

Accepted Manuscript

Probing 3'-ssDNA Loop Formation in *E. coli* RecBCD/RecBC-DNA Complexes using Non-natural DNA: A Model for "Chi" Recognition Complexes

C. Jason Wong, Rachel L. Rice, Nathan A. Baker, Tao Ju, Timothy M. Lohman

PII: S0022-2836(06)00879-5  
DOI: doi:[10.1016/j.jmb.2006.07.016](https://doi.org/10.1016/j.jmb.2006.07.016)  
Reference: YJMBI 58471

To appear in: *Journal of Molecular Biology*

Received date: 24 April 2006  
Revised date: 6 July 2006  
Accepted date: 7 July 2006



Cite this article as: , Probing 3'-ssDNA Loop Formation in *E. coli* RecBCD/RecBC-DNA Complexes using Non-natural DNA: A Model for "Chi" Recognition Complexes, *Journal of Molecular Biology* (2006), doi:[10.1016/j.jmb.2006.07.016](https://doi.org/10.1016/j.jmb.2006.07.016)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Probing 3'-ssDNA Loop Formation in *E. coli* RecBCD/RecBC-DNA Complexes using Non-natural DNA: A Model for “Chi” Recognition Complexes**

C. Jason Wong<sup>†#</sup>, Rachel L. Rice<sup>†</sup>, Nathan A. Baker<sup>†</sup>, Tao Ju<sup>‡</sup>  
and Timothy M. Lohman<sup>†\*</sup>

<sup>†</sup>Department of Biochemistry and Molecular Biophysics

Washington University School of Medicine

660 S. Euclid Ave., Box 8231

Saint Louis, MO 63110-1093, USA

and

<sup>‡</sup>Department of Computer Science and Engineering

517 Lopata Hall

Washington University

St. Louis, MO 63130, U.S.A.

Running Title: 3' Loop Formation in RecBCD-DNA Complexes

Keywords: helicase; fluorescence; motor protein; recombination; thermodynamics

\*Address correspondence to:

Department of Biochemistry and Molecular Biophysics

Washington University School of Medicine

660 S. Euclid Ave.

St. Louis, MO 63110

314-362-4393

# Current address: Sections of Microbiology and Molecular and Cell Biology

University of California, Davis,

Davis, CA 95616

ACCEPTED MANUSCRIPT

## Abstract

The equilibrium binding of *E. coli* RecBC and RecBCD helicases to duplex DNA ends containing varying lengths of polyethylene glycol (PEG) spacers within pre-formed 3'-single-stranded (ss) DNA ((dT)<sub>n</sub>) tails were studied. These studies were designed to test a previous proposal that the 3'-(dT)<sub>n</sub> tail can be looped out upon binding RecBC and RecBCD for 3'-ssDNA tails with  $n \geq 6$  nucleotides. Equilibrium binding of protein to unlabeled DNA substrates with ends containing PEG-substituted 3'-ssDNA tails was examined by competition with a Cy3-labeled reference DNA which undergoes a Cy3 fluorescence enhancement upon protein binding. We find that the binding affinities of both RecBC and RecBCD for a DNA end are unaffected upon substituting PEG for the ssDNA between the sixth and the final two nucleotides of the 3'-(dT)<sub>n</sub> tail. However, placing PEG at the end of the 3'-(dT)<sub>n</sub> tail increases the binding affinities to their maximum values (i.e. the same as binding constants for RecBC or RecBCD to a DNA end with only a 3'-(dT)<sub>6</sub> tail). Equilibrium binding studies of a RecBC mutant containing a nuclease domain deletion, RecB<sup>Δnuc</sup>C<sup>1</sup> suggest that looping of the 3'-tail (when  $n \geq 6$  nucleotides) occurs even in the absence of the RecB nuclease domain, the nuclease domain stabilizes such loop formation. Computer modeling of the RecBCD-DNA complexes suggests that the loop in the 3'-ssDNA tail may form at the RecB/RecC interface. Based on these results we suggest a model for how a loop in the 3'-ssDNA tail might form upon encounter of a "Chi" recognition sequence during unwinding of DNA by the RecBCD helicase.

## Introduction

DNA helicases are a class of motor proteins that couple the energy from nucleoside triphosphate (NTP) binding and hydrolysis to unwind double-stranded (ds) DNA to form single-stranded (ss) DNA intermediates required for DNA metabolic processes, such as replication, recombination and repair<sup>2; 3; 4; 5</sup>. Mutations in human enzymes with helicase activity result in genetic disorders such as the Werner syndrome<sup>6; 7</sup> and the Bloom syndrome<sup>8; 9</sup>.

*E. coli* RecBCD is a heterotrimeric enzyme consisting of the RecB (134 kDa), RecC (129 kDa) and RecD (67 kDa) subunits. It is a multifunctional enzyme, possessing ds and ssDNA exonuclease, ssDNA endonuclease, DNA-dependent ATPase and helicase activities and is responsible for the majority of recombination and dsDNA break repair in *E. coli*<sup>10; 11</sup>. These RecBCD-facilitated recombination and repair events are regulated by an eight-nucleotide recombination hotspot sequence (Chi site: 5'-GCTGGTGG-3'). RecBCD generates a 3'-ended ssDNA intermediate upon encountering a Chi site during the course of DNA unwinding, and facilitates the loading of RecA protein onto this ssDNA<sup>12</sup>. The RecA-bound ssDNA filament then forms a joint molecule with a homologous DNA to initiate a recombination event.

RecBCD has also been called a “bipolar DNA helicase” because both the RecB and RecD subunits possess helicase activity in isolation, but display opposite unwinding polarities; RecB is a 3' to 5' helicase<sup>13</sup> while RecD is a 5' to 3' helicase<sup>14</sup>. This suggests that in order for a RecBCD heterotrimer to unwind duplex DNA in the same net direction, the RecB subunit will translocate along the 3'-ended strand while the RecD subunit will translocate along the 5'-ended strand. This suggestion is consistent with both the chemical crosslinking studies of RecBCD-DNA complexes indicating that the

RecB subunit can be crosslinked to the 3'-ended strand while the RecC and RecD subunits can be crosslinked to the 5'-ended strand<sup>15</sup>, and the crystal structure of RecBCD bound to a DNA end<sup>16</sup>. The RecBC enzyme, lacking the RecD subunit, also functions as a helicase and is capable of facilitating homologous recombination *in vivo*<sup>17; 18</sup>. However, even though the nuclease site is located on the 30-kDa C-terminal domain of the RecB subunit<sup>19</sup>, nuclease activity is greatly attenuated in RecBC<sup>20; 21</sup>, hence the RecD subunit activates the RecB nuclease activity.

Both RecBCD and RecBC enzymes initiate unwinding from duplex DNA ends, including blunt-ended duplexes<sup>22; 23</sup>. Our previous studies on the equilibrium binding of RecBC and RecBCD helicases to duplex DNA ends containing pre-existing ss-(dT)<sub>n</sub> tails of varying lengths found that the binding affinities of both RecBC and RecBCD increase as the 3'-(dT)<sub>n</sub> tail length increases from zero to six nucleotides, but then decrease sharply as the 3'-(dT)<sub>n</sub> tail length is further increased from six to 20 nucleotides<sup>24</sup>. Isothermal titration calorimetry (ITC) experiments with RecBC show the decrease in binding affinity for 3'-(dT)<sub>n</sub> tail length with  $n \geq 6$  is due to an unfavorable entropic contribution. Thus, we proposed a 3'-tail looping model for  $n \geq 6$  nucleotides in which the helicase interacts with both the ss/dsDNA junction and somewhere on the 3'-tail between the ss/dsDNA junction and the end of the tail. Hence, the portion of the 3'-ssDNA tail located between these two sites would be looped out and may not interact with the helicase upon binding. To test this hypothesis, we substituted part of the 3'-(dT)<sub>n</sub> tail with varying lengths of polyethylene glycol (PEG) and compared the binding affinities of RecBC and RecBCD for these DNA molecules to the binding affinities for DNA molecules containing unmodified 3'-(dT)<sub>n</sub> tails of equivalent lengths. We have also

examined if the nuclease domain of the RecB subunit interacts with the 3'-ssDNA tail by studying the equilibrium binding of a RecBC mutant in which the nuclease domain of RecB has been deleted (RecB<sup>Δnuc</sup>C)<sup>1</sup> to DNA molecules containing either 3'-(dT)<sub>n</sub> tails or PEG-substituted 3'-tails.

## Results

### ***DNA substrate design***

Our 3'-tail looping model hypothesizes that the RecBC and RecBCD helicases interact with both the ss/dsDNA junction and some other region of the 3'-(dT)<sub>n</sub> tail to form a loop in the ssDNA when the pre-existing 3'-ssDNA tail has  $n \geq 6$  nucleotides. Therefore, replacement of the stretch of nucleotides between the sixth and final two nucleotides of the 3'-ssDNA tail with polyethylene glycol (PEG) should not affect the free energy of binding (binding constant) because this region of the 3'-tail is not predicted to interact with the helicase. Looping of the 3'-tail is the proposed basis for the increasingly unfavorable entropic contribution to the binding free energy change that results in the decrease in the equilibrium constants for RecBC and RecBCD binding to a DNA end containing 3'-(dT)<sub>n</sub> tail with  $n \geq 6$  nucleotides. Therefore, if looping of the 3'-tail is eliminated, the binding constant should not decrease for RecBC/RecBCD binding to DNA molecules possessing 3'-ssDNA tails with  $n \geq 6$  nucleotides. Another hypothesis of the 3'-tail looping model is that no loop can be formed at the ss/dsDNA junction because it interacts with the helicase. However, if a loop can be formed at the beginning of the 3'-(dT)<sub>n</sub> tail, then replacing the first through sixth nucleotides of the 3'-(dT)<sub>n</sub> tail with PEG should not affect the binding constant because this section of 3'-tail would be looped out thus not interact with the helicase.

We have tested these predictions by measuring the equilibrium constants for RecBC and RecBCD binding to DNA ends containing PEG at different positions within the 3'-(dT)<sub>n</sub> tail (DNA **IV** through **VI** series in Figure 1a). A previous study indicates that a hexaethylene glycol chain ((EG)<sub>6</sub>) has an average contour length equivalent to a trinucleotide of ssDNA<sup>25</sup>. We therefore compared the binding constants for a series of PEG-substituted DNA molecules to the binding constants for unmodified DNA molecules possessing the equivalent lengths of ss-(dT)<sub>n</sub> tails.

Our experiments were performed with the same 60-bp Cy3-labeled reference DNA (**I** in Figure 1a) and two series of non-fluorescent competitor DNA molecules (**II** and **III** in Figure 1a) for the reasons discussed previously<sup>24</sup>. A series of three PEG-modified non-fluorescent competitor DNA molecules (**IV**, **V** and **VI** in Figure 1a) were also studied. The DNA **IV** series molecules contain *m* repeating units of ethylene glycol (EG) (one repeating unit is defined as -(CH<sub>2</sub>CH<sub>2</sub>O)-) between the first 3'-(dT)<sub>6</sub> and the final (dT)<sub>2</sub> nucleotides at the 3'-end. The DNA **V** series molecules contain *m* repeating units of EG spacers after the first 3'-(dT)<sub>6</sub> nucleotides with no additional nucleotides at the 3'-end while the DNA **VI** series contains 12 units of EG spacers at the beginning of the 3'-tail followed by ss-(dT)<sub>n</sub>. In both the DNA **IV** and **V** molecules, a region of (dT)<sub>6</sub> was placed in front of the PEG region to maximize the interactions between RecBC or RecBCD and the ss/dsDNA junction because both RecBC and RecBCD bind optimally to DNA possessing a 3'-(dT)<sub>6</sub> tail<sup>24</sup>. The chemical structures of these modified DNA molecules are shown in Figure 1b through d.

### ***Effects of PEG-substituted 3'-ssDNA tails on RecBCD binding to DNA ends***

We have previously shown that RecBC and RecBCD exist as a stable heterodimer and heterotrimer, respectively, under the solution conditions used in our equilibrium binding studies<sup>24</sup>. Using sedimentation equilibrium methods (see Materials and Methods) we have also shown that RecB<sup>Anuc</sup>C exists as a stable heterodimer at 25°C over the concentration range from 90 to 230 nM in buffer M plus 10 mM MgCl<sub>2</sub> and 30 or 750 mM NaCl, which are the conditions used in the current study (data not shown). As in our previous studies<sup>24</sup>, we have used the fluorescence competition method<sup>26</sup> to investigate the equilibrium binding of RecBC, RecBCD and RecB<sup>Anuc</sup>C to DNA ends with PEG-substituted 3'-ssDNA tails.

As a test of the 3'-tail looping model, we measured equilibrium constants ( $K_{BCD}$ ) for RecBCD binding to the PEG-substituted DNA **IV**, **V** and **VI** series under the same solution conditions (buffer M plus 200 mM NaCl, 25°C) used in our previous study of RecBCD and compared these to  $K_{BCD}$  for RecBCD binding to DNA possessing unmodified 3'-(dT)<sub>n</sub> tails of equivalent lengths (DNA **II** series). The equilibrium constant for RecBCD binding to the end of reference DNA **I** was measured previously<sup>24</sup> ( $K_{BCD,R} = (7.5 \pm 0.4) \times 10^7 \text{ M}^{-1}$ ), and thus we obtained  $K_{BCD}$  for RecBCD binding to the PEG-substituted non-fluorescent DNA **IV** through **VI** series using a competitive binding assay as described previously<sup>24</sup>. For each non-fluorescent DNA molecule in the **IV**, **V** and **VI** series, three separate titrations were performed in which the concentration of fluorescent reference DNA **I** was maintained at 20 nM while the concentration of the non-fluorescent competitor DNA was varied in each titration. Data from all three titrations were analyzed globally by non-linear least square (NLLS) methods using

equations (11) and (13) as described in Materials and Methods to obtain  $K_{BCD}$  for binding to the end of each competitor DNA (Table 1).

Figure 2a shows the results of representative competition experiments performed with a reference DNA **I** (20 nM) and DNA **IV** molecule with  $m = 12$ . The solid lines represent simulations based on equations (11) and (13) and the best fit value of  $K_{BCD}$  (Table 1). The values of  $K_{BCD}$  for the DNA **IV** through **VI** series are plotted in Figure 2b as a function of equivalent 3'-ssDNA tail length, along with  $K_{BCD}$  (Table 1) for DNA with unmodified 3'-(dT)<sub>*n*</sub> tails (DNA **II** series) as measured previously<sup>24</sup>.

The DNA **IV** substrates have either 12 or 24 units of EG placed between the first (dT)<sub>6</sub> and final (dT)<sub>2</sub> on the 3'-tail, which is equivalent to a total 3'-ssDNA tail length of 14 and 20 nucleotides, respectively. As shown in Figure 2b and Table 1,  $K_{BCD}$  for these two DNA **IV** molecules ( $2.5 \pm 0.5 \times 10^7 \text{ M}^{-1}$  and  $0.53 \pm 0.05 \times 10^7 \text{ M}^{-1}$  for  $m = 12$  and 24, respectively) are the same as  $K_{BCD}$  for DNA molecules with regular 3'-(dT)<sub>*n*</sub> tails (DNA **II** series) of equivalent lengths ( $2.3 \pm 0.6 \times 10^7 \text{ M}^{-1}$  and  $0.51 \pm 0.03 \times 10^7 \text{ M}^{-1}$  for  $n = 15$  and 20, respectively). The DNA **V** substrates, with either six or 18 units of EG placed at the end of its 3'-(dT)<sub>6</sub> tail, have a total 3'-ssDNA tail length equivalent to nine or 15 nucleotides, respectively. As shown in Figure 2b and Table 1,  $K_{BCD}$  for these two DNA **V** molecules ( $22 \pm 3 \times 10^7 \text{ M}^{-1}$  and  $21 \pm 3 \times 10^7 \text{ M}^{-1}$  for  $m = 6$  and 18, respectively) are the same as  $K_{BCD}$  for a DNA with a 3'-(dT)<sub>6</sub> tail ( $23 \pm 3 \times 10^7 \text{ M}^{-1}$ ), which is the maximum  $K_{BCD}$  for a DNA with only a regular 3'-(dT)<sub>*n*</sub> tail.  $K_{BCD}$  for the DNA **V** series remains constant at  $22 \pm 3 \times 10^7 \text{ M}^{-1}$  as the number of units of EG increases from  $m = 6$  to  $m = 18$  while  $K_{BCD}$  for the DNA **IV** series decreases ~ five-fold as  $m$  increases from 12 to 24 (Table 1). The DNA **VI** molecules, with either (dT)<sub>2</sub> or (dT)<sub>6</sub> placed after 12 units

of EG on the 3'-tail, have a total 3'-ssDNA tail length equivalent to eight or 12 nucleotides, respectively.  $K_{BCD}$  for these two DNA **VI** molecules ( $1.8 \pm 0.4 \times 10^7 \text{ M}^{-1}$  and  $0.55 \pm 0.06 \times 10^7 \text{ M}^{-1}$  for  $n = 2$  and 6, respectively, Table 1) are lower than  $K_{BCD}$  for a blunt DNA end ( $3 \pm 0.4 \times 10^7 \text{ M}^{-1}$ ) by a factor of  $\sim 2/3$  and  $1/6$ , respectively.

### ***Effects of PEG-substituted 3'-ssDNA tails on RecBC binding to DNA ends***

As was found with RecBCD, the equilibrium constant ( $K_{BC}$ ) for RecBC binding to a DNA end also decreases sharply as the length of the 3'-(dT)<sub>n</sub> tail increases from six to 20 nucleotides<sup>24</sup>. These data also support the 3'-tail looping model and indicate that the RecD subunit is not required for such effects. To test the looping model for RecBC, we measured  $K_{BC}$  for RecBC binding to the DNA **IV** through **VI** molecules by performing three competition titrations for each non-fluorescent DNA molecule in which the concentration of reference DNA **I** was kept constant at 20 nM while the concentration of the non-fluorescent DNA molecule was varied in each titration (data not shown). Data from all three titrations were analyzed globally using equations (11) and (13) as described in Materials and Methods to obtain  $K_{BC}$  (Table 2).

The resulting values of  $K_{BC}$  for the DNA **IV** molecules with  $m = 6$  or 24, which have 3'-ssDNA tail lengths equivalent to a total of 11 or 20 nucleotides, respectively, are shown in Figure 3 and Table 2, along with  $K_{BC}$  for the DNA **II** series obtained previously<sup>24</sup>. As was observed with RecBCD, replacing the ssDNA in the middle stretch of the 3'-tail with PEG (DNA **IV** series) does not change  $K_{BC}$ , whereas placing PEG at the end of the 3'-tail (DNA **V** series) increases  $K_{BC}$  to the same value as  $K_{BC}$  for a DNA end with only a 3'-(dT)<sub>6</sub> tail (Figure 3). Substituting the first through the sixth nucleotides on the

3'-(dT)<sub>n</sub> tail (DNA **VI** series) also decreases  $K_{BC}$  to a value lower than  $K_{BC}$  for a blunt DNA end (Figure 3).

The 3'-ssDNA tail looping model suggests that the middle region of the 3'-ssDNA tail located between the sixth nucleotide and the 3'-end should not interact with RecBC or RecBCD. The fact that  $K_{BC}$  and  $K_{BCD}$  for the DNA **IV** series are the same, within experimental error, as  $K_{BC}$  and  $K_{BCD}$  for the DNA **II** series molecules containing the equivalent length of 3'-(dT)<sub>n</sub> tail is consistent with this prediction. The observed increase in  $K_{BC}$  and  $K_{BCD}$  when the final two nucleotides of the 3'-ssDNA tail are removed (DNA **V** series) agrees with our hypothesis that the 3'-end is needed for loop formation.

### ***Effects of 3'-(dT)<sub>n</sub> ssDNA tails on RecB<sup>Δnuc</sup>C binding to DNA ends***

Although the recent crystal structure of RecBCD in complex with a 19 bp hairpin duplex<sup>16</sup> indicates potential contacts between the ss/dsDNA junction and both RecB and RecC subunits, the 3'-ssDNA region is not sufficiently long (4 nucleotides) to suggest how the end of a 3'-tail that is over six nucleotides long might interact with the protein. The preferential digestion of the 3'-end of the unwound DNA by the RecBCD nuclease activity before it encounters a Chi site<sup>27</sup> indicates an interaction between the end of the 3'-tail and the nuclease domain of RecB must occur at some point during the unwinding reaction. We therefore considered the 30-kDa C-terminal nuclease domain of the RecB subunit as a possible candidate for interaction with the end of a long 3'-ssDNA tail. To examine this we studied the binding of DNA ends to a RecBC complex containing a RecB polypeptide with the 30 kDa nuclease domain deleted, referred to as RecB<sup>Δnuc</sup>C.

Equilibrium binding of RecB<sup>Δnuc</sup>C to reference DNA **I** is very tight under low monovalent salt conditions (30 mM NaCl in buffer M plus 10 mM MgCl<sub>2</sub>) and results in a ~45% enhancement in the Cy3 fluorescence signal (Figure 4a). Under these tight binding conditions, the relative fluorescence change ( $\Delta F_{\text{obs}}$  as defined in equation (9)) increases linearly with increasing [RecB<sup>Δnuc</sup>C] until a sharp breakpoint is reached at a molar ratio of two RecB<sup>Δnuc</sup>C per DNA **I** (Figure 4a), consistent with one molecule of RecB<sup>Δnuc</sup>C bound per DNA end at saturation. Importantly, the sharp breakpoint at a RecB<sup>Δnuc</sup>C/DNA ratio of two indicates that our preparation of RecB<sup>Δnuc</sup>C protein is 100% active in DNA binding.

In order to lower the binding constant of RecB<sup>Δnuc</sup>C for reference DNA **I** so that it can be measured accurately, the monovalent salt concentration was increased to 100 mM NaCl (in buffer M plus 10 mM MgCl<sub>2</sub> at 25°C). Titrations were performed at three different concentrations of DNA **I** (Figure 4b). These three titrations were analyzed using the MBDF method<sup>28; 29; 30</sup> to determine the relationship between  $\Delta F_{\text{obs}}$  and the average extent of RecB<sup>Δnuc</sup>C bound to the DNA. Figure 4c shows that  $\Delta F_{\text{obs}}$  is directly proportional to the average number of molecules of RecB<sup>Δnuc</sup>C bound per DNA **I** molecule ( $[\text{RecB}^{\Delta\text{nuc}}\text{C}]_{\text{bound}}/[\text{DNA}]_{\text{total}}$ ) and extrapolates to a value of 2 at  $\Delta F_{\text{obs}} = 0.45$ , indicating a stoichiometry of two molecules of RecB<sup>Δnuc</sup>C per DNA molecule at saturation. Global NLLS analysis of the three titration curves using equation (12) as described in Materials and Methods yields an equilibrium binding constant,  $K_{\Delta\text{nuc,R}} = (4.5 \pm 0.5) \times 10^7 \text{ M}^{-1}$  for RecB<sup>Δnuc</sup>C binding to each end of the reference DNA **I**, which is the same, within experimental error, as the binding constant of RecBC for DNA **I** ( $K_{\text{BC,R}} = (4.8 \pm 0.3) \times 10^7 \text{ M}^{-1}$ )<sup>24</sup>. The fact that the same equilibrium constant describes all three

titrations very well and that the Cy3 fluorescence enhancement is the same for RecB<sup>Δnuc</sup>C binding to each DNA end indicates that each RecB<sup>Δnuc</sup>C molecule binds independently to each DNA end and the binding affinity of RecB<sup>Δnuc</sup>C for each DNA end is identical within experimental uncertainties. These results also indicate that the RecB<sup>Δnuc</sup>C enzyme does not undergo any change in its assembly state either when free or upon binding DNA over the protein concentration range examined.

Competition experiments were performed to obtain the equilibrium constant ( $K_{\Delta\text{nuc}}$ ) for RecB<sup>Δnuc</sup>C binding to the ends of the series of non-fluorescent DNA **II** molecules as described above for RecBCD and RecBC (data not shown) using DNA **I** as the reference DNA molecule. The resulting values of  $K_{\Delta\text{nuc}}$  are presented in Table 3 and plotted in Figure 5a as a function of 3'-(dT)<sub>n</sub> tail length, along with  $K_{\text{BC}}$  for the same DNA **II** series (Table 2). As shown in Figure 5a, when  $n \leq 6$  nucleotides,  $K_{\Delta\text{nuc}}$  is the same, within experimental uncertainties, as  $K_{\text{BC}}$ . Both  $K_{\Delta\text{nuc}}$  and  $K_{\text{BC}}$  increase by a factor of ~20 as the length of the 3'-tail increases from zero to six nucleotides. However, although  $K_{\Delta\text{nuc}}$  decreases as the 3'-tail length increases beyond six nucleotides,  $K_{\Delta\text{nuc}}$  decreases only by a factor of ~4, as opposed to the ~20-fold decrease exhibited by  $K_{\text{BC}}$ . Hence, removal of the nuclease domain does influence the binding constant for  $n \geq 6$  nucleotides although a decrease is still observed.

To further understand the cause for the different tail length dependences exhibited by  $K_{\Delta\text{nuc}}$  and  $K_{\text{BC}}$  when  $n > 6$  nucleotides, we measured  $K_{\Delta\text{nuc}}$  for the PEG-substituted DNA **IV** and **V** series. Using the same competition methods described above and DNA **I** as the reference molecule, we measured  $K_{\Delta\text{nuc}}$  for DNA **IV** molecules with  $m = 12$  and 24, which should have 3'-tail lengths equivalent to a total of 14 and 20 nucleotides,

respectively. We also measured  $K_{\Delta\text{nuc}}$  for the DNA **V** series molecules with  $m = 6$  and 18, which should have 3'-tail lengths equivalent to a total of nine and 15 nucleotides, respectively. The values of  $K_{\Delta\text{nuc}}$  are presented in Table 3 and plotted in Figure 5b as a function of equivalent 3'-tail length.

As shown in Figure 5b,  $K_{\Delta\text{nuc}}$  for the DNA **IV** series ( $21 \pm 2 \times 10^7 \text{ M}^{-1}$  and  $9 \pm 2 \times 10^7 \text{ M}^{-1}$  for  $m = 12$  and 24, respectively), which has PEG in the middle of its 3'-tail, are the same as those for the DNA **II** series containing an unmodified 3'-(dT)<sub>n</sub> tail of equivalent length ( $19 \pm 2 \times 10^7 \text{ M}^{-1}$  and  $10 \pm 2 \times 10^7 \text{ M}^{-1}$  for  $n = 15$  and 20, respectively). This result was the same as that observed for RecBCD (Figure 2b) and RecBC (Figure 3).  $K_{\Delta\text{nuc}}$  for RecB<sup>Δnuc</sup>C binding to a DNA end with PEG placed at the end of the 3'-tail (DNA **V** series) ( $38 \pm 3 \times 10^7 \text{ M}^{-1}$  and  $38 \pm 4 \times 10^7 \text{ M}^{-1}$  for  $m = 6$  and 18, respectively) has the same value as  $K_{\Delta\text{nuc}}$  for a DNA **II** molecule with only a 3'-(dT)<sub>6</sub> tail ( $38 \pm 3 \times 10^7 \text{ M}^{-1}$ ) (Figure 5b), the same behavior as was observed for RecBCD (Figure 2b) and RecBC (Figure 3). Figure 5b also shows that  $K_{\Delta\text{nuc}}$  for binding to a DNA **V** molecule with  $m = 6$  is the same as  $K_{\Delta\text{nuc}}$  for binding to a DNA **V** molecule with  $m = 18$ . This behavior is very different from the decrease observed in  $K_{\Delta\text{nuc}}$  when different lengths of PEG spacers are located in the interior of the 3'-ssDNA tail as in the DNA **IV** series molecules. These data suggest that loops can still form within a pre-existing 3'-(dT)<sub>n</sub> tail with  $n > 6$  nucleotides upon binding RecB<sup>Δnuc</sup>C to a DNA end, although the loop probability appears to be affected.

### ***Effects of 5'-(dT)<sub>n</sub> ssDNA tails on RecB<sup>Δnuc</sup>C binding to DNA ends***

We also examined if removal of the nuclease domain influences the effects of the 5'-(dT)<sub>n</sub> tail length on the energetics of RecB<sup>Δnuc</sup>C binding to a DNA end by measuring  $K_{\Delta\text{nuc}}$  for binding to the non-fluorescent DNA **III** series containing pre-existing 5'-(dT)<sub>n</sub> tails with  $n$  varying from zero to 20. The same competition approach, using DNA **I** as the reference DNA, yields the values of  $K_{\Delta\text{nuc}}$  presented in Table 4 and plotted in Figure 6. Values of  $K_{\text{BC}}$  for binding to the same series of DNA molecules measured under the same solution conditions are also plotted in Figure 6 for comparison. These data show that  $K_{\Delta\text{nuc}}$  and  $K_{\text{BC}}$  are the same, within experimental uncertainties, for the DNA **III** series. Hence removal of the nuclease domain does not affect the energetics of RecBC binding to the 5'-(dT)<sub>n</sub> tail.

### **Discussion**

Both RecBCD and RecBC can initiate unwinding from a blunt duplex DNA end as well as from duplex DNA ends possessing short ssDNA extensions<sup>22; 23</sup>. In our previous equilibrium studies of RecBC and RecBCD binding to duplex DNA ends in the absence of ATP, we observed that both  $K_{\text{BC}}$  and  $K_{\text{BCD}}$  increase as the length of a pre-formed 3'-(dT)<sub>n</sub> tail increases from zero to six nucleotides. However, when the length of the 3'-(dT)<sub>n</sub> tail increases beyond six nucleotides, both  $K_{\text{BC}}$  and  $K_{\text{BCD}}$  decrease sharply<sup>24</sup>. ITC studies performed with RecBC indicate that the decrease in  $K_{\text{BC}}$  for  $n > 6$  nucleotides is accompanied by an unfavorable entropic contribution. These results led us to propose that the pre-formed 3'-ssDNA tail forms a loop upon binding RecBCD or RecBC when  $n > 6$  nucleotides<sup>24</sup>. No evidence for such looping in pre-formed 5'-(dT)<sub>n</sub> tails was observed.

### **Formation of 3'-ssDNA tail loop in RecBCD and RecBC-DNA complexes**

The 3'-tail looping model<sup>24</sup> predicts that the helicase interacts with two regions of the pre-existing 3'-(dT)<sub>n</sub> tail when  $n \geq$  six nucleotides (Figure 7a). The first point of interaction is expected to occur at the ss/dsDNA junction, including the first six nucleotides in the 3'-ss-DNA tail, while the second interaction point is expected to be somewhere near the end of the 3'-(dT)<sub>n</sub> tail. Hence, the stretch of DNA located between these two points of interaction is not expected to interact with the helicase. Our observation that replacing the stretch of DNA between the sixth and the last two nucleotides of the 3'-(dT)<sub>n</sub> tail with PEG does not affect  $K_{BC}$  or  $K_{BCD}$  (Figure 2b and Figure 3) is consistent with this prediction (Figure 7b).

The 3'-tail looping model also hypothesizes that the increasingly unfavorable entropic contribution to binding upon increasing the 3'-(dT)<sub>n</sub> tail length for  $n >$  six nucleotides reflects the looping of the 3'-tail. Hence, if ssDNA looping is eliminated, then  $K_{BCD}$  and  $K_{BC}$  should remain unchanged as the length of the 3'-(dT)<sub>n</sub> tail increases from six to 20 nucleotides and should be higher than the values of  $K_{BCD}$  and  $K_{BC}$  for DNA with unmodified 3'-(dT)<sub>n</sub> tail of equivalent length. Our observations that placing PEG at the end of the 3'-tail increases both  $K_{BCD}$  and  $K_{BC}$  to their maximum values supports this proposal (Figure 2b and Figure 3). These data further suggest that the two (possibly one) nucleotides at the 3'-end of the ss-DNA tail are needed for loop formation (Figure 7c) and thus appear to be part of the second interaction site with the helicase.

Our data also suggest that the proposed ssDNA loop does not form at the ss/dsDNA junction (i.e., before the 3'-ssDNA tail enters the enzyme). Support for this conclusion comes from the fact that replacing the first through sixth nucleotides on the

3'-(dT)<sub>n</sub> tail with PEG results in a binding constant lower than that for binding to a blunt DNA end.

### **Looped vs. Non-looped structures**

Since the formation of a looped 3'-ssDNA tail should be energetically unfavorable, relative to an unlooped structure, a simple equilibrium scheme, such as that in Scheme 1 (Figure 8a) is not consistent with our results, as discussed below. In Scheme 1, B represents RecBC (or RecBCD), BD<sub>L</sub> represents the looped complex and BD<sub>N</sub> represents the non-looped complex (Figure 8a). The binding polynomial for Scheme 1 is given by equation (1),

$$P = 1 + 2(K_L + K_N)B_f + (K_L + K_N)^2 B_f^2 \quad (1)$$

where B<sub>f</sub> is the free RecBC or RecBCD concentration, K<sub>L</sub> is the equilibrium constant for formation of a looped complex, K<sub>N</sub> is the equilibrium constant for formation of a non-looped complex, and K<sub>LN</sub><sup>D</sup> describes the equilibrium between the looped and the non-looped states as defined in equation (2).

$$K_{LN}^D = \frac{[BD_N]}{[BD_L]} = \frac{K_N}{K_L} \quad (2)$$

The observed equilibrium constant for the binding of B to one DNA end (K<sub>obs</sub>) can then be expressed in terms of K<sub>L</sub> and K<sub>N</sub> as in equation (3).

$$K_{obs} = K_L + K_N \quad (3)$$

K<sub>L</sub> is expected to decrease as the length of the 3'-(dT)<sub>n</sub> tail increases from six to 20 nucleotides because the looping of a longer tail is energetically more unfavorable than that of a shorter tail. Therefore, if only looped complexes are formed when B binds a DNA end containing a 3'-(dT)<sub>n</sub> tail over six nucleotides long, then K<sub>L</sub> should equal the

values of  $K_{BC}$  (Table 2) or  $K_{BCD}$  (Table 1) for binding to the DNA **II** series with  $n > 6$  nucleotides, thus providing an upper limit for the value  $K_L$ . Since  $K_N$  is expected to be independent of the 3'-(dT)<sub>n</sub> tail length for  $n > 6$ , the value of  $K_{BC}$  (Table 2) or  $K_{BCD}$  (Table 1) for binding to the DNA **II** substrate with  $n = 6$  nucleotides provides a lower limit for  $K_N$ .

$K_L$  decreases as  $n$  increases from eight to 20 nucleotides (from  $(19 \pm 2$  to  $0.89 \pm 0.09) \times 10^7 \text{ M}^{-1}$  for RecBC binding to the DNA **II** series (Table 2)) while  $K_N$  remains unchanged ( $40 \pm 3 \times 10^7 \text{ M}^{-1}$  for RecBC binding to a DNA **II** substrate with  $n = 6$  (Table 2)). Hence, Scheme 1 predicts that  $K_{LN}^D$  should increase ~20-fold as  $n$  increases from eight to 20 nucleotides since formation of the non-looped state would be favored as the length of the 3'-ssDNA tail increases. Scheme 1 also predicts that  $K_{obs} \geq K_N$  for binding to a DNA end containing 3-(dT)<sub>n</sub> tail longer than six nucleotides. However, our results show that  $K_{BC}$  for binding to the DNA **II** series with  $n > 6$  (ranges from  $(19 \pm 2$  to  $0.89 \pm 0.09) \times 10^7 \text{ M}^{-1}$ ) (Table 2) is always lower than  $K_{BC}$  for a DNA end with 3'-(dT)<sub>6</sub> tail ( $40 \pm 3 \times 10^7 \text{ M}^{-1}$ ) (Table 2). Similar results are also observed in  $K_{BCD}$  for binding to the DNA **II** series with  $n > 6$  (Table 1). Thus Scheme 1 is not consistent with our experimental data.

However, our results can be explained by either of two models. The first is the equilibrium model shown in Scheme 2 (Figure 8b), in which the free protein exists in equilibrium between two conformations,  $B_L$  and  $B_N$ . In this scheme, a looped complex is formed when conformation  $B_L$  binds to a DNA end while a non-looped complex is formed when conformation  $B_N$  binds to a DNA end. The equilibrium between  $B_L$  and  $B_N$  is defined by the equilibrium constant  $K_{LN}$  in equation (4).

$$K_{LN} = \frac{[B_N]}{[B_L]} \quad (4)$$

The binding polynomial describing Scheme 2 is given in equation (5),

$$P = 1 + 2 \left( \frac{K_L + K_N K_{LN}}{1 + K_{LN}} \right) B_f + \left( \frac{K_L + K_N K_{LN}}{1 + K_{LN}} \right)^2 B_f^2 \quad (5)$$

where  $B_f = B_L + B_N$  (see Materials and Methods). From this we obtain the expressions for  $K_{obs}$  and  $K_{LN}^D$  given in equations (6) and (7), respectively.

$$K_{obs} = \frac{K_L + K_N K_{LN}}{1 + K_{LN}} \quad (6)$$

$$K_{LN}^D = \frac{K_{LN} K_N}{K_L} \quad (7)$$

For the case of RecBC, if  $K_N$  equals  $K_{BC}$  for binding to a DNA end with a 3'-(dT)<sub>6</sub> tail ( $\sim 4.0 \times 10^8 \text{ M}^{-1}$ ) and  $K_L$  equals  $K_{BC}$  for binding to the DNA **II** series with  $n > 6$  (ranging from  $\sim 19$  to  $1 \times 10^7 \text{ M}^{-1}$ ), then a value of 0.01 or less for  $K_{LN}$  would result in  $K_{obs}$  having a similar value to  $K_L$ . The same is true for RecBCD. Therefore, our results could be explained if the free protein exists in equilibrium between two forms,  $B_L$  and  $B_N$ , with the  $B_L$  conformation favored over the  $B_N$  conformation.

The second possibility is that the looped structure is not a true equilibrium state, but rather is a kinetically trapped state. The formation of a non-looped complex between RecBC or RecBCD and a DNA end with a long 3'-ssDNA tail would require the tail to thread through the helicase until the ss/dsDNA junction can interact with the helicase. It is therefore possible that the ssDNA tail may fail to thread its entire length, getting stuck at the RecB/RecC interface, and therefore form a looped complex. Hence, the looped

complex could be caused by the kinetic inaccessibility of the non-looped state. Our current data cannot distinguish between these last two possibilities.

### ***RecB nuclease domain affects the interaction of RecBC with the end of the 3'-ssDNA tail***

Our results indicate that the nuclease domain of RecB does not influence the energetics of the interactions of RecBC with the first six nucleotides of the 3'-tail. However, when the length of the 3'-(dT)<sub>n</sub> tail is increased beyond six nucleotides,  $K_{\Delta\text{nuc}}$  exhibits a much more gradual decrease than  $K_{\text{BC}}$  (Figure 5a). Our PEG-substituted DNA data (Figure 5b) indicate that some 3'-tail looping does occur when RecB<sup>Δnuc</sup>C binds a DNA with a 3'-(dT)<sub>n</sub> tail longer than six nucleotides. Hence, the higher value of  $K_{\Delta\text{nuc}}$  over  $K_{\text{BC}}$  for the same DNA when  $n > 6$  nucleotides can be explained in either of two ways. The first possibility is that the 3'-ssDNA loop formed in a RecB<sup>Δnuc</sup>C-DNA complex is different from the loop formed in a RecBC-DNA complex. The second possibility is that the population of RecB<sup>Δnuc</sup>C-DNA complexes has a smaller fraction of looped states than does the population of RecBC-DNA complexes. The absence of the RecB nuclease domain could facilitate the threading of ssDNA tails through the helicase or it could increase the value of  $K_{\text{LN}}$  and/or  $K_{\text{L}}$  as discussed above. In either case, these data suggest that the RecB nuclease domain is not likely the site of interaction with the end of the 3'-ss-DNA tail, although it can influence the formation of loop on the 3'-tail.

In the RecBCD-DNA crystal structure<sup>16</sup>, the RecB nuclease domain is ~70 Å from the end of the end of the four nucleotide 3'-ssDNA tail melted out by RecBCD. Based on modeling studies, we estimate that a ss-DNA tail of ~15 nucleotides would be required to reach the nuclease domain. However, the nuclease domain is connected to the

helicase domain via a long flexible linker (~70 amino acid residues), thus it is possible that the nuclease domain could be in a different position in solution that would enable it to interact with the helicase domain of RecB and/or the RecC subunit and influence the kinetics and/or equilibria for formation of the looped state.

The effects of the RecB nuclease domain on loop formation in RecBCD were not studied here. Even though  $K_{BCD}$  also decreases sharply when the 3'-tail is increased from six to 20 nucleotides, the RecB nuclease domain may influence loop formation in RecBCD differently due to the presence of RecD. Since RecBCD preferentially digests the unwound 3'-ssDNA strand before interacting with a Chi site, it is possible that the nuclease domain of RecB could stabilize loop stabilization in a RecBCD-DNA complex via direct contact with the end of the 3'-ssDNA.

### ***Is there a functional role for 3'-ssDNA tail looping?***

Our data suggest that loops in the 3'-ssDNA tails can form *in vitro* upon binding of RecBCD or RecBC to duplex DNA ends that possess pre-existing 3'-ssDNA tails longer than six nucleotides. Large ssDNA loops spanning several hundred nucleotides in length in the 3'-strand have been observed in electron microscopic (EM) studies of DNA unwinding by RecBCD but not RecBC<sup>31; 32; 33</sup>. The current interpretation is that these loops result from the fact that RecBCD is composed of two DNA motors that translocate on opposite strands of the DNA duplex, but in the same net direction because RecB is a 3' to 5' translocase while RecD is a 5' to 3' translocase. Hence, if RecB is a slower helicase than RecD, then this would result in formation of a loop within the unwound 3'-ssDNA tail<sup>33</sup>. Our results suggest a possible alternative explanation for the formation of such loops in that the nucleation of a loop on the 3'-strand may be inherent to how DNA

binds the RecB subunit in the initiation complex. However, since such loops are not observed in studies of RecBC-catalyzed DNA unwinding<sup>33</sup> we do not favor this hypothesis.

Alternatively, the loops suggested by our binding studies may be related to loops that have been hypothesized to form in the 3'-ssDNA tail after RecBCD has encountered a "Chi" site. Singleton et al.<sup>16</sup> and Spies et al.<sup>34</sup> have suggested that a loop might form within the unwound 3'-ssDNA upon interaction with a Chi site. Once the "Chi" site becomes bound to the RecC subunit (amino acid residues 647-663 within RecC<sup>35</sup>), then subsequent DNA unwinding by RecB would produce a loop in the unwound 3'-ssDNA tail. Although no "Chi" sequences exist within the DNA substrates that we have examined, it is possible that 3'-ssDNA loops that are suggested from our data follow a similar path within the enzyme. We suggest that a DNA end possessing a pre-existing 3' single strand that is greater than six nucleotides is unable to enter the ss-DNA binding site within the RecC subunit (see below). As such, a loop will form in the 3'-ssDNA tail and this loop will increase for DNA ends possessing longer 3'-ssDNA tails.

### ***Models for RecBCD complexes with duplex DNA ends possessing long ssDNA tails and a looped "Chi" recognition complex***

The results presented here and previously<sup>15; 24; 36</sup> as well as the RecBCD-DNA crystal structure<sup>16</sup> indicate that the RecBCD and RecBC enzymes interact with both the duplex DNA and the first six nucleotides of the 3'-ssDNA at the ds/ssDNA junction. Based on this, we hypothesize that in order for a loop in the 3'-ssDNA to form, the 3'-ssDNA tail must exit the enzyme somewhere after the first six nucleotides and then re-enter the enzyme to form the second set of contacts with the enzyme. We therefore

examined the crystal structure of the RecBCD-DNA duplex in an attempt to identify potential pathways by which a loop in the 3'-ssDNA tail might form within a RecBC or RecBCD complex.

Starting with the RecBCD-DNA structure<sup>16</sup>, we used computational molecular modeling approaches (see Materials and Methods) to search for “pathways” within the RecBCD structure that would accommodate ssDNA extensions from the ends of the four nucleotide 3'-ssDNA tail and the 5'-ssDNA tail that are observed in the crystal structure (see Materials and Methods). We identified approximately 25 such paths; however, only 13 of these started at the DNA fork, extended through the protein, and ended at the surface of the protein. Six of these 13 paths were selected for further examination (see Figure 9a) based on satisfying one of the following three criteria: (1)- paths that were previously identified in the crystal structure<sup>16</sup>; (2)- paths that are consistent with our experimental binding data<sup>24</sup>; or (3)- paths that might allow formation of a bulge loop, defined as a path in which the ssDNA exits and re-enters the protein along two paths that are within 5 Å of each other. For these six paths, we then extended both the four-nucleotide 3'-ssDNA and the four-nucleotide 5'-ssDNA tails observed in the crystal structure by sequentially adding one deoxythymidylate at a time to the ends of each ssDNA tail (see Materials and Methods).

*Extension of the 5'-ssDNA-* Using the above criteria only one path extended from the end of the 5'-strand. Based on our DNA binding data, RecBC shows optimal binding to a duplex DNA end possessing a six-nucleotide 5'-ssDNA tail, whereas RecBCD shows optimal binding to a duplex DNA end possessing a ten-nucleotide 5'-ssDNA tail<sup>24</sup>. The existing four-nucleotide 5'-ssDNA tail in the crystal structure (shown in orange in Figure

9b) was therefore extended (as described above) by an additional six nucleotides. However, due to the fact that a significant portion of the RecD polypeptide is not observable in the RecBCD crystal structure, there were no protein constraints to guide the modeling of the 5'-ssDNA strand beyond this length. The first two nucleotides added are shown in cyan, while the last four nucleotides added are shown in white. The model in Figure 9b indicates that a six-nucleotide 5'-ssDNA tail will only reach the surface of the RecC polypeptide, but not enter into the RecD polypeptide. This is consistent with our observation that RecBC shows optimal binding to a duplex DNA end possessing a six-nucleotide 5'-ssDNA tail<sup>24</sup>. The model in Figure 9b shows that a 10-nucleotide 5'-ssDNA tail reaches, at least partially, into the DNA binding site within RecD<sup>24</sup>.

*Extension of the 3'-ssDNA tail-* Five of the six potential ssDNA paths identified above originated from the end of the 3'-ssDNA observed in the crystal structure. Path 1 (see Figure 9a) is the shortest path, requiring only 2 additional nucleotides (6 nucleotides total length from the duplex) to reach the protein surface. Paths 2 and 3 (see Figure 9a) both exit the protein at the RecB/RecC interface, requiring a total of 12 and 14 nucleotides, respectively, to reach the protein surface. Paths 4 and 5 (see Figure 9a) extend through RecB and RecC, pass the proposed Chi recognition site, and exit the protein near the nuclease domain. Path 4 requires a total of 18 nucleotides, and extends around the outside of the nuclease domain. Path 5, requiring a total of 17 nucleotides, extends through the nuclease domain to the active site (Asp 1067, Lys 1082, and Asp 1080). Paths 4 and 5 were previously identified by Singleton et al.<sup>16</sup> in their description of the RecBCD-DNA crystal structure.

Our modeling suggests that the proposed pathway for the 3'-ssDNA tail that reaches the nuclease domain requires a ssDNA tail of ~18 nucleotides (Figure 9b). Interestingly, in this model, the sixth nucleotide of the 3'-ssDNA tail just reaches the RecB/RecC interface (Figure 9b). Since our equilibrium binding data suggest that the ssDNA loop forms only when the pre-existing 3'-ssDNA tail is longer than six nucleotides, this result suggests that the ssDNA might exit the enzyme at this RecB/RecC interface in order to form the loop. Paths 2 and 3 were the only paths satisfying the constraint that the ssDNA exits and re-enters the protein at the same location (within 5 Å of each other) and thus we focused on these paths to construct a looped ssDNA. We chose path 2 as the path for ssDNA to exit the protein and path 3 as the path for ssDNA to re-enter the protein. The ssDNA modeled within path 3 was also extended in the direction towards the nuclease domain. The resulting model for the looped structure is shown in Figure 9c.

We envision the model shown in Figure 9c to represent the looped structures that form in the RecBC/RecBCD-DNA complexes formed with DNA ends containing pre-existing 3'-ssDNA tails that we infer from the binding studies reported here. That is, in these structures, the 3'-end of the ssDNA remains stuck at the RecB/RecC interface. In Figure 9d, we have extended the 3'-end of the ss-DNA to continue along path 4 until it reaches the region of the RecC subunit proposed to contain the “Chi” recognition site<sup>35</sup> (shown in magenta in Figure 9d). We suggest that the complex in Figure 9d would represent the looped structure formed after recognition of a “Chi” sequence during DNA unwinding, whereas Figure 9b would represent the situation during DNA unwinding before a “Chi” site was recognized. Once the “Chi” site is recognized, the 3'-ssDNA tail

would become bound tightly to the RecC subunit at the “Chi” recognition site, preventing the 3'-ssDNA from reaching the nuclease domain of RecB. However, since DNA unwinding would continue, the unwound ss-DNA would need to exit the enzyme, thus forming the proposed loop. As suggested by Spies and Kowalczykowski<sup>34</sup>, the nuclease domain of RecB would then interact directly with RecA protein and facilitate its loading on to the 3'-ssDNA tail via the looped region. This RecA filament will form which subsequently invades homologous duplex DNA to initiate recombination.

## Materials and Methods

### Buffers

Buffers were made from reagent grade chemicals using double-distilled water that was further deionized with a Milli-Q purification system (Millipore Corp., Bedford, MA). Buffer C contains 20 mM potassium phosphate (pH 6.8), 0.1 mM 2-mercaptoethanol (2-ME), 0.1 mM EDTA, 10% (v/v) glycerol. Buffer M contains 20 mM MOPS-KOH (pH 7.0), 1 mM 2-ME, 5% (v/v) glycerol. The concentration of MgCl<sub>2</sub> stocks was determined by measuring the refractive index of a stock solution in water using a Mark II refractometer (Leica Inc., Buffalo, NY) and a standard table relating refractive index to [MgCl<sub>2</sub>]<sup>37</sup>.

### Proteins

*E. coli* RecB and RecC proteins were purified and reconstituted to form RecBC as described<sup>38</sup>. RecBCD protein was purified as described elsewhere<sup>39</sup>. RecBC and RecBCD concentrations were determined spectrophotometrically in buffer C using extinction coefficients of  $\epsilon_{280} = 3.9 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$ <sup>38</sup> and  $\epsilon_{280} = 4.5 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$ <sup>39</sup>, respectively. RecB<sup>Anuc</sup>C was reconstituted from RecB<sup>Anuc</sup> and RecC proteins by mixing

equimolar concentrations of purified RecB<sup>Δnuc</sup> and RecC proteins in buffer C and incubated on ice for 10 minutes. RecB<sup>Δnuc</sup> was expressed in *E. coli* strain V186<sup>40</sup> carrying pNM52 (*lacI*<sup>q</sup>)<sup>41</sup> and pMY100 (*recB*<sup>Δnuc+</sup>)<sup>1</sup> (pMY100 was a gift from Dr. Douglas Julin, University of Maryland, MD). The cells were grown, lysed and processed through to the ammonium sulfate precipitation step as described<sup>42</sup>, followed by chromatography purification that is the same as the RecB purification<sup>38</sup>. RecB<sup>Δnuc</sup> has an extinction coefficient of  $\epsilon_{280} = 1.4 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$  in buffer C, which was determined by comparing the absorbance spectra of aliquots of RecB<sup>Δnuc</sup> in buffer C to spectra taken in 6 M guanidinium-HCl. The extinction coefficient of RecB<sup>Δnuc</sup> in 6 M guanidinium-HCl at 280 nm was calculated from amino acid sequence as described<sup>43</sup>. Similar measurements performed with a 1:1 mixture of RecB<sup>Δnuc</sup> and RecC in buffer C yielded an extinction coefficient of  $\epsilon_{280} = 3.4 \times 10^5 \text{ M}^{-1} \text{ cm}^{-1}$  for RecB<sup>Δnuc</sup>C in buffer C. All protein concentrations reported refer to the RecBC or RecB<sup>Δnuc</sup>C heterodimer or RecBCD heterotrimer. Bovine serum albumin (BSA) was from Roche (Indianapolis, IN) and its concentration was determined using an extinction coefficient of  $\epsilon_{280} = 4.3 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$  in buffer C<sup>24</sup>. All proteins were dialyzed into the particular reaction buffer before use. Dialyzed RecBC, RecB<sup>Δnuc</sup>C and RecBCD were stored at 4°C for up to five days, since a loss of activity (~15%) was observed after five days at 4°C.

### **Oligodeoxynucleotides**

Oligodeoxynucleotides were synthesized using an ABI model 391 synthesizer (Applied Biosystems, Foster City, CA) using reagents and phosphoramidites from Glen Research (Sterling, VA). A first purification step of each single-stranded oligodeoxynucleotide was performed using polyacrylamide gel electrophoresis under

denaturing conditions followed by removal of the DNA from the gel by electroelution<sup>44</sup>. The resulting oligodeoxynucleotides were then further purified chromatographically by reverse phase HPLC using an XTerra MS C18 column (Waters, Milford, MA). The concentration of each DNA strand was determined by completely digesting the strand with phosphodiesterase I (Worthington, Lakewood, NJ) in 100 mM Tris-Cl pH 9.2, 3 mM MgCl<sub>2</sub>, at 25°C and measuring the absorbance of the resulting mixture of mononucleotides at 260 nm as described<sup>45</sup>. The extinction coefficients at 260 nm used in this analysis are: 15340 M<sup>-1</sup> cm<sup>-1</sup> for AMP, 7600 M<sup>-1</sup> cm<sup>-1</sup> for CMP, 12160 M<sup>-1</sup> cm<sup>-1</sup> for GMP, 8700 M<sup>-1</sup> cm<sup>-1</sup> for TMP<sup>46</sup> and 5000 M<sup>-1</sup> cm<sup>-1</sup> for Cy3 (Glenn Research). Duplex DNA substrates were prepared by mixing equimolar concentrations (usually 3 μM) of the appropriate DNA strands in reaction buffer, which was subsequently heated to 90°C for five minutes followed by slow cooling to 25°C. Reference DNA **I** (Figure 1a) was formed from strands **1** and **2** (Figure 1e); competitor DNA **II** was formed from strands **3** and **4**; competitor DNA series **III** was formed from strands **5** and **6**; competitor DNA series **IV** was formed from strands **7** and **8**; competitor DNA series **V** was formed from strands **9** and **10**; and competitor DNA series **VI** was formed from strands **11** and **12**. The sequences of the oligodeoxynucleotides used in this study are given in Figure 1e.

### **Analytical sedimentation equilibrium**

Analytical sedimentation equilibrium experiments with RecB<sup>Δnuc</sup>C were performed in an Optima XL-A analytical ultracentrifuge using an An50Ti rotor and Epon charcoal-filled six-channel centerpieces (Beckman Instruments, Fullerton, CA) in the same way as for RecBCD and RecBC<sup>24</sup>. Sedimentation equilibrium was achieved within 24 hours. Data were edited using WinREEDIT (<http://www.biotech.uconn.edu/auf/>) to

extract the data between the sample meniscus and the bottom of the sample cell. The edited data were then analyzed by NLLS methods using WinNONLIN (<http://www.biotech.uconn.edu/auf>). The data were fit to equation (8), which describes the behavior for sedimentation of a single ideal species:

$$A_T = A_{\text{ref}} \exp \left( \frac{M(1-\bar{v}\rho)\omega^2(r^2 - r_{\text{ref}}^2)}{2RT} \right) + b \quad (8)$$

where  $A_T$  is the total absorbance at radial position  $r$ ,  $A_{\text{ref}}$  is the absorbance of the single species at reference radial position  $r_{\text{ref}}$ ,  $M$  is the molecular mass of the single species,  $\bar{v}$  is the partial specific volume calculated from protein sequences<sup>47</sup> (0.7341 ml g<sup>-1</sup> at 25°C for RecB<sup>Δnuc</sup>C),  $\rho$  is the solvent density,  $\omega$  is the angular velocity,  $R$  is the gas constant,  $T$  is the absolute temperature and  $b$  is the baseline offset. Sedimentation equilibrium experiments of 1:1 molar mixtures of RecB<sup>Δnuc</sup> and RecC (with [RecB<sup>Δnuc</sup>C] ranging from 90 to 230 nM) were performed in buffer M plus 10 mM MgCl<sub>2</sub>, 30 mM NaCl (or 750 mM NaCl) at 25°C and indicated the presence of only a single species with  $M = (242 \pm 11)$  kDa, consistent with the presence of a stable 1:1 RecB<sup>Δnuc</sup>C complex under these solution conditions.

### **Fluorescence Titrations**

Fluorescence titrations were performed as described previously<sup>24</sup> using a PTI QM-4 fluorometer (Photon Technology International, Lawrenceville, NJ) equipped with a 75 W Xe lamp. All slit widths were set at 0.5 mm. The temperature of sample in the 10-mm pathlength Type 3 quartz fluorometer cuvette (3mL) (NSG Precision Cells Inc., Farmingdale, NY) was controlled using a Lauda RM6 recirculation water bath (Brinkmann, Westbury, NY). Stirring was maintained throughout each experiment using

a P-73 cylindrical cell stirrer with a diameter of 8 mm (NSG Precision Cells Inc). The corrected Cy3 fluorescence intensity ( $F_{i,\text{corr}}$ ) after the  $i$ -th addition of protein and the initial corrected Cy3 fluorescence of the reference DNA ( $F_{0,\text{corr}}$ ) were obtained as described previously<sup>24</sup> and they are related to the observed relative fluorescence change ( $\Delta F_{\text{obs}}$ ) as in equation (9):

$$\Delta F_{\text{obs}} = \frac{F_{i,\text{corr}} - F_{0,\text{corr}}}{F_{0,\text{corr}}} \quad (9)$$

$\Delta F_{\text{obs}}$  reaches its maximum value ( $\Delta F_{\text{max}}$ ) when both ends of the reference DNA are bound with protein. Hence  $\Delta F_{\text{obs}}/\Delta F_{\text{max}}$  ( $0 \leq \Delta F_{\text{obs}}/\Delta F_{\text{max}} \leq 1$ ) equals the average number of protein molecules bound per DNA end, and thus the average number of protein molecules bound per DNA molecule is given by ( $2\Delta F_{\text{obs}}/\Delta F_{\text{max}}$ ) (see equation (11))

### ***Equilibrium binding of RecB<sup>Anuc</sup>C to Cy3-labeled reference DNA***

Equilibrium titrations of RecB<sup>Anuc</sup>C binding to a Cy3-labeled reference DNA were analyzed using a model used previously to describe RecBC and RecBCD binding to the ends of reference DNA<sup>24</sup>. In this model, RecB<sup>Anuc</sup>C, RecBC or RecBCD (hereafter collectively referred to as B) binds to each end of reference DNA (D) with the same binding constant,  $K_R$ , because the reference DNA has nearly identical ends. The binding polynomial for this model, which has two independent and identical binding sites, is given in equation (10):

$$P = 1 + 2K_R B_f + K_R^2 B_f^2 \quad (10)$$

where  $B_f$  is the free concentration of protein.

The average number of protein molecules bound per DNA molecule is given by equation (11):

$$\frac{B_{\text{bound}}}{D_T} = \frac{2K_R B_f}{1+K_R B_f} = 2 \frac{\Delta F_{\text{obs}}}{\Delta F_{\text{max}}} \quad (11)$$

where  $B_{\text{bound}} = ([DB] + 2[B_2D])$ ,  $[DB]$  is the concentration of D with only one of its ends bound by B and  $[B_2D]$  is the concentration of D with both of its ends bound by B.

As derived previously<sup>24</sup>,  $\Delta F_{\text{obs}}/\Delta F_{\text{max}}$  can be expressed explicitly in terms of total protein concentration ( $B_T$ ), total reference DNA concentration ( $D_T$ ) and  $K_R$  as in equation (12):

$$\frac{\Delta F_{\text{obs}}}{\Delta F_{\text{max}}} = \frac{1 + K_R (B_T + 2D_T) - \sqrt{4K_R B_T + (1 - K_R B_T + 2K_R D_T)^2}}{4K_R D_T} \quad (12)$$

Experimental fluorescence titrations, plotted as  $\Delta F_{\text{obs}}$  vs.  $[B_T]$ , were obtained at three different reference DNA concentrations,  $D_T$ , and analyzed by global NLLS analysis using equation (12) to obtain the best fit values of  $K_R$  and  $\Delta F_{\text{max}}$ .

### **Competition methods to determine equilibrium binding to non-fluorescent DNA**

Equilibrium constants for RecBCD, RecBC and RecB<sup>Δnuc</sup>C binding to non-fluorescent DNA molecules (N) were obtained from analysis of competition binding studies<sup>26</sup>. The analysis of competition titration data has been described in detail previously<sup>24</sup> and the same analysis is used here in this study. Briefly, three separate titration experiments were performed at three different non-fluorescent competitor DNA concentrations ( $N_1$ ,  $N_2$  and  $N_3$ ). In each titration experiment, a constant concentration of competitor DNA ( $N_1$ ,  $N_2$  or  $N_3$ ) was added to a cuvette containing a Cy3 labeled reference DNA at 20 nM and then titrated with protein. Since the competitor DNA molecules used in this study has nearly identical ends (DNA **II** through **VI** series in Figure 1a), B should bind to both ends of N with binding constant,  $K_N$ . Then, the total

and free protein concentration ( $B_T$  and  $B_f$ , respectively) can be related to the total non-fluorescent competitor DNA concentration ( $N_T$ ),  $K_N$ ,  $D_T$  and  $K_R$  as shown in equation (13):

$$B_T = B_f \left( 1 + 2 \left( \frac{K_N N_T}{(1 + K_N B_f)^2} + \frac{K_R D_T}{(1 + K_R B_f)^2} \right) + 2B_f \left( \frac{K_N^2 N_T}{(1 + K_N B_f)^2} + \frac{K_R^2 D_T}{(1 + K_R B_f)^2} \right) \right) \quad (13)$$

Data from the three titration experiments (at competitor DNA concentrations,  $N_1$ ,  $N_2$  and  $N_3$ ) were analyzed simultaneously using equations (11) and (13) and the “implicit fitting” NLLS algorithm in Scientist (Micromath, St. Louis, MO) without the need to obtain an explicit expression for  $B_f$ . In this analysis, the value of  $K_N$  was allowed to float, while  $K_R$  and  $\Delta F_{\max}$  were fixed at the values determined from the analysis of independent titrations with reference DNA in the absence of competitor. The uncertainties for the independently determined values of  $K_R$  and  $\Delta F_{\max}$  were propagated into the reported uncertainties in  $K_N$ .

In the text, we denote the site binding constants simply as  $K_{BCD}$ ,  $K_{BC}$  or  $K_{\Delta nuc}$ , referring to RecBCD, RecBC and RecB <sup>$\Delta nuc$</sup> C binding, respectively, without specifically designating the particular type of end, which should be apparent from the context of the discussion. All NLLS analyses were performed using Scientist (Micromath, St. Louis, MO) and all uncertainties are reported at the 68% confidence limit ( $\pm$  one standard deviation).

### **Binding polynomial for Scheme 2**

The binding polynomial for Scheme 2 (Figure 8b) is obtained by first expressing the total DNA concentration ( $D_T$ ) in terms of all the molecular species as given in equation (14):

$$D_T = D_f + 2[BD] + [B_2D] \quad (14)$$

where  $D_f$  is the free DNA concentration.  $[BD]$  and  $[B_2D]$  consist of both looped and non-looped complexes as given in equations (15) and (16), respectively:

$$[BD] = [B_L D] + [B_N D] \quad (15)$$

$$[B_2D] = [(B_L)_2 D] + [(B_N)_2 D] + 2[B_L D B_N] \quad (16)$$

where  $[B_L D]$  is the concentration of DNA with only one of its ends bound by  $B_L$  to form a looped state and  $[B_N D]$  is the concentration of DNA with only one of its ends bound by  $B_N$  to form a non-looped state.  $[(B_L)_2 D]$  is the concentration of DNA with both ends bound by  $B_L$ ,  $[(B_N)_2 D]$  is the concentration of DNA with both ends bound by  $B_N$ , and  $[B_L D B_N]$  is the concentration of DNA with one of its ends bound by  $B_L$  while the other end is bound by  $B_N$ . By substituting equations (15) and (16) into equation (14) and using the definitions of  $K_L$  and  $K_N$  as given in equation (17),

$$K_L = \frac{[B_L D]}{[B_L][D_f]} \quad \text{and} \quad K_N = \frac{[B_N D]}{[B_N][D_f]} \quad (17)$$

one can express  $D_T$  in terms of  $D_f$ ,  $K_L$ ,  $K_N$ ,  $B_L$  and  $B_N$  as given in equation (18):

$$D_T = D_f + 2(K_L B_L D_f + K_N B_N D_f) + (K_L^2 B_L^2 D_f + K_N^2 B_N^2 D_f + 2K_L B_L K_N B_N D_f) \quad (18)$$

Since  $K_{LN} = B_N/B_L$  (defined in equation (4)), equation (18) can be rewritten as given in equation (19):

$$D_T = D_f \left( 1 + 2B_L (K_L + K_{LN} K_N) + B_L^2 (K_L + K_{LN} K_N)^2 \right) \quad (19)$$

Since  $B_f = (B_L + B_N) = B_L(1 + K_{LN})$ , equation (19) can be expressed in terms of  $B_f$  and  $K_{LN}$  as given in equation (20):

$$D_T = D_f \left( 1 + 2 \left( \frac{K_L + K_{LN} K_N}{1 + K_{LN}} \right) B_f + \left( \frac{K_L + K_{LN} K_N}{1 + K_{LN}} \right)^2 B_f^2 \right) \quad (20)$$

Hence the binding polynomial for Scheme 2 is:

$$P = 1 + 2 \left( \frac{K_L + K_{LN} K_N}{1 + K_{LN}} \right) B_f + \left( \frac{K_L + K_{LN} K_N}{1 + K_{LN}} \right)^2 B_f^2 \quad (21)$$

### **Computational Methods**

Energy minimizations and molecular dynamics simulations were carried out using the AMBER 99 non-polarizable force field<sup>48</sup> in the AMBER 8 molecular dynamics package (<http://amber.scripps.edu>)<sup>49</sup>. The starting RecBCD-DNA structure (1W36)<sup>16</sup> was obtained from the Protein Data Bank<sup>50</sup>. The heavy atoms from Ser 1122 of RecC and Arg 607 from RecD that are missing in the structure were added using PDB2PQR (<http://pdb2pqr.sf.net>)<sup>51</sup>. No attempt was made to model any portion of the RecD subunit (69 residues) that is not observable in the crystal structure. The program PDB2PQR<sup>51</sup> was used to calculate the protonation states of titratable residues (Arg, Asp, Glu, Lys, His, Cys, Ser, and Thr) using the PROPKA method<sup>52</sup> and to add hydrogen atoms and optimize hydrogen-bonding. Unfavorable steric contacts in the crystal structure were removed by minimizing the energy of the protein using a steepest descent algorithm until convergence was reached (RMSD of 0.1 cal mol<sup>-1</sup> Å<sup>-1</sup>). Solvent influences were included via the generalized Born implicit solvent model<sup>53</sup> using a long-range electrostatic cutoff of 40 Å and a Lennard-Jones cutoff of 40 Å. The system was equilibrated by increasing the temperature from 50 K to 300 K over a 70 ps interval.

*Identifying potential paths for extending the ssDNA tails-* In the published crystal structure<sup>16</sup>, RecBCD is bound to a 43-nucleotide DNA hairpin that forms a 19-base pair

duplex with a five-nucleotide hairpin turn. In this structure, four nucleotides at the blunt end of the DNA are not involved in base pairing with their complementary base, thus forming four-nucleotide long 3' and 5' ssDNA tails at the ss/ds DNA junction. All potential paths for ssDNA that start from the 5' or 3'-ends of the ssDNA within the crystal structure were identified using an iterative thinning algorithm<sup>54</sup>. The iterative thinning process was started with a probe-accessibility map constructed from the RecBCD structure using the Shrake-Rupley algorithm<sup>55</sup> as implemented in APBS (<http://apbs.sf.net/>)<sup>56</sup> and a 0.5 Å probe. This small probe radius was chosen to provide an over-sampling of potential paths (to be filtered using the methods described below) and to compensate for the relatively low resolution (1.0 Å) of the accessibility map grid. The probe accessibility data was transformed into potential ssDNA paths using morphological thinning<sup>57; 58</sup>. The voxel set for this thinning was chosen as the union of the probe-accessible points surrounding RecBCD, as described above. The thinning procedure was carried out by iteratively removing boundary voxels from the voxel set while preserving those voxels required to maintain connected components and shape information<sup>54</sup>. This iterative thinning procedure reduced the original set of probe-accessible points into one closed, discrete surface enclosing the RecBCD structure and a collection of discrete curves extending from the surface to the interior, representing potential ssDNA paths. Finally, to remove spurious features resulted from thinning an irregular set, path branches shorter than 30 Å were pruned using morphological dilation and erosion<sup>54</sup>.

*Building the DNA chains-* Once the potential ssDNA paths were identified, the ssDNA was extended by sequentially adding thymine nucleotides to the 3' and 5'-single-

stranded ends. The nucleotides, with standard bond lengths and angles, were added using the LEaP program in the AMBER 8 molecular dynamics package<sup>49</sup>. After each nucleotide addition, the energy of the extended part of the DNA was minimized as described above with a restraint of  $10 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ , while the structures of the polypeptides and DNA present in the crystal structure were held fixed. Each round of minimization was followed by 50 ps of molecular dynamics with  $10 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$  restraints on all atoms outside a  $10 \text{ \AA}$  radius of the phosphorus atom of the new nucleotide. This short sampling was included to remove any unfavorable steric contacts resulting from the addition of the new nucleotide. A representative structure from this simulation was chosen based on the criterion that the ssDNA strand continues in the direction of the selected path. This representative structure was then used as the starting structure to add the next nucleotide.

### **Acknowledgements**

We thank T. Ho for synthesis and purification of DNA. We also thank Dr. Gerald Smith and Dr. Doug Julin for providing plasmids and cell lines and Dr. Gerald Smith for valuable discussions, and his continued interest and encouragement. This research was supported in part by NIH grant GM45948 (TML).

## References

1. Yu, M., Souaya, J. & Julin, D. A. (1998). 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 U S A* **95**, 981-6.
2. Matson, S. W., Bean, D. W. & George, J. W. (1994). DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. *Bioessays* **16**, 13-22.
3. Lohman, T. M. & Bjornson, K. P. (1996). Mechanisms of helicase-catalyzed DNA unwinding. *Annu Rev Biochem* **65**, 169-214.
4. von Hippel, P. H. & Delagoutte, E. (2001). A general model for nucleic acid helicases and their "coupling" within macromolecular machines. *Cell* **104**, 177-90.
5. Lohman, T. M., Hsieh, J., Maluf, N. K., Cheng, W., Lucius, A. L., Brendza, K. M., Korolev, S. & Waksman, G. (2003). DNA Helicases, motors that move along nucleic acids: lessons from the SF1 helicase superfamily. *The Enzymes* **23**, 303-369.
6. Gray, M. D., Shen, J. C., Kamath-Loeb, A. S., Blank, A., Sopher, B. L., Martin, G. M., Oshima, J. & Loeb, L. A. (1997). The Werner syndrome protein is a DNA helicase. *Nat Genet* **17**, 100-3.
7. Suzuki, N., Shimamoto, A., Imamura, O., Kuromitsu, J., Kitao, S., Goto, M. & Furuichi, Y. (1997). DNA helicase activity in Werner's syndrome gene product synthesized in a baculovirus system. *Nucleic Acids Res* **25**, 2973-8.

8. Watt, P. M. & Hickson, I. D. (1996). Failure to unwind causes cancer. Genome stability. *Curr Biol* **6**, 265-7.
9. van Brabant, A. J., Stan, R. & Ellis, N. A. (2000). DNA helicases, genomic instability, and human genetic disease. *Annu Rev Genomics Hum Genet* **1**, 409-59.
10. Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. & Rehrauer, W. M. (1994). Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol Rev* **58**, 401-65.
11. Smith, G. R. (2001). Homologous recombination near and far from DNA breaks: alternative roles and contrasting views. *Annu Rev Genet* **35**, 243-74.
12. Anderson, D. G. & Kowalczykowski, S. C. (1997). The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a chi-regulated manner. *Cell* **90**, 77-86.
13. Boehmer, P. E. & Emmerson, P. T. (1992). The RecB subunit of the *Escherichia coli* RecBCD enzyme couples ATP hydrolysis to DNA unwinding. *J Biol Chem* **267**, 4981-7.
14. Dillingham, M. S., Spies, M. & Kowalczykowski, S. C. (2003). RecBCD enzyme is a bipolar DNA helicase. *Nature* **423**, 893-7.
15. Ganesan, S. & Smith, G. R. (1993). Strand-specific binding to duplex DNA ends by the subunits of the *Escherichia coli* RecBCD enzyme. *J Mol Biol* **229**, 67-78.
16. Singleton, M. R., Dillingham, M. S., Gaudier, M., Kowalczykowski, S. C. & Wigley, D. B. (2004). Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432**, 187-93.

17. Chaudhury, A. M. & Smith, G. R. (1984). A new class of *Escherichia coli* recBC mutants: implications for the role of RecBC enzyme in homologous recombination. *Proc Natl Acad Sci U S A* **81**, 7850-4.
18. Thaler, D. S., Sampson, E., Siddiqi, I., Rosenberg, S. M., Thomason, L. C., Stahl, F. W. & Stahl, M. M. (1989). Recombination of bacteriophage lambda in recD mutants of *Escherichia coli*. *Genome* **31**, 53-67.
19. Yu, M., Souaya, J. & Julin, D. A. (1998). Identification of the nuclease active site in the multifunctional RecBCD enzyme by creation of a chimeric enzyme. *J Mol Biol* **283**, 797-808.
20. Palas, K. M. & Kushner, S. R. (1990). Biochemical and physical characterization of exonuclease V from *Escherichia coli*. Comparison of the catalytic activities of the RecBC and RecBCD enzymes. *J Biol Chem* **265**, 3447-54.
21. Chen, H. W., Randle, D. E., Gabbidon, M. & Julin, D. A. (1998). Functions of the ATP hydrolysis subunits (RecB and RecD) in the nuclease reactions catalyzed by the RecBCD enzyme from *Escherichia coli*. *J Mol Biol* **278**, 89-104.
22. Taylor, A. F. & Smith, G. R. (1985). Substrate specificity of the DNA unwinding activity of the RecBC enzyme of *Escherichia coli*. *J Mol Biol* **185**, 431-43.
23. Korangy, F. & Julin, D. A. (1993). Kinetics and processivity of ATP hydrolysis and DNA unwinding by the RecBC enzyme from *Escherichia coli*. *Biochemistry* **32**, 4873-80.
24. Wong, C. J., Lucius, A. L. & Lohman, T. M. (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**, 765-82.

25. Williams, D. J. & Hall, K. B. (1996). RNA hairpins with non-nucleotide spacers bind efficiently to the human U1A protein. *J Mol Biol* **257**, 265-75.
26. Jezewska, M. J. & Bujalowski, W. (1996). A general method of analysis of ligand binding to competing macromolecules using the spectroscopic signal originating from a reference macromolecule. Application to Escherichia coli replicative helicase DnaB protein nucleic acid interactions. *Biochemistry* **35**, 2117-28.
27. Anderson, D. G. & Kowalczykowski, S. C. (1997). The recombination hot spot chi is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes Dev* **11**, 571-81.
28. Lohman, T. M. & Bujalowski, W. (1988). Negative cooperativity within individual tetramers of Escherichia coli single strand binding protein is responsible for the transition between the (SSB)<sub>35</sub> and (SSB)<sub>56</sub> DNA binding modes. *Biochemistry* **27**, 2260-5.
29. Lohman, T. M. & Bujalowski, W. (1991). Thermodynamic methods for model-independent determination of equilibrium binding isotherms for protein-DNA interactions: spectroscopic approaches to monitor binding. *Methods Enzymol* **208**, 258-90.
30. Lohman, T. M. & Mascotti, D. P. (1992). Nonspecific ligand-DNA equilibrium binding parameters determined by fluorescence methods. *Methods Enzymol* **212**, 424-58.
31. Taylor, A. & Smith, G. R. (1980). Unwinding and rewinding of DNA by the RecBC enzyme. *Cell* **22**, 447-57.

32. Braedt, G. & Smith, G. R. (1989). Strand specificity of DNA unwinding by RecBCD enzyme. *Proc Natl Acad Sci U S A* **86**, 871-5.
33. Taylor, A. F. & Smith, G. R. (2003). RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature* **423**, 889-93.
34. Spies, M., Dillingham, M. S. & Kowalczykowski, S. C. (2005). Translocation by the RecB Motor Is an Absolute Requirement for  $\chi$ -Recognition and RecA Protein Loading by RecBCD Enzyme. *J Biol Chem* **280**, 37078-87.
35. Arnold, D. A., Handa, N., Kobayashi, I. & Kowalczykowski, S. C. (2000). 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**, 469-79.
36. Farah, J. A. & Smith, G. R. (1997). The RecBCD enzyme initiation complex for DNA unwinding: enzyme positioning and DNA opening. *J Mol Biol* **272**, 699-715.
37. Wolf, A. V., Brown, M. G. & Prentiss, P. G. (1974). Concentrative properties of aqueous solutions: conversion tables. In *CRC handbook of chemistry and physics : a ready-reference book of chemical and physical data* 55th edit. (Weast, R. C., ed.), pp. D-209. CRC Press, Cleveland.
38. Lucius, A. L., Wong, C. J. & Lohman, T. M. (2004). Fluorescence stopped-flow studies of single turnover kinetics of E.coli RecBCD helicase-catalyzed DNA unwinding. *J Mol Biol* **339**, 731-50.
39. Lucius, A. L., Vindigni, A., Gregorian, R., Ali, J. A., Taylor, A. F., Smith, G. R. & Lohman, T. M. (2002). DNA unwinding step-size of E. coli RecBCD helicase

- determined from single turnover chemical quenched-flow kinetic studies. *J Mol Biol* **324**, 409-28.
40. Chaudhury, A. M. & Smith, G. R. (1984). Escherichia coli recBC deletion mutants. *J Bacteriol* **160**, 788-91.
41. Gilbert, H. J., Blazek, R., Bullman, H. M. & Minton, N. P. (1986). Cloning and expression of the Erwinia chrysanthemi asparaginase gene in Escherichia coli and Erwinia carotovora. *J Gen Microbiol* **132** ( Pt 1), 151-60.
42. Boehmer, P. E. & Emmerson, P. T. (1991). Escherichia coli RecBCD enzyme: inducible overproduction and reconstitution of the ATP-dependent deoxyribonuclease from purified subunits. *Gene* **102**, 1-6.
43. Lohman, T. M., Chao, K., Green, J. M., Sage, S. & Runyon, G. T. (1989). Large-scale purification and characterization of the Escherichia coli rep gene product. *J Biol Chem* **264**, 10139-47.
44. Wong, I., Chao, K. L., Bujalowski, W. & Lohman, T. M. (1992). DNA-induced dimerization of the Escherichia coli rep helicase. Allosteric effects of single-stranded and duplex DNA. *J Biol Chem* **267**, 7596-610.
45. Holbrook, J. A., Capp, M. W., Saecker, R. M. & Record, M. T., Jr. (1999). Enthalpy and heat capacity changes for formation of an oligomeric DNA duplex: interpretation in terms of coupled processes of formation and association of single-stranded helices. *Biochemistry* **38**, 8409-22.
46. Gray, D. M., Hung, S. H. & Johnson, K. H. (1995). Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes. *Methods Enzymol* **246**, 19-34.

47. Laue, T. M., Shah, B. D., Ridgeway, T. M. & Pelletier, S. L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In *Analytical ultracentrifugation in biochemistry and polymer science* (Harding, S. E., Rowe, A. J. & Horton, J. C., eds.), pp. 90-125. Royal Society of Chemistry, Cambridge England.
48. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *J Am Chem Soc* **117**, 5179-5197.
49. Case, D. A., Darden, T.A., Cheatham, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Wang, B., Pearlman, D.A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J.W., Ross, W.S., Kollman, P.A. (2004). AMBER 8. *University of California, San Francisco*.
50. Deshpande, N., Adness, K. J., Bluhm, W. F., Merino-Ott, J. C., Townsend-Merino, W., Zhang, Q., Knezevich, C., Xie, L., Chen, L., Feng, Z., Green, R. K., Flippen-Anderson, J. L., Westbrook, J., Berman, H. M. & Bourne, P. E. (2005). The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucleic Acids Res* **33**, D233-7.
51. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* **32**, W665-7.

52. Li, H., Robertson, A. D., Jensen J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics* **61**, 704-721.
53. Onufriev, A., Bashford, D. & Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **55**, 383-94.
54. Ju, T., Baker, M.L., Chiu, W. (submitted). Computing and pruning surface skeletons of volume images with application in shape description. *Pattern Recognition*.
55. Shrake, A. & Rupley, J. A. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol* **79**, 351-71.
56. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**, 10037-41.
57. Bertrand, G. (1995). A Parallel Thinning Algorithm for Medial Surfaces. *Pattern Recognition Letters* **16**, 979-986.
58. Borgfors, G., Nystrom, I. & Di Baja, G. S. (1999). Computing skeletons in three dimensions. *Pattern Recognition* **32**, 1225-1236.

## Figure Legends

**Figure 1.** DNA molecules used in the equilibrium binding studies. (a) Schematic representations of Cy3-labeled reference DNA **I** and the non-fluorescent series of competitor DNA (DNA **II** through **VI**), each possessing a 60 bp duplex region. (b) Structure of the Cy3 fluorophore and its covalent attachment to the phosphate group on the 5'-end of reference DNA **I** via a 3-carbon linker. (c) Structure of the polyethylene glycol spacer containing  $m$  repeating units of ethylene glycol ((EG) $_m$ ) and its covalent attachment to the phosphate group within the 3'-(dT) $_n$  tail of DNA **IV** and DNA **VI**. (d) Structure of the (EG) $_m$  spacer and its covalent attachment to the phosphate group at the end of the 3'-(dT) $_n$  tail of DNA **V**. (e) Sequences of the DNA strands used to form the duplex DNA molecules shown schematically in panel (a). Reference DNA **I** was formed from strands **1** and **2**; the series of DNA **II** molecules was formed from strands **3** and **4**; the series of DNA **III** molecules was formed from strands **5** and **6**; the series of DNA **IV** molecules was formed from strands **7** and **8**; the series of DNA **V** molecules was formed from strands **9** and **10**; the series of DNA **VI** molecules was formed from strands **11** and **12**.

**Figure 2.** Effects of replacing regions of the pre-existing 3'-(dT) $_n$  tail with PEG on the equilibrium constants ( $K_{BCD}$ ) for RecBCD binding to a duplex DNA end. (a) Equilibrium competition titrations to determine  $K_{BCD}$  for RecBCD binding to the ends of a non-fluorescent DNA **IV** molecule possessing 3'-(dT) $_6$ (EG) $_m$ (dT) $_2$  tails with  $m = 12$ .

Mixtures of Cy3 labeled reference DNA **I** (20 nM) and a DNA **IV** molecule with  $m = 12$  were titrated with RecBCD in buffer M plus 200 mM NaCl at 25°C and the relative Cy3 fluorescence enhancement ( $\Delta F_{\text{obs}}$  defined in equation (11) in Materials and Methods) plotted as a function of total [RecBCD]. Three separate titrations were performed in which a constant total reference DNA **I** concentration was used but the total concentration of the competitor DNA **IV** molecule was varied in each titration; reference DNA **I** only (20 nM) (●); reference DNA **I** (20 nM) and competitor DNA **IV** ( $m = 12$ ) at 70 nM (○) and 210 nM (■). Solid lines are simulations using equations (11) and (13) and the best fit values of  $K_{\text{BCD}}$  (Table 1). (b) Values of  $K_{\text{BCD}}$  for binding to the ends of the DNA **IV** series molecules containing 3'-(dT)<sub>6</sub>(EG)<sub>*m*</sub>(dT)<sub>2</sub> (with  $m = 12$  or 24) (■); for the DNA **V** series containing 3'-(dT)<sub>6</sub>(EG)<sub>*m*</sub> (with  $m = 6$  or 18) (●); and for the DNA **VI** series containing 3'-(EG)<sub>12</sub>(dT)<sub>*n*</sub> (with  $n = 2$  or 6) (▲) (Table 1) are plotted as a function of equivalent 3'-tail length, along with  $K_{\text{BCD}}$  for the DNA **II** series molecules containing unmodified 3'-(dT)<sub>*n*</sub> tails (with  $n = 0$  to 20 nucleotides) (○) (Table 1).

**Figure 3.** Effects of replacing part of the pre-existing 3'-(dT)<sub>*n*</sub> tail with PEG on RecBC binding to a duplex DNA end. Values of  $K_{\text{BC}}$  for binding to the ends of the DNA **IV** series containing 3'-(dT)<sub>6</sub>(EG)<sub>*m*</sub>(dT)<sub>2</sub> (with  $m = 6$  or 24) (■);  $K_{\text{BC}}$  for the DNA **V** series containing 3'-(dT)<sub>6</sub>(EG)<sub>*m*</sub> (with  $m = 6$  or 18) (●); and  $K_{\text{BC}}$  for DNA **VI** series containing 3'-(EG)<sub>12</sub>(dT)<sub>*n*</sub> (with  $n = 2$  or 6) (▲) (Table 2) are plotted as a function of equivalent 3'-tail length, along with  $K_{\text{BC}}$  for the DNA **II** series containing unmodified 3'-(dT)<sub>*n*</sub> tails (○) (Table 2). Experiments were performed in buffer M, 10 mM MgCl<sub>2</sub> plus 100 mM NaCl at 25°C.

**Figure 4.** Binding of RecB<sup>Δnuc</sup>C to the fluorescent reference DNA **I**. (a) Titrations of reference DNA **I** (total concentration of 20 nM (●) and 60 nM (□)) with RecB<sup>Δnuc</sup>C under stoichiometric (high affinity) conditions in buffer M, 10 mM MgCl<sub>2</sub> plus 30 mM NaCl at 25°C. Cy3 fluorescence enhancement ( $\Delta F_{\text{obs}}$ ) is plotted as a function of the  $[\text{RecB}^{\Delta\text{nuc}}\text{C}]_{\text{total}}/[\text{DNA}]_{\text{total}}$ . Excitation wavelength was 515 nm and fluorescence emission was monitored at 563 nm. (b) Determination of the equilibrium constant for RecB<sup>Δnuc</sup>C binding to the ends of reference DNA **I**. Three separate titration experiments were performed at different total DNA **I** concentrations of 10 nM (●), 20 nM (○) and 50 nM (■) (buffer M, 10 mM MgCl<sub>2</sub> plus 100 mM NaCl at 25°C) and  $\Delta F_{\text{obs}}$  is plotted as a function of total  $[\text{RecB}^{\Delta\text{nuc}}\text{C}]$ . Global NLLS analysis of these data using equation (12) as described in Materials and Methods yields the equilibrium constant for RecB<sup>Δnuc</sup>C binding to each end of reference DNA **I**,  $K_{\Delta\text{nuc,R}} = (4.5 \pm 0.5) \times 10^7 \text{ M}^{-1}$  and  $\Delta F_{\text{max}} = 0.45 \pm 0.04$ . Solid lines are simulations using equation (12) and the best fit values of  $K_{\Delta\text{nuc,R}}$  and  $\Delta F_{\text{max}}$ . (c)  $\Delta F_{\text{obs}}$  is directly proportional to the average moles of RecB<sup>Δnuc</sup>C bound per mole of DNA **I**. The dependence of  $\Delta F_{\text{obs}}$  on  $[\text{RecB}^{\Delta\text{nuc}}\text{C}]_{\text{bound}}/[\text{DNA}]_{\text{total}}$  was determined from the titration experiments in (b) performed at 10 and 50 nM DNA **I** as described<sup>28; 29; 30</sup>. The solid line is a linear fit of the dependence of  $\Delta F_{\text{obs}}$  on  $[\text{RecB}^{\Delta\text{nuc}}\text{C}]_{\text{bound}}/[\text{DNA}]_{\text{total}}$  while the dashed line is a linear extrapolation of the solid line to  $\Delta F_{\text{max}} = 0.45$  and  $[\text{RecB}^{\Delta\text{nuc}}\text{C}]_{\text{bound}}/[\text{DNA}]_{\text{total}} = 2$ .

**Figure 5.** Influence of the nuclease domain of RecB on the dependence of the end binding constant on the length of the pre-existing 3'-(dT)<sub>n</sub> tails. All experiments were performed in buffer M plus 10 mM MgCl<sub>2</sub> and 100 mM NaCl at 25°C. (a) Comparisons of the effects of the length of the pre-existing 3'-(dT)<sub>n</sub> tails on the equilibrium constants for RecB<sup>Δnuc</sup>C ( $K_{\Delta\text{nuc}}$ ) and RecBC ( $K_{\text{BC}}$ ) binding to the ends of duplex DNA. Values of  $K_{\Delta\text{nuc}}$  (●) and  $K_{\text{BC}}$  (□) for binding to the ends of the non-fluorescent DNA **II** series (Table 3) are plotted as a function of  $n$ . (b) Effects of substituting parts of the pre-existing 3'-(dT)<sub>n</sub> tails with PEG on  $K_{\Delta\text{nuc}}$ . Values of  $K_{\Delta\text{nuc}}$  for the DNA **IV** series with 3'-(dT)<sub>6</sub>(EG)<sub>m</sub>(dT)<sub>2</sub> tails ( $m = 12$  or  $24$ ) (□) and  $K_{\Delta\text{nuc}}$  for the DNA **V** series with 3'-(dT)<sub>6</sub>(EG)<sub>m</sub> tails ( $m = 6$  or  $18$ ) (○) are plotted as a function of equivalent 3'-tail length, along with  $K_{\Delta\text{nuc}}$  for the DNA **II** series (●) (Table 3).

**Figure 6.** Deletion of the nuclease domain of RecB does not affect the influence of the 5'-ssDNA tail length on the binding of RecBC to a DNA end. Equilibrium constants ( $K_{\Delta\text{nuc}}$ ) for RecB<sup>Δnuc</sup>C binding to the ends of the DNA **III** series possessing pre-existing 5'-(dT)<sub>n</sub> tails with  $n$  varying between zero and 20 nucleotides (●) are plotted as a function of 5'-tail length ( $n$ ), along with equilibrium constants ( $K_{\text{BC}}$ ) for RecBC binding to the ends of the same DNA **III** series (□) (Table 4). All experiments were performed in buffer M, 10 mM MgCl<sub>2</sub> plus 100 mM NaCl at 25°C.

**Figure 7.** Cartoons depicting the binding of RecBCD to DNA ends possessing different types of pre-existing 3'-tails. (a) Looping of the 3'-ssDNA tail occurs when RecBCD binds to a DNA end containing a 3'-(dT)<sub>n</sub> tail with  $n > 6$  nucleotides. (b) Replacing the

section of 3'-(dT)<sub>n</sub> tail between the sixth and the last two nucleotides with PEG does not affect loop formation because this part of 3'-ssDNA tail does not interact with RecBCD.

(c) Looping is eliminated when PEG is placed at the end of the 3'-tail.

**Figure 8.** Possible equilibrium schemes for the binding of RecBC to a DNA end possessing a pre-existing 3'-(dT)<sub>n</sub> tail with  $n > \text{six}$  nucleotides. (a) In Scheme 1, the binding of RecBC (B) can result in the formation of either a looped (BD<sub>L</sub>) or non-looped (BD<sub>N</sub>) state. (b) In Scheme 2, RecBC is assumed to exist in two conformations (B<sub>L</sub> and B<sub>N</sub>) prior to binding DNA. The looped complex forms when B<sub>L</sub> binds to a DNA end possessing a pre-existing 3'-(dT)<sub>n</sub> tail with  $n > \text{six}$  nucleotides, while the binding of B<sub>N</sub> to a DNA end results in a non-looped complex. The nuclease domain (B<sup>N</sup>) and helicase domain (B<sup>H</sup>) of RecB are indicated.

**Figure 9.** Potential pathways for the 3' and 5'-ssDNA tails within the RecBCD-DNA complex and a model for the RecBCD-“Chi” recognition complex. (a) Representation of the six potential paths for the extended ssDNA tails as described in the text. (b) Proposed model for the non-looped complex in which the 3'-ssDNA tail is threaded through both the RecB and RecC subunits and exits near the RecB nuclease domain. The 4-nucleotide 3' and 5'-ssDNA tails in the RecBCD-DNA crystal structure were extended along the proposed channels within the RecBCD enzyme identified by Singleton et al.<sup>16</sup>. RecB subunit is in gold, RecC subunit is in blue, RecD subunit is in green and the original 19-bp duplex DNA used in the formation of crystal is shown in orange. Six nucleotides were added to the 5'-tail to a total length of ten nucleotides while 14 nucleotides were

added to the 3'-tail to a total length of 18 nucleotides. The fifth and sixth nucleotides of the 3' and 5'-ssDNA tails are shown in cyan, while the seventh through 18<sup>th</sup> nucleotides of the 3'-ssDNA tail and the seventh to tenth nucleotides of the 5'-ssDNA tail are shown in white. (c) Proposed model for the looped 3'-ssDNA tail structure (BD<sub>L</sub>) in Figure 8, Scheme 2) that can form upon binding RecBC or RecBCD to DNA possessing a long pre-formed 3'-ssDNA tail. The 3'-ssDNA tail follows path 2 to exit the enzyme and then re-enters the enzyme via path 3. (d) Proposed model for the 3'-ssDNA looped structure that might form after recognition of a "Chi" site. The 3'-ssDNA follows paths 2, 3 and 4 to reach the region of the RecC subunit that contains the amino acid residues (shown in magenta) that have been proposed to be involved in "Chi" recognition<sup>35</sup>.

## Tables

**Table 1.** Equilibrium constants ( $K_{BCD}$ ) for RecBCD binding to the ends of non-fluorescent DNA **II**, **IV**, **V** and **VI** series (buffer M plus 200 mM NaCl at 25°C).

3'- ssDNA tail length (nt)	$K_{BCD}$ ( $10^7 M^{-1}$ ) for DNA <b>II</b> <sup>a</sup>	3'-tail composition of DNA <b>II</b>	$K_{BCD}$ ( $10^7 M^{-1}$ ) for DNA with PEG- substituted 3'-tail ( <b>IV</b> , <b>V</b> and <b>VI</b> )	3'-tail composition of DNA with PEG- substituted 3'-tail ( <b>IV</b> , <b>V</b> and <b>VI</b> )
0	$3.0 \pm 0.4$	no tail		
2	$5.2 \pm 0.4$	3'-(dT) <sub>2</sub>		
4	$15 \pm 1$	3'-(dT) <sub>4</sub>		
6	$23 \pm 3$	3'-(dT) <sub>6</sub>		
8	$16 \pm 2$	3'-(dT) <sub>8</sub>	$1.8 \pm 0.4$	3'-(EG) <sub>12</sub> (dT) <sub>2</sub>
9	ND		$22 \pm 3$	3'-(dT) <sub>6</sub> (EG) <sub>6</sub>
10	$6.4 \pm 0.8$	3'-(dT) <sub>10</sub>		
12	ND		$0.55 \pm 0.06$	3'-(EG) <sub>12</sub> (dT) <sub>6</sub>
14	ND		$2.5 \pm 0.5$	3'-(dT) <sub>6</sub> (EG) <sub>12</sub> (dT) <sub>2</sub>
15	$2.3 \pm 0.6$	3'-(dT) <sub>15</sub>	$21 \pm 3$	3'-(dT) <sub>6</sub> (EG) <sub>18</sub>
20	$0.51 \pm 0.03$	3'-(dT) <sub>20</sub>	$0.53 \pm 0.05$	3'-(dT) <sub>6</sub> (EG) <sub>24</sub> (dT) <sub>2</sub>

<sup>a</sup>Data from <sup>24</sup>

**Table 2.** Equilibrium constants ( $K_{BC}$ ) for RecBC binding to the ends of non-fluorescent DNA **II**, **IV**, **V** and **VI** series (buffer M, 10 mM MgCl<sub>2</sub> plus 100 mM NaCl at 25°C).

3'-ssDNA tail length (nt)	$K_{BC}$ ( $10^7$ M <sup>-1</sup> ) for DNA <b>II</b> <sup>a</sup>	3'-tail composition of DNA <b>II</b>	$K_{BC}$ ( $10^7$ M <sup>-1</sup> ) for DNA with PEG-substituted 3'-tail ( <b>IV</b> and <b>V</b> )	3'-tail composition of DNA with PEG-substituted 3'-tail ( <b>IV</b> and <b>V</b> )
0	1.6 ± 0.3	no tail		
2	4.3 ± 0.4	3'-(dT) <sub>2</sub>		
4	19 ± 2	3'-(dT) <sub>4</sub>		
6	40 ± 3	3'-(dT) <sub>6</sub>		
8	19 ± 3	3'-(dT) <sub>8</sub>	0.77 ± 0.08	3'-(EG) <sub>12</sub> (dT) <sub>2</sub>
9	ND		39 ± 3	3'-(dT) <sub>6</sub> (EG) <sub>6</sub>
10	8.8 ± 0.9	3'-(dT) <sub>10</sub>		
11	ND		8.8 ± 0.9	3'-(dT) <sub>6</sub> (EG) <sub>6</sub> (dT) <sub>2</sub>
12	ND		0.31 ± 0.06	3'-(EG) <sub>12</sub> (dT) <sub>6</sub>
15	ND		36 ± 3	3'-(dT) <sub>6</sub> (EG) <sub>18</sub>
20	0.89 ± 0.09	3'-(dT) <sub>20</sub>	0.87 ± 0.07	3'-(dT) <sub>6</sub> (EG) <sub>24</sub> (dT) <sub>2</sub>

<sup>a</sup>Data from <sup>24</sup>

**Table 3.** Equilibrium constants ( $K_{\Delta\text{nuc}}$ ) for RecB <sup>$\Delta\text{nuc}$</sup> C binding to the ends of non-fluorescent DNA **II**, **IV** and **V** series (buffer M plus 10 mM MgCl<sub>2</sub> and 100 mM NaCl at 25°C).

3'-ssDNA tail length (nt)	$K_{\Delta\text{nuc}}$ ( $10^7 \text{ M}^{-1}$ ) for DNA <b>II</b>	3'-tail composition of DNA <b>II</b>	$K_{\Delta\text{nuc}}$ ( $10^7 \text{ M}^{-1}$ ) for DNA with PEG-substituted 3'-tail ( <b>IV</b> and <b>V</b> )	3'-tail composition of PEG-substituted DNA ( <b>IV</b> and <b>V</b> )
0	$1.5 \pm 0.2$	no tail		
2	$3.9 \pm 0.3$	3'-(dT) <sub>2</sub>		
4	$15 \pm 2$	3'-(dT) <sub>4</sub>		
6	$37 \pm 3$	3'-(dT) <sub>6</sub>		
8	$38 \pm 3$	3'-(dT) <sub>8</sub>		
9	ND		$38 \pm 3$	3'-(dT) <sub>6</sub> (EG) <sub>6</sub>
10	$34 \pm 3$	3'-(dT) <sub>10</sub>		
14	ND		$21 \pm 2$	3'-(dT) <sub>6</sub> (EG) <sub>12</sub> (dT) <sub>2</sub>
15	$19 \pm 2$	3'-(dT) <sub>15</sub>	$38 \pm 4$	3'-(dT) <sub>6</sub> (EG) <sub>18</sub>
20	$10 \pm 2$	3'-(dT) <sub>20</sub>	$9 \pm 2$	3'-(dT) <sub>6</sub> (EG) <sub>24</sub> (dT) <sub>2</sub>

**Table 4.** Equilibrium constants for RecB<sup>Δnuc</sup>C ( $K_{\Delta\text{nuc}}$ ) and RecBC ( $K_{\text{BC}}$ ) binding to the ends of non-fluorescent DNA **III** series (buffer M plus 10 mM MgCl<sub>2</sub> and 100 mM NaCl at 25°C).

5'-(dT) <sub>n</sub> tail length ( <i>n</i> )	$K_{\text{BC}}$ ( $10^7 \text{ M}^{-1}$ ) for DNA <b>III</b>	$K_{\Delta\text{nuc}}$ ( $10^7 \text{ M}^{-1}$ ) for DNA <b>III</b>
0	$1.6 \pm 0.3$	$1.6 \pm 0.3$
2	$2.2 \pm 0.4$	ND
4	$3.8 \pm 0.4$	$4.1 \pm 0.5$
6	$4.8 \pm 0.5$	$5.0 \pm 0.5$
8	$4.7 \pm 0.4$	$5.1 \pm 0.5$
10	$5.0 \pm 0.5$	$4.6 \pm 0.4$
20	$4.7 \pm 0.5$	$4.8 \pm 0.5$

Figure 1

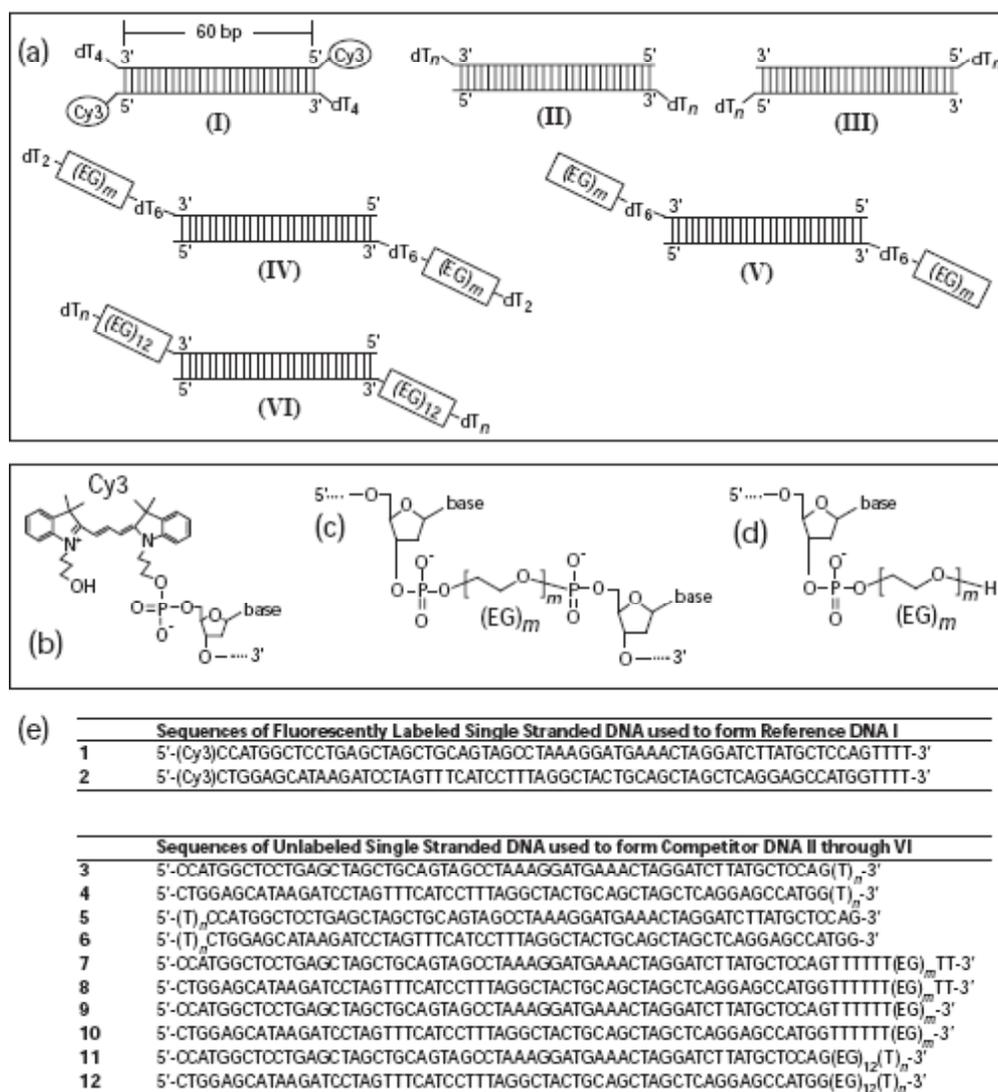


Figure 1

Figure 2

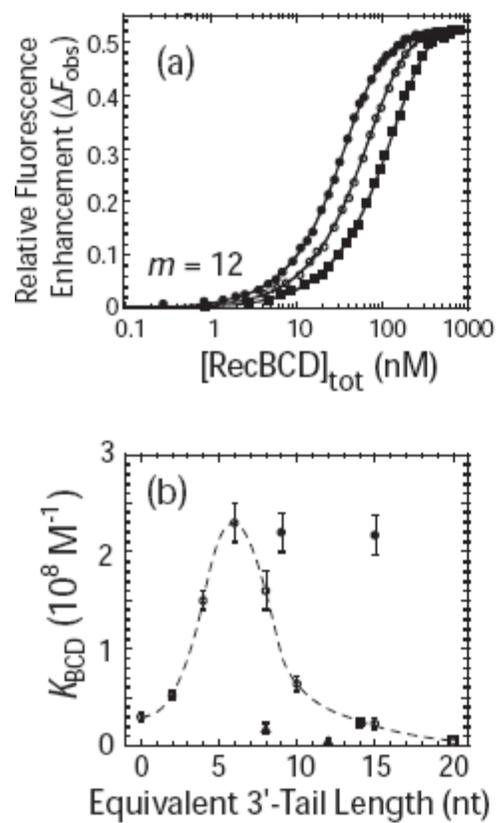


Figure 2

Figure 3

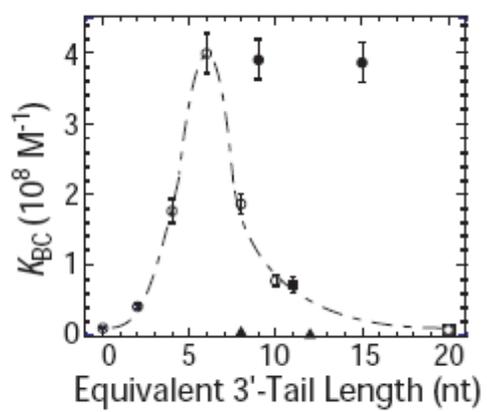


Figure 3

Figure 4

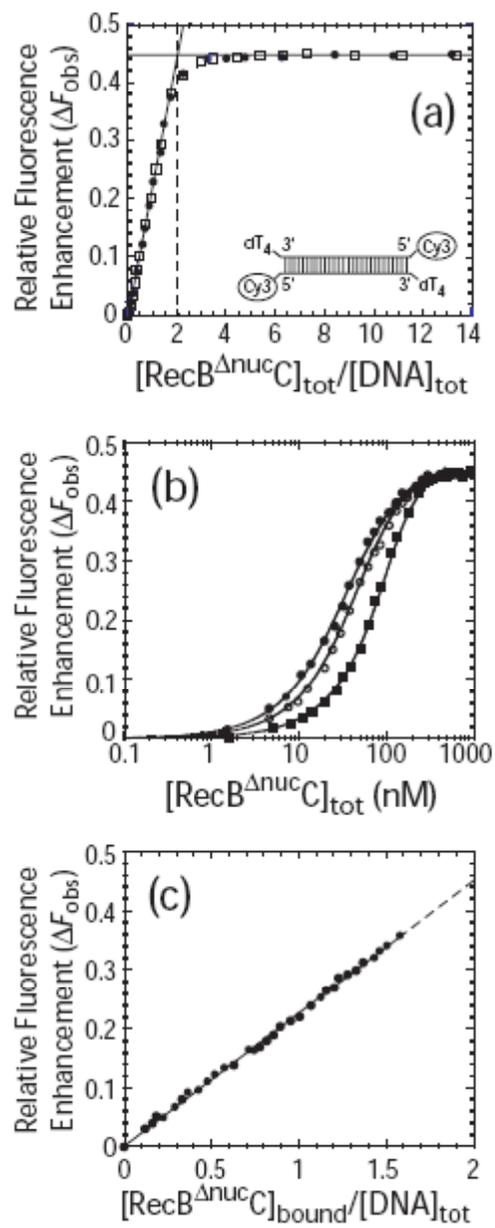


Figure 4

Figure 5

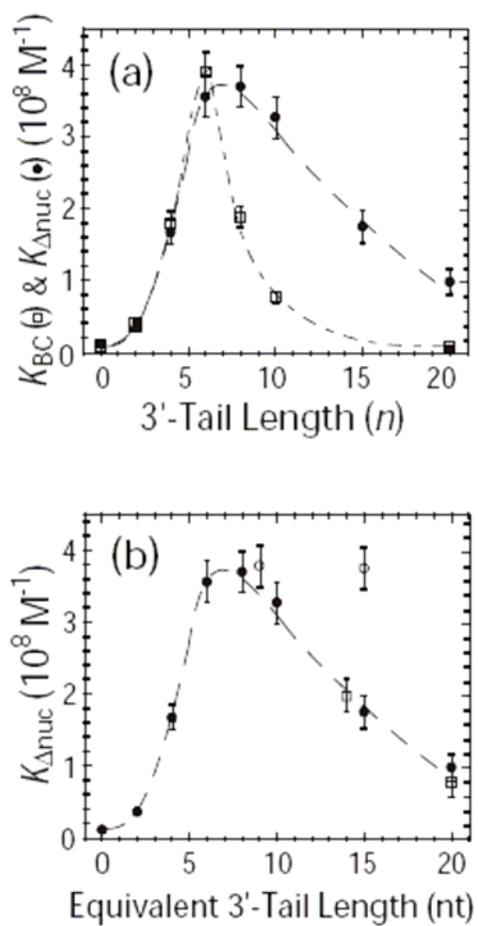


Figure 5

Figure 6

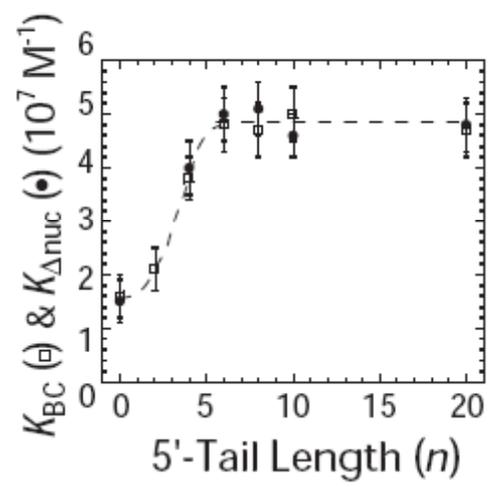


Figure 6

Figure 7

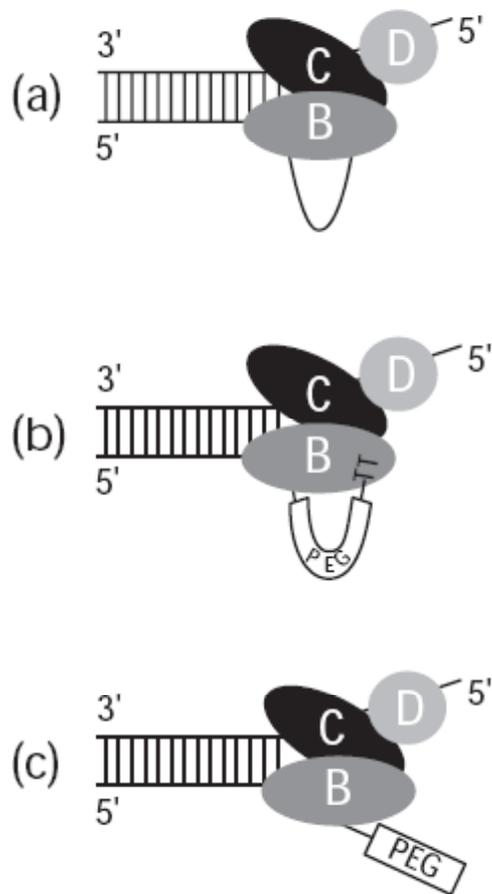


Figure 7

Figure 8

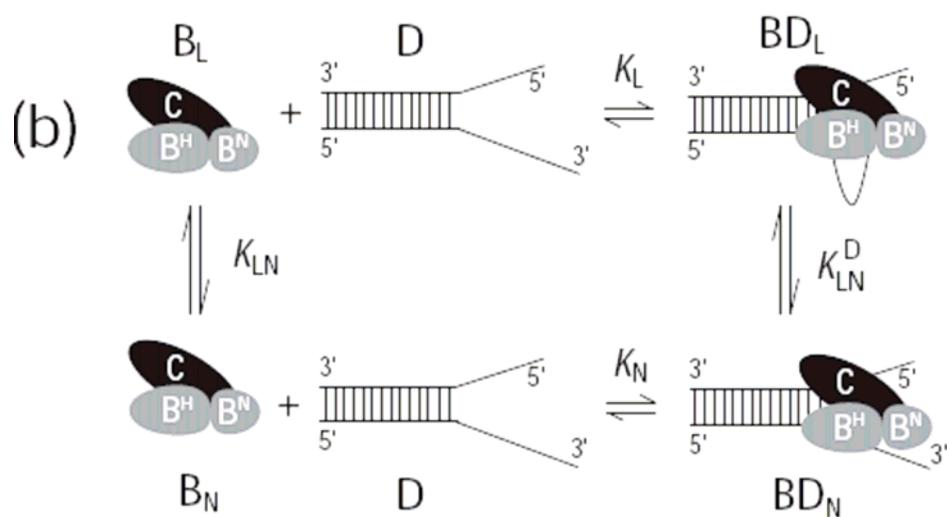
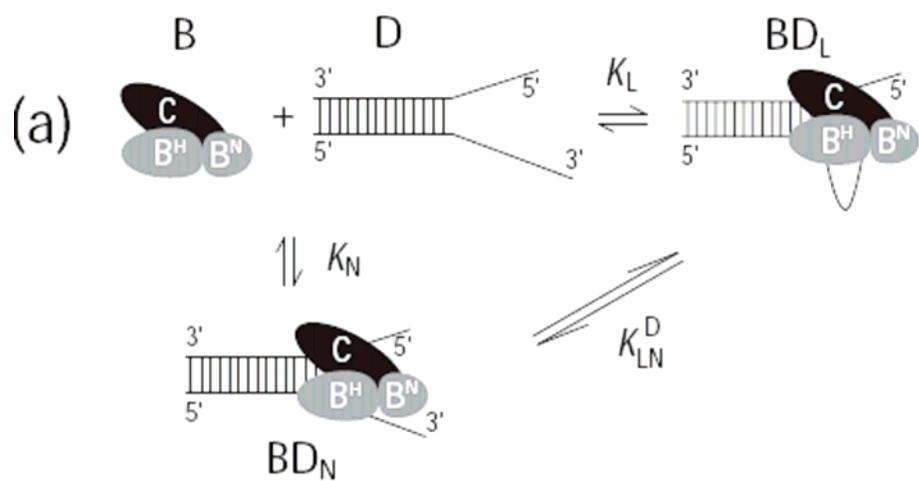


Figure 8

Figure 9

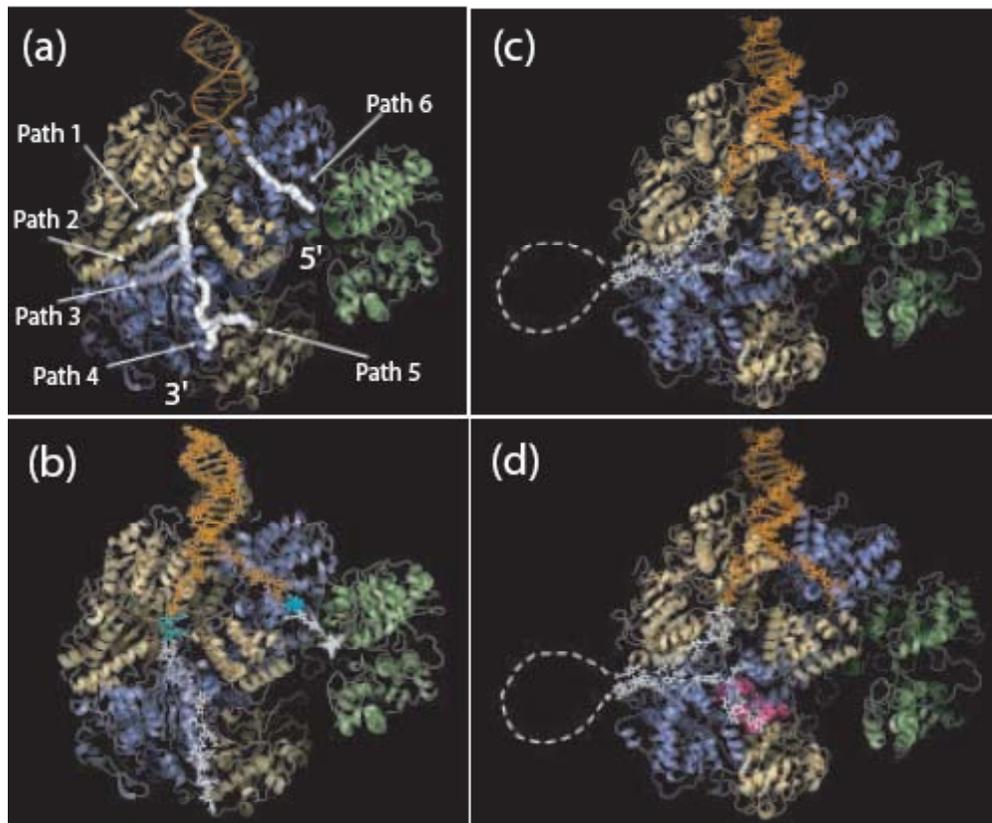


Figure 9