STRUCTURAL BIOLOGY

RecA filament maintains structural integrity using ATP-driven internal dynamics

Sung Hyun Kim,[1,2] TakKyoon Ahn,[1] Tao Ju Cui,[2] Sweeny Chauhan,[2] Jaeyoung Sung,[3] Chirlmin Joo,[2] Doseok Kim[1]*

At the core of homologous DNA repair, RecA catalyzes the strand exchange reaction. This process is initiated by a RecA loading protein, which nucleates clusters of RecA proteins on single-stranded DNA. Each cluster grows to cover the single-stranded DNA but may leave 1- to 2-nucleotide (nt) gaps between the clusters due to three different structural phases of the nucleoprotein filaments. It remains to be revealed how RecA proteins eliminate the gaps to make a seamless kilobase-long filament. We develop a single-molecule fluorescence assay to observe the novel internal dynamics of the RecA filament. We directly observe the structural phases of individual RecA filaments and find that RecA proteins move their positions along the substrate DNA to change the phase of the filament. This reorganization process, which is a prerequisite step for interjoining of two adjacent clusters, requires adenosine triphosphate hydrolysis and is tightly regulated by the recombination hotspot, Chi. Furthermore, RecA proteins recognize and self-align to a 3-nt-period sequence pattern of TGG. This sequence-dependent phase bias may help the RecA filament to maintain structural integrity within the kilobase-long filament for accurate homology search and strand exchange reaction.

INTRODUCTION

RecA and its homologs are critical in maintaining the genetic integrity of the cell (*1*–*3*). In bacteria, RecA is essential for the cell viability against DNA damage (*4*). Its eukaryotic homologs are instrumental in preventing genetic disorders such as cancer (*5*). RecA mediates homologous recombination by forming a helical filament on single-stranded (ss) DNA (*6*, *7*). In vivo, the RecA loading machinery, such as RecBCD and RecFOR, initiates the RecA-mediated DNA strand exchange by generating ssDNA at a DNA damage site and by loading RecA monomers onto this nascent ssDNA (*8*–*10*). In case of RecBCD, it nucleates clusters of RecA proteins as it translocates along the DNA from 3′ to 5′. However, these RecA clusters grow mainly from 5′ to 3′ by consecutive addition of monomers to the 3′ end (*11*). This opposite directionality between RecBCD translocation and RecA filament growth may result in multiple patches of a filament on the same ssDNA that need to merge.

The ssDNA molecule lies at the central axis of the RecA filament and is lengthened by 1.5-fold compared to the canonical B-form DNA (*7*, *12*). A RecA monomer within the filament occupies a group of three bases (triplet) and generates a gap separating the triplet from its neighboring triplets (fig. S1) (*12*). Whereas the three bases of a triplet make the canonical B-form–like structure, the large separation between the triplets leads to the 1.5-fold elongation. The triplet serves as a fundamental structural unit for homology searching and strand exchange reaction (*13*). Because each RecA monomer in the filament occupies three bases of DNA, there can be three different structural phases of the filament in terms of the relative position (1 or 2 nt shifted from each other) between the RecA monomer and the DNA (Fig. 1 and fig. S1) (*14*, *15*).

If the RecA loading machinery makes more than one nucleation cluster on one ssDNA as it translocates, then there may remain 1- or 2-nucleotide (nt) gaps between adjacent patches due to the difference in the structural phases of the nucleation clusters. How this faulty structure is consolidated for the success of further strand exchange reaction is not understood. Such a disconnected filament structure is observed only when a nonhydrolysable adenosine triphosphate (ATP) analog, ATPγS, is used (*16*). This hints that a RecA filament is capable of switching its structural phase using an unknown ATP-coupled process.

Here, we develop a single-molecule fluorescence assay to investigate the structural phases of the RecA filament (*17*). By exploiting the unique nucleotide triplet structure of the embedded DNA within the RecA filament, we directly measure the three different structural phases of a RecA filament in real time via fluorescence resonance energy transfer (FRET) (*18*). We find that, powered by ATP hydrolysis, a RecA filament can change its phase presumably via allosteric changes of RecA proteins within a filament and resulting lateral movement along its substrate DNA. This internal filament dynamic is influenced by the Chi sequence (the recombination hotspot, 5′-GCTGGTGG-3′), making the filament biased to a specific phase. This Chi sequence–induced bias is further extended to a more general sequence-dependent behavior of the RecA filament that recognizes the 3-nt-period sequence pattern of TGG, which may serves as a mechanistic basis to keep a kilobase-long filament in a single phase.

RESULTS

Single-molecule detection of the structural phases of RecA filament

To visualize three different structural phases of a RecA filament, we designed a single-molecule assay by exploiting a unique structure of DNA within the filament. Chen *et al.* showed that a group of 3 nt that are bound by a RecA monomer is separated from the neighboring groups by a 0.78-nm gap (fig. S1A) (*12*). This unique structural feature enabled us to determine the phase of a filament by counting the number of gaps between any two points on DNA: If the two points are separated by (3N − 1) nt or (3N − 2) nt, where N is a positive integer, then the number of gaps between the two points changes depending on the phase of the filament. For example, as illustrated in Fig. 1A and fig. S1B, when two positions are 8 nt apart, the number of the gaps is either two or three depending on the phase. Likewise, with a 10-nt separation, the number

[1]Department of Physics and Interdisciplinary Program of Integrated Biotechnology, Sogang University, Seoul, Republic of Korea. [2]Department of Bionanoscience, Kavli Institute of NanoScience, Delft University of Technology, 2628 CJ Delft, Netherlands. [3]Department of Chemistry, Chung-Ang University, Seoul, Republic of Korea.
*Corresponding author. Email: doseok@sogang.ac.kr

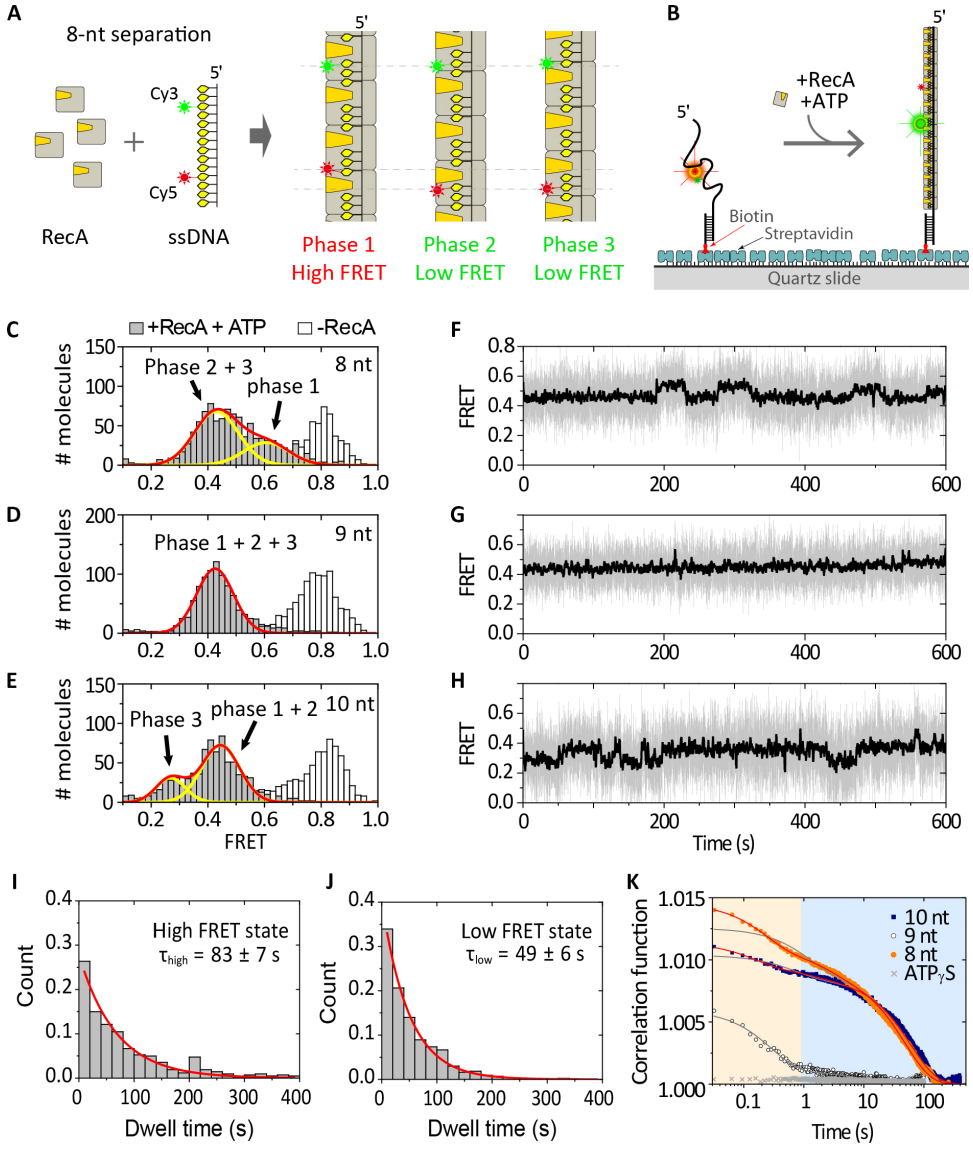

Fig. 1. Real-time observation of RecA filament phase shifting dynamics. (**A**) Schematic of phase detection of RecA filament. The number of the gaps between a dye pair is determined by the phase of the filament and can be read by FRET efficiency. (**B**) Schematic of DNA immobilization on the polyethylene glycol–coated quartz slide via a streptavidin-biotin linker. (**C** to **E**) Population distributions of single-molecule FRET from (C) an 8-nt, (D) a 9-nt, and (B) a 10-nt separation of the dye pair. Gray bars are the distribution observed in the presence of 1 μM RecA and 1 mM ATP, and empty bars are obtained from bare DNA in the absence of RecA. Solid lines are Gaussian fits. (**F** to **H**) Representative single-molecule traces obtained from (F) an 8-nt, (G) a 9-nt, and (H) a 10-nt dye separation. Time traces with 32-ms time resolution (gray) were depicted with their moving average with a 1-s time window (black). (**I** to **J**) Dwell time distributions of high (I) and low (J) FRET states with a 10-nt dye separation. Solid lines are single-exponential fits. A manually given threshold value was applied to each trace to determine the transition points, and 150 traces were processed to build a dwell time histogram. The average dwell times with SEs were obtained from three independent data sets. (**K**) Averaged time correlation functions of single-molecule traces obtained with 8 nt (orange filled circle), 9 nt (black open circle), and 10 nt (blue filled box) in the presence of ATP and 10 nt in the presence of ATPγS (gray crosses). Solid lines are two exponential (gray) and three exponential fits (red) of the correlation functions (fit parameters are summarized in table S1).

of the gaps is either three or four. On the other hand, when separated by 9 nt, the number of the gaps is three, irrespective of the phase. Therefore, the distance of the two points can be a direct measure of the phase of the filament.

To measure the distance between two points using FRET, we labeled ssDNA with a pair of donor and acceptor fluorophores (Fig. 1A and fig. S1B). We hybridized the ssDNA with biotinylated ssDNA and immobilized it on a quartz surface using a biotin-streptavidin conjugation (Fig. 1B). We recorded fluorescence signals using total internal reflection microscopy and determined FRET efficiency from the ratio of the two fluorescent signals (*19*). When two fluorophores were separated by 8 nt (Fig. 1A), high FRET would arise from one phase (phase 1 with two gaps) and low FRET from the other two phases (phases 2 and 3 with three gaps). For a population analysis, we collected fluorescence intensities of individual dye pairs from typically ~1000 DNA molecules for 1 s. We observed two populations (*E*, ~0.45 and 0.60), which were biased to low FRET with the ratio of 2.3:1 (low/high) (Fig. 1C). When two fluorophores were separated by 10 nt, more populated was in the high FRET state (low/high ratio = 1:2.7; Fig. 1E) that arose from two phases. On the other hand, only one population was observed with a 9-nt separation as expected (Fig. 1D).

Note that the RecA filament keeps its extended/active conformation in our ATP-rich buffer condition; therefore, we exclude the possibility that filament pitch length changes due to nucleotide cofactor exchange (*20*).

A RecA filament changes its structural phase using ATP energy

We then questioned whether a RecA filament changes its phase via unknown internal dynamics. To track how the FRET efficiency of each molecule evolved over time, we recorded fluorescence images with a frame rate at 30 Hz. Individual filaments showed switching between two FRET states when formed on the constructs with an 8- and a 10-nt separation (Fig. 1, F and H, and fig. S2, A to C). In contrast, no clear transitions were observed from a 9-nt separation (Fig. 1G and fig. S2B), because all three phases were degenerated in a single FRET state. The process was also suppressed when ATPγS, a nonhydrolysable ATP analog, was used as a cofactor (fig. S3), indicating that this process is triggered by ATP hydrolysis.

We determined the kinetic rate of the phase switching by analyzing the dwell time of each FRET state determined by a threshold algorithm applied to individual single-molecule time traces from the construct of a 10-nt separation. The dwell time histograms were fitted with a single-exponential function with mean dwell times of $\Delta\tau_{low} = 49 \pm 6$ s ($k_{low\ to\ high}$, 0.02 s^{-1}) and $\Delta\tau_{high} = 83 \pm 7$ s ($k_{high\ to\ low}$, 0.012 s^{-1}) for the low and high FRET states, respectively (Fig. 1, I and J). Alternatively, we obtained similar kinetic rates (fig. S4; $k_{low\ to\ high} = 0.019 \pm 0.002\ s^{-1}$, $k_{high\ to\ low} = 0.011 \pm 0.001\ s^{-1}$) from hidden Markov analysis of the time traces (for details, see Materials and Methods) (*21*). Consistent with the population bias observed in Fig. 1E, the filament stayed longer in the high FRET state. Although there are two phases degenerated in the high FRET state, there was no sign of a rising phase in the dwell time histogram (Fig. 1I), suggesting that the transitions occur in a random forward-and-reverse manner rather than sequential or directional transitions between the states.

We further analyzed the single-molecule traces by using time-correlation analysis (*22*). We observed two decay processes in the correlation functions obtained from the constructs with 8- and 10-nt separations (Fig. 1K). With the 9-nt construct, the slow process at a time scale of 1 to 100 s was not detectable, yet the fast process was still present. Both the slow and the fast kinetics were entirely suppressed when ATPγS is used (Fig. 1K), implying that both dynamics were ATP-driven. The slow process lies in the same time scale with the kinetics identified in the dwell time analysis (Fig. 1, I and J) and is hence interpreted as successful phase shifts. The fast dynamics of around 0.3 to 2.5 s may be considered to come from failed attempts to reach another phase.

The Chi sequence makes one phase favored over the other two

When RecBCD processes a damaged DNA before RecA loading, it generates a ss-overhang with the Chi sequence at the 3′ end (*23*–*26*). However, it is unclear whether the Chi sequence is merely a flag that indicates the start point of the strand exchange reaction or it is further related to the enzymatic activity of RecA. We questioned whether the Chi sequence at the 3′ end might influence the phase of a RecA filament (*14*). To mimic the ssDNA having the Chi sequence generated by RecBCD, we placed the Chi sequence near the 3′ end and a dye pair with a 10-nt separation at 9, 10, and 11 nt away from the Chi sequence (Fig. 2A) (*27*, *28*). Measured

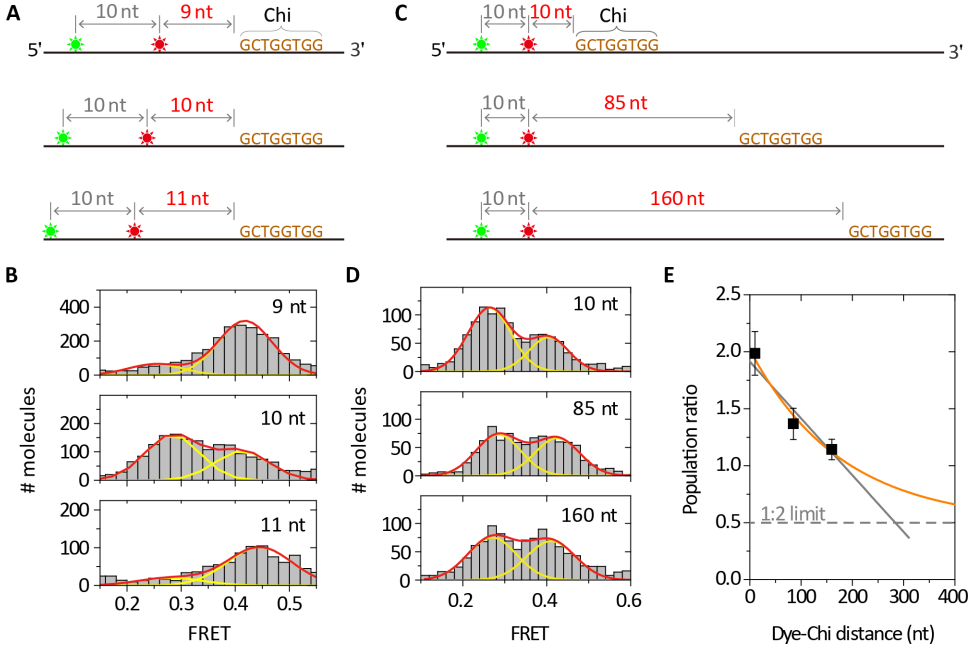

Fig. 2. Chi sequence–dependent phase bias. (**A**) DNA designs with different distances between the Chi sequence and the dye pair. Interdye separation was kept at 10 nt, whereas dye-to-Chi distances were increased by 1 nt from 9 to 11 nt. (**B**) Population distributions obtained from the 9-, 10-, and 11-nt dye-to-Chi distances. The measured population ratios (low to high) were 1:5, 1.6:1, and 1:5 for 9-, 10-, and 11-nt dye-to-Chi distances, respectively. Note that these three DNA samples are exactly the same as the one used in Fig. 1E (10 nt) except for the varying distance of the Chi sequences. (**C**) DNA designs for longer dye-to-Chi distances of 10, 85, and 160 nt. (**D**) The population distributions obtained from the 10-, 85-, and 160-nt dye-to-Chi distances. The measured population ratios (low to high) were 2.0:1, 1.4:1, and 1.1:1 for 10-, 85-, and 160-nt dye-to-Chi distances, respectively. (**E**) The population ratios of the low FRET peak to the high FRET peak (black rectangle). Error bars are SEs from three independent data sets. Gray and orange solid lines are linear and single-exponential fits to the data. Dashed line represents the 1:2 population ratio limit of the random phase distribution. All the measurements in (B) to (E) were carried out in the presence of 1 mM ATP.

distributions were biased to the high FRET state (9 and 11 nt) and the low FRET state (10 nt) (Fig. 2B). These results are distinctly different from the 1:2 ratio expected from the random distribution and the ratio obtained from the polythymine sequence in Fig. 1E. This demonstrates that the Chi sequence makes one phase favored over the other two.

We hypothesized that the phase shift dynamics occurs locally, leaving multiple phase domains within a single continuous RecA filament, because a transition of the entire RecA filament is unlikely to happen because it requires simultaneous hydrolysis of all the bound ATP, which would destabilize the entire filament. If there are phase domains existing, then one might expect that the phase bias by the Chi sequence would diminish along the length of the filament because a distant point from the Chi sequence would have a higher chance to be in a different phase domain with the Chi sequence. Consequently, we questioned over what range of distance the Chi sequence would influence the phase selection. We prepared longer ss-overhangs with the Chi sequence 10, 85, and 160 nt away from the dye pair (Fig. 2C). The distance between the Chi sequence and the dye pair was chosen such that the preferred phase by the Chi sequence would always show low FRET bias as is in the case in Fig. 2B. Consistent with the phase domain hypothesis, the bias in the population gradually decreased with the distance between the Chi and the fluorophores (Fig. 2D). Although we were limited in making the

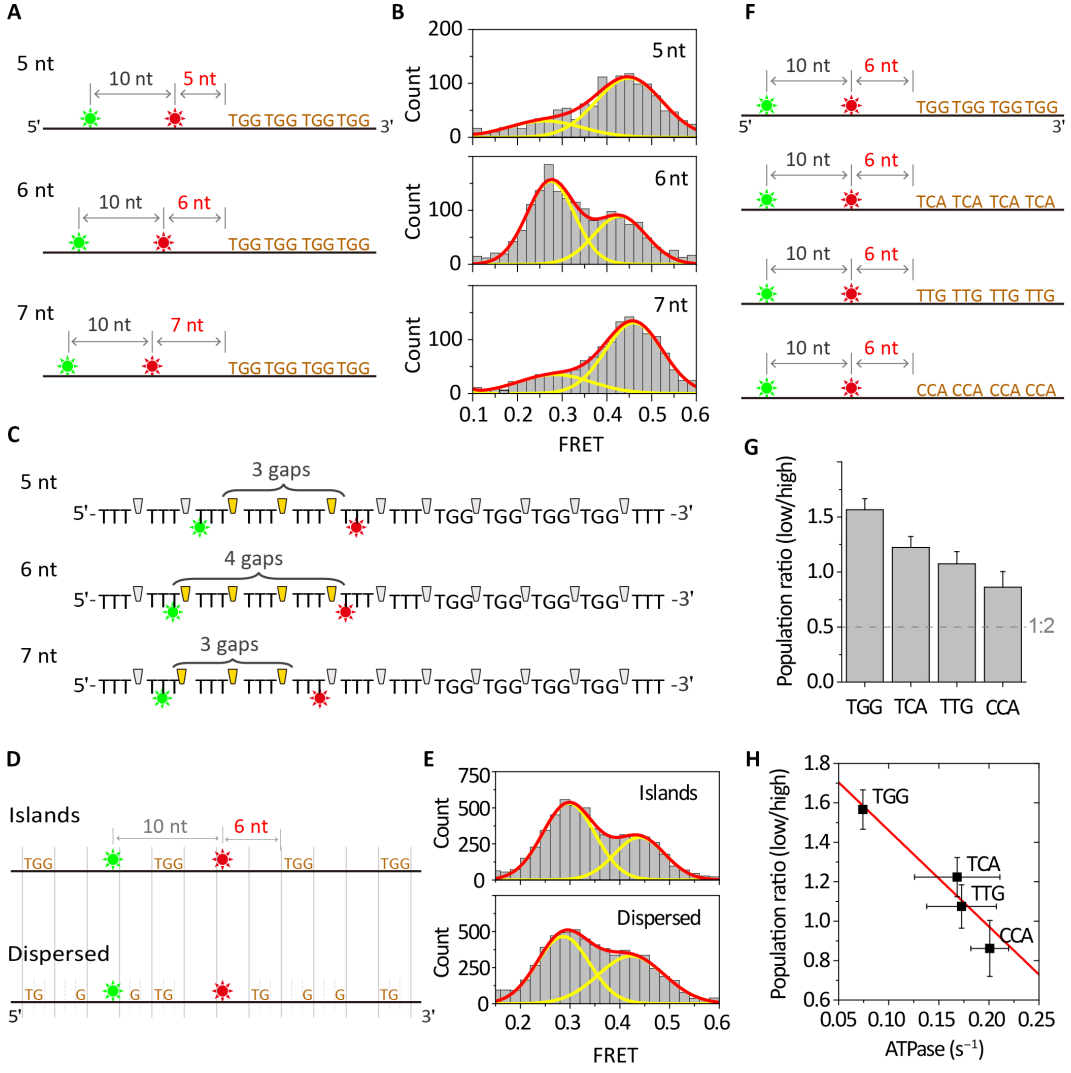

Fig. 3. Phase bias due to 3-nt-period sequence pattern. (**A**) DNA designs with different distances (5, 6, and 7 nt) between TGG repeats and a dye pair. The interdye separation was kept at 10 nt. (**B**) The population distributions obtained from the 5-, 6-, and 7-nt dye-to-TGG separations. The population ratios (low to high FRET state) determined by Gaussian fitting of each peak are 1:3.3, 1.5:1, and 1:3.3 for 5-, 6-, and 7-nt dye-to-Chi distances, respectively. (**C**) Schematic of the base grouping and gap generation in which the bases are grouped into "TGG" that account for the data in (B). Each DNA shows a different number of the gaps between the two dyes depending on their dye-to-TGG distance. (**D**) DNA designs with different base distributions. In the sequence named "islands," the four triplets of TGG were distributed along the DNA, whereas their positions from the dye pairs are kept in phase. In the sequence "dispersed," the guanine bases were displaced further, whereas their positions are kept in phase. (**E**) The population distributions obtained from the islands and dispersed. The population ratios (low to high FRET state) determined by Gaussian fitting of each peak are 1.8 and 1.6 with islands and dispersed, respectively. (**F**) DNA designs with different triplet sequences placed 6 nt apart from the dye pair of a 10-nt separation. (**G**) The population ratio (low FRET to high FRET) observed from the four triplet sequences in (F). Dashed line indicates the 1:2 ratio limit expected from the random distribution. (**H**) Correlation plot between the population ratios in (G) and the ATP hydrolysis rates measured from the filament formed on the triplet repeats. The ATP hydrolysis rates were adapted from the study of Kim *et al.* (*38*). All the measurements were carried out in the presence of 1 mM ATP.

Chi-containing polythymine sequences longer than 200 base pairs (bp) due to practical difficulties, single-exponential fit to the data predicted the characteristic decay length to be ~180 nt (that is, the distance where the effect of bias decays by half, Fig. 2E, orange solid line). Alternatively, a linear extrapolation of the population ratios estimated the influence of the Chi reaching as far as ~300 nt (Fig. 2E, gray solid line).

TGG sequence pattern with a 3-nt period induces phase bias

RecA is known to have sequence dependence in binding affinity and enzymatic activity (*29*–*32*). We sought to find out whether phase selection by the Chi sequence reflects general sequence dependence (*33*, *34*). The TGG repeat associates with RecA proteins with high affinity and is the most frequently found pattern of codon usage (GTG) in the *Escherichia coli* genome (*35*–*38*). When we placed four TGG repeats 5, 6, and 7 nt away from the dye pair (Fig. 3A), we observed biased FRET histograms. By checking the positions of gaps in all three phases for the three DNA constructs (fig. S5), we deduced that the RecA filament preferentially forms a phase that generates a gap between G at 5′ side and T at 3′ side (Fig. 3C).

When we scattered the TGG triplets along the polythymine DNA but kept them in the same phase (Fig. 3D, islands), we observed that the biased phase still persisted (Fig. 3E, top). We further examined the phase bias by breaking up the triplet composition of TGG by scattering guanine bases along the DNA while keeping them in the same phase (Fig. 3E, dispersed). The phase bias still persisted (Fig. 3E, bottom), suggesting that the positioning of the gaps (hence, breaking the base stacking of 5′-G and 3′-T) is essential for the sequence recognition rather than the grouping of the triplets. These results indicate that the dynamic process of internal filament reorganization is controlled by 3-nt-period patterns embedded in the sequence.

To check whether the observed phase bias was a special feature of the TGG pattern or whether other triplet sequence combinations could also affect the phasing dynamics of RecA, we examined three other triples sequences in addition to the TGG repeats: TCA, TTG, and CCA (Fig. 3F). We placed the triplet sequence 6 nt apart from a dye pair of a 10-nt separation. Phase biases were observed from all three triplet sequences (fig. S6). The levels of bias varied with the sequence, as can be read off from the differences in population ratios (Fig. 3G), while exhibiting the strongest preference for the TGG triplets. The phase bias levels of the three sequences were no higher than that of TGG. The phase bias was strongly negatively correlated with the ATP hydrolysis rate (Fig. 3H) (*38*), implying that the RecA filament consumes fewer ATP molecules when formed on a preferential sequence in a proper phase.

Phase of a newly nucleated filament can be affected by the RecA loading machinery

We showed that the phase of a RecA filament is determined by the Chi sequence near the 3′ end. However, the initial phase of a newly nucleated filament may also be influenced by a RecA loading machinery, which is situated at the 5′ end of a filament (*25*). To mimic the situation

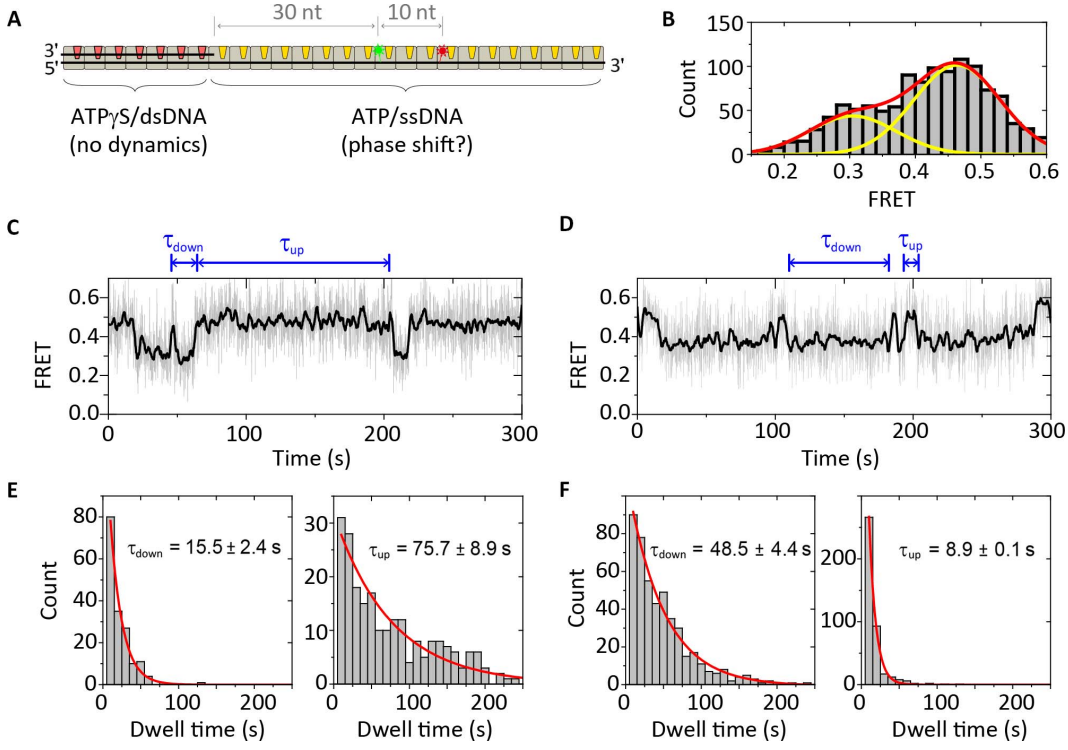

Fig. 4. Suppression of phase shift dynamics from 5′ side of a RecA filament. (**A**) Schematic of the heterologous RecA filament construct. A portion of the RecA filament was formed with ATPγS [formed on double-stranded DNA (dsDNA) region] and was connected to the filament in the 3′ direction formed with ATP (on ssDNA region). The interdye separation was 10 nt. (**B**) Single-molecule FRET distribution obtained from the heterologous RecA filaments. Population ratio (low to high) was 1:2.3. (**C** and **D**) Representative single-molecule time traces of (C) up-stable and (D) down-stable filaments. Time traces with 32-ms time resolution (gray lines) were depicted with their moving average with a 1-s time window (black lines). (**E**) Dwell time distribution of up state (left) and down state (right) from the transitions found in up-stable traces. The average dwell times were obtained by single-exponential fit to the data (red lines) from three independent data sets. (**F**) Dwell time distribution of down state (left) and up state (right) from the transitions found in down-stable traces. The average dwell times were obtained by single-exponential fit to the data (red lines) from three independent data sets.

in which a filament is held in a preset phase from its 5′ end, we prepared a RecA filament consisting of two moieties: The 3′ side of which was formed with ATP, whereas the other moiety at 5′ side was formed with ATPγS (Fig. 4A and fig. S7A, see Materials and Methods for the details) (*11*, *20*). Here, the 5′ side moiety was not allowed to change the phase but could interact with the 3′ side moiety. To avoid any phase bias effect other than that from the 5′ side moiety, we made the ss-region of poly-thymine lacking the Chi sequence or TGG repeats. We placed a dye pair of a 10-nt separation on the 3′ side moiety, 30 nt away from the junction of the two moieties. From the ATPγS/ATP heterologous filament, we observed 1:2.3 ratio in the population distribution (Fig. 4B), which was similar to that of randomly phased filaments observed in Fig. 1E (ratio, 1:2.7). Note that this observation by itself does not mean that there is no influence from the 5′ to 3′ direction because the phase of the 5′ side moiety itself can be randomly distributed. To check whether the kinetically frozen 5′ side moiety has influence on its 3′ side neighbor, we monitored FRET time trajectories of individual molecules. Unlike the dynamic transitions observed in the ATP-only filaments shown in Fig. 1 (F and H), the ATPγS/ATP heterologous filaments tend to stay in one FRET state among the two (Fig. 4, C and D, and fig. S7, B and C), confirming that the 5′ side moiety with ATPγS has influence on the phase shift dynamics of the 3′ side moiety. The traces were still able to visit the other FRET state but only briefly—more than fivefold shorter than the stable state (Fig. 4, E and F). Thus, although the 3′ side moiety retained its dynamic nature as it continually hydrolyzes ATP, its phase transition dynamics was biased such that the filament would be quickly pushed back to its biased phase due to the presence of the 5′ side moiety. These results suggest that the RecA loading machinery that remains in contact with a newly patched filament may have an influence on the phase of the loaded filament patch.

DISCUSSION

We showed, for the first time, that a RecA filament reorganizes its internal structure. A RecA filament is divided into phase domains that may shift their positions relatively to the substrate DNA. This internal dynamics is essential to eliminate any gap between two adjacent filaments. This process requires ATP hydrolysis, which answers the long-standing conundrum as to why RecA monomers in the presynaptic filament continuously hydrolyze ATP without any apparent outcome (*39*, *40*). This dynamic process is controlled by the DNA sequence such that the RecA filament aligns in phase with the Chi sequence and the 3-nt-period TGG pattern. The 3-nt-period pattern of TGG is present in the *E. coli* genome, and the Chi sequences in the genome are mostly in phase with this pattern (*35*–*37*), suggesting that RecA might have evolved to exploit these sequence signatures encoded in the *E. coli* genome to keep its internal structure flawless.

Our findings provide new insights into the RecA presynaptic filament formation, the preceding stages of the strand exchange reaction. On the basis of our findings, we propose a conceivable model of how RecA proteins make a long filament on a damaged DNA in concert with RecBCD (Fig. 5). At a double-strand break site, RecBCD processes a damaged dsDNA to generate ssDNA overhang with the Chi sequence at the 3′ end (*8*, *10*, *27*, *28*). RecBCD recruits RecA proteins to create a small filament patch on the ssDNA overhang (*41*). According to the observations in Fig. 2, this nucleated filament would align to the Chi sequence and remains in the preferred phase. The large mismatch in the speed between the fast translocation of RecBCD (~400 nt/s) (*25*) and the slow RecA growth at the 5′ end (~1 nt/s at the physiological

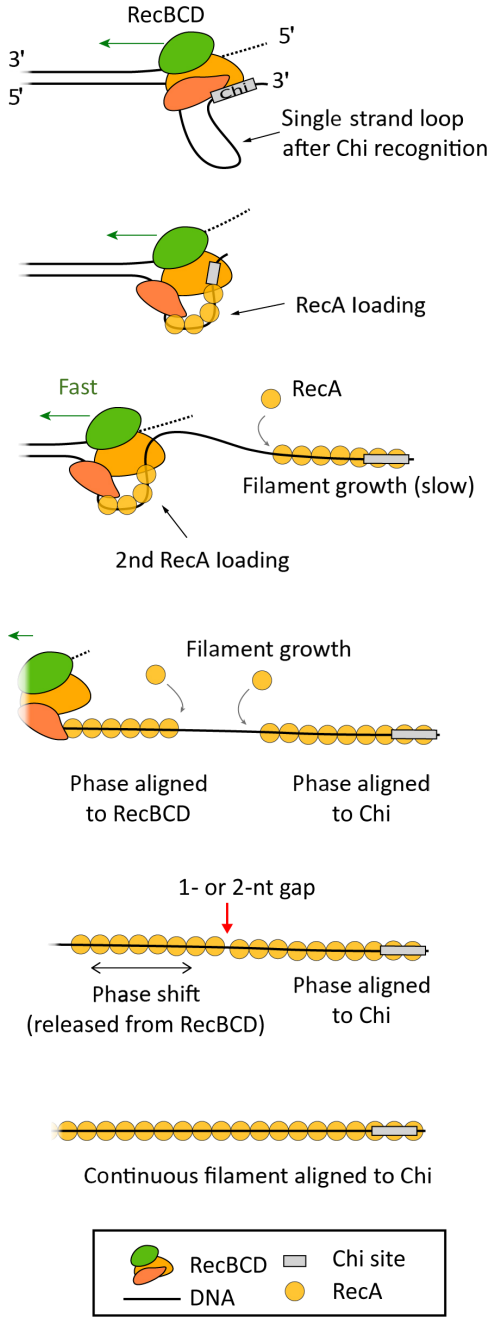

Fig. 5. A model for the RecA filament formation working in concert with a RecA loading machinery.

RecA concentration) (*11*) may generate a large bare ssDNA region between the RecBCD and the nucleated filament patch. Consequently, RecBCD may load a second patch of a filament away from the first one, and these two patches can be in different phases. When the two filament patches grow and cover the entire ss-region, the phase difference leaves a 1- or 2-nt gap between the two patches. Because the first patch is aligned to the Chi sequence at its 3′ end, the two patches would merge only when the second patch changes its phase to match that of the first. Because the phase dynamics of first patch is locked by the Chi sequence, it is mostly the second patch that changes its phase to match to the first. Also, for a rapid end-joining of the two patches, a fast

release of the second patch from the RecBCD is expected as we have shown that the phase shift dynamics can also be affected from the 5′ side (Fig. 4). Once the gap is eliminated, the two patches merge to form a longer single filament. The 3-nt-period TGG pattern in *E. coli* genome further ensures to keep the merged kilobase-long filament in a single phase (Fig. 3) (*35*, *36*). These multiple rounds of loading, extension, and joining processes are exploited as a strategic way to quickly make a long RecA filament seamless and may possibly be a general mechanism in other homologous recombinase proteins.

MATERIALS AND METHODS

DNA, protein, and buffer condition

All DNA oligomers were purchased from Integrated DNA Technologies. The sequence of the oligomer is 5′-TGGCGACGGCAGCGAGGC-$(dT)_{29}$-T*-$(dT)_n$-T*-$(dT)_{29\text{-}n}$, where T* denotes the amine-modified thymine base for dye labeling and n denotes the number of bases between the amine-modified bases (*7*–*9*). The 18 bases at the 5′ end were to hybridize to its complementary sequences, which have biotin at their 3′ end. For the sequence-dependent experiments, we inserted four repeats of TGG sequence: 5′-TGGCGACGGCAGCGAGGC-$(dT)_{29\text{-}m}$-T*-$(dT)_9$-T*-$(dT)_m$-$(TGG)_4$-$(dT)_5$, where m is either 4, 5, or 6. Likewise, the Chi sequence is placed in the polythymine tail: 5′-TGGCGACGGCAGCGAGGC-$(dT)_{29}$-T*-$(dT)_9$-T*-$(dT)_m$-GCTGGTGG-$(dT)_5$, where m is either 9, 10, or 11. The Chi-containing sequences with longer polythymine tails of 226-bp and 10-nt dye-to-Chi distance were made by ligating two ssDNAs: (i) 5′-TGGCGACGGCAGCGAGGC-$(dT)_{24}$-T*-$(dT)_9$-T*-$(dT)_9$-GCTGGTGG-$(dT)_6$ and (ii) 5′-$(dT)_{150}$. Similar Chi-containing sequences with the dye-to-chi distances of 85 nt and 160 nt were made by ligating (i) 5′-TGGCGACGGCAGCGAGGC-$(dT)_{24}$-T*-$(dT)_9$-T*-$(dT)_{23}$ and (ii) 5′-$(dT)_m$-GCTGGTGG-$(dT)_n$, where m is 61 and 136, and n is 81 and 6, respectively. The ligated mixtures were purified by denatured polyacrylamide gel electrophoresis.

Fluorescence labeling was performed by mixing the amine-modified DNA oligomers with the dye molecules functionalized with *N*-hydroxysuccinimide ester. The amine-to-dye ratio was 1:20. The reaction was carried out in 100 mM $Na_2B_4O_7$ solution at pH 8.5 for 6 hours. Unlabeled dyes were removed by ethanol precipitation methods. The labeling efficiency obtained from the absorbance was typically close to 100%.

RecA protein was purchased from New England BioLabs and used without further purification. The reaction buffer for filament formation and phase measurement contains 25 mM tris-OAc, 100 mM NaOAc, 10 mM MgOAc, glucose oxidase (1 mg/ml), 0.8 % (w/v) dextrose, 3 mM Trolox (6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid), and catalase (0.04 mg/ml). One millimolar ATP or 1 mM ATPγS was added as specified in the text or figure caption. The ATP concentration was kept at 1 mM to avoid the conformational transition of the RecA filament (*20*). The buffer pH was kept at 7.5. Chemicals were purchased from Sigma unless otherwise specified.

Single-molecule FRET assay

The hybridized partial duplex DNA was immobilized on a quartz glass slide. The slide surface was coated with polyethylene glycol to suppress nonspecific binding of proteins. The dyes labeled on the DNA were excited by a 532-nm laser (CrystaLaser) via total internal reflection, and the fluorescence was collected detected with an objective lens (Olympus, NA 1.2 water immersion). The fluorescence signal was divided spectrally for the donor (Cy3) and the acceptor (Cy5) and imaged on an electron-multiplied charged-coupled device (EMCCD). Fluorescence intensity was determined from the CCD image, and the FRET efficiency, *E*, was calculated as the ratio of the acceptor intensity to the sum of the donor and acceptor intensities

$$E = \frac{F_A + L \times F_D}{F_A + F_D}$$

Here, F_A and F_D are fluorescence intensities measured in the acceptor and the donor channels, respectively. L is the fraction of the donor signal observed in the acceptor change due to the spectral overlap of the two fluorophores and determined empirically by measuring DNA molecules labeled with donor only.

ATP/ATPγS heteroduplex filament formation

To make a heteroduplex filament consisting of RecA/ATPγS moiety at the 5′ side and RecA/ATP moiety at the 3′ side used in Fig. 4, we first formed a RecA filament on a partial duplex DNA in the presence of ATPγS. As shown in our previous studies (*11*, *20*), ATPγS enabled the filament to form on both ss- and ds-regions of the duplex DNA (fig. S7A) in contrast of the RecA/ATP filament that formed only on an ssDNA region of the duplex DNA (for example, see Fig. 1B). The removal of RecA and ATPγS from the solution led slow dissociation of RecA monomers from the ss-region but not from ds-region of the duplex DNA. Complete disassembly of the filament from ss-region typically took ~30 min, whereas the ATPγS moiety on ds-region was stable for several hours in our buffer condition. We then grew RecA/ATP filament from the remaining RecA/ATPγS moiety toward the ss-region by introducing fresh RecA proteins and ATP, resulting in a formation of a heteroduplex filament of connected RecA/ATPγS and RecA/ATP moieties.

Correlation analysis

For each single-molecule trace, we calculated correlation function defined as

$$G(\tau) = \frac{\langle E(0) \times E(\tau) \rangle}{\langle E \rangle^2}$$

where $E(\tau)$ is the FRET efficiency at time τ, and the brackets indicate time average. Because the number of transitions in a single-molecule time trace was not enough to get a good statistical average, we averaged the correlation functions obtained from different molecules but in an otherwise identical condition.

For the two-state Markov chain, in which a system is subjected to transit between two discrete states with constant transition probabilities, the correlation function can be described as (*22*)

$$G(\tau) = \frac{k_{12}}{k_{21}} e^{-(k_{12}+k_{21})\tau}$$

where k_{ij} is the transition rate from state i to j. Note that the kinetic rate determined by the correlation analysis is the sum of two rates. Therefore, if one of the two rates is sufficiently larger than the other, then the decay rate of the correlation function becomes nearly the same as the rate of the fast kinetics.

Hidden Markov method and dwell time analysis

For a quantitative analysis of the kinetics observed in a real-time single-molecule measurement, we determined the transition points with either a threshold algorithm or hidden Markov modeling. For the threshold algorithm, the transition points were defined as the points that the signal crosses the threshold value, which is manually given as a middle point between two FRET states. For the hidden Markov modeling, we followed the methods introduced by McKinney *et al.* (*21*). Once the transition points were identified, we obtained the dwell times for each FRET state.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/3/9/e1700676/DC1

fig. S1. Single-molecule RecA filament phase detection assay.
fig. S2. Single-molecule FRET time traces in the presence of ATP.
fig. S3. Single-molecule FRET time traces in the presence of ATPγS.
fig. S4. Dwell time distribution determined by hidden Markov method.
fig. S5. Possible FRET distributions of the three DNA construct at each phase.
fig. S6. The population distributions obtained from different triplet sequences in Fig. 3F.
fig. S7. RecA ATPγS/ATP heterologous filament assay.
table S1. Fit parameters obtained from the correlation functions in Fig. 1K.

REFERENCES AND NOTES

1. P. R. Bianco, R. B. Tracy, S. C. Kowalczykowski, DNA strand exchange proteins: A biochemical and physical comparison. *Front. Biosci.* **3**, D570–D603 (1998).
2. S. C. Kowalczykowski, Initiation of genetic recombination and recombination-dependent replication. *Trends Biochem. Sci.* **25**, 156–165 (2000).
3. S. L. Lusetti, M. M. Cox, The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu. Rev. Biochem.* **71**, 71–100 (2002).
4. E. M. Witkin, Ultraviolet mutagenesis and inducible DNA repair in *Escherichia coli*. *Bacteriol. Rev.* **40**, 869–907 (1976).
5. A. R. Venkitaraman, Tracing the network connecting BRCA and Fanconi anaemia proteins. *Nat. Rev. Cancer* **4**, 266–276 (2004).
6. C. E. Bell, Structure and mechanism of *Escherichia coli* RecA ATPase. *Mol. Microbiol.* **58**, 358–366 (2005).
7. E. H. Egelman, A. Stasiak, Electron-microscopy of RecA-DNA complexes: Two different states, their functional significance and relation to the solved crystal structure. *Micron* **24**, 309–324 (1993).
8. M. S. Dillingham, S. C. Kowalczykowski, RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol. Mol. Biol. Rev.* **72**, 642–671 (2008).
9. M. Spies, S. C. Kowalczykowski, The RecA binding locus of RecBCD is a general domain for recruitment of DNA strand exchange proteins. *Mol. Cell* **21**, 573–580 (2006).
10. B. Michel, H. Boubakri, Z. Baharoglu, M. LeMasson, R. Lestini, Recombination proteins and rescue of arrested replication forks. *DNA Repair* **6**, 967–980 (2007).
11. C. Joo, S. A. McKinney, M. Nakamura, I. Rasnik, S. Myong, T. Ha, Real-time observation of RecA filament dynamics with single monomer resolution. *Cell* **126**, 515–527 (2006).
12. Z. C. Chen, H. J. Yang, N. P. Pavletich, Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature* **453**, 489–494 (2008).
13. J. Y. Lee, T. Terakawa, Z. Qi, J. B. Steinfeld, S. Redding, Y. Kwon, W. A. Gaines, W. Zhao, P. Sung, E. C. Greene, Base triplet stepping by the Rad51/RecA family of recombinases. *Science* **349**, 977–981 (2015).
14. A. A. Volodin, R. D. Camerini-Otero, Influence of DNA sequence on the positioning of RecA monomers in RecA-DNA cofilaments. *J. Biol. Chem.* **277**, 1614–1618 (2002).
15. A. A. Volodin, E. A. Smirnova, T. N. Bocharova, R. D. Camerini-Otero, Phasing of RecA monomers on quasi-random DNA sequences. *FEBS Lett.* **546**, 203–208 (2003).
16. M. T. J. van Loenhout, T. van der Heijden, R. Kanaar, C. Wyman, C. Dekker, Dynamics of RecA filaments on single-stranded DNA. *Nucleic Acids Res.* **37**, 4089–4099 (2009).
17. T. Ha, A. G. Kozlov, T. M. Lohman, Single-molecule views of protein movement on single-stranded DNA. *Annu. Rev. Biophys.* **41**, 295–319 (2012).
18. T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin, S. Weiss, Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6264–6268 (1996).
19. R. Roy, S. Hohng, T. Ha, A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507–516 (2008).
20. S. H. Kim, K. Ragunathan, J. Park, C. Joo, D. Kim, T. Ha, Cooperative conformational transitions keep RecA filament active during ATPase cycle. *J. Am. Chem. Soc.* **136**, 14796–14800 (2014).
21. S. A. McKinney, C. Joo, T. Ha, Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91**, 1941–1951 (2006).
22. H. D. Kim, G. U. Nienhaus, T. Ha, J. W. Orr, J. R. Williamson, S. Chu, Mg^{2+}-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4284–4289 (2002).
23. D. G. Anderson, S. C. Kowalczykowski, The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ-regulated manner. *Cell* **90**, 77–86 (1997).
24. D. A. Dixon, S. C. Kowalczykowski, Homologous pairing in vitro stimulated by the recombination hotspot, Chi. *Cell* **66**, 361–371 (1991).
25. M. Spies, P. R. Bianco, M. S. Dillingham, N. Handa, R. J. Baskin, S. C. Kowalczykowski, A molecular throttle: The recombination hotspot χ controls DNA translocation by the RecBCD helicase. *Cell* **114**, 647–654 (2003).
26. P. R. Bianco, S. C. Kowalczykowski, The recombination hotspot Chi is recognized by the translocating RecBCD enzyme as the single strand of DNA containing the sequence 5′-GCTGGTGG-3′. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 6706–6711 (1997).
27. A. F. Taylor, D. W. Schultz, A. S. Ponticelli, G. R. Smith, RecBC enzyme nicking at chi sites during DNA unwinding: Location and orientation-dependence of the cutting. *Cell* **41**, 153–163 (1985).
28. A. S. Ponticelli, D. W. Schultz, A. F. Taylor, G. R. Smith, Chi-dependent DNA strand cleavage by RecBC enzyme. *Cell* **41**, 145–151 (1985).
29. M. Amaratunga, A. S. Benight, DNA sequence dependence of ATP hydrolysis by RecA protein. *Biochem. Biophys. Res. Commun.* **157**, 127–133 (1988).
30. R. Bar-Ziv, A. Libchaber, Effects of DNA sequence and structure on binding of RecA to single-stranded DNA. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9068–9073 (2001).
31. E. Biet, J.-S. Sun, M. Dutreix, Conserved sequence preference in DNA binding among recombination proteins: An effect of ssDNA secondary structure. *Nucleic Acids Res.* **27**, 596–600 (1999).
32. P. Wittung, C. Ellouze, F. Maraboeuf, M. Takahashi, B. Nordèn, Thermochemical and kinetic evidence for nucleotide-sequence-dependent RecA-DNA interactions. *Eur. J. Biochem.* **245**, 715–719 (1997).
33. R. Rajan, J. W. Wisler, C. E. Bell, Probing the DNA sequence specificity of *Escherichia coli* RECA protein. *Nucleic Acids Res.* **34**, 2463–2471 (2006).
34. R. B. Tracy, S. C. Kowalczykowski, In vitro selection of preferred DNA pairing sequences by the *Escherichia coli* RecA protein. *Genes Dev.* **10**, 1890–1903 (1996).
35. V. Biaudet, M. El Karoui, A. Gruss, Codon usage can explain GT-rich islands surrounding Chi sites on the *Escherichia coli* genome. *Mol. Microbiol.* **29**, 666–669 (1998).
36. T. Colbert, A. F. Taylor, G. R. Smith, Genomics, Chi sites and codons: 'Islands of preferred DNA pairing' are oceans of ORFs. *Trends Genet.* **14**, 485–488 (1998).
37. R. B. Tracy, F. Chédin, S. C. Kowalczykowski, The recombination hot spot chi is embedded within islands of preferred DNA pairing sequences in the *E. coli* genome. *Cell* **90**, 205–206 (1997).
38. S. H. Kim, C. Joo, T. Ha, D. Kim, Molecular mechanism of sequence-dependent stability of RecA filament. *Nucleic Acids Res.* **41**, 7738–7744 (2013).
39. W. A. Bedale, M. Cox, Evidence for the coupling of ATP hydrolysis to the final (extension) phase of RecA protein-mediated DNA strand exchange. *J. Biol. Chem.* **271**, 5725–5732 (1996).
40. J. P. Menetski, D. G. Bear, S. C. Kowalczykowski, Stable DNA heteroduplex formation catalyzed by the *Escherichia coli* RecA protein in the absence of ATP hydrolysis. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 21–25 (1990).
41. J. J. Churchill, D. G. Anderson, S. C. Kowalczykowski, The RecBC enzyme loads RecA protein onto ssDNA asymmetrically and independently of χ, resulting in constitutive recombination activation. *Genes Dev.* **13**, 901–911 (1999).

Acknowledgments

Funding: This research was supported by the National Research Foundation (NRF) grant No. 2017R1D1A1B03031150. J.S. was funded by the Creative Research Initiative Project program (NRF grant No. 2015R1A3A2066497). C.J. was funded by European Research Council (ERC) under the European Union's Seventh Framework Programme [FP7/2007-2013]/ERC grant agreement no. [309509]. **Author contributions:** S.H.K. and C.J. designed the research. S.H.K., T.A., S.C., and T.J.C. performed the measurements. S.H.K., C.J., J.S., and D.K. analyzed the data. S.H.K., C.J., and D.K. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interest. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Submitted 7 March 2017
Accepted 8 August 2017
Published 6 September 2017
10.1126/sciadv.1700676

Citation: S. H. Kim, T. Ahn, T. J. Cui, S. Chauhan, J. Sung, C. Joo, D. Kim, RecA filament maintains structural integrity using ATP-driven internal dynamics. *Sci. Adv.* **3**, e1700676 (2017).