doi:10.1016/j.jmb.2006.07.016 J. Mol. Biol. (2006) 362, 26–43

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

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[1], Rachel L. Rice[1], Nathan A. Baker[1], Tao Ju[2] and Timothy M. Lohman[1]*

[1]*Department of Biochemistry and Molecular Biophysics Washington University School of Medicine, 660 S. Euclid Ave Box 8231, St Louis MO 63110-1093, USA*

[2]*Department of Computer Science and Engineering 517 Lopata Hall, Washington University, St Louis MO 63130, USA*

**Corresponding author*

The equilibrium binding of *Escherichia 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 ($(dT)_n$) tails was studied. These studies were designed to test a previous proposal that the 3′-$(dT)_n$ 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′-$(dT)_n$ tail. However, placing PEG at the end of the 3′-$(dT)_n$ 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′-$(dT)_6$ tail). Equilibrium binding studies of a RecBC mutant containing a nuclease domain deletion, $RecB^{\Delta nuc}C$, suggest that looping of the 3′-tail (when $n \geq 6$ nucleotides) occurs even in the absence of the RecB nuclease domain, although 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.

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

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.[1–4] Mutations in human enzymes with helicase activity result in genetic disorders such as the Werner syndrome[5,6] and the Bloom syndrome.[7,8]

Escherichia 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*.[9,10] 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

Present address: C. Jason Wong, Sections of Microbiology and Molecular and Cell Biology, University of California, Davis, Davis, CA 95616, USA.

Abbreviations used: ds, double-stranded; ss, single-stranded; PEG, polyethylene glycol; EG, ethylene glycol; ITC, isothermal titration calorimetry; NLLS, non-linear least-squares.

E-mail address of the corresponding author: lohman@biochem.wustl.edu

ssDNA intermediate upon encountering a Chi site during the course of DNA unwinding, and facilitates the loading of RecA protein onto this ssDNA.[11] 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[12] while RecD is a 5′ to 3′ helicase.[13] 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 cross-linking studies of RecBCD–DNA complexes indicating that the RecB subunit can be cross-linked to the 3′ ended strand while the RecC and RecD subunits can be cross-linked to the 5′ ended strand,[14] and the crystal structure of RecBCD bound to a DNA end.[15] The RecBC enzyme, lacking the RecD subunit, also functions as a helicase and is capable of facilitating homologous recombination *in vivo*.[16,17] However, even though the nuclease site is located on the 30 kDa C-terminal domain of the RecB subunit,[18] nuclease activity is greatly attenuated in RecBC,[19,20] hence the RecD subunit activates the RecB nuclease activity.

Both RecBCD and RecBC enzymes initiate unwinding from duplex DNA ends, including blunt-ended duplexes.[21,22] Our previous studies on the equilibrium binding of RecBC and RecBCD helicases to duplex DNA ends containing pre-existing ss-$(dT)_n$ tails of varying lengths found that the binding affinities of both RecBC and RecBCD increase as the 3′-$(dT)_n$ tail length increases from zero to six nucleotides, but then decrease sharply as the 3′-$(dT)_n$ tail length is further increased from six to 20 nucleotides.[23] Isothermal titration calorimetry (ITC) experiments with RecBC show the decrease in binding affinity for 3′-$(dT)_n$ 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′-$(dT)_n$ 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′-$(dT)_n$ 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 ($RecB^{\Delta nuc}C$)[24] to DNA molecules containing either 3′-$(dT)_n$ tails or PEG-substituted 3′-tails.

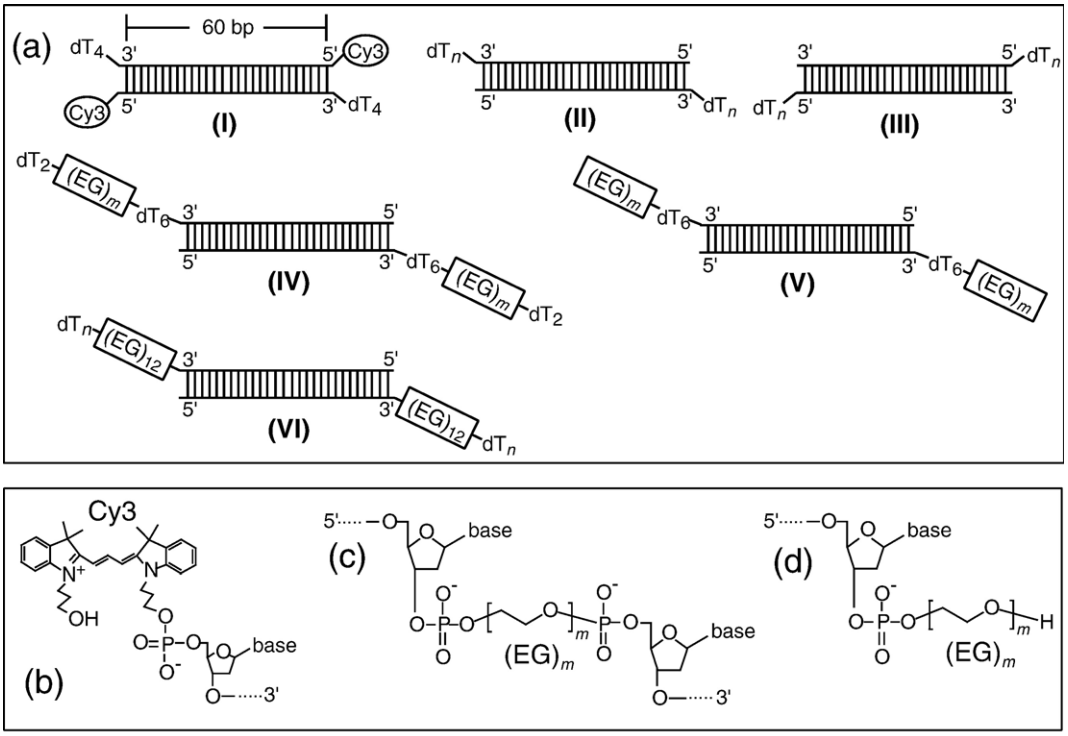
Results

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′-$(dT)_n$ 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′-$(dT)_n$ 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′-$(dT)_n$ tail, then replacing the first through sixth nucleotides of the 3′-$(dT)_n$ 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′-$(dT)_n$ tail (DNA **IV** through **VI** series in Figure 1(a)). A previous study indicates that a hexaethylene glycol chain ($(EG)_6$) has an average contour length equivalent to a trinucleotide of ssDNA.[25] 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-$(dT)_n$ tails.

Our experiments were performed with the same 60 bp Cy3-labeled reference DNA (**I** in Figure 1(a)) and two series of non-fluorescent competitor DNA molecules (**II** and **III** in Figure 1(a)) for the reasons discussed previously.[23] A series of three PEG-modified non-fluorescent competitor DNA molecules (**IV**, **V** and **VI** in Figure 1 were also studied. The DNA **IV** series molecules contain m repeating units of ethylene glycol (EG) (one repeating unit is defined as $-CH_2CH_2O-$) between the first 3′-$(dT)_6$ and the final $(dT)_2$ nucleotides at the 3′ end. The DNA **V** series molecules contain m repeating units of EG spacers after the first 3′-$(dT)_6$ 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-$(dT)_n$. In both the DNA **IV** and **V** molecules, a region of $(dT)_6$ was placed in front of the PEG region to maximize the



(e)

	Sequences of Fluorescently Labeled Single Stranded DNA used to form Reference DNA I
1	5′(Cy3)CCATGGCTCCTGAGCTAGCTGCAGTAGCCTAAAGGATGAAACTAGGATCTTATGCTCCAGTTTT-3′
2	5′-(Cy3)CTGGAGCATAAGATCCTAGTTTCATCCTTTAGGCTACTGCAGCTAGCTCAGGAGCCATGGTTTT-3′

	Sequences of Unlabeled Single Stranded DNA used to form Competitor DNA II through VI
3	5′-CCATGGCTCCTGAGCTAGCTGCAGTAGCCTAAAGGATGAAACTAGGATCTTATGCTCCAG$(T)_n$-3′
4	5′-CTGGAGCATAAGATCCTAGTTTCATCCTTTAGGCTACTGCAGCTAGCTCAGGAGCCATGG$(T)_n$-3′
5	5′-$(T)_n$CCATGGCTCCTGAGCTAGCTGCAGTAGCCTAAAGGATGAAACTAGGATCTTATGCTCCAG-3′
6	5′-$(T)_n$CTGGAGCATAAGATCCTAGTTTCATCCTTTAGGCTACTGCAGCTAGCTCAGGAGCCATGG-3′
7	5′CCATGGCTCCTGAGCTAGCTGCAGTAGCCTAAAGGATGAAACTAGGATCTTATGCTCCAGTTTTTT$(EG)_m$TT-3′
8	5′-CTGGAGCATAAGATCCTAGTTTCATCCTTTAGGCTACTGCAGCTAGCTCAGGAGCCATGGTTTTTT$(EG)_m$TT-3′
9	5′-CCATGGCTCCTGAGCTAGCTGCAGTAGCCTAAAGGATGAAACTAGGATCTTATGCTCCAGTTTTTT$(EG)_m$-3′
10	5′-CTGGAGCATAAGATCCTAGTTTCATCCTTTAGGCTACTGCAGCTAGCTCAGGAGCCATGGTTTTTT$(EG)_m$-3′
11	5′-CCATGGCTCCTGAGCTAGCTGCAGTAGCCTAAAGGATGAAACTAGGATCTTATGCTCCAG$(EG)_{12}(T)_n$-3′
12	5′-CTGGAGCATAAGATCCTAGTTTCATCCTTTAGGCTACTGCAGCTAGCTCAGGAGCCATGG$(EG)_{12}(T)_n$-3′

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 ethlyene 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 (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.

interactions between RecBC or RecBCD and the ss/dsDNA junction because both RecBC and RecBCD bind optimally to DNA possessing a 3′-$(dT)_6$ tail.[23] The chemical structures of these modified DNA molecules are shown in Figure 1(b) 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.[23] Using sedimentation equilibrium methods (see Materials and Methods) we have also shown that $RecB^{\Delta nuc}C$ exists as a stable heterodimer at 25 °C over the concentration range from 90 nM to 230 nM in buffer M plus 10 mM $MgCl_2$ and 30 or 750 mM NaCl, which are the conditions used here (data not shown). As in our previous studies,[23] we have used the fluorescence competition method[26] to investigate the equilibrium binding of RecBC, RecBCD and $RecB^{\Delta nuc}C$ 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′-$(dT)_n$ tails of equivalent lengths (DNA **II** series). The equilibrium constant for RecBCD binding to the end of reference DNA **I** was measured previously[23] ($K_{BCD,R} = (7.5\pm0.4)\times10^7$ 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.[23] 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 squares (NLLS) methods using equations (11) and (13) as described in Materials and Methods to obtain the K_{BCD} value for binding to the end of each competitor DNA (Table 1).

Figure 2(a) shows the results of representative competition experiments performed with a reference DNA **I** (20 nM) and DNA **IV** molecule with $m=12$. The continuous 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 2(b) as a function of equivalent 3′-ssDNA tail length, along with K_{BCD} (Table 1) for DNA with unmodified 3′-$(dT)_n$ tails (DNA **II** series) as measured previously.[23]

The DNA **IV** substrates have either 12 or 24 units of EG placed between the first $(dT)_6$ and final $(dT)_2$ on the 3′-tail, which is equivalent to a total 3′-ssDNA tail length of 14 and 20 nucleotides, respectively. As shown in Figure 2(b) and Table 1, K_{BCD} for these two DNA **IV** molecules $(2.5(\pm0.5)\times10^7$ M^{-1} and $0.53(\pm0.05)\times10^7$ M^{-1} for $m=12$ and 24, respectively) are the same as K_{BCD} for DNA molecules with regular 3′-$(dT)_n$ tails (DNA **II** series) of equivalent lengths $(2.3(\pm0.6)\times10^7$ M^{-1} and $0.51(\pm0.03)\times10^7$ 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′-$(dT)_6$ tail, have a total 3′-ssDNA tail length equivalent to nine or 15 nucleotides, respectively. As shown in Figure 2(b) and Table 1, the K_{BCD} value for these two DNA **V** molecules $(22(\pm3)\times10^7$ M^{-1} and $21(\pm3)\times10^7$ M^{-1} for $m=6$ and 18, respectively) are the same as K_{BCD} for a DNA with a 3′-$(dT)_6$ tail $(23(\pm3)\times10^7$ $M^{-1})$, which is the maximum K_{BCD} value for a DNA with only a regular 3′-$(dT)_n$ tail. The K_{BCD} value for the DNA **V** series remains constant at $22(\pm3)\times10^7$ M^{-1} as the number of units of EG increases from $m=6$ to $m=18$ while the K_{BCD} value for the DNA **IV** series decreases ~ fivefold as m increases from 12 to 24 (Table 1). The DNA **VI** molecules, with either $(dT)_2$ or $(dT)_6$ placed after 12 units of EG on the 3′-tail,have a total 3′-ssDNA tail length equivalent to eight or 12 nucleotides, respectively. The K_{BCD} value for these two DNA **VI** molecules $(1.8(\pm0.4)\times10^7$ M^{-1} and $0.55(\pm0.06)\times10^7$ M^{-1} for $n=2$ and 6, respectively; Table 1) are lower than the K_{BCD} for a blunt DNA end $(3(\pm0.4)\times10^7$ $M^{-1})$ by a factor of ~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′-$(dT)_n$ tail increases from six to 20 nucleotides.[23] 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 the 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 the K_{BC} value (Table 2).

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 **II**[a]	3′-Tail composition of DNA **II**	K_{BCD} (10^7 M^{-1}) for DNA with PEG-substituted 3′-tail (**IV**, **V** and **VI**)	3′-Tail composition of DNA with PEG-substituted 3′-tail (**IV**, **V** and **VI**)
0	3.0±0.4	No tail		
2	5.2±0.4	3′-$(dT)_2$		
4	15±1	3′-$(dT)_4$		
6	23±3	3′-$(dT)_6$		
8	16±2	3′-$(dT)_8$	1.8±0.4	3′-$(EG)_{12}(dT)_2$
9	ND		22±3	3′-$(dT)_6(EG)_6$
10	6.4±0.8	3′-$(dT)_{10}$		
12	ND		0.55±0.06	3′-$(EG)_{12}(dT)_6$
14	ND		2.5±0.5	3′-$(dT)_6(EG)_{12}(dT)_2$
15	2.3±0.6	3′-$(dT)_{15}$	21±3	3′-$(dT)_6(EG)_{18}$
20	0.51±0.03	3′-$(dT)_{20}$	0.53±0.05	3′-$(dT)_6(EG)_{24}(dT)_2$

[a] Data from Wong *et al.*[23]

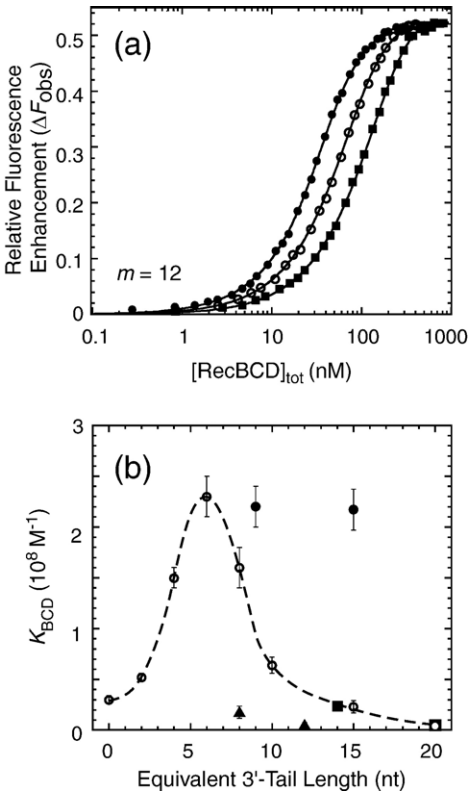


Figure 2. Effects of replacing regions of the pre-existing 3′-$(dT)_n$ tail with PEG on the equilibrium constants (K_{BCD}) for RecBCD binding to a duplex DNA end. (a) Equilibrium competition titrations to determine K_{BCD} for RecBCD binding to the ends of a non-fluorescent DNA **IV** molecule possessing 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 (ΔF_{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 (■). Continuous lines are simulations using equations (11) and (13) and the best fit values of K_{BCD} (Table 1). (b) Values of K_{BCD} for binding to the ends of the DNA **IV** series molecules containing 3′-$(dT)_6(EG)_m(dT)_2$ (with $m=12$ or 24) (■); for the DNA **V** series containing 3′-$(dT)_6(EG)_m$ (with $m=6$ or 18) (●); and for the DNA **VI** series containing 3′-$(EG)_{12}$ $(dT)_n$ (with $n=2$ or 6) (▲) (Table 1) are plotted as a function of equivalent 3′-tail length, along with the K_{BCD} for the DNA **II** series molecules containing unmodified 3′-$(dT)_n$ tails (with $n=0$ to 20 nucleotides) (○) (Table 1).

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.[23] As was observed with RecBCD, replacing the ssDNA in the middle stretch of the 3′-tail with PEG (DNA **IV** series) does not change the K_{BC}, whereas placing PEG at the end of the 3′-tail (DNA **V** series) increases the K_{BC} to the same value as K_{BC} for a DNA end with only a 3′-$(dT)_6$ tail (Figure 3). Substituting the first through the sixth nucleotides on the 3′-$(dT)_n$ 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 the K_{BC} and K_{BCD} values for the DNA **IV** series are the same, within experimental error, as the K_{BC} and K_{BCD} values for the DNA **II** series molecules containing the equivalent length of 3′-$(dT)_n$ 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′-$(dT)_n$ ssDNA tails on RecB$^{\Delta nuc}$C binding to DNA ends

Although the recent crystal structure of RecBCD in complex with a 19 bp hairpin duplex[15] indicates potential contacts between the ss/dsDNA junction and both RecB and RecC subunits, the 3′-ssDNA region is not sufficiently long (four 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[27] 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$^{\Delta nuc}$C.

Equilibrium binding of RecB$^{\Delta nuc}$C to reference DNA **I** is very tight under low monovalent salt conditions (30 mM NaCl in buffer M plus 10 mM $MgCl_2$) and results in a ~45% enhancement in the Cy3 fluorescence signal (Figure 4(a). Under these tight binding conditions, the relative fluorescence change (ΔF_{obs} as defined in equation (9)) increases linearly with increasing [RecB$^{\Delta nuc}$C] until a sharp breakpoint is reached at a molar ratio of two RecB$^{\Delta nuc}$C per DNA **I** (Figure 4(a)), consistent with one molecule of RecB$^{\Delta nuc}$C bound per DNA end at saturation. Importantly, the sharp breakpoint at a RecB$^{\Delta nuc}$C/DNA ratio of two indicates that our preparation of RecB$^{\Delta nuc}$C protein is 100% active in DNA binding.

In order to lower the binding constant of RecB$^{\Delta nuc}$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 $MgCl_2$ at 25 °C). Titrations were performed at three different concentrations of DNA **I** (Figure

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 $MgCl_2$ plus 100 mM NaCl at 25 °C)

3′-ssDNA tail length (nt)	K_{BC} (10^7 M^{-1}) for DNA **II**[a]	3′-Tail composition of DNA **II**	K_{BC} (10^7 M^{-1}) for DNA with PEG-substituted 3′-tail (**IV** and **V**)	3′-Tail composition of DNA with PEG-substituted 3′-tail (**IV** and **V**)
0	1.6±0.3	No tail		
2	4.3±0.4	3′-$(dT)_2$		
4	19±2	3′-$(dT)_4$		
6	40±3	3′-$(dT)_6$		
8	19±3	3′-$(dT)_8$	0.77±0.08	3′-$(EG)_{12}(dT)_2$
9	ND		39±3	3′-$(dT)_6(EG)_6$
10	8.8±0.9	3′-$(dT)_{10}$		
11	ND		8.8±0.9	3′-$(dT)_6(EG)_6(dT)_2$
12	ND		0.31±0.06	3′-$(EG)_{12}(dT)_6$
15	ND		36±3	3′-$(dT)_6(EG)_{18}$
20	0.89±0.09	3′-$(dT)_{20}$	0.87±0.07	3′-$(dT)_6(EG)_{24}(dT)_2$

[a] Data from Wong *et al.*[23]

4(b)). These three titrations were analyzed using the MBDF method[28–30] to determine the relationship between ΔF_{obs} and the average extent of $RecB^{\Delta nuc}C$ bound to the DNA. Figure 4(c) shows that ΔF_{obs} is directly proportional to the average number of molecules of $RecB^{\Delta nuc}C$ bound per DNA **I** molecule ($[RecB^{\Delta nuc}C]_{bound}/[DNA]_{total}$) and extrapolates to a value of 2 at $\Delta F_{obs}=0.45$, indicating a stoichiometry of two molecules of $RecB^{\Delta nuc}C$ 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, $K_{\Delta nuc,R}=4.5(\pm0.5)\times10^7$ M^{-1} for $RecB^{\Delta nuc}C$ 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** ($K_{BC,R}=4.8(\pm0.3)\times10^7$ M^{-1}).[23] The fact that the same equilibrium constant describes all three titrations very well and that the Cy3 fluorescence enhancement is the same for $RecB^{\Delta nuc}C$ binding to each DNA end indicates that each $RecB^{\Delta nuc}C$ molecule binds independently to each DNA end and the binding affinity of $RecB^{\Delta nuc}C$ for each DNA end is identical within experimental uncertainties. These results also indicate that the $RecB^{\Delta nuc}C$ enzyme does not undergo any change in its assembly state either when free or upon binding DNA over the protein concentration range examined.

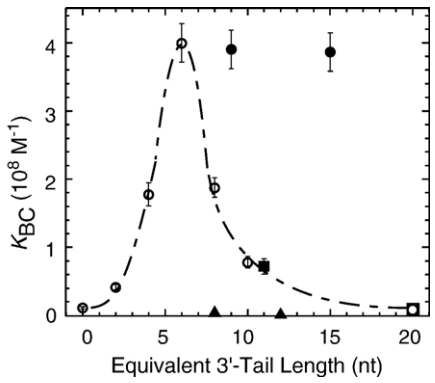


Figure 3. Effects of replacing part of the pre-existing 3′-$(dT)_n$ tail with PEG on RecBC binding to a duplex DNA end. Values of K_{BC} for binding to the ends of the DNA **IV** series containing 3′-$(dT)_6(EG)_m(dT)_2$ (with $m=6$ or 24) (■); K_{BC} for the DNA **V** series containing 3′-$(dT)_6(EG)_m$ (with $m=6$ or 18) (●); and K_{BC} for DNA **VI** series containing 3′-$(EG)_{12}$ $(dT)_n$ (with $n=2$ or 6) (▲) (Table 2) are plotted as a function of equivalent 3′-tail length, along with K_{BC} for the DNA **II** series containing unmodified 3′-$(dT)_n$ tails (○) (Table 2). Experiments were performed in buffer M, 10 mM $MgCl_2$ plus 100 mM NaCl at 25 °C.

Competition experiments were performed to obtain the equilibrium constant ($K_{\Delta nuc}$) for $RecB^{\Delta nuc}C$ 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 $K_{\Delta nuc}$ are presented in Table 3 and plotted in Figure 5(a) as a function of 3′-$(dT)_n$ tail length, along with K_{BC} for the same DNA **II** series (Table 2). As shown in Figure 5(a), when $n\leq6$ nucleotides, $K_{\Delta nuc}$ is the same, within experimental uncertainties, as K_{BC}. Both $K_{\Delta nuc}$ and K_{BC} increase by a factor of ~20 as the length of the 3′-tail increases from zero to six nucleotides. However, although $K_{\Delta nuc}$ decreases as the 3′-tail length increases beyond six nucleotides, $K_{\Delta nuc}$ decreases only by a factor of ~4, as opposed to the ~20-fold decrease exhibited by K_{BC}. Hence, removal of the nuclease domain does influence the binding constant for $n\geq6$ nucleotides although a decrease is still observed.

To further understand the cause for the different tail length dependences exhibited by $K_{\Delta nuc}$ and K_{BC} when $n>6$ nucleotides, we measured $K_{\Delta nuc}$ 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 $K_{\Delta nuc}$ 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 the $K_{\Delta nuc}$ value 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 $K_{\Delta nuc}$ are presented in Table 3 and plotted in Figure 5(b) as a function of equivalent 3′-tail length.

As shown in Figure 5(b), the $K_{\Delta nuc}$ value for the DNA **IV** series ($21(\pm2)\times10^7$ M^{-1} and $9(\pm2)\times10^7$ 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′-$(dT)_n$ tail of equivalent length ($19(\pm2)\times10^7$ M^{-1} and $10(\pm2)\times10^7$ M^{-1} for $n=15$ and 20, respectively). This result was the same as that observed for RecBCD (Figure 2(b)) and RecBC (Figure 3). The $K_{\Delta nuc}$ for $RecB^{\Delta nuc}C$ binding to a DNA end with PEG placed at the end of the 3′-tail (DNA **V** series) ($38(\pm3)\times10^7$ M^{-1} and $38(\pm4)\times10^7$ M^{-1} for $m=6$ and 18, respectively) has the same value as the $K_{\Delta nuc}$ for a DNA **II** molecule with only a 3′-$(dT)_6$ tail ($38(\pm3)\times10^7$ M^{-1}) (Figure 5(b)), the same behavior as was observed for RecBCD (Figure 2(b)) and RecBC (Figure 3). Figure 5(b) also shows that the $K_{\Delta nuc}$ for binding to a DNA **V** molecule with $m=6$ is the same as the $K_{\Delta nuc}$ value for binding to a DNA **V** molecule with $m=18$. This behavior is very different from the decrease observed in $K_{\Delta nuc}$ 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′-$(dT)_n$ tail with $n>6$ nucleotides upon binding $RecB^{\Delta nuc}C$ to a DNA end, although the loop probability appears to be affected.

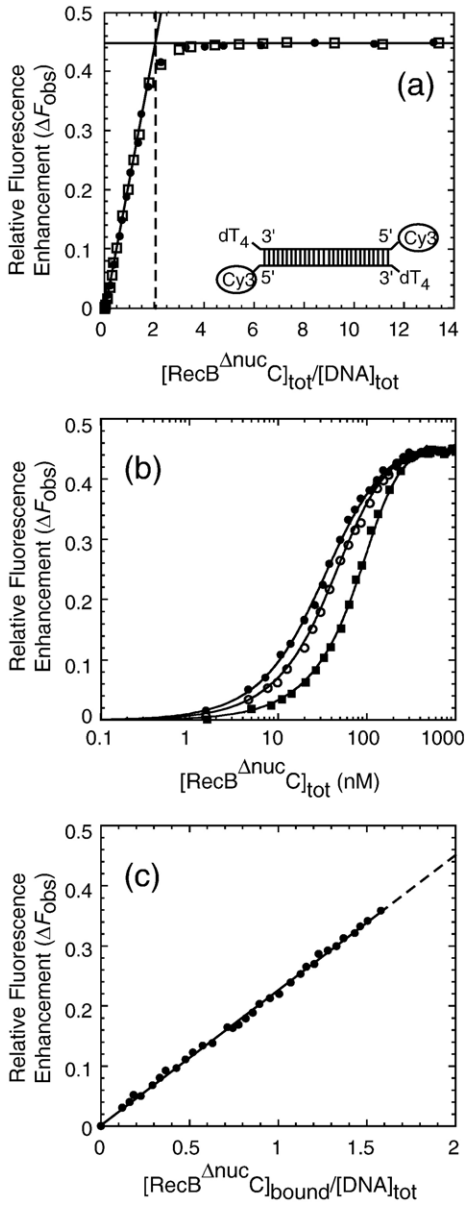


Figure 4. Binding of $RecB^{\Delta nuc}C$ to the fluorescent reference DNA **I**. (a) Titrations of reference DNA **I** (total concentration of 20 nM (●) and 60 nM (□)) with $RecB^{\Delta nuc}C$ under stoichiometric (high affinity) conditions in buffer M, 10 mM $MgCl_2$ plus 30 mM NaCl at 25 °C. Cy3 fluorescence enhancement (ΔF_{obs}) is plotted as a function of the $[RecB^{\Delta nuc}C]_{total}/[DNA]_{total}$. Excitation wavelength was 515 nm and fluorescence emission was monitored at 563 nm. (b) Determination of the equilibrium constant for $RecB^{\Delta nuc}C$ 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 $MgCl_2$ plus 100 mM NaCl at 25 °C) and ΔF_{obs} is plotted as a function of total $[RecB^{\Delta nuc}C]$. Global NLLS analysis of these data using equation (12) as described in Materials and Methods yields the equilibrium constant for $RecB^{\Delta nuc}C$ binding to each end of reference DNA **I**, $K_{\Delta nuc,R}=4.5(\pm0.5)\times10^7$ M^{-1} and $\Delta F_{max}=0.45\pm0.04$. Continuous lines are simulations using equation (12) and the best fit values of $K_{\Delta nuc,R}$ and ΔF_{max}. (c) ΔF_{obs} is directly proportional to the average moles of $RecB^{\Delta nuc}C$ bound per mole of DNA **I**. The dependence of ΔF_{obs} on $[RecB^{\Delta nuc}C]_{bound}/[DNA]_{total}$ was determined from the titration experiments in (b) performed at 10 and 50 nM DNA **I** as described.[28–30] The continuous line is a linear fit of the dependence of ΔF_{obs} on $[RecB^{\Delta nuc}C]_{bound}/[DNA]_{total}$ while the broken line is a linear extrapolation of the continuous line to $\Delta F_{max}=0.45$ and $[RecB^{\Delta nuc}C]_{bound}/[DNA]_{total}=2$.

Effects of 5′-$(dT)_n$ ssDNA tails on $RecB^{\Delta nuc}C$ binding to DNA ends

We also examined if removal of the nuclease domain influences the effects of the 5′-$(dT)_n$ tail length on the energetics of $RecB^{\Delta nuc}C$ binding to a DNA end by measuring $K_{\Delta nuc}$ for binding to the non-fluorescent DNA **III** series containing pre-existing 5′-$(dT)_n$ tails with n varying from zero to 20. The same competition approach, using DNA **I** as the reference DNA, yields the values of $K_{\Delta nuc}$ presented in Table 4 and plotted in Figure 6. Values of K_{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 $K_{\Delta nuc}$ and K_{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′-$(dT)_n$ 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.[21,22] In our previous equilibrium studies

Table 3. Equilibrium constants ($K_{\Delta nuc}$) for RecB$^{\Delta nuc}$C binding to the ends of non-fluorescent DNA **II**, **IV** and **V** series (buffer M plus 10 mM $MgCl_2$ and 100 mM NaCl at 25 °C)

3′-ssDNA tail length (nt)	$K_{\Delta nuc}$ (10^7 M^{-1}) for DNA **II**	3′-Tail composition of DNA **II**	$K_{\Delta nuc}$ (10^7 M^{-1}) for DNA with PEG-substituted 3′-tail (**IV** and **V**)	3′-Tail composition of PEG-substituted DNA (**IV** and **V**)
0	1.5±0.2	No tail		
2	3.9±0.3	3′-$(dT)_2$		
4	15±2	3′-$(dT)_4$		
6	37±3	3′-$(dT)_6$		
8	38±3	3′-$(dT)_8$		
9	ND		38±3	3′-$(dT)_6(EG)_6$
10	34±3	3′-$(dT)_{10}$		
14	ND		21±2	3′-$(dT)_6(EG)_{12}(dT)_2$
15	19±2	3′-$(dT)_{15}$	38±4	3′-$(dT)_6(EG)_{18}$
20	10±2	3′-$(dT)_{20}$	9±2	3′-$(dT)_6(EG)_{24}(dT)_2$

of RecBC and RecBCD binding to duplex DNA ends in the absence of ATP, we observed that both K_{BC} and K_{BCD} increase as the length of a pre-formed 3′-$(dT)_n$ tail increases from zero to six nucleotides. However, when the length of the 3′-$(dT)_n$ tail increases beyond six nucleotides, both K_{BC} and K_{BCD} decrease sharply.[23] ITC studies performed with RecBC indicate that the decrease in K_{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.[23] No evidence for such looping in pre-formed 5′-$(dT)_n$ tails was observed.

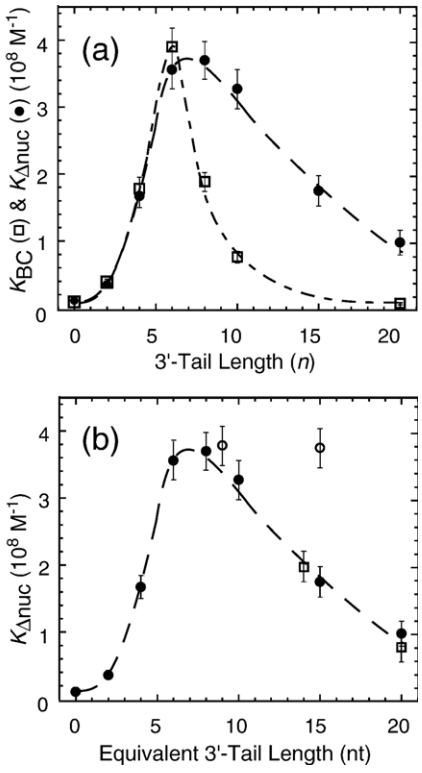


Figure 5. Influence of the nuclease domain of RecB on the dependence of the end binding constant on the length of the pre-existing 3′-$(dT)_n$ tails. All experiments were performed in buffer M plus 10 mM $MgCl_2$ and 100 mM NaCl at 25 °C. (a) Comparisons of the effects of the length of the pre-existing 3′-$(dT)_n$ tails on the equilibrium constants for RecB$^{\Delta nuc}$C ($K_{\Delta nuc}$) and RecBC (K_{BC}) binding to the ends of duplex DNA. Values of $K_{\Delta nuc}$ (●) and K_{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′-$(dT)_n$ tails with PEG on $K_{\Delta nuc}$. Values of $K_{\Delta nuc}$ for the DNA **IV** series with 3′-$(dT)_6(EG)_m(dT)_2$ tails ($m = 12$ or 24) (□) and $K_{\Delta nuc}$ for the DNA **V** series with 3′-$(dT)_6(EG)_m$ tails ($m = 6$ or 18) (○) are plotted as a function of equivalent 3′-tail length, along with $K_{\Delta nuc}$ for the DNA **II** series (●) (Table 3).

Formation of 3′-ssDNA tail loop in RecBCD and RecBC-DNA complexes

The 3′-tail looping model[23] predicts that the helicase interacts with two regions of the pre-existing 3′-$(dT)_n$ tail when $n \geq$ six nucleotides (Figure 7(a)). 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′-$(dT)_n$ 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′-$(dT)_n$ tail with PEG does not affect K_{BC} or K_{BCD} (Figure 2(b) and Figure 3) is consistent with this prediction (Figure 7(b)).

The 3′-tail looping model also hypothesizes that the increasingly unfavorable entropic contribution to binding upon increasing the 3′-$(dT)_n$ 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′-$(dT)_n$ 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′-$(dT)_n$ 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 2(b) and Figure 3). These data further suggest that the two (possibly one) nucleotides at the 3′-end of the ss-DNA tail

Table 4. Equilibrium constants for RecB$^{\Delta nuc}$C ($K_{\Delta nuc}$) and RecBC (K_{BC}) binding to the ends of non-fluorescent DNA **III** series (buffer M plus 10 mM $MgCl_2$ and 100 mM NaCl at 25 °C)

5′-(dT)$_n$ tail length (n)	K_{BC} (10^7 M^{-1}) for DNA **III**	$K_{\Delta nuc}$ (10^7 M^{-1}) for DNA **III**
0	1.6±0.3	1.6±0.3
2	2.2±0.4	ND
4	3.8±0.4	4.1±0.5
6	4.8±0.5	5.0±0.5
8	4.7±0.4	5.1±0.5
10	5.0±0.5	4.6±0.4
20	4.7±0.5	4.8±0.5

are needed for loop formation (Figure 7(c)) 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)$_n$ tail with PEG results in a binding constant lower than that for binding to a blunt DNA end.

Looped *versus* 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 8(a)) is not consistent with our results, as discussed below. In Scheme 1, B represents RecBC (or RecBCD), BD_L represents the looped complex and BD_N represents

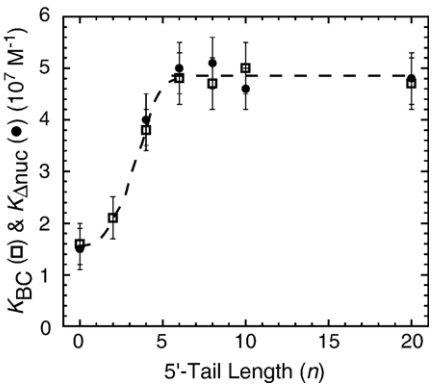


Figure 6. Deletion of the nuclease domain of RecB does not affect the influence of the 5′-ssDNA tail length on the binding of RecBC to a DNA end. Equilibrium constants ($K_{\Delta nuc}$) for RecB$^{\Delta nuc}$C binding to the ends of the DNA **III** series possessing pre-existing 5′-(dT)$_n$ tails with n varying between zero and 20 nucleotides (●) are plotted as a function of 5′-tail length (n), along with equilibrium constants (K_{BC}) for RecBC binding to the ends of the same DNA **III** series (□) (Table 4). All experiments were performed in buffer M, 10 mM $MgCl_2$ plus 100 mM NaCl at 25 °C.

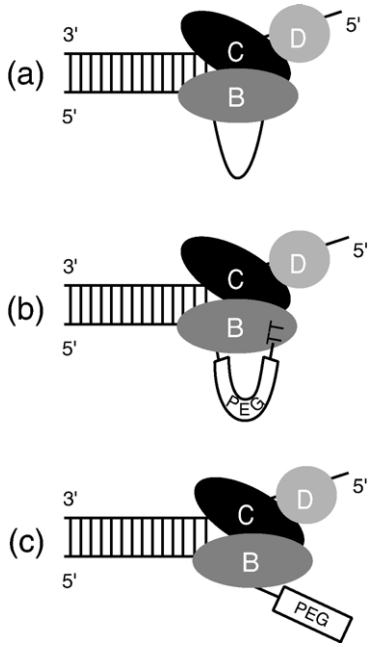


Figure 7. Cartoons depicting the binding of RecBCD to DNA ends possessing different types of pre-existing 3′-tails. (a) Looping of the 3′-ssDNA tail occurs when RecBCD binds to a DNA end containing a 3′-(dT)$_n$ tail with n > six nucleotides. (b) Replacing the section of 3′-(dT)$_n$ tail between the 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.

the non-looped complex (Figure 8(a)). The binding polynomial for Scheme 1 is given by equation (1):

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

where B_f is the free RecBC or RecBCD concentration, K_L is the equilibrium constant for formation of a looped complex, K_N is the equilibrium constant for formation of a non-looped complex, and K_{LN}^D describes the equilibrium between the looped and the non-looped states as defined in equation (2):

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

The observed equilibrium constant for the binding of B to one DNA end (K_{obs}) can then be expressed in terms of K_L and K_N as in equation (3):

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

K_L is expected to decrease as the length of the 3′-(dT)$_n$ 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)$_n$ tail over six nucleotides long, then K_L 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)$_n$ 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(±2) to 0.89(±0.09)×10^7 M^{-1} for RecBC binding to the DNA **II** series (Table 2)) while K_N remains unchanged (40(±3)×10^7 M^{-1} for RecBC binding to a DNA **II** substrate with $n=6$ (Table 2)). Hence, Scheme 1 predicts that K^D_{LN} 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)$_n$ 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(±2) to 0.89(±0.09)×10^7 M^{-1}) (Table 2) is always lower than K_{BC} for a DNA end with 3′-(dT)$_6$ tail (40(±3)×10^7 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 8(b)), 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):

$$K_{LN} = \left[\frac{B_N}{B_L}\right] \tag{4}$$

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

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

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

$$K_{obs} = \frac{K_L + K_N K_{LN}}{1 + K_{LN}} \tag{6}$$

$$K^D_{LN} = \frac{K_{LN} K_N}{K_L} \tag{7}$$

For the case of RecBC, if K_N equals K_{BC} for binding to a DNA end with a 3′-(dT)$_6$ tail (~4.0×10^8 M^{-1}) and K_L equals K_{BC} for binding to the DNA **II** series with $n>6$ (ranging from ~19×10^7 M^{-1} to 1×10^7 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)$_n$ tail is increased beyond six nucleotides, $K_{\Delta nuc}$ exhibits a much more gradual decrease than K_{BC} (Figure 5(a)). Our PEG-substituted DNA data (Figure 5(b)) indicate that some 3′-tail looping does occur when RecB$^{\Delta nuc}$C binds a DNA with a 3′-(dT)$_n$ tail longer than six nucleotides. Hence, the higher value of $K_{\Delta nuc}$ over K_{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$^{\Delta nuc}$C–DNA complex is different from the loop formed in a RecBC–DNA complex. The second possibility is that the population of RecB$^{\Delta nuc}$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_{LN} and/or K_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,[15] 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 ssDNA 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.

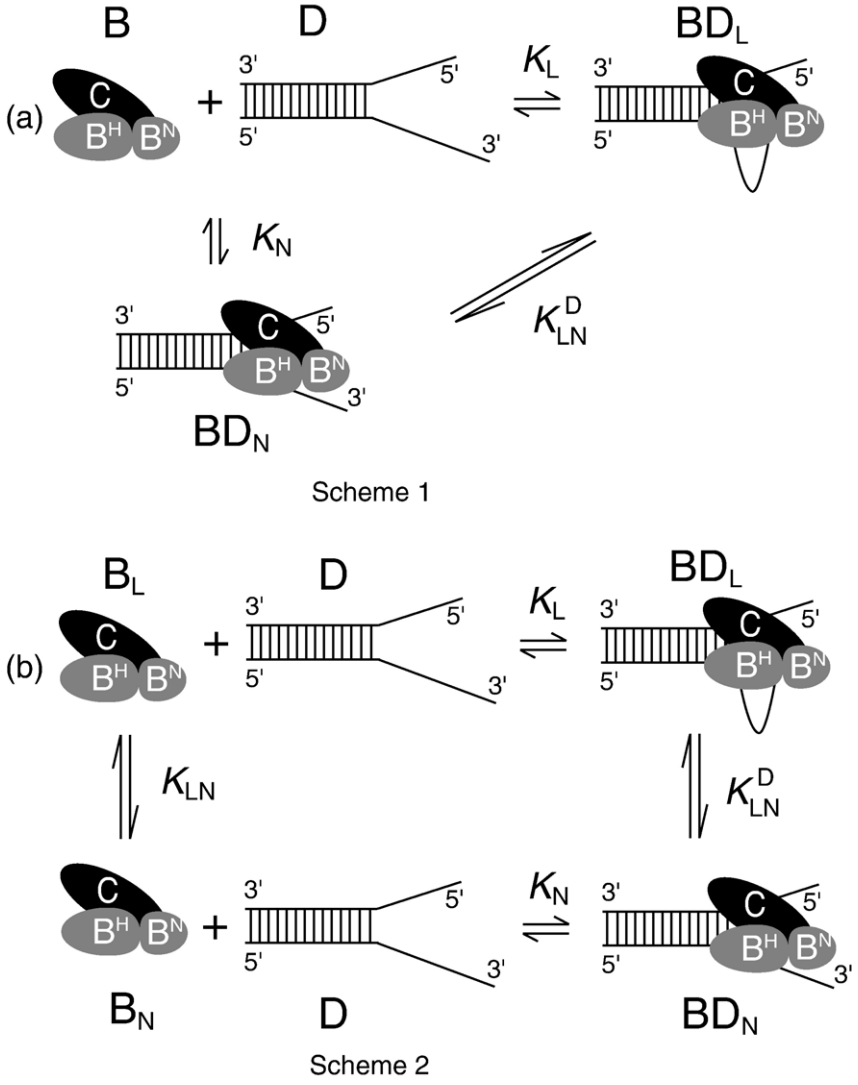


Figure 8. Possible equilibrium schemes for the binding of RecBC to a DNA end possessing a pre-existing 3′-$(dT)_n$ tail with n > six nucleotides. (a) In Scheme 1, the binding of RecBC (B) can result in the formation of either a looped (BD_L) or non-looped (BD_N) state. (b) In Scheme 2, RecBC is assumed to exist in two conformations (B_L and B_N) prior to binding DNA. The looped complex forms when B_L binds to a DNA end possessing a pre-existing 3′-$(dT)_n$ tail with n > six nucleotides, while the binding of B_N to a DNA end results in a non-looped complex. The nuclease domain (B^N) and helicase domain (B^H) of RecB are indicated.

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.[31–33] 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.[33] 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[33] 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.*[15] and Spies *et al.*[34] 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[35]), 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 ssDNA 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[14,23,36] as well as the RecBCD–DNA crystal structure[15] 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,[15] 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 9(a)) based on satisfying one of the following three criteria: (1) paths that were previously identified in the crystal structure;[15] (2) paths that are consistent with our experimental binding data;[23] 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.[23] The existing four-nucleotide 5′-ssDNA tail in the crystal structure (shown in orange in Figure 9(b)) 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 9(b) 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.[23] The model in Figure 9(b) shows that a ten-nucleotide 5′-ssDNA tail reaches, at least partially, into the DNA binding site within RecD.[23]

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 9(a)) is the shortest path, requiring only two additional nucleotides (six nucleotides total length from the duplex) to reach the protein surface. Paths 2 and 3 (see Figure 9(a)) 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 9(a)) 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 (Asp1067, Lys1082, and Asp1080). Paths 4 and 5 were previously identified by Singleton *et al.*[15] 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 9(b)). Interestingly, in this model, the sixth nucleotide of the 3′-ssDNA tail just reaches the RecB/RecC interface (Figure 9(b)). Since our

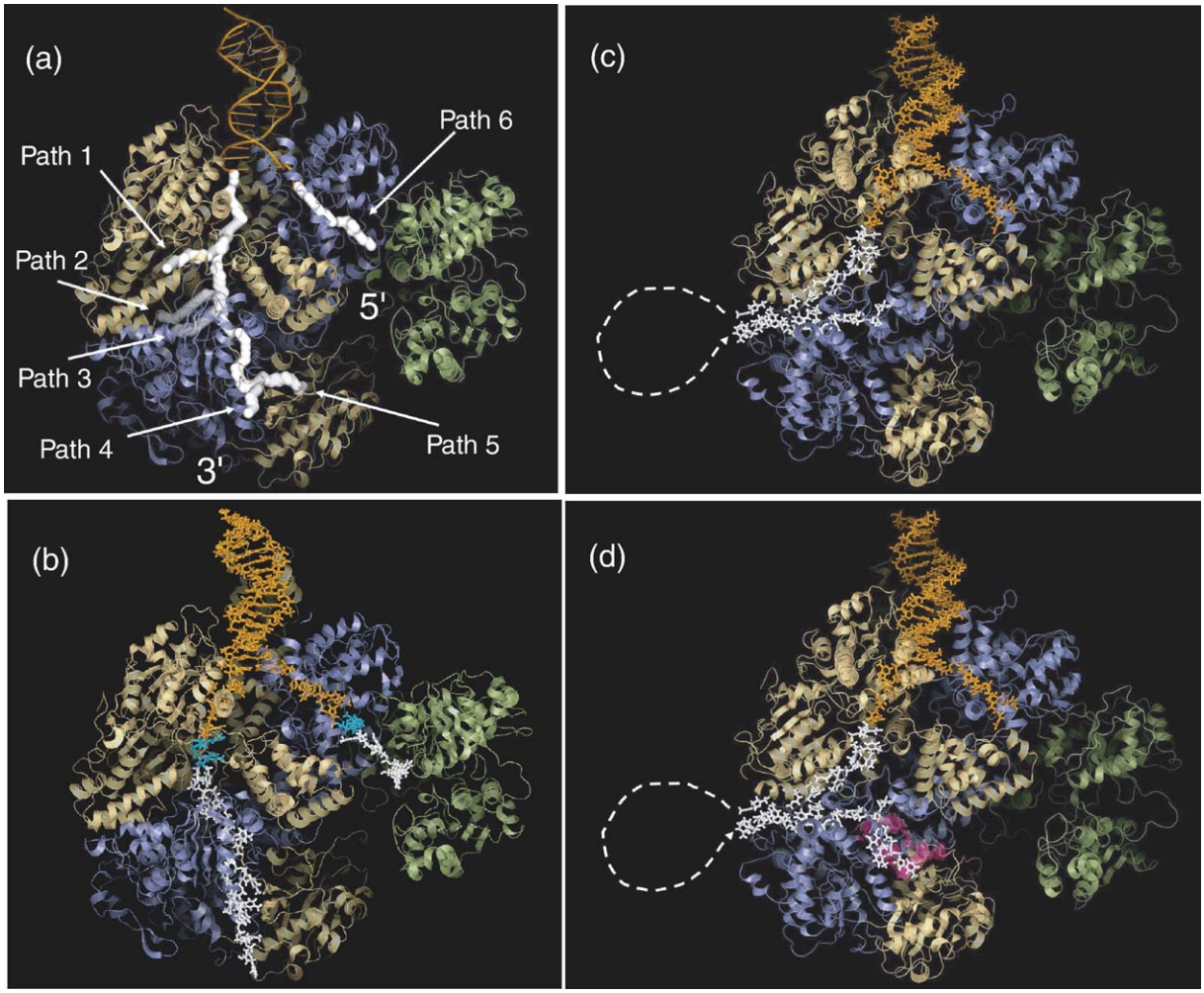


Figure 9. Potential pathways for the 3′ and 5′-ssDNA tails within the RecBCD-DNA complex and a model for the RecBCD–Chi recognition complex. (a) Representation of the six potential paths for the extended ssDNA tails as described in the text. (b) Proposed model for the non-looped complex in which the 3′-ssDNA tail is threaded through both the RecB and RecC subunits and exits near the RecB nuclease domain. The four-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.*[15] The 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 the 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 18th 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_L) 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.[35]

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 9(c).

We envision the model shown in Figure 9(c) 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 9(d), we have extended the 3′ end of the ssDNA to continue along path 4 until it reaches the region of the RecC subunit proposed to contain the Chi recognition site[35] (shown in magenta in Figure

9(d)). We suggest that the complex in Figure 9(d) would represent the looped structure formed after recognition of a Chi sequence during DNA unwinding, whereas Figure 9(b) 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 ssDNA would need to exit the enzyme, thus forming the proposed loop. As suggested by Spies & Kowalczykowski,[34] 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 $MgCl_2$ 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 [$MgCl_2$].[37]

Proteins

E. coli RecB and RecC proteins were purified and reconstituted to form RecBC as described.[38] RecBCD protein was purified as described elsewhere.[39] RecBC and RecBCD concentrations were determined spectrophotometrically in buffer C using extinction coefficients of $\varepsilon_{280}=3.9\times10^5$ M^{-1} cm^{-1} [38] and $\varepsilon_{280}=4.5\times10^5$ M^{-1} cm^{-1},[39] respectively. $RecB^{\Delta nuc}C$ was reconstituted from $RecB^{\Delta nuc}$ and RecC proteins by mixing equimolar concentrations of purified $RecB^{\Delta nuc}$ and RecC proteins in buffer C and incubated on ice for 10 min. $RecB^{\Delta nuc}$ was expressed in *E. coli* strain V186[40] carrying pNM52 ($lacI^q$)[41] and pMY100 ($recB^{\Delta nuc+}$)[1] (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,[42] followed by chromatorgraphy purification that is the same as the RecB purification.[38] $RecB^{\Delta nuc}$ has an extinction coefficient of $\varepsilon_{280}=1.4\times10^5$ M^{-1} cm^{-1} in buffer C, which was determined by comparing the absorbance spectra of aliquots of $RecB^{\Delta nuc}$ in buffer C to spectra taken in 6 M guanidinium-HCl. The extinction coefficient of $RecB^{\Delta nuc}$ in 6 M guanidinium-HCl at 280 nm was calculated from amino acid sequence as described.[43] Similar measurements performed with a 1:1 mixture of $RecB^{\Delta nuc}$ and RecC in buffer C yielded an extinction coefficient of $\varepsilon_{280}=3.4\times10^5$ M^{-1} cm^{-1} for $RecB^{\Delta nuc}C$ in buffer C. All protein concentrations reported refer to the RecBC or $RecB^{\Delta nuc}C$ heterodimer or RecBCD heterotrimer. Bovine serum albumin (BSA) was from Roche (Indianapolis, IN) and its concentration was determined using an extinction coefficient of $\varepsilon_{280}=4.3\times10^4$ M^{-1} cm^{-1} in buffer C.[23] All proteins were dialyzed into the particular reaction buffer before use. Dialyzed RecBC, $RecB^{\Delta nuc}C$ 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.[44] 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 $MgCl_2$, at 25 °C and measuring the absorbance of the resulting mixture of mononucleotides at 260 nm as described.[45] The extinction coefficients at 260 nm used in this analysis are: 15340 M^{-1} cm^{-1} for AMP, 7600 M^{-1} cm^{-1} for CMP, 12160 M^{-1} cm^{-1} for GMP, 8700 M^{-1} cm^{-1} for TMP[46] and 5000 M^{-1} cm^{-1} 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 1(a)) was formed from strands 1 and 2 (Figure 1(e)); 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 1(e).

Analytical sedimentation equilibrium

Analytical sedimentation equilibrium experiments with $RecB^{\Delta nuc}C$ 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.[23] Sedimentation equilibrium was achieved within 24 h. Data were edited using WinREEDIT† 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†. The data were fit to equation (8), which describes the behavior for sedimentation of a single ideal species:

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

where A_T is the total absorbance at radial position r, A_{ref} is the absorbance of the single species at reference

† http://www.biotech.uconn.edu/auf/

radial position r_{ref}, M is the molecular mass of the single species, $\bar{v}$ is the partial specific volume calculated from protein sequences[47] (0.7341 ml g^{-1} at 25 °C for $RecB^{\Delta nuc}C$), ρ is the solvent density, ω 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 $RecB^{\Delta nuc}$ and RecC (with [$RecB^{\Delta nuc}C$] ranging from 90 to 230 nM) were performed in buffer M plus 10 mM $MgCl_2$, 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 $RecB^{\Delta nuc}C$ complex under these solution conditions.

Fluorescence titrations

Fluorescence titrations were performed as described[23] 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 (3 mL) (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 ($F_{i,corr}$) after the ith addition of protein and the initial corrected Cy3 fluorescence of the reference DNA ($F_{0,corr}$) were obtained as described[23] and they are related to the observed relative fluorescence change (ΔF_{obs}) as in equation (9):

$$\Delta F_{obs} = \frac{F_{i,corr} - F_{0,corr}}{F_{0,corr}} \tag{9}$$

ΔF_{obs} reaches its maximum value (ΔF_{max}) when both ends of the reference DNA are bound with protein. Hence $\Delta F_{obs}/\Delta F_{max}$ ($0 \le \Delta F_{obs}/\Delta F_{max} \le 1$) 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 ($2\Delta F_{obs}/\Delta F_{max}$) (see equation (11)).

Equilibrium binding of $RecB^{\Delta nuc}C$ to Cy3-labeled reference DNA

Equilibrium titrations of $RecB^{\Delta nuc}C$ 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.[23] In this model, $RecB^{\Delta nuc}C$, RecBC or RecBCD (hereinafter 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):

$$P = 1 + 2K_R B_f + K_R^2 B_f^2 \tag{10}$$

where B_f is the free concentration of protein.

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

$$\frac{B_{bound}}{D_T} = \frac{2K_R B_f}{1 + K_R B_f} = 2\frac{\Delta F_{obs}}{\Delta F_{max}} \tag{11}$$

where $B_{bound} = ([DB] + 2[B_2D])$, [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,[23] $\Delta F_{obs}/\Delta F_{max}$ 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):

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

Experimental fluorescence titrations, plotted as ΔF_{obs} *versus* [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 ΔF_{max}.

Competition methods to determine equilibrium binding to non-fluorescent DNA

Equilibrium constants for RecBCD, RecBC and $RecB^{\Delta nuc}C$ binding to non-fluorescent DNA molecules (N) were obtained from analysis of competition binding studies.[26] The analysis of competition titration data has been described in detail[23] and the same analysis is used here. 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 here has nearly identical ends (DNA **II** through **VI** series in Figure 1(a)), 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):

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

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 ΔF_{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 ΔF_{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 $RecB^{\Delta nuc}C$ 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 (± one standard deviation).

Binding polynomial for Scheme 2

The binding polynomial for Scheme 2 (Figure 8(b)) 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:

$$[BD] = [B_LD] + [B_ND] \quad (15)$$

$$[B_2D] = [(B_L)_2D] + [(B_N)_2D] + 2[B_LDB_N] \quad (16)$$

where [B_LD] is the concentration of DNA with only one of its ends bound by B_L to form a looped state and [B_ND] 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_LDB_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):

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

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

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

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

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

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[48] in the AMBER 8 molecular dynamics package‡.[49] The starting RecBCD-DNA structure (1W36)[15] was obtained from the Protein Data Bank.[50] The heavy atoms from Ser1122 of RecC and Arg607 from RecD that are missing in the structure were added using PDB2PQR§.[51] 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[51] was used to calculate the protonation states of titratable residues (Arg, Asp, Glu, Lys, His, Cys, Ser, and Thr) using the PROPKA method[52] 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^{-1} Å^{-1}). Solvent influences were included *via* the generalized Born implicit solvent model[53] 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,[15] 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/dsDNA 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.[54] The iterative thinning process was started with a probe-accessibility map constructed from the RecBCD structure using the Shrake–Rupley algorithm[55] as implemented in APBS‖[56] 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.[57,58] 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.[54] 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.[54]

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.[49] After each nucleotide addition, the energy of the extended part of the DNA was minimized as described above with a restraint of 10 kcal mol^{-1} Å^{-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 kcal mol^{-1} Å^{-1} restraints on all atoms outside a 10 Å radius of the phosphorus atom of the new

‡ http://amber.scripps.edu
§ http://pdb2pqr.sf.net
‖ http://apbs.sf.net/

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 (T.M.L.).

References

1. 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.
2. Lohman, T. M. & Bjornson, K. P. (1996). Mechanisms of helicase-catalyzed DNA unwinding. *Annu. Rev. Biochem.* **65**, 169–214.
3. von Hippel, P. H. & Delagoutte, E. (2001). A general model for nucleic acid helicases and their "coupling" within macromolecular machines. *Cell*, **104**, 177–190.
4. Lohman, T. M., Hsieh, J., Maluf, N. K., Cheng, W., Lucius, A. L., Brendza, K. M. *et al.* (2003). DNA helicases, motors that move along nucleic acids: lessons from the SF1 helicase superfamily. *The Enzymes*, **23**, 303–369.
5. Gray, M. D., Shen, J. C., Kamath-Loeb, A. S., Blank, A., Sopher, B. L., Martin, G. M. *et al.* (1997). The Werner syndrome protein is a DNA helicase. *Nature Genet.* **17**, 100–103.
6. 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. *Nucl. Acids Res.* **25**, 2973–2978.
7. Watt, P. M. & Hickson, I. D. (1996). Failure to unwind causes cancer. Genome stability. *Curr. Biol.* **6**, 265–267.
8. van Brabant, A. J., Stan, R. & Ellis, N. A. (2000). DNA helicases, genomic instability, and human genetic disease. *Annu. Rev. Genomics Hum. Genet.* **1**, 409–459.
9. 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**, 401–465.
10. Smith, G. R. (2001). Homologous recombination near and far from DNA breaks: alternative roles and contrasting views. *Annu. Rev. Genet.* **35**, 243–274.
11. 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.
12. 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**, 4981–4987.
13. Dillingham, M. S., Spies, M. & Kowalczykowski, S. C. (2003). RecBCD enzyme is a bipolar DNA helicase. *Nature*, **423**, 893–897.
14. 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.
15. 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**, 187–193.
16. 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. USA*, **81**, 7850–7854.
17. 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.
18. 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.
19. 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**, 3447–3454.
20. 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.
21. 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**, 431–443.
22. Korangy, F. & Julin, D. A. (1993). Kinetics and processivity of ATP hydrolysis and DNA unwinding by the RecBC enzyme from *Escherichia coli*. *Biochemistry*, **32**, 4873–4880.
23. 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**, 765–782.
24. 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. USA*, **95**, 981–986.
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**, 265–275.
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**, 2117–2128.
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**, 571–581.
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)35 and (SSB)56 DNA binding modes. *Biochemistry*, **27**, 2260–2265.
29. Lohman, T. M. & Bujalowski, W. (1991). Thermodynamic methods for model-independent deter-

mination of equilibrium binding isotherms for protein-DNA interactions: spectroscopic approaches to monitor binding. *Methods Enzymol.* **208**, 258–290.
30. Lohman, T. M. & Mascotti, D. P. (1992). Nonspecific ligand-DNA equilibrium binding parameters determined by fluorescence methods. *Methods Enzymol.* **212**, 424–458.
31. Taylor, A. & Smith, G. R. (1980). Unwinding and rewinding of DNA by the RecBC enzyme. *Cell*, **22**, 447–457.
32. Braedt, G. & Smith, G. R. (1989). Strand specificity of DNA unwinding by RecBCD enzyme. *Proc. Natl Acad. Sci. USA*, **86**, 871–875.
33. Taylor, A. F. & Smith, G. R. (2003). RecBCD enzyme is a DNA helicase with fast and slow motors of opposite polarity. *Nature*, **423**, 889–893.
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**, 37078–37087.
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**, 469–479.
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-eference Book of Chemical and Physical Data* (Weast, R. C., eds), 55th edit., 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**, 731–750.
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**, 409–428.
40. Chaudhury, A. M. & Smith, G. R. (1984). *Escherichia coli* recBC deletion mutants. *J. Bacteriol.* **160**, 788–791.
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)**, 151–160.
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**, 10139–10147.
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**, 7596–7610.
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**, 8409–8422.
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. *et al.* (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. *et al.* (2004). *AMBER 8,* University of California, San Francisco.
50. Deshpande, N., Addess, K. J., Bluhm, W. F., Merino-Ott, J. C., Townsend-Merino, W., Zhang, Q. *et al.* (2005). The RCSB Protein Data Bank: a redesigned query system and relational database based on the mmCIF schema. *Nucl. Acids Res.* **33**, D233–D237.
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. *Nucl. Acids Res.* **32**, W665–W667.
52. Li, H., Robertson, A. D. & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct. Funct. Bioinformat.* **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: Struct. Funct. Genet.* **55**, 383–394.
54. Ju, T., Baker, M. L. & Chiu, W. (2006). Computing a family of skeletons of volumetric models for shape description. In *Lecture Notes in Computer Science (Proceedings of GMP 2006)* (Kim, M.-S. & Shimada, K., eds), vol. 4077, pp. 235–247.
55. Shrake, A. & Rupley, J. A. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351–371.
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. USA*, **98**, 10037–10041.
57. Bertrand, G. (1995). A parallel thinning algorithm for medial surfaces. *Pattern Recognit. Letters*, **16**, 979–986.
58. Borgefors, G., Nystrom, I. & Di Baja, G. S. (1999). Computing skeletons in three dimensions. *Pattern Recognit.* **32**, 1225–1236.

Edited by D. E. Draper

(Received 24 April 2006; received in revised form 6 July 2006; accepted 7 July 2006)
Available online 15 July 2006