DNA binding to RecD: role of the 1B domain in SF1B helicase activity

Kayarat Saikrishnan, Stuart P Griffiths,
Nicola Cook, Robert Court and
Dale B Wigley*

Cancer Research UK Clare Hall Laboratories, The London Research
Institute, Herts, UK

The molecular mechanism of superfamily 1B α helicases remains unclear. We present here the crystal structure of the RecD2 helicase from *Deinococcus radiodurans* at 2.2-Å resolution. The structure reveals the folds of the 1B and 2B domains of RecD that were poorly ordered in the structure of the *Escherichia coli* RecBCD enzyme complex reported previously. The 2B domain adopts an SH3 fold which, although common in eukaryotes, is extremely rare in bacterial systems. In addition, the *D. radiodurans* RecD2 structure has aided us in deciphering lower resolution (3.6 Å) electron density maps for the *E. coli* RecBCD enzyme in complex with a long DNA substrate that interacts with the RecD subunit. Taken together, these structures indicated an important role for the 1B domain of RecD, a β -hairpin that extends from the surface of the 1A domain and interacts with the DNA substrate. On the basis of these structural data, we designed a mutant RecD2 helicase that lacks this pin. The 'pin-less' mutant protein is a fully active ssDNA-dependent ATPase but totally lacks helicase activity.

The EMBO Journal (2008) 27, 2222–2229. doi:10.1038/emboj.2008.144; Published online 31 July 2008

Subject Categories: structural biology

Keywords: DNA helicase; DNA repair; RecD

Introduction

There is a huge variety of helicase enzymes and efforts have been made to divide these into families of related enzymes. Initial efforts utilised conserved sequence motifs to characterise a number of helicase superfamilies (Gorbalenya and Koonin, 1993). However, recent structural and biochemical data have allowed further refinement and extension of the original superfamilies (Singleton *et al*, 2007).

Superfamily 1 enzymes fall into two classes: those that translocate along their nucleic acid substrates with a 3'–5' polarity (SF1A enzymes) and those that translocate in a 5'–3' direction (SF1B enzymes). A great deal is known about the mechanism of the SF1A enzymes such as PcrA, Rep and UvrD both from a structural perspective (Subramanya *et al*, 1996;

Korolev *et al*, 1997; Velankar *et al*, 1999) and from biochemistry (Soultanas *et al*, 2000; Dillingham *et al*, 2001; Tomko *et al*, 2007).

Rather less is known about the mechanism of SF1B enzymes. Several enzymes of this family have important functions in both eukaryotes and prokaryotes. Structural data have been obtained for the SF1B RNA helicase Upf1 in complexes with phosphate, ADP and the non-hydrolysable ATP analogue, ADPNP (Cheng *et al*, 2006), although a structure with bound RNA remains lacking. These structures reveal a conformational change that accompanies binding of ATP and which is very similar to that which occurs during catalysis in SF1A helicases such as PcrA (Velankar *et al*, 1999).

The best characterised DNA helicases of this family are probably the T4 phage enzyme Dda and RecD from *Escherichia coli*. Work on Dda has shown that the enzyme works as a monomer (Nanduri *et al*, 2002), although these monomers can 'cooperate' to increase their activity (Byrd and Raney, 2004). Studies on RecD are complicated by the fact that this enzyme is just one component of the three subunit RecBCD complex and most biochemical studies have been on this intact complex, which contains two helicase subunits: RecB (SF1A enzyme) and RecD (SF1B enzyme). Structural data have revealed how the two motors of opposite polarity cooperate to unwind DNA in a highly processive manner and also suggested how the nuclease activity of the RecB subunit might be regulated by the RecD protein (Singleton *et al*, 2004). However, although the DNA substrate used for the crystallisation extended into the RecB motor domains, it was too short to reach the RecD subunit. Consequently, although the RecBCD complex provided the first snapshot of the structure of a SF1B enzyme, the details of how DNA interacts with RecD could not be visualised. Furthermore, the 1B and 2B domains of RecD were poorly ordered in the RecBCD complex and so could not be built with confidence in the structure.

To address these shortcomings, we have taken two approaches. The first of these is to extend the DNA substrate so that it now crosses the RecD motor domains. In a complementary approach, we have also determined the crystal structure of a RecD homologue from *Deinococcus radiodurans* at high resolution. Taken together, these structures reveal the way that ssDNA interacts with the enzyme as well as the structures of the important 1B and 2B domains of RecD, indicating the likely involvement of the 1B domain in helicase activity. On the basis of these structural data, we have designed a mutant lacking the 1B domain and show that, although this protein retains wild-type ssDNA-dependent ATPase activity, it has lost helicase function completely.

Results

Crystal structure of a RecD family helicase from *D. radiodurans*

Isolated *E. coli* RecD protein is not well behaved in solution unless as a part of the RecBCD complex. Although activity of

*Corresponding author. Cancer Research UK, Clare Hall Laboratories, The London Research Institute, Blanche Lane, South Mimms, Potters Bar, Herts EN6 3LD, UK. Tel.: +44 1707 625930; Fax: +44 1707 625803; E-mail: Dale.Wigley@cancer.org.uk

Received: 7 May 2008; accepted: 30 June 2008; published online: 31 July 2008

the protein can be characterised (Dillingham *et al*, 2003), we wished to have a RecD family enzyme that might be easier to work with, especially as a basis for mutagenesis and structural studies. Sequence alignments have revealed that there are two clades of RecD enzymes, those that have RecBC partners in the cells and those which do not (Rocha *et al*, 2005). We choose to use *D. radiodurans* RecD2 for these studies, an enzyme of the class that lacks RecBC partners. Previous work on *D. radiodurans* RecD2 has characterised the enzyme as a 5'-3' helicase with a preference for substrates with a 5'-tail (Wang and Julin, 2004). We hoped that this simpler system would allow us to determine the structure of the domains in *E. coli* RecD that are disordered in the RecBCD structure and also provide a framework for biochemical studies to unravel the mechanism of the SF1 5'-3' helicases.

An N-terminal truncation of *D. radiodurans* RecD2 (Δ NRecD2) was made on the basis of a proteolytic fragment seen during purification of the full-length protein (data not shown). The 150 residues at the N terminus are poorly conserved between RecD2 enzymes and are missing in RecD1 enzymes such as that in *E. coli*. The truncated enzyme was overexpressed in *E. coli* and purified to homogeneity. The truncated enzyme has helicase activity that is indistinguishable from that of the full-length protein (unpublished data). Crystals of Δ NRecD2 were obtained and the structure was determined at 2.2 Å.

As expected, the overall structure of the protein is similar to *E. coli* RecD (Figure 1). The N-terminal region of the protein shows essentially the same fold but some of the secondary structural elements are better defined. Similarly, the structure of the ATPase domains of *D. radiodurans* RecD2 was essentially the same as for *E. coli* RecD but with a slight alteration of their relative disposition. However, the 1B and 2B domains that were poorly ordered in the *E. coli* enzyme were now well ordered in the *Deinococcus* protein (Figure 1). Domain 1B is very small (10 residues) and forms a short hairpin that protrudes from the surface of the 1A domain. Domain 2B is larger (70 residues) and adopts a SRC homology domain 3 (SH3) domain fold, which is extremely unusual for a bacterial protein. The 1B and 2B domains have important functions in helicase activity in 3'-5' SF1 enzymes such as PcrA (Velankar *et al*, 1999; Soultanas *et al*, 2000) and Rep (Brendza *et al*, 2005) and we envisage this is probably also the case for SF1 5'-3' helicases. Knowledge of the structure of the 1B and 2B domains in RecD2 assisted us in interpreting the weak density associated with these domains in the RecD

subunit of the RecBCD structure. The 1B domain could be discerned unambiguously (see below). However, although we could pick out that the overall fold of the 2B domain in RecD was also an SH3 fold, we were unable to build a model that could be refined unambiguously (data not shown) so we have omitted this domain from our refinement of the RecBCD complex presented below.

Interactions between the 5'-tail of the DNA substrate and RecD

As a first step towards processing double-strand breaks in DNA, RecBCD binds to the broken DNA end to form an initiation complex. Biochemical and biophysical studies have shown that, on formation of the initiation complex, as many as six base pairs of the duplex end are unwound by RecBCD even in the absence of ATP (Farah and Smith, 1997; Wong *et al*, 2005). However, only four base pairs were unwound in the crystal structure of the RecBCD DNA complex that we determined previously (Singleton *et al*, 2004). The reason for this appears to be that the contacts between the 'arm' of the RecB subunit and the single/double strand junction of the substrate span 15 base pairs. Consequently, upon complex formation, of the 19 base pairs that form the duplex stem of the DNA, 15 base pairs are required to span the duplex-binding site leaving only the remaining 4 base pairs to be unwound and enter the enzyme. To overcome this length constraint, a longer DNA substrate was designed with 21 base pairs of duplex. In addition, a four nucleotide single-stranded 5'-overhang was introduced as it has been shown that similar DNA substrates bind to RecBCD tightly (Wong *et al*, 2005).

The crystal structure of the complex between this new DNA substrate and RecBCD was determined at a resolution of 3.6 Å (Figure 2). As seen previously (Singleton *et al*, 2004), upon unwinding, the 3'-ssDNA is fed into the tunnel formed by RecB helicase and N-terminal region of RecC, whereas the 5'-ssDNA enters the tunnel formed by RecD and the C-terminal region of RecC. The structure confirms that RecBCD is able to unwind six base pairs even in the absence of ATP (Farah and Smith, 1997; Wong *et al*, 2005) and six bases at the 3'-end of the DNA are evident, spanning the RecB subunit. However, unambiguous assignment of the electron density for the 5'-tail was complicated by the 1B domain of the RecD subunit that was disordered in the previous crystal structure but is ordered in this structure (see below). To resolve this ambiguity, 5-iodouracil was used at seven of the nine 5'-terminal nucleotide positions. The iodine positions were identified using difference Fourier maps and ensured the correct register of the DNA tails. Fifteen base pairs of duplex are the same as observed for the previous crystal structure (Singleton *et al*, 2004). Difference electron density for the DNA in this region is presented in the Supplementary data.

The unwinding of six base pairs from a blunt end results in the 3'-tail being fully engaged with the helicase motor domains of RecB, ready for translocation to initiate (Figure 2B). However, with a blunt-ended substrate, the 5'-tail is not sufficiently long to interact with the RecD motor. Consequently, the first few steps of DNA unwinding and translocation have to be catalysed by RecB resulting in the 5'-tail being pushed through the 5'-channel until it engages the RecD motor. In the structure presented here, the four

Figure 1 Comparison of the structures of *E. coli* RecD1 (a part of the RecBCD structure presented herein) and the truncated *D. radiodurans* RecD2 (Δ NRecD2). Domain 2B of *E. coli* RecD1 is disordered. The ATPase domains 1A and 2A are shown in red and green, respectively.

Figure 2 (A) Overall structure of the RecBCD complexed with the longer DNA substrate. The DNA substrate used in the previous structure (Singleton *et al*, 2004) is depicted in blue with the additional bases in the present structure shown in magenta. (B) Close up view of the DNA-binding site for the 3'-tail as it crosses the RecB subunit. The protein and DNA are coloured using the same scheme as in (A). (C) Close up view of the DNA-binding site for the 5'-tail as it crosses RecC (residues coloured in cyan) and onto the RecD subunit (green residues). The protein and DNA are coloured using the same scheme as in (A).

nucleotide overhang on the 5'-end combined with the unwinding of six base pairs of the DNA duplex results in a single-stranded 5'-tail with a length of 10 nt. This 5'-tail is long enough to interact with RecD, although it is still too short to engage fully with both motor domains (Figure 2C). This observation is consistent with experiments on RecBCD complexes containing an inactive RecB subunit so that only the RecD subunit is contributing to helicase activity (Dillingham *et al*, 2005). This mutant complex prefers 5'-overhangs with the best unwinding seen with a substrate that is most similar to that used in this crystal structure.

Path of the 5'-tail

Although extensive biochemical and structural characterisation of the SF1A helicases have provided an in-depth understanding of how these motors unwind and translocate in a 3'-5' direction, rather less is known about the SF1B helicases. In particular, the molecular basis for the 5'-3' directionality remains unclear. Structural studies have revealed that ssDNA binds to SF1A helicases across the 1A and 2A domains spanning the domain interface in a 3'-5' direction (Korolev *et al*, 1997; Velankar *et al*, 1999). Translocation results from movements of the motor domains along the ssDNA tail with the 2A domain leading and the 1A domain following. Two simple variations on this model could result in translocation with the reverse polarity (Figure 3A). In one model, the enzyme binds to ssDNA in an orientation opposite to that of 3'-5' helicases, that is, with the 2A domain towards the 3'-end and 1A domain facing the 5'-end. During translocation, the 2A domain would lead with the 1A domain following. In the alternative model, the motor domains bind to ssDNA in the same orientation as that of a SF1A helicase, but the direction of translocation is reversed. Hence, in principle, reversing either the orientation of binding or the direction of translocation could result in a helicase with the opposite directionality. The 5'-tail of the DNA substrate has been shown to pass along a channel in the RecBCD complex

(Singleton *et al*, 2004), revealing that the ssDNA binds to the RecD subunit with the same polarity as SF1A enzymes. Consequently, it was proposed that directionality was determined by an altered direction of translocation rather than a reversed binding orientation but experimental evidence supporting that proposal remained to be obtained.

This new structure of RecBCD with ssDNA bound to RecD now provides that experimental evidence. The 10 nt 5'-tail continues through the 5'-tunnel, first traversing the C-terminal domain of RecC and then entering the part of the channel formed by the interface between RecB and RecD. On entering the channel, the ssDNA first interacts with the 1A domain of RecD and then bridges the interface with the 2A domain. This structure represents the first molecular picture of the interactions made by a SF1B helicase with ssDNA. It confirms that ssDNA binds to the RecD motor domains in the same orientation as that seen for SF1A motors, implying that the direction of movement of the motor domains along the bound ssDNA will need to be opposite to effect reversed polarity of translocation.

The role of the motif 1a pocket

In the SF1A helicases such as PcrA, structural and biochemical data have shown that residues in motif 1a have a crucial function in helicase activity (Velankar *et al*, 1999; Dillingham *et al*, 2001). In particular, Phe64 undergoes a conformational change between the substrate and product complexes that opens and closes a pocket on domain 1A that becomes occupied transiently by a base of the translocating single strand of DNA (Figures 3C, D). This gating mechanism is a key component in the translocation process (Velankar *et al*, 1999; Dillingham *et al*, 2001). By contrast, although SF1B helicases have a conserved motif 1a, the phenylalanine is replaced by a conserved proline residue. As a consequence, the pocket is no longer regulated and is in a permanently open state. In the present structure, we see that the motif 1a pocket is occupied by a base (Figure 3B). Interestingly, the

Figure 3 (A) Two models for ways to convert helicase directionality from 3'–5' to 5'–3'. The panel at the top shows the DNA bound in the reverse orientation, whereas the panel below shows the alternative in which translocation is reversed instead. (B) Location of the motif 1a pocket in RecD. The base in the pocket is coloured orange. Other colours follow the same scheme as Figure 2. (C) The equivalent view of the PcrA substrate complex (Velankar *et al*, 1999). Note that in this structure, the motif 1a pocket is blocked by F64. (D) The equivalent view of the PcrA product complex (Velankar *et al*, 1999). In this structure, the motif 1a pocket is open because F64 has swung out of the way allowing a DNA base (shown in orange) to occupy the pocket.

consensus motif 1a of SF2 enzymes also contains a proline rather than a phenylalanine (Gorbalenya and Koonin, 1993).

DNA binding to a pin

In our previous structure of RecBCD (Singleton *et al*, 2004), the polypeptide chain of the RecD subunit was disordered in two regions corresponding to regions that would be the 1B and 2B domains in other helicases such as PcrA (Subramanya *et al*, 1996). The extended 5'-tail of the DNA substrate in the present structure passes across the 1B region of RecD. As discussed above, electron density corresponding to the protein and DNA was distinguished by labelling of the DNA with iodouracil. It became evident that the structure of the 1B domain of RecD was more ordered than in previous structures and formed a β -hairpin extending from the surface of the 1A domain (Figure 4) in the same way that the equivalent region forms a hairpin in RecD2 (Figure 1). Furthermore, the bound DNA runs along the side of this structure, suggesting that it might be an important part of the helicase mechanism. Similar 'pin' or 'wedge' structures have a function in the mechanical splitting of the DNA substrate in many other helicases (Hargreaves *et al*, 1998; Roe *et al*, 1998; Velankar *et al*, 1999; Ariyoshi *et al*, 2000; Singleton *et al*, 2001, 2004; Buttner *et al*, 2007). However, the location and form of the pin vary between different enzymes. To test whether this pin might contribute to helicase activity in RecD, we decided to make a mutant enzyme in which the pin was truncated.

A pin-less RecD2 mutant is a defective helicase but retains wild-type ATPase activity

Using sequence alignments of different RecD and RecD-like enzymes (Figure 4) together with the crystal structures presented here, we were able to design a mutant of the wild-type *D. radiodurans* RecD2 enzyme that lacks the pin. This protein was overexpressed and purified from *E. coli*.

Basic helicase assays revealed that the mutant enzyme was deficient in helicase activity compared with wild type (Figure 5) to such an extent that even by increasing the level of enzyme 2000-fold we were unable to detect any activity. Hence the helicase activity is diminished by at least 3–4 orders of magnitude.

To analyse the reasons for the deficiency, we also looked at DNA binding and ATPase properties of the mutant (Figure 6). These studies revealed that the DNA-binding affinity of the mutant was impaired slightly. However, once saturated with DNA, the ATPase activity was the same as wild type. Consequently, helicase and ATPase activities have become uncoupled in the mutant enzyme. It has been shown that disruption of the helicase activity in PcrA by mutations in the 1B and 2B domains uncoupled helicase from translocase activity (Soultanas *et al*, 2000). As the mutant RecD2 retains full DNA-dependent ATPase activity it is likely that the protein also retains translocase activity, although this remains to be tested.

Discussion

The only other SF1B helicase for which we have structural information is the RNA helicase Upf1 (Cheng *et al*, 2006). Upf1 and RecD are similar in regions that correspond to the helicase domains (1A and 2A) but differ significantly outside these regions. Upf1 has a large N-terminal extension with a completely different structure to that of RecD enzymes and lacks a 2B domain entirely. Upf1 does have a 1B domain albeit rather larger and with a different fold to the 1B domain of RecD. However, this domain could function as a pin in a manner analogous to the 1B domain of RecD. Indeed, it has been shown that deletion of this domain results in a protein that is deficient in RNA binding, although this was not quantified (Cheng *et al*, 2006). The helicase activity of the

Figure 4 (A) Sequence comparison of RecD family enzymes in the region of the pin domain (blocked in yellow). Enzyme sequences are *E. coli*, *Salmonella typhimurium* (*S. typhi*), *Pseudomonas aeruginosa* (*P. aeru*), *Vibrio cholerae* (*V. chol*), *D. radiodurans* (*D. radio*) and *Haemophilus influenzae* (*H. infl*). (B) Structure of the pin in *D. radiodurans* RecD2, shown in yellow. (C) Difference electron density map showing the location of the iodine atoms (tan) and the pin domain (grey). The protein and DNA are coloured using the same scheme as Figure 2. Domain 1A is coloured green and domain 1B (pin) is in yellow.

Figure 5 DNA helicase assays using full-length wild-type RecD2 and the pin-less mutant. The substrate used was a 20 base-pair duplex with a 12 base 5'-tail. The top panel shows a time course with wild-type and pin-less mutant. The panel below shows the end products of a 16 min assay at increasing concentrations of the protein. A boiled control lane is also shown.

mutant protein was also not characterised, although it was shown that the protein was unable to support nonsense-mediated mRNA decay *in vivo*.

The role of the 1B domain pin in the helicase activity of RecD2 is quite striking. However, it has become apparent that helicases are modular (Singleton and Wigley, 2002) and that helicase activity results from a combination of three structur-

al components—an ATP-dependent nucleic acid translocation motor, substrate specificity domains and a duplex-splitting pin or wedge (Soultanas *et al*, 2000). Similar pin or wedge structures have been observed in many other helicases, including PcrA (Velankar *et al*, 1999), RecG (Singleton *et al*, 2001), RuvAB (Hargreaves *et al*, 1998; Roe *et al*, 1998; Ariyoshi *et al*, 2000), Hel308 (Buttner *et al*, 2007) and even elsewhere within the RecBCD complex (Singleton *et al*, 2004). The role of these structures is to exert an effect in a mechanical way to split DNA or RNA duplexes. The truncation of the pin in RecD2 produces an enzyme that is no longer a helicase but is a fully active ssDNA-dependent ATPase. Previous studies on the SF1A helicase PcrA have shown that helicase and translocation activities can be uncoupled by mutations in the 1B and 2B domains that interrupt interaction between the enzyme and the duplex region of the substrate (Soultanas *et al*, 2000). These data led to the realisation that the so-called 'helicase' motifs are actually motifs characteristic of an ATP-dependent nucleic acid translocase rather than of a helicase *per se*. The data presented here suggest this is also the case for SF1B enzymes such as RecD2 as ATPase and helicase activities can be uncoupled without compromising any of the conserved motifs.

Although the role of the pin is quite clear in RecD2, its function is less obvious in RecD within the RecBCD complex. In this context, RecD is presented with ssDNA originating from a duplex that has already been split by RecC further up the DNA-binding site. However, RecD1 has been shown to be a 5'-3' helicase when not a part of RecBCD (Dillingham *et al*, 2003). It may be that the pin has an important function under these circumstances, although it is not known whether RecD1 functions in isolation *in vivo*.

One unexpected finding from the structure of the Δ NRecD2 helicase structure was that domain 2B has an SH3 fold. The SH3 domain of RecD2 forms a discrete domain that is inserted within the overall RecA-like fold of domain 2A. SH3 domains are quite common in eukaryotic signal-transduction systems and were identified initially in the tyrosine kinase SRC. The SH3 domain is a peptide recognition module that binds peptides that are usually rich in proline residues because the bound peptide adopts a strained conformation that is favoured by prolines (Zarrinpar *et al*, 2003). The function of the binding is usually as a part of a regulation mechanism that activates or represses the proteins (Young *et al*, 2001). This raises the intriguing possibility that this domain might also function to bind peptides in RecD. However, what peptide would the SH3 domain bind to and why? There are two long-standing observations on RecBCD activity that might be relevant. The first is that the nuclease activity of the RecB subunit is repressed in RecBC but can be activated when RecD is a part of the complex (Korangy and Julin, 1993). The crystal structure revealed why the nuclease activity of RecB is repressed in the RecBCD complex (Singleton *et al*, 2004). The active site of the nuclease domain is blocked by a flexible linker peptide connecting the C-terminal nuclease domain of RecB to the N-terminal helicase domains, which also passes close to the RecD subunit. We postulated that this linker peptide might interact with RecD in some way that would release it from the nuclease-active site to relieve the inhibition. The surprising discovery that RecD contains an SH3 domain now suggests a mechanism for that interaction. The linker peptide does contain a number of proline

Figure 6 (A) DNA gel mobility shift assays using wild-type RecD2 and pin-less mutant proteins. The DNA used was either the same substrate used for the helicase assays or an oligo dT of 15 bases in length. (B) Graphical representation of the data in (A). (C) ATPase activity of wild-type and pin-less mutant proteins.

residues that could be ligands for the SH3 domain but any possible interaction remains only speculation at this stage.

A second possibility arises from single molecule experiments on the RecBCD complex that investigate the effect on the RecBCD complex when it encounters a Chi sequence (Spies *et al*, 2007). On normal DNA sequences, the RecBCD progresses rapidly typically at several hundred base pairs per second. However, when RecBCD encounters the eight base Chi sequence, it pauses briefly before continuing at roughly half of the pre-Chi rate. The reduction in translocation speed results from a switching from RecD being the lead motor prior to Chi, to the slower RecB subunit leading the complex after encountering Chi. It was proposed that, by a mechanism which remains unclear, the RecD subunit disengages from the DNA and appears to have no further function in the translocation after Chi. The observation that domain 2B contains a potential peptide-binding SH3 fold raises the possibility that, in response to a conformational change induced by Chi binding and/or recognition, this domain could bind to a region of the protein complex and hence lock the RecD subunit in an inactive conformation preventing it from contributing further to the translocation process. Of course, despite being an intriguing possibility, we have no data at present to demonstrate that the SH3 domain of RecD actually binds peptides. Although we have tried to produce the SH3 domain of *E. coli* RecD as a discrete peptide, the domain is poorly soluble (data not shown) precluding binding experiments at the present stage.

Interestingly, the DNA-binding domain of HIV-1 integrase (Vink *et al*, 1993) has been shown to have an SH3-like fold (Eijkelenboom *et al*, 1995; Lodi *et al*, 1995). This domain binds to double-stranded DNA but the structure of a complex with DNA has not been determined. There are also structures of two closely related archaeal proteins with an SH3 fold that

bind to double-stranded DNA, Sso7d (Gao *et al*, 1998) and Sac7d (Robinson *et al*, 1998). Binding of these proteins to DNA involves insertion into the minor groove, inducing a sharp bend in the helical axis. The position of the SH3-like 2B domain in the RecD and RecD2 structures suggests that this domain will also have a function in DNA binding in these enzymes. However, the structural data for RecBCD show that the domain binds single-stranded rather than double-stranded DNA. Consequently, it is not clear how this binding mode would relate to the Sso7d structures that bind to double-stranded DNA.

Materials and methods

Purification of wild-type and mutant *D. radiodurans* RecD2 proteins

The wild-type RecD2 gene is very GC rich, which caused problems with PCR reactions. Consequently, the gene was synthesised using the most frequently recognised codons in *E. coli* but to encode the same amino-acid sequence (Celtek, Nashville, US). This gene was then cloned into pET22b (Novagen) with a hexahistidine C-terminal tag and expressed in *E. coli* BL21. The cells were grown at 37°C until OD₆₀₀ 0.6 then induced with 1 mM IPTG for 3 h. Purification was by HisTrap FF, Hi Trap heparin and monoQ columns (GE Healthcare).

The N-terminal truncation of *D. radiodurans* RecD2 (Δ RecD2) was prepared by PCR, cloned into pET22b and expressed in *E. coli* BL21. The truncated protein lacked the first 150 amino acids. Protein expression and purification were similar to that of full-length protein. Selenomethionine-containing protein was expressed in B843 cells grown in LeMaster medium containing selenomethionine. The cells were grown for 8 h, induced with 1 mM IPTG and grown for a further 10 h. Purification was similar to that of the wild-type protein.

The pin-less RecD2 mutant protein gene was prepared by PCR. Residues 412–419 of the protein were replaced with a glycine residue. The gene was cloned into pET22b, then expressed and purified as described for the full-length protein.

Crystallisation and structure determination of the truncated *D. radiodurans* RecD2

Protein was buffer exchanged for crystallisation in 10 mM Tris pH 7.5, 1 mM DTT, 100 mM NaCl and concentrated to 8.5 mg/ml. Crystals of RecD2 grew in hanging drops set up above reservoir solution of 13–18% ethanol and 100 mM Tris pH 8.5.

Diffraction data were collected from the crystals of *D. radiodurans* Δ RecD2 frozen in 30% ethylene glycol at 100 K. A 2.2 Å native data set was collected on beamline ID23.1 at the European Synchrotron Radiation Facility. Data sets from the crystals derivatised with ethyl mercuric phosphate and sodium tungstate, respectively, and selenomethionine-substituted protein were collected in-house on a Rigaku RU-H3R generator and MAR345 detector system. The data sets were processed with XDS (Kabsch, 1993). The crystals belong to space group $P2_12_12_1$ with unit cell dimensions of $a = 54.6$, $b = 89.6$, $c = 131.7$ Å and contain one molecule in the asymmetric unit. Subsequent data reduction and other crystallographic procedures utilised the CCP4 program suite (Collaborative Computational Project Number 4, 1994) unless stated otherwise. The positions of heavy atoms in the two single site derivative crystals (Hg and tungstate) were located from a combination of isomorphous and anomalous difference Patterson maps but the selenomethionine data were too non-isomorphous to be useful for phasing and instead were just used to confirm methionine locations. Heavy atom positions were refined and phases were calculated from isomorphous and anomalous differences by the program SHARP (de La Fortelle and Bricogne, 1997). The initial phase estimates had a figure of merit of 0.35 but were improved by density modification. Initial rounds of structure refinement were carried out with CNS (Brunger, 2007). The final cycles of refinement were carried out using Refmac5 (Collaborative Computational Project Number 4, 1994) involving TLS refinement and restrained maximum likelihood refinement of position and atomic B-factor. Electron density for the first 43 residues at the N terminus was extremely poor and these residues were not built into the structure. Water molecules were identified by visual inspection of the $F_o - F_c$ difference electron density map. The final refined model contains 517 protein residues and 265 water molecules. Data and refinement statistics are provided in Table 1.

Purification, crystallisation and structure determination of the *E. coli* RecBCD–DNA complex

E. coli RecBCD was purified from *ΔrecBD* strain harbouring pETDuet-recB, pRSFDuet-recC and pCDFDuet-recD constructs. The cells were grown to OD₆₀₀ 0.6 at 37°C and then induced with 1 mM IPTG and grown overnight at 18°C. The protein and the hairpin DNA used for crystallisation (5′-IIIIAICIAATGCGAGCACT GCTATTCCCTAGCAGTCTCGCATIAGAIA-3′ where I stands for 5-iodouracil) were purified and processed as described previously (Singleton *et al*, 2004). The crystals were obtained using the hanging drop method with 100 mM Hepes pH 7.0, 300 mM calcium acetate, 6–8% PEG 20 000 as reservoir solution. Diffraction quality crystals were obtained by microseeding.

A 3.6 Å diffraction data set for the *E. coli* RecBCD–DNA complex was collected on beamline ID14.4 at the European Synchrotron Radiation Facility. Diffraction data were integrated using MOSFLM (Leslie, 1992) and scaled and merged using SCALA (Collaborative Computational Project Number 4, 1994). The crystal form was similar to that reported previously (Singleton *et al*, 2004), with space group $P2_12_12_1$ and unit cell dimensions of $a = 133.8$, $b = 192.9$ and $c = 334.8$ Å. Consequently, the same set of data was flagged for calculation of free R-factor that was used during refinement of the previous structure. Positions of the two molecules in the asymmetric unit were located by the molecular replacement program PHASER with a single molecule of RecBCD–DNA complex (PDB ID: 1W36) as search model. Because of the low resolution of the data, only limited refinement of the structure was carried out. A refinement protocol was implemented by monitoring free R-factor as suggested for low-resolution structures (DeLa Barre and Brunger, 2006). CNS was used to carry out rigid body refinement of the various subdomains of the molecule followed by NCS-restrained B-group refinement. A few cycles of positional refinement were carried out to eliminate short contacts and minimise bad geometry. The map calculated from the refined structure was used to manually rebuild large changes in protein structure and to build the longer 5′- and 3′-tails of the hairpin DNA. Model building of the tails of the DNA was made possible by strong electron density obtained for the 5-iodouridine bases. The iodine positions were identified from a difference Fourier map between data sets obtained from a crystal after exposure to low and high doses of X-ray irradiation (Gigant *et al*, 2005).

Helicase assays

Helicase assays utilised a synthetic substrate comprising two oligonucleotides to make a 20 base-pair duplex with a 12 base 5′ tail. The oligonucleotides used were 5′-TACAGCTACCTAGTCGATG TGCATACTACGGC-3′ and 5′-GCCGTAGTATGCACATCGAC-3′. The shorter oligonucleotide was 5′-end labelled. The substrate (1 nM) and protein (1 nM) were incubated in helicase buffer (100 mM NaCl, 10 mM MgCl₂, 0.1 mg/ml BSA, 0.1 mM DTT, 20 mM Tris acetate pH 7.5) at 30°C for 5 min before adding ATP (1 mM) to start the reaction. At set time points, samples were removed and the reaction was stopped by the addition of quench buffer (0.2% SDS, 40 mM EDTA, 10% glycerol, 0.1% bromophenol blue). The products were analysed by gel electrophoresis and visualised by exposure to a phosphorimager plate.

DNA gel mobility shift assays

These were carried as described previously (Dillingham *et al*, 2001).

ATPase assays

These assays were conducted as described previously (Seybert and Wigley, 2004).

Table 1 Crystallographic statistics

<i>D. radiodurans</i> RecD2	Native	Hg	W	Se
Resolution (Å)	20–2.2	20–3.2	20–2.8	20–3.0
Completeness (%)	98.0 (98.0)	97.7 (90.0)	97.1 (79.2)	97.4 (83.8)
R_{symm} (%)	3.8 (32.1)	6.7 (26.3)	2.4 (16.0)	74.1 (21.6)
R_{deriv} (%)	—	31.4	14.8	31.8
No. of sites	—	1	1	—
Phasing power	—	0.3	0.9	—
Mean figure of merit	0.35	—	—	—
<i>E. coli</i> RecBCD	IodoDNA	IodoDNA (low dose)	IodoDNA (high dose)	—
Resolution (Å)	30–3.6	30–4.0	30–4.0	—
Completeness (%)	96.7 (97.5)	99.6 (100)	93.1 (63.3)	—
R_{symm} (%)	7.5 (35.1)	10.0 (23.8)	11.0 (29.1)	—
Final models	<i>D. radiodurans</i> RecD2	<i>E. coli</i> RecBCD	—	—
R-factor (%)	23.1	24.6	—	—
R_{free} (%)	28.6	30.0	—	—
r.m.s.d. bond length (Å)	0.009	0.011	—	—
r.m.s.d. bond angles (deg)	1.3	1.7	—	—

Figures in parentheses indicate highest resolution shell.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

References

- Ariyoshi M, Nishino T, Iwasaki H, Shinagawa H, Morikawa K (2000) Crystal structure of the Holliday junction DNA in complex with a single RuvA tetramer. *Proc Natl Acad Sci USA* **97**: 8257–8262
- Brendza KM, Cheng W, Fischer CJ, Chesnik MA, Niedziela-Majka A, Lohman TM (2005) Autoinhibition of *Escherichia coli* Rep monomer helicase activity by its 2B subdomain. *Proc Natl Acad Sci USA* **102**: 10076–10081
- Brunger AT (2007) Version 1.2 of the crystallography and NMR system. *Nat Protoc* **2**: 2728–2733
- Buttner K, Nehring S, Hopfner K-P (2007) Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat Struct Mol Biol* **14**: 647–652
- Byrd AK, Raney KD (2004) Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA. *Nat Struct Mol Biol* **11**: 494–496
- Cheng Z, Muhlrud D, Lim MK, Parker R, Song H (2006) Structural and functional insights into the human Upf1 helicase core. *EMBO J* **26**: 253–264
- Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760–763
- de La Fortelle E, Bricogne G (1997) Maximum likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol* **276**: 472–494
- DeLa Barre B, Brunger AT (2006) Considerations for refinement of low-resolution crystal structure. *Acta Crystallogr D Biol Crystallogr* **62**: 923–932
- Dillingham MS, Soultanas P, Wiley P, Webb MR, Wigley DB (2001) Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. *Proc Natl Acad Sci USA* **98**: 8381–8387
- Dillingham MS, Spies M, Kowalczykowski SC (2003) RecBCD enzyme is a bipolar DNA helicase. *Nature* **423**: 893–897
- Dillingham MS, Webb MR, Kowalczykowski SC (2005) Bipolar DNA translocation contributes to highly processive DNA unwinding by RecBCD enzyme. *J Biol Chem* **280**: 37069–37077
- Eijkelenboom PAM, Puras Lutzke RA, Boelens R, Plasterk RH, Kaptein R, Hård K (1995) The DNA-binding domain of HIV1 integrase has an SH3-like fold. *Nat Struct Biol* **2**: 807–810
- Farah JA, Smith GR (1997) The RecBCD enzyme initiation complex for DNA unwinding: enzyme positioning and DNA opening. *J Mol Biol* **272**: 699–715
- Gao Y, Su S, Robinson H, Padmamabadhan S, Lim L, McCrary BS, Edmondson SP, Shriver JW, Wang AH (1998) The crystal structure of the hyperthermophile chromosomal protein Sso7d bound to DNA. *Nat Struct Biol* **5**: 782–786
- Gigant B, Wang C, Ravelli RBG, Roussi F, Steinmetz MO, Curmi PA, Sobel A, Knossow M (2005) Structural basis for the regulation of tubulin by vinblastine. *Nature* **435**: 519–522
- Gorbalenya AE, Koonin EV (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr Opin Struct Biol* **3**: 419–429
- Hargreaves D, Rice DW, Sedelnikova SE, Artymiuk PJ, Lloyd RG, Rafferty JB (1998) Crystal structure of *E. coli* RuvA with bound DNA Holliday junction at 6 Å resolution. *Nat Struct Biol* **5**: 441–446
- Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* **26**: 795–800
- Koranyi F, Julin DA (1993) Kinetics and processivity of ATP hydrolysis and DNA unwinding by the RecBC enzyme from *Escherichia coli*. *Biochemistry* **32**: 4873–4880
- Korolev S, Hsieh J, Gauss GH, Lohman TM, Waksman G (1997) Major domain swiveling revealed by the crystal structures of

Acknowledgements

We thank EMBO for a fellowship (KS) and the ESRF for access to X-ray beamlines. This study was funded by Cancer Research UK.

- complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* **90**: 635–647
- Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, no. 26
- Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, Gronenborn AM (1995) Solution structure of the DNA-binding domain of HIV-1 integrase. *Biochemistry* **34**: 9826–9833
- Nanduri B, Byrd AK, Eoff RL, Tackett AJ, Raney KD (2002) Pre-steady-state unwinding by bacteriophage T4 Dda helicase reveals a monomeric molecular motor. *Proc Natl Acad Sci USA* **99**: 14722–14727
- Robinson H, Gao Y, McCrary BS, Edmondson SP, Shriver JW, Wang AH (1998) The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature* **392**: 202–205
- Rocha EP, Cornet E, Michel B (2005) Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet* **1**: e15
- Roe SM, Barlow T, Brown T, Oram M, Keeley A, Tsaneva IR, Pearl LH (1998) Crystal structure of an octameric RuvA–Holliday junction complex. *Mol Cell* **2**: 361–372
- Seybert A, Wigley DB (2004) Distinct roles for ATP binding and hydrolysis at individual subunits of an archaeal clamp loader. *EMBO J* **23**: 1360–1371
- Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB (2004) Crystal structure of RecBCD reveals a machine for processing DNA breaks. *Nature* **432**: 187–193
- Singleton MR, Dillingham MS, Wigley DB (2007) Structure and mechanism of helicases and nucleic acid translocases. *Ann Rev Biochem* **76**: 23–50
- Singleton MR, Scaife S, Wigley DB (2001) Structural analysis of DNA replication fork reversal by RecG. *Cell* **107**: 79–89
- Singleton MR, Wigley DB (2002) Modularity and specialisation in superfamily 1 and 2 helicases. *J Bacteriol* **184**: 1819–1826
- Soultanas P, Dillingham MS, Wiley P, Webb MR, Wigley DB (2000) Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism. *EMBO J* **19**: 3799–3810
- Spies M, Amitani I, Baskin RJ, Kowalczykowski SC (2007) RecBCD enzyme switches lead motor subunits in response to chi recognition. *Cell* **131**: 694–705
- Subramanya HS, Bird LE, Brannigan JA, Wigley DB (1996) Crystal structure of a DExx box DNA helicase. *Nature* **384**: 379–383
- Tomko EJ, Fischer CJ, Niedziela-Majka A, Lohman TM (2007) A nonuniform stepping mechanism for *E. coli* UvrD translocation along single stranded DNA. *Mol Cell* **26**: 335–347
- Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB (1999) Crystal structures of complexes of PcrA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**: 75–84
- Vink C, Oude Groeneger AAM, Plasterk RH (1993) Identification of the catalytic and DNA-binding region of the human immunodeficiency virus type I integrase protein. *Nucleic Acids Res* **21**: 1419–1425
- Wang J, Julin DA (2004) DNA helicase activity of the RecD protein from *Deinococcus radiodurans*. *J Biol Chem* **279**: 52024–52032
- Wong J, Lucius AL, Lohman TM (2005) Energetics of DNA end binding by *E. coli* RecBC and RecBCD helicases indicate loop formation in the 3′-single-stranded DNA tail. *J Mol Biol* **352**: 765–782
- Young MA, Gonfloni S, Superti-Furga G, Roux B, Kuriyan J (2001) Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**: 115–126
- Zarrinpar A, Bhattacharyya RB, Lim WA (2003) The structure and function of proline recognition domains. *Sci STKE* **179**: re8