



Published in final edited form as:

Nature. 2020 October ; 586(7831): 801–806. doi:10.1038/s41586-020-2820-9.

Mechanism of strand exchange from RecA-DNA synaptic and D-loop structures

Haijuan Yang¹, Chun Zhou^{1,3}, Ankita Dhar¹, Nikola P. Pavletich^{1,2,*}

¹Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA.

²Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA.

³Current address: Zhejiang University School of Medicine, Zhejiang Sheng, China, 310000

Abstract

The strand exchange reaction is central to homologous recombination. It is catalyzed by the RecA family of ATPases that form a helical filament with single-stranded DNA (ssDNA) and ATP. This filament binds to a donor double-stranded DNA (dsDNA) to form synaptic filaments that search for homology, and then catalyze the exchange of the complementary strand to form a new heteroduplex, or a D-loop if homology is limited^{1,2}. Here we report the Cryo-EM analysis of synaptic mini filaments with both non-complementary and partially-complementary dsDNA, and structures of RecA–D-loop complexes containing a 10 or 12 base pair heteroduplex at 2.8 and 2.9 Å, respectively. The RecA C-terminal domain (CTD) binds to dsDNA and directs it to the L2 loop, which inserts into and opens the duplex. The opening propagates through RecA sequestering the homologous strand at a secondary DNA-binding site, freeing the complementary strand to sample pairing with the ssDNA. Duplex opening has a significant probability of stopping at each RecA step, with the as yet unopened dsDNA portion binding to another CTD. Homology suppresses this process through heteroduplex pairing cooperating with secondary site-ssDNA binding to extend dsDNA opening. This mechanism locally limits the length of ssDNA sampled for pairing if homology is not encountered, and it may provide for the formation of multiple synapses separated substantially on the donor dsDNA, increasing the probability of encountering homology.

Homologous recombination plays essential roles in the generation of genetic diversity and the repair of DNA double strand breaks and related lesions. In the strand exchange reaction, the RecA ATPase is activated by binding to ATP and ssDNA, forming a helical assembly, termed presynaptic filament, of 3 nucleotides (nts) per RecA^{1–3}. The presynaptic filament then binds to donor dsDNA, forming synaptic filaments that sample for ssDNA-

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

* *Correspondence and material requests:* pavletich@mskcc.org .

Author Contributions H.Y. designed and carried out the DNA binding experiments, prepared RecA mini filaments and strand-exchange reactions and collected, processed and analyzed the cryo-EM data, C.Z. carried out the initial cryo-EM characterization of the RecA mini filaments, A.D. carried out the DNA binding experiments, prepared RecA proteins and strand-exchange reactions and collected cryo-EM data, and N.P.P. analyzed the data and wrote the manuscript.

Author Information The authors declare no competing financial interests.

dsDNA homology in a poorly understood process. Homology gives rise to a postsynaptic filament where the complementary strand of the donor is paired with the primary ssDNA in a heteroduplex (Extended Data Fig. 1a). A secondary DNA-binding site, defined by mutagenesis, is thought to bind to both the donor dsDNA and its displaced homologous strand⁴⁻⁷.

RecA consists of a flexibly-linked 30-residue N-terminal helix (α N) that anchors adjacent protomers⁸, a 64-residue C-terminal domain (CTD), and a 240-residue helicase/ATPase core that binds to the primary ssDNA and heteroduplex³ (Extended Data Fig. 1b). A previous crystal structure of the presynaptic filament³, obtained with a mini filament of five fused RecAs end-mutated to prevent further polymerization, showed that while the ssDNA is overall underwound and stretched, locally, over 3 nts (triplets), it has a B-type DNA conformation. Together with the structure of the mini filament containing a heteroduplex but lacking the displaced strand, this indicated that RecA holds the ssDNA substrate in a conformation that can locally sample Watson-Crick pairing with the donor dsDNA³.

What has not been understood is the formation of synaptic filaments, and the corollary questions of how RecA searches for homology and how it destabilizes the double-stranded nature of the donor dsDNA to free the complementary strand for homology sampling^{1,9}. To address these questions, we constructed a fusion protein of nine *Escherichia coli* RecA protomers corresponding to one and a half turns of the filament.

Non-homologous synaptic filaments

We first collected cryo-EM data of the nine-RecA mini filament assembled with 27 nt oligo(dT) ssDNA and non-hydrolyzable ATP γ S in the presence of 14-fold molar excess of a 67 bp non-homologous dsDNA, the length and concentration of which were optimized using an in vitro DNA-binding assay (Extended Data Fig. 1c-f). The consensus reconstruction of 972,675 particles extended to 2.8 Å by the gold-standard fourier shell correlation (FSC) procedure (Extended Data Fig. 2a-d, Table 1). The map showed that the primary ssDNA binds in the filament center as with the presynaptic crystal structure³ (Fig. 1a), but it also revealed two features not present in the crystal structure: (i) patchy density at and around secondary-site residues that was consistent with ssDNA and which extended along the length of the filament (henceforth S2 site), and (ii) trace amounts, at a level ~10 % of the protein density level, of double helical density at each RecA that was consistent with dsDNA (Fig. 1b). The dsDNA extended from the RecA CTD at the filament periphery towards the center, with its axis bisecting the primary ssDNA axis. The CTD's involvement in dsDNA binding was predicted by an NMR chemical shift perturbation analysis of the isolated CTD polypeptide^{10,11}.

We next used iterative 3D classification with partial signal subtraction to identify the fraction of particles that contained dsDNA at each RecA position (RecA protomers labeled with superscript letters, starting with A at the 3' end of the primary ssDNA). Because the filament ends are overall poorly ordered, and the CTDs at the 3' end extend the farthest away into the solvent, we could not classify particles at the RecA^A position, and classification at RecA^B was incomplete. At the remaining RecA positions, the fraction of particles that

contained duplexes ranged from 9 % to 18 % (thereafter duplex occupancy; Extended Data Fig. 2e–f). The highest duplex occupancies of 18 % and 15 % were at RecA^C and RecA^H, respectively, whereas the lowest ones were 9 % at RecA^F, RecA^G and RecA^I. The lower duplex occupancy in the mid-portion of the filament likely results from the limited dsDNA accessibility of the CTDs there (Extended Data Fig. 2g).

Analysis of the pattern of duplexes per particle showed that there was on average 0.9 duplexes per particle, ranging from no duplexes (41 % of particles) to six duplexes (0.01 %; Extended Data Fig. 2f, Supplementary Spreadsheet 1). Initial 3D classification with partial signal subtraction of select multi-duplex combinations showed that they contained duplex pairs that appeared to be connected through S2 density, or combinations of a pair(s) with single duplexes flanking the S2 density-connected pair(s) (Extended Data Fig. 3a–d).

We thus grouped particles into 28 sets representing all combinations of duplex pairs for positions B to I, and performed partial signal subtraction and 3D classification with masks covering the respective duplex pairs and the intervening S2 density region (Supplementary Fig. 2). Analysis of their reconstructions showed that duplexes bind to RecA in two conformations, with a ~28° angle between their dsDNA axes. One conformation has the dsDNA tilted substantially relative to the filament axis, with the filament-proximal end pointing 5', while the other is tilted slightly in the opposite direction (henceforth 5'- and 3'-tilt; Extended Data Fig. 2d).

The majority of the duplexes at the 3' end of the mini filament tended to have the 5' tilt, but this preference was scrambled towards the 5' filament end (Fig. 1c, Supplementary Spreadsheet 2). The 5' end also had classes with conspicuously short duplexes that barely reached their respective CTD, in addition to classes with normal duplexes (Extended Data Fig. 4a). Such short duplexes, which presumably represent a dsDNA end, were not observed at the 3' end.

For each pair of duplexes, the most abundant class tended to have both strong S2 density in between duplexes and at the S2-duplex connections, and also the 5' tilt for the 3' duplex and 3' tilt for the 5' duplex. These tilts direct the two duplexes towards the S2 density that links them (Fig. 1e–j; Supplementary Fig. 2, Spreadsheet 2, and Supplementary Discussion).

For any given 3' duplex, the number of particles in S2-connected classes decreased with increasing RecA-RecA spacing to its 5' mate (Fig. 1d; Extended Data Fig. 4b). Because of the aforementioned caveats, including low overall duplex occupancy at the filament center, the trend is not regular enough for accurate simulation. Nevertheless, it suggests that there is on average a ~20 % decrease in the number of S2-connected pairs at each RecA step.

Taken together, these observations suggest that in our mini filament system the initial dsDNA-CTD binding occurs preferentially at the 3' end of the filament in the 5'-tilt conformation. The initial binding opens up the DNA, and the local opening of the duplex propagates preferentially in the 5' direction, likely driven by the S2 site binding to the homologous strand (Supplementary Discussion). As the opening propagates, there is a significant probability of the as yet unopened part of the dsDNA binding to another CTD. This precludes further S2 engagement, thus terminating strand-separation to produce S2-

connected duplex pairs. This model accounts for short duplexes being limited to the 5' end of the mini filament, where duplex opening could get close to the dsDNA end. It also accounts for the scrambling of the tilt preference at the 5' end, where some of duplexes would result from a terminated opening and have a 3' tilt, while others would arise from initial dsDNA binding and have a 5' tilt.

The decrease in the probability of S2-connected duplex pairs with increasing separation raises the possibility that RecA-dsDNA synapses in the absence of homology are limited in length locally¹²⁻¹⁴. To confirm that this decrease is not due to the limited length of the dsDNA used, we collected cryo-EM data (1,592,037 particles) from a strand exchange reaction that had a longer, 120 bp donor dsDNA. 3D classification analyses recapitulated the aforementioned observations from the 67 bp dsDNA reaction (Fig. 1k), except for a higher overall duplex occupancy as expected from the longer 120 bp dsDNA having nearly twice the number of RecA binding sites (Extended Data Figs 5a-c). Most importantly, the data also exhibited a marked decrease in the frequency of S2-paired duplexes with increasing RecA-RecA distance (Fig. 11, Extended Data Fig. 4c, Supplementary Fig. 3). This indicates the roughly 20 % probability of strand-separation terminating at each RecA step is an intrinsic feature of the synaptic mini filament under our experimental conditions.

The reconstructions of S2-connected paired-duplex particles had no density for the second strand of the strand-separated portion of the donor dsDNA (Figs 1e-j). This suggested that homology sampling occurs in a D-loop like DNA structure, where the complementary strand is mobile and thus available for base pairing with the primary ssDNA.

Partially homologous synaptic filaments

To address homology sampling, we collected cryo-EM data with a mini filament where the primary ssDNA contained 10 nts of homology to the middle of the 67 bp dsDNA. Based on the postsynaptic crystal structure³, the heteroduplex at the primary site would extend from RecA^B to RecA^E. Indeed, the 2.4 Å consensus reconstruction of 1,697,726 particles revealed patches of low-level density for the complementary strand pairing with the ssDNA at the expected positions (Fig. 2a, Extended Data Fig. 6a-c). The homologous data set exhibited a similar preference for 5' tilts at the 3' end of the filament as the non-homologous reaction, but had a higher average duplex occupancy (Fig. 2b, Extended Data 6d). Crucially, the decrease in the frequency of S2-paired duplexes with increasing RecA-RecA distance was substantially diminished, except at the RecA protomers outside the region of homology (Fig. 2c, Extended Data Fig. 6e). This suggests that heteroduplex formation stabilizes the strand-separated state of the donor dsDNA, likely through the pairing energy and also by keeping the complementary strand away from the homologous strand being sequestered by the S2 site.

3D classification of paired duplexes revealed patchy density for the complementary strand of a heteroduplex between many S2-connected pairs that encompass duplex^D and duplex^G, which spatially flank the homology region on the ssDNA, and even some pairs that encompass only one of the two positions (Fig. 2d to j, Supplementary Fig. 4). These classes were predominantly in the 5'/3'-tilted conformation. Because of the particle heterogeneity

apparent in the patchiness of heteroduplex density, we 3D sub-classified, with partial signal subtraction outside the complementary strand, the 19,730 particles in the S2-connected duplex^D-duplex^G class. This identified a class of 7,990 particles that appeared to have a bona fide D-loop^{DG}, with a 10 bp heteroduplex having overall continuous complementary-strand density that was connected to the duplexes (Fig. 2k). A class of 2,684 particles lacked complementary-strand density, presumably because the duplex opened up where there was no homology. The remaining particles appeared heterogeneous, with weak, patchy complementary strand density that could not be further evaluated by sub-classification (Extended Data Fig. 6f–h).

Postsynaptic filaments with D-loop DNA

To investigate the structural details of dsDNA opening and D-loop formation, we designed, based on the D-loop^{DG} density, a dsDNA substrate with an 11-bp region of homology that contained a bubble of 9 mismatched base pairs (Fig. 3a). We reasoned that the unpaired complementary strand would readily pair with the primary ssDNA and guide the flanking duplexes to the proper CTDs, allowing the formation of a homogeneous D-loop. Indeed, the 2.7 Å map from 399,184 particles showed that CTD^D and CTD^G contained a duplex at essentially full occupancy (Extended Data Fig. 7c).

The 2.8 Å refined model contains a 15-bp duplex^D, a 17-bp duplex^G, 10 nts of complementary strand in a heteroduplex, and 10 nts of homologous strand bound by RecA (Fig. 3a, Extended Data Table 1). As expected from the postsynaptic crystal structure³, the heteroduplex is arranged in 3¹/₃ base-pair triplets, each bound by the helicase core L1 and L2 loops and the α helices that follow them.

Duplex^D adopts the 5' tilt and duplex^G the 3' tilt. The 3'-tilt conformation is associated a 13° rotation of CTD^G relative to the helicase core, while the 5'-tilt has a smaller CTD^D rotation of -3° in the opposite direction (Extended Data Fig. 8a). When superimposed on their respective CTDs, the axes of the two duplexes differ by an additional 16° angle (Extended Data Fig. 8b).

Each duplex abuts an L2 loop, whose Phe206 side chain stacks with the aromatic face of the last base pair and terminates the double helix (Fig. 3b–c; Extended Data Fig. 8c–d). However, the different duplex orientations are associated with non-equivalent L2 loops, which in turn have distinct backbone conformations and use a different set of additional residues to extend their contacts to the duplex end (Fig. 3d; full contacts discussed in Extended Data Fig. 8 legend).

The CTD-dsDNA interactions are very analogous in the two duplexes. They extend over 6 base pairs and involve a loop-helix motif that is structurally coupled to an adjacent hairpin motif. These insert into the minor groove of the DNA and expand it (Fig. 3e–f, Extended Data Fig. 8e–f and legend).

The path the homologous strand tracks, and thus our definition of the S2 site, consists of the long β6 strand (residues 226 to 232) of the helicase core on one side, and the L2 loop on the other (Fig. 3g; Extended Data Fig. 8g). The S2 site–DNA contacts repeat every 5 nts, in

contrast to the primary site–DNA contacts that repeat every 3 nts or heteroduplex base pairs. The two sites also differ in their extents because the 5' end of the homologous S2 ssDNA is below (in the 5' filament direction) L2^E, while its 3' end is above L2^C (Fig. 3g). Thus, while the 10 bps of heteroduplex extend along a filament segment equivalent to 3¹/₃ RecA repeats, the 10 nts of homologous strand extend along a segment equivalent to two RecA repeats.

The homologous-strand conformation is characterized by extensive stacking interactions by nucleotide bases as well as sugar groups, both within the DNA and between the DNA and RecA. RecA hydrogen bonds to the phosphodiester backbone are sporadic, and only a subset is repetitive (Fig. 3g).

Immediately after the opening of duplex^D, the Ade28 base group is sandwiched between the last base pair of the duplex and Met202 of L2^C, while its phosphodiester backbone is contacted by Arg226 of β6^D (Fig. 3g, bottom). Arg226 also makes bidentate hydrogen bonds to L2^D Glu207, an interaction conserved in all RecA protomers. This interaction would stabilize the L2 loop conformation, suggesting that the E207Q mutation that eliminates secondary-site binding^{15,16} acts indirectly.

Cyt27, the next nucleotide in the 5' filament direction, is sandwiched between L2^C and L2^D residues and the E207^D-R226^D salt bridge, while the following Cyt26 and Cyt25, which stack together, pack against L2^D and hydrogen bond to β6^E backbone and side chain groups (Fig. 3g; Extended Data Fig. 8g and legend). Cyt24 and Cyt23 also stack together. They bind into a tight gap between β6^E and L2^D, which contact both the phosphodiester and base groups of the pair in a pinch-like arrangement (Fig. 3g; Extended Data Fig. 8g and legend). The extensive contacts suggest that this is a key aspect of RecA's secondary site DNA-binding activity. Thereafter, Cyt22 is bound similarly to Cyt27, five nucleotides over (Fig. 3g).

Cyt21 to Cyt19, the last three nucleotides before duplex^G, are too far to contact β6^E as they stack along an alternate curved path around L2^E on their way to the duplex. They have relatively high temperature factors and have been built based on unsharpened maps (Fig. 3g; Extended Data Fig. 8g).

To investigate how general these aspects of this D-loop are, we also determined the structure of 9-RecA bound to a DNA containing a 12-bp bubble with 12 bps of homology positioned to direct the duplexes to CTD^D and CTD^H (Fig. 4a; Extended Data Fig. 7d). The 2.9 Å-refined structure showed that in addition to the 12 bp bubble, 2 base pairs, one from each duplex, are opened up. The structure thus has 14 homologous-strand nucleotides extending along three S2 repeats and 12 bps of heteroduplex in four triplets, flanked by a flipped-out nucleotide on each side (Fig. 4a, Extended Data Fig. 8h, Extended Data Table 1). This D-loop^{DH} structure recapitulates the key aspects of D-loop^{DG}, including the overall S2 site contacts, and the L2^C and L2^F loop conformations and their stacking with the end of their respective duplex (Fig. 4a–b; Extended Data Fig. 8h and legend). Most of these findings likely also apply to the eukaryotic RAD51 homolog, which has a putative N-terminal DNA-binding domain located at the filament periphery as with the RecA CTD (Extended Data Fig. 9).

These structures indicate that the L2 loop plays a major role in the initial opening of the dsDNA. Loops or hairpins with aromatic or hydrophobic amino acids are common protein motifs that insert into dsDNA to separate its strands locally^{17,18}. To extend the opening, RecA additionally uses the adjacent S2 site to bind to the homologous strand, in effect stabilizing in the single-stranded state of the nucleotides that follow the L2 loop insertion. This suggests that the first duplex to bind preferentially adopts the 5'-tilt conformation as the first nucleotide after the L2 loop insertion immediately binds to the S2 site. By contrast, the 3'-tilt conformation has a 3-nt spacer from the duplex end to the start of significant S2 contacts, and it likely represents the second duplex binding that terminates the opening.

Conclusion

Our data reveals that the donor dsDNA binds to the CTD domain, which directs the dsDNA to clash with the L2 loop. The L2 loop then inserts into the dsDNA and initiates the opening of the duplex, with the S2 site propagating the opening preferentially in the 3' to 5' direction. Each S2 site binds to and opens up a 5-nt strand of the donor dsDNA compared to the primary site binding to 3 nts per RecA, a ratio beneficial for finding homology. Duplex opening has a significant probability of stopping at each RecA step. Homology suppresses this, allowing the opening of the dsDNA to extend across the homologous region. This is likely a mechanism that locally limits the length of ssDNA sampled for pairing if homology is not encountered, thus allowing the formation multiple synapses, separated substantially on the donor dsDNA^{13,19,20}, that can increase the probability of the filament encountering the correct register of the two sequences.

METHODS

Protein engineering, expression and purification.

To make the nine-RecA fusion protein, three more copies of *Escherichia coli recA* gene were inserted into the previously described overexpression vector containing six copies of the *E. coli recA* gene with the same linker length, N-terminal and C-terminal mutations that disrupt oligomerization³. In addition to the original N-terminal 12 histidine residues tag, a C-terminal FLAG tag was inserted after the last *recA* gene to facilitate the purification of the full-length nine-RecA fusion protein. The nine RecA fusion protein was overexpressed and purified as described³ except that after eluting the protein off the Ni²⁺ resin (GE Healthcare), the full-length protein was further enriched with pull down with ANTI-FLAG® M2 Affinity gel (Sigma) before fractionated with anion-exchange chromatography (Mono Q, GE Healthcare). Peak fractions were pooled and aliquoted for storage in 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10 mM dithiothreitol (DTT) and 10% glycerol at -80 °C.

DNA substrates.

DNA oligonucleotides used in the cryo-EM experiments were purchased from IDT as standard PAGE gel purified oligonucleotides with the exception of the 120-nt oligonucleotides that were ordered as Ultramer oligonucleotides due to their lengths. dsDNA molecules were all blunt ended and were further purified with PAGE gel after annealing to ensure no ssDNA was present in the dsDNA samples. For the electrophoretic mobility

shift assays, fluorophores were attached to the 5' end of the oligonucleotide. HPLC-purified oligonucleotides labeled with Alexa 488 or Alexa 647 were purchased from IDT, and the 120-nt oligonucleotide labeled with 6-FAM from Sigma.

Electrophoretic mobility shift assay of dsDNA binding.

To visualize RecA binding to ssDNA and dsDNA simultaneously, we used fluorophores of different excitation wavelength for ssDNA (Alexa 647, excitation wavelength of 650 nm) and for dsDNA (Alexa 488, excitation wavelength of 492 nm; 6-FAM, excitation wavelength 495 nm). First, presynaptic filaments were prepared by incubating the nine-RecA fusion protein with 27-nt ssDNA at an equimolar concentration (1.6 μM) on ice for 15 minutes in 20 mM Tris, pH 8.0, 67 mM NaCl, 3 mM $\text{Mg}(\text{OAc})_2$, 0.3mM $\text{ATP}\gamma\text{S}$, 2 mM DTT and 5% (v/v) glycerol. Then, dsDNA (ranging from 18 to 120 bps) at indicated molar ratios were added to the presynaptic filament and incubated for 10 additional minutes. Free DNA was separated from the RecA-bound DNA on 0.8% Agarose gels in 0.5x TB buffer supplemented with 3 mM $\text{Mg}(\text{OAc})_2$. The gels were imaged for fluorescence detection of Alexa 647-labeled ssDNA at 635 nm, of Alexa 488 or 6-FAM-labeled dsDNA at 473 nm with Typhoon FLA 9000 (GE healthcare). The overlay was created using ImageJ software (Fujifilm). The source gels are shown in Supplementary Fig. 1.

Cryo-EM sample preparation and data collection.

On the day of making grids for the cryo-EM experiments, the nine-RecA fusion protein from the Mono Q peak was concentrated to $\sim 3.9 \text{ mg ml}^{-1}$ and purified by gel filtration chromatography (Superose 6 Increase 5/150 GL, GE Healthcare) equilibrated in 20 mM Tris-HCl, pH 8.0, 200 mM NaCl and 2 mM DTT. Only the peak fraction was used for making grids. Briefly, 1.6 μM nine-RecA fusion protein was incubated with 1:1 molar ratio of ssDNA (27 nucleotides long, 1.6 μM ssDNA molecules) in 20 mM Tris-HCl, pH 8.0, 3 mM $\text{Mg}(\text{OAc})_2$, 0.3–1 mM $\text{ATP}\gamma\text{S}$ at 100–150 mM NaCl for 15 minutes at room temperature to allow for presynaptic filament formation. Then the salt was reduced to 87–100 mM to facilitate dsDNA binding and the desired dsDNA was added at 1.2 or 14 molar ratios to the presynaptic filament to start the pairing reaction. At desired time point (20 seconds to 30 minutes), the sample (3–4 μl) was applied to glow discharged UltrAuFoil 300 mesh R1.2/1.3 grids (Quantifoil). Grids were immediately blotted for 1–1.5 seconds at room temperature and 100% humidity and plunge-frozen in liquid ethane using a FEI Vitrobot Mark IV. All data were collected with Titan Krios microscopes operated at 300 kV. The dataset with the (dT)₂₇ and 14-fold molar excess of non-homologous 67 bp dsDNA were collected from samples frozen at 30 seconds to 5 minutes after adding dsDNA and were acquired over 6 sessions at the MSKCC Cryo-EM facility Krios with a Gatan K2 Summit camera with a 1.09 Å pixel size and dose rate of 10.0 electrons per pixel per second. Each 8 second exposure was dose-fractionated into 40 frames and contained a total dose of 67 electrons per Å². Comparison of the individual data sets from the time course of 30 seconds to 5 minutes did not show any meaningful differences in duplex occupancy or duplex patterns (Supplementary Spreadsheet 1). The dataset with the 27 mer ssDNA (5'-(dT)₁₄CGCTCGCCCA(dT)₃-3') bearing 10 bp homology to the 14-fold molar excess 67 bp dsDNA were collected from samples frozen at 3 and 7 minutes at the a Krios microscope of the HHMI cryo-EM facility. Movies were recorded with a Gatan K3 camera with a

1.078 Å pixel size and 20.0 electrons per pixel per second. Each 3 second movie was dose-fractionated into 40 frames and contained a total dose of ~52 electrons per Å². The datasets with the (dT)₂₇ and 14-fold molar excess non-homologous 120 bp dsDNA sample were collected from samples frozen at 3 and 5 minutes with a Krios microscope of the HHMI cryo-EM facility, using a Gatan K3 camera operating in CDS mode with a 1.078 Å pixel size and 8.0 electrons per pixel per second. Each ~7.8 second movie was dose-fractionated into 50 frames and contained a total dose of 54 electrons per Å². The datasets with the 9 and 12nt D-loops were collected from samples equilibrated for 30 minutes after the addition of 1.2 molar-excess of bubble dsDNA. They were collected at an NYSBC Krios microscope with a Gatan K2 camera with 1.096 Å pixel size and 8.5 and 10 electrons per pixel per second. Each 8 second movie was dose-fractionated into 40 frames and contained a total dose of 57 and 67 electrons per Å², respectively.

Cryo-EM image processing.

The super-resolution movies were initially aligned with MOTIONCOR2²¹, and the contrast transfer function (CTF) parameters were estimated with CTFFIND4²². All 2D/3D classifications, partial signal subtraction, 3D refinements, local resolution estimation and other image processing were carried out with RELION-3²³. Bayesian beam induced motion correction, scale and B-factors for radiation-damage weighting, and per particle refinement of CTF parameters were also applied with RELION-3²⁴. Particle polishing followed by CTF refinement were iterated twice, as further iterations did not yield improvements in resolution. All reported map resolutions are from gold-standard refinement procedure with the FSC=0.143 criterion after post-processing by applying a soft mask.

To identify particles containing duplexes at each CTD position by 3D classification, we used a mask around the dsDNA density to avoid confounding factors such as dsDNA on adjacent RecAs, or filament's conformational flexibility influencing the classification. Because the density level inside the mask was very low, we subtracted the signal of the consensus map outside the dsDNA mask from each particle, and we performed 3D classification of the partially signal-subtracted particles without alignment. For each CTD position, we calculated a soft mask from a 15 bp dsDNA model manually positioned into density from the consensus reconstruction that was low-pass filtered at ~6 Å to improve the dsDNA contours. For CTDs in the middle of the filament with lower duplex occupancy and poorly-defined duplex density, the duplex was shifted from a better-defined CTD using translation-rotation operations derived from a 9-RecA model constructed based on the crystal structure of the 5-RecA-ssDNA mini filament³. Following the initial round of 3D classification into two classes with a broad duplex mask, further 3D classification of density-restored particles from the duplex-positive class at each RecA position indicated that the duplexes bind to RecA at two different conformations. Based on this, 3D classification with partial signal subtraction without alignment was iterated with improved masks made from coordinates that contained both duplex conformations at each position. The weight of the experimental data was reduced (RELION-3 tau_fudge_factor) to facilitate convergence and reduce the influence of density features not directly related to the presence of a duplex. The tau_fudge_factor was adjusted depending on the number of particles in the data set, by aiming for ~8 Å resolution in 3D classification (typically a value of 1 to 2 for ~1.5

million particles). We avoided filling the volume outside the solvent mask with zero density (RELION-3 zero_mask option), as this resulted in a higher fraction of particles falsely assigned to the duplex containing class. We could not reliably identify duplex containing classes at RecA^A, because not only this terminal RecA is poorly ordered, but also its CTD is at its least ordered part at the extreme tip of the filament. RecA^B had similar drawbacks. Even though 3D classification appeared to identify a duplex-containing class, it had far fewer particles compared to the adjacent well-ordered RecA^C, and on further classification or 3D refinement more than half of these particles ended-up in classes with poorly-defined double helical density for the duplex. RecA^I, at the other end of the filament, is also less ordered, but its CTD is among its better-ordered parts as it points towards the mid-portion of the filament. At the end of the +/- duplex classification process, each particle was assigned a binary code of 8 digits representing duplex occupancy at RecAs B to I (as a “_rlnComment” entry in the Relion particle star file) to facilitate manipulations such as selecting particles with a particular pattern of duplexes (for example, particles that contain duplex^C and duplex^G but no intervening duplexes could be selected by matching the pattern “x10001xx”, where “x” could be either 1 or 0).

For the representative 3D classification of particles with more than 2 duplexes of Extended Data Fig. 3, we used the 120 bp dsDNA data set, because such combinations are infrequent and this data set had the highest duplex occupancy. The particles were partially signal subtracted and 3D classified without alignment using masks that contained not only the respective duplexes but also the S2 density (its mask made from a pdb model of ssDNA) in between the most distal duplexes. Because of the low particle numbers and to make the classification more responsive to density features such as S2 DNA, the weight of the experimental data was increased through the RELION-3 tau_fudge_factor, aiming for a resolution of better than 8 to 10 Å. The individual classes were then 3D refined, and low-pass filtered to ~6–8 Å to facilitate comparison of classes with widely different resolutions, as the S2 density tends to be broken up at the higher resolutions that result from the RecA protein.

For the 3D classification of paired-duplex particles (Supplementary Fig. 2–4) the signal subtraction and solvent masks contained, in addition to the duplexes and S2 density, portions of the β6 strand and L2 loops of RecA protomers at and in between the duplexes. The weight of the data was increased to achieve a resolution of 4–6 Å during classification (tau_fudge_factor values of 8, 16, 32 and 64 for >120,000, ~90,000, ~45,000 and ~20,000 particles, respectively), and the calculations were run until convergence (changes in particles per class < 0.01 %), typically for 100–400 iterations. All classes that had at least 10 % of the particles in each run were 3D refined, and the overall resolution limits shown on the figures are by the gold-standard fourier shell correlation (FSC) procedure. For determining whether 2 duplexes were connected to continuous intervening S2 density, the maps of the different classes were also low-pass filtered to a common low resolution (5.6 Å to 8 Å) and inspected at a uniform density level to reduce the need for judgment calls.

For the 3D sub-classification of the heteroduplex-containing duplex^D-duplex^G particles in the partially-homologous data set, the signal subtraction and solvent masks contained only the complementary strand, and they were constructed based on the structure of the

heteroduplex model. Because the density within the mask was a very small fraction of total density, classification required the weight of the data to be increased substantially (tau_fudge_factor of 256), and also the zeroing of the density outside the solvent mask, otherwise, all the particles ended up in one class. For the final reconstructions of the non-homologous 120 bp dsDNA and homologous 67-bp strand exchange reaction data, as well as the 9-RecA complexes with D-loop^{DG} and D-loop^{DH}, we performed two focused 3D refinements using soft masks covering the four terminal RecA protomers (RecA^{ABCD} and RecA^{FGHI}; Extended Data Fig. 7).

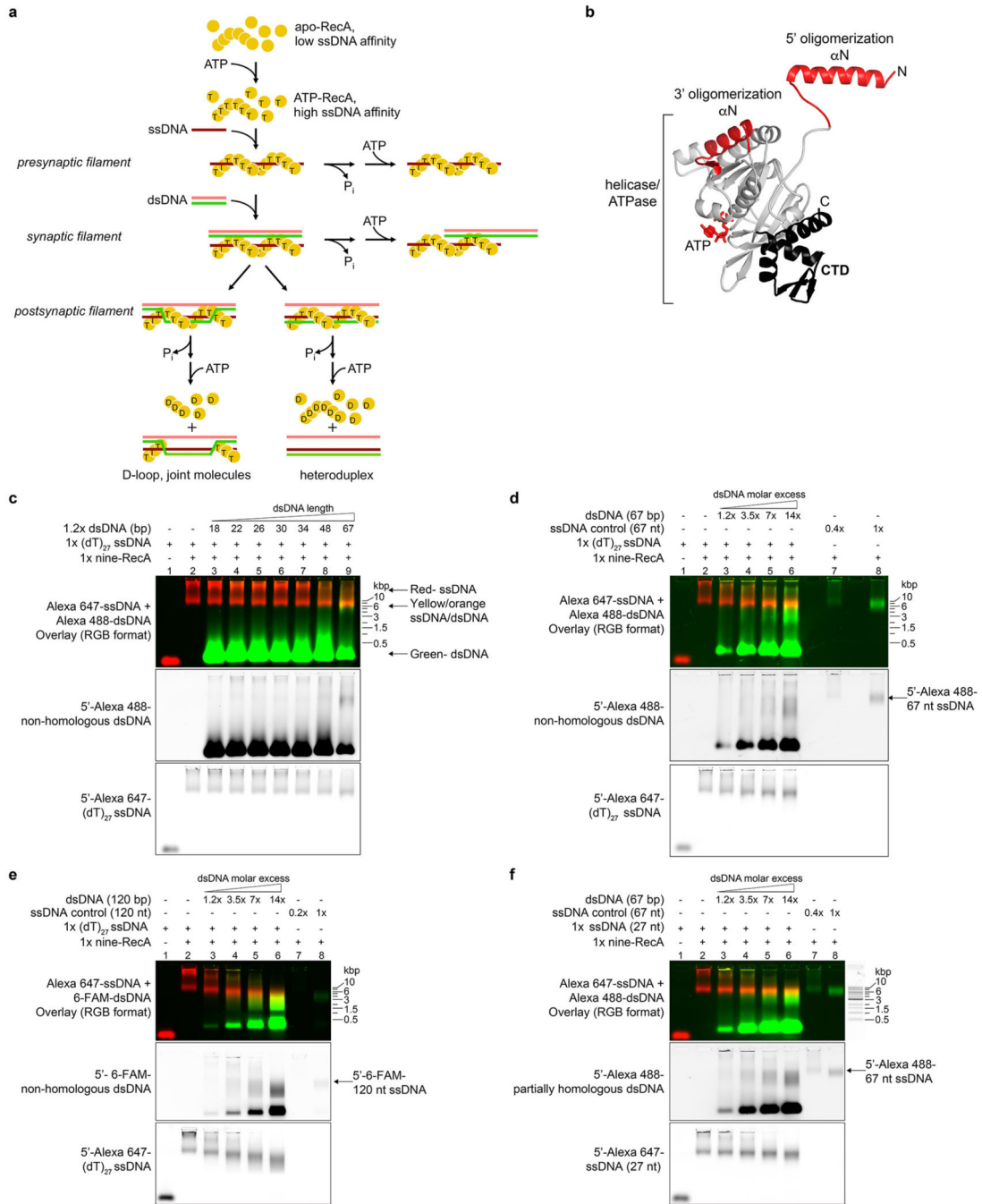
Cryo-EM structure refinement.

Model refinement was done with REFMAC5 modified for cryo-EM²⁵ and with PHENIX²⁶. For the 120 bp dsDNA and homologous 67 bp strand exchange reaction data, as well as the 9-RecA complexes with D-loop^{DG} and D-loop^{DH}, the focused maps were aligned on their respective consensus reconstruction using CCP4²⁷, and their RecA^{ABC} and RecA^{GHI} portions combined with the mid-portion (RecA^{DEF}) of the consensus map using the composite sfcalc option of REFMAC5²⁵. The single set of structure factors and corresponding maps were then used to refine the models first in real space with PHENIX²⁶, and then in reciprocal space with REFMAC5. The initial model was built based on the crystal structure of the 5-RecA-ssDNA-ADPAI(F)₄-Mg²⁺ complex and for corresponding heteroduplex complex³. The S2 ssDNA of the strand exchange reaction models is overall ill-defined at the resolution limit of the sharpened map, likely due to the heterogeneity of the bound DNA and also because the ends of the low-occupancy duplexes overlap with the S2 ssDNA. Their S2 ssDNA was placed into the unsharpened map from the high-resolution structure of the 9-RecA-D-loop^{DG} using the filament repeat. To validate the D-loop^{DG} and D-loop^{DH} models, the two half maps of the three reconstructions, post-processed with the --half_maps option of the relion_postprocess program, were combined as above into separate half1 and half2 composite maps. A partially refined model was then refined against the half1 structure factors until convergence with the same protocol as with the full-data composite map. For D-loop^{DG} half1 refinement the starting R_f , overall FSC, and highest-resolution shell FSC were 33.1 %, 0.814, and 0.557, respectively, with final values of 28.0 %, 0.869, and 0.612, and changes in these values between the 49th and 50th refinement cycles of 0.0 %, 0.000, and 0.001. For D-loop^{DH} half1 refinement the starting R_f , overall FSC, and highest-resolution shell FSC were 34.7 %, 0.788, and 0.087, respectively, with final values of 28.4 %, 0.859, and 0.579, and no changes in the last refinement cycle. The final models were then used to calculate the validation FSC curves against the half2 structure factors (Extended Data Fig. 7c-d).

Data availability.

The refined coordinates and corresponding cryo-EM maps, including the consensus and focused reconstructions, composite maps used in refinement and maps of Fig. 1e-j and 2d-k have been deposited with the Protein Data Bank and the Electron Microscopy Data Bank under accession codes PDB-7JY6 and EMDB-22522, PDB-7JY8 and EMDB-22524, PDB-7JY9 and EMDB-22525, PDB-7JY7 and EMDB-22523, for the strand exchange reactions containing non-homologous 120 bp dsDNA and homologous 67 bp dsDNA, and the D-loop^{DG} and D-loop^{DH} complexes, respectively.

Extended Data



Extended Data Figure 1 | RecA-catalyzed strand-exchange reaction.

a, Schematic of the strand exchange reaction. RecA is shown as yellow spheres, with the letters D and T respectively indicating ADP and ATP-bound forms of RecA. The ssDNA is indicated by dark brown lines, the donor dsDNA by double lines that are colored green for the complementary strand and red for the homologous strand. The role of ATP hydrolysis, which is activated on ssDNA binding, is incompletely understood. ATP hydrolysis reverts

the RecA-RecA relationship to a state that is inactive in ssDNA binding²⁸. However, the primary ssDNA likely does not diffuse far because (i) the RecA protomers, which can remain associated in a concentration-dependent manner through the α N helix of one RecA interacting with the helicase domain of the 5' RecA (oligomerization motifs shown in **b**), would be topologically wrapped around the ssDNA, and (ii) ATP exchanging for ADP due to the high ratio of ATP to ADP in the cell²⁹. In effect, ATP hydrolysis by the synaptic filament and subsequent exchange of ADP for ATP may serve to dissociate dsDNA while appearing to not affect ssDNA binding³⁰, even though the ADP state cannot bind to ssDNA. In the absence of strand exchange, this likely results in the donor dsDNA re-binding stochastically at a different register (shown hypothetically as a shifted dsDNA), continuing the search for homology. ATP hydrolysis by a fully exchanged postsynaptic filament results in the release of the new heteroduplex and the displaced homologous strand of the donor, while partial homology results in postsynaptic filaments that contain D-loops and other joint ssDNA-dsDNA molecules, with the ssDNA portions that have not exchanged reconstituted with RecA after ATP re-binding. ATP hydrolysis is not an inherent requirement for strand exchange (except for the release of products), as short dsDNA molecules can exchange in the presence of the non-hydrolysable ATP analogs³¹⁻³³. With longer, physiologically relevant substrates, however, ATP hydrolysis is needed for bypassing heterology and for the extension of initial joint molecules, or branch migration^{34,35}. This presumably involves the release of the portions of the donor duplex that have not exchanged, followed by their resampling in a new round of the reaction. With long DNA substrates, about ~100 ATP molecules are hydrolyzed per base pair (bp) exchanged *in vitro*³⁶. The direction of branch migration with ATP hydrolysis has been reported to be in the 5' to 3' direction with circular ssDNA³⁵. With linear ssDNA, however, the reverse polarity is suggested by the finding that ssDNA having 3'-end homology reacts more efficiently than that with 5'-end homology³⁷. This has been attributed to RecA polymerizing on ssDNA preferentially in the 5' to 3' direction³⁸. It is not clear to what extent the directionality of branch migration with ATP hydrolysis is related to the local opening of dsDNA without ATP hydrolysis, which this study finds occurs preferentially in the 3' to 5' direction of the mini filament. Since the mini-filament consists of fused RecA protomers, it does not reflect the effects a preferential polarity of RecA polymerization may have on the directionality of strand exchange. Also, our strand exchange reactions do not include the single-stranded DNA binding protein SSB that is involved in strand exchange *in vivo* and may sequester released DNA strands.

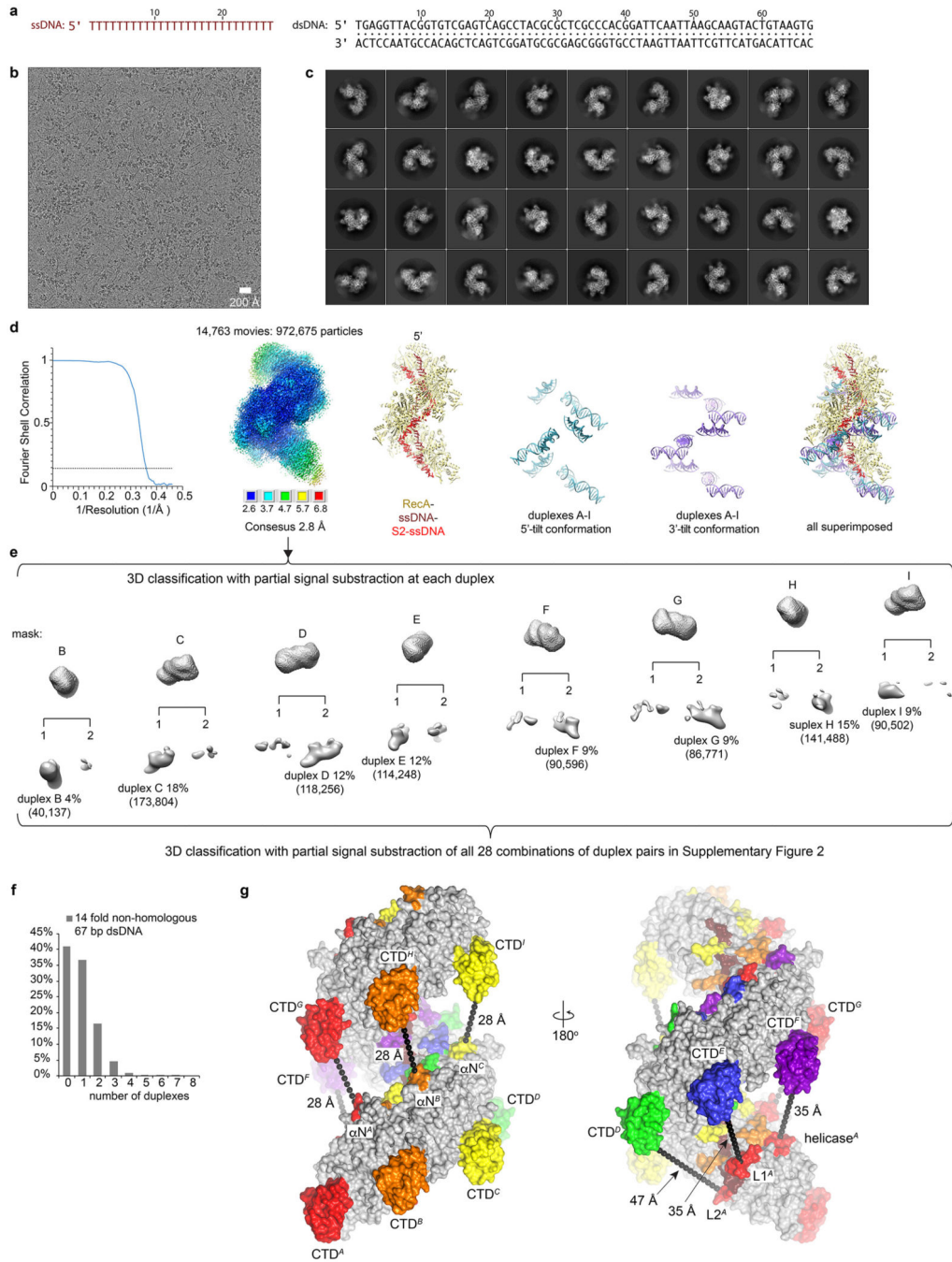
b, RecA monomer structure from the presynaptic mini filament²⁸. The α N oligomerization motif that interacts with the 5' RecA, and the site on the helicase domain that interacts with the α N of the 3' RecA are colored red. The CTD is black. ATP is shown in sticks. As reported²⁸, ssDNA binding cooperates with ATP binding to induce the conformational change from the inactive to the active filament states. The active filament conformation has a distinct RecA-RecA relationship that is stabilized by the ATP getting sandwiched between adjacent RecAs, and by two of the three nucleotides in each triplet binding to flanking RecAs. Even though the presynaptic filament binds to primary ssDNA with an overall stoichiometry of 3 nts per RecA, each nucleotide triplet is bound by three RecAs, and, conversely, each RecA contacts three nucleotides²⁸.

c, Electrophoretic mobility shift assay evaluating different lengths of non-homologous dsDNA binding to the presynaptic mini filament of nine-RecA-(dT)₂₇-ATP γ S. dsDNA, with

lengths ranging from 18 bps to 67 bps, was added at a 1.2 molar excess to the mini filament as described in Methods. Top, overlay of the gel scanned at the two wavelengths for the two different fluorophores. Signal from Alexa 647-ssDNA is shown in red and signal from Alexa 488-dsDNA is shown in green. Middle, signal from Alexa 488-dsDNA alone. Bottom, signal from Alexa647-ssDNA alone. While the presynaptic filament formed readily (lane 2), short dsDNA had no detectable signal under these conditions (lanes 3–7, 18–34 bp). A weak signal was detected at 48 bps of DNA (lane 8) and increased further at 67 bp (lane 9), the longest dsDNA we tested in this series.

d, Concentration titration of the non-homologous 67 bp dsDNA used in the cryo-EM analysis binding to the presynaptic mini filament. Top, overlay of the gel scanned at the two wavelengths (colored as in **c**). A clear trend of increased binding is evident as the dsDNA concentration increased from 1.2 molar excess to 14 molar excess (lane 3–6) to the presynaptic filament. To confirm that the green signal is from the binding of dsDNA and not a single strand that dissociated from the dsDNA, we also tested Alexa 488-labeled 67 nucleotide ssDNA at the same nucleotide concentration as the (dT)₂₇ (lane 7, 0.63 μM), or at the same molar-ratio to (dT)₂₇ (lane 8, 1.6 μM). DNA molecular weight markers are marked as in **c**.

e-f, Concentration titrations with 120 bp non-homologous dsDNA (**e**) and 67-bp partially homologous dsDNA (**f**) used in the cryo-EM analyses, performed as in **d**. Because we could not procure Alexa 488-labeled 120 nt DNA, we instead used the corresponding 6 FAM-labeled DNA (Sigma). The experiments of **c-f** were repeated three times with similar results (Supplementary Fig. 1). The DNA molecular weight markers are marked to the right of each top panel, in units of thousands of base pairs (Kbp).



Extended Data Figure 2 | Cryo-EM analysis of the strand exchange reaction with non-homologous 67 bp dsDNA.

a, Sequences of the 27 nt ssDNA (left, brown) and the 67 bp non-homologous dsDNA (right, black) used in the strand exchange reaction.

b, Micrograph from the reaction containing nine-RecA, (dT)₂₇, ATPγS, and non-homologous 67 bp dsDNA. The micrograph is similar to the rest of the 14,762 micrographs except for variations in particle numbers, ice thickness and other parameters across the grid.

c, Representative 2D classes of the particles after polishing. Box size is 279 Å. 2D classifications, performed two to three successive times prior to polishing resulted in similar classes, except for classes with low-quality 2D projections that were discarded.

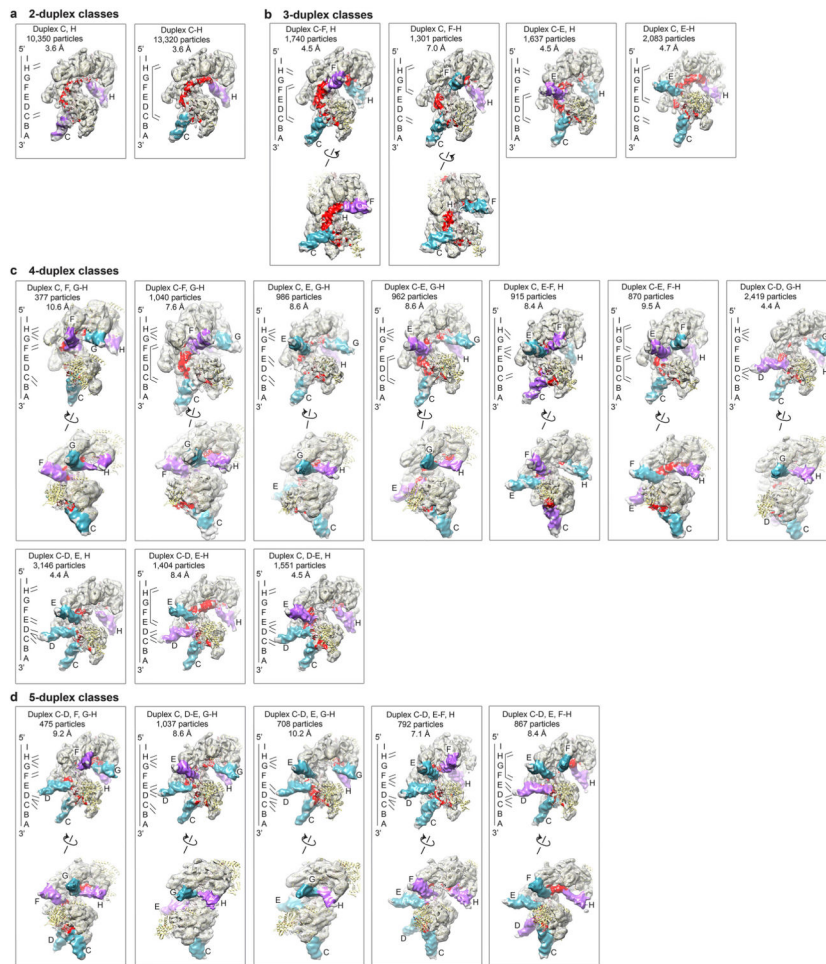
d, Chart shows the gold-standard FSC plot of the consensus reconstruction. Dashed line marks the FSC cutoff of 0.143. Second from the left is the consensus reconstruction map colored by local resolution estimated with the RELION3 post-processing program. The resolution range is mapped to the colors in the inset below the map; the terminal RecA proteins are less ordered than the rest. Third, cartoon representation of the refined model of the consensus refinement. As in Figure 1a, primary ssDNA is colored in brown, S2 ssDNA red. The nine-RecA protein is colored uniformly khaki for simplicity. Fourth and fifth, cartoon representations of duplexes A to I in the 5'- and 3'-tilt conformations, respectively colored cyan and purple, in the same relative orientation as the refined model. Lastly, duplexes with both tilts are superimposed on the protein to highlight the difference in the 5' and 3' tilts. The 5'- and 3'-tilted duplexes were combined to generate the masks for the 3D classifications as described in methods.

e, The masks used for 3D classification with partial signal subtraction at each duplex are at the top, and the maps of the 3D classes at the bottom. For each RecA position, the classes with duplexes are labeled with percentage and particle numbers (in parentheses). Because of the poor order of the terminal RecAs, and in particular at the 3' end of the filament where the CTDs extend the farthest away from the filament, we could not reliably classify particles at CTD^A, and for the same reason the penultimate CTD^B was an outlier with a low 4 % duplex occupancy (hereinafter we will be referring to individual RecA protomers with letters, starting with A from the 3' end of the primary ssDNA). At the 5' end, even though RecA^I was overall poorly ordered, duplex^I-containing particles could readily be identified, as CTD^I points towards the mid-portion of the filament, and its density is significantly better defined than CTD^A and CTD^B that point away from the filament's 3' end. Masks and maps of the 3D classification for all 28 combination of duplex pairs are shown in Supplementary Figure 2 and their details listed in Supplementary Spreadsheet 2.

f, Histogram of the number of duplexes per particles. Chart shows the percentage of particles that have the indicated number of duplexes for this dataset. The data set was collected once. Also see Supplementary Spreadsheet 1 for details.

g, The accessibility of the CTDs to dsDNA is highest at the 3' end of the filament, where the DNA-binding tips of CTD^A to CTD^C point into empty solvent and the dsDNA can approach from a roughly half-spherical volume (left panel). CTD^A is even more accessible as it has no neighboring RecA 3' to it. Moving towards the 5' end, the CTDs start getting increasingly encumbered by the presence of RecA protomers 3' to them. Thus, CTD^D gets slightly hindered by the RecA^A L2 loop that is 47 Å away (right panel; C_α-C_α distance from CTD Gly288 to L2 loop Gly200 in a direction that would approximately bisect the axis of dsDNA). CTD^E is more encumbered, as, in addition to the RecA^B L2 loop, is within 35 Å of the RecA^A L1 loop (right panel; C_α-C_α distance from CTD Gly288 to L1 loop Glu158). And, CTD^F is obstructed by not only the RecA^C L2 and RecA^B L1 loops, but also by the helicase domain of RecA^A, which is within 35 Å (right panel; C_α-C_α distance from CTD Gly288 to helicase Ala131). CTD^G and CTD^H are encumbered the most, by the full turn of filament 3' to them (left panel). Their DNA-binding tip is 28 Å away from the N-terminal helices of RecA^A and RecA^B, respectively (C_α-C_α distance from CTD Gly288 to Lys19 of

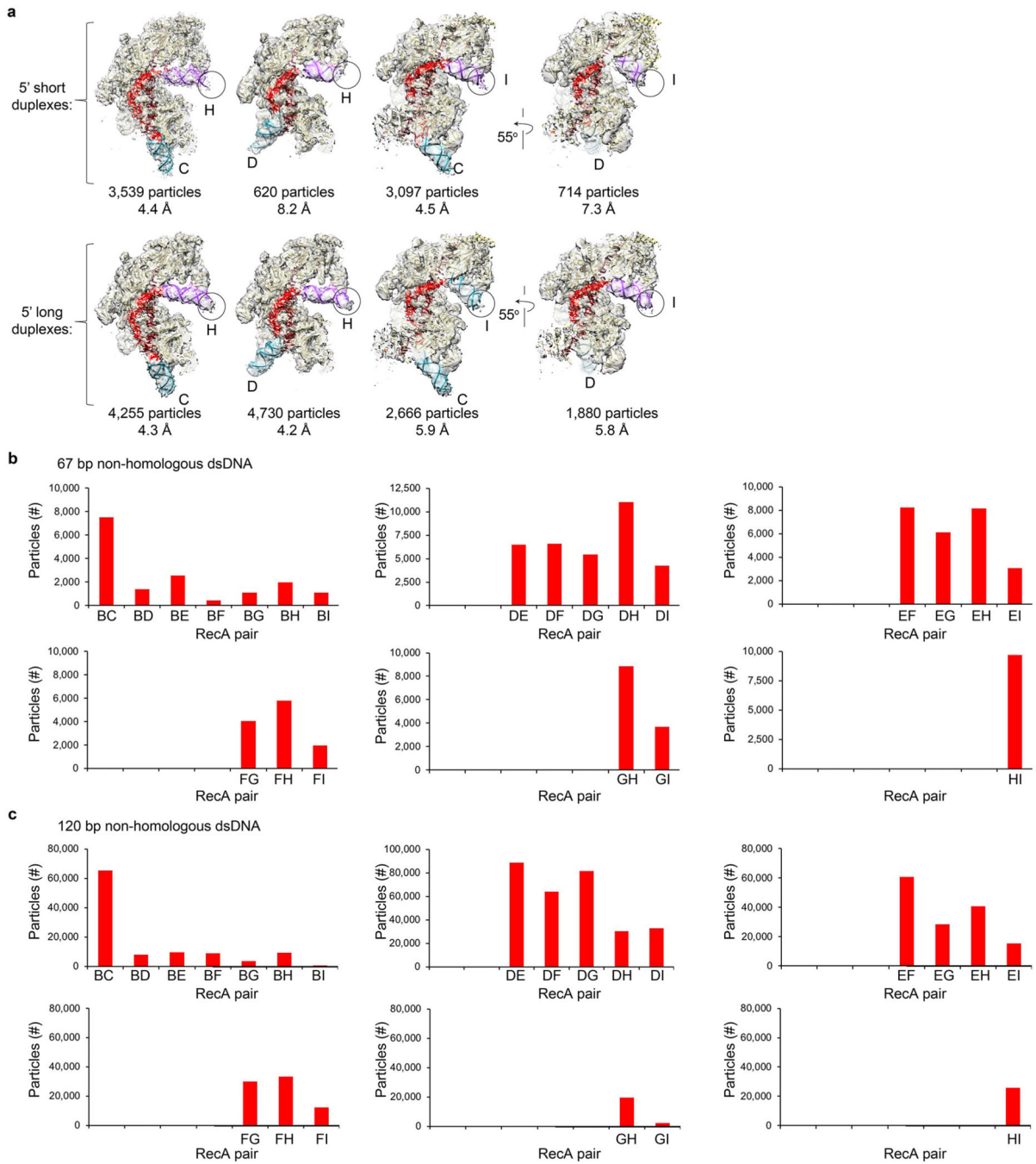
α N), a distance that is only fractionally larger than the ~ 20 Å width of a DNA duplex. The terminal CTD^I is similarly close to the N-terminal helix of RecA^C, although the absence of a 5-neighboring RecA would substantially increase its accessibility to dsDNA compared to those of CTD^G and CTG^H. Figure shows molecular surface of the 9-RecA filament with the aforementioned structural elements colored for each RecA as in Figure 1a and labeled. Black dotted lines indicate the shortest RecA-RecA distances (marked) that would approximately bisect the axis of dsDNA bound at each CTD. Primary ssDNA is colored brown. The homologous ssDNA is not shown for clarity. View in right panel is rotated by 180°, roughly half a turn of the filament, about the vertical axis to show the environment of CTD^D, CTD^E and CTD^F that are obscured in the left view.



Extended Data Figure 3 | Initial 3D classification of multi-duplex classes.

a-d, Select multi-duplex combinations containing duplex^C from the 120 bp dsDNA reaction were 3D classified with partial signal subtraction and masks specific for the duplex combinations as described in methods. Shown here are the reconstructions of select classes after 3D refinement. Duplexes, their connectivity, particle number and map resolution according to the gold-standard fourier shell correlation (FSC) procedure are shown on top of each map. Maps are colored by the duplex tilt (5' tilt cyan, 3' purple), S2 DNA red. Duplex positions are labeled. A cartoon showing a simplified version of the duplex pattern

for each map is shown on the left of each class. Some maps are rotated to give a clear view of obscured duplexes in the top view with rotation axis indicated. The maps are organized by the number of duplexes in each classification.



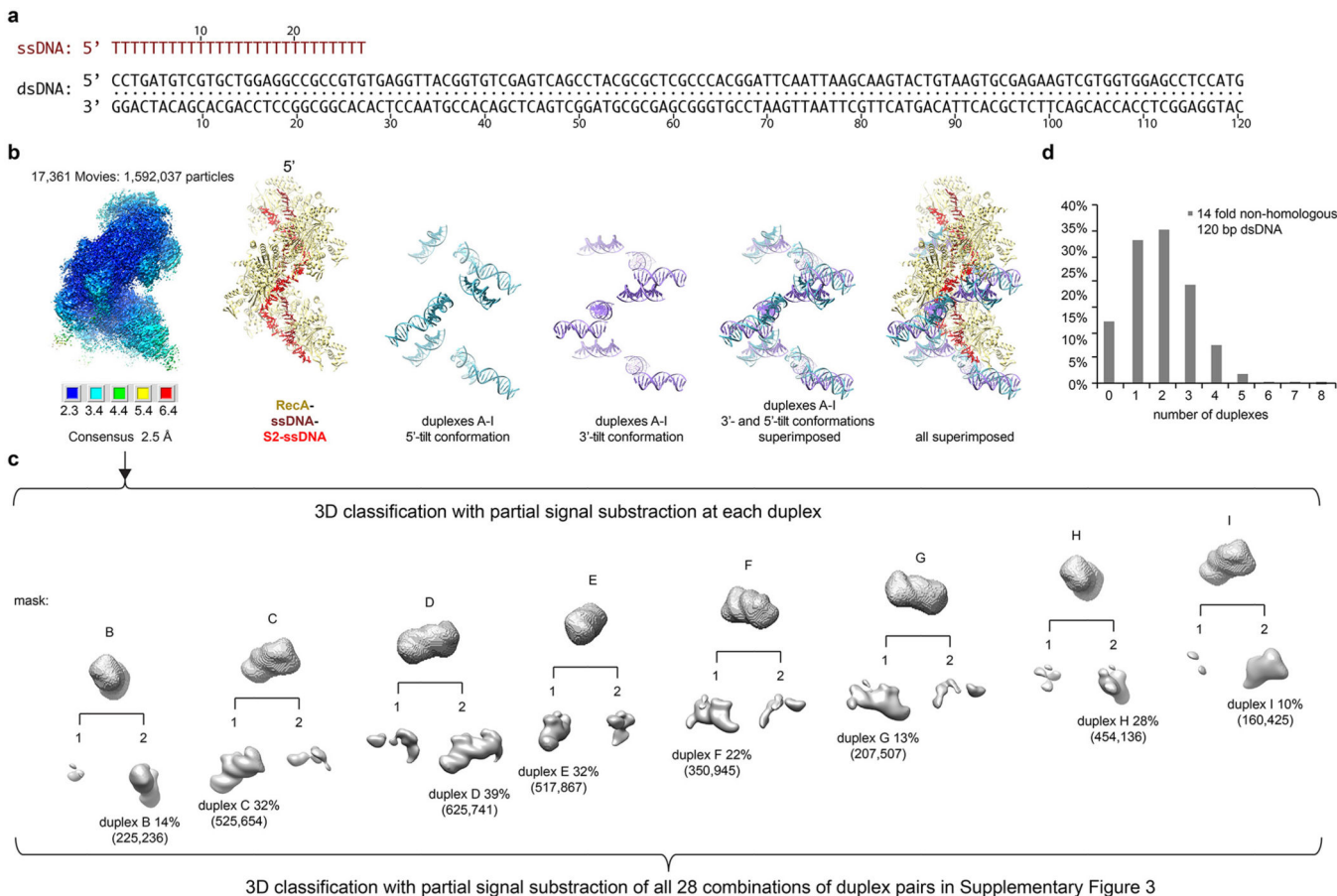
Extended Data Figure 4 | Cryo-EM Analysis of S2-connected duplex pairs from non-homologous dsDNA reactions.

a, The 5' end of the filament, but not the 3' end produces classes with conspicuously short duplexes. Reconstruction after 3D refinement of paired duplexes that have duplex^H or duplex^I. Top, maps of classes with short duplexes. Bottom, maps of classes from the same

3D classification but with regular, long duplexes. Black circle highlights the difference in the duplex length. Maps are colored by the duplex tilt (5' cyan, 3' purple), S2 DNA red. Overall resolution, from gold-standard refinement procedure, and particle numbers are also shown.

b, Charts show the number of particles with S2-connected duplex pairs for the non-homologous 67 bp dsDNA dataset. Each chart is of the series that contains the 1st duplex in common. Pairs of duplexes starting at RecA^C are omitted, as they are shown in Fig 1d. The data set was collected once.

c, Charts show the number of particles with S2-connected duplex pairs for the non-homologous 120 bp dsDNA dataset as in **b**. Pairs of duplexes starting at RecA^C are omitted as they are shown in Fig 1l. The data set was collected once.



Extended Data Figure 5 | Cryo-EM analysis of the strand exchange reaction with non-homologous 120 bp dsDNA.

a, Sequences of the 27 nt ssDNA (top, brown) and the 120 bp non-homologous dsDNA (bottom, black) used for this data set.

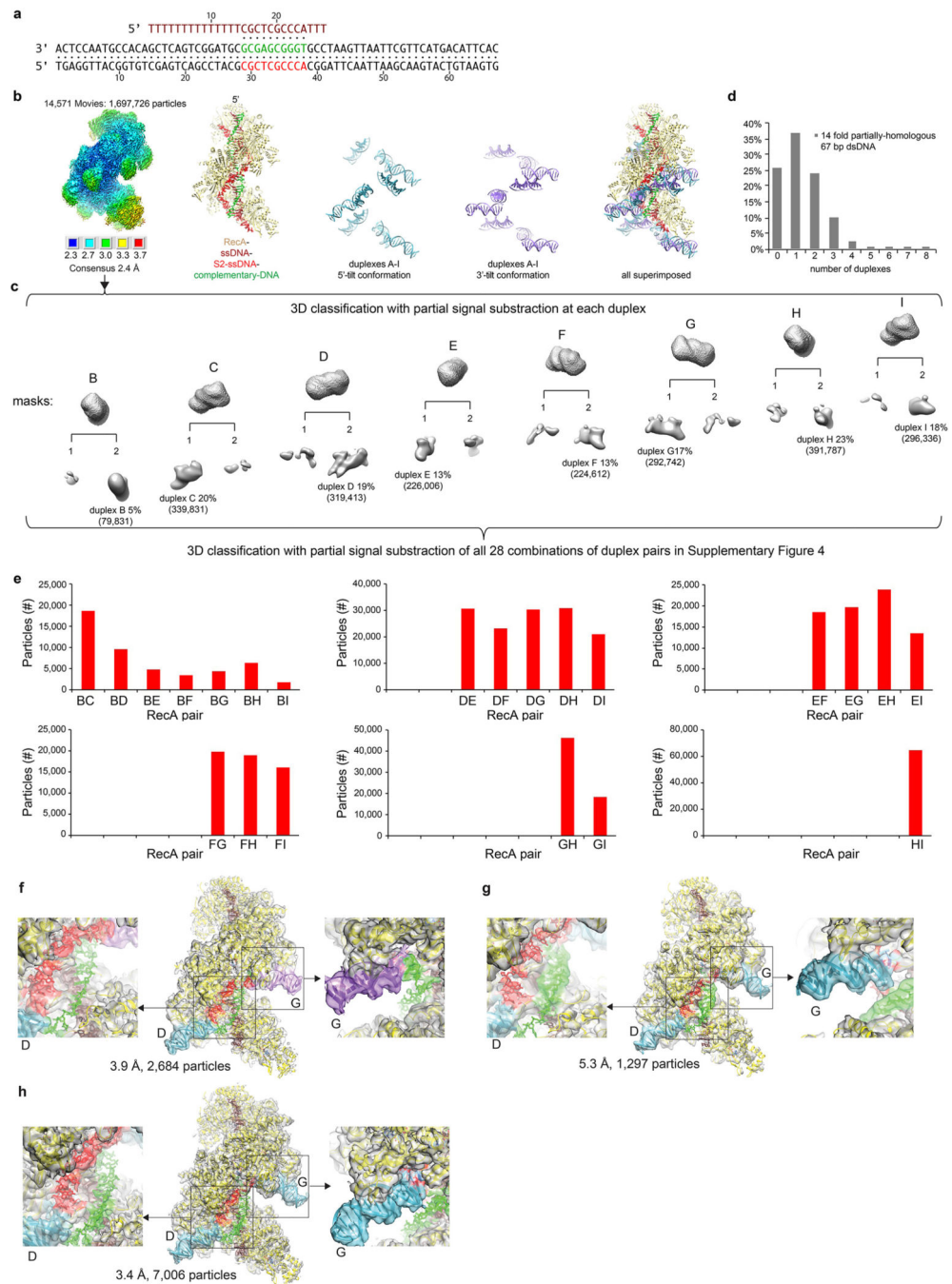
b, Starting from the left, consensus reconstruction map of this dataset colored by local resolution estimated with the RELION3 post-processing program. The resolution range is mapped to the colors in the inset below the map. Next, cartoon representation of the refined model of the consensus refinement. Primary ssDNA is colored in brown, S2 ssDNA red, and RecA is khaki. Cartoon representation of duplexes A to I in the 5'-tilted (cyan) and 3'-tilted conformations (purple), followed by the superposition of the two. Last, all cartoons

superimposed highlighting the tilt difference, and the relative organization of the DNA elements on the filament.

c, Masks and maps of the 3D classification for individual duplexes as in Extended Data Fig 2e. The overall duplex occupancy of up to 39 % is higher, with 1.9 duplexes per particle on average, as expected from the presence of nearly twice the number of binding sites corresponding to a duplex compared to the 67-bp non-homologous dsDNA reaction. The mid-filament duplex^F and duplex^G were again outliers with low relative occupancy (22 % and 13 %, respectively), consistent with the crowded filament mid-portion having low accessibility for dsDNA. And, as before, duplex^B and duplex^I at the poorly-ordered filament termini resulted in apparently low occupancy (14 % and 10 %, respectively).

Masks and maps of the 3D classification for all 28 combination of duplex pairs are shown in Supplementary Figure 3 and their details listed in Supplementary Spreadsheet 4. Paired-duplex 3D classification showed a distribution qualitatively very similar to that of the 67 bp dsDNA reaction, including the preference for 5' tilts at the 3' but not 5' end of the filament.

d, Chart of the percentage of particles that have the indicated number of duplexes for this dataset. The data set was collected once. Also see Supplementary Spreadsheet 3 for more details.



Extended Data Figure 6 | Cryo-EM analysis of the strand exchange reaction with partially-homologous 67 bp dsDNA.

a, Sequences of the primary 27-nt ssDNA (top, brown) and of the 67-bp donor dsDNA (bottom) containing a 10-nt segment of homology to the ssDNA. The dsDNA region of homology is colored green and red for the complementary and homologous strands, respectively. The directions of the DNA strands, and every 10th nucleotide are labeled. Dots indicate complementarity.

b, Consensus reconstruction as in Extended Data Fig. 2d, except complementary DNA can now be seen and is included in the model (green).

c, Masks and maps of the 3D classification for individual duplexes as in Extended Data Fig. 2e. Masks and maps of the 3D classification for all 28 combination of duplex pairs are shown in Supplementary Figure 4 and their details listed in Supplementary Spreadsheet 6.

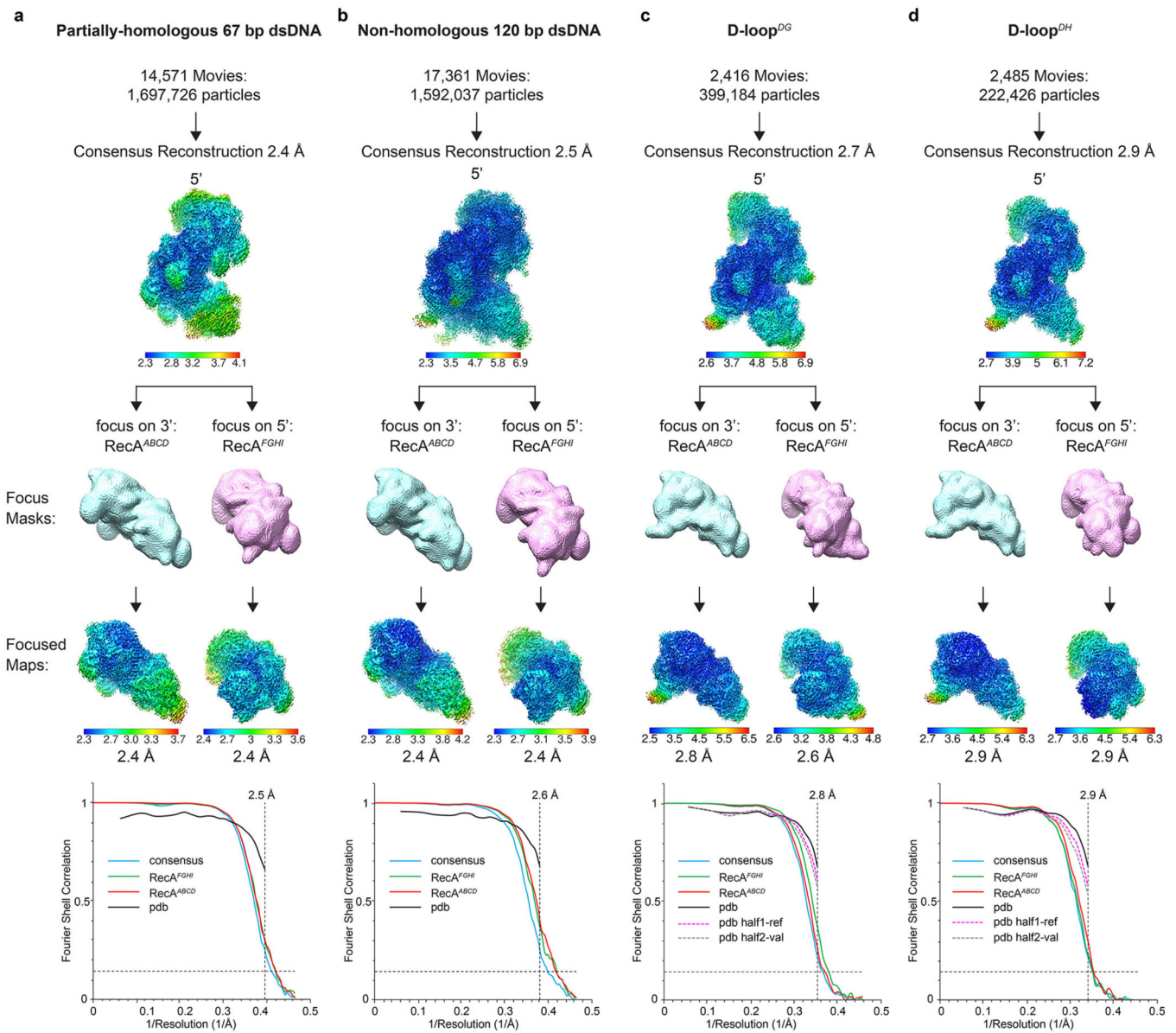
d, Chart of the percentage of particles that have the indicated number of duplexes for this dataset. The data set was collected once. Also see Supplementary Spreadsheet 5 for details.

e, Charts show the number of particles with the indicated S2-connected duplex pairs for this dataset. Pairs of duplexes starting at RecA^C are shown in Fig 2c. See Supplementary Spreadsheet 6. The data set was collected once.

f, Reconstruction after 3D refinement of the minor class from the classification of the 5'/3' tilted duplex^D-duplex^G particles. This class essentially has no heteroduplex density.

g, The same 3D sub-classification analysis of the class with 5'/5'-tilted duplexes (10,703 particles) identified only 1,297 particles with some heteroduplex density. Their reconstruction, shown in the figure, had overall weak density for both the homologous and complementary strands, and the connections to both duplexes were very weak. We presume that this class was still heterogeneous, but it could not be further sub-classified due to limited numbers of particles. This suggests that the D-loop forms preferentially in the 5'/3'-tilt conformation.

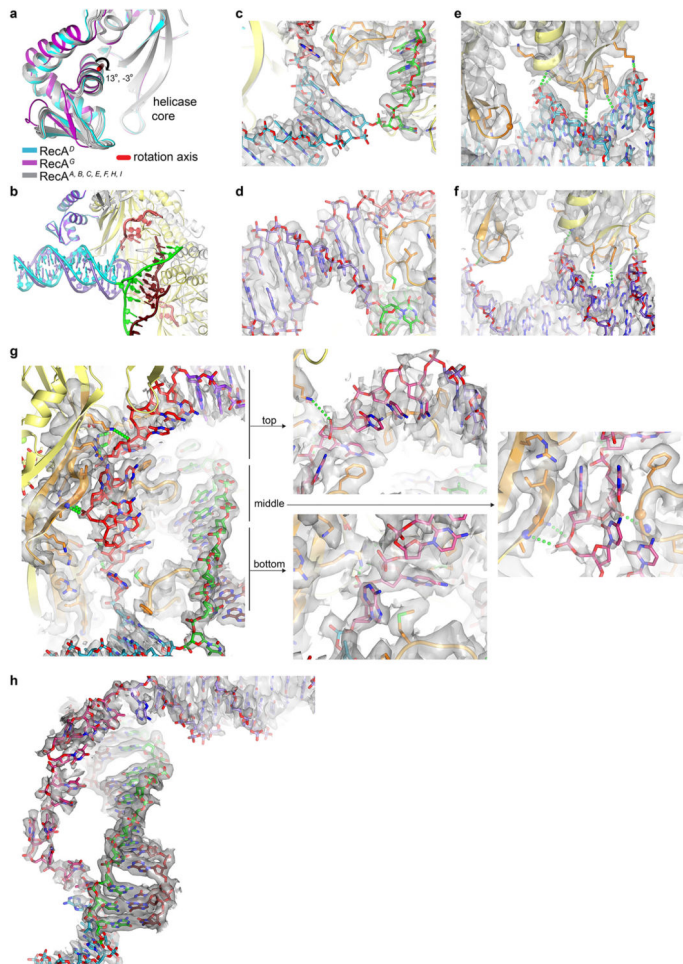
h, Figure shows reconstruction after 3D refinement of the major class from the 3D sub-classification of the 5'/5' tilted duplex^D-duplex^G particles. The density for the complementary strand of the heteroduplex is even weaker than that of **g**, and it is discontinuous. We could not further sub-classify this class, presumably due to extensive heterogeneity and limited numbers of particles.



Extended Data Figure 7 | Flowcharts of focused reconstructions for the data sets with partially-homologous 67-bp dsDNA, non-homologous 120 bp dsDNA, D-loop^{DG} and D-loop^{DH}.

a-d, Data processing of **a**, partially homologous 67 bp dsDNA reaction, **b**, non-homologous 120 bp dsDNA reaction, **c**, 9-RecA–D-loop^{DG} complex, and **d**, 9-RecA–D-loop^{DH} complex. The consensus reconstruction map and the two focused refinement maps for each dataset are colored by local resolution estimated with the RELION3 postprocess program. The resolution range is mapped to the colors in the inset next to each map. Bottom, the graphs show gold-standard FSC plots of the consensus reconstruction (blue) and two focused refinement maps (green and red), as well as the FSC curves between the composite map and the refined model (black, labeled pdb), between the composite map combining the first of the two half maps of each reconstruction and the model refined against this map (dashed pink curve, labeled half1-ref), and between the composite map combining the second of the two half maps and the model refined against the first composite half

map for validation (dashed gray curve, labeled half2-val). Horizontal dashed line marks the FSC cutoff of 0.143 and the vertical dashed line indicates the resolution of the REFMAC5 refinement.



Extended Data Figure 8 | Models and density of D-loop^{DG} and D-loop^{DH}.

Discussion of RecA-DNA contacts. The L2 loop-duplex contacts involve non-equivalent RecA protomers due to the different duplex orientations. Thus, duplex^D abuts the L2 loop of the adjacent RecA^C and additionally stacks with Gly204 backbone atoms, whereas duplex^G abuts the L2 loop of RecA^E, two RecAs over, and packs with the Met202 side chain (Fig. 3b–c). The CTD-dsDNA interactions are very similar in the two duplexes, except those at duplex^G consistently have slightly longer distances. The contacts from the loop-helix and hairpin motifs expand the minor groove of the DNA to 15.2 Å for duplex^D and 14.6 Å for duplex^G. The loop and the amino-terminus of the loop-helix motif (residues 297 to 302) make a set of hydrogen bonds to backbone phosphate groups of both strands (backbone amide of Lys302 and side chain of Lys297) while the side chain amide group of Gln300 hydrogen bonds to a thymidine O2 (duplex^D) or guanidine N3 (duplex^G) groups. Crucially, the Gln300 side chain is also in a π - π stack with the Trp290 side chain from the hairpin motif (residues 286 to 290 with the sequence Lys-Ala-Gly-Ala-Trp). With both side chains thus rigidified, they fit snugly in the minor groove and make multiple van der Waals

contacts to the ribose groups, with the amino group of the Trp290 side chain also hydrogen bonding to an N3 group of an adenine (duplex^D) or guanine (duplex^G). The tip of the hairpin (Ala-Gly-Ala portion) partially inserts in the minor groove as well, with the preceding Lys286 within contact distance of the phosphodiester backbone. A second hairpin at the end of the long $\beta 6$ strand of the helicase core is positioned above the adjacent major groove, and Lys232 contacts the phosphodiester backbone of duplex^G, but the corresponding contact is not made to duplex^D, which is farther away due to its 5' tilt. Among the CTD-duplex contacts, the K286N and K302N mutations were shown to cause defects in UV-repair in vivo and in binding to and pairing with dsDNA, although they were interpreted as affecting the secondary DNA-binding activity³⁹.

The S2 site contacts to the homologous strand are overall more extensive and the density stronger at the duplex^D-proximal two thirds of the strand than near duplex^G. The contacts start immediately after the opening of duplex^D by L2^C (these Ade28 contacts are discussed in the main text). The base group of the next residue, Cyt27, is sandwiched between the L2^C Met202 and L2^D Pro206 side chains, while its ribose group packs against the E207^D-R226^D salt bridge (Fig. 3g, bottom). Cyt26 then packs on one side with the L2^D Pro206 side chain and Gly204 and Asn205 backbone groups, and on the other side with Cyt25. The Cyt25 phosphodiester group in turn hydrogen bonds to the Ala230 backbone amide and Arg227 side chain groups, both from $\beta 6^E$ (Fig. 3g, middle). The next two nucleotides stack together and as a pair fit snugly into a tight gap between the backbones of $\beta 6^E$ and L2^D, as if they are pinched (Fig. 3g, middle). Here, $\beta 6^E$ side chain and backbone groups of Ile228 and Gly229 pack with Cyt24, while L2^D backbone groups from Phe203 and Gly204 pack with Cyt23. In addition, the phosphodiester group of Cyt24 hydrogen bonds to the backbone amide of L2^D Asn205, and that of Cyt23 to the side chain of $\beta 6^E$ Arg226. Thereafter, Cyt22 has RecA contacts and relative position very similar to Cyt27, five nucleotides away in the 3' filament direction. The one difference is the Cyt27-L2^C Met202 packing is replaced by Cyt22-L2^D Phe203, owing to the alternate conformation that L2^C adopts as it book-ends duplex^D (Fig. 3g, top). The Cyt22 base is also within contact distance of $\beta 6^E$ Lys245, where the K245N mutation was reported to affect homologous pairing⁴⁰.

The D-loop^{DH} structure recapitulates the key aspects of D-loop^{DG}. These include the conformations of the L2^C and L2^F loops and their stacking with their respective duplex^D and duplex^H, the overall S2 ssDNA backbone conformation, and the $\beta 6$ -L2 pinch of a nucleotide pair, of which it has two (Fig. 4b). One, at Thy26-Thy27, is essentially identical to that of D-loop^{DG}, while the other, at Thy21-Thy22 has Thy21 in a slightly different conformation, as it is next to the 3 nt spacer that transitions from S2-binding to duplex^H. The D-loop^{DH} structure also exhibits the same pattern of contacts at the transitions from the duplexes to the opened up homologous strands. The first-opened up base immediately after duplex^D (Ade31) packs with L2^C, while its phosphodiester backbone is contacted by Arg226 of $\beta 6^D$. The opposing, flipped out base of the complementary strand has poor density and does not appear to make any RecA contacts. And, as with D-loop^{DG}, the 3 nts just before duplex^H (Thy20-Thy19-Ade18) are poorly ordered, and make few contacts as they follow an alternate path around their respective L2^F.

a, On dsDNA binding, the CTD domains undergo a rotation about an axis (red stick) near residue 269 and roughly perpendicular to the direction of the filament. The 9 RecA protomers of the 9-RecA fusion protein were superimposed by aligning their helicase core

domains. The RecA^A, RecA^B, RecA^C, RecA^E, RecA^F, RecA^H, and RecA^I CTD domains (gray) do not exhibit any conformational changes, whereas those of RecA^D and RecA^G rotate (curved arrow) in opposite directions, by -3° and 13° respectively.

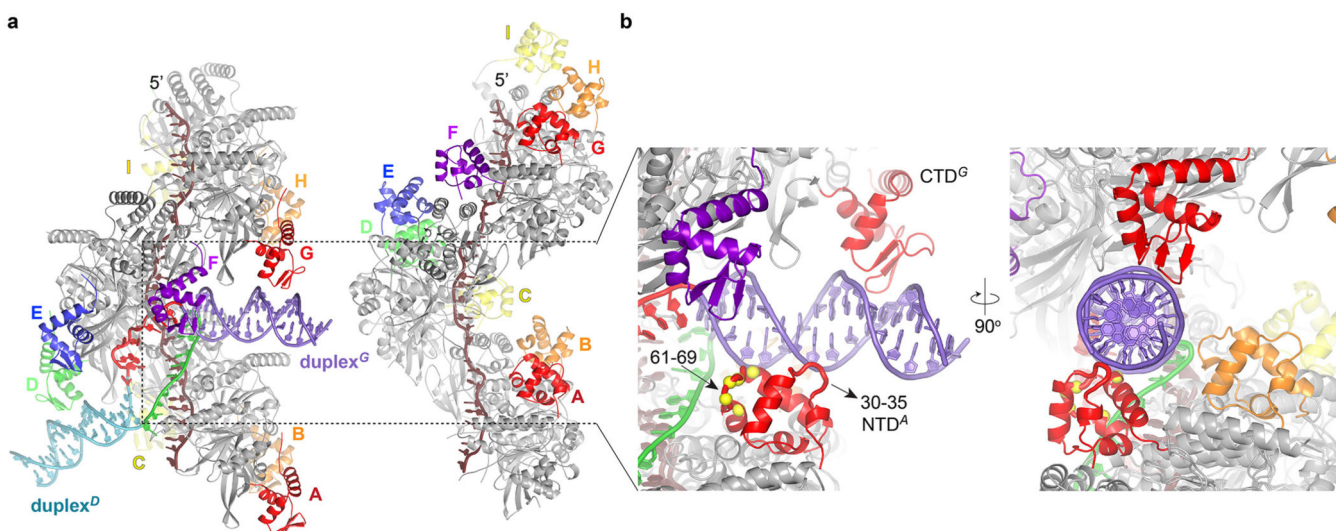
b, Cartoon representation showing the superposition RecA^D and RecA^G on their CTD domains to highlight the different tilts of duplex^D and duplex^G relative to their already rotated CTD domains. Colored as in Figure 3d (which shows the superposition of the RecA protomers on their helicase domains). The rest of the RecAs are colored yellow and gray.

c-d, Density of duplex^D and duplex^G from the D-loop^{DG} maps used in REFMAC5, as described in methods, in the same orientation as Fig 3b and 3c.

e-f, Density of the interactions of duplex^D and duplex^G with their respective CTDs from the D-loop^{DG} REFMAC5 refinement the same orientation as Fig 3e and 3f.

g, Density of the S2 site structural elements and all DNA from the D-loop^{DG} maps used in REFMAC5. Orientations as in Fig 3g.

h, Density of the DNA only from D-loop^{DH} maps used in REFMAC5 in the same orientation as Fig 4b. RecA protomers and their density are omitted for clarity.



Extended Data Figure 9 | Rad51, the eukaryotic RecA homolog, likely functions similarly to RecA in strand exchange.

Rad51 lacks the RecA CTD but instead contains an N-terminal domain (NTD) that has been implicated in dsDNA binding by chemical shift perturbation data of the isolated NTD domain⁴¹. While the RAD51 NTD is structurally unrelated to the RecA CTD, it occupies an analogous position at the filament periphery^{42–44}, except that it is oriented with its solvent exposed surface pointing to the 5' end of the filament instead of the 3' end that the RecA CTD points to. Because of this, the NTDs at the 5' end of the RAD51 filament are more accessible for initial dsDNA binding compared to those at the 3' end, the opposite of the RecA filament.

a, Figure shows a side-by-side comparison of the RecA D-loop^{DG} structure (left) and a model of a 9-protomer Rad51-ssDNA presynaptic filament (right). The Rad51 model was constructed from the coordinates of the 4.4 Å cryo-EM structure of a three-Rad51 segment bound to 9 nts of ssDNA (PDBID 5H1B) by successively applying the transformation that

relates the middle-Rad51 to the 5'-most Rad51. Because of small differences in the relative orientation of adjacent protomers, the RecA and Rad51 filaments are superimposable only locally (while individual helicase domains superimpose with an r.m.s.d. of 1.9 Å for 201 of 245 RecA C_α atoms, three-protomer segments superimpose with an r.m.s.d. of 2.2 Å for 531 C_α atoms). For the side-by-side comparison, the two filaments were superimposed on the central, 3-protomer segment (RecA^D to RecA^F). The proteins and cofactors are colored uniformly gray, except their respective CTD and NTD domains are colored rainbow as in Figure 1a, the primary ssDNA is brown for both, and the rest of the RecA DNA molecules are colored as in Figure 3a. The NTD and CTD domains are labeled.

b, Close-up views of the RecA Dloop^{DG} and the 9-Rad51 model superimposed on the three 3'-most RecA^A-RecA^C, focusing on the RecA duplex^G in an orientation similar to **a** (right view rotated 90° about the vertical axis). This superposition brings duplex^G of RecA in close proximity to the RAD51 NTD^A, which is nearly one turn of the filament away, in the 3' direction, from CTD^G owing to the different locations of the Rad51 NTDs. The backbone amide nitrogen atoms reported to be involved in dsDNA binding⁴¹ are shown as yellow spheres (Ile61, Lys64, Gly65, Ile66 and Ala69). These are located at a loop (shown in thick tube) and the N-terminus of the helix that follows. These structural elements are positioned relative to the DNA duplex analogously to the loop-helix motif of the RecA CTD, although they approach the DNA from the opposite direction. An adjacent Rad51 loop (residues 30–35, also shown as a thick tube) is also in close proximity to the RecA duplex^G of the superposition.

Extended Data Table 1.

Cryo-EM data collection, model refinement and validation statistics.

	Partially-homologous 67 bp dsDNA (PDB-7JY6) (EMDB-22522)	Non-homologous 120 bp dsDNA (PDB-7JY8) (EMDB-22524)	D-loop ^{DG} (PDB-7JY9) (EMDB-22525)	D-loop ^{DH} (PDB-7JY7) (EMDB-22523)
Data collection and processing				
Magnification	81,000	81,000	105,000	105,000
Voltage (kV)	300	300	300	300
Electron exposure (e-/Å ²)	51.6	53.7	56.6	66.6
Defocus range (µm)	1.0–2.5	0.8–2.5	1.0–2.5	1.0–2.5
Pixel size (Å)	1.078	1.078	1.096	1.096
Symmetry imposed	C1	C1	C1	C1
Initial particle images (no.)	1,697,726	1,592,037	399,184	222,426
Final particle images (no.)	1,697,726	1,592,037	399,184	222,426
Map resolution (Å)				
Consensus reconstruction	2.4	2.5	2.7	2.9
Focus ABCD reconstruction	2.4	2.4	2.8	2.9
Focus FGHI reconstruction	2.4	2.4	2.6	2.9
FSC threshold	0.143	0.143	0.143	0.143

	Partially-homologous 67 bp dsDNA	Non-homologous 120 bp dsDNA	D-loop^{DG}	D-loop^{DH}
	(PDB-7JY6)	(PDB-7JY8)	(PDB-7JY9)	(PDB-7JY7)
	(EMDB-22522)	(EMDB-22524)	(EMDB-22525)	(EMDB-22523)
Map resolution range (Å)				
Consensus reconstruction	2.3–4.1	2.3–6.9	2.6–6.9	2.1–12
Focus ABCD reconstruction	2.3–3.7	2.3–4.2	2.6–6.5	2.1–63
Focus FGHI reconstruction	2.4–3.5	2.3–3.9	2.7–4.8	2.1–63
Refinement				
Initial model used (PDB code)	3CMW	3CMW	3CMX	3CMX
Model resolution (Å)	2.5	2.5	2.8	2.9
FSC threshold	0.70	0.68	0.71	0.68
Model resolution range (Å)	195–2.5	195–2.5	197–2.8	196–2.9
Map sharpening <i>B</i> factor (Å ²)	–73	–71	–71	–64
Model composition				
Non-hydrogen atoms	23,389	24,059	24,858	25,127
Protein residues	2,967	2,967	2,964	2,967
DNA residues	38	72	111	123
Cofactors	9	9	9	9
B factors (Å²)				
Protein	78.0	92.9	75.8	111.1
DNA	89.0	185.7	151.7	222.7
Cofactors	41.5	47.8	29.6	65.6
R.m.s. deviations				
Bond lengths (Å)	0.006	0.006	0.006	0.006
Bond angles (°)	1.57	1.55	1.52	1.53
B factors main chain (Å ²)	3.37	3.73	2.47	2.54
B factors side chain (Å ²)	3.49	4.77	2.58	2.83
Validation				
MolProbity score	1.68	1.62	1.54	1.75
Clashscore	1.06	0.93	0.75	1.0
Poor rotamers (%)	7.58	5.40	5.70	7.84
Ramachandran plot				
Favored (%)	96.71	95.96	96.54	96.03
Allowed (%)	3.29	3.70	3.12	3.56
Disallowed (%)	0.17	0.34	0.34	0.41
R _{work} (%)	27.60	28.89	26.60	26.75
Average FSC	0.89	0.89	0.90	0.89

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

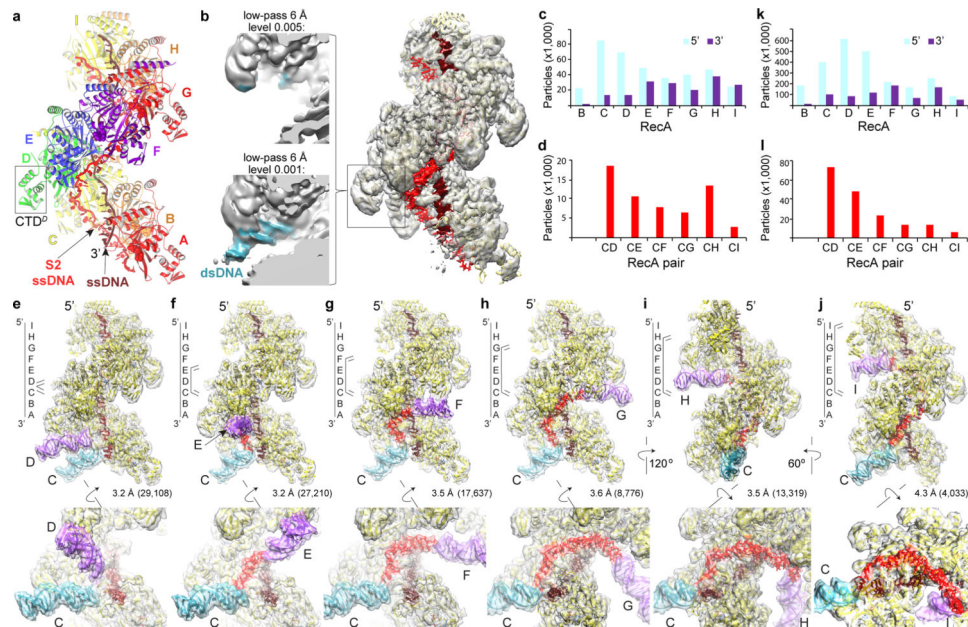
We thank the staff of the MSKCC Cryo-EM facility, the NYSBC Simons Electron Microscopy Center, and the HHMI Cryo-EM facility for help with data collection. Supported by HHMI and National Institutes of Health grant CA008748.

References

- Bell JC & Kowalczykowski SC RecA: Regulation and Mechanism of a Molecular Search Engine: (Trends in Biochemical Sciences, June 2016, Vol. 41, No. 6, 491–507). Trends in biochemical sciences 41, 646, doi:10.1016/j.tibs.2016.05.006 (2016). [PubMed: 27283513]
- Cox MMMotoring along with the bacterial RecA protein. Nature reviews. Molecular cell biology8, 127–138, doi:10.1038/nrm2099 (2007). [PubMed: 17228330]
- Chen Z, Yang H & Pavletich NP Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. Nature 453, 489–484, doi:10.1038/nature06971 (2008). [PubMed: 18497818]
- Zlotnick A, Mitchell RS & Brenner SL recA protein filaments bind two molecules of single-stranded DNA with off rates regulated by nucleotide cofactor. J Biol Chem 265, 17050–17054 (1990). [PubMed: 2211610]
- Mazin AV & Kowalczykowski SC The specificity of the secondary DNA binding site of RecA protein defines its role in DNA strand exchange. Proceedings of the National Academy of Sciences of the United States of America 93, 10673–10678, doi:10.1016/j.pcc.2009.03.005 (1996). [PubMed: 8855238]
- Wittung P, Bazemore LR, Takahashi M, Norden B & Radding C. Second-site RecA-DNA interactions: lack of identical recognition. Biochemistry 35, 15349–15355, doi:10.1021/bi961755e (1996). [PubMed: 8952486]
- Mazin AV & Kowalczykowski SC The function of the secondary DNA-binding site of RecA protein during DNA strand exchange. The EMBO journal 17, 1161–1168, doi:10.1093/emboj/17.4.1161 (1998). [PubMed: 9463393]
- Story RM, Weber IT & Steitz TA The structure of the E. coli recA protein monomer and polymer. Nature 355, 318–325, doi:10.1038/355318a0 (1992). [PubMed: 1731246]
- Greene ECDNA Sequence Alignment during Homologous Recombination. J Biol Chem291, 11572–11580, doi:10.1074/jbc.R116.724807 (2016). [PubMed: 27129270]
- Aihara Het al.An interaction between a specified surface of the C-terminal domain of RecA protein and double-stranded DNA for homologous pairing. Journal of molecular biology274, 213–221, doi:10.1006/jmbi.1997.1403 (1997). [PubMed: 9398528]
- Kurumizaka H, Ikawa S, Sarai A & Shibata T. The mutant RecA proteins, RecAR243Q and RecAK245N, exhibit defective DNA binding in homologous pairing. Archives of biochemistry and biophysics 365, 83–91, doi:10.1006/abbi.1999.1166 (1999). [PubMed: 10222042]
- Hsieh P, Camerini-Otero CS & Camerini-Otero RD The synapsis event in the homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of DNA. Proceedings of the National Academy of Sciences of the United States of America 89, 6492–6496, doi:10.1073/pnas.89.14.6492 (1992). [PubMed: 1631148]
- Forget AL & Kowalczykowski SC Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search. Nature 482, 423–427, doi:10.1038/nature10782 (2012). [PubMed: 22318518]
- Qi Zet al.DNA sequence alignment by microhomology sampling during homologous recombination. Cell160, 856–869, doi:10.1016/j.cell.2015.01.029 (2015). [PubMed: 25684365]
- Gourves AS, Defais M & Johnson NP Equilibrium binding of single-stranded DNA to the secondary DNA binding site of the bacterial recombinase RecA. J Biol Chem 276, 9613–9619, doi:10.1074/jbc.M004855200 (2001). [PubMed: 11121401]
- Cazaux Cet al.Investigation of the secondary DNA-binding site of the bacterial recombinase RecA. J Biol Chem273, 28799–28804, doi:10.1074/jbc.273.44.28799 (1998). [PubMed: 9786879]
- Min JH & Pavletich NP Recognition of DNA damage by the Rad4 nucleotide excision repair protein. Nature 449, 570–575, doi:10.1038/nature06155 (2007). [PubMed: 17882165]

18. Scrima A et al. Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* 135, 1213–1223, doi:10.1016/j.cell.2008.10.045 (2008). [PubMed: 19109893]
19. Jain SK, Cox MM & Inman RB Occurrence of three-stranded DNA within a RecA protein filament. *J Biol Chem* 270, 4943–4949, doi:10.1074/jbc.270.9.4943 (1995). [PubMed: 7876269]
20. Lee AJ, Sharma R, Hobbs JK & Walti C. Cooperative RecA clustering: the key to efficient homology searching. *Nucleic Acids Res* 45, 11743–11751, doi:10.1093/nar/gkx769 (2017). [PubMed: 28977583]
21. Zheng S Q et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature methods* 14, 331–332, doi:10.1038/nmeth.4193 (2017). [PubMed: 28250466]
22. Rohou A & Grigorieff N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology* 192, 216–221, doi:10.1016/j.jsb.2015.08.008 (2015). [PubMed: 26278980]
23. Zivanov J et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* 7, doi:10.7554/eLife.42166 (2018).
24. Zivanov J, Nakane T & Scheres SHW A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* 6, 5–17, doi:10.1107/S205225251801463X (2019).
25. Brown A et al. Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr D Biol Crystallogr* 71, 136–153, doi:10.1107/S1399004714021683 (2015). [PubMed: 25615868]
26. Afonine P V et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68, 352–367, doi:10.1107/S0907444912001308 (2012). [PubMed: 22505256]
27. Winn M D et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67, 235–242, doi:S0907444910045749[pil]10.1107/S0907444910045749 (2011). [PubMed: 21460441]
28. Chen Z, Yang H & Pavletich N P Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* 453, 489–484, doi:10.1038/nature06971 (2008). [PubMed: 18497818]
29. Tran Q H & Uden G. Changes in the proton potential and the cellular energetics of *Escherichia coli* during growth by aerobic and anaerobic respiration or by fermentation. *European journal of biochemistry* 251, 538–543, doi:10.1046/j.1432-1327.1998.2510538.x (1998). [PubMed: 9492330]
30. Lee J Y, Qi Z & Greene E C ATP hydrolysis Promotes Duplex DNA Release by the RecA Presynaptic Complex. *J Biol Chem* 291, 22218–22230, doi:10.1074/jbc.M116.740563 (2016). [PubMed: 27587394]
31. Menetski J P, Bear D G & Kowalczykowski S C Stable DNA heteroduplex formation catalyzed by the *Escherichia coli* RecA protein in the absence of ATP hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America* 87, 21–25, doi:10.1073/pnas.87.1.21 (1990). [PubMed: 2404275]
32. Rosselli W & Stasiak A. Energetics of RecA-mediated recombination reactions. Without ATP hydrolysis RecA can mediate polar strand exchange but is unable to recycle. *Journal of molecular biology* 216, 335–352, doi:10.1016/S0022-2836(05)80325-0 (1990). [PubMed: 2147722]
33. Stasiak A Z, Rosselli W & Stasiak A. RecA-DNA helical filaments in genetic recombination. *Biochimie* 73, 199–208, doi:10.1016/0300-9084(91)90203-d (1991). [PubMed: 1883882]
34. Rosselli W & Stasiak A. The ATPase activity of RecA is needed to push the DNA strand exchange through heterologous regions. *The EMBO journal* 10, 4391–4396 (1991). [PubMed: 1836761]
35. Jain S K, Cox M M & Inman R B On the role of ATP hydrolysis in RecA protein-mediated DNA strand exchange. III. Unidirectional branch migration and extensive hybrid DNA formation. *J Biol Chem* 269, 20653–20661 (1994). [PubMed: 8051165]
36. Cox M M Why does RecA protein hydrolyse ATP? *Trends in biochemical sciences* 19, 217–222, doi:10.1016/0968-0004(94)90025-6 (1994). [PubMed: 8048163]
37. Konforti B B & Davis R W The preference for a 3' homologous end is intrinsic to RecA-promoted strand exchange. *J Biol Chem* 265, 6916–6920 (1990). [PubMed: 2182632]

38. Konforti BB & Davis RW DNA substrate requirements for stable joint molecule formation by the RecA and single-stranded DNA-binding proteins of Escherichia coli. *J Biol Chem* 266, 10112–10121 (1991). [PubMed: 2037569]
39. Kurumizaka H et al. A possible role of the C-terminal domain of the RecA protein. A gateway model for double-stranded DNA binding. *J Biol Chem* 271, 33515–33524, doi:10.1074/jbc.271.52.33515 (1996). [PubMed: 8969216]
40. Kurumizaka H, Ikawa S, Sarai A & Shibata T. The mutant RecA proteins, RecAR243Q and RecAK245N, exhibit defective DNA binding in homologous pairing. *Archives of biochemistry and biophysics* 365, 83–91, doi:10.1006/abbi.1999.1166 (1999). [PubMed: 10222042]
41. Aihara H, Ito Y, Kurumizaka H, Yokoyama S & Shibata T. The N-terminal domain of the human Rad51 protein binds DNA: structure and a DNA binding surface as revealed by NMR. *Journal of molecular biology* 290, 495–504, doi:10.1006/jmbi.1999.2904 (1999). [PubMed: 10390347]
42. Conway AB et al. Crystal structure of a Rad51 filament. *Nature structural & molecular biology* 11, 791–796, doi:10.1038/nsmb795 (2004).
43. Short JM et al. High-resolution structure of the presynaptic RAD51 filament on single-stranded DNA by electron cryo-microscopy. *Nucleic Acids Res* 44, 9017–9030, doi:10.1093/nar/gkw783 (2016). [PubMed: 27596592]
44. Xu J et al. Cryo-EM structures of human RAD51 recombinase filaments during catalysis of DNA-strand exchange. *Nature structural & molecular biology* 24, 40–46, doi:10.1038/nsmb.3336 (2017).



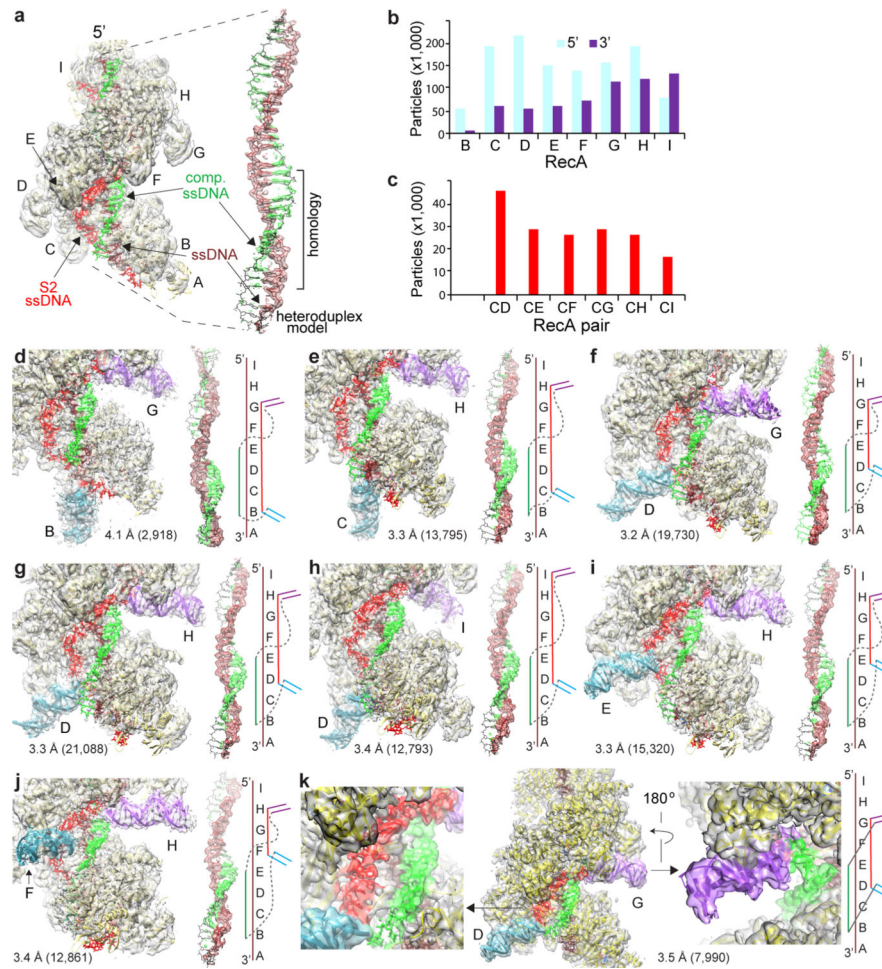


Figure 2 | Cryo-EM analysis of the strand exchange reaction with 67 bp dsDNA containing a 10-nt region of homology.

a, Semi-transparent maps of the unsharpened cryo-EM density from the consensus reconstruction (left), and of the masked heteroduplex density from the post-processed 2.4 Å map (right). The density and model of the complementary strand in the heteroduplex are in green, and the rest are colored as in Figure 1b. **b**, Number of particles that have 5' (cyan) or 3'-tilted (purple-blue) tilted duplexes. **c**, Number of particles with S2-connected duplex pairs (duplex^C series). **d-j**, Representative 3D refined classes that contain patchy heteroduplex density between the indicated S2-connected duplex pairs. The density and schematics, drawn as in Figure 1e, now also show the complementary strand in green. **k**, Overall (center) and close-up views of the class containing a continuous D-loop spanning RecA^D-RecA^G after 3D refinement. Charts are from a single data set.

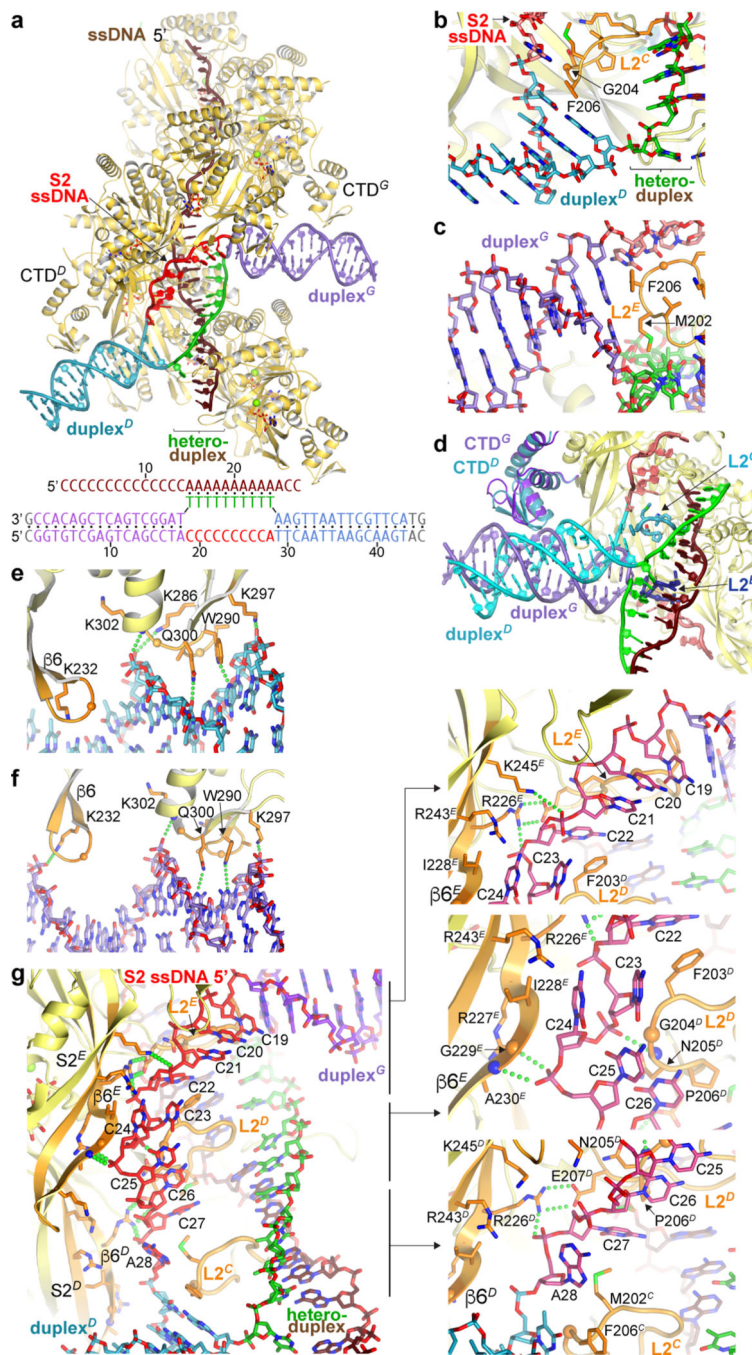


Figure 3 | Structure of D-loop^{DG}.
a. Cartoon representation and sequence of D-loop^{DG}. Protein is yellow, DNA is colored as indicated (gray nucleotides in sequence are disordered), and ATPγS is in sticks. **b.** Close-up view of duplex^D stacking against L2^C. All L2 loop side chains are shown, but only those that contact the duplex are labeled. **c.** Close-up view of duplex^G stacking against L2^E, rotated ~180° from **a.** **d.** Superposition of the helicase cores of RecA^D and RecA^G highlighting differences in duplex^D and duplex^G tilts and L2 loops. Colored as in **a.**, except the CTD domains and L2 loops are colored as their respective duplexes. Cartton is limited to the

duplex-proximal DNA portions, and to only one set of RecA protomers outside RecA^D and RecA^G. **e-f**, Close-up views of the CTD-dsDNA contacts in the minor groove of the DNA. Green dotted lines indicate hydrogen-bond contacts. **g**, Left, overall view of the S2 ssDNA, colored as in **a**, except the S2 site $\beta 6$ and L2 are orange. Right, close-up views of the interactions.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

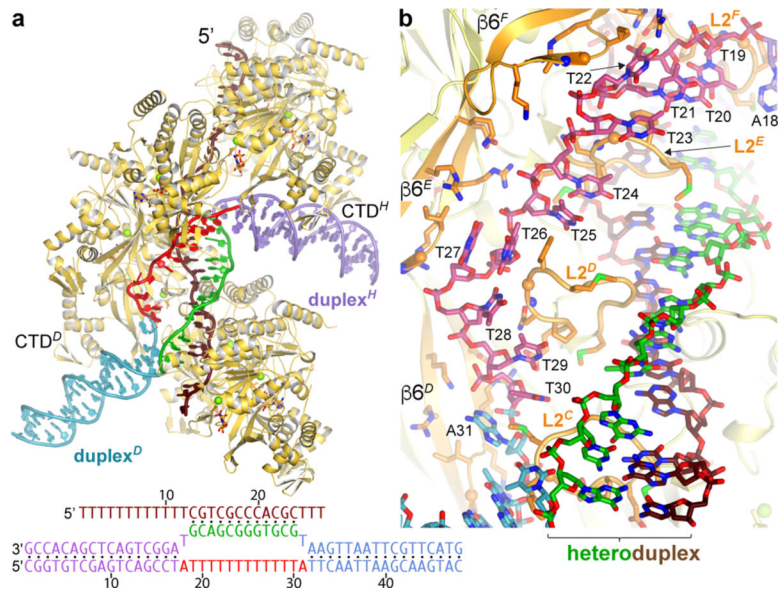


Figure 4 | Structure of D-loop^{DH}.
a, Cartoon representation of D-loop^{DH} and its sequence. Coloring and view as in Fig. 3a, except for a -30° rotation about the horizontal axis. **b**, Closeup view of the S2 ssDNA and S2 protein residues as in Fig. 3g.