Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks

Martin R. Singleton[1]*, Mark S. Dillingham[2]*, Martin Gaudier[1], Stephen C. Kowalczykowski[3] & Dale B. Wigley[1]

[1]*Cancer Research UK Clare Hall Laboratories, The London Research Institute, Blanche Lane, South Mimms, Potters Bar, Herts. EN6 3LD, UK*
[2]*National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK*
[3]*Sections of Microbiology and Molecular & Cellular Biology, Center for Genetics and Development, University of California at Davis, Davis, California 95616, USA*

* These authors contributed equally to this work

RecBCD is a multi-functional enzyme complex that processes DNA ends resulting from a double-strand break. RecBCD is a bipolar helicase that splits the duplex into its component strands and digests them until encountering a recombinational hotspot (Chi site). The nuclease activity is then attenuated and RecBCD loads RecA onto the 3′ tail of the DNA. Here we present the crystal structure of RecBCD bound to a DNA substrate. In this initiation complex, the DNA duplex has been split across the RecC subunit to create a fork with the separated strands each heading towards different helicase motor subunits. The strands pass along tunnels within the complex, both emerging adjacent to the nuclease domain of RecB. Passage of the 3′ tail through one of these tunnels provides a mechanism for the recognition of a Chi sequence by RecC within the context of double-stranded DNA. Gating of this tunnel suggests how nuclease activity might be regulated.

Double-strand breaks in DNA are potentially lethal to a cell if not repaired. They arise in several ways, including DNA damage from ionizing radiation and as a result of single-strand nicks that produce a double-strand break as a replication fork passes by them (reviewed in ref. 1).

In eubacteria, double-strand breaks are repaired mainly by the homologous recombination pathway in a process initiated by the RecBCD/AddAB family of enzymes[1]. RecBCD acts in a complicated way using a variety of different enzyme activities that are associated with the complex (Fig. 1). A combination of helicase and nuclease activities digests away the DNA until a Chi ('crossover hotspot instigator') site is encountered. Chi has the sequence 5′-GCTGGTGG-3′ and is recognized as single-stranded DNA (ssDNA) but within a double-stranded context (that is, as the duplex is unwound[2–4]). At this point the protein pauses[5] and the nuclease activity is both reduced[6] and its polarity switched[7]. The final cleavage event on the 3′ tail occurs at or within a few bases to the 3′ side of Chi[8], and the 3′ tail is then protected from further digestion. Finally, RecBCD enzyme loads RecA protein onto the 3′ tail to initiate recombination by means of the RecA pathway[9].

The RecBCD enzyme comprises three subunits arranged as a 330-kDa heterotrimer that is fully functional without the need for further oligomerization[10]. The extraordinary range of enzyme activities catalysed by RecBCD can be attributed to the three subunits as follows: RecB is a 3′–5′ helicase and multi-functional nuclease[11,12], RecC recognizes Chi[13] and RecD is a 5′–3′ helicase[14,15].

To gain a better understanding of the multifaceted mechanism of this DNA-processing machine we have determined the crystal structure of a complex of *Escherichia coli* RecBCD enzyme bound to a blunt-ended DNA hairpin. In addition to revealing the architecture of the enzyme, the structure shows how the two motor activities and the nuclease are coupled, allowing the recognition of Chi within the context of double-stranded DNA and suggesting a mechanism by which this regulates the polarity of the nuclease activity.

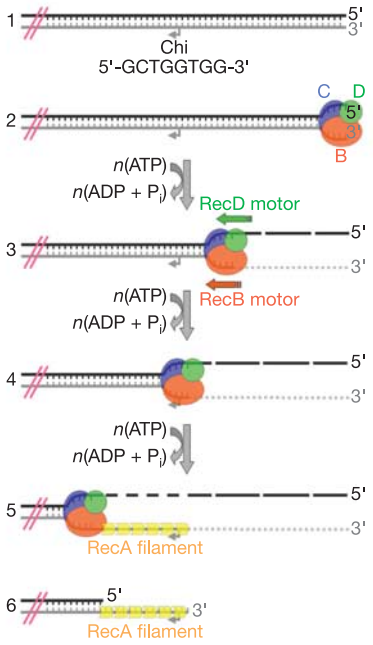

Figure 1 The processing of double-strand breaks by RecBCD enzyme. RecB is coloured orange, RecC is blue and RecD is green. In stage 1, a double-strand break is created as a result of DNA damage or a collapsed replication fork. The location of the eight base Chi site (sequence as shown) is represented by a bent arrow. In stage 2, RecBCD binds to the end of the DNA duplex and initiates unwinding. In stage 3, ATP-dependent DNA unwinding progresses and is coupled to the endonucleolytic digestion of both DNA strands. Although the 3′ tail is cleaved frequently, the 5′ tail is cleaved much less often. In stage 4, on encountering a Chi site in the 3′ terminated strand, the enzyme pauses and digestion of the 3′ tail ceases. In stage 5, cleavage of the 5′ tail becomes more frequent and RecA protein is loaded on the 3′ tail. In stage 6, RecBCD protein dissociates leaving a RecA-coated 3′ tail that can initiate homologous recombination.

Structure of the bound DNA

RecBCD has a high affinity for blunt DNA ends[16] with a footprint extending about 20 base pairs[17]. Consequently, we designed a self-complementary 43-mer oligonucleotide that forms a 19-base pair duplex with a five-base hairpin turn at one end and a blunt end at the other. As predicted, the DNA is bound with the blunt end of the duplex into the RecBCD complex. Remarkably, although the first 15 base pairs of the duplex from the hairpin form regular B-form DNA, the last four base pairs at the blunt end have been unwound by the protein. Experiments involving the chemical modification of RecBCD–DNA complexes have shown that DNA is unwound in the initiation complex even in the absence of ATP[18].

RecB protein: endonuclease and 3′–5′ helicase

The RecB subunit is a helicase with 3′–5′ directionality[19]. Sequence analysis has identified RecB as a member of the helicase Superfamily 1 (SF1)[20] in common with several other helicases, such as PcrA and Rep, whose structures have been determined[21,22]. Accordingly, the structure of the amino-terminal region of the RecB subunit is similar to that of other SF1 helicases (Fig. 2a) with regions equivalent to the canonical 1A, 1B, 2A and 2B domains. The structures of all four of these domains are very similar to their equivalents in PcrA (with a root-mean-square deviation (r.m.s.d.) for each domain of 1.9–3.2 Å for equivalent Cα positions) but domain 1B contains an additional large insertion compared with PcrA. These residues form a discrete 'arm' structure that extends from the surface of the complex alongside the duplex DNA substrate. The relative orientation of domains 1A and 2A is similar to that expected for the enzyme in the absence of ATP[23]. However, domains 1B and 2B are in orientations that are quite different from those of any of the PcrA or Rep structures and seem to fulfil very different roles. For example, domain 2B of RecB forms a major interface with the RecC protein (see below) but this domain contacts duplex DNA in PcrA.

The 3′ tail of the bound DNA runs across domain 2A of the RecB subunit with the last three bases in a similar conformation to that seen for the equivalent residues in the ssDNA tail in the PcrA–DNA complex[23]. Domains 1A and 2A are the canonical helicase motor domains responsible for the ATP-dependent translocation of the protein along ssDNA[23–25]. Similar motor domains are found in both SF1 and SF2 helicases[26].

Domain 3, the RecB nuclease domain, is connected to the remainder of the protein by a long linker region of about 70 amino acids. This linker region has been shown to be sensitive to proteolytic cleavage[12]. As suggested previously[27], the fold of this domain is similar to that of the core of the λ exonuclease[28] (r.m.s.d. of 3.5 Å for Cα positions). Residue Asp 1080, known to be important for nuclease activity[29], is located at a position equivalent to the active site of λ exonuclease. There are three conserved acidic residues and a lysine residue in the active site (Glu 1020, Asp 1067, Asp 1080 and Lys 1082). Electron density that we attribute to a bound calcium ion from the crystallization medium is also

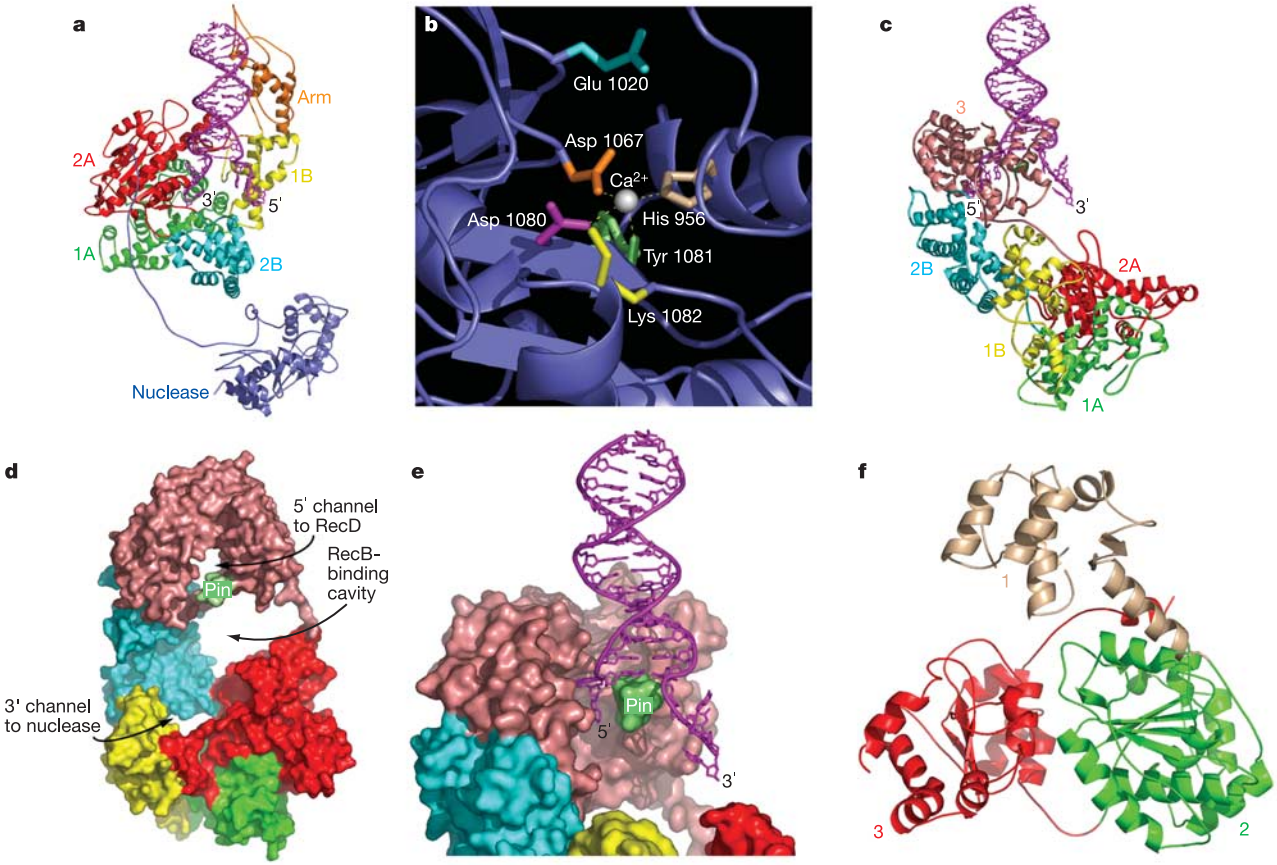

Figure 2 Structures of the individual RecBCD subunits. **a**, Domain structure of the RecB subunit. **b**, Close-up of the active site of the nuclease. The calcium ion (grey sphere) is coordinated to the side chains of three residues (His 956, Asp 1067 and Asp 1080) and the main-chain carbonyl of Tyr 1081. **c**, Domain structure of the RecC subunit. The pin region is highlighted. **d**, Space-filling representation of RecC, showing the channels through the protein in the same colour scheme as in **c**. **e**, Close-up view of the pin region in RecC, showing how the DNA duplex is split across this feature of the RecC protein. **f**, Domain structure of the RecD subunit. Domains 2 and 3 are equivalent to the canonical 1A and 2A domains of other SF1 helicases. The images in **a**, and also those in Figs 2 and 4, were created with PyMOL (http://www.pymol.org).

present in this region (Fig. 2b). It is likely that this calcium ion is bound at the position where a magnesium ion would normally bind in the active site of the protein, because calcium has been shown to be an inhibitor of the nuclease activity[30].

RecC protein: duplex splitting and recognition of Chi

The overall fold of RecC is the most remarkable of the subunits (Fig. 2c) and contains three large channels that run through the protein (Fig. 2d). The largest of these accommodates the 2B domain of RecB and provides a major interface between the proteins. The other two channels are pathways along which the single-stranded tails of the DNA run to, or from, the two helicase subunits (see below). The protein can be divided into three domains. Quite unexpectedly, domains 1 and 2 have the same fold as canonical SF1 helicases (r.m.s.d. ~3.5 Å for equivalent Cα positions of PcrA), despite a lack of sequence similarity including any of the characteristic helicase motifs. This strongly suggests that the RecC subunit evolved from a helicase ancestor but has now lost this helicase function. Furthermore, the similarity to helicase domains suggests a potential ssDNA-binding site that could provide a mechanism for the recognition of Chi as it passes through this site (see below).

Domain 3 of RecC is connected to the rest of the protein by a 20-residue linker that also forms the sides of the channel into which domain 2B of RecB binds (Fig. 3a). The fold of this domain has not previously been observed, according to the DALI server[31]. Domain 3 makes intimate contacts with each of the two separated strands of the DNA substrate running in different directions either side of a 'pin' that protrudes from the surface of the protein and serves to split the duplex (Fig. 2e). The 5′ tail passes through another channel in RecC and heads towards the motor domains of RecD.

The open structure of RecC explains why it is readily proteolysed in solution. The major cleavage site of RecC is located after residue 804; this would release the C-terminal domain 3 of the protein, as suggested previously[32]. Deleting this domain eliminates RecD assembly within the RecBCD complex and this seems to relate to a small area of contacts with domain 2 of RecD, although the major interface between these proteins is between domain 1 of RecD and subdomain 2B of RecC (Fig. 3).

RecD protein: 5′–3′ helicase

Although identified as a ssDNA-dependent ATPase several years ago[33] and shown to contain several characteristic helicase motifs[20], it was only recently shown that the RecD subunit has 5′–3′ helicase activity[14,15]. The structure of the RecD subunit is the first of a SF1 helicase with 5′–3′ directionality. The protein comprises three domains (Fig. 2f). Domain 1 forms the principal interface between RecD and domain 2B of RecC. Domains 2 and 3 are similar to the motor domains (1A and 2A) of other SF1 helicases (r.m.s.d. 3.9 and 3.3 Å on Cα positions, respectively). There is a disordered region of about 55 amino acids in domain 3, which would correspond to domain 2B of PcrA. The rest of domain 3 is also highly mobile, with higher *B* factors than for the rest of the structure.

Although most SF1 helicases have 3′–5′ polarity, some (such as RecD and Dda) have the opposite polarity[14,34]. There are two simple ways in which a SF1 helicase could be adapted to alter the directionality to 5′–3′, either to reverse the direction of translocation along the bound DNA or to bind the DNA in the opposite orientation. The path taken by the bound ssDNA tail across RecC would present the 5′ tail to the helicase domains in an orientation similar to that seen for other helicases, suggesting that the alteration in directionality of helicases in SF1 is based on the direction of movement along the bound DNA rather than the directionality of ssDNA binding itself. It has been shown that residues in motif 1a have a pivotal function in directional translocation in SF1 helicases[35], and it is probably significant that these residues are absent from RecD.

The RecBCD complex

The structures of the individual subunits of RecBCD reveal why they are difficult to express and prepare, unless complexed together. The RecB and RecC proteins, in particular, are wrapped tightly around one another. In isolation, both proteins have exposed linker regions that are susceptible to proteolysis but protected when the proteins are bound to each other. This intimate embrace of the proteins is essential for their function and allows the ssDNA strands to be shepherded through the enzyme, providing a temporal link between their separation and subsequent endonucleolytic digestion.

The bound DNA makes extensive contacts with the RecB and RecC subunits. The 'arm' structure of the RecB protein interacts with duplex DNA ahead of the fork about 12 base pairs from the junction. The RecC protein contacts both strands of DNA and splits them before feeding the 3′ strand to the RecB protein and the 5′ tail

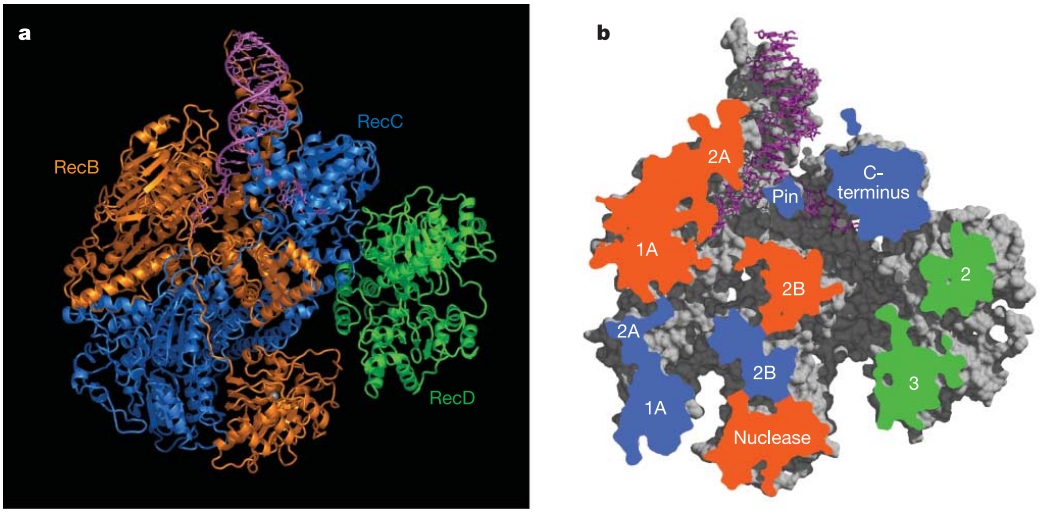

Figure 3 Structure of the RecBCD–DNA complex. **a**, The entire RecBCD–DNA complex. The bound DNA is coloured magenta and the bound calcium ion is a grey sphere. **b**, Cutaway view showing the channels through the RecBCD complex. The cutaway surface of the RecB subunit is orange, that of RecC is blue and that of RecD is green. The bound DNA is coloured magenta. Numbers refer to the domains of the appropriate subunits. Domains 2 and 3 of RecD are equivalent to canonical helicase domains 1A and 2A. Both **b** and Fig. 4a were created with MSMS[49] and rendered with Raster3D[50].

to the RecD subunit. Consequently, the protein complex covers the first 16 bases of the 5′-terminated strand of the DNA and 13 bases on the 3′-terminated strand. These results are consistent with DNase I footprinting studies of the initiation complex that showed protection of 20 or 21 bases on the 5′ tail and 16 or 17 bases on the 3′ tail[17]. Furthermore, these studies showed that ultraviolet radiation crosslinked the 3′ tail to the RecB protein, which is consistent with the crystal structure. The 5′ tail was crosslinked to both the RecC and RecD subunits. Although we observe extensive contacts between the RecC subunit and the bound DNA, there are no contacts with the RecD subunit in our structure, although, as unwinding of the DNA progresses, the 5′ tail would be fed from the RecC subunit towards the motor domains of RecD. This large footprint also relates to experiments that examined the activity of RecBC on gapped duplex substrates[36]. These experiments showed that RecBC can pass over single-stranded gaps of up to 23 bases provided that the 3′ strand is intact. This property of the protein most probably relates to the arm structure of RecB that contacts duplex DNA ahead of the junction, allowing the protein to step over intervening single-strand regions.

RecBCD has both 3′–5′ and 5′–3′ helicase activities, contributed by the RecB and RecD subunits, respectively. These subunits drive DNA unwinding by acting as ssDNA motors, pulling the two antiparallel strands of the DNA across the pin of the RecC subunit and thus splitting the duplex. RecB is closely related to the well-studied SF1 helicase, PcrA. The N-terminal region of RecB is the same as PcrA with the exception of an insertion of about 115 residues within domain 1B (the 'arm') that contacts the duplex ahead of the junction in a manner similar to contacts in the PcrA–DNA complex[23], except that this contact is provided by the 2B domain in PcrA. In PcrA, the strength of this contact is dependent on ATP binding[25], and it may be that this is also true for RecBCD. The present structure was determined in the absence of ATP and the *B* factors of the residues in the arm are considerably higher than the average for the rest of the protein, suggesting that this part of the structure is more mobile. Furthermore, in PcrA, the duplex is split across a 'pin' that protrudes from domain 2B in a manner similar to the pin in RecBCD, except that in the latter case the pin is contributed by the RecC subunit. Consequently, we observe an apparent conservation of mechanism but contributed by different parts of the structure.

The two helicase motors can apparently work independently of one another. Electron microscopy data have shown that loops can be extruded from the complex when the helicase activity of either subunit is inactivated[15]. A ssDNA loop is also extruded from the protein on encountering a Chi site[5].

Isolated RecB protein has been shown to have weak 3′–5′ helicase activity[19], but the processivity of this activity is greatly enhanced when complexed with RecC[12]. The crystal structure reveals how RecC contributes to the helicase activity both directly, through the pin that splits the DNA duplex at the junction, and indirectly, by sequestering the 3′ tail of the substrate and preventing reannealing. Passage of the DNA through the complex would also provide a steric block to dissociation of the protein, a tactic frequently adopted by other processive enzymes in DNA metabolism.

PcrA has been shown to use one ATP molecule for each base pair that it translocates and, by inference, during helicase action as well[24]. By contrast, RecBCD uses between two and three ATP per base pair separated[16,37] although RecBC, or RecBCD containing an ATPase-deficient RecD subunit, uses only one ATP per base pair[37]. This suggests that one ATP molecule is used by each helicase motor during translocation, a decrease in efficiency that is apparently tolerated in exchange for the increased processivity of RecBCD in comparison with RecBC, or the ability of the holoenzyme to bypass single-stranded gaps.

The discovery that RecC might be a defunct helicase raises some important implications about the interaction of DNA with the complex. The channel for the 3′-terminated strand passes from RecC, across the RecB motor domains, then back through RecC until it emerges adjacent to the nuclease domain of RecB (Fig. 3b). In passing through RecC, the channel runs across the top of the 'helicase' domains, which corresponds to the ssDNA-binding site in helicases. Consistent with this proposal is the observation that the channel is lined with conserved basic and exposed hydrophobic residues, as would be expected for a site that binds ssDNA. The ssDNA-binding site in SF1 helicases spans about eight bases[22,23], and it is probably significant that the Chi sequence is also eight bases in length (5′-GCTGGTGG-3′). Recognition of ssDNA by helicases is non-specific with regard to sequence. This is achieved by providing complementary charge for the phosphodiester backbone and exposed aromatic residues to stack with the DNA bases. Contacts with the functional groups on the bases are avoided, because this would necessarily result in unwanted sequence specificity. However, the introduction of such contacts could transform a canonical

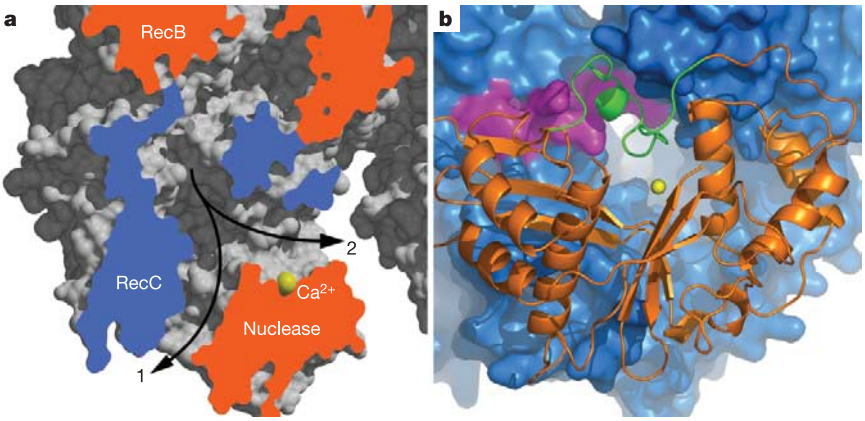

Figure 4 Alternative exits from the 3′ tunnel. **a**, Cutaway view of the exit channels running each side of the nuclease domain. The calcium ion at the nuclease active site is coloured yellow. For the purposes of this figure, the loop that blocks the channel has been omitted. There are two exit channels from the RecC subunit. One of these (labelled 1) bypasses the nuclease site. Access to the nuclease active site through channel 2 is blocked by a helix in the structure. **b**, The interface between the RecC subunit and the RecB nuclease domain viewed across the RecB nuclease active site. RecC is shown in blue as a space-filling representation, with the region affected in the RecC* mutants highlighted in magenta. The RecB nuclease domain is overlaid as an orange ribbon and the bound calcium ion as a yellow sphere. Access to the nuclease active site from the channel is blocked by a loop from the nuclease domain that includes an α-helix (residues 909–930, coloured green).

helicase ssDNA translocation site into a ssDNA 'scanning' site that would provide a convenient mechanism by which RecC could recognize a Chi sequence as it passed through the channel.

There is a group of RecC mutants with altered Chi recognition[38–40]. The mutations (residues 647–663) map to a region at the opening of the channel that accommodates the 3′ tail, at the interface with the nuclease domain of RecB (Fig. 4a). It has been shown that on Chi recognition the last DNA cleavage event is no more than four to six bases to the 3′ side of the Chi site[8]. This suggests that the end of the Chi recognition site should be 15–25 Å from the nuclease active site depending on the degree of extension of the ssDNA strand. Residues 647–663 are located 20–25 Å from the acidic triad in the nuclease domain. These mutations are frameshifts across several residues rather than point mutations, mostly introducing deletions in the sequence with undetermined consequences on the local structure. Furthermore, this region is the linker between the 2A and 2B domains, so mutations here could have significant consequences on the overall conformation of the protein. It is therefore difficult to draw detailed information from these mutants other than identifying this region as important for Chi recognition. The complex pathway taken by the ssDNA suggests a mechanism for the observation that Chi sequences are recognized as ssDNA but arising from double-stranded DNA that is unwound by the enzyme[2–4]. During unwinding, the channels, and hence also the Chi-binding site, would be occupied with ssDNA, thus blocking the binding of oligonucleotides containing Chi sequences when added *in trans* to the substrate after DNA unwinding has begun[41].

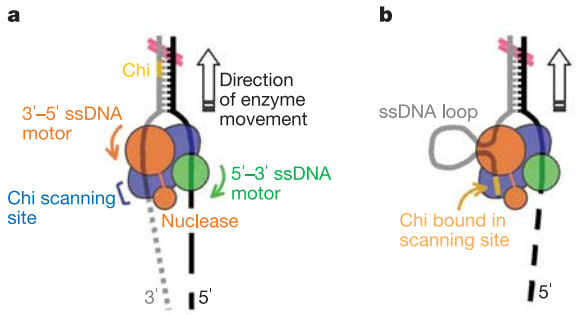

Figure 5 Diagram outlining the changes in RecBCD that occur after encountering a Chi site. **a**, Before Chi, the enzyme progresses along duplex DNA using the bipolar motors of the RecB (orange) and RecD (green) subunits. The 3′ tail is digested processively but the 5′ tail is cut much less frequently. **b**, On encountering Chi, the RecC subunit (blue) binds tightly to the 3′ tail, preventing further digestion of this strand. The 5′ tail is now able to access the nuclease site more frequently and is degraded more fully. The enzyme continues to advance along the DNA resulting in a loop out from the RecB subunit that can be loaded with RecA protein.

Regulation of nuclease activities

The nuclease activities of RecB are carefully regulated in several ways, including switching cleavage between 5′ and 3′ tails, 3′ nuclease attenuation after Chi, and nuclease activation by RecD. The location of the nuclease active site allows processive hydrolysis of the 3′ tail as it emerges from the RecC subunit. However, proximity of the 5′ tail would enable it to compete with the 3′ tail for binding at the nuclease, although this would be less frequent owing to the less favourable disposition of the 5′ tail, which would access the site from the opposite direction. Nuclease digestion of the 3′ tail would be attenuated after Chi simply as a consequence of binding tightly to the RecC subunit, resulting in more frequent cleavage of the 5′ tail resulting from less hindered access. Although we do not need to invoke a conformational change in the protein after it has encountered Chi, we cannot exclude this possibility.

Interestingly, there are two exits from the 3′ tunnel as it emerges from RecC (Fig. 4a). One of these passes along the back of the nuclease domain, bypassing the nuclease active site. The other is blocked by an α-helix from RecB (Fig. 4b). This would provide a simple mechanism for controlling the nuclease activity of the complex such that access to the active site would be blocked if this helix remained as positioned in this structure, forcing the strand to exit without passing through the nuclease. The helix is connected to the rest of the protein by two long flexible linkers, suggesting that it might be able to swing out of the way. The RecD subunit is involved in stimulating the nuclease activity[42] and, although the trigger for movement of this helix remains unclear, domain 3 of RecD is in close proximity.

A mechanism for processing DNA ends

The crystal structure allows us to understand RecBCD activity in better molecular detail (Fig. 5). Once bound to a blunt end, the initiation complex is formed in which the double strands of the duplex are opened up and fed towards the motor subunits. The DNA is unwound by the combined activities of the two helicase subunits

Table 1 **Crystallographic statistics**

Data collection	Native	TaBr1	TaBr2	Se1	IodoU
Resolution (Å)	30–3.1	30–3.8	30–5.5	20–3.2	30–3.1
Completeness (%)	99.8	91.4	91.3	99.5	95.8
R_{symm} (%)	7.1	14.9	9.3	11.3	8.5
R_{deriv} (%)	–	16.2	22.3	13.5	17.4
No. sites	–	10	10	119	18
Phasing power	–	0.32	0.97	0.60	0.28

Anomalous data statistics	Se1 (peak)	Se2 (inflection)	Se3 (remote)	TaBr3 (peak)
Resolution (Å)	30–3.2	30–3.4	30–3.6	20–5.0
Wavelength (Å)	0.97942	0.97960	0.93927	1.25485
Completeness (%)	99.5	99.7	99.7	90.3
R_{symm} (%)	11.3	11.9	14.3	6.8
Phasing power	0.44	0.63	0.27	–

Final model	Value
R factor (All data excluding free *R* set) (%)	24.2
R_{free} (5% of data) (%)	29.6
R.m.s.d. bond length (Å)	0.018
R.m.s.d. bond angle (deg)	2.16

that pull the DNA strands across the pin of RecC, thus splitting the duplex. The 5′ tail is fed to the RecD helicase and then onto the nuclease domain of RecB for digestion. The 3′ tail is fed along a channel within the protein complex that emerges at the nuclease active site. Consequently, this strand is digested processively as the protein complex progresses along the DNA. The 5′ tail is also digested, but less frequently because it is not situated as favourably as the 3′ tail and cannot compete as effectively for the nuclease active site. As the 3′ tail is fed through the RecC subunit it is able to scan the DNA for a Chi sequence. On encountering Chi, the RecC subunit binds tightly to this site in the 3′ tail, thus preventing further digestion of this strand, with the final cleavage event taking place at, or close to, the Chi site. The 5′ tail is now able to access the nuclease site more readily and is therefore cleaved more frequently. Finally, RecBCD loads RecA protein onto the 3′ tail by a mechanism that is not yet fully understood but involves the nuclease domain of RecB[43] before dissociating from the DNA. □

Methods

Protein expression and purification, and DNA preparation

Full-length RecB, RecC and RecD were expressed together with the use of the pPB520 and pPB800 plasmids[11] in a *ΔrecBCD E. coli* strain (V330) containing the $LacI^q$ overexpression plasmid pMS421. The RecBCD complex was purified by ammonium sulphate fractionation (0–50% w/v), followed by Fast Flow Q (Amersham) anion-exchange chromatography. RecBCD-containing fractions were dialysed overnight and further purified by heparin–Sepharose (Amersham), Mono-Q (Amersham) and gel-filtration (Superdex 200; Amersham) chromatography. The DNA hairpin was prepared by heating the self-complementary oligonucleotide (5′-TCTAATGCGAGCACTGCTATTCCCTAG CAGTGCTCGCATTAGA-3′) to 95 °C for 5 min followed by rapid cooling in ice. The DNA was then purified by anion-exchange chromatography and desalted by gel filtration.

Crystallization and structure determination

Protein was concentrated to 10 mg ml^{-1} in 10 mM Tris-HCl pH 7.5, 100 mM NaCl and 1 mM dithiothreitol, with DNA present in a 1.3:1 molar excess. Hanging drops were set up above reservoir solutions of 100 mM HEPES pH 7.0, 200 mM calcium acetate, 4–8% (w/v) poly(ethylene glycol) 20000. Crystal nucleation was initiated by microseeding, and crystals grew to a full size of about $400 \times 100 \times 60\,\mu m^3$ within 5 days.

Crystals were cryoprotected by rapid sequential transfer through the reservoir solution with the addition of ethylene glycol in 5% (v/v) steps to a final concentration of 30%. Diffraction data were collected on beamlines 14.1 and 14.4 at the European Synchrotron Radiation Facility, Grenoble. The crystals were of space group $P2_12_12_1$ with unit cell dimensions $a = 133$ Å, $b = 187$ Å, $c = 335$ Å and contained two RecBCD–DNA complexes in the asymmetric unit.

Diffraction data (Table 1) were integrated and scaled together with the XDS package[44]. The positions of ten Ta_6Br_{12} clusters were determined by analysis of anomalous diffraction data collected at the Ta L_{III} edge (TaBr3) using the program Shake 'n Bake[45]. However, these data were too non-isomorphous to be useful for multiple isomorphous replacement (MIR) phasing with native data. Consequently, data from two lower-occupancy soaks were collected and used for MIR phasing (TaBr1 and TaBr2). Phases calculated from these data sets allowed the location of 18 iodine sites in data from a crystal grown with iodouracil-substituted oligonucleotide and 119 selenium sites in selenomethionine-substituted protein. Heavy-atom parameters were refined and phases calculated with the program SHARP[46], using a combination of two runs, one with regular MIRAS phasing (including the data from Se1 as a derivative and using a spherically averaged approximation for the tantalum clusters) and the other with multiwavelength anomalous diffraction phasing for the three selenium data sets collected from the same crystal (Se1, Se2 and Se3). Phases from each run were combined, resulting in a set of phase estimates with an overall mean figure of merit of 0.33. These initial phases were improved by density modification with twofold non-crystallographic symmetry (NCS) averaging in DM[47]. These maps permitted location of the DNA and building of most of one complex. Initial refinement was performed by using the CNS package[48] with NCS constraints. Further rounds of model building were undertaken between refinement cycles. Towards the end of refinement the NCS constraints were relaxed and parts of the structure that did not obey local symmetry were refined independently. The final model has an R factor of 24.2% ($R_{free} = 29.6\%$).

Received 1 July; accepted 2 September 2004; doi:10.1038/nature02988.

1. Kowalczykowski, S. C. Initiation of genetic recombination and recombination-dependent replication. *Trends. Biochem. Sci.* **25,** 156–165 (2000).
2. Ponticelli, A. S., Schultz, D. W., Taylor, A. F. & Smith, G. R. Chi-dependent DNA strand cleavage by the RecBC enzyme. *Cell* **41,** 145–151 (1985).
3. Taylor, A. F., Schultz, D. W., Ponticelli, A. S. & Smith, G. R. RecBC enzyme nicking at Chi sites during DNA unwinding: location and orientation-dependence of the cutting. *Cell* **41,** 153–163 (1985).
4. Bianco, P. R. & Kowalczykowski, S. C. The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5′-GCTGGTGG-3′. *Proc. Natl Acad. Sci. USA* **94,** 6706–6711 (1997).
5. Spies, M. *et al.* A molecular throttle: the recombination hotspot Chi controls DNA translocation by the RecBCD helicase. *Cell* **114,** 647–654 (2003).
6. Dixon, D. A. & Kowalczykowski, S. C. The recombination hotspot Chi is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* RecBCD enzyme. *Cell* **73,** 87–96 (1993).
7. Anderson, D. G. & Kowalczykowski, S. C. The recombination hot spot Chi is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes Dev.* **11,** 571–581 (1997a).
8. Taylor, A. F. & Smith, G. R. Strand specificity of nicking of DNA at Chi sites by RecBCD enzyme. *J. Biol. Chem.* **270,** 24459–24467 (1995b).
9. Anderson, D. G. & Kowalczykowski, S. C. The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a Chi-regulated manner. *Cell* **90,** 77–86 (1997b).
10. Taylor, A. F. & Smith, G. R. Monomeric RecBCD enzyme binds and unwinds DNA. *J. Biol. Chem.* **270,** 24451–24458 (1995a).
11. Boehmer, P. E. & Emmerson, P. T. *Escherichia coli* RecBCD enzyme: inducible overproduction and reconstitution of the ATP-dependent deoxyribonuclease from purified subunits. *Gene* **102,** 1–6 (1991).
12. Yu, M., Souaya, J. & Julin, D. A. The 30 kDa C-terminal domain of the RecB protein is critical for the nuclease activity, but not the helicase activity, of the RecBCD enzyme from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **95,** 981–986 (1998).
13. Handa, N., Ohashi, S., Kusano, K. & Kobayashi, I. Chi*, a chi-related 11-mer sequence partially active in an *E. coli recC1004* strain. *Genes Cells* **2,** 525–536 (1997).
14. Dillingham, M. S., Spies, M. & Kowalczykowski, S. C. RecBCD enzyme is a bipolar DNA helicase. *Nature* **423,** 893–897 (2003).
15. Taylor, A. F. & Smith, G. R. RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature* **423,** 889–893 (2003).
16. Roman, L. J. & Kowalczykowski, S. C. Characterisation of the adenosinetriphosphatase activity of the *Escherichia coli* RecBCD enzyme: relationship of ATP hydrolysis to the unwinding of duplex DNA. *Biochemistry* **28,** 2873–2881 (1989).
17. Ganesan, S. & Smith, G. R. Strand-specific binding to duplex DNA ends by the subunits of the *Escherichia coli* RecBCD enzyme. *J. Mol. Biol.* **229,** 67–78 (1993).
18. Farah, J. A. & Smith, G. R. The RecBCD enzyme initiation complex for DNA unwinding: enzyme positioning and DNA opening. *J. Mol. Biol.* **272,** 699–715 (1997).
19. Boehmer, P. E. & Emmerson, P. T. The RecB subunit of the *Escherichia coli* RecBCD enzyme couples ATP hydrolysis to DNA unwinding. *J. Biol. Chem.* **267,** 4981–4987 (1992).
20. Gorbalenya, A. E. & Koonin, E. V. Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3,** 419–429 (1993).
21. Subramanya, H. S., Bird, L. E., Brannigan, J. A. & Wigley, D. B. Crystal structure of a DExx box helicase. *Nature* **384,** 379–383 (1996).
22. Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M. & Waksman, G. Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. *Cell* **90,** 635–647 (1997).
23. Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97,** 75–84 (1999).
24. Dillingham, M. S., Wigley, D. B. & Webb, M. R. Unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. *Biochemistry* **39,** 205–212 (2000).
25. Soultanas, P., Dillingham, M. S., Wiley, P., Webb, M. R. & Wigley, D. B. Uncoupling DNA translocation and helicase activity in PcrA: direct evidence for an active mechanism. *EMBO J.* **19,** 3799–3810 (2000).
26. Singleton, M. R. & Wigley, D. B. Modularity and specialisation in Superfamily 1 and 2 helicases. *J. Bacteriol.* **184,** 1819–1826 (2002).
27. Aravind, L., Makarova, K. S. & Koonin, E. V. Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.* **28,** 3417–3432 (2000).
28. Kovall, R. & Matthews, B. W. Toroidal structure of λ-exonuclease. *Science* **277,** 1824–1827 (1997).
29. Yu, M., Souaya, J. & Julin, D. A. Identification of the nuclease active site in the multifunctional RecBCD enzyme by creation of a chimeric enzyme. *J. Mol. Biol.* **283,** 797–808 (1998).
30. Rosamond, J., Telander, K. M. & Linn, S. Modulation of the action of the *RecBC* enzyme of *Escherichia coli* K-12 by Ca^{2+}. *J. Biol. Chem.* **254,** 8648–8652 (1979).
31. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233,** 123–138 (1993).
32. Amundsen, S. K., Taylor, A. F. & Smith, G. R. Domain of RecC required for assembly of the regulatory RecD subunit into the *Escherichia coli* RecBCD holoenzyme. *Genetics* **161,** 483–492 (2002).
33. Chen, H. W., Ruan, B., Yu, M., Wang, J. & Julin, D. A. The RecD subunit of the RecBCD enzyme from *Escherichia coli* is a single-stranded DNA dependent ATPase. *J. Biol. Chem.* **272,** 10072–10079 (1997).
34. Kuhn, B., Abdel-Monem, M., Krell, H. & Hoffmann-Berling, H. Evidence for two mechanisms for DNA unwinding catalyzed by DNA helicases. *J. Biol. Chem.* **254,** 11343–11350 (1979).
35. Dillingham, M. S., Soultanas, P., Wiley, P., Webb, M. R. & Wigley, D. B. Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. *Proc. Natl Acad. Sci. USA* **98,** 8381–8387 (2001).
36. Bianco, P. R. & Kowalczykowski, S. C. Translocation step size and mechanism of the RecBC DNA helicase. *Nature* **405,** 368–372 (2000).
37. Korangy, F. & Julin, D. A. Efficiency of ATP hydrolysis and DNA unwinding by the RecBC enzyme from *Escherichia coli*. *Biochemistry* **33,** 9552–9560 (1994).
38. Schultz, D. W., Taylor, A. F. & Smith, G. R. *Escherichia coli* RecBC pseudorevertants lacking Chi recombinational hotspot activity. *J. Bacteriol.* **155,** 664–680 (1983).
39. Arnold, D. A., Bianco, P. R. & Kowalczykowski, S. C. The reduced levels of Chi recognition exhibited by the $RecBC^{1004}D$ enzyme reflect its recombination defect *in vivo*. *J. Biol. Chem.* **273,** 16476–16486 (1998).
40. Arnold, D. A., Handa, N., Kobayashi, I. & Kowalczykowski, S. C. A novel, 11 nucleotide variant of Chi, Chi*: One of a class of sequences defining the *Escherichia coli* recombination hotspot Chi. *J. Mol. Biol.* **300,** 469–479 (2000).
41. Kulkarni, A. & Julin, D. A. Specific inhibition of the *E. coli* RecBCD enzyme by Chi sequences in single-stranded oligodeoxynucleotides. *Nucleic Acids Res.* **32,** 3672–3682 (2004).
42. Palas, K. M. & Kushner, S. R. Biochemical and physical properties of exonuclease V from *Escherichia coli*. *J. Biol. Chem.* **265,** 3447–3454 (1990).
43. Churchill, J. J. & Kowalczykowski, S. C. Identification of the RecA protein-loading domain of RecBCD enzyme. *J. Mol. Biol.* **297,** 537–542 (2000).

44. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800 (1993).
45. Weeks, C. M. & Miller, R. Optimising Shake-and-bake for proteins. *Acta Crystallogr. D* **55,** 492–500 (1999).
46. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276,** 472–494 (1997).
47. Collaborative Computing Project No. 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50,** 760–763 (1994).
48. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 904–925 (1998).
49. Sanner, M. F., Spehner, J. C. & Olson, A. J. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38,** 305–320 (1996).
50. Merritt, E. A. & Bacon, D. J. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.* **277,** 505–524 (1997).

Acknowledgements We thank V. Ramakrishnan and J. Lowe for advice on the use and supply of tantalum bromide clusters, S. Halford for discussions about nucleases, C. Flensburg for advice on using SHARP and a pre-release version of the program, V. Dearing for oligonucleotide synthesis and purification, and D. Hall and E. Gordon for time and assistance on ESRF beamlines. This work was supported by Cancer Research UK (D.B.W.), NIH (S.C.K.) and a Wellcome Trust Travelling Research Fellowship (M.S.D.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.W. (dale.wigley@cancer.org.uk). The coordinates have been deposited at the Protein Databank with accession code 1W36.