BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

Molecular mechanism underlying ATP-induced conformational changes in nucleoprotein filament of Mycobacterium smegmatis RecA

G P Manjunath, Neelesh Soni, Pavana L Vaddavalli,
Dipeshwari J Shewale, M. S. Madhusudhan, and K Muniyappa

Biochemistry, **Just Accepted Manuscript** • DOI: 10.1021/acs.biochem.5b01383 • Publication Date (Web): 25 Feb 2016

ACS Publications

Biochemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Molecular mechanism underlying ATP-induced conformational changes in nucleoprotein filament of *Mycobacterium smegmatis* RecA

G P Manjunath[1, 2*], Neelesh Soni[3], Pavana L Vaddavalli[2], Dipeshwari J Shewale[2], M S Madhusudhan[3] and K Muniyappa[1]

1. Department of Biochemistry, Indian Institute of Science (IISc), Bangalore-560012, India

2. Center of Excellence in Epigenetics, Indian Institute of Science Education and Research (IISER), Pune-411008, India

3. Department of Biological Sciences, Indian Institute of Science Education and Research (IISER), Pune-411008, India

** To whom correspondence should be addressed*

Address for Correspondence:

Indian Institute of Science Education and Research (IISER),

Homi Bhabha Road, Pashan, Pune-411008; India

Tel: +91-20-25908194; Fax: +91-20-25899790;

Email: manjunathgp@iiserpune.ac.in

Funding information: GPM was supported by a research fellowship awarded by IISc, Bangalore and subsequently by the Center of Excellence in Epigenetics (CoEE) grant awarded by Department of Biotechnology, New Delhi to Prof Sanjeev Galande. Research in KM lab is supported by Center of Excellence and Innovation Program awarded by Department of Biotechnology, New Delhi.

Keywords: Homologous Recombination, Allostery, DNA Strand Exchange, ATP hydrolysis, DNA binding

Abbreviations: ATP -Adenosine triphosphate, SPR-Surface plasmon resonance, EMSA-Electrophoretic mobility shift assay, CD-Circular dichroism, EM-Electron microscopy

Abstract

RecA plays a central role in bacterial DNA repair, homologous recombination and restoration of stalled replication forks by virtue of its active extended nucleoprotein filament. Binding of ATP and its subsequent recognition by the carboxamide group of a highly conserved Glutamine (Gln196 in MsRecA) has been implicated in the formation of active RecA nucleoprotein filaments. Although, the mechanism of ATP-dependent structural transitions in RecA has been proposed based on low-resolution electron microscopic reconstructions, the precise sequence of events that constitute these transitions were poorly explored. Based on biochemical and crystallographic analyses of MsRecA variants carrying mutations in highly conserved Gln196 and Arg198 residues, we propose that the disposition of the inter-protomer interface is the structural basis of allosteric activation of RecA. Furthermore, this study accounts for the contributions of several-conserved amino acids to ATP hydrolysis and to the transition from collapsed to extended filament forms in *Mycobacterium smegmatis* RecA (MsRecA). In addition

to their role in the inactive compressed state, the study reveals a role for Gln196 and Arg198 along with Phe219 during ATP hydrolysis in the active extended nucleoprotein filament. Finally, our data suggests that the primary, but not the secondary nucleotide-binding site in MsRecA isomerizes into the ATP binding site present in the extended nucleoprotein filament.

Escherichia coli RecA (EcRecA) is the founding member of a growing class of proteins that couple energy from ATP hydrolysis to perform mechanical functions [1-3]. RecA plays important roles in DNA repair, homologous recombination and restoration of stalled replication forks by virtue of biochemical activities attributed to the catalytically active, extended RecA nucleoprotein filaments. These filaments are formed by the polymerization of RecA monomers on single stranded DNA (ssDNA) in the presence of ATP [4-6]. The right-handed helical filaments have one monomer of RecA bound to three nucleotides of ssDNA or three base pairs of double stranded DNA (dsDNA), with 6 monomers of RecA per helical turn. The extended active filaments have a helical pitch of ~ 95Å [7-9] whereas the inactive compressed filaments, formed in the presence of ADP or by RecA alone, with or without DNA, and have a helical pitch of ~ 65-75 Å [7].

The conversion of the inactive compressed RecA filament to an active extended one is an essential regulatory step in the context of the RecA catalytic cycle. It is initiated by the binding of ATP in the primary ATP binding pocket of RecA located in the central catalytic domain[10-14]. This domain consists of three invariant structural elements; the P loop (residues 66-73), the walker B motif (residues 140-144), a single residue Glu96 and two residues, Gln194 and Arg196[1], that make hydrogen bonds with the γ-phosphate of the bound ATP (Supplementary figure 1) [15]. Several studies, both biochemical and those based on crystal structures of *E. coli* and mycobacterial RecA proteins, implicate Gln194/196 and Arg196/198 as sensors of ATP bound in the catalytic core of RecA [15-20]. This inference is supported by the high degree of conservation evident at this site [21]. Crystal structure of *Mycobacterium smegmatis* RecA (MsRecA) in

[1] Amino acid notation in Ec and MsRecA differ due to presence of two extra amino acids in N-terminal of MsRecA. Henceforth, the first number refers to Ec and the second to MsRecA notation.

complex with dATP revealed that Gln194/196 makes contact with γ-phosphate of dATP using its carboxamide group [18]. Superimposition of the apo and the dATP bound forms of MsRecA showed that the α-carbon of Gln194/196 deviates by 1Å or more [18, 20]. Quantum mechanical calculations indicate that Glu194/196 is in close proximity to the β-phosphate of the ADP resulting from hydrolysis of ATP in the extended RecA filament [22], indicating that these residues play important roles in the extended filament conformation as well.

To investigate the role of Gln194/196 and Arg196/198, we generated MsRecA variants carrying mutations in these highly conserved residues. These variants contained substitutions of Gln196 by Alanine ($MsRecA^{Q196A}$), Glutamic acid ($MsRecA^{Q196E}$) or Asparagine ($MsRecA^{Q196N}$), and Arg198 by Lysine ($MsRecA^{R198K}$). Biochemical characterization of these mutant proteins showed that while these mutations had no measurable effect on ATP-binding; they were defective in activation of ATPase activity by ssDNA and high concentrations of sodium chloride. Mutant variants of MsRecA were able to catalyze D-loop formation to a limited extent, indicating that they were competent in binding single and double stranded DNA. However, they were unable to catalyze strand exchange between single and double-stranded DNA. Crystallographic analysis and molecular modeling of MsRecA protein and its variants in the presence of ATP or its analogues revealed that these mutations caused defects in several molecular steps that follow ATP binding in MsRecA.

Materials and methods

Materials

Oligonucleotides were purchased from SIGMA-GENOSYS India. DNA polymerase, restriction endonucleases and T4 polynucleotide kinase were purchased from New England Bio labs. Radio labeled ATP was purchased from NEN Life science products and PEI cellulose

chromatography plates were obtained from MERCK. ATPγS and ATP were purchased from Merck and GE Life Sciences, respectively. M13 bacteriophage DNA was prepared in the laboratory according to published protocols [23].

Site-directed mutagenesis

Site-directed mutations in MsRecA were made using the sense-antisense method of mutagenesis [24]. Briefly, *M. smegmatis recA* gene was amplified using the primers spanning the nucleotides coding for both the Gln196 or Arg198 residues and a sequence 3' downstream. The primers annealed at the 3' end contained mismatches either at Gln196 or Arg198. The second primer (Mut. primer) contains translationally silent mutations that introduce a Hind III site (Supplementary table 1). The identity of the constructs was confirmed by DNA sequencing.

Protein purification, crystallization and structure refinement

M. smegmatis SSB and *M. tuberculosis* LexA were purified according to published protocols [25, 26]. *M. smegmatis* RecA and its mutant forms were purified to near homogeneity essentially as described earlier [27]. Detailed protocol for purification of MsRecA and its variants is included in the accompanying supporting information. Protein preparations contained no detectable exo- or endonuclease activities. Concentration of the proteins was determined using dye binding method [28]. Crystallization and structure refinement of MsRecA and its variants as well as their complexes with ATP/ATP analogues has been described in an earlier publication [20]. Details of the PDB structures used in this study are described in Table 1.

Generation of MsRecA presynaptic filament model

A homology model for the structure of presynaptic filament of MsRecA structure was constructed using the Modeller v9.14 suite of programs [29-31]. The structure of *E. coli* RecA presynaptic filament was used as the template (PDB code: 3CMT). Target template alignment

was performed using Clustal Omega server [32] with default parameters. Ten models were constructed utilizing the Modeller "refine very slow" option. The best scoring model was selected based on the DOPE potential [33]. Initially, models of RecA dimers were constructed with five monomers and subsequent addition of one protomer to complete a turn of the filament using symmetry considerations. The PDB file for the model of presynaptic MsRecA filament has been uploaded as part of the supplementary information.

Molecular modeling of intermediate structures between collapsed and extended RecA filament.

Intermediate structures between compressed and extended RecA filaments were generated using the FATCAT server [34-36] by linear interpolation. The intermediate structures were studied using the *E. coli* RecA structures alone, as residues that are important in the transition from compressed to extended state are nearly identical in both proteins (Supplementary figure 1). Dimers of EcRecA in the compressed (extracted from PDB: 1XMS) and the extended (extracted from PDB: 3CMT) states were used as input structures to the FATCAT server. All other parameters of the FATCAT server remained unaltered from their default values. The FATCAT server generates intermediate structures with Cα atomic positions. An in-house script that uses rigid structure superimpositions was used to complete these intermediate structures with explicit coordinates of all atoms in the protein.

Circular dichroism spectroscopy

The CD spectra of MsRecA and its variant forms were measured in the far UV region from 260 to 200 nm at 25°C using the CD spectrophotometer (model J750, Jasco, Tokyo Japan). Reaction mixture containing 25 mM Tris-HCl (pH 7.0), 1 mM DTT and 100 mM KCl and 5 μM MsRecA or its variant forms were pre incubated with varying concentrations of ATPγS at 37°C for 10 min

in a reaction volume of 200 µL. The spectra were recorded using a quartz cuvette with constant stirring; a 1 mm slit and a 5 nm/min scan speed. The spectra were averaged over three scans and the elipticity was reproducible within an error of ± 5%.

Surface Plasmon Resonance (SPR) measurements

SPR measurements were performed using BIAcore 2000 (GE biosciences). The 60-mer oligonucleotide containing 3' terminal biotin was immobilized in the flow cells of a streptavidin-coated chip with flow cell number 1 serving as control (Supplementary table 1). Usually, 1100 relative resonance units (RU) were immobilized. Various concentrations of MsRecA or its variant forms in binding buffer [20 mM Hepes (pH 7.0), 12 mM $MgCl_2$, 1.4 mM DTT, 0.1mM ATPγS and 50 mM NaCl] were passed across the surface for 5 min at a flow rate of 5 µL/min. The surface was regenerated using solutions containing 0.1% SDS, followed by 100 mM NaCl. Each protein concentration was injected thrice into the chip to ensure reproducibility. Affinity and kinetic analysis was performed using BIAevaluation software, ver. 3.0. The experiment was repeated twice to reduce noise.

D-loop assay

The formation of D-loops by MsRecA or its variant forms was performed as described [37]. Reaction mixtures (20 µL) containing 33 mM Tris-HCl (pH 7.0), 12 mM $MgCl_2$, 0.1 mM ATPγS and 1 µM (nucleotides) of an 83-mer oligonucleotide complimentary to M13 phage DNA-ODN1 (Supplementary Table 1) was incubated at 37°C for 5 min with increasing concentrations of MsRecA or its mutant forms. Oligonucleotide ODN1 was radiolabelled at its 5' end using polynucleotide kinase (PNK). The formation of D-loops was initiated by the addition of 40µM M13 replicative form I DNA, and the reaction mixture was incubated at 37°C for 5 minutes. The reaction was terminated by the addition of 2% SDS and deproteinised with 0.2 mg/ml proteinase

K. The samples were subjected to electrophoresis in a 0.8% agarose gel in 89 mM Tris/borate buffer (pH 8.3) at 2V/cm for 4 hours. The gels were treated with a 7% tricholoroacetic acid solution for 15 min, dried onto a 3 mm filter paper (Whatman™) and exposed to phosphorimager plates.

Three-strand exchange assay

The reaction mixture (20 µL) contained 20 mM Tris-HCl, (pH 7.0), 1 mM DTT, 0.1 mM ATP/ATPγS and 10 mM $MgCl_2$ at 37°C. The 83-mer ssDNA substrate ODN1 (Supplementary table 1) was pre-incubated with MsRecA, or its variant forms, in the presence of ATPγS or ATP for 10 min at 37°C. The reactions were initiated by the addition of the ^{32}P-labeled 83-mer dsDNA formed by annealing ODN1 with its complimentary sequence ODN2. The reactions were terminated at the indicated time intervals with the addition of EDTA and SDS to final concentrations of 20 mM and 1%, respectively. 5µL gel loading dye (30% Glycerol, 0.08% bromophenol blue, 0.08% xylene cyanol FF) was added to each reaction. This was followed by electrophoresis on 10% polyacrylamide gel in 13.2 mM Tris-acetate buffer (pH 7.4) at 12 V/cm for 6 hours at 4°C. The gels were dried onto a 3 mm filter paper (Whatman™), exposed to phosphorimager plates and visualized by autoradiography.

RecA mediated LexA cleavage

RecA mediated cleavage of LexA in the presence of ATPγS and ssDNA was monitored *in vitro*. Briefly, 2.5µM MsRecA or its mutant variants were incubated for 20 minutes at 37°C, in a buffer containing 20mM Tris-HCl (pH 7.0), 2mM $MgCl_2$, 1mM DTT and where indicated, either 0.1mM ATPγS alone or in the presence 30µM M13ssDNA. 10 µM MtLexA was incubated in 20mM Tris-HCl (pH 7.0), 5mM $MgCl_2$, 1mM DTT and 0.1mM ATPγS at 37°C for 20 minutes. The reaction was initiated by mixing the two samples and the mixture was incubated at 37°C for

60 minutes. The reaction was terminated by the addition of SDS-PAGE loading dye. LexA cleavage was monitored on a 15% SDS PAGE and visualized by silver staining.

Results

Gln196 and Arg198 are essential for ATP hydrolysis, but not ATP binding.

Photo crosslinking by exposure to ultra violet (UV) light has been used to quantify nucleotide binding affinities of several proteins including RecA [16]. Quantification of radiolabelled ATP bound by MsRecA and its variants showed that the same amount of ATP was cross-linked to each of the proteins (Supplementary figure 3A). We have previously shown the presence of dATP in crystal structures of MsRecA and its variants [20].

Kinetics of ATP hydrolysis by $MsRecA^{wt}$ and its mutant proteins was assayed by measuring the release of inorganic phosphate from ATP in the absence of any cofactor (Supplementary figure 3B), presence of ssDNA (Supplementary figure 3C) or 1.8 M NaCl (Supplementary figure 3D) [38]. Consistent with a previous report [16], mutant MsRecA proteins showed no increase in ATPase activity in the presence of ssDNA or 1.8M NaCl, indicating that the functional role of Gln194/196 and Arg196/198 is conserved among bacterial RecA proteins.

Carboxamide group at position 196 is essential for structural transitions associated with ATP binding in MsRecA

The amino acid sequence of loop L2 is highly conserved in all eubacterial RecA proteins with several amino acids displaying near complete conservation (Figure 1A). This region of RecA has been shown to undergo conformation changes in response to ATP binding [39]. Therefore, we followed conformation changes in MsRecA and its variants in the presence of ATPγS using circular dichroism spectroscopy. This method that has been used to study conformation changes in RecA upon binding of DNA [40, 41] as well as that of DNA binding loop L2 upon binding ATP

or its analogues [39]. The CD spectrum of MsRecA shows a dichroic minimum at 209 and 220 nm (Figure 1C). Nearly identical spectra were obtained for MsRecA variant proteins, indicating that the mutant proteins have no gross structural abnormalities as a result of the substitutions (Figure 1D-E). The CD profile of $MsRecA^{wt}$ showed secondary structure transitions associated with ATPγS binding, centered at around 220 nm (Figure 1C), similar to those displayed by isolated loop L2 in the presence of ATP and its analogues [39]. On the other hand, MsRecA variant proteins were unable to display significant spectral changes in the presence of ATPγS (Figure 1D-G), indicating that nucleotide binding is not accompanied by changes in secondary structure of the mutant variants. While, we do not rule out the possibility that changes in the CD spectra upon addition of ATPγS may be an indirect result, when considered in the context of isolated loop L2 acquiring an extended β-sheet in the presence of ATPγS [39], it is likely that these transitions are contributed, at least in part, by changes in the secondary structure of loop L2 (Figure 1C).

Subsequent to/as a result of the changes in the secondary structure of loop L2; a conserved Phenylalanine (Phe217/219) is postulated to interact with a hydrophobic pocket in the neighboring protomer [11], thereby transmitting the information about ATP binding in the central core of RecA to the inter-protomer interface (Figure 2A-C). Upon inspection of the crystal structure of MsRecA bound to dATP, we observed a small but clear change in the orientation of α-helix 8 as well as the aromatic ring of Phe219 (Figure 2D). Closer inspection of the immediate environment surrounding Phe219 revealed that the observed rotation of 20 degrees is the maximum rotation possible at this position due to steric considerations. In contrast, Tyr218/220 showed no change and was completely superimposable in the apo (cyan) and ATP (brown) bound states, indicating that these changes were specific to Phe217/219. On the other hand we observed no change upon dATP binding in crystal structures of mutant variants of MsRecA,

indicating that this movement of Phe217/219 is tightly correlated with the interaction of the carboxamide group at position 194/196 with the γ-phosphate of dATP (Compare figure 2D with 2E-G). These results support the inference that the interaction of Phe217/219 with the hydrophobic pocket in the neighboring protomer requires the presence of both ATP and a carboxamide group at position 194/196.

Gln194/196, Arg196/198 and Phe217/219 regulate ATP hydrolysis in the extended nucleoprotein filament.

In addition to communicating the presence of ATP in the nucleotide-binding pocket, the crystal structure of presynaptic EcRecA as well as our model of the MsRecA presynaptic filament suggests that Gln194/196, Arg196/198 and Phe217/219 participate directly during ATP hydrolysis in the presynaptic filament (Figure 3). These interactions between Gln194/196, Arg196/198 and Phe217/219 as well as the ATP sandwiched between two RecA protomers are mediated by cation-pi and anion-pi interactions, distinct from the hydrogen bond network seen in the compressed state of RecA involving the same interacting partners. We postulate that the transition of RecA filament into an extended conformation causes Phe217/219 to leave the hydrophobic environment created by Thr150/152 and Ile155/157 in the neighboring protomer. Presence of ATP creates a negative environment in the vicinity of Phe219, which attracts the Gln196 and Arg198 side chains from the neighboring protomer. This movement induces contact between Gln194/196 and Phe217/219, which in turn causes the Phe217/219 ring to form a more stable cation-pi interaction, bringing the edge of the Phe217/219 ring towards the Glu96/98 (Figure 3A). The partial positive charge attracts the Glu96/98, forming anion-pi interaction while simultaneously forming anion-pi interaction with γ-phosphate of the bound ATP (Figure 3B). Phe217/219 restrains Glu96/98 and γ-phosphate of the ATP in an optimal position for a

nucleophilic attack by Glu96/98 leading to ATP hydrolysis. These results could potentially explain the relatively faster kinetics of ATP hydrolysis in the presynaptic filament as the absence of such interactions in the compressed state implies that the positioning of Glu96/98 is a function of thermal fluctuations and as such inefficient, accounting for slower kinetics of ATP hydrolysis. Our model suggests that these molecular events are unlikely to occur in MsRecA variants (Figure 3C) hinting at a role for Gln194/196 and Arg196/198 during ATP hydrolysis in the presynaptic RecA filament. This conclusion is supported by the high degree of conservation observed in side chains that make contact with Gln194/196 and Arg196/198 in the presynaptic filament, as any change in their disposition is likely to result in disruption of these interactions (Table 2) and consequently affect RecA functions.

Mutations at Gln196 and Arg198 affect stability of RecA-ssDNA nucleoprotein complexes.

The ATP-dependent ssDNA-binding activities of MsRecA and its mutant variants were assessed using nitrocellulose filter binding assay, electrophoretic mobility shift assay (EMSA) and surface plasmon resonance (SPR) measurements. Consistent with the previous studies, $MsRecA^{wt}$ displayed stable ssDNA-binding activity, yielding a binding curve with an endpoint corresponding to a stoichiometric ratio of 3 nucleotides per RecA monomer in both mobility shift and nitrocellulose filter binding assay (Supplementary figure 4). In the same assay, ssDNA-binding activity of Gln196 and Arg198 mutants was much weaker, compared to the $MsRecA^{wt}$. While $MsRecA^{Q196A}$ and $MsRecA^{Q196E}$ failed to bind ssDNA, we were able to detect formation of nucleoprotein complex by $MsRecA^{Q196N}$ that disintegrated while being electrophoresed through polyacrylamide gel (Supplementary figure 4B). These results suggest that, $MsRecA^{Q196N}$ was capable of interacting with ssDNA, however the resulting nucleoprotein filaments were unstable and therefore, not detected in nitrocellulose filter binding or mobility shift assays.

To explore this possibility further, SPR measurements were carried out (Figure 4). The resonance units increased rapidly with the flow of MsRecA^{wt} over the chip, and reached a plateau at 1 μM followed by a gradual decrease indicating dissociation of the complex upon injection of the buffer at 120 sec (Figure 4A). Next, MsRecA mutant proteins were individually applied and the changes in the resonance units were recorded. We observed that with the exception of MsRecA^{Q196E} (Figure 4E), the mutant proteins were able to interact with ssDNA (Figure 4B-D). In comparison with MsRecA^{wt}, the kinetics of ssDNA-binding by the variant proteins was relatively slower, although the stoichiometric ratios of binding remained 3-nucleotide per molecule of RecA. The profile of the sensogram was comparable for all proteins that bound ssDNA, at sub saturating concentrations of RecA. MsRecA^{wt} saturated the immobilized ssDNA at high concentrations (500 and 1000nM) as evident by the stable resonance units. On the other hand, MsRecA variants were unable to achieve stable resonance, indicating that a stable complex formation was not achieved by these variants.

The poor stability of these protein-DNA complexes prevented us from exploring their conformation further experimentally. We, therefore, employed molecular modeling to understand the differences in ssDNA binding by MsRecA and its variants (Figure 5). Comparison of the compressed and the presynaptic filament of MsRecA revealed that the residues that make salt bridges with the phosphate backbone of ssDNA are not affected by the presence of ATP in the primary nucleotide-binding site (Figure 5A). Furthermore, the positively charged groove present on the inner surface of MsRecA filament remains unaltered by the binding of ATP, accounting for ssDNA binding by both MsRecA^{wt} and its variants that fail to transition into extended conformation (Figure 5B). However, comparative energetics of the protomer interfaces (Supplementary table 2) indicates that the protomer interface is relatively more stable in the

extended conformation of MsRecA^{wt}. Furthermore, the inter protomer interface is stabilized by five additional H-bonds in the extended filament conformation (Figure 5C) (Supplementary tables 3 and 4), resulting in a greatly stabilized association between adjacent protomers. The resulting cooperative binding of MsRecA protein on ssDNA accounts for the stable nucleoprotein filament formation as opposed to non-cooperative interaction observed with the mutant variants (Figure 4).

MsRecA variants are capable of initiating three-strand exchange but fail to take the process to completion

Since the mutant variants of MsRecA were capable of binding ssDNA, we wanted to investigate if they can initiate strand exchange. We employed D-loop assay for this purpose, as D-loop is the first heteroduplex DNA intermediate generated between two recombining DNA molecules. Furthermore, D-loop formation by RecA requires ATP binding but not hydrolysis [42]. While EcRecA is able to incorporate up to 50% of labeled ssDNA into D loops [43], mycobacterial RecA proteins are relatively inefficient in their ability to promote D loop formation [27]. MsRecA^{wt} displayed ability to catalyze D loop formation consistent with published reports [27]. Surprisingly, however, mutant MsRecA proteins were also able to catalyze the formation of D-loops, albeit at a much reduced level (Figure 6). We ruled out the possibility that this limited activity may be a result of EcRecA that may have co purified with mutant variants of MsRecA, by expressing these variants in *E. coli* strain JC10289 {Δ (*srl-recA*) *306::Tn10*}. However, it is pertinent to note that reactions containing MsRecA^{Q196E} that failed to bind ssDNA in all three assays, also showed the formation of D-loops. It is, therefore, plausible that the mechanism by which MsRecA variants promote D-loop formation may be distinct from that employed by MsRecA^{wt} (Figure 6 B-D). Although these results indicate that variants of MsRecA are capable of

catalyzing limited D-loop formation, the mechanism by which these D-loops are generated and their parity or lack thereof, with D-loops generated during the course of strand exchange by MsRecA^{wt}, warrants further investigation.

We next tested the ability of mutant MsRecA proteins to catalyze strand exchange between ssDNA and linear duplex DNA (Figure 7). DNA substrates with a relatively shorter length and a low probability of forming secondary structures were used to exclude the possibility that RecA nucleation is interrupted due to the presence of such structures [44]. RecA nucleoprotein filament was assembled on ssDNA (83-mer; ODN1) with stoichiometric amounts of RecA protein (Figure 7A). To prevent dissociation of the nucleoprotein filament of RecA-ssDNA, a poorly hydrolyzed analog of ATP, ATPγS was also used. After the addition of homologous 83 bp dsDNA, the reactions were carried out for the time intervals as indicated at the top of each panel. MsRecA^{wt} was able to catalyze complete strand transfer with fast kinetics of heteroduplex formation in the presence of ATP and ATPγS (Figure 7B second row). In contrast, the mutant MsRecA proteins failed to catalyze strand exchange in the presence of ATP or ATPγS, indicating that transition to an extended filament conformation is essential for strand exchange even over short DNA fragments (Figure 7B, rows 3,4 and 5).

We further confirmed the inability of the mutant MsRecA variants to acquire the extended active nucleoprotein conformation using co-protease assay. The presence of the extended filament conformation has been shown to be essential for co-protease activity of RecA [45, 46]. While the MsRecA^{wt} protein promoted cleavage of LexA repressor (Figure 8, lane 4) in the presence of both ATPγS and ssDNA, none of the variants were able to do so (Figure 8, lanes 5-16). These results are consistent with the notion that the nucleoprotein filaments formed by MsRecA in the absence of ATP or by variants of MsRecA are structurally distinct from those

formed by MsRecA^{wt} in the presence of ATP, and are likely to be closer to the inactive and collapsed nucleoprotein filaments formed by RecA in the absence of ATP.

The primary ATP binding domain isomerizes to result in the ATP binding site in the extended filament.

Two nucleotide-binding sites have been reported in the compressed/collapsed filament state of MsRecA [18, 47]. While one of these represents the canonical ATP binding site found in all eubacterial RecA proteins (Figure 9A), the second binding site has been reported only in MsRecA so far (Figure 9B) [47]. The primary binding site (Figure 9A) contains all structural elements necessary for binding and hydrolysis of ATP. In contrast, absence of side chains that could make the nucleophilic attack necessary for ATP hydrolysis suggests that the ATP present in the second site is not hydrolysable. Ordering of the C-terminal (MsRecA notation 329-349) results in the formation of the second ATP binding site (Figure 9B). Residues 102-105 of the first ATP binding site interact with the ATP in the second binding site, hinting at an allosteric coupling between the two sites. However, it is unclear if residues in the first site influence formation of the second site or vice versa.

While we were able to identify intermediate structures that could result in the formation of the ATP binding pocket in the extended RecA nucleoprotein filament from the primary ATP binding site (Figure 9C), we were unable to do the same using the second ATP binding site as the initiating conformation using FATCAT server (Figure 9D). Our model suggests that the primary, but not the secondary, ATP binding site can isomerize to result in ATP binding site present in the extended RecA filament and that these sites share common structural elements.

Discussion

Comparison of CD profiles (Figure 1) and crystal structures (Supplementary figure 2) of MsRecA^{wt} and its variants showed that substitution of Gln196 did not result in gross structural abnormalities. Our observations indicate that Gln194/196 is involved in sensing the γ-phosphate of bound ATP, acting as an allosteric effector**,** switching RecA between active and inactive states depending on the occupancy status of the nucleotide-binding pocket. In addition, these observations reveal new details about the catalytic cycle of RecA protein.

Biochemical and biophysical experiments conducted over the last two decades have helped generate a broad outline of the molecular events that constitute the RecA catalytic cycle [10, 16, 48-50]. The interaction of the carboxamide group of Gln194/196 with ATP leads to a conformation change in the DNA binding loop L2, causing it to acquire an extended β-sheet conformation [21]. The information is then transmitted along the peptide backbone of L2, to α-helix 8 downstream of L2, resulting in the insertion of a conserved Phenylalanine (Phe217/219), into a hydrophobic pocket in the neighboring RecA molecule [11]. This insertion results in changes in the cross-subunit interactions that enhance cooperative filament assembly as well as an increase in the pitch of the RecA nucleoprotein filament from ~ 70 Å in the compressed state to ~ 95 Å in the extended state. Upon ATP hydrolysis Gln194/196 returns to its position in the inactive state resulting in the relaxation of inter-subunit interactions.

The substitutions of Gln196 and Arg198 with non-conservative mutations abolished the structural transitions associated with nucleotide binding as evident by circular dichroism (Figure 1D-G). Furthermore, these mutants were unable to communicate the presence of ATP to the inter-protomer interface via α-helix 8 and Phe217/219 and therefore, incapable of acquiring the extended filament conformation (Figure 2).

EM reconstructions of EcRecA nucleoprotein filament, in the presence and absence of ATP, have revealed several structural differences in the compressed and extended nucleoprotein filament [49]. The subunit interactions in the compressed and extended states are significantly different including a large rotation of subunits relative to each other. The result of these structural changes is the positioning of the primary ATP binding site at the interface of two subunits, which has implications for cooperative hydrolysis of ATP. This model proposes that Phe217/219 interacts with the ATP bound in the neighboring subunit and thus coordinates cooperative ATP hydrolysis. While elements of this model concerning the location of the nucleotide binding site were confirmed by the X-ray structure of EcRecA bound to ssDNA [50], we provide evidence in support of a role for Phe217/219 during ATP hydrolysis in the extended filament. Gln194/196 and Arg196/198 position Phe217/219, which in turn positions Glu96/98 to make a nucleophilic attack on the β-γ phosphate bond of the bound ATP (Figure 3A). These interactions are mediated by the cation-pi interactions (Figure 3B), which are absent in the mutant variants of MsRecA (Figure 3C). Such interactions, particularly involving Phenylalanine, have been shown to have catalytic roles in several proteins [51]. These interactions are generally involved in positioning the bases within active sites of enzymes [52] and have an advantage over polar interactions (hydrogen bonding and salt bridges) as the hydrophobic ring can be protected by a hydrophobic pocket when not involved in positioning the base in the active sites.

Although the MsRecA^{Q196A}, MsRecA^{Q196N} and MsRecA^{R198K} proteins were devoid of functional activities of MsRecA^{wt}, these mutants were able to bind ssDNA under conditions of SPR (Figure 4). MsRecA^{Q196E} did not show any binding to ssDNA in either SPR or nitrocellulose filter binding assay (Figure 4E). The structure of presynaptic RecA filament suggests that inter-protomer interface is stabilized by H-bonds (Figure 5C), including those contributed by

Gln194/196 and Arg196/198 [50]. This conclusion is supported by our finding that protomer interface is highly destabilized in the collapsed conformation (Supplementary table 2). As a result such complexes can only be detected by techniques such as SPR that do not physically perturb the complex. Our data supports the inference that, ATP binding enhances ssDNA binding by an increase in cooperative inter-protomer interactions as a result of transition to an extended filament conformation, rather than an increase in the inherent DNA binding affinity of RecA [53]. The inability of MsRecA^{Q196E} to bind ssDNA can be explained, as the negative charge on the glutamate repels the γ-phosphate of the ATP, resulting in a structural distortion of the ATP binding pocket. These distortions are clearly visible in the structure of MsRecA^{Q196E} bound to dATP [20].

While, MsRecA is able to catalyze formation of D-loops in the presence of ATPγS, the mutant proteins displayed highly reduced but measurable levels of D loop formation (Figure 6). However, this limited ability did not extend to conversion of these early intermediates into heteroduplex DNA (Figure 7), indicating that while the inactive collapsed filament of RecA is perhaps competent in initiating strand exchange, it is unable to promote the formation of the heteroduplex DNA due to an inability to acquire the extended filament conformation. This conclusion is supported by the inability of MsRecA variants to support cleavage of LexA (Figure 8), which requires an extended RecA filament conformation [46].

Unlike the primary ATP binding site (Figure 9D), attempts to identify intermediate/s that may lead to formation of the presynaptic ATP binding site from the second ATP binding domain were unsuccessful. Results of our modelling seem to indicate that the second nucleotide-binding site does not participate directly in the catalytic cycle of RecA and is perhaps dispensable for RecA functions. This conclusion is supported by the fact that, while deletion of residues equivalent to

those that form the second ATP binding site (MsRecA notation 329-349), in EcRecA causes change in the pH optima for strand exchange, it does not result in inability to promote strand exchange[54].

In summary, we provide experimental evidence to support the models proposed by De Zutter *et al.,* [11] and VanLoock *et al.,* [49] explaining structural transitions associated with ATP binding in RecA. We observed changes in secondary structure of MsRecA as well as a small but significant rotation of aromatic side chain of Phe217/219 as a result of ATPγS binding in $MsRecA^{wt}$. These changes were absent in MsRecA variants (Figure 2D-G), establishing a causal relationship between nucleotide binding, its recognition by carboxamide group of Gln194/196 and subsequent conformation changes including change in the orientation of Phe217/219. In addition, we propose that interaction of Phe217/219 with ATP (Figure 3A) provides a mechanistic explanation for the faster kinetics of ATP hydrolysis in the extended filaments. Finally, our results suggest that the primary, but not the secondary ATP binding site can isomerize to result in the ATP binding site in the nucleoprotein filament (Figure 9D) of MsRecA.

Acknowledgements

We wish to thank Dr. Elaine O Davis at MRC, London for the gift of bacterial expression (pFM18) construct for MtLexA. GPM is thankful to Profs. Sanjeev Galande and L S Shashidhara at IISER Pune India for their support and encouragement.

Supporting Information Available

Detailed protocol for purification of MsRecA and its variants is included in the associated supplementary information. In addition, protocols for ATP hydrolysis, nitrocellulose filter binding, electrophoretic mobility shift assay (EMSA) and molecular modelling are included.

Additional results in the supplementary information include results of ATP binding and hydrolysis, EMSA and nitrocellulose filter binding assay.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. GPM and KM designed the experiments, analyzed the results and wrote the manuscript; GPM, NS and MSM performed structural analysis and molecular modelling. PLV and DJS contributed to preparation of the manuscript.

References:

[1] MacFarland, K. J., Shan, Q., Inman, R. B., and Cox, M. M. (1997) RecA as a motor protein. Testing models for the role of ATP hydrolysis in DNA strand exchange, *J Biol Chem 272*, 17675-17685.

[2] Cox, M. M. (2003) The bacterial RecA protein as a motor protein, *Annu Rev Microbiol 57*, 551-577.

[3] Cox, M. M. (2007) Motoring along with the bacterial RecA protein, *Nat Rev Mol Cell Biol 8*, 127-138.

[4] Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., and Rehrauer, W. M. (1994) Biochemistry of homologous recombination in Escherichia coli, *Microbiol Rev 58*, 401-465.

[5] Roca, A. I., and Cox, M. M. (1997) RecA protein: structure, function, and role in recombinational DNA repair, *Prog Nucleic Acid Res Mol Biol 56*, 129-223.

[6] McGrew, D. A., and Knight, K. L. (2003) Molecular design and functional organization of the RecA protein, *Crit Rev Biochem Mol Biol 38*, 385-432.

[7] Egelman, E. H., and Stasiak, A. (1986) Structure of helical RecA-DNA complexes. Complexes formed in the presence of ATP-gamma-S or ATP, *J Mol Biol 191*, 677-697.

[8] Stasiak, A., Egelman, E. H., and Howard-Flanders, P. (1988) Structure of helical RecA-DNA complexes. III. The structural polarity of RecA filaments and functional polarity in the RecA-mediated strand exchange reaction, *J Mol Biol 202*, 659-662.

[9] Stasiak, A., and Egelman, E. H. (1994) Structure and function of RecA-DNA complexes, *Experientia 50*, 192-203.

[10] Logan, K. M., Forget, A. L., Verderese, J. P., and Knight, K. L. (2001) ATP-mediated changes in cross-subunit interactions in the RecA protein, *Biochemistry 40*, 11382-11389.

[11] Kelley De Zutter, J., Forget, A. L., Logan, K. M., and Knight, K. L. (2001) Phe217 regulates the transfer of allosteric information across the subunit interface of the RecA protein filament, *Structure 9*, 47-55.

[12] Hewat, E. A., Ruigrok, R. W., and DiCapua, E. (1991) Activation of recA protein: the pitch of the helical complex with single-stranded DNA, *EMBO J 10*, 2695-2698.

[13] Lee, A. M., and Singleton, S. F. (2006) Intersubunit electrostatic complementarity in the RecA nucleoprotein filament regulates nucleotide substrate specificity and conformational activation, *Biochemistry 45*, 4514-4529.

[14] Yu, X., and Egelman, E. H. (1992) Direct visualization of dynamics and co-operative conformational changes within RecA filaments that appear to be associated with the hydrolysis of adenosine 5'-O-(3-thiotriphosphate), *J Mol Biol 225*, 193-216.

[15] Story, R. M., and Steitz, T. A. (1992) Structure of the recA protein-ADP complex, *Nature 355*, 374-376.

[16] Kelley, J. A., and Knight, K. L. (1997) Allosteric regulation of RecA protein function is mediated by Gln194, *J Biol Chem 272*, 25778-25782.

[17] Datta, S., Prabu, M. M., Vaze, M. B., Ganesh, N., Chandra, N. R., Muniyappa, K., and Vijayan, M. (2000) Crystal structures of Mycobacterium tuberculosis RecA and its complex with ADP-AlF(4): implications for decreased ATPase activity and molecular aggregation, *Nucleic Acids Res 28*, 4964-4973.

[18] Datta, S., Krishna, R., Ganesh, N., Chandra, N. R., Muniyappa, K., and Vijayan, M. (2003) Crystal structures of Mycobacterium smegmatis RecA and its nucleotide complexes, *J Bacteriol 185*, 4280-4284.

[19] Krishna, R., Prabu, J. R., Manjunath, G. P., Datta, S., Chandra, N. R., Muniyappa, K., and Vijayan, M. (2007) Snapshots of RecA protein involving movement of the C-domain and different conformations of the DNA-binding loops: crystallographic and comparative analysis of 11 structures of Mycobacterium smegmatis RecA, *J Mol Biol 367*, 1130-1144.

[20] Prabu, J. R., Manjunath, G. P., Chandra, N. R., Muniyappa, K., and Vijayan, M. (2008) Functionally important movements in RecA molecules and filaments: studies involving mutation and environmental changes, *Acta Crystallogr D Biol Crystallogr 64*, 1146-1157.

[21] Hortnagel, K., Voloshin, O. N., Kinal, H. H., Ma, N., Schaffer-Judge, C., and Camerini-Otero, R. D. (1999) Saturation mutagenesis of the E. coli RecA loop L2 homologous DNA pairing region reveals residues essential for recombination and recombinational repair, *J Mol Biol 286*, 1097-1106.

[22] Reymer, A., Babik, S., Takahashi, M., Norden, B., and Beke-Somfai, T. (2015) ATP Hydrolysis in the RecA-DNA Filament Promotes Structural Changes at the Protein-DNA Interface, *Biochemistry 54*, 4579-4582.

[23] Cunningham, R. P., Shibata, T., DasGupta, C., and Radding, C. M. (1979) Single strands induce recA protein to unwind duplex DNA for homologous pairing, *Nature 281*, 191-195.

[24] Li, S., and Wilkinson, M. F. (1997) Site-directed mutagenesis: a two-step method using PCR and DpnI, *Biotechniques 23*, 588-590.

[25] Reddy, M. S., Guhan, N., and Muniyappa, K. (2001) Characterization of single-stranded DNA-binding proteins from Mycobacteria. The carboxyl-terminal of domain of SSB is essential for stable association with its cognate RecA protein, *J Biol Chem 276*, 45959-45968.

[26] Movahedzadeh, F., Colston, M. J., and Davis, E. O. (1997) Characterization of Mycobacterium tuberculosis LexA: recognition of a Cheo (Bacillus-type SOS) box, *Microbiology 143 (Pt 3)*, 929-936.

[27] Ganesh, N., and Muniyappa, K. (2003) Characterization of DNA strand transfer promoted by Mycobacterium smegmatis RecA reveals functional diversity with Mycobacterium tuberculosis RecA, *Biochemistry 42*, 7216-7225.

[28] Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal Biochem 72*, 248-254.

[29] Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints, *J Mol Biol 234*, 779-815.

[30] Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2006) Comparative protein structure modeling using Modeller, *Curr Protoc Bioinformatics Chapter 5*, Unit 5 6.

[31] Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U., and Sali, A. (2007) Comparative protein structure modeling using MODELLER, *Curr Protoc Protein Sci Chapter 2*, Unit 2 9.

[32] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol Syst Biol 7*, 539.

[33] Shen, M. Y., and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures, *Protein Sci 15*, 2507-2524.

[34] Ye, Y., and Godzik, A. (2004) FATCAT: a web server for flexible structure comparison and structure similarity searching, *Nucleic Acids Res 32*, W582-585.

[35] Ye, Y., and Godzik, A. (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists, *Bioinformatics 19 Suppl 2*, ii246-255.

[36] Nguyen, M. N., and Madhusudhan, M. S. (2011) Biological insights from topology independent comparison of protein 3D structures, *Nucleic Acids Res 39*, e94.

[37] Ganesh, N., and Muniyappa, K. (2003) Mycobacterium smegmatis RecA protein is structurally similar to but functionally distinct from Mycobacterium tuberculosis RecA, *Proteins 53*, 6-17.

[38] Pugh, B. F., and Cox, M. M. (1988) High salt activation of recA protein ATPase in the absence of DNA, *J Biol Chem 263*, 76-83.

[39] Voloshin, O. N., Wang, L., and Camerini-Otero, R. D. (2000) The homologous pairing domain of RecA also mediates the allosteric regulation of DNA binding and ATP hydrolysis: a remarkable concentration of functional residues, *J Mol Biol 303*, 709-720.

[40] Kumar, K. A., Mahalakshmi, S., and Muniyappa, K. (1993) DNA-induced conformational changes in RecA protein. Evidence for structural heterogeneity among nucleoprotein filaments and implications for homologous pairing, *J Biol Chem 268*, 26162-26170.

[41] Takahashi, M., Kubista, M., and Norden, B. (1989) Binding of recA protein to Z-form DNA studied with circular and linear dichroism spectroscopy, *J Biol Chem 264*, 8568-8574.

[42] Cox, M. M., and Lehman, I. R. (1981) recA protein of Escherichia coli promotes branch migration, a kinetically distinct phase of DNA strand exchange, *Proc Natl Acad Sci U S A 78*, 3433-3437.

[43] Shibata, T., DasGupta, C., Cunningham, R. P., and Radding, C. M. (1980) Homologous pairing in genetic recombination: formation of D loops by combined action of recA protein and a helix-destabilizing protein, *Proc Natl Acad Sci U S A 77*, 2606-2610.

[44] Reddy, M. S., Vaze, M. B., Madhusudan, K., and Muniyappa, K. (2000) Binding of SSB and RecA protein to DNA-containing stem loop structures: SSB ensures the polarity of RecA polymerization on single-stranded DNA, *Biochemistry 39*, 14250-14262.

[45] Little, J. W., Edmiston, S. H., Pacelli, L. Z., and Mount, D. W. (1980) Cleavage of the Escherichia coli lexA protein by the recA protease, *Proc Natl Acad Sci U S A 77*, 3225-3229.

[46] Gruenig, M. C., Renzette, N., Long, E., Chitteni-Pattu, S., Inman, R. B., Cox, M. M., and Sandler, S. J. (2008) RecA-mediated SOS induction requires an extended filament conformation but no ATP hydrolysis, *Mol Microbiol 69*, 1165-1179.

[47] Krishna, R., Manjunath, G. P., Kumar, P., Surolia, A., Chandra, N. R., Muniyappa, K., and Vijayan, M. (2006) Crystallographic identification of an ordered C-terminal domain and a second nucleotide-binding site in RecA: new insights into allostery, *Nucleic Acids Res 34*, 2186-2195.

[48] Skiba, M. C., and Knight, K. L. (1994) Functionally important residues at a subunit interface site in the RecA protein from Escherichia coli, *J Biol Chem 269*, 3823-3828.

[49] VanLoock, M. S., Yu, X., Yang, S., Lai, A. L., Low, C., Campbell, M. J., and Egelman, E. H. (2003) ATP-mediated conformational changes in the RecA filament, *Structure 11*, 187-196.

[50] Chen, Z., Yang, H., and Pavletich, N. P. (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures, *Nature 453*, 489-484.

[51] Philip, V., Harris, J., Adams, R., Nguyen, D., Spiers, J., Baudry, J., Howell, E. E., and Hinde, R. J. (2011) A survey of aspartate-phenylalanine and glutamate-phenylalanine interactions in the protein data bank: searching for anion-pi pairs, *Biochemistry 50*, 2939-2950.

[52] Schwans, J. P., Sunden, F., Lassila, J. K., Gonzalez, A., Tsai, Y., and Herschlag, D. (2013) Use of anion-aromatic interactions to position the general base in the ketosteroid isomerase active site, *Proc Natl Acad Sci U S A 110*, 11308-11313.

[53] De Zutter, J. K., and Knight, K. L. (1999) The hRad51 and RecA proteins show significant differences in cooperative binding to single-stranded DNA, *J Mol Biol 293*, 769-780.

[54] Lusetti, S. L., Wood, E. A., Fleming, C. D., Modica, M. J., Korth, J., Abbott, L., Dwyer, D. W., Roca, A. I., Inman, R. B., and Cox, M. M. (2003) C-terminal deletions of the

Escherichia coli RecA protein. Characterization of in vivo and in vitro effects, *J Biol Chem 278*, 16372-16380.

Figure Legends

Figure 1: Ordering of loop L2 is dependent on recognition of γ-phosphate by carboxamide group of Gln196. (A) The loop L2 (MsRecA Residue 195-213) is highly conserved among eubacterial RecA sequences. The y-axis represents the bit score; a score of one indicates that the residue is invariant among the sequences analyzed. Numbers on the X-axis refer to residue numbers according to MsRecA notation. (B) L2 loop from Apo (green; PDB code 2zrn) and presynaptic structural model (brown) of MsRecA were superimposed. Additional H-bonds formed in the presynaptic state are shown in cyan. (C) The CD spectra of MsRecA (5 μM) alone (dotted line), in the presence of 10 μM ATPγS (broken line) or 25 μM ATPγS (continuous line) are represented in terms of molar elipticity. (D-G) CD spectra of MsRecA variants in the presence of ATPγS were recorded similar to MsRecA and displayed no change upon incubation with ATPγS.

Figure 2: Movement of α-helix 8 and Phe219 correlates with the formation of H-bonds between γ-phosphate of ATP and carboxamide group at position 196 (A) α-helix 8 (MsRecA residue 216-221) containing Phe217/219 is highly conserved among eubacterial RecA sequences. The y-axis represents the bit score; a score of one indicates that the residue is invariant among the sequences analyzed. Numbers on the X-axis refer to residue numbers according to MsRecA notation. (B) Phe217/219 at the C-terminal end of α-helix 8, makes contact with a hydrophobic pocket in the neighboring protomer made up of residues Thr150/152 and Ile155/157 (C) Residues comprising of Loop L2 and α-helix 8 transmit information from ATP binding pocket to the inter-protomer surface. (D) Crystal structures of apo (green; PDB code: 2zrn) and ATPγS bound (brown; PDB

code: 2zre) MsRecA were superimposed to detect changes in the disposition of α-helix 8 and Phe217/219 upon ATP binding. The aromatic side chain of Phe217/219 rotates by 20 degrees-the maximum possible rotation in this position. Tyr 220 is highlighted to indicate that the changes are specific to Phe219. (E-G) No changes were observed in the disposition of Phe217/219 in MsRecA variants in response to ATPγS binding (please refer to table 1 for PDB codes of the crystal structures).

Figure 3: Gln196, Arg198 and Phe219 regulate ATP hydrolysis in the presynaptic filament. (A) Presynaptic filament structure of MsRecA was modeled based on the structure of EcRecA bound to ssDNA (PDB code: 3CMW). F219 ring in ATP unbound (blue) and ATP bound (pink) state shows rotation. Mesh representation around the side chains shows the volume occupied by the side chains. (B) Molecular orbital diagram showing side chain movements in the ATP bound state of MsRecA. Negative charge on the ATP causes Gln196 and Arg198 to move towards ATP, that induces contact between molecular orbitals of Gln196 and Phe219 in the neighboring protomer. To facilitate this interaction, Phe219 rotates towards Gln196 making a more stable cation-pi interaction. Partial positive charge on Phe219 ring brings Glu98 towards it making anion-pi interaction. Movement of Glu98 facilitates nucleophilic attack on ATP leading to ATP hydrolysis. (C) MsRecA mutant variants are unable to influence the disposition of Phe219.

Figure 4: Binding of MsRecA and its mutant variants to single-stranded DNA (SPR). (A-D) Surface plasmon resonance (SPR) measurements; ssDNA-binding was measured using BIAcore 2000, as described under materials and methods. The y-axis shows the response units (RU) as a function of time represented on the x-axis (sec). Only three concentrations of RecA are shown for clarity. The arrow indicates the time at which buffer was injected into the chip and the protein

starts dissociating from ssDNA (E) Comparison of dissociation constants (μM) for MsRecA and its variants obtained from filter binding and SPR studies.

Figure 5: ssDNA-binding interface remains unaltered in both collapsed and extended MsRecA filaments (A) MsRecA structures in compressed (brown) and extended (green) form are superimposed to highlight residues that interact with ssDNA. ssDNA was modeled based on the structure of the EcRecA nucleoprotein filament (PDB code: 3CMW). The disposition of residues that interact with ssDNA remains unchanged in compressed and extended state. (B) Space filling model of the compressed MsRecA filament (PDB code: 2zrn); the basic residues lining the inner surface of the filament are indicated in blue. (C) H-bonds between neighboring protomers are shown in blue. Inter-protomer interface is stabilized by five additional H-bonds in the extended filament conformation; see also supplementary tables 3 and 4.

Figure 6: MsRecA variants can promote limited D-loop formation. The formation of D-loop by MsRecA and its mutant forms was assayed as described under materials and methods. Lanes 1 to 5 correspond to reaction mixtures containing no protein, 0.5 μM, 1 μM, 2 μM and 4 μM of either MsRecA or its variants.

Figure 7: MsRecA variants are unable to promote three-strand exchange. (A) A schematic depicting the experimental design. (B) 100 nM (molecules) of 83-mer ssDNA was incubated with 3 μM of either wild type or mutant MsRecA in a buffer containing 20 mM Tris-HCl (pH 7.0), 10 mM $MgCl_2$, 0.1 mM ATPγS or ATP and 1 mM DTT at 37°C for 10 min. Strand exchange was initiated by the addition of 100 nM (molecules) of 83-mer duplex assembled by annealing oligonucleotide I and II where oligo I is labeled at its 5‘ end. Aliquots were taken at the end of the reaction, deproteinised and analyzed using 10% polyacrylamide gel.

Figure 8: MsRecA variants do not promote cleavage of LexA. The co-protease activity of MsRecA was monitored using LexA cleavage as described under materials and methods. The reaction was performed in the absence of cofactors, in the presence of ATPγS, or in the presence of both ATPγS and M13ssDNA as indicated at bottom of the gel. Molecular weight markers in kDa are indicated in lane 1.

Figure 9: Comparison of ATP binding sites in RecA. (A) The primary nucleotide-binding site of MsRecA is located in the central catalytic domain. The carboxamide group of residues Gln196 and Arg198 make hydrogen bonds with the γ phosphate of ATP. Glu98 is positioned to make the nucleophilic attack necessary for ATP hydrolysis (PDB code: 2zrm). (B) Ordering of the C-terminal results in the formation of a second ATP binding site in MsRecA. Lys258 and Gln259 make hydrogen bonds with the γ-phosphate of ATP. ATP hydrolysis is unlikely in this site as there is no acidic side chain equivalent to Glu98 that can make the necessary nucleophilic attack (PDB code: 2G88) (C) Extended nucleoprotein filament of MsRecA was modeled based on the structure of EcRecA bound to ssDNA (PDB code: 3CMW). The ATP binding site in the presynaptic filament is formed at the interface of two neighboring protomers (green and brown). Lys250 and 252 hold the γ phosphate in position and Glu98 makes the nucleophilic attack necessary for ATP hydrolysis. (D) The primary ATP binding site isomerizes to ATP binding site in the extended nucleoprotein filament. One monomer is shown in surface representation. Different intermediate stages (light blue and orange) of the neighboring monomer, from compressed state (green) to extended state (pink) are shown in ribbon representation.

S. No	Description	PDB code	Reference
1	MsRecA IV	2zrn	20
2	MsRecA–dATP IV	2zrm	20

3	MsRecA Q196A IV	2zrh	20
4	MsRecA Q196A–dATP IV	2zrk	20
5	MsRecA Q196N IV	2zrc	20
6	MsRecA Q196N–dATP IV	2zrf	20
7	MsRecA Q196E IV	2zr0	20
8	MsRecA Q196E–dATP IV	2zr9	20
9	MsRecA-dATP complex	2G88	47
10	EcRecA-ssDNA-ADPAlF$_4$	3CMW	49

Table 1: RecA structures used in this study for crystallographic analysis and molecular modelling.

S. No	Amino Acid	Mutant/Experimental Method	Phenotype
1	Ser145/147	*recAS145C*	*No Protein*
		recA1647	*rec*$^{+}$, *prt*^{c}
		recAS145C/F217C	*No Protein*
2	Ala147/148	*recAN125*	*rec*$^{-}$, *prt*$^{-}$
3	Ser172/174	*No known mutants*	*Unknown*
4	Asn193/195	*Saturation Mutagenesis*	*rec*$^{-}$
5	Thr210/212	*Saturation Mutagenesis*	*rec*$^{-}$, *rec*$^{-/+}$
6	Ala214/216	*No known mutants*	*Unknown*
7	Tyr218/220	*Multiple*	*rec*$^{-}$, *rec*$^{-/+}$, *prt*^{c}, *prt*$^{+}$

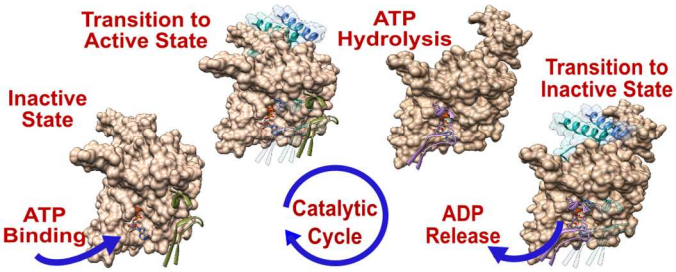
Table 2: Mutation of residues that make direct contact with Glu194/196 and Arg196/198 in the extended nucleoprotein filament affects RecA activity. First residue number refers to *E. coli* and the second to *M. smegmatis* RecA. No protein - protein is unstable and hence not detected; *rec*$^{+}$- No change in RecA activity; *rec*$^{-}$ - loss of RecA activity; *rec*$^{+/-}$ - reduction in RecA activity; *prt*$^{-}$ -

loss of coprotease activity; prt^{C}- constitutive co protease activity. The entries in this table were extracted from McGrew and Knight (2003).

For Table of Contents Use Only

Molecular mechanism underlying ATP-mediated transition to extended filament conformation in *Mycobacterium smegmatis* RecA protein.

G P Manjunath[1, 2*], Neelesh Soni[3], Pavana Lakshmi Vaddavalli[2], Dipeshwari J Shewale[2], M S Madhusudhan[3] and K Muniyappa[1]



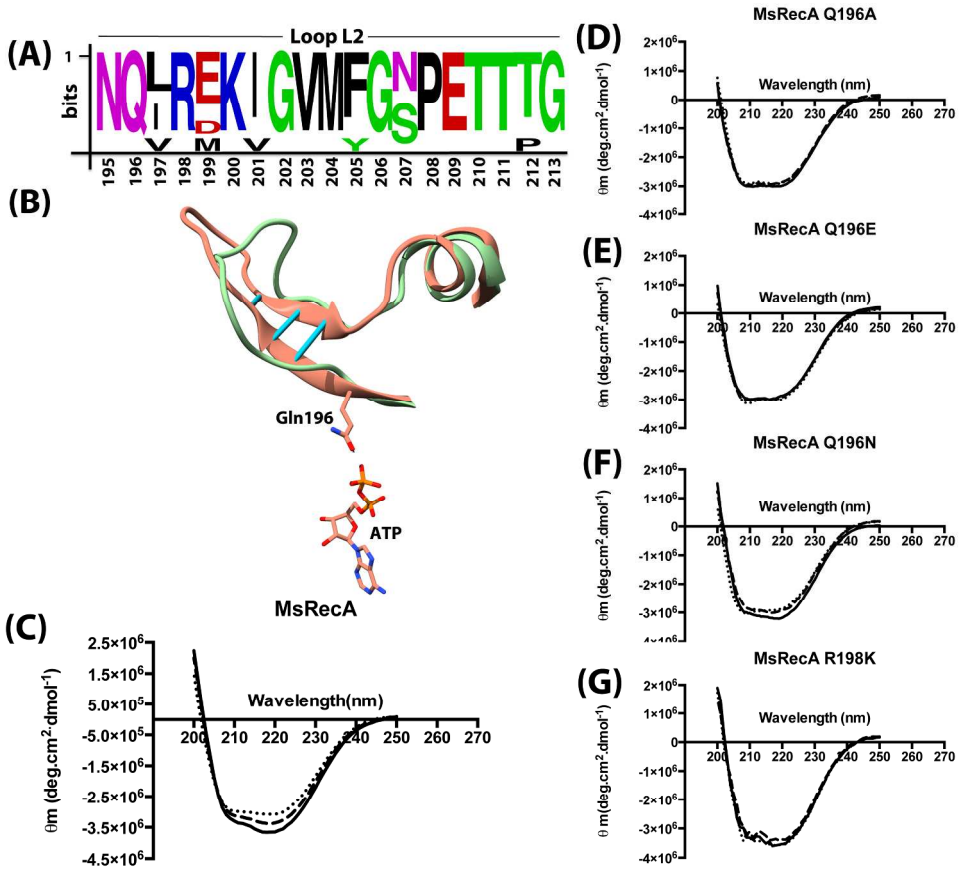


Figure 1: Ordering of loop L2 is dependent on recognition of γ-phosphate by carboxamide group of Gln196. (A) The loop L2 (MsRecA Residue 195-213) is highly conserved among eubacterial RecA sequences. The y-axis represents the bit score; a score of one indicates that the residue is invariant among the sequences analyzed. Numbers on the X-axis refer to residue numbers according to MsRecA notation. (B) L2 loop from Apo (green; PDB code 2zrn) and presynaptic structural model of MsRecA were superimposed. Additional H-bonds formed in the presynaptic state are shown in cyan. (C) The CD spectra of MsRecA (5 μM) alone (dotted line), in the presence of 10 μM ATPγS (broken line) or 25 μM ATPγS (continuous line) are represented in terms of molar elipticity. (D-G) CD spectra of MsRecA variants in the presence of ATPγS were recorded similar to MsRecA and displayed no change upon incubation with ATPγS.

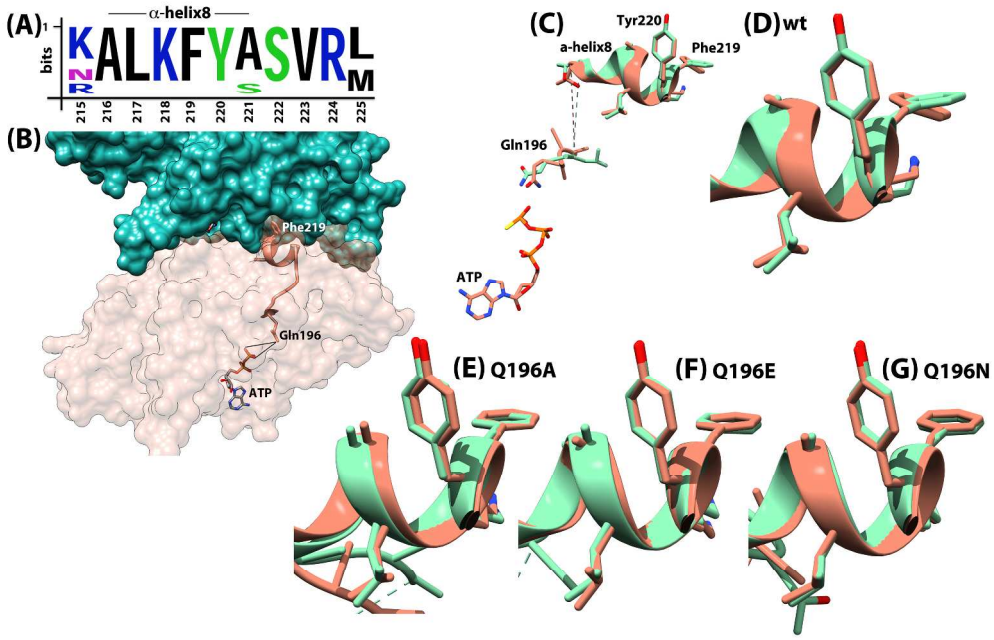
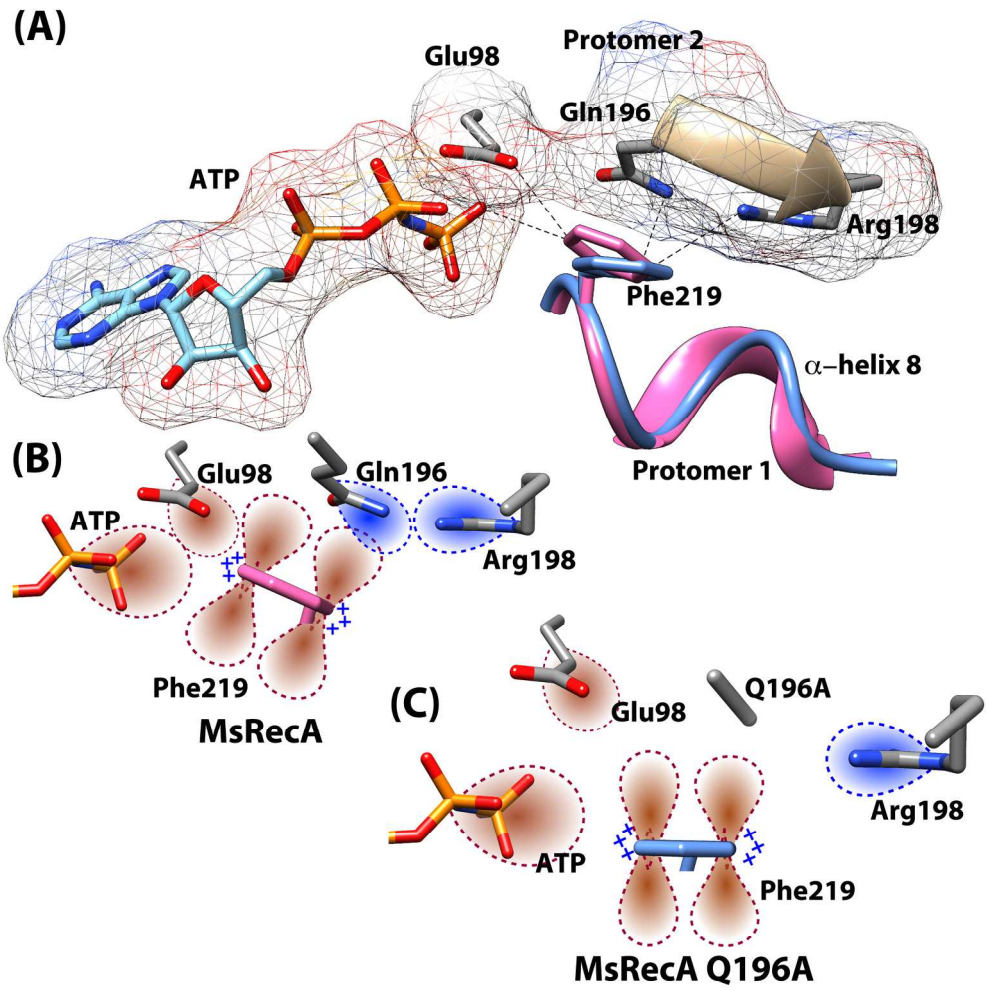
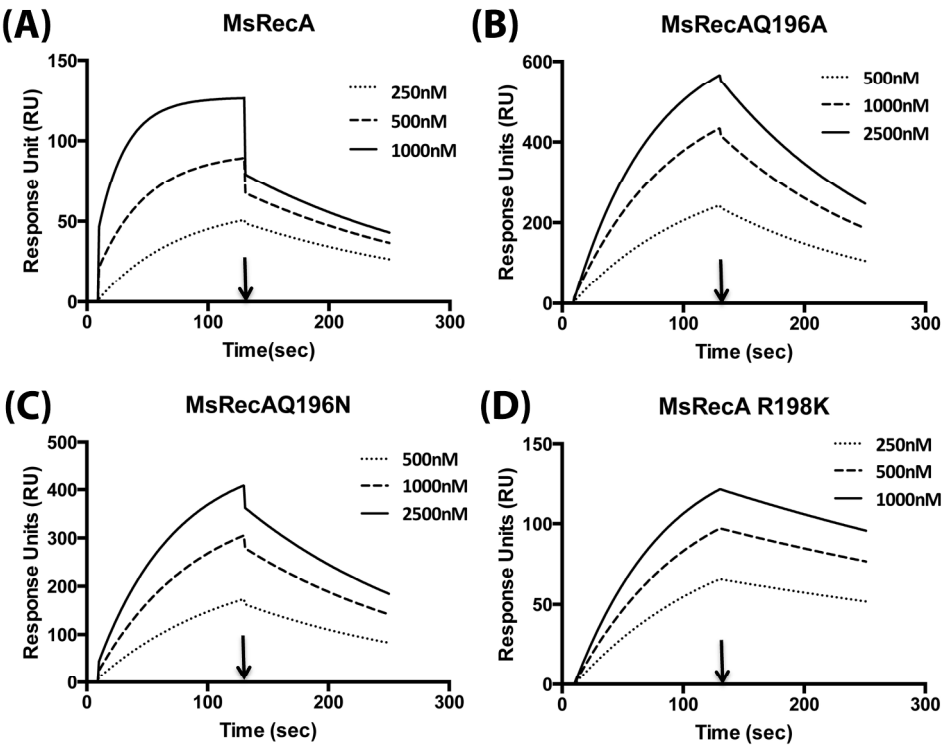


Figure 2: Movement of α-helix 8 and Phe219 correlates with the formation of H-bonds between γ-phosphate of ATP and carboxamide group at position 196 (A) α-helix 8 (MsRecA residue 216-221) containing Phe217/219 is highly conserved among eubacterial RecA sequences. The y-axis represents the bit score; a score of one indicates that the residue is invariant among the sequences analyzed. Numbers on the X-axis refer to residue numbers according to MsRecA notation. (B) Phe217/219 at the C-terminal end of α-helix 8, makes contact with a hydrophobic pocket in the neighboring protomer made up of residues Thr150/152 and Ile155/157 (C) Residues comprising of Loop L2 and α-helix 8 transmit information from ATP binding pocket to the inter-protomer surface. (D) Crystal structures of apo (green; PDB code 2zrn) and ATPγS bound (brown; 2zre) MsRecA were superimposed to detect changes in the disposition of α-helix 8 and Phe217/219 upon ATP binding. The aromatic side chain of Phe217/219 rotates by 20 degrees-the maximum possible rotation in this position. Tyr 220 is highlighted to indicate that the changes are specific to Phe219. (E-G) No changes were observed in the disposition of Phe217/219 in MsRecA variants in response to ATPγS binding (please refer to table 1 for PDB codes of the crystal structures).



Gln196, Arg198 and Phe219 regulate ATP hydrolysis in the presynaptic filament. (A) Presynaptic filament structure of MsRecA was modeled based on the structure of EcRecA bound to ssDNA (PDB code 3CMW). F219 ring in ATP unbound (blue) and ATP bound (pink) state shows rotation. Mesh representation around the side chains shows the volume occupied by the side chains. (B) Molecular orbital diagram showing side chain movements in the ATP bound state of MsRecA. Negative charge on the ATP forces Gln196 and Arg198 to move towards ATP, that induces contact between molecular orbitals of Gln196 and Phe219 in the neighboring protomer. To facilitate this interaction, Phe219 rotates towards Gln196 making a more stable cation-pi interaction. Partial positive charge on Phe219 ring brings Glu98 towards it making anion-pi interaction. Movement of Glu98 facilitates nucleophilic attack on ATP leading to ATP hydrolysis. (C) MsRecA mutant variants are unable to influence the disposition of Phe219.
103x106mm (600 x 600 DPI)



(E)

	Method	MsRecA	MsRecA Q196A	MsRecA Q196E	MsRecA Q196N	MsRecA R198K
1	Filter Binding	2.3	N.D.	N.D.	N.D.	N.D.
2	Surface Plasmon Resonance	0.13	2.4	N.D.	1.75	2.0

Figure 4: Binding of MsRecA and its mutant variants to single-stranded DNA (SPR). (A-D) Surface plasmon resonance (SPR) measurements; ssDNA-binding was measured using BIAcore 2000 as described under materials and methods. The y-axis shows the response units (RU) as a function of time represented on the x-axis (sec). Only three concentrations of RecA are shown for clarity. The arrow indicates the time at which buffer was injected into the chip and the protein starts dissociating from ssDNA (E) Comparison of dissociation constants (μM) for MsRecA and its variants obtained from filter binding and SPR studies.
193x209mm (600 x 600 DPI)

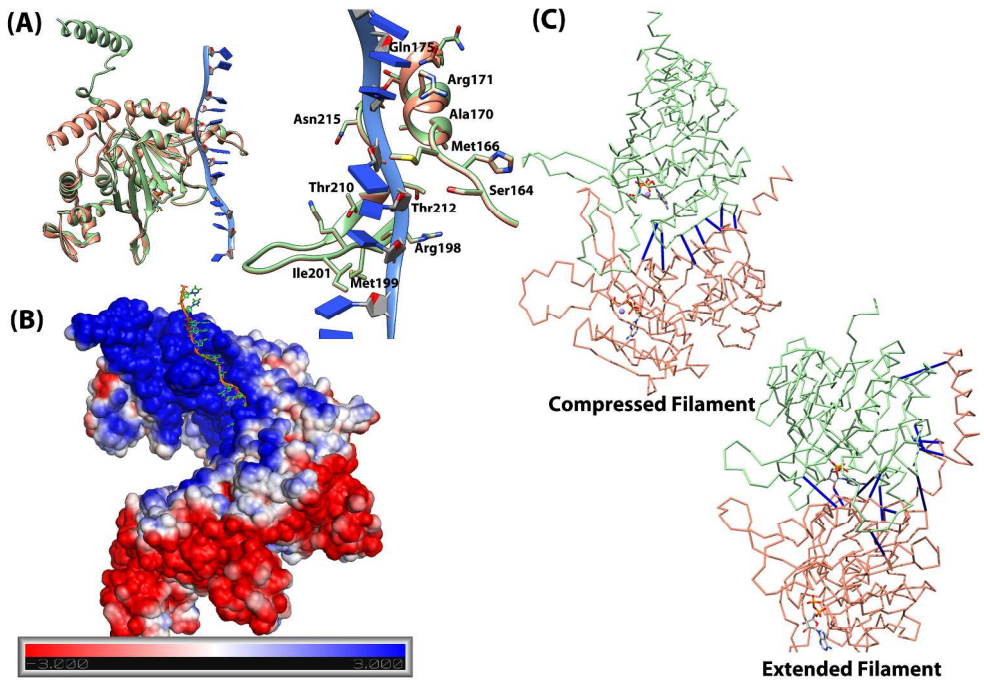


Figure 5: ssDNA-binding interface remains unaltered in both collapsed and extended MsRecA filaments (A) MsRecA structures in compressed (brown) and extended (green) form are superimposed to highlight residues that interact with ssDNA. ssDNA was modeled based on the structure of the EcRecA nucleoprotein filament (PDB code 3CMW). The disposition of residues that interact with ssDNA remains unchanged in compressed and extended state. (B) Space filling model of the compressed MsRecA filament (PDB code 2zrn); the basic residues lining the inner surface of the filament are indicated in blue. (C) H-bonds between neighboring protomers are shown in blue. Inter-protomer interface is stabilized by five additional H-bonds in the extended filament conformation; see also supplementary tables 3 and 4.
129x93mm (600 x 600 DPI)

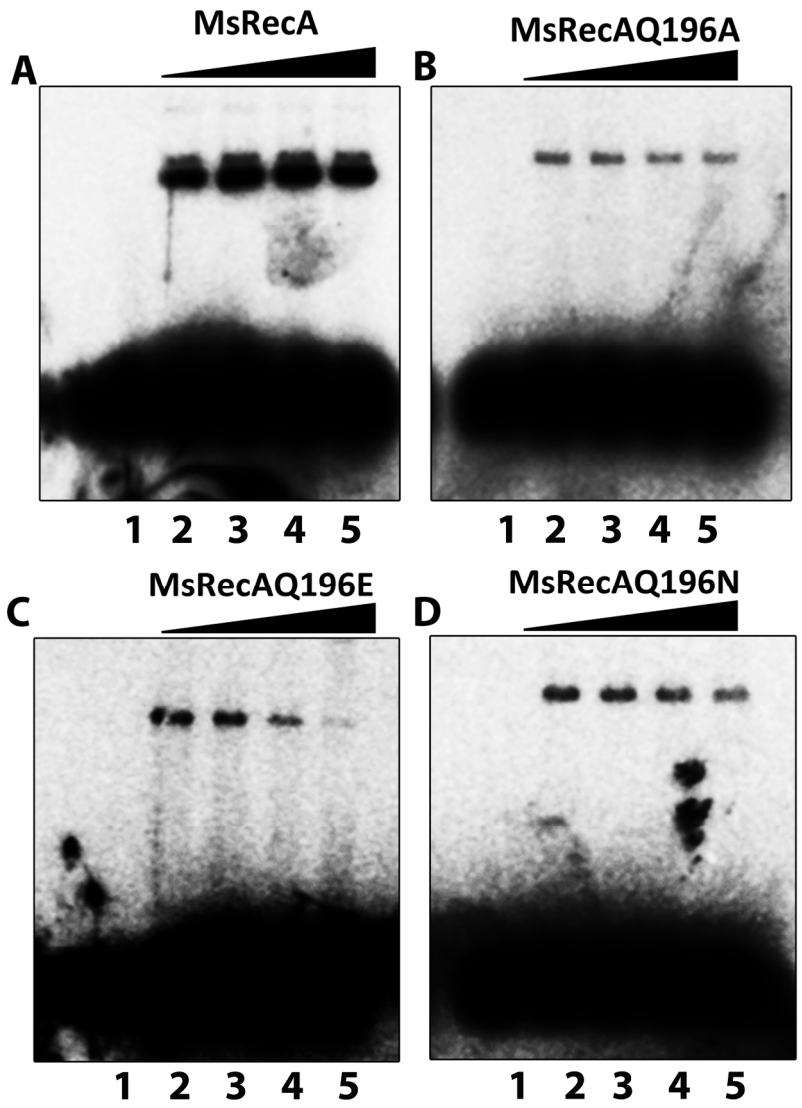


Figure 6: MsRecA variants can promote limited D-loop formation. The formation of D-loop by MsRecA and its mutant forms was assayed as described under materials and methods. Lanes 1 to 5 correspond to reaction mixtures containing no protein, 0.5 μM, 1 μM, 2 μM and 4 μM of either MsRecA or its variants.
113x161mm (600 x 600 DPI)

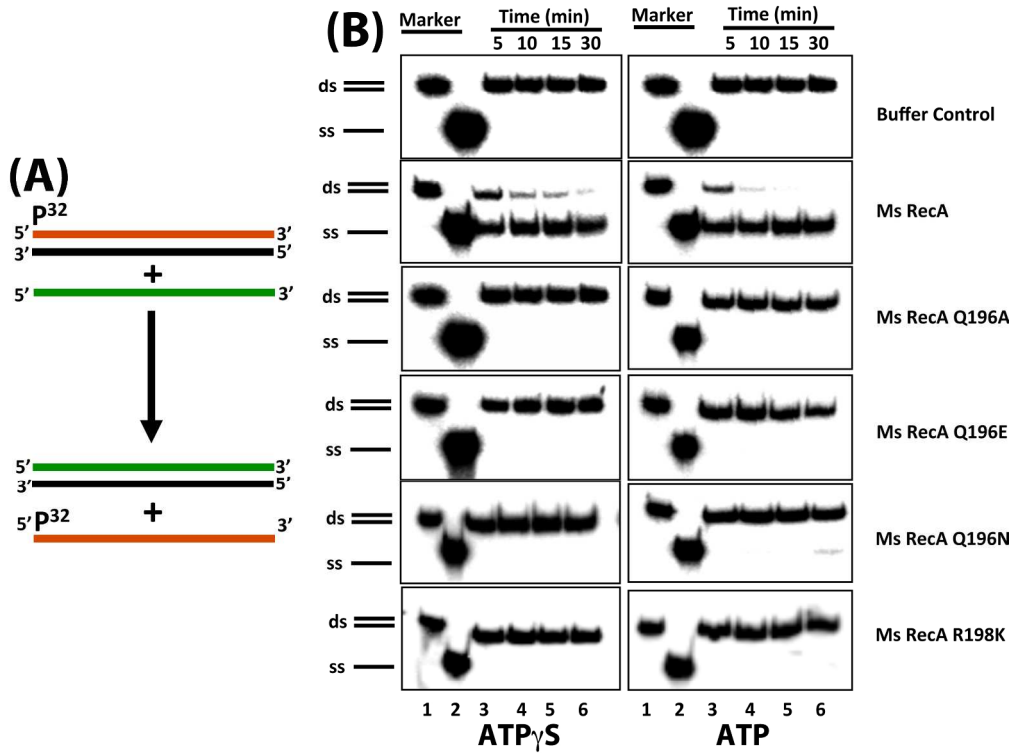


Figure 7: MsRecA variants are unable to promote three-strand exchange. (A) A schematic depicting the experimental design. (B) 100 nM (molecules) of 83-mer ssDNA was incubated with 3 μM of either wild type or mutant MsRecA in a buffer containing 20 mM Tris-HCl (pH 7.0), 10 mM MgCl2, 0.1 mM ATPγS or ATP and 1 mM DTT at 37°C for 10 min. Strand exchange was initiated by the addition of 100 nM (molecules) of 83-mer duplex assembled by annealing oligonucleotide I and II where oligo I is labeled at its 5‵ end. Aliquots were taken at the end of the reaction, deproteinised and analyzed using 10% polyacrylamide gel.

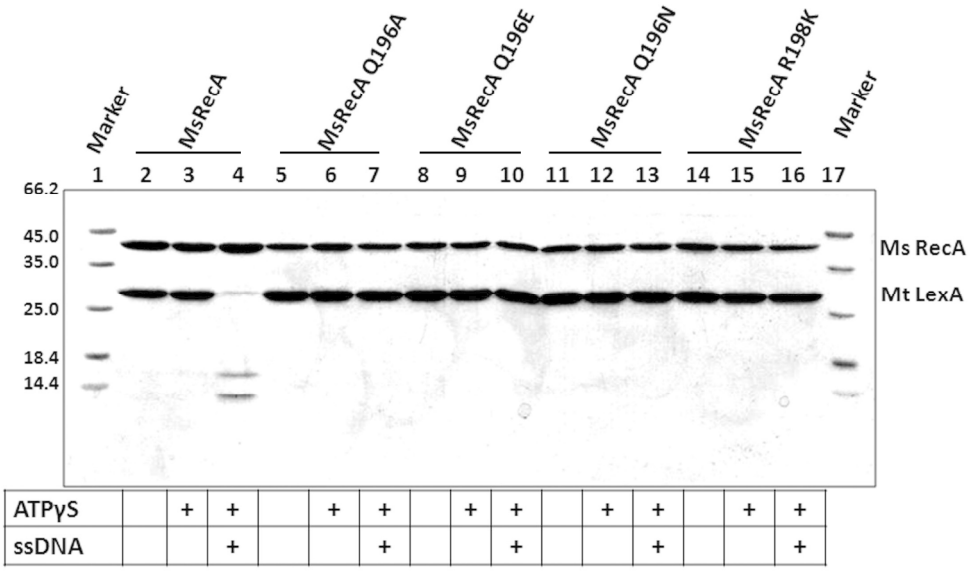


Figure 8: MsRecA variants do not promote cleavage of LexA. The co-protease activity of MsRecA was monitored using LexA cleavage as described under materials and methods. The reaction was performed in the absence of cofactors, in the presence of ATPγS, or in the presence of both ATPγS and M13ssDNA as indicated at bottom of the gel.

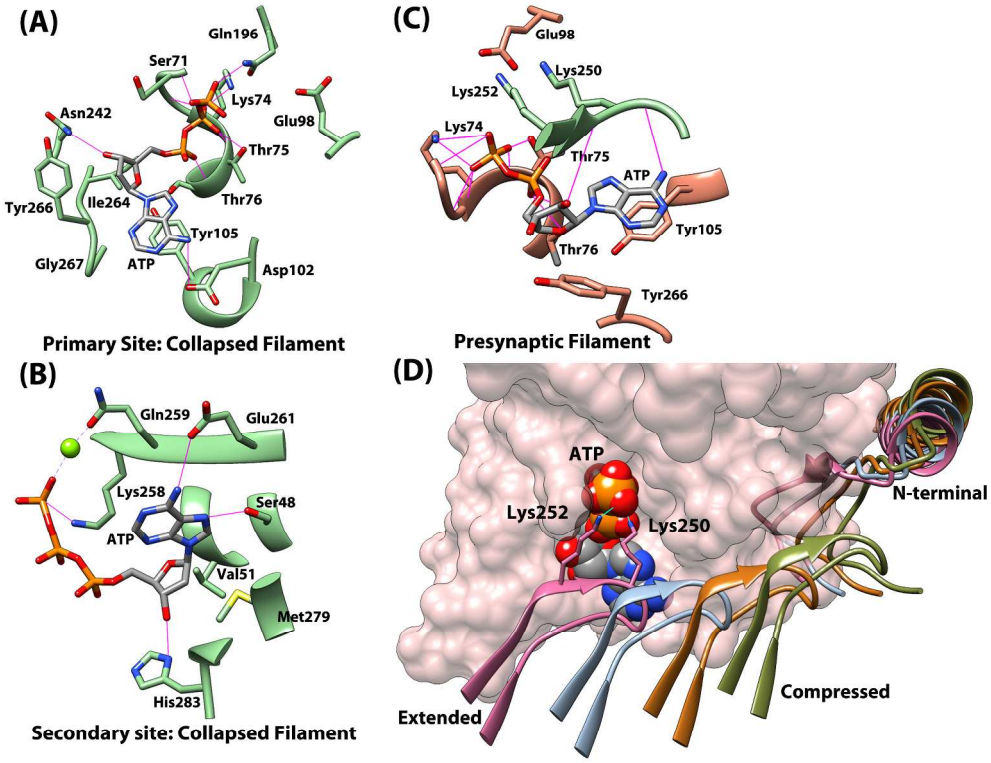
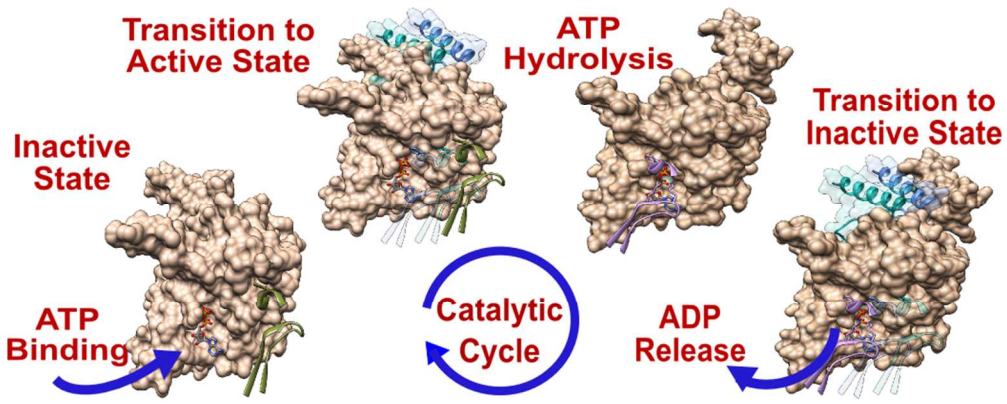


Figure 9: Comparison of ATP binding sites in RecA. (A) The primary nucleotide-binding site of MsRecA is located in the central catalytic domain. The carboxamide group of residues Gln196 and Arg198 make hydrogen bonds with the γ phosphate of ATP. Glu98 is positioned to make the nucleophilic attack necessary for ATP hydrolysis (PDB code 2zrm). (B) Ordering of the C-terminal results in the formation of a second ATP binding site in MsRecA. Gln259 and Lys258 make hydrogen bonds with the γ-phosphate of ATP. ATP hydrolysis is unlikely in this site as there is no acidic side chain equivalent to Glu98 that can make the necessary nucleophilic attack (PDB code 2G88) (C) Extended nucleoprotein filament of MsRecA was modeled based on the structure of EcRecA bound to ssDNA (PDB code 3CMW). The ATP binding site in the presynaptic filament is formed at the interface of two neighboring protomers (green and brown). Lys250 and 252 hold the γ phosphate in position and Glu98 makes the nucleophilic attack necessary for ATP hydrolysis. (D) The primary ATP binding site isomerizes to ATP binding site in the extended nucleoprotein filament. One monomer is shown in surface representation. Different intermediate stages (light blue and orange) of the neighboring monomer, from compressed state (green) to extended state (pink) are shown in ribbon representation.



For Table of Contents Graphic Only
89x35mm (299 x 299 DPI)