The intrinsic ability of double-stranded DNA to carry out D-loop and R-loop formation

Takehiko Shibata a,⇑, Wakana Iwasaki b, Kouji Hirota a

a Department of Chemistry, Graduate School of Science, Tokyo Metropolitan University, Hachioji, Tokyo, Japan
b Laboratory for Translation Structural Biology, RIKEN Center for Biosystems Dynamics Research, Tsurumi, Yokohama, Japan

Abstract

Double-stranded (ds)DNA, not dsRNA, has an ability to form a homologous complex with single-stranded (ss)DNA or ssRNA of homologous sequence. D-loops and homologous triplexes are homologous complexes formed with ssDNA by RecA/Rad51-family homologous-pairing proteins, and are key intermediates of homologous (genetic/DNA) recombination. R-loop formation independent of transcription (R-loop formation in trans) was recently found to play roles in gene regulation and development of mammals and plants. In addition, the crRNA-Cas effector complex in CRISPR-Cas systems also relies on R-loop formation to recognize specific target. In homologous complex formation, ssDNA/ssRNA finds a homologous sequence in dsDNA by Watson-Crick base-pairing, crRNA-Cas effector complexes appear to actively melt dsDNA to make its bases available for annealing to crRNA. On the other hand, in D-loop formation and homologous-triplex formation, it is likely that dsDNA recognizes the homologous sequence before the melting of its double helix by using its intrinsic molecular function depending on CH2 at the 2'-position of the deoxyribose, and that the major role of RecA is the extension of ssDNA and the holding of dsDNA at a position suitable for homology search. This intrinsic dsDNA function would also play a role in R-loop formation. The dependency of homologous-complex formation on 2'-CH2 of the deoxyribose would explain the absence of homologous complex formation by dsRNA, and dsDNA as sole genome molecule in all cellular organisms.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

It is a longstanding question why DNA, not RNA, is the sole genomic material of all cellular organisms. Possible explanations to answer this question include the chemical and biological instability of RNA, its A-form structure as its double-strand form of RNA and possible ribozyme activity. The A-form dsRNA has small grooves that prevent sequence-dependent interactions with proteins, but dsRNA has been shown to adopt also the A' form that contains wider major groove to accommodate sequence-specific interactions with proteins [1]. Encoded sequences may induce a ribozyme activity that causes the degradation of the RNA genome, but DNA enzymes have been constructed [2]. Thus, these explanations are not qualitative differences between DNA and RNA. The ability of double-stranded (ds)DNA to form homologous complexes with single-stranded (ss)DNA or ssRNA provides a more definitive answer.

Homologous complexes are products of a reaction in which a region of a strand of dsDNA is replaced by the ssDNA or ssRNA of fully or almost fully identical sequence (homologous sequence), thus forming a hybrid duplex (less than 10 base-pairs [bps] to several kbps in length) with the complementary strand of the dsDNA (Fig. 1). Homologous complexes include D-loops [3,4] (Fig. 1, B and E), R-loops [5] (Fig. 1, B and D) and “homologous triplexes” [6] (Figs. 1C and 2A). Homologous-complex formation with either ssDNA or ssRNA are generally observed in the cells of all living organisms. Though dsRNA is found in various cells, intermolecular homologous-complex formation with dsRNA has not thus far been reported in vivo. It is intriguing that dsRNA is generated during RNA interference but processed into ssRNA before the recognition of the complementary sequence in its target RNAs. RNA interference is a cellular system of post transcriptional gene regulation and repression of transposons (See Ref. [7], an example of review).

Homologous complexes with single-stranded DNA were first postulated as essential intermediates for homologous (genetic) recombination (general genetic recombination; homologous DNA recombination) [3,8,9], and proven to the case by the discovery that homologous-complex formation is catalyzed by either RecA [10,11] or Rad51 [12], both of which are encoded by essential genes for homologous recombination [13,14]. Homologous recombination is a basic genetic function in all living organisms. It is essential to cell proliferation in vertebrates [15], including humans. The double-strand structure allows dsDNA to maintain genome information in duplicate, thus enabling the accurate correction or repair of replication errors and DNA lesions by using the complementary strand as a template. When both strands of dsDNA are broken (double-strand break), homologous recombination facilitates precise repair by using dsDNA bearing the homologous sequence as a template (See Ref. [16]). Homologous recombination reorganizes genes between genomes inherited from parents, and thus, increases genetic variation, which increases the efficiency of natural selection [17]. The mechanism of homologous recombination has been reviewed numerous times (for example, see Refs. [16,18]).

Homologous recombination of RNA genomes (RNA viruses) is rare and depends on a template-switch (or “copy-choice”) mechanism using ssRNA rather than an RNA version of a homologous-complex intermediate (See Ref. [19] for review; see also Ref. [20]). It has been reported that dsRNA is more thermally stable than dsDNA [21]. This may explain the difference between dsDNA and dsRNA in terms of their ability to form homologous complexes.

The absence of any evidence of homologous-complex formation with dsRNA, even after decades of study, suggests a qualitative advantage for dsDNA over dsRNA.

Homologous (DNA) complexes with ssRNA (i.e. R-loops) have been regarded as genome-destabilizing byproducts of transcription in which the transcripts (RNA) associate with the template DNA strand [22]. Recent studies have revealed that R-loops play an important role in cellular function, as described below.

The crucial question related to homologous-complex formation is how the ssDNA or ssRNA recognizes the complementary sequence within dsDNA. A naive possibility is that the melting (disruption of double-strand structure) of the dsDNA is followed by annealing with homologous ssDNA or ssRNA to form hybrid. Studies on RecA have revealed that homologous sequences are recognized before the melting of the dsDNA, in the presence of ATP [23–25]. Unlike the annealing of complementary ssDNA or ssRNA, all bases in dsDNA that form stable Watson-Crick (W-C) base pairs are not readily available to recognize sequence complementarity with ssDNA or ssRNA.

In this minireview, we describe the biological functions of homologous-complex formation with ssDNA or ssRNA as well as the biochemical and structural features of homologous-complex formation with ssDNA or ssRNA by proteins that catalyze formation. We also discuss the possible molecular mechanisms suggested by molecular-structural analyses and recommend the use of computational simulation methods to further investigate outstanding questions.

2. D-loops, R-loops, and homologous triplexes as homologous (DNA) complexes with ssDNA or ssRNA

2.1. D-loops and R-loops

D-loops and R-loops are typical homologous complexes in which the replaced strand of dsDNA (Fig. 1A) separates from a hybrid duplex and forms a loop (Fig. 1, B and E). D-loop formation by RecA is enhanced by negative (right-hand) supercoil, which all natural covalently closed circular dsDNA (ccc-dsDNA) isolated from cells owns [26]. The D-loop or R-loop, formed with ccc-dsDNA, is stabilized by the relaxation of its negative supercoil [27]. Under conditions that denature the double helix, ccc-dsDNA carrying a D-loop or an R-loop in relaxed form is dissociated into ccc-dsDNA with negative supercoil and free ssDNA or ssRNA. A D-loop in dsDNA that has a strand break outside the site of the D-loop is spontaneously dissociated by branch migration induced by the rotational thermal movement of dsDNA at almost 10^4 bps/sec at 37 °C [28].

Topoisomerases change the topological state of ccc-dsDNA and its homologous complex. However, topoisomerases do not appear to directly stimulate or inhibit RecA-catalyzed homologous-complex formation. Topoisomerases are very useful when studying the mechanism of homologous-complex formation, since untwisting (often expressed by unwinding) and rewinding of the double helix during D-loop or R-loop formation can be sensitively measured as topological changes that are quantified by the use of topoisomerase and ccc-dsDNA (See Refs. [23,29–31]).

2.2. Homologous triplex

The other type of homologous complex is the homologous triplex, which is a type of parallel triplex [32] (Figs. 1C and 2A), in
which the replaced DNA strand is held by base-pair-specific hydrogen bonds (not by Hoogsteen hydrogen bonds) in the major groove of the hybrid W-C duplex that is newly formed between the invading ssDNA or ssRNA and the complementary strand of the parental dsDNA (Fig. 1C). The homologous triplex has been shown to be a homologous complex formed by RecA in vitro between ssDNA and dsDNA oligomer with hairpin at an end [6]. The homologous triplex between relaxed ccc-dsDNA with ssDNA has been shown to be formed by a protein, Mhr1, which is essential to yeast mitochondrial homologous recombination [29]. Unlike the case of the D-loop described above, when the relaxed ccc-dsDNA carrying the homologous triplex formed by Mhr1 in the presence of a topoisomerase is dissociated into ssDNA and dsDNA by high temperatures or alkaline pH after the inactivation of the proteins, the ccc-dsDNA has been demonstrated to remain relaxed without detectable topological change [29].

Note that the homologous triplex is different from the products of triplex formation between ssDNA or ssRNA and dsDNA in the absence of disruption of the W-C base pairs of the parental dsDNA. The triplex formed in the absence of disruption of the parental dsDNA does not contain the hybrid W-C duplex (Fig. 2C). Typical triplexes are formed between polypurine-polypyrimidine ssRNA and polypurine-polypyrimidine dsDNA, with the third polypurine-polypyrimidine RNA strand held antiparallel to the polypurine-polypyrimidine strand of the dsDNA by Hoogsteen hydrogen bonds in the major groove of the dsDNA [33] (Fig. 2, D and E). The formation of this type of triplex has various sequence constraints (See Ref. [34]). Unlike triplex formation without hybrid duplex formation, homologous triplex formation requires two parallel strands with an almost identical sequence, and accepts various base-sequences (Figs. 1C and 2A).

3. Proteins that catalyze homologous-complex formation with ssDNA and their biological function

3.1. D-loop formation by RecA/Rad51-family homologous-pairing proteins in homologous recombination

The RecA homologous-pairing protein (RecA protein, RecA) is the prototype of the RecA/Rad51-family homologous-pairing proteins that catalyze D-loop formation from dsDNA and homologous ssDNA in vitro in an ATP-dependent mode [10,11]. RecA/Rad51-family proteins include eubacterial RecA, eukaryotic orthologue Rad51 [35], archaeal RadA [36], and Dmc1 [37]. Dmc1 is the meiosis-specific paralogue of Rad51 that is essential to reductive segregation in meiosis for gamete formation in sexual reproduction of eukaryotes.

Homologous-complex formation by RecA/Rad51/Dmc1/RadA consists of two steps: homologous pairing and branch migration [38]. Homologous pairing is a quick (within a second at 37 °C) reaction that forms the core of the homologous complex, a roughly 15 bp hybrid duplex [39,40]. Homologous pairing requires ATP but does not require its hydrolysis. Branch migration is an ATP-hydrolysis-dependent slow reaction that elongates the hybrid duplex unidirectionally by thousands of bps at the rate of approximately 4 bps per second (at 37 °C) [38]. The RecA-family proteins form a right-handed helical filament around ssDNA or dsDNA [30,41,42] (Fig. 3, A and B). Homologous pairing does not require filament formation, but branch migration depends on it [43].

It is intriguing to note that the three-dimensional (3D) structure of RecA/Rad51/Dmc1/RadA’s core domain is almost identical to that of the DNA helicase domain of T7 replicative helicase-primase [44] and is also very similar to that of the DNA helicases of the same group [45]. DNA helicases are enzymes that melt dsDNA into complementary ssDNAs in an ATP-hydrolysis-dependent mode.

In eukaryotic cells, dsDNA is folded into chromatin, wherein it wraps around histone octamers and forms a beads-on-a-string structure [46]. Nucleosomes on the dsDNA prevent D-loop forma-
tion by Rad51, but Dmc1 forms D-loops preferentially in the nucleosome-depleted regions [47]. Rad54, a protein that interacts with Rad51 and is required for homologous recombination, stimulates remodeling of the chromatin structure to enable Rad51 to form D-loops [48]. A histone chaperone, Nap1, interacts with Rad54 and stimulates Rad54-mediated eviction of the linker histone H1, further stimulating D-loop formation by Rad51-Rad54 in dsDNA in higher-ordered chromatin with the linker histone H1 [49]. Thus, even in the cell, D-loop formation occurs at nucleosome-free regions of dsDNA.

Fig. 2. Triplexes. A. Crystal structure of a parallel DNA triplex at G GC [32]. This parallel triplex is included in the homologous triplex formed between the invading ssDNA (pink; red strand in Fig. 1C) and the parental homologous dsDNA (sky blue; blue strand in Fig. 1C). This panel is an example of base pairings in the triplex. All combinations of C CG, A AT and T TA in addition to G GC exists in the homologous triplex. Note that the pink strand forms a hybrid W-C duplex with the complementary strand of the parental dsDNA (sky blue), and has the same polarity and sequence as the replaced strand of the parental dsDNA (sky blue) during homologous-triplex formation (See Fig. 1C). B. Crystal structure of an antiparallel DNA triplex at G GC [32]. The top strand with G (pink) and the strand with G of the parental dsDNA (sky blue) are antiparallel and paired with Hoogsteen hydrogen bonds. C. Antiparallel triplex formed from polypyrimidine ssRNA and polypurine-polypyrromidine W-C dsDNA without disruption of the double-strand structure. D. Antiparallel triplex at U in ssRNA (pink) paired with the A:T bp in dsDNA (sky blue). E. Antiparallel triplex at C in ssRNA (pink) paired with the GC W-C base pair in dsDNA (sky blue). The polypyrimidine RNA strand is parallel to the polypurine DNA strand and is antiparallel to the polypyrromidine DNA strand. Thus, this type of triplex cannot form between the transcript and the template dsDNA (See Ref. [34]). Panels A and B were generated from PDB 272D [32], and panels D and E are from PDB 1R3X [102]. All the figures of the structures in this article were prepared with PyMOL (PyMOL Molecular Graphics System, Schrödinger, LLC). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
3.2. Homologous complex formation by ATP-independent homologous-pairing proteins

Homologous triplexes have been identified as in vitro products formed in the absence of ATP by budding yeast mitochondrial protein Mhr1. MHR1 is an essential gene for homologous recombination, DNA replication, and DNA partitioning (segregation) into daughter cells in budding yeast mitochondria, which lack the RecA/Rad51 orthologue. In budding yeast, mitochondrial DNA is replicated through a rolling-circle mode initiated at the homologous complex formed by Mhr1 (See Ref. [50] for review,[51]). Although major homologous recombination of genomic DNA strictly requires either RecA [13] or Rad51 [14], homologous recombination of plasmid DNA and homologous recombination of inverted repeats are detected even in the absence of active RecA and Rad51. These minor recombination events are promoted by eubacterial RecT and eukaryotic Rad52 in vivo. RecT and Rad52 catalyze homologous-complex formation in the absence of ATP in vitro [52,53]. RecT and Rad52 belong to the same structural family of proteins [54], but Mhr1 belongs to a structurally distinct group and has a 3D structure similar to that of RecA/Rad51 [55].

3.3. TRF2 (telomere-repeat-binding factor 2) for telomere-loop formation

A telomere loop is a large lariat-like dsDNA loop of variable size (several to dozens of kbps in human cultured cells) that forms at the ends of chromosomes called telomeres. Telomere loops protect the chromosomal termini. For example, in humans, telomeres are thousands of repeats of the 5'-TTAGGG-3'/3'-AATCCC-5' sequence. In the telomere loop, the 3' ssDNA tail of the repeats invades an inner repeat to form a homologous complex [56]. TRF1 and TRF2 are major telomere-binding proteins that recognize the telomere-repeat sequence. In vivo, TRF2 is required for telomere-loop formation, but TRF1 is not [57]. In vitro, TRF2, but not TRF1, promotes telomere-loop formation of dsDNA with ssDNA (both of which have telomere repeat sequences) in an ATP-independent mode [31,56,58]. Note that the negative supercoils and telomere sequences in substrate dsDNA stimulate spontaneous formation in the absence of protein, resulting in a high background of TRF2-promoted formation in vitro [31].

4. Homologous complex formation with ssRNA

4.1. In vivo R-loop formation in trans

Various tools are available for analyzing R-loops in an entire genome at a single-base resolution. For example, DNA-RNA hybrid immunoprecipitation (RDIP or DRIP) coupled with high-throughput sequencing is an effective tool. The combination of RDIP and strand-specific sequencing of complementary DNA (RDIP-seq), or single-strand DNA ligation-based library construction after DRIP combined with next generation sequencing (ssDRIP-seq), enables us to locate strand-specific R-loops in an entire genome ([59] and see Ref. [60]). The presence of DNA-RNA protein Mhr1. MHR1 is an essential gene for homologous recombination, DNA replication, and DNA partitioning (segregation) into daughter cells in budding yeast mitochondria, which lack the RecA/Rad51 orthologue. In budding yeast, mitochondrial DNA is replicated through a rolling-circle mode initiated at the homologous complex formed by Mhr1 (See Ref. [50] for review,[51]). Although major homologous recombination of genomic DNA strictly requires either RecA [13] or Rad51 [14], homologous recombination of plasmid DNA and homologous recombination of inverted repeats are detected even in the absence of active RecA and Rad51. These minor recombination events are promoted by eubacterial RecT and eukaryotic Rad52 in vivo. RecT and Rad52 catalyze homologous-complex formation in the absence of ATP in vitro [52,53]. RecT and Rad52 belong to the same structural family of proteins [54], but Mhr1 belongs to a structurally distinct group and has a 3D structure similar to that of RecA/Rad51 [55].
hybrid duplexes is confirmed by sensitivity to RNase H, which specifically degrades RNA in DNA-RNA hybrid duplexes [61]. Studies using these new tools have shown that R-loops are abundant in the genomes of human [59] and plant (See Ref. [60]) cells, and have revealed that R-loops, rather than acting as lesions to destabilize genomes, are involved in gene regulation for various genetic and developmental functions. Most R-loops are thought to be formed co-transcriptionally (in association with transcription; see Ref. [62] for review), but it has recently been shown that some R-loops are formed independent of transcription. This type of R-loop formation is called R-loop formation in trans [59].

Studies on a long non-coding RNA (IncRNA) called APOLO (AUXIN-REGULATED PROMOTER LOOP) in Arabidopsis thaliana have shown that this IncRNA forms R-loops in trans at multiple distant-target genetic loci for lateral root development [63]. Seeking a second, independent approach to obtain experimental support, the authors of this study introduced, in addition to DRIP, a new DRIP-independent protocol called RNA isolation by DNA purification (RIDP), which uses biotinylated DNA probes to isolate the DNA of specified loci. APOLO-IncRNA transcription is activated by a phytohormone, auxin, during lateral root development. APOLO-IncRNA-mediated R-loop formation modulates chromatin loops at the target loci (at least 200 loci). R-loop formation at each target locus occurs at a pair of target sequences that are distant on the DNA but in close proximity via the chromatin loop, resulting in decrease of LHP1 (plant Polycomb Repressive Complex 1 component), decrease in repressive mark (histone methylation [H3K27me3]), opening of the chromatin loop, and activation of transcription at the target locus. The invasion of the target site by APOLO IncRNA can thus coordinate the transcription of spatially non-associated auxin-responsive genes [63]. APOLO IncRNA forms R-loops at regions containing a consensus sequence, GAAGAAC/G/ C, which is required for R-loop formation [63]. At present, it is not known which protein catalyzes R-loop formation with APOLO IncRNA.

4.2. R-loop formation by the crRNA-Cas-effector complex in CRISPR-Cas immune systems

CRISPR (clusters of regularly interspaced short palindromic repeats)-Cas (CRISPR-associated) systems are adaptive immune systems that act against invading viruses and plasmids found in prokaryotes and archaea. When DNA viruses and plasmids invade the cell, their DNA is processed into short DNA sequences, which are integrated in the form of spacer sequences in the CRISPR loci as a memory of the infection. crRNA (CRISPR RNA) is a processed transcript of CRISPR (for example, 61 nucleotides for Escherichia coli) that binds to Cas protein(s) to form a crRNA-Cas-effector complex in which the crRNA works as a guide RNA that recruits the effector complex to the homologous sequence (called a “protospacer”) of its target dsDNA by forming an R-loop. This R-loop formation results in the degradation of the foreign DNA. Two crRNA-Cas-effector complexes that have been extensively studied are: Cascade (a CRISPR-associated complex for antiviral defense) in E. coli and other eubacteria (Class 1, a multi-effector protein complex) and the effector complexes of Cas9 (Class 2, a single effector protein), which is used in gene-editing technology. The details and classification of CRISPR-Cas systems have been reviewed multiple times [64-66].

crRNA-Cas-effector complexes (containing guide RNA) first recognize a PAM (protospacer adjacent motif) sequence in target DNA by protein-DNA interaction that does not involve guide RNA. R-loops are then formed to complete target recognition, wherein the spacer sequence of the guide RNA and the complementary sequence (target strand) of the protospacer sequence form a hybrid duplex. PAM comprises a 2 to 3 bp sequence adjacent to either the 5’ end (Cascade, Fig. 1D) or the 3’ end (Cas9) of the non-target strand of the protospacer in the target dsDNA. This R-loop formation induces the cleavage of the target DNA.

5. Mechanisms of homologous-complex formation with ssDNA/ssRNA

RecA/Rad51-family proteins share, with a minor difference, a common mechanism for D-loop formation. Most research on the reaction steps of D-loop formation has focused on RecA. In D-loop formation, RecA first binds to ssDNA, even in the presence of dsDNA [67,68], then forms a right-handed spiral filament around the DNA at a rate of 6 RecA protomers per turn [42] (Fig. 3, A and B). In the presence of ATP or of a nonhydrolyzable ATP analogue, ssDNA in the filament is extended 1.5 fold longer than B-form dsDNA with the same base sequence (active filament [69]). The 3D structure of the RecA/Rad51-ssRNA active filament in the presence of a nonhydrolyzable ATP-analogue has been determined by X-ray crystallography [42] (Fig. 3, A and B) and by high-resolution cryo-electron microscopy [70,71]. The active RecA-ssDNA filament (extended filament [42,69]) structure shows that ssDNA binds to the primary DNA-binding site that contains the folded L1 and L2 loop regions (Fig. 3C). The L1 and L2 loops are ordered in the inactive RecA-ssDNA filament (in the absence of DNA and ATP analogue [72]). In the active RecA/Rad51-ssRNA complex, the extension of ssDNA is not uniform. The RecA protomer binds stoichiometrically to the ssDNA, with one protomer to three nucleotides, and pairs of folded L2 loops of all adjacent RecA protomers sandwich 3 nucleotides (nucleotide triplet) of DNA (Fig. 3C), in which the base distance is slightly larger than that of B-form DNA (3.83 Å to 4.66 Å vs. 3.5 Å). The base distance between triplets is further extended (7.8 Å) by the L2 loop (Fig. 3C). Cryo-electron microscopic observation has revealed the similar structure of the ssDNA and the L2 loop in the active Rad51-ssDNA filament [70].

By binding to ssDNA in the presence of ATP or a nonhydrolyzable ATP-analogue, RecA is activated for sequence-independent binding to dsDNA [67]. The RecA-ssDNA filament then slides along the dsDNA [39] to search for a homologous sequence between the ssDNA in the RecA-filament and the dsDNA. Once a homologous sequence is found, a homologous complex-core (nascent homologous complex) is formed [67], without ATP-hydrolysis [38]. The binding sites of ATP are located between the adjacent RecA protomers, and ATP-binding to the RecA-ssDNA filament modulates the spatial configuration of the adjacent RecA protomer. This spatial configuration change results in the extension of the filament associated with the extension of the bound ssDNA (compare [42] with [72]) as well as the activation of the dsDNA binding to the filament. The binding surface of dsDNA, consisting of several amino-acid residues (called the gateway), is located on the C-terminal domain of RecA and at the boundaries between adjacent RecA protomers on the extended RecA filament [73], and refs. cited) (Fig. 3).

The smallest detectable homologous complex formed by RecA/Rad51 contains a 6–8 bp hybrid duplex [40,74,75], and the smallest stable homologous complex contains a 15 bp hybrid duplex [40]. Alongside the nucleotide triplets in the RecA/Rad51-ssDNA filament, the hybrid duplex is elongated from 6 – 8 bps to 15 bps in 3 bp increments, in the absence of ATP-hydrolysis [39,40].

The nascent homologous complex could be a homologous triplet [5] (See 2.2). The hybrid duplex in the nascent complex is spontaneously elongated in association with D-loop growth by the rotational stress generated by the negative supercoils of the
substrate ccc-dsDNA (See Refs. [43,76]), or by an ATP-hydrolysis-dependent function of RecA [77].

The gateway leads the bound dsDNA to the L2 loop (Fig. 3, B and C), where the dsDNA encounters ssDNA in the RecA filament. Homologous-sequence recognition between dsDNA and ssDNA then takes place in the vicinity of the L2 loop. A recent study revealed that the active unit for ATP-dependent homologous pairing is a pair of adjacent RecA protomers and that filament formation is not essential to this process [43]. ATP-hydrolysis-dependent branch migration depends on filament formation (See Ref. [78] for review).

5.2. Role of RecA/Rad51 in homologous complex formation

The primary model for homologous-sequence recognition is the melting of dsDNA at the gateway to make the dsDNA bases available to form W-C base pairs with ssDNA. Melting induces untwisting of the double helix. Several studies have described the dsDNA and ATP or ATP-analogue-dependent untwisting of dsDNA by RecA under conditions for homologous-complex formation. Untwisting of dsDNA bound to RecA-heterologous ssDNA filament had been observed in the presence of an nonhydrolyzable analogue, supporting this melting-before-annealing model [79], but the later studies in the presence of ATP revealed that the untwisting of dsDNA bound to RecA-ssDNA filaments requires homology between dsDNA and the ssDNA in the RecA-filaments [23,25], supporting homologous-sequence recognition in the absence of the melting of dsDNA. Since RecA, ssDNA, and dsDNA form a huge network of molecules with an irregular shape under these conditions [80], the individual RecA-dsDNA filaments had not been identified by either electron microscopy or crystallographic analysis.

Although the initial binding takes a longer time [81] by using the gateway [82], RecA can bind to dsDNA in the absence of ssDNA (in the presence of ATP). Electron micrographs taken under conditions for dsDNA-untwisting in the absence of ssDNA by RecA show well-ordered filaments [30,41]. The crystal structure of RecA-dsDNA filament formed in the presence of an ATP-analogue is almost identical to that of active RecA-ssDNA filament, and dsDNA bound at the center of RecA-dsDNA filaments at the primary DNA-binding sites is untwisted but conserves base pairing. No base pair is disrupted in the filament [42]. Note that the dsDNA untwisting so far observed does not require ATP-hydrolysis, and it is induced by binding to the primary DNA-binding site in the extended RecA-filament induced by ATP-binding.

The features of dsDNA bound to the gateway or RecA-ssDNA spiral filament had not been revealed until recently. Yang et al. solved by use of cryo-electron microscope, high resolution 3D structures of active RecA-ssDNA filaments consisting of nine protomers and dsDNA bound to gateway [83]. They analyzed the complexes formed by the RecA-ssDNA and heterologous dsDNA and by RecA-ssDNA and homologous dsDNA. Based on these structures, the authors proposed a mechanistic model for D-loop formation: dsDNA bound to the gateways of the RecA-ssDNA spiral filament encounters an L2 loop, and this interaction causes the local melting of the double helix (up to 15 bps or more), followed by annealing with the ssDNA at the center [83]. This model appears to be well supported by various 3D structures revealed by their analysis, but does not explain why DNA, but not RNA, can form homologous complex with ssDNA or ssRNA. Note that Yang collected the cryo-electron microscopic images in the presence of a nonhydrolyzable ATP-analogue, the condition where the untwist of dsDNA by RecA bound to heterologous ssDNA had been observed [79] (For discussion, see above and below).

The sole model that explains why only DNA can form homologous complexes was proposed based on the unique Nuclear Magnetic Resonance (NMR) 3D structure of ssDNA oligomers bound to homologous-complex-forming proteins. The NMR analysis (transfer NOE) has revealed the unique, ATP-analogue-dependent 3D structure of extended ssDNA oligomers induced by transient binding to RecA/Rad51 (Fig. 4A). In this structure, the distance between adjacent bases is extended 1.5 fold of B-form DNA (Fig. 4B). B-form DNA is stabilized by stacking between adjacent bases. Instead of base-stacking, the extended ssDNA structure is stabilized by CH-π (CH-pi) interaction, an attractive molecular force [84], between the CH2 moiety at the 2' position on the deoxynucleobase and the base of the following nucleotide [85,86] (Fig. 4A). A similar extended ssDNA structure was induced also by all tested proteins (RecT, Mhr1, and RecO) that catalyze homologous-complex formation between dsDNA and ssDNA in an ATP-independent fashion [87]. Thus, the extended DNA structure stabilized by CH-π interaction is a general intermediate of the homologous-complex formation of dsDNA and ssDNA. Considering the variation in the 3D structure of homologous-complex forming proteins, the DNA structure stabilized by CH-π interaction in the general intermediate suggests that the homologous complex formation is an intrinsic DNA function rather than a function of specific proteins [87,88].

Models of extended dsDNA stabilized by CH-π interaction were constructed [86]. In normal dsDNA, base pairs randomly opens and the mean lifetime of W-C base pairs at 35 °C is in the range of milliseconds [89]. In the extended dsDNA, dsDNA bases in base pairs opened can be flipped out spontaneously by the interconversion of sugar pucker, and can thus interact with the bases of nearby ssDNA. This interaction explains the mechanism for homologous-sequence recognition between dsDNA and ssDNA without the melting of the double helix [86,88] (Fig. 4C). The CH-π interaction between a sugar and a base is unique to DNA and does not exist in RNA, and thus, the interaction explains the unique ability of dsDNA to form homologous complexes. In Fig. 4C, both strands of dsDNA are assumed to be elongated, but when the CH-π interaction occurs in one of the two strands, it results in DNA bending [90]. Thus, this model works with bent dsDNA. Lipfert et al. found that the characteristic transition rate of RNA during plectonemic buckling is two orders of magnitude slower than that of dsDNA [91]. This finding suggests that dsDNA has a high enough flexibility to allow sharp local bending of dsDNA (on a scale of approximately 5 nm in length). The two orders of magnitude difference between dsDNA and dsRNA also explain the absence of homologous-complex formation by dsRNA. The cryo-electron microscopic study published by Yang shows a 3D structure that appears RecA-ssDNA filament that forms complex with a heterologous bent dsDNA [83].

The absence of active melting of dsDNA by RecA in nascent homologous-complex formation is supported by kinetic analysis of homologous-complex formation by RecA [92]. Using oligo DNA labeled with fluorescent dyes, Xiao et al. found that nascent homologous-complex formation from dsDNA and extended ssDNA on an active RecA-ssDNA filament was driven entropically rather than enthalpically [92]. Their results support the conclusion that RecA-catalyzed homologous-complex formation relies on dynamic intrinsic DNA functions [88,92].

5.3. Uncatalyzed homologous-complex formation

A recent magnetic-tweezers experiment revealed that the mechanical stretching of ssDNA induces homologous-triplex formation with homologous dsDNA, not with heterologous dsDNA, at room temperature in the absence of any protein. The ssDNA in the homologous complex was shown to be sensitive to a dsDNA-specific endonuclease, indicating that ssDNA forms a hybrid duplex with the complementary strand of dsDNA [93]. This finding shows that the active melting of dsDNA is not required for homologous-complex formation, and supports a view that homologous complex formation is an intrinsic DNA function.
Regarding the uncatalyzed homologous-triplex formation by mechanically stretched ssDNA and dsDNA described above [93], we can speculate that RecA/Rad51 plays a critical role in homologous-complex formation via the extension of ssDNA and the holding of dsDNA at a location suitable for homology-search interaction with the extended ssDNA. The homologous sequence in the extended ssDNA may be recognized by dsDNA through base-flipping enabled by transient CH-π interaction-dependent dsDNA extension [86,88] and/or bending [90].

5.4. Role played by the crRNA-Cas-effector complex in R-loop formation

Single-molecule analysis using rotor-bead tracking (RBT) or an equivalent technique has shown that R-loops are formed within 100 ms by Cas9 [94]. R-loop formation by Cascade is faster than by Cas9, and the R-loops formed by Cascade are more stable than those formed by Cas9 [95]. R-loop formation involves transient discrete intermediates, including initial DNA-RNA hybridization within the seed region (11 bps; Figs. 1D and 5A) of the protospacer sequence adjacent to the PAM sequence of the target DNA [94].

The need for high-resolution imaging of the 3D structures of crRNA-Cas-effector complexes with or without target dsDNA has been met by X-ray crystallography (Cascade [Class 1, Type I] [96,97], RNA-silencing Cmr complex [Class 1, Type III] [98]), cryo-electron microscopy (Cascade [99,100]), and by both techniques together (Cas9 [Class 2, Type II] [101]). These techniques all reveal differences in R-loop structure between class 1 (Cascade and RNA-silencing Cmr complex) crRNA-Cas-effector complexes and class 2 (Cas9) crRNA-Cas-effector complexes, as well as common unique 3D structures that point to the recognition mechanism of the protospacer sequence in the target dsDNA.

The 3 bp PAM sequence of target DNA is recognized by interaction with a subunit of Cascade. The PAM-adjacent protospacer region of the dsDNA is then bent approximately 90° and the two strands are separated. At this point, a wedge is inserted to disrupt the first base pair and to help flip out the first several bases from both DNA strands [99,100] (Fig. 5A). Thus, PAM recognition leads to the bending of dsDNA, resulting in the spontaneous melting of the dsDNA, which is captured by the crRNA to form the seed bubble. The seed bubble is elongated along a crRNA guide in 6 base increments to form a full R-loop [97,99].

In the R-loops formed by the Class 1 Cas-effector complex (Cascade and RNA silencing Cmr complex), the crRNA hybrid duplex and the target strand of the protospacer region of the target DNA are not a canonical double helix, but rather an extensively untwisted ladder-like structure. The R-loop consists of 5 segments, each including a 5 bp region followed by a disrupted base pair [97,99,100] (Figs. 1D and 5A). Every sixth unpaired nucleotide of crRNA binds tightly to one of the 6 well-organized Cas7 protein subunits of the cascade along the crRNA [96]. The unfolded crRNA structure is not changed before or after R-loop formation [100].

Upon recognition of a PAM sequence in dsDNA, Cas9 (Class 2, Type II) also binds dsDNA, providing enough stress to melt the dsDNA and stimulate R-loop formation with crRNA, though the bend is only about 30° [99,101] (Fig. 5B). Cas9 has no wedge structure, showing that the wedge found in Cascade is not a general structure required to melt dsDNA. In dsDNA bound to Cas9, the non-target strand of the protospacer dsDNA adjacent to a PAM sequence is held separate from the crRNA by extensive interaction with the protein [101] (Fig. 5B). The R-loop formed by Cas9 is a continuous 20 bp A-form DNA-RNA-hybrid duplex [101], in contrast to the segmented ladder-like structure formed by Cascade.

6. Summary and perspectives

It has been suggested as described above that, in R-loop formation by crRNA-Cas-effector complexes, proteins play an active role
This raises the intriguing question of whether recognition of homologous sequence between ssDNA/ssRNA and dsDNA without the preceding melting of dsDNA double helix is a general principle for homologous complex formation. This possibility has not been excluded as a mechanism for R-loop formation by crRNA-Cas-effector complexes.

Currently available analytical techniques are not sufficient to test and evaluate the various hypotheses for homology recognition and homologous-complex formation between ssDNA and dsDNA before the melting of the double helix. Such hypotheses include the contribution of base-flipping by bending and/or by transient extension of dsDNA, the role played by CH-π interaction, and the contribution of RecA/Rad51. Current techniques are limited to X-ray crystallography, high resolution cryo-electron microscopy, NMR analyses and single molecule analysis, and these techniques do not have sufficient time- or spatial-resolution, or direct information about the order of events and time elapsed in a reaction sequence. On the other hand, molecular dynamic simulation by computational means has enormous potential as a more effective way to evaluate these hypotheses. It should be possible to simulate homology recognition on a larger time scale, since the rate of spontaneous base-flipping depends on the rate of thermal W-C base pair opening and the mean lifetime of W-C base pairs at 35 °C is in the range of milliseconds [89]. This will require significant development in the field of computer simulation well beyond existing methods.

References

[1] Tanaka Y, Fujii S, Hiroaki H, Sakata T, Tanaka T, Uesugi S, et al. A’-form RNA double helix in the single crystal structure of (UUCACUUCCGCUUC). Nucl Acids Res 1999;27:949–55.
[2] Li Y, Breakey RR. Deoxyribozymes: new players in the ancient game of biocatalysis. Curr Opin Struct Biol 1999;9(3):315–23. https://doi.org/10.1016/S0959-440X(99)80042-6.
[3] Holloman WK, Wiegand R, Hoesli C, Radding CM. Uptake of homologous single-stranded fragments by superhelical DNA: a possible mechanism for initiation of genetic recombination. Proc Natl Acad Sci 1975;72(6):2394–8. https://doi.org/10.1073/pnas.72.6.2394.
[4] Liu LF, Wang JC. On the degree of unwinding of the DNA helix by ethidium. Biochim Biophys Acta 1975;395(4):405–12. https://doi.org/10.1016/0006-2977(75)90064-3.
[5] Thomas M, White RL, Davis RW. Hybridization of RNA to double-stranded DNA: formation of R-loops. Proc Natl Acad Sci 1976;73(7):2294–8. https://doi.org/10.1073/pnas.73.7.2294.
[6] Rao BJ, Chu SK, Radding CM. Homologous recognition and triplex formation promoted by RecA protein between duplex oligonucleotides and single-stranded DNA. J Mol Biol 1993;229:328–43.
[7] Curtbro MJ, Martienssen RA. Conserved chromosomal functions of RNA interference. Nat Rev Genet 2020;21(5):311–31. https://doi.org/10.1038/s41576-019-0203-6.
[8] Whitehouse HLK. A theory of crossing-over by means of hybrid deoxyribonucleic acid. Nature 1963;199(4898):1034–40. https://doi.org/10.1038/1991034a0.
[9] Holliday R. A mechanism for gene conversion in fungi. Genetic Res Camb 1964;5:282–304.
[10] Shibata T, DasGupta C, Cunningham RP, Radding CM. Purified RecA protein catalyzes homologous pairing of superhelical DNA and single-stranded fragments. Proc Natl Acad Sci 1979;76(4):1638–42. https://doi.org/10.1073/pnas.76.4.1638.
[11] McEntee K, Weinstock GM, Lehman IR. Initiation of general recombination catalyzed in vitro by the recA protein of Escherichia coli. Proc Natl Acad Sci 1979;76(6):2615–9. https://doi.org/10.1073/pnas.76.6.2615.
[12] Sung P. Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. Science 1994;265(5176):1241–3. https://doi.org/10.1126/science.265.5176.1241.
[13] Clark AJ, Margulies AD. Isolation and characterization of recombination-deficient mutants of Escherichia coli K12. Proc Natl Acad Sci 1965;53(2):451–9. https://doi.org/10.1073/pnas.53.2.451.
[14] Nakai S, Matsumoto S. Two types of radiation-sensitive mutants in yeast. Mutat Res 1967;4:129–36.
[15] Sonoda E, Sasaki MS, Buerstedde J-M, Bezzubova O, Shinohara A, Ogawa H, et al. Rad51 deficient vertebrate cells accumulate chromosomal breaks prior to cell death. EMBO J 1998;17:598–608.
[16] Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. The double-strand-break repair model for recombination. Cell 1983;33(1):25–35. https://doi.org/10.1016/0092-8674(83)90131-9.
Shibata T, W. Iwasaki and K. Hirota Computational and Structural Biotechnology Journal 18 (2020) 3350–3360

[17] Goddard MR, Godfry HCJ, Burt A. Sex increases the efficiency of natural selection in experimental yeast populations. Nature 2005;434(7033):636–40. https://doi.org/10.1038/nature03405

[18] San Filippo J, Sung P, Klein H. Mechanism of eukaryotic homologous recombination. Annu Rev Biochem 2008;77(1):229–57. https://doi.org/10.1146/annurev.biochem.77.061306.125255

[19] Dumas K, Hug N, Vindigni A, Bouvet P, Paoletti J, Gilson E, Giraud-Panis M-J. Mhr1, a yeast mitochondrial ATP-independent homologous pairing protein. J Biol Chem 2009;284(14):9341–53. https://doi.org/10.1074/jbc.M900323200

[20] Amiard S, Doudeau M, Pinte S, Poulet A, Lenain C, Faivre-Moskalenko C, Angelov D, Hug N, Vindigni A, Bouvet P, Paoletti J, Gilson E, Giraud-Panis M-J. A topological mechanism for TRF2-dependent strand invasion. Nucl Struct Mol Biol 2007;14(2):147–54. https://doi.org/10.1038/nsmb1192

[21] Shimohara A, Ogawa H, Ogawa T. Rad51 protein involved in repair and recombination. Genes Dev 1998;12:1248–53. https://doi.org/10.1101/gad.6.12.1248

[22] Machida S, Takaku M, Ikura M, Sun J, Suzuki H, Kobayashi W, Kimbara S, Iino T, Ando T. ATP-dependent unwinding of double helix in closed circular DNA by RecA protein of E. coli. Nature 1982;298(5879):86–9. https://doi.org/10.1038/298086a0

[23] Ohtani T, Shibata T, Iwabuchi M-A, Watabe H-O, Iino T, Ando T. ATP-dependent unwinding of double helix in closed circular DNA by RecA protein of E. coli. Nature 1982;298(5879):86–9. https://doi.org/10.1038/298086a0

[24] West SC, Cassuto E, Howard-Flanders P. Homologous pairing can occur before DNA strand separation in general genetic recombination. Nature 1981;290(5801):29–33. https://doi.org/10.1038/290029a0

[25] Yu X, Engelham EH. The RecA hexamer is a structural homologue of ring helicases. Nat Struct Mol Biol 1997;4(2):101–4. https://doi.org/10.1038/41250

[26] Lager K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nuclease core particle at 2.8 Å resolution. Nature 1997;389:251–60. https://doi.org/10.1038/3801

[27] Kobayashi W, Takaku M, Machida S, Tachinawa H, Maehara K, Okawa Y, Kuruma E. Chromatin architecture may dictate the target site for DMC1, building for RAD51, during homologous pairing. Sci Rep 2016;6(1). https://doi.org/10.1038/srep24228

[28] Alexeev A, Mazin A, Kowalczykowski SC. Rad54 protein possesses chromatin-remodeling activity that is stimulated by the Rad51–ssDNA nucleoprotein filament. Nucl Struct Mol Biol 2003;10(3):182–6. https://doi.org/10.1016/s1097-2041

[29] Michida S, Takaku M, Ikura M, Sun J, Suzuki H, Kobayashi W, Kinomura A, Osakabe A, Tachinawa H, Horikoshi Y, Fukuuto A, Matsuda R, Uka T, Tashiro S, Ikura M. Kuzumaki Y, Wang X. Triple-loops: formation, function, and relevance to cell stress. Cell Stress 2019;3:38–46.

[30] Allinson DF, Wang E, Williams JGK, Osber L, Radding CM. Stability and properties of double and triple helices: dramatic effects of RNA or DNA backbone composition. Science 1992;258:1466–6. https://doi.org/10.1126/science.1279808

[31] Shinohara T, Ikawa S, Iwasaki W, Hiraki T, Hikima T, Mikawa T, Arai N, Kamiya K, Latrasse D, Blein T, Liu C, Benhamed M, Crespi M. R-loop mediated trans chromatin remodeling activity. J Cell Sci 2015;128(1):2. https://doi.org/10.1242/jcs.180822

[32] Flory J, Tsang SS, Muniyappa K. Isolation and visualization of active RecA–ssDNA/dsDNA structures. Nature 2008;453(7194):489–94. https://doi.org/10.1038/nature06941

[33] Ariel F, Lucero L, Christ A, Mammarella MF, Jegu T, Veluchamy A, Mariappan P, Noirot P, Kolodner RD. DNA strand invasion promoted by Escherichia coli RecT protein. J Biol Chem 1998;273(20):12274–80. https://doi.org/10.1074/jbc.273.20.12274

[34] Kagawa W, Kuzumaki Y, Shibata T, Fukui S, Nureko O, Shibata T, Yosokuya S. Target structure of 3D homologous pairing domain from the human Rad52 recombinase in the undeformed complex. Mol Cell 2002;2(2):359–71. https://doi.org/10.1016/s1097-2041(02)00587-7

[35] Lopes A, Amarrir-Bouhrami J, Faure G, Petit MA, Guereés R. Detection of novel recombinases in bacterial rDNA operons. J Mol Biol 2017;429(4):757–65. https://doi.org/10.1016/j.jmb.2015.01.029

[36] Prasai K, Robinson LC, Scott RS, Tatchell K, Harrison L. Evidence for double-strand break mediated mitochondrial DNA recombination in Saccharomyces cerevisiae. Nucl Acids Res; 2017.

[37] Noirot P, Kolodner RD. DNA strand invasion promoted by Escherichia coli RecT protein. J Biol Chem 1998;273(20):12274–80. https://doi.org/10.1074/jbc.273.20.12274

[38] Dunn K, Chrysogelos S, Griffith J. Electron microscopic visualization of RecA–ssDNA/dsDNA structures. J Struct Biol 2001;136(2):222–4. https://doi.org/10.1006/jsbi.2001.2099

[39] Ragunathan K, Joo C, Ha T. Real-time observation of strand exchange reaction with high spatiotemporal resolution. Structure 2011;19(8):1064–73. https://doi.org/10.1016/j.str.2011.06.009

[40] Qi Z, Redding Sj, Lee J, Lai X, Kwon B, Yiu N, Haines W, Sung P, Greene E. DNA sequence alignment by microhomology sampling during homologous recombination. Cell 2015;160(5):856–69. https://doi.org/10.1016/j.cell.2015.01.029

[41] Duan K, Chrysogelos S, Griffith J. Electron microscopic visualization of RecA–DNA filaments: evidence for a cyclic extension of duplex DNA. Cell 1998;28(4):757–65. https://doi.org/10.1006/cell.1998.0345

[42] Chen Z, Yang H, Pavletich NP. Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. Nature 2008;453(7194):494–9. https://doi.org/10.1038/nature06842

[43] Yamaguchi Y, Suzuki-Nagata K, Shibata T, Mikawa T. Nonfilament-forming RecA dimer catalyzes homologous joint formation. Nucl Acids Res 2018;46(1):1-10. https://doi.org/10.1093/nar/gky1085

[44] Short JM, Liu Y, Chen S, Soni N, Madhusudhan MS, Shivji MK, et al. High-resolution structure of the presynaptic RAD51 filament on single-stranded DNA by electron cryo-microscopy. Nucleic Acids Res 2016;44:9017–30.
