Molecular BioSystems

PAPER

Cite this: *Mol. BioSyst.*, 2017, **13**, 2697

Mechanism of the formation of the RecA–ssDNA nucleoprotein filament structure: a coarse-grained approach

Goutam Mukherjee, Arumay Pal and Yaakov Levy*

In prokaryotes, the RecA protein catalyzes the repair and strand exchange of double-stranded DNA. RecA binds to single-stranded DNA (ssDNA) and forms a presynaptic complex in which the protein polymerizes around the ssDNA to form a right-handed helical nucleoprotein filament structure. In the present work, the mechanism for the formation of the RecA–ssDNA filament structure is modeled using coarse-grained molecular dynamics simulations. Information from the X-ray structure was used to model the protein itself but not its interactions; the interactions between the protein and the ssDNA were modeled solely by electrostatic, aromatic, and repulsive energies. For the present study, the monomeric, dimeric, and trimeric units of RecA and 4, 8, and 11 NT-long ssDNA, respectively, were studied. Our results indicate that monomeric RecA is not sufficient for nucleoprotein filament formation; rather, dimeric RecA is the elementary binding unit, with higher multimeric units of RecA facilitating filament formation. Our results reveal that loop region flexibility at the primary binding site of RecA is essential for it to bind the incoming ssDNA, that the aromatic residues present in the loop region play an important role in ssDNA binding, and that ATP may play a role in guiding the ssDNA by changing the electrostatic potential of the RecA protein.

Received 2nd August 2017,
Accepted 27th October 2017

DOI: 10.1039/c7mb00486a

rsc.li/molecular-biosystems

Introduction

In an organism, when damage to dsDNA occurs through metabolic processes or external effects, such as radiation, a DNA repair process begins. The repair of the damaged dsDNA is carried out by the homologous recombination process. This recombination process is constantly active throughout the entire lifespan of prokaryotes and eukaryotes,[1] and defects in the dsDNA repair process cause cancer. Proteins, such as RecA or Rad51 for prokaryotes and eukaryotes respectively, are involved in this repair process.[1,2] The recombination process occurs in three stages: (i) the assembly of the protein and ssDNA to form a protein–ssDNA complex; (ii) dsDNA binding to the complex and the initiation of complementary strand search; and (iii) strand exchange.[3]

One of the main reactions in this recombination process is strand exchange between ssDNA and dsDNA. Once damage occurs in dsDNA, ssDNA forms in that region. One of the ssDNA strands gets degraded, and the other strand interacts with the RecA protein in a highly cooperative manner to from a right-handed helical nucleoprotein filament that is also called a presynaptic filament.[4–18] The formation of the nucleoprotein filament takes place in the presence of an adenosine triphosphate (ATP) molecule that is sandwiched between two RecA monomers[5] (Fig. 1). The ATP-bound protein has a higher affinity for ssDNA binding, whereas the hydrolyzed state of the ATP (*i.e.*, the adenosine diphosphate-bound (ADP-bound)

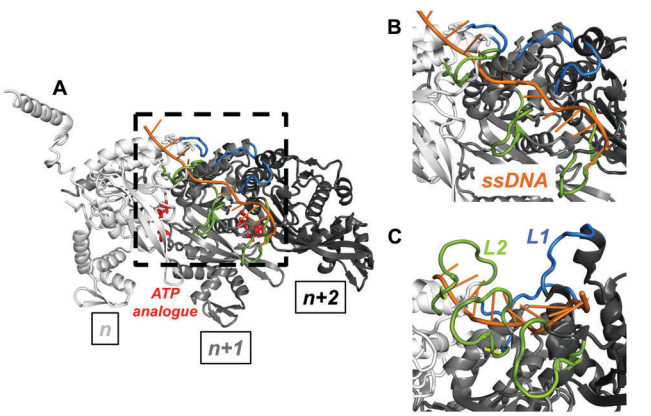

Fig. 1 The complex of the trimeric RecA with ssDNA. The three monomers of RecA are shown in grey and the ssDNA in orange. (A) Pictorial representations of the L1 and L2 loops in the RecA(trimer)–ssDNA complex. (B and C) Show a close-up view of the dashed square. Loops L1 and L2 are shown in blue and green, respectively. The ATP analogues are shown in red. The structures were made using the PDB code 3CMU.

Department of Structural Biology, Weizmann Institute of Science, Rehovot, 76100, Israel. E-mail: Koby.Levy@weizmann.ac.il

state of the protein) strongly affects the thermodynamic and kinetic stability of the RecA–ssDNA complex. Previous experimental studies confirmed that the nucleoprotein filament adopts a stretched conformation in the ATP-bound state, whereas in the ADP-bound state, the conformation changes to a compressed state (also known as the inactive state).[10]

In the next step, the filament extension process begins *via* the association of more RecA monomers with the ssDNA in the $5' \rightarrow 3'$ direction, and hence the displaced strand is also released in that direction.[19] Traditionally, the location at which ssDNA binds with the RecA protein is known as the primary binding site. A recent structural analysis revealed that the RecA–ssDNA complex has a secondary binding site as well, where the donor dsDNA binds weakly during the presynaptic phase.[5] In the synaptic phase, the RecA–ssDNA filament structure starts interacting with dsDNA by searching for the complementary strand within the bound dsDNA.[2,6,20,21] This search process takes place in the presence of ATP. The reaction between RecA–ssDNA and dsDNA occurs at the primary binding site, and the growing DNA heteroduplex remains in place at the primary binding site. However, the displaced ssDNA strand from the donor dsDNA system binds more tightly at the secondary binding site and thus provides stability to the overall system. After the strand exchange reaction, the filament is disassembled by the hydrolysis of ATP at the interface between the RecA units.[5,6,14,15,22]

Over the past two decades, several groups have endeavored to understand the mechanism by which the filament forms.[7–9,12,13] From the experimental perspective, it is clear that the filament formation process takes place in two steps: (i) nucleation and (ii) filament extension.[8] However, the mechanism of the nucleation step (*i.e.*, the minimum number of monomers required to start nucleation) is still not clearly understood. Based on circular dichroism spectroscopy, gel filtration, and electron microscopy experiments, it was concluded that a monomeric unit of RecA is the basic unit for filament formation.[7] In contrast, other advanced experimental techniques showed that a RecA-dimer[8,12] or a RecA-pentamer[13] may serve as the fundamental unit for filament formation. Thus, the origin of the nucleation mechanism between RecA and ssDNA remains unclear and requires study by a computational approach.

In this work, we probe the nucleation mechanism using coarse-grained modeling of the RecA protein and ssDNA systems. We took the structures of RecA-monomer, RecA-dimer, and RecA-trimer with ssDNA molecules of length 4 NT, 8 NT, and 11 NT, respectively, to investigate the nucleation mechanism. We investigate the degree of specificity in the interactions between ssDNA and RecA by examining the various binding modes in addition to the one found in the X-ray structure. We thus ask to what extent the RecA is frustrated for interaction with ssDNA and whether their interactions[23] are governed by a funneled energy landscape.[24,25] Furthermore, the RecA–ssDNA complex (PDB ID: 3CMU) clearly shows the presence of protein loops at the primary binding site where ssDNA binds. Therefore, we performed the coarse-grained simulations by varying the flexibility of the loop present near the primary binding site. The role of ATP in complex formation was investigated as well.

Materials and methods

Coarse-grained model for RecA–ssDNA interactions

We used a newly developed coarse-grained model to explore the mechanism of the RecA nucleation process at the molecular level.[26] In this model, each protein residue is represented by two beads placed at the C_α and C_β positions (only C_α for Gly).[26–28] For the charged amino acids (K, R, D, and E), a positive/negative unit charge is placed at the C_β positions. The protein is simulated by a native topology-based model in which the Lennard-Jones potential is used to represent native contact interactions. For ssDNA, each nucleotide is modeled by three beads to represent phosphate, sugar, and base. The potential energy of interaction between a protein residue and an ssDNA nucleotide is modeled from three components: (i) electrostatic interactions between the C_β beads of the charged residues (K, R, D, and E) and the negatively charged phosphate beads of the ssDNA; (ii) stacking interactions between the C_β beads of the aromatic residues (W, F, Y, and H) and the ssDNA base bead; and (iii) repulsive interactions between all other beads of the protein and ssDNA. The details of the potentials used for the protein, the ssDNA, and their interactions are described in detail in ref. 26.

The RecA protein consists of 352 amino acids and has three structural domains: an N-terminal domain (residues 1–33); a central core (residues 34–268); and a C-terminal domain (residues 269–352). Monomer, dimer, and trimer RecA units were identified from the crystal structure of the RecA hexamer (PDB ID: 3CMU). Missing residues in these units were modeled using the Iterative Threading ASSEmbly Refinement (I-TASSER) server (http://zhanglab.ccmb.med.umich.edu/I-TASSER/).[29,30] In this manner, five output structures were created and the one whose root-mean-square deviation (RMSD $\sim$ 0.5 Å) from the corresponding X-ray crystal structure of RecA was lowest, was used in the simulations. To study the assembly mechanism of RecA–ssDNA, simulations were started from dissociated RecA and ssDNA. We adopted the same procedure as described by Mishra and Levy in their work[26] for the prediction purpose in each system (RecA-monomer, -dimer and -trimer), that is, the corresponding free ssDNA molecules (4, 8, and 11 NT long, respectively) were placed randomly in different positions all around the protein.

Simulation protocol

In the RecA crystal structure, the RecA monomers were assembled along a crystallographic 6_1-screw axis to form a helical filament structure with $\sim$6 monomers per helical turn.[5] ATP analogues (ADP–aluminum fluoride–Mg, ADP–ALF–Mg, a non-hydrolysable ATP analogue that mimics the transition state of ATP hydrolysis) are present at the interfaces between two RecA subunits.[5] To incorporate the effect of the ATP analog, positively charged C_β beads lying within 5 Å of any atom of the ATP analog or of another monomeric unit of RecA are neutralized to zero charges (Lys72 and Lys599 for RecA(dimer) and also Lys421 and Lys948 for RecA(trimer)) since these residues are not accessible for interaction with the incoming ssDNA strand. Apart from this, the loop structure near the ssDNA binding site was rendered

flexible during the simulation by removing the native contact pairs between the beads of loops L1 (residues 156–166) and L2 (residues 197–213). The deleted pairs that correspond to the flexible region were modeled as excluded volume interactions.

The dynamics of the protein and ssDNA were simulated using Langevin dynamics. The RecA–ssDNA complex was studied at a salt concentration of 10 mM using a dielectric constant of 10 at a temperature of 0.10 (arbitrary units). At this temperature, the protein does not undergo unfolding but remains flexible (the folding/unfolding transitions occur at a temperature of about 1.0). We used a lower dielectric constant than normal (70 for water) as well as a relatively lower temperature to achieve conditions in which the bound state of the complex was more populated and thermodynamically more favorable than the dissociated state. Electrostatic interactions play a major role in RecA–ssDNA interactions,[3,5] hence performing simulations at a lower dielectric constant facilitates the electrostatic interactions between the protein and ssDNA beads. For each system, a total of 600 simulations (100 for each of the six initial configurations where the ssDNA faced a different patch of the RecA surface) were performed to achieve extensive sampling of the association mechanisms. Each simulation consisted of 10^7 molecular dynamics steps, each with a time step of 0.005. Finally, the last 20% of each trajectory was used for the analysis in which the conformation was collected every 20 000 steps resulting in 100 conformations from each trajectory. Thus, a total of 60 000 conformations from each of the RecA–ssDNA systems were analyzed.

Structural similarity parameter

The degree of conformational similarity (D) between the simulated RecA–ssDNA complex structure and the crystal form is measured using eqn (1):

$$D\ (\text{Å}) = \frac{1}{N_{\text{protein}} N_{\text{ssDNA}}} \sum_{i}^{N_{\text{protein}}} \sum_{j}^{N_{\text{ssDNA}}} \left(\left| r_{ij} - r_{ij}^{0} \right| \right) \qquad (1)$$

where N_{protein} is the number of C_β beads among the interface residues, that is, among the positively charged or aromatic residues lying within a certain cut-off distance from any N_{ssDNA}, phosphate or base bead. Different cut-off distances of 12 Å, 12 Å, or 10 Å were used for monomer, dimer, and trimer systems, respectively, to maintain a roughly constant number of protein beads (N_{protein}) arranged around the ssDNA. The r_{ij} term is the measured distance between each of the selected protein beads and each of the phosphate and base beads of the model ssDNA and r_{ij}^0 is the experimental value of the same[26] (eqn (1)).

To obtain a better structural classification of the model *vis-à-vis* the experimental structure, D_1 and D_2 values were measured by dividing the protein–ssDNA interfacial C_β beads (both positively charged and aromatic) into two groups—a group closer to the 3′ end and a group closer to the 5′ end—and then calculating the similarity parameters measured between the ssDNA and these two groups. The D_1 and D_2 values were calculated for the RecA-monomer using the group of residues (H163, R169, K216 and F217) and (R196 and K198), respectively. For dimeric RecA, the groups of residues were (H163, R169, R176, K216, F217, Y218 and R525) and (R196, K198, H512, R518, R545, K547, K565 and F566) for D_1 and D_2 calculations, respectively. For trimeric RecA, the groups of residues were (H163, R169, R176, R196, K198, R518, R525 and K565) and (H512, R545, K547, H861, R867, R894 and K896) for D_1 and D_2 calculations, respectively.

Results and discussion

Structural analysis of ssDNA binding at the primary binding site of RecA

To understand the mechanism of nucleation formation and protein–ssDNA interaction, we performed structural analyses of the coarse-grained molecular dynamics (CGMD) simulations and plotted the distribution of the D_1 and D_2 values.

Fig. 2(A–C) presents these data for the three coarse-grained model systems: monomeric RecA, dimeric RecA, and trimeric RecA complexes with ssDNA of length 4 NT, 8 NT, or 11 NT, respectively. The contour maps in Fig. 2(A–C) indicate the probability of the distributions of different configurations of the ssDNA in different regions of the RecA protein, which in turn relates to the free energy of the different binding states of the complex modeled by protein–ssDNA electrostatic and aromatic interactions. The maps of the binding ensembles of RecA with ssDNA illustrate the existence of various binding ssDNA modes that are the outcome of transient interactions

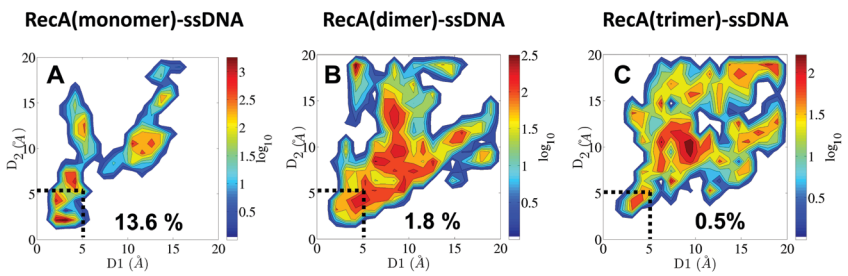

Fig. 2 Energy landscape of RecA–ssDNA binding. Conformational ensemble for complexes between RecA(monomer), RecA(dimer), and RecA(trimer) with ssDNA of length (A) 4 NT, (B) 8 NT, and (C) 11 NT. Population density in regions having low D_1 and D_2 values ($\lesssim$5 Å) corresponds with greater structural similarities between the predicted protein–ssDNA binding interface and the crystal structure (the relative population of that region is shown in %). The color bar corresponds to the free energy which is estimated by $-\log(P)$, where P is the probability of sampling conformations with each value of D_1 and D_2. Greater population (red color) means greater sampling.

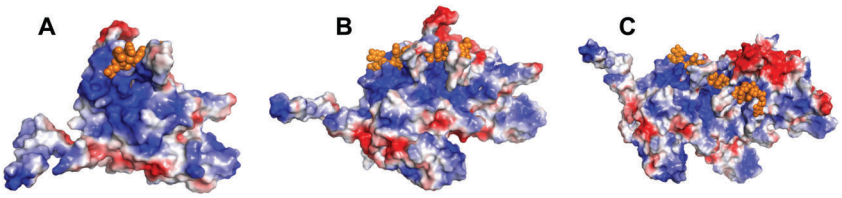

Fig. 3 Electrostatic surfaces of (A) monomeric, (B) dimeric and (C) trimeric RecA. Positively and negatively charged surfaces are shown in blue and red, respectively. ssDNA is shown in orange.

with positively charged residues or aromatic side-chains on the surface of the RecA proteins. The native binding mode is defined by low D values. Here we consider D_1 and D_2 values within 5 Å as native-like assemblies. Fig. 2 suggests that the population of native-like configurations of RecA(dimer)–ssDNA is smaller than that obtained for RecA(monomer)–ssDNA. This is indicated by the smaller sampling of states with low D values for the dimeric RecA than the monomeric RecA. Similarly, the population of native-like configurations is smaller for the RecA(trimer)–ssDNA (Fig. 2C) than for the monomeric (Fig. 2A) and dimeric complexes (Fig. 1B). This could be attributed to the increasing complexity of ssDNA binding at the primary binding site of the RecA protein when moving from monomeric to trimeric RecA.

The force-field used in the CGMD model comprises two attractive energy terms for protein–ssDNA interactions: (i) electrostatic interactions between positively charged C_β protein beads and negatively charged ssDNA phosphate beads; and (ii) aromatic interactions between aromatic C_β protein beads and ssDNA bases. The electrostatic surfaces of the RecA-monomer, -dimer, and -trimer (Fig. 3) show the presence of a positively charged (blue) surface other than the primary binding site of the protein. This may explain the population at high D_1 and D_2 values for the three systems. As the monomers assemble with each other to form a dimer or a trimer, the number of positively charged surfaces on the protein grows and the tendency of the model ssDNA to bind at different surfaces increases, and therefore the population density of the model complex at the correct binding site reduces (low values of D_1 and D_2). Similar D_1 *vs.* D_2 distribution plots were also obtained using the same model for other protein–ssDNA systems.[26]

Energetics of RecA–ssDNA binding

Electrostatic and aromatic interactions are crucial stabilizers of ssDNA and RecA protein binding.[3,5,31] To explore the energy landscape of RecA–ssDNA complexes, the total interaction energy (E_{tot}; the sum of electrostatic and aromatic energies) is plotted against the D order parameter for the three systems (Fig. 4(A–C)).

The existence of several binding modes for ssDNA when interacting with monomeric RecA is seen in the plot of E_{tot} against D for the conformational ensemble of RecA(monomer)–ssDNA (Fig. 4A), which is similar to the heterogeneous assembly seen when plotting D_1 against D_2 (Fig. 2A) for the same ensemble. The diverse binding of ssDNA is also observed for dimeric and trimeric RecA (Fig. 2(B and C)). However, the total energy landscapes for the RecA(dimer)–ssDNA and RecA(trimer)–ssDNA are funnel-like (Fig. 4(B and C)), whereas the corresponding energy landscape for the RecA(monomer) has a non-funneled shape (Fig. 4A). This non-funneled characteristic is illustrated by the relatively high energy of the near native conformations of RecA(monomer)–ssDNA ($D \leq 5$ Å). The total energy of the RecA(monomer)–ssDNA for different values of the D measure suggests that the interaction of ssDNA with monomeric RecA is nonspecific. We may therefore conclude that the monomer is not the elementary unit for nucleoprotein filament formation; rather, dimeric RecA is required to initiate the formation of the RecA–ssDNA assembly.

Separating the total energy of the interface between RecA and the ssDNA into its electrostatic (Fig. 4(D–F)) and aromatic (Fig. 4(G–I)) components shows that the former contributes more to the total energy than the latter (the largest electrostatic and aromatic energies in the native-like assembly are about 3 and 2 kcal mol^{-1} per nucleotide, respectively). The average protein–ssDNA electrostatic interaction computed by the present model for the RecA–ssDNA conformations of the three systems (*i.e.*, monomeric, dimeric and trimeric RecA) having 5 Å or less D values are -1.9 ± 0.3, -1.9 ± 0.4, and -1.8 ± 0.6 kcal mol^{-1} per nucleotide, respectively. These values further confirm that electrostatic interactions play an important role in the total energy, as has been found for other biomolecular recognition processes.[3,5,31] Moreover, the electrostatic contribution to ssDNA–RecA binding does not depend on the ssDNA length, and its strength remains very similar even if the binding results in a non-native assembly.

The role of aromatic residues, and particularly those in loops L1 and L2 near the primary binding site of RecA, is not fully understood. The average aromatic interaction energies for native-like structures of the RecA-monomer, -dimer, and -trimer that were computed using the present model are -0.09 ± 0.2, -1.3 ± 0.3, and -1.6 ± 0.2 kcal mol^{-1} per nucleotide, respectively. These data clearly indicate that increasing the number of RecA units in the association increases the strength of the aromatic interactions (but not of the electrostatic interactions) between ssDNA and RecA and hence improves the nucleating agent for filament formation.

The role of aromatic interactions in the RecA–ssDNA assembly was explored experimentally. Particularly, the role of the Phe203 residue in loop L2 of RecA was studied by site-directed mutagenesis of the other nineteen amino acid residues.[31] The results clearly showed that the activity of wild-type RecA is partially reduced when Phe is replaced with the aromatic residues Tyr

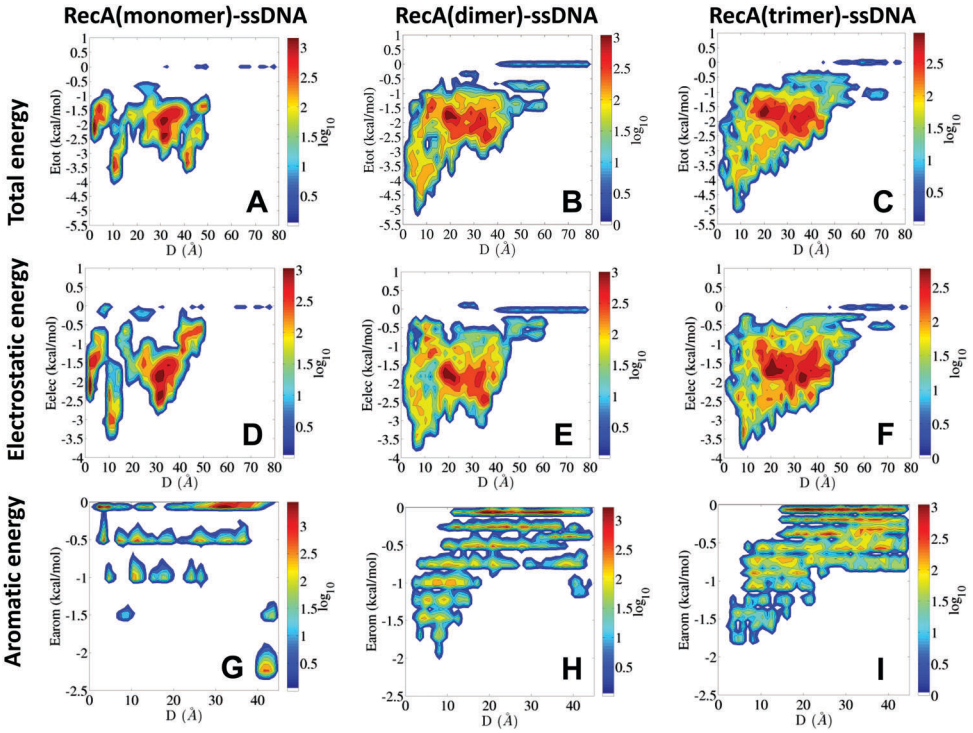

Fig. 4 Comparison of the binding energies per nucleotide for monomeric (left column), dimeric (middle column), and trimeric (right column) RecA. Binding energy per nucleotide is plotted against *D*, with the total, electrostatic, and aromatic energies shown in the top, middle, and bottom rows, respectively. The color bar corresponds to the population of the conformation with values of *D* and binding energy (total energy, electrostatic energy or aromatic energy). The probability, *P*, is presented by calculating −log(*P*), where greater population (red color) means greater sampling.

or Trp, whereas the activities of all other mutant proteins were reduced significantly or completely.[31] This reflects the significance of the aromatic residue Phe203 in the L2 loop region for the activity of RecA. Moreover, fluorescence quenching experiments on F203W mutant RecA–ssDNA/dsDNA complexes also provided evidence of indirect interactions between Phe203 and DNA bases.[32] The structural analysis of the lowest aromatic energy for a model containing a native-like RecA–ssDNA system showed that introducing loop flexibility brings the aromatic residues, His163 in loop L1 and Phe203 in loop L2, closer to the bases of ssDNA. Several other experimental studies showed that the L2 loop region in RecA plays a vital role in ssDNA binding. A 20 AA long peptide derived from the L2 region is capable of strongly binding with both ssDNA and dsDNA.[33] The Phe203 residue present in the L2 loop region is highly conserved and a spectroscopic study also confirmed that the aromatic residue Phe203 is functionally important.[32]

From an experimental study, Wittung *et al.*[34] showed that RecA has a net preference for polynucleotide thymine DNA strands, p(dT), compared with polynucleotide adenine DNA strands, p(dA). The X-ray structure of the RecA–ssDNA complex reveals the presence of aromatic residues in the loop region that lies far from the DNA bases. In contrast, the L1 and L2 loops present in the primary binding site of ssDNA create steric clashes and thus prevent the incoming DNA strand from accessing the protein binding site. Therefore, the movement of loops is essential for the protein–ssDNA binding phenomenon. Using the present method, it was possible to successfully predict the structure of the protein–ssDNA complex only when loop flexibility was introduced into the CGMD model, because loop flexibility brings aromatic residues present in the loop region closer to the DNA bases.

Further examination of the energy landscapes for the electrostatic (Fig. 4D) and aromatic energies (Fig. 4G) of RecA(monomer)–ssDNA against *D* indicates that neither exhibit a funneled shape. However, the corresponding energy landscapes for RecA(dimer)–ssDNA (electrostatic: Fig. 4E; aromatic: Fig. 4H) and RecA(trimer)–ssDNA (electrostatic: Fig. 4F; aromatic: Fig. 4I) are funneled. Although electrostatic energy contributes more to the stability of the protein–ssDNA interface, the funnel shape of the aromatic energy landscape is steeper. This may indicate that aromatic interactions dominate the specificity of the interactions between ssDNA and the RecA.

Role of protein loop flexibility and ATP in ssDNA binding

The central core domain of RecA contains two loops, L1 and L2, which are close to the binding sites of ssDNA. Each loop contains one aromatic residue (His163 in L1 and Phe203 in L2). One of the most interesting aspects of RecA is the conservation of loop regions L1 and L2 in bacterial RecA.[3] From the X-ray structure, it is clear that the side chains of His163 and Phe203 do not directly interact with the ssDNA, but the L1 and L2 regions may clash with an incoming ssDNA. Hence, the movement of loops L1 and L2 is essential when RecA binds ssDNA. A previous all-atom molecular dynamics simulation of the RecA–ssDNA system observed that binding caused a 4–6 Å change in the

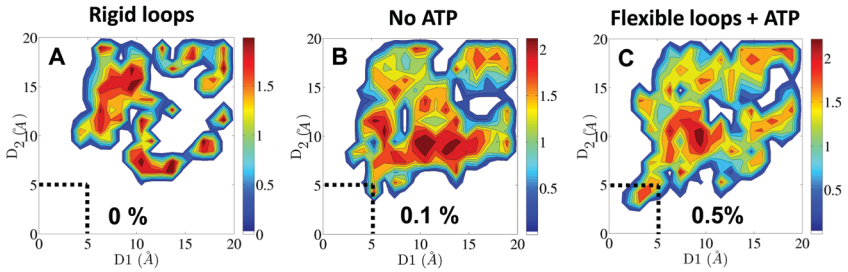

Fig. 5 Predicted conformational ensemble of RecA(trimer)–ssDNA when the mobilities of L1 and L2 are restricted (A), when ATP binding is not included (B), and when both loop flexibility and ATP binding are included (C). The greater population found with low values ($\leq$5 Å) of D_1 and D_2 in (B) and (C) suggests that ssDNA can approach the binding site only when the loops undergo conformational changes that make space for ssDNA to bind with RecA. The relative population of the region with low values of D_1 and D_2 (marked with the dashed lines) is shown in %.

root-mean-square difference of the L1 and L2 loop regions.[35] In the present coarse-grained model, the flexibility of L1 and L2 was enhanced by eliminating a few interactions between them. This enhanced flexibility model successfully predicts RecA–ssDNA binding. The L1 and L2 loops (Fig. 5) might be viewed as serving as a gate that must be open for RecA to bind ssDNA.

To evaluate the importance of loop flexibility, we also studied the interaction between RecA and ssDNA in its absence, and found that loop rigidity reduces the model's predictive power. The D_1 *vs.* D_2 plot of the RecA-trimer (rigid loop) binding ssDNA (Fig. 5A) indicates that it fails to capture the native structure. The lowest D value is increased from 3.3 Å to 6.9 Å for the trimeric RecA–ssDNA complex. Therefore, in order to bind ssDNA to the primary binding site of RecA, loop flexibility is mandatory.

In addition to flexibility, ATP binding is essential to the RecA–ssDNA binding process. In the crystal structure, instead of ATP, an analogous mixture of ADP–ALF4–Mg^{2+} was used to mimic the transition state of ATP hydrolysis. The ATP binding site is located at the interface between RecA monomers. We examined the effects of ATP on the binding of ssDNA to RecA(trimer) by considering the two binding pockets for the ATP analogs at the interface of neighboring RecA monomers. The effect of ATP binding was effectively modeled by neutralizing the positively charged residues within 5 Å of any atom of the ATP analog (see the Methods section). In the absence of the ATP analogs, there is a decrease in the population of native-like conformations of RecA-trimer–ssDNA. This is seen in Fig. 5B, which shows negligible populations of configurations with low values of D_1 and D_2 ($\leq$5 Å). This result supports the experimental findings of low ssDNA affinity for RecA in the absence of ATP and thus the importance of ATP for the assembly of ssDNA with RecA (Fig. 5C). The simple model of ATP binding may suggest that the ATP guides ssDNA by changing the electrostatic potential of the RecA protein.

Conclusions

The mechanism for the nucleation process in RecA–ssDNA filament formation was studied here using a coarse-grained model for protein interactions with ssDNA. The interactions between RecA and the ssDNA were modeled by electrostatic interactions between charged amino acids and the phosphate groups of the DNA and by aromatic interactions between aromatic residue side chains and DNA bases. The simulations illustrated that the surface of RecA is frustrated and the ssDNA can interact with different sites on the surface. The degree of molecular frustration is smaller for dimeric and trimeric RecA than for their monomeric counterpart. Additionally, the model shows that loop flexibility is essential for an incoming ssDNA strand to bind with RecA. Starting from the unbound structure, only a few ssDNA were able to bind to the RecA-dimer or -trimer when the loop residues were not flexible. However, the ability to bind RecA improved dramatically when the loop residues were allowed to move, thus indicating the importance of the role of the loop flexibility in ssDNA binding. Aromatic residues in the loop region play a crucial role in ssDNA binding. A non-funneled shape energy landscape of RecA-monomer suggests the non-specificity of ssDNA binding, and therefore the RecA-monomer is unlikely to serve as an elementary unit for a nucleoprotein filament. The funneled shape energy landscape for either dimeric or trimeric RecA indicates that dimeric RecA is required to initiate the formation of the RecA–ssDNA assembly. Finally, the ssDNA affinity for RecA was found to be low in the absence of ATP and it may be that the ATP guides ssDNA by changing the electrostatic potential of the RecA protein.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the Israel Science Foundation and by the Minerva Foundation, with funding from the Federal German Ministry for Education and Research. YL is the Morton and Gladys Pickman professional chair in Structural Biology.

References

1 A. Saladin, C. Amourda, P. Poulain, N. Ferey, M. Baaden, M. Zacharias, O. Delalande and C. Prevost, *Nucleic Acids Res.*, 2010, **38**, 6313–6323.

2 D. R. Yang, B. Boyer, C. Prevost, C. Danilowicz and M. Prentiss, *Nucleic Acids Res.*, 2015, **43**, 10251–10263.
3 I. P. Bugreeva, D. V. Bugreev and G. A. Nevinsky, *FEBS J.*, 2005, **272**, 2734–2745.
4 C. Danilowicz, A. Peacock-Villada, J. Vlassakis, A. Facon, E. Feinstein, N. Kleckner and M. Prentiss, *Nucleic Acids Res.*, 2014, **42**, 526–533.
5 Z. Chen, H. Yang and N. P. Pavletich, *Nature*, 2008, **453**, 489–494.
6 M. Prentiss, C. Prevost and C. Danilowitz, *Crit. Rev. Biochem. Mol. Biol.*, 2015, **50**, 453–476.
7 R. Masui, T. Mikawa, R. Kato and S. Kuramitsu, *Biochemistry*, 1998, **37**, 14788–14797.
8 A. L. Forget, M. M. Kudron, D. A. McGrew, M. A. Calmann, C. A. Schiffer and K. L. Knight, *Biochemistry*, 2006, **45**, 13537–13542.
9 R. Roy, T. M. L. A. G. Kozlov and T. Ha, *Nature*, 2009, **461**, 1092–1097.
10 S. H. Kim, K. Ragunathan, J. Park, C. Joo, D. Kim and T. Ha, *J. Am. Chem. Soc.*, 2014, **136**, 14796–14800.
11 G. P. Manjunath, N. Soni, P. L. Vaddavalli, D. J. Shewale, M. S. Madhusudhan and K. Muniyappa, *Biochemistry*, 2016, **55**, 1850–1862.
12 J. C. Bell, J. L. Plank, C. C. Dombrowski and S. C. Kowalczykowski, *Nature*, 2012, **491**, 274–278.
13 C. Joo, S. A. McKinney, M. Nakamura, I. Rasnik, S. Myong and T. Ha, *Cell*, 2006, **126**, 515–527.
14 S. L. Lusetti and M. M. Cox, *Annu. Rev. Biochem.*, 2002, **71**, 71–100.
15 A. J. Conover, C. Danilowicz, R. Gunaratne, V. W. Coljee, N. Kleckner and M. Prentiss, *Nucleic Acids Res.*, 2011, **39**, 8833–8843.
16 M. T. J. van Loenhout, T. van der Heijden, R. Kanaar, C. Wyman and C. Dekker, *Nucleic Acids Res.*, 2009, **37**, 4089–4099.
17 C. Prevost and M. Takahashi, *Q. Rev. Biophys.*, 2003, **36**, 429–453.
18 M. M. Cox, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 127–138.
19 G. Bertucat, R. Lavery and C. Prevost, *Biophys. J.*, 1999, **77**, 1562–1576.
20 D. Yang, B. Boyer, C. Prevost, C. Danilowicz and M. Prentiss, *Nucleic Acids Res.*, 2015, **43**, 10251–10263.
21 A. Peacock-Villada, D. Yang, C. Danilowicz, E. Feinstein, N. Pollock, S. McShan, V. Coljee and M. Prentiss, *Nucleic Acids Res.*, 2012, **40**, 10441–10451.
22 A. V. Mazin and S. C. Kowalczykowski, *EMBO J.*, 1998, **17**, 1161–1168.
23 D. U. Ferreiro, E. A. Komives and P. G. Wolynes, *Q. Rev. Biophys.*, 2014, **47**, 285–363.
24 J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, *Proteins: Struct., Funct., Genet.*, 1995, **21**, 167–195.
25 J. N. Onuchic and P. G. Wolynes, *Curr. Opin. Struct. Biol.*, 2004, **14**, 70–75.
26 G. Mishra and Y. Levy, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 5033–5038.
27 O. Givaty and Y. Levy, *J. Mol. Biol.*, 2009, **385**, 1087–1097.
28 Y. Levy, A. Caflisch, J. N. Onuchic and P. G. Wolynes, *J. Mol. Biol.*, 2004, **340**, 67–79.
29 J. Y. Yang, R. X. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang, *Nat. Methods*, 2015, **12**, 7–8.
30 A. Roy, A. Kucukural and Y. Zhang, *Nat. Protoc.*, 2010, **5**, 725–738.
31 K. Hortnagel, O. N. Voloshin, H. H. Kinal, N. Ma, C. Schaffer-Judge and R. D. Camerini-Otero, *J. Mol. Biol.*, 1999, **286**, 1097–1106.
32 F. Maraboeuf, O. Voloshin, R. D. CameriniOtero and M. Takahashi, *J. Biol. Chem.*, 1995, **270**, 30927–30932.
33 L. J. Wang, O. N. Voloshin, A. Stasiak, A. Stasiak and R. D. Camerini-Otero, *J. Mol. Biol.*, 1998, **277**, 1–11.
34 P. Wittung, C. Ellouze, F. Maraboeuf, M. Takahashi and B. Norden, *Eur. J. Biochem.*, 1997, **245**, 715–719.
35 C. Carra and F. A. Cucinotta, *J. Mol. Model.*, 2011, **17**, 133–150.