Conformational flexibility revealed by the crystal structure of a crenarchaeal RadA

Antonio Ariza, Derek J. Richard[1], Malcolm F. White[1] and Charles S. Bond*

Division of Biological Chemistry and Molecular Microbiology, School of Life Sciences, University of Dundee, Dow St, Dundee, DD1 5EH, UK and [1]Centre for Biomolecular Sciences, University of St Andrews, North Haugh, St Andrews, KY16 9ST, UK

ABSTRACT

Homologous recombinational repair is an essential mechanism for repair of double-strand breaks in DNA. Recombinases of the RecA-fold family play a crucial role in this process, forming filaments that utilize ATP to mediate their interactions with single- and double-stranded DNA. The recombinase molecules present in the archaea (RadA) and eukaryota (Rad51) are more closely related to each other than to their bacterial counterpart (RecA) and, as a result, RadA makes a suitable model for the eukaryotic system. The crystal structure of *Sulfolobus solfataricus* RadA has been solved to a resolution of 3.2 Å in the absence of nucleotide analogues or DNA, revealing a narrow filamentous assembly with three molecules per helical turn. As observed in other RecA-family recombinases, each RadA molecule in the filament is linked to its neighbour via interactions of a short β-strand with the neighbouring ATPase domain. However, despite apparent flexibility between domains, comparison with other structures indicates conservation of a number of key interactions that introduce rigidity to the system, allowing allosteric control of the filament by interaction with ATP. Additional analysis reveals that the interaction specificity of the five human Rad51 paralogues can be predicted using a simple model based on the RadA structure.

INTRODUCTION

Homologous recombination (HR) is a ubiquitous mechanism by which segments of DNA can be exchanged between DNA molecules with related sequences. Consequently, HR plays a central role in both creating genetic diversity and in DNA repair. The process of HR involves the association of a region of single-stranded DNA (ssDNA) from one parent molecule with double-stranded DNA (dsDNA) from the other parent. This strand invasion step is catalysed by ATP-dependent enzymes of the RecA family (RecA in bacteria, Rad51 in eukaryota and RadA in archaea), which initiate the formation of a nucleoprotein filament where the protein mediates the interaction between ssDNA and dsDNA (1,2).

Enzymes involved in DNA repair in the archaea have attracted interest owing to their greater similarity to analogous proteins in the eukaryota, rather than their bacterial counterparts (3). The case of RecA/RadA/Rad51 is an illuminating example, with *Sulfolobus solfataricus* RadA (*Ss*RadA) sharing 42% sequence identity with human Rad51 (*Hs*Rad51) while *Escherichia coli* RecA (*Ec*RecA) shares only 21% identity. This pattern of conservation is reflected in the domain structures of the proteins (4,5). The core ATPase domain of the Rad51/RecA family is the ∼27 kDa RecA-fold domain. Divergence between RecA and RadA/Rad51 occurs at the termini; while RecA has a short 10-residue N-terminal 'domain' consisting of a single α-helix, RadA and Rad51 possess an ∼7 kDa α-helical bundle domain at their N-terminus. Additionally, RadA/Rad51 completely lacks the C-terminal α+β domain observed in RecA. The RadA/Rad51 N-terminal and RecA C-terminal domains are proposed to have analogous roles involved in DNA interaction (6).

Structural characterization of RecA-family members by many complementary techniques has yielded information on a variety of conformational states in the presence and absence of nucleotide co-factors and DNA. RecA has been observed in right-handed helical filaments with ∼6 molecules per turn [electron microscopy (EM) (7,8) and crystal structures (9–13)] and closed-ring forms [EM (14)] with 3- or 6-fold symmetry. Rad51 has been observed as extended and compressed filaments in the presence of ATP and ATPγS (a non-hydrolysable ATP analogue), respectively (15,16). The eukaryotic meiotic recombinase, Dmc1, originally believed to possess activity as

an octameric closed-ring (17,18), has since been shown to form a helical filament on DNA in its active form (19). Filaments and rings of archaeal RadA molecules have also been observed by EM (20,21), and more recently in crystal structures [*Methanococcus voltae* (*Mv*RadA): filament (22); *Pyrococcus furiosus* (*Pf*RadA): heptameric ring (23)]. Analysis of the results of these various studies suggests that the recombinational activity of the RecA-family recombinases is determined by their conformational flexibility and controlled by their ATP binding and hydrolysing activity (24).

A common feature in all these assemblies is the presence of a specific interaction between neighbouring subunits, often termed the oligomerization motif (13,23,25). This motif is found in the linker region between the N-terminal and ATPase subunits and consists of a short segment of conserved sequence that forms a β-strand, which zips onto the core β-sheet of the neighbouring ATPase domain. In Rad51/RadA, a well-conserved phenylalanine residue in this motif docks in a hydrophobic pocket on the ATPase domain. This mode of interaction is also observed in the crystal structure of *Hs*Rad51 complexed with a segment of the BRCA2 peptide (25), implying a role for BRCA2 in controlling nucleoprotein filament assembly/disassembly.

Higher eukaryotes have evolved a series of Rad51 paralogues in addition to Rad51, which have an important role in recombinational double-strand break repair. The human enzymes Rad51B, Rad51C, Rad51D, XRCC2 and XRCC3 have been observed to form complexes [B·C·D·2, C·3 and 3·C (26,27)] whose specific roles include mediation of the strand-exchange activity of Rad51 (26) and even Holliday junction resolution (28). Interestingly, although these paralogues are expected to interact using a similar oligomerization motif to Rad51, the residue in the position usually occupied by Phe is variable, perhaps as a specificity determinant for complex formation (27).

Comparative EM studies of the crenarchaeal *Ss*RadA, yeast and human Rad51 (*Sc*Rad51 and *Hs*Rad51, respectively) and *Ec*RecA (21) suggest that RadA provides an excellent model for eukaryotic Rad51 and have helped to elucidate details of the role of RadA/Rad51 in strand exchange during recombination. Specifically, it has been observed as octameric rings (in the absence of adenine nucleotides) and as extended and compressed helical filaments (in the presence of ATP or ATPγS) with ssDNA or dsDNA. Most recently, an atomic force microscopy (AFM) study of *Ss*RadA has shown that it can also form fine helical filaments in the absence of DNA or nucleotide co-factor (29).

In this work, we present the crystal structure of full-length *Ss*RadA, which forms a filament-like arrangement of RadA molecules interacting via their oligomerization motifs. Analysis of the intersubunit interfaces in comparison with related RadA, Rad51 and RecA structures provides novel insight into the structural basis of its constrained flexibility and also of amino acid sequence-based interaction specificity of the Rad51-family recombinases.

MATERIALS AND METHODS

Cloning, expression and purification

The gene encoding *Ss*RadA (accession no. Q55075) was cloned from *S.solfataricus* strain P2 by PCR using the oligonucleotides 5′-GTCGGATCCCCATGGCAAATGAAG-TTGAACAGAAAAAGAATATAAAAACC and 5′-CCGGG-GATCCGTCGACTTACTCTTCCGCATCCCTTATTCC and subcloned into the pET19b vector (Invitrogen) using the NcoI and BamHI sites. Overexpression of *Ss*RadA in *E.coli* strain Rosetta (pLysS) DE3 (Stratagene) was attained by induction with 0.5 mM isopropyl-β-ᴅ-thiogalactopyranoside for 3 h, 37°C at 350 r.p.m. Cells were pelleted by centrifugation, washed and then resuspended in 20 mM Tris–HCl, pH 7.6, 150 mM NaCl, 1 mM aminoethylbenzene sulfonyl fluoride and 1 mM DTT. Lysis was obtained by sonication and cell debris was removed by centrifugation at 40 000 *g* for 30 min at 5°C. Cleared cell lysate was then heated to 70°C for 30 min to denature *E.coli* proteins followed by centrifugation at 40 000 *g* for 30 min at 5°C. Analysis of the supernatant by SDS–PAGE showed a band corresponding to the predicted molecular weight (36 kDa). Further purification was achieved by diluting the supernatant 4-fold with 20 mM Tris–HCl, pH 7.5, 1 mM EDTA and 1 mM DTT and applying to a Resource S cation exchange column (Amersham Pharmacia). A linear gradient (0–1000 mM NaCl) was used to elute cationic proteins. Fractions were analysed by SDS–PAGE and those corresponding to RadA were pooled, concentrated and adjusted to 250 mM NaCl before being loaded onto a 26/70 gel-filtration column (Superdex 200 Hi-Load, Amersham Pharmacia). The two major distinct peaks eluted from the column were both analysed by SDS–PAGE and both correspond to *Ss*RadA. Both peaks were used in subsequent experiments, but only material from the first peak to elute produced crystals.

Crystallization and X-ray data collection

Pure protein was concentrated to 7 mg ml⁻¹ in 20 mM Tris–HCl buffer, pH 7.5 with 250 mM NaCl and screened for crystallization using JenaBiosciences HTS screens in 96-well format sitting drop plates (Greiner). Needle-shaped crystals appeared in a single condition containing 100 mM Tris–HCl, pH 8.0 and 40% *tert*-butanol. Optimization yielded diffraction quality crystals grown at 4°C using the hanging drop vapour diffusion method against a reservoir containing 0.5 ml 0.1 M Tris–HCl, pH 9.5 with 10%(v/v) *tert*-butanol, and drops consisting of 0.5 µl of protein and 0.5 µl reservoir. Hexagonal rods grow to a maximum size of 0.1 × 0.1 × 0.4 mm³ over a period of 2 weeks. Crystals are cryoprotected by brief washing (∼30 s) in reservoir solution supplemented with 25% 2-methyl-2,4-pentanediol immediately before flash-cooling under an X-stream cryohead (Oxford Cryosystems). Initially, 104.5° of data were collected using a Rigaku Micromax-007 X-ray generator equipped with an R-Axis-IV++ area detector. The weak diffraction data collected allowed structure solution by molecular replacement (see below), but led to unstable refinement. Subsequent data collected at beamline ID14-EH4 (ESRF, Grenoble) on a selenomethionine-substituted crystal (although not at the Se X-ray absorption edge) were a compromise between maximizing diffraction intensity and minimizing radiation damage to the sample, resulting in a dataset 98% complete to 3.2 Å.

Data processing and structure solution

Details are given in Table 1. Data processing, reduction and analysis of systematic absences [CRYSTALCLEAR/

Table 1. Diffraction data and refinement statistics

	In-house	ESRF
Data		
Space group	$P\,3_1\,2\,1$	$P\,3_1\,2\,1$
Cell dimensions: a, c (Å)	98.42, 99.27	98.15, 99.19
Resolution (Å)	3.0	3.2
Measured reflections	71647	70999
Unique reflections	11474 (1122)	9023 (696)
Completeness (%)	99.5 (99.5)	96.3 (74.8)
Redundancy	6.2 (6.3)	10.0 (8.0)
$I/\sigma(I)$	5.2 (2.4)	13.1 (2.1)
R_{sym}	0.302 (0.636)	0.119 (0.321)
Refinement		
R_{free} (4.75% of reflections)	0.343	0.238
R_{cryst}	0.279	0.174
RMSD bonds (Å)	—	0.011
RMSD angles (°)	—	1.23
Cruickshank's DPI	—	0.408
Total number of atoms		
Protein	—	2286
Water	—	43
Chloride	—	1
Average temperature factors (Å^2)		
Protein (total, TLS component, residual)	—	83, 31, 52
Water	—	57
Chloride	—	63

D*TREK (MSC Corporation)] indicated a space group of $P\,3_1\,2\,1$ or its enantiomorph. Diffraction from these crystals is very weak despite a low mosaicity (0.2°), suggesting the possibility of one molecule per asymmetric unit [68% solvent, Matthews coefficient (30) 3.9] rather than two (36%, 1.9). This was confirmed by molecular replacement [MOLREP and CCP4 (31)] using chain A of the ATPase domain of *Hs*Rad51 [1N0W (25), including side chains], which produced a single significant solution in space group $P\,3_1\,2\,1$. Rigid-body refinement (REFMAC5) led to electron density maps that revealed protein electron density features not represented in the search model. Prime-and-switch density modification [RESOLVE (32)] improved the map, allowing us to assign side chains to most residues in the ATPase domain and build some residues that were not present in the search model [O (33) and COOT (34)]. Cycles of simulated annealing [CNS (35)], density modification (RESOLVE) and model building allowed us to extend the N-terminus into new electron density. The N-terminal domain is highly disordered, resulting in diffuse electron density, but alignment of the NMR structure of the Rad51 N-terminal domain [1B22 (6)] with the clearer features was a significant aid in interpreting this electron density. The subsequent availability of coordinates for *Pf*RadA [1PZN (23)], which included one copy of the enzyme's N-terminal domain, confirmed that our structural assignment was correct. Owing to the low resolution and high disorder, a conservative approach was taken such that refinement restraints were kept tight. Owing to unstable refinement, the synchrotron data were collected and used from this point. Cycles of TLS refinement (10 rounds), positional and B-factor refinement were performed with REFMAC5, using simple Wilson scaling (bulk solvent correction proving unreliable). Water molecules and a chloride ion were added to the model based on the presence of appropriate electron density features in σ_A-weighted difference maps at 3σ and 2mFo-dFc maps at 1σ, together with suitable interactions with protein atoms.

The final model consists of residues 10–259 and 282–324 (the true C-terminus), one chloride ion and 43 water molecules. The refined temperature factors are very high, reflecting the disorder observed in the structure. The N-terminal domain (residues 10–69) has an average temperature factor 122 Å^2 (68 Å^2 TLS component; 54 Å^2 residual) while the remainder of the molecule averages 76 Å^2 (22 Å^2; 54 Å^2). No residues are found outside the allowed regions of a Ramachandran plot [PROCHECK (36)]. Coordinates and diffraction data have been deposited with the Protein Data Bank [PDB (37)] with accession code 2bke.

Superposition of structures from the PDB was performed using LSQMAN (38). The AREAIMOL and CONTACT programs were used to analyse interaction surface areas and contacts.

Recent AFM analysis of *Ss*RadA concludes that *Ss*RadA can form extended helical filaments in the absence of DNA and nucleotide co-factor (29). The authors also describe crystals which support this hypothesis. The similarity of unit cell dimensions reported by Lee *et al*. (29) to the *Ss*RadA structure described here suggests they may be from the same trigonal crystal form, rather than from a different hexagonal form.

RESULTS

Monomer structure

The crystal structure of *Ss*RadA (Figure 1a) represents the intact polypeptide chain, which is composed of two domains connected by a linker region. The N-terminal domain (residues 10–69) forms a bundle of four α-helical segments: α1 (22–31), α2 (36–41), α3 (44–50) and α4 (55–68). The linker region (residues 70–84) contains a short extended section (72–75; strand β0) and an α-helix (α5; 76–83), leading into the C-terminal ATPase domain (residues 85–324). This large α/β domain has a RecA fold composed around a nine-stranded β-sheet (strand order 5-4-6-7-3-8-9-10-11), in which all strands except two (β9 and β11) are parallel, flanked by four α-helices.

The N-terminal helical bundle of *Ss*RadA superimposes well with those from other related structures, with the least similar structure (*Hs*Rad51; 1B22) having an average root-mean-square deviation (RMSD) of 2.2 Å for 42 Cα atoms. (The sequence identity normalized to the shorter *Ss*RadA sequence is 28%). The structure of the C-terminal domain is also similar to other RadA/Rad51 structures, reflecting their high sequence identity (e.g. 45% with *Hs*Rad51). Two segments of sequence (230–236 and 268–292) were not observed in *Hs*Rad51, while only the latter is absent in *Ss*RadA (residues 260–281). Superpositions with both *Hs*Rad51 (1.0 Å for 199 Cα atoms) and *Ec*RecA (1.7 Å for 167 Cα atoms) reflect their homology. The ATPase active site in our structure follows the same architecture as in other recombinases. At 3.2 Å resolution, it is difficult to differentiate *cis* and *trans* peptides, so we have modelled the Asp210–Ser211 peptide in the Walker-B motif in the *cis* conformation in common with other higher resolution RecA-family structures, although a *trans* peptide would also satisfy the electron density. There is no nucleotide in our crystal, and a chloride ion is situated in the ATPase site, close to the positions of the
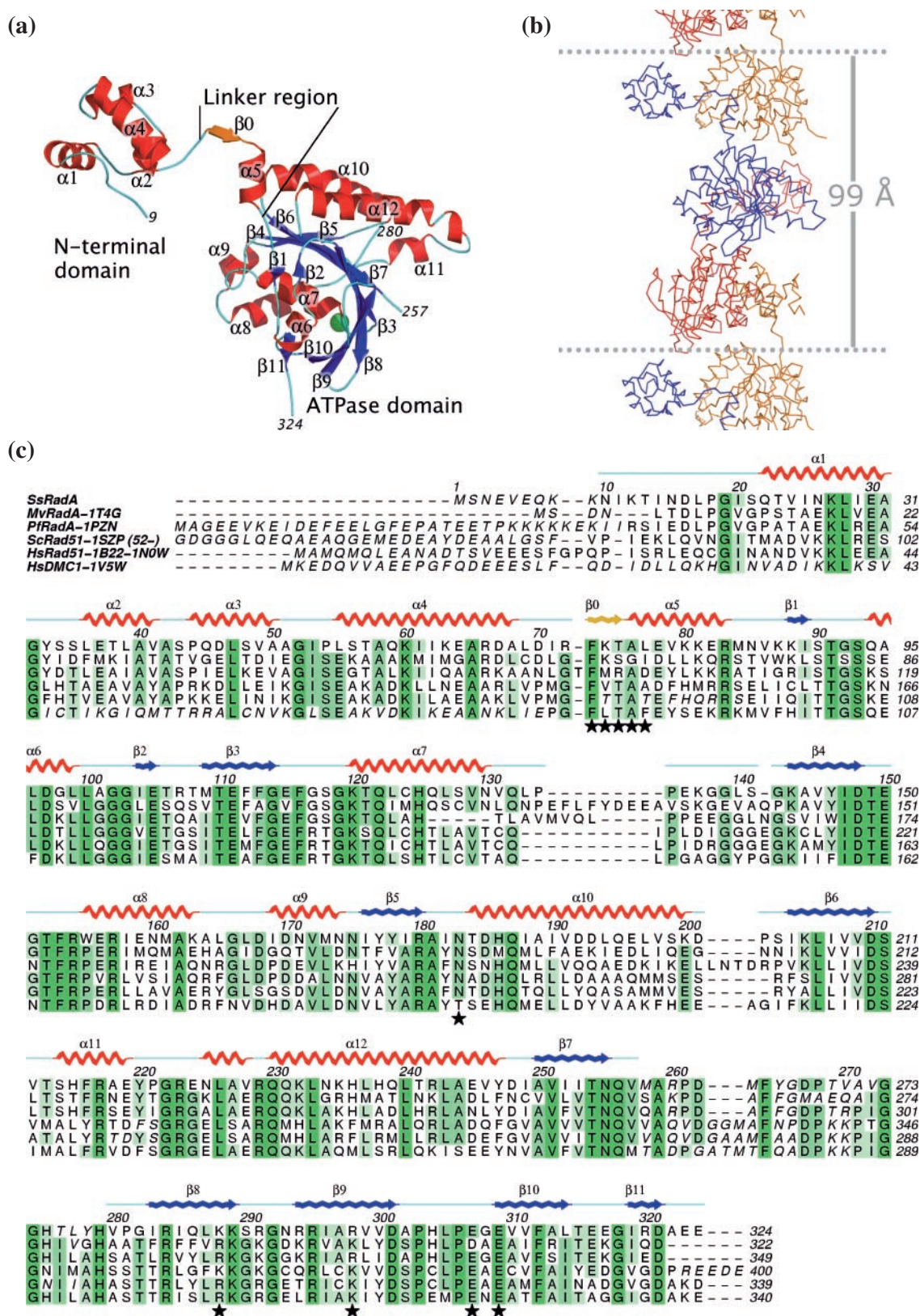
Figure 1. The structure of *Ss*RadA. (**a**) The RadA monomer (**b**) The RadA filament observed in the crystal viewed perpendicular to the 3_1 axis. [All molecular graphics composed with PYMOL (43)]. (**c**) A structure-based sequence alignment of *Ss*RadA with other RadA/Rad51 structures. Italic font indicates residues not observed in the crystal structures. Numbering and secondary structure assignment [DSSP (44)] pertain to *Ss*RadA. Black asterisks highlight residues explicitly mentioned in the text, composed using INDONESIA (38) and ALINE (available from the authors).

α- and β-phosphates of an ADP molecule superimposed from the *Ec*RecA structure (39).

Filament-like quaternary structure

The packing of *Ss*RadA in the space group $P\,3_1\,2\,1$ results in seven contacts with symmetry-related molecules. Two of these contacts account for >80% of the buried surface area (2400 Å^2 per monomer) on crystallization and result in filament-like structures running along the crystallographic 3_1 axis (Figure 1b). The contacts between neighbouring, anti-parallel filaments (related by the crystallographic 2-fold axis) are extremely modest at 563 Å^2 per monomer divided among five different points of contact.

The extensive contacts between monomers in the same filament structure can be divided into three parts: the interactions of (i) the N-terminal domain (burying 356 Å^2), (ii) the linker region (1268 Å^2) containing the conserved β-zip motif (786 Å^2) and (iii) the ATPase domain (844 Å^2) of one subunit, with the ATPase domain of the adjacent subunit. In common with the other intact structures of RecA-family recombinases, the linker region provides the critical interaction which glues RadA molecules into a chain, while allowing flexibility in the ATPase interactions (Figure 2). The short segment of sequence from residues 73–77 in *Ss*RadA (sequence FKTAL), *Hs*BRCA2 (FHTAS) and *Ec*RecA (IMRLG) has identical backbone conformations (*Ss*RadA–*Hs*BRCA2: 1.8 Å for 25 backbone atoms; *Ss*RadA–*Es*RecA: 3.1 Å, for 23 atoms). They form a short β-strand (β0), which zips onto the edge of the central β-sheet of the ATPase domain of the other subunit, adding a 10th strand anti-parallel to β3. Residues at this edge of the ATPase domain form hydrophobic pockets, which accommodate the side chains of highly conserved residues Phe73 and Ala76 (Figure 2a).

DISCUSSION

Our structure contains three RadA molecules per turn in a helical filament, despite the absence of nucleotide or DNA. The observation by AFM (29) of similar narrow filaments suggests that this is a solution property of the *Ss*RadA molecule, rather than a result of crystallization.

Conformational flexibility

For the RadA/Rad51 molecule to be able to sample the various reported states (rings and filaments), one would expect that many gross conformational changes have to take place. However, comparison of the SsRadA structure with the other structures reveals that a number of structural elements are well-conserved and that significant differences are much more limited, implying a well-constrained flexibility.

The domain structure of RadA/Rad51 suggests that it is composed of two rigid bodies (the N-terminal helical bundle and the C-terminal ATPase domain) joined by a flexible, extended linker region. However, analysis of our structure and comparison with other RadA structures indicate that the linker region itself, when complexed to another RadA molecule, forms a rigid motif encompassing strand β0 and helix α5. A cluster of charged residues is found in RadA sequences in this linker region between positions

67 and 78 (*Ss*RadA numbering), forming a network of electrostatic interactions resulting in a compact local structure (Figure 2b). The significance of this may be related to the enhanced folding stability at high temperatures conferred by salt bridges (40,41) that would temper flexibility in this thermostable protein.

From sequence analysis, it is clear that there are differences between RecA and RadA/Rad51 in the role that their N-termini play in forming intermolecular interactions: The N-terminal α-helix of RecA adds to the tertiary structure of the neighbouring subunit by lining up parallel to the other flanking helices, incrementing the helix array on one side of the sheet to three (Figure 2c). Such an interaction is not observed in the RadA/Rad51 structures as they have a helical bundle domain rather than a single helix at this point, and helix α4 runs perpendicular to the flanking helices of the neighbouring ATPase domain. Furthermore, while the conserved Phe residue of RadA/Rad51 is bound in a deep pocket composed of residues of the other subunit, RecA has a less bulky Ile in this position and the N-terminal helix from the same subunit forms part of its binding pocket. The interaction of BRCA2 with *Hs*Rad51 is significantly more different to both RadA and RecA (Figure 2c).

Our comparisons show that not only is the β-zip interaction conserved, but that all RadA structures involve similar interactions of the N-terminal domain with the neighbouring ATPase domain. Hydrogen bonds are formed between the backbone of the turn between helices α1 and α2 of the N-terminal domain and a well-conserved Asn (183 in *Ss*RadA) on the ATPase domain, as well as hydrophobic interactions between nearby residues (Figure 2d and e). The substitution of this Asn to Thr in most Dmc1 sequences, albeit conservative, may play a role in the different conformational properties of Dmc1.

The apparent restraint of conformational freedom imposed by the interactions listed above results in the only remaining point of torsional flexibility lying in the loop between helix α5 (in the linker) and sheet β1 (in the ATPase domain). Comparison of the *Ss*RadA, *Pf*RadA and *Mv*RadA structures shows that this is indeed the case, and residues 84–86 (in *Ss*RadA) represent the single location of significant torsional difference.

To achieve 3_1 symmetry, the *Ss*RadA filament is 'overwound' with respect to a 6_1 filament; similarly, the 7-fold rotational symmetry observed in crystals of *Pf*RadA is caused by 'underwinding'. Accompanying the torsional difference around position 85, the interactions between neighbouring ATPase domains differ significantly in the various conformational forms. If one ATPase domain from the different forms is superposed, the neighbouring ATPase domains are disposed differently to each other (Figure 3). The structure of *Pf*RadA, like the *Ec*RecA filament, has an open ATP-binding site, whereas the *Sc*Rad51 and *Mv*RadA filaments involve residues from both subunits to form the ATP site, as observed in EM reconstructions of active *Ec*RecA filaments. To transform between these active and inactive conformations, the relative rotation between adjacent ATPase domains is ∼30°. In the *Ss*RadA filament, this rotation is increased by a further 60° resulting in an occlusion of the ATPase site by residues from helices α10 to α12. In the active conformation, it is residues from the 'ATP cap' (between β9 and β10) that lie closest to this position.
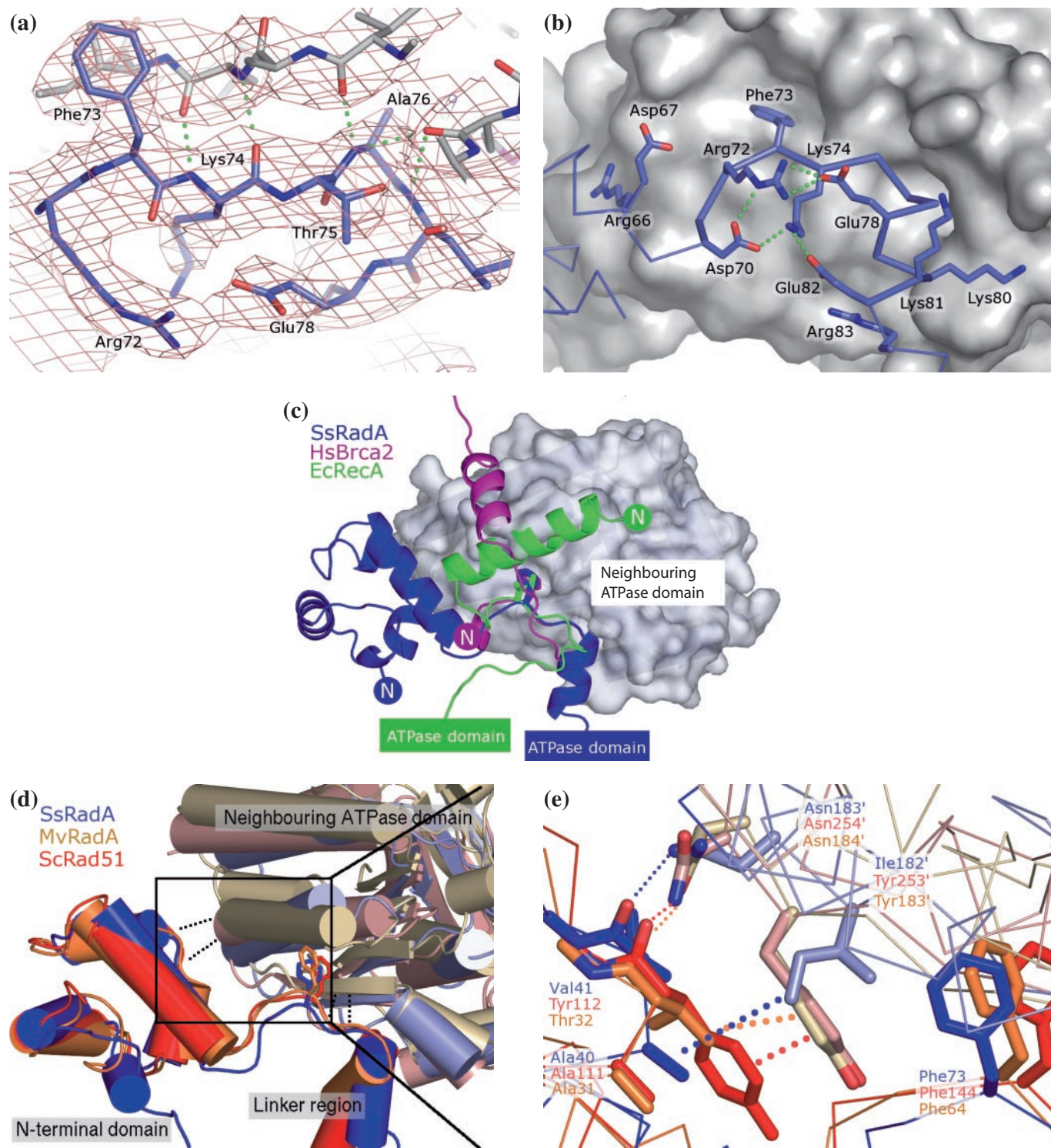
Figure 2. The oligomerization strand. (**a**) σ_A-weighted electron density (1.2 σ; purple mesh) of the region surrounding Phe73. Atoms from different monomers are coloured differently (grey/blue). (**b**) A cluster of salt bridges stabilizes the *Ss*RadA oligomerization motif. (**c**) Superposition of the oligomerization strands of *Ss*RadA (blue), *Ec*RecA (green) and *Hs*Rad51/BRCA2 (magenta). The common ATPase domain of the interacting subunit is shown as a grey surface. (**d** and **e**) Conserved interactions between the N-terminal domain of *Ss*RadA, *Mv*RadA and *Sc*Rad51, with the neighbouring ATPase domain.

Interaction specificity among Rad51 paralogues

Yeast two-hybrid experiments used to screen constructs consisting of the N-terminal and linker regions of the *Hs*Rad51 paralogues, Rad51B, Rad51C, Rad51D, XRCC2 and XRCC3 against their C-terminal domains, indicate that the following complexes exist: Rad51B:Rad51C:Rad51D:XRCC2, Rad51C:XRCC3 and XRCC3:Rad51C, where the colon signifies the N-terminus of the preceding protein interacting with
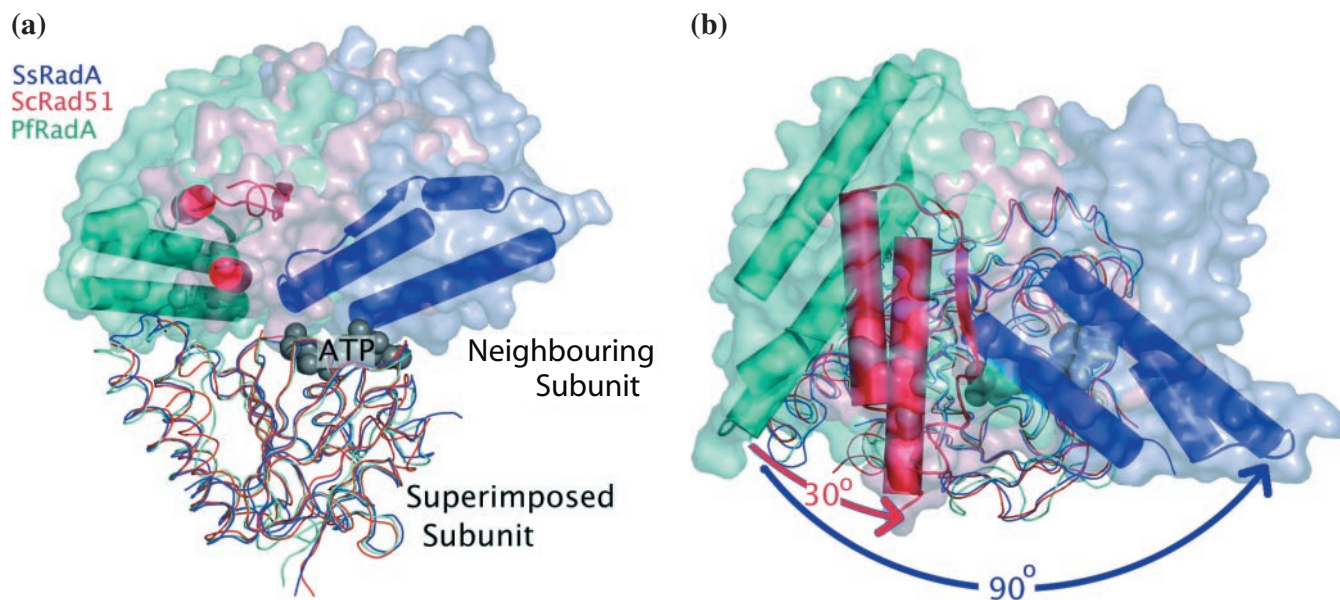
Figure 3. Interactions between ATPase subunits for *Ss*RadA (blue), *Sc*Rad51 (red) and *Pf*RadA (green). Structures were superimposed on one subunit (shown as backbone trace). The neighbouring subunits are shown as semi-transparent surfaces, with a solid cartoon representation of residues between helices α10 and α12. (**a**) Side view. (**b**) Top view.

Table 2. Pairwise interaction scores for *Hs*Rad51 paralogues

Linker residue (position 73)	Rad51B: VHYRSLE	Rad51C: VYFQLLF	ATPase domain Rad51D: LQVGAQT	XRCC2: LFVTSLM	XRCC3: VRGCKKV
Rad51B: M	−0.02	**3.33**[a]	0.28	1.26	−2.86
Rad51C: K	−2.20	−3.96	**−2.02**	−4.27	**−6.00**[a,c]
Rad51D: V	0.12	**4.47**[b]	0.75	3.62[a]	−2.04
XRCC2: F	0.98	**5.97**[b]	1.43	4.77	−1.84
XRCC3: I	0.33	**5.00**[a]	0.69	4.09	−2.38

The residue type where *Ss*RadA has Phe73 is given in the left column and pocket-forming residues given in top row; bold font indicates highest scoring interactions.
[a]Expected interactions (27).
[b]False positive.
[c]False negative.

the C-terminus of the succeeding one (27). As discussed previously, the major contribution to the β-zip interaction between the linker region of one molecule and the ATPase domain of its neighbour in RadA comes from backbone–backbone hydrogen bonding. The only significant sequence-dependent (i.e. side chain–side chain) interactions come from Phe73 and Ala76, the most highly conserved elements of the Rad51-family linear motif. As Ala76 is also conserved among the Rad51 homologues, it can have no influence on specificity, but the residue at position 73 (*Ss*RadA numbering) is variable.

In order to investigate the possibility that the residue type at this position is largely responsible for the interaction specificity, we performed a pairwise comparison of complementarity of this residue with those forming the binding pocket, for all pairs of paralogues. Investigation of the *Ss*RadA structure showed that the side chain of a residue in position 73 (Phe in *Ss*RadA) could interact with those of residues 145 (Val in *Ss*RadA), 177 (Tyr), 179 (Ile), 190 (Ile), 193 (Asp), 194 (Leu) and 197 (Leu). Using a sequence alignment of the equivalent parts of each of the paralogues, we produced a list of

residue–pair interactions for each protein pair. Each residue pair was then assigned a score [using the matrix employed by INTERPRETS (42)] and the scores were summed for each protein pair (Table 2). Surprisingly for such a simple model, there is significant agreement with the experimental result: of all possible interactions of the linker region of Rad51B, that with Rad51C scores highest, and similarly Rad51C:Rad51D and XRCC3:Rad51C are scored most favourably, while the Rad51D:XRCC2 interaction is second most favourable. Only the Rad51C:XRCC3 interaction is completely incompatible with this model. Inspection of the residue types suggests that favourable complementarity occurs where both sets of residues are largely hydrophobic, while poor complementarity occurs for charged residues. The Lys residue of Rad51C is highly unlikely to interact with the selected residues for XRCC3 owing to the presence of three positively charged residues, suggesting that either our alignment does not represent the XRCC3 structure well enough or that a different mode of interaction is used in the Rad51C:XRCC3 association.

In summary, the *Ss*Rada crystal structure reveals a conformation of fine filaments in the absence of nucleotides that is in agreement with AFM studies (29). Comparison with eukaryotic Rad51 paralogues supports a surprisingly simple model for their interaction specificity, opening up the possibility of predicting such interactions. Comparison with crystal structures of other RadA/Rad51-family members has allowed us to determine that significant conformational flexibility is only apparent in a surprisingly small part of the monomer's structure and to identify residues that may be implicated in conformational restraint highlighting the dynamic but restrained flexibility of this system.

ACKNOWLEDGEMENTS

REFERENCES

1. Bianco,P.R., Tracy,R.B. and Kowalczykowski,S.C. (1998) DNA strand exchange proteins: a biochemical and physical comparison. *Front. Biosci.*, **3**, D570–D603.
2. Lusetti,S.L. and Cox,M.M. (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu. Rev. Biochem.*, **71**, 71–100.
3. White,M.F. (2003) Archaeal DNA repair: paradigms and puzzles. *Biochem. Soc. Trans.*, **31**, 690–693.
4. Seitz,E.M., Brockman,J.P., Sandler,S.J., Clark,A.J. and Kowalczykowski,S.C. (1998) RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. *Genes Dev.*, **12**, 1248–1253.
5. Komori,K., Miyata,T., Daiyasu,H., Toh,H., Shinagawa,H. and Ishino,Y. (2000) Domain analysis of an archaeal RadA protein for the strand exchange activity. *J. Biol. Chem.*, **275**, 33791–33797.
6. Aihara,H., Ito,Y., Kurumizaka,H., Yokoyama,S. and Shibata,T. (1999) The N-terminal domain of the human Rad51 protein binds DNA: structure and a DNA binding surface as revealed by NMR. *J. Mol. Biol.*, **290**, 495–504.
7. Egelman,E.H. and Stasiak,A. (1986) Structure of helical RecA–DNA complexes. Complexes formed in the presence of ATP-gamma-S or ATP. *J. Mol. Biol.*, **191**, 677–697.
8. Yu,X. and Egelman,E.H. (1992) Structural data suggest that the active and inactive forms of the RecA filament are not simply interconvertible. *J. Mol. Biol.*, **227**, 334–346.
9. Datta,S., Prabu,M.M., Vaze,M.B., Ganesh,N., Chandra,N.R., Muniyappa,K. and Vijayan,M. (2000) Crystal structures of *Mycobacterium tuberculosis* RecA and its complex with ADP-AlF(4): implications for decreased ATPase activity and molecular aggregation. *Nucleic Acids Res.*, **28**, 4964–4973.
10. Datta,S., Krishna,R., Ganesh,N., Chandra,N.R., Muniyappa,K. and Vijayan,M. (2003) Crystal structures of *Mycobacterium smegmatis* RecA and its nucleotide complexes. *J. Bacteriol.*, **185**, 4280–4284.
11. Xing,X. and Bell,C.E. (2004) Crystal structures of *Escherichia coli* RecA in a compressed helical filament. *J. Mol. Biol.*, **342**, 1471–1485.
12. Rajan,R. and Bell,C.E. (2004) Crystal structure of RecA from *Deinococcus radiodurans*: insights into the structural basis of extreme radioresistance. *J. Mol. Biol.*, **344**, 951–963.
13. Story,R.M., Weber,I.T. and Steitz,T.A. (1992) The structure of the *E.coli* recA protein monomer and polymer. *Nature*, **355**, 318–325.
14. Yu,X. and Egelman,E.H. (1997) The RecA hexamer is a structural homologue of ring helicases. *Nature Struct. Biol.*, **4**, 101–104.
15. Conway,A.B., Lynch,T.W., Zhang,Y., Fortin,G.S., Fung,C.W., Symington,L.S. and Rice,P.A. (2004) Crystal structure of a Rad51 filament. *Nature Struct. Mol. Biol.*, **11**, 791–796.
16. Yu,X., Jacobs,S.A., West,S.C., Ogawa,T. and Egelman,E.H. (2001) Domain structure and dynamics in the helical filaments formed by RecA and Rad51 on DNA. *Proc. Natl Acad. Sci. USA*, **98**, 8419–8424.
17. Passy,S.I., Yu,X., Li,Z., Radding,C.M., Masson,J.Y., West,S.C. and Egelman,E.H. (1999) Human Dmc1 protein binds DNA as an octameric ring. *Proc. Natl Acad. Sci. USA*, **96**, 10684–10688.
18. Kinebuchi,T., Kagawa,W., Enomoto,R., Tanaka,K., Miyagawa,K., Shibata,T., Kurumizaka,H. and Yokoyama,S. (2004) Structural basis for octameric ring formation and DNA interaction of the human homologous-pairing protein Dmc1. *Mol. Cell*, **14**, 363–374.
19. Sehorn,M.G., Sigurdsson,S., Bussen,W., Unger,V.M. and Sung,P. (2004) Human meiotic recombinase Dmc1 promotes ATP-dependent homologous DNA strand exchange. *Nature*, **429**, 433–437.
20. McIlwraith,M.J., Hall,D.R., Stasiak,A.Z., Stasiak,A., Wigley,D.B. and West,S.C. (2001) RadA protein from *Archaeoglobus fulgidus* forms rings, nucleoprotein filaments and catalyses homologous recombination. *Nucleic Acids Res.*, **29**, 4509–4517.
21. Yang,S., Yu,X., Seitz,E.M., Kowalczykowski,S.C. and Egelman,E.H. (2001) Archaeal RadA protein binds DNA as both helical filaments and octameric rings. *J. Mol. Biol.*, **314**, 1077–1085.
22. Wu,Y., He,Y., Moya,I.A., Qian,X. and Luo,Y. (2004) Crystal structure of archaeal recombinase RadA: a snapshot of its extended conformation. *Mol. Cell*, **15**, 423–435.
23. Shin,D.S., Pellegrini,L., Daniels,D.S., Yelent,B., Craig,L., Bates,D., Yu,D.S., Shivji,M.K., Hitomi,C., Arvai,A.S. *et al.* (2003) Full-length archaeal Rad51 structure and mutants: mechanisms for Rad51 assembly and control by BRCA2. *EMBO J.*, **22**, 4566–4576.
24. Wyman,C., Ristic,D. and Kanaar,R. (2004) Homologous recombination-mediated double-strand break repair. *DNA Repair (Amst.)*, **3**, 827–833.
25. Pellegrini,L., Yu,D.S., Lo,T., Anand,S., Lee,M., Blundell,T.L. and Venkitaraman,A.R. (2002) Insights into DNA recombination from the structure of a Rad51–BRCA2 complex. *Nature*, **420**, 287–293.
26. Masson,J.Y., Tarsounas,M.C., Stasiak,A.Z., Stasiak,A., Shah,R., McIlwraith,M.J., Benson,F.E. and West,S.C. (2001) Identification and purification of two distinct complexes containing the five Rad51 paralogs. *Genes Dev.*, **15**, 3296–3307.
27. Miller,K.A., Sawicka,D., Barsky,D. and Albala,J.S. (2004) Domain mapping of the Rad51 paralog protein complexes. *Nucleic Acids Res.*, **32**, 169–178.
28. Liu,Y., Masson,J.Y., Shah,R., O'Regan,P. and West,S.C. (2004) Rad51C is required for Holliday junction processing in mammalian cells. *Science*, **303**, 243–246.
29. Lee,M.H., Leng,C.H., Chang,Y.C., Chou,C.C., Chen,Y.K., Hsu,F.F., Chang,C.S., Wang,A.H. and Wang,T.F. (2004) Self-polymerization of archaeal RadA protein into long and fine helical filaments. *Biochem. Biophys. Res. Commun.*, **323**, 845–851.
30. Matthews,B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491–497.
31. Collaborative Computational Project,N. (1994), The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.
32. Terwilliger,T.C. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 965–972.
33. Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
34. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
35. Brunger,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. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
36. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283–291.
37. Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.F.,Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.
38. Kleywegt,G.J. and Jones,T.A. (1997) Detecting fold motifs and similarities in protein structures. *Methods Enzymol.*, **277**, 525–545.
39. Story,R.M. and Steitz,T.A. (1992) Structure of the recA protein–ADP complex. *Nature*, **355**, 374–376.

40. Perutz,M.F. and Raidt,H. (1975) Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature*, **255**, 256–259.

41. Zhou,H.X. (2002) Toward the physical basis of thermophilic proteins: linking of enriched polar interactions and reduced heat capacity of unfolding. *Biophys. J.*, **83**, 3126–3133.

42. Aloy,P. and Russell,R.B. (2003) InterPreTS: protein interaction prediction through tertiary structure. *Bioinformatics*, **19**, 161–162.

43. DeLano,W.L. (2002) *The PyMOL Molecular Graphics System.* DeLano Scientific, CA, USA.

44. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.