Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Functionally important movements in RecA molecules and filaments: studies involving mutation and environmental changes

J. Rajan Prabu,[a] G. P. Manjunath,[b] Nagasuma R. Chandra,[c]* K. Muniyappa[b] and M. Vijayan[a]*

[a]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India, [b]Department of Biochemistry, Indian Institute of Science, Bangalore 560 012, India, and [c]Bioinformatics Centre, Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore 560 012, India

Correspondence e-mail: nchandra@serc.iisc.ernet.in, mv@mbu.iisc.ernet.in

Received 12 July 2008
Accepted 5 September 2008

PDB References: MsRecA mutants and complexes, 2zr7; 2zrp; 2zrn; 2zro; 2zrm; 2zrh; 2zri; 2zrj; 2zrk; 2zrl; 2zrc; 2zrd; 2zre; 2zrf; 2zrg; 2zrb; 2zr0; 2zra; 2zr9.

The crystal structures of mutants of *Mycobacterium smegmatis* RecA (MsRecA) involving changes of Gln196 from glutamine to alanine, asparagine and glutamic acid, wild-type MsRecA and several of their nucleotide complexes have been determined using mostly low-temperature and partly room-temperature X-ray data. At both temperatures, nucleotide binding results in a movement of Gln196 towards the bound nucleotide in the wild-type protein. This movement is abolished in the mutants, thus establishing the structural basis for the triggering action of the residue in terms of the size, shape and the chemical nature of the side chain. The 19 crystal structures reported here, together with 11 previously reported MsRecA structures, provide further elaboration of the relation between the pitch of the 'inactive' RecA filament, the orientation of the C-terminal domain with respect to the main domain and the location of the switch residue. The low-temperature structures define one extreme of the range of positions the C-terminal domain can occupy. The movement of the C-terminal domain is correlated with those of the LexA-binding loop and the loop that connects the main and the N-terminal domains. These elements of molecular plasticity are made use of in the transition to the 'active' filament, as evidenced by the recently reported structures of RecA–DNA complexes. The available structures of RecA resulting from X-ray and electron-microscopic studies appear to represent different stages in the trajectory of the allosteric transformations of the RecA filament. The work reported here contributes to the description of the early stages of this trajectory and provides insight into structures relevant to the later stages.

1. Introduction

RecA protein plays a crucial role in recombination, repair and the SOS response (McGrew & Knight, 2003). The three-dimensional structure of the protein was first elucidated by X-ray analysis of the crystal structure of RecA from *Escherichia coli* (*Ec*RecA; Story *et al.*, 1992). A partially refined complex of the protein with ADP was also then reported (Story & Steitz, 1992). In common with RecA from other sources studied subsequently, *Ec*RecA crystallized in space group $P6_1$, with a RecA filament generated by a 6_1 screw axis as the basic unit of aggregation. Electron-microscopic studies showed that it corresponded to the RecA filament that exists prior to activation by ATP and subsequent DNA binding within it, often referred to as an 'inactive' filament (Flory *et al.*, 1984; Egelman & Stasiak, 1993; Yu *et al.*, 2001; VanLoock *et al.*, 2003). Thus, the crystal structures provide information not

only on the molecular structure of RecA but also on its filamentous aggregation under physiological conditions.

The next major advances in the structural study of RecA resulted from the X-ray analysis of the protein from *Mycobacterium tuberculosis* (*Mt*RecA) and its nucleotide complexes in this laboratory (Datta *et al.*, 2000; Datta, Ganesh *et al.*, 2003). This work resulted in a detailed characterization of RecA–nucleotide interactions. The two DNA-binding loops, the shorter L1 loop and the longer L2 loop, were not defined in the *Ec*RecA structures. They were defined in one or the other of the *Mt*RecA structures and hence their disposition in the filament could be characterized. The structures also brought to light differences between *Ec*RecA and *Mt*RecA, particularly in the charges on the filament surface and the interfilament interactions. Work on *Mt*RecA was followed by that on RecA from *M. smegmatis* (*Ms*RecA), a nonpathogenic homologue of *M. tuberculosis*. X-ray studies on *Ms*RecA and its nucleotide complexes provided further insights into RecA–nucleotide interactions, particularly regarding the movement of a critical residue, Gln196, on nucleotide binding (Datta, Krishna *et al.*, 2003). In a subsequent study, a fully ordered C-terminal domain was observed for the first time in the crystal structure of a complex of *Ms*RecA (Krishna *et al.*, 2006). In all previous studies, the C-terminal stretch designated as C* involving residues 331–349 had been disordered. This ordering of the C-terminal region was accompanied by the generation of a second ATP-binding site. In the filament, the site communicates with the first binding site of an adjacent molecule, suggesting a new route to allosteric regulation.

In the meantime, several new forms of *Ec*RecA and its nucleotide complexes were reported (Xing & Bell, 2004). A major finding to emanate from these structure determinations was the correlation between the changes in the helical pitch of the filament and the orientation of the C-terminal domain (C-domain) with respect to the main domain (M-domain). A telling example of this correlation was further provided by the structure of the protein from *Deinococcus radiodurans* (*Dr*RecA; Rajan & Bell, 2004). This observation was further elaborated by the structure analysis of several crystal forms of *Ms*RecA (Krishna *et al.*, 2007). The compatibility of the domain movement with the structures of Rad51 (Pellegrini *et al.*, 2002) and RadA (Wu *et al.*, 2004), the eukaryotic orthologues, was also explored. Most recently, the crystal structures of *E. coli* RecA–ssDNA and RecA–heteroduplex DNA filaments have been solved using covalently linked oligomers of truncated RecA molecules (Chen *et al.*, 2008). These structures provided the first high-resolution visualization of DNA-bound RecA filaments and led to valuable insights into the mechanism of action of RecA.

In addition to providing information on internal movements in the molecule relevant to protein action, the work on the nucleotide complexes of *Ms*RecA (Datta, Krishna *et al.*, 2003) strongly indicated that Gln196, which interacts with the bound nucleotide and is also the first residue in the DNA-binding loop L2, functions as a trigger for transmitting the effect of nucleotide binding to the DNA-binding region. Invariably, nucleotide binding causes movement of Gln196. In this scenario, studies of *Ms*RecA mutants with Gln replaced by Ala, Asn and Glu at position 196 were undertaken. In consonance with our identification of Gln196 as the switch residue, the mutant proteins do not exhibit enhanced ATPase activity in the presence of DNA (Manjunath & Muniyappa, unpublished results). Also, in the structures the mutants do not exhibit a directed movement of residue 196 on nucleotide binding. Unlike in the previous studies, X-ray data from the crystals of the mutants and their complexes were collected at low temperature, which in itself appeared to cause structural changes in the molecule. Subsequently, the wild-type protein and a couple of its complexes were analyzed using low-temperature data, while a couple of complexes of the mutants and one of the mutants were studied at room temperature. These structures, together with previously reported structures of *Ms*RecA, provide new insights into the internal movements in the RecA molecule with and without change in the pitch of the helical filament, in addition to confirming the role of Gln196 as a switch residue. A detailed comparative study of crystal structures from different sources leads to a clear elucidation of the nature and the extent of the plasticity of the molecule in 'inactive' filaments. The availability of the recently determined X-ray structures of DNA-bound oligomeric RecA constructs also permits an analysis of the extent to which this plasticity is used in the assembly of such 'active' filaments. The available native, nucleotide-bound and DNA-bound RecA structures, taken together, provide valuable insights into the trajectory of the allosteric transformations associated with the action of RecA.

2. Materials and methods

2.1. Construction of Gln196 variants and their purification

Mutants were introduced in *Ms*RecA to replace Gln196 with Ala, Glu or Asn using the sense–antisense method of site-directed mutagenesis (Li & Wilkinson, 1997) with *Ms*RecA cloned into vector pThioA (Novagen) as template. Primers were designed such that in addition to the desired mutation, a silent mutation creating a unique *Hin*dIII site was also introduced. Plasmids were screened for the desired mutation using the unique *Hin*dIII site. Amino-acid substitutions were further confirmed using DNA sequencing.

The plasmids were transformed into *E. coli* JC10289 (ΔRecA) strain and expressed under the *trp* promoter using IPTG induction. Wild-type *Ms*RecA protein and its variants were purified to homogeneity as described previously (Ganesh & Muniyappa, 2003). The protein concentrations were determined using Bradford's method. The RecA preparations were found to be free of any contaminating exonuclease or endonuclease activity.

2.2. Crystallization and data collection

The mutants and their complexes were crystallized from a hanging drop of 6 mg ml^{-1} protein in 80 m*M* citrate–phosphate buffer pH 6.9 containing 80 m*M* NaCl, 40 m*M* ammonium acetate, 20 m*M* sodium citrate and 6% polyethylene

glycol 3350 equilibrated against 30% polyethylene glycol 3350, 200 m*M* ammonium acetate in sodium citrate buffer pH 5.8. ADP, ATPγs and dATP obtained from Amersham Biosciences were used in cocrystallization experiments, in which 2 m*M* of

Table 1
X-ray crystal data, refinement and model statistics.

Values in parentheses are for the highest resolution shell. See Table 2 for nomenclature.

	*Ms*RecA II′	*Ms*RecA–dATP II′	*Ms*RecA IV	*Ms*RecA–ADP IV	*Ms*RecA–dATP IV	*Ms*RecA A IV	*Ms*RecA A–ADP IV	*Ms*RecA A–ATPγS IV	*Ms*RecA A–dATP IV	*Ms*RecA A–dATP II′
Space group	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$
Unit-cell parameters										
a (Å)	103.61	103.75	101.73	102.18	101.89	101.93	102.08	101.89	102.11	103.37
c (Å)	74.41	74.57	75.44	76.13	75.49	75.95	74.41	75.3	74.22	75.39
Resolution range (Å)	30–3.6	30–3.3	30–3.3	30–2.9	30–2.8	30–3.2	30–3.3	30–2.6	30–3.2	30–3.7
Matthews coefficient V_M (Å^3 Da^{-1})	3.0	3.1	3.0	3.0	3.0	3.0	3.0	2.9	2.9	3.1
Solvent content (%)	59.5	59.7	58.6	59.2	58.7	59.0	58.2	58.6	58.1	59.8
Unique reflections	5133 (761)	6986 (1008)	6726 (922)	10128 (1447)	11096 (1601)	7501 (1082)	6715 (947)	13807 (1995)	7364 (1077)	4582 (711)
Multiplicity	2.3 (2.3)	3.3 (3.3)	6.3 (5.3)	7.1 (6.7)	7.2 (7.1)	6.4 (6.4)	6.9 (5.8)	4.8 (4.0)	6.5 (6.4)	2.1 (2.0)
Data completeness (%)	95.8 (97.6)	100 (100)	99.0 (95.5)	99.9 (99.5)	100 (100)	99.8 (100)	99.4 (97.5)	100 (99.8)	99.8 (100)	92.0 (97.1)
Mean $I/\sigma(I)$	7.1 (1.6)	7.6 (2.1)	13.7 (4.1)	17.8 (3.7)	16.9 (3.9)	12.8 (4.2)	12.8 (3.4)	18.5 (3.0)	13.9 (4.2)	4.4(1.5)
R_{merge}† (%)	14.2 (49.6)	10.5 (49.0)	14.5 (40.7)	11.7 (50.5)	11.2 (50.9)	15.5 (43.1)	16.9 (48.2)	6.9 (37.1)	14.0 (43.6)	16.6 (48.1)
Refinement and model statistics										
R factor (%)	22.1	19.4	20.9	20.4	21.5	19.5	21.8	21.8	20.3	20.8
R_{free}‡ (%)	27.8	26.6	27.0	24.5	26.7	25.3	27.6	25.2	25.0	27.9
R.m.s. deviations from ideal										
Bonds (Å)	0.009	0.008	0.009	0.008	0.009	0.009	0.009	0.008	0.008	0.010
Angles (°)	1.7	1.6	1.8	1.7	1.8	1.6	2.1	1.6	1.8	1.7
Dihedral angles (°)	24.5	24.0	22.7	23.5	23.4	22.7	22.9	23.3	23.3	24.5
Improper angles (°)	1.2	2.3	1.2	0.9	2.2	1.0	1.4	1.8	2.3	3.0
Ramachandran plot statistics (% of residues)										
Favoured regions	84.4	84.4	84.6	90.2	93.6	87.9	82.3	93.5	88.6	85.2
Allowed regions	15.6	15.6	15.4	9.8	6.4	12.1	17.7	6.5	11.4	14.8
Disallowed regions	0	0	0	0	0	0	0	0	0	0
Ordered residues	6–159, 164–197, 212–330	6–159, 165–198, 212–331	5–200, 210–330	6–202, 209–334	6–198, 209–334	5–198, 207–331	5–200, 210–330	5–161, 165–198, 209–331	5–198, 209–334	6–160, 167–197, 212–330

	*Ms*RecA N IV	*Ms*RecA N–ADP IV	*Ms*RecA N–ATPγS IV	*Ms*RecA N–dATP IV	*Ms*RecA N–dATP II′	*Ms*RecA E IV	*Ms*RecA E–ATPγS	*Ms*RecA E–dATP IV	*Ms*RecA E II′
Space group	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$	$P6_1$
Unit-cell parameters									
a (Å)	101.69	102.17	102.03	102.09	103.62	101.36	107.08	101.95	103.48
c (Å)	75.78	76.02	75.02	75.23	75.71	74.71	72.99	75.15	75.06
Resolution range (Å)	30–3.1	30–3.1	30–2.9	30–3.0	30–3.5	30–3.0	30–3.1	30–2.5	30–3.25
Matthews coefficient V_M (Å^3 Da^{-1})	3.0	3.0	3.0	3.0	3.1	2.9	3.1	2.9	3.1
Solvent content (%)	58.7	59.2	58.5	58.7	60.1	57.8	61.3	58.6	59.7
Unique reflections	8075 (1093)	8220 (1172)	9902 (1418)	8662 (1263)	5943 (862)	8808 (1281)	8701 (1265)	15503 (2247)	7218 (1026)
Multiplicity	8.0 (5.5)	4.3 (3.1)	3.5 (3.2)	3.8 (3.6)	3.2 (3.2)	5.9 (5.8)	4.7 (4.5)	5.4 (4.8)	3.8 (3.8)
Data completeness (%)	98.5 (92.6)	99.3 (98.0)	99.5 (98.7)	95.8 (97.5)	99.9 (100)	99.6 (100)	99.2 (99.9)	99.9 (100)	99.0 (98.2)
Mean $I/\sigma(I)$	16.1 (3.5)	13.6 (2.6)	10.3 (2.1)	10.3 (2.4)	9.5 (2.2)	13.7 (3.0)	13.0 (3.0)	16.8 (3.0)	10.6 (2.4)
R_{merge}† (%)	13.6 (49.1)	10.7 (39.7)	12.8 (49.9)	13.0 (48.8)	12.7 (49.4)	14.2 (48.8)	11.6 (45.2)	8.2 (46.4)	8.9 (49.2)
Refinement and model statistics									
R factor (%)	22.4	20.4	22.1	21.9	20.9	21.9	21.8	21.0	22.8
R_{free}‡ (%)	27.0	24.9	26.3	27.1	27.6	26.4	27.9	25.1	28.7
R.m.s. deviations from ideal									
Bonds (Å)	0.009	0.009	0.008	0.009	0.009	0.008	0.010	0.007	0.009
Angles (°)	1.7	2.3	1.8	2.2	1.8	2.0	1.8	2.2	1.7
Dihedral angles (°)	23.7	22.3	23.9	23.4	24.5	23.0	23.0	23.1	24.1
Improper angles (°)	1.0	1.7	2.4	2.5	2.3	1.4	2.4	1.5	1.1
Ramachandran plot statistics (% of residues)									
Favoured regions	90.8	88.3	92.7	88.8	83.5	90.0	88.1	94.7	86.7
Allowed regions	9.2	11.7	7.3	11.2	15.5	10.0	11.9	5.3	13.3
Disallowed regions	0	0	0	0	1	0	0	0	0
Ordered residues	5–198, 207–334	5–199, 208–334	5–198, 209–334	5–198, 209–233, 238–330	6–159, 165–199, 212–330	5–198, 209–330	6–162, 165–198, 212–330	5–198, 210–331	6–160, 166–197, 212–330

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th intensity measurement of a reflection, $\langle I(hkl)\rangle$ is the average intensity value of that reflection and the summation is over all measurements. ‡ 10% of reflections were used for the R_{free} calculations.

the appropriate nucleotide and 2 m*M* $MgCl_2$ were included in the drop. Wild-type *Ms*RecA and its complexes were also crystallized under identical conditions for comparison. 30% ethylene glycol was used as a cryoprotectant for data collection at 100 K. X-ray diffraction data from the native as well as the complex crystals were collected at room temperature and low temperature using a 345 mm MAR imaging plate mounted on a Rigaku RU-200 X-ray generator. The crystal-to-detector distance was maintained at 250 mm with an oscillation angle of 1° for all crystals. The crystals diffracted to resolutions in the range 2.5–3.7 Å. The recorded images were processed and scaled using *MOSFLM* (Leslie, 2006; Powell, 1999) and *SCALA* (Evans, 1993) from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994). Data-collection statistics are given in Table 1.

2.3. Structure refinement

The protein atoms of a native *Ms*RecA structure (PDB code 1ubc) were used as the starting model for refinement, as all the structures are isomorphous to one another. The structures were refined in an identical manner using *CNS* (Brünger *et al.*, 1998). Initially, the structures were subjected to rigid-body refinement, treating the three domains (N, 1–30; M, 31–270; C, 271–330) of the protein as rigid groups. This was followed by positional refinement and simulated annealing. Both $2F_o - F_c$ and $F_o - F_c$ electron-density maps were calculated at this stage. These maps permitted unambiguous initial positioning of nucleotide and the interpretation of DNA-binding loop L1 (158–168) and some residues in loop L2 (196–211) in some instances. Model building was carried out using *Coot* (Emsley & Cowtan, 2004). Iterative model building and refinement were carried out until convergence of R and R_{free}. Grouped B factors in main chains and side chains for the protein residues were refined and the bulk-solvent correction was applied. The geometric parameters for ADP, ATPγs, dATP and PO_4 used in the refinement were obtained from the *PRODRG* database (Schüttelkopf & van Aalten, 2004). Water molecules were built into the electron-density maps where the peaks were visible at contours of at least 2.5σ in $F_o - F_c$ and 0.8σ in $2F_o - F_c$ electron-density maps. The *Ms*RecA E ATPγS data were twinned with a twinning fraction of 0.28 and twin law h, $-h - k$, $-l$. The data were detwinned using the detwin utility in *CNS* and used in refinement. Refinement parameters for all the structures are given in Table 1.

2.4. Geometric analysis and modeling

The stereochemical quality of the structures was validated using *MOLPROBITY* (Davis *et al.*, 2007). A distance of less than 3.6 Å between the donor and acceptor atoms and an angle of greater than 90° at the H atom were used as criteria for delineating hydrogen bonds. Contact distances were calculated using *CONTACT* from *CCP*4 (Collaborative Computational Project, Number 4, 1994). Structural alignments were carried out using *ALIGN* (Cohen, 1997). Figures were prepared using *PyMOL* (DeLano, 2002) and *MOLSCRIPT* (Kraulis, 1991).

3. Results

Crystals of the three mutants Q196A, Q196N and Q196E grew under conditions that were somewhat different from the conditions under which the wild-type crystals had been grown in previous studies (Datta, Krishna *et al.*, 2003; Krishna *et al.*, 2006, 2007). Crystals involving the mutants were flash-frozen for X-ray data collection, as this procedure significantly improved the resolution of the data compared with that of data collected previously at room temperature from *Ms*RecA crystals and their complexes. During the course of structure analysis, it appeared that the change in temperature itself caused some structural changes in the molecule. In order to avoid any ambiguities in the analysis, wild-type crystals were grown under conditions identical to those under which the crystals of the mutants had been grown. Data were collected at room temperature as well as at low temperature. Furthermore, two nucleotide complexes of the wild-type protein were also analyzed using low-temperature data. Crystals of the mutants and their complexes did not diffract well at room temperature. However, using trial and error, room-temperature data could be collected from crystals of the Q196E mutant and of the dATP complexes of Q196A and Q196N.

The crystal structures analyzed in the present study, all of which belong to space group $P6_1$, are listed in Table 1. All the structures refined using low-temperature data have resolutions in the range 2.5–3.3 Å. This range is somewhat better than those of the previously reported mycobacterial RecA structures. The room-temperature structures reported here have somewhat lower resolutions of 3.3 and 3.7 Å. In general,

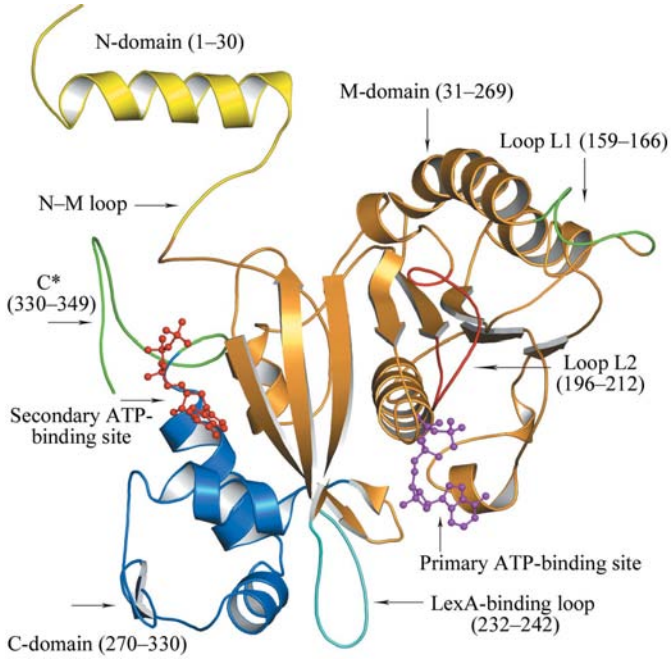

Figure 1
Molecular structure of RecA. The N-, M- and C-domains are coloured yellow, orange and blue, respectively. The C-terminal stretch designated as C*, the DNA-binding loops L1 and L2 and the LexA-binding loop are coloured differently. C* is most often disordered, while the three loops exhibit different extents of disorder in different structures.

mycobaterial RecA crystals diffract poorly. However, the resolutions attained are adequate to justify the conclusions derived from a comparative study of the structures.

3.1. Overall features

As in other structures reported previously, the 349-residue *Ms*RecA folds into three domains (Fig. 1). The N-terminal domain (N-domain), made up of residues 1–30, consists of a long α-helix followed by an extended region. The main domain (M-domain; residues 31–270) has a 'P-loop-containing NTPase fold' according to the SCOP classification (Murzin *et al.*, 1995) and is made up of an eight-stranded β-sheet flanked by α-helices. The C-terminal domain (C-domain; residues 271–349) is an independent globular unit containing three helices and three β-strands. The N-domain, together with part of the M-domain, is involved in filament formation. The C-domain is also involved in molecular aggregation and is known to have a regulatory role (Eggler *et al.*, 2003; Aihara *et al.*, 1997; Lusetti *et al.*, 2003). The M-domain is the business end of the molecule. It contains the nucleotide-binding region including the P-loop (residues 68–76), the DNA-binding loops L1 (residues 159–166) and L2 (residues 196–212) and the LexA-binding loop (residues 232–242). Different regions of the molecule, mainly loops, are known to exhibit disorder. In particular, the 20 C-terminal residues are only ordered in one crystal structure. L1 and L2 exhibit considerable conformational variability and are ordered to different extents in different structures. The

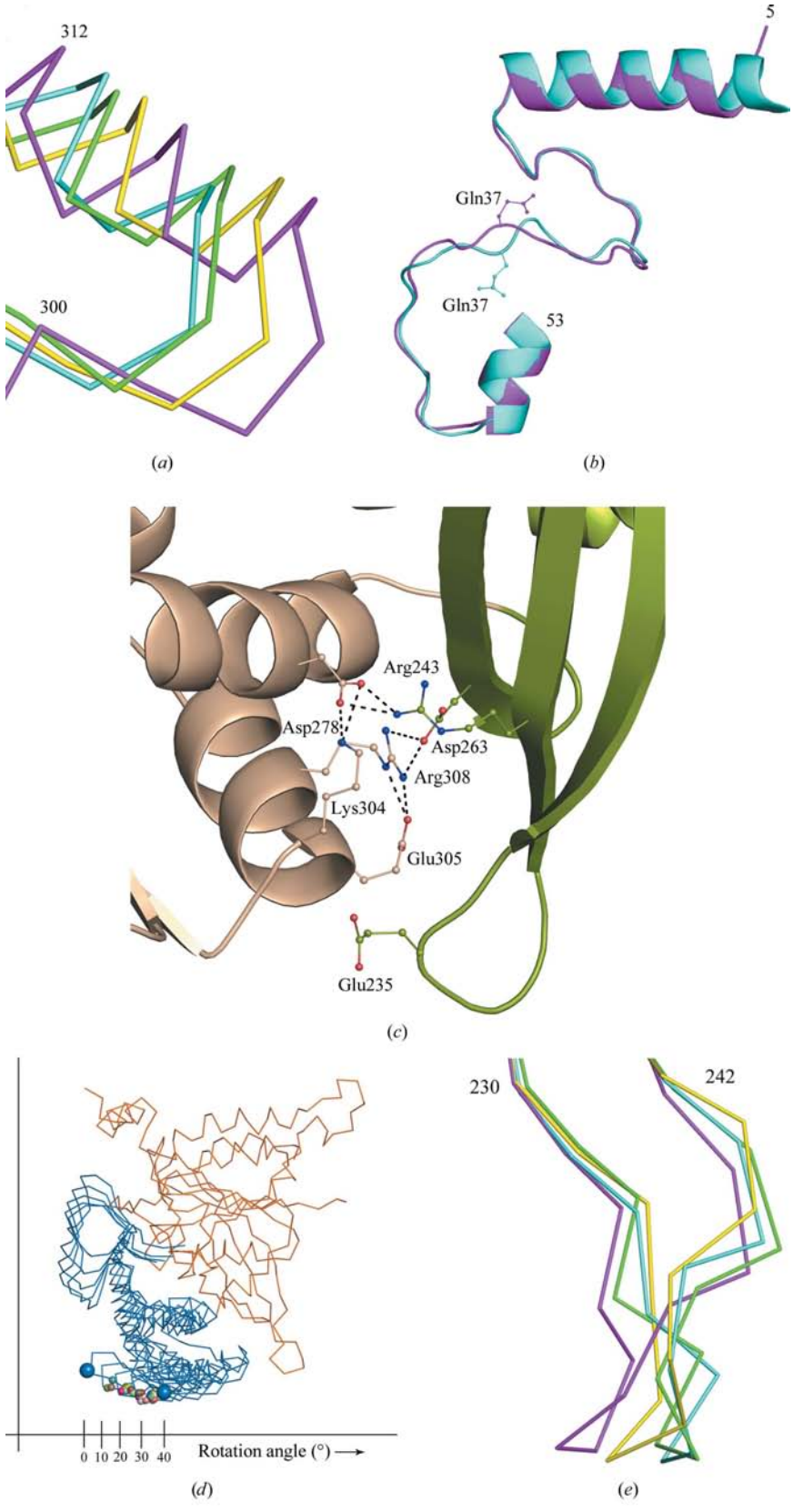

Figure 2
Movement of the C-domain and its consequences. (*a*) The relative positions of the region of the C-domain closest to the LexA-binding loop in forms I (cyan), II′ (yellow), III (green) and IV (magenta). The M-domain is used for superposition. Unless otherwise stated, forms I, II′, III and IV refer to wild-type native crystals. (*b*) The N–M loop in forms I (cyan) and IV (magenta). The side chain of Gln37, which flips, is also shown. (*c*) The charge cluster between the M- and C-domains in form IV. (*d*) The computed trajectory of the C-domain (blue) movement with respect to the M-domain (orange). The large balls represent the extreme positions of residue 297. The positions of the same residue in different crystal structures are shown as small balls. (*e*) Relative positions of the LexA-binding loop in forms I (cyan), II (yellow), III (green) and IV (magenta). The movement of the LexA-binding loop is in a direction inclined by about 50° to that of the C-domain.

ordered residues in the structures reported here are listed in Table 1.

3.2. Pitch of the filament and domain movement

11 crystal structures of *Ms*RecA and its complexes have been reported previously, all of which were determined at room temperature in this laboratory. On the basis of the unit-cell parameters, one of which corresponds to the pitch of the helical filament, they have been classified into the following three types: form I (a = 108.2–108.4, c = 73.0–73.9 Å), form II (a = 103.0–103.6, c = 75.1–77.1 Å) and form III (a = 102.9–103.7, c = 71.7–73.6 Å). The native crystals that were originally used for structure determination belonged to form III. The nucleotide complexes prepared under the same conditions also belonged to form III. Somewhat better crystals of native *Ms*RecA belonging to form I could be grown under slightly altered conditions. Crystals of nucleotide complexes prepared using these altered conditions belonged to either form I or form II. In one case, that involving a complex with dATP, form I as well as form II crystals could be obtained using the altered conditions. Interestingly, form I readily transforms to form II with a slight change in water content when the environmental humidity is reduced. In this context, it is pertinent to note that the water content of form I crystals is around 62%, while that of form II crystals varies between 59.3% and 60.9%. The water content of form III crystals is still lower at around 58.5%.

The C-domain has a somewhat different orientation with respect to the M-domain in the three forms. The two domains are not linked at the edge of either (Krishna *et al.*, 2007; Fig. 1). Therefore, when the terminal region of the C-domain moves away from the bulk of the M-domain, the opposite ('bottom') end of the C-domain moves toward the LexA-binding loop of the M-domain. The C-terminal region is farthest from and the opposite end closest to the M-domain in form I crystals. The opposite is true of form II crystals. The difference in the angle of orientation between the two domains ranges from 7.6° to 10.2° when form I and form II crystals are compared. The orientation of the C-domain is between these two extremes in form III crystals.

The crystals used in the present study belong to form II at room temperature. However, at low temperature, when they are flash-frozen, crystals of the wild-type protein, the mutants and their complexes transform to what may be best described as form IV, except in one case. Form IV crystals have unit-cell parameters in the range a = 101.4–102.2, c = 74.2–76.1 Å. As illustrated in Fig. 2(*a*), they are characterized by a larger C-domain movement than that that occurs between forms I and II. The angular separation of the domain from its position in form I and form II is about 15° and 6°, respectively. This represents the largest C-domain movement in RecA crystals so far; the largest movement in the other direction is in *Dr*RecA. The only structure determined at low temperature that does not belong to form IV is the ATPγS complex of the Q196E mutant. The crystals of this complex have unit-cell parameters that are close to those of form I, but the orientation of the C-domain is between those found in form III and form II.

Table 2
*Ms*RecA structures.

Description	Crystallization condition	PDB code
*Ms*RecA structures solved in the present study		
*Ms*RecA II′	Condition 3	2zr7
*Ms*RecA–dATP II′	Condition 3	2zrp
*Ms*RecA IV	Condition 3	2zrn
*Ms*RecA–ADP IV	Condition 3	2zro
*Ms*RecA–dATP IV	Condition 3	2zrm
*Ms*RecA A IV	Condition 3	2zrh
*Ms*RecA A–ADP IV	Condition 3	2zri
*Ms*RecA A–ATPγS IV	Condition 3	2zrj
*Ms*RecA A–dATP IV	Condition 3	2zrk
*Ms*RecA A–dATP II′	Condition 3	2zrl
*Ms*RecA N IV	Condition 3	2zrc
*Ms*RecA N–ADP IV	Condition 3	2zrd
*Ms*RecA N–ATPγS IV	Condition 3	2zre
*Ms*RecA N–dATP IV	Condition 3	2zrf
*Ms*RecA N–dATP II′	Condition 3	2zrg
*Ms*RecA E II′	Condition 3	2zrb
*Ms*RecA E IV	Condition 3	2zr0
*Ms*RecA E–ATPγS	Condition 3	2zra
*Ms*RecA E–dATP IV	Condition 3	2zr9
*Ms*RecA structures reported previously		
*Ms*RecA III	Condition 1	1ubc
*Ms*RecA–ADP III	Condition 1	1ube
*Ms*RecA–ATPγS III	Condition 1	1ubf
*Ms*RecA–dATP III	Condition 1	1ubg
*Ms*RecA–dATP I	Condition 2	2g88
*Ms*RecA–dATP II	Condition 2	2odn
*Ms*RecA–ATPγS II	Condition 2	2odw
*Ms*RecA–LH II	Condition 2	2oe2
*Ms*RecA–DP I	Condition 2	2oep
*Ms*RecA(SSB) III	Condition 2	2oes
*Ms*RecA I	Condition 2	2ofo

The room-temperature wild-type crystals grown in the present study, which belongs to form II, are hereafter referred to as *Ms*RecA II′. The prime is used to distinguish them from the form II crystals grown under slightly different conditions reported in the previous study. The dATP complex grown and studied under identical conditions is referred to as *Ms*RecA II′–dATP. The low-temperature wild-type form IV crystals and their complexes are referred to as *Ms*RecA IV, *Ms*RecA–dATP IV and *Ms*RecA–ADP IV. The three mutant crystal structures refined using low-temperature data are designated *Ms*RecA N IV, *Ms*RecA A IV and *Ms*RecA E IV. When referring to the complexes, the symbols for the nucleotides are added to the name. The low-temperature ATPγS complex of the Q196E mutant is referred to as *Ms*RecA E–ATPγS as there is confusion regarding the crystal form. The mutant structures, including complexes, determined at room temperature belong to form II′. Therefore, when referring to them, II′ is used in the names instead of IV. For clarity, the designated names of the *Ms*RecA structures determined as part of the present study, together with those of previously reported structures, are listed in Table 2. They form an extensive database for analysis.

3.3. Trajectory of domain movement and plasticity of the molecule

A striking result that emanates from a critical examination of the 30 structures involving *Ms*RecA is undoubtedly the

plasticity resulting from the movement of the C-domain as a whole with respect to the M-domain. In terms of the mutual orientation of the two domains, form I represents one extreme with the 'bottom' half of the C-domain farthest from the LexA-binding loop and form IV the other extreme with the two structural features closest to each other. Form III and form II have a mutual orientation of the two domains that is between the two extremes. The movement from form I to form IV is accompanied by changes in the interdomain interactions. The hydrogen bonds present in form I include those made by the side chain of Gln37 in the loop connecting the M- and N-domains with Ile55 of the M-domain and Ile331 of the C-domain (Gln37 NE···Ile55 O and Gln37 NE···Ile331 O). These are broken during the movement from form I to form IV, with an interesting structural consequence. As shown in Fig. 2(*b*), the side chain of Gln37 flips over to the opposite side, with a movement as large as 3.5 Å in the C^{α} position. This movement ripples along the entire loop. While going from form I to form IV, Lys304, Glu305 and Arg308 of the C-domain move a couple of angstroms closer to residues Glu235, Arg243 and Asp263 of the M-domain, creating a charge cluster in the interface between the M- and C-domains (Fig. 2*c*). The newly formed interdomain hydrogen bonds in this region are Arg243 NH2···Asp278 OD1, Arg243 NH2···Asp278 OD2, Arg308 NH2···Asp263 OD1, Arg308 NH1···Asp263 OD1, Arg308 NH2···Glu305 OE2, Arg308 NE···Glu305 OE2, Lys304 NZ···Asp278 OD1, Lys304 NZ···Asp278 OD2 and Asp263 O···Ser271 N.

Among the known structures of RecA, *Dr*RecA has the C-domain farthest from the LexA-binding loop. The position of the domain in the various *Ec*RecA structures lies between those in *Dr*RecA and *Ms*RecA form I. The disposition of the domain in the various *Mt*RecA structures is similar to that in *Ms*RecA form I. Interestingly, the C-domains in different structures can be related to one another reasonably well by rotations about a common axis. The positions of the C-domain at periodic intervals when it is rotated about the common axis starting from the position in *Dr*RecA to a position where unacceptable steric contacts with the M-domain sets in are illustrated in Fig. 2(*d*). The C-domains observed in different RecA structures fall at various points along this trajectory. Among the known structures, the angular separation of the C-domain between the two extremes, namely *Dr*RecA and form IV of *Ms*RecA, is over 37°. The C-domain cannot move much closer to the LexA-binding loop from its position in form IV on account of steric hindrance. Calculations with an ordered C* (residues 331–349) indicate that C* would clash with the loop connecting the M- and N-domains if the C-domain moves in the opposite direction beyond its position in *Dr*RecA. Thus, the positions of the domain in *Dr*RecA and form IV *Ms*RecA appear to approximately define the opposite ends of the range of possible C-domain movement in 'inactive' filaments.

Another region which exhibits structural variability among the crystal forms is the LexA-binding loop. The formation of the charge cluster involving the C- and M-domains when going from form I to form IV results in a movement towards each other of the bottom half of the C-domain and the region of the C-domain containing the LexA-binding loop (Fig. 2*e*). However, a steric clash between Glu235 and Glu305 does not permit movement of the loop in a direction opposite to that of the C-domain. Consequently, the tip of the loop moves in a direction inclined by nearly 50° with respect to the direction of the movement of the region of the C-domain closest to the LexA-binding loop. The DNA-binding loops L1 and L2 also exhibit considerable flexibility. They, and particularly L2, are wholly or partly disordered in many structures. However, the variability in the conformation of these loops does not appear to be correlated with the movement of the C-domain.

To sum up, the 30 structures involving *Ms*RecA, including the 19 structures reported here and the 13 crystal structures of RecA from other sources, help to elucidate the plasticity of the RecA molecule. The most important element of this plasticity is the ability of the C-domain to undergo systematic rotation with respect to the M-domain. The plasticity also involves movement of the LexA-binding loop and changes in the

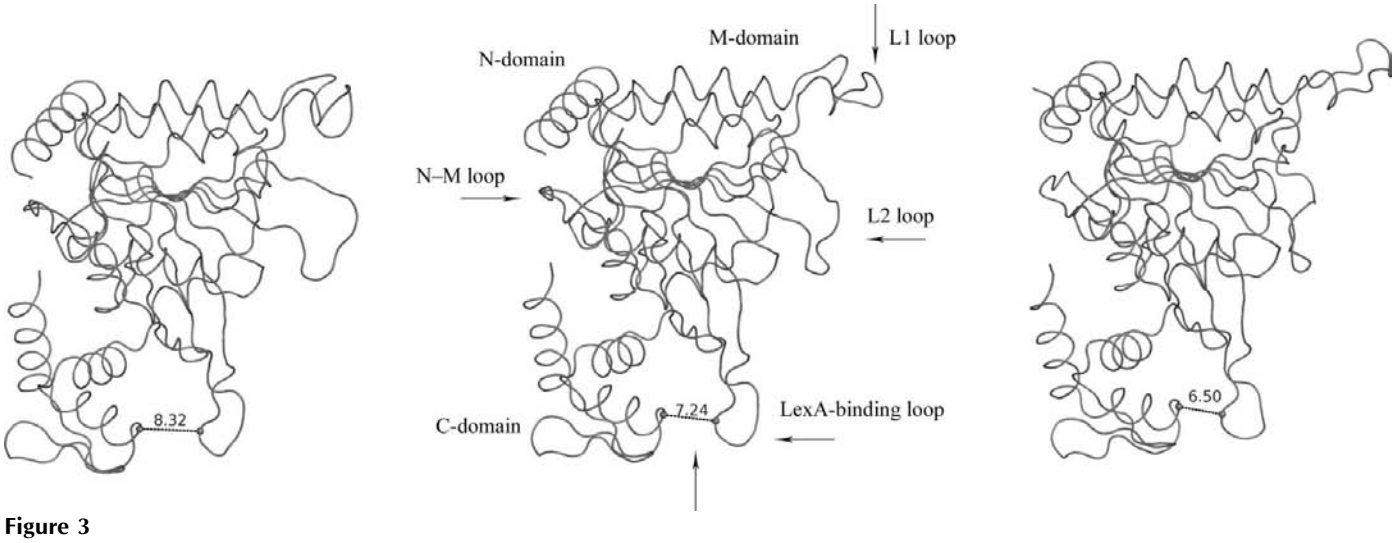

Figure 3
Snapshots in the trajectory of transformation from form I to form IV. See text for details.

conformation of the loop that connects the M- and N-domains. These three elements of plasticity exhibit a measure of correlation between them. The flexibility of the DNA-binding loops, however, does not appear to be correlated with that of these three elements.

Transformations from form I to form II and from form II to form IV have already been experimentally observed in *Ms*RecA crystals. In order to explore the possible trajectory of transformation from one extreme to the other, form I was morphed to form IV using the Yale Morph Server (Krebs & Gerstain, 2000). The algorithm used for morphing interpolates the coordinates in a chosen number of steps to evenly reduce the distance between the starting and final structures, while carrying out energy minimization at each step in order to obtain energetically and stereochemically acceptable states. Three representative snapshots during the transition from form I to form IV are shown in Fig. 3. Admittedly, only the experimentally determined initial and final states are entirely reliable. However, the exercise demonstrates the possibility of a smooth transition between the two states and provides insights into the correlated movement of the different segments of the molecule during the transition.

3.4. Movement of the switch residue and its relation to the C-domain movement

Another movement that has received particular attention is that of Gln196. It was indeed its suggested role as a switch residue in transmitting the information on nucleotide binding to the DNA-binding regions that constituted the motivation for studying mutants involving this residue. The present study helps to shed light on its movement as a function of filament assembly even in the absence of a nucleotide ligand. The relative location of the C^{α} position of the residue in the four forms when the other three structures are superposed on the form I structure is shown in Fig. 4. The relative positions of Gln196 in forms I, II and IV exhibit good correlation with the movements of the C-domain in these forms, but that in form III does not. It appears that form III stands apart from the other forms. Form I and form II grow under the same condition (condition 2). Form II′ crystals grown under condition 3 readily transform to form IV on reduction of temperature. Admittedly, data have not been collected from form III crystals at low temperature. However, it is interesting to note that form III has the lowest solvent content of the three room-temperature forms. It also has the shortest *c* dimension and hence the shortest pitch for the helical filament of the four forms.

Not only the native crystals but also the complexes with ATPγS, dATP and ADP obtained in crystallization condition 1 belong to form III. All three complexes exhibit a shift of Gln196 C^{α} by more than 1 Å from its position in the native crystals. It is indeed this observation that led to the identification of this residue as a switch in the transmission of the effect of nucleotide binding to the DNA-binding regions. Crystallization condition 2 yields form I native crystals and form I and form II derivative crystals. In form I derivative crystals, the position of Gln196 C^{α} is displaced towards the binding site from its position in native form I crystals by more than 1 Å in each case. Native form II′ crystals and those of a complex with dATP grown under crystallization condition 3 are available. Here, Gln196 C^{α} in the complex again moves by more than 1 Å with respect to its position in the native crystals. In form IV crystals, movement of Gln196 C^{α} on complexation is also observed; it moves by 0.8 Å in the ADP complex and by 0.9 Å in the dATP complex. Thus, in every single instance among crystals grown under the same conditions Gln196 moves towards the binding site on nucleotide binding.

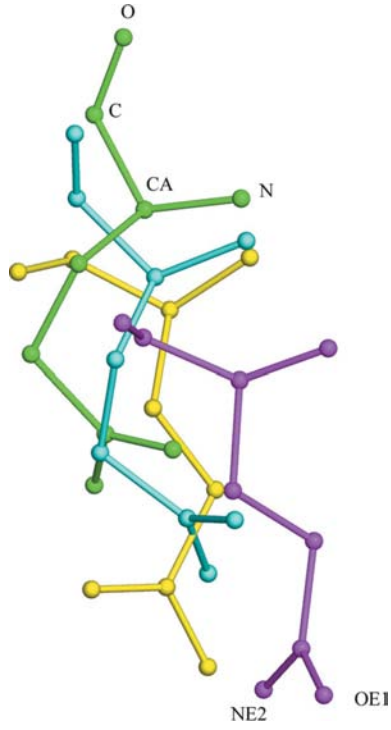

Figure 4
The relative positions of Gln196 in forms I (cyan), II (yellow), III (green) and IV (magenta). The M-domain was used for superposition.

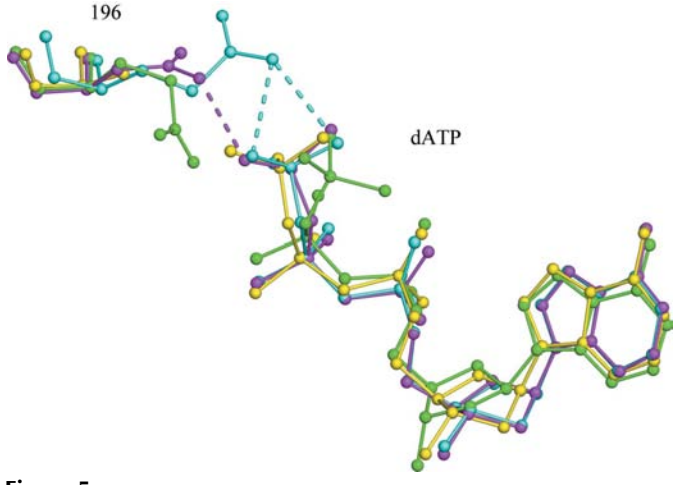

Figure 5
Superposition of residue 196 and the nucleotide in the dATP complexes *Ms*RecA–dATP IV (cyan), *Ms*RecA A–dATP IV (yellow), *Ms*RecA N–dATP IV (magenta) and *Ms*RecA E–dATP IV (green).

Crystals grown under condition 2 provides an interesting lead on the possible correlation between the movement of Gln196 and that of the C-domain. Under this condition only form I native crystals are formed. However, of the four complex crystals grown under the same conditions, two belong to form I itself, while the remaining two belong to form II. Thus, in the case of the latter complexation results not only in the expected movement of Gln196, but also in a substantial concomitant movement of the C-domain. Even when there is no difference in the crystal form, as when comparing form I native crystals and the two complex crystals belonging to the same form, there is a C-domain movement in the same direction, but the magnitude of the change is very low at 1.2–1.5°. The same is true when form III native crystals and their nucleotide complexes are compared, with a movement of 1.3–2.5°. Native form II′ crystals and crystals of one of its complexes grown under identical conditions (condition 3) are available. Here, the movement of the C-domain is 1°. The C-domain movement in the crystals of the two available complexes belonging to form IV from the position in the form IV native crystals is 1° or less and the direction of movement does not correlate well with that in the other forms. This could well be because, as indicated previously, the C-domain is already at the edge of its range of possible movement in form IV. Thus, on the whole nucleotide binding appears to lead to a movement of the C-domain in the same direction. When binding results in change of form, as in form I to form II, the movement is large. When it does not, the movement is 1° or 2° in most cases. When one considers the movement of a domain as a whole, these values could be significant. Thus, there appears to be a correlation between the displacement of residue 196 resulting from nucleotide binding and the movement of the C-domain.

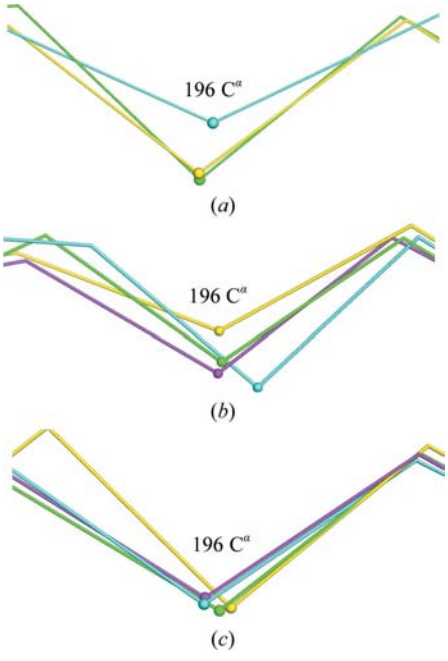

Figure 6
Movement of C^{α} of residue 196 with respect to the apo structures (cyan) and in complexes with ADP (yellow), ATPγS (magenta) and dATP (green) in crystals involving (*a*) wild-type *Ms*RecA, (*b*) Q196A and (*c*) Q196N.

3.5. Internal movements in the mutants and their complexes

All three native mutant crystals belong to form IV, as do the crystals of the three nucleotide complexes of Q196A and Q196N. The complexes of Q196E behave somewhat anomalously. The ATPγS derivative moves to a form which has unit-cell parameters corresponding to form I and a C-domain orientation between those in form III and form II. In the dATP complex, residue 196 exhibits some movement but, as illustrated in Fig. 5, the location of the γ-phosphate in the ligand is different from that observed in all other relevant structures. This anomalous behaviour of the complexes of Q196E could well be a consequence of the negative charge on the side chain, which is located close to the negatively charged phosphate group of the nucleotide ligand. In any case, the complexes of this mutant cannot be meaningfully compared with those of the other mutants and the wild-type protein.

The locations of residue 196 in form IV wild-type and complex crystals as well as in Q196A and Q196N and their complexes are shown in Fig. 6. In the wild-type complexes, residue 196 exhibits the expected shift. However, no systematic shifts are observed in the complexes of the mutants. In the complexes involving Q196A the residue moves in the opposite direction. The residue moves in different directions in the three complexes of Q196N. As mentioned previously, the complexes of Q196E exhibit anomalous behaviour. This emphasizes the criticality of the geometry and the chemical nature of the side chain. Ala is a small nonpolar residue and, as illustrated in Fig. 5 in the case of dATP complexes, it cannot form a hydrogen bond to the nucleotide. Asn has the same chemical nature as Gln, but the side chain is slightly shorter. It makes only one hydrogen bond to the γ-phosphate, whereas the side-chain amide group of Gln, which is centrally located with respect to two O atoms of the γ-phosphate, is involved in two such hydrogen bonds. Glu has the same size as Gln, but has a different chemical nature. In fact, understandably, its side chain and the γ-phosphate move away from each other. None of the three residues in the mutants exhibit the expected movement. Thus, Gln is presumably the only residue which has the required size, shape and chemical nature.

Interestingly, the correlated movement of residue 196 and the C-domain is relatively unaffected by the mutations. As in the case of the wild-type protein, crystals of the mutants and their complexes grown under condition 3 belong to form IV at low temperature. Data could be collected at room temperature from only three crystals: native Q196E and the dATP complexes of Q196A and Q196N. Like wild-type crystals handled at room temperature, they belong to form II′. When the room-temperature (form II′) and the low-temperature (form IV) structures of the three crystals were compared, it was found that C^{α} of residue 196 moves in the expected direction by 0.5–1.6 Å, while the C-domain moves by 6.5–8.6°. These movements are qualitatively similar to those exhibited by wild-type crystals when going from form II′ to form IV.

Thus, mutation at residue 196 affects the movement in response to nucleotide binding but not those related to the change in filament dimensions.

A comparison of the nucleotide-binding region demonstrates that mutation at residue 196 does not introduce any major change in the region. In the mutants, the r.m.s. deviation in C^{α} positions of the 14 residues involved in binding from those in the wild-type protein varies between 0.35 and 0.41 Å. Even deviations in the side chains are not large. These deviations are as expected largest in the side chain of the mutated residue. The near-conservation of the binding site on mutation further suggests that the major difference between the wild-type protein and the mutants is only in the movement of residue 196 on nucleotide binding and the consequent biological effects.

3.6. Compatibility with the structure of the active filament

Until very recently, much of the structural information on DNA-bound 'active' filaments was derived from electron-microscopic studies (Flory *et al.*, 1984; Egelman & Stasiak, 1993; Yu *et al.*, 2001; VanLoock *et al.*, 2003). The recent crystal structure determinations of *Ec*RecA–DNA filaments, while essentially conforming to the results of electron-microscopic studies, provide a near-atomic resolution picture of the 'active' filament and RecA–DNA interactions (Chen *et al.*, 2008). The filaments are made up of tetramers, pentamers and hexamers of covalently linked RecA molecules. A low-resolution crystal structure of the tetramer without bound DNA has also been determined. All the structures contain a bound ATP analogue. 18 C-terminal residues corresponding to the C* stretch were removed from each molecule. Furthermore, to prevent polymerization, the N-domain was deleted from the N-terminal molecule of each oligomer and specific mutations were introduced in the C-terminal molecule. However, these truncations and mutations do not prevent the formation of biologically relevant filaments similar to those observed in electron micrographs.

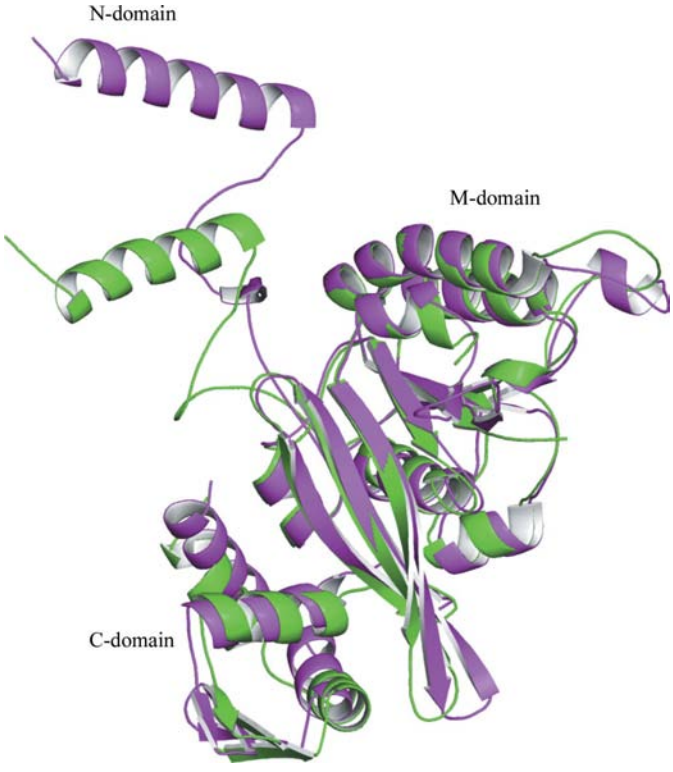

Figure 7
The position of the N-domain with respect to the rest of the molecule in 'inactive' (green) and 'active' (magenta) filaments.

The intrinsic plasticity of the RecA molecule deduced from the structures of the proteins from *M. smegmatis* and other sources is effectively made use of in the transition from the 'inactive' filament to the 'active' filament. In the 'active' filament the C-domain moves by about 20° within the range defined in the studies reported here. The established flexibility of the loop connecting the M- and N-domains has also been made use of in the transition from the 'inactive' filament to the 'active' filament. The disposition of the N-domain with respect to the M-domain is substantially different in two types of filaments (Fig. 7). The transition from one to the other is achieved through a conformational change in the loop that connects them.

The DNA-binding loops L1 and L2 present an interesting situation. They exhibit extraordinary conformational flexibility and often disorder in 'inactive' filaments (Story *et al.*, 1992; Story & Steitz, 1992; Datta, Krishna *et al.*, 2003; Rajan & Bell, 2004; Xing & Bell, 2004; Krishna *et al.*, 2006, 2007). However, they are highly ordered and their conformation is substantially homogeneous in 'active' filaments. This is presumably caused by DNA binding in a process roughly similar to the recently observed convergence of paratope conformation in related germline antibodies on antigen binding (Sethi *et al.*, 2006). For reasons which are not clear, a similar situation is also obtained in the case of the LexA-binding loop.

4. Discussion

The tertiary and the quaternary structures of RecA and its action are characterized by a high degree of complexity and subtlety. Its activity cycle involves binding to ATP, changing to an 'active' form, binding ssDNA and then dsDNA, hydrolysis of ATP to ADP and reversion to the 'inactive' form. The activity cycle is triggered by binding to ATP. The results reported here, particularly the studies on the mutants, establish the structural basis of the triggering mechanism with Gln196 as the switch. Mutants involving residue 196 abolish the enhancement of ATPase activity in the presence of DNA, which is crucial for the biological action of RecA. Studies on *Ec*RecA have previously demonstrated the loss of the ATP-induced increase in DNA-binding affinity on mutation of this residue (Kelley & Knight, 1997). The extensive work on *Ms*RecA and its nucleotide complexes has conclusively demonstrated that in the wild-type structures nucleotide binding results in the movement of residue 196 in a specific direction which apparently triggers the activity cycle. Such movements do not occur when Gln is mutated to Ala, Glu or

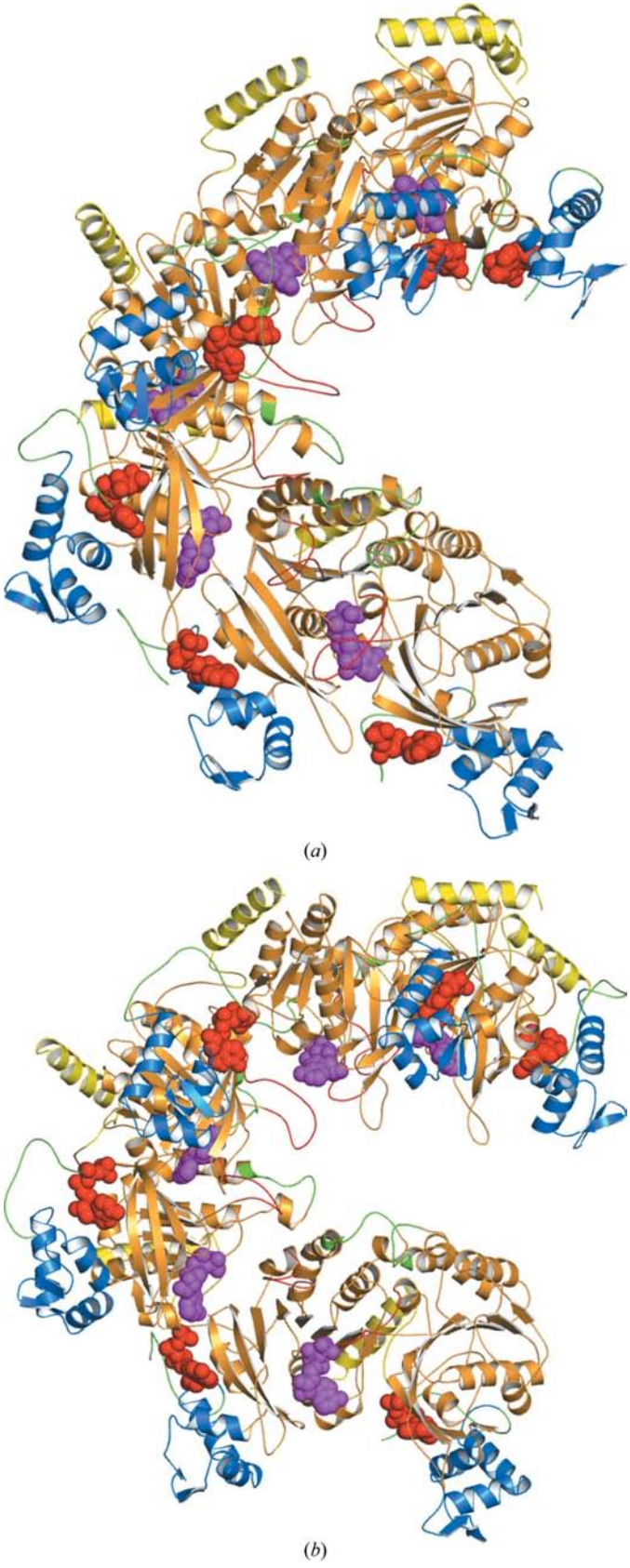

Figure 8
Six consecutive subunits in (*a*) the modelled 'active' filament and (*b*) the 'inactive' filament. The colour code for different regions of the molecule is the same as that used in Fig. 1. The ATP molecules at the primary site and the secondary site are shown in magenta and red, respectively.

Asn. Thus, in terms of size, shape and chemical nature only Gln appears to be suitable for this specific purpose.

The filaments found in the crystal structures of DNA-bound RecA (Chen *et al.*, 2008) differ from natural 'active' filaments in a couple of respects. The adjacent molecules in the former are covalently linked and are also of limited length, unlike natural filaments. Perhaps more importantly, the molecules in the crystal filaments do not contain the C-terminal stretch C*, which has been shown to be important in regulation (Eggler *et al.*, 2003; Aihara *et al.*, 1997; Lusetti *et al.*, 2003). Nevertheless, these filaments undoubtedly provide an excellent representation of RecA–DNA interactions and the organization of molecules in the 'active' filament. However, for purposes of comparison, a model of the 'active' filament containing six molecules was generated using information from the crystal structure (Chen *et al.*, 2008). One molecule from the middle of the RecA–dsDNA complex was chosen as the motif. The C* stretch and the ATP molecule located at the nucleotide-binding site generated concurrently with the ordering of C*, as observed in the *Ms*RecA–dATP complex (Krishna *et al.*, 2006), were added onto the motif by superposing the C-domains of the two molecules. The model of the fully ordered RecA molecule thus generated was used to construct a model of the six-molecule filament (Fig. 8*a*) employing the helical parameters derived from the structure of the RecA–dsDNA complex. A model of the 'inactive' filament involving the whole molecule was similarly constructed (Fig. 8*b*). This model is understandably very similar to the filaments observed in the crystal structures of RecA from different sources.

We have shown previously how the nucleotide at the primary ATP-binding site can communicate with that at the secondary ATP-binding site, generated by the ordering of C*, in an adjacent molecule in the 'inactive' filament through conserved residues in a possible route to allosteric regulation (Krishna *et al.*, 2006). We have also suggested that the two binding sites could be connected in the active filament by a

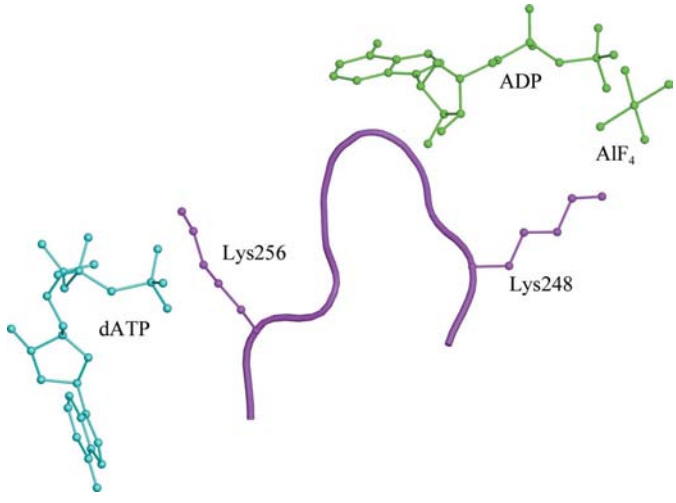

Figure 9
The locations of dATP at the secondary binding site, as in *Ms*RecA dATP I, and that of ADP·AlF_4 in an adjacent molecule, as in the *Ec*RecA–dsDNA complex, about loop 248–257 (in *Ec*RecA numbering) in the 'active' filament.

highly conserved loop (248–257 in *Ec*RecA numbering) between two β-strands. This connectivity exists in the model of the 'active' filament based on the crystal structure of the RecA–dsDNA complex. Lys256 (in *Ec*RecA numbering) belonging to this stretch can interact with the ATP in the second nucleotide-binding site, while Lys248 (in *Ec*RecA numbering), which is implicated in cooperative DNA binding (Nguyen *et al.*, 1993), is situated close to the primary nucleotide-binding site in an adjacent molecule in the 'active' filament (Fig. 9). Another interesting feature of the modelled 'active' filament is the proximity of C* in the molecule to the C-domain of an adjacent molecule in the filament. Extensive mutational studies involving residues in the C-terminal stretch constituting C* have indicated a modulating role for this region of the molecule (Eggler *et al.*, 2003; Aihara *et al.*, 1997; Lusetti *et al.*, 2003). Indeed, a regulatory role has been suggested for the C-domain as a whole. The proximity of the C-domains of the adjacent molecules in the filament, mediated through C*, could possibly provide a structural basis for this regulation.

Computations using the Yale Morph Server indicate that the 'inactive' filament could readily transform into the 'active' one without any break or unacceptable steric clashes in the trajectory of transformation. This appears to suggest that there could be a smooth transition from the 'inactive' to the 'active' filament using the plasticity of the molecule elaborated earlier in this paper. The different crystal structures of RecA from various sources, electron-microscopic studies on the 'inactive' and the 'active' RecA filaments and the recently reported structures of RecA–DNA complexes provide snapshots along this trajectory of transformation from inactive to active filaments. The results reported here are particularly relevant to the early stages of this trajectory; they also relate to and provide valuable insights into the later stages.

The diffraction data were collected at the X-ray Facility for Structural Biology at the Molecular Biophysics Unit supported by the Department of Science and Technology, Government of India. The Supercomputer Education and Research Centre of the Institute and the graphics facility supported by the Department of Biotechnology (DBT) were used for computations. MV is supported by a Distinguished Biotechnologist award of the DBT.

References

Aihara, H., Ito, Y., Kurumizaka, H., Terada, T., Yokoyama, S. & Shibata, T. (1997). *J. Mol. Biol.* **274**, 213–221.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chen, Z., Yang, H. & Pavletich, N. P. (2008). *Nature (London)*, **453**, 489–494.
Cohen, G. E. (1997). *J. Appl. Cryst.* **30**, 1160–1161.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Datta, S., Ganesh, N., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2003). *Proteins*, **50**, 474–485.
Datta, S., Krishna, R., Ganesh, N., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2003). *J. Bacteriol.* **185**, 4280–4284.
Datta, S., Prabu, M. M., Vaze, M. B., Ganesh, N., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2000). *Nucleic Acids Res.* **28**, 4964–4973.
Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). *Nucleic Acids Res.* **35**, W375–W383.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific LLC, San Carlos, California, USA. http://www.pymol.org.
Egelman, E. H. & Stasiak, A. (1993). *Micron*, **24**, 309–324.
Eggler, A. L., Lusetti, S. L. & Cox, M. (2003). *J. Biol. Chem.* **278**, 16389–16396.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.
Flory, J., Tsang, S. S. & Muniyappa, K. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 7026–7030.
Ganesh, N. & Muniyappa, K. (2003). *Proteins*, **53**, 6–17.
Kelley, J. A. & Knight, K. L. (1997). *J. Biol. Chem.* **272**, 25778–25782.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Krebs, W. G. & Gerstain, M. (2000). *Nucleic Acids Res.* **28**, 1665–1675.
Krishna, R., Manjunath, G. P., Kumar, P., Surolia, A., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2006). *Nucleic Acids Res.* **34**, 2186–2195.
Krishna, R., Prabu, J. R., Manjunath, G. P., Datta, S., Chandra, N. R., Muniyappa, K. & Vijayan, M. (2007). *J. Mol. Biol.* **367**, 1130–1144.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Li, S. & Wilkinson, M. F. (1997). *Biotechniques*, **23**, 588–590.
Lusetti, S. L., Shaw, J. J. & Cox, M. M. (2003). *J. Biol. Chem.* **278**, 16381–16388.
McGrew, D. A. & Knight, K. L. (2003). *Crit. Rev. Biochem. Mol. Biol.* **38**, 385–432.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Nguyen, T. T., Muench, K. A. & Bryant, F. R. (1993). *J. Biol. Chem.* **268**, 3107–3113.
Pellegrini, L., Yu, D. S., Lo, T., Anand, S., Lee, M., Blundell, T. L. & Venkitaraman, A. R. (2002). *Nature (London)*, **420**, 287–293.
Powell, H. R. (1999). *Acta Cryst.* D**55**, 1690–1695.
Rajan, R. & Bell, C. E. (2004). *J. Mol. Biol.* **344**, 951–963.
Schüttelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Cryst.* D**60**, 1355–1363.
Sethi, D. K., Agarwal, A., Manivel, V., Rao, K. V. S. & Salunke, D. M. (2006). *Immunity*, **24**, 429–438.
Story, R. M. & Steitz, T. A. (1992). *Nature (London)*, **355**, 374–376.
Story, R. M., Weber, I. T. & Steitz, T. A. (1992). *Nature (London)*, **355**, 318–325.
VanLoock, M. S., Yu, X., Yang, S., Lai, A. L., Low, C., Campbell, M. J. & Egelman, E. H. (2003). *Structure*, **11**, 187–196.
Wu, Y., He, Y., Moya, L. A., Qian, X. & Luo, Y. (2004). *Mol. Cell*, **15**, 423–435.
Xing, X. U. & Bell, C. E. (2004). *J. Mol. Biol.* **342**, 1471–1485.
Yu, X., Jacobs, S. A., West, S. C., Ogawa, T. & Egelman, E. H. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 8419–8424.