bulge formation at a distant site. We followed bulge formation by monitoring the number of nucleotides accumulated between the sites shown in Figure 2A ($l_{i,j} = |nt_i - nt_j|, j = i + 1$). Figure 3B presents the time evolution of $l_{1,2}$ (the length of the ssDNA stored between sites 1 and 2) and $l_{3,4}$ (the length of the ssDNA stored between sites 3 and 4). The migration of the bulge in several events along the representative trajectory occurred by dissolution of the bulge between sites 3 and 4, which was accompanied by bulge formation between sites 1 and 2.

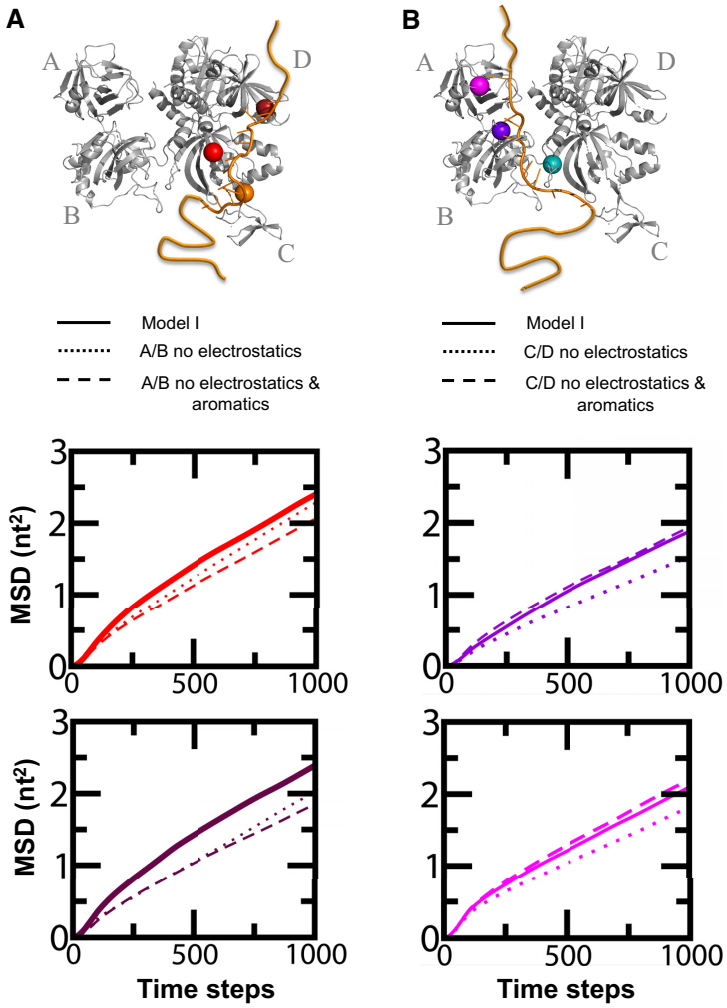

Figure 6. Contribution of the different RPA domains for the ssDNA diffusion. Diffusion of ssDNA is tested along RPA in which some domains are modified by either neutralizing the positively charged residues at the interface with ssDNA or by eliminating both positively charged residues and aromatic residues at the interface. ssDNA diffusion along RPA with modified A and B domains (**A**) or modified C and D domains (**B**). Two types of modifications were studied: in the first model, all the interfacial positively charged residues are neutral (dotted lines) and in the second model both the charged residues as well as the aromatic residues were eliminated (dashed lines). The implication of the latter model is that the ssDNA does not interact with the modified domains besides the excluded volume interactions. For the case that the modified domains are C and D, the diffusion was probed by sites on domains A and B (sites 5 and 6, see Figure 2A) and were compared to the corresponding diffusion coefficients from model I (see Figure 4). For the case that the modified domains are A and B, the diffusion was probed by sites on domains C and D (sites 1 and 2, see Figure 2A) and were compared to the corresponding diffusion coefficients from model I (see Figure 4).

The effect of interface stability on ssDNA diffusion speed

The sliding of ssDNA along its interface with the protein may involve the formation of bulges at different locations. The bulges are transient (i.e. they form and dissociate) and their presence may reduce the cooperativity of ssDNA sliding along the protein surface. The stability of the RPA–ssDNA interface is expected to affect the probabilities of bulge formation and thus the sliding speed and mechanism. We therefore studied diffusion utilizing three models of RPA characterized by identical electrostatic contributions to interface stability but different affinities to ssDNA following manipulation of the strength of the aromatic interactions between the bases and the aromatic residues (see Materials and Method). The aromatic interactions (i.e. between the ϕ group of the protein and the base groups of the ssDNA, see scheme in Figure 4) are strongest for model I and weakest for model III.

To quantify the effect of interface stability on the sliding mechanism and speed, we plotted the mean square displacement (MSD) as a function of time (Figure 4, Supplementary Figure S5) at each of the six selected sites on the RPA interface (see Figure 2A) for models I–III. Faster diffusion is evident as the interface becomes weaker. The linear diffusion coefficient, D_1, for ssDNA along wild-type RPA (model I) is 2×10^{-3} nt^2/time step (being $\sim 4 \times 10^4$ nt^2/s, given that a simulation time step is equivalent to $\sim$50 ps, see SI Appendix, Supplementary Figure S2) which is about three orders of magnitude smaller than the D_1 obtained from simulations for diffusion of a transcription factor along dsDNA (4), in agreement with the experimental values (8,12). The value of the D_1 from the coarse-grained simulations is about

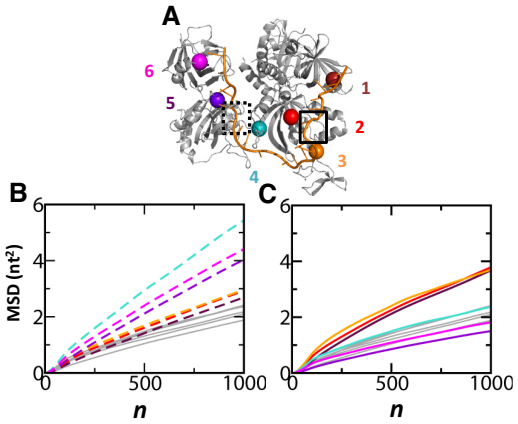

Figure 7. Effect of point mutations on the linear diffusion of ssDNA along RPA. (**A**) Local mutations of two aromatic residues (to positively charged residues) are introduced at two regions marked by a square: one variant involves mutations of residues W180/F217 (located at the dashed square) and the other one involves mutations of residues W358/F411 (located at the solid square). (B and C) The corresponding MSD plots of ssDNA diffusion along the two RPA mutants using the coarse-grained simulations, probed at the six selected sites on RPA as shown by their respective colors. The MSD for the variants with the two point mutations made in regions indicated by the dashed and solid squares are shown by dashed (**B**) and solid (**C**) lines, respectively.

an order of magnitude higher than the experimentally measured D_1 of 10^3–5×10^3 nt^2/s at temperature range of 10–37°C (8).

Figure 4 shows the distribution of ssDNA length between sites 1 and 2 ($l_{1,2}$), 2 and 3 ($l_{2,3}$), 3 and 4 ($l_{3,4}$) and 4 and 5 ($l_{4,5}$). We observed that some of the distributions are broad and $l_{i,j}$ has a range of values with finite probability, which suggests the possibility of bulge formation. It is interesting to note that the distribution width of $l_{3,4}$ is wider than those of $l_{2,3}$ and $l_{4,5}$ (Figure 4, middle panels), indicating that the largest sized bulges can occur between sites 3 and 4 on the RPA. More importantly, the distribution of $l_{3,4}$ is widest for model I, suggesting that the greater the stability of the interface, the greater the probability of formation of a larger bulge that includes excess ssDNA. However, when we reduce the aromatic interaction strength in the other two models, $l_{3,4}$ shows a relatively narrower distribution.

The effect of bulges on the cooperative motion of ssDNA along the RPA surface can be examined by measuring correlation between the nucleotide indexes interacting with the different sites 1–6. Figure 4 (bottom panels) shows the degree of correlation that characterizes the diffusion of the ssDNA through the six selected sites in the three RPA variants. The correlation matrix reflects that the interface of model I (wild-type RPA) is divided into two regions. The degree of correlation within sites 1–3 and within sites 4–6 is high (colored red) but the correlation between sites 1–3 and sites 4–6 is very low. The poor correlation between these two regions is consistent with the accumulation of a larger bulge between sites 3 and 4 that breaks diffusion coupling through the interface. As interface stability decreases, the overall diffusion of ssDNA along the six selected RPA sites becomes more correlated (Figure 4). The greater cooperativity between sites 1–3 and sites 4–6 is consistent with higher diffusion coefficients (Figure 4, upper panel) and a lower formation probability for long bulges (Figure 4, middle panels). The convergence of the simulations was verified by running another independent set of 50 simulations for models I and III which show consistent ssDNA diffusion to that concluded from the original sets of simulations (see SI Appendix, Supplementary Figure S5).

Contribution of long range electrostatic interactions to 1D diffusion

The origin of the weaker cooperativity of ssDNA diffusion on RPA in the presence of strong interactions at the interfaces between aromatic residues and ssDNA bases may be linked to their short-range nature. Accordingly, these strong aromatic interactions may disrupt continuous ssDNA sliding dynamics because they act as deep and narrow energetic traps. To address the hypothesis that bulges preferentially form around deep energetic minima, we designed three additional models (IV–VI) in which the RPA–ssDNA interface is stabilized solely by long-range electrostatic interactions. In these models, all the interfacial residues are positively charged and interact with the phosphates of ssDNA (Figure 5, scheme). Faster diffusion of ssDNA along RPA is observed in the presence of these long-range electrostatic interactions (Figure 5, model IV) compared with the wild-type model in which the interface is stabilized by both short- and long-range interactions (Figure 4, model I). Bulges also form when using the model with the homogenous interface that is stabilized wholly by long-range electrostatic interactions (Figure 5A, model IV), however, the probabilities of bulge formation are lower in comparison with the case of heterogeneous interfacial residues (Figure 4, model I). Furthermore, there is greater correlation between the nucleotide indices interacting with the different sites on the interface for the mutant model (Figure 5, model IV) relative to wild-type RPA (Figure 4, model I), which indicates much more cooperative diffusion of ssDNA along mutant RPA in the former case.

To further examine the role of long-range interactions in increasing cooperativity of diffusion and, thereby, its speed, we designed another model in which the RPA–ssDNA interface is governed by electrostatic interactions that have a much shorter range (Figure 5A, model V). We note that the energetic gain from forming an attractive electrostatic interaction in the long- or short-range models (Figure 5, models IV and V) is very similar. Converting the electrostatic interactions into short-range interactions resulted in slower diffusion and poorer correlation between the dynamics adopted at each of the six sites (Figure 5, model V compared with model IV). This result substantiates our finding that the short-range nature of aromatic interactions slows the diffusion of ssDNA along RPA.

Another possible mechanism for slowing ssDNA diffusion may be the frequent requirement for the aromatic bases facing the aromatic residues to 'flip' as the ssDNA moves along the heterogeneous surface and the interactions change from aromatic to electrostatic, or vice versa. To investigate this possibility, we considered a model of the

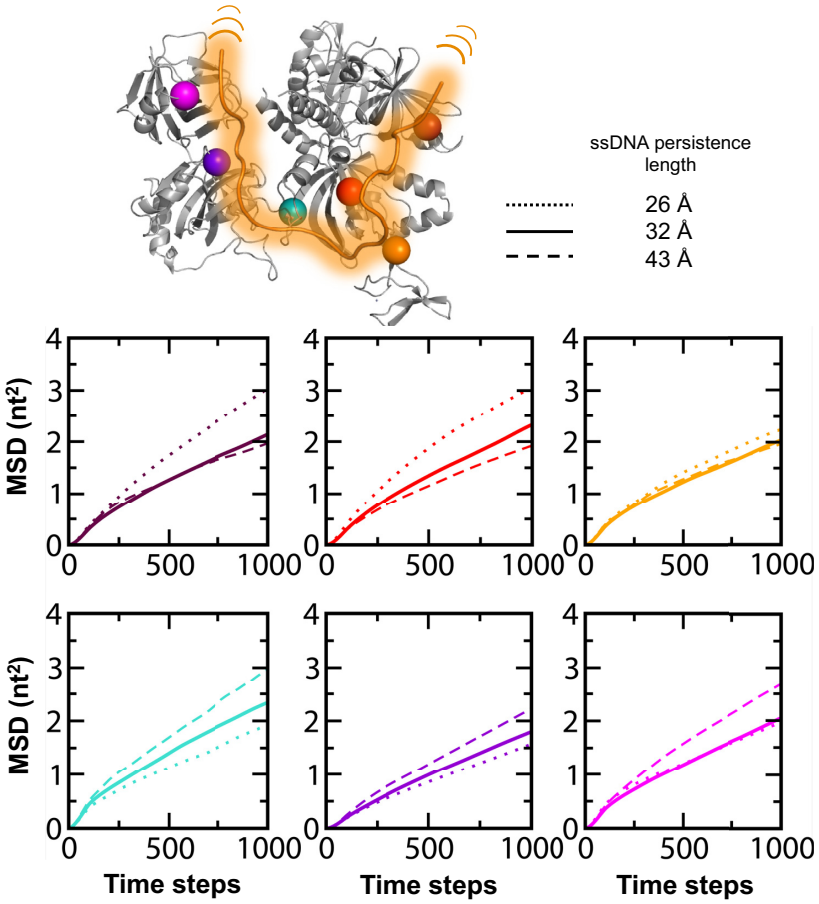

Figure 8. The effect of the persistence length of the ssDNA on its diffusion along RPA. The persistence length is modified by changing the strength of the dihedral angle between four consecutive phosphate beads. In the original model, the strength of this dihedral angle is 0.7 (corresponds to persistence length of 32 Å, solid line). Two additional ssDNA models were studied with dihedral strength of 0.2 (persistence length of 26 Å, dotted line) and 2.0 (persistence length of 43 Å, dashed line). The MSD curves suggest that the rigidity of the ssDNA may have local effect on the diffusion. It is shown that increasing the ssDNA flexibility (dotted lines) results in faster diffusion through sites 1–3. On the other hand, reducing the ssDNA flexibility (dashed lines) results in faster diffusion through sites 4–6. This local sensitivity to the persistence length of the ssDNA may correspond to the local conformation the ssDNA adopts at these sites and that changing the flexibility may not be compatible with these local structures. This suggests how mutating the ssDNA may affect the diffusion via change in persistence length.

RPA–ssDNA interface in which the negative point charges are placed alternately on the phosphate and base beads of ssDNA (Figure 5, model VI). As the ssDNA diffuses along a purely positively charged RPA, the phosphate and base beads must reorient, which is qualitatively similar to the flipping of a base as ssDNA moves along the heterogeneous protein surface. We found that reorientation does not significantly affect the diffusion coefficients or cooperativity between the six sites (Figure 5, model VI compared with model IV). Consistently with the similar cooperativity between models IV and VI, there is no strong difference between the probabilities of bulge formation in these two models in which the RPA–ssDNA interface is dominated solely by electrostatic interactions. While high cooperativity is indicative of ssDNA diffusion via sliding mechanism, low cooperativity supports the reptation mechanism. A crude estimate of the diffusion mechanism can be obtained using the cooperativity matrices (Figures 4 and 5) that suggest that model I follows 80% reptation and 20% sliding and model IV follows 26% reptation and 74% sliding.

Contribution of the different RPA domains to the ssDNA diffusion

The RPA interface for diffusion of ssDNA comprises four OB domains labeled A-D, whose structures as well as the chemical properties of their surface are not identical. Given the extended and linear conformation of the ssDNA when bound to RPA, one may consider diffusion along smaller interface composed of fewer OB domains. To address this question, we designed variants of RPA in which domains A and B or domains C and D are modified. Two types of modifications were considered. In the first modification, all the interfacial residues in A and B (or in C and D) were eliminated. In this case, the ssDNA has no stabilizing interactions with the modified domains and the ssDNA diffusion can only be along the remaining domains (i.e., C and D or A and B, respectively). In the second modification, only the positively charged residues were eliminated on A and B (or on C and D) so the diffusion along these domains cannot be supported by the long-range electrostatic interactions.

The diffusion of ssDNA along domains A or B are quite similar regardless of the nature of the interactions ssDNA has with domains C and D. Accordingly, the diffusion coefficients of diffusion along domains A and B is similar in cases that ssDNA interacts tightly with domains C and D or has no affinity to these domains (Figure 6B). The ability of ssDNA to diffuse along fewer OB domains with similar diffusion coefficient to wild-type RPA is consistent with experimental results that showed diffusion along a truncated variant of RPA that lacks OB domains C, D and E (8). Similarly, diffusion of ssDNA along domains C and D is almost insensitive to the strength of interactions between the ssDNA and domains A and B (Figure 6A). The ability to diffuse along smaller interface of RPA demonstrates that diffusion of ssDNA can take place along fewer OB domains. Furthermore, the similarity in the diffusion coefficients for diffusion along wild-type RPA or along two adjacent OB domains reflect that the diffusion is fragmented and governed by the local environment introduced by each domain. The slightly lower diffusion coefficients along C and D domains when the interactions of ssDNA with domains A and B is weaker can be explained by the lower probability of forming bulges between domains B and C that disrupt the communication between these regions and disrupt ssDNA motion. This effect is greater when both aromatic and electrostatic interactions are eliminated between ssDNA and domains A and B (Figure 6A). The diffusion along domains A and B is less affected by eliminating interactions with C and D (Figure 6B). This can be explained by their higher affinity to ssDNA (also reflected by their slower diffusion in wild-type RPA compared to diffusion of the C and D domains) and therefore ssDNA diffusion along A and B is less perturbed by changes in the other domains.

Effect of local mutations and conformational flexibility

Point mutations at the RPA–ssDNA interface can further affect the degree of coupling and cooperativity in ssDNA diffusion. When ssDNA propagates along a homogenous interface governed by electrostatic interactions, a consistent increase in the diffusion rate is observed throughout the interface compared with diffusion along the heterogeneous RPA–ssDNA interface. It is still not known how a local modification at the interface will affect the diffusion characteristics. We therefore designed two additional variants of RPA, each of which comprised two point mutations of aromatic residues to Lys. One variant included mutations of W358 and F411 (located near sites 2 and 3) and the other variant included mutations of residues W180 and F217 (located near sites 4 and 5) (see Figure 7). The mutations near sites 2 and 3 accelerate the sliding of ssDNA along sites 1–3 relative to wild-type RPA. However, these mutations have no significant effect on ssDNA diffusion through sites 4–6 compared with wild-type RPA (Figure 4, model I). A similar observation is made when the mutations are near sites 4 and 5. These mutations accelerate the diffusion of ssDNA along sites 4–6 relative to wild-type RPA but do not affect the sliding of ssDNA along sites 1–3. These findings suggest that changes occur to the local movement of ssDNA near the mutated sites, however, no changes occur in the global motion of ssDNA across all six-sites, which further implies that these aromatic residues act as anchoring sites for ssDNA. Similarly, the ssDNA sequence may affect the local diffusion via change in the ssDNA persistence length, indicating that the ssDNA rigidity may affect its migration (Figure 8). Restricting RPA flexibility is also found to slow down the diffusion (See SI Appendix, Supplementary Figure S6).

CONCLUSIONS

The molecular mechanism of the experimentally reported ssDNA diffusion along RPA (8) is studied here using computational models. The diffusion of 60 nt ssDNA along the RPA protein was obtained using a coarse-grained model that was earlier shown to capture the structures of various ssDNA–ssDBP complexes (28,30). The diffusion of ssDNA along RPA is accompanied by the formation of bulges; thus following the reptation (i.e. sliding-with bulges) mechanism. The bulges are stochastically formed and they are transient. There are some sites on the protein surface at which the probability of bulge formation is greater than others, depending on the interface topology and the interfacial residues. Moreover, the lifetime of each bulge and the length of the ssDNA stored at each site varies. For RPA, the bulges store 1–7 nt of ssDNA, consistently with experimental (20) and computational (29) observations for *E. coli* SSB. Formation of bulges causes the diffusion to be more fragmented, which may serve as an efficient means to diffuse along the extensive and elongated interface of RPA–ssDNA. Consequently, the cooperativity of the diffusion of the ssDNA throughout the interface it forms with RPA is affected by the intermittent dynamics introduced by the bulges. We note that the diffusion of ssDNA via defects at the interface with RPA is reminiscent to the sliding mechanism in nucleosomes that was shown recently to propagate via twist defects (42). Furthermore, one may speculate that the potentially long bulge, particularly between OB-B and OB-C may mediate the experimentally self-exchange of the ssDNA between different RPA molecules (26,43).

The wild-type RPA–ssDNA interface is chemically heterogeneous, as it is stabilized by both electrostatic interactions (e.g. Lys, Arg or His with the ssDNA phosphate groups) and aromatic interactions (e.g. Trp, Phe, or Tyr with the ssDNA base groups). Mutating all the aromatic residues to positively charged residues results in an energetically smoother landscape for ssDNA diffusion. Long-lived bulges are less common when ssDNA diffuses along such a purely electrostatic RPA interface, presumably because they dissociate more rapidly due to the long-range nature of the electrostatic interactions. Consequently, ssDNA diffusion is more cooperative and therefore faster for the sliding-without-bulges mechanism. Electrostatic interactions are essential for the progression of ssDNA. Aromatic interactions contribute to the stability of the RPA–ssDNA interface, however, their short-range nature means that they decrease cooperativity between the sites and thus reduce diffusion speed. Accordingly, aromatic interactions play an important role in controlling the dynamics of the RPA–ssDNA complex. One may note that the heterogeneous RPA–ssDNA interface comprises approximately equal numbers of aromatic and electrostatic residues and

that and they are spread along the entire interface. This balance between the aromatic and electrostatic forces confers high affinity on ssDNA–RNA interactions at the interface while ensuring ssDNA mobility at its interface with RPA.

To summarize, our study describes the molecular mechanisms involved in the 1D diffusion of ssDNA along its RPA binding interface in the context of extensive ssDNA–RPA interactions. We showed that the sliding of ssDNA along RPA does not require simultaneous breaking of all the contacts; rather the ssDNA breaks only a few contacts with the RPA interface, which results in the formation of bulges as the ssDNA moves. This direct evidence for the involvement of bulges in the 1D diffusion of ssDNA along its RPA binding interface substantiates the hypothesis that a reptation mechanism is utilized for SSB protein movement along ssDNA.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by the Kimmelman Center for Macromolecular Assemblies. Y.L. holds The Morton and Gladys Pickman professional chair in Structural Biology. We thank Harry M. Greenblatt for assitance with atomistic simulations of RPA. GM acknowledges support from the computational facility from HPC2013 cluster of IIT Kanpur.

FUNDING

Israel Science Foundation grant number 1583/17. DST through an INSPIRE faculty fellowship grant number DST/INSPIRE/04/2014/002085. Funding for open access charge: Internal funding.
Conflict of interest statement. None declared.

REFERENCES

1. Givaty,O. and Levy,Y. (2009) Protein sliding along DNA: dynamics and structural characterization. *J. Mol. Biol.*, **385**, 1087–1097.
2. Marklund,E.G., Mahmutovic,A., Berg,O.G., Hammar,P., van der Spoel,D., Fange,D. and Elf,J. (2013) Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 19796–19801.
3. Ando,T. and Skolnick,J. (2014) Sliding of proteins non-specifically bound to DNA: brownian dynamics studies with Coarse-Grained Protein and DNA Models. *PLoS Comput. Biol.*, **10**, e1003990.
4. Marcovitz,A. and Levy,Y. (2012) *Sliding Dynamics Along DNA: A Molecular Perspective*. RSC.
5. Shereda,R.D., Kozlov,A.G., Lohman,T.M., Cox,M.M. and Keck,J.L. (2008) SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.*, **43**, 289–318.
6. Roy,R.A.G., Kozlov,T.M.L. and Ha,T. (2009) SSB protein diffusion on single-stranded DNA stimulates RecA filament formation. *Nature*, **461**, 1092–1097.
7. Ha,T., Kozlov,A.G. and Lohman,T.M. (2012) Single-molecule views of protein movement on single-stranded DNA. *Annu. Rev. Biophys.*, **41**, 295–319.
8. Nguyen,B., Sokoloski,J., Galletto,R., Elson,E.L., Wold,M.S. and Lohman,T.M. (2014) Diffusion of human replication protein A along single-stranded DNA. *J. Mol. Biol.*, **426**, 3246–3261.
9. Kozlov,A.G. and Lohman,T.M. (2002) Kinetic mechanism of direct transfer of Escherichia coli SSB tetramers between single-stranded DNA molecules. *Biochemistry*, **41**, 11611–11627.
10. Spenkelink,L.M., Lewis,J.S., Jergic,S., Xu,Z.Q., Robinson,A., Dixon,N.E. and van Oijen,A.M. (2019) Recycling of single-stranded DNA-binding protein by the bacterial replisome. *Nucleic. Acids. Res.*, **47**, 4111–4123.
11. Berg,O.G., Winter,R.B. and von Hippel,P.H. (1981) Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **20**, 6929–6948.
12. Blainey,P.C., Luo,G., Kou,S.C., Mangel,W.F., Verdine,G.L., Bagchi,B. and Xie,X.S. (2009) Nonspecifically bound proteins spin while diffusing along DNA. *Nat. Struct. Mol. Biol.*, **16**, 1224–1229.
13. Mirny,L., Slutsky,M., Wunderlich,Z., Tafvizi,A., Leith,J. and Kosmrlj,A. (2009) How a protein searches for its site on DNA: the mechanism of facilitated diffusion. *J. Phys. A-Math. Theor.*,**42**, 43.
14. Tafvizi,A., Huang,F., Fersht,A.R., Mirny,L.A. and van Oijen,A.M. (2011) A single-molecule characterization of p53 search on DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 563–568.
15. Halford,S.E. and Marko,J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic. Acids. Res.*, **32**, 3040–3052.
16. Raghunathan,S., Kozlov,A.G., Lohman,T.M. and Waksman,G. (2000) Structure of the DNA binding domain of E. coli SSB bound to ssDNA. *Nat. Struct. Biol.*, **7**, 648–652.
17. Roy,R., Kozlov,A.G., Lohman,T.M. and Ha,T. (2007) Dynamic structural rearrangements between DNA binding modes of E. coli SSB protein. *J. Mol. Biol.*, **369**, 1244–1257.
18. Lohman,T.M. and Ferrari,M.E. (1994) Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. *Annu. Rev. Biochem.*, **63**, 527–570.
19. Kuznetsov,S.V., Kozlov,A.G., Lohman,T.M. and Ansari,A. (2006) Microsecond dynamics of protein-DNA interactions: direct observation of the wrapping/unwrapping kinetics of single-stranded DNA around the E. coli SSB tetramer. *J. Mol. Biol.*, **359**, 55–65.
20. Zhou,R., Kozlov,A.G., Roy,R., Zhang,J., Korolev,S., Lohman,T.M. and Ha,T. (2011) SSB functions as a sliding platform that migrates on DNA via reptation. *Cell*, **146**, 222–232.
21. Suksombat,S., Khafizov,R., Kozlov,A.G., Lohman,T.M. and Chemla,Y.R. (2015) Structural dynamics of E. coli single-stranded DNA binding protein reveal DNA wrapping and unwrapping pathways. *Elife*, **4**, e08193.
22. Lee,K.S., Marciel,A.B., Kozlov,A.G., Schroeder,C.M., Lohman,T.M. and Ha,T. (2014) Ultrafast redistribution of E. coli SSB along long single-stranded DNA via intersegment transfer. *J. Mol. Biol.*, **426**, 2413–2421.
23. Zou,Y., Liu,Y.Y., Wu,X.M. and Shell,S.M. (2006) Functions of human replication protein A (RPA): From DNA replication to DNA damage and stress responses. *J. Cell. Physiol.*, **208**, 267–273.
24. Chen,R. and Wold,M.S. (2014) Replication protein A: Single-stranded DNA's first responder Dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. *Bioessays*, **36**, 1156–1161.
25. Fan,J. and Pavletich,N.P. (2012) Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. *Genes Dev.*, **26**, 2337–2347.
26. Gibb,B., Ye,L.F., Gergoudis,S.C., Kwon,Y., Niu,H., Sung,P. and Greene,E.C. (2014) Concentration-dependent exchange of replication protein A on single-stranded DNA revealed by single-molecule imaging. *PLoS One*, **9**, e87922.
27. Kunzelmann,S., Morris,C., Chavda,A.P., Eccleston,J.F. and Webb,M.R. (2010) Mechanism of interaction between single-stranded DNA binding protein and DNA. *Biochemistry-US*, **49**, 843–852.
28. Mishra,G. and Levy,Y. (2015) Molecular determinants of the interactions between proteins and ssDNA. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 5033–5038.
29. Maffeo,C. and Aksimentiev,A. (2017) Molecular mechanism of DNA association with single-stranded DNA binding protein. *Nucleic Acids Res.*, **45**, 12125–12139.
30. Mukherjee,G., Pal,A. and Levy,Y. (2017) Mechanism of the formation of the RecA-ssDNA nucleoprotein filament structure: a coarse-grained approach. *Mol. Biosyst.*, **13**, 2697–2703.

31. Levy,Y., Wolynes,P.G. and Onuchic,J.N. (2004) Protein topology determines binding mechanism. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 511–516.
32. Azia,A. and Levy,Y. (2009) Nonnative electrostatic interactions can modulate protein folding: molecular dynamics with a grain of salt. *J. Mol. Biol.*, **393**, 527–542.
33. Morriss-Andrews,A., Rottler,J. and Plotkin,S.S. (2010) A systematically coarse-grained model for DNA and its predictions for persistence length, stacking, twist, and chirality. *J. Chem. Phys.*, **132**, 035105.
34. Freeman,G.S., Hinckley,D.M. and de Pablo,J.J. (2011) A coarse-grain three-site-per-nucleotide model for DNA with explicit ions. *J. Chem. Phys.*, **135**, 165104.
35. Ouldridge,T.E., Louis,A.A. and Doye,J.P. (2011) Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.*, **134**, 085101.
36. Chakraborty,D., Horio,N. and Thirumalai,D. (2018) Sequence-Dependent three interaction site model for single- and double-stranded DNA. *J. Chem. Theory Comput.*, **14**, 3763–3779.
37. Tinland,B., Pluen,A., Sturm,J. and Weill,G. (1997) Persistence length of single-stranded DNA. *Macromolecules*, **30**, 5763–5765.
38. Murphy,M.C., Rasnik,I., Cheng,W., Lohman,T.M. and Ha,T. (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.*, **86**, 2530–2537.
39. Chen,H., Meisburger,S.P., Pabit,S.A., Sutton,J.L., Webb,W.W. and Pollack,L. (2012) Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 799–804.
40. Rutledge,L.R., Campbell-Verduyn,L.S. and Wetmore,S.D. (2007) Characterization of the stacking interactions between DNA or RNA nucleobases and the aromatic amino acids. *Chem. Phys. Lett.*, **444**, 167–175.
41. Andrews,C.T., Campbell,B.A. and Elcock,A.H. (2017) Direct comparison of amino acid and salt interactions with double-stranded and single-stranded DNA from explicit-solvent molecular dynamics simulations. *J. Chem. Theory Comput.*, **13**, 1794–1811.
42. Brandani,G.B., Niina,T., Tan,C. and Takada,S. (2018) DNA sliding in nucleosomes via twist defect propagation revealed by molecular simulations. *Nucleic Acids Res.*, **46**, 2788–2801.
43. Pokhrel,N., Origanti,S., Davenport,E.P., Gandhi,D., Kaniecki,K., Mehl,R.A., Greene,E.C., Dockendorff,C. and Antony,E. (2017) Monitoring Replication Protein A (RPA) dynamics in homologous recombination through site-specific incorporation of non-canonical amino acids. *Nucleic Acids Res.*, **45**, 9413–9426.