

# SCIENTIFIC REPORTS



OPEN

## Efficient modification of $\lambda$ -DNA substrates for single-molecule studies

Yoori Kim<sup>1</sup>, Armando de la Torre<sup>1</sup>, Andrew A. Leal<sup>1</sup> & Ilya J. Finkelstein<sup>1,2</sup> 

Single-molecule studies of protein-nucleic acid interactions frequently require site-specific modification of long DNA substrates. The bacteriophage  $\lambda$  is a convenient source of high quality long (48.5 kb) DNA. However, introducing specific sequences, tertiary structures, and chemical modifications into  $\lambda$ -DNA remains technically challenging. Most current approaches rely on multi-step ligations with low yields and incomplete products. Here, we describe a molecular toolkit for rapid preparation of modified  $\lambda$ -DNA. A set of PCR cassettes facilitates the introduction of recombinant DNA sequences into the  $\lambda$ -phage genome with 90–100% yield. Extrahelical structures and chemical modifications can be inserted at user-defined sites via an improved nicking enzyme-based strategy. As a proof-of-principle, we explore the interactions of *S. cerevisiae* Proliferating Cell Nuclear Antigen (yPCNA) with modified DNA sequences and structures incorporated within  $\lambda$ -DNA. Our results demonstrate that *S. cerevisiae* Replication Factor C (yRFC) can load yPCNA onto 5'-ssDNA flaps, (CAG)<sub>13</sub> triplet repeats, and homoduplex DNA. However, yPCNA remains trapped on the (CAG)<sub>13</sub> structure, confirming a proposed mechanism for triplet repeat expansion. We anticipate that this molecular toolbox will be broadly useful for other studies that require site-specific modification of long DNA substrates.

Single-molecule imaging and manipulation approaches have greatly expanded our understanding of protein-nucleic acid interactions<sup>1</sup>. More recently, the single-molecule toolkit has grown to include hybrid methods that combine fluorescence imaging (e.g., single-molecule FRET) with force manipulation (e.g., optical or magnetic tweezers) for multi-modal data acquisition<sup>2–4</sup>. Many of these approaches require very long (>10 kb), chemically modified DNA substrates. For example, long DNA substrates are frequently used as flexible, micron-long tethers in optical and magnetic tweezers experiments<sup>4–7</sup>. Moreover, long DNA molecules elevate the biochemical reactions away from glass coverslips and other surfaces, reducing the potential for non-specific interactions<sup>5–7</sup>. DNA derived from bacteriophage  $\lambda$  ( $\lambda$ -DNA) offers several advantages for single-molecule studies. It is a long (48.5 kb) substrate, permitting the observation of protein movement over kilobase-length distances. In addition, synthetic DNA handles can be ligated to the *cosL* and *cosR* ssDNA overhangs. Finally, high-quality recombinant DNA can be purified from lysogenic cells in large quantities, precluding the need to maintain high titer viral stocks.

Site-specific  $\lambda$ -DNA modification strategies generally fall within one of four categories: (i) restriction enzyme cleavage and ligation<sup>4,8</sup>; (ii) recombinase-mediated modifications<sup>9,10</sup>; and (iii) insertion of extrahelical structures via oligonucleotide mimics, or (iv) nicking endonuclease (nickase)-based oligo replacement<sup>11–15</sup>. To date, restriction enzyme cleavage and multi-step ligation *in vitro* is one of the most frequently used methods for modifying  $\lambda$ -DNA. However, this approach is technically challenging because multi-step intra-molecular ligation is inefficient. Restriction enzyme-based cloning is also limited due to the few unique restriction sites within the 48.5 kb phage genome. Site-specific recombination *in vitro* is another promising approach for inserting exogenous DNA structures into  $\lambda$ -DNA<sup>9,10</sup>. However, this requires two unique recombinase sites to be cloned into  $\lambda$ -DNA. The recombinant DNA then needs to be packaged into phage particles, and the virus titer amplified prior to infection of an *E. coli* host. The recombinase enzymes would also need to be produced recombinantly as they are not commercially available. In sum, existing strategies for site-specific  $\lambda$ -DNA modification are either low efficiency or require time-consuming DNA and protein purification.

<sup>1</sup>Department of Molecular Biosciences and Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, Texas, 78712, USA. <sup>2</sup>Center for Systems and Synthetic Biology, The University of Texas at Austin, Austin, Texas, 78712, USA. Correspondence and requests for materials should be addressed to I.J.F. (email: [ifinkelstein@cm.utexas.edu](mailto:ifinkelstein@cm.utexas.edu))

Introducing site-specific synthetic oligonucleotides and extrahelical structures poses additional challenges in such a long DNA substrate. Such structures are usually inserted via a nicking endonuclease (nickase) based strategy<sup>12, 16–19</sup>. This strategy requires a nicking enzyme to create two or more closely spaced nicks (~10–30 nt apart) on the same strand of the DNA duplex. After nicking, the DNA substrate is annealed with a synthetic oligonucleotide that replaces the short fragments produced by the nickase. However, nickase-based insertion is inefficient for wild type  $\lambda$ -DNA because nickases that produce closely-spaced nicks also cleave hundreds of additional sites on both the top and bottom DNA strands, fragmenting the DNA substrate and reducing the overall yield of full-length DNA molecules. Because this strategy relies on naturally occurring nick sites, it also precludes site-specific incorporation of extrahelical structures at user-specified positions along the DNA substrate.

Here, we develop a method for rapidly modifying and purifying recombinant  $\lambda$ -phage DNA. We use *in vivo* recombineering to target any segment of a lysogenic phage, abrogating the need for restriction sites and ligation. Using this approach, we develop a molecular toolkit for inserting exogenous DNA sequences into the  $\lambda$ -phage genome with >90% efficiency. Site-specific DNA incorporation at three unique sites is confirmed via both ensemble and single-molecule fluorescence assays. We also demonstrate a strategy for inserting non-replicative extrahelical DNA structures at these sites. We explore the utility of these DNA structures by demonstrating site-specific loading of *S. cerevisiae* Proliferating Cell Nuclear Antigen (yPCNA) by the *S. cerevisiae* clamp-loader complex Replication Factor C (yRFC). Our results show that yPCNA can be loaded on a 5'-single stranded DNA (ssDNA) flap, a (CAG)<sub>13</sub> triplet nucleotide repeat (TNR), and homoduplex DNA. While yPCNA can diffuse freely from the flap and along homoduplex DNA, it remains trapped within the (CAG)<sub>13</sub> repeat, adding evidence to a model where stationary yPCNA promotes TNR expansion<sup>20–22</sup>. In sum, we anticipate that this molecular toolkit will be broadly useful for both ensemble and single-molecule studies that require site-specific modification of long DNA substrates.

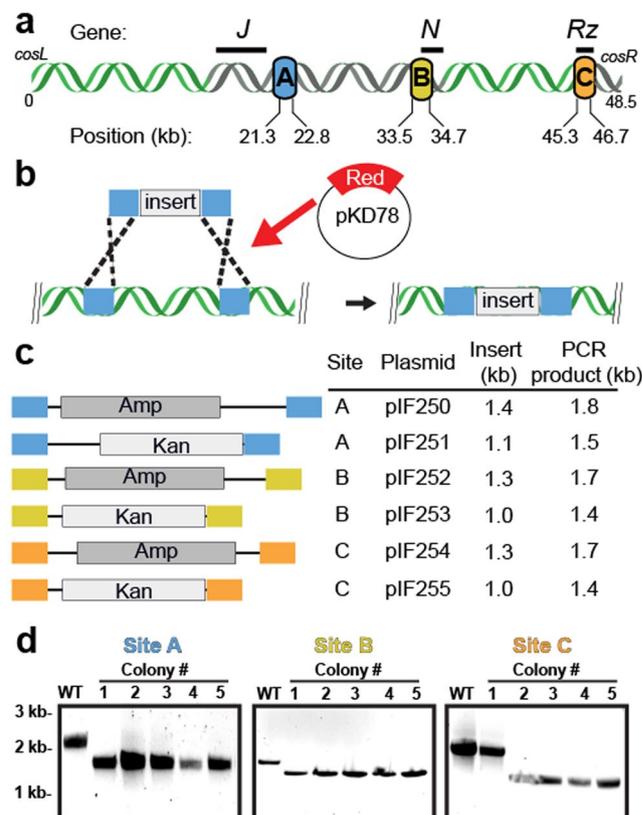
## Results

**A molecular toolkit for modifying  $\lambda$ -phage DNA.** We sought to develop a method that fulfilled three criteria: (i) multiple exogenous DNA sequences can be inserted into  $\lambda$ -phage DNA with nearly 100% efficiency, (ii) the insertion positions are not limited by availability of unique restriction sites, and (iii) chemical modifications and extra-helical structures can be introduced efficiently and with minimal handling. To satisfy the first design criterion, we surveyed the literature to identify two large segments encompassing 37% (17.6 kilobase; kb) of the  $\lambda$ -phage genome that are dispensable for both lytic and lysogenic growth (Fig. 1a). The first dispensable segment is 16 kb long and is between gene products *J* and *N*<sup>23, 24</sup>. The second 1.6 kb segment, is between gene product *Rz* and the *cosR* end<sup>25</sup>. Next, we chose three sites—designated sites A, B, and C (Fig. 1a)—because they are each separated by ~12 kb (~4  $\mu$ m) and are thus readily resolved via single-molecule fluorescence microscopy (see below). These three sites provide convenient site-specific modifications within  $\lambda$ -phage DNA.

To satisfy the second design criterion, we developed a series of plasmids for efficient recombineering-based insertion of exogenous DNA into the  $\lambda$ -phage genome. Recombineering is a recombination-based method for genome modification *in vivo* (Fig. 1b)<sup>26, 27</sup>. Recombineering proteins, supplied on an arabinose-inducible helper plasmid, catalyze the insertion of a linear DNA substrate between two homologous sites in the phage genome<sup>28</sup>. The phage is maintained as an inducible lysogen in *E. coli*, precluding the need to amplify and maintain high-titer phage particles<sup>26</sup>. We synthesized a series of plasmids with targeting cassettes for sites A, B, and C (Fig. 1c). The cassettes consist of a user-specified DNA and an antibiotic resistance gene flanked by ~200 bp of homology to the insertion site (see Supplemental Information). Increasing the length of the homologous arms facilitates efficient recombineering<sup>29</sup>. Resistance to an antibiotic marker further selects the desired product. The recombineering reaction efficiency was scored via colony PCR and agarose gel electrophoresis. Each of the three insertion cassettes replaces the native sequence with a ~1–1.5 kb insert (Fig. 1c). As expected, the recombineering efficiency approached nearly 100% at each of the three insertion sites (Fig. 1d, 100%, N = 10/10; 100%, N = 10/10; and 90%, N = 9/10 colonies for sites A, B, and C, respectively). To amplify the recombinant  $\lambda$ -DNA, lytic growth was induced at a non-permissive temperature (>42 °C). Unfolding of the temperature-sensitive repressor (ci857*ind* 1) initiates phage replication and packaging<sup>24, 30</sup>. An additional amber mutation in the *S* gene (*Sam* 7) delays cell lysis and produces large burst sizes (>200 phage particles per cell), increasing the overall yield of phage DNA<sup>23, 31</sup>. Insertion of the cassette was confirmed via PCR, restriction enzyme digestion, and single-molecule imaging (see below).

As a first proof-of-principle, high-throughput DNA curtains were used to measure the binding positions of Lac repressor (LacI) proteins on recombinant  $\lambda$ -DNA harboring a high-affinity Lac operator sequence. In this assay, a fluid lipid bilayer is deposited on the surface of a micro-patterned flowcell<sup>32–34</sup>. Biotinylated DNA is immobilized on the lipid bilayer via a biotin-streptavidin linkage and buffer flow is used to organize hundreds of DNA molecules at chromium (Cr) diffusion barriers (Fig. 2a)<sup>32, 34</sup>. LacI was incubated with the DNA and fluorescently labeled with anti-FLAG antibody-conjugated quantum dots (QDs). The DNA was stained with a fluorescent intercalating dye (YOYO-1) and 89% (N = 89/100) of the DNA molecules were full-length (Fig. 2b). This measurement likely underestimates the percentage of full-length DNA because intercalating dyes induce photodamage and accelerate the accumulation of broken molecules. As expected, LacI bound specifically to the recombinant operator site (Fig. 2c) at each of our three target sites<sup>35</sup>. These results demonstrate that high quality recombinant  $\lambda$ -DNA provides an ideal substrate for single-molecule imaging.

Next, we developed a strategy for inserting multiple recombinant DNA sequences—each flanked by a different antibiotic cassette—into the same  $\lambda$ -DNA substrate. All colonies (N = 10/10) harbored both insertion cassettes, as scored by colony PCR. To confirm the quality of these DNA substrates for single-molecule experiments, we imaged LacI binding to DNA curtains harboring a Lac operator in both sites B & C. As expected, 92% (N = 483/525) of all LacI molecules bound to either one of the two insertion sites (Fig. 2d). The broad LacI binding distributions largely stem from experimental uncertainty in the precise surface tethering of DNA molecules

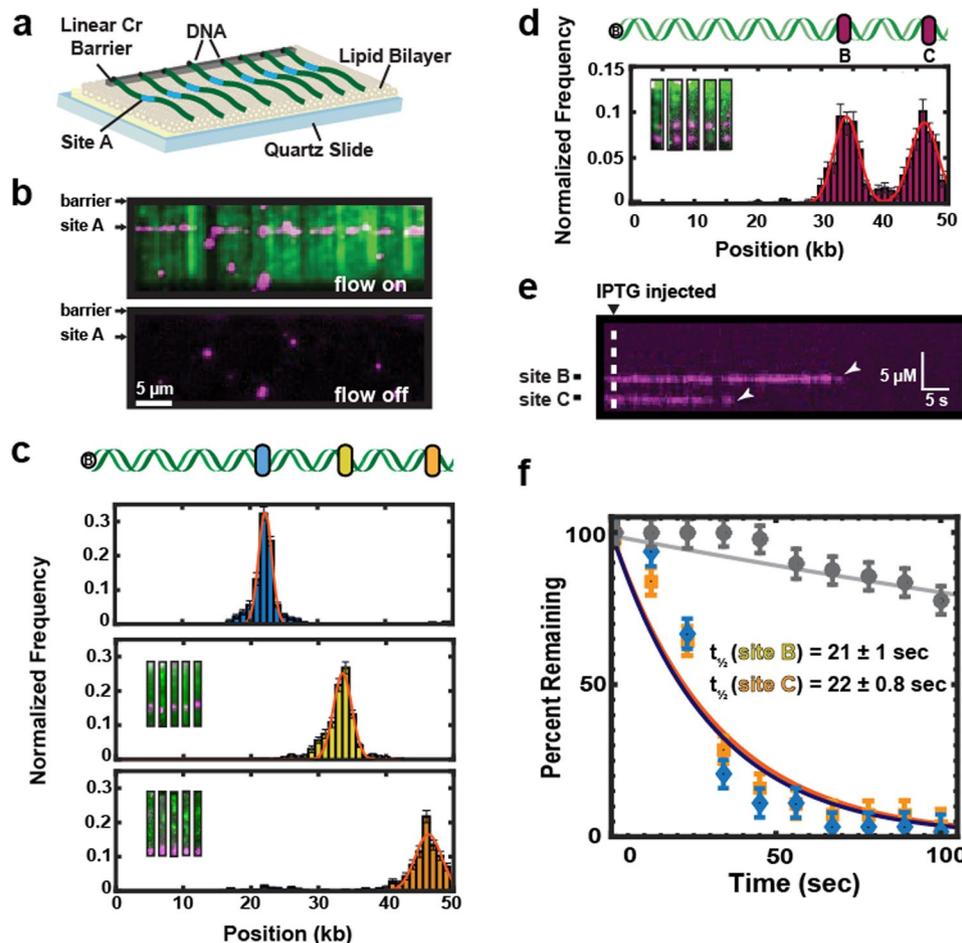


**Figure 1.** Recombineering at three unique positions within the  $\lambda$ -phage genome. **(a)** Insertion cassettes were designed for three sites targeting dispensable segments of the  $\lambda$ -phage genome (shown in gray). These cassettes are 21.3 kb, 33.5 kb and 45.3 kb away from *cosL* (designated A, B and C, respectively). **(b)** Schematic of the Red-based recombineering. An arabinose-inducible plasmid supplies the three Red genes. **(c)** Cassettes for efficient recombineering into each of the sites shown in **(a)**. The cassettes carry an antibiotic resistance gene that is flanked on both sides by  $\sim$ 200 bp of homology to  $\lambda$ -DNA. **(d)** Recombineering efficiency was scored by colony PCR followed by agarose gels analysis. Successful recombineering generates a smaller PCR fragment. We observed 90–100% insertion efficiency at each of the three target sites.

(Fig. 2c and d). To construct binding histograms, we assume that all DNA molecules are tethered at the center of the diffusion barriers, which are  $\sim$ 1–1.5  $\mu$ m wide<sup>32</sup>. When compiling binding distributions from hundreds of such DNA molecules (spanning several flowcells), this approximation results in an apparent broadening of the resulting histogram. This broadening has also been observed for both site- and sequence-specific DNA-binding proteins in similar flow-stretched assays<sup>36–40</sup>. We further confirmed that LacI is bound to the operator site by injecting 1 mM IPTG into the flowcell. IPTG caused rapid dissociation of all LacI molecules with a half-life of  $21 \pm 1.0$  seconds ( $N = 63$ ; error indicates 95% C.I.) for site B and  $22 \pm 0.8$  seconds ( $N = 56$ ) for site C (Fig. 2e and f). In contrast, the half-life of LacI in the absence of IPTG was  $327 \pm 10$  seconds ( $N = 49$ ). We conclude that recombinant  $\lambda$ -DNA can be rapidly modified at multiple sites for single-molecule studies.

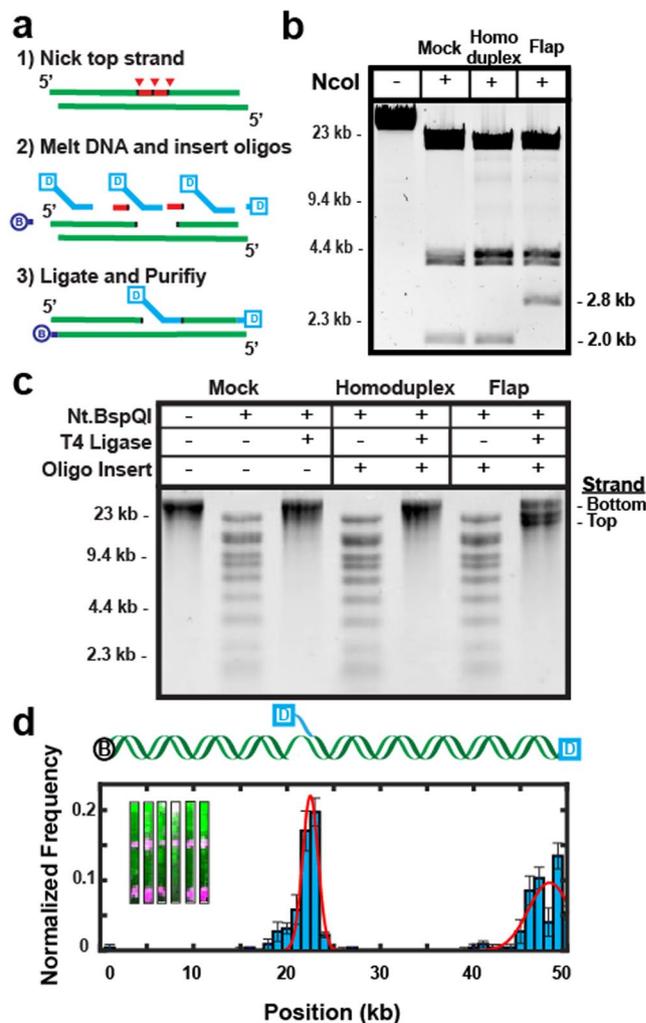
**Insertion of modified oligonucleotides into  $\lambda$ -DNA.** Single-molecule studies frequently require the insertion of modified bases or fluorophores at precise positions along the DNA substrates. We developed an improved nicking-enzyme based strategy for inserting synthetic oligonucleotides at precise positions along  $\lambda$ -DNA (Fig. 3a). We designed a nicking cassette that includes three consecutive BspQI recognition sequences separated by 12–13 bp spacers (see Supplemental Information). The DNA is nicked with Nt.BspQI (nickase) enzyme that creates closely spaced discontinuities along one of the two DNA strands.  $\lambda$ -DNA harbors ten additional widely spaced BspQI recognition sequences, thereby limiting nicks and double-strand breaks along the rest of the DNA substrate. The nicked DNA is heated to liberate short single-stranded DNA fragments. A synthetic oligonucleotide is inserted into the backbone by annealing and ligation<sup>11</sup>. After nicking with Nt.BspQI and heating, the gapped DNA is slowly cooled with a 100-fold excess of the desired insertion oligo. The annealed and ligated  $\lambda$ -DNA is further purified by gel filtration to remove enzymes and excess oligonucleotides. Conversion of restriction sites (NcoI, NotI) to single-stranded DNA within the nicking cassette facilitates rapid quantification of the insertion efficiency via restriction digest mapping.

We validated this approach by inserting extrahelical structures into a nicking cassette that was cloned into  $\lambda$ -DNA using the recombineering strategy described above. We first inserted a 5'-ssDNA flap oligonucleotide that annealed to 26 nt of the gapped ssDNA. Insertion of the flap oligonucleotide abolishes a single NcoI site within the cassette, converting two small NcoI DNA fragments (2.0 kb + 0.8 kb) to a larger 2.8 kb band on a native agarose gel



**Figure 2.** Single-molecule imaging of Lac repressor (LacI) dynamics on recombiner  $\lambda$ -DNA substrates. (a) Schematic of a single-tethered DNA curtain assay. DNA molecules (green) are immobilized on a fluid lipid bilayer and arranged at micro-fabricated Chromium (Cr) barriers. One of the insertion sites (site A) is indicated in blue. (b) TIRFM image of LacI (magenta) binding to a lac operator inserted at site A on a recombiner  $\lambda$ -DNA substrate (green). (c) Binding distributions of LacI to a lac operator sequence inserted at one of the three sites within  $\lambda$ -DNA. Solid lines indicate a Gaussian fit to the data with the indicated mean  $\pm$  standard deviation. Blue: site A ( $21.7 \pm 1.0$  kb,  $N = 558$ , blue). Yellow: site B ( $34.0 \pm 1.5$  kb, mean  $N = 538$ ). Orange: site C ( $46.0 \pm 2.3$  kb,  $N = 467$ ). (d) Insertion of two lac operator cassettes within a single  $\lambda$ -DNA construct. Inset: single-molecule imaging of the recombinant DNA shows LacI at both positions B & C. Binding histograms show nearly equal occupancy at both lac operator sites. Red line: double Gaussian fit ( $34.0 \pm 2.0$  kb and  $46.2 \pm 2.2$  kb, respectively;  $N = 525$ ). (e) As expected, LacI dissociates from both operator sites after injection of  $0.4$  mM Isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG). (f) Quantification of the LacI lifetimes upon IPTG injection. The solid lines are single-exponential fits to the data (blue diamonds,  $t_{1/2} = 21 \pm 1.0$  sec,  $N = 63$ , and orange squares,  $t_{1/2} = 22 \pm 0.8$  sec,  $N = 56$  at positions B & C, respectively). The solid gray line is the single-exponential fit of the LacI lifetime data, in the absence of IPTG (grey circles,  $t_{1/2} = 327 \pm 10$  s,  $N = 49$ ). The error bars indicate the standard deviation obtained via bootstrap analysis.

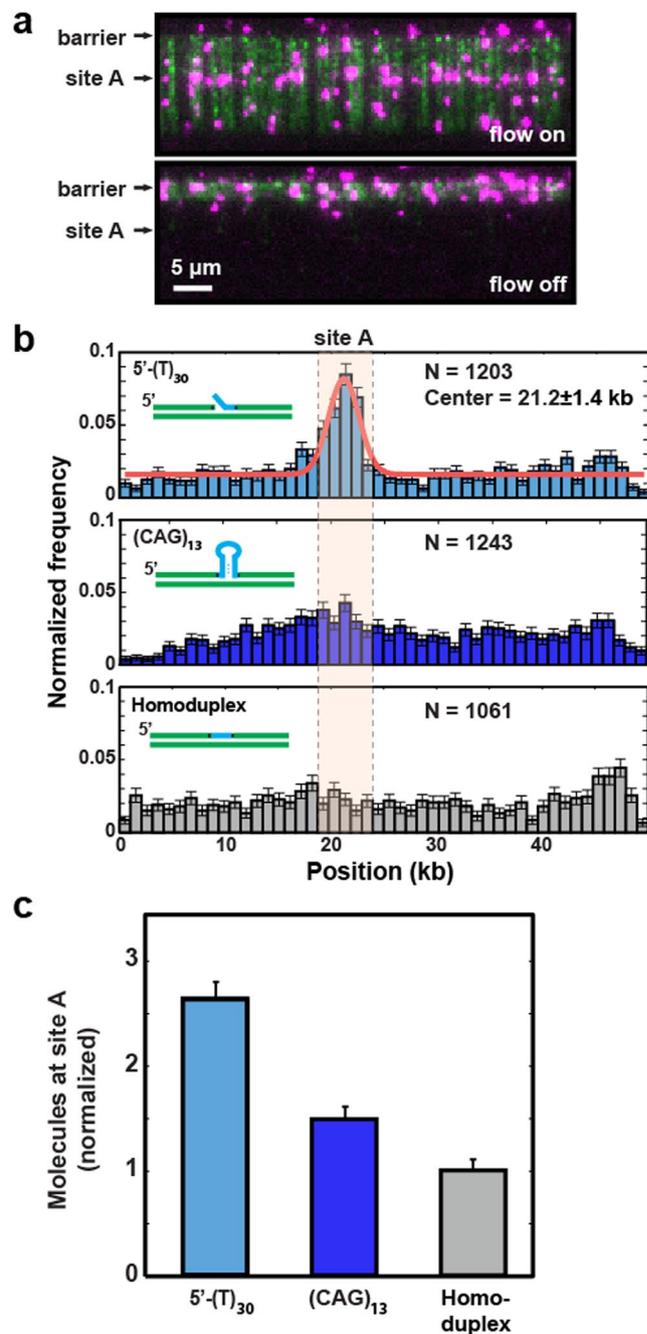
(Fig. 3b). We saw 100% cleavage of the flap-proximal NcoI site with both untreated homoduplex DNA and with a complementary phosphorylated oligonucleotide that perfectly replaced the excised cassette (i.e., mock-treated substrate). However, only  $\sim 1\%$  of the DNA was NcoI-sensitive after 5'-ssDNA flap incorporation. Next, alkaline agarose gel electrophoresis was used to assay the quality of the ligated DNA substrate (Fig. 3c). The 48.5-kb DNA runs as a single large band on a denaturing gel. Nicking with Nt.BspQI produces a distinct ladder of bands, which are removed after ligation. As expected, ligating the nicked DNA substrate with a complementary phosphorylated oligonucleotide (mock-treated DNA) produces a substrate that is indistinguishable from homoduplex  $\lambda$ -DNA. Importantly, adding the 5'-flap oligonucleotide produces a DNA substrate with three distinct bands on the alkaline agarose gel; the full-length bottom strand and the 21.3 kb and 26.7 kb discontinuous top strands (Fig. 3c). The quality of the DNA substrate and site-specific flap incorporation was also evaluated via single-molecule fluorescence imaging (Fig. 3d). Single-tethered DNA curtains were assembled with  $\lambda$ -DNA that has a biotin on the *cosL* end and a digoxigenin on both the inserted flap and the *cosR* end. Digoxigenin was fluorescently labeled via anti-digoxigenin-functionalized QDs. The fluorescent signal was localized to the flap and DNA ends (Fig. 3d, inset). A histogram of the binding positions showed two peaks, one at the distal DNA end and a second at site



**Figure 3.** Construction of DNA substrates containing extrahelical structures. **(a)** Schematic of nickase-mediated insertion of a synthetic oligonucleotide (Digoxigenin (dig)-labeled 5'-flap). **(b)** Diagnostic digest of a recombinered  $\lambda$ , demonstrated here by the 2.8 kb band in lane 4 which is digested into two smaller bands in lanes 2 and 3 (bottom band not shown). **(c)** Denaturing alkaline agarose gel of three separate reactions (a "mock-treated" DNA where the nicked DNA is replaced with a complementary oligo, a homoduplex DNA, and a 5'-ssDNA flap) to determine ligation efficiency. The last lane shows a larger band (bottom strand) and two smaller bands (top strand) that are created upon insertion of the flap. **(d)** Top: schematic of a DNA substrate containing a dig-labeled 5'-ssDNA flap and a dig-terminated free DNA end. Bottom: binding distribution of anti-dig quantum-dots to this DNA substrate. Solid lines indicate a double Gaussian fit to the data with the indicated mean  $\pm$  standard deviation (5'-ssDNA flap:  $22 \pm 0.8$  kb; 3'-DNA end:  $47.5 \pm 2$  kb;  $N = 225$ ). Insets: representative DNA molecules (green) labeled with quantum dots (magenta) at position A and at the free DNA ends.

A, where the ssDNA flap was incorporated (Fig. 3d). Importantly, 82% ( $N = 82/100$ ) of the fluorescently stained DNA molecules were full-length, confirming that the DNA was not fragmented after the oligo insertion reaction. In sum, our ensemble biochemical and single-molecule results confirm efficient and site-specific incorporation of synthetic oligonucleotides into full-length  $\lambda$ -DNA.

**Observing *S. cerevisiae* PCNA ( $\gamma$ PCNA) dynamics on modified DNA substrates.** To further demonstrate the utility of modified  $\lambda$ -DNA substrates for single-molecule studies, we determined how  $\gamma$ RFC loads  $\gamma$ PCNA on various DNA structures. RFC loads PCNA by using the energy of ATP hydrolysis to open and close the clamp ring<sup>41–43</sup>. Although RFC preferentially loads PCNA on primer-template junctions, it can also catalyze loading on nonspecific homoduplex dsDNA. RFC has also been proposed to load PCNA on extrahelical triplet nucleotides repeats (TNRs), ultimately driving TNR expansion<sup>20–22</sup>. However, nearly all prior biochemical studies have used a truncated  $\gamma$ RFC $\Delta N$ , where the DNA-binding patch of Rfc1 has been deleted<sup>44</sup>. This is because a dsDNA-binding motif within the large RFC subunit (Rfc1) binds dsDNA non-specifically, complicating ensemble biochemical experiments<sup>45</sup>. Therefore, we first assayed  $\gamma$ PCNA loading by the more physiologically relevant full length wild type (wt)  $\gamma$ RFC<sup>46</sup>.



**Figure 4.** yRFC loads yPCNA on various extrahelical DNA structures. **(a)** Fluorescent images of yPCNA on DNA substrates with a 5'-ssDNA flap inserted at site A. Top: with buffer flow, bottom: without flow. DNA is stained with YOYO-1 (green) and yPCNA is labeled with anti-FLAG antibody conjugated QDs (magenta). Stopping buffer flow retracts both yPCNA and DNA to the Cr barrier, confirming that yPCNA is bound to the DNA and is not on the surface (bottom panel). **(b)** Binding distribution histogram of yPCNA/yRFC on  $\lambda$ -DNA molecules containing a 5'-ssDNA flap (top), (CAG)<sub>13</sub> repeat (middle), and homoduplex DNA (bottom). Red line: fit to a Gaussian distribution (center = 21.2 kb  $\pm$  1.4 kb; st. dev.). Pink box: a 5 kb window that captures 99% of all yPCNAs at site A (see text for details). **(c)** The mean number of yPCNA molecules loaded at site A (within the pink box), as determined by bootstrap analysis of the histograms in **(b)**. Error bars represent the st. dev. of the mean from the bootstrap analysis.

We first explored the relative specificity of yPCNA loading onto DNA substrates harboring a 30-nt 5'-ssDNA flap, a (CAG)<sub>13</sub> repeat, or homoduplex dsDNA (Fig. 4). For fluorescent labeling, yPCNA was over-expressed with a triple FLAG epitope tag on the N-terminus, which is far away from the DNA binding site<sup>42,47</sup>. The slowly hydrolyzable ATP $\gamma$ S was used in the pre-incubation and loading reactions to capture the yRFC-yPCNA complex at the loading site. ATP $\gamma$ S prevents yPCNA ring closing and release from yRFC<sup>42,44,48</sup>. yPCNA was fluorescently labeled *in situ* with anti-FLAG functionalized quantum dots (QDs), as described previously for human

RFC-PCNA (hRFC-hPCNA)<sup>49</sup>. We confirmed that yPCNA loading was yRFC-dependent (not shown). The yRFC-C(ATP $\gamma$ S)-yPCNA complex concentration was also titrated down to yield one or fewer fluorescent yPCNAs per DNA molecule.

As expected, yRFC preferentially loads yPCNA at the 5'-ssDNA flap relative to homoduplex DNA, which is consistent with prior studies (Fig. 4a and b)<sup>50</sup>. Turning off buffer flow retracted both the DNA and yPCNA to the Cr barrier, confirming that yPCNA was loaded on the DNA (Fig. 4a, bottom). We also observed mild enrichment of yPCNA at a (CAG)<sub>13</sub> repeat (Fig. 4b, middle) relative to homoduplex DNA (Fig. 4b, bottom). To better characterize the loading efficiency on these structures, we calculated the relative enrichment of yPCNA in a 5-kb window spanning the nicking cassette on the three recombinant DNA molecules (Fig. 4c). This window was selected because it captures ~99% (~3 standard deviations of the Gaussian fit in Fig. 4b, top) of all site-specifically bound yPCNA molecules. Loading at the 5'-ssDNA flap showed 2.7-fold enrichment over homoduplex DNA, whereas loading at the (CAG)<sub>13</sub> was 1.5-fold higher than homoduplex DNA. Because the enrichment at (CAG)<sub>13</sub> was very modest, we also tested the yPCNA binding to (CAG)<sub>13</sub> inserted at site B and at site A with a flipped tethering geometry (Fig. S1). The same mild yPCNA enrichment was observed at the expected (CAG)<sub>13</sub> position in all three DNA substrates (Fig. S1). Moreover, this enrichment was statistically significant (p-value:  $1 \times 10^{-4}$ ) when compared against all possible 5 kb windows across the entire 48.5 kb DNA substrate (see Statistical Methods and Fig. S1c). We conclude that wt yRFC has a mild preference for loading yPCNA onto (CAG)<sub>13</sub> repeats relative to homoduplex DNA.

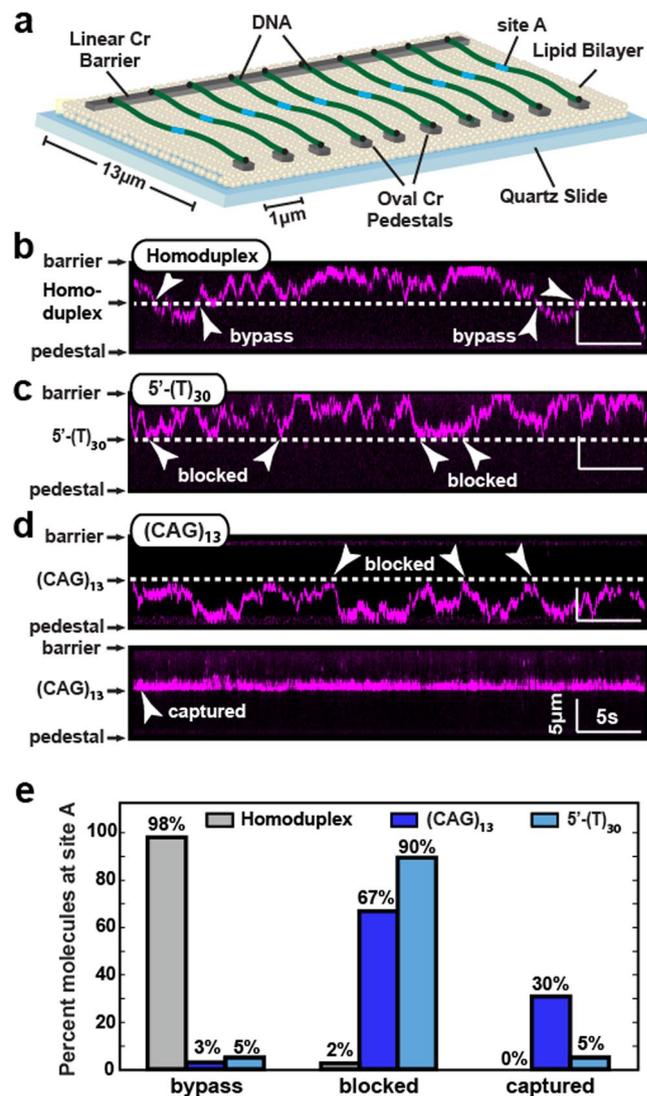
PCNA has a 3.4 nm-diameter inner opening that is wider than the diameter of B-form DNA<sup>42,45</sup>. This inner opening facilitates sliding past small mismatches and bulges, but may be blocked by larger extrahelical structures<sup>50</sup>. To directly test this hypothesis, we used double-tethered DNA curtains to observe yPCNA diffusion on various DNA substrates (Fig. 5a). In this assay, one end of the DNA was labeled with a biotin and the second end was labeled with a digoxigenin (dig). Pedestals that were located 13  $\mu$ m away from the diffusion barriers were decorated with anti-dig antibodies. The antibody-coated pedestals provide a second attachment point for dig-labeled DNA molecules<sup>32,34</sup>. The DNA molecules that are tethered between the barriers and pedestal remain fully extended without the need for additional buffer flow (Fig. 5a). As buffer flow can bias protein diffusion on DNA, double-tethered curtains provide an ideal platform for monitoring yPCNA dynamics. We first observed yPCNA diffusion on homoduplex DNA. For these experiments, ATP was added to the yRFC-yPCNA pre-incubation reaction. ATP hydrolysis catalyzes closing of the yPCNA ring and release of yRFC from DNA. To ensure complete yRFC removal from DNA, the flowcells were further washed with BSA buffer containing 300 mM NaCl. Following yRFC removal, the yPCNA molecules diffused freely over the entire length of the double-tethered DNA curtains (Fig. 5b).

Ring-like proteins can diffuse on DNA via a combination of two different modes: 1D sliding and/or hopping. During 1D sliding, the protein tracks the helical backbone of the dsDNA. In contrast, hopping is characterized by a series of correlated microscopic detachment and reattachment events. To differentiate between sliding and hopping, we measured yPCNA diffusion coefficients at increasing ionic strengths (Fig. S2). A higher ionic strength increases electrostatic screening between a protein and DNA, thereby reducing the fraction of time that a protein is in contact with the DNA. This results in increased diffusion coefficients at higher ionic strengths. yPCNA diffusion was weakly dependent on the ionic strength, in agreement with a previous study with hPCNA (Fig. S2)<sup>49</sup>. The ionic strength-dependent change in diffusion coefficients was  $0.33 \pm 0.04 \mu\text{m}^2 \text{sec}^{-1} \text{mM}^{-1}$  and  $0.25 \pm 0.06 \mu\text{m}^2 \text{sec}^{-1} \text{mM}^{-1}$  for hPCNA and yPCNA, respectively. This result indicates that like hPCNA, yPCNA also moves predominantly by 1D sliding along the helical pitch of the DNA substrate<sup>49</sup>.

We next monitored yPCNA diffusion on DNA harboring a 5'-ssDNA flap or a (CAG)<sub>13</sub> repeat (Fig. 5c,d). yPCNA can approach the flap from one of two orientations: (i) in the *cosL*  $\rightarrow$  *cosR* direction yPCNA encounters a 4-nt ssDNA gap prior to the flap and (ii) in the *cosR*  $\rightarrow$  *cosL* direction yPCNA encounters the 5'-T<sub>30</sub> tail. Out of 40 yPCNA molecules 90% (N = 36/40) were blocked by the flap in either orientation (N = 18 for *cosL*  $\rightarrow$  *cosR* and N = 18 for *cosR*  $\rightarrow$  *cosL*). The remaining molecules were either captured at the nickase cassette (5%, N = 2/40) or directly bypassed (5%, N = 2/40) the flap structure. This observation may stem from incomplete re-ligation of the flap oligo at this site. These data suggest that the flap is too large to be accommodated within the yPCNA pore. The (CAG)<sub>13</sub> repeat also blocked yPCNA that was loaded on homoduplex DNA upstream or downstream of site A (Fig. 5d,e). Remarkably, we also observed that 30% (N = 11/36) of the yPCNA molecules loaded directly at the (CAG)<sub>13</sub> were completely stationary (Fig. 5e). Moving the (CAG)<sub>13</sub> repeat to a different site in the DNA substrate also moved the location of the stationary yPCNA molecules (Fig. S3a,b). The distribution of freely diffusing, stationary, and (CAG)<sub>13</sub>-blocked yPCNAs was statistically indistinguishable across all three (CAG)<sub>13</sub>-containing DNA substrates (Fig. S3c). We conclude that yPCNA diffusion is blocked by large extrahelical structures and that yRFC loads yPCNA within the (CAG)<sub>13</sub> repeat. This result is consistent with a mechanism where TNR-bound PCNA interacts with DNA mismatch repair factors to promote TNR expansion<sup>20</sup>.

## Discussion

Here we developed a recombinering-based molecular toolbox for rapid and site-specific incorporation of exogenous DNA sequences and structures in multiple positions in  $\lambda$ -DNA. In this work, 1–1.5 kb cassettes were inserted at three positions along the  $\lambda$ -DNA. Both longer and shorter modifications are also possible, provided that dispensable segments of the phage genome are targeted for modification (Fig. 1). For such applications, we developed a series of ampicillin- and kanamycin-resistant cassettes that can accommodate large insertion fragments that are flanked by 200 bp of homology to three distinct sites within the phage DNA. These markers facilitate positive selection, ultimately yielding nearly 100% recombination efficiency. If required, the drug markers can be removed by flanking the cassettes with FRT sites and inducing Flp recombinase *in vivo*<sup>26,28</sup>. Moreover, the DNA substrate may be shortened or expanded, provided that the  $\lambda$ -DNA remains within 78–105% (38 to 53 kb) the length of the wild-type phage genome. This requirement is essential for efficient packaging of the DNA into



**Figure 5.**  $\gamma$ PCNA diffusion on  $\lambda$ -DNA containing various extrahelical structures. **(a)** An illustration of the double-tethered DNA curtains. Chromium (Cr) pedestals are coated with anti-digoxigenin antibodies, and buffer flow is used to immobilize the dig-labeled  $\lambda$ -DNA between the linear barriers and Cr pedestals. Kymograph of diffusing  $\gamma$ PCNA molecule on DNA substrates with **(b)** homoduplex DNA **(c)** a 5'-ssDNA flap or a (CAG)<sub>13</sub> repeat **(d)**. The characteristic changes in diffusion behavior of  $\gamma$ PCNA are indicated with arrows, and the dashed lines indicate site A. **(e)** Percentage of molecules showing either bypass, blocked, or captured behavior at site A. At least 35 DNA molecules were analyzed and classified into each of three categories (N = 40, 36, and 40 for the flap, (CAG)<sub>13</sub>, and homoduplex DNA substrates).

phage capsids<sup>24</sup>. This flexibility guarantees that both large and small insertions can be made within the phage genome.

We regularly achieved 90–100% recombineering efficiency using PCR constructs directed towards one of three unique sites within the  $\lambda$ -phage genome. However, we also noticed that deviating from a few best practices could produce low yields or off-target insertion products. Cryopreserving lysogens with the recombineering helper plasmid resulted in high off-target recombination rates. Substantially higher yields were achieved when the lysogens were freshly transformed with the helper plasmid and used immediately for Red system induction in mid-log cells (at an OD<sub>600</sub> ~ 0.5 or below). Additionally, recombineering efficiency was higher when the PCR products were both DpnI-treated to remove residual template DNA and gel purified prior to electroporation. After electroporation, a minimum 4-hour outgrowth should be allowed before plating the cells on antibiotic selective plates. Finally, recombineering at site A yielded slow-growing cells that required an overnight outgrowth in LB and up to two days for colonies to appear on antibiotic selective plates. Following these simple guidelines routinely resulted in highly specific modification at the desired locus.

We also demonstrated an improved nickase-based strategy for incorporating extrahelical structures at defined positions along the DNA duplex. Because this approach requires two or more closely spaced nicks, it is most commonly used with plasmid-based modifications<sup>11, 16, 17</sup>. This approach is less suitable for wild type  $\lambda$ -DNA

because it requires nicking enzymes that cleaves the substrate at hundreds of sites on both strands, leading to double-stranded breaks and low yields of full-length DNA<sup>16,17</sup>. Our approach refines the previously published strategies in four key ways: (i) a nicking cassette can be inserted into any dispensable segment of the genome, (ii) nicking at three adjacent sites exposes 39 nt of single stranded DNA, providing sufficient space to insert large oligonucleotides (e.g., long ssDNA gaps, RNA-DNA structures, and other chemically modified oligonucleotides), (iii) using a nickase with few recognition sites minimizes damage to the DNA substrate, and (iv) incorporation efficiency can be rapidly scored via digestion with one of several restriction enzymes.

Using this approach, we demonstrated that  $\gamma$ RFC preferentially loads  $\gamma$ PCNA on a 5'-ssDNA flap and a (CAG)<sub>13</sub> repeat relative to homoduplex DNA.  $\gamma$ PCNA can diffuse on homoduplex DNA but is blocked by both DNA structures. These observations suggest that  $\gamma$ PCNA cannot simultaneously accommodate both the homoduplex DNA track (2 nm diameter for B-form DNA) and the additional DNA structures within its inner ring. Molecular dynamics simulations and ensemble biochemical studies suggest that homopolymeric poly-T oligonucleotides occupy an ensemble of extended structures (e.g., 2.7 nm end-to-end distance for a (T)<sub>12</sub> oligonucleotide; we used a (T)<sub>30</sub><sup>51</sup>). Similarly, structural and biochemical studies have established that relatively short (CNG)<sub>n=3-10</sub> repeats are in a hairpin-stem structure with mismatched A-A base pairs that are flanked by two Watson-Crick G-C base pairs<sup>52-55</sup>. However, above a critical threshold of trinucleotide repeats, (CNG)<sub>n>11</sub>, the oligonucleotide transitions from a simple hairpin-stem model to bis- (or multi-branched) hairpin-like structures and bulged out ssDNA loops<sup>53,55</sup>. This dynamic, multi-branched structure likely explains why a (CAG)<sub>13</sub> repeat acts as both a loading site and a barrier to  $\gamma$ PCNA diffusion. Finally,  $\gamma$ PCNA that is directly loaded onto the (CAG)<sub>13</sub> is trapped at the lesion. These results shed light on PCNA diffusion dynamics and further highlight the utility of this molecular toolkit for both single-molecule and ensemble biochemical studies.

## Materials and Methods

**Proteins and DNA.** FLAG-epitope labeled LacI was purified from BL21(DE3) as described previously<sup>56</sup>. Biotinylated anti-FLAG M2 antibodies were purchased from Sigma and streptavidin-conjugated quantum dots (QDs) were purchased from Life Technologies. All oligonucleotides were ordered from IDT and are summarized in Table S1. Plasmids containing an antibiotic cassette flanked by two 200-basepair segments that are homologous to the  $\lambda$ -phage genome were custom synthesized by BioMatik. Plasmid maps and sequences are available via Benchling (<https://benchling.com/ifinkelstein>). All plasmid modifications were carried out using the Q5 inverse PCR mutagenesis kit (NEB).

**S. cerevisiae Replication Factor C ( $\gamma$ RFC).** The plasmids pLant2b-RFC-AE and pET11-RFC-BCD were kindly provided by Manju Hingorani<sup>46</sup>. The two plasmids were co-transformed into BL21(DE3) ArcticExpress cells (Agilent). A single colony was inoculated into 680  $\mu$ L of LB and grown overnight at 37 °C. 200  $\mu$ L of these cells were then seeded into 2 L LB with 50  $\mu$ g mL<sup>-1</sup> kanamycin and 50  $\mu$ g mL<sup>-1</sup> carbenicillin. The culture was grown at 30 °C to an OD<sub>600</sub> ~ 0.6, cooled on ice with swirling to 16 °C and induced with 0.5 mM IPTG. Induction continued for 16 hours at 12 °C. Cells were harvested by centrifuging for 15 minutes (3,300 RCF at 4 °C) and resuspended in 20 mL of lysis buffer (30 mM HEPES [pH 7.5], 0.25 mM EDTA, 5% (v/v) glycerol). The resuspended cell paste were either frozen in liquid nitrogen and stored at -80 °C, or prepared for lysis by adding 250 mM NaCl, 1 mM phenylmethanesulfonyl fluoride (PMSF, Sigma-Aldrich), and 1x HALT protease inhibitor (Thermo-Fisher). Cells were lysed in a homogenizer (Avestin), and centrifuged (140,000 RCF, at 4 °C) for 35 minutes. A home-made 7 mL SP FF column (resin from GE Healthcare) was equilibrated with 35 mL of buffer A (30 mM HEPES [pH 7.5], 0.25 mM EDTA, 5% (v/v) glycerol) + 200 mM NaCl and the clarified lysate was loaded onto the equilibrated column at 0.5 mL min<sup>-1</sup>. The column was then washed with 35 mL of buffer A + 250 mM NaCl and the protein was eluted with a 100 mL gradient to buffer B (30 mM HEPES [pH 7.5], 0.25 mM EDTA, 5% (v/v) glycerol, 1 M NaCl) at 1 mL min<sup>-1</sup>.  $\gamma$ RFC-containing fractions were pooled and developed through a 1 mL Q HP column (GE Healthcare) pre-equilibrated with buffer A + 100 mM NaCl. The combined sample from the SP column was loaded onto the Q column at a rate of 0.7 mL min<sup>-1</sup>. The column was washed with 10 mL of buffer A + 110 mM NaCl at 0.8 mL min<sup>-1</sup>, and then eluted with an 18 mL gradient into buffer B at 0.8 mL min<sup>-1</sup>. The eluted Q column fractions were analyzed on a 10% SDS-PAGE gel, and the fractions containing  $\gamma$ RFC were combined, frozen in liquid nitrogen, and stored at -80 °C. The  $\gamma$ RFC concentration was determined by comparison to a BSA titration curve using SDS-PAGE.

**Proliferating Cell Nuclear Antigen (PCNA).** A plasmid expressing *Saccharomyces cerevisiae* His<sub>6</sub>-PCNA was kindly provided by Francisco Blanco<sup>57</sup>. A triple FLAG epitope was introduced at the N-terminus via inverse PCR mutagenesis (NEB) using primers YK\_PCNA01 and YK\_PCNA02 (see Table S1) to generate pIF105. pIF105 was transformed into BL21(DE3) codon plus RIL cells. A colony was inoculated into 30 mL LB with 50  $\mu$ g mL<sup>-1</sup> kanamycin and 34  $\mu$ g mL<sup>-1</sup> chloramphenicol, and grown overnight at 37 °C. Ten mL of the overnight culture were seeded into 1 L LB and grown in the presence of both antibiotics. When the culture reached an OD<sub>600</sub> ~ 0.5, 0.8 mM IPTG was added and induction continued at 37 °C for 4 hours. Cells were harvested by centrifugation at 3,000 RCF for 10 minutes, and resuspended in 50 mL lysis buffer (50 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.5 mM TECP, 10% (v/v) glycerol) with 1x HALT protease inhibitor. Cells were lysed by sonication on ice and centrifuged at 95,000 RCF for 30 minutes. Imidazole was added to the supernatant to a final concentration of 30 mM. A 5 mL HisTrap HP column (GE Healthcare) was pre-equilibrated with 50 mL lysis buffer and the lysate was loaded onto the column and washed with 50 mL of Ni-buffer (50 mM Tris-HCl [pH 7.6], 150 mM NaCl, 10% glycerol (v/v), 30 mM imidazole). PCNA was eluted with a ~110 mL gradient to Ni-buffer + 500 mM imidazole over 120 minutes. PCNA-containing fractions were identified via 12% SDS-PAGE, dialyzed into storage buffer (20 mM Tris-HCl [pH 7.6], 100 mM NaCl, 10% (v/v) glycerol, 1 mM DTT) and concentrated using a centrifugal filter (10 kDa Amicon, Millipore). Small aliquots were frozen in liquid nitrogen and stored at -80 °C.

**Recombineering  $\lambda$ -phage lysogens.** Red-based *in vivo* recombination was used to construct recombinant  $\lambda$ -phage DNA<sup>26,58</sup>. First, a lysogen was created by packaging  $\lambda$ -phage DNA (cI857*ind 1 Sam 7*; NEB) into empty phage particles (MaxPlax Lambda Packaging Kit, EpiCentre). The packaged phage was used to infect *E. coli* strain 7723 and lysogens were identified via colony PCR and sensitivity to growth at a restrictive temperature (42–45 °C). The resulting cells (IF189) were transformed with pKD78, which harbors the Red recombineering system under the control of an arabinose-inducible promoter<sup>28</sup>. The insertion cassettes were PCR amplified from helper plasmids (Fig. 1) with custom primers (AD027 and AD028 for site A, AD012 and AD013 for site B, or AD016 and AD017 for site C; see Table S1) using *Taq* DNA polymerase (NEB# M0320S). PCR reactions were treated with 1 U DpnI (NEB #R0176s) at 37 °C for 1 hour and purified by gel extraction (Qiagen Kit #28704) to remove residual plasmid template, which can create false positives during recombineering<sup>26</sup>. The gel-extracted DNA was re-suspended in Milli-Q water to a final concentration of 100–150  $\mu\text{g}\mu\text{L}^{-1}$  and used as the targeting DNA for the recombineering reaction.

Fresh electrocompetent cells were prepared for every recombineering reaction. A 5 mL LB culture of strain IF189 ( $\lambda$ -lysogen) transformed with pKD78 was grown overnight at 30 °C in the presence of 10  $\mu\text{g}\mu\text{L}^{-1}$  chloramphenicol. The following day, 350  $\mu\text{L}$  of cells were used to inoculate a fresh 35 mL culture of LB containing the same concentration of antibiotic. When the cells reached an O.D.<sub>600</sub> ~ 0.5, the Red recombinase system was induced by adding 2% L-arabinose (GoldBio) and incubated for an additional 1 hour at 30 °C. Cells were harvested at 4,500 RCF for 7 min, washed three times in ice-cold Milli-Q H<sub>2</sub>O, and finally resuspended in 200  $\mu\text{L}$  of H<sub>2</sub>O<sup>59</sup>. Electrocompetent cells were kept on ice and used immediately for the recombineering reaction.

For recombineering, 50–150 ng of targeting DNA was electroporated at 18 kV cm<sup>-1</sup> in 0.1 cm cuvettes using a micropulser (Biorad #165-210). Cells were immediately resuspended in 1 mL of SOC and then transferred to culture tubes containing 10 mL LB broth. After a 4 hour outgrowth at 30 °C, 100  $\mu\text{L}$  of the culture was plated onto LB agar plates containing either 30  $\mu\text{g}\text{mL}^{-1}$  carbenicillin or 30  $\mu\text{g}\text{mL}^{-1}$  kanamycin. Colonies were checked for successful incorporation of recombinant DNA via colony PCR with oligos AD031 and AD032 for site A, AD022 and AD023 for site B, or AD024 and AD025 for site C (see Table S1).

**Purifying phage DNA from lysogens.** Expression of recombinant phage was induced via heat shock. A single lysogenic colony was grown in 50 mL of LB broth with the appropriate antibiotic overnight at 30 °C. 5 mL of this starter culture was used to inoculate 500 mL of LB the following morning. When the flask reached an O.D.<sub>600</sub> ~ 0.6, the temperature was rapidly raised to 42 °C in a water bath. The culture was placed at 45 °C in a shaking incubator for 15 minutes and then transferred to a 37 °C incubator for two hours. To liberate the phage particles, cells were harvested by centrifugation at 3,000 RCF for 30 minutes and lysed via re-suspension in 10 mL of SM buffer (50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 8 mM MgSO<sub>4</sub>) + 2% chloroform, and rotated at 37 °C for 30 min. A subsequent 1 hour incubation with 50 ng  $\mu\text{L}^{-1}$  DNaseI (Sigma# D2821) and 30 ng  $\mu\text{L}^{-1}$  RNaseA (Sigma# R6513) degraded the bacterial genomic DNA and RNA. The clarified lysate, containing soluble phage capsids, was obtained by centrifugation for 15 minutes at 6,000 RCF and 4 °C, and further diluted with 40 mL of SM buffer. Phage capsids were precipitated by incubating for 1 hour with 10 mL ice-cold buffer L2 (30% PEG 6000, 3 M NaCl) and then harvest by centrifugation at 10,000 RCF for 10 minutes at 4 °C. The phage pellet was washed with 1 mL of buffer L3 (100 mM Tris-HCl [pH 7.5], 100 mM NaCl, 25 mM EDTA) and then re-suspended with 3 mL of buffer L3, followed by an equal volume of buffer L4 (4% SDS). The phage capsid proteins were further digested by incubation with 100 ng  $\mu\text{L}^{-1}$  of proteinase K (NEB #P8012S) for 1 hour at 55 °C. SDS was precipitated with 3 mL buffer L5 (3 M potassium acetate [pH 7.5]), and the cloudy solution was clarified by centrifugation at 15,000 RCF for 30 minutes at 4 °C. The soluble phage DNA was passed over a pre-equilibrated Qiagen tip-500 column (Qiagen #10262), washed with buffer QC (1.0 M NaCl, 50 mM MOPS [pH 7.0], 15% isopropanol) and eluted with 15 mL buffer QF (1.25 M NaCl, 50 mM Tris [pH 8.5], 15% isopropanol). Finally, the DNA was precipitated with the addition of 10.5 mL of 100% isopropanol, rinsed in 70% ethanol and re-dissolved in TE buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA) to a final DNA concentration of 200–500 ng  $\mu\text{L}^{-1}$ . We routinely obtained ~250  $\mu\text{g}$  of pure  $\lambda$ -DNA from a single purification.

**Inserting synthetic oligonucleotides into  $\lambda$ -DNA.** Recombinant  $\lambda$ -DNA was obtained from strain IF189, which was modified and purified as described above, and 25  $\mu\text{g}$  of the DNA was incubated with 150 U of Nt.BspQI (NEB# R0644S) in a 250  $\mu\text{L}$  reaction with 1X buffer 3.1 (NEB #B7203) at 55 °C for 1 hour. The reaction was halted with 1 U of proteinase K (NEB #P8107S) for 1 hour at 55 °C. The nicked DNA was mixed with a 100-fold molar excess of the desired insert oligo (AD006 for 5'-flap, YK105 for (CAG)<sub>13</sub>, MB32 for mock insert, Table S1), along with a 10-fold excess of *cosL* and *cosR*-complementary oligos (IF003 and IF004 for *cosL* → *cosR*, IF001 and IF002 for *cosR* → *cosL*, Table S1). The solution was heated to 70 °C for 15 minutes followed by slow cooling to 22 °C in a thermocycler at a rate of -0.5 °C min<sup>-1</sup>. The annealed mixture was supplemented with 6000 U of T4 DNA ligase (NEB #M0202L) and 1 mM ATP and further incubated overnight at room temperature in a final reaction volume of 300  $\mu\text{L}$ . A 50  $\mu\text{L}$  aliquot was taken for alkaline agarose gel and restriction enzyme digest analysis. The remaining 250  $\mu\text{L}$  was inactivated with high salt (1 M NaCl) and then passed through a 120 mL Sephacryl S-1000 column (GE #17-0476-01), in TE running buffer (10 mM Tris-HCl [pH 8.0], 1 mM EDTA) plus 150 mM NaCl, to separate the modified  $\lambda$  DNA from excess oligos and enzymes. Purified DNA was stored at 4 °C.

To verify incorporation of the insert, a restriction enzyme digests was performed with NcoI, an enzyme whose recognition site is abolished upon incorporation of the insert into our recombineered  $\lambda$ -DNA. 5  $\mu\text{L}$  of the above mentioned ligation reaction was digested with 20 U of NcoI (NEB #R13193) in a 30  $\mu\text{L}$  reaction with 1X Cutsmart buffer (NEB #B7204) for 1 hour at 37 °C and run on a 0.8% agarose gel stained with ethidium bromide. To check for complete ligation, a 5  $\mu\text{L}$  aliquot was run over a 0.6% alkaline agarose gel at 15 mV for 20 hours at 4 °C. The gel was then submerged in neutralization solution (1 M Tris-HCl [pH 7.6], 1.5 M NaCl) for one hour, and stained in a

solution of 10 mg mL<sup>-1</sup> ethidium bromide for an additional hour. The stained gel was visualized using a Typhoon FLA 9500 laser scanner (GE).

**Imaging Lac repressor (LacI) on DNA curtains.** To obtain distributions of triple FLAG epitope tagged LacI binding to recombiner Lac operator  $\lambda$ -DNA, we incubated 120 nM LacI with 10–25 ng  $\mu$ L<sup>-1</sup> of biotinylated  $\lambda$ -DNA in a 150  $\mu$ L volume of BSA buffer (40 mM Tris-HCl [pH 7.6], 50 mM NaCl, 2 mM MgCl<sub>2</sub>, 0.2 mg mL<sup>-1</sup> BSA, 1 mM DTT) supplemented with 100 mM NaCl, for 5 min on ice. This mixture was diluted to a 1 mL final volume with BSA buffer and injected into flowcells, containing a streptavidin-activated lipid bilayer, to produce single-tethered DNA curtains bound with LacI at the Lac operator site. LacI was then fluorescently labeled *in situ* with an injection of 1.5 nM of anti-FLAG antibodies (Sigma-Aldrich #F9291) conjugated to fluorescent quantum dots (QDs, Life Tech #Q10163MP). The antibodies and QDs were pre-conjugated by mixing in a test tube for 10 minutes on ice prior to injection into the flowcell.

**Loading PCNA on DNA curtains.** To observe the binding distribution of yPCNA on DNA, 0.8 nM yRFC was mixed with 1 nM yPCNA (concentrations reported for the trimer) in BSA buffer supplemented with 100  $\mu$ M ATP $\gamma$ S (Roche). The yPCNA-yRFC(ATP $\gamma$ S) complex was injected into flowcells containing single-tethered DNA curtains, and incubated for 6 minutes at room temperature. Following this incubation, yPCNA was labeled *in situ* by injecting 1.5 nM of biotinylated anti-FLAG antibodies conjugated to fluorescent streptavidin-conjugated quantum dots (QDs, emitting at 705 nm, Thermo-Fisher). The antibodies and QDs were pre-conjugated in a test tube on ice prior to injection into the flowcell.

The same protocol was used to measure yPCNA diffusion dynamics with the flowing modifications. First, 1 mM ATP was used instead of 100  $\mu$ M ATP $\gamma$ S during the yPCNA-yRFC(ATP) incubation. Second, 5 minutes after antibody-coupled QDs were injected into the flowcell, residual yRFC and free QDs were washed out by injecting 200  $\mu$ L of BSA buffer +300 mM NaCl over 1 minute. yPCNA diffusion data was acquired for 10 minutes after this yRFC washing step.

**Single-Molecule Microscopy.** DNA curtains were assembled in custom-made flowcells consisting of a microscope slide patterned with 1–2  $\mu$ m-wide Chromium barriers to lipid diffusion. These features were deposited using standard UV lithography, as described previously<sup>32</sup>. A jeweler's drill press (Servo Products, Model #7110) and diamond-plated bits (Jewelry Tools, Product ID DIB-211.00) were used to drill two inlet ports along the long face of the quartz slides. Next, 5 mm-wide channels were cut in double-sided sticky tape (3 M #665) and sandwiched between the microfabricated quartz slide and a glass coverslip (#50-948-951, Fisher Scientific). This sandwich was held together with binder clips and baked at 150 °C for 1 hour. Finally, nanoports (IDEX #N-333) were glued to the drilled side of the assembled flowcell using a hobbyist hot glue gun (Superbond, #HE-750).

Single-molecule fluorescence images were collected with a Nikon Ti-E microscope in a prism-TIRF configuration. The inverted microscope setup allowed for the sample to be illuminated by a 488 nm laser light (Coherent) through a quartz prism (Tower Optical Co.). The laser was adjusted to deliver 40 mW of power at the front face of the prism. Experiments were conducted on a floating TMC optical table to avoid spatial drift. A 60x water immersion objective lens (1.2 NA, Nikon), two EM-CCD cameras (Andor iXon DU897, –80 °C) and NIS-Elements software (Nikon) were used to collect the data with a 50 ms exposure time. Two-color imaging was conducted using a 638 nm dichroic beam splitter (Chroma). Frames were saved as TIFF files without compression for further image analysis in FIJI, a distribution of the open-source ImageJ software (NIH)<sup>60</sup>.

**Data Analysis.** *Particle tracking.* Fluorescent particles were tracked in FIJI with a custom-written particle tracking script (available upon request)<sup>60</sup>. The resulting trajectories were analyzed in MATLAB R2014a (Mathworks). The fluorescent particle in each frame was fit to a two-dimensional Gaussian function to obtain its position with sub-pixel resolution. The series of positions of a given particle were used to obtain trajectories<sup>56,61</sup>.

yPCNA-binding distributions were determined by calculating the distance between the center of DNA-bound fluorescent particles and the chromium barrier. yPCNAs were deemed to be on the DNA substrate if they retracted to the chromium barrier when buffer flow was toggled off and on. All binding distributions were collected from at least three flowcells. When indicated, these were fit to a Gaussian distribution using the MATLAB curve-fitting toolbox.

**Statistical methods.** Error bars on the yPCNA binding distributions were calculated in MATLAB using bootstrap analysis with replacement<sup>62</sup>. Statistical tests were conducted in the PAST3 software package<sup>63</sup> or in MATLAB (Mathworks). A Pearson's chi-squared test was used to determine whether the different DNA substrates have statistically significant effects on yPCNA sliding (Tables S2 and S3). The significance threshold was set at 0.05 in all tests.

Wild type yRFC nonspecifically loads yPCNA on homoduplex DNA<sup>43</sup>. This makes it difficult to determine whether the mild enrichment at the (CAG)<sub>13</sub> repeat is statistically significant for any single target DNA. Therefore, we used the two-tailed t test to determine whether loading yPCNA at the (CAG)<sub>13</sub> repeat on all three DNA substrates has the same normalized frequency as loading on all homoduplex DNA sequences (the null hypothesis). For this test, the entire DNA substrate was divided into eight windows, each containing 5 kb of DNA. A 5 kb window captures >99% of all target-bound yPCNAs within our spatial resolution, as determined by analysis of the 5'-ssDNA flap structure (Fig. 4). Thus, each modified DNA substrate has a single window with a (CAG)<sub>13</sub> repeat and 7 windows with nonspecific homoduplex DNA. A two-tailed t-test demonstrated that the mild enrichment in the (CAG)<sub>13</sub> window is statistically significant relative to yPCNA binding in all other homoduplex windows (Fig. S3).

**Measuring diffusion coefficients.** yPCNA diffusion coefficients were measured on double-tethered DNA curtains that extend DNA molecules between two micro-fabricated chromium features in the absence of buffer flow<sup>32,34</sup>. The average separation between the two chromium features was 13 μm (~80% extension relative to B-form DNA). Trajectories of individual molecules were used to calculate the one-dimensional (1D) mean squared displacement (MSD) as a function of time interval using:

$$MSD(n\Delta t) = \frac{1}{N-n} \sum_{i=1}^{N-n} (y_{i+n} - y_i)^2 \quad (1)$$

where  $N$  is the total number of frames in the trajectory,  $n$  is the number of frames for a given time interval,  $\Delta t$  is the time between frames, and  $y_i$  is the yPCNA position at frame  $i$ . The MSD was calculated for the first 10 time intervals ( $\Delta t = 0.05$  s to 0.5 s) and plotted as a function of  $\Delta t$  to generate the line:

$$MSD(\Delta t) = 2D\Delta t \quad (2)$$

where  $D$  is the diffusion coefficient. Plots were fit to a line in MATLAB and the slope was used to calculate diffusion coefficients of individual yPCNA molecules. Diffusion coefficients were calculated for  $\geq 30$  molecules per NaCl concentration, and are reported as a mean  $\pm$  standard error (S.E.).

## References

- Bustamante, C. In singulo Biochemistry: When Less Is More. *Annual Review of Biochemistry* **77**, 45–50, doi:10.1146/annurev.biochem.012108.120952 (2008).
- Ha, T. Single-molecule methods leap ahead. *Nat Meth* **11**, 1015–1018, doi:10.1038/nmeth.3107 (2014).
- Ngo, T. T. M., Zhang, Q., Zhou, R., Yodh, J. G. & Ha, T. Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility. *Cell* **160**, 1135–1144, doi:10.1016/j.cell.2015.02.001 (2015).
- Kemmerich, F. E. *et al.* Simultaneous Single-Molecule Force and Fluorescence Sampling of DNA Nanostructure Conformations Using Magnetic Tweezers. *Nano Lett.* **16**, 381–386, doi:10.1021/acs.nanolett.5b03956 (2016).
- Candelli, A., Wuite, G. J. L. & Peterman, E. J. G. Combining optical trapping, fluorescence microscopy and micro-fluidics for single molecule studies of DNA–protein interactions. *Physical Chemistry Chemical Physics* **13**, 7263–2853, doi:10.1039/c0cp02844d (2011).
- Cordova, J. C., Das, D. K., Manning, H. W. & Lang, M. J. Combining single-molecule manipulation and single-molecule detection. *Curr. Opin. Struct. Biol.* **28**, 142–148, doi:10.1016/j.sbi.2014.08.010 (2014).
- Chemla, Y. R. High-resolution, hybrid optical trapping methods, and their application to nucleic acid processing proteins. *Biopolymers* **105**, 704–714, doi:10.1002/bip.22880 (2016).
- Liu, S. *et al.* A Viral Packaging Motor Varies Its DNA Rotation and Step Size to Preserve Subunit Coordination as the Capsid Fills. *Cell* **157**, 702–713, doi:10.1016/j.cell.2014.02.034 (2014).
- Thorpe, H. M. & Smith, M. C. *In vitro* site-specific integration of bacteriophage DNA catalyzed by a recombinase of the resolvase/invertase family. *Proc. Natl. Acad. Sci. USA* **95**, 5505–5510, doi:10.1073/pnas.95.10.5505 (1998).
- Bell, J. C., Plank, J. L., Dombrowski, C. C. & Kowalczykowski, S. C. Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. *Nature* **491**, 274–8, doi:10.1038/nature11598 (2012).
- Luzzietti, N., Knappe, S., Richter, I. & Seidel, R. Nicking enzyme-based internal labeling of DNA at multiple loci. *Nat. Protocols* **7**, 643–653, doi:10.1038/nprot.2012.008 (2012).
- Luzzietti, N. *et al.* Efficient preparation of internally modified single-molecule constructs using nicking enzymes. *Nucl. Acids Res.* **39**, e15–e15, doi:10.1093/nar/gkq1004 (2011).
- Berezney, J. P. & Saleh, O. A. Locked nucleic acid oligomers as handles for single molecule manipulation. *Nucl. Acids Res.* **42**, e150–e150, doi:10.1093/nar/gku760 (2014).
- Zohar, H., Hetherington, C. L., Bustamante, C. & Muller, S. J. Peptide Nucleic Acids as Tools for Single-Molecule Sequence Detection and Manipulation. *Nano Lett.* **10**, 4697–4701, doi:10.1021/nl102986v (2010).
- Zohar, H. & Muller, S. J. Labeling DNA for single-molecule experiments: methods of labeling internal specific sequences on double-stranded DNA. *Nanoscale* **3**, 3027–3039, doi:10.1039/c1nr10280j (2011).
- Kuhn, H. & Frank-Kamenetskii, M. D. Labeling of unique sequences in double-stranded DNA at sites of vicinal nicks generated by nicking endonucleases. *Nucl. Acids Res.* **36**, e40–e40, doi:10.1093/nar/gkn107 (2008).
- Loparo, J. J., Kulczyk, A. W., Richardson, C. C. & van Oijen, A. M. Simultaneous single-molecule measurements of phage T7 replisome composition and function reveal the mechanism of polymerase exchange. *Proceedings of the National Academy of Sciences* **108**, 3584–3589, doi:10.1073/pnas.1018824108 (2011).
- Yardimci, H. *et al.* Bypass of a protein barrier by a replicative DNA helicase. *Nature* **492**, 205–209, doi:10.1038/nature11730 (2012).
- Gorman, J. *et al.* Single-molecule imaging reveals target-search mechanisms during DNA mismatch repair. *Proc Natl Acad Sci USA* **109**, e3074–e3083, doi:10.1073/pnas.1211364109 (2012).
- Pluciennik, A. *et al.* Extrahelical (CAG)/(CTG) triplet repeat elements support proliferating cell nuclear antigen loading and MutL $\alpha$  endonuclease activation. *Proc. Natl. Acad. Sci. USA* **110**, 12277–12282, doi:10.1073/pnas.1311325110 (2013).
- Iyer, R. R., Pluciennik, A., Napierala, M. & Wells, R. D. DNA triplet repeat expansion and mismatch repair. *Annual Review of Biochemistry* **84**, 199–226, doi:10.1146/annurev-biochem-060614-034010 (2015).
- Schmidt, M. H. M. & Pearson, C. E. Disease-associated repeat instability and mismatch repair. *DNA Repair (Amst.)* **38**, 117–126, doi:10.1016/j.dnarep.2015.11.008 (2016).
- Hendrix, R. W. *Lambda II*. (Cold Spring Harbor Laboratory, 1983).
- Chauthaiwale, V. M., Therwath, A. & Deshpande, V. V. Bacteriophage lambda as a cloning vector. *Microbiol Rev* **56**, 577–591 (1992).
- Hernandez, V. J., Edlind, T. D., Young, R. F. & Ihler, G. M. The DNA between Rz and cosR in bacteriophage lambda is nonessential. *Gene* **33**, 363–365, doi:10.1016/0378-1119(85)90245-8 (1985).
- Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: A Homologous Recombination-Based Method of Genetic Engineering. *Nat Protoc* **4**, 206–223, doi:10.1038/nprot.2008.227 (2009).
- Pines, G., Freed, E. F., Winkler, J. D. & Gill, R. T. Bacterial Recombineering: Genome Engineering via Phage-Based Homologous Recombination. *ACS Synth Biol* **4**, 1176–1185, doi:10.1021/acssynbio.5b00009 (2015).
- Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645, doi:10.1073/pnas.120163297 (2000).
- Liu, P., Jenkins, N. A. & Copeland, N. G. A Highly Efficient Recombineering-Based Method for Generating Conditional Knockout Mutations. *Genome Res.* **13**, 476–484, doi:10.1101/gr.749203 (2003).
- George, H. J., Watson, R. J., Harbrecht, D. F. & DeLorbe, W. J. A Bacteriophage  $\lambda$  cI857 Cassette Controls  $\lambda$  PL Expression Vectors at Physiological Temperatures. *Nat Biotech* **5**, 600–603 (1987).

31. Chang, C. Y., Nam, K. & Young, R. S gene expression and the timing of lysis by bacteriophage lambda. *J Bacteriol* **177**, 3283–3294, doi:10.1128/jb.177.11.3283-3294.1995 (1995).
32. Gallardo, I. F. *et al.* High-Throughput Universal DNA Curtain Arrays for Single-Molecule Fluorescence Imaging. *Langmuir* **31**, 10310–10317, doi:10.1021/acs.langmuir.5b02416 (2015).
33. Robison, A. D. & Finkelstein, I. J. Rapid prototyping of multichannel microfluidic devices for single-molecule DNA curtain imaging. *Anal. Chem.* **86**, 4157–4163, doi:10.1021/ac500267v (2014).
34. Finkelstein, I. J. & Greene, E. C. Supported lipid bilayers and DNA curtains for high-throughput single-molecule studies. *Methods Mol. Biol.* **745**, 447–461, doi:10.1007/978-1-61779-129-1\_26 (2011).
35. Simons, A., Tils, D., von Wilcken-Bergmann, B. & Müller-Hill, B. Possible ideal lac operator: *Escherichia coli* lac operator-like sequences from eukaryotic genomes lack the central G X C pair. *Proc Natl Acad Sci USA* **81**, 1624–1628, doi:10.1073/pnas.81.6.1624 (1984).
36. Finkelstein, I. J., Visnapuu, M.-L. & Greene, E. C. Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature* **468**, 983–987, doi:10.1038/nature09561 (2010).
37. Price, A. C. *et al.* DNA Motion Capture Reveals the Mechanical Properties of DNA at the Mesoscale. *Biophysical Journal* **108**, 2532–2540, doi:10.1016/j.bpj.2015.04.022 (2015).
38. Lee, J. Y., Finkelstein, I. J., Arciszewska, L. K., Sherratt, D. J. & Greene, E. C. Single-molecule imaging of FtsK translocation reveals mechanistic features of protein-protein collisions on DNA. *Mol. Cell* **54**, 832–843, doi:10.1016/j.molcel.2014.03.033 (2014).
39. Myler, L. R. *et al.* Single-molecule imaging reveals the mechanism of Exo1 regulation by single-stranded DNA binding proteins. *Proc. Natl. Acad. Sci. USA* **113**, e1170–e1179, doi:10.1073/pnas.1516674113 (2016).
40. Brown, M. W. *et al.* Dynamic DNA binding licenses a repair factor to bypass roadblocks in search of DNA lesions. *Nat Commun* **7**, 10607, doi:10.1038/ncomms10607 (2016).
41. Fien, K. & Stillman, B. Identification of replication factor C from *Saccharomyces cerevisiae*: a component of the leading-strand DNA replication complex. *Mol. Cell. Biol.* **12**, 155–163, doi:10.1128/MCB.12.1.155 (1992).
42. Krishna, T. S. R., Kong, X.-P., Gary, S., Burgers, P. M. & Kuriyan, J. Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* **79**, 1233–1243, doi:10.1016/0092-8674(94)90014-0 (1994).
43. Majka, J. & Burgers, P. M. J. The PCNA-RFC families of DNA clamps and clamp loaders. *Prog. Nucleic Acid Res. Mol. Biol.* **78**, 227–260, doi:10.1016/S0079-6603(04)78006-X (2004).
44. Gomes, X. V. & Burgers, P. M. J. ATP Utilization by Yeast Replication Factor I. ATP-Mediated Interaction with DNA and with Proliferating Cell Nuclear Antigen. *J. Biol. Chem.* **276**, 34768–34775, doi:10.1074/jbc.M011631200 (2001).
45. Bowman, G. D., O'Donnell, M. & Kuriyan, J. Structural analysis of a eukaryotic sliding DNA clamp–clamp loader complex. *Nature* **429**, 724–730, doi:10.1038/nature02585 (2004).
46. Finkelstein, J., Antony, E., Hingorani, M. M. & O'Donnell, M. Overproduction and analysis of eukaryotic multiprotein complexes in *Escherichia coli* using a dual-vector strategy. *Analytical Biochemistry* **319**, 78–87, doi:10.1016/S0003-2697(03)00273-2 (2003).
47. Dieckman, L. M., Freudenthal, B. D. & Washington, M. T. PCNA structure and function: insights from structures of PCNA complexes and post-translationally modified PCNA. *Subcell. Biochem.* **62**, 281–299, doi:10.1007/978-94-007-4572-8\_15 (2012).
48. Gomes, X. V., Schmidt, S. L. G. & Burgers, P. M. J. ATP Utilization by Yeast Replication Factor II. Multiple Stepwise ATP Binding Events are Required to Load Proliferating Cell Nuclear Antigen onto Primed DNA. *J. Biol. Chem.* **276**, 34776–34783, doi:10.1074/jbc.M011743200 (2001).
49. Kochaniak, A. B. *et al.* Proliferating cell nuclear antigen uses two distinct modes to move along DNA. *J. Biol. Chem.* **284**, 17700–17710, doi:10.1074/jbc.M109.008706 (2009).
50. Yao, N., Hurwitz, J. & O'Donnell, M. Dynamics of  $\beta$  and Proliferating Cell Nuclear Antigen Sliding Clamps in Traversing DNA Secondary Structure. *J. Biol. Chem.* **275**, 1421–1432, doi:10.1074/jbc.275.2.1421 (2000).
51. Biyani, M. & Nishigaki, K. Single-Strand Conformation Polymorphism (SSCP) of Oligodeoxyribonucleotides: An Insight into Solution Structural Dynamics of DNAs Provided by Gel Electrophoresis and Molecular Dynamics Simulations. *J Biochem* **138**, 363–373, doi:10.1093/jb/mvi149 (2005).
52. Mariappan, S. V. *et al.* Solution structures of the Huntington's disease DNA triplets, (CAG)<sub>n</sub>. *J. Biomol. Struct. Dyn.* **15**, 723–744, doi:10.1080/07391102.1998.10508988 (1998).
53. Amrane, S. *et al.* Length-dependent energetics of (CTG)<sub>n</sub> and (CAG)<sub>n</sub> trinucleotide repeats. *Nucl. Acids Res.* **33**, 4065–4077, doi:10.1093/nar/gki716 (2005).
54. Lee, B. J. *et al.* Structure and Dynamics in DNA Looped Domains: CAG Triplet Repeat Sequence Dynamics Probed by 2-Aminopurine Fluorescence. *Biochemistry* **46**, 10756–10766, doi:10.1021/bi7005674 (2007).
55. Degtyareva, N. N., Barber, C. A., Reddish, M. J. & Petty, J. T. Sequence Length Dictates Repeated CAG Folding in Three-Way Junctions. *Biochemistry* **50**, 458–465, doi:10.1021/bi101756e (2011).
56. Finkelstein, I. J., Visnapuu, M.-L. & Greene, E. C. Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature* **468**, 983–987, doi:10.1038/nature09561 (2010).
57. De Biasio, A. *et al.* Reduced Stability and Increased Dynamics in the Human Proliferating Cell Nuclear Antigen (PCNA) Relative to the Yeast Homolog. *PLoS ONE* **6**, e16600, doi:10.1371/journal.pone.0016600 (2011).
58. Oppenheim, A. B., Rattray, A. J., Bubunenko, M., Thomason, L. C. & Court, D. L. *In vivo* recombineering of bacteriophage  $\lambda$  by PCR fragments and single-strand oligonucleotides. *Virology* **319**, 185–189, doi:10.1016/j.virol.2003.11.007 (2004).
59. Sambrook, J. & Russell, D. W. Transformation of *E. coli* by Electroporation. *Cold Spring Harb Protoc* **2006**, pdb.prot3933 (2006).
60. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682, doi:10.1038/nmeth.2019 (2012).
61. Yildiz, A. *et al.* Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization. *Science* **300**, 2061–2065, doi:10.1126/science.1084398 (2003).
62. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (Chapman and Hall/CRC, 1993).
63. Hammer, Ø., Harper, D. A. & Ryan, P. D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontologia Electronica* **4**, 1–9 (2001).

## Acknowledgements

We are indebted to Manju Hingorani, Bruce Stillman, and Francisco Blanco for sharing plasmids. We are grateful to members of the Finkelstein lab for carefully reading the manuscript. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was supported by the National Science Foundation (1453358 to I.J.F.), the Institute of General Medical Sciences of the National Institutes of Health (GM097177 and GM120554 to I.J.F.), National Cancer Institute (P01 CA092584 to I.J.F.), CPRIT (R1214 to I.J.F.), and the Welch Foundation (F-1808 to I.J.F.). I.J.F. is a CPRIT Scholar in Cancer Research. Yoori Kim is a Howard Hughes Medical Institute international graduate student fellow.

### Author Contributions

Y.K., A.D. and I.J.F. designed the experiments and wrote the main manuscript text. Y.K., A.D. and A.A.L. performed the experiments. A.D. and A.A.L. contributed to Figures 1–3 and Y.K. prepared Figures 4 and 5, as well as the supplemental figures. Y.K. and A.D. contributed equally to the preparation of the manuscript. All authors have reviewed and given approval to the submitted manuscript.

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-01984-x](https://doi.org/10.1038/s41598-017-01984-x)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017