Allosteric movements in eubacterial RecA

Anu V. Chandran · M. Vijayan

Received: 10 May 2012 / Accepted: 3 October 2012 / Published online: 23 October 2012
© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag Berlin Heidelberg 2012

Abstract The action of RecA, an important eubacterial protein involved in recombination and repair, involves the transition from an inactive filament in the absence of DNA to an active filament formed in association with DNA and ATP. The structure of the inactive filament was first established in *Escherichia coli* RecA (*EcRecA*). The interaction of RecA with non-hydrolysable ATP analogues and ADP has been thoroughly characterized and the DNA binding loops visualized based on the crystal structures of the RecA proteins from *Mycobacterium tuberculosis* (*MtRecA*) and *Mycobacterium smegmatis* (*MsRecA*). A switch residue, which triggers the transformation of the information on ATP binding to the DNA binding regions, has been identified. The 20-residue C-terminal stretch of RecA, which is disordered in all other relevant crystal structures, has been defined in an *MsRecA*-dATP complex. The ordering of the stretch is accompanied by the generation of a new nucleotide binding site which can communicate with the original nucleotide binding site of an adjacent molecule in the filament. The plasticity of *MsRecA* and its mutants involving the switch residue has been explored by studying crystals grown under different conditions at two different temperatures and, in one instance, at low humidity. The structures of these crystals and those of *EcRecA* and *Deinococcus radiodurans* RecA (*DrRecA*) provide information on correlated movements involving different regions of the molecule. These correlated movements appear to be important in the allosteric transitions of RecA during its action.

Keywords Recombination and repair · Molecular plasticity · DNA and nucleotide binding · Correlated intra-molecular movements · Mycobacterial RecA

Introduction

RecA is a multifunctional protein that binds ATP and DNA. It plays a central role in recombination, and in eubacteria it is primarily involved in DNA repair, including the SOS response (McGrew and Knight 2003). RecA molecules exist as helical filaments even when not bound to DNA. Upon binding to DNA, which is preceded by binding to ATP, RecA molecules continue to form slightly longer filaments in complex with ATP and DNA, which are often referred to as ‘active’ filaments, in contrast to the slightly compressed filaments formed in the absence of DNA that are referred to as ‘inactive filaments’ (Flory et al. 1984; Ganesh and Muniyappa 2003). The molecular structure of RecA was first reported by Story et al. (1992), based on their crystal structure analysis of the *Escherichia coli* enzyme (*EcRecA*); at the same time these authors determined the partially refined structure of a complex of the enzyme with ADP. The characterization of these structures established the tertiary structure of the protein, the way it associates into an ‘inactive’ filament in the space group P6₁ and the location of the nucleotide binding site. However, detailed characterization of the RecA–nucleotide interactions remained to be elucidated. The two DNA binding loops and the C-terminal stretch of the molecule were disordered in the crystal structure. In addition, the availability of only one well-refined structure also meant that the plasticity of the molecule, necessary for its allosteric action, remained unexplored. Subsequent X-ray studies on *Mycobacterium tuberculosis* RecA (*MtRecA*) were undertaken by Datta et al. (2000, 2003a), as part of a long-term research programme on mycobacterial proteins, which in due course developed into a national effort (Arora et al. 2011). The research on *MtRecA* was subsequently extended to the homologous protein from *Mycobacterium smegmatis* (*MsRecA*) (Datta et al. 2003b), and the work on these two homologous proteins and further investigations on *EcRecA* led to, among other things, the characterization of possible allosteric movements within the protein molecule.

A. V. Chandran · M. Vijayan (✉)
Molecular Biophysics Unit, Indian Institute of Science,
Bangalore 560012, India
e-mail: mv@mbu.iisc.ernet.in

Structure and aggregation

As in the case of the *E. coli* enzyme, *MtRecA*, the first mycobacterial RecA to be X-ray analysed, crystallizes in space group $P6_1$, and filaments are formed around the six-fold screw axis (Datta et al. 2000). Indeed, the RecA molecules from other sources that were studied subsequently were also found to crystallize in the same space group (Datta et al. 2003b; Rajan and Bell 2004). Thus, not only the crystallographically determined molecular structure, but also the crystal structures themselves are biologically relevant.

The structure of *MtRecA* was analysed in the native crystals as well as in the crystals of its complex with ADP-AlF₄, ATP γ S and dATP, all of which are non-hydrolysable analogues of ATP and ADP (Datta et al. 2000, 2003a). *MtRecA* has approximately a 60% sequence identity with *EcRecA*. Unlike in *EcRecA*, the functional *MtRecA* molecule is produced by the post-translational proteolytic removal of a segment of the polypeptide stretch from the middle of the chain and the subsequent joining of the free ends (Davis et al. 1992; Kumar et al. 1996). However, *EcRecA* and *MtRecA* have essentially the same structure (Datta et al. 2003a), as illustrated in Fig. 1, although

Fig. 1 The structure of *Mycobacterium tuberculosis* (*MtRecA*). The N-, M- and C-domains (see section [Structure and aggregation](#) for description of domains) are denoted in *blue*, *green* and *pink*, respectively. Different functionally important regions in the M-domain are indicated. Coordinates of the dATP complex [Protein Data Bank (PDB) ID: 1MO6] were used to prepare the bulk of the figure. The loops L1 and L2 were taken from the ATP γ S-Mg complex (1MO5) and the ATP γ S complex (1MO4), respectively, of *MtRecA*. This and the subsequent images were prepared using the Python-enhanced molecular graphics tool PyMOL (W.L. DeLano: The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA; available at: <http://www.pymol.org>)

differences in their fine detail are present. *MsRecA*, analysed subsequently along with its complexes, also has the same structure. The basic structure consists of three domains. The N-domain (residues 1–30), which consists of a long helix followed by an extended stretch, projects out of the main M-domain (residues 31–269), which has a ‘P-loop containing NTPase fold’ according to the SCOP classification (Murzin et al. 1995). The fold essentially consists of a twisted β sheet flanked by helices. The M-domain is followed by a 59-residue C-domain which consists of three helices and three β -strands (Datta et al. 2000). Interestingly, the last 20 or so residues of the C-domain, which have been shown to be important in allosteric regulation, are disordered except in one instance (see further in text) in crystal structures.

The M-domain is the business end of the molecule. It contains the P-loop region (residues 67–74), which is important for nucleotide binding. Residues in loop region 225–241 are known to be involved in LexA binding in the SOS response (Dutreix et al. 1989; Story et al. 1992; Yu and Egelman 1993), while loops L1 (157–165) and L2 (195–210) are known to be responsible for recognizing and binding DNA in the RecA filament. These DNA binding loops were not defined in the original *EcRecA* structures. The electron density of L1 and L2 has been defined in the *MtRecA* complexes of Mg-ATP γ S (1MO5) and ATP γ S (1MO4), respectively, and L1 and L2 have been found in some of the *MsRecA* structures as well. The loops project into the central groove of the RecA filaments (Fig. 2) which

Fig. 2 Location of the L1 (*yellow*) and L2 (*black*) loops when viewed down the 6_1 screw axis of the *MtRecA* filament. The six molecules are colored differently. The filament has been generated by applying 6_1 symmetry to the coordinates used for preparing Fig. 1

is the location of the bound DNA. The N- and C-domains are involved in the higher order aggregation of RecA into filaments and bundles of filaments. *EcRecA* and mycobacterial RecAs aggregate into filaments in nearly the same manner. *EcRecA* filaments are known to form bundles (Egelman and Stasiak 1986, 1988), and it has been suggested that the function of bundles could be to suppress filament formation on inappropriate substrates so that the initiation of the SOS pathway on such substrates is prevented (Story et al. 1992). Bundle formation appears to be much weaker in mycobacterial RecA. For example, while there are ten inter-filament hydrogen bonds in the *EcRecA* crystal structure (PDB ID: 2REB; Story et al. 1992), there is only one in the structure of *MtRecA* (1G19; Datta et al. 2000). Shape complementarity between filaments is also poor in the *MtRecA* structure. Furthermore, the surface of the mycobacterial RecA filaments is substantially negatively charged (Fig. 3) relative to the *EcRecA* filaments due to a few amino acid substitutions. Repulsion between the same charges also appears to weaken the interactions among filaments which are in consonance with the observed difference

Fig. 3 Electrostatic surface representation of filaments of *Escherichia coli* RecA (*EcRecA*; **a**) and *MtRecA* (**b**). Adaptive Poisson–Boltzmann Solver (APBS) software (Baker et al. 2001) was used in conjunction with PyMOL to generate the figure. Coordinates of Form II *EcRecA* (1U94) and *MtRecA*-dATP complex (1MO6) were used to prepare the figure

in the aggregation–disaggregation behaviour of *EcRecA* and *MtRecA* (Vaze and Muniyappa 1999).

RecA–nucleotide interactions and the switch residue

The details of RecA–nucleotide interactions were first characterized by X-ray analysis of the complexes of *MtRecA* with ADP (1MO3), ATP γ S (1MO4) and dATP (1MO6) (Datta et al. 2003a). They were further explored in crystallographic studies on similar complexes involving *MsRecA*, which exhibits 89% sequence identity with *MtRecA* (Datta et al. 2003b). The binding site of the nucleotides can be roughly divided into three regions: the phosphate binding region (P), the sugar binding region (S) and the base binding region (B) (Fig. 4). As expected, the P region involves the P loop (residues 68–75), particularly (in the *MsRecA* numbering scheme) Ser71, Ser72, Gly73, Lys74, Thr75 and Thr76. Interestingly, Gln196 (195 in *MtRecA*), the first residue in a DNA binding loop, is also part of the P region. The residues that interact with the sugar include Asn242 and Tyr105. Tyr105, in addition to interacting with a sugar hydroxyl group, stacks against the base. Another residue involved in interacting with the base is Asp102. The structures of the nucleotide complexes of the *EcRecA* (1XMV and 1XMS, Xing and Bell 2004) and *Deinococcus radiodurans* RecA (*DrRecA*; 1XP8; Rajan and Bell 2004), which became subsequently available, suggest that the nucleotide binding region in eubacterial RecAs is essentially conserved.

The geometry of the binding site is substantially well preserved irrespective of the nature of the ligand. The same

Fig. 4 Surface representation of the nucleotide binding pocket of *MtRecA* (1MO6). The bound dATP molecule is shown in stick representation. The switch residue is colored differently. *P* Phosphate binding region, *S* sugar binding region, *B* base binding region

geometry is observed in *MtRecA* and *MsRecA*. The location of the base is also reasonably well-preserved. However, differences do occur in the location of the sugar and, in a more pronounced manner, in the location of the phosphate tail. In spite of the very high sequence homology between *MtRecA* and *MsRecA*, the locations of a given ligand in their respective complexes exhibit some differences. These differences appear to be in consonance with the observed small differences in the biological activity of the protein from the two mycobacteria (Ganesh and Muniyappa 2003).

One residue which consistently moves in response to nucleotide binding is Gln196 in *MsRecA*. The side chain amide group of this residue interacts with the phosphate tail of the nucleotide. The residue moves by about 1 Å or slightly more towards the nucleotide in the complexes of *MsRecA* from its position in the native structure (Fig. 5). The action of RecA, including its association with DNA, is triggered by its binding to ATP. A major question is how the information on nucleotide binding is conveyed to the DNA binding region. Interestingly, Gln196 belongs to a DNA binding loop. The movement of this residue upon nucleotide binding appears to suggest that it could well be the switch that triggers the transmission of the information mentioned above. The relevant crystal structures show that the corresponding residues in *MtRecA* and *EcRecA* (Gln195 and Gln194, respectively) move in a similar manner upon nucleotide binding. Indeed, it has been demonstrated that mutation of this residue in *EcRecA* leads to the loss of the ATP-induced increase in DNA binding affinity (Kelley and Knight 1997). The mutations in *MsRecA* also abolish the

enhancement of ATPase activity in the presence of DNA (Manjunath 2009).

Identification of a fully ordered C-terminal domain and the generation of a second nucleotide binding site

In 2003, Lusetti et al. (2003a, b) suggested that the C-terminal domain has a regulatory role in the catalytic activity of RecA. However, their exploration of the structural basis of this role was hampered as around 20 of the last residues of this domain remained undefined in all of the crystal structures. In this context, it is interesting that Krishna et al. (2006) observed a fully ordered C-domain, including the last 20 residues, in the crystal structure of a complex of *MsRecA* with dATP (2G88). The structure of another form of the same complex had already been determined by our group. The new form was obtained from the same set of crystallization experiments, but with a higher concentration of ammonium sulphate in the crystallization medium. The interest generated by this structure, observed in the new form, shifted the focus of further effort from *MtRecA* to *MsRecA*.

The C-terminal stretch as seen for the first time (329–349; referred to as C*) has a compact folded structure and interacts with the rest of the molecule through a number of hydrogen bonds. The ordering of C* is accompanied by a small movement of the C-domain towards the M-domain. More importantly, this ordering results in a concurrent generation of a new nucleotide binding site (dATP II) (Fig. 6a) with good electron density for dATP. The residues and their disposition in the new binding site are similar to those in the first nucleotide binding site (dATP I) and in the well-characterized sites in other relevant proteins, indicating the probable biological relevance of the site. An exploration of the pockets and cavities in the molecule using the CAST programme (Liang et al. 1998) indicated the presence of this site, while no site was found when the exploration was carried out in the absence of C*. Thermodynamic parameters also appear to confirm the presence of new site (Krishna et al. 2006).

In the RecA filament, formed around the 6₁ screw axis in the crystal structure, dATP II in each subunit is close to dATP I in an adjacent subunit. A communication pathway between the dATP molecules at the two sites has also been established (Fig. 6b) and found to involve the 102–105 stretch, belonging to the first site, and Gln259 of the second site. A magnesium ion bridges the γ -phosphate and Gln259 at the second site. The Asp102 hydrogen bonds with the adenine base of the first site as well as with the Gln259 of the adjacent molecule. Tyr105, which stacks on the base, is also part of the 102–105 stretch. Asp102, Tyr105 and Gln259 are highly conserved residues. The ordering of C*

Fig. 5 The location of Gln196 in native *Mycobacterium smegmatis* (*MsRecA*; green; 1UBC) and when complexed with dATP (magenta; 1UBG). The M-domains of the two structures were superposed to generate the figure. The molecular graphics programme Coot (Emsley et al. 2010) was used in all superpositions

Fig. 6 **a** The ordered C-terminal stretch (yellow, C*) in an *MsRecA*–dATP complex (2G88). Among other features, the second nucleotide binding site (dATP II) is indicated. **b** The communication path between

dATP II and dATP I of the adjacent molecule. The 258–262 and the 100–107 stretches belong to different molecules

and the generation of the second nucleotide binding site are accompanied by several small changes in the structure, some of which could strengthen the communication between the two sites and that between the nucleotide and the DNA binding region (Krishna et al. 2006). In light of the earlier biochemical results on the role of the C-domain, including C*, on the action of RecA, the observation of the ordered C-domain, a second ATP binding site and the communication link between this site and the first site in an adjacent molecule appears to be important in the allostery in RecA filaments.

Concerted movements in the molecule

While we were pursuing our work on mycobacterial RecAs, the structures of three new forms of *EcRecA* (Xing and Bell 2004) and its nucleotide complexes (Bell and Xing 2004) became available. These structures revealed, among other things, a correlation between the helical pitch of the filament, which is the length of the c axis in the hexagonal crystals, and the orientation of the C-domain with respect to the M-domain. The pitch of the helix (the c dimension of the crystal) is 82.7 Å in the original crystals of *EcRecA* (*EcRecA* I (2REB); Story et al. 1992). The three new forms [*EcRecA* II (1U94), *EcRecA* III (1U98) and *EcRecA* IV

(1U99); Xing and Bell 2004] have c dimensions close to 74 Å. The ‘bottom’ half of the C-domain in the newly reported forms moves away from the M-domain, in comparison to its position in the *EcRecA* I, as the pitch of the helix decreases (Fig. 7). The determination of the structure of *DrRecA* provided further evidence for the correlation (1XP8; Rajan and Bell 2004). In *DrRecA*, the relevant region of the C-domain moves still farther away from the M-domain when the pitch reduces still further to 67.5 Å. In view of these observations, we judged it worthwhile to further explore this correlation and the correlation among different types of movements, including those involving the switch residue, the pitch of the filament and the orientation of the C-domain. For this purpose, the structures of *MsRecA*, three mutants involving substitutions of the switch residue and their nucleotide complexes grown under slightly different crystallization conditions were studied at room temperature and low temperature (100K) (Krishna et al. 2007; Prabu et al. 2008). In one instance, a low humidity (LH) form generated by lowering the humidity around the crystal was also studied. In the process, 30 different crystal structures involving *MsRecA* were examined (Table 1).

The 30 structures referred to above can be classified into four forms on the basis of their unit cell dimensions (Table 2). They were crystallized in three slightly different conditions. The form of the crystal is indicated as the last character in the

Fig. 7 Orientation of the C-domain with respect to the M-domain in *Deinococcus radiodurans* (*DrRecA*; brown; 1XP8), *EcRecA* III (cyan; 1U98), *EcRecA* I (magenta; 2REB), *MsRecA* I (blue; 2OFO), *MsRecA* III (orange; 1UBC), *MsRecA* II' (light blue; 2ZR7) and *MsRecA* IV (green; 2ZRN). Residues 308 in *DrRecA*, 295 in *EcRecA* and 297 in *MsRecA* are equivalent. The orientation of the C-domain in the only known crystal form of the *MtRecA* is similar to that in *MsRecA* I. The M-domains in the different structures superpose well. For example, the root means square deviations in the C $^{\alpha}$ position among the superposed M-domains from different forms of *MsRecA* are less than 1 Å

structures listed in the first column of Table 1, and the condition is given in the second column. Most of the early work was done using crystals grown under conditions 1 and 2. However, the mutants could be crystallized only under condition 3. For comparison, native crystals were also grown under this condition. Crystals very similar to Form II crystals were obtained under condition 3. In Table 1, these crystals are designated as Form II', with the prime added to II to indicate the difference in crystallization condition. Crystals grown under the new condition were studied at low temperature as well, which resulted in the identification of a new form, Form IV. In Table 1, the native crystals are denoted by *MsRecA* followed by the form to which they belong. In the case of mutants, the symbol of the substituted amino acid comes after *MsRecA*, followed by the name of the ligand, if present. Thus, *MsRecA* N-dATP IV denotes the dATP complex of the Q196N mutant which crystallizes in Form IV. One set of crystals was grown in the presence of the C-terminal polypeptide stretch of the single-stranded DNA binding (SSB) protein; only the native crystals grew. SSB is added in parenthesis when denoting these crystals. The low humidity form included in the analysis has an 'LH' in its name.

Table 1 Different crystal structures involving *Mycobacterium smegmatis* RecA that have been used in the study of concerted movements in RecA

Crystal structures	Crystallization condition	Protein Data Bank ID
<i>MsRecA</i> III	Condition 1	1UBC
<i>MsRecA</i> -ADP III	Condition 1	1UBE
<i>MsRecA</i> -ATP γ S III	Condition 1	1UBF
<i>MsRecA</i> -dATP III	Condition 1	1UBG
<i>MsRecA</i> -dATP I	Condition 2	2G88
<i>MsRecA</i> -dATP II	Condition 2	2ODN
<i>MsRecA</i> -ATP γ S II	Condition 2	2ODW
<i>MsRecA</i> -LH II	Condition 2	2OE2
<i>MsRecA</i> -ADP I	Condition 2	2OEP
<i>MsRecA</i> (SSB) III	Condition 2	2OES
<i>MsRecA</i> I	Condition 2	2OFO
<i>MsRecA</i> II'	Condition 3	2ZR7
<i>MsRecA</i> -dATP II'	Condition 3	2ZRP
<i>MsRecA</i> IV	Condition 3	2ZRN
<i>MsRecA</i> -ADP IV	Condition 3	2ZRO
<i>MsRecA</i> -dATP IV	Condition 3	2ZRM
<i>MsRecA</i> A IV	Condition 3	2ZRH
<i>MsRecA</i> A-ADP IV	Condition 3	2ZRI
<i>MsRecA</i> A-ATP γ S IV	Condition 3	2ZRJ
<i>MsRecA</i> A-dATP IV	Condition 3	2ZRK
<i>MsRecA</i> A-dATP II'	Condition 3	2ZRL
<i>MsRecA</i> N IV	Condition 3	2ZRC
<i>MsRecA</i> N-ADP IV	Condition 3	2ZRD
<i>MsRecA</i> N-ATP γ S IV	Condition 3	2ZRE
<i>MsRecA</i> N-dATP IV	Condition 3	2ZRF
<i>MsRecA</i> N-dATP II'	Condition 3	2ZRG
<i>MsRecA</i> E II'	Condition 3	2ZRB
<i>MsRecA</i> E IV	Condition 3	2ZR0
<i>MsRecA</i> E-ATP γ S	Condition 3	2ZRA
<i>MsRecA</i> E-dATP IV	Condition 3	2ZR9

Ms, *Mycobacterium smegmatis*

The native and the three derivative crystals grown under condition 1 belong to Form III. Forms I, II and III occur

Table 2 Classification of different *MsRecA* crystal structures based on the cell dimensions

<i>Mycobacterium smegmatis</i> RecA crystal structures	a	c
Form I	108.2–108.4 Å	73.0–73.9 Å
Form II	103.0–103.6 Å	75.1–77.1 Å
Form III	102.9–103.7 Å	71.7–73.6 Å
Form IV	101.4–102.2 Å	74.2–76.1 Å

All crystal structures belong to the space group P6 $_1$

when crystals are grown under condition 2. Interestingly, Form I and Form II crystals are obtained simultaneously when *MsRecA* and dATP are co-crystallized under this condition. Furthermore, Form I readily transforms to Form II when the water content of the crystal is reduced by decreasing the relative humidity around the crystal. The water content of Form I crystals is around 62 %, whereas that of Form II crystals varies between 59.3 and 60.9 %. The water content of Form III crystals is still lower at 58.5 %. Only Form II' crystals are obtained under condition 3, and they transform to Form IV at low temperature. *MsRecA* E-ATP γ S assumes a structure which cannot be readily assigned to any one of the four crystal forms, and this is the only one of the 30 structures which does not belong to one or the other of the four forms.

Understandably, the DNA binding loops which protrude into the central groove exhibit considerable flexibility. However, the movement of these loops cannot be readily correlated with movements in other regions of the molecule. As in the case of *EcRecA*, *MsRecA* also exhibits variation in the orientation of the C-domain among the different forms (Fig. 7). However, unlike in the case of *EcRecA*, the correlation between the orientation of the C-domain and the pitch of the helix is not strong in *MsRecA* structures. On the other hand, any change in the orientation of the C-domain is correlated with a change in that of the LexA binding loop; there is also a reasonably good correlation between a change in the orientation of the C-domain and a change in the location of the switch residue (Prabu et al. 2008). The 'bottom' half of the C-domain and the LexA loop are farthest apart in Form I (Fig. 8a) and closest to each other in Form IV (Fig. 8b). Gln37 appears to have an important role in causing the movement of the C-domain from the position in Form I to that in Form IV (Fig. 8c). In Form I, the side chain of this residue belonging to the loop that connects the N- and M-domains forms hydrogen bonds with the main chain carbonyl group of Ile55 of the M-domain and that of Ile331 of the C-domain. The hydrogen bonds are broken in Form IV, and the side chain of Gln37 flips over to the opposite side, with a movement of 3.5 Å in the C α position. The movement ripples along the loop and contributes to the movement of the 'bottom' half of the C-domain towards the LexA binding loop, leading to the formation of a charge cluster connecting the two regions in Form IV. Steric interactions prevent the two regions from approaching far too close to each other.

As illustrated in Fig. 8d, the position of Gln196 differs among the four crystal forms. The direction of these changes and the directions of the movement of the C-domain correlate well in Forms I, II and IV. Form III deviates from the pattern. The tendency is for the switch residue to move towards its position in the nucleotide complex when the 'bottom' half of the C-domain comes closer to the M-

domain. This tendency is found in the *EcRecA* structures as well. In each *MsRecA* form, Gln196 moves towards the tail of the nucleotide when it binds to the protein. This movement correlates with a small, though perceptible, movement of the C-domain in all the forms. Thus, the movements of the switch residue and the C-domain appear to be correlated. In addition, the switch residue moves in response to nucleotide binding. The situation is best illustrated by comparing Forms I and IV. The position of the switch residue represented by its C α in Form IV is displaced with respect to that in Form I by 1.75 Å. The residue moves in nearly the same direction by 1.31 Å when dATP binds to the molecule in Form I. The movement of the residue in the molecule upon dATP binding is lower at 0.88 Å in Form IV.

The observed movement of residue 196 in the mutants (Q196A, Q196N and Q196E) is of interest in relation to its role as a switch residue. The first mutant involves a change in the chemical nature and the size of the side chain; in the second mutant, the chemical nature remains the same, but the size of the side chain is smaller; in the third mutant, the chemical nature changes, but the size of the side chain remains the same. The crystal structures of all the mutants and their complexes belong to Form IV. As an example, the interactions of dATP with residue 196 in the Form IV wild-type structure and the three mutant structures are illustrated in Fig. 9. In the Q196E complex, the carboxylate group of Glu196 and the terminal phosphate of the ligand understandably move away from each other as both are negatively charged. The alanyl side chain, which in any case cannot form a hydrogen bond, is positioned far away from the terminal phosphate in the complex with Q196A. In the Q196N complex, the side chain nitrogen atom forms a hydrogen bond with one phosphate oxygen, while the nitrogen atom of the glutamyl residue forms two hydrogen bonds with two phosphate oxygens in the complex involving the wild-type protein. Thus, the glutamyl side chain has the right functional group and size to form an appropriate interaction with the terminal phosphate group of the nucleotide. As noted earlier, in every single case, nucleotide binding causes residue 196 to move towards the phosphate chain of the nucleotide. No such systematic movements occur in the case of the mutants. Thus, the research with mutants further indicates the role of Gln196 as a switch residue.

Conclusion

The work reviewed here demonstrates, among other things, the mobility of the C-domain, the LexA binding loop, the switch residue and the loop connecting the N- and the M-domains. The movement of the switch residue in response to nucleotide binding adds another biologically relevant dimension to the plasticity of the RecA molecule. The

Fig. 8 The ‘bottom’ half of the C domain and the LexA binding loop in Form I (2OFO) (a) and Form IV (2ZRN) (b) *MsRecA* crystals. c The N–M loop in Forms I (blue) and IV (green). The side chain of Gln37 which flips is also shown. d Gln196 in Form I (blue; 2OFO), Form II' (light blue; 2ZR7), Form III (orange; 1UBC) and Form IV (green; 2ZRN). The location and orientation of the residue are comparable to those in Fig. 5

generation of a second nucleotide binding site on the ordering of C* and the connectivity between the first nucleotide

binding site in one molecule and the second site in an adjacent molecule in the filaments provide a pathway for

Fig. 9 The switch residue 196 and the nucleotide in the dATP complexes of *MsRecA*-dATP IV (cyan; 2ZRM), *MsRecA* A-dATP IV (yellow; 2ZRK), *MsRecA* N-dATP IV (magenta; 2ZRF) and *MsRecA* E-dATP IV (green; 2ZR9). Reproduced with permission of the International Union of Crystallography from Prabu et al. (2008). (Available at: <http://journals.iucr.org/>)

Fig. 10 The DNA binding loops L1 (a) and L2 (b) in active and inactive filaments of eubacterial RecAs. The main chain trace of the loop region after superposition of the M-domains is shown. The loops in the active filament (green) were taken from the *EcRecA*₅-ssDNA-ADP-AlF₄ (3CMW) and those in the inactive filaments (red) are from the *MsRecA* and *MtRecA* structures in which L1 (1MO5, 1UBG, 2OFO, 2OE2, 2OEP, 2ZRN, 2ZRO, 2ZRM, 2ZRH, 2ZRI, 2ZRK, 2ZRC, 2ZRD, 2ZRE, 2ZRF, 2ZR0) and L2 (1MO4, 1UBG, 2OFO, 2OE2) are ordered. L1 is ordered and has the same conformation in most of the *MsRecA* Form IV structures

allosteric communication that enables the transformations in molecular aggregation during the action of RecA. Indeed, the intrinsic plasticity of the molecule, elucidated through the studies discussed here, is used in the transition from the inactive to the active filament, as evidenced by the presence of a complex structure consisting of a covalently linked oligomer of truncated *Ec*RecA molecules with DNA and an ATP analogue (Chen et al. 2008; Prabu et al. 2008). The most striking difference in the plasticity of the molecule in the inactive and the active filaments is understandably in relation to the DNA binding loops. These are disordered in the *Ec*RecA and *Dr*RecA structures and can be seen in some of the *Ms*RecA and *Mt*RecA structures. They exhibit considerable conformational differences among themselves even when ordered. On the contrary, all of these loops are ordered in the active filament, and they have the same conformation in all of the molecules in the filament even when they are not related crystallographically (Fig. 10).

The gene segment for the 18 C-terminal residues of *Ec*RecA, corresponding to C* were removed when constructing the ‘active’ filaments represented by the *Ec*RecA-DNA-nucleotide complex referred to above (Chen et al. 2008). However, a model of an untreated ‘active’ filament could be readily built by incorporating this stretch into the structure using the geometry of C* observed in the appropriate *Ms*RecA-dATP complex. C* in the model is now adjacent to the C-domain of a neighboring molecule in the ‘active’ filament. This observation, namely, the proximity of the C-domains of adjacent molecules in the filament mediated through C*, appears to provide a structural framework for the known regulatory role of the C-domain including C* (Prabu et al. 2008). Calculations using the Yale Morph server (Krebs and Gerstein 2000) suggest that the ‘inactive’ filament could readily transform into the ‘active’ filament without any break or unacceptable steric clashes in the trajectory of transformation. Thus, a smooth transition from ‘active’ to ‘inactive’ filaments appears to be possible using the internal motions of the RecA molecule elucidated mainly through studies on the inactive filaments.

The internal motions of the RecA molecule discussed here have been derived primarily from studies on *Ms*RecA and, to some extent, *Ec*RecA. The movement of the switch residue in response to nucleotide binding is seen in *Ms*RecA, *Ec*RecA and *Mt*RecA. The movement of the C-domain occurs in *Ms*RecA and *Ec*RecA. This movement correlates with the helical pitch of the filament in *Ec*RecA, but not in *Ms*RecA. The movements of the C-domain and the LexA loop appear to be correlated in *Ms*RecA. Correlation between the movements of the C-domain and the switch residue is seen in *Ms*RecA and *Ec*RecA. The structures of *Mt*RecA that have been X-ray analyzed to date all belong to the same crystal form. A comprehensive exploration involving the different crystal forms and environmental conditions

of the *Mt*RecA molecule remains to be carried out. Nevertheless, available evidence appears to suggest that the internal movements in the RecA molecule from different eubacterial species have commonalities as well as differences.

Acknowledgements Many collaborators, graduate students and post-doctoral fellows have contributed to the work reviewed here. Their names appear in the appropriate references. The work is supported by the Department of Biotechnology, Government of India. AVC is a CSIR Junior Research Fellow. MV is a DAE Homi Bhabha Professor.

Conflict of interest None

References

- Arora A, Chandra NR, Das A, Gopal B, Mande SC, Prakash B, Ramachandran R, Sankaranarayanan R, Sekar K, Suguna K, Tyagi AK, Vijayan M (2011) Structural biology of *Mycobacterium tuberculosis* proteins: the Indian efforts. *Tuberculosis* 91 (5):456–468. doi:10.1016/j.tube.2011.03.004
- Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 98(18):10037–10041. doi:10.1073/pnas.181342398
- Bell CE, Xing X (2004) Crystal structures of *Escherichia coli* RecA in complex with MgADP and MnAMP-PNP. *Biochemistry* 43:16142–16152. doi:10.1021/bi048165y
- Chen Z, Yang H, Pavletich NP (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453:489–494. doi:10.1038/nature06971
- Datta S, Prabu MM, Vaze MB, Ganesh N, Chandra NR, Muniyappa K, Vijayan M (2000) Crystal structures of *Mycobacterium tuberculosis* RecA and its complex with ADP-AlF₄: implications for decreased ATPase activity and molecular aggregation. *Nucleic Acids Res* 28:4964–4973. doi:10.1093/nar/28.24.4964
- Datta S, Ganesh N, Chandra NR, Muniyappa K, Vijayan M (2003a) Structural studies on *Mt*RecA-nucleotide complexes: insights into DNA and nucleotide binding and the structural signature of NTP recognition. *Proteins* 50:474–485. doi:10.1002/prot.10315
- Datta S, Krishna R, Ganesh N, Chandra NR, Muniyappa K, Vijayan M (2003b) Crystal structures of *Mycobacterium smegmatis* RecA and its nucleotide complexes. *J Bacteriol* 185:4280–4284. doi:10.1128/JB.185.14.4280-4284.2003
- Davis EO, Jenner PJ, Brooks PC, Colston MJ, Sedgwick SG (1992) Protein splicing in the maturation of *M. tuberculosis* RecA protein: a mechanism for tolerating a novel class of intervening sequence. *Cell* 71:201–210. doi:10.1016/0092-8674(92)90349-H
- Dutreix M, Moreau PL, Bailone A, Galibert F, Battista J, Walker GC, Devoret R (1989) New recA mutations that dissociate the various RecA protein activities in *Escherichia coli* provide evidence for an additional role for RecA protein in UV mutagenesis. *J Bacteriol* 171:2415–2423
- Egelman EH, 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. doi:10.1016/0022-2836(86)90453-5
- Egelman EH, Stasiak A (1988) Structure of helical RecA-DNA complexes II. Local conformational changes visualized in bundles of RecA-ATPγS filaments. *J Mol Biol* 200(2):329–349. doi:10.1016/0022-2836(88)90245-8

- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Cryst D* 66:486–501. doi:10.1107/S0907444910007493
- Flory J, Tsang SS, Muniyappa K (1984) Isolation and visualization of active presynaptic filaments of RecA protein and Single-Stranded DNA. *Proc Natl Acad Sci USA* 81:7026–7030. doi:10.1073/pnas.81.22.7026
- Ganesh N, Muniyappa K (2003) *Mycobacterium smegmatis* RecA protein is structurally similar to but functionally distinct from *Mycobacterium tuberculosis* RecA. *Proteins* 53:6–17. doi:10.1002/prot.10433
- Kelley JA, Knight KL (1997) Allosteric regulation of RecA protein function is mediated by Gln194. *J Biol Chem* 272(41):25778–25782. doi:10.1074/jbc.272.41.25778
- Krebs WG, Gerstein M (2000) The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res* 28:1665–1675. doi:10.1093/nar/28.8.1665
- Krishna R, Manjunath GP, Kumar P, Surolia A, Chandra NR, Muniyappa K, 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(8):2186–2195. doi:10.1093/nar/gkl107
- Krishna R, Prabu JR, Manjunath GP, Datta S, Chandra NR, Muniyappa K, 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(4):1130–44. doi:10.1016/j.jmb.2007.01.058
- Kumar RA, Vaze MB, Chandra NR, Vijayan M, Muniyappa K (1996) Functional characterization of the precursor and spliced forms of RecA protein of *Mycobacterium tuberculosis*. *Biochemistry* 35(6):1793–1802. doi:10.1021/bi9517751
- Liang JH, Edelsbrunner H, Woodward C (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* 7:1884–1897. doi:10.1002/pro.5560070905
- Lusetti SL, Shaw JJ, Cox MM (2003a) Magnesium ion-dependent activation of the RecA protein involves the C terminus. *J Biol Chem* 278:16981–16988. doi:10.1074/jbc.M212916200
- Lusetti SL, Wood EA, Fleming CD, Modica MJ, Korth J, Abbott L, Dwyer DW, Roca AI, Inman RB, Cox MM (2003b) C-terminal deletions of the *E. coli* RecA protein: characterization of in vivo and in vitro effects. *J Biol Chem* 278:16972–16980. doi:10.1074/jbc.M212917200
- Manjunath GP (2009) *Mycobacterium smegmatis* RecA and SSB: Structure function relationships, interaction with cofactors and accessory proteins. PhD thesis. Indian Institute of Science, Bangalore
- McGrew DA, Knight KL (2003) Molecular design and functional organization of the RecA protein. *Crit Rev Biochem Mol Biol* 38:385–432. doi:10.1080/10409230390242489
- Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247:536–540. doi:10.1016/S0022-2836(05)80134-2
- Prabu JR, Manjunath GP, Chandra NR, Muniyappa K, Vijayan M (2008) Functionally important movements in RecA molecules and filaments: studies involving mutation and environmental changes. *Acta Crystallographica Section D* 64(11):1146–1157. doi:10.1107/S0907444908028448
- Rajan R, Bell CE (2004) Crystal structure of RecA from *Deinococcus radiodurans*: insights into the structural basis of extreme radioresistance. *J Mol Biol* 344(4):951–963. doi:10.1016/j.jmb.2004.09.087
- Story RM, Weber IT, Steitz TA (1992) The structure of the *E. coli* recA protein monomer and polymer. *Nature* 355:318–325. doi:10.1038/355318a0
- Vaze MB, Muniyappa K (1999) RecA protein from *Mycobacterium tuberculosis* possesses pH dependent homologous DNA pairing and strand exchange activities: implications for allele exchange in mycobacteria. *Biochemistry* 38:3175–3186. doi:10.1021/bi9819125
- Xing X, Bell CE (2004) Crystal structures of *Escherichia coli* RecA in a compressed helical filament. *J Mol Biol* 342:1471–1485. doi:10.1016/j.jmb.2004.07.091
- Yu X, Egelman EH (1993) The LexA repressor binds within the deep helical groove of the activated RecA Filament. *J Mol Biol* 231:29–40. doi:10.1006/jmbi.1993.1254