



# HHS Public Access

Author manuscript

*Phys Rev E Stat Nonlin Soft Matter Phys.* Author manuscript; available in PMC 2016 August 04.

Published in final edited form as:

*Phys Rev E Stat Nonlin Soft Matter Phys.* 2013 March ; 87(3): . doi:10.1103/PhysRevE.87.032702.

## Tension on dsDNA bound to ssDNA-RecA filaments may play an important role in driving efficient and accurate homology recognition and strand exchange

Julea Vlassakis, Efraim Feinstein, Darren Yang, Antoine Tilloy, Dominic Weiller, Julian Kates-Harbeck, Vincent Coljee, and Mara Prentiss

Harvard University, Department of Physics, Cambridge, MA, 02138\*

### Abstract

It is well known that during homology recognition and strand exchange the double stranded DNA (dsDNA) in DNA/RecA filaments is highly extended, but the functional role of the extension has been unclear. We present an analytical model that calculates the distribution of tension in the extended dsDNA during strand exchange. The model suggests that the binding of additional dsDNA base pairs to the DNA/RecA filament alters the tension in dsDNA that was already bound to the filament, resulting in a non-linear increase in the mechanical energy as a function of the number of bound base pairs. This collective mechanical response may promote homology stringency and underlie unexplained experimental results.

### I. INTRODUCTION

Sexual reproduction and DNA damage repair often include homologous recombination facilitated by RecA family proteins [1, 2]. In homologous recombination, a single stranded DNA molecule (ssDNA) locates and pairs with a sequence matched double-stranded DNA molecule (dsDNA). In the first step of the process, the incoming ssDNA binds to site I in RecA monomers, resulting in a helical ssDNA-RecA filament with 3 base pairs/monomer and ~ 6 monomers/helical turn [3]. This helical ssDNA-RecA filament then searches dsDNA molecules for homologous sequences by rapidly binding and unbinding dsDNA to site II in RecA [3]. Thus, the sequence of the ssDNA in the searching filament is fixed, and the system then searches through the available dsDNA to find a sequence match for that ssDNA. The binding of dsDNA to site II is very unstable, so if the dsDNA is not homologous to the ssDNA bound to site I, the dsDNA rapidly unbinds from the ssDNA-RecA filament. If the dsDNA is homologous, strand exchange should occur, probably via base-flipping that transfers the Watson-Crick pairing of the complementary strand from the outgoing DNA strand bound to site II to the incoming DNA strand bound to site I [4]. This strand exchange reduces the unbinding rate for the dsDNA [5]. RecA is an ATPase, but *in vitro* homology recognition and strand exchange can occur without ATP hydrolysis [6–8]. Thus, each step in the homology search/strand exchange process is fully reversible.

\*Electronic address: prentiss@fas.harvard.edu

During the homology search and strand exchange process, dsDNA bound to RecA is extended significantly beyond the B-form length [9]. Recent theoretical work proposed that the free energy penalty associated with extension may promote rapid unbinding of non-homologous sequences, but the free energy penalty was assumed to be a linear function of the number of bound triplets and the kinetic trapping due to near homologs was not considered [10]. Earlier work had also suggested that the dsDNA extension promotes base-flipping [11] and reduces kinetic trapping since the lattice mismatch between extended dsDNA and B-form dsDNA presents a steric barrier to interactions between unbound dsDNA and bound ssDNA which implies that the dsDNA must bind to the filament in order to interact with the ssDNA [12]. These studies assumed the dsDNA in the DNA/RecA filament is uniformly extended; however, the X-ray crystal structures of the dsDNA in the final post-strand exchange state and the ssDNA in the homology searching state both consist of base pair triplets in a nearly B-form conformation separated by large rises as illustrated in Figure 1a. The rises occur at the interfaces between adjoining RecA monomers [3], as illustrated in Fig. 2. The functional role of the non-uniform extension has been unclear.

In this paper we present a simple model that calculates the extension of each base pair triplet in a dsDNA. Using this model, we calculate the free energy changes associated with progression through the homology recognition/strand exchange process. The results of that calculation suggest a resolution to the long standing question of why strand exchange is free energetically favorable even though the Watson-Crick pairing in the initial and final states is the same and the DNA/protein contacts in the ssDNA-RecA filament and final post-strand exchange state are nearly the same.[3] The model also makes several significant qualitative predictions, the most significant being the suggestion that the collective behavior of the triplets due to their attachment to the phosphate backbones leads to a free energy that is a non-linear function of the number of consecutive bound triplets. As a result of this non-linearity, total binding energy has a minimum as a function of the number bound triplets in a given conformation. After that minimum is reached, adding more triplets that given conformation becomes free energetically unfavorable.

Such a change in sign in the binding energy as a function of the number of bound triplets can never occur in a theory where the energy is a purely linear function of the number of bound triplets since the binding of any base pair anywhere in the system is equally likely regardless of the state of any of the other triplets in the system. In a system with a binding energy that is a linear function of the number of correctly paired bound triplets, if homologous triplets can initially bind to the system, then additional homologous triplets will always continue to bind. Thus, binding will readily progress across a non-homologous triplet. As we will discuss in detail in this work, in a system with a linear energy and more than  $\sim 4$  binding sites, either homologs will be too unstable or near homologs will be too stable.

In contrast, the non-linearity may provide more rapid and accurate homology recognition than is available for systems using linear energies because the non-linearity requires that dsDNA binding to the ssDNA-RecA proceed iteratively triplet by triplet through a series of checkpoints which inhibit the progression of strand exchange past a non-homologous triplet. At each checkpoint, the progression of strand exchange to more stable binding conformations is only free energetically favorable if a sufficient number of contiguous

homologous base pair triplets are bound to the ssDNA/RecA filament in the appropriate conformations. The general qualitative features of the homology recognition based on the non-linear energy follow from basic properties of the simple model and are insensitive to the parameters chosen. These features include the following: 1. that homology recognition will proceed iteratively through consecutive triplets 2. that strand exchange reversal is much more favorable at the ends of the filament than at the center 3. that there will be two checkpoints that cannot be passed unless the bound dsDNA contains a sufficient number of contiguous homologous base pairs in the appropriate conformations. Though the general features of the model are very robust, the exact number of contiguous homologous bp required to progress past a particular checkpoint depends strongly on the choice of model parameters. Analytical modeling and numerical simulations suggest that there are a small range of parameters that allow the free energies predicted by the model presented in this paper to provide homology recognition which is both fast and accurate [13, 14].

Simulation results suggest that though the initial binding of  $\sim 9$  base pairs (bp) is free energetically accessible, adding more bound triplets is not favorable, and adding more than 15 bp is enormously unlikely unless the first checkpoint is passed. [14] The first major checkpoint requires that  $\sim 9$  of the  $\sim 15$  bp that initially bind to the filament are contiguous and homologous. If the initial  $\sim 15$  bp do not contain  $\sim 9$  contiguous homologous base pairs, the dsDNA cannot make a transition to the more stably bound intermediate state; therefore, the weakly bound dsDNA will almost immediately unbind from the filament. This checkpoint rapidly rejects all but 50 of the  $\sim 10,000,000$  possible binding positions in a bacterial genome. If the initial  $\sim 15$  bp do include  $\sim 9$  contiguous homologous base pairs, the system can make a transition to a metastable intermediate state, which allows more base pairs to be added to the filament. The second major checkpoint occurs when  $\sim 18$  contiguous bp are bound to the filament in the metastable intermediate state. If all of the bp are contiguous and homologous, the system can make a transition to the final post-strand exchange state. Otherwise, the long regions of accidental homology will slowly reverse strand exchange and unbind. Given the statistics of bacterial genomes, passing the homology requirement for the second checkpoint would guarantee that the correct match had been found. These predictions are in good agreement with known experimental results that measure the stability of strand exchange products as a function of the number of contiguous bound base pairs [15].

In this work, we will not attempt to optimize the parameters of the model in order to provide rapid and accurate homology recognition. Rather, we will consider why homology recognition systems in which the energies are a linear function of the number of bound base pairs can either provide rapid unbinding of non-homologs or stable binding of complete homologs, but not both if the number of binding sites in the system is  $>\sim 4$ . We will then discuss how qualitative features of the non-linear free energy predicted by the model allow strand exchange to avoid kinetic trapping in near mismatches while also permitting homologs to progress completely to strand exchange in systems where the number of binding sites is  $> 4$ .

## A. General Issues in Self-Assembly Based on the Pairing of Arrays of Matching Binding Sites

In efficient self-assembly/recognition systems that create correct assembly by matching linear arrays of binding sites, correctly paired arrays of binding sites must remain stably bound, whereas incorrectly paired arrays must rapidly unbind even if the incorrect pairing contains only one single mismatched binding site. For a system in thermodynamic equilibrium, the populations in different binding configurations are determined simply by their binding energies. Thus, in a system where every array consists of  $N$  binding sites if  $U(m, N)$  is the binding energy when  $m$  of the binding sites are correctly paired, accurate recognition requires that  $\exp[-(U(m, N) - U(N, N))/(kT)] \ll 1 \forall m < N$ , where  $k$  is Boltzmann's constant and  $T$  is the Kelvin temperature. Furthermore, in a system with a temperature  $T$ , the requirement that mismatched pairings rapidly unbind implies that  $U(m, N) > -kT \forall m < N$ , whereas if the correctly bound ones are to remain stably bound  $U(N, N)$  must be  $\ll -kT$ . If the energy is a linear function of the number bound and a mismatch contributes zero free energy then  $U(N, N) = -N\epsilon kT$  and  $U(N-1, N) = -(N-1)\epsilon kT$ , so the condition  $\exp[-(U(N-1, N) - U(N, N))/(kT)] \ll 1$  requires  $\epsilon \gg 1$ , which also implies that the homolog will remain stably bound. For simplicity consider  $\epsilon = 3$ , which implies  $U(N, N) = -N\epsilon kT = -3NkT$ . Substitution into the requirement for the unbinding of near homolog requires that and  $U(N-1, N) = -(3N-3)kT > -kT \Rightarrow N < 4/3$ . Thus, the requirements for rapid and accurate recognition can only be met if  $N = 1$ . The requirements become more stringent if the specificity ratio is more strict than  $1/20$ . As we will discuss below, accurate homology recognition in a bacterial genome requires accurate recognition over a length of more than 12 bp, which implies  $N > 4$  since 12 base pairs is 4 triplets.

Some of the problems with realizing accurate recognition in systems at thermodynamic equilibrium were recognized by John Hopfield in the 1970's, inducing him to propose a kinetic proofreading system that requires an irreversible process. [16] In such systems, the energy of the bound state can be very deep without making the energy of the searching state deep because of the irreversible step that transfers the system from the searching state to the bound state. Even in Hopfield's system there is a tradeoff between search speed and accuracy since greater sequence discrimination requires greater unbinding probabilities for homologs. The increased binding probability for homologs increases the searching time because the correct binding site must be revisited many times before the homolog makes the irreversible transition to the bound state. Earlier work had proposed that RecA based homology recognition could proceed via kinetic proofreading, [5], but homology recognition *in vitro* is known to proceed without an irreversible step. [6-8] In this work, we will consider the non-linearities in the free energy as a function of the number of contiguous bound base pairs that arise as a function of the differential extension of dsDNA bound to RecA and how those non-linearities can promote rapid and accurate homology recognition without requiring irreversibility by making transitions between successive bound states contingent on the state and relative position of the bound base pairs.

## B. Homology Recognition and Strand Exchange are Likely to Proceed in Units of Base pair Triplets

Previous experimental results suggest that homology recognition occurs via base flipping of the complementary strand bases between their initial pairing with the outgoing strand and their final pairing with incoming strand. [5, 17, 18]. Experimental results have already shown that strand exchange does progress in triplets. [19] We propose that within each triplet the nearly B-form structure preserves sufficient stacking to allow homology discrimination to exploit the energy difference between the Watson-Crick pairing of homologs and heterologs, while the large rises between triplets result in mechanical stress that plays several important functional roles in homology recognition and strand exchange. Furthermore, if the stacking makes it free energetically favorable for the triplets to flip as a group, the lost of Watson-Crick pairing due to a single mismatch in a triplet may make strand exchange unfavorable for the entire triplet, as would be the case if Watson-Crick pairing alone determined the free energy of the strand exchanged state. If a single mismatch makes strand exchange of a triplet unfavorable, then testing in triplets implies that in a random sequence the possibility of an accidental mismatch is  $1/64$ , whereas if the search were done by comparing single base pairs the probability of an accidental match would be  $1/4$ . Homology stringency and searching speed would be greatly reduced if the search were not done in triplets. Once a sufficient number of base pair triplets have undergone strand exchange, the complementary strand backbone relocates to the position shown in the X-ray structure of the final state [5, 18], as illustrated in Figures 1c and 2.

## C. General Model for dsDNA Bound to a RecA Filament

Certain underlying assumptions and qualitative features are important to the model proposed in this paper: 1. that in the absence of hydrolysis the extension of the protein filament is unaffected by dsDNA binding [20] 2. that homology recognition and strand exchange occur in quantized units of base pair triplets [3, 19] 3. that the incoming and outgoing strands consist of nearly B-form triplets separated by large rises where the strand is bound to the filament due to strong charged interactions between the backbone phosphates and positive residues in the protein [3, 18, 21] 4. that the complementary strand is bound to the filament dominantly via Watson-Crick pairing, resulting in large stress on the base pairs unless the dsDNA is in the final post-strand exchanges state where the L1 and L2 loops provide significant mechanical support. [3, 22]

The model utilizes work presented by deGennes that calculated the force required to shear dsDNA [23]. We extend deGennes' model to the triplet structures in the initial, intermediate, and final dsDNA conformations. In the model the actual three dimensional helical structure is converted into a one dimensional system. In the simple one dimensional model,  $L_R$  is rise between the triplets in the incoming and outgoing strands and a single variable,  $\gamma$ , characterizes the equilibrium spacing between phosphates when the complementary strand is in a particular state. Thus, for a particular state, the difference between the equilibrium spacing is given by  $(1 - \gamma)L_R$ .

## D. Predicted Extension and Energy

As shown in Fig. 3, the extensions of rises between base pair triplets in a strand bound directly to the protein are given by  $v_{N,i}$  and for a system with  $N$  triplets bound RecA,

$$v_{N,i} = iL_R. \quad (1)$$

The  $u_{N,j}$  specify the extensions of the rises in the complementary strand. At equilibrium, the net force on each  $u_{N,j}$  must be zero; therefore, for  $j = 2$  to  $N$

$$Q((u_{N,j+1} - u_{N,j} - \gamma L_R) - (u_{N,j} - u_{N,j-1} - \gamma L_R)) + R(u_{N,j} - v_{N,j}) = 0, \quad (2)$$

where  $R$  and  $Q$  are the spring constants for the base pairs and the backbones, respectively. These values for  $R$  and  $Q$  may be substantially different from those for individual dsDNA base pairs when the dsDNA is not bound to RecA because of the interactions between the charged phosphates and the protein and because dsDNA is grouped in triplets where the stacking between the triplets is strongly disrupted; however, it is still likely that  $R \ll Q$  as it is in naked dsDNA since the interactions between the bases on opposite strand is significantly weaker than the interaction between the phosphates in the backbone of the same strand.

The boundary condition on the last triplet  $u_{N,1}$  requires that

$$+Q(u_{N,2} - u_{N,1} - \gamma L_R) + R(u_{N,1} - v_{N,1}) = 0 \quad (3)$$

The values of  $u_{N,i}$  can be found using Equations ??-??. The angle between base pairs and the DNA helical axis is shown as  $\theta_{bp}$  in Fig. 3.

In the continuous limit where the discrete subscript  $i$  is replaced by a continuous variable  $x$ , the equations have an analytical solution

$$v_{N,x} = x + A_N(x) \sinh(\chi x / L_R) \quad (4)$$

where  $\chi = \text{Sqrt}[R/(2Q)]$  is the deGennes length for RecA bound dsDNA and the constant  $A$  is found by using the boundary conditions for the ends, yielding

$$A_N(x) [R \sinh(\chi x / L_R) + Q \chi \cosh(\chi x / L_R)] = Q(\gamma - 1)L_R. \quad (5)$$

In the limit where  $1/(2\chi) \gg 1$

$$A_N(x) = \frac{(\gamma-1)L_R}{\cosh(\chi N/2)}. \quad (6)$$

These assumptions and features lead to a nuanced picture of the distribution of tension during strand exchange. The lattice mismatch between the complementary strand and its pairing partners is largest at the ends of the filament as shown in Fig. 4; therefore, the base pair tension is largest at the end of the filament. Furthermore, the lattice mismatch at the ends increases significantly with the number of bound triplets as shown in Fig. 4 and 5. The angle  $\theta_{bp}$  is greatest at the ends of the molecule. When fewer than  $\sim 30$  bp are bound to the filament, the tension on the base pairs at the ends increases rapidly as more RecA bound triplets are added, as shown in Fig. 4 and 5; however,  $\theta_{bp}$  and the tension on the base pairs near the center decreases as more triplets are added, as shown in Fig. 4 and 5. In contrast with the tension on the base pairs, which is largest at the ends of the complementary strand, the tension on the rises in the complementary strand is largest at the center and smallest at the ends. These general qualitative features are not sensitive to  $R$ ,  $Q$  or  $\gamma$  as long as  $R \ll Q$ .

Given the values of  $u_{N,i}$  the mechanical energy of the system, which is of the form  $1/2k(x-x_0)^2$ , may be calculated from:

$$E_{mech}(N) = \left[ \sum_{i=1}^N \frac{1}{2} R (u_{N,i} - v_{N,i})^2 + \sum_{i=2}^N \frac{1}{2} Q (u_{N,i} - u_{N,i-1} - \gamma L_R)^2 \right] \quad (7)$$

When  $R \ll Q$  this expression simplifies to:

$$E_{mech}(2) = 2 \left[ \frac{1}{2} R \frac{(L_R(1-\gamma))^2}{2} \right] \quad (8)$$

$$E_{mech}(3) = 2 \left[ \frac{1}{2} R (L_R(1-\gamma))^2 \right] \quad (9)$$

These energy terms represent the stress on base pairs at the ends of the molecule due to the lattice mismatch. When  $R \ll Q$ , for moderate numbers of bound base pairs, this energy is stored primarily in the extended base pairs. This idea is particularly important because it implies that an increase in the extension of the rises in the complementary strand can reduce the free energy by reducing the tension on the base pairs even if the increase in extension of the rises requires some energy. Similarly, in the RecA structure the transition from the intermediate state to the final state may reduce the tension on the base pairs because the interactions between the base pairs and the L1 and L2 loops may increase the equilibrium extension of the rises in the complementary strand.

Using the continuous limit allows us to generate scaling laws for the extension as a function of the total number of bound base pairs. In the continuous limit, the non-linear contribution to the free energy is given by

$$E_{non-linear}(N) = \frac{1}{2} \frac{(1-\gamma)^2 L_R^2}{\chi^2 \cosh^2[\chi N/2]} \sum_{i=-N/2}^{i=N/2} \sinh^2[\chi i]. \quad (10)$$

Thus, when  $\chi \ll 1$  the non linear energy term has the following scaling

$$E_{non-linear}(N) \propto (1-\gamma)^2 L_R^2 \frac{\sinh(\chi N) - \chi N}{1 + \cosh^2[\chi N/2]}. \quad (11)$$

In the limit where the number of base pairs bound is much less than the deGennes length so  $\chi^N \ll 1$  the energy the non-linear energy scaling is

$$E_{non-linear}[N] \propto (1-\gamma)^2 L_R^2 \chi^2 N^2. \quad (12)$$

Thus, when  $N$  is small the energy increases as the square of the number of bound triplets, consistent with the exact results for the discrete case given in equations ??,?? and ?. In contrast, when  $N$  is larger than the deGennes length, the non-linear energy term approaches zero and the mechanical energy increases linearly with increasing  $N$ . In this limit, the base pairs at the center of the filament are no longer under tension. Thus, adding a triplet to the end of the filament effectively adds another triplet to the unstressed center rather than increasing the stress on all of the bound triplets.

The total energy of the system includes the mechanical energy calculated above and the non-mechanical binding energy per RecA monomer,  $E_{bind}$ . Assuming the free energy of an unbound dsDNA is zero and the free energy gain upon binding a triplet is independent of  $N$  and  $i$  then the non-mechanical contribution to the binding energy for  $N$  triplets is  $NE_{bind}$ . When the first RecA monomer binds,  $E_{total}[1] = E_{bind}$  which is a constant negative value. In contrast, when  $N > 1$ , the stress on the molecule yields a total energy of

$$E_{total}[N] = E_{mech}[N] + N E_{bind} \quad (13)$$

which changes in sign and magnitude depending on  $L_R$ ,  $R$ ,  $Q$  and  $\gamma$ .

## E. Modeling the Homology Search and Strand Exchange Process

Recent experimental results suggest that the homology recognition/strand exchange process uses four major dsDNA conformations: 1. B-form dsDNA, which is the structure of the



dsDNA when it is not bound to the protein 2. an initial sequence independent searching state with the dsDNA bound to the RecA where the complementary strand bases paired with outgoing strand 3. an intermediate sequence dependent strand exchanged state in which the complementary strand bases are paired with the incoming strand where the complementary strand backbone is in a position where its bases can flip between pairing with the incoming and complementary strands 4. the final state known from the X-ray structure where the complementary strand bases are paired with the incoming strand bases and the phosphate backbone is in the position shown in the x-Ray structure. [18]

Consistent with strand exchange proceeding through base flipping of triplets, recent experimental results have suggested that outgoing strand bases are arranged in B-form triplets separated by large rises such that the complementary strand bases can readily rotate between pairing with the outgoing strand and pairing with the incoming strand. [21]. The crystal structure shows that the incoming strand is located near the center of the helical DNA/protein structure whereas the residues associated with the binding of the outgoing strand are much farther away from the center, as illustrated in Fig. 1; however, if strand exchange occurs via the base flipping of base pair triplets, the spacing within triplets must be approximately the same for all three strands. Thus, given that the total extension of the outgoing strand backbone is much larger than the total extension of the incoming strand backbone, the rises in the outgoing strand must be much larger than the rises in the incoming strand, as illustrated in Fig. 2.

In the one dimensional model considered here where  $\gamma$  is the only parameter characterizing each dsDNA conformation,  $\gamma$  is smallest for the initial searching state, where the complementary strand is paired with the very highly extended outgoing strand, consistent with experimental results that the dsDNA in the initial bound state has a large differential extension between the outgoing and complementary strands that prevents more base pairs from binding to the filament unless the dsDNA undergoes strand exchange. [21] The  $\gamma$  for the intermediate state is slightly larger because the complementary strand is paired with the less extended incoming strand. This value is consistent with experimental results that suggest that the strand exchange of homologous triplets is favorable because it reduces the differential tension on the dsDNA. [18] The  $\gamma$  for the final state is almost 1, because of the support provided for the rises by the L1 and L2 loops of the protein that occupy the rises in the complementary strand, consistent with the known experimental result that the binding of at least  $\sim 80$  bp in the final state is free energetically favorable. [22] In order to probe the role of tension during homology recognition strand exchange we apply the model to the various stages of strand exchange using the following parameters:  $\gamma = 0.4, 0.75, 0.8, 0.9875$  and  $E_{bind}$  equal to  $0, -0.75kT, -0.8kT$  and  $-0.125kT$  per homologous triplet in the unbound, initial bound, intermediate and final states, respectively. These  $\gamma$  values are inspired by X-ray structure DNA RecA filaments, the known properties of B-form dsDNA, experimental results on the stability of strand exchange products [15], and results of numerical simulations that optimize homology recognition. [14]

## II. RESULTS AND DISCUSSION

### A. Changes in Mechanical Tension Allows Strand Exchange to be Free Energetically Favorable if and only if 6 Contiguous Homologous bp Undergo Strand Exchange Insensitive to Model Parameters

Early RecA recognition systems assumed that strand exchange is always free energetically favorable, just as Watson-Crick pairing of unpaired ssDNA is always favorable, but this assumption is not correct. In self-assembly based on ssDNA/ssDNA pairing, the pairing of matched bases is free energetically favorable and the pairing of mismatched bases is approximately neutral. Thus, in ssDNA/ssDNA pairing systems, such as DNA origami, assembly of matching sites is always free energetically favorable because the correct Watson-Crick pairing reduces the free energy below that for the system where no bases were bound. In contrast, if only Watson-Crick pairing is considered, self-assembly based on strand exchange is free energetically neutral for matched bases and free energetically unfavorable for mismatched bases because the system begins in a state with all of the bases in the complementary strand correctly paired with the corresponding bases in the outgoing strand. Thus, previous models of strand exchange that considered only the Watson-Crick pairing were faced with a paradox: at thermal equilibrium strand exchange is only free energetically neutral, so at best only 50% of correct pairings would end up in the strand exchanged state. The other 50% would end up paired with their initial partners; however, *in vivo* strand exchange proceeds to completion. Including protein contacts did not solve the problem since the protein contacts in the initial ssDNA-RecA filament are almost identical to the contacts in the final post-strand exchange state [3]; however, the model presented in this paper resolves the problem, as discussed below.

In the model considered here, the coupling along the dsDNA backbone makes any change in the dsDNA conformation (such as adding or base-flipping a triplet) alter the positions and extensions of all other bound triplets. This effect is illustrated in Fig. 4, which shows the rises and extensions of the base pairs calculated from the model when an additional triplet binds in site II of the protein when all of the bound triplets are bound in site II. The triplet highlighted with the arrow experiences increasing tension as triplets are added to the end of the sequence. Furthermore, as shown in Fig. 5, the lattice mismatch at the ends of the filament increases monotonically as more triplets are added. In contrast, the lattice mismatch for the first triplet out from the center of the sequence decreases as base pairs are added (solid black line) because the added triplets extend the inner rises. As the number of bound triplets approaches the deGennes length for RecA bound dsDNA, the lattice mismatch at the ends of the filament approaches a constant and the central rises in the complementary strand approach  $L_R$ .

The mechanical stress model considered here suggests strand exchange is free energetically favorable if and only if strand exchange transfers a rise because the reduction in free energy is the result of the reduction in mechanical stress due to a reduction in the stress on the base pairs due to decrease in the lattice mismatch between the complementary strand and its Watson-Crick pairing partner. Thus, when the dsDNA first binds to the searching filament and all of the triplets are in the initial state, base flipping of a single triplet does not lower

the energy of the system since a rise is not transferred even if that triplet is perfectly homologous; however, once one homologous triplet is strand exchanged, it becomes favorable to transfer a contiguous homologous triplet. Thus, the transfer of the Watson-Crick pairing of the complementary strand from the outgoing strand to the incoming strand is not free energetically favorable except in sequence regions containing at least six base pairs of contiguous homology. This minimum number of contiguous homologous bases required to pass the first checkpoint is a basic qualitative feature of the model that is extremely insensitive to the model parameters chosen. This first checkpoint can provide very rapid unbinding of almost all initial dsDNA pairings, which is highly advantageous for rapid searching. Though the model specifies that 6 contiguous bases is the minimum number required to pass the checkpoint insensitive to the parameters used in the model, the actual number of base pairs required to pass the first checkpoint is sensitive to parameters.

Experimental results suggest that strand exchange is only marginally stable when  $\sim 9$  contiguous homologous bp undergo strand exchange, [15] suggesting that rapid unbinding will occur for all tested sequences except for  $1/4^9 \sim 2 \times 10^{-6}$  which represents only  $\sim 50$  possible positions for a bacterial genome with a length of 10,000,000 bp. All other sequences will rapidly unbind from the RecA filament because the binding energy for sequence independent searching state is very weak when only  $\sim 9$  bp are bound and adding more base pairs to the searching state would increase rather than decrease the binding energy.

## B. Calculations of the Free Energy as Function of the Number of Bound Triplets in Each dsDNA Conformation

In order to understand the progression through homology recognition and strand exchange, it is important to consider the free energies for all of the dsDNA conformations involved, not just the initial bound state and the intermediate state. Of course the process is kinetic, so the energies of the transition states play a vital role; however, for simplicity we will only show the energies of the various bound conformations using  $\gamma$  values based on numerical simulations. [14] Fig. 6 shows  $E_{total}$  as a function of number of base pairs in the initial bound state, the intermediate state and the final post-strand exchange state when all of the triplets are in the same conformation. Except for the initial binding of  $\sim 9$  to 15 bp, all of the triplets are rarely in the same conformation. Thus, the free energy curves for states with all of the triplets same conformation rarely represent the free energy of the system; however, it is clear that reductions in the mechanical energy of the bound dsDNA can drive strand exchange for homologs, as we discuss below.

## C. Non-linearities in the Free Energy Allows Homologs to Progress to Complete Strand exchange

While it is energetically favorable to add base pairs to the initial bound state for small numbers of base pairs, the quadratic term of  $E_{total}$  from Equation ?? rapidly increases as a function of increasing number of base pairs, making the binding of a large number of base pairs in the initial state unfavorable. This is because of the significant tension due to the lattice mismatch between the complementary strand and the outgoing strand; consequently, for the parameters based on simulation, no more than  $\sim 15$  bp can bind to the filament in the

initial searching state because the energy required greatly exceeds  $kT$ . Thus, once  $\sim 15$  bp are bound to the filament, the system is in a highly free energetically unfavorable state which will force it to choose between the following: 1. unbinding from the filament 2. strand exchange, which is unfavorable for non-homologs. Again, the prediction that there will be an checkpoint in the progression of strand exchange from the initial bound state to the intermediate state is insensitive to the model parameters. The parameters only determine whether more than 6 contiguous homologous base pairs are required to progress past the checkpoint.

Homologs can rapidly progress to complete strand exchange if the weak initial binding holds long enough for  $\sim 9$  homologous contiguous base pairs to undergo strand exchange, which stabilizes the binding for those homologous bp and allows strand exchange to progress. [13, 14] The non-linearity in the free energy makes the strand exchange of consecutive homologous triplets increasingly favorable as long as the number of bound base pairs is  $< \sim 30$ ; consequently, the non-linearity in the free energy makes strand exchange reversal more improbable as a more contiguous homologous base pairs are strand exchanged. Furthermore, the non-linearity makes strand exchange at the center of the filament increasingly unfavorable as the number of bound triplets increases as shown in Fig. 8, while still allowing strand exchange reversal to remain possible at the ends of the filament. Again, these qualitative features are basic properties of the model that are highly insensitive to the model parameters. These qualitative features allow true homologs to progress to complete strand exchange even though non-homologs readily unbind. In contrast, for a system with a linear free the probability that strand exchange will be reversed for a given triplet is independent of the number of other triplets bound and of the position of the particular triplet in the filament. As a result, such systems either suffer rapid unbinding of homologs or strong kinetic trapping in near homologs, as discussed above for the general case of a system with a linear binding energy and greater than 4 binding sites.

#### D. dsDNA Tension Drives the Transition from the Intermediate State to the Final State

The free energy penalty due to binding in the intermediate state is so large that even a perfect homolog cannot bind more than  $\sim 18$  bp in the intermediate state. The only way to continue to add base pairs to the filament is to make a transition to the final post-strand exchange state where the L1 and L2 loops provide significant mechanical support for the rises. This transition reduces the tension sufficiently to allow more triplets to bind in the initial bounds state. If the transition does not occur, strand exchange will reverse because the intermediate state energy is unfavorable. Thus, the collectivity in the behavior of the dsDNA can have a significant effect in enforcing homology stringency at this final step in the strand exchange process by enforcing the following: 1. that the number of base pairs that can be bound in the intermediate state is limited 2. that transfer of a single triplet from the intermediate state to the final state is never favorable 3. that transferring non-contiguous triplets from the intermediate state to the final state is unfavorable even if they are homologs 4. that the transfer of pairs of triplets is not favorable until  $\sim 18$  bp are bound in the intermediate state. Properties 1–3 are shared with the transition from the initial bound state to the intermediate state, but the fourth property is different. It may arise from some combination of two features: 1. the linear energy in the final state may be less favorable than

the linear energy in the intermediate state 2. there is a significant boundary penalty associated with the deformation of the backbone that occurs when the system is partially in the intermediate state and partially in the final state. In either case, then the transition from the intermediate state to the final state will only become free energetically favorable when the favorable non-linear term becomes dominant over the other terms. If, the final state only becomes free energetically favorable when  $\sim 15$  contiguous base pairs are present in the final state, then the non-linearity in the energies of the intermediate and final states in the energy may provide additional discrimination against regions of accidental homology since for a particular given searching ssDNA sequence the odds of this occurring with a given searching sequence are  $\sim 1.4 \times 10^{-11}$ /base pair. Thus, even in a 10 million bp genome, the probability that such an accidental homology is present in a given bacterial genome is  $\sim 1/10^4$ . We have considered a few bacterial genomes and found that the sequences are indeed random in this sense, with the exception of repeated genes. Thus, many bacterial genomes contain no accidental mismatches consisting of 18 contiguous bp; therefore, no further homology checking is required if 18 contiguous bp exactly match. The important statistical property is consistent with experimental results that strand exchange products do not become stable until more than  $\sim 18$  bp have undergone strand exchange.

#### **E. For dsDNA in the final state the free energy has a minimum as a function of $N$**

If the final state has a much higher  $\gamma$  than the intermediate state, then adding base pairs will remain free energetically favorable for much larger  $N$ ; however, as long as the mechanical contribution remains a non-linear function of  $N$ , the free energy as a function of  $N$  will achieve a minimum for some  $N$ , after which adding more base pairs becomes free energetically unfavorable. Eventually, as  $N$  approaches the deGennes length, the mechanical energy will become a linear function of  $N$ . In this case, the energy cost of adding an additional triplet remains constant as a function of  $N$ . If this cost is small in comparison with  $kT$  and unbinding is forbidden, the model suggests that the length of the strand exchange product can increase without limit.

#### **F. The Effect of the Non-Linearity on the Strand Exchange of a Mismatched Triplet**

It has previously been assumed that the free energy penalty for strand exchange of a triplet is approximately equal to the loss of Watson-Crick pairing for that triplet, with a possible additional factor due to the effect of the mismatch on the pairing of the two neighboring bases which ranges from  $\sim 1.5$  to  $\sim 4kT$ . [24] In contrast, for a system with the non-linearity considered here if the initially bound base pairs contain a single mismatch, then strand exchange may be significantly more unfavorable because the unfavorable free energy contribution due to this mismatch must include not only the Watson-Crick pairing energy for that base pair and its neighbors, but also the increased mechanical stress on the two matched base pairs. This stress not only makes a direct contribution to the free energy penalty, but it can also increase the stacking penalty by distorting the bonds between the two homologous base pairs which lowers their Watson-Crick pairing energy. A detailed structural calculation would be required to correctly assess all of these factors. In what follows, we will assume that the free energy penalty for the strand exchange of a mismatched base is approximately equal to the Watson-Crick pairing loss as long as the number of  $< 18$  bp are bound to the filament.

### G. dsDNA Tension Inhibits Progression of Strand Exchange Past a Mismatch

In a system with a linear free energy as a function of the number of bound homologous triplets, adding more homologous triplets is always favorable even if the last triplet added were non-homologous, resulting in enormous kinetic trapping. In contrast, the non-linear energy inhibits binding of additional base pairs after a non-homologous triplet has bound, as illustrated in Fig. 7. The dashed black line shows the curve for a perfect homolog adding a triplet to the initial bound state if all of the other bound triplets have undergone strand exchange. For up to  $\sim 18$  bp thermal energy is sufficient to bind additional base pairs. The dashed gray line shows the free energy penalty for adding a homologous triplet if the last triplet added was non-homologous. The free energy penalty is only slightly larger than the penalty for a homolog; however, the solid gray line shows that the penalty for adding a second triplet is very large, even though both triplets added after the non-homolog were in fact homologous. For comparison, the solid black line shows the energetic favorability of strand exchange of a homologous triplet from the initial binding state. This graph shows that the non-linearity makes adding additional triplets to the initial state is unfavorable once a mismatched triplet has bound, even when the subsequent base pairs are homologous.

### H. Possible Explanations of Biological Results

We have already discussed the proposal that the energetic non-linearity explains why strand exchange is free energetically favorable even though the sequences of the incoming and outgoing strands are the same and the protein contacts in the initial searching ssDNA-RecA filament are similar to those in the final post-strand exchange state.

In addition, experimental results have shown that a rapid initial interaction incorporating  $\sim 15$  base pairs is followed by a slower progression of strand exchange that occurs in triplets [19]. Figure 6 suggests that the binding of dsDNA to site II is favorable for fewer than 9 bases and requires only a few  $kT$  of energy for fewer than 15 bases, whereas for more bases the binding is highly free energetically unfavorable.

Furthermore, FRET based studies indicate that homology recognition may be accurate for short sequences, but inaccurate for longer sequences [17]. A separate study showed that strand exchange pauses at sequence mismatches [25, 26], and we argue that such pauses lead to the unbinding of shorter non-homologous sequences because the binding of the dsDNA to the filament occurs sequentially. In the model presented here the pause in strand exchange at a mismatch results from the free energy cost of transferring the non-homologous triplet to the intermediate state as well as the cost of progressing past a mismatched triplet. Spontaneous unbinding of the entire strand exchange product becomes unlikely as the sequence lengthens because so many free energetically unfavorable transitions are required. If the strand exchange product becomes too long, the unbinding time exceeds the recognition time available to the organism; however, as discussed above, accidental mismatches that extend beyond 18 bp rarely exist *in vivo*. *In vivo*, strand exchange does progress through regions of non-homology once a sufficiently long strand exchange product is formed, but ATP hydrolysis is required. [27, 28]

Finally, it is also well known that in the presence of ATP hydrolysis the size of the strand exchange product increases monotonically until it reaches a limit of  $M \sim 80$  bp [22], where  $M$  is the number of bound dsDNA base pairs. Strand exchange then continues to progress, but  $M$  remains constant because the heteroduplex dsDNA unbinds from the lagging edge of the filament at the same average rate that new dsDNA binds to site II [22, 29]. Since the dsDNA can freely unbind from the filament, free energy minimization implies  $M$  will remain  $\sim M_{freemin}$ . Additional effects associated with dynamics may explain why the strand exchange window moves along the dsDNA with  $M \sim M_{freemin}$  rather than remaining stationary [30]. In contrast with the experimental results obtained in the presence of hydrolysis, experimental results obtained in the absence of hydrolysis show that the length of the strand exchange window can continue without bound.[22] In the model presented here, if the number of base pairs bound is small the mechanical energy penalty associated with adding base pairs is a quadratic function of energy for small numbers already bound; however, the base pairs redistribute the stress between the backbones, so that eventually the base pairs in the center of the filament are not under stress. In this case, the penalty for adding another base pair triplet to the end becomes a linear function of the number bound rather than a quadratic function. Thus, when a sufficient number of base pairs have undergone strand exchange, the energy required to add an additional triplet to the filament is constant, independent of the number bound. If, as suggested by this model, the constant energy decrease due to additional DNA protein contacts is approximately equal to the constant energy increase associated with the added mechanical stress, then the filament can extend forever because both energies are independent of the number of base pairs already bound.

### I. Additional Features in Three Dimensions

In the real RecA system steric factors are associated with the mismatch between the 150 bp persistence length of dsDNA and the strong bending of dsDNA in the 18 bp/turn helical RecA filament. The local rigidity of the dsDNA may play a role in limiting the initial binding length to  $\sim 9$  bp since that many base pairs can interact with the ssDNA-RecA without significant bending. After some dsDNA triplets are bound, the rigidity may also play a role in preventing non-contiguous triplets from being added to the filament. The nearest unbound triplet is already very near to the filament because it is attached to the bound triplets by the phosphate backbones which cannot extend much more than 0.5 nm/bp. Thus, the phosphates are in a position to interact strongly with positively charged residues on the protein which can provide sufficient free energy for the required bending. In contrast, the second neighboring triplet will be separated by a larger distance which reduces the interaction with the protein and requires more bending. A detailed structural calculation would be required to correctly evaluate these effects, but both effects would further support rapid and accurate homology recognition. In the simple one dimensional model discussed here, the free energy effects of the bending can be included in the  $\gamma$  for the initial bound state, but the additional degrees of freedom would alter the coupling between the initial bound state  $\gamma$  and the  $\gamma$  for the intermediate state.

In addition, in the final post-strand exchange state interactions with the L1 and L2 loops may be more favorable for homologous triplets than non-homologous triplets due to steric

factors. Thus, the final state could have a sequence dependent linear contribution to the free energy that was not considered in this model, but may provide additional homology stringency.

### III. CONCLUSION

We have proposed a simple mechanical model for the stress distribution on dsDNA bound to ssDNA-RecA filaments. The model suggests that a change in the conformation of one bound triplet can change the conformation of all of the other bound triplets; consequently, the total energy is a non-linear function of the number of bound base pairs. The model makes several significant qualitative and quantitative predictions. The most important qualitative prediction is that a change in the configuration of one bound triplet changes configuration of all the other bound triplets. The collective behavior of the triplets leads to a free energy that is a non-linear function of the number of consecutive bound triplets, where the binding of additional triplets becomes increasingly unfavorable as the number of bound triplets increases. This non-linearity is important because in systems with more than  $\sim 4$  binding sites, neither thermodynamic equilibrium nor kinetic proofreading can combine accurate and efficient homology recognition when the energy is a linear function of the number of correct pairings. In contrast, an unfavorable non-linear energy combined with the a favorable linear energy due to DNA/protein contacts can promote rapid and accurate homology recognition by making initial sequence independent binding interactions favorable for up to  $\sim 9$  base pairs, while preventing any additional base pairs from binding unless the bound base pairs include 9 contiguous homologous base pairs. If the initially bound base pairs do not contain 9 contiguous homologous base pairs, adding more base pairs to the filament is highly improbable regardless of whether or not the additional base pairs are homologous. This effect combined with the statistics of the sequence distribution of bacterial genomes implies that of all but the  $\sim 50$  out of 10,000,000 possible pairings will rapidly unbind.

In addition, the non-linearity forces the addition of triplets to the filament to proceed sequentially from the initially binding, where adding more than two base pair triplets after a mismatch is highly unlikely, even if the additional base pairs are sequence matched. Furthermore, the model suggests true homologs can proceed to complete strand exchange because the strand exchange of contiguous homologous base pair triplets reduces the tension on the dsDNA. The tension reduction associated with the strand exchange of the initial  $\sim 9$  contiguous homologous base pairs allows more base pairs to bind to the filament due to two effects: 1. the reduction in dsDNA tension reduces the free energy penalty for adding more triplets to the initial bound state 2. the decrease in the energy of the bound dsDNA due to strand exchange reduces the unbinding rate of the dsDNA by making unbinding more unfavorable. As strand exchange progresses from the  $\sim 9$  initial contiguous base pairs, the binding of the strand exchanged state still remains weak enough to be reversed unless  $\sim 18$  contiguous bp make the transition to the final post-strand exchange state, a transition which is not favorable for  $< \sim 18$  bp. These effects could provide exact sequence recognition for bacterial genomes except for repeated genes. The actual speed and accuracy of homology recognition depends on detailed values of parameters as well as additional factors associated with the three dimensional geometry of the filament, so we do not provide detailed estimates here; however, the simple model presented here provides mechanisms for overcoming



fundamental limitations encountered in systems with more than 4 binding sites where the binding energy is a linear function of the number of correctly paired binding sites.

In sum, the energy non-linearity produces three crucial advantages that are unavailable to systems with a strictly linear energy: 1. the initial interaction is limited to  $\sim 15$  bp beyond which binding of dsDNA triplets to the filament cannot progress without a sequence dependent transition of 9 contiguous homologous base pairs to the strand exchanged state 2. nearly immediate unbinding of any sequence that does not contain at least 9 contiguous homologous base pairs 3. a large free energy penalty that prevents strand exchange from progressing past a sequence mismatch even if the mismatch is followed by homologous triplets 4. a large free energy penalty that makes the transition from the intermediate state to the final state unfavorable until  $\sim 18$  contiguous base pairs make the transition to the final state. These features provide much more rapid and accurate homology recognition than systems using linear energies: in systems with linear energies addition and strand exchange of a homologous triplet is always favorable; therefore, in systems with linear energies even short regions of accidental homology can produce substantial trapping times, as demonstrated by both analytical modeling and numerical simulations. [14] Qualitative features of the model provide possible explanations for well known but previously unexplained features for homology recognition and strand exchange and suggest that the bond rotations that appear in the overstretching of naked dsDNA may have a role in strand exchange.

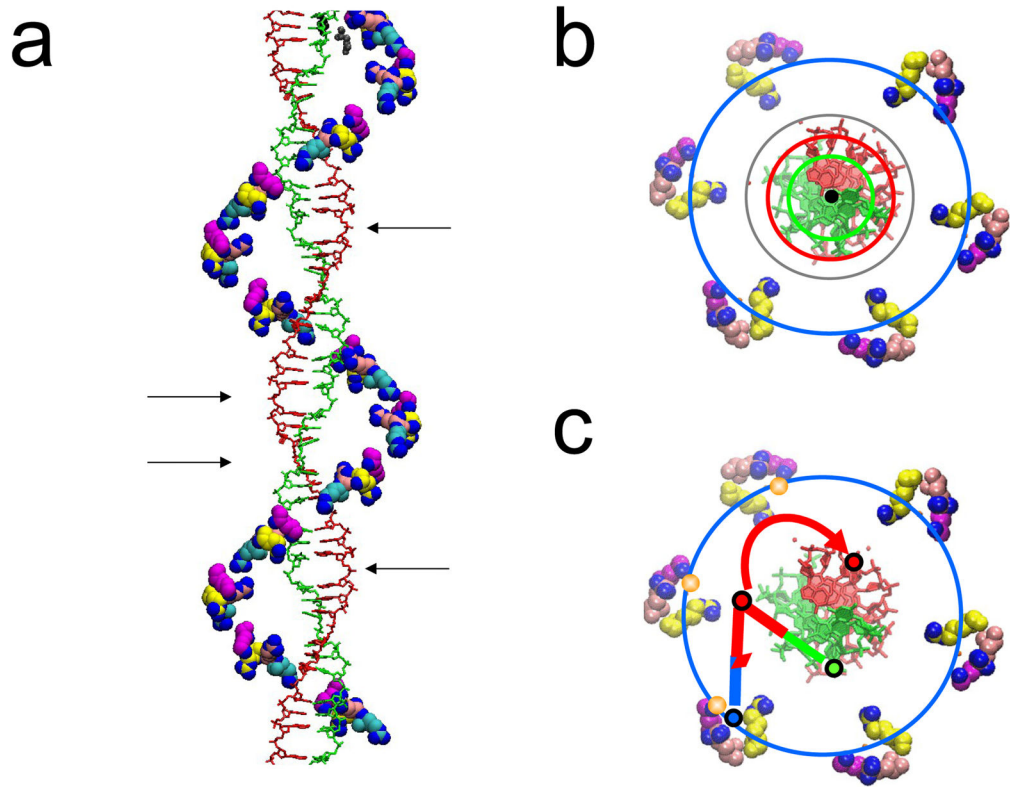
## Acknowledgments

We would like to thank Douglas Bishop, Yuen-Ling Chan, and Chantal Prévost for helpful conversations about the X-ray structure of DNA bound to RecA. We would also like to thank Chantal Prévost for the PDB file of a complete RecA helix.

## References

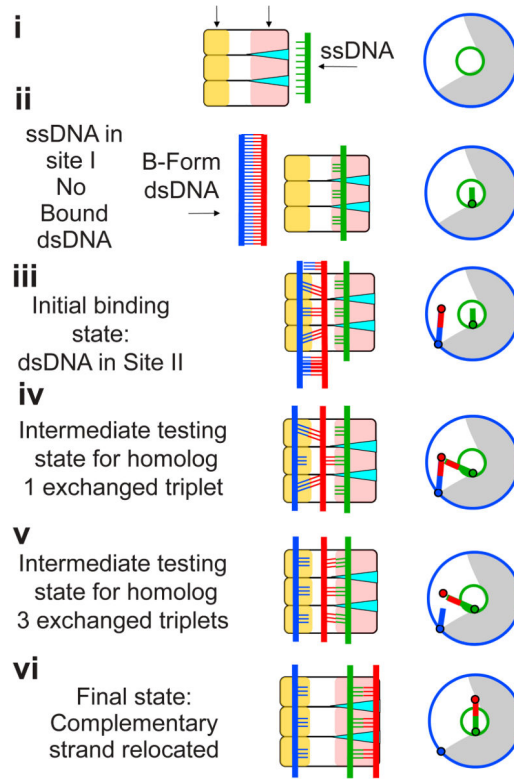
1. Roca A, Cox M. *Mol Biol.* 1990; 25:415.
2. Kowalczykowski S, Eggleston A. *Annu Rev Biochem.* 1994; 63:991. [PubMed: 7979259]
3. Chen Z, Yang H, Pavletich N. *Nature.* 2008; 453:489. [PubMed: 18497818]
4. Folta-Stogniew E, OMalley S, Gupta R, Anderson K, Radding C. *Mol Cell.* 2004; 15:965. [PubMed: 15383285]
5. Xiao J, Lee A, Singleton S. *ChemBioChem.* 2006; 7:1265. [PubMed: 16847846]
6. Mazin A, Kowalczykowski S. *Proc Natl Acad Sci USA.* 1996; 93:10673. [PubMed: 8855238]
7. Menetski J, Bear D, Kowalczykowski S. *Proc Natl Acad Sci USA.* 1990; 87:21. [PubMed: 2404275]
8. Rosselli W, Stasiak A. *J Mol Biol.* 1990; 216:335. [PubMed: 2147722]
9. Stasiak A, Capua ED, Koller T. *J Mol Biol.* 1981; 151:557. [PubMed: 7040675]
10. Savir Y, Tlusty T. *Mol Cell.* 2010; 40:388. [PubMed: 21070965]
11. Mazin A, Kowalczykowski S. *Genes Dev.* 1999; 13:2005. [PubMed: 10444598]
12. Klapstein K, Chou T, Bruinsma R. *Biophys J.* 2004; 87:1466. [PubMed: 15345529]
13. Kates-Harbeck J, Tilloy A, Prentiss M. to be published.
14. Feinstein E, Prentiss M. to be published.
15. Hsieh P, Camerini-Otero C, Camerini-Otero R. *Proc Natl Acad Sci USA.* 1992; 89:6492. [PubMed: 1631148]
16. Hopfield J. *Proc Natl Acad Sci USA.* 1974; 71:4135. [PubMed: 4530290]

17. Bazemore L, Folta-Stogniew E, Takahashi M, Radding C. Proc Natl Acad Sci USA. 1997; 94:11863. [PubMed: 9342328]
18. Peacock-Villada A, Yang D, Danilowicz C, Feinstein E, Pollack N, McShan S, Coljee V, Prentiss M. Nucl Acids Res. 2012; 40:10441. [PubMed: 22941658]
19. Rangunathan K, Joo C, Ha T. Structure. 2011; 19:1064. [PubMed: 21827943]
20. Karplus, M. private communication.
21. Danilowicz C, Feinstein E, Conover A, Coljee V, Vlassakis J, Chan Y, Bishop D, Prentiss M. Nucl Acids Res. 2011; 40:1717. [PubMed: 22013164]
22. van der Heijden T, Modesti M, Hage S, Kanaar R, Wyman C, Dekker C. Mol Cell. 2008; 30:530. [PubMed: 18498754]
23. Gennes PGD. CR Acad Sci IV-Phys. 2001; 2:1505.
24. SantaLucia J Jr. Proc Natl Acad Sci USA. 1998; 95:1460. [PubMed: 9465037]
25. Lee A, Xiao J, Singleton S. J Mol Biol. 2006; 360:343. [PubMed: 16756994]
26. Sagi, aSD.; Tlusty, T. Nucl Acids Res. 2006; 34:5021. [PubMed: 16990254]
27. Rosselli W, Stasiak A. EMBO J. 1991; 10:4391. [PubMed: 1836761]
28. Kim J, Cox M, Inman R. J Biol Chem. 1992; 267:16444. [PubMed: 1644828]
29. Fan HF, Cox M, Li HW. PLoS ONE. 2011; 6:e21359. [PubMed: 21765895]
30. Conover A, Danilowicz C, Gunaratne R, Coljee V, Kleckner N, Prentiss M. Nucl Acids Res. 2011; 39:8833. [PubMed: 21768124]

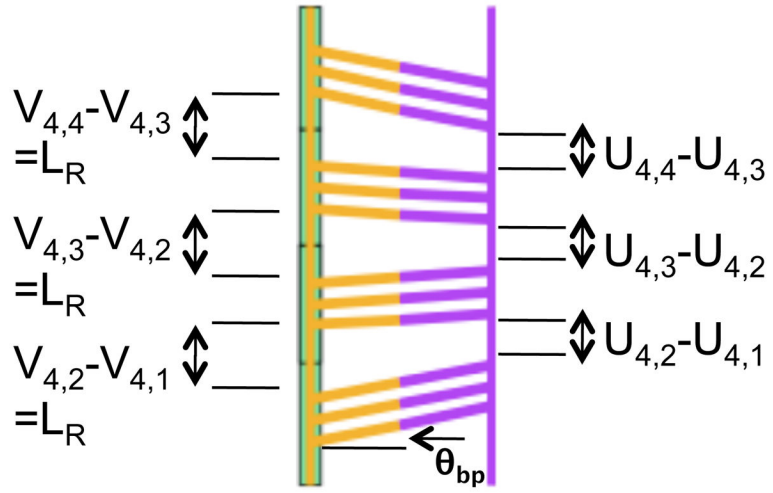


**Figure 1.**

a. Side view of the X-ray structure of the dsDNA post-strand exchange (final) state with complementary (on the right at the arrow) and incoming strands shown as green and red stick renderings. The arrow points to a few rises between two triplets in the dsDNA structure. The VMD (32) renderings of RecA crystal structure 3CMX (8) show site II residues Arg226 (pink), Lys227 (cyan), Arg243 (yellow), and Lys245 (magenta) with charged nitrogen atoms (blue). The cyan triangle indicates the approximate position of an outgoing strand phosphate b. top view of the same structure with circles indicating the radii occupied by the incoming (green), outgoing (blue), and complementary strands (red for final state and gray for intermediate state). c. top view showing the base pairing in the homology recognition/strand exchange process superimposed on the actual X-ray structure.



**Figure 2.** Schematic of Interactions between dsDNA and the ssDNA-RecA filament as strand exchange progresses for a homolog with side views shown in the central part of each panel and top views shown in the right part of each panel where the grey region indicates the space occupied by the protein DNA/protein and the circles indicate the radii occupied by the three strands in their final post-strand exchange positions. In the schematic, the filament consisting of three RecA proteins, showing site I (pale red), site II (pale orange), and the support for the rises provided by the L1 and L2 loops (cyan), with (i) unbound ssDNA (ii) ssDNA bound in site I and unbound B-form dsDNA, (iii) ssDNA bound in site I and dsDNA bound in site II; The outgoing strand (far left), complementary strand (bound to the outgoing strand in purple), and incoming strand (bound to the protein on the right) are shown in blue, red, and green, respectively. (iv) central triplet has undergone strand exchange, resulting in a decrease in lattice mismatch and a decrease in the stress on the bp (v) all three triplets shown have undergone strand exchange (vi) ssDNA bound in site II and dsDNA bound in site I in the final assembled state which has even less bp stress than the strand exchanged state due to increase mechanical support for the rises.



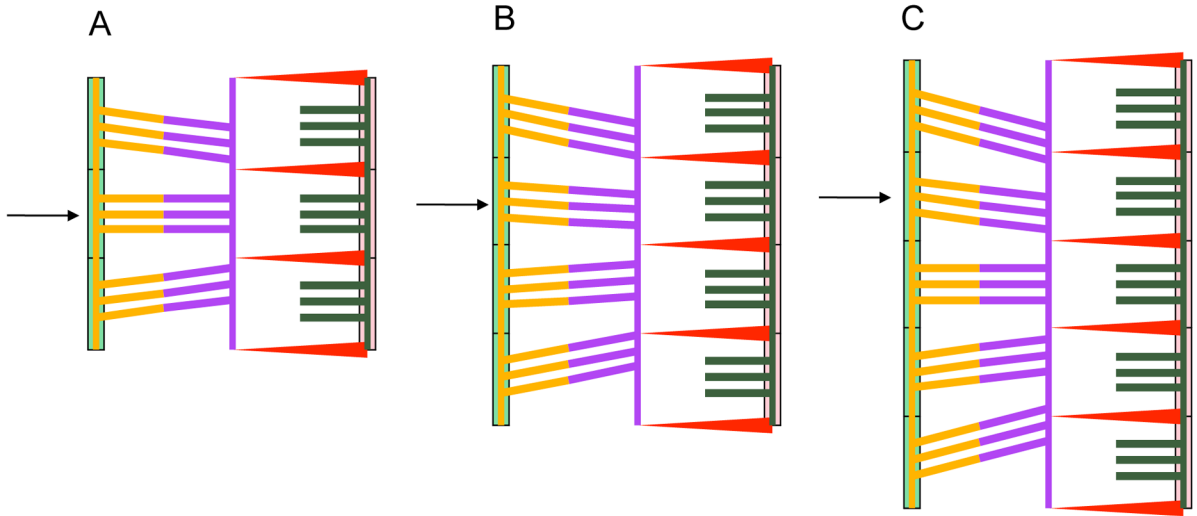
**Figure 3.** Schematic of dsDNA bound to site II of the protein showing the model parameters.

Author Manuscript

Author Manuscript

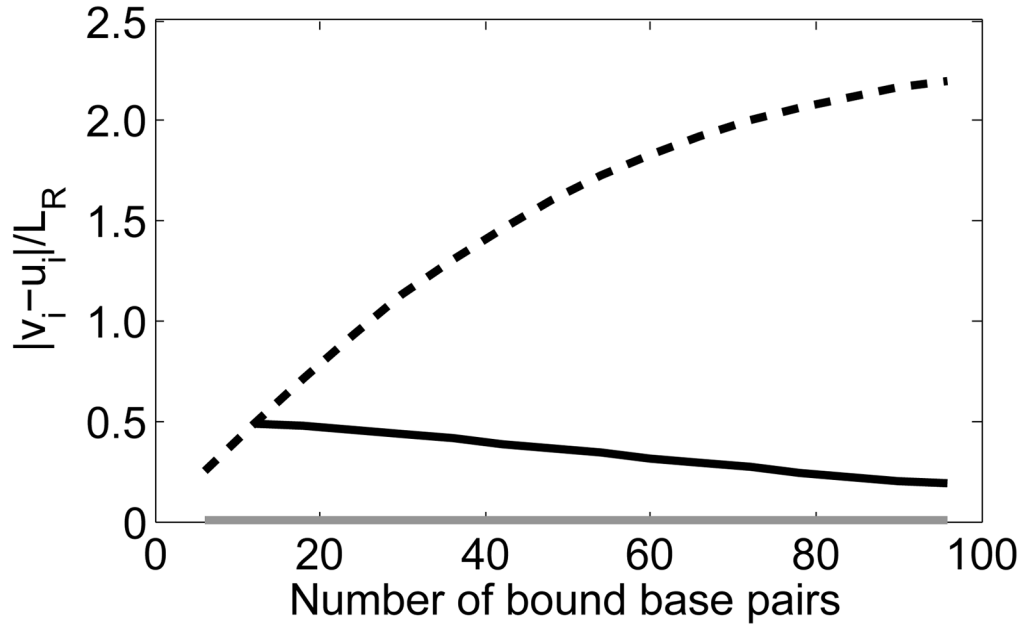
Author Manuscript

Author Manuscript



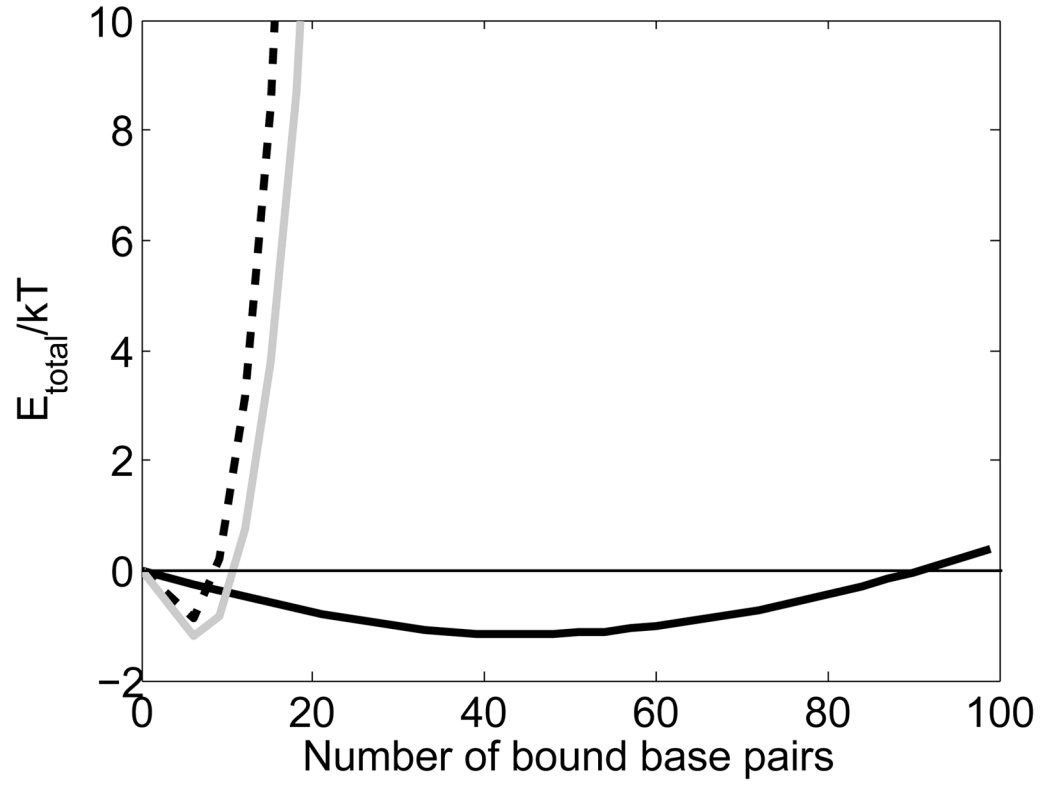
**Figure 4.**

Effects of adding triplets to the initial bound state as calculated by the model. (A) Three triplets shown in the bound state with an arrow pointing to the second one. (B) Adding a triplet to the initial state changes the tension in the other triplets. (C) Again tension on the triplets change as a fifth triplet is added to the initial state.



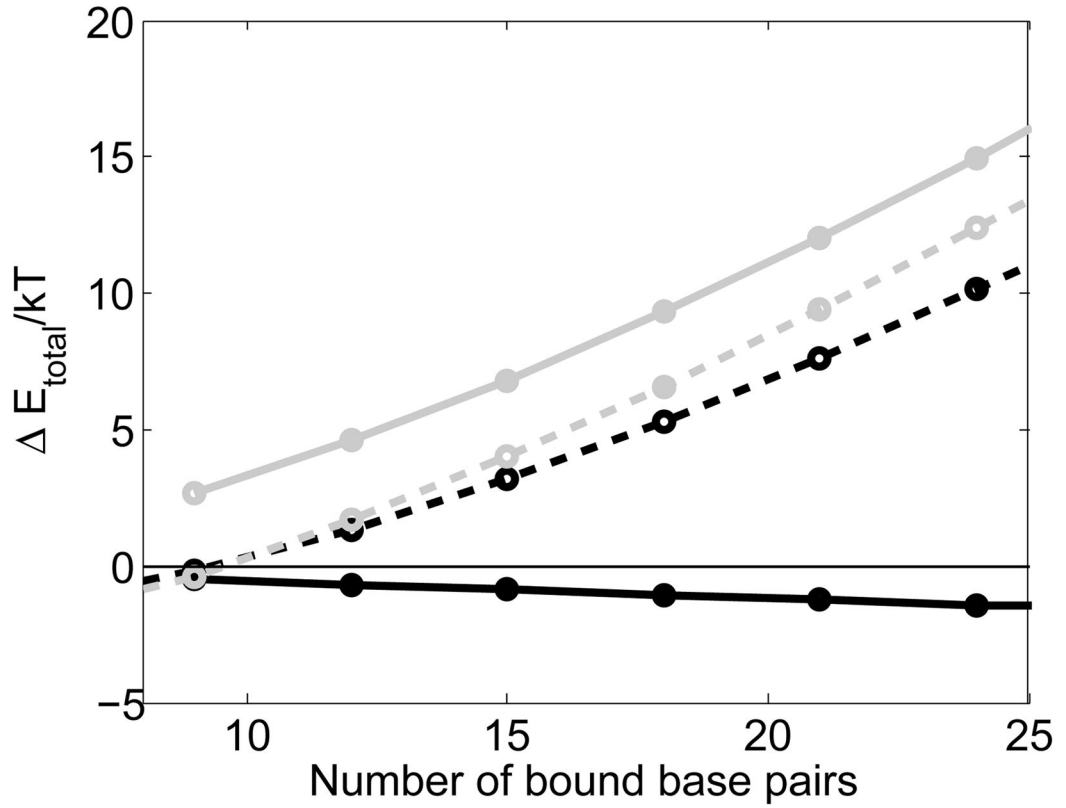
**Figure 5.**

The dashed, solid gray, and solid black lines show the absolute value of  $v_i - u_i$  for the outer triplet, central triplet, and the first triplet out from the center respectively when the number of bound triplets is odd.



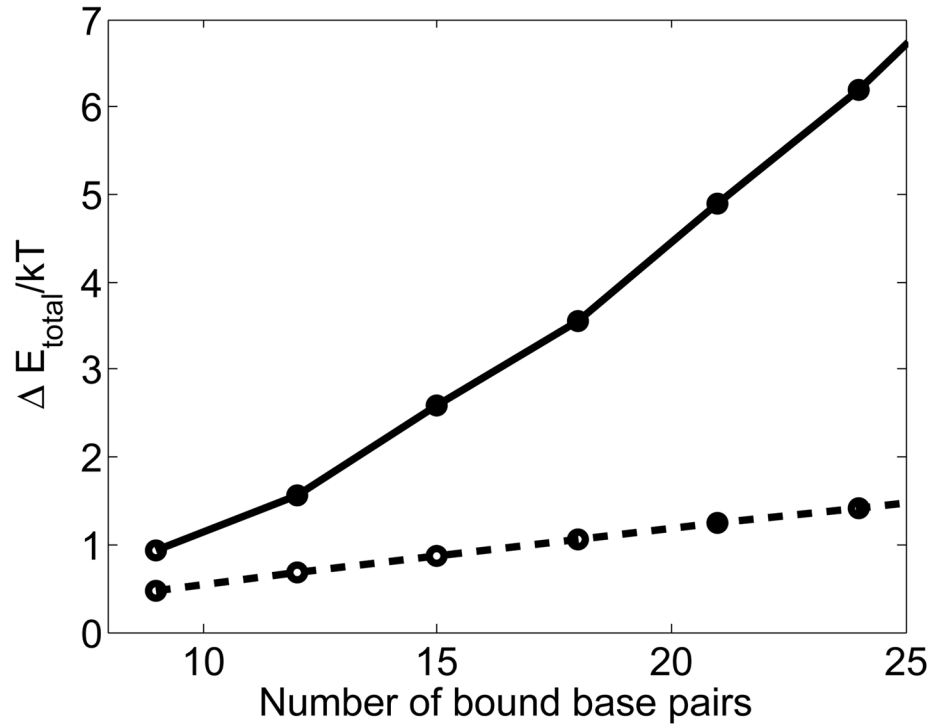
**Figure 6.** Total energy as a function of bound base pairs all in the initial bound state (dashed line), the intermediate state (solid gray line) and the final post-strand exchange state (solid black line).





**Figure 7.**

The free energy penalty for adding another triplet in the initial bound state as a function of the number of base pairs in the strand exchanged state. The dashed black line shows the penalty when all of the triplets are homologous. The dashed gray line shows the penalty for adding an additional triplet to the initial bound state after a triplet that is non-homologous. The solid gray line shows the penalty for adding a second triplet after a non-homolog where the first triplet after the non-homolog was homologous. The solid black line shows the energetic favorability of strand exchanging a homologous triplet in the initial bound state.



**Figure 8.**

For a completely homologous dsDNA with all triplets in the intermediate strand exchanged state, the free energy penalty for a single triplet reversing strand exchange by making a transition from the intermediate strand exchanged state to the initial bound state as a function of then number of base pairs when the base flipping triplet is in the center of the filament (solid line) or at the end of the filament (dashed line)