

Physics of RecA-Mediated Homologous Recognition

Kevin Klapstein,* Tom Chou,* and Robijn Bruinsma[†]

*Department of Biomathematics, [†]Department of Physics, University of California-Los Angeles, Los Angeles, California

ABSTRACT To accomplish its DNA strand exchange activities, the *Escherichia coli* protein RecA polymerizes onto DNA to form a stiff helical nucleoprotein filament within which the DNA is extended by 50%. Homology search and recognition occurs between ssDNA within the filament and an external dsDNA molecule. We show that stretching the internal DNA greatly enhances homology recognition by increasing the probability that the homologous regions of a stretched DNA molecule and a parallel, unstretched DNA molecule will be “in register” at some position. We also show that the stretching and stiffness of the filament act together to ensure that initiation of homologous exchange between the substrate DNA molecules at one position precludes initiation of homologous exchange at any other position. This prevents formation of multiple exchange site “topological traps” which would prevent completion of the exchange reaction and resolution of the products.

INTRODUCTION

RecA is a 38-kDa *Escherichia coli* protein that plays a key role both in DNA repair and in the exchange of genetic material by promoting DNA strand exchange. RecA or an RecA homolog has been found in every species in which it has been sought (Roca and Cox, 1997). RecA-mediated strand exchange is important in maintenance of the genome and essential for sexual reproduction. To facilitate strand exchange, RecA polymerizes onto both single- and double-stranded DNA (ssDNA and dsDNA) to form a right-handed helical filament ~ 10 nm in diameter (Heuser and Griffith, 1989) with a 6-monomer-per-turn repeat length (Yu and Egelman, 1992; Takahashi and Norden, 1994).

The extended filament, formed with an ATP cofactor (Heuser and Griffith, 1989; Yu and Egelman, 1992), is the active form of RecA for strand exchange. Extended filaments are very stiff. The persistence length of the extended filaments formed with ssDNA is $\xi_{\text{ssDNA-RecA}} \simeq 860$ nm (Hegner et al., 1999), ~ 16 times that of dsDNA. DNA within the extended filament is stretched by 50% relative to B-form DNA. Although many different proteins with diverse functions act on DNA, most act as monomers or components of oligomers containing relatively few other elements. RecA is unusual in forming a stable protein filament to accomplish its activity.

The presence of this structure is puzzling. Stretching DNA by 50% is energetically expensive, requiring $\sim 0.7 k_B T$ per basepair for dsDNA. The presence of the filament also seems likely to hinder the close contact between the DNA substrates necessary for sequence comparison. Moreover, stretching one substrate DNA molecule relative to the other seems to present a serious obstacle to RecA activity by complicating the process of aligning regions of homology

between them. Fig. 1 shows how a stretched DNA molecule is unable to remain “in register” with a homologous region on an unstretched DNA molecule. If they are homologously aligned at one base(pair), the neighboring base(pair) is “out of register” by the difference in base(pair) spacing between the two molecules. For RecA filaments, the stretching is by 50%. Starting from a homologously aligned base(pair), the next-to-neighboring base on the stretched molecule is an entire base out-of-register. These facts would seem to inhibit the strand exchange activity.

Other DNA processing proteins, including those with recombination or repair activity, function without needing this structure, yet the RecA family have preserved the filament structure over 2.5 billion years of evolution in species as diverse as *E. coli* and *Homo sapiens*. The aim of this article is to use basic physical considerations to understand the role of RecA filaments in RecA function.

FACILITATING HOMOLOGOUS ALIGNMENT

Incompatible interbase spacings

Homology recognition requires that many consecutive basepairs be recognized as complementary. Identifying only a single pair of complementary bases is insufficient, yet stretching the DNA within the filament limits simultaneous comparison to a single basepair. Surprisingly, stretching the DNA within the filament does not impede, but rather *accelerates* the initial alignment of the homologous regions of the DNA substrates.

To see how, consider two B-form DNA molecules having a region of homology. The distance between consecutive basepairs of B-form DNA is $a_0 \simeq 3.4$ nm. If these molecules are parallel, then for some position of one molecule relative to the other, the homologous regions will be homologously aligned. As shown in Fig. 2 A, every basepair throughout the region of homology is then homologously aligned.

Submitted January 5, 2004, and accepted for publication May 19, 2004.

Address reprint requests to Tom Chou, Dept. of Biomathematics, UCLA, Box 951766, Los Angeles, CA 90095-1766. Tel.: 310-206-2787; E-mail: tomchou@ucla.edu.

© 2004 by the Biophysical Society

0006-3495/04/09/1466/12 \$2.00

doi: 10.1529/biophysj.104.039578

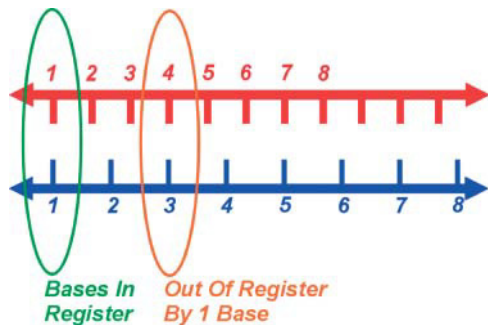


FIGURE 1 Unequal spacing between basepairs obstructs homologous alignment. Spacing between consecutive bases on the bottom strand is 50% larger than on the top strand, as for RecA filaments. The molecules are homologously aligned at base 1. By base 3 on the stretched strand they are an entire base out-of-register.

Fig. 2 *B* shows the effect if we displace one molecule relative to the other by a_0 , the distance between consecutive basepairs. Now, none of the basepairs are homologously aligned. When two DNA molecules have identical spacing between consecutive bases, being out-of-register anywhere results in being out-of-register everywhere. Homologous alignment between two DNA molecules with identically spaced bases is an all-or-nothing phenomenon.

Next, consider the case when one of the molecules is stretched. The spacing between consecutive basepairs is ηa_0 ,

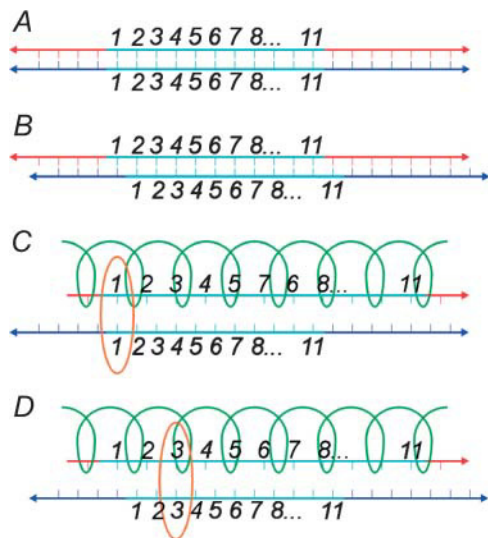


FIGURE 2 Alignment between DNA molecules. The blue regions are homologous. (A) Consecutive bases are a distance a_0 apart in both molecules. All homologous bases are homologously aligned and in-register. (B) The bottom strand has been moved to the right a distance a_0 . Now none of the bases are homologously aligned. (C) The upper molecule is coated with RecA, so bases are $3a_0/2$ apart. The first base in the region of homology is homologously aligned (orange circle), but all others are out-of-register. (D) Moving the bottom strand to the right a distance a_0 does not destroy the alignment, but merely moves its location to the third base in the region of homology (orange circle). All other bases remain out-of-register.

where the stretching factor $\eta > 1$. For RecA-coated DNA, $\eta = 3/2$. This is shown in Fig. 2, *C* and *D*. The stretching has two effects on the homologous alignment. First, as shown in Fig. 2 *C*, a homologously aligned basepair will be the only homologously aligned pair. The molecules immediately get out-of-register and the alignment is lost. Second, if we again displace one of the molecules a distance a_0 , we get Fig. 2 *D*. Now, unlike the previous case, a homologous alignment is preserved. The basepair originally homologously aligned is now out of alignment, but two nearby bases have moved into homologous alignment. The behavior is analogous to the operation of a Vernier scale or a slide rule. Unlike the situation when both molecules have the same basepair spacing, the homologous alignment is stable with respect to changes in the relative positions of the two DNA molecules. This is a key property which is essential to our reasoning. Although homologous alignment between one stretched and one unstretched DNA molecule occurs at only one basepair, there will always be one homologously aligned basepair ready to initiate strand exchange.

Homology alignment and recognition model

To show how this is of value for RecA activity, we need a model of how homologous alignment and recognition occur. We define two substrates, \mathcal{R} and \mathcal{D} . \mathcal{R} is a RecA-coated ssDNA molecule J bases long. \mathcal{D} is a very long B-form dsDNA molecule which contains a region homologous to \mathcal{R} .

We make several comments about the energetics involved in recombination. First, since the process can occur without ATP hydrolysis, any suggested mechanism must function in the absence of an energy source. This means diffusion and thermal fluctuations must be sufficient to drive the process. It also means that the energy of the substrates \mathcal{R} and \mathcal{D} must be higher than the energy of the hybrid DNA molecule and displaced DNA strand, but lower than the energy of the products of an attempt to exchange strands nonhomologously.

To compare bases, the DNA molecules must be brought into close physical proximity, so some segment of \mathcal{D} must enter the filament of \mathcal{R} through the helical groove. There are binding sites for more than one strand of DNA within the RecA filament, meaning the potential energy of this invading segment decreases as it enters the filament.

Homology recognition requires that the invading segment be stretched to keep its bases in register with those of \mathcal{R} . We assume that this happens simultaneously with the movement of the invading segment into the filament. This will have an energy cost of $\sim 0.7 k_B T$ per basepair. For more than a few basepairs this will be large compared to the energies of the thermal fluctuations which drive the process. This can be overcome if the energy from binding the invading segment within the filament remains close to the energy needed to stretch it at all times as it enters the filament. This is shown

qualitatively in Fig. 3. In that case, the net energy change of the invading segment as it simultaneously moves into the filament and is stretched by 50% will remain small. The combined process, and its reverse, are then easily accessible to thermal fluctuations.

The requirement that the stretching and binding energies nearly cancel at the end points of the reaction would ensure only that the thermodynamic difference between these states is small. We have additionally stipulated that the stretching of the invading segment occur simultaneously with its approach to the binding site within the filament and that the stretching and binding energies very nearly cancel throughout the entire process. These features ensure that not only is the thermodynamic difference between the states small, but that the energy barrier between them is small as well. This allows the kinetics of the transition between these states to be very fast, even when the transitions are driven only by thermal fluctuations.

Finally, consider the energetics involved in checking for complementarity between a base of the DNA within \mathcal{R} and an adjacent base on the invading segment by attempting to exchange basepairing partners. This could be accomplished by rotating the bases relative to the sugar phosphate backbone (Nishinaka et al., 1997, 1998), which is accessible to thermal energies. Stretching of the DNA will have largely eliminated base-stacking interactions, so the energy barrier will be dominated by the need to break the hydrogen bonds of the original Watson-Crick basepairs. This will be easier for A:T basepairs, with two hydrogen bonds each, than for G:C basepairs, with three, so we expect this part of the reaction to be dominated by A:T basepairs (Gupta et al., 1999). This is also accessible to thermal energies.

Process steps

We divide the process of aligning and identifying a homology between \mathcal{R} and \mathcal{D} into six steps.

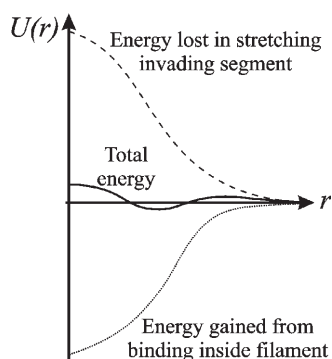


FIGURE 3 Qualitative sketch of the potential energy U of an invading segment as a function of a reaction coordinate r corresponding roughly to the distance of the invading segment from the binding site within the filament. (Dashed) Energy needed to stretch the invading segment; (dotted) energy gained from binding inside the filament; and (solid) total reaction energy.

1. Random initial contact between \mathcal{R} and \mathcal{D} .
2. Rotation into a parallel but random and non-sequence-specific orientation.
3. Introduction of “invading segments” into the filament to position bases of \mathcal{D} close enough to bases of \mathcal{R} to allow for comparison and testing for complementarity.
4. Testing for complementarity.
 - a. Rejection of noncomplementary alignments.
 - b. Stabilization of complementary alignments.
5. Testing for homology.
 - a. Expulsion of the invading segment in the absence of sufficient stabilization by additional complementary alignments.
 - b. Extension of the invading segment when stabilized by enough additional complementary alignments.
6. Extension of the hybrid DNA from a homologously aligned *exchange nucleation point*.

Step 1, the initial contact between the substrates, will occur at a point as shown in Fig. 4 A. Since this is a result of diffusion, the initial point of contact will be random. It has been noted in the literature that there is a weak, nonspecific (electrostatic) attraction between the substrates (Karlín and Brocchieri, 1996). This exerts a torque on the substrates, accomplishing step 2 by pulling them toward a loose, nonspecific parallel alignment as shown in Fig. 4 B.

In a one-dimensional search, a simple model would be to wind a long section of \mathcal{D} into the filament nonspecifically and search for homology by sliding longitudinally within the filament. This avoids repeatedly winding short invading segments of \mathcal{D} in and out of the filament, but has other serious defects. These include experimental evidence that excludes significant sliding during the search process (Adzuma, 1998), and the observation that the longitudinal sliding of such large molecules would be very slow.

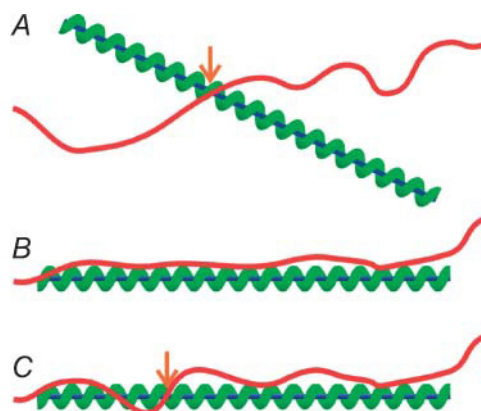


FIGURE 4 (Red curve) dsDNA molecule (\mathcal{D}); (blue line) ssDNA; and (green helix) RecA filament (\mathcal{R}). (A) Contact at a point (orange arrow). An attractive interaction between \mathcal{D} and \mathcal{R} exerts a torque around this point. (B) The nonspecific parallel orientation produced by the torque in A. (C) An invading segment of \mathcal{D} enters the RecA filament of \mathcal{R} through the filament groove (orange arrow).

This would also create a “trap”. A long section of \mathcal{D} will only wind into the filament if its potential energy is lower inside than outside. In that case, it will not be readily removed from the filament once it has been wound in. The result is that any DNA heterologous to the DNA in \mathcal{R} acts as a suicide substrate by winding into the filament and blocking entry by homologous DNA. Impediment of the homologous recombination reaction with the human RecA homolog hRad51 under low salt conditions has been observed as a result of binding by heterologous dsDNA to the exterior of the nucleoprotein filament (Sigurdsson et al., 2001). The situation would be greatly exacerbated if the heterologous dsDNA were bound to the nucleoprotein filament more tightly, as would be the case for heterologous dsDNA bound within the filament. This poisoning of the reaction does not happen in vitro and would be lethal in vivo. For these and other reasons, we reject this model.

The above considerations compel us to postulate a model involving steps 3–5, repeatedly winding short invading segments of \mathcal{D} into and out of the filament as shown in Fig. 4 C. The invading segment must be stretched simultaneously with its movement into the filament so that it remains in register with the DNA within \mathcal{R} . As described in the previous subsection, this can be driven by the thermal fluctuations of \mathcal{D} . Step involves testing for complementarity by some process such as the rotation of the A and T bases (Gupta et al., 1999) relative to the sugar phosphate backbone (Nishinaka et al., 1997, 1998). Hybrid basepairs formed this way must have a lower energy than the original basepairs, most likely as a result of the geometry to which they are constrained by the filament.

If the aligned bases of the hybrid molecule are not complementary, they cannot form a hybrid pair. Without a new basepairing, the energy of the rotated state will be higher than the energy of the original pairing, and the original pairing will quickly reform. In this manner, step 4a is accomplished.

If the aligned bases of the hybrid molecule are complementary, they can form a hybrid pair. We assume that the energy of the hybrid pair is only slightly lower than that of the original basepair. In this way the hybrid pair will be slightly stabilized as required by step 4b, but will eventually be disrupted by thermal fluctuations if further stabilization is not achieved. The return to the original pairing will not be completely prevented, but it will be delayed.

If the original basepairings reform as in step 4a, thermal fluctuations will quickly remove the invading segment from the filament. The search for homology between \mathcal{R} and \mathcal{D} has failed at this particular point along their lengths. If a hybrid pair forms as in step 4b, the slight stabilization of the position of the invading segment within the filament will allow time in which thermal fluctuations can introduce more of \mathcal{D} into the filament, lengthening the invading segment.

There are then two possibilities. If, as in step 5a, \mathcal{R} and \mathcal{D} are not homologously aligned at this point, there will be only

occasional, fortuitous complementary alignments of A and T bases. This will provide little or no additional stabilization of the position of the invading segment within the filament. Such minimal stabilization will be insufficient to prevent the eventual disruption of these serendipitous hybrid pairs and thermal fluctuations will remove the invading segment from the filament.

Step 5b occurs if \mathcal{R} and \mathcal{D} are homologously aligned at this point, in which case every A:T pair of \mathcal{D} which enters the filament can make a small further contribution to stabilizing the presence of the invading segment within the filament. This allows time for even more of \mathcal{D} to enter the filament as a result of thermal fluctuations, adding further stabilization. This region now constitutes a nucleation point for the strand exchange reaction from which the entire homologous region of \mathcal{D} will be wound into the filament. This is a very stable state, and will persist long enough for the G:C pairs to break their three hydrogen bonds and exchange basepairing partners, forming hybrid G:C pairs and completing the strand exchange reaction.

Advantages from stretching one molecule

The minimal energy costs of steps through allow \mathcal{D} and \mathcal{R} to be rapidly and efficiently searched for a homologous alignment. If there exists a point of homologous alignment, it will be found and will serve as a nucleation site from which the exchange reaction will extended throughout the entire homology. This reduces the problem to a question of how probable it is that there is a point of homologous alignment. Stretching of the DNA by RecA plays a crucial role by enhancing this likelihood.

Target size enhancement

To have a point of homologous alignment between two DNA molecules requires that they lie within a certain range of relative positions. We define the *target size* σ for homologous alignment as the range of longitudinal positions of one DNA molecule relative to another which result in the homologous alignment of at least one basepair. For two B-form DNA molecules, this requires great precision. The molecules must lie within $\pm a_0/2$ of an exact alignment so that $\sigma = a_0$. If they do lie in this range then all the bases are homologously aligned, but this is not desirable. Later, we show that an extended region of homologous alignment poses serious problems.

This contrasts with the situation between \mathcal{D} and \mathcal{R} . When \mathcal{R} is stretched by a factor $\eta_{\mathcal{R}}$, the target size is greatly augmented. This is because of the phenomenon illustrated in Fig. 2. Recall what we saw there: given a homologous alignment between a base on an unstretched DNA molecule and a base on a stretched DNA molecule, moving one of the molecules relative to the other changed the location of the homologous alignment to a different pair of bases, but it did

not destroy the alignment. The target size for a homologous alignment between a stretched DNA molecule and unstretched DNA molecule is the maximum range of positions of one of the molecules relative to the other which maintains some point of homologous alignment between them. For a region of homology J bases in length, the resulting expression (compare to Appendix A) is

$$\sigma = [(J - 1)(\eta_{\mathcal{R}} - 1) + 1]a_0. \quad (1)$$

For two unstretched DNA molecules, $\eta_{\mathcal{R}} = 1$ and

$$\sigma = a_0 = 0.34 \text{ nm}, \quad (2)$$

exactly as expected. Note that this is a fixed, constant value which is unchanged by the length of the homology.

Now consider what happens if we use the known RecA value of $\eta_{\mathcal{R}} = 3/2$. Equation 1 then gives us

$$\sigma = (J + 1)a_0/2. \quad (3)$$

Here at last we see in a clear, quantitative form the enormous advantage of stretching the DNA within \mathcal{R} . This expression scales linearly with the homology length. A modest 200 base homology gives

$$\sigma \simeq 100 \quad a_0 = 34 \text{ nm}. \quad (4)$$

This huge target size, already a 100-fold increase relative to two unstretched DNA molecules, offers an enormous advantage to the homology recognition process. It is achieved only because of the stretching of the DNA within the RecA filament.

Reaction rate

Knowing the target size allows us to estimate the reaction rate. The fact that the target size is proportional to the length of the region of homology produces an interesting and somewhat surprising result. In the absence of sliding (Adzuma, 1998) the reaction rate cannot exceed the diffusion limit, so the maximum “on rate” for the reaction is the Debye-Smoluchowski rate. Our target is a section of \mathcal{D} , and is therefore a cylinder of length σ , but we approximate this by a spherical target of radius $\sigma/2$. We also estimate the diffusion constant of \mathcal{R} , which is a cylinder of length \mathcal{R} , by the diffusion constant for a sphere with a radius of $\ell/2$. With these approximations we find

$$k_a \simeq \frac{2k_B T}{3\eta} \left(1 - \frac{1}{\eta_{\mathcal{R}}} \right). \quad (5)$$

This result is independent of the length of the homology and the spacing between consecutive base(pair)s in unstretched B-form DNA. It depends only on the temperature T , on the viscosity of the solution η , and on the stretching factor $\eta_{\mathcal{R}}$. The stretching factor for DNA within the RecA filament is $\eta_{\mathcal{R}} = 3/2$, and $\eta \simeq 10^{-3}$ Poiseuille at 20°C, so we have $k_a \simeq 5.3 \times 10^8 \text{ Mol}^{-1}\text{s}^{-1}$.

The reaction rate for homologous alignment and recognition provides an opportunity to test the enhanced target size hypothesis. In the absence of target size enhancement the reaction rate should be roughly proportional to $1/J$. Target size enhancement predicts that it should be relatively insensitive to J . Measurement of the reaction rates as a function of the length of \mathcal{R} should clearly distinguish between these alternatives.

PREVENTING MULTIPLE HOMOLOGOUS ALIGNMENTS

Topological trapping

If homologous strand exchange between two substrates is initiated at two or more separate points it will result in a problematic topological trapping of the reaction. To extend a region of hybrid DNA, at least one strand of the external dsDNA must wind into the RecA filament (Honigberg and Radding, 1998). If exchange between the substrates is initiated at two separate points, this motion produces compensating counterturns of the dsDNA around the outside of the filament, as shown in Fig. 5. Extending the hybrid DNA increases the number of counter turns while decreasing the length of dsDNA that forms them, which rapidly decreases the radius of the counterturns. This makes them very energetically expensive to produce, which eventually stops extension of the hybrid DNA.

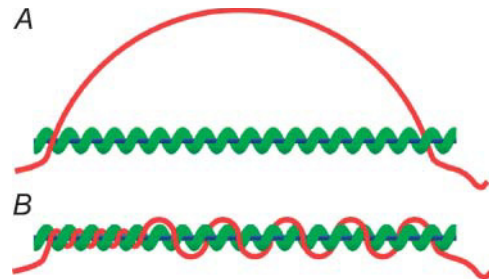


FIGURE 5 Topological trapping resulting from initiation of strand exchange at two separate points. (Red curve) dsDNA molecule (\mathcal{D}); (blue line) ssDNA; and (green helix) RecA filament (\mathcal{R}). (A) Homologous alignment at alignment regions 1 (left) and 2 (right). These are shown only at their initial point of closest approach but may be of arbitrary length, e.g., region 1 may extend far to the left of what is shown. (B) Extending alignment region 1 to the right requires that \mathcal{D} be wound into the filament of \mathcal{R} . This entails rotation of \mathcal{D} and \mathcal{R} , forming counterturns of \mathcal{D} around \mathcal{R} . Because \mathcal{D} is fixed relative to \mathcal{R} at region 2, the counterturns are trapped between regions 1 and 2.

Structures resembling topological traps have been previously observed by electron microscopy and their potential significance for RecA activity has been discussed (Rice et al., 2001; Shan et al., 1996). We believe that the observed structures are most likely heterologous topological traps, in which the points of interaction between the two substrates are chiefly heterologous alignments or false-hits with at most one homologous alignment between any pair of substrates. A consideration of heterologous topological traps would require a very different treatment, and is beyond the scope of this article. We address here only homologous topological traps, in which all interactions between the substrates are homologous. For the remainder of this article, *topological trap* is to be understood to mean homologous topological trap only.

Such topological trapping can occur only if homologous strand exchange is able to begin at two points. If initiation of homologous strand exchange between the substrates at one point somehow prevents the initiation of homologous strand exchange at any other point along the same two substrates, topological trapping will never occur. To exploit this fact, the strand exchange machinery must impart to the entire length of the homology the information that homologous strand exchange has been initiated between them. The problem thus becomes finding a means of communicating over large distances the fact that no further homologous strand exchange process should be initiated.

We propose that a key function of the extended filament structure is to prevent topological trapping. Both the stiffness of the filament and the stretching of the DNA within it are essential to accomplishing this. Here, we give a qualitative explanation of how this works. A more detailed treatment appears in Appendix B.

Consider homologous exchange between \mathcal{D} and \mathcal{R} which has extended to encompass a segment of the substrates which we call the *first region*. Symmetry makes it sufficient to consider only the sections of the substrates to one side of the first region. Number the base(pair)s in ascending order to the right, beginning with the rightmost base(pair) in the first region. If the second contact is at base N on \mathcal{D} , it will be homologously aligned only if it is at basepair N on \mathcal{R} . Formation of a second region produces a double-hit loop, half of which is composed of ssDNA from \mathcal{R} and the remainder of dsDNA from \mathcal{D} . The second region is distinct if the RecA filament passes through the double-hit loop at least once, otherwise it only extends into the first region. We estimate the minimum work to form a double-hit loop for different N and show that the resulting Boltzmann factor is too small to permit the structure to form.

When the RecA filament passes through the double-hit loop $M = 1$ times we consider N small. This forms *structure 1*, shown in Fig. 6 A. As \mathcal{R} is very stiff it will behave as a rigid rod for small N , so we ignore any bending of \mathcal{R} . To align the N^{th} base(pair)s of \mathcal{D} and \mathcal{R} then requires stretching and bending \mathcal{D} . We ignore the work necessary to bend \mathcal{D} and

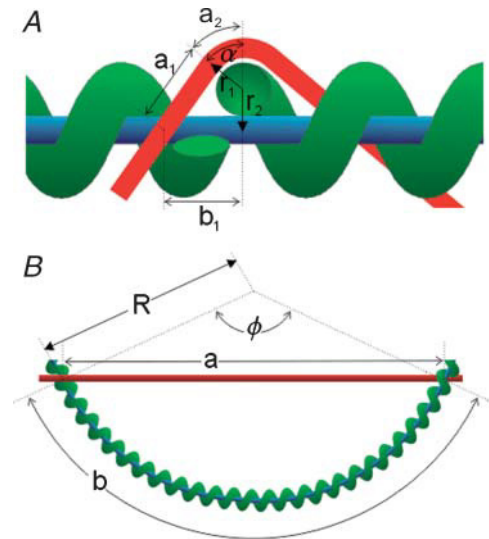


FIGURE 6 Two homologous alignments between the same substrates. (Red curve) dsDNA molecule (\mathcal{D}); (blue line) ssDNA; and (green helix) RecA filament (\mathcal{D}). (A) Structure 1 (small N). The helix is cut to allow clearer labeling of the figure. The red and blue curves form the double-hit loop. The RecA filament (green) passes through this once. Because \mathcal{R} is much stiffer than \mathcal{D} , we assume \mathcal{R} does not bend significantly on this length scale. (B) Structure 2 (large N). Now the RecA filament passes through the double-hit loop many times. Because \mathcal{R} is longer here, we include bending with a radius of curvature R . We ignore the bending of \mathcal{D} necessary to enter the RecA filament.

consider only the work needed to stretch it. This underestimates the work, producing a lower bound.

The minimum work $W^*(\eta_{\mathcal{R}})$ required to form structure 1 is calculated in Appendix B and plotted as a function of the stretching factor $\eta_{\mathcal{R}}$ in Fig. 7. W^* increases with $\eta_{\mathcal{R}}$, making structure 1 more difficult to form. For RecA, $\eta_{\mathcal{R}} = 1.5$ and the work is minimized by $\alpha^* = 0.84$ and $N^* \approx 25$. We call this the *minimal structure 1*. The energy of this is $\sim W^* \approx 170 k_B T$. The Boltzmann factor is $e^{-E_i/k_B T}$, so the probability of forming structure 1 is on the order of e^{-170} . Formation of

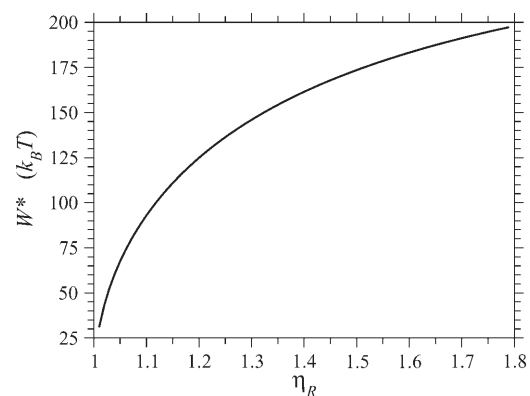


FIGURE 7 The minimum work W^* required to form structure 1 as a function of the stretching factor $\eta_{\mathcal{R}}$ of the DNA within substrate \mathcal{R} relative to B-form DNA.

a second homologous alignment therefore does not occur for small N as a result of thermal fluctuations.

When the RecA filament passes through the double-hit loop $M \geq 2$ times we must include the bending of \mathcal{R} , but we ignore the bending of \mathcal{D} as it enters or leaves the RecA filament. This shortens the path for \mathcal{D} and lengthens the radius of curvature for \mathcal{R} , both of which reduce the work, giving a lower limit on the work required to achieve the second alignment. This is *structure 2*, shown in Fig. 6 B. The minimum work W^* to form structure 2 is calculated in Appendix B.

If $30 \leq N \leq 100$, the expression for W^* is complicated by the division of the parameter space into different regions. Fig. 8, A and B, are more informative. Fig. 8 A shows $W^*(N)$ with the persistence length of \mathcal{R} fixed at the physiological value of $\xi_{\mathcal{R}} = 860$ nm for various values of $\eta_{\mathcal{R}}$. The curves all have a similar shape, with the amplitudes increasing with increasing $\eta_{\mathcal{R}}$. For small N , the minimal structure involves stretching \mathcal{D} without significantly bending \mathcal{R} . In this case,

$$W^* \approx F_0 a_0 (\eta_{\mathcal{D}-1}) N = F_0 a_0 (\eta_{\mathcal{R}} - 1) N, \quad (6)$$

where F_0 is a known constant. Consequently, the left edges of these curves in Fig. 8 A are spaced linearly in proportion to $\eta_{\mathcal{R}}$. For the physiological value of $\eta_{\mathcal{R}} = 1.5$, the minimal

structure at first involves only stretching \mathcal{D} , and W^* forms a straight line which increases with N as in Eq. 6. This is because the larger N is, the more basepairs in \mathcal{D} must be stretched. As N increases, the length of the segment of \mathcal{R} between regions 1 and 2 increases and the work required to bend \mathcal{R} decreases. At $N \simeq 48$, it becomes comparable to the work required to stretch \mathcal{D} and the minimal structure becomes a combination of bending \mathcal{R} and stretching \mathcal{D} . As N increases further, the work to bend \mathcal{R} drops further, whereas the work to stretch \mathcal{D} continues to increase so bending \mathcal{R} becomes a steadily larger part of the process. W^* curves downward as this happens, and by $N \simeq 60$ the minimal structure involves only bending \mathcal{R} and no stretching \mathcal{D} . From here on, W^* decreases when N increases as $W^* \propto N^{-2}$.

The location of the transition from the stretching \mathcal{D} regime with $W^* \propto N$ to the bending \mathcal{R} regime with $W^* \propto N^{-2}$ is influenced by $\eta_{\mathcal{R}}$. For larger $\eta_{\mathcal{R}}$, greater work is required to sufficiently stretch a given number of basepairs, making W^* larger for larger values of $\eta_{\mathcal{R}}$. The work required to stretch \mathcal{D} thus becomes comparable to the work required to bend \mathcal{R} at smaller values of N , and the peak value of W^* occurs at smaller N for larger values of $\eta_{\mathcal{R}}$.

Fig. 8 B shows $W^*(N)$, with $\eta_{\mathcal{R}}$ fixed at the physiological value of $\eta_{\mathcal{R}} = 1.5$, for various values of $\xi_{\mathcal{R}}$. The notable points here are that increasing $\xi_{\mathcal{R}}$ makes it more difficult to bend \mathcal{R} , and therefore W^* increases as $\xi_{\mathcal{R}}$ increases. This also means that higher values of $\xi_{\mathcal{R}}$ push the transition from the stretching \mathcal{D} regime to the bending \mathcal{R} regime to higher values of N . For $\xi_{\mathcal{R}} = 500$ nm the transition occurs at values of N which are off the left side of Fig. 8 B.

From these figures and from the calculation we see that for N , W^* remains too high for structure 2 to form as a result of thermal fluctuations. This is only the case because of the large values of $\xi_{\mathcal{R}}$ and $\eta_{\mathcal{R}}$, since smaller values of either or both of these decrease W^* and make the structure more accessible to random thermal processes. Since our calculation has produced only a lower limit on W^* we can be confident that this conclusion is valid for values of N which are at least this large.

For N , the minimal structure is dominated by bending \mathcal{R} , and we can ignore stretching of \mathcal{D} . By contrast, we can no longer ignore the work required to separate \mathcal{D} and \mathcal{R} against the nonspecific attractive force which initially brought them into alignment. This work will be proportional to the length of the substrates between regions one and two. We use ε for the constant of proportionality. The work to form structure 2 in this case is calculated in Appendix B. The function is found to have a minimum with respect to N at

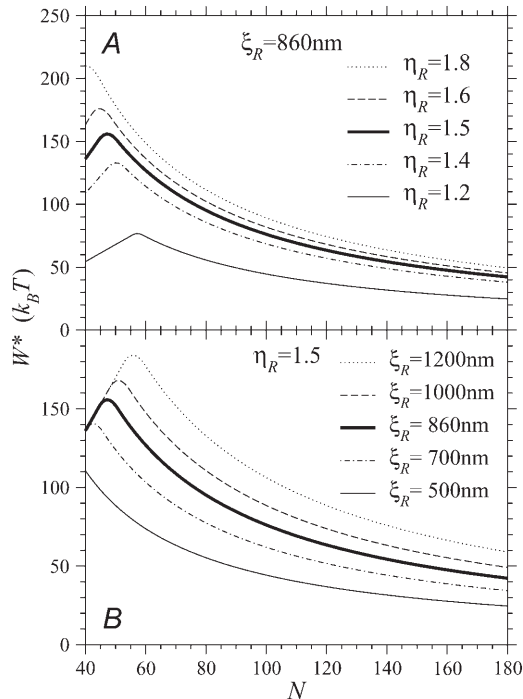


FIGURE 8 W^* as a function of N . (A) For various stretching factors $\eta_{\mathcal{R}}$ with the physiological persistence length of $\xi_{\mathcal{R}} = 860$ nm. The thick curve shows the physiological case $\eta_{\mathcal{R}} = 1.5$. (B) For various persistence lengths $\xi_{\mathcal{R}}$ with the physiological stretching factor of $\eta_{\mathcal{R}} = 1.5$. The thick curve shows the physiological case $\xi_{\mathcal{R}} = 860$ nm.

$$N^* = \frac{2}{a_0 \eta_{\mathcal{R}}} \sqrt{\left(\frac{\xi_{\mathcal{R}}}{\varepsilon}\right) \left(5 - \sqrt{\frac{30}{\eta_{\mathcal{R}}} - 5}\right)} + 1. \quad (7)$$

This value for N produces the minimal structure 2 for large N . The behavior of N^* in Eq. 7 with respect to η_R , as shown in Fig. 9. N^* has a maximum with respect to η_R . This occurs at some value η_R^* which is found in the Appendices to be $\eta_R^* \simeq 1.58$, strikingly close to the physiological value of $\eta_R = 1.5$. This value maximizes the distance between the first and second regions for which the work required to form structure 2 is minimized.

Upon using N^* in the expression for work, we obtain the minimum value

$$W^* = 4 k_B T \sqrt{(\xi_R \varepsilon) \left(5 - \sqrt{\frac{30}{\eta_R} - 5} \right)}, \quad (8)$$

shown in Fig. 10. The work W^* increases as the square root of ε . For the physiological values of η_R and ξ_R , W^* already reaches $\sim 50 k_B T$ by $\varepsilon \simeq 0.2 \text{ nm}^{-1}$, so structure 2 will not form as a result of thermal motions for the physiological values of η_R and ξ_R when $\varepsilon \geq 0.2 \text{ nm}^{-1}$. W^* also increases as the square root of ξ_R , and increases with η_R in a more complicated fashion. Sufficiently small values of η_R or ξ_R would produce values of W^* which would be more accessible to thermal energies.

The probability that a second, local region of homology aligns N basepairs away is $\propto N^{-6}$. This small probability forms an “entropic” barrier to formation of a second region of homology. However, since $\eta_R \simeq \eta_R^*$, the value of N for which the energetic obstacle to alignment is smallest, N^* , is made as large as possible, maximizing the entropic obstacle to alignment. The two processes are “tuned” to work in a complementary fashion, providing a further form of selection pressure for the maintenance of stretching of the DNA by the RecA filament and, specifically, to a value close to the physiological value of $\eta_R = 1.5$.

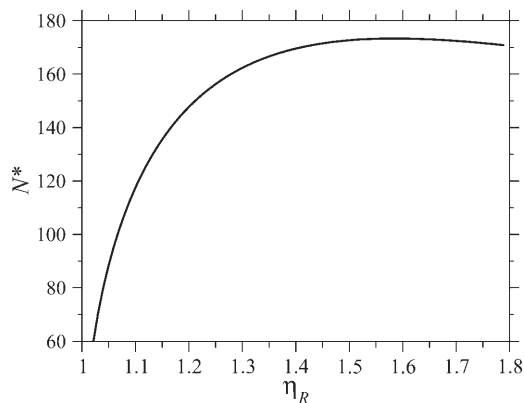


FIGURE 9 N^* , the value of N at which the energy required to form structure 2 is minimized, as a function of the stretching factor η_R . Note that N^* has a maximum with respect to η_R at $\eta_R \simeq 1.58$, very near the physiological value of $\eta_R = 1.5$.

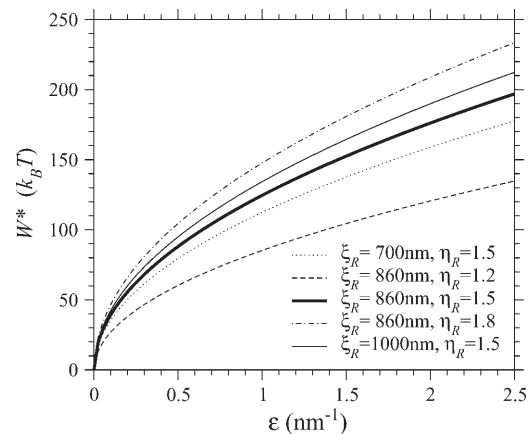


FIGURE 10 The minimum work W^* to form structure 2 for $N > 60$ as a function of the constant of proportionality ε for various stretching factors η_R and persistence lengths ξ_R . The thick curve shows the physiological case $\eta_R = 1.5$ and $\xi_R = 860 \text{ nm}$.

SUMMARY AND CONCLUSION

RecA-facilitated homologous recombination derives two advantages from the stiff extended filament and the stretching of the DNA within the filament. The first is a great increase in the efficiency of the homology search and recognition process. This is a consequence of the increase in σ , the target size for homologous alignment between the substrates. For a region of homology J bases in length, stretched by a factor η_R , we have

$$\sigma = [(J - 1)(\eta_R - 1) + 1]a_0, \quad (9)$$

where a_0 is the spacing of B-form DNA. For $\eta_R \simeq 3/2$ this is $\sigma = (J + 1)a_0/2$. This huge σ allows large segments of the substrates to be checked for homology without the need for sliding. The second advantage is the prevention of homologous topological trapping. Molecules undergoing homologous strand exchange are kept in close proximity by the region of hybrid DNA being formed. This greatly enhances the probability that they will contact each other at additional points. Without the stretching of the DNA within the extended filament, these secondary contacts would often be in homologous alignment and capable of initiating a second homologous strand exchange reaction. This would lead to a trapped state in which a region of counter-wound DNA is trapped between two regions of hybrid DNA, preventing completion of the exchange reaction.

The extended filament prevents homologous alignment at secondary contacts. Homologously aligned secondary contacts can only form through some combination of stretching the DNA external to the filament and bending the filament itself. For moderate distances from the point at which homologous exchange is occurring, thermal fluctuations are incapable of sufficiently bending or stretching the filament

for a second hit to occur. For larger distances, thermal fluctuations are also unlikely to separate the locally aligned strands held together by nonspecific electrostatic forces. The interplay between stiffness of the filament and the stretching of the DNA within it ensures that homologous strand exchange between two substrate molecules is initiated at only one point.

Both the stiffness of the filament and the extension of the DNA within it are necessary features of the recombination apparatus. Without them, locating and aligning regions of homology between two DNA molecules would be a slow and inefficient process, and the exchange reaction would be prone to topological traps which would prevent completion of the reaction and resolution of the products. These effects provide a selection pressure to preserve the extended filament as a feature of homologous DNA recombination facilitated by RecA and its homologs.

APPENDIX A: KINETIC PERFECTION IN THE HOMOLOGY SEARCH

Consider \mathcal{D} and \mathcal{R} as defined in the text. The center-to-center base(pair) spacing is a_0 in \mathcal{D} and $\eta_{\mathcal{R}} a_0$ in \mathcal{R} , so an N basepair segment has length

$$L_{\mathcal{D}} = (N - 1)a_0, \quad (\text{A1})$$

$$L_{\mathcal{R}} = (N - 1)\eta_{\mathcal{R}}a_0. \quad (\text{A2})$$

If \mathcal{D} and \mathcal{R} intersect at a point and rotate around this point until parallel, at most one base within \mathcal{R} and a basepair in \mathcal{D} will be homologously aligned. We wish to determine if this parallel orientation produces an alignment.

To define “aligned” we consider only the longitudinal positions of \mathcal{D} and \mathcal{R} . A base on \mathcal{R} is aligned with whichever basepair on \mathcal{D} is closest to it in the longitudinal direction. When the aligned base and basepair are homologous they are homologously aligned.

To quantify this, first note that the center-to-center distance between consecutive basepairs of \mathcal{D} is a_0 . Let the longitudinal distance between the centers of the k^{th} basepair on \mathcal{D} and the l^{th} base on \mathcal{R} be $\delta_{k,l}$. These are aligned if

$$|\delta_{k,l}| < \frac{a_0}{2}. \quad (\text{A3})$$

Denote the positions of initial contact by $x_{\mathcal{D}}$ along \mathcal{D} and by $x_{\mathcal{R}}$ along \mathcal{R} . The parallel orientation is achieved by rotating around $x_{\mathcal{D}}$ and $x_{\mathcal{R}}$, so these completely determine the relative positions of \mathcal{D} and \mathcal{R} once they are parallel. It is sufficient to fix $x_{\mathcal{R}}$ and ask what values of $x_{\mathcal{D}}$ produce a homologous alignment in the parallel orientation. If the position $x_{\mathcal{D}}$ is moved a distance d along \mathcal{D} , \mathcal{D} will be displaced relative to \mathcal{R} by this same distance d in the parallel orientation. The range of $x_{\mathcal{D}}$ which produces a homologous alignment is therefore the same as the range of longitudinal positions of \mathcal{D} relative to \mathcal{R} which will produce an homologous alignment.

Consider \mathcal{D} and \mathcal{R} as shown in Fig. 11 A. Number the basepairs beginning with 1 at the leftmost basepair in the region of homology. Here, \mathcal{D} is as far to the left as possible while maintaining a homologous alignment between \mathcal{D} and \mathcal{R} . The homologous alignment is between base and basepair 1, and the center of basepair 1 is $a_0/2$ to the left of the center of base 1. Fig. 11 B shows an enlarged view of this.

Moving \mathcal{D} to the right by $a_0/2$ gives Fig. 11 C, where the centers of basepair 1 and base 1 are exactly aligned. Moving \mathcal{D} to the right by $(\eta_{\mathcal{R}} - 1)a_0$ then produces Fig. 11 D, in which the centers of basepair 2 and base 2 are exactly aligned. Each displacement of \mathcal{D} to the right by

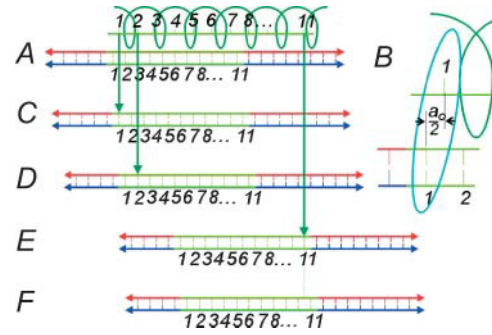


FIGURE 11 Dark green helix, RecA filament, light green line inside helix, ssDNA (\mathcal{R}); red and blue lines with light green segments, dsDNA (\mathcal{D}). \mathcal{R} is arbitrarily chosen to be 11 bases in length. The light green region of \mathcal{D} is homologous to \mathcal{R} . (A) \mathcal{D} is as far to the left as possible while maintaining a homologous alignment between \mathcal{D} and \mathcal{R} (between base and basepair 1). (B) Closeup of the homologously aligned base and basepair in A. (C) Effect of shifting \mathcal{D} to the right a distance $a_0/2$. (D) Effect of shifting \mathcal{D} further to the right, this time by a distance a_0 . (E) Effect of nine more consecutive shifts of \mathcal{D} to the right by a distance a_0 . (F) Effect of a final shift of \mathcal{D} to the right by a distance $a_0/2$.

$(\eta_{\mathcal{R}} - 1)a_0$ now increments by 1 the base and basepair whose centers are exactly aligned. Starting with Fig. 11 C and repeating this motion $(J - 1)$ times produces Fig. 11 E, in which the centers of basepair J and base J are exactly aligned. A final movement of \mathcal{D} to the right by $a_0/2$ produces Fig. 11 F, where \mathcal{D} is as far to the right relative to \mathcal{R} as is possible while still maintaining a homologous alignment between them.

The target size σ is the range of longitudinal positions of \mathcal{D} relative to \mathcal{R} which produces a homologous alignment between them. This is the change in the position of \mathcal{D} in going from Fig. 11 A–F, which is given in Eq. 1. If $\eta_{\mathcal{R}}$ has the known RecA value of $\eta_{\mathcal{R}} = 3/2$, we get $\sigma = (J + 1)a_0/2$, which scales as the length of the region of homology.

From σ we can estimate the reaction rate. With no sliding (Adzuma, 1998), diffusion limits the maximum “on rate” k_a for the reaction to the Debye-Smoluchowski rate. Our target is cylindrical, but the magnitude should be reasonably approximated if we substitute our target size for the diameter of a spherical target, $r \rightarrow \sigma/2$, giving $k_a \simeq 2\pi D_3 \sigma$, where D_3 is the three-dimensional diffusion constant. The length of \mathcal{R} is

$$\ell = (J - 1)\eta_{\mathcal{R}}a_0 \simeq J\eta_{\mathcal{R}}a_0. \quad (\text{A4})$$

Upon substituting this for the diameter of a spherical molecule, $r \rightarrow \ell/2$, the three-dimensional diffusion constant in a solvent with viscosity η , becomes $D_3 = k_B T / (3\pi\eta\ell)$. Assuming J is reasonably large, $\sigma \simeq J(\eta_{\mathcal{R}} - 1)a_0$, and k_a is given by Eq. 5.

APPENDIX B: DOUBLE-HIT PROBABILITY IN THE RecA RECOMBINATION SYSTEM

If N is small we consider structure 1 as shown in Fig. 6 A. Define $\eta_{\mathcal{D}}$ and $\eta_{\mathcal{R}}$ as the extension of \mathcal{D} and \mathcal{R} relative to the length of B-form DNA. Let a_0 be the spacing of basepairs in B-form DNA and denote the persistence lengths of \mathcal{D} and \mathcal{R} by $\xi_{\mathcal{D}}$ and $\xi_{\mathcal{R}}$, respectively. These have the numerical values $a_0 = 0.34$ nm, $\xi_{\mathcal{D}} \simeq 53$ nm, and $\xi_{\mathcal{R}} \simeq 860$ nm. We model the RecA filament as a cylinder of radius r_{RecA} whose axis follows a helical path of radius $r_2 \simeq 3$ nm. Treating \mathcal{D} as a cylinder of radius $r_{\mathcal{D}}$, the closest approach of the center of \mathcal{D} to the center of \mathcal{R} will be the sum of their radii, which we denote by $r_1 = r_{\text{RecA}} + r_{\mathcal{D}} \simeq 2.5$ nm.

We wish to determine the minimum work required to form structure 1. We ignore bending of \mathcal{R} on this length scale and also ignore the work required to bend \mathcal{D} . We calculate the work solely from the stretching of \mathcal{D} , making our calculation of the work a lower bound.

The angle α can be varied to find the minimal-energy form of structure 1, the form produced with the minimum possible work, subject to two constraints. Steric hindrance between the RecA filament and \mathcal{D} at the points where it enters the RecA filament requires $\alpha \geq \pi/5$, whereas other physical considerations show that the minimum energy can occur only for $\alpha \leq \pi/2$.

Both regions can simultaneously be in homologous alignment only if N satisfies

$$2(a_1 + a_2) = a_0 \eta_D (N - 1) \quad 2b_1 = a_0 \eta_R (N - 1). \quad (\text{B1})$$

Since $a_2 = r_1 \alpha$, trigonometry demands

$$a_1 = \frac{r_1 \cos \alpha + r_2}{\sin \alpha} \quad \text{and} \quad b_1 = \frac{r_2}{\tan \alpha} + \frac{r_1}{\sin \alpha}. \quad (\text{B2})$$

Using Eqs. B1 and B2 we find

$$(N - 1) = \frac{2}{a_0 \eta_R} \left(\frac{r_1 + r_2 \cos \alpha}{\sin \alpha} \right) \\ \eta_D = \frac{\eta_R (r_1 \alpha \sin(\alpha) + r_1 \cos(\alpha) + r_2)}{r_1 + r_2 \cos \alpha}. \quad (\text{B3})$$

We make the simplification of assuming that the force to stretch a dsDNA is

$$F_D = \begin{cases} 0 & \text{for } \eta_D < 1 \\ F_0 & \text{for } \eta_D \geq 1 \end{cases}. \quad (\text{B4})$$

Here, $F_0 \approx 20 \text{ k}_B T/\text{nm}$, but we leave this parameter free for the present.

Above $\eta_D \approx 1.8$, Eq. B4 is not valid. Here, the dsDNA melts whereas the force required to stretch it rises rapidly. Further stretching breaks the sugar phosphate backbones of the DNA strands. Although our calculation may produce values of $\eta_D > 1.8$, we are not concerned. We only wish to show that the minimal form of structure 1 does not form as a result of random thermal fluctuations, and our calculation will still accomplish this.

Using Eq. B4 for F_D , the work required to stretch the dsDNA to a final extension η_R (in units of $k_B T$) is

$$W_D = \begin{cases} 0 & \text{for } \eta_D < 1 \\ F_0(\eta_D - 1)a_0(N - 1) & \text{for } \eta_D \geq 1 \end{cases}.$$

Under our approximations, this is the only contribution to the total work. Using the expressions in Eq. B3, W becomes a function of the single variable α as

$$W(\alpha) = \begin{cases} 0 & \text{for } \eta_D < 1 \\ \frac{2F_0 \csc \alpha}{\eta_R} (-r_1 + r_2 \eta_R + (r_1 \eta_R - r_2) \cos \alpha + r_1 \alpha \eta_R \sin \alpha) & \text{for } \eta_D \geq 1 \end{cases}. \quad (\text{B5})$$

Minimizing this with respect to α we find $\alpha^* = \text{arcsec}(\eta_R)$, which gives us the equations

$$W^* = 2F_0 \left(r_2 \sqrt{1 - \frac{1}{\eta_R^2}} + r_1 \text{arcsec}(\eta_R) \right) \\ \eta_D^* = \frac{\eta_R (r_1 + r_2 \eta_R + r_1 \sqrt{\eta_R^2 - 1} \text{arcsec}(\eta_R))}{(r_1 \eta_R + r_2)} \\ N^* = \frac{2}{a_0} \left(\frac{r_1 \eta_R + r_2}{\eta_R \sqrt{(\eta_R)^2 - 1}} \right) + 1. \quad (\text{B6})$$

With the known values of the constants, Eq. B6 becomes

$$W^*(\eta_R) = 24 k_B T \sqrt{1 - \frac{1}{\eta_R^2}} + 20 k_B T \text{arcsec}(\eta_R), \quad (\text{B7})$$

which is plotted in Fig. 7. At the physiological value of $\eta_R = 3/2$ for the RecA system, $\alpha^* = 0.84$, $\eta_D^* = 2.1$, $N^* = 25$, and $W^* = 174 \text{ k}_B T$. The probability of the system being in the minimal form of structure 1 is on the order of e^{-174} . This vanishing probability persists for $\eta_D = \eta_D^* \simeq 1.8$.

For somewhat larger N we consider structure 2 as shown in Fig. 6 B. The angle α no longer enters the calculation directly, and we deal with the angle ϕ . The parameters are subject to the restrictions $1 \leq \eta_R \leq 1.8$, $\eta_D \leq \eta_R$, and $\xi_D \leq \xi_R$.

The exchange regions can simultaneously be in homologous alignment if N satisfies

$$a = a_0 \eta_D (N - 1) \quad \text{and} \quad b = a_0 \eta_R (N - 1), \quad (\text{B8})$$

from which we find $\eta_D/\eta_R = a/b$.

The radius of curvature R for structure \mathcal{R} is related to the opening angle ϕ by

$$b = R\phi, \quad (\text{B9})$$

whereas trigonometry gives

$$a = 2R \sin\left(\frac{\phi}{2}\right). \quad (\text{B10})$$

Upon using the above,

$$\eta_D = \frac{2\eta_R \sin\left(\frac{\phi}{2}\right)}{\phi}. \quad (\text{B11})$$

Using Eqs. B8 and B9 we also find

$$R = \frac{a_0 \eta_R (N - 1)}{\phi}. \quad (\text{B12})$$

We can now calculate the work required to form structure 2. We will vary N and ϕ to minimize this. We can then vary η_R and ξ_R (subject to $\eta_R \geq \eta_D$ and $\xi_R \geq \xi_D$) to examine their effects on the system. The work to form

structure 2 comes from three terms: stretching \mathcal{D} , bending \mathcal{R} , and separating \mathcal{D} from \mathcal{R} against the nonspecific attractive force by which they were initially aligned.

The force required to stretch \mathcal{D} to η_D times its B-form contour length is approximately

$$F_D = \begin{cases} 0 & \text{for } \eta_D < 1.0 \\ F_0 & \text{for } 1.0 \leq \eta_D \leq 1.8. \\ \infty & \text{for } 1.8 \leq \eta_D \end{cases}. \quad (\text{B13})$$

Since $\eta_D \geq 1.8$ is unphysical we impose $F_D(\eta_D \geq 1.8) = \infty$ to ensure that this does not occur. The work to stretch \mathcal{D} is thus

$$W_D = \begin{cases} 0 & \text{for } \eta_D < 1 \\ F_0(\eta_D - 1)a_0(N - 1) & \text{for } 1 \leq \eta_D \leq 1.8 \\ \infty & \text{for } 1.8 \leq \eta_D \end{cases} \quad (\text{B14})$$

Using Eq. B11 and expressing the regime boundaries in terms of ϕ , this becomes (in units of $k_B T$)

$$W_D = \begin{cases} 0 & \text{for } \left(\frac{\sin\left(\frac{\phi}{2}\right)}{\frac{\phi}{2}} \right) \eta_R < 1 \\ F_0 \left(\left(\frac{\sin\left(\frac{\phi}{2}\right)}{\frac{\phi}{2}} \right) \eta_R - 1 \right) a_0(N - 1) & \text{for } 1 \leq \eta_R \left(\frac{\sin\left(\frac{\phi}{2}\right)}{\frac{\phi}{2}} \right) \leq 1.8 \\ \infty & \text{for } 1.8 < \eta_R \left(\frac{\sin\left(\frac{\phi}{2}\right)}{\frac{\phi}{2}} \right) \end{cases} \quad (\text{B15})$$

The work to bend \mathcal{R} into the circular arc in structure 2 is

$$W_R = \frac{1}{2} \kappa_R \int \left(\frac{1}{R(s)} \right)^2 ds = \frac{1}{2} \kappa_R \left(\frac{1}{R} \right)^2 b. \quad (\text{B16})$$

Using Eqs. B9 and B12 and the fact that $\kappa_R \simeq \xi_R$ (in units of $k_B T$), gives

$$W_R = \frac{\xi_R}{2} \left(\frac{\phi^2}{a_0 \eta_R (N - 1)} \right). \quad (\text{B17})$$

There is a nonspecific attractive interaction between \mathcal{D} and \mathcal{R} . The work to pull \mathcal{D} and \mathcal{R} apart is approximately proportional to the length of \mathcal{R} between the exchange regions. For intermediate N we ignore an energetic contribution $\propto \epsilon k_B T a_0 \eta_R N$, which underestimates the work and produces a lower bound,

$$W(\phi) = \begin{cases} W_R & \text{for } \left(\frac{\sin\left(\frac{\phi}{2}\right)}{\frac{\phi}{2}} \right) \eta_R < 1 \\ W_D + W_R & \text{for } 1 \leq \eta_R \left(\frac{\sin\left(\frac{\phi}{2}\right)}{\frac{\phi}{2}} \right) \leq 1.8 \\ \infty & \text{for } 1.8 < \eta_R \left(\frac{\sin\left(\frac{\phi}{2}\right)}{\frac{\phi}{2}} \right) \end{cases} \quad (\text{B18})$$

Upon minimizing the work $W(\phi)$ with respect to the angle ϕ , we find ϕ^* and $W(\phi^*)$ equal to

$$a_0 F_0 (N - 1) \eta_R \left\{ \left(\frac{\eta_R - 1}{\eta_R} \right) + \left(\frac{1}{120} \max \left[0, \min \left[10 - 2 \sqrt{\frac{30}{\eta_R} - 5}, 10 - \frac{120 k_B T \xi_R}{a_0^2 F_0 (N - 1)^2 \eta_R^2} \right] \right] \right. \right. \\ \left. \left. \times \left(\frac{240 k_B T \xi_R}{a_0^2 F_0 (N - 1)^2 \eta_R^2} - 20 \max \left[0, \min \left[10 - 2 \sqrt{\frac{30}{\eta_R} - 5}, 10 - \frac{120 k_B T \xi_R}{a_0^2 F_0 (N - 1)^2 \eta_R^2} \right] \right] \right) \right\}. \quad (\text{B19})$$

This function is plotted in Fig. 8, *A* and *B*, using standard values for a_0 , F_0 , and $k_B T$. From Fig. 8, *A* and *B*, it is clear that by $N \approx 60$ the minimal form of structure 2 is dominated by bending \mathcal{R} . For large N , (≥ 60), the interaction energy can no longer be ignored, but now we always have $(\sin(\frac{\phi}{2})/\frac{\phi}{2})\eta_R < 1$. The total work thus simplifies to

$$W = \frac{4 k_B T \left(5 - \sqrt{\frac{30}{\eta_R} - 5} \right) \xi_R}{a_0 (N - 1) \eta_R} + \epsilon k_B T a_0 \eta_R (N - 1). \quad (\text{B20})$$

The separation between exchange regions which minimizes this and the corresponding minimum work is

$$N^* = \frac{2}{a_0 \eta_R} \sqrt{\left(\frac{\xi_R}{\epsilon} \right) \left(5 - \sqrt{\frac{30}{\eta_R} - 5} \right)} \quad (\text{B21})$$

and

$$W^* = 4 k_B T \sqrt{\left(\frac{\xi_R}{\epsilon} \right) \left(5 - \sqrt{\frac{30}{\eta_R} - 5} \right)}. \quad (\text{B22})$$

These are plotted in Figs. 9 and 10, respectively. We also note that N^* is monotonic with respect to the ξ_R and ϵ , but has a maximum with respect to η_R at $\eta_R^* = \frac{5}{4}(3 + \sqrt{3}) \simeq 1.58$.

K.K. and T.C. acknowledge support from the National Science Foundation through grant DMS-0206733. K.K. and R.B. acknowledge support from the University of California-Los Angeles Dean's Funds.

REFERENCES

- Adzuma, K. 1998. No sliding during homology search by RecA protein. *J. Biol. Chem.* 273:31565–31573.
- Gupta, R. C., E. Folta-Stogniew, S. O'Malley, M. Takahashi, and C. M. Radding. 1999. Rapid exchange of A:T basepairs is essential for recognition of DNA homology by human Rad51 recombination protein. *Mol. Cell.* 4:705–714.
- Hegner, M., S. B. Smith, and C. Bustamante. 1999. Polymerization and mechanical properties of single RecA-DNA filaments. *Proc. Natl. Acad. Sci. USA.* 96:10109–10114.
- Heuser, J., and J. Griffith. 1989. Visualization of RecA protein and its complexes with DNA by quick-freeze/deep-etch electron microscopy. *J. Mol. Biol.* 210:473–484.
- Honigberg, S., and C. Radding. 1998. The mechanics of winding and unwinding helices in recombination: torsional stress associated with strand transfer promoted by RecA protein. *Cell.* 54:525–532.
- Karlin, S., and L. Brocchieri. 1996. Evolutionary conservation of RecA genes in relation to protein structure and function. *J. Bacteriol.* 178:1881–1894.
- Nishinaka, T., Y. Ito, S. Yokoyama, and T. Shibata. 1997. An extended DNA structure through deoxyribo-base stacking induced by RecA protein. *Proc. Natl. Acad. Sci. USA.* 94:6623–6628.
- Nishinaka, T., A. Shinohara, Y. Ito, S. Yokoyama, and T. Shibata. 1998. Basepair switching by interconversion of sugar puckers in DNA extended by proteins of RecA-family: a model for homology search in homologous genetic recombination. *Proc. Natl. Acad. Sci. USA.* 95:11071–11076.
- Roca, A. I., and M. M. Cox. 1997. RecA protein: structure, function, and role in recombinational DNA repair. *Prog. Nucleic Acids Res. Mol. Biol.* 56:129–223.
- Rice, K. P., A. L. Egger, P. Sung, and M. M. Cox. 2001. DNA pairing and strand exchange by the *Escherichia coli* RecA and yeast Rad51 proteins without ATP hydrolysis. *J. Biol. Chem.* 276:38570–38581.
- Shan, Q., M. M. Cox, and R. B. Inman. 1996. DNA strand exchange promoted by RecA K72R. *J. Biol. Chem.* 271:5712–5724.
- Sigurdsson, S., K. Trujillo, B.-W. Song, S. Stratton, and P. Sung. 2001. Basis for avid homologous DNA strand exchange by human Rad51 and RPA. *J. Biol. Chem.* 276:8798–8806.
- Takahashi, M., and B. Norden. 1994. Structure of RecA-DNA complex and mechanism of DNA strand exchange reaction in homologous recombination. *Adv. Biophys.* 30:1–35.
- Yu, X., and E. H. Egelman. 1992. Structural data suggest that the active and inactive forms of the RecA filament are not simply interconvertible. *J. Mol. Biol.* 227:334–346.