Alignment of 3 (but Not 4) DNA Strands within a RecA Protein Filament*

Michael M. Cox†

From the Department of Biochemistry, University of Wisconsin-Madison, Madison, Wisconsin 53706

Homologous genetic recombination is required for recombinational DNA repair, which is involved in the successful segregation of chromosomes at cell division, and contributes to the generation of genetic diversity in all organisms. In order to recombine, the homologous sequences in two chromosomes must first be aligned. Alignment is preceded in most cases by the processing of one of the chromosomes to generate a region of single-stranded DNA. By mechanisms that are still unclear, a search for homology is then conducted in three dimensions which ends with homologous alignment. A central issue within this problem is the structure of the potentially novel DNA pairing intermediates that mediate a chromosome-chromosome interaction leading to a productive genetic exchange.

Structures in which 3 or even 4 DNA strands are interwound over a significant length, DNA triplexes and quadruplexes, have played a major role in thinking about homologous genetic recombination for decades. Stable triplex and quadruplex DNA structures are known (1–3), but all have special requirements for sequence and/or strand orientation. The structures proposed to occur during recombination must form regardless of the primary structure or base content of the sequence itself, and strands with identical sequences must be oriented parallel to each other. These constraints test the limits of current information about DNA structure. A putative DNA triplex structure proposed as a recombination intermediate, called R-form DNA, has recently been the target of a range of experimental and computer modeling efforts (4–6). Of necessity, most of the experimental manifestations of these ideas have focused on the RecA protein and the DNA strand exchange reactions it promotes. The status of these investigations is summarized here. Additional discussion of these topics may be found in several recent reviews (5, 7–13).

The RecA protein promotes an efficient DNA strand exchange reaction in vitro involving 3 or 4 strands (Fig. 1). RecA first forms a nucleoprotein filament on a circular ssDNA or a gapped duplex (DNA1). The length of DNA1 is increased by 50% or more, and any duplex portions of DNA1 are underwound by about 40% within this filament, with the presumed elimination of base-stacking interactions (14–16). A homologous linear duplex DNA (DNA2) is then paired with DNA1 to form what is sometimes termed a joint molecule. A rapid but limited DNA strand exchange can occur in the presence of an ATP analog, ATP$\gamma$S,1 that is not hydrolyzed by RecA. This phase of the reaction can occur anywhere on DNA2, typically produces 1–3 kilobase pairs of hybrid DNA, and is mediated by binding energy within the filament (17–20). This nascent hybrid DNA is extended in a final reaction phase that requires ATP hydrolysis and proceeds 5’ to 3’ relative to the ssDNA or the single-stranded gap in DNA1 (23). The ends of DNA2 where this unidirectional process normally begins and terminates are often referred to as the proximal and distal ends, respectively.

The R-form DNA Triplex

In 1984, Paul Howard-Flanders and colleagues (24) advanced the notion that 3 or 4 DNA strands could be aligned within the major groove of a RecA nucleoprotein filament. Evidence has accumulated in the intervening years that at least 3 DNA strands can be bound within the helical groove of a RecA filament (25–28). Additional reports provided evidence that a stable DNA triplex could be detected under some conditions after RecA protein had been removed (29–31), and much subsequent work has focused on these protein-free structures.

Joint molecules formed in RecA reactions on the distal end of DNA2 (where further strand exchange is constrained), and then deproteinized, have a reported stability comparable to double-stranded DNA (29, 31, 32). Detailed chemical probing and nuclease protection studies provided evidence that all 3 DNA strands in the deproteinized joint molecules were in a complex. Proposed triplex pairing schemes initially featured hydrogen bonds to the N7 of purines (29). However, guanine N7 is not protected from methylation by dimethyl sulfate in the putative triplex joints, and prior methylation of guanine N7 has no apparent effect on pairing (32, 33). In addition, duplex DNA containing purine analogs that replace N7 with carbon (7-deazaguanine and 7-deazadenine) in place of guanine and/or adenine are excellent substrates for DNA strand exchange (34). This indicates that interactions with purine N7 are not critical to a rate-limiting step in pairing or strand exchange. Methylation of N9 adenine or N4 cytosine did affect the stability of the joint molecules (32), leading to the triplex pairing proposal in Fig. 2 (32, 35). An alternative proposal by Zhurkin et al. (6) is similar but retains one triplex with a hydrogen bond at guanine N7. All of the current models are strikingly similar to a proposal made by Lacks in 1966 (36).

A homologous DNA strand will not associate spontaneously with a B-form duplex in solution at 37°C to create an R-form triplex. One assumption in these studies is that formation of an R-form triplex entails some hysteresis, with RecA protein providing the binding energy required to overcome a high activation barrier for formation of an otherwise stable structure. A consideration of possible DNA pairing pathways (Fig. 3) reveals a potential for two different forms of triplex DNA. Using terminology introduced by Radding and colleagues (37), these can be called “pre-strand switch” and “post-strand switch” triplexes. A pre-strand switch triplex would have DNA2 remaining Watson-Crick base-paired, with the single strand of DNA1 arrayed in its major groove. A post-strand switch triplex would have DNA1 base-paired with the complementary strand from DNA2, with the other strand from DNA2

* This minireview will be reprinted in the 1995 Minireview Compendium, which will be available in December, 1995.
† To whom correspondence should be addressed: Dept. of Biochemistry, University of Wisconsin-Madison, 420 Henry Mall, Madison, WI 53706. Tel.: 608-262-1181; Fax: 608-265-2603; E-mail: coxlabs@mac.wisc.edu.
§ This abbreviation used is: ATP$\gamma$S, adenosine 5’-O-(thiotriophosphate); ssDNA, single-stranded DNA; dsDNA, double-stranded DNA.

3 W. A. Bedale and M. M. Cox, submitted for publication.
4 S. K. Jain and M. M. Cox, unpublished results.

© 1995 by The American Society for Biochemistry and Molecular Biology, Inc.
Printed in U.S.A.

26021
Minireview: Alignment of 3 (but Not 4) DNA Strands within RecA Protein Filament

Alignment of 3 (but Not 4) DNA Strands within RecA Protein Filament

arrayed in the major groove of the hybrid DNA. Both structures can be modeled by Fig. 2, although the designation of which strands are involved in Watson-Crick base-pairing change. As pointed out by others (13, 38), pre-strand switch and post-strand switch triplexes defined in this way are not readily interconverted. If pairing is initiated by formation of a pre-strand switch triplex with the single-strand substrate in the major groove of the duplex, rotating strands to yield a hybrid duplex will place the displaced strand in the minor groove of the hybrid DNA product. Similarly, to form a post-strand switch triplex, the initial interaction between the DNA sub-
no evidence that psoralen derivatives would link a strand paired in the major groove as in Fig. 2 with the other 2 strands, and the absence of triplex structure in the electron microscope may indicate a selective destabilization of the triplex by surface tension or other considerations introduced by the spreading procedures.

A recent study of the properties of oligonucleotides with a capacity for intramolecular pairing and triplex formation has added some new evidence for the existence of R-form triplexes in solution (44). Thermal denaturation of the intramolecular triplexes, monitored spectrophotometrically, reveals a temperature-dependent increase in absorption that may reflect a triplex to duplex transition at temperatures of 25–30°C. In this study, triplexes with two free 3' ends were shown to be more stable than those with two free 5' ends, potentially explaining the enhanced stability of distal joints relative to proximal joints. However, the structures studied by Shchylkina et al. (44) did not exhibit a thermal stability approaching that reported for the triplexes derived from deproteinized joint molecules assembled by RecA action (29, 31, 32).

Although many issues remain to be resolved, continued study of deproteinized joint molecules, initiated by Camerini-Otero, Radding, and their colleagues (9, 10, 12, 29–32), seems likely to reveal a novel DNA structure that will form the basis for continued research on DNA-DNA interaction in recombination for years to come.

### DNA Quadruplexes in 4-Strand Exchange Reactions

The original Howard-Flanders proposal included up to 4 strands within the RecA filament, interwound in a quadruplex (24). The pairing scheme was based on one first presented by McGavin (45) and extended by Wilson (46). The McGavin structure has, in turn, provided the inspiration for much speculation about potential recombination intermediates in systems where two duplexes are presumed to be the substrates (46–49). Although the ongoing discussion often refers to work done with RecA, the extension of the Howard-Flanders formalism for RecA to include quadruplexes has not been well-supported by subsequent investigations. Many lines of evidence suggest that RecA-mediated DNA strand exchange in particular, and recombination in general, is fundamentally designed to initiate with 3 DNA strands.

1. Prokaryotic recombination systems include elaborate mechanisms involving multiple nuclease and helicase activities (RecBCD, RecJ, RecQ, etc.) designed to convert double-stranded DNA to single-stranded DNA, presumably to initiate recombination (50–52).

2. Nucleases and helicases have been implicated quite generally in situations requiring homologous recombination (conjugational recombination, transduction, plasmid recombination, and recombinational DNA repair) (52). In prokaryotes, there are no circumstances in which recombination has clearly been shown to require a duplex-duplex interaction mediated by RecA or a RecA-like protein.

3. If the function of RecA is to align two duplex DNA's, a facile binding of RecA to duplex DNA might be expected. Instead, RecA protein binds to ssDNA; binding to dsDNA occurs only after a very long lag at neutral pH values (53–55). This molecular design can be rationalized as a way to target RecA to regions where recombinational DNA repair is required (56) or to single-stranded substrates prepared for recombination by the enzymes mentioned above.

4. Studies probing the coordination of multiple DNA strands in the RecA filament consistently show a facile incorporation of 3 strands, but a very weak or nonexistent capacity to take up a fourth strand (25, 26). The physical evidence that binding of DNA in the RecA filament groove is limited to 3 strands is extensive and growing (13, 28).

5. RecA protein promotes DNA strand exchange reactions involving 4 strands, but they begin as 3-strand reactions. The single-strand gap that must be present in DNA1 is not only a loading point for assembly of the RecA filament, it is also the initiation point for the strand exchange reaction. DNA2 must overlap the gap by 50–100 base pairs to observe an optimal 4-strand exchange reaction (57, 58). Once DNA1 and DNA2 are aligned in a 3-stranded region, strand exchange can hypothetically proceed into the 4-stranded region without a further requirement for a 4-strand (quadruplex) interaction.

6. The limited strand exchange mediated by binding energy in a RecA filament in the presence of ATP·S will accommodate 3, but not 4, DNA strands (59). The same is true of the limited strand exchange promoted by a RecA mutant protein that binds but does not hydrolyse ATP (RecA K72R). 2 A strand exchange reaction involving two duplexes therefore exhibits an absolute requirement for ATP hydrolysis, while a limited 3-strand exchange can occur without ATP hydrolysis.

7. DNA2 must not only overlap the gap in DNA1, it must overlap on its proximal end so that the normal polarity of RecA-mediated DNA strand exchange will carry the reaction into the 4-strand region. In 3-strand reactions, DNA pairing occurs on the distal end of DNA2 that can transiently generate paired regions containing triplex or hybrid DNA that can be thousands of base pairs in length (37, 60). In 4-strand reactions where DNA2 overlaps the gap on its distal end, DNA pairing is limited to the 3-stranded region even when ATP is hydrolyzed (61).

8. A plethora of recent and related models proposed for the disposition of three separate binding sites for individual DNA strands in the RecA filament (12, 33, 39, 59, 62–64) have no 4-binding site counterparts. The proposed structures that have evolved for R-form triplexes (Fig. 2) leave no obvious place for a fourth strand. One could postulate an interaction with the free N7 position of the purines, but duplex DNA with guanine replaced throughout by 7-deazaguanine is as good a substrate for 4-strand exchanges as it is for 3-strand exchanges (34). It is particularly hard to envision a quadruplex with duplex-duplex interactions in the respective major grooves as a mechanism for the initial DNA pairing if, as the data increasingly indicate, the initial DNA-DNA interactions in 3-strand exchanges occur in the minor groove of the duplex (38).

In support of the alternative view, three observations are often cited as evidence for the existence of a 4-stranded DNA pairing intermediate.

1. Since DNA1 is stretched and underwound within the RecA filament, DNA2 must be underwound in order to align with it. In 3-strand reactions, a homology-dependent underwinding of DNA2 is readily detected (65, 66). If DNA2 is a nicked circle, the nick can be ligated while it is paired with a circular single-stranded DNA to trap any underwinding that may have occurred. The covalently closed and underwound DNA2 circle has a superhelical density consistent with pairing interactions extending over thousands of base pairs (66). Duplex-duplex pairing is much harder to detect, and the weak signals produced can be ambiguous. If DNA1 is a gapped duplex circle and homology with DNA2 is limited to the duplex portion of DNA1, ligation of DNA2 in a similar assay does not reveal underwinding of DNA2 that might signal duplex-duplex pairing (57, 58). An alternative assay uses supercoiled DNA2 and measures underwound DNA that is resistant to Escherichia coli topoisomerase I by virtue of its interaction with DNA1 within the filament. This assay produces a strong homology-dependent signal in the 3-strand reaction and a much weaker signal in the 4-strand reaction (57, 58). If DNA2 is relaxed at the beginning
of the experiment, only the 3-strand reaction yields a signal indicative of a pairing reaction. The weak signal found for the 4-strand reaction in one of 3 assays may indicate a duplex-duplex interaction. Alternatively, the requirement for DNA2 to be supercoiled (57, 58, 67) may lead to limited strand separation in a few molecules of DNA2, with only 1 strand being incorporated into the RecA filament. In the latter case, the weak signal would come from interactions involving only 3 of the 4 strands.

2. A RecA-mediated 4-strand exchange reaction can bypass a double-strand break in DNA2 (68). Whereas the first fragment of DNA2 initiates DNA strand exchange in the single-strand gap of DNA1, the second fragment must initiate at a site where DNA1 is presumably duplex. The mechanism of this bypass reaction has not been explored adequately. A few hundred base pairs of DNA1 beyond the break in DNA2 would need to provide a single-strand gap in DNA1 of ample size to permit the second fragment of DNA2 to initiate exchange as a 3-strand reaction.

3. The Holliday junction resolvase will not cleave Holliday junctions unless they occur within a 6–12-base pair region of the 4 strands. The Holliday junction resolvase will not cleave Holliday junctions unless they occur within a 6–12-base pair region of the 4 strands. The Holliday junction resolvase will not cleave Holliday junctions unless they occur within a 6–12-base pair region of the 4 strands.