Title

RecA: Regulation and Mechanism of a Molecular Search Engine

Journal

Trends in Biochemical Sciences, 41(6)

ISSN

0968-0004

Authors

Bell, Jason C

Kowalczykowski, Stephen C

Publication Date

2016-06-01

DOI

10.1016/j.tibs.2016.04.002

Peer reviewed

HHS Public Access

Author manuscript

Trends Biochem Sci. Author manuscript; available in PMC 2017 June 01.

Published in final edited form as:

Trends Biochem Sci. 2016 June ; 41(6): 491–507. doi:10.1016/j.tibs.2016.04.002.

RecA: Regulation and Mechanism of a Molecular Search Engine

Jason C. Bell[†] and **Stephen C. Kowalczykowski**[*]

Department of Microbiology & Molecular Genetics and Department of Molecular & Cellular Biology, University of California, Davis, California 95616, USA

Abstract

Homologous recombination maintains genomic integrity by repairing broken chromosomes. The broken chromosome is partially resected to produce single-stranded DNA that is used to search for homologous dsDNA. This homology-driven 'search and rescue' is catalyzed by a class of DNA strand exchange proteins that are defined in relation to *Escherichia coli* RecA, which forms a filament on single-stranded DNA. Here, we review the regulation of RecA filament assembly and the mechanism by which RecA quickly and efficiently searches for and identifies a unique homologous sequence amongst a vast excess of heterologous DNA. Given that RecA is the prototypic DNA strand exchange protein, its behavior affords insight into the actions of eukaryotic RAD51 orthologs and their regulators, BRCA2 and other tumor suppressors.

Keywords

RAD51; Recombination; Homology search; Self-assembly; Single-molecule imaging; Total internal reflection fluorescence (TIRF); Förster Resonance Energy Transfer (FRET)

RecA: a molecular search engine

Escherichia coli RecA is the defining member of an ancient and ubiquitous clade of DNA strand exchange proteins that are essential for homologous recombination [1]. This clade consists of three distinct families: RecA, RAD51 and the RAD51-paralogs (Figure 1A) [1]. RecA is found in all free-living bacteria, and is the most slowly evolving gene involved in DNA metabolism, with an average sequence conservation of ~60–70% across the entire Bacterial domain of life [2]. Many bacteriophage also have homologs from phylogenetically-defined RecA subfamilies: the Phage SAR1, Phage SAR2, and Phage UvsX, although only phage T4 UvsX has been isolated and studied biochemically (the others were only identified in metagenomic data) [3, 4]. The RecA-SAR1 subfamily is likely an ancient paralog of RecA not found in cultured/studied bacteria. Plants—but neither

[*]Correspondence: Stephen C. Kowalczykowski, University of California, Department of Microbiology & Molecular Genetics, One Shields Ave. Briggs Hall - Rm. 310, Davis, CA 95616-8665, Phone: 530-752-5938, Fax: 530-752-5939, ; Email: sckowalczykowski@ucdavis.edu

[†]Present address: Department of Biochemistry, Stanford University, Stanford, California 94305, USA.

animals nor fungi—have both mitochondrial and chloroplast specific variants of RecA of unknown functional importance, indicating that these genes were acquired through endosymbiotic transfer by an ancestor common to eukaryotes, but then lost during evolutionary divergence [4]. All Archaea and Eukarya share two classes of RecA homologs [1]. The RADα class is functionally homologous to RecA and includes archaeal RadA, eukaryotic RAD51 and eukaryotic DMC1, the last of which functions specifically during meiosis [5, 6]. The RADβ class is a divergent class of proteins that genetically function with RAD51 and are collectively known as the RAD51 paralogs. These paralogs are RAD51B, RAD51C, RAD51D, XRCC2, and XRCC3 proteins in mammals, Rad55 and Rad57 as well as Csm2 and Psy3 in fungi, and RadB in archaea; recently, RadA of *E. coli* (not to be confused with RadA of Archaea), which is a member of this RADβ class, was shown to be a novel branch migration enzyme [7]. Many of the eukaryotic paralogs form various sub-complexes [8]. Though recent biochemical analysis has demonstrated that the yeast Rad55/57 complex promotes homologous recombination by antagonizing an anti-recombination helicase (Srs2) [9], the biochemical function(s) of the human paralogs remain unclear.

Failure to properly regulate the assembly of human RAD51 causes an accumulation of mutations that accelerate tumorigenesis in individuals with genetic defects in BRCA2 and its epistasis group, which includes the five RAD51 paralogs (RAD51B, RAD51C, RAD51D, XRCC2, and XRCC3), PALB2, and some proteins of the Fanconi Anemia core complex [8, 10]. This accelerated mutation and tumorigenesis is due to the loss of homologous recombination-dependent repair, often in conjunction with—or downstream of—DNA repair pathways that correct specific types of damage, including crosslinks, alkylations, bulky adducts, ssDNA gaps, and double-strand breaks [11]. Although this review will focus on *E. coli* RecA, many of the biochemical properties of RecA are fundamentally similar to human RAD51 [5], and studies on RecA continue to inform our understanding of the molecular details of the core homologous recombination machinery across all domains of life [12].

In vivo, RecA preferentially binds to single-stranded DNA (ssDNA) that is generated through the resection of double-stranded DNA (dsDNA) breaks, or the formation of ssDNA gaps, which arise when replication forks stall or collapse. To initiate homologous recombination, helicases and/or nucleases process a broken chromosome or stalled replication fork to produce ssDNA, which is rapidly coated with—and continuously sequestered by—ssDNA binding proteins (SSB, in bacteria). An essential step during homologous recombination is the assembly of a presynaptic filament of RecA on the ssDNA of the broken chromosome, which in turn uses the ssDNA sequence to search for a homologous region in the dsDNA genome. Importantly, bacteria are either haploid or merodiploid (partially diploid) organisms, and therefore use the sister chromatid as the repair template during replication but before the cell has divided. Formation of the RecA filament triggers the SOS response through the autocatalytic proteolysis of the LexA repressor via a direct LexA-RecA-ssDNA complex. In addition to inducing the expression of DNA repair proteins and activating translesion polymerases, the SOS response delays cell division, temporarily increasing the copy number of the chromosome and causing physiological filamentation of *E. coli*. Recent *in vivo* super-resolution imaging of RecA

filaments demonstrates that the RecA filament (or bundles of filaments) transverses the mother and daughter cells in response to DNA damage [13].

RecA filament assembly in response to a dsDNA break is not spontaneous, but rather tightly regulated by RecBCD, a helicase/nuclease that processes dsDNA breaks and directly loads RecA onto ssDNA upon encountering a bacterial self-recognition sequence called χ (*crossover hotspot instigator*, Chi, 5′-GCTGGTGG-3′), which is spaced approximately every 4–5 kbp throughout the *E. coli* genome [14]. When the χ sequence is not present, RecBCD will rapidly and processively degrade linear dsDNA (up to 30 kbp during an average processing event in vitro, but ~10 kbp in vivo [15]) to defend against bacteriophage infection. Interestingly, the degradation products generated by RecBCD can be used by the CRISPR system to identify foreign DNA sequences during 'adaptation', when protospacer sequences are acquired [16]. During DNA replication, RecA filaments are assembled on ssDNA gaps and breaks that form when replication forks stall or collapse upon encountering either chemical damage (nicks, crosslinks, adducts, etc.), or in some cases, transcriptional impediments (e.g., collision with RNA polymerases) [17]. These so-called daughter strand gaps are processed not by RecBCD (which requires a nearly blunt DNA end to initiate), but rather by the concerted action of RecQ helicase and RecJ nuclease, which produces long regions (i.e., several thousand nucleotides) of SSB-coated ssDNA tails or gaps (i.e., a ssDNA region flanked by either one or two segments of dsDNA) [18, 19]. Unlike RecBCD, neither RecQ nor RecJ are reported to directly interact with RecA during processing; rather, RecA is loaded onto SSB-ssDNA by the RecF, RecO, and RecR proteins [20–22]. In this context, both the RecFOR and RecOR complexes regulate RecA filament assembly by enhancing nucleation and growth of the filament through structural perturbation of the SSB-coated ssDNA nucleoprotein (i.e., protein-DNA) complex and/or altering the kinetics of SSB sliding/wrapping on ssDNA [23, 24]. Here, we aim to provide a synthesis of our current understanding of the mechanism of RecA, which is illuminated by both a rich history of genetic and biochemical research, as well as recent single-molecule experiments.

The search for *rec* genes: identification and biochemical isolation of RecA

Clark and Margulies first identified *recA* in 1965 in a genetic screen where they mutagenized F^- bacteria and screened for strains unable to recombine with an Hfr donor strain [25]. Subsequent genetic analysis revealed that *recA* is essential for the RecBCD- and RecF-pathways of homologous recombination and that *recA* mutants are extraordinarily pleiotropic, affecting DNA repair, SOS mutagenesis, induction of λ prophage, cell division, and chromosomal segregation [26]. A decade after the discovery of the *recA* gene, the protein was first purified and a wide range of biochemical activities were reported, including DNA binding, *trans*activation of auto-proteolysis of the LexA and λ phage repressors, ATP hydrolysis, the ability to form filaments on DNA, and the capacity to exchange homologous DNA strands [26–28]. Owing to a rich history of elegant genetic and biochemical investigation of RecA, many of the biochemical and biophysical properties of RecA can be summarized in relation to the physiology of *E. coli* (Table 1).

RecA is a 38 kDa protein that is basally expressed at approximately 1,000 monomers per cell and forms right-handed nucleoprotein filaments on ssDNA and dsDNA, albeit

substantially more slowly on the latter. The *in vivo* concentration of RecA is ~1 μM in the absence of DNA damage; however, the concentration is increased approximately 10-fold during the SOS response [28]. RecA has a potent ATP hydrolysis activity in the presence of DNA, a result of a complex and dynamic cycle of filament assembly and disassembly on DNA [29]. The nucleoside triphosphate-binding site lies between adjacent monomers in a filament (Figure 1B), and the hydrolyzed state of the nucleotide strongly affects the stability of the RecA-DNA complex [30]. The ATP-bound form has a high affinity for DNA, while the ADP bound form has a lower affinity, and therefore, is both thermodynamically and kinetically less stable. The ATP-bound nucleoprotein filament serves as a surface catalyst for the search and capture of a homologous sequence of DNA, a process known as synapsis. Once a region of homology is found, the ssDNA strands on the homologous chromosomes are exchanged, producing heteroduplex DNA (i.e., each single-strand originated from a different chromosome). Although RecA must bind ATP to form an active filament, hydrolysis is not required to catalyze synapsis or exchange the DNA strands to produce heteroduplex DNA; however, hydrolysis is required for filament disassembly [29, 31]. The progression of these steps—synapsis and strand exchange—produces a complex, metastable intermediate of recombination known as a joint molecule that is then repaired through a combination of DNA synthesis, ligation and resolution [11].

The search for structural clues: insight from electron microscopy and X-ray crystallography

Although the first crystal structure of RecA was solved in 1992 by Story and Steitz [32], detailed structural information of the protein in complex with ATP and either ssDNA or dsDNA was remarkably elusive. For nearly three decades, most of the structural information on the RecA nucleoprotein filament came from high-resolution electron microscopy, which yielded tremendous insight into the morphological changes in the nucleoprotein structures during recombination (Figure 1C and 1D) [33–35]. Under physiological conditions, ssDNA folds into a multitude of structures that are stabilized by localized annealing of short regions of complementarity, called secondary structure, which inhibit RecA filament formation [23]. DNA secondary structure is largely overcome by the SSB protein, which binds tightly to ssDNA, effectively denaturing the secondary structure [36]. SSB-coated ssDNA is compacted relative to dsDNA and is morphologically distinct from RecA filaments because each tetramer of SSB wraps a segment of 30–70 nucleotides around itself (Figure 1D and Figure 2A) [35, 37]. When a molecule of ssDNA is completely and contiguously coated with RecA in the presence of ATPγS (adenosine 5′-[gamma-thio]triphosphate, a non-hydrolyzable ATP analog), the ssDNA is extended by 150–160% relative to a molecule of dsDNA of the same length, has one RecA monomer bound for approximately three nucleotides, and has a helical pitch of 9–10 nm (~6 monomers per turn) with a diameter of ~10 nm [38]. When the filament is strictly in the ADP-bound form, the pitch is reduced to 6-7 nm and the diameter expands to ~12 nm [39]. When RecA filaments are formed with ATP and hydrolysis results in ADP accumulation, RecA filaments are heterogeneous in pitch even along a single filament, consistent with a model of localized, cooperative ATP hydrolysis [38]. The failure of mutant RecA proteins (exemplified best by RecA142 [40]) to stretch ssDNA to the canonical filament length, despite their ability to bind to ssDNA and

hydrolyze ATP, is one of the strongest morphological factors that correlates with an inability to catalyze synapsis and DNA strand exchange, showing that stretching DNA is an essential component of DNA strand exchange [29, 41].

In 2008, Chen and Pavletich solved the crystal structure of both the presynaptic (ssDNA-bound) and post-synaptic (dsDNA-bound) RecA filaments, revealing surprising mechanistic details (Figure 1B and Figure 2A) [30]. While thought to be relatively uniform (or isotropic) throughout the complex, the DNA stretched inside the RecA filament is in fact segmented into nucleotide triplets that maintain approximately normal B-form dimensions (~3.2–3.5 Å for dsDNA & 3.5–4.2 Å for ssDNA) but are separated by an inter-nucleotide extension of ~7–8 Å (Figure 2B) [30]. Biochemical analysis had previously demonstrated that ATP binding is necessary and sufficient to promote DNA strand exchange, whereas hydrolysis of ATP is dispensable [31]. This observation substantiated the hypothesis that synapsis is "simply" a bimolecular collisional process between the RecA-ssDNA nucleoprotein filament and the homologous genomic locus (see subsequent "Search optimization through parallel processing: RecA finds homology by simultaneously sampling many transient and weak contacts"), rather than a process that is coupled to some putative ATP-dependent directed motion. The pre- and post-synaptic crystal structures demonstrated how this is possible [30]. By using the free energy of ATP-binding to stretch and untwist the ssDNA within the RecA filament, binding of the incoming dsDNA to the secondary site within the filament is stabilized through base pairing interactions only when a sufficient amount of homology is attained (as few as 8 bases). By maintaining these B-form triplet states (Figure 2B), the RecA filament creates a series of genetic 'words' that are held in register to one another throughout a contiguous filament allowing the nucleotide triplets to test for localized homology through normal, albeit transient, base pairing [30]. Each RecA monomer interacts with the entire triplet closest to itself in the structure, as well as with two more nucleotides, one from each of the preceding and following triplets. As a result, each nucleotide triplet is bound by three monomers, and each monomer interacts with five nucleotides, but because the first and fifth nucleotides are obligatorily 'shared' with adjacent monomers, the net stoichiometric ratio (or binding site size) is three nucleotides per RecA. In a similar fashion, the dsDNA in the postsynaptic filament is also stretched into nearly perfect Watson-Crick base-paired triplets [30].

The structural elucidation of the RecA-ssDNA complex was a substantial breakthrough that advanced our understanding of how DNA is recognized at the base-pair level during genetic recombination. However, how RecA filament assembly is regulated and whether the collisional, diffusion-driven model for homology search was a correct or feasible model remained open questions that traditional, ensemble biochemical approaches could only partially address.

The search for dynamics: single-molecule microscopy brings RecA assembly and search mechanisms into focus

Single-molecule techniques are a diverse and powerful toolbox with which to probe biological function with high spatiotemporal resolution. The first reported experiment to use

a single-molecule assay to probe the dynamics of RecA was performed in 1998 by John Marko's research group using a glass fiber attached to one end of a single dsDNA molecule and a bead held by a micropipette at fixed tension at the other end [42]. By pulling on the molecule with the micropipette, the authors could accurately measure the force required to stretch the DNA in the presence and absence of RecA by observing the displacement (i.e., bending) of the glass fiber, which acted as a force transducer. Polymerization of RecA onto the dsDNA resulted in extension of the DNA molecule, which was observed in real-time. Other groups reported similar experiments shortly thereafter, where a single DNA molecule was attached to the surface of a piezo-driven stage at one end and a bead at the other [43, 44]. The bead was held in place by an optical trap that was calibrated to measure either the change in force upon displacement, or the displacement under constant force. Like the previous study, the authors were able to measure extension of a single RecA-DNA molecule in real-time as well as the force-extension relationship of filaments; however, they expanded upon the previous work by investigating the role of various nucleotide cofactors in filament assembly and by monitoring filament disassembly. Despite the advance presented by single-molecule manipulation, the ability to distinguish between the nucleation and growth phase of filament assembly remained elusive—in part because the DNA molecule used was very long and multiple filaments could simultaneously nucleate and grow on a single molecule of DNA.

In 2006, two independent groups used very different single-molecule assays to differentiate between these kinetic parameters of nucleation and growth. The first method used Total Internal Reflection Fluorescence (TIRF) microscopy to detect FRET (Förster Resonance Energy Transfer) between a donor (Cy3) fluorophore that had been attached to the 5′-end of a dsDNA-ssDNA junction and an acceptor (Cy5) fluorophore that had been incorporated internally into a ssDNA overhang 13 nucleotides away (although the total length of the overhang was greater). In the absence of RecA, the short persistence length (and therefore high flexibility) of ssDNA results in high FRET efficiency, whereas formation of the filament results in an increased persistence length that diminishes the FRET efficiency approximately 4-fold or greater by physically separating the donor and acceptor molecules [45]. When RecA filaments were at steady-state in this assay, approximately four or five quantized FRET states of varying efficiency were observed depending on the length of the ssDNA substrate. Each quantized state was dependent on the number of RecA monomers bound between the donor and acceptor molecules and the authors were able to extract kinetic parameters from thousands of transitions between states, on hundreds of individual molecules, counting the frequency of addition and subtraction of monomers in time.

At essentially the same time, a complementary direct visualization assay was developed [46]. Individual clusters of RecA filaments could be observed directly while forming and growing on dsDNA, which was attached at one end to a polystyrene bead immobilized in an optical trap within a multi-channel, laminar flow cell [47]. An epifluorescence microscope was used to visualize fluorescent RecA bound to dsDNA tethered to the bead, which could be transferred between channels through the movement of the microscope stage. This approach is remarkably powerful due to the relative ease with which basic biochemical parameters like concentration, cofactor, or salt can be varied, while simultaneously maintaining tight control of the incubation time for each individual molecule. These

experiments provided three major insights into the mechanism of RecA filament assembly. First, that the nucleation kinetics of individual clusters forming on dsDNA follow a power dependence, where the rate increases proportionally to the 4th to 5th power of RecA concentration ($k_{nuc\text{-}dsDNA} \propto [RecA]^{\sim 4–5}$), indicating that approximately 4–5 monomers of RecA are the nucleation unit when binding to dsDNA. Second, that RecA filament growth is bidirectional on dsDNA, where the rate of filament growth at internal locations on the DNA is approximately 2-fold faster than filaments growing at the distal end of the DNA molecule. Third, that RecA filaments could be formed in the presence of ATP and stabilized by either ATPγS or ATP:Ca^{2+}, providing direct evidence that the nucleotide cofactor can be exchanged within the filament without disassembly.

Searching for control in the face of competition: Regulation of RecA filament assembly on SSB-coated ssDNA

Within its cellular context, RecA filaments must nucleate and grow on ssDNA in direct competition with SSB, which rapidly and contiguously sequesters ssDNA generated during replication and nucleolytic processing of damaged DNA. This competition serves two basic functions: first, it suppresses unwanted and potentially detrimental recombination. Second, SSB denatures localized DNA secondary structure that impedes contiguous RecA filament formation, which in turn reduces the efficiency of DNA strand exchange. Although methods for measuring the nucleation and growth of other filament forming proteins—actin and tubulin—have existed for several decades, the complexity of forming a filament on a linear template in the presence of a contiguous kinetic competitor made quantitative measurements of RecA assembly difficult. To accomplish this, long ssDNA molecules were attached to a surface within a flow cell and the nucleation and growth of filaments in the presence of SSB was directly imaged using TIRF microscopy, which enabled the unexpectedly long time-lapse experiments required to observe filament formation [24].

Several features of RecA filament assembly and regulation were revealed by this approach. First, the nucleation rate of RecA increased in proportion to the second power of the RecA concentration ($k_{nuc\text{-}ssDNA} \propto [RecA]^{\sim 2}$), indicating that a dimer of RecA—the minimal oligomer that can bind ATP—is the critical nucleus required for stable filament formation [24, 30]. This smaller critical nucleus required for binding to ssDNA (dimers) relative to dsDNA (tetramers to hexamers) is presumably due to the higher affinity of RecA for ssDNA, an interpretation that is supported by the observation that nucleation on dsDNA was extremely sensitive to low concentrations of salt, while nucleation on SSB-coated ssDNA was largely insensitive over this same range [24, 46, 48]. Second, RecA filament growth on ssDNA is rapid and bidirectional, albeit 50–60% faster in the 5′→3′ direction, and proceeds through monomer addition to the ends of the filaments coupled to a conformational change that is slower than the diffusion limit. Third, nucleoside triphosphate-binding activates a high-affinity state, acting as a conformational effector that increases the success of a nucleation event by stabilizing capture of the ssDNA transiently released from SSB during sliding or unwrapping. This is supported by the observation that RecA nucleation is faster when ATP is replaced with either ATPγS or dATP, which is hydrolyzed ~30% faster than ATP. The observed increase in nucleation rate is energetically linked to the increased affinity

of RecA for ssDNA when bound to the nucleotide or analog. Therefore, both the ability to hydrolyze ATP and the rate at which it is hydrolyzed are dispensable properties with respect to assembly, although they are essential properties for disassembly during DNA strand exchange. Fourth, both nucleation and growth of RecA are severely depressed at physiological pH (i.e., ~7.5). If RecA filament assembly were intrinsically at its maximum, then the cell would have little means by which to regulate its activity. This finding therefore rationalizes the necessity of mediator proteins—RecFOR and RecOR—to potentiate filament assembly by binding to SSB via RecO to trap ssDNA transiently released during SSB sliding or unwrapping (Figure 2C) [24]. Although RecOR and RecF had long been known to stimulate RecA filament formation, the mechanism of this stimulation remained unclear. In the single-molecule experiments described above, the addition of RecOR reduced the lag time for RecA filament nucleation and stimulated the rate of growth. The addition of RecF to the RecOR-stimulated reaction further reduced the lag time, but had no effect on growth. Hence these experiments were able to clearly distinguish, for the first time, the relative contributions of the so-called ‘recombination mediator proteins’ in the kinetic regulation and stimulation of RecA filament formation [24].

Although the intracellular pH is highly regulated in *E. coli,* it can be altered [49, 50]. For example, exposure to lipophilic acids such as propionic acid or sodium benzoate induce a stable reduction in the intracellular pH to ~6.5–6.8 that is toxic to cells if prolonged [49, 51, 52]. The expression profile of many bacterial genes, including the SOS response genes, is altered by environmental pH [53, 54]. In fact, lowering the intracellular pH was used to reactivate a pH-sensitive mutant of RecA (*RecA142*) that is inactive at normal physiological pH, but active at lower pH [53, 54]. Therefore, it is interesting to speculate that the alteration of filament assembly by solution conditions could be one of the many mechanisms by which bacteria respond to environmental stress, wherein the chemical environment (i.e., pH, osmolality, metabolites, etc.) might directly modulate the DNA damage response. Recent single molecule FRET experiments have further demonstrated that the pH-induced rate changes to filament growth and disassembly are moderate; however, the effect is amplified owing to both an acceleration of binding and a reduction in dissociation as pH decreases [55]. Similarly, it was proposed that the enhanced activity of RecA in the presence of dATP may be a reflection of an *in vivo* mechanism of biochemical regulation [56]. This interpretation is supported by the observation that the intracellular pool of dATP is increased during the bacterial DNA damage-induced SOS response due to a shift in ribonucleotide and deoxyribonucleotide concentrations mediated by the upregulation of ribonucleotide reductase (RNR) [57]. Nonetheless, the primary biological mechanism for filament regulation in *E. coli* is through the concerted action of RecFOR and RecOR, the latter of which binds to the auto-regulatory C-terminal tail of SSB [20, 21, 23, 24].

Search optimization through parallel processing: RecA finds homology by simultaneously sampling many transient and weak contacts

Once a RecA filament has formed on a resected chromosome, it must use the sequence information encoded within the single strand of DNA within the nucleoprotein filament to find a homologous dsDNA region and then exchange the individual strands [29]. This

biochemical process is unique to the RADα sub-class, but elements of this process are now recognized as being employed by CRISPR systems to uniquely target any sequence in dsDNA using a protein-bound ssRNA, rather than ssDNA, sequence. Although the search mechanism might seem simple and intuitive, the task of finding a homologous locus within the short time window required for DNA repair is challenging. This difficulty arises due to the vast excess of heterologous DNA within the genome—a challenge that scales exponentially with genome size, but may be overcome through organizing principles that establish chromatin territories within the eukaryotic nucleus [58, 59]. To access this information, RecA uses two DNA binding sites: a primary site that stretches the ssDNA within the nucleoprotein filament and a weaker affinity secondary DNA binding site that lies proximal to the ssDNA within the interior of the filament groove [60]. When homologous dsDNA comes in close contact with the stretched ssDNA within the groove of the filament, it can randomly sample homology through either canonical or non-canonical base pairing until a sufficient threshold (8 bases) stabilizes the interaction and permits DNA strand exchange products to be long-lived (~26 base pairs) [61]. DNA strand exchange is favored by virtue of the higher affinity displayed by the secondary site for the product of DNA strand exchange (the displaced ssDNA) rather than the substrate (the complementary dsDNA) [60]. This was clarified by the structural studies described earlier in this Review [30]; however, the mechanism by which RecA kinetically searches for its homologous target was, until recently, largely a black box. [29, 62][63–65]

By contrast, how a site-specific DNA binding protein can find a particular DNA sequence is relatively well understood, as many DNA-binding proteins (e.g., transcription factors and restriction enzymes) and specific structural motifs (zinc-fingers, TALENs, etc.) that enable a protein to bind to a specific sequence have been extensively characterized both kinetically and structurally [66–69]. These processes are largely driven by a diffusive, random walk (shown schematically in Figure 3A) until a collision between the protein and its target site results in a successful binding event [70–74]. How quickly a target can be found is dependent on the degrees of freedom of the random walk (i.e., one-dimensional sliding along a lattice, two-dimensional diffusion on a plane, or three-dimensional diffusion in space) and whether the diffusion is facilitated and/or directed (i.e., coupled to an accelerating factor such as an electrostatic potential) or confined (Figure 3B) [71, 72, 75].

More than three decades ago, Otto Berg, Robert Winter, and Peter von Hippel proposed three major diffusion-driven mechanisms that could explain the partitioning kinetics between specific, high-affinity binding to a target site and non-specific, low-affinity binding to random DNA (Figure 3C)[71–75]. These general search mechanisms can be thought of as a model for all protein-DNA interactions, with only a handful of special-case exceptions. First, in one-dimensional diffusion, a protein can slide along a DNA lattice (i.e., a DNA polymer composed of repetitive and identical units) by randomly stepping away from—and back towards—its initial binding site. Sliding is theoretically limited to sampling only a small segment of DNA because—on average—the protein would randomly step backwards as frequently as it steps forward, requiring long lifetimes for large mean squared displacements. Second, proteins can “hop” along DNA by iteratively binding and dissociating. Each release will be followed by three-dimensional diffusion; however, because DNA behaves as a rod over short distances, rather than a point source, the probability of rebinding to an adjacent

site close to the point of dissociation is statistically high. Therefore, this correlation between unbinding and rebinding results in a statistically probable “hop” within ~10–20 bp [70]. Alternatively, *in vivo*, because the genome is tightly compacted within a confined space [76–78], there is a high probability of a new binding event being far away when measured along the linear molecule (e.g., tens of kilobase pairs to several megabase pairs), but which is very close in three-dimensional space. Such an event is simply normal dissociation followed by rebinding (and therefore microscopically indistinguishable from ‘hopping’), but because of the dense coiled nature of DNA, the term “jumping” has been attributed to events associated with re-binding events that are very far away from the point of dissociation on the linear chromosome [79]. Finally, if a protein has at least two or more DNA-binding domains, then it can iteratively bind and release via one domain while remaining tethered to the DNA through another—a phenomenon called “inter-segmental transfer” or “looping”. This transfer can occur between segments that are close in the linear arrangement of DNA or between segments that are widely separated in the genome. In each of these models, the target site is short and rare, but is bound tightly and specifically when it is found and is often linked to a regulatory function such as binding to a regulatory metabolite or protein-partner [71].

The RecA-DNA complex is not a simple, compact entity; rather it is a long partially flexible filament that has a persistence length (i.e., a measure of polymer stiffness) of ~900 nm for the static ATPγS filament [44], which is approximately twenty times greater than dsDNA (45–50 nm [80, 81]) and several hundred times greater than ssDNA (0.6–3 nm [44, 82, 83]) (Table 1). So if diffusion of a RecA filament is severely restricted through physical confinement, how then does RecA find its homologous partner? Although it was over three decades ago that RecA was found to assimilate single-stranded DNA into homologous duplex DNA [84, 85], the details of the kinetic intermediates formed and the transient physical mechanics of the RecA-DNA complexes remained elusive until several groups recently used modern single-molecule assays to study the process. Historically, this was because – despite the existing familiarity with the concepts of facilitated diffusion – under the extant ensemble reaction conditions, the rate of homologous pairing is not limited by the rate of the homology search [29, 62]; therefore, it was not possible to elucidate the mechanism of the search process itself, despite clever attempts [63–65].

To detect DNA pairing in real-time using direct visualization, Forget and Kowalczykowski incubated fluorescent RecA-ssDNA nucleoprotein filaments with λ-phage DNA (48.5 kbp). In the initial attempts, the dsDNA was attached at one end to the surface of a flow cell, extended by solution flow to just under contour length, and then the opposite DNA end was attached to the surface [86]. When imaged using TIRF microscopy, the vast majority of the stretched molecules did not form stable complexes, which was perplexing because when the molecules were incubated together under identical conditions and subsequently attached to the surface, pairing efficiency was high. Interestingly, a subset of molecules that were either tethered only at one end, or proximally tethered in such a way that the ends of the DNA molecule were close to each other, demonstrated stable pairing when the reaction was incubated without flow (Figure 3D). These observations suggested that the three-dimensional conformation of the dsDNA target was instrumental in the homology search. To further test this hypothesis, a dual beam optical trap was used to micro-manipulate a single

molecule of dsDNA between two beads so as to vary the end-to-end distance in a controlled manner [86, 87]. When a λ dsDNA molecule was extended to beyond ~16 μm (the length at which the DNA was at its theoretical B-form length), homologous pairing was not observed, in agreement with the TIRF experiments; however, when the ends of the molecule were brought together so that the dsDNA could adopt a more randomly coiled polymer, the efficiency of RecA pairing increased monotonically as the end-to-end distance was decreased (Figure 3E). Furthermore, transient interactions at off-target, non-homologous sites were observed to induce DNA looping events that were short lived (< 10 seconds) and were released as the DNA molecule was extended though micromanipulation. These observations provided direct evidence that a single RecA-ssDNA nucleoprotein filament simultaneously samples different sites within the same dsDNA while the filament is searching for homology, and also that efficient searching occurred when the target dsDNA adopted a three-dimensional, random coil that allowed multiple contacts with the filament. This process was described as a “inter-segmental contact sampling”, wherein a single RecA filament of a theoretically unlimited length, makes many transient and weak interactions with the randomly coiled dsDNA containing the target site; both the polyvalent binding and the behavior of coiled DNA ensure that segments of the dsDNA target are statistically confined to, and quickly sampled within, the three dimensional space around the filament (Figure 3C) [86]. This inter-segmental contact sampling is analogous to parallel processing, where only the local concentration of the dsDNA target and the net length of the RecA nucleoprotein filament limit the number of interactions that can be simultaneously tested.

But are random three-dimensional collisions sufficient for the homology search? Ragunathan and Ha used single-molecule FRET to probe the dynamic fluctuations of RecA-nucleoprotein filaments on short, oligonucleotide-length filaments and dsDNA targets (Figure 3F) [88, 89]. They observed rapid fluctuations consistent with one-dimensional sliding of the filament along the dsDNA, with an average interaction time of 0.5–10 seconds and a sampling site size of 60–300 base pairs [89]. It is very likely that this rapid, one-dimensional sliding is another important aspect of homology sampling at each contact point along the RecA filament; classic analysis by Berg and von Hippel showed that the optimal search strategy employs a combination of three dimensional searching and one-dimensional sliding [66, 70, 71]. Ultimately, it will be essential to understand these reactions in the context of the cellular process. Live imaging of *E. coli* showed that RecA filaments span the length of dividing cells and that they can laterally aggregate into bundles [13]. *In vivo*, the homology search took approximately 50 minutes [13].

Recognition of DNA Sequence Homology: Testing Involves Concerted Sampling of Short Duplet or Triplet Matches

Recognition of DNA sequence homology by the RecA filament was detected in an elegant set of experiments where one end of a dsDNA molecule was tethered to the surface of a flow cell with opposite end attached to a magnetic bead, while a second transverse molecule of a RecA-ssDNA nucleoprotein filament was held in a dual-beam optical trap (Figure 4A) [90]. The RecA filament could be physically micro-manipulated, where it could be pushed against or dragged across the target dsDNA molecule. Transient homologous recognition events

could be measured due to the high sensitivity and time resolution of the experiment; these pairing events were only observed when the dsDNA molecule was supercoiled through the rotation of the paramagnetic trap. Supercoiling was long known to facilitate RecA pairing in D-loop assays (an assay wherein an ssDNA oligonucleotide coated with RecA is paired with supercoiled DNA), by stabilizing the paired heteroduplex product [85].

Although pairing intermediates of the RecA-ssDNA filament bound to the incoming dsDNA have not been crystalized, integration of RecA biochemistry, structural analysis, and molecular dynamics was used to generate a molecular model of the pairing process (Figure 4B). Elements of the model included both the position of positively charged residues within the nucleoprotein filament that are thought to comprise the second strand DNA binding site, also called site II (Figure 4C), the energetic costs of dsDNA extension upon binding to the filament, and the energy 'payback' when new Watson-Crick base pairs formed in a homologous product [91, 92]. A numerical simulation of the process offers a gratifying molecular perspective of the DNA pairing and recognition process (see Supplemental Video)

To measure the transient kinetic interactions between the RecA filament and dsDNA molecules containing varying degrees of heterologous sequences, a TIRF-based approach was used by the Greene lab. 'Curtains' of either Rad51- or RecA-coated ssDNA were tethered between nanofabricated barriers, and short fluorescently-labeled duplex oligonucleotides were incubated with the filaments (Figure 4D and Figure 4E). This approach allowed the authors to measure the kinetic lifetimes of the pairing intermediates, and thereby test the model for homology recognition described by Mara Prentiss and colleagues [93, 94]. Interestingly, when dsDNA substrates containing seven nucleotides or less of microhomology were used, the authors saw only transiently paired complexes; however, adding an eighth homologous nucleotide resulted in a dramatic increase in both pairing efficiency and binding energy. Adding an additional ninth nucleotide of homology contributed to another dramatic increase in binding energy, constituting a transition from homology sampling to pairing (Figure 4F) [94]. Interestingly, the stability of paired intermediates increases in three nucleotide increments, and this energetic profile is conserved between RecA, Rad51 and Dmc1 [95]. Therefore, RecA also discriminates against heterologous sequences by effectively ignoring homologous sequences shorter than seven nucleotides, using a fundamental eight-nucleotide homology test during the molecular search process [94]. Together, both a reduction in dimensionality (through intersegment contact sampling) and a reduction in complexity (through microhomology sampling), provide an elegant and satisfying solution to a classical and fundamental problem in molecular search theory.

Whether these processes are sufficient to explain the *in vivo* homology search in larger eukaryotic cells remains unclear. Although only 12, 13 or 17 nucleotides are required to define a unique sequence in the *E. coli*, *S. cerevisiae*, and human genomes, respectively [94], mammalian cells have nuclear volumes approximately 1000 times greater than a yeast nucleus or a bacterial cell [58, 59, 96]. Complicating the problem of an expanded search volume and larger genome is that eukaryotic DNA is densely packaged into chromatin that is tightly regulated and organized into stable 'territories' within a single cell's nucleus. In yeast, *in vivo* labeling experiments have demonstrated centromeres and telomeres only

explore about 3–5% of the nuclear volume, indicating that once the chromosomes are packaged into the nucleus, they are relatively well confined [96]. This in itself would not be a problem, but FISH and chromatin capture experiments have demonstrated that homologous chromosomes can occupy territories that are far apart from one another, suggesting that the homology search process could be greatly impeded by physical distance [77, 78]. This problem is mitigated after DNA replication due to sister chromatid-cohesion. When exposed to DNA damage such as ionizing radiation or inducible endonucleolytic dsDNA breaks, the motility of eukaryotic chromosomes increases; however, the extent of motility seems to vary greatly between organisms and cell cycle, with some loci inherently more mobile than others [96_99] [100_102]. These *in vivo* chromosome mobility assays are dependent on ploidy, where and how the loci is labeled, and length of observation; nonetheless, it is clear that cells exposed to DNA damage exhibit an increase in chromosome mobility, in principle, enabling loci to explore larger volumes. The precise factors that contribute to this ‘nuclear jiggling’ and how these dynamics contribute to chromosome pairing remain controversial and are still being defined. Defining and characterizing the factors that are required to accelerate and facilitate the eukaryotic homology search, either through conformational changes in global or local chromatin structure or directed, motor-dependent motion, will be important steps in understanding how the recombination machinery finds its homologous target (see Outstanding Questions Box). We expect that advances in super-resolution microscopy and single-molecule manipulation and detection methods will be essential in future studies probing the dynamics of the search process in living cells, and that those findings will impact not only our understanding of recombination and DNA repair, but will also directly inform the development of template-directed gene therapies using RNA-guided CRISPRs.

OUTSTANDING QUESTIONS BOX

- How do other mediator proteins and regulatory motor proteins alter the dynamic behavior (assembly or disassembly) of the RecA/RAD51 filament?
- What is the role of SMC (Structural Maintenance of Chromosomes) proteins in recombination-dependent homology searches?
- How is the homology search affected by the length of the ssDNA within the filament: at some characteristic length, will the search be hindered by more off-target sampling?
- How does RAD51 search for homology in the context of chromatin and heterochromatin?
- How does the physical organization of the genome influences search and recombination bias?
- How are repair outcomes influenced by differential euchromatin and heterochromatin states?
- To what extent do histone chaperones and chromatin-remodeling enzymes affect DNA pairing and strand-exchange by RAD51?

- What other chromatin modifications and remodelers are required for recombination?
- How does chromatid cohesion impose sister bias and is this merely a consequence of physical proximity, or are other biochemical mechanisms at play?
- How do chromatin-associated RNAs (e.g., R-loops, non-coding RNA, etc.) affect recombination?

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Research in the laboratory was supported by grants from the National Institutes of Health, GM41347, GM62653, GM64745, and CA154920, and from the Department of Defense, CDMRP, Breast Cancer Research Program, W81XWH-13-1-0322 to SCK and National Institutes of Health Training Grant T32 GM007377 to JCB. We are grateful to the members of my laboratory for their careful reading of a draft of this manuscript.

References

1. Lin Z, et al. Origins and evolution of the recA/RAD51 gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. Proc Natl Acad Sci U S A. 2006; 103:10328–10333. [PubMed: 16798872]
2. Rocha EP, et al. Comparative and evolutionary analysis of the bacterial homologous recombination systems. PLoS Genet. 2005; 1:e15. [PubMed: 16132081]
3. Wu D, et al. Stalking the fourth domain in metagenomic data: searching for, discovering, and interpreting novel, deep branches in marker gene phylogenetic trees. PLoS ONE. 2011; 6:e18011. [PubMed: 21437252]
4. Chintapalli SV, et al. Reevaluation of the evolutionary events within recA/RAD51 phylogeny. BMC Genomics. 2013; 14:240. [PubMed: 23574621]
5. Bianco PR, et al. DNA strand exchange proteins: A biochemical and physical comparison. Front Biosci. 1998; 3:D570–603. [PubMed: 9632377]
6. Brown MS, Bishop DK. DNA strand exchange and RecA homologs in meiosis. Cold Spring Harb Perspect Biol. 2015; 7:a016659. [PubMed: 25475089]
7. Cooper DL, Lovett ST. Recombinational branch migration by the RadA/Sms paralog of RecA in Escherichia coli. eLife. 2016; 5
8. Zelensky A, et al. Mediators of homologous DNA pairing. Cold Spring Harb Perspect Biol. 2014; 6:a016451. [PubMed: 25301930]
9. Liu J, et al. Rad51 paralogues Rad55-Rad57 balance the antirecombinase Srs2 in Rad51 filament formation. Nature. 2011; 479:245–248. [PubMed: 22020281]
10. Prakash R, et al. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. Cold Spring Harb Perspect Biol. 2015; 7:a016600. [PubMed: 25833843]
11. Kowalczykowski SC. An Overview of the Molecular Mechanisms of Recombinational DNA Repair. Cold Spring Harb Perspect Biol. 2015; 7
12. Bell JC, Kowalczykowski SC. Mechanics and Single-Molecule Interrogation of DNA Recombination. Annu Rev Biochem. 2016; 85
13. Lesterlin C, et al. RecA bundles mediate homology pairing between distant sisters during DNA break repair. Nature. 2014; 506:249–253. [PubMed: 24362571]
14. Dillingham MS, Kowalczykowski SC. RecBCD enzyme and the repair of double-stranded DNA breaks. Microbiol Mol Biol Rev. 2008; 72:642–671. [PubMed: 19052323]

15. Cockram CA, et al. Quantitative genomic analysis of RecA protein binding during DNA double-strand break repair reveals RecBCD action in vivo. Proc Natl Acad Sci U S A. 2015; 112:E4735–4742. [PubMed: 26261330]
16. Levy A, et al. CRISPR adaptation biases explain preference for acquisition of foreign DNA. Nature. 2015; 520:505–510. [PubMed: 25874675]
17. Kogoma T. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. Microbiol Mol Biol Rev. 1997; 61:212–238. [PubMed: 9184011]
18. Morimatsu K, Kowalczykowski SC. RecQ helicase and RecJ nuclease provide complementary functions to resect DNA for homologous recombination. Proc Natl Acad Sci U S A. 2014; 111:E5133–5142. [PubMed: 25411316]
19. Handa N, et al. Reconstitution of initial steps of dsDNA break repair by the RecF pathway of *E. coli*. Genes Dev. 2009; 23:1234–1245. [PubMed: 19451222]
20. Umezu K, Kolodner RD. Protein interactions in genetic recombination in Escherichia coli. Interactions involving RecO and RecR overcome the inhibition of RecA by single-stranded DNA-binding protein. J Biol Chem. 1994; 269:30005–30013. [PubMed: 7962001]
21. Morimatsu K, Kowalczykowski SC. RecFOR proteins load RecA protein onto gapped DNA to accelerate DNA strand exchange: a universal step of recombinational repair. Mol Cell. 2003; 11:1337–1347. [PubMed: 12769856]
22. Morimatsu K, et al. RecFOR proteins target RecA protein to a DNA gap with either DNA or RNA at the 5′ terminus: implication for repair of stalled replication forks. J Biol Chem. 2012; 287:35621–35630. [PubMed: 22902627]
23. Bell JC, et al. Imaging and energetics of single SSB-ssDNA molecules reveal intramolecular condensation and insight into RecOR function. eLife. 2015; 4:e08646. [PubMed: 26381353]
24. Bell JC, et al. Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. Nature. 2012; 491:274–278. [PubMed: 23103864]
25. Clark AJ, Margulies AD. Isolation and Characterization of Recombination-Deficient Mutants of Escherichia Coli K12. Proc Natl Acad Sci U S A. 1965; 53:451–459. [PubMed: 14294081]
26. Kowalczykowski SC, et al. Biochemistry of homologous recombination in *Escherichia coli*. Microbiol Rev. 1994; 58:401–465. [PubMed: 7968921]
27. McEntee K, et al. Identification and radiochemical purification of the *recA* protein of *Escherichia coli* K-12. Proc Natl Acad Sci U S A. 1976; 73:3979–3983. [PubMed: 792880]
28. Kuzminov A. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage λ. Microbiol Mol Biol Rev. 1999; 63:751–813. [PubMed: 10585965]
29. Kowalczykowski SC. Biochemistry of genetic recombination: Energetics and mechanism of DNA strand exchange. Annu Rev Biophys Biophys Chem. 1991; 20:539–575. [PubMed: 1831022]
30. Chen Z, et al. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. Nature. 2008; 453:489–484. [PubMed: 18497818]
31. Menetski JP, et al. Stable DNA heteroduplex formation catalyzed by the *Escherichia coli* RecA protein in the absence of ATP hydrolysis. Proc Natl Acad Sci U S A. 1990; 87:21–25. [PubMed: 2404275]
32. Story RM, et al. The structure of the *E. coli* recA protein monomer and polymer. Nature. 1992; 355:318–325. [PubMed: 1731246]
33. Egelman EH, Yu X. The location of DNA in RecA-DNA helical filaments. Science. 1989; 245:404–407. [PubMed: 2667137]
34. Stasiak A, Egelman EH. Structure and Dynamics of recA Protein-DNA Complexes as Determined by Image Analysis of Electron Micrographs. Biophys J. 1986; 49:5–7. [PubMed: 19431645]
35. Griffith JD, et al. Visualization of SSB-ssDNA complexes active in the assembly of stable RecA-DNA filaments. Cold Spring Harb Symp Quant Biol. 1984; 49:553–559. [PubMed: 6397310]
36. Kowalczykowski SC, Krupp RA. Effects of *Escherichia coli* SSB protein on the single-stranded DNA-dependent ATPase activity of *Escherichia coli* RecA protein. Evidence that SSB protein facilitates the binding of RecA protein to regions of secondary structure within single-stranded DNA. J Mol Biol. 1987; 193:97–113. [PubMed: 2953903]

37. Raghunathan S, et al. Structure of the DNA binding domain of E. coli SSB bound to ssDNA. Nat Struct Biol. 2000; 7:648–652. [PubMed: 10932248]
38. Egelman EH, Stasiak A. Structure of helical RecA-DNA complexes. Complexes formed in the presence of ATP-gamma-S or ATP. J Mol Biol. 1986; 191:677–697. [PubMed: 2949085]
39. Heuser J, Griffith J. Visualization of RecA protein and its complexes with DNA by quick-freeze/deep-etch electron microscopy. J Mol Biol. 1989; 210:473–484. [PubMed: 2693735]
40. Kowalczykowski SC, et al. Biochemical events essential to the recombination activity of *Escherichia coli* RecA protein. I Properties of the mutant RecA142 protein. J Mol Biol. 1989; 207:719–733. [PubMed: 2527303]
41. Kowalczykowski SC. Biochemical and biological function of *Escherichia coli* RecA protein: behavior of mutant RecA proteins. Biochimie. 1991; 73:289–304. [PubMed: 1883888]
42. Léger JF, et al. RecA binding to a single double-stranded DNA molecule: a possible role of DNA conformational fluctuations. Proc Natl Acad Sci U S A. 1998; 95:12295–12299. [PubMed: 9770480]
43. Shivashankar GV, et al. RecA polymerization on double-stranded DNA by using single-molecule manipulation: the role of ATP hydrolysis. Proc Natl Acad Sci U S A. 1999; 96:7916–7921. [PubMed: 10393922]
44. Hegner M, et al. Polymerization and mechanical properties of single RecA-DNA filaments. Proc Natl Acad Sci U S A. 1999; 96:10109–10114. [PubMed: 10468570]
45. Joo C, et al. Real-time observation of RecA filament dynamics with single monomer resolution. Cell. 2006; 126:515–527. [PubMed: 16901785]
46. Galletto R, et al. Direct observation of individual RecA filaments assembling on single DNA molecules. Nature. 2006; 443:875–878. [PubMed: 16988658]
47. Amitani I, et al. Watching individual proteins acting on single molecules of DNA. Methods Enzymol. 2010; 472:261–291. [PubMed: 20580968]
48. Handa N, et al. Single molecule analysis of a red fluorescent RecA protein reveals a defect in nucleoprotein filament nucleation that relates to its reduced biological functions. J Biol Chem. 2009; 284:18664–18673. [PubMed: 19419960]
49. Kitko RD, et al. Osmolytes contribute to pH homeostasis of Escherichia coli. PLoS ONE. 2010; 5:e10078. [PubMed: 20386696]
50. Wilks JC, Slonczewski JL. pH of the cytoplasm and periplasm of Escherichia coli: rapid measurement by green fluorescent protein fluorimetry. J Bacteriol. 2007; 189:5601–5607. [PubMed: 17545292]
51. Repaske DR, Adler J. Change in intracellular pH of Escherichia coli mediates the chemotactic response to certain attractants and repellents. J Bacteriol. 1981; 145:1196–1208. [PubMed: 7009571]
52. Dri AM, Moreau PL. Control of the LexA regulon by pH: evidence for a reversible inactivation of the LexA repressor during the growth cycle of Escherichia coli. Mol Microbiol. 1994; 12:621–629. [PubMed: 7934886]
53. Zaitsev EN, Kowalczykowski SC. Enhanced monomer-monomer interactions can suppress the recombination deficiency of the recA142 allele. Mol Microbiol. 1999; 34:1–9. [PubMed: 10540281]
54. Maurer LM, et al. pH regulates genes for flagellar motility, catabolism, and oxidative stress in Escherichia coli K-12. J Bacteriol. 2005; 187:304–319. [PubMed: 15601715]
55. Kim SH, et al. Dynamic growth and shrinkage govern the pH dependence of RecA filament stability. PLoS ONE. 2015; 10:e0115611. [PubMed: 25608006]
56. Menetski JP, Kowalczykowski SC. Enhancement of *Escherichia coli* recA protein enzymatic function by dATP. Biochemistry. 1989; 28:5871–5881. [PubMed: 2673351]
57. Barbe J, et al. ATP hydrolysis during SOS induction in Escherichia coli. J Bacteriol. 1986; 167:1055–1057. [PubMed: 3528124]
58. Miné-Hattab J, Rothstein R. Increased chromosome mobility facilitates homology search during recombination. Nat Cell Biol. 2012; 14:510–517. [PubMed: 22484485]

59. Dion V, et al. Increased mobility of double-strand breaks requires Mec1, Rad9 and the homologous recombination machinery. Nat Cell Biol. 2012; 14:502–509. [PubMed: 22484486]
60. Mazin AV, Kowalczykowski SC. The specificity of the secondary DNA binding site of RecA protein defines its role in DNA strand exchange. Proc Natl Acad Sci U S A. 1996; 93:10673–10678. [PubMed: 8855238]
61. Hsieh P, et al. The synapsis event in the homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of DNA. Proc Natl Acad Sci U S A. 1992; 89:6492–6496. [PubMed: 1631148]
62. Julin DA, et al. On the mechanism of pairing of single- and double-stranded DNA molecules by the recA and single-stranded DNA-binding proteins of *Escherichia coli*. J Biol Chem. 1986; 261:1025–1030. [PubMed: 3511041]
63. Gonda DK, Radding CM. By searching processively RecA protein pairs DNA molecules that share a limited stretch of homology. Cell. 1983; 34:647–654. [PubMed: 6616624]
64. Gonda DK, Radding CM. The mechanism of the search for homology promoted by recA protein. Facilitated diffusion within nucleoprotein networks. J Biol Chem. 1986; 261:13087–13096. [PubMed: 3020024]
65. Adzuma K. No sliding during homology search by RecA protein. J Biol Chem. 1998; 273:31565–31573. [PubMed: 9813072]
66. Berg OG, et al. How do genome-regulatory proteins locate their DNA target sites? Trends Biochem Sci. 1982; 7:52–55.
67. Lavery R. Recognizing DNA. Q Rev Biophys. 2005; 38:339–344. [PubMed: 16515738]
68. Record, MTJ.; Spolar, RS. Some thermodynamic principles of nonspecific and site-specific protein-DNA interactions. In: Revzin, A., editor. The Biology of Nonspecific DNA Protein Interactions. CRC Press; 1990. p. 33-69.
69. Sauer RT, et al. Lambda repressor: a model system for understanding protein-DNA interactions and protein stability. Adv Protein Chem. 1990; 40:1–61. [PubMed: 2195849]
70. Berg, OG. The role of nonspecific interactions for gene-regulatory proteins in their search for specific target sites. In: Revzin, A., editor. The Biology of Nonspecific DNA Protein Interactions. CRC Press; 1990. p. 71-85.
71. von Hippel PH, Berg OG. Facilitated target location in biological systems. J Biol Chem. 1989; 264:675–678. [PubMed: 2642903]
72. Berg OG, et al. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1 Models and theory. Biochemistry. 1981; 20:6929–6948. [PubMed: 7317363]
73. Winter RB, et al. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3 The Escherichia coli lac repressor--operator interaction: kinetic measurements and conclusions. Biochemistry. 1981; 20:6961–6977. [PubMed: 7032584]
74. Winter RB, von Hippel PH. Diffusion-driven mechanisms of protein translocation on nucleic acids. 2 The Escherichia coli repressor--operator interaction: equilibrium measurements. Biochemistry. 1981; 20:6948–6960. [PubMed: 6274381]
75. Berg OG, von Hippel PH. Diffusion-controlled macromolecular interactions. Annu Rev Biophys Biophys Chem. 1985; 14:131–160. [PubMed: 3890878]
76. Hadizadeh Yazdi N, et al. Variation of the folding and dynamics of the Escherichia coli chromosome with growth conditions. Mol Microbiol. 2012; 86:1318–1333. [PubMed: 23078205]
77. Cremer T, Cremer M. Chromosome territories. Cold Spring Harb Perspect Biol. 2010; 2:a003889. [PubMed: 20300217]
78. Lieberman-Aiden E, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009; 326:289–293. [PubMed: 19815776]
79. Halford SE, Marko JF. How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res. 2004; 32:3040–3052. [PubMed: 15178741]
80. Smith SB, et al. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. Science. 1996; 271:795–799. [PubMed: 8628994]
81. Baumann CG, et al. Ionic effects on the elasticity of single DNA molecules. Proc Natl Acad Sci U S A. 1997; 94:6185–6190. [PubMed: 9177192]

82. Murphy MC, et al. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. Biophys J. 2004; 86:2530–2537. [PubMed: 15041689]
83. Saleh OA, et al. Nonlinear low-force elasticity of single-stranded DNA molecules. Phys Rev Lett. 2009; 102:068301. [PubMed: 19257640]
84. McEntee K, et al. Initiation of general recombination catalyzed in vitro by the recA protein of Escherichia coli. Proc Natl Acad Sci U S A. 1979; 76:2615–2619. [PubMed: 379861]
85. Shibata T, et al. Purified Escherichia coli recA protein catalyzes homologous pairing of superhelical DNA and single-stranded fragments. Proc Natl Acad Sci U S A. 1979; 76:1638–1642. [PubMed: 156361]
86. Forget AL, Kowalczykowski SC. Single-molecule imaging of DNA pairing by RecA reveals a three-dimensional homology search. Nature. 2012; 482:423–427. [PubMed: 22318518]
87. Forget AL, et al. Exploring protein-DNA interactions in 3D using in situ construction, manipulation and visualization of individual DNA dumbbells with optical traps, microfluidics and fluorescence microscopy. Nat Protoc. 2013; 8:525–538. [PubMed: 23411634]
88. Ragunathan K, et al. Real-time observation of strand exchange reaction with high spatiotemporal resolution. Structure. 2011; 19:1064–1073. [PubMed: 21827943]
89. Ragunathan K, et al. RecA filament sliding on DNA facilitates homology search. eLife. 2012; 1:e00067. [PubMed: 23240082]
90. De Vlaminck I, et al. Mechanism of homology recognition in DNA recombination from dual-molecule experiments. Mol Cell. 2012; 46:616–624. [PubMed: 22560720]
91. Yang D, et al. Integrating multi-scale data on homologous recombination into a new recognition mechanism based on simulations of the RecA-ssDNA/dsDNA structure. Nucleic Acids Res. 2015; 43:10251–10263. [PubMed: 26384422]
92. Prentiss M, et al. Structure/function relationships in RecA protein-mediated homology recognition and strand exchange. Crit Rev Biochem Mol Biol. 2015; 50:453–476. [PubMed: 26459995]
93. Jiang L, Prentiss M. RecA-mediated sequence homology recognition as an example of how searching speed in self-assembly systems can be optimized by balancing entropic and enthalpic barriers. Phys Rev E Stat Nonlin Soft Matter Phys. 2014; 90:022704. [PubMed: 25215755]
94. Qi Z, et al. DNA sequence alignment by microhomology sampling during homologous recombination. Cell. 2015; 160:856–869. [PubMed: 25684365]
95. Lee JY, et al. DNA RECOMBINATION. Base triplet stepping by the Rad51/RecA family of recombinases. Science. 2015; 349:977–981. [PubMed: 26315438]
96. Miné-Hattab J, Rothstein R. DNA in motion during double-strand break repair. Trends Cell Biol. 2013; 23:529–536. [PubMed: 23867212]
97. Cho NW, et al. Interchromosomal homology searches drive directional ALT telomere movement and synapsis. Cell. 2014; 159:108–121. [PubMed: 25259924]
98. Dion V, et al. Cohesin and the nucleolus constrain the mobility of spontaneous repair foci. EMBO Rep. 2013; 14:984–991. [PubMed: 24018421]
99. Dion V, Gasser SM. Chromatin movement in the maintenance of genome stability. Cell. 2013; 152:1355–1364. [PubMed: 23498942]
100. Lee CS, et al. Chromosome position determines the success of double-strand break repair. Proc Natl Acad Sci U S A. 2016; 113:E146–154. [PubMed: 26715752]
101. Nakajima Y, Haber JE. Chromosomes at loose ends. Nat Cell Biol. 2016; 18:257–259. [PubMed: 26911910]
102. Strecker J, et al. DNA damage signalling targets the kinetochore to promote chromatin mobility. Nat Cell Biol. 2016; 18:281–290. [PubMed: 26829389]
103. Kubitschek HE, Friske JA. Determination of bacterial cell volume with the Coulter Counter. J Bacteriol. 1986; 168:1466–1467. [PubMed: 3536882]
104. Blattner FR, et al. The complete genome sequence of Escherichia coli K-12. Science. 1997; 277:1453–1462. [PubMed: 9278503]
105. Reyes-Lamothe R, et al. Stoichiometry and architecture of active DNA replication machinery in Escherichia coli. Science. 2010; 328:498–501. [PubMed: 20413500]
106. Kornberg, A.; Baker, TA. DNA Replication. W. H. Freeman and Co; 1992.

107. Iyer VN, Rupp WD. Usefulness of benzoylated naphthoylated DEAE-cellulose to distinguish and fractionate double-stranded DNA bearing different extents of single-stranded regions. Biochim Biophys Acta. 1971; 228:117–126. [PubMed: 4926026]
108. Wang TC, Chen SH. Similar-sized daughter-strand gaps are produced in the leading and lagging strands of DNA in UV-irradiated E. coli uvrA cells. Biochem Biophys Res Commun. 1992; 184:1496–1503. [PubMed: 1590807]
109. Pennington JM, Rosenberg SM. Spontaneous DNA breakage in single living Escherichia coli cells. Nat Genet. 2007; 39:797–802. [PubMed: 17529976]
110. Krasin F, Hutchinson F. Repair of DNA double-strand breaks in Escherichia coli, which requires recA function and the presence of a duplicate genome. J Mol Biol. 1977; 116:81–98. [PubMed: 338918]
111. Sinden RR, Cole RS. Repair of cross-linked DNA and survival of Escherichia coli treated with psoralen and light: effects of mutations influencing genetic recombination and DNA metabolism. J Bacteriol. 1978; 136:538–547. [PubMed: 361714]
112. Lohman TM, Ferrari ME. Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities. Annu Rev Biochem. 1994; 63:527–570. [PubMed: 7979247]
113. Kowalczykowski SC. Interaction of recA protein with a photoaffinity analogue of ATP, 8-azido-ATP: determination of nucleotide cofactor binding parameters and of the relationship between ATP binding and ATP hydrolysis. Biochemistry. 1986; 25:5872–5881. [PubMed: 3539181]
114. Weinstock GM, et al. Hydrolysis of nucleoside triphosphates catalyzed by the recA protein of Escherichia coli. Steady state kinetic analysis of ATP hydrolysis. J Biol Chem. 1981; 256:8845–8849. [PubMed: 6455429]
115. Menetski JP, Kowalczykowski SC. Interaction of recA protein with single-stranded DNA. Quantitative aspects of binding affinity modulation by nucleotide cofactors. J Mol Biol. 1985; 181:281–295. [PubMed: 3981638]
116. Zaitsev EN, Kowalczykowski SC. Binding of double-stranded DNA by *Escherichia coli* RecA protein monitored by a fluorescent dye displacement assay. Nucleic Acids Res. 1998; 26:650–654. [PubMed: 9421529]
117. Perrin, J. Les Atomes. Constable & Company Ltd; 1916.
118. Neumann FR, et al. Targeted INO80 enhances subnuclear chromatin movement and ectopic homologous recombination. Genes Dev. 2012; 26:369–383. [PubMed: 22345518]

Trends Box

- Single-molecule methods have yielded new insight into how RecA filaments form and find homologous DNA.
- RecA nucleation and growth on SSB-coated ssDNA is kinetically regulated by the RecFOR and RecOR complexes, which both microscopically and macroscopically alter the SSB-ssDNA nucleoprotein fiber
- RecA finds homology through many random, weak and transient interactions made in parallel and ignores very short regions of microhomology resulting in a reduction of both dimensionality and complexity.

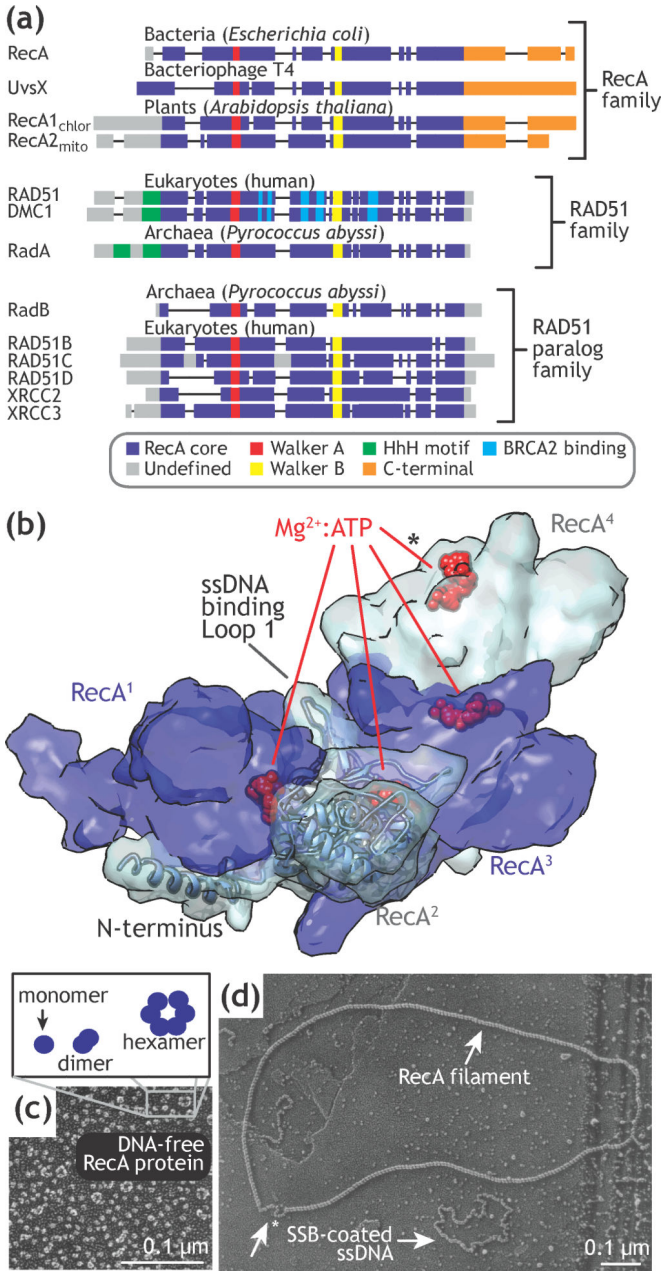

Figure 1. RecA conservation and structure

A) Conserved domains of the RecA, RAD51, and RAD51 paralog families. **B)** Segment of RecA filament showing the Mg:ATP binding site at the monomer-monomer interface. The asterisk indicates a half-site; for simplicity, only one of ssDNA binding loops per monomer is shown. *Adapted from* PDB 3CMW (RecA) [30]. **C)** Electron microscopy of DNA-free RecA showing heterogeneous oligomerization states of RecA. **D)** Electron microscopy of RecA filament-formation on SSB-coated circular ssDNA. Note the dramatic extension of the RecA filament relative to the compaction of the SSB-coated ssDNA. **C) and D)** *adapted from* [39].

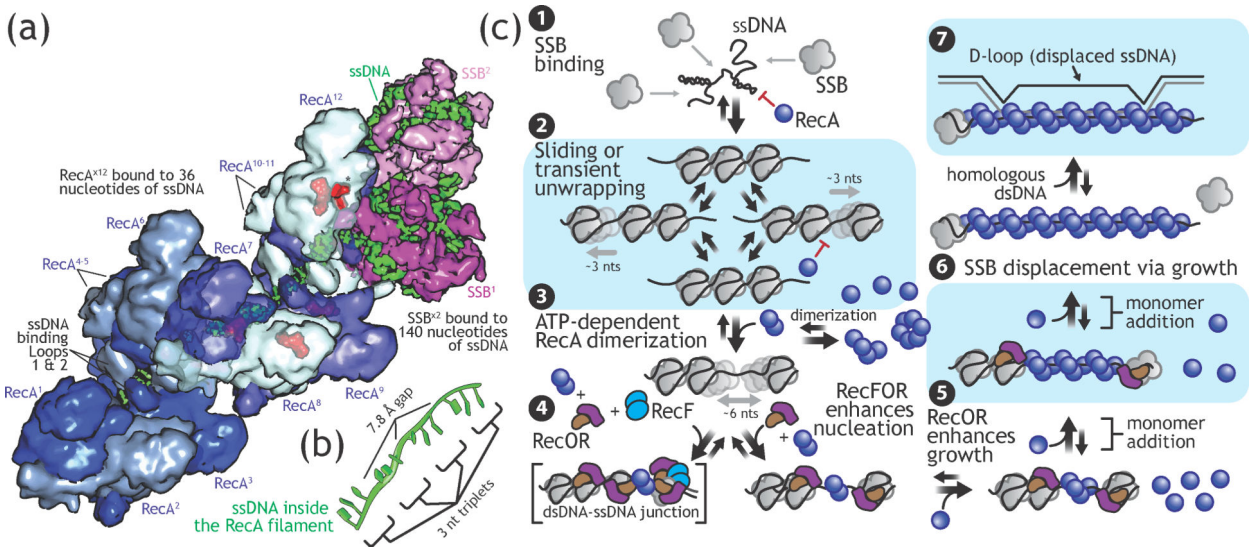

Figure 2. Mechanism of RecA nucleation and growth on SSB-coated ssDNA

A) Structural model of twelve-monomers of RecA bound to 36 nucleotides of ssDNA adjacent to two tetramers of SSB bound to 140 nucleotides of ssDNA demonstrating the vast differences in site size and accessibility of the ssDNA bound to each protein. *Adapted from* PDB 3CMW (RecA) and PDB 1EYG (SSB) [30, 37]. Mg:ATP (red) is visible at the monomer-monomer interface for $RecA^{7}$-$RecA^{12}$. **B)** Single-stranded DNA within the filament is stretched into nucleotide triplets that can maintain Watson-Crick interactions during homologous pairing. **C) 1,** SSB binds rapidly to ssDNA, removing secondary structure that impedes RecA-mediated DNA strand exchange. **2,** SSB kinetically blocks RecA filament formation. **3,** Nucleation of RecA onto rare and transient microscopic gaps requires ATP-dependent dimerization, making nucleation infrequent. **4,** RecOR binds to the C-terminal tails of SSB, microscopically altering the SSB-ssDNA complex, but not displacing it. This 'stoichiometric remodeling' creates microscopic gaps that are large enough and long-lived enough for RecA to stably bind, enhancing nucleation. Similarly, RecF further, in coordination with RecOR, enhances RecA nucleation at dsDNA-ssDNA junctions. **5,** RecA filament growth through monomer addition is impeded by SSB, though less so than nucleation; however, in the presence of RecOR, RecA filament growth is stimulated ~3-fold. **6,** The RecA filament grows monotonically and displaces SSB from ssDNA. *Adapted from* [24]. **7**, The RecA-ssDNA filament catalyzes pairing and strand exchange with a homologous dsDNA molecule, resulting in an intermediate, three-stranded molecule called a D-loop (or 'displacement loop'); SSB binds to the displaced strand to stabilize the D-loop (not shown).

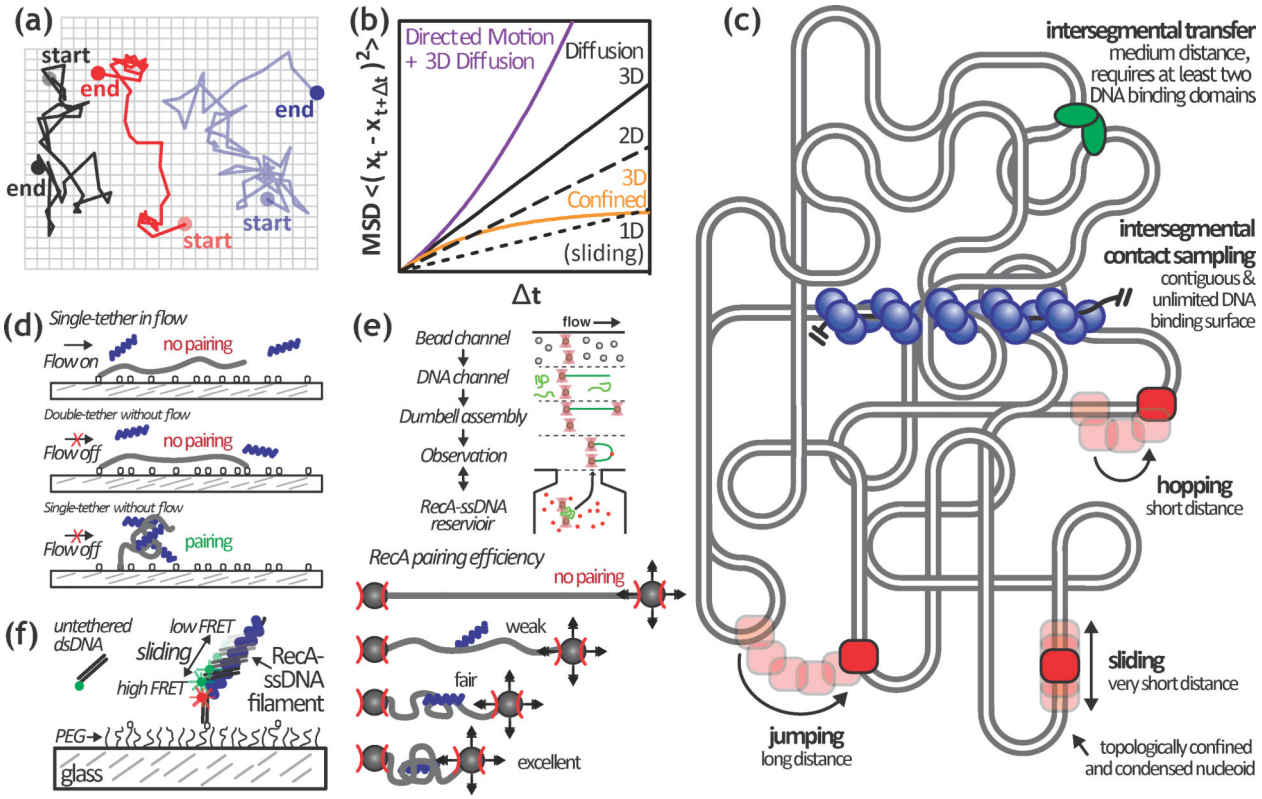

Figure 3. Diffusion-driven mechanism of RecA-mediated homology search

A) Diagram of three particles demonstrating random walk diffusion represented on a two-dimensional plane. *Adapted from reference* [117]. **B)** Plot showing the mean squared displacement as a function of time for one-, two-, and three-dimensional diffusion compared to three-dimensional diffusion in a confined space without or with directed motion *Adapted from* [99, 118]. **C)** Cartoon depicting different modes by which proteins find their targets by sliding, hopping, jumping, intersegmental transfer, and intersegmental contact sampling [70–72, 75]. Single-molecule methods used to measure RecA-mediated homology search: **D)** Total internal reflection fluorescence (TIRF) microscopy used to visualize ssDNA-RecA filaments pairing with λ phage DNA [86]. **E)** DNA micromanipulation experiments demonstrating RecA pairing efficiency increases as the DNA is allowed to adopt three-dimensional, random-coil configurations [86]. **F)** Single-molecule Förster Resonance Energy Transfer (FRET) experiments used to demonstrate microscale sliding of RecA filaments [89].

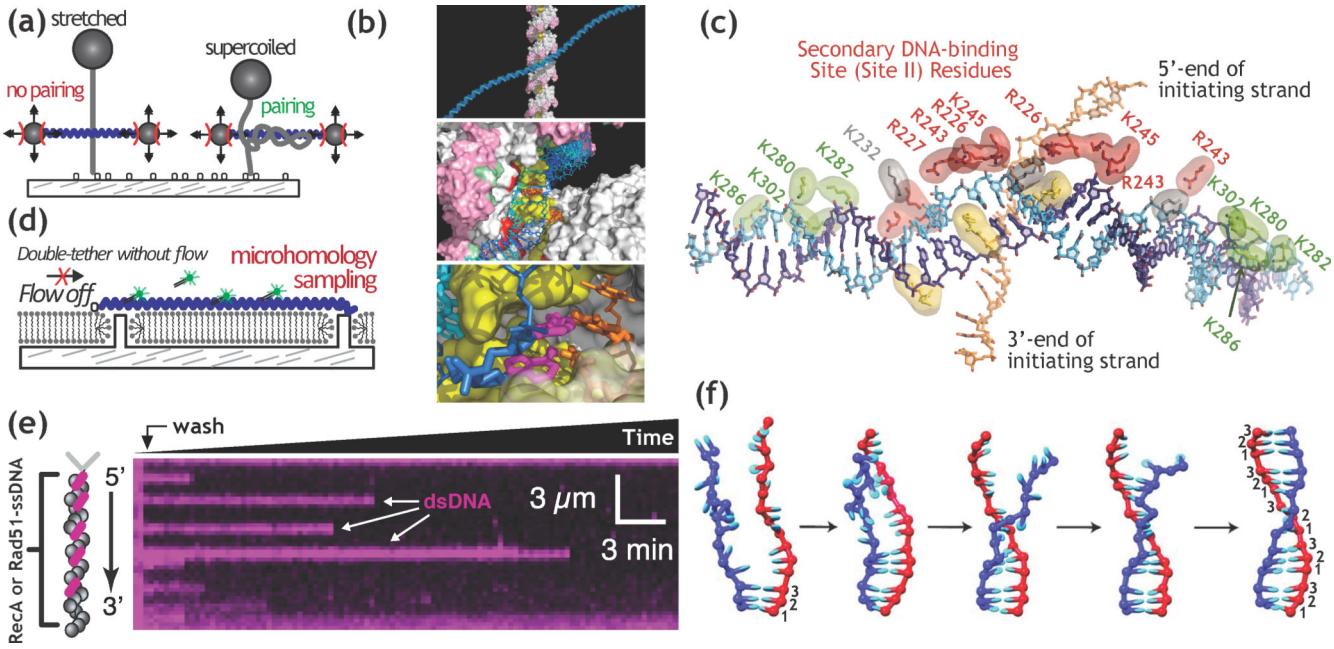

Figure 4. Reduction in complexity through microhomology sampling

A) "Dual-molecule" method simultaneously using magnetic and optical traps to measure filament-DNA contacts formed on stretched or supercoiled DNA [90]. **B)** Molecular modeling and simulation of dsDNA interactions with the RecA-ssDNA nucleoprotein filament during initial contact, binding to the secondary DNA binding site (site II) and homology sampling. *Adapted from* [91]. **C)** Molecular model of a RecA pairing intermediate [91, 92]. Labeled amino acids indicate positively charged residues that constitute the secondary DNA binding site that binds to and stabilizes the incoming homologous dsDNA. **D)** Schematic and **E)** representative data from a single-molecule 'DNA curtain' experiment measuring RecA-mediated DNA 'microhomology sampling' between a long RecA-ssDNA filament and short, fluorescently-labeled dsDNA molecules [94]. **F)** Molecular modeling (for simplicity, only two strands are shown) of the transient kinetic intermediates formed during the initial microhomology sampling (8 base pairs or less) and subsequent DNA pairing reaction (9 base pairs or greater), resulting in the post-synaptic, Rad51/RecA-Stretched DNA intermediate that is stretched into segments of three base pairs containing normal Watson-Crick spacing, interspersed with 7–8 Å gaps formed by intercalating hydrophobic residues (e.g., Met164 and Ile99). *Adapted from* [95].

Table 1

RecA Filament Assembly & Synapsis During Replication and Homologous Recombination

Rates & Relevant Physical & Physiological Measurements		Ref
Volume of an E. coli cell	1 femtoliter (1×10^{-15} L)	[103]
E. coli genome size	4.7 Mbp (4.7×10^{6} bp)	[104]
Replication rate	650–800 bp/sec	[105]
ssDNA generated at replication fork, *(average Okazaki fragment size)*	1000–2000 nucleotides	[106]
Average size of daughter strand gaps	100–800 nucleotides	[107, 108]
dsDNA breaks per division, *(average)* *(maximum tolerated)*	0.1–1 less than 3	[28, 109] [110]
DNA cross-links per division, *(maximum tolerated)*	50–70	[111]
Oxidative lesions per division	~2000	[28]
Rate of RecBCD resection	1000–2000 bp/sec	[14]
Average χ (Chi) frequency	1 per 4,500 bp	[14]
Average length of dsDNA resection by RecBCD, *(in vitro) (in vivo)*	30,000 bp 10,000 bp	[14] [15]
SSB site size per tetramer	30–70 nucleotides	[37, 112]
RecA site size per monomer	3 nucleotides	[29]
Persistence length of dsDNA	~50 nm	[80]
Persistence length of ssDNA	~1 nm	[80, 82]
Persistence length of RecA-ssDNA	~900 nm	[44]
Radius of gyration (R_G) for λ dsDNA (48.5 kbp)	~900 nm	
RecA nucleation time (rate) [1]***, spontaneous*** ***RecOR-mediated*** ***RecFOR-mediated***	10–60 min (1–6 nuclei/hr) 5–30 min (2–12 nuclei/hr) 2–10 min (6–30 nuclei/hr)	[24]
RecA growth rate [1]***, spontaneous*** ***RecOR-mediated*** ***RecFOR-mediated***	0.3–1.3 RecA monomers/sec 2–6 RecA monomers/sec 2–6 RecA monomers/sec	[24]
RecA K_D for ATP ***RecA K_D for ATP, + ssDNA***	~15 μM ~2.5 μM (ssDNA)	[113]
RecA K_M for ATP, + ssDNA ***+ dsDNA***	~20 μM (ssDNA); ~100 μM (dsDNA)	[114]
RecA k_{cat} for ATP, + ssDNA ***for dATP, +ssDNA***	~21 per min per RecA ~33 per min per RecA	[56]
RecA + ssDNA salt titration midpoint [2]***,*** *0.1 mM nucleotide cofactor*	255 mM NaCl ~400 mM NaCl (+ATP) 165 mM NaCl (+ADP)	[115]
RecA + dsDNA salt titration midpoint [3]***,*** *1 mM nucleotide cofactor*	~300 mM NaCl (+ATP) 190 mM NaCl (+ADP)	[116]

[1] Nucleation and growth rates reported were measured in the presence of ATPγS (1 μM RecA, 2 mM ATPγS, pH 7.5, 37°C, ~8,000 nt substrate) [24]. The numbers in this table represent our best estimate for physiologically relevant nucleation and growth rates in the presence of ATP (instead

of ATPγS), which is ten times slower for nucleation and two-thirds slower for growth. However, these rates are affected by temperature, pH, excluded volume, and concentrations of proteins, and mono-, di-, and tri-valent salts in nonlinear ways. Nonetheless, these estimates are consistent with ensemble experiments with RecFOR [22] and in vivo imaging of RecA bundle appearance and growth [13].

[2] 20 mM Tris Acetate, pH 7.5, 4 mM $Mg(OAc)_2$, 25°C

[3] 20 mM MES, pH 6.2, 10 mM $MgCl_2$, 25°C