RICE UNIVERSITY

Structural Characterization and Kinetic Evaluation of RecA-Mediated Strand Exchange

by

Jie Xiao

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Scott F. Singleton, Assistant Professor, Advisor,
Department of Chemistry and Department of Biochemistry and Cell Biology

Yousif Shaloo, Assistant Professor, Chair,
Department of Biochemistry and Cell Biology

Kathleen S. Matthews, Stewart Professor,
Department of Biochemistry and Cell Biology

Edward P. Nikonowicz, Associate Professor,
Department of Biochemistry and Cell Biology

R. Bruce Weisman, Professor, Department of Chemistry

HOUSTON, TEXAS
JULY, 2002
ABSTRACT

Structural Characterization and Kinetic Evaluation of

RecA-mediated Strand Exchange

by

Jie Xiao

The RecA protein of *Escherichia coli* plays essential roles in homologous recombination. *In vitro*, the protein mediates DNA strand exchange between single-stranded (ssDNA) and homologous double-stranded DNA (dsDNA) molecules that serves as a model system for the *in vivo* process. During this process, a key intermediate comprised of three DNA strands simultaneously bound to a RecA filament (RecA-tsDNA complex) forms. To date, the questions about the high-resolution structure of this intermediate and how the it forms have not been addressed.

We present a systematic characterization of the helical geometry of the three DNA strands of the RecA-tsDNA complex using fluorescence resonance energy transfer (FRET). Measurements of the helical parameters for the RecA-tsDNA complex revealed that all three DNA strands adopt extended and unwound conformations similar to those of RecA-bound dsDNA. The structural data are consistent with the hypothesis that this complex is a late, post-strand-exchange intermediate with the outgoing strand shifted by about three base pairs with respect to its registry with the incoming and complementary
strands. Furthermore, the bases of the incoming and complementary strands are displaced away from the helix axis toward the minor groove of the heteroduplex, and the bases of the outgoing strand lie in the major groove of the heteroduplex.

We then monitored the formation of the RecA·tsDNA complex in real time using a fluorescent base analog. Time-dependent changes of polarized emission from the fluorophore demonstrated this process involves at least three phases. The first phase is strongly dependent on substrate concentration and reaction temperature. Kinetic simulation revealed a sequential four-steps mechanism as the best description for the reaction. The association rate constant approaches the magnitude of $10^7$ M$^{-1}$s$^{-1}$. Thermodynamic analysis indicates the reaction is entropy favorable and a large activation energy is necessary for the association step. A comparison of the energy diagrams between a homologous, partially homologous and a completely heterologous dsDNA substrates suggests that the RecA protein discriminates homology both kinetically and thermodynamically. Based on these data, a structural and mechanistic model for RecA-mediated strand exchange was constructed.
In remembrance of my father

&

To my beloved mother
ACKNOWLEDGMENTS

This is the moment where I would like to express my greatest appreciation for all the people who have helped me, both intellectually and emotionally, through my years here at Rice. No words could possibly express how grateful I am for having all the care, encouragement and guidance from the wonderful people around me, yet I will do my best give the proper credit to them.

I owe a great debt of gratitude to my advisor Dr. Scott Singleton. His expertise in fluorescence spectroscopy and enzyme kinetics and his perfectionism in scientific writing and presentations have benefited and prepared me to become the scientist I am today. It has been a great experience and privilege to learn and work with him, whether it be in running experiments in the lab, talking about the projects or even going to a conference with him.

I feel fortunate to have Dr. Yousif Shamoo, Dr. Kathleen Matthews, Dr. Edward Nikonowicz and Dr. Bruce Weisman as my committee members. Their scientific expertise and high efficiency in dealing with my every single request have been amazing. All my progress reviews with them have been pleasant and inspirational for my work. I feel I couldn’t have any better committee members.

I am also very grateful to my fellow colleagues: Dr. Alberto Roca for his thorough understanding of the RecA protein and great experimental skills; Dr. Michael Berger for being the “mechanic” of equipment and for keeping the lab lively; Andrew
Lee for continuing the kinetic project and assisting my research; and Neil Sharma for cloning *recA* gene constructs and being the most persistent researcher I have ever seen. Special thanks go to Rebecca Simonette for being expert in protein purification and a great friend in my life.

I also want to acknowledge other fine people in the Department for making my experience in Rice an absolutely pleasant one: Dr. Fred Rudolph for being the most humble and warm-hearted chairman I have ever met; and Dolores Schwartz and Jo Gerla for being very helpful and kind in making me feel I am always welcome at the office.

Finally, I would like to thank my husband Bo for his love, encouragement and patience with me. As for my father and mother, I can not say enough for their sacrifice and endless love that they gave me. They are the dearest people of my life, and I owe my every success to them.
Table of Contents

Abstract ii
Acknowledgements v
Table of Contents vii
List of Figures ix
List of Tables xi

Chapter 1 Introduction 1
Recombinational DNA Repair in DNA Metabolism 6
In Vitro Recombination: the DNA Three-Strand Exchange Reaction 10
RecA Protein is Active as a Multimeric Helical Filament 17
Towards a Mechanism for Stand Exchange: A Two DNA binding Site Model for RecA Protein 20
Structure of RecA Protein Filaments 25
The RecA Protein Two DNA binding Site Model Revisited 28
Mode of Homology Recognition 31
Is the Synaptic Complex A DNA Triplex? 34
Is R-DNA the Key to Homology Recognition? 37
The Base Pair Model for Homology Recognition 40
Formation of the Heteroduplex DNA 41
Kinetics of Homology Searching and Pairing 42
The Role of the RecA Protein in Strand Exchange 45
Experimental Goals 46

Chapter 2 Elucidating a Key Intermediate in Homologous DNA Strand Exchange: Structural Characterization of the RecA-triple-Stranded DNA Complex Using Fluorescence Resonance Energy Transfer 48
Introduction 49
Theory 54
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results</td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
<td>93</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td></td>
<td>118</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Construction and Evaluation of a Kinetic Scheme for RecA-Mediated DNA Strand Exchange in the Absence of ATP</td>
<td>134</td>
</tr>
<tr>
<td>Hydrolysis</td>
<td></td>
<td>135</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
<td>136</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td>195</td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
<td>218</td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Summary</td>
<td>222</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Cartoon depicting the role of the active RecA-ssDNA filaments</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Unrooted phylogeny of RecA/RadA/DMC1 superfamily</td>
<td>7</td>
</tr>
<tr>
<td>1.3</td>
<td>Potential pathways for the nonmutagenic re-establishment of inactivated replication forks in bacteria</td>
<td>9</td>
</tr>
<tr>
<td>1.4</td>
<td>General model for homologous recombination</td>
<td>11</td>
</tr>
<tr>
<td>1.5</td>
<td>Cartoons depicting strand exchange reaction using long bacteriophage-derived DNAs</td>
<td>14</td>
</tr>
<tr>
<td>1.6</td>
<td>Schematic drawing of RecA-mediated strand exchange</td>
<td>16</td>
</tr>
<tr>
<td>1.7</td>
<td>Schematic representation of the structure of the RecA protein and RecA(dsDNA) filament</td>
<td>19</td>
</tr>
<tr>
<td>1.8</td>
<td>A two DNA binding site model for RecA-mediated strand exchange</td>
<td>23</td>
</tr>
<tr>
<td>1.9</td>
<td>A modified two DNA binding site model for RecA-mediated strand exchange</td>
<td>29</td>
</tr>
<tr>
<td>1.10</td>
<td>Base-pair and base-triple models of homology recognition</td>
<td>32</td>
</tr>
<tr>
<td>1.11</td>
<td>Details of hydrogen-bonding among corresponding bases of the three strands</td>
<td>35</td>
</tr>
<tr>
<td>2.1</td>
<td>Chemical structures of fluorophores and DNA sequences</td>
<td>52</td>
</tr>
<tr>
<td>2.2</td>
<td>Schematic representation of DNA helix geometry model</td>
<td>56</td>
</tr>
<tr>
<td>2.3</td>
<td>Fluorescent DNA reaction systems</td>
<td>58</td>
</tr>
<tr>
<td>2.4</td>
<td>Agarose gel mobility shift assay</td>
<td>63</td>
</tr>
<tr>
<td>2.5</td>
<td>Spectral data characterizing the fluorescein-Rox reaction systems</td>
<td>67</td>
</tr>
<tr>
<td>2.6</td>
<td>Energy transfer efficiency between fluorescein and Rox at different base-pair separations</td>
<td>72</td>
</tr>
<tr>
<td>2.7</td>
<td>Global nonlinear least square analysis of E-n data for three RecA·tsDNA</td>
<td>86</td>
</tr>
<tr>
<td>2.8</td>
<td>Energy transfer efficiency between fluorescein and Rox in the presence of ADP-ALF4</td>
<td>88</td>
</tr>
<tr>
<td>2.9</td>
<td>Reduced x^2 values plotted as a function of the fitted parameters</td>
<td>92</td>
</tr>
<tr>
<td>2.10</td>
<td>Models depicting possible arrangement of the bases corresponding to the tree DNA strands</td>
<td>113</td>
</tr>
<tr>
<td>3.1</td>
<td>Chemical structure of 6MI and sequences of oligonucleotides</td>
<td>138</td>
</tr>
</tbody>
</table>
Figure 3.2  Cartoon drawings illustrating experimental designs  
Figure 3.3  Agarose gel mobility shift assay  
Figure 3.4  Steady-state spectra of the 6MI-labeled DNA and the RecA(DNA) complexes in homologous and heterologous reactions  
Figure 3.5  Apparent dissociation constant measurement on Reaction I  
Figure 3.6  Three typical stopped-flow traces of Reaction I, C and O  
Figure 3.7  Overview of the concentration-dependence of Reaction I  
Figure 3.8  Concentration-dependence of the total amplitude changes and the reciprocal of relaxation times of Reaction I, C and O  
Figure 3.9  Overview of concentration dependence of heterologous reactions  
Figure 3.10  Concentration-dependence of the total amplitude changes and reciprocal of relaxation times of Reaction Het1 and Reaction Im  
Figure 3.11  Overview of the temperature-dependence of Reaction I  
Figure 3.12  Response factors of each species in Reaction I, C and O  
Figure 3.13  Thermodynamic analysis of strand exchange  
Figure 3.14  A comparison of the first phase kinetics between this and previous studies  
Figure 3.15  Free energy of the transient state of the first step in Reaction Im and Het1  
Figure 3.16  A molecular model depicting the structural and kinetic aspects of the strand exchange
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Spectral characteristics of fluorescein and Rox measured in different complexes</td>
<td>66</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>The energy transfer efficiency and distance between fluorescein and Rox in five reaction systems</td>
<td>71</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Fitted parameters using equation (6) for each reaction system</td>
<td>75</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>Summary of experimental helical geometry parameters and dye-linker structural parameters for five complexes</td>
<td>76</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Characteristic combination of emission maximum, relative fluorescence intensity and anisotropy for each RecA complex</td>
<td>156</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Exponential curve fit parameters for strand exchange reactions with increasing substrate concentration</td>
<td>167</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Exponential curve fit parameters for strand exchange reactions with increasing heterologous substrate concentration</td>
<td>178</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>Exponential curve fit parameters for Reaction I at different temperatures</td>
<td>179</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>Rate constants of Reaction I, C, and O at different substrate concentrations solved by global simulation based on the sequential four steps mechanism</td>
<td>186</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>Rate constants and equilibrium constants of Reaction I at different temperatures solved by simulation based on the sequential four steps mechanism</td>
<td>192</td>
</tr>
</tbody>
</table>
CHAPTER 1

Introduction
In living organisms, there is a dynamic balance between the need to preserve genomic information and the need to generate genetic diversity. The repair of damaged DNA is essential to the maintenance of heritable genetic information, while the variation of that information drives evolutionary adaptation. In *Escherichia coli*, the RecA protein mediates events in both pathways (Figure 1.1). Because its functions are conserved from bacteriophage to humans mammals, the RecA protein’s study has provided a paradigm for understanding these essential biological processes (for review, see 1). Although extensive in vitro biochemical studies over the last twenty years have characterized this remarkable protein’s structure and diverse activities, the molecular mechanisms underlying the complex biological processes in which it participates are not yet understood (2).

The discovery of the *recA* gene was marked by the paper “Isolation and Characterization of Recombination-deficient Mutants of *Escherichia coli* K12” more than three decades ago (3). An analysis of the conjugational recombination phenotypes of two *E. coli* mutants led the authors to hypothesize that “one gene determines an enzyme, which catalyzes one of the steps in recombination and one of the steps in the replacement of photoproducts in ultraviolet-damaged DNA”. Further studies on the formation of φX174 recombinant DNA corroborated the hypothesis that the *recA* gene was involved in recombination (4,5). The ensuing investigations revealed that *recA* mutants were pleiotropic, affecting not only homologous recombination but also DNA repair, DNA
Figure 1.1 Cartoons depicting the activated RecA-ssDNA nucleoprotein filament formed in the presence of ATP, and the four classes of activities mediated by the activated filament.
mutagenesis, cell division, and phage induction (6). Elucidating the basis of this complex phenotype proved to be challenging. However, an important clue was provided by the discovery that the product of the recA gene, the RecA protein, activates the λ cI repressor for autoproteolysis with concomitant deactivation of its DNA binding capacity (7). Shortly thereafter, similar essential roles for the RecA protein were demonstrated in the inactivation of the LexA repressor (8), which depresses more than 30 genes giving rise to the S.O.S. response (9), and activation of the UmuD protein (10-12) which is required for mutagenic translesion DNA synthesis by PolV (UmuD'2C) (13-15). In each of these cases, RecA apparently facilitates the autocatalytic cleavage of each protein.

Upon the discovery that the RecA protein plays essential roles in the derepression of other genes, it appeared likely that the RecA protein’s roles in the biological events mentioned above was simply to affect the levels of expression of those other genes. With the cloning of the recA gene (16) and subsequent purification of the RecA protein (17-21), it was discovered that RecA protein displays an in vitro DNA-dependent ATPase activity (17,22). In combination with the results of Holloman and Radding implicating RecA protein in the pairing of a single strand DNA fragment with superhelical DNA to produce a displacement loop (D-loop) (5), this activity emphasized the importance of RecA protein’s interaction with DNA and suggested a direct involvement of RecA protein in recombination and/or repair. This possibility was further supported by the observed effect of a temperature-sensitive recA allele on recombination in the absence of
protein and RNA synthesis (23). These results stimulated experiments demonstrating that RecA protein catalyzed the renaturation of complementary ssDNAs (24) and the pairing of two homologous DNA molecules (25-27). The most compelling evidence for RecA protein’s role in recombinational events in biological contexts is provided by observation on RecA protein’s \textit{in vitro} strand exchange activities. Likewise, RecA protein’s facilitation of this reaction serves as a model of the protein’s roles in recombinational events.

It is now clear that the \textit{recA} genetic pleiotropy reflects a remarkable duality of RecA protein activity (Figure 1.1) (28). On the one hand, RecA protein specifically facilitates the autoproteolysis of certain repressors. On the other hand, RecA protein is a DNA-dependent ATPase that promotes homologous pairing of DNA molecules. These activities allow RecA protein to carry out at least three \textit{in vivo} DNA metabolic functions (28-31). First, RecA controls the SOS response for DNA damage tolerance (32). In this capacity, the formation of a complex (a “nucleoprotein filament”) between RecA and single-stranded DNA (ssDNA) serves as an initiating signal for the derepression of many DNA repair genes (33) and the activation of translesion DNA synthesis (34,35). Second, the RecA protein mediates homologous genetic recombination (36), a process that may allow lateral gene transfer between species (37) and has been implicated in several putative mechanisms of adaptive (stationary phase) mutation in bacterial populations (38,39). Third, RecA plays an indispensable role in DNA damage repair, participating
directly in double-strand break repair (DSBR) (40) and stalled replication fork restart (29,41,42). RecA protein’s roles in the second and third physiological processes are functionally related in that they require the pairing of a ssDNA molecule with a molecule of double-stranded DNA (dsDNA) containing a strand nearly identical in sequence to that of the ssDNA. These DNA metabolic functions are conserved across phylogenetic kingdoms, and structural and functional homologs of the RecA protein are found in all living organisms from bacteriophage to mammals (Figure 1.2) (for a review, see 36). The wide distribution and high conservation of this protein over evolutionary space emphasizes the important functions of this protein (29,36,43-45)

RECOMBINATIONAL DNA REPAIR IN DNA METABOLISM

Based on the accumulation of evidence over the years, RecA protein’s likely biological functions can be divided – at least conceptually – into four categories depending on their effect on genomic information (Preservation vs. Variation) and the protein’s role (Direct vs. Indirect) (Figure 1.1). A comparative evaluation of the four categories suggests that RecA protein’s active roles are characterized by the interaction between a nucleoprotein filament and dsDNA in the context of recombinational processes. As first pointed out by Campbell in his summary of an influential 1984 meeting at Cold Spring Harbor (46), the real biological function of recombination systems is enigmatic and often ignored. Nine years later, Cox elegantly summarized two
Figure 1.2 Unrooted phylogeny of the RecA/RadA DMC1 superfamily from (Leipe et al., 2000). (Red) Eukaryota; (green) Archaea; (blue) Bacteria; (pink) Bacteriophages. Names in boxes identify the individual protein families. Where applicable, source organisms are identified by four-letter abbreviations. (Aepe) Aeropyrum pernix; (Aqaе) A. aeolicus; (Arfu) Archaeoglobus fulgidus; (Arth) A. thaliana; (Basu) Bacillus subtilis; (Bobu) Borrelia burgdorferi; (T7) bacteriophage T7; (T4) bacteriophage T4; (Cael) C. elegans; (CDnaB_Odsi) Odontella sinensis chloroplast; (CDnaB_Popu) Porphyra purpurea chloroplast; (Chtr) Chlamydia trachomatis; (Ecol) E. coli; (Glma) Glycine max; (Hain) Haemophilus influenzae; (Hepy) H. pylori; (Hosa) Homo sapiens; (Lema) Leishmania major; (Meja) Methanococcus jannaschii; (Meth) Methanobacterium thermoautotrophicum; (Mumu) Mus musculus; (Myge) Mycoplasma genitalium; (Mytu) Mycobacterium tuberculosis; (Plch) P. chabaudi; (Rhma) Rhodothermus marinus; (Sace) Saccharomyces cerevisiae; (SPP1) Bacillus subtilis bacteriophage SPP1; (Suso) Sulfolobus solfataricus; (Sy68) Synechocystis PCC6803; (Teth) Tetrahymena thermophila; (Thma) T. maritima; (Trpa) Treponema pallidum.
competing perspectives underlying the thinking at the time: the recombination paradigm and the repair paradigm (47). The former view supposed that the primary function of recombination was the creation of genetic diversity, while the latter view supposed that DNA repair was the primary function and that recombination events arose as a side-product of repair. Although perspectives emphasizing the recombination paradigm remain influential (36), the repair paradigm has become the central perspective (42) as stated in a recent progress report from six independent laboratories (48): “The major function of homologous genetic recombination in bacteria, and a major function in virtually all cells, is the nonmutagenic recombinational DNA repair of stalled or collapsed replication forks.”

Recombinational DNA repair is an exact method of DNA repair in which a recombination protein binds to ssDNA resulting from the demise of a replication fork, and pairs the damaged DNA with intact, homologous dsDNA to initiate repair without mutation (Figure 1.3). Kuzminov proposed a model for reactivating stalled replication forks in bacteria that accounted for a diverse array of results regarding *E. coli* replication and recombination (49). Following this model, the idea developed that replication fork reactivation is an important function that operates frequently during normal growth of bacteria (42,50-53).

Recombinational repair likely evolved in the context of DNA metabolism to overcome replication fork arrest and maintain chromosome integrity (37,42,48,54). The
**Figure 1.3** Some potential pathways for the nonmutagenic re-establishment of inactivated replication forks in bacteria. The pathways shown illustrate two of the important situations during normal cell growth that may result in replication fork demise, encounter with a DNA lesion or a DNA strand break. Reactivation involves the two main homologous genetic recombination pathways. The processes shown are broadly based on some published studies and discussions at recent national meetings; however, many of the details shown are speculative. The configurations of DNA strands shown in the intermediates are neither representative of all the proposals for fork reactivation nor intended to represent anyone's ideas of the most likely paths. (Reprinted from Cox et al., 2000, Nature, 404, 37-41)
universal conservation of the RecA family of recombination proteins emphasizes their importance in DNA maintenance and implies their presence in the earliest forms of life (55,56).

Due to the essential nature of DNA repair and replication fork maintenance, bacteria have developed multiple DNA processing complexes which modulate RecA activity and complicate direct observation of the intricate mechanisms of DNA metabolism (48,57-59). Thus, the molecular mechanisms of RecA protein’s role remain controversial (41,60), despite the fact that the important functions of bacterial DNA recombination systems, and RecA protein in particular, in repairing stalled replication forks are nearly universally recognized.

**IN VITRO RECOMBINATION: THE DNA THREE-STRAND EXCHANGE REACTION**

Models for homologous genetic recombination generally divide the process into several distinct phases (Figure 1.4) (for review, see 31,61). Pioneering work to delineate the molecular mechanism led to the recognition of the probable importance of single strands in recombination (26,62). Thus, the first phase ("synapsis") consists of a pairing reaction between two homologous dsDNA molecules – at least of one of which contains a single stranded region – leading to the formation of joint molecules containing heteroduplex regions in one or both of the molecules. In the next phase the heteroduplex regions are moved or extended in a reaction known as "branch migration." Such
Figure 1.4 General model for homologous recombination. See text for details. (Adapted from Bianco et al., 1998, Front Biosci, 3, D570-603)
heteroduplex regions may be several thousand base pairs long. Subsequent steps ("resolution") serve to convert the branched structures to recombinant products.

In an effort to discern how a homologous single strand might invade dsDNA, Radding and coworkers demonstrated that superhelical DNA spontaneously took up homologous ssDNA fragments under certain in vitro solution conditions to form joint molecules that contain displacement loops (D-loops) (63,64). Subsequently, Holloman and Radding found that a mixture of superhelical DNA of phage φX174 and fragments of homologous ssDNA underwent recombination in spheroplasts of E. coli (5); moreover, the recombination efficiency was dependent on the presence of a wild-type recA gene and could be rescued in recA− hosts by pre-forming the joint complexes described above. The laboratories of Radding and Lehman then independently demonstrated that purified RecA protein was capable of pairing a ssDNA fragment with a homologous superhelical DNA to form D-loops (18,25). When combined with the earlier in vivo findings regarding recA mutants, these results firmly established a direct role for RecA protein in the first phase of homologous recombination.

Most of the work to delineate the mechanism by which RecA protein promotes heteroduplex joint formation emphasized importance of ssDNA in the initial pairing. Indeed, ssDNA plays an important role as both an effector and a substrate in RecA protein activities. In its effector role, ssDNA activates both the ATPase activity (17,22) and the coprotease activity (65) of RecA protein. As discussed above, the pairing
reaction requires ssDNA, and the reaction proceeds optimally under conditions in which
the amount of RecA protein is stoichiometric with the amount of ssDNA nucleotides (nts)
(26,62). Furthermore, ssDNA stimulates RecA protein to bind dsDNA (18) and change
its topology or helical geometry (19,66), even in the absence of homology. Based on
these observations, Shibata et al. (1979) inferred that the binding of ssDNA to RecA
occurred prior to synapsis and the properties of the RecA-ssDNA complex governed its
subsequent interaction with dsDNA (26). Although the model was not universally
accepted at that time (25), most recombination models now propose the initial, pre-
synaptic interaction of RecA with ssDNA followed by synapsis of the complex with a
homologous dsDNA (67) (For a recent discussion, see 68).

The RecA protein-dependent D-loop, or pairing, assay provided an important in
vitro model system for early events in homologous recombination. Moreover, the assay
permitted a remarkably thorough characterization of the D-loop formation reaction with
respect to its substrate specificity, solution conditions preferences, etc. within a short
period of time (25,27,62,66,69-76). For example, the Lehman and Radding groups found
that homologous pairing promoted by RecA protein does not share the requirement for a
superhelical DNA substrate demonstrated by protein-free uptake of ssDNA by
superhelical DNA (18,25). These experiments were carried out using substrates derived
from bacteriophage DNA (circular, single-stranded) and linearized replicative form DNA
as shown in Figure 1.5.
Figure 1.5: Cartoons depicting substrates — circular single-stranded DNA (cssDNA) and linear duplex DNA (dsDNA) — synaptic intermediates, and products of strand exchange reaction with bacteriophage-derived DNAs.
In the context of long regions of homologous DNA substrate pairs, extension of a D-loop involves migration of a crossover junction, the point at which one strand’s Watson-Crick pairing partner switches from one DNA molecule to the other. This process, termed branch migration, would be expected to occur after RecA protein-promoted homologous pairing. Radding and coworkers observed that D-loop structures generated by RecA protein contained heteroduplex regions that extended for thousands of base pairs, implying that branch migration does occur either spontaneously or as facilitated by the action of RecA protein (77). The Howard-Flanders and Lehman laboratories subsequently demonstrated that RecA protein does indeed promote the exchange of strands between DNA substrates, not simply their homology-dependent pairing (71,75). By using substrate dsDNA containing a free end, Cox and Lehman demonstrated that RecA protein could promote a complete strand exchange reaction, yielding nicked circular dsDNA and linear ssDNA as products, wherein the pairing and branch migration phases were kinetically distinct (71). Importantly, all reactants, intermediates, and products in such an assay are easily separated and readily identified using gel electrophoresis (71). Most current hypotheses describing the participation of RecA protein in biological recombination events rely substantially on our understanding of such DNA three-strand exchange reactions promoted by RecA protein (Figure 1.6).

The in vitro observations described above indicate that RecA protein can promote all phases of homologous recombination: synapsis, branch migration, and resolution.
**Figure 1.6** Schematic drawing of RecA-mediated strand exchange. During the presynapsis phase, the RecA protein (gray balls) polymerize on an incoming single-stranded DNA (blue circle) to form a nucleoprotein filament. This nucleoprotein filament then pairs on a homologous double-stranded DNA (composed of one complementary strand, red, and one identical strand, green, which is referred as the outgoing strand) to form a RecA-triple-stranded DNA complex (RecA-tsDNA) during the synapsis phase. In this RecA-tsDNA complex, the original base-paired complementary strand switch its partner and forms a heteroduplex DNA with the incoming strand. The above three steps require the presence of ATP but not its hydrolysis. Finally, through branch migration and in the presence of ATP hydrolysis the outgoing strand is displaced.
Hence, unlike replication and transcription, the cardinal events of homologous recombination can be accomplished by a single polypeptide of M, 37,842 (352 amino acids) (78). Based on the characteristics of ssDNA uptake by superhelical DNA in the absence of (63,64), Radding, Shibata, and their coworkers enumerated four components necessary for the DNA strand exchange reaction to occur: (i) unfolding of the ssDNA to make its bases available for pairing, (ii) unwinding of the dsDNA substrate, (iii) formation of base pairs between homologous regions of the ssDNA and unwound dsDNA, and (iv) formation of a product dsDNA (termed the "heteroduplex") containing the substrate ssDNA, rather than rewinding of the original dsDNA (26,66). This statement led to the question of how such a complex reaction could be catalyzed by a single protein of modest size.

RECA PROTEIN IS ACTIVE AS A MULTIMERIC HELICAL FILAMENT

Purified RecA protein was found to migrate as tetramers and higher order aggregates in glycerol gradient sedimentation experiments (17), a finding subsequently corroborated by EM visualization of RecA protein-only filaments (79,80). Early in vitro studies found that RecA protein’s optimal activities could only be achieved when the number of molecules of protein in solution were nearly equivalent to the number of nucleotides (for ssDNA) or base pairs (for dsDNA) of DNA (i.e., the protein was required at stoichiometric, rather than catalytic, concentrations). Moreover, the
dependence of each relative activity on RecA protein concentration was sigmoidal (18,25,26). These observations led to the hypothesis that the ssDNA must be fully covered by RecA protein in order to properly pair with its homologous dsDNA target (26,71). Electron microscopy (EM) subsequently provided direct evidence that RecA protein does indeed polymerize on both gapped, circular double-stranded DNA (81,82) and circular, single-stranded DNA (80,83) to form filamentous complexes (Figure 1.7). These in vitro studies were corroborated by in vivo studies demonstrating that RecA protein levels reached \( \approx 10^5 \) copies per cell (84) and that defective recA genes introduced into cells containing the wild-type recA gene showed dominance (85,86). The negative complementation observed in the latter experiments is a signature for a protein that functions as a multimer.

The conclusion that RecA protein functions as a homopolymer provided part of the answer to the question as to how a single protein could mediate homologous recombination. Specifically, the right-handed helical filament formed by RecA polymerization on DNA is well suited for the tasks of homology-dependent pairing and strand exchange over extended lengths of DNA. Moreover, the organization of multiple identical polypeptide chains into a filament provides a structure in which sequence specificity is determined by DNA-DNA interactions, rather than protein-DNA interactions optimized for each member of a set of unique polypeptides (87).
Figure 1.7 Schematic representation of the three-dimensional crystal structure of the *E.coli* RecA protein monomer (A), polymer (B) (adapted from Story RM et al., *Nature*, 1992, 355, 318-325) and the transparent surface electron microscopic model of the helical nucleoprotein filament formed by the RecA protein on double-stranded DNA (adapted from Science cover, 259, 1993. Ogawa et al., *Science*, 1993, 259, 1896-1899). The RecA monomer consists of three domains: a core domain with eight b-strands bounded by eight a-helices, an amino domain with a large a-helix and a short b-strand and a carboxyl domain with three a-helices and three b-sheets. The RecA polymer is a right-handed filament with a 6 k società axis (shown in white). The nucleoprotein filament observed under EM is about 120 Å wide with 6 RecA monomers per turn. The dsDNA (red) inside the filament is extended about 50% and untwisted 17° comparing to its standard B-form.
Much of the understanding of the structure of RecA nucleoprotein filaments came from EM studies of RecA-dsDNA filaments, which are characterized by a pitch of about 95 Å with 6.2 monomers per helical turn (88). Although the presynaptic filaments formed by RecA on ssDNA were much more dynamic and prone to collapse (80,83), active presynaptic filaments were eventually isolated and EM characterization suggested that they were similar in morphology to the RecA-dsDNA complexes (89). Interestingly, EM also revealed that the contour lengths of RecA-dsDNA filaments, formed in the presence of ATP (or its slowly hydrolyzed analog, ATPγS), were 50–60% longer than those of the underlying protein-free replicative form DNAs from bacteriophage (82,83,90), and allowed the determination of the helical rise per base pair, \( h = 5.2 \pm 0.18 \) Å (82). Thus, synapsis requires the homology-dependent pairing between the strand in the presynaptic filament and naked dsDNA, substrates whose interbase spacing is substantially different (5.2 Å vs. 3.4 Å). Given the structural disparity between the two DNA substrates, a key question arose: How is homology recognition accomplished?

**TOWARDS A MECHANISM FOR STRAND EXCHANGE: A TWO DNA BINDING SITE MODEL FOR RECA PROTEIN**

Radding, Shibata, and their coworkers suggested the importance of a ternary complex formed by RecA protein, ssDNA, and dsDNA (26,61,69,70,77). They postulated that RecA protein brings ssDNA and dsDNA into proximity and then partially (or locally) unwinds the dsDNA in order to allow the search for homology (69,70,77).
Although early evidence suggested that RecA protein promoted the unwinding of dsDNA (62,66), it was demonstrated that the strands of the duplex were not separated prior to homologous pairing (27,76) and the unwinding data were later interpreted (91) to be consistent with the structure of RecA protein-bound dsDNA, which is characterized by a reduction in twist accompanied by interbase elongation (82,83,88,90,92,93). In contrast, the likely intermediacy of the ternary complex was well substantiated by filter binding, electrophoretic mobility shift, density gradient sedimentation, and EM assays (18,24-27,69,77). Thus, the formation of a ternary RecA-ssDNA-dsDNA complex that preceded homologous alignment became accepted as a key step in DNA strand exchange and homologous recombination (28,73,94).

Thus, one part of the answer to the question of how molecular recognition of homology is accomplished is that the presynaptic filament has a second binding site for dsDNA. This inference was corroborated by the observation that a RecA-bound dsDNA is immune to pairing with RecA-bound ssDNA (95,96). Moreover, hexagonal crystals of RecA protein formed at low pH (97) indicated that the spiral filaments consist of RecA monomers arranged head-to-tail in equivalent positions, rather than in a sequence of dimmers or tetramers (98,99). Taken together with reconstructed images of RecA-dsDNA filaments from EM suggesting that the helical complexes were too large to allow two filaments to juxtapose two DNA molecules (98), the structural data reinforced the concept that each RecA monomer in a helical filament has binding sites for two DNA
molecules. These considerations led Howard-Flanders and his coworkers to construct a two DNA binding site molecular model for RecA-mediated strand exchange (98,99) (Figure 1.8). In this model, the incoming ssDNA occupies "site I" of the RecA filament during the presynapsis phase. Binding of one strand of dsDNA at "site II" initiates the synapsis step. The RecA protein may rotate the bases of the incoming ssDNA and the complementary strand of the duplex to form a new heteroduplex DNA. These ideas were almost immediately substantiated by EM images of pairing intermediates (100,101). From this point on, the groundwork for a mechanistic understanding of the RecA-mediated strand exchange reaction has been laid, and most subsequent work has followed this same idea (102-104).

In the context of the DNA strand exchange reaction mediated by the RecA protein, the two DNA binding site hypothesis provides a model for the physiological processes related to homologous recombination. The strand exchange reaction can be separated into four conceptually distinct phases: presynapsis, synapsis, heteroduplex formation, and branch migration (for reviews, see 29,31,61) (Figure 1.6). Presynapsis is the sequence-independent assembly of multiple RecA monomers on the incoming ssDNA to form a multimeric right-handed helical nucleoprotein filament with the DNA strand bound at site I of RecA protein. During synapsis, a dsDNA molecule, composed of strands whose sequences are identical (the outgoing strand) and complementary to that of the incoming strand, is aligned in a sequence-specific manner with the homologous
Figure 1.8 A two DNA binding site model for RecA-mediated strand exchange proposed by Howard-Flanders et al. (Howard-Flanders et al., Nature, 1984, 309, 215-220). The reaction starts with a ssDNA binds to site I of the RecA protein, then this initial complex binds to one strand of a homologous dsDNA at site II. The RecA protein rotates the bases of the incoming and complementary strands to form heteroduplex DNA and displaces the outgoing strand. Please note that in this scheme, the incoming ssDNA recognize homology through the major groove of the duplex DNA.
ssDNA in the nucleoprotein filament forming a three-stranded nucleoprotein complex (RecA·tsDNA). As discussed by Howard-Flanders et al. (1984), the alignment presumably depends upon sequence-specific interstrand hydrogen bonds between DNA bases. Although the two strands of the dsDNA substrate could be separated to allow Watson-Crick pairing, several studies failed to detect dsDNA strand separation prior to pairing (27,76,91,105), and models in which the ssDNA pairs in the major groove of the fully base-paired dsDNA were favored (98,99). Within this nascent three-stranded complex, a rapid and partial strand switch ensues, yielding a “joint molecule” containing a relatively short region of heteroduplex product. Synapsis and heteroduplex formations are rapid and non-polar with respect to the initiating nucleoprotein filament. The first three phases require the presence of ATP (or its analogs, such as ATPγS or ATP·AlF₄⁻) but not its hydrolysis. Finally, the heteroduplex DNA is extended until products are formed in a relatively slow and highly polar branch migration reaction that requires ATP hydrolysis.

Over the past twenty years, extensive work has been carried out in order to understand the strand exchange reaction. In particular, numerous investigations to delineate the structure of the presynaptic filament (106-109) and to elucidate the role of ATP hydrolysis in branch migration (110-123) have allowed relatively clear understandings to be achieved. Relevant reviews can be found elsewhere (29,124-130). However, knowledge of the precise structures and mechanistic strategies involved in the
middle two phases, synapsis and heteroduplex formation, is still conspicuously lacking. This deficit is not surprising given the multiphasic nature of strand exchange and the multiple functions of the RecA protein. Because the two phases determine how the homology of the dsDNA is searched and recognized and how the DNA strand switch is initiated, a better understanding of these two phases will undoubtedly shed light on the corresponding cellular process of homologous genetic recombination.

For the reasons discussed above, we designed our reaction systems deliberately aiming only at the synapsis and heteroduplex formation steps by using specific reaction conditions as described in Chapters 2 and 3. We were able to dissect the process and answer some key questions in regard to the structural and mechanistic mode of the RecA-mediated strand exchange reaction. In order to provide a context for our work, we reviewed previous studies focusing on the structural and kinetic aspects of the synapsis and heteroduplex formation steps of the reaction.

STRUCTURE OF RECA PROTEIN FILAMENTS

Over the past 20 years, considerable progress has been made in determining the structures of the RecA protein as well as those of the RecA-ssDNA and the RecA-dsDNA filaments. EM provided the first images of the RecA nucleoprotein filament (reviewed in 124). It was observed that the cofactor had a drastic effect on the observed pitch of the helical filament. A "collapsed" filament with a helical pitch of ca. 75 Å occurs in the absence of a cofactor or with ADP. The collapsed nucleoprotein filament is similar to the
spiral RecA protein-only filament. In contrast, an “extended” filament (95-Å pitch) is observed with nonhydrolyzable analogs of ATP. The extended conformation is the active form of the RecA nucleoprotein filament since the quaternary complex formed between RecA protein, ssDNA, ATP and Mg$^{2+}$ in vitro is the species that pairs and exchanges homologous DNA strands (89,114,131,132). The crystal structure of a RecA-ADP complex provided the first high resolution view (2.3 Å) of the protein-only and RecA-ADP filaments (Figure 1.7), corroborating many of the structural inferences derived from lower resolution EM images (133). However, the conformation of the filament in the crystal is probably that of the inactive collapsed filament since the pitch (83 Å) is not as extended as that seen when DNA is bound (134). The active conformation is achieved only in the presence of both DNA (either single- or double-stranded) and ATP (or a slowly- or non-hydrolyzable analog, such as ATPγS or ADP·AlF$_4^−$, respectively).

EM of RecA-dsDNA complexes has provided the most consistent and highest resolution images of RecA nucleoprotein filaments. As introduced above, early EM images revealed that the contour lengths of RecA-dsDNA filaments were 50 - 60% longer than that of the underlying protein-free replicative form DNAs from bacteriophage (82,83,90), and allowed the determination of a helical rise per base pair, $h = 5.2 \pm 0.18$ Å (82).

While EM has allowed RecA nucleoprotein filaments to be directly visualized and a likely DNA binding location to be identified (see below), it has not provided for the
direct characterization of the DNA strands buried inside. A key limitation in this regard is the relative masses of the filament components: the two DNA strands in a RecA-dsDNA complex account for less than 5% of the total mass of the complex (135,136). Nevertheless, a comparison of the number of turns of the overall nucleoprotein filament images with the number of bp's in the substrate dsDNA revealed that the average pitch of the dsDNA was 18.6 bp ($\Omega_n = 19.4^\circ$) (88). Using an assay to monitor changes in the topological linking number of supercoiled DNA, Radding and coworkers first demonstrated a RecA-induced reduction in the twist of substrate dsDNA (66). Subsequently, several labs have used related assays to quantify the extent of unwinding, reporting helical repeats of 18.6 ± 1.3 (92), 17 ± 1 (93), 21 ± 5 (90), and 17.9 ± 0.3 bp (137). Thus, in spite of a number of critical assumptions and limitations in the topological assays (for a discussion see 92,93,137,138,139,140), a remarkable agreement with the prediction derived from EM images has been achieved.

It has been proposed that the stretched and partially unwound DNA structure imposed by RecA binding is instrumental for its strand exchange activity (for representative current perspectives, see 44,126,141,142,143). Based on this idea, it is speculated that the three DNA strands inside the RecA-tsDNA complex, a key intermediate formed during the strand exchange reaction, would also be extended and unwound (144). An accurate description of the DNA structure in this complex should be able to elucidate the mechanistic strategy employed by the RecA protein and its
homologs. The RecA-tsDNA complex that forms during the strand exchange reaction has been visualized by EM (100,101). However, despite the success of the structural characterization of the RecA-ssDNA and RecA-dsDNA complexes using EM and classical biochemical methods, the quantitative structural details of the RecA-tsDNA complex have not been elucidated in previous studies, partially due to its complexity and heterogeneity (145).

**THE RECA PROTEIN TWO DNA BINDING SITE MODEL REVISITED**

While EM has been invaluable for the direct visualization of RecA nucleoprotein filaments, no data have allowed unambiguous direct structural characterization of the DNA within. Using EM, Egelman & Yu (136) observed density differences between the RecA-ssDNA and RecA-dsDNA complexes that identify the likely location of one DNA strand. Specifically, the EM images reveal that both strands bind deep within the helical groove of the RecA protein filament in sites that physically overlap. Thus, one strand of the dsDNA inside the RecA filament occupies the same position that ssDNA does. By combining this structural data with the ideas that the RecA-ssDNA and RecA-dsDNA filaments represent the initial substrate binary complex and final product binary complex (103,146) in the strand exchange reaction, respectively, a revised molecular mechanism can be envisaged (Figure 1.9). This mechanism is consistent with the general features of the Howard-Flanders two DNA binding site model described above. Some experimenters
Figure 1.9 A modified two DNA binding site model for RecA-mediated strand exchange proposed by Mazin et al. 1996, PNAS. The reaction starts with a ssDNA binds to site I of the RecA protein, then this initial complex binds to both strands of a homologous dsDNA at site II. The RecA protein rotates the bases of the incoming and complementary strands to form heteroduplex DNA in site I and displaces the outgoing strand in site II. Upon ATP hydrolysis, the outgoing strand is released from site II.
have proposed that a third binding site (site III) is necessary to accommodate the other strand of the duplex DNA (147-150).

It is now generally agreed that the function of binding site I in strand exchange involves its abilities to accommodate the incoming ssDNA and maintain the extended DNA structure. On the other hand, the investigation of the function of site II (or sites II and III) in terms of its roles in strand exchange are ongoing. In this context, Mazin et al. (1996) proposed that the presence of site II in the RecA filament serves two functions in strand exchange: at the beginning of the reaction, site binds to dsDNA weakly, thereby facilitating the homology searching process; after strand exchange, the secondary site binds the newly-displaced outgoing ssDNA more tightly to stabilize the reaction (103). The first putative role for site II is supported by other work suggesting that the binding of the second DNA molecule to site II is fast – independent of sequence homology – whereas rapid dissociation is favored in the absence of complementarity (151,152). The second putative role for site II is supported by the experimental inference that strand exchange in the absence of ATP hydrolysis is mediated by binding energy within the filament (117,132,153).

To completely unravel the mysteries of which site is occupied by which DNA strand, and how the strands at different sites are exchanged, the structure and formation of the RecA-tsDNA complex, a key intermediate for this process, have to be understood. As described above, though the RecA-tsDNA complex has been visualized under EM, its
structure was far less well documented than the RecA-ssDNA or dsDNA complexes. Thus, the need for our structural and kinetic characterization of the RecA·tsDNA complex is obvious.

MODE OF HOMOLOGY RECOGNITION

The central question as to how the homology between the RecA-bound ssDNA and the target dsDNA is recognized remains intriguing. Following the initial proposal by Howard-Flanders and coworkers (1984), two limiting models have been proposed: the base-triplet and the base-pair models (Figure 1.10) (for a representative review, see 150). In the base-triplet model, non-Watson-Crick hydrogen bonding contacts between the bases of the incoming ssDNA and the intact base pairs of the target duplex DNA are made in order to identify the homologous sequences (141,154-157). In the base-pair model, however, the target dsDNA is at least partially melted prior to pairing, then the homologous sequences are recognized by forming regular Watson-Crick interactions between the complementary strand of the duplex DNA and the incoming ssDNA (158,159). Importantly, the two models require different mechanistic strategies and kinetics. For the base-pair model, the second and third phases of the strand exchange reaction, homologous pairing and initiation of strand switch, respectively, are concurrent (160). In contrast, for the base-triplet model, homologous pairing precedes the strand switch and a unique intermediate is required.
Figure 1.10 Base-pair and base-triple models of homology recognition. The gray ovals represent the RecA protein (different shades represent different binding sites), and the incoming, complementary and outgoing DNA strands are color-coded as blue, red and green, respectively. The yellow bars between bases represent Watson-Crick hydrogen bonds and the black dash lines represent non-Watson-Crick hydrogen bonds.
The delineation of the synaptic complex structure and the mechanistic strategy employed by RecA protein is important for understanding how the following two homology recognition requirements are fulfilled: the recognition must provide sequence specificity while allowing the presynaptic filament bound to non-homologous dsDNA to be recycled. The contact between the incoming ssDNA and the target dsDNA has to be specific enough to maintain the fidelity of homologous recombination, yet when homology is not identified, the two DNA molecules should be readily dissociated so that the filament is able to launch another round of searching. Conceptually, the base-pair model more easily fulfills the first requirement whereas the base-triplet model is more feasible for the latter one. The proposed non-Watson-Crick hydrogen bonding between the incoming ssDNA and the duplex DNA in the base-triplet model does not possess the same sequence specificity as the base-pair model (161). For example, the thymine of the incoming strand can form nearly identical hydrogen bonds with a GC base pair as with an AT pair in the pre-exchange triplet proposed by Rao et al. (154). On the other hand, the much weaker interactions (162) within the putative base triplets would prevent the nucleoprotein filament from being trapped, allowing the rapid scanning of a large population of dsDNA molecules for homology (141).

During RecA protein-mediated strand exchange, the three DNA strands must associate with one another in order to allow molecular recognition of homology and strand switching. This suggests that a synaptic RecA·tsDNA complex must exist – at
least transiently – as an important intermediate for this process. Indeed, visualization of RecA-tsDNA complexes by EM (100,101,163,164) suggests that the homology search occurs prior to dsDNA strand separation within a RecA protein-bound three-stranded DNA complex. Thus, all of the putative synaptic intermediates for the alternative mechanisms are triple-helical joint molecules in the sense that all three strands are RecA protein-bound and wound around a common helical axis. However, the intermediates are not equivalent in terms of the hydrogen bonding contacts between strands. Thus, we will refer to the complexes generally as “three-stranded” (e.g. RecA-tsDNA). Those complexes in which hydrogen bonding is present among all three strands, with at least one set of non-Watson-Crick bonds, will be referred to as “triplexes”. The complexes in which hydrogen bonding only occurs between two strands using Watson-Crick bonding exclusively will be referred to as “nontriplexes”.

**IS THE SYNAPTIC COMPLEX A DNA TRIPLEX?**

Because the structure of the synaptic RecA-tsDNA complex provides the basis for homology recognition, it has elicited great attention. Lacks first proposed a DNA triplex, in which the bases of the dsDNA substrate are paired by normal Watson-Crick hydrogen bonds while the ssDNA lies in the major groove of the duplex parallel to the identical (outgoing) strand and paired with the major-groove edge of the intact bp’s of dsDNA by non-Watson-Crick hydrogen bonds, as the tsDNA structure (Figure 1.11) (165). This
Figure 1.11 Details of hydrogen-bonding among corresponding bases of the three strands (adapted from Bertucat et al., Biophys J, 1999, 77, 1562-1576). The bases of the three strands are color-coded as in panel A. From top to bottom, the triplets are C:GC, G:CG, A:TA and T:AT. Please note that the minor groove triplets (left) are converted to major groove triplets (right), and vice versa.
idea was later reintroduced in the context of the Howard-Flanders molecular model for RecA protein-mediated strand exchange (98,99).

Although early EM work allowed a RecA-tsDNA complex to be visualized (100,101,164,166), its structural features could not be characterized. Camerini-Otero and his coworkers pointed out that triplex and non-triplex (i.e., a heteroduplex and a displaced strand) intermediate joints should be functionally distinct (167). In particular, these authors argued that a triplex joint should be refractory to dissociation by spontaneous branch migration (i.e., protein-independent exchange of hydrogen bonds) and that such deproteinized structures should be demonstrably stable. Thus, the isolation and characterization of tsDNA complexes after removal of RecA protein from synaptic joints affirmed the idea that a DNA triplex was the key intermediate in homology recognition (154,167-173). Importantly, these complexes must be prepared under specific solution conditions using long DNAs with the presence of topological or heterologous constraints (168,169,174). Moreover, the deproteinized joints are not uniform (168) and have stabilities that depend on the position of homology relative to heterologous sequences (distal, medial, or proximal) (169).

Specific DNA triplex model structures were proposed by the laboratories of Camerini-Otero (167) and Radding (154,171). The structures of the constituent triplets were similar to those originally proposed for the homologous recombination intermediate by Lacks (165). This triplex structure was subsequently dubbed "recombination-form
DNA” (R-DNA) by Zhurkin, Camerini-Otero, and their colleagues (155), and is characterized by the following features: (i) all three strands are sequence-specifically hydrogen bonded, (ii) the third strand is parallel to its identical strand in the dsDNA, (iii) the third strand lies in major groove of the dsDNA, (iv) the complex is not restricted with respect to sequence class at which it can form, and (v) the complex has a uniform structure. Based on computer simulations, Zhurkin et al. (155) proposed that both extended (rise = 5.1 Å) and collapsed (rise = 3.4 Å) R-DNA structures were viable and that the third strand was stabilized by the sequence-specific electrostatic interactions with the major groove edge of the bp’s in the duplex DNA. That such triplets could be the basis of an energetically reasonable RecA-bound DNA triplex has been further supported by subsequent computer simulations (141,171,175) as well as the observation of protein-free R-form triplexes formed by oligonucleotides (176-178).

**IS R-DNA THE KEY TO HOMOLOGY RECOGNITION?**

Much of the work described above is consistent with the hypothesis that a RecA protein-bound triplex DNA (R-DNA) is the intermediate that allows the molecular recognition of homology. However, several concurrent works demonstrated that the modification of N7 on the major groove edge of guanine does not affect pairing or strand exchange (154,159,179). Moreover, the results of Adzuma’s chemical and enzymatic probing of the RecA-ttsDNA complex (not deproteinized) were inconsistent with expectations derived from a protein-bound triplex DNA (159). Thus, while Camerini-
Otero and coworkers interpreted data for the deproteinized DNA complexes in terms of a pre-strand-exchange R-DNA (155,167,172,180), Radding and coworkers interpreted similar data to be consistent with a post-strand-exchange R-DNA (171,174). In a pre-exchange triplex, the incoming ssDNA makes additional hydrogen bonds with the intact dsDNA substrate (154), while in the post-exchange triplet, the complementary strand has switched its base pair partner to the incoming strand, and it is the outgoing strand that makes the non-Watson-Crick contact with the heteroduplex DNA (170).

The base triplets proposed to comprise the pre- and post-exchange triplexes were similar but differed in the strands to which the bases are assigned. The structure proposed by Radding and coworkers involved the incoming and complementary strands base-paired in the Watson-Crick sense and the outgoing strand associated through contacts in the major groove of the heteroduplex. By virtue of the major groove placement of the outgoing strand, the incoming strand must have inserted itself via the minor groove of the substrate dsDNA (Figure 1.4B). In contrast, the structure proposed by Camerini-Otero and coworkers involved the location of incoming strand in major groove of the substrate dsDNA, a conformation consistent with homologous pairing before strand separation.

The laboratory of Dervan (181) pointed out that stereochemical constraints would prevent the facile interconversion of pre-exchange and post-exchange R-DNA triplexes. Indeed, the initiation of strand exchange in minor groove of the substrate dsDNA would
lead to the displacement of the outgoing strand into the major groove, while initiation of exchange in the major groove would lead to a heteroduplex minor groove location for the displaced outgoing strand (181). Thus, the investigation of the existence and relevance of R-DNA was delimited by the question as to whether homology recognition initiates in the minor or major groove of the substrate dsDNA.

This question is of additional significance for the base-triplet model because the functional groups of an base pair differ in the major and minor grooves (167,182,183). Thus, for a base-triplet mechanism, the corresponding structures of the triplets and the kinetic influence of key functional groups must differ between the two grooves. The major groove possesses a less symmetric array of functional groups as a basis for discrimination, while the minor groove can distinguish only between guanine and other bases (due to the unique guanine N2 amino group) (161). Based on this consideration, to maintain the fidelity of homologous recombination, the major groove should be favored to discriminate homology if such a pre-exchange triplet exists. Indeed, such an intermediate was proposed by Rao et al. (154). Nevertheless, computational modeling of the putative triplets indicates that both major groove and minor groove triplets are energetically feasible and capable of discriminating homology (141,155).

The question of minor or major groove invasion has been addressed in several laboratories over the past decade by detecting the presence of the outgoing strand in the minor or major groove of the newly-formed heteroduplex DNA. In general, the previous
studies generated rather controversial results. The hypothesis that homology is recognized via the major groove has been supported by computational modeling (155) and radioprobeing experiments (184), while the alternative hypothesis – recognition of homology from the minor groove – has been supported by chemical modification (159,171), atomic mutagenesis (179), ligand displacement (182,183), and affinity modification (158,181,185) experiments as well as computational modeling (141). Indeed, the latter set of results are consistent with a structural model in which the RecA·tsDNA complex is an advanced, post-strand-exchange intermediate wherein the outgoing strand has been unpaired from the original dsDNA but remains bound to the nucleoprotein filament. However, this conclusion has recently been questioned on the basis that the interpretation of the experimental results depends on the structural model of the RecA·tsDNA complex (184). A direct structural characterization of the DNA constituents of the RecA·tsDNA strand exchange intermediate would offer insight into this mechanistic issue, thereby providing indirect information on how homologous recombination and recombinational DNA repair are mediated in vivo.

THE BASE PAIR MODEL FOR HOMOLOGY RECOGNITION

The base-pair model is indirectly supported by the extension and unwinding of the duplex DNA by the RecA·ssDNA filament in regardless of its homology (66), the accessibility of the outgoing strand by chemical probes (159), the inability to photocrosslink among the three DNA strands (158) and the increased strand exchange
efficiency of AT-rich dsDNA substrates (160,186,187). One interesting and insightful mechanism proposed by Radding and coworkers is the A·T base pair flipping model in accordance with the original base pair model. The A·T base pair flipping hypothesis suggests that the key to kinetic homology discrimination is the breathing of A·T base pairs in the dsDNA substrate, and the consequent ability of the A and T bases to flip out of the helix and form new base pairs with the complementary bases on the incoming ssDNA (160). One notable thing about this model is that it does not require the opening of the whole duplex DNA prior to pairing: a few base pairs flipping may be enough to allow the homology be identified and act as a nucleus for the following elongation of heteroduplex DNA.

One of the key issues that leads to the undetermined homology recognition mechanism is that the structure of the RecA·tsDNA complex, a key intermediate formed during the strand exchange, has not been reported to date. By determining the conformation of the three DNA strands inside the nucleoprotein filament and how the three strands are organized with respect to each other, the controversies concerning the alternative base-pair/base-triplet mechanisms and the minor/major groove invasion models could be resolved.

**FORMATION OF THE HETERODUPEX DNA**

In their landmark 1984 paper, Howard-Flanders and coworkers advanced the notion that the mechanism of the formation of the heteroduplex DNA would involve the
RecA-guided base rotation of the incoming ssDNA (99). The most compelling evidence for this mechanism of heteroduplex formation comes from a recent NMR study (188). This report suggested that the conversion of the sugar puckers between the N- and S-type leads to the switching of the complementary partners of the two DNA molecules. However, a recent computational study confirmed the potential importance of the base motions but also demonstrated that heteroduplex formation could occur without changes in nucleoside sugar pucker (141). Moreover, the sugar-pucker model cannot address the question whether the homology recognition is mediated via the minor or major groove of the substrate dsDNA based on the reversible conversion of the two types of the sugar pucker. The A:T base pair flipping model proposed by Gupta et al. (1999), suggests that the base pair switching first starts from the less hydrogen-bonded A:T base pairs, and the homology recognition and heteroduplex DNA formation are integrated into one step (160, 186, 187). This integration is the natural consequence of the base-pair model as discussed above.

**KINETICS OF HOMOLOGY SEARCHING AND PAIRING**

While understanding the DNA structures during RecA-mediated strand exchange is instrumental, the kinetics aspects of this reaction cannot be ignored. The alternative models for homology recognition necessarily imply the existence of different transient intermediates during the active process of strand exchange. By asking questions about how fast the nucleoprotein filament scans homology and mediate strand exchange, and
what are the possible impact of DNA lesions such as mismatches on this process, useful insights could be offered for a more complete understanding of homologous DNA recombination in vivo.

Historically, the kinetic study of the strand exchange has been characterized by several false starts. Early studies using long (3 to 10 kb) natural DNA sequences derived from different replication forms of bacteriophage m13 or ΦX174 suggested that the search was facilitated by the concentration of the DNA within large networks (coaggregates) that are formed by multiple short-lived contacts of the filament with the dsDNA (189-191). In this context, the search was characterized as being processive (192), first order (193), and, under certain circumstances (e.g., low RecA-ssDNA filament to dsDNA molecule ratio), rate limiting for strand exchange. Moreover, the formation of networks limited the range of the search for homology such that it was largely confined to neighboring duplex molecules (194). The searching rate has been estimated at about 0.2 kb/s for a nucleoprotein filament formed on a 6-kb ssDNA molecule (194). However, given the short cell cycle (20 mins) and the genome size (4700 kb) of E. coli, the RecA-ssDNA filament must scan the genome DNA at a rate greater than 4 kb/s in regardless of the size of the ssDNA inside of the RecA filament. Thus, the DNA strand exchange that occurs between viral DNAs in vitro was a poor model for the efficient search that occurs in vivo (194). This disagreement further led to the suggestion that the formation of the
large coaggregates may be an artifact due to ionic strength (195,196) and the topological complexities imposed by the long DNA substrates (197,198).

By using relatively short oligodeoxyribonucleotides (ODNs), later work avoided the long-range interactions as described above, and the kinetics of strand exchange has been evaluated in a much more quantitative way. It was demonstrated that the homology searching is not the rate-limiting step for the DNA strand exchange reaction (199-202) and that it is apparently second order (199,201). A recent work by Adzuma demonstrated that the homology search by a RecA-ss40mer filament on a 3kb duplex DNA does not appear to be mediated by facilitated diffusion or sliding (203). In the earliest attempts to define a kinetic scheme for strand exchange using ODN substrates, Camerini-Otero and coworkers reported a bimolecular association rate constant for synapsis, \( k = 10^3 \text{ M}^{-1}\text{s}^{-1} \) (199). Subsequent experiments conducted using different assays under various solution conditions by the Radding laboratory reported that the bimolecular association rate constant was ca. \( 10^6 \text{ M}^{-1}\text{s}^{-1} \) (201,202). Similar apparent kinetics were reported by Shaner and coworkers (200); however, this study reported only observed relaxation times rather than true kinetic rate constants. The differences between the Camerini-Otero and Radding reports have not been reconciled. The latter authors made a more careful evaluation of the true kinetics by simulating the time-dependent concentration changes of the reaction species. In order to simplify the calculations, however, Bazemore et al. (1997) made the assumption that the observed signal changes were solely due to the
concentration changes of the reactants and products, *i.e.*, no contribution of the intermediates to the observed signal was considered. This assumption was demonstrated to be invalid in their own following-up experiments (160,202). Thus, a true kinetic scheme evaluated in terms of a molecular mechanism for the RecA-mediated strand exchange reaction is still lacking, and our kinetic studies in Chapter 3 specifically targeted the delineation of such a scheme.

**THE ROLE OF THE RECA PROTEIN IN STRAND EXCHANGE**

The RecA protein was initially thought to have a rather active role in the strand exchange reaction, wherein it would act like an enzyme which *promotes* the exchange of two identical DNA strands between two DNA molecules (for representative reviews, see 31,87). Several laboratories have postulated that the RecA protein exerts its role actively by protein-DNA interactions. Specifically, by stretching or melting the DNA molecules, the RecA protein prepares them for homologous alignment, and by directing the base pair rotation, the RecA protein drives the strand exchange reaction to completion. To date, there is no direct evidence unambiguously supporting an active role of the protein in the strand exchange reaction.

Interestingly, a recent work measuring the binding of the RecA protein to stretched dsDNA revealed that the RecA protein binds to stretched dsDNA preferentially over canonical B-DNA (204). The stretched dsDNA can naturally occur during thermal fluctuation, and the binding of the RecA protein could stabilize the high-energy state of
the stretched, naked dsDNA. Similarly, a work exploring the role of sugar pucker conversion as the basis for formation of the heteroduplex found that the energy barrier for the conversion is sufficiently low that the base rotation can be induced just by thermal motions (188). Recent calculations of RecA-bound triplexes led the authors to conclude that RecA takes advantage of the intrinsic properties of stretched triplexes rather than actively promoting strand exchange (141). Thus, an active role of the RecA protein in driving the base pair switching does not appear necessary.

In summary of these works, it is possible that the RecA protein mediates rather than promotes the strand exchange reaction. Indeed, its roles may simply involve directing the proper approach of the substrate dsDNA, trapping it in an extended form, and providing an appropriate microenvironment in which the DNA molecules make contact with each other freely and switch partners depending on available thermal energy. The governing force would then be the DNA-DNA interactions instead of protein-DNA interactions.

**EXPERIMENTAL GOALS**

Our long-term objective is the characterization of site-specific interactions among the filament constituents in order to reach a molecular understanding of recombination and recombinational repair by the construction and elaboration of three-dimensional models of the intermediates in the reactions. Based on the discussion above, we framed two specific experimental questions:
• What are the structures of the DNA strands bound within the RecA-tsDNA complex formed during the strand exchange reaction?

• What are the kinetics underlying the strand exchange reaction?

By addressing the first question, we anticipate providing direct insights into the mechanism of the homology recognition. By addressing the second question, we should be able to obtain substantial information about the mechanistic strategy employed by the RecA protein. By combining these two together, we endeavored to construct a mechanistic scheme for the strand exchange reaction mediated by the RecA protein and its homologs, which, in turn, would offer valuable information on how homologous recombination and recombinational DNA repair are mediated in vivo.
CHAPTER 2

Elucidating a Key Intermediate in Homologous DNA Strand Exchange: Structural Characterization of the RecA-Triple-Stranded DNA Complex Using Fluorescence Resonance Energy Transfer

*DNA Structure During RecA-mediated Strand Exchange*
INTRODUCTION

The clearest evidence for RecA's role in recombinational events in biological contexts is derived from the in vitro DNA strand exchange reaction facilitated by RecA. In addition, much of our current understanding of structure-function relationships for RecA originates from investigations of this reaction. During such a DNA strand exchange reaction, three DNA strands must be at least transiently juxtaposed by a multimeric RecA filament. A knowledge of the structures of the intervening RecA-DNA complexes would be invaluable for ascertaining the mechanistic strategy by which RecA effects recombination.

Over the past twenty years, considerable progress has been made in determining the structures of the RecA protein-only and protein-DNA filaments. Electron microscopy (EM) of RecA-dsDNA complexes has provided the most consistent and highest resolution images of RecA nucleoprotein filaments, allowing them to be directly visualized and a likely DNA binding location to be (135,136). A combination of EM and classical biochemical methods has provided a consistent and well-accepted structural model wherein the dsDNA is stretched by ca. 50% and partially unwound to a right-handed helix with ca. 19 bp per turn (82,88,90,92,93,137). It has been proposed that the stretched and partially unwound DNA structure imposed by RecA binding is instrumental in the process of strand exchange (for representative current perspectives, see 44,126,141,142,143).
The work summarized above has been instrumental in characterizing the RecA-ssDNA and RecA-dsDNA nucleoprotein filaments. These complexes can be thought of as the first binary (protein-DNA) substrate complex and the final binary product complex of the strand exchange reaction. The association of the former complex, with the ssDNA extended and underwound with respect to B-DNA, with dsDNA results in the formation of a RecA-triple-stranded-DNA complex (RecA-tsDNA), which serves as the key intermediate for the reaction. EM of RecA-DNA complexes trapped during sequential stages of in vitro strand exchange have allowed visualization of genetic recombination intermediates (100,101). In spite of this success, no structures of RecA-tsDNA complexes that resolves protein monomers or DNA strands have been reported.

One long-term objective of characterizing the structures of the filament components is to reach a molecular understanding of recombination and recombinational repair by the construction and elaboration of three-dimensional models of the intermediates in the reactions. Moreover, by developing an accurate description of the DNA structure during strand exchange, it should be possible to elucidate the mechanistic strategy employed by the RecA protein and its homologs. To date, even the basic elements of that strategy are not universally agreed upon. For example, the answer to the question of whether RecA mediates homology recognition via major- or minor-groove contacts with the substrate dsDNA remains elusive. Although the issue has been addressed in several laboratories over the past decade, those experiments have produced
controversial results whose interpretation depends on the structural model of the RecA-tsDNA complex (for recent discussion, see 141,184). Direct structural characterization of the DNA constituents of the RecA-tsDNA strand exchange intermediate would offer insight into these mechanistic issues as well as providing valuable information on how homologous recombination and recombinational DNA repair are mediated in vivo.

In the work described herein, we systematically employed fluorescence resonance energy transfer (FRET) efficiency measurements to characterize the helical geometry of the RecA-tsDNA complex in RecA-mediated strand exchange. The experiments relied on FRET between the donor fluorescein, conjugated to an amino-modified deoxythymidine (dT\(^A\), Figure 2.1a), and the similarly conjugated acceptor carboxy-X-rhodamine (Rox). Since the distance between the two dyes is a function of the base pair separation and the helical parameters (i.e., helical rise and twist angle), measuring the energy transfer efficiency between the two dyes enabled us to measure the distance between the two dyes and, hence, the DNA's helical geometry. We envisage several potential advantages of this method for DNA structural characterization in nucleoprotein filaments. First, it is important to note that both key helical geometry parameters are determined simultaneously in the same experiment. Second, because the fluorescence signals originate from the bound DNA strands, the measurements directly characterize the local DNA structure rather than the overall filament. Finally, the non-invasive
Figure 2.1 Chemical structures of fluorescein, carboxy-X-rhodamine (Rox), and deoxythymidineamine (dTA) (a), and sequences of single-stranded DNA molecules (b) and double-stranded DNA molecules (c) used in the study. All ssDNA and upper-strand dsDNA sequences are written from 5’ to 3’, while the lower, complementary strands of dsDNA read from 3’ to 5’. E: fluorescein-conjugated dTA; R: Rox-conjugated dTA.
labeling of the DNA strands with fluorophores also enables the FRET measurements to be made under physiologically relevant solution conditions, which introduce the least amount of perturbation into the RecA-DNA complex.

Prior to undertaking an analysis of the RecA·tsDNA complex, we used FRET efficiency measurements to characterize a RecA·dsDNA complex. The resulting helical geometry parameters agreed with the well-accepted structural model of RecA-bound dsDNA, providing an important confirmation of the FRET-based method. Because the FRET measurements using oligonucleotides provide accurate helical geometry information over distances comparable to the radius and pitch of the RecA nucleoprotein filament, the approach described here can reveal structural information complementary to that derived from bacteriophage DNAs with length scales 10–100-fold larger.

In the measurements using RecA·tsDNA complexes, one of the three DNA strands was labeled with Rox, while one of the remaining strands was labeled with fluorescein (Figure 2.1b and c). Hence, all three DNA strands within the strand exchange intermediate were characterized. All three strands are extended and unwound to a similar extent as RecA-bound dsDNA. The data are clearly in accord with recognition of homology occurring via minor groove contact, since the bases of the incoming and complementary strands are displaced away from the helix axis toward the minor groove of the heteroduplex while the bases of the outgoing strand lie in the major
groove of the heteroduplex. The resulting structural model can account for much of the previously published data regarding the mechanism of homologous pairing.

**THEORY**

**Fluorescence Resonance Energy Transfer**

The efficiency of Förster energy transfer ($E$) between a FRET donor-acceptor pair is quantitatively related to the inverse sixth power of the distance ($R$) between the two dyes:

$$E = \frac{R_0^6}{R_0^6 + R^6}$$  \hspace{1cm} (1)

where $R_0$ is the Förster critical distance (in Å) between donor and acceptor at which the energy transfer efficiency is 50% \( (205) \). Therefore, a slight change in the distance between the donor and acceptor leads to a large change in the energy transfer efficiency. FRET has been successfully employed as a “molecular ruler” in a variety of biomolecular systems \( (\text{for representative reviews, see 206,207-211}) \). In the RecA field, FRET has been applied to a description of the kinetic steps \( (200,201) \) and polarity \( (186) \) of RecA-mediated strand exchange.

For a DNA double helix with a FRET donor and acceptor conjugated to opposite strands, the distance between the two dyes should reflect the helical parameters such as rise per base pair \( (h) \) and twist angle per base pair \( (\Omega_n) \). \( (\text{The symbols and nomenclature describing the DNA geometry are those suggested by 212}) \). Indeed, it was demonstrated
by the elegant experiments of Clegg and coworkers that FRET signals display a characteristic modulation when the fluorophores "walk" around the helix (213). Therefore, by measuring the energy transfer efficiency between the two dyes, the structural information characterizing the DNA helix could be obtained. In our study, we systematically shifted the position of Rox along one DNA strand while maintaining a constant position of fluorescein on another strand. This arrangement allows the energy transfer efficiency between the two dyes to be expressed as a function of the characteristic helical parameters. These helical parameters can be readily obtained by fitting the experimental data to a simple DNA helix model (213), which is described briefly below.

**Geometric model of DNA helix**

For dsDNA formed from oligonucleotides, the geometry of the DNA helix can be represented in cylindrical coordinates (Figure 2.2). A unique, straight helical $z$-axis, which is perpendicular to the base pair plane, is assumed. The distance $R$ between donor and acceptor is expressed as a function of the number of base pairs ($n$) separating the donor (d) and acceptor (a), the helical parameters $h$ and $\Omega_\alpha$, and a set of parameters describing the geometry of the linker connecting the dyes to the DNA:

$$R^2 = r_0^2 + r_s^2 + (nh + \Delta h)^2 - 2r_0r_s \cos(n\Omega_\alpha + \Delta \Omega_\alpha)$$  \hspace{1cm} (2)
Figure 2.2 Schematic representation of DNA helix geometry model. The central helix axis $z$ is perpendicular to the base pair plane. The distance $R$ between donor and acceptor is related to the following parameters by equation 3: $h$ (the helical rise per base pair), $\Omega_n$ (helical twist angle per base pair), $r_d$ (the length of the vector extending from the $z$-axis to the center of the fluorescence donor dye), $r_a$ (the length of the corresponding vector to the acceptor dye), $\Delta h$ (the vertical distance between the two dyes if they were attached to the same base pair), and $\Delta \Omega$ (the angle between vectors $r_a$ and $r_d$ if the two dyes were attached to the same base pair). Although the planes containing the vectors $r_a$ and $r_d$ are parallel, they are not necessarily coincident with the base pair plane. The model does not require the two dye-labeled DNA strands to be paired (Watson-Crick or otherwise) as long as the Rox-labeled strand is characterized by a regular helical geometry.
where the values of $r_a$ and $r_d$ are the lengths of the vectors pointing from the helix axis to the centers of dyes projected on the base pair plane; $\Delta h$ is the vertical distance separating the two dye centers if they were attached to the same base pair; and $\Delta \Omega_a$ is the angle between $r_a$ and $r_d$ if the two dyes were attached to the same base pair. For B-form dsDNA, the helical axis is through the center of each base pair plane. Thus, the two vectors $r_a$ and $r_d$ are pointing directly from the base pair center to the center of the dyes. For other DNA helices, including RecA-DNA complexes, the helical axis may be displaced away from the base pair center. Hence, the vectors $r_a$ and $r_d$ include the contribution of the distance from the helix axis to the base pair center. Importantly, this general DNA helix geometry model does not necessarily require two DNA strands to be paired as long as the Rox-labeled strand is ordered relative to the position of fluorescein on the other strand. The base pair separation $n$ between donor and acceptor is the only extrinsic variable in equation (2), and can be systematically varied throughout a series of DNA molecules. The energy transfer efficiency $E$ between donor and acceptor then can be expressed as a function of $n$ by combining equations (1) and (2):

$$E = \frac{R_0}{R_0 + \left[ r_a^2 + r_d^2 + (n h + \Delta h)^2 - 2 r_a r_d \cos(n \Omega_a + \Delta \Omega_a) \right]^2}$$

(3)

Experimental design

We designed five different but directly related reaction systems to facilitate the FRET measurements on different DNA and RecA-DNA complexes (Figure 2.3). Of the
**Figure 2.3** Fluorescent DNA reaction systems. Each DNA strand used in each system only contains one fluorescein or one Rox. The apparent multiple labeling of Rox on the 45mer serves the illustrative purpose of indicating the relative separations between the fluorophores. The actual set of DNA molecules used is shown in Figure 2.1. In the dsDNA system, a dual-labeled partial dsDNA FR, formed between the fluorescein-labeled 74mer and its complementary Rox-labeled 45mers 45R, was used. In the RecA·dsDNA system, the same dsDNA FR, was used to form the complex with the RecA protein (gray ovals). In the RecA·tsDNA-C and RecA·tsDNA-O systems, a presynaptic filament was first formed between the RecA protein and a 74mer (74X or 74C). The partial dsDNA FR, was used as the homologous dsDNA substrate. In the RecA·tsDNA-I system, the presynaptic filament was formed on a series of Rox-labeled 45mers, and a full-length fluorescein-labeled dsDNA Fds74 was used as the homologous substrate. The concentration of RecA protein was sufficient to coat the full length of the 74-bp dsDNA. I: incoming strand; C: complementary strand; O: outgoing strand.
Figure 2.3
five systems, the dsDNA series was designed to provide information about the dye-linker structures, the RecA-dsDNA series was designed to test FRET-derived helical geometry parameters against those reported in the literature, and the RecA·tsDNA−I, −C, and −O series were designed to extract the helical geometry of the incoming, complementary, and outgoing strands, respectively. Each system contains a fluorescein-labeled 74-nt oligodeoxynucleotide (ODN) and a series of six Rox-labeled 45mers (45R_i; see Figure 2.1). All ODNs in the 45R_i series have the same sequence; only the thymidine position to which Rox was attached was varied, and each unique attachment position is indicated by the subscript on 45R_i (Figure 2.1).

As an initial control, a set of partial dsDNA molecules (FR_i) was labeled with both fluorescein and Rox (Figure 2.1). Each partial dsDNA FR_i has 7- and 22-nt overhangs at the 5' and 3' ends, respectively. The fluorescein-labeled dT in each FR_i is the nucleoside in the 22-nt overhang adjacent to the 3' end of the base-paired region. In aqueous solution, this set of dsDNAs was assumed to adopt a standard B-form double-helical structure. This RecA-free dsDNA system was designed to isolate the dye-related parameters for usage in the other four RecA-involved systems.

In the second system, the same partial dsDNA FR_i was used to facilitate the polymerization of the RecA protein on the middle dsDNA region. The DNA helical parameters of the RecA·dsDNA complex (Figure 2.3) were measured using this system.
The other three systems (RecA·tsDNA-C, RecA·tsDNA-O, and RecA·tsDNA-I; Figure 2.3) allowed measurements of the helical parameters for the complementary strand (C), outgoing strand (O), or incoming strand (I) of the RecA·tsDNA complex. The partial dsDNA FR, described above was used as the homologous dsDNA substrate in reactions where the Rox-labeled strand participates as the complementary or outgoing strand (RecA·tsDNA-C and -O, respectively). For RecA·tsDNA-C, the sequence of the fluorescein-labeled strand in the partial dsDNA FR, is identical to a RecA-coated 74mer (74X). When the RecA·74X filament pairs on FR, the fluorescein-labeled strand is the outgoing strand while the Rox-labeled strand (45R,) is the complementary strand. This arrangement allows the structural information of the complementary strand to be ascertained. For RecA·tsDNA-O, the incoming ssDNA was 74C, the sequence of which is complementary to the fluorescein-labeled strand of the dsDNA FR, . Therefore, the Rox-labeled strand (45R,) becomes the outgoing strand. For RecA·tsDNA-I, a full-length fluorescein single-labeled dsDNA (ds74F) was utilized. A Rox-labeled 45mer (45R,) was used as the incoming ssDNA to pair with ds74F. The sequence of this incoming 45R, is identical to a 45-base sequence in the middle of the non-fluorescein-labeled ssDNA (74X) of ds74F. Thus, the fluorescein-labeled strand becomes the complementary strand when the RecA-mediated strand exchange takes place. This systematic labeling of Rox on each strand in the three RecA·tsDNA systems enabled us
to collect structural information directly from the complementary strand, outgoing strand, and incoming strand of the strand exchange intermediate.

RESULTS

Characteristics of DNA strand exchange with fluorophore-labeled ODNs

We first examined the possibility that the conjugation of fluorophores to the ODNs might impede the binding of RecA or strand exchange. Different reactions using dye-labeled DNA substrates as well as unlabeled DNA substrates were performed under identical conditions, and equal amounts of each reaction, either deproteinized or untreated, were fractionated on 3% Metaphor agarose gels (Figure 2.4). The untreated reactants showed equal amounts of involvement of the labeled or unlabeled dsDNA substrates into the reaction complexes. No release of the outgoing strand was observed. After treatment to remove RecA protein, both labeled and unlabeled reactions showed similar extents of product formation. In addition, when a heterologous 45mer, either Rox-labeled or unlabeled (h45 and h45R, respectively), was used as an incoming ssDNA to pair with a full-length dsDNA (ds74 or ds74F), it was found that the strand exchange reaction did not occur. These results demonstrated that the labeling of fluorophores did not affect the occurrence and completion of the reactions under our reaction conditions, and that the strand exchange reaction we observed is homology-specific. Our conclusion
Figure 2.4  Agarose gel mobility shift assay for RecA-mediated strand exchange. Lane 1: ssDNA 45R₀; Lane 2: RecA·45R₀; Lane 3: ds74F; Lane 4 and 6: ds74F + RecA·45R₀; Lane 5 and 7: ds74F + RecA·45R₀, deproteinized; Lane 8: ds74F + h45R; Lane 9: ds74F + h45R, deproteinized; Lane 10: FR₀ + RecA·74C; Lane 11: FR₀ + RecA·74C, deproteinized. Lane 12: FR₀; Lane 13: RecA·74C; Lane 14: 74C. In lane 6 and 7, the RecA protein used to form the presynapsis filament is at stoichiometric amount of a 74mer, while in lane 4 and 5, the RecA protein is at stoichiometric amount of the 45mer. The final concentrations of the reactants (dsDNA and RecA·ssDNA filament) are at 0.1 µM (molecule).
that strand exchange is not affected by labeling the DNA strands with fluorescent dyes is consistent with published reports (200,201).

**Spectral characteristics of the reaction systems**

In order to relate the energy transfer efficiency quantitatively with the helical parameters that give rise to the distance between a FRET donor and acceptor, both the energy transfer efficiency and the Förster critical distance $R_0$ of the pair has to be known (see equation (3)). The energy transfer efficiency varies with the placement of the dye, while $R_0$ is characteristic of the position-independent spectral properties of the dyes. This, in turn, requires a knowledge of the relative orientation of the two dyes ($\kappa^2$), the donor quantum yield in the absence of the acceptor ($\Phi_d$), and the overlap integral $J$ that characterizes the resonance between the donor and acceptor transition dipoles.

The orientation factor $\kappa^2$ can not be directly measured. However, rapid motion of the dyes during their excited state lifetimes, combined with random static orientation of the dye transition moments, simplifies the interpretation of energy transfer measurements by allowing the assumption of a constant orientation factor ($\kappa^2 = 2/3$). The quality of the assumption can be evaluated by independent fluorescence anisotropy measurements. When a dual-labeled dsDNA sample is excited at 492 nm, only the fluorescein fluorescence is detected at 520 nm. When it is excited at 585 nm, only Rox fluorescence is detected at 605 nm. Therefore, the fluorescence anisotropy of each fluorescence probe
can be measured independently, even in the dual-labeled duplexes. Within error, the
anisotropies of both fluorescein and Rox were independent of the dye’s position of
attachment throughout the length of the DNA molecule in each system, suggesting a
constant $\kappa^2$ value. The average anisotropies of fluorescein and Rox are shown in Table
2.1. In general, fluorescein has a much lower anisotropy than Rox in each system.
Because the fluorescence lifetimes of both dyes are similar, and the dyes are attached to
the same macromolecule, fluorescein must undergo considerably faster rotational
diffusion, or rotate rapidly within a greater cone angle, than Rox (the fundamental
anisotropy of both dyes is almost 0.40). Because of the lower anisotropy of fluorescein,
and the lack of any indication that either fluorophore is located in a preferred static
orientation relative to the dsDNA molecule, we assume a random relative orientation of
the dyes with rapid reorientational motion and a constant value $\kappa^2 = 2/3$. Such an
assumption has been validated in a variety of experimental systems (207,213-224).

We measured the quantum yield of fluorescein in the absence of Rox ($\Phi_F$) and the
spectral overlap integral for each system. An example of the spectral overlap between
fluorescein and Rox measured on two dye-conjugated ssDNA samples is shown in Figure
2.5a. The spectral overlap between fluorescein and Rox is smaller than those between
fluorescein and tetramethylrhodamine (201) or hexachlorofluorescein (200), which, in
turn, results in a smaller $R_0$ value. Importantly, however, a smaller $R_0$ value reduces the
Table 2.1. Spectral characteristics of fluorescein (F) and Rox (R) measured in dsDNA, and in the RecA-dsDNA and RecA-tssDNA complexes

<table>
<thead>
<tr>
<th></th>
<th>dsDNA</th>
<th>RecA-dsDNA</th>
<th>RecA-tssDNA-C</th>
<th>RecA-tssDNA-I</th>
<th>RecA-tssDNA-O</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Phi_F$</td>
<td>0.76</td>
<td>0.83</td>
<td>0.77</td>
<td>(0.77) $^a$</td>
<td>(0.77) $^a$</td>
</tr>
<tr>
<td>$J \cdot 10^{-15}$ (M$^{-1}$cm$^{-1}$nm$^3$)</td>
<td>2.0</td>
<td>1.8</td>
<td>2.0</td>
<td>(2.0)</td>
<td>(2.0)</td>
</tr>
<tr>
<td>$R_0$ (Å)</td>
<td>53.6</td>
<td>53.2</td>
<td>53.5</td>
<td>(53.5)</td>
<td>(53.5)</td>
</tr>
<tr>
<td>$\varepsilon_F^{492}$ (M$^{-1}$cm$^{-1}$)</td>
<td>41000</td>
<td>57000</td>
<td>60000</td>
<td>(60000)</td>
<td>(60000)</td>
</tr>
<tr>
<td>$\varepsilon_R^{585}$ (M$^{-1}$cm$^{-1}$)</td>
<td>78000</td>
<td>72000</td>
<td>65000</td>
<td>(65000)</td>
<td>(65000)</td>
</tr>
<tr>
<td>$\varepsilon_R^{492} / \varepsilon_R^{585}$</td>
<td>0.053</td>
<td>0.059</td>
<td>0.056</td>
<td>0.053</td>
<td>0.064</td>
</tr>
<tr>
<td>$r_F^{(492, 520)}$</td>
<td>0.10</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22 $^b$</td>
<td>0.22</td>
</tr>
<tr>
<td>$r_R^{(585, 605)}$</td>
<td>0.22</td>
<td>0.28</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
</tbody>
</table>

The tabulated parameters derived from spectroscopy are the fluorescein quantum yield ($\Phi_F$), the spectral overlap integral ($J$), the Förster critical distance ($R_0$), the molar absorption coefficient ($\varepsilon$) of each fluorophore (subscript) at a specific wavelength in nm (superscript), the ratio of the molar absorption coefficients for Rox at 492 and 585 nm ($\varepsilon_R^{492} / \varepsilon_R^{585}$), and the anisotropy ($r$) of each fluorophore (subscript) acquired at specified excitation and emission wavelengths (superscript). The coefficients were measured by comparing their absorbance or excitation spectra under reaction conditions with those of the fluorescein or Rox-labeled ssDNA. Relative uncertainties in the parameters tabulated in the first five rows were generally less than 5% of the tabulated values as judged using standard deviations for replicate experiments. The relative uncertainties in the $\varepsilon_R^{492} / \varepsilon_R^{585}$ values were less than 10% in every case. The anisotropy values were essentially independent of the relative positions of fluorescein and Rox, and the absolute uncertainties in the anisotropy values (as judged by one standard deviation derived from all six complexes for each system) were ± 0.015. $^a$ The spectral parameters shown in parentheses for the RecA-tssDNA–I and –O complexes were assumed to be unchanged from those measured for the RecA-tssDNA–C complex. $^b$ The fluorescein anisotropy values measured for the RecA-tssDNA–I complex with Rox at positions $i = 0$ and $i = 4$ were significantly lower than other values and were excluded from the tabulated average. The cause and effect of the abnormal anisotropy values are discussed in the text.
Figure 2.5  Spectral data characterizing the fluorescein-Rox reaction systems. (a) The spectral overlap between fluorescein emission spectrum (λ_ex=492 nm, dashed line) and the absorption spectra of Rox (solid line) in 25 mM Tris-OAc, pH 7.5, 5% glycerol, 1 mM EDTA. The spectra were recorded for fluorescein-labeled dsDNA FX and Rox-labeled 45mer 45R₀ (see Figure 2.1). (b) The emission spectra used to calculate the energy transfer efficiency for partial duplex FR₁₅. *Gray line:* emission spectrum of FR₁₅ with excitation at 492 nm. An apparent emission enhancement around 605 nm was observed, indicating the occurrence of energy transfer between fluorescein and Rox. *Dashed line:* normalized emission spectrum of the fluorescein-only labeled dsDNA FX with excitation at 492 nm. There is no shift of the emission peak near 520 nm for the dual-labeled dsDNA FR₁₅ emission spectrum compared to that of fluorescein-only labeled FX, indicating an absence of strong electronic coupling between fluorescein and Rox. *Thick black line:* sensitized emission spectrum calculated by subtraction of the fluorescein-only emission (dashed line) from the emission spectrum of the dual-labeled dsDNA sample FR₁₅ (gray line). *Thin black line:* emission spectrum of dsDNA FR₁₅ with excitation at 585nm. The (Ratio)ₐ is determined from intensities of the sensitized emission spectrum (thick black line) and the Rox emission spectrum (thin dark line), and the corresponding energy transfer efficiency is calculated using equation 19.
sensitivity of our FRET measurements to the cross-interaction between complexes in solution, and generates well-resolved emission spectra, in which spectral distortion is minimal. We verified that the normalized emission spectra of fluorescein or Rox in the presence of the other fluorophore are identical to those in its absence (one exception will be discussed below). This suggests that the energy distributions of the donor or acceptor emission are not affected by the presence of complementary acceptor or donor, respectively. These characteristics of the fluorophore pair employed here provide sensitive information on relatively short-range, local conformational changes. The calculated $R_0$ values for the reaction systems are listed in Table 2.1.

We also measured the extinction coefficients of fluorescein and Rox at their respective excitation wavelengths, as well as the anisotropy of each dye at different base pair separations in each system. The results are listed in Table 2.1. The extinction coefficients of both dyes are constant, independent of their positions of conjugation to the DNA strands, indicating similar chromophore environments. These extinction coefficients are used in the calculation of the energy transfer efficiency. Representative fluorescein and Rox emission spectra necessary to perform such calculations are illustrated in Figure 2.5b for dsDNA. The sensitized emission spectrum of Rox (excitation at 492 nm) – corrected by removal of fluorescein background emission – and the emission spectrum of Rox (direct excitation at 585 nm) were used to obtain $(Ratio)_A$.
according to equations (17) and (18). This ratio is related to the energy transfer efficiency by the extinction coefficients according to equation (19).

**Energy transfer efficiency measurements using dsDNA**

We first measured the energy transfer efficiency on 0.1 \( \mu \text{M} \) RecA-free dsDNA with fluorescein and Rox conjugated to the opposite strands. The values of \( E \) and \( R \) at each \( n \) are listed in Table 2.2. A plot of the energy transfer efficiencies between fluorescein and Rox versus the number of base pairs separating the two dyes is displayed in Figure 2.6a (open squares). The data points were fitted to the \( E-n \) function (equation (3)), yielding a unique solution regardless of initial parameter values. With the assumption of B-form dsDNA helical parameters (\( h = 3.4 \) Å and \( \Omega_h = 36^\circ \)) (225), the best-fitted parameters were found to be: \( r_s^2 + r_d^2 = 960 \) Å\(^2\), \( 2r_s r_d = 140 \) Å\(^2\), \( \Delta h = -11 \) Å, and \( \Delta \Omega_h = 60^\circ \) (Table 2.3). The pair of parameters, \( r_s \) and \( r_d \), describing the radial distances between each fluorophore and the helix axis, can be calculated by the simultaneous solution of the pair of equations for the two combined fitting parameters, \( r_s^2 + r_d^2 \) and \( 2r_s r_d \). For the protein-free dsDNA, the radial distances are \( 31 \pm 3 \) Å and \( 2 \pm 2 \) Å (Table 2.4). Because the distances are symmetric with respect to exchange in the two combined fitting parameters, the helix geometry model does not allow the unique assignment of individual dye-linker distances. These dye-linker parameters are geometrically reasonable but are clearly different from the parameters expected for fully extended
Table 2.2. The energy transfer efficiency ($E$) and the distance ($R$) between the fluorescein and Rox at different base pair separations in five reaction systems

<table>
<thead>
<tr>
<th>Rox Position</th>
<th>dsDNA</th>
<th>RecA·dsDNA</th>
<th>RecA·tsDNA-C</th>
<th>RecA·tsDNA-O</th>
<th>RecA·tsDNA-I</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E$</td>
<td>$R (\text{Å})$</td>
<td>$E$</td>
<td>$R (\text{Å})$</td>
<td>$E$</td>
</tr>
<tr>
<td>0</td>
<td>0.80 ± 0.02</td>
<td>42 ± 6</td>
<td>0.37 ± 0.03</td>
<td>58 ± 7</td>
<td>0.17 ± 0.01</td>
</tr>
<tr>
<td>4</td>
<td>0.94 ± 0.04</td>
<td>34 ± 26</td>
<td>0.74 ± 0.01</td>
<td>45 ± 1</td>
<td>0.45 ± 0.02</td>
</tr>
<tr>
<td>7</td>
<td>0.96 ± 0.03</td>
<td>32 ± 20</td>
<td>0.54 ± 0.01</td>
<td>52 ± 2</td>
<td>0.33 ± 0.02</td>
</tr>
<tr>
<td>12</td>
<td>0.76 ± 0.05</td>
<td>44 ± 12</td>
<td>0.20 ± 0.01</td>
<td>67 ± 3</td>
<td>0.13 ± 0.00</td>
</tr>
<tr>
<td>15</td>
<td>0.56 ± 0.01</td>
<td>52 ± 2</td>
<td>0.10 ± 0.00</td>
<td>76 ± 1</td>
<td>0.07 ± 0.01</td>
</tr>
<tr>
<td>20</td>
<td>0.24 ± 0.03</td>
<td>65 ± 10</td>
<td>0.02 ± 0.01</td>
<td>99 ± 25</td>
<td>0.03 ± 0.01</td>
</tr>
</tbody>
</table>

Energy transfer efficiency ($E$) was calculated using equation (19) and the distance ($R$) was calculated using equation (1). Each set of data is represented as the mean of at least three independent experiments ± standard deviation.
Figure 2.6 Energy transfer efficiency ($E$) between fluorescein and Rox at different base-pair separations ($n$) in the B-form dsDNA and RecA·dsDNA (a), the RecA·tsDNA–C and –O ((b) and (c)), and the RecA·tsDNA–I (d) reaction systems. In each case, the symbols represent the actual data points and the solid lines are the best-fit curves according to nonlinear least squares analysis using equation 3. (a) The critical Förster distances $R_0$ were set at 53.6 Å for dsDNA (open squares, gray line) and 53.2 Å for RecA·dsDNA (filled circles, black line). For B-form dsDNA, the helical pitch per base pair $h$ at 3.4 Å and helical twist angle per base pair $\Omega_h$ at 36° were used to obtain the fitted dye-linking parameters: $r_a^2 + r_d^2 = 960 \text{ Å}^2$, $2r_a \cdot r_d = 140 \text{ Å}^2$, $\Delta h = -10.6 \text{ Å}$, and $\Delta \Omega = 60°$. For RecA·dsDNA, the vertical distance between the two dyes was taken from the B-form dsDNA as $\Delta h = -10.6 \text{ Å}$. The fitted parameters are: $r_a^2 + r_d^2 = 2000 \text{ Å}^2$, $2r_a \cdot r_d = 280 \text{ Å}^2$, $h = 4.9 \text{ Å}$, $\Omega_h = 17.3°$, and $\Delta \Omega = 12°$. (b) For the RecA·tsDNA–C (filled squares, red line) and RecA·tsDNA–O (filled circles, green line) complexes, the base registry was taken to be $\Delta n = 0$. The critical Förster distance $R_0$ was set at 53.5 Å, and the vertical distance between the two dyes was taken from the B-form dsDNA as $\Delta h = -10.6 \text{ Å}$. The fitted parameters for RecA·tsDNA–C are: $r_a^2 + r_d^2 = 3400 \text{ Å}^2$, $2r_a \cdot r_d = 370 \text{ Å}^2$, $h = 4.3 \text{ Å}$, $\Omega_h = 18.6°$, and $\Delta \Omega = 290°$; for RecA·tsDNA–O, the parameters are: $r_a^2 + r_d^2 = 4290 \text{ Å}^2$, $2r_a \cdot r_d = 480 \text{ Å}^2$, $h = 4.2 \text{ Å}$, $\Omega_h = 31.1°$, and $\Delta \Omega = 200°$. (c) For the RecA·tsDNA–C (red) and RecA·tsDNA–O (green) complexes, the base registry was taken to be $\Delta n = -3$. The fitted
parameters for RecA·tsDNA–C are: \( r_a^2 + r_d^2 = 3590 \, \text{Å}^2 \), \( 2r_a \cdot r_d = 830 \, \text{Å}^2 \), \( h = 4.9 \, \text{Å} \), \( \Omega_h = 13.3^\circ \), and \( \Delta \Omega = 30^\circ \); for RecA·tsDNA–O, the parameters are: \( r_s^2 + r_d^2 = 4400 \, \text{Å}^2 \), \( 2r_s \cdot r_d = 550 \, \text{Å}^2 \), \( h = 5.2 \, \text{Å} \), \( \Omega_h = 21.9^\circ \), and \( \Delta \Omega = 360^\circ \). (d) For the RecA·tsDNA–I (filled triangles, blue line), the best fit helical parameters are: \( r_a^2 + r_d^2 = 4800 \, \text{Å}^2 \), \( 2r_a \cdot r_d = 1900 \, \text{Å}^2 \), \( h = 1.9 \, \text{Å} \), \( \Omega_h = 11.0^\circ \), and \( \Delta \Omega = 340^\circ \).
Figure 2.6
<table>
<thead>
<tr>
<th>Δn</th>
<th>h (Å)</th>
<th>Ωₙ (°)</th>
<th>r₁² + r₂² (Å²)</th>
<th>2rₙrₙ (Å²)</th>
<th>ΔΩ (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dsDNA</td>
<td>4.9 ± 0.1</td>
<td>+17.3± 4.1</td>
<td>2000 ± 90</td>
<td>280 ± 80</td>
<td>12 ± 27</td>
</tr>
<tr>
<td>RecA-dsDNA</td>
<td>1.9 ± 3.7</td>
<td>+11.0± 3.3</td>
<td>4800±1600</td>
<td>1900±1700</td>
<td>340 ± 5</td>
</tr>
<tr>
<td>RecA-tsDNA-I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-5</td>
<td>6.0 ± 0.2</td>
<td>+16.5± 1.7</td>
<td>4200 ± 90</td>
<td>1000 ± 60</td>
<td>72 ± 6</td>
</tr>
<tr>
<td>-4</td>
<td>5.6 ± 0.1</td>
<td>+19.1± 1.7</td>
<td>4360 ± 60</td>
<td>720 ± 60</td>
<td>38 ± 6</td>
</tr>
<tr>
<td>-3</td>
<td>5.2 ± 0.1</td>
<td>+21.9± 1.7</td>
<td>4400 ± 50</td>
<td>550 ± 60</td>
<td>360 ± 7</td>
</tr>
<tr>
<td>-2</td>
<td>4.8 ± 0.1</td>
<td>+25.1± 1.5</td>
<td>4380 ± 50</td>
<td>470 ± 60</td>
<td>310 ± 9</td>
</tr>
<tr>
<td>-1</td>
<td>4.5 ± 0.1</td>
<td>+28.4± 1.4</td>
<td>4330 ± 50</td>
<td>450 ± 60</td>
<td>250 ± 13</td>
</tr>
<tr>
<td>RecA-tsDNA-O</td>
<td>4.2 ± 0.1</td>
<td>+31.1± 1.5</td>
<td>4290 ± 50</td>
<td>480 ± 60</td>
<td>200 ± 15</td>
</tr>
<tr>
<td>1</td>
<td>4.0 ± 0.1</td>
<td>+33.0± 1.4</td>
<td>4250 ± 50</td>
<td>530 ± 60</td>
<td>150 ± 15</td>
</tr>
<tr>
<td>2</td>
<td>3.7 ± 0.1</td>
<td>+34.4± 1.5</td>
<td>4200 ± 50</td>
<td>590 ± 60</td>
<td>110 ± 16</td>
</tr>
<tr>
<td>3</td>
<td>3.6 ± 0.1</td>
<td>+35.5± 1.5</td>
<td>4160 ± 50</td>
<td>640 ± 60</td>
<td>60 ± 17</td>
</tr>
<tr>
<td>4</td>
<td>3.4 ± 0.1</td>
<td>+36.8± 1.8</td>
<td>4100 ± 50</td>
<td>700 ± 80</td>
<td>15 ± 12</td>
</tr>
<tr>
<td>5</td>
<td>3.5 ± 0.1</td>
<td>+43.6± 1.0</td>
<td>4100 ± 60</td>
<td>1500 ± 300</td>
<td>260 ± 20</td>
</tr>
<tr>
<td>-5</td>
<td>5.3 ± 2.1</td>
<td>+10.7± 2.8</td>
<td>3430 ± 1460</td>
<td>1500 ± 320</td>
<td>80 ± 55</td>
</tr>
<tr>
<td>-4</td>
<td>5.1 ± 1.7</td>
<td>+11.9± 3.8</td>
<td>3560 ± 1000</td>
<td>1120 ± 530</td>
<td>60 ± 42</td>
</tr>
<tr>
<td>-3</td>
<td>4.9 ± 1.3</td>
<td>+13.3± 4.9</td>
<td>3590 ± 730</td>
<td>830 ± 580</td>
<td>30 ± 28</td>
</tr>
<tr>
<td>-2</td>
<td>4.7 ± 1.0</td>
<td>+15.0± 6.3</td>
<td>3550 ± 580</td>
<td>610 ± 540</td>
<td>5 ± 15</td>
</tr>
<tr>
<td>-1</td>
<td>4.6 ± 0.8</td>
<td>+17.0± 8.1</td>
<td>3500 ± 500</td>
<td>450 ± 470</td>
<td>330 ± 27</td>
</tr>
<tr>
<td>RecA-tsDNA-C</td>
<td>4.3 ± 0.7</td>
<td>+18.6± 10.8</td>
<td>3390 ± 510</td>
<td>370 ± 470</td>
<td>290 ± 64</td>
</tr>
<tr>
<td>1</td>
<td>4.6 ± 0.1</td>
<td>+41.9± 3.9</td>
<td>3100 ± 80</td>
<td>390 ± 270</td>
<td>90 ± 36</td>
</tr>
<tr>
<td>2</td>
<td>4.3 ± 0.1</td>
<td>+41.9± 3.9</td>
<td>3100 ± 80</td>
<td>670 ± 290</td>
<td>30 ± 16</td>
</tr>
<tr>
<td>3</td>
<td>4.1 ± 0.1</td>
<td>+44.5± 1.0</td>
<td>3000 ± 90</td>
<td>900 ± 300</td>
<td>340 ± 11</td>
</tr>
<tr>
<td>4</td>
<td>4.0 ± 0.1</td>
<td>+44.8± 0.7</td>
<td>3000 ± 90</td>
<td>1100 ± 300</td>
<td>300 ± 9</td>
</tr>
<tr>
<td>5</td>
<td>3.8 ± 0.1</td>
<td>+45.0± 0.6</td>
<td>2900 ± 90</td>
<td>1260 ± 300</td>
<td>250 ± 8</td>
</tr>
</tbody>
</table>

The parameters presented here are the best solutions regardless of initial values varied over a wide range. Each set of parameters is represented as the mean of at least three independent experiments ± standard deviation. For dsDNA, the helical rise per base pair d and helical twist angle per base pair Ωₙ are fixed at 3.4 Å and 36 °, respectively. The fitted parameter Δh was used in the other four RecA-involved reaction systems to in order to obtain the best fits of the other five parameters. For RecA-tsDNA-O and -C, the base pair separation is varied from −5 (5' end shift) to 5 (3' end shift), and the corresponding fitted parameters are shown.
Table 2.4 Summary of experimental helical geometry parameters and dye-linker structural parameters for B-DNA and the RecA-bound dsDNA and tsDNA complexes

<table>
<thead>
<tr>
<th></th>
<th>$h$ (Å)$^a$</th>
<th>$\Omega_b$ (°)$^a$</th>
<th>$r_d$ (Å)$^b$</th>
<th>$r_s$ (Å)$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual Analysis</strong>$^a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B-dsDNA</td>
<td>3.4</td>
<td>+36</td>
<td>31 ± 3</td>
<td>2 ± 2</td>
</tr>
<tr>
<td>RecA·dsDNA</td>
<td>4.9 ± 0.1</td>
<td>+17 ± 4</td>
<td>44 ± 1</td>
<td>3 ± 1</td>
</tr>
<tr>
<td>RecA·tsDNA–C$^c$</td>
<td>4.9 ± 1.3</td>
<td>+13 ± 5</td>
<td>59 ± 6</td>
<td>7 ± 5</td>
</tr>
<tr>
<td>RecA·tsDNA–I</td>
<td>1.9 ± 3.7</td>
<td>+11 ± 3</td>
<td>68 ± 13</td>
<td>14 ± 14</td>
</tr>
<tr>
<td>RecA·tsDNA–O$^c$</td>
<td>5.2 ± 0.1</td>
<td>+22 ± 2</td>
<td>66 ± 1</td>
<td>4 ± 1</td>
</tr>
<tr>
<td><strong>Global Analysis</strong>$^b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RecA·tsDNA–C$^c$</td>
<td>4.8 ± 0.4</td>
<td>+15 ± 4</td>
<td>60 ± 2</td>
<td>7 ± 2</td>
</tr>
<tr>
<td>RecA·tsDNA–I</td>
<td>4.8 ± 0.4</td>
<td>+15 ± 4</td>
<td>53 ± 2</td>
<td>5 ± 2</td>
</tr>
<tr>
<td>RecA·tsDNA–O$^c$</td>
<td>4.8 ± 0.4</td>
<td>+15 ± 4</td>
<td>68 ± 2</td>
<td>5 ± 2</td>
</tr>
</tbody>
</table>

$^a$ The helical rise ($h$) and helical twist ($\Omega_b$) per nucleotide were obtained as described in the text and the legend for Figure 5. The values for the protein-free dsDNA were taken as the canonical values (in parentheses), while the value for the RecA-DNA complexes are from Table 3. The uncertainties are represented by one standard deviation as described for Table 3. The $r$ parameters describe the radial distances of the donor (subscript “d”) and acceptor (subscript “a”) fluorophores from the helix axis as described in the legend for Figure 2. The values of $r$ were calculated from the two combined fitting parameters, ($r_s^2 + r_d^2$) and 2·$r_s$r_d (Table 3), by solution of two simultaneous equations as described in the text. The uncertainties in the $r$ values were calculated by propagation of the standard deviations in the combined fitting parameters (Table 3).

$^b$ The parameters were obtained by global data analysis in which $h$ and $\Omega_b$ were adjustable parameters fitted simultaneously to the data for all three strands while the respective dye-linker structure parameters (e.g., $r_s^2 + r_d^2$ and 2·$r_s$r_d) were separately fitted to each strand’s E–n data (Figure 6) as described in the text and the legend for Figure 6. The uncertainties for the two helical geometry parameters and the two combined dye-linker structure parameters were estimated from a reduced $\chi^2$ analysis like that described in Figure 7. The values of $r$ were calculated from the two combined fitting parameters, $r_s^2 + r_d^2$ and 2·$r_s$r_d, by solution of two simultaneous equations as described in the text. The uncertainties in the $r$ values were calculated by propagation of the estimated errors in the combined fitting parameters.

$^c$ The helical parameters listed for the complementary and outgoing strands assume a base registry shift of $\Delta n = -3$ for the outgoing strand (see the text).
linkers. For linkers in fully extended all-anti conformations coplanar with the thymine heterocycles, a simple calculation using canonical bond length and bond angle data to describe the distance from the helix axis to the dyes predicts dye-linker radial distances \(r_s = r_d = 25 \text{ Å}\) and fitting parameters \((r_n^2 + r_d^2 = 1250 \text{ Å}^2, 2r_n r_d = 1250 \text{ Å}^2, \Delta h = 0 \text{ Å and } \Delta \Omega_n = 180^\circ)\) different from those observed. One explanation for this difference is that one of the linkers is folded to allow the fluorophore to approach the DNA more closely. Conformations preferring dye-DNA interactions have been observed previously for both 5-tetramethylrhodamine (213,226,227) and cyanine-3 (228). The net result is that one of the dyes appears closer to the helical axis than predicted for a fully extended linker.

**Satisfaction of a saturation binding condition for RecA-DNA complexes**

In the four RecA-involved reaction systems, excess amounts of one non-fluorescent reaction species over the other fluorescent components were necessary to insure that greater than 90% of the fluorophore-labeled DNA molecules converted to the final product. This criterion is important for accurate FRET measurements. The FRET efficiency is highly sensitive to the distance between the fluorescence donor and acceptor: If one member of the FRET pair exists in both the bound and unbound states, the apparent energy transfer efficiency would represent contributions from each of the species. Hence, FRET efficiency measurements would be unable to characterize accurately the structure of the single complex of interest. For example, in the
RecA·dsDNA complex, when the RecA concentration was at its minimum stoichiometric concentration (1 monomer per 3bp), only about 70% of dsDNA molecules were bound. The energy transfer efficiencies measured under this condition were irregular and imprecisely determined, and reasonable curve fits could not be guaranteed. In contrast, when the RecA concentration was raised such that over 90% of the dsDNA molecules were bound, the data were highly reproducible. The fitting correlation coefficient using the $E-n$ function on these saturated dsDNA increased significantly. Based on these considerations, we measured the apparent dissociation constants of all our RecA·DNA complexes by treating them as simple one-step, reversible binding reactions.

The apparent dissociation constant ($K_d$) for each complex was measured by fitting fluorometric titrations, and the necessary concentration of each reaction component was determined as follows. In the RecA·dsDNA, RecA·tsDNA-C, and RecA·tsDNA-O reactions, the concentration of the partial dsDNA FR$_i$ was set at 0.1 μM while the concentration of the other reactant was adjusted to achieve the desired extent of product formation. For RecA·dsDNA, the RecA concentration was 4 μM. For RecA·tsDNA-C and RecA·tsDNA-O, the concentration of the RecA·ssDNA filament was 0.73 μM. In RecA·tsDNA-I, we used a different strategy. The reaction occurs between a RecA-bound Rox-labeled ssDNA (45R$_i$) and a fluorescein-labeled full-length dsDNA (ds74F), and both of the species emit fluorescence. In order to reduce the complexity inherent in
calculating the energy transfer efficiency between the FRET pair when one dye is in excess, the concentration of both species were maintained at 0.5 μM to achieve the 90% reaction completion. Taken together, the agarose gel mobility shift assays and the thermodynamic analysis using fluorescence titrations confirm that greater than 90% of the observed FRET signals originated from the species of interest.

**DNA helical parameters for RecA-bound dsDNA**

The energy transfer efficiencies for the set of RecA-bound dsDNA molecules were measured in the second reaction system. Excess RecA protein (4 μM), saturated with ATPγS (0.5 mM), was incubated with 0.1 μM partial dsDNA FR, for 1 h at 37 °C, and the emission spectra of fluorescein and Rox were recorded. This amount of RecA·ATPγS was used to insure that greater than 90% of the dsDNA was bound by the RecA protein. The energy transfer efficiencies were calculated and plotted against base pair separation (Figure 2.6a, filled circles). The data were fitted with the E-n function as described in equation (3). The dye-linker parameter Δh obtained from B-form dsDNA above was fixed, leaving the other helical parameters variable. The fitting yielded a unique solution regardless of initial parameter values. The fitted helical parameters are: Δh = 4.9 ± 0.1 Å, Ω = 17 ± 4° (Table 2.3). The original dsDNA FR, is presumed to adopt B-form conformation in solution with a helical rise of 3.4 Å and a helical twist angle of 36°. As such, the fitted helical parameters for the RecA-bound dsDNA reveal an
apparent conformational change. The dsDNA molecule was stretched by more than 40% of its original B-form length and was unwound by 19°. Within experimental uncertainty, the helical rise and twist angle are in quantitative agreement with the values of \( h = 5.2 \pm 0.18 \, \text{Å} \) (82) and \( \Omega_\theta = 19.4 \pm 1.4° \) (92) measured previously. This successful application of quantitative interfluorophore distance measurements demonstrates that the FRET method is valid and provides an accurate complement to EM methods.

**DNA helical parameters for RecA-tsDNA complexes**

The energy transfer efficiencies between fluorescein and Rox in the RecA-mediated strand exchange products were measured for the three RecA-tsDNA complexes. For RecA-tsDNA-C and -O, the same dual-labeled partial dsDNA FR\(_i\), described above, was used as the homologous dsDNA substrate. During strand exchange, one strand of the homologous dsDNA FR\(_i\) serves as the complementary strand while the other one serves as the outgoing strand depending on the sequence of the incoming RecA-ssDNA filament. In these two cases, the fluorophores are located on the outgoing and complementary strands. In order to analyze the data, the base-pair registry (\( \Delta n \)) between the two strands was required to determine the effective separation of the fluorophores (\( n \)). For the dsDNA and RecA-dsDNA structures, we assumed that the strands were paired in the complexes and that \( \Delta n = 0 \). We initially assumed that bases on the two strands of FR\(_i\) remained in the same base pair plane in the RecA-tsDNA complex. Thus, the base-pair
registry between fluorescein and Rox remained the same as depicted in Figure 2.6a ($\Delta n = 0$). Based on this assumption, the measured energy transfer efficiency between the two dyes for both complexes was plotted against their original base pair separations, and the $E$-$n$ function (equation (3)) was used to fit the experimental data (Figure 2.6b, red squares for RecA·tsDNA-C and green circles for RecA·tsDNA-O). The best-fit helical parameters are: $h = 4.3$ Å, $\Omega_n = 19^\circ$ for the complementary strand and $h = 4.2$ Å, and $\Omega_n = 31^\circ$ for the outgoing strand (Table 2.3, $\Delta n = 0$). Despite the similarity of the helical rise parameters for the two strands, the helical twist angle of the outgoing strand is significantly different from that of the complementary strand in the RecA·tsDNA complex. This observation would lead to the conclusion that the originally base-paired complementary and outgoing strands are not interwound in this strand exchange intermediate. However, the helical rise parameters for both characterized strands of the RecA·tsDNA complex were substantially different from that of the RecA·dsDNA complex. This observation led us to question the validity of the assumption regarding the base-pair registry ($\Delta n = 0$).

A recent paper reported that the outgoing strand inside a post-strand-exchange RecA·tsDNA filament may be shifted towards its 5' end three to four bases with respect to the corresponding bases in the heteroduplex (184). Based on this prediction, we refitted our experimental data by systematically shifting the outgoing strand towards
either its 5' end ($\Delta n < 0$) or its 3' end ($\Delta n > 0$). The best-fit helical parameters for each shift are listed in Table 2.3. Based on the nonlinear least squares analysis, using reduced $\chi^2$ as the fitting criterion, a registry shift from $-1$ to $-5$ produced quality fits with indistinguishable fitting goodness. We were unable to establish a quantitative criterion to select an absolute registry shift from $-1$ to $-5$ bases. However, the helical rise and twist angle parameters for both the complementary and outgoing strands were indistinguishable from one another or from those of RecA-bound dsDNA (Table 2.3) when $\Delta n = -3$ (Figure 2.6c). This result is consistent with the published report (184). Moreover, the internal consistency of strands’ helical parameters would minimize topological stress in the RecA-tsDNA complex. One important implication of the 5' registry shift of the outgoing strand is the same as that mentioned above based on the assumption $\Delta n = 0$. Namely, the complementary and outgoing strands are not paired in this strand exchange intermediate.

In the RecA-tsDNA-I system (Figure 2.6d, blue triangles), a full-length fluorescein-labeled dsDNA (ds74F) was utilized. A set of Rox-labeled 45mers (45R$_n$) was used as the incoming ssDNA to pair with ds74F. The sequence of this incoming ssDNA is identical to a 45-base sequence in the middle of the non-labeled strand of ds74F. Consequently, Rox labels the incoming strand. The best-fit helical parameters to the $E$-$n$ data were found to be: $h = 1.9 \, \text{Å}$, and $\Omega_h = 11^\circ$ (Table 2.3). The apparent
helical rise per residue is much smaller than that of B-form dsDNA. Moreover, the residual uncertainty in the fit parameters is large ($\sigma(h) = \pm 3.7 \, \text{Å}$). Further analysis of the spectral data revealed that the fluorescein and Rox emission spectra for RecA:45R$_o$-ds74F and RecA:45R$_i$-ds74F show positions of maximum intensity that are blue-shifted about 1 nm compared to those spectra obtained for each fluorophore in isolation. The spectral maxima of the dyes at every other position in every other complex are consistent. Such spectral changes contraindicate the application of equation (1) to the data for complexes RecA:45R$_o$-ds74F and RecA:45R$_i$-ds74F (207,211). Förster theory is based on weak electronic coupling between donor and acceptor fluorophores (205). These phenomena indicate that there may be strong electronic coupling (229) between the fluorescein and the Rox at positions $i = 0$ and $i = 4$ for the RecA-tsDNA-I complex. Interestingly, the anisotropies of the fluorescein in these two complexes were significantly lower (0.16 ± 0.01) than those from the other RecA-tsDNA complexes (0.22 ± 0.01) in the same system. While the nature of the relationships between the putative strong electronic coupling, the spectral changes, and the reduced anisotropy values is beyond the scope of this report, the unique photophysical characteristics of the first two members of the RecA-tsDNA-I series undoubtedly result in the low and imprecise $h$ value.

Such strong electronic coupling between the two dyes may be caused by the particular junction structure in this system. In the RecA-tsDNA-I complex, the incoming ssDNA is shorter than the dsDNA substrate: there are 7 and 22 bp flanking homologous
sequences in the dsDNA substrate. After strand exchange, the outgoing strand cannot be completely displaced because of the flanking base-paired regions. The junction between the paired and the displaced segments of the outgoing strand may create an environment in which the fluorescein on the complementary strand and the Rox at position zero or four on the incoming strand are brought close together, and they interact strongly as a result. This idea was tested by a control experiment, in which a fluorescein-labeled partial dsDNA (FX) was paired with the RecA·45Ri filament. Because the outgoing strand now has the same length as the incoming strand, a D-loop junction should not be formed. Indeed, the spectral shifts and abnormally low anisotropies were no longer observed. This idea could be exploited if we were able to use the alternative partial dsDNA FX for the RecA·tsDNA-I complex. Unfortunately, because the apparent dissociation constant between this partial dsDNA FX and the RecA·45Ri filament is relatively high (0.03 μM), we were not able to observe 90% complex formation at reasonable concentrations of both reactants. In addition, we were unable to choose two alternative thymidine positions for Rox-labeling on the incoming 45mer. Based on these observations, we conclude that problems associated only with the data collected for RecA·tsDNA-I with Rox at \( i = 0 \) and \( i = 4 \) positions led to the small values and relatively large errors of the helical parameters obtained for this complex.
Although the helical rise $h$ was not determined with high precision for the
RecA·tsDNA–I series, both $h$ and $\Omega_h$ are similar to those parameters obtained for
RecA·tsDNA–C and –O with a registry shift $\Delta n = -3$. Moreover, each helical geometry
parameter for the three systems was indistinguishable (accounting for experimental
uncertainty) from the weighted mean value of each parameter (derived from all three
RecA·tsDNA data sets), $h = 5.2 \pm 0.1$ Å and $\Omega_h = 18 \pm 3^\circ$. Based on the conformational
similarity among the three strands, we used a global data analysis in which the helical rise
and twist angle were adjustable parameters fitted simultaneously to the data for all three
strands (the data points corresponding to the first two $i$ values for the RecA·tsDNA–I
complex having been omitted), while the respective dye-linker parameters were
separately fitted to each strand’s $E=n$ data (Figure 2.7). This algorithm provided the best
helical rise and twist angle for the simultaneous description of all three strands. We
found that the best fit parameters, $h = 4.8 \pm 0.4$ Å and $\Omega_h = 15 \pm 4^\circ$, for a three-base
registry shift (Table 2.4) were in good quantitative agreement the weighted mean
parameters. These data are consistent with the conclusion that, within uncertainty limits,
all three strands inside the RecA filament share an identical helical geometry.

**Probing the potential effect of ATPγS hydrolysis**

Prior to undertaking a more careful structural interpretation of these results, we
wished to account for the uncertainty in the helical parameters. We envisaged two main
Figure 2.7  Global nonlinear least squares analysis of $E-n$ data for three RecA-tsDNA reaction systems using the same helical rise and twist angle for all three strands, while leaving the dye-labeling parameters variable within each series. The data symbols and best-fit curves are color-coded as described in Figure 2.5. The first two data points of the RecA-tsDNA–I series were excluded in this global fitting for reasons discussed in the text. A unique and common helix structure ($h = 4.8 \pm 0.4 \, \text{Å}, \, \Omega_h = 15 \pm 4^\circ$) were found to fit all three strands simultaneously.
sources of uncertainty: those arising from experimental conditions and those arising from the analytical methodology. For the former set, possible structural perturbations from ADP, the nucleotide product of ATPγS hydrolysis, were investigated. Because the RecA protein filament is known to have different structures and DNA-binding properties in the presence of ATPγS and ADP (230), the presence of a substoichiometric amount of ADP could impact the observed FRET efficiencies. The same set of reactions described above was performed under the same conditions using the non-hydrolyzable ATP analog ADP·AlF₄ in place of ATPγS. The measured energy transfer efficiency using ADP·AlF₄ were systematically about 20% lower than those measured in the presence of ATPγS (Figure 2.8). However, the trends in the E-n data remained the same. Most significantly, curve fitting using equation (3) generates essentially identical DNA helical parameters with slightly different dye-linker parameters for each system. This demonstrated that the existence of ADP resulting from ATPγS hydrolysis has negligible impact on the structure of the DNA helix in our reaction time scale.

**Evaluation of potential analytical sources of uncertainty in the helical parameters**

As described above, a second potential source of uncertainty is the analytical methodology employed. Within this context, we identified three particular elements warranting further investigation. First, calculation of the Förster critical distance $R_0$ required the assumption that the orientation factor is constant, $\kappa^2 = 2/3$. Second, curve-
Figure 2.8  Energy transfer efficiency ($E$) between fluorescein and Rox at different base-pair separations ($n$) in the RecA-dsDNA (a) and the RecA-tsDNA–O (b) reaction systems using ADP–AlF$_4$ instead of ATP$\gamma$S.  (a) The fitted parameters are: $r_a^2 + r_d^2 = 2285.5$ Å$^2$, $2r_ar_d = 314.1$ Å$^2$, $h = 5.0$ Å, $\Omega_h = 15.0^\circ$, and $\Delta \Omega = 55^\circ$.  (b) The fitted parameters (without the three base pair shift, comparing with Figure 2.5b) are: $r_a^2 + r_d^2 = 49778$ Å$^2$, $2r_ar_d = 612$ Å$^2$, $h = 4.3$ Å, $\Omega_h = 33^\circ$, and $\Delta \Omega = 182^\circ$. 
fitting produced the most robust fits upon the assumption of a constant dye-linker parameter, \( \Delta h = -10.6 \text{ Å} \) (determined from the protein-free B-DNA data). Finally, nonlinear least square analysis of the \( E-n \) function using multiple adjustable parameters could lead to unreliable parameter values. Each of these issues was evaluated in turn to evaluate the accuracy of the helical parameters.

Although the assumption of a constant \( \kappa^2 \) value corresponding to randomly oriented fluorophore has been validated in a number of FRET systems, we estimated the maximum and minimum values of \( \kappa^2 \) in our system to explore how much uncertainty was introduced by assuming a \( \kappa^2 \) value of 2/3. The \( \kappa^2 \) factor cannot be measured experimentally, but the upper (\( \kappa^2_{\text{max}} \)) and lower (\( \kappa^2_{\text{min}} \)) limits of \( \kappa^2 \) can be obtained from the measured limiting anisotropy of the donor and acceptor and the calculated axial depolarization factors using a published procedure (231). The \( \kappa^2_{\text{max}} \) and \( \kappa^2_{\text{min}} \) for the fluorescein and Rox in the RecA-bound dsDNA were calculated at 3.5 and 0.064 as described in Materials and Methods, which provided a wide range for \( R_0 \): \( 0.68R_0 < R_0 < 1.32R_0 \).

The calculation of \( \kappa^2_{\text{max}} \) and \( \kappa^2_{\text{min}} \) using equation (10) provides a worst-case estimation, which usually overestimates the effects of \( \kappa^2 \) on the calculated distance (209). For fluorophores with mixed extents of polarization, \( r_0 < 0.3 \), the error in distance is thought to be below 10% (209). Although the range of possible Förster distances is wide,
the uncertainty in the determination of the average distance between fluorescein and Rox will be smaller than suggested by the analysis of the orientation factor $\kappa^2$ value. Indeed, employing either $R_{0,\text{min}}$ or $R_{0,\text{max}}$ in place of $R_0$ in our analysis led to dramatically increased error ranges of the fitted parameters and significantly decreased goodness of fitting in all five systems. Combined with the independent anisotropy analysis presented above, the loss of fitting robustness indicates that the true $\kappa^2$ is substantially different from $\kappa_{\text{min}}^2$ and $\kappa_{\text{max}}^2$, and validates our use of $\kappa^2 = 2/3$.

In all five systems, the same $E$-$n$ function was used to fit the relationship between the measured energy transfer efficiency and the DNA structure. The dsDNA system was first fitted by assuming a helical rise of 3.4 Å and a twist angle at 36°. The dye-labeling parameter, $\Delta h = -10.6$ Å, which is the difference between the vertical positions of the two dyes if they were attached to the same base pair, was kept constant in the fitting of the other four systems. Although this approach is based on the reasonable assumption that the dye-linker structures remain essentially unaltered across all five systems, its impact on the fitting results was explored by allowing $\Delta h$ to vary. The same curve fitting procedure, using the $E$-$n$ function, was performed on each RecA-DNA complex while the value of $\Delta h$ was systematically varied from $-5$ to $-15$ Å. The range was chosen based on the nucleotide positions labeled by the fluorophores and the possible geometric arrangement of the dye linkers. It was found that the fitted parameters in each system
only vary about 10 to 15% when $\Delta h$ is varied from $-5$ Å to $-15$ Å. This range has already been accounted for by the standard deviation weighting procedure of the $E-n$ function fitting in each system. Therefore, we drew the conclusion that potential variability of the dye-linker parameter $\Delta h$ exerted a minimal impact on our fitting results.

The most informative results derived from the analysis of FRET efficiencies described above are the $h$ and $\Omega_h$ helical descriptors; however, these are only two of five adjustable parameters necessary to fit the helical geometry model to the $E-n$ data. A key issue in the interpretation of the helical parameters is their robustness as quantitative structural descriptors. Nonlinear least squares analysis of the data sets using the helical geometry model allowed this issue to be explored. As described in Materials and Methods, the values of all five parameters were simultaneously varied for each reaction system to minimize the reduced $\chi^2$. The parameter space around all five parameters for each of the three RecA-DNA complexes was searched using a manual grid search. The results of this analysis are displayed as two-dimensional cross-sectional plots of $\chi^2$ versus $h$ and $\Omega_h$ values (Figure 2.9a and b). These plots demonstrate the quadratic nature of the $\chi^2$ surface near the optimum for each data set's parameters and verify that the fitting routines were not trapped in a local minimum. Moreover, the steepness of the surfaces near the optima provides an estimate of how reliably determined each best-fit parameter is. For the helical rise per residue, the $h$ parameter values describing the RecA-dsDNA,
Figure 2.9 Reduced $\chi^2$ values plotted as a function of the fit parameters $h$ (a) and $\Omega_n$ (b) for the RecA-dsDNA (black), RecA-tsDNA–I (blue), RecA-tsDNA–C (red), and RecA-tsDNA–O (green) complexes. For the helical rise per residue, the $h$ values describing the RecA-dsDNA, RecA-tsDNA–C, and RecA-tsDNA–O complexes are unequivocally determined. In contrast, the $h$-value of the RecA-tsDNA–I complex is much less well determined, a fact that recapitulates the statistical imprecision in fitting this parameter (see Table 3). For the helical twist angle, the $\Omega_n$ values are approximately equally well defined for all four reaction systems.
RecA·tsDNA-C, and RecA·tsDNA-O complexes are unequivocally determined. In contrast, the \( h \)-value of the RecA·tsDNA-I complex is much less well determined, a fact that recapitulates the statistical imprecision in fitting this parameter (Table 2.3). For the helical twist angle, the \( \Omega \) parameter values are approximately equally well defined for all four reaction systems. The plots in Figure 2.9 allow us to assign conservative round upper estimates to the uncertainties in the best-fit \( h \) and \( \Omega \) values as \( \pm 0.4 \) Å and \( \pm 4^\circ \), respectively. The same analytical method applied to the parameters derived from the global analysis (\( \chi^2 \) plots not shown) led to the same uncertainty estimates.

Upon consideration of all three sources of experimental uncertainty described above, it is apparent that each source yields uncertainty levels that are adequately accounted for by the statistical measures of precision previously tabulated (Table 2.4). Hence, we used these latter values as reasonable estimates of the level of experimental uncertainty in each helical parameter.

**DISCUSSION**

We have presented a systematic implementation of FRET to characterize the structure of the RecA·tsDNA intermediate in RecA-mediated strand exchange. The FRET pair, fluorescein and Rox, was chosen to label a series of DNA molecules, and the energy transfer efficiency between them was measured in different DNA structures. The energy transfer efficiency is quantitatively related to the distance between the two dyes,
which is correspondingly related to the particular DNA structure. By solving the
relationship between the energy transfer efficiency and the DNA structure, we were able
to extract structural information, such as the helical rise and twist angle, for each
particular RecA-DNA complex. The data are consistent with the conclusion that all the
DNA strands in both RecA-dsDNA and RecA-tsDNA are similarly stretched and
unwound \( (h = 5 \text{ Å} \text{ and } \Omega_o = 20^\circ) \) relative to B-DNA. As discussed below, the data can be
further analyzed to resolve key structural features of the strand exchange intermediates.

**DNA structure inferred from interfluorophore distances**

The systematic application of FRET described above provides an overlapping set
of interfluorophore distances which are partly determined by the structures of the DNA
strands to which the dyes are conjugated. The interfluorophore distances are also
sensitive to the distance between the dye and the conjugated thymine heterocycle, and the
extraction of structural information beyond helical geometry parameters requires a
knowledge of the conformation of the dye linker arms. Unfortunately, the dyes are
conjugated to the DNA via flexible alkyl chains in order to provide sufficient rotational
freedom to justify the approximation \( \kappa^2 = 2/3 \) and alleviate the otherwise requisite
knowledge of the orientations of the transition dipoles of the two dyes (207).

In order to resolve this dilemma and allow the use of interfluorophore distances as
exact structural constraints, the geometry of dye-DNA conjugates have been resolved for
the dye pairs fluorescein-cyanine-3 (228) and fluorescein-5-tetramethylrhodamine (213,226,227). Whereas the conjugates employed here involve covalent attachment to thymine moieties at internal nucleotide positions, in both cited cases the dyes are coupled to the DNA via deoxyribose moieties at terminal positions. Nevertheless, a consideration of the previous works is instructive and relevant. In general, NMR structure determination (228), crystallographic structure analysis (226), computational modeling (226,232), and fluorescence anisotropy measurements (217,227) in combination with interfluorophore distance modeling (226,228) have led to the following conclusions: the linker to the negatively charged fluorescein dye adopts a conformation that is maximally extended, while the linker to the other dye (either cyanine-3 or 5-tetramethylrhodamine) adopts conformations that allow the positively charged polycyclic moiety to stack on the DNA helix end (for further discussion, see the review by 208).

While the helix geometry model does not allow the unique assignment of dye-linker conformations or lengths, we can extract the pair of radial distances, \( r_a \) and \( r_b \), from the linear combination of the combined fitting parameters, \( (r_a^2 + r_b^2) \) and \( 2r_ar_b \). The measurement of FRET efficiencies using the protein-free control B-DNA molecule revealed that the distance from the helix center to one dye is only about 2 Å, while the distance to the other fluorophore is about 31 Å (Table 2.4). This observation is consistent with published data (see above) suggesting that tetramethylrhodamine (analogous to Rox) is likely to coil back close to the DNA helix while fluorescein is maximally extended.
laterally from the DNA helix. Corroborative evidence is derived from the anisotropy measurements demonstrating that Rox adopts a more rigid conformation than fluorescein (Table 2.1). Because Rox resides in the major groove (Rox is attached to C5 of thymine) and the distance between Rox and the central helix axis of B-DNA is shorter than the radius of a typical B-DNA base pair (≈ 5 Å) (161), Rox most likely adopts a position near the floor of the major groove (226) within ≈ 2 Å of the helix axis. Hence, it is reasonable to assign tentatively the shorter \( r_s \) and longer \( r_e \) dye-linker distance to Rox and fluorescein, respectively.

Interestingly, this general case of one long dye-linker and one short dye-linker was observed in all the reaction systems (Table 2.4). However, for each of the RecA-DNA complexes, the value for one of the dye-linker distances was significantly longer than can be accounted for by an extended alkyl linker. This result may arise due to (i) a systematic problem with the fitting procedure, (ii) an altered DNA structure near one of the dyes, or (iii) a combination of both. If the anomalously long dye-linker distance is a result of a systematic problem, the meaning of the distances and their application would be suspect. It is important to note that the helical geometry parameters, not being correlated with the radial distance parameters (see above), are not invalidated by this possibility. On the other hand, if the apparent anomaly is the result of an altered DNA structure, then the dye-linker distance values would be useful. We argue that this latter case is the more likely scenario.
Four experimental observations argue against the possibility of a systematic error in the data collection and analysis. First, the spectral characteristics (Table 2.1) and FRET efficiencies (Table 2.2) for the RecA-DNA complexes are reasonable and highly reproducible. Second, the dye-linker distance values in the case of the protein-free DNA are physically reasonable and in agreement with previous determinations. Third, the helical geometry parameters, \( h \) and \( \Omega_n \), for the RecA·dsDNA complex, also obtained by curve fitting, are in good agreement with those obtained by other methods. Finally, similar values for the dye-linker parameters are obtained for the RecA·tsDNA complexes whether the data are analyzed individually or globally (Table 2.4).

One important alternative to a systematic experimental error is the possibility that the dye-linker radial distance accurately reflects an altered DNA geometry near the fluorophore. Three sets of facts suggest that this is a possibility. First, the experimental design places the thymidine to which fluorescein is conjugated one nucleotide position away from the end of the designated 45-nt region of homology (Figure 2.3). This design inadvertently places the site of fluorescein at a potentially flexible junction between structural domains. Second, there is evidence from restriction endonuclease reactivity (180), affinity modification (233) and computational modeling (184) that the junctions between a RecA·tsDNA structure and dsDNA (or RecA·dsDNA) may adopt an irregular structure. Finally, there is evidence that a fluorescein label can interact noncovalently with RecA protein in a nucleoprotein filament (234). In addition, our own observation of
anomalous spectrofluorometric behavior of fluorescein in two of the RecA·tsDNA–I complexes are consistent with an unusual junction structure (see above). Taken together, these data suggest that the fluorescein located at a point of DNA structural transition may be positioned further than expected from the helical axis at a site favored by protein-dye interactions.

Based on the evidence and reasoning presented above, we conclude that the radial distance parameters, $r_s$ and $r_{so}$ are valid and meaningful. Moreover, for each complex, we assign the longer distance to fluorescein and the shorter distance to Rox in analogy with the assignments for protein-free dsDNA (Table 2.4). While the assignment of the radial distances to specific fluorophores is not a prerequisite for a description of the helical geometry of the DNA nor the subsequent structural interpretation, as mentioned in the opening paragraph of this section, a knowledge of the location of the fluorophores relative to the DNA provides a necessary foundation for drawing conclusions about the relative positions of the DNA strands. We reiterate, however, that accurate helical geometry data is directly obtained from the parameters, $h$ and $\Omega_{so}$, and does not depend on the fluorophore dye-linker structures.

**Structural features of RecA-bound dsDNA**

After confirming that interfluorophore energy transfer measurements using this system of fluorescein- and Rox-labeled DNA strands could satisfactorily describe the
canonical B-DNA geometry, the energy transfer efficiency was measured on the same series of dsDNA molecules after addition of excess RecA protein. For each of the six dsDNA molecules, an increase of the fluorescein emission peak and decrease of Rox emission were observed, indicating that the dsDNA was stretched and/or unwound by the binding of RecA protein. The fitted helical rise for the RecA-bound dsDNA is $4.9 \pm 0.1 \, \text{Å}$ and the helical twist angle is $17 \pm 4^\circ$. The significant difference in the structure of this RecA-bound dsDNA in comparison to standard B-form dsDNA revealed the simultaneous extension and unwinding of the DNA molecule by the binding of the RecA protein. As discussed below, the values of these two helical geometry parameters determined simultaneously from FRET efficiency measurements are in excellent agreement with those parameter values measured individually using other methods. Furthermore, the analysis of interfluorophore distances in the context of the helical geometry model provides an important modification to a widely accepted model of the RecA-dsDNA nucleoprotein filament.

Over the past 20 years, considerable progress has been made in determining the structures of the RecA protein as well as those of the RecA-ssDNA and the RecA-dsDNA filaments. EM provided the first images of the RecA nucleoprotein filament (reviewed in 124). It was observed that the cofactor had a drastic effect on the observed pitch of the helical filament. A "collapsed" filament with a helical pitch of ca. 75 Å occurs in the absence of a cofactor or with ADP. In contrast, an "extended" filament (95-Å pitch) is
observed with nonhydrolyzable analogs of ATP. The extended conformation is the active form of the RecA nucleoprotein filament since the quaternary complex formed between RecA protein, ssDNA, ATP and Mg$^{2+}$ \textit{in vitro} is the species that pairs and exchanges homologous DNA strands (131,132). The crystal structure of a RecA-ADP complex provided the first high resolution view (2.3 Å) of the protein-only filament (133). However, the conformation of the filament in the crystal is probably that of the inactive collapsed filament since the pitch (83 Å) is not as extended as that seen when DNA is bound (134). The active conformation is achieved only in the presence of both DNA (either single- or double-stranded) and ATP (or a slowly- or non-hydrolyzable analog, such as ATP$\gamma$S or ADP·AlF$_4^-$, respectively). In the ensuing discussion, only complexes – and their related structural parameters – formed in the presence of ATP$\gamma$S (or ADP·AlF$_4^-$) will be addressed.

EM of RecA-dsDNA complexes has provided the most consistent and highest resolution images of RecA nucleoprotein filaments. Early EM images revealed that the contour lengths of RecA-dsDNA filaments were 50 - 60% longer than that of the underlying protein-free replicative form DNAs from bacteriophage (82,83,90), and allowed the determination of a helical rise per base pair, $h = 5.2 \pm 0.18$ Å (82). This value is identical within experimental uncertainty to that derived herein ($4.9 \pm 0.1$ Å) from FRET efficiencies for fluorophore pairs separated by no more than 20 bp.
While EM has allowed RecA nucleoprotein filaments to be directly visualized and a likely DNA binding location to be identified (136), it has not provided for the direct characterization of the DNA strands buried inside. A key limitation in this regard is the relative masses of the filament components: the two DNA strands in a RecA-dsDNA complex account for less than 5% of the total mass of the complex (136). Nevertheless, a comparison of the number of turns of the overall nucleoprotein filament images with the number of bp's in the substrate dsDNA revealed that the average pitch of the dsDNA was 18.6 bp ($\Omega_h = 19.4^\circ$) (88). Using an assay to monitor changes in the topological linking number of supercoiled DNA, Radding and coworkers first demonstrated a RecA-induced reduction in the twist of substrate dsDNA (19,66). Subsequently, several labs have used related assays to quantify the extent of unwinding, reporting helical repeats of $18.6 \pm 1.3$ (92), $17 \pm 1$ (93), $21 \pm 5$ (90), and $17.9 \pm 0.3$ bp (137). Thus, in spite of a number of critical assumptions and limitations in the topological assays (for a discussion see 92,93,137,138,139,140), a remarkable agreement with the prediction derived from EM images has been achieved. The helical repeat of the RecA-bound dsDNA derived from the FRET efficiency analysis [$360 / (17 \pm 4) = 21 \pm 5.0$ bp/turn] is in excellent agreement with this collection of literature values.

The quantitative agreement between the FRET-derived helical geometry parameters for RecA-bound dsDNA and the published characterizations of RecA-DNA filaments provides an important validation of the FRET efficiency measurements for the
quantitative elucidation of RecA-DNA complex structures. Moreover, because the helical geometry parameters were determined from interfluorophore separations of no more than 20 nucleotides, these parameters provide important confirmation that the average helical parameters derived from EM and topological measurements are reflected by the local DNA structure. Hence, the combination of global and local structural analyses corroborates the hypothesis that the DNA structure is uniform throughout the nucleoprotein filament.

**Functional implications of the RecA-bound dsDNA structure**

The apparent local structural uniformity of the DNA is important in the context of RecA-DNA interactions. As discussed previously, the fact that the DNA adopts the same helical geometry as the RecA filament suggests that the observed stoichiometry of 3 bp per RecA monomer arises from structural coincidence of the protein and DNA within the helical filament. The conclusion that contacts between the protein and DNA within the complex would be equivalent for every RecA monomer is supported by the uniformity of the local structure of the DNA (Egelman & Stasiak, 1989).

Published results suggest that the RecA protein interacts with dsDNA along its minor groove (90,183,235). One possible result of the stretched and underwound DNA structure imposed by RecA binding is to widen the minor groove and move its floor closer to the cylindrical surface contour of the DNA molecule (141). Such structural
changes would likely provide for more facile interaction between the interior surface of the RecA protein filament and the exposed floor of the minor groove of DNA, as described for TATA box-binding protein (142,236) and architectural proteins (237) binding their cognate dsDNA sequences. Although it is generally accepted that RecA binds DNA without high sequence specificity, the protein demonstrates some sequence preferences in its binding to single-stranded DNA (103,238-242). The putative existence of direct contacts between RecA protein side chains and DNA functional groups is supported by the aforementioned structural coincidence of the protein filament with the bound DNA. Given the open interior of the RecA filament (133), the formation of such interactions would require the base pairs of DNA to be displaced away from the central helix axis.

Based on the conclusions described above for the B-DNA dye-linker conformations, Rox most likely sits on the major groove edge of the thymine-adenine base pair with a distance of 2 Å to the base-pair center. Moreover, if the base pairs are displaced away from the helix axis, the axis must be located within the major groove (161). This requirement results from the fact that the DNA strands in the RecA-dsDNA complex are right-handed (as indicated by the positive sign of the helical twist angle). Therefore, the maximum distance from the helix axis to the base-pair center is the sum of the distance from the helix center to Rox (r_s = 3 Å) and the distance from Rox to the base-pair center (2 Å). This approximation yields a distance of 5 ± 3 Å between the
helix axis and each base-pair center. Hence, the base-pair centers are likely displaced away from the helix axis and a hollow helix center may be present. Consistent with this hypothesis, the application of a geometric model requiring coincidence of the base-pair centers and helix axis generates an unreasonable fit (not shown).

The interfluorophore distance data presented herein are consistent with the widely accepted model of stretched and underwound RecA-bound dsDNA, but the data further suggest that the base pairs are not coaxially stacked. The possibility that the base pairs are radially displaced from the helix axis is not incompatible with the helical geometry parameters, $h = 5 \text{ Å}$ and $\Omega_h = 20^\circ$. Egelman and coworkers have shown that the radial distance between the base pair centers and the filament axis is a function of the local separation between phosphate groups along each DNA strand (136). In turn, the helical pathlength along the phosphate backbone is constrained by stereochemical forces and has a maximum value of 7.6 Å. Moreover, as the models in Figure 4 of Egelman & Yu (1989) indicate, relatively small changes in the local separation between phosphate groups (e.g., 5.5 → 7.2 Å) are correlated with substantial changes in helical radius of the DNA (e.g., 6 → 15 Å). Importantly, DNA structure models entirely consistent with the combination of constraints imposed by the measured helical geometry parameters, stereochemical backbone limitations, and significant radial displacement of the base pairs have been predicted for stretched dsDNA (142) as well as for dsDNA in RecA.
nucleoprotein filaments (136). Our data are consistent with the DNA structural models
by previous authors presented (136,142). The implications of such a DNA structure for
the RecA·tsDNA strand exchange intermediate are discussed further below.

**Structural features of three DNA strands in the strand-exchange intermediate**

We used the same method to measure the energy transfer efficiencies as a
function of base-pair separation for each strand in the RecA·tsDNA complex representing
the strand exchange intermediate. In each complex, Rox was labeled on one strand of
the three strands and its position was varied throughout the strand. This enabled us to
isolate the helical geometry information for that particular strand relative to the
fluorescein-labeled strand.

For the complementary and outgoing strands, the requirement that the
corresponding bases on the complementary and outgoing strands share the same base pair
plane results in significantly different fitted twist angles for the two strands. In turn, this
difference excludes the possibility that the two strands are interwound in a cohelical
manner. This inference is challenged by the following problems. First, a separation of
the two strands would require that the topological tension between the originally
interwound strands be released. Yet, it is well known that the RecA protein does not
demonstrate helicase activity (123), and under our reaction conditions, ATP hydrolysis is
absent. Second, if the outgoing strand is not cohelical with the RecA filament or the
complementary strand, the protein monomers must make discontinuous contacts with one of the two strands. To date, the sort of structural discontinuities in the protein filament required for such contacts have not been reported. Lack of a satisfactory solution to these two problems led us to reconsider the structural assumption.

These issues were readily resolved by shifting the outgoing strand three bases toward its 5' end, as predicted in a recent report (184). For a three base registry change between the outgoing and complementary strands, the best-fit helical parameters are $h = 4.9 \pm 1.3 \text{ Å}$, $\Omega_n = 13 \pm 5^\circ$ for the complementary strand, and $h = 5.2 \pm 0.1 \text{ Å}$, $\Omega_n = 22 \pm 2^\circ$ for the outgoing strand. The conformation of the two strands now exhibits marked similarity. When considered together with the helical parameters of the incoming strand ($h = 1.9 \pm 3.7 \text{ Å}$, $\Omega_n = 11 \pm 3.3^\circ$), we noted that all three sets of helical parameters fall into the category of the RecA-bound extended and unwound DNA strands (weighted mean parameters, $h = 5.2 \pm 0.1 \text{ Å}$ and $\Omega_n = 18 \pm 3^\circ$).

Indeed, the data sets of all three RecA-tDNA complexes can be fit simultaneously by a common helical geometry (Figure 2.7). The best-fit common helical parameters for a registry shift of three bases for the outgoing strand, $h = 4.8 \pm 0.4 \text{ Å}$, and $\Omega_n = 15 \pm 4^\circ$, are consistent with the weighted mean parameters. The global analysis results are consistent with a structural model in which all three RecA-bound DNA strands are cohelical within the RecA filament. This is the first direct evidence that all three
DNA strands in the RecA·tsDNA complex are simultaneously extended and unwound and share a common helical geometry with the RecA-bound dsDNA (Table 2.4).

EM has been invaluable in imaging RecA protein-only and protein-DNA filaments. Micrographs of RecA-DNA complexes trapped during sequential stages of in vitro strand exchange have allowed visualization of genetic recombination intermediates (100,101). Moreover, it has been proposed that the stretched and partially unwound DNA structure imposed by RecA binding is instrumental in the process of strand exchange (43,99,101,141,143,155,243,244). In spite of this success, no structures of RecA·tsDNA complexes to the resolution of the protein monomer have been reported, in part because the conditions of the reaction do not allow continuous coverage of the DNA by RecA and the necessary sample fixation results in compaction of the complexes (100,101). Thus, hypotheses regarding the molecular mechanism of strand exchange have been largely evaluated in the absence of direct structural data to this point.

Radding and coworkers used the topological assay described above to demonstrate that there is extensive unwinding of the substrate dsDNA during the early stages of DNA strand exchange initiated by a RecA·ssDNA nucleoprotein filament (19,66). Several groups have corroborated and extended the early results (245-247). More recently, two laboratories have employed related assays to quantify the extent of unwinding of the substrate dsDNA, reporting helical repeats of ca. 19 (248) and 27 (249) bp per turn in the region of homologous pairing. The helicity determined herein for all
three strands of the RecA·tsDNA complex \[360 / (18 \pm 3) = 20 \pm 3.3 \text{ bp/turn}\] agrees closely with the smaller reported value (248). Although many limitations in the use of topological assays have been discussed (see above), it is not clear whether the larger reported value (249) is a result of a systematic difference in those authors' experiments (e.g., the inclusion of ADP in the reaction buffer) or may simply reflect a large inherent uncertainty in such an analysis (90). We conclude that all three strands in the RecA·tsDNA complex share a common helicity and that the average helical twist parameter for those strands \(\Omega = 18 \pm 3\) is indistinguishable from that of the better characterized RecA·dsDNA complex.

To the best of our knowledge, no laboratory has published a measurement of the helical rise or pitch for any of the DNA strands in a RecA·tsDNA complex. However, Malkov and coworkers concluded that the distribution of complementary and outgoing strand breaks resulting from the Auger radiodecay of a site-specific \(^{125}\text{I}\) label were consistent with an extended helical rise for both strands (184). The average helical rise between nucleotides of each of the three strands \(h = 5.2 \pm 0.1 \text{ Å}\) confirms the expectation that the extension of the DNA in the RecA·tsDNA complex is indistinguishable from that observed for RecA-bound dsDNA.
Implications for the structure of the strand exchange intermediate

Using agarose gel mobility shift assays, we demonstrated that the conjugation of the dye-linker structures to DNA on the major groove edge of thymine has no effect on the efficiency of strand exchange. Moreover, the assay results allow us to conclude that the strand exchange intermediate retains all three DNA strands bound to the protein filament. This is consistent with the conclusion drawn from previous studies (118,119,122,130,132,200) that no strand exchange products are released from the filament in the absence of ATP hydrolysis. Of course, the gel assay result is also consistent with the observation of strong FRET signals when using the fluorophore-conjugated outgoing strand as the excited state energy donor or acceptor. Finally, the analysis above demonstrated that an outgoing strand registry shift of three nucleotides is required for reasonable, internally consistent structural parameters. Taken together, these conclusions support only one model for the RecA·tsDNA complex we have probed. Namely, that the RecA·tsDNA complex must represent an advanced, post-strand-exchange intermediate wherein the outgoing strand has been unpaired from the original dsDNA but remains bound to the nucleoprotein filament. Importantly, the combination of interstrand distances (inferred from interfluorophore distances, Table 2.2) and registry shift of the outgoing strand exclude the possibility that a triplex DNA – involving at least one set of novel, non-Watson-Crick hydrogen-bond interactions between the bases of an
intact duplex and those of a third strand – represents this RecA·tsDNA complex. Elements of the structural model are consistent with recent data relying on a variety of biochemical methods including chemical modification (159,249), affinity modification (158,181,185), topology testing (248,250), fluorescence (200,201), gel mobility shift (251), and radioprobing (184).

The next interesting question to address involves the structural organization of the three strands within the RecA filament. Given the common helical parameters of all three strands and their structural coincidence with the protein filament, the simplest assumption for the relative positions of the three strands is that they have a common helix axis and are interwound with each other. This arrangement presents minimal topological problems and does not require any energy input from ATP hydrolysis. However, the determination of the location of the three strands in relation to the helix center requires a knowledge of the conformation of the dye linker arms.

Based on the conclusions regarding the B-DNA dye-linker conformations and the approximations described for the RecA·dsDNA complex, the length of the smaller dye-linker radial projection \( r_s \) likely provides information about the position of Rox and the strand to which it is conjugated. An analysis of the radial distances between each Rox and the helix center \( r_s \); Table 2.4) indicate that the Rox of the outgoing, complementary, and incoming strands are located \( 4 \pm 1, 7 \pm 5, \) and \( 14 \pm 14 \) Å, respectively, from the common helix axis. There is considerable uncertainty in two of the three \( r_s \) values and it is not
possible to distinguish the three strands' radial positions with confidence. Nevertheless, when taken together with the magnitudes of the \( r_n \) parameters, the results of the global analysis suggest that \( r_n > 0 \) for all three strands (Table 2.4). As discussed above, the position of Rox is likely within 2 Å of the thymine to which it is conjugated. Hence, the maximum distance from the helix center to the base center of a strand is the sum of the distance from the helix center to Rox and the distance from Rox to the base center (2 Å). We conclude that the distances from the helix center to each of the three strands likely fall within the approximate range, 5 – 10 Å. Thus, the radial positions of Rox labels on each strand strongly suggests that each strand is displaced away from the common helical axis. This finding is in qualitative agreement with the structure for the RecA-dsDNA complex proposed above.

From a consideration of physical constraints imposed by the RecA filament and stereochemical constraints imposed by the covalent bonds in the phosphate backbone (see above), it has been estimated that the phosphate backbone of a RecA-bound DNA strand will possess a helical radius between 6 and 17 Å (136). This limitation coincides with the probable radial distances from the helix center to the DNA strands described above. Therefore, the base pairs composed from the incoming and complementary strands are likely located on the surface of a cylinder with a radius of 5 – 10 Å, with the base pairs displaced into the minor groove of the right-handed double-helix (see above). Moreover, the outgoing strand must be located in the major groove of the incoming-complementary
heteroduplex at a similar radial displacement (Figure 2.10b). Notably, this arrangement of strands would leave the minor groove of the complementary-incoming heteroduplex DNA available for protein-DNA interactions. Previous work showed that the RecA protein binds dsDNA along its minor groove (90,183,235). Because the heteroduplex DNA in the RecA·tsDNA complex exhibits an extended and unwound conformation indistinguishable from that of RecA-bound dsDNA in our FRET work, it is likely that RecA protein also binds to the heteroduplex along its minor groove. This inference is based on the theory that similar RecA-dsDNA interactions will produce similar conformational changes in the bound DNA molecule and is consistent with the idea that the RecA·dsDNA is similar to the product of strand exchange (103).

Implications for mechanistic aspects of homology recognition

In the context of the post-strand-exchange RecA·tsDNA complex described here, the location of the outgoing strand within the major groove of the complementary-incoming heteroduplex necessarily implies that homology recognition was initiated in the minor groove of the substrate dsDNA. The question of whether RecA mediates homology recognition via major- or minor-groove contacts with the substrate dsDNA has been addressed in several laboratories over the past decade. To a large extent, those experiments have produced controversial results. Indeed, the hypothesis that homology is recognized via the major groove has been supported by computational modeling (155)
Figure 2.10 Models depicting possible arrangements of bases corresponding to the three DNA strands in the RecA-tsDNA complex. The gray circles, with radii corresponding to 17 Å (see the text) represent crosssections of the RecA-tsDNA nucleoprotein filament. The bases of the incoming (I, blue), complementary (C, red), outgoing (O, green) strands are shown. (a) Model for the post-strand-exchange heteroduplex and outgoing strand accounting for major groove homology recognition as proposed by Malkov and coworkers (Malkov et al., 2000). The heteroduplex DNA (comprised of the incoming and complementary strands) is located in the center of the nucleoprotein cylinder (white cross) with the outgoing strand in the minor groove of the heteroduplex DNA near the surface of the cylinder. The indicated distances between an $^{125}$I atom attached to C5 of cytosine (incoming strand) and the C1' atoms of the complementary and outgoing strands correspond to those proposed to account for previous radioprobing data (see the text and Figure 4(d) of Malkov et al., 2000). (b) Model for the post-strand-exchange heteroduplex and outgoing strand consistent with this FRET work. The dashed circles represent the radial distances of 5 Å and 10 Å (see the text). The heteroduplex DNA is near the surface of the cylinder while the outgoing strand is interwound in the major groove of the heteroduplex. This position of the outgoing strand is consistent with minor groove homology recognition, in contrast to the proposal of Malkov et al. Nevertheless, the FRET-derived model adequately accounts for the two distance requirements derived from those authors' radioprobing experiments.
and radioprobing experiments (184), while the alternative hypothesis – recognition of homology from the minor groove – has been supported by chemical modification (159,171), atomic mutagenesis (179), ligand displacement (182), and affinity modification (158,181,185) experiments as well as computational modeling (141). Our data are clearly in accord with recognition of homology occurring via minor groove contacts. Moreover, the accompanying structural model can account for much of the previously published data.

The experimental evidence related to the issue of whether homology recognition occurs from the major or minor groove of the substrate dsDNA has been reviewed recently in the context of arguments favoring the former (184) and latter (141,158,181) hypotheses. In particular, in the context of their elegant experiments, Malkov et al. (2000) were able to rationalize much of the published data in terms of a model favoring major-groove homology recognition. These authors’ own results were derived from a powerful biochemical approach to the homology recognition issue using the novel technique of radioprobing with the Auger radiodecay of $^{125}\text{I}$ (252). Based on the distribution of DNA strand breaks localized near a site-specific [5-$^{125}\text{I}$]dC residue, Malkov and coworkers were able to determine relative distances between the radioactive iodine atom and various nucleotide positions on the outgoing and incoming strands. In the context of a model wherein the heteroduplex DNA is found near the central filament axis and the outgoing strand forms an extended helix on the surface of the heteroduplex,
two of the strongest pieces of evidence for the minor groove location of the outgoing strand were the observations of (1) “a low overall frequency of breaks of the outgoing strand compared to the complementary strand”, and (2) “a flat pattern and 3' shift of the cleavage of the outgoing strand” (184). According to those authors, this evidence was best reconciled with a RecA-tsDNA structure wherein the outgoing strand is displaced into the minor groove of the more central heteroduplex, indicating that homology must have been detected in the major groove of the substrate duplex.

The first condition requires that the distance between the iodine atom at C5 of a complementary strand cytosine and C1' of the outgoing strand nucleotide is around two-fold longer than the corresponding distance to C1' of the complementary strand nucleotide (Figure 2.10a). The second condition requires that the contour length of the complementary and outgoing strands be substantially different. The pair of evidentiary data presented by Malkov et al. (2000) is not sufficient for uniquely supporting the location of the outgoing strand in the major groove of the heteroduplex. In fact, a model in which the bases of the heteroduplex dsDNA, rather than the bases of the outgoing strand, are displaced away from the central filament axis can adequately account for both key observations of Malkov et al. (2000) (Figure 2.10b). Given the magnitudes of the $r_i$ parameters described herein (Table 2.4), we conclude that the bases of the heteroduplex DNA are displaced away from the central helix axis and that the outgoing strand forms an extended helix intertwined along the major groove of the heteroduplex. Such a groove
location is consistent with the conclusions of the Dervan and Adzuma groups, and suggests that homology recognition leading to the observed post-strand-exchange intermediate occurred via minor groove contacts with the substrate dsDNA.

While these structural conclusions are consistent with the results of many other researchers in the field as well as internally consisted with our own data, we emphasize that the mechanistic conclusion is inferred from the steady-state structure of an intermediate following homology recognition and strand exchange. Ultimately, the detection of homology and exchange of DNA strands are active molecular processes whose mechanistic details remain elusive. The results presented here provide a structural framework upon which to base subsequent mechanistic hypotheses. Moreover, the remarkable generality of real-time fluorescence-based assays may offer an approach to tease apart the intricate mechanistic details of homology recognition and strand exchange mediated by the RecA protein.

Summary

We have presented a systematic analysis of FRET efficiencies to infer interfluorophore distances describing the structures of the DNA strands within RecA-DNA complexes. The helical geometry parameters of the RecA·dsDNA complex, $h = 4.9 \pm 0.1 \, \text{Å}$ and $\Omega_h = 17 \pm 4^\circ$, are consistent with structural conclusions derived from EM and other classic biochemical methods. Hence, the direct FRET measurements are accurate
and suitable for the solution-phase elucidation of RecA-DNA complex structures. In addition, the interfluorophore distance measurements unexpectedly suggest that the DNA base pairs are displaced away from the central helix axis, providing for an open helix center. The key structural conclusions regarding the RecA-tsDNA complex can be summarized as: (1) all three strands are extended and unwound to a similar extent relative to B-DNA, and all show \( h \)- and \( \Omega_n \)-values that are similar to RecA-bound dsDNA; (2) the outgoing strand is shifted by about three nt positions with respect to its registry with the complementary and incoming strands; (3) the complex represents a late, post-strand-exchange intermediate; (4) the bases of the incoming and complementary strands are displaced away from the helix axis toward the minor groove of the heteroduplex; and (5) the bases of the outgoing strand lie in the major groove of the heteroduplex. Taken together, the data corroborate an important role for interactions with the minor groove of dsDNA, both between protein and DNA for RecA-bound dsDNA and heteroduplex DNA and between the nucleoprotein filament and the substrate dsDNA in the recognition of sequence homology.

**MATERIALS AND METHODS**

**RecA Protein**

*E. coli* RecA was purified as described (253) and stored in aqueous buffer (25 mM Tris-HCl, pH 7.5, 1 mM DTT, 5% glycerol) at -80 °C. The protein concentration
was determined spectrophotometrically using an extinction coefficient at 279 nm of 0.59 mg⁻¹ml⁻¹ (254).

**Oligodeoxyribonucleotides**

ODNs containing single fluorescent labels were synthesized and HPLC-purified by Integrated DNA Technologies. ODNs without fluorescent labels were synthesized on an automated nucleic acids synthesis system (PerSeptive Biosystems, Model 8905) and purified using 12% denaturing polyacrylamide gel electrophoresis. The DNA bands were viewed by short-wavelength UV shadowing and eluted from gel slices using a Schleicher & Schuell Elutrap in 50 mM Tris-Borate, pH 8.0, 2 mM EDTA. The DNA was then recovered by ethanol precipitation followed by dialysis against Milli-Q water, the pH of which was adjusted to 7.5. The dialyzed ODN samples were then lyophilized and stored at −80 °C. The concentrations of the ODNs were determined using their extinction coefficients at 260 nm calculated by the nearest neighbor method from nucleoside monomer and dimmers (255). The absorbance contributions of labeled fluorophores at 260 nm were subtracted using the extinction coefficients of the fluorophores at 260 nm (fluorescein: 19361 M⁻¹cm⁻¹, Rox: 38065 M⁻¹cm⁻¹) (256). All concentrations of ODNs were expressed in molecules.

Double-stranded DNA was prepared by mixing equal molar concentrations of the two complementary strands in 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA.
The mixtures were incubated at 80 °C for 4 min followed by a slow cooling back down to room temperature. All duplexes were analyzed on 12% native polyacrylamide gel to insure that no excess single strands were present.

**Reaction conditions**

For the reaction systems investigated, reactions and pre-reaction incubations were carried out at 37 °C in Rx6 Buffer (25 mM Tris·HOAc, pH 7.5, 2 mM DTT, and 6 mM Mg(OAc)$_2$) or Rx2 Buffer (25 mM Tris·HOAc, pH 7.5, 2 mM DTT, and 2 mM Mg(OAc)$_2$) unless otherwise specified.

**Agarose gel mobility shift assay**

Reactions using either dye-labeled DNA substrates or unlabeled DNA substrates performed under identical reaction conditions were incubated in Rx6 Buffer for 1 h, then equal amounts of each reaction, either untreated or treated with 0.1 mg/mL proteinase K and 0.5% SDS at 37 °C for 25 min to remove RecA, were loaded on a 3% Metaphor agarose gel (FMC Bioproducts) in chilled 50 mM Tris-borate, pH 7.5, 2 mM EDTA, 2 mM Mg(OAc)$_2$. The gel was run under “Rapid Run” conditions (as described in the manufacturer’s protocol), at 4 °C for 2 h. The gel was stained with SyberGold (Molecular Probes) at room temperature for 20 min and viewed under an UV lamp.
Steady-state fluorescence measurements

All fluorescence data were collected on a SLM 8100 spectrofluorometer (SLM-Aminco, Illinois) with a 3-mm x 3-mm cuvette. The excitation and emission beams were polarized with Glan-Thompson polarizing prisms set at the magic angle values (excitation 0° and emission 54.7°) to eliminate polarization artifacts. The fluorescein and Rox emission spectra were collected with excitation wavelengths of 492 nm and 585 nm, respectively, unless otherwise noted. In both cases, the bandpass was set at 2 nm for the excitation monochromator and 4 nm for the emission monochromator. Scan rate was set at 0.95 nm/sec with 1 nm per step. Correction for the inner filter effect was not necessary since the absorbance values at the excitation wavelength were less than 0.05. All spectra were corrected for background using an identical control reaction without fluorophores. Instrumental response was also corrected by comparing to the emission spectra of a standard fluorophore using the following formula (209,211):

\[
S_{\text{sam}}^{\text{corr}}(\lambda_{\text{ex}}, \lambda) = \frac{[S_{\text{sam}}^{\text{obs}}(\lambda_{\text{ex}}, \lambda) - S_{\text{buf}}^{\text{obs}}(\lambda_{\text{ex}}, \lambda)] \cdot S_{\text{std}}^{\text{corr}}(\lambda_{\text{ex}}, \lambda)}{S_{\text{std}}^{\text{obs}}(\lambda_{\text{ex}}, \lambda)}
\]

(4)

where \( S(\lambda_{\text{ex}}, \lambda) \) is a wavelength \( \lambda \) dependent emission spectrum with excitation at \( \lambda_{\text{ex}} \). The subscripts indicate the spectrum is taken on the sample interested (\text{sam}), buffer (\text{buf}) or the standard fluorophore (\text{std}) used to correct for instrumental response. The superscripts indicate the spectrum is observed on the SLM8100 fluorometer (\text{obs}) or is the spectrum after correction (\text{corr}). Fluorescein (F-1300, Molecular probes) in 0.1 M NaOH was
used as the standard fluorophore. The correct emission spectrum of fluorescein was provided by Molecular Probes.

For the dsDNA system, the emission spectra of fluorescein and Rox were recorded using 0.1 µM duplex in Rx2 Buffer at 37 °C. RecA protein (4.0 µM final concentration) and ATPγS (0.2 mM final concentration) were then added to the cuvette, the mixture was incubated at 37 °C for 1 h, and the emission spectra of fluorescein and Rox were taken again for the RecA-dsDNA system. For the RecA-tsDNA-C and -O series, 0.73 µM incoming 74mer (74X or 74C) was first incubated with 18.0 µM RecA protein in Rx2 Buffer supplemented with 0.5 mM ATPγS at 37 °C for 10 min, then the dual-labeled partial dsDNA (FRi) was added at a final concentration of 0.1 µM. Aliquots of a 500-mM Mg(OAc)2 stock were also added together with the dsDNA to bring the final concentration of Mg2+ to 6 mM (i.e., Rx6 Buffer conditions). This concentration of Mg2+ (6 mM) was chosen to balance the effects of the precipitation of RecA protein at high Mg2+ concentration and the ability of Mg2+ to facilitate completion of the strand exchange (data not shown). The solutions were incubated at 37 °C for 1 h before the emission spectra of fluorescein and Rox were collected. For the RecA-tsDNA-I series, 0.5 µM ssDNA Ri was incubated with 18.5 µM RecA protein in Rx2 Buffer at 37 °C for 10 min, and fluorescein-labeled full-length dsDNA ds74F was added to a final concentration of 0.5 µM along with aliquots of Mg(OAc)2 to a final Mg2+ concentration
of 6 mM (i.e., Rx6 Buffer conditions). The solutions were incubated for 1 h at 37 °C before the emission spectra of fluorescein and Rox were recorded.

For the reactions performed using ADP·AlF₄ instead of ATPγS, the following reaction conditions were used. For RecA·dsDNA, 4.0 μM RecA protein and 0.1 μM partial dsDNA FR, were incubated at 37 °C for 1 h in Rx2 Buffer supplemented with 5 mM ADP, 0.4 mM Al(NO₃)₃ and 10 mM NaF before the emission spectra of fluorescein and Rox were collected. For the RecA·tsDNA systems, the incoming ssDNA was first incubated at 37 °C for 10 min with an appropriate amount of the RecA protein in Rx2 Buffer supplemented with 5 mM ADP, 0.4 mM Al(NO₃)₃ and 10 mM NaF, and then the homologous dsDNA substrate was added. Aliquots of 500 mM Mg(OAc)₂ stock were also added together with the dsDNA to bring the final concentration of Mg²⁺ to 6 mM (Rx6 Buffer conditions).

**Steady-state fluorescence anisotropy measurements**

All steady-state fluorescence anisotropy measurements were performed in L format using Glan-Thompson polarizers placed in the excitation and emission channels of the SLM 8100 spectrofluorometer. The fluorescence anisotropy of the sample was calculated using the following equation (209):

\[
r = \frac{I_{VV} - G \cdot I_{VH}}{I_{VV} + 2G \cdot I_{VH}}
\]  (5)
where $I$ is the fluorescence intensity and the first and second subscripts refer to vertical ($V$) polarization of the excitation and vertical ($V$) or horizontal ($H$) polarization of the emitted light, respectively. The factor $G = I_{HV}/I_{HH}$ corrects for the different sensitivity of the emission monochromator for vertically and horizontally polarized light. For each sample, the anisotropy was measured 10 times with 1-s measuring times, and the values were averaged. Fluorescein and Rox anisotropy data were measured with emission wavelengths of 520 nm and 605 nm, respectively.

**Determination of the limiting anisotropy**

The limiting anisotropy $r_l$ was measured by constructing Perrin plots in which the steady-state polarization was measured for various viscosities (209,257)

$$\frac{1}{r} = \frac{1}{r_l} + \frac{1}{A} \cdot \frac{1}{\eta}$$  \hspace{1cm} (6)

where $r$ is the measured sample anisotropy; $A$ is a constant which is related to the lifetime of the fluorophore, the volume of the rotating unit and the temperature at the measurement; and $\eta$ is the viscosity of the sample. The solvent viscosity of a fluorescein-only labeled or a Rox-only labeled sample was varied at 30 °C by adding aliquots of 62% stock sucrose solution. Published viscosity values for aqueous sucrose solutions at 30 °C were used (258). The emission spectra of the fluorophores were not altered as the sucrose concentration was increased. Perrin plots were constructed, and the
limiting anisotropies of both dyes were obtained by extrapolating the fitted curve to reach
the y-intercept according to Perrin equation.

**Determination of the quantum yield of fluorescein**

The quantum yield of the donor fluorescein $\Phi_F$ was determined for a fluorescein-
labeled DNA sample in all reaction systems by the comparative method (209,211):

$$\Phi_F = \frac{\Phi_{std} \cdot A_{std} \cdot \int S_F(\lambda_{ex}, \lambda)d\lambda}{A_F \cdot \int S_{std}(\lambda_{ex}, \lambda)d\lambda}$$  \hspace{1cm} (7)

where $\Phi_{std}$ is the quantum yield of a standard fluorophore, $\int S_{std}(\lambda_{ex}, \lambda)d\lambda$ and $\int S_F(\lambda_{ex}, \lambda)d\lambda$ are the areas under the corrected emission spectra of the standard fluorophore and fluorescein. $A_{std}$ and $A_F$ are the absorbance of the standard fluorophore and the fluorescein sample at a given excitation wavelength using the same concentration at which their fluorescence intensities were measured.

**Determination of Förster critical distance $R_0$**

The Förster critical distance $R_0$ is characteristic of the position-independent
spectral property of the dyes. This, in turn, requires the measurements of the donor
quantum yield in the absence of the acceptor ($\Phi_0$), the overlap integral $J$ that characterizes the resonance between the donor and acceptor dipoles (211):

$$R_0^6 = 8.875 \times 10^{-5} \frac{\kappa^2 \cdot J \cdot \Phi_d}{n^4}$$  \hspace{1cm} (8)
where $\kappa^2$ is the orientation factor, which has a value of $2/3$ when the donor and the acceptor rotate rapidly in a short time compared to the donor lifetime. $\Phi_d$ is the donor quantum yield in the absence of acceptor and $n$ is the refractive index of the medium that is normally measured as 1.4 for aqueous buffer. The overlap integral $J$ characterizes the resonance between the donor and acceptor dipoles and is evaluated by integration of the mutual area of the overlap between the donor emission spectrum $S_d(\lambda_{ex}, \lambda)$ and the acceptor absorption spectrum $\varepsilon_a(\lambda)$ using the following equation (211):

$$J = \frac{\int [S_d(\lambda_{ex}, \lambda) \cdot \varepsilon_a \cdot \lambda^4] d\lambda}{\int S_d(\lambda_{ex}, \lambda) d\lambda}$$  \hspace{1cm} (9)

**Determination of the maximum and minimum value of $R_0$**

When both depolarization factors are positive, $\kappa_{\text{max}}^2$ and $\kappa_{\text{min}}^2$ can be calculated from the following equation:

$$\kappa_{\text{max}}^2 = \frac{2}{3} \cdot (1 + d_d^x + d_a^x + 3 \cdot d_d^y \cdot d_a^y)$$
$$\kappa_{\text{min}}^2 = \frac{2}{3} \cdot [1 - \frac{1}{2} \cdot (d_d^x + d_a^x)]$$  \hspace{1cm} (10)

where $d_d^x$ and $d_a^x$ are the axial depolarization factors for the donor and acceptor, respectively. The axial depolarization factors $d^x$ can be calculated from the limiting anisotropies ($r_i$) and fundamental anisotropies ($r_0$) of the donor and acceptor:

$$d^x = \frac{r_i}{\sqrt{r_0}}$$  \hspace{1cm} (11)
The value of $d^*$ represents the depolarization factor due to segmental motion of the donor ($d^*_d$) or acceptor ($d^*_a$), but not the depolarization due to overall rotational diffusion of the protein. The overall rotational diffusion is not important because it does not change the donor-acceptor orientation.

After the limiting anisotropy of the fluorophores are obtained, maximum and minimum value of the Förster critical distance $R_0$ can be calculated using the following formula:

$$R_{0,\text{min}} = \left(\frac{3}{2} \kappa_{\text{min}}^{\frac{3}{2}}\right)^{\frac{1}{3}} \cdot R_0$$

$$R_{0,\text{max}} = \left(\frac{3}{2} \kappa_{\text{max}}^{\frac{3}{2}}\right)^{\frac{1}{3}} \cdot R_0$$

(12)

RecA-DNA complex dissociation constant measurements

For the RecA-dsDNA complex, the concentration of a fluorescein-labeled partial dsDNA FX was kept constant at 0.1 $\mu$M in 400 $\mu$L of Rx2 Buffer at 37 °C, while small aliquots of RecA stock solution were added. The RecA final concentration was varied from 1 $\mu$M to 9 $\mu$M. For the RecA-tsDNA-C and -O systems, a RecA·74F presynaptic filament was first formed between 0.1 $\mu$M 74F ssDNA and 3 $\mu$M RecA protein in 400 $\mu$L Rx6 Buffer and incubated at 37 °C for 10 min, then small aliquots of concentrated partial plain dsDNA (XX) stock of 9.85 $\mu$M were added. The final concentration of XX was varied from 0.02 $\mu$M to 0.6 $\mu$M. For the RecA-tsDNA-I system, a Rox-labeled ssDNA 45R$_{20}$ was used to form the RecA·45R$_{20}$ presynaptic filament. In 400 $\mu$L Rx6 Buffer, 0.1
μM 45R_{20} ssDNA and 3 μM RecA were incubated at 37 °C for 10 min, then small aliquots of concentrated plain dsDNA (ds74) stock of 7.6 μM were added. The final concentration of ds74 was varied from 0.02 μM to 0.5 μM. Between each addition of titrant, the reactions were allowed to equilibrate for a time period of 15 to 20 min before the emission spectrum and anisotropy of the fluorophore were recorded. Fluorescein emission data were collected from 510 to 530 nm, and Rox emission data were collected from 595 to 615 nm. Each spectrum represented an average of three repetitions. The fluorescence intensity was calculated by integrating the area under the spectral curve after the spectrum has been corrected for addition dilution and buffer background. A RecA·DNA binding isotherm was then constructed by plotting the fluorescence intensities at different titration points against the titrant’s concentration. The dissociation constant was determined by fitting the experimental data to the theoretical curve using the following equation derived from the mass balance equations for the rapid, one-step reversible association of molecules A and B:

\[
F_i = F_0 + \frac{\Delta F \cdot n}{2B_0} \cdot \left[ A + \frac{B}{n} + K_r - \sqrt{(A + \frac{B}{n} + K_r)^2 - 4A \cdot \frac{B}{n}} \right]
\]  

(13)

In this equation, \( F_i \) is the total fluorescence intensity; \( F_0 \) is the initial fluorescence when no titrant has been added; \( \Delta F \) is the total fluorescence change defined by the difference between the end fluorescence and initial fluorescence; \( A \) is the concentration of the titrant; \( B_0 \) is the concentration of the substrate being titrated; \( n \) is the stoichiometry...
number between the titrant and the substrate; and $K_d$ is the apparent dissociation constant between titrant R and substrate $B_0$.

For the RecA·dsDNA complex, the concentration of the dsDNA $B_0$ is expressed as
the following:

$$B_0 = \frac{C_{dsDNA} \cdot 74}{n}$$  

(14)

where $C_{dsDNA}$ is the concentration of the dsDNA expressed in molecule, 74 is the
length of the dsDNA, and $n$ is the stoichiometry number, which indicates how many
nucleotides were bound by each RecA monomer.

For the RecA·tsDNA complexes, $B_0$ is the concentration of the dual-labeled partial
dsDNA FR, in molecule, $A$ is the concentration of the preformed RecA·ssDNA filament
in molecules and $n$ equals one because presumably each RecA·ssDNA filament only
binds one homologous dsDNA molecule. Because RecA protein has a high affinity for
ssDNA (259), and the dissociation of RecA from linear ssDNA is blocked in the presence
of ATPγS (260), we used stoichiometric amounts of RecA protein to form the presynaptic
filament with ssDNA and counted it as one species in the reaction.

After the apparent dissociation constant for each system was obtained, a binding
isotherm plot was constructed. For RecA·dsDNA, RecA·tsDNA-C and –O, the following
equation is used to determine the concentration of the non-fluorescent species required to
keep at least 90% of the fluorescent species involved in the reaction:
\[ A = \theta \cdot B_0 + \frac{\theta}{1 - \theta} \cdot K_a \]  
(15)

where \( A \) is the concentration of the species in excess, \( B_0 \) is the concentration of the limiting fluorescent species kept at a fixed concentration, and \( \theta \) is the desired fractional completion of the reaction (\( \theta \geq 0.9 \) for all reactions), and \( K_a \) is the apparent dissociation constant determined from equation (13).

For RecA-tsDNA-I, the following equation was used, which is a simplified version of equation (15) when \( A \) equals \( B_0 \):

\[ A = \frac{\theta}{(1 - \theta)^\gamma} \cdot K_a \]  
(16)

Anisotropy measurements were also used to determine the apparent dissociation constants. The anisotropy of the fluorophore at each titration point was plotted against titrant concentration and the same curve fitting was performed using equation (13) except that the fluorescence terms have been changed to anisotropy.

**Energy transfer efficiency measurements**

The energy transfer efficiencies between fluorescein and Rox were calculated by measuring the \( (\text{Ratio})_A \), which is a normalized measurement of the enhancement of the fluorescence from the acceptor Rox due to FRET: (207)

\[
(\text{Ratio})_A = \frac{\int_{600}^{610} [S_{FR}(\lambda_{492}, \lambda) - a \cdot S_F(\lambda_{492}, \lambda)] d\lambda}{\int_{600}^{610} S_{FR}(\lambda_{585}, \lambda) d\lambda}
\]  
(17)
where $S(\lambda, \lambda)_{c}$ is a wavelength $\lambda$ dependent emission spectrum with excitation wavelength indicated at the subscript of the first $\lambda$. The subscripts $FR$ or $F$ of each $S(\lambda, \lambda)_{c}$ indicates the emission spectrum is taken on the sample containing both fluorescein and Rox ($FR$) or just fluorescein ($F$). The denominator of the equation is the fluorescence intensity of the acceptor Rox at 605 nm due to direct excitation at 585 nm in a reaction sample containing both donor and acceptor. The numerator is the fluorescence intensity of the same sample at Rox emission peak 605 nm with excitation at 492 nm after the fluorescein emission background at 605 nm has been subtracted by a normalizing factor $a$, which is expressed in the following formula:

$$
a = \frac{\int_{515}^{525} S_{FR}(\lambda_{492}, \lambda) d\lambda}{\int_{515}^{525} S_{c}(\lambda_{492}, \lambda) d\lambda}
$$

The denominator of equation (18) is the fluorescence intensity of fluorescein at 520 nm with excitation at 492 nm in the absence of Rox. The numerator is the fluorescence intensity of fluorescein at 520 nm with excitation at 492 nm in the presence of Rox.

This $(\text{Ratio})_{\lambda}$ is linearly dependent on the energy transfer efficiency $E$ and the donor fluorescein labeling fraction $d^{*}$ on the oligonucleotides:

$$
E = \frac{(\text{Ratio})_{\lambda} \cdot \varepsilon_{\text{NS}}^{\text{NS}} - \varepsilon_{\text{NS}}^{\text{NS}}}{\varepsilon_{\text{NS}}^{\text{NS}} \cdot d^{*}}
$$

where $\varepsilon$ is the molar absorption coefficient of the fluorophore (indicated by the subscripts) at a given wavelength (indicated by the superscripts). $(\text{Ratio})_{\lambda}$ normalizes
the measured sensitized FRET signal for the concentration and for the apparent quantum yield of the acceptor. In all experiments, the donor fluorescein labeling percentage $d'$ was verified to be 1 by comparing the labeled dye concentration with the oligonucleotides concentrations, which were calculated using their extinction coefficients at their absorbance maxima respectively.

**FRET efficiency versus base-pair separation data analysis**

For each reaction system, the energy transfer efficiency, $E$, was plotted as a function of the corresponding base pair separation, $n$. To determine values of $r_\alpha^2 + r_d^2$, $2r_ar_d$, $h$, $\Omega_n$, $\Delta\Omega$, and $\Delta h$ that best fit the data for each system, nonlinear least squares fitting to the $E$-$n$ function (Equation 3) was used with five (or four) of the parameters varied to minimize $\chi^2$. All curve fitting and subsequent analysis was performed using proFit 5.5.1 (Quantum Soft, Zürich, Switzerland). For each data set, an initial Monte Carlo search was performed with at least $2 \times 10^6$ iterations to search parameter space randomly for the best starting values. These starting parameters were then used as initial values in Levenberg-Marquardt fits to generate the best-fit parameters. For dsDNA, a B-form conformation was assumed ($h = 3.4$ Å, $\Omega_n = 36^\circ$ per base pair), and the other four parameters were variable. For RecA-dsDNA and RecA-tsDNA complexes, the dye-labeling parameter $\Delta h$ generated from B-form dsDNA was used while the other five parameters, $r_\alpha^2 + r_d^2$, $2r_ar_d$, $h$, $\Omega_n$ and $\Delta\Omega$ were left variable for best fitting. All the results
reported in this study are the best solutions regardless of initial values varied over a wide range.
CHAPTER 3

Construction and Evaluation of a Kinetic Scheme for RecA-mediated DNA Strand Exchange in the Absence of ATP Hydrolysis

Rapid Discrimination of DNA Strands and Double-stranded DNA Unwinding Accompany Homology Testing During the Earliest Phases of RecA-directed Strand Exchange
INTRODUCTION

We have characterized the structure of the RecA(tsDNA) complex, which is a key intermediate in RecA-mediated strand exchange (Chapter 2). Characterization of this strand exchange intermediate offered insight into its biological counterparts, which, in turn, provided valuable information on how homologous recombination and recombinational DNA repair are mediated in vivo (261). In this chapter, we will address the mechanistic mode of assembly and resolution of the RecA(tsDNA) intermediate, which have not been elucidated previously in molecular terms. Specifically, we are interested in the kinetics of the formation of this RecA(tsDNA) complex.

In the study described here, RecA-mediated DNA strand exchange was monitored in the presence of ATPγS using oligonucleotides (ODN) substrates containing a fluorescent base analog. The use of intrinsic DNA fluorescence in the context of short DNA molecules enables the observation of short-range interactions and local environmental change during the process. Based on a presteady-state kinetic analysis, we will present data suggesting that the RecA filament rapidly discriminates among the three DNA strands, induces a conformational change in the dsDNA, and tests for homology during the first detectable kinetic phase of strand exchange.

In the earliest attempts to define a kinetic scheme for strand exchange using ODN substrates, Camerini-Otero and coworkers reported a bimolecular association rate
constant for synapsis, $k_{r1} = 10^3 \text{ M}^{-1}\text{s}^{-1}$ (199). Subsequent experiments conducted using different assays under different solution conditions by the Radding laboratory reported that the bimolecular association rate constant was ca. $10^6 \text{ M}^{-1}\text{s}^{-1}$ (201). Similar apparent kinetics were reported by Shaner and coworkers (200); however, this study reported only observed relaxation times rather than true kinetic rate constants. The differences between the Camerini-Otero and Radding’s reports have not been reconciled. The latter authors made a more careful evaluation of the true kinetics by simulating the time-dependent concentration changes of the reaction species. In order to simplify the calculations, however, Bazemore et al. (1997) made the assumption that the observed signal changes were solely due to the concentration changes of the reactants and products, i.e., no contribution of the intermediates to the observed signal was considered. This assumption was demonstrated to be invalid in their own following-up experiments (160,186,187). Thus, a true kinetic scheme evaluated in terms of a molecular mechanism for the RecA-mediated strand exchange reaction is still lacking. Viewed from this perspective, the need for further chemical-kinetic investigation of RecA is clear.

**MATERIALS AND METHODS**

**Materials**

ATPγS was from Roche (Nutley, NJ). EDTA was from Amresco (Solon, OH).

NaCl, DTT, Tris, and Mg(OAc)$_2$ were from Fisher Scientific (Pittsburgh, PA).
RecA Protein

_E. coli_ RecA was purified as described (253) with minor modifications and stored in aqueous buffer (25 mM Tris-HCl, pH 7.5, 1 mM DTT, 5% glycerol) at −80 °C. Physical purity was estimated to be > 96% using Coomassie Blue stained polyacrylamide gels and quantitated using ImageQuant 3.22 imaging software. Protein concentration was determined spectrophotometrically using an extinction coefficient at 279 nm of 0.59 mg⁻¹ml⁻¹ (254). Purified RecA protein had a steady-state poly(dT)-dependent ATPase activity, measured at 37 °C using a spectrophotometric ATP hydrolysis assay (262,263), that was identical to that reported \( k_{\text{cat}} = 30 \pm 2 \text{ min}^{-1} \) (264).

Oligodeoxyribonucleotides

The sequences of the 30mer ODNs are derived from pGEM4Z (GenBank accession number X65305) coordinates 1685-1723 (185) and listed in Figure 3.1. ODNs with fluorescent labels were purchased commercially from Trilink Biotechnologies, Inc. Each 6MI-incorporated ODN was RP-HPLC purified. ODNs without fluorescent labels were purchased and PAGE purified from Integrated DNA Technologies, Inc. The concentrations of the ODNs were determined using their extinction coefficients at 260 nm calculated by the nearest neighbor method from nucleoside monomer and dimmers (255), and the contribution of absorbance of 6MI at 260 nm in the 6MI-incorporated ODNs were corrected for the extinction coefficient at 260 nm for 6MI \( (\text{4900 M}^{-1}\text{cm}^{-1}) \)
**Figure 3.1** Chemical structure of 6-methylisoxanthopterin (6MI) (A, shown in gray) and sequences of oligonucleotides (B) used in this study with 6MI indicated as bold g. For comparison reasons, a G:C base pair is also shown in panel A. In panel B, the base composition (BC) of each ODNs is indicated as Y (pyrimidine-rich), R (purine rich) or N (mixed bases). The superscript star indicates the ODNs is 6MI-incorporated.
(265). For example, the extinction coefficients at 260 nm for the nonfluorescent, d6MI-30mers were $2.47 \times 10^2 \text{ M}^{-1}\text{cm}^{-1}$ and $2.41 \times 10^5 \text{ M}^{-1}\text{cm}^{-1}$, respectively. All concentrations of ODNs were expressed in molecules.

Double-stranded DNA was prepared by mixing equal molar concentrations of the two complementary strands in 10 mM Tris·HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA. The mixture was heated at 80 °C for 4 min followed by a slow cooling back down to room temperature.

**Reaction conditions.**

A standard strand exchange reaction is performed as follows: 0.1 μM (in molecule) dye-labeled reactant (either the R(ssDNA) filament or the dsDNA) reacts with 0.5 μM non-labeled reactant (either the dsDNA or R(ssDNA) filament) in 25 mM Tris·OAc, pH 7.5, 6 mM Mg(OAc)$_2$, 1 mM DTT and 0.5 mM ATPγS at 37°C. The concentration of the non-labeled reactant and the reaction temperature may be varied as indicated in the text. For all reactions, the R(ssDNA) filament is preformed between the RecA protein and the ssDNA at a ratio of 2.5 nucleotides per RecA monomer in 25 mM Tris·OAc, pH 7.5, 2 mM Mg(OAc)$_2$ and 0.5 mM ATPγS for 15 mins prior to react with dsDNA.

**Agarose gel mobility shift assay**

Different reactions using dye-labeled DNA substrates as well as plain DNA substrates at the standard reaction conditions were incubated at 37 °C for 30 mins, then
equal amounts of each reaction were loaded on a 3% Metaphor agarose gel (FMC Bioproducts) in chilled 50 mM Tris-borate, pH 7.5, 2 mM EDTA, 2 mM Mg(OAc)$_2$.

The gel was run under rapid run conditions (manufacturer’s protocol) at 4 °C for two hours. The gel was stained with SybrGold (Molecular Probes) at room temperature for 20 min and viewed under an UV lamp.

**Steady-state fluorescence measurements**

All fluorescence data were collected on a SLM8100 spectrofluorometer (SLM-Aminco, Illinois) equipped with a thermostated cell holder attached with a circulated waterbath. A 3mm x 3mm cuvette (Nelson) unless otherwise indicated was used to minimize the reaction volumes. The bandpass was set at 4 nm for the excitation monochromator and 4 nm for the left emission monochromator. Scan rate was set at 0.95 nm/sec with 1 nm per step. Correction for the inner filter effect was not necessary since the absorbance values at the excitation wavelength were less than 0.05. The 6MI emission spectra were collected from 350 nm to 600 nm on the Emission left channel with excitation at 340 nm. The excitation and emission beams were polarized with Glan-Thompson polarizing prisms set at the parallel and perpendicular positions to reconstruct the total emissions $I_{tot}(\lambda)$ at each emission wavelength $\lambda$ according to the following equation (209):

$$I_{tot}(\lambda) = I_{VV}(\lambda) + 2G(\lambda) \cdot I_{VH}(\lambda)$$ (1)
where \( I \) is the fluorescence intensity and the first and second subscripts refer to vertical (\( V \)) polarization of the excitation and vertical (\( V \)) or horizontal (\( H \)) polarization of the emitted light, respectively. The factor \( G(\lambda) = I_{VH}(\lambda)/I_{HH}(\lambda) \) corrects for the different sensitivity of the emission monochromator for vertically and horizontally polarized light.

All spectral data were corrected for background by using an appropriate nonfluorescent control reaction and were fitted to a log normal distribution to extract the emission peak data (266):

\[
I(\lambda) = \frac{I_0}{\exp\left\{ \frac{\ln 2}{\ln 2 - \ln^2[p(\lambda - \lambda_{\text{max}})/(\rho^2 - 1)]]} \right\}}
\]

where \( I_0 \) is the intensity observed at the wavelength of maximum intensity \( \lambda_{\text{max}} \) and \( \Gamma \) is the full width of the spectrum at half-maximum intensity \( I_0/2 \). The asymmetry of the distribution is described by the parameter \( \rho \), determined by nonlinear least-squares fitting.

The total emission intensity value of each sample was also collected through a 420 nm longpass filter (Oriel 59480) inserted into the emission right channel with excitation at 340 nm in order to mimic the stopped-flow data collection mode. The excitation and emission beams were polarized with Glan-Thompson polarizing prisms set at the parallel and perpendicular positions to reconstruct the total emissions \( I_{\text{tot}} \) according to equation 1 except the emission wavelength term \( \lambda \) was ignored. It was demonstrated that the
relative total emission intensity measured by using the longpass filter did not exhibit significantly difference from the value generated by the log normal distribution function (equation 2).

For RecA-bound ssDNA complex, 1.2 μM RecA protein was added to 0.1 μM ssDNA sample containing 25 mM Tris·OAc, pH 7.5, 0.5 mM ATPγS, 1 mM DTT and 2 mM Mg(OAc)$_2$. The solution was allowed to stand at 37 °C for 15 mins before the spectrum was taken. It was demonstrated that at these concentration of RecA protein, all the DNA molecules in the solution was bound by the RecA protein, and further addition of RecA protein did not change the spectra. For RecA-bound dsDNA complex, 24 μM RecA protein was used to insure a complete binding of dsDNA molecule. For the strand exchange reaction, 0.1 μM ssDNA was first incubated with 1.2 μM RecA protein in reaction buffer (25 mM Tris·OAc, pH 7.5, 1 mM DTT, 2 mM Mg(OAc)$_2$ and 0.5 mM ATPγS) at 37 °C for 15 min, then either 0.5 μM homologous dsDNA or 2.0 μM heterologous dsDNA was added to the solution. Aliquots of 500 mM Mg(OAc)$_2$ stock were also added together with the dsDNA to bring the concentration of Mg$^{2+}$ to a final concentration of 6 mM. This concentration of Mg$^{2+}$ (6 mM) was chosen to balance both the effects of the precipitation of RecA protein at high Mg$^{2+}$ concentration and the ability of Mg$^{2+}$ to allow the strand exchange reaction to complete (data not shown). The mixture was incubated at 37 °C for 15 mins before the emission spectra of 6MI were taken.
Steady-state fluorescence anisotropy measurements

All steady-state fluorescence anisotropy measurements were performed in R format using Glan-Thompson polarizers placed in the excitation and emission channels of the SLM 8100 spectrofluorometer. The excitation wavelength was set at 340 nm and a 420 nm long pass filter (Oriel 59480) was inserted into the right emission channel to mimic the data collection mode of stopped-flow spectrofluorimeter. The corrected fluorescence anisotropy of the sample was reconstructed after background subtraction:

\[
    r = \frac{(I_{\text{vv}}^F - I_{\text{vv}}^N) - G \cdot (I_{\text{vv}}^F - I_{\text{nn}}^F)}{(I_{\text{vv}}^F - I_{\text{nn}}^F) + 2(I_{\text{nn}}^F - I_{\text{nn}}^N)} \tag{3}
\]

where the superscript indicate if the value was measured on a fluorescent sample (F) or a corresponding nonfluorescent sample (N). The factor \(G = I_{\text{nv}}^F / I_{\text{nn}}^F\) corrects for the different sensitivity of the emission monochromator for vertically and horizontally polarized light. For each sample, the anisotropy was measured 10 times with 1 s measuring time each and the values were averaged.

RecA-DNA complex dissociation constant measurements

The RecA(S1) presynaptic filament was first formed between 0.1 µM fluorescent S1 and 1.2 µM RecA protein in 400 µL reaction buffer in a 2-mm x 10-mm cuvette oriented with the 2-mm path length parallel to the excitation beam to minimize inner filter effects. The mixture was incubated at 37 °C for 15 min, then small aliquot of
concentrated Mg(OAc)$_2$ of 500 mM together with dsDNA D0 of 10 μM were added. The final concentration of Mg(OAc)$_2$ is 6 mM. The final concentration of D0 was varied from 0.05 μM to 3 μM. Between each addition of titrant, the reactions were allowed to equilibrate for a time period of 5 mins before the total emission and anisotropy of the reaction were recorded through the emission right channel with a 420 nm longpass filter. The recorded total emission intensity and anisotropy were corrected for background (measured on an appropriate nonfluorescent reaction RecA(S0) + D0) and dilution. Then a binding isotherm between the RecA(S1) filament and D0 was constructed. The apparent dissociation constant $K_d$ was determined by fitting the experimental data to the theoretical curve using the following equation derived from the mass balance equations for the rapid, one-step reversible association of molecules A and B:

$$I_t = I_0 + \frac{\Delta I}{2B_0} \cdot [A + B_0 + K_d - \sqrt{(A + B_0 + K_d)^2 - 4A \cdot B_0}]$$

(4)

In this equation, $I_t$ is the total fluorescence intensity; $I_0$ is the initial fluorescence when no titrant has been added; $\Delta I$ is the total fluorescence change defined by the difference between the end fluorescence and initial fluorescence; $A$ is the concentration of the dsDNA D0; $B_0$ is the concentration of the RecA(S1) filament being titrated.

The dependence of $\chi^2$ on the apparent dissociation constant $K_d$ was calculated using the following equation:
\[ \chi^2 = \sum_{i=1}^{n} \frac{(I_i(K_0) - I_i)^2}{\sigma^2(I_i)} \]  

where the denominator is the square of the difference between the fitted fluorescence intensity at a given \( K_0 \) value \( I_i(K_0) \) and the experimental measured fluorescence intensity \( I_i \). The numerator is the square of the standard deviation of the experimentally measured fluorescence intensity \( I_i \).

**Stopped-flow spectrofluorometry**

A SX.18MV stopped-flow spectrofluorometer (dead time \( \leq 2 \) msec, Applied Photophysics Ltd, Leatherhead, UK) in fluorescence anisotropy mode to was used to monitor the time-dependent fluorescence emission upon mixing RecA protein with ODN. The signal from both the parallel and perpendicular channels through a 420 nm long pass filter (Corion) with excitation at 340 nm were recorded. A logarithmic time scale was used in order to view the early events. For a typical reaction I with RecA(S1) filament and dsDNA D0 at the final concentrations of 0.1 \( \mu \text{M} \) and 0.5 \( \mu \text{M} \), respectively, the reaction condition is as follows: the presynaptic filament was preformed between 2.4 \( \mu \text{M} \) RecA protein and 0.2 \( \mu \text{M} \) S1 in 25 mM Tris-OAc, pH 7.5, 1 mM DTT, 2 mM Mg(OAc)\(_2\) and 0.5 mM ATP\(_7\)S in the drive syringe at 37 °C for 15 mins prior to mixing. The other drive syringe is filled with 1.0 \( \mu \text{M} \) D0 in 25 mM Tris-OAc, pH 7.5, 1 mM DTT, 10 mM Mg(OAc)\(_2\) and 0.5 mM ATP\(_7\)S and incubated at 37 °C for 5 mins before mixing. After mixing, the final concentrations of the RecA(I), dsDNA and Mg(OAc)\(_2\) are at 0.1 \( \mu \text{M} \), 0.5
μM and 6 mM, respectively. For reaction C and O, the RecA(ssDNA) filament were kept at 0.5 μM or higher (as indicated), while the concentration of fluorescent dsDNA (D1 or D2) was kept at 0.1 μM after mixing. All the solutions were prepared using degassed and filtered buffer and water.

**Stopped-flow data analysis**

For reaction C and O, the background signal (obtained from an identical reaction with no fluorophore present) was subtracted from the fluorescent reaction in order to minimize the light-scattering background brought by the high concentration of RecA protein. For reaction I, it was demonstrated that the background correction is not necessary because of the less concentrated nucleoprotein filament. Only the signal collected through the perpendicular channel was analyzed and reported in this study. Non-linear least-square fitting analysis of the kinetic traces was accomplished using KaleidaGraph (Synergy Software, PA). Each reaction trace was fitted with a single or multiple exponential decay function where is suitable:

\[ F_i = F_e - \sum_{\Delta F_i}^{\tau_i} \exp\left( \frac{-t}{\tau_i} \right) \]  

(6)

where \( F_i \) is the observed fluorescence at time \( t \), \( F_e \) is the fluorescence of the reaction at the end of the reaction, \( \Delta F_i \) is the amplitude change of phase \( i \), and \( \tau_i \) is the time constant for phase \( i \). For each reaction, up to eight runs were recorded and averaged. The 95% confidence interval of each averaged trace was calculated and used to weight the data in
the exponential curve fitting. We also excluded some data points according to the following criterion: First, the data points recorded before 2 ms (machines' deadtime for mixing) was excluded; Second, we compared each data point with its moving average (the average of adjacent six data points, three earlier, three later). If the difference is greater than its corresponding confidence interval, the data point will be excluded. This treatment results in about 1% data exclusion of 1000 data points, and rarely results in exclusion for data points collected later than 100 ms. The decision of which exponential was appropriate was made by analyzing the goodness of fitting residuals and the effort of trying to fit data with the lowest order exponential.

**Kinetic Simulation**

Non-linear least-square regression analysis of the kinetic traces under a given mechanism was accomplished using Dyanfit (BioKin, Ltd.) (267).

**Thermodynamic analysis**

The Arrhenius plot was constructed by using the following equation:

\[
\ln k = \ln A - \frac{E_a}{RT}
\]

(7)

where \( k \) is the rate constant for the reaction step to be analyzed, \( \ln A \) is a constant of integration, \( E_a \) is the activation energy and corresponds to the standard enthalpy of reaction \( \Delta H^\circ \) in the van't Hoff equation. R is the gas constant with a constant value of 8.314 J·mol\(^{-1}\)·K\(^{-1}\), and \( T \) is temperature in Kelvin.
The van’t Hoff plot was constructed using the following equation:

$$\ln K = \frac{\Delta S^0}{R} - \frac{\Delta H^0}{RT}$$  \hspace{1cm} (8)

where $K$ is the equilibrium constant for the reaction step to be analyzed. $\Delta S^0$ and $\Delta H^0$ are the standard entropy and enthalpy changes of the reaction step, respectively.

The free energy change $\Delta G$ was calculated using the either the equilibrium constant $K$ or the enthalpy change $\Delta H^0$ and entropy change $\Delta S^0$:

$$\Delta G^\circ = -RT \ln K = \Delta H^0 - T \Delta S^0$$  \hspace{1cm} (9)

The activation free energy $\Delta G^\ddagger$ was calculated using the Eyring equation:

$$\Delta G^\ddagger = -RT \ln \frac{kh}{k_BT}$$  \hspace{1cm} (10)

where $k$ is the rate constant of the reaction step; $h$ is the Planck’s constant with a value of $6.626 \times 10^{-34}$ J·s; $k_B$ is the Boltzmann constant with a value of $1.38 \times 10^{-23}$ J·K$^{-1}$.

**RESULTS**

**Experimental design**

In order to characterize the kinetics of the synapsis and heteroduplex formation phases of RecA-mediated strand exchange, we employed a number of changes from the traditional methods of analysis. First, we limited the events occurring during reaction to as few as possible. We achieved this by pre-forming the RecA(I) nucleoprotein filament, thus eliminating the presynapsis phase as a complicating factor. In addition, we
employed ATPγS, a slowly hydrolyzed analog of ATP, to eliminate the influence of ATP turnover on the events we monitored. As a further simplification, we employed synthetic oligodeoxyribonucleotides (ODNs) in place of long, bacteriophage-derived DNA molecules. The ODN sequences used here were derived from those capable of exhibiting RecA-DNA pairing activity in the presence of ATPγS (185). Moreover, we have recently demonstrated that ODNs from this sequence family as short as 30 nts were competent for activating the ATPase activity of the RecA filament. The use of relatively short DNAs (30mers) avoids the topological constraints of long DNAs. In addition, the use of ODNs emphasizes the influence of short-range RecA monomer-DNA interactions over long range effects such as cooperatively. In addition, standard chemical synthesis of ODNs allows the directed placement of intrinsic DNA fluorophores at unique sites in the DNA molecules, and the use of short ODNs maximized the relative amount of fluorophore in a non-fluorescent ODN background, providing for optimum signal-to-noise. A final alteration was the use of polarized emission to reduce the possible influence of polarization artifacts introduced by changes in the rotational lifetime of the fluorophore. For each fluorescent reaction, the signal is recorded simultaneously in two emission channels with each polarizer orientated either parallel or perpendicular to the polarizer inserted in the excitation pathway.
We designed three different but directly related homologous reaction systems (Figure 3.2) to facilitate characterization of the RecA-mediated strand exchange reaction in a strand-specific manner. Each system is comprised of a 30-nt incoming strand (I strand) and 30-bp duplex (C·O), comprised of strands complementary (C) and identical (O, for outgoing) to that of I. The reaction systems differ only in the strand location of the 6MI fluorophore and each system is named after the 6MI-incorporated strand. The simultaneous analysis of data from the three complementary systems allows us to analyze time-dependent changes in each of the three DNA strands involved in RecA-mediated exchange. We also designed two heterologous reaction systems (Figure 3.2) in which the strand labeled with 6MI is either in the RecA(I) filament or in the dsDNA substrate.

**Characteristics of DNA Strand Exchange with 6MI-labeled ODNs**

We first examined the possibility that 30-nt ODNs containing fluorescent labeling might impede the binding of RecA or strand exchange. Three homologous dye-labeled reactions (Reaction I: R(S1)+ D0; Reaction C: R(S2) + D1; and Reaction O: R(S0) + D1) and their corresponding non-labeled reactions were performed under standard reaction conditions and were analyzed by using agarose gel mobility shift assay (Figure 3.3). We found that the RecA-bound reactants showed equivalent amounts of involvement of 6MI-containing and nonfluorescent substrates into the reaction complexes. These results demonstrate that the ODNs are competent for DNA strand
Figure 3.2 Cartoon drawings illustrating the experimental designs for homologous DNA strand exchange reactions and heterologous reactions. The letters \( R_{xn} \) in the three homologous reactions indicated either the incoming (blue), the complementary (red) or the outgoing (green) strand is labeled with 6MI, respectively. The fluorophore 6MI is indicated as a yellow star in each strand. In heterologous reaction Het1, the incoming strand is 6MI-labeled while in reaction Het2, the dsDNA is 6MI-labeled. The gray ovals represent the RecA protein.
Figure 3.3 Different strand exchange reactions using 6MI-incorporated DNA substrates D1 or S1 series (lane 2: D1; lane 3: Reaction O; lane 9: Reaction C; lane 13: Reaction Het1; lane 16: Reaction I) as well as non-labeled DNA substrates (lane 1: D0; lane 4: Reaction O; lane 5: RecA-bound S0; lane 6: S0; lane 7: S2; lane 8: RecA-bound S2; lane 10: Reaction C; lane 11: heterologous ssDNA h45 (see Figure 3.1 for sequence); lane 12: RecA-bound h45; lane 14: Reaction Het1; lane 17: Reaction I). The Reaction C, O and I, either using nonfluorescent or fluorescent substrates, are incubated at 37 °C for 20 mins under standard reaction conditions (0.1 μM dsDNA substrate, 0.5 μM RecA-ssDNA filament in 25 mM Tris·OAc, pH 7.5; 5% glycerol, 6 mM Mg(OAc)2, 0.5 mM ATPγS, 1 mM DTT), then equal amount of each reaction were fractionated by electrophoresis through Metaphor agarose gel at Rapid Running condition.
exchange and that the presence of 6MI has no impact on the extent of reaction. In addition, no release of DNA strand was observed. This suggests that the products of strand exchange, the I-C heteroduplex DNA and the O strand, remain bound to the RecA filament under these conditions. Most importantly, when one I strand characterized with no sequence homology with the dsDNA was employed as the incoming strand (Reaction Het2, R(h2) + D1), a clear separation between the bands corresponding to R(I) filament and the band to dsDNA was observed. This observation contrasts with those described for the homologous reactions. Thus, even the 30-nt substrates require sequence homology for efficient strand exchange.

**Spectral Characteristics of the Reaction Systems**

In order to relate time-dependent changes in 6MI fluorescence with structural changes during the reaction, the spectral characteristics of the substrates and products were needed. We collected steady-state fluorescence emission spectra as well as polarized emission data (fluorescence anisotropy) for RecA-free DNA molecules, RecA-bound ssDNA and dsDNA, and the RecA-bound triple-stranded DNA complexes (tsDNA) containing alternately 6MI-labeled strands. In Figure 3.4A and B, one set of typical spectra collected using the ODNs related to S1 was shown, and the spectra collected using other three sets of ODNs with different sequence identities were also shown in Figure 3.4 from C to H. Table 3.1 shows the total emission intensity
Figure 3.4 Steady-state spectra of the 6MI-labeled DNA and the RecA(DNA) complexes in homologous and heterologous reactions. Panel A and B are the spectra of the reactions using S1 and D1 series; Panel C and D are the reactions using S3 and D2 series; Panel E and F are the reactions using S5 and D4 series; Panel G and H are the reactions using S7 and D5 series. The left panels are the reactions with the ssDNA labeled: open orange triangles, ssDNA; filled pink diamonds, RecA(ssDNA); filled blue squares, Reaction I (RecA(S1)+D0 for A, RecA((S3)+D0 for C, RecA(S5)+D3 for E and RecA(S7)+D3 for G); Black crosses, Reaction Het1 (RecA(S1)+H2 for A, RecA(S3)+H2 for C, RecA(S5)+D0 for E, and RecA(S7)+D0 for G). Right panels are the reaction with the dsDNA labeled: open orange triangles, dsDNA; filled pink diamonds, RecA(dsDNA); open red circles, Reaction C (RecA(S2)+D1 for B, RecA(S0)+D2 for D, RecA(S4)+D4 for F and RecA(S6)+D5 for H); open green squares, Reaction O (RecA(S0)+D1 for B, RecA(S2)+D2 for D, RecA(S4)+D4 and RecA(S6)+D5
for H); black crosses, Reaction Het2 (RecA(h2)+D1 for B, RecA(h2)+D2 for D, RecA(S4)+D4 for F and RecA(S4)+D5 for H). The data for A, B, C and D were total intensity data and panel E, F, C and G were collected through the perpendicular channel. All the data were corrected for background (measured on an appropriate non-fluorescent sample). For clarity reasons, only 20% of the data points are shown. The solid curves connecting the data points are the fitted spectrum according to the log normal distribution function as described in Materials and Methods (equation 2).
Table 3.1. Characteristic combination of emission maximum ($\lambda_{\text{max}}$), relative fluorescence intensity ($F_{\text{rel}}$) and anisotropy ($r$) of each DNA and RecA-DNA complexes with 6MI labeled in a DNA strand with different base compositions

<table>
<thead>
<tr>
<th>Complex</th>
<th>Identity of 6MI-labeled DNA Strand</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1 $a$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{\text{max}}$ (nm)</td>
</tr>
<tr>
<td>ssDNA</td>
<td>424</td>
</tr>
<tr>
<td>R(ssDNA)</td>
<td>414</td>
</tr>
<tr>
<td>dsDNA</td>
<td>421</td>
</tr>
<tr>
<td>R(dsDNA)</td>
<td>420</td>
</tr>
<tr>
<td>Reaction I</td>
<td>415</td>
</tr>
<tr>
<td>Reaction C</td>
<td>423</td>
</tr>
<tr>
<td>Reaction O</td>
<td>415</td>
</tr>
<tr>
<td>Heter 1 $c$</td>
<td>414</td>
</tr>
<tr>
<td>Heter 2</td>
<td>420</td>
</tr>
</tbody>
</table>

$^a$ The values reported here were measured on four 6MI-labeled DNA strands and their corresponding complexes. S1 is composed of two thirds of pyrimidine bases while S2 is composed of two thirds of purine bases. S5 and S7 have equal base compositions (see Figure 1 for sequences). $^b$ Emission maximum $\lambda_{\text{max}}$ was obtained from the fitted emission spectrum of each complex using the log normal distribution function (equation 3). In all cases, the experimental uncertainty on $\lambda_{\text{max}}$ is less than 1 nm. $^c$ Relative fluorescence were obtained by normalizing the measured fluorescence of each complex to that of the corresponding duplex DNA. $^d$ The experimental uncertainties on the all anisotropy values ($r$) are plus or minus 0.02. $^e$ Reaction Heter 1 is the reaction in which the 6MI-labeled nucleoprotein filament R(S) pairs with a heterologous duplex DNA H1; Reaction Heter 2 is the reaction in which a heterologous nucleoprotein filament R(h1) pairs with the 6MI-labeled duplex.
(characterized by $I_{\text{tot}}$), emission maximum ($\lambda_{\text{max}}$), and anisotropy ($r$) for the four sets of 6MI-labeled ODNs.

As shown in Figure 3.4A, hybridization of a 6MI-containing 30mer S1 (orange triangle, Figure 3.4A) with its complement to form a 6MI-labeled dsDNA D1 (orange triangle, Figure 3.4B) results in a 90% quench of fluorescence. When RecA protein was added to the ssDNA S1, a two-fold fluorescence enhancement was observed (Figure 3.4A, pink diamonds). A similar fluorescence enhancement was observed upon the addition of the RecA protein to dsDNA D1 (Figure 3.4B, pink diamonds). The sensitivity of 6MI fluorescence to base pairing and base stacking interactions has been noted earlier (268). The increase in fluorescence intensity is consistent with the well-documented DNA base destacking that occurs in the extended RecA(DNA) filament (244,249,269)

The total emission intensity of the RecA(tsDNA) complexes showed a strong dependence on homology as well as the strand location of the 6MI fluorophore. Upon addition of a homologous dsDNA D0 to the 6MI-labeled R(S1) filament (Figure 3.4A, pink diamonds), a 75% fluorescence quench was observed (Reaction I, Figure 3.4A, blue squares), while the addition of a heterologous dsDNA H2 results in only a 25% quench (Reaction Het1, Figure 3.4A, black crosses). Similarly, when a homologous RecA(S0) filament was added to a fluorescent dsDNA substrate D1 (orange triangles in Figure 3.4B), a 13-fold fluorescence enhancement was observed from the dsDNA substrate
(Reaction O, Figure 3.4 B, green squares). When another homologous RecA(S2) filament was added to the same dsDNA D1, a somewhat smaller fluorescence enhancement was observed (Reaction C, red circles, Figure 3.4B). In contrast to the large fluorescence enhancement observed in the homologous reactions, the total fluorescence increased by less than 50% when a heterologous RecA(h2) filament was added to dsDNA D1 (Reaction Het2, Figure 3.4B, black crosses). Interestingly, the $I_{\text{tot}}$ values determined for the two homologous reactions I and C are similar to one another and substantially smaller than that for the reaction O. Similar patterns of total intensity changes was observed when other 6MI-labeled strands with different sequence identities was used (Table 3.1).

In addition to the dramatic emission intensity changes described above, differences in the wavelength-dependent emission profiles (characterized by $\lambda_{\text{max}}$) were apparent (Figure 3.4). A spectral shift of about 10 nm was observed when the RecA protein binds to the ssDNA S1 (Figure 3.4A, compare orange triangles with pink diamonds), while binding of the RecA protein to the dsDNA D1 did not result a significant spectral shift (less than 1 nm, Figure 3.4B, compare orange triangles with pink diamonds). However, the hybridization of a complementary strand to S1 causes a blue shift of about 3 nm (compare orange triangles in Figure 3.4A and B).

One general pattern we observed in the steady-state is the spectral separation of the C and O strands. As shown in Figure 3.4, the emission peak of O strand (green
squares in Figure 3.4B) is blue-shifted about 8 nm from the C strand, while that of the C (strand red circles in Figure 3.4B) is close to the region of RecA-free ssDNA or dsDNA. A similar pattern was observed in the reactions using other 6MI-labeled 30mers (S3, S5 and S7, see Figure 3.4 and Table 3.1). Considering that before mixing with RecA-ssDNA, the originally base-paired C and O strands display indistinguishable fluorescence properties, the apparent difference of the two strands after mixing clearly demonstrated that the RecA (I) filament distinguishes the two strands according to their sequences in relative to the I strand.

We also found in the majority of cases that the spectral maximum of Reaction O and I are all blue-shifted towards the region of RecA-bound ssDNA (Table 3.1). The coincidence of the emission peak positions of Reaction O and RecA(I) can be explained by the well documented fact that the strand O behaves more like a ssDNA remaining bound to the RecA filament after strand exchange (31,117). Although the I strand should be base-paired with the C strand after the strand exchange reaction, which implies that the local environment for the two strands should be similar, we noticed that the emission peak positions of the two strands are quite different.

A set of interesting observations related to the apparent discrepancy described above is that the RecA-free ssDNA and dsDNA have very similar emission peak positions (see Table 3.1: averaged values: 424 ± 1 nm for ssDNA and 422 ± 1 nm for dsDNA), regardless of the sequence and/or base composition of the 6MI-containing
ODNs. However, when RecA protein is involved, the emission peak of the same complex formed on the four different 6MI-containing ODNs are quite different from each other. Only one exception was observed, namely that reaction C has a constant emission maximum near 423 nm regardless of the ODNs’ sequence.

It was previously reported that the fluorescence intensity of 6MI is mainly affected by base stacking while protein-binding or base-pairing mainly affects the spectral maximum (268). Thus, the ODNs’ identity-dependent emission peak positions in the context of the RecA complexes may result from the combination effect of protein contacts (different site binding and/or different affinity) and base contacts. Based on the resemblance of the less variable and similar emission peak positions of C strand to ssDNA and dsDNA and the spread of the emission peak positions of Reaction I, we believe that the difference in the emission peak positions between strand C and I may provide evidence for extra protein contact by the I strand. Moreover, among the three strands in the final RecA(tsDNA) complex, the C strand may have least contact with the protein judging by its red shift to the region of the RecA-free ssDNA or dsDNA. When the dsDNA is heterologous to the RecA(I) filament, no apparent spectral shifts of the 6MI-containing ODNs were observed (Figure 3.4, black crosses).

The formation of the various complexes could also be characterized by measuring the anisotropy of the fluorescence as the mobility of the 6MI base becomes restricted (Table 3.1). In contrast to the characteristic $I_{tot}$ and $\lambda_{max}$ parameters for each complex or
reaction system, the fluorescence anisotropy values fell within a more selected range. The fluorescence anisotropy of 6MI in ssDNA increased upon hybridization with its complementary strand. In the case of 6MI-labeled DNA bound to RecA, the fluorescence anisotropy increased still further to a value \((r \geq 0.3)\) that was independent of whether ssDNA, dsDNA, or tsDNA was bound to RecA. Interestingly, though the heterologous reactions did not show a characteristic fluorescence intensity nor spectral shift as the homologous reactions, increased anisotropy of the heterologous reactions indicate a sequence nonspecific RecA(tsDNA) complex may be present.

**Apparent dissociation constant measurement**

An apparent equilibrium dissociation constant for the RecA-tsDNA complex \((K_d)\) was measured using Reaction I. A tight-binding isotherm revealed a small dissociation constant, \(K_d = 0.005 \pm 0.003 \mu M\) (Figure 3.5). This indicates that at our standard reaction condition (0.1 \(\mu M\) fluorescent reactant over 0.5 \(\mu M\) nonfluorescent reactant), 99% of the fluorescent reactant is converted into the final product, namely, the RecA(tsDNA) complex. We also noticed that there is a large uncertainty associated with the apparent \(K_d\). The large uncertainty of the measured \(K_d\) could be reduced if the concentration of the RecA(S1) filament was comparable to the measured \(K_d\). However, at such low concentration the fluorescence intensity of the reaction became too weak to be monitored accurately. In order to have a better estimation of the upper limit of the apparent \(K_d\), we
Figure 3.5 Apparent dissociation constant ($K_d$) measurement on reaction I. In this measurement, the RecA(S1) filament was preformed between 0.1 μM S1 and 1.2 μM RecA protein at 37°C for 15 mins in 25 mMTris-OAc, pH 7.5, 5% Glycerol, 0.5 mM ATPγS, 2 mM Mg(OAc)$_2$ and 1 mM DTT. Aliquot of concentrated stock solution of dsDNA D0 was added and the total fluorescence through a 420 nm longpass filter was recorded at each titration point. Concentrated stock solution of Mg(OAc)$_2$ was also added along the first addition of dsDNA to raise the final concentration of Mg$^{2+}$ to 6 mM. Each data point was corrected for background measured on an identical non fluorescent reaction and dilution. The solid curves connecting the data points is the fitted curve according to equation (4). The insert shows the dependence of $\chi^2$ on the measured apparent $K_d$ values according to equation (5). It is obvious that the $\chi^2$ increases shapely when the measured apparent $K_d$ exceeds 0.01 μM while remained almost flat at lower $K_d$. 
analyzed the dependence of $\chi^2$ on $K_d$ (Figure 3.5, insert) and an upper limit of $K_d$ at about 0.01 $\mu$M (98% product conversion at the standard reaction condition) was found. This estimated upper limit was used in the following studies.

**Time-dependent fluorescence change of the reaction systems**

The fluorescence change of 6MI-incorporated strand exchange reactions (Reaction I, C and O) over time was monitored upon rapid mixing of appropriate reactants. One typical trace of each reaction is plotted in Figure 3.6. In Reaction I, a rapid and large fluorescence quench of 6MI was observed during the reaction time course (Figure 3.6A). When a RecA·S0 filament was mixed with a 6MI-labeled dsDNA D1, depending on the final location of the 6MI-labeled strand (O or C strand), different extents of fluorescence enhancement for each strand’s label were observed (Figure 3.6D and G). In general, the total relative amplitude changes for each reaction agree well with what was observed in steady-state fluorescence measurements.

We wanted to determine whether the kinetics observed above were generally characteristic of RecA-mediated strand exchange between 30mer DNAs. Three 6MI-labeled ODNs with different sequences were used (S1, S3 and S5). The same stopped-flow experiments under identical conditions were conducted. The kinetic traces collected using these three ODNs exhibit similar relaxation times and relative amplitudes changes, especially the characteristic mutual exclusion of the first and second phases in Reaction
Figure 3.6 Three typical stopped-flow traces of Reaction I (A), Reaction C (D, RecA(s0)+D1) and Reaction O (G) under standard reaction conditions (0.1 μM fluorescent reactant mixing with 0.5 μM nonfluorescent reactant in 25 mMTris-OAc, pH 7.5, 5% Glycerol, 0.5 mM ATPγS, 6 mM Mg(OAc)₂ and 1 mM DTT ). The data were collected and plotted in logarithm scale in order to view the early events. For each trace, at least six individual runs were collected and averaged. The data were treated for outliner exclusions as described in Materials and Methods and fitted with equation (6) by using the 95% confidence interval as weight. The best fitted curve for each reaction was shown as the black solid curve plotted on the top of scattered data points. Panel B, C, E, F, H and I are the normalized residuals resulting from different orders of exponential fittings: E and H are from single exponential fittings; B, F and I are from double-exponential fitting; and C is from three-exponential fittings. All the plotted residuals are normalized by dividing the direct residuals resulting from the exponential fittings by its corresponding 95% confidence interval in order to amplify the residuals at longer time scales. It is apparent that for Reaction I, a three is better than a double-exponential function while for Reaction C and O, a double is better than a single-exponential function. Further increase of exponential function order did not result in significantly improved fitting goodness.
Figure 3.6
C and O. This demonstrates that the kinetics we observed for the S1 series provide general rather than sequence-specific information about the RecA-mediated strand exchange reaction under our reaction conditions. Based on this result, we only used the data from the S1 series in the subsequent analyses, and did not distinguish the kinetics observed on the two Reaction Cs (RecA(S0) + D2 and RecA(S2) + D1).

Because all the reactions were performed under pseudo-first order reaction conditions (one reactant in large molar excess over the other), we used exponential functions as described in equation 6 to analyze the real-time fluorescence data. For Reaction I, we found that the residuals of a double-exponential function deviated from zero in a non-random manner, especially in the range from 10 s to 100 s (Figure 3.6B). However, a three-exponential function resulted in significantly lower residuals (Figure 3.6C). When a four-exponential function was used, no significant improvement of the fitting goodness was observed (data not shown). Therefore, a three-exponential function was selected for the initial description of the time-course observed in Reaction I. The same analysis is performed on Reaction C and O, and a double-exponential function was found to adequately account for the signal changes observed in both cases. Taken together, these multiexponential fitting results suggest that at least three phases are required to describe the RecA-mediated strand exchange reaction (Table 3.2, first row of each reaction). The fastest phase was observed in Reaction I and O but not in Reaction
Table 3.2: Exponential curve fit parameters for strand exchange reactions with increasing substrate concentrations

<table>
<thead>
<tr>
<th></th>
<th>Curve Fit Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[S] µM</td>
</tr>
<tr>
<td>Reaction I (^a)</td>
<td>R(S1) + D0</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction C (^b)</td>
<td>R(S0) + D2</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction O (^c)</td>
<td>R(S0) + D1</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Values reported are the parameters that describe the 3-exponential curve fit of the raw data generated by KaleidaGraph. Amplitude changes (\( \Delta F \)) are the true values reported by KaleidaGraph for the curve fits. \(^b\) In reaction I, the duplex DNA concentration is varied as indicated in the [S] column while the concentration of the preformed R(S1) filament is maintained at 0.1 µM; In reaction C and O, the concentration of the preformed nucleoprotein filament R(S0) is varied as indicated while the duplex DNA D1 or D2 is maintained at 0.1 µM. All concentrations are expressed in molecules. \(^c\) Figure 1 and 2 shows the components of the reactions indicated here.
C. In contrast, the second phase was present in Reaction I and C, but was absent in Reaction O. The third phase, however, was observed for all three reactions.

**Concentration-dependence of strand exchange reaction**

To investigate the concentration-dependence of the observed relaxation times and amplitudes, we varied the concentration of the non-fluorescent species over a fixed concentration of the fluorescent species in three homologous reactions (Reaction I, C and O). The resulting relaxation times ($\tau$) and fluorescent intensity changes ($\Delta F$) obtained at different substrate concentrations by the exponential curve fitting described above were tabulated in Table 3.2. A typical set of the concentration dependence time courses of Reaction I is shown in Figure 3.7A.

It is apparent that the total amplitude change ($\Delta F^{\text{tot}}$) for all the three reactions did not change significantly over the concentration range used (Figure 3.8A), indicating that even at the lowest substrate concentration (0.5 $\mu$M substrate over 0.1 $\mu$M fluorescent substrate) the reaction essentially goes to completion. This is consistent with the steady-state apparent $K_a$ measurement.

In contrast to the concentration-independent total amplitude changes, the relaxation times of all the three reactions are dependent on the substrates’ concentration (Figure 3.8B, C and D). The reciprocal relaxation times for the first phases of Reaction I and O exhibit nearly identical linear dependencies on the substrate’s concentration.
Figure 3.7 Overview of the concentration-dependence of Reaction I. The concentration of the RecA(S1) filament was maintained constant at 0.1 μM, while the concentration of dsDNA D0 was at 0.5 to 0.7, 1.0, 1.5 and 2.5 μM from top to bottom respectively. For clarity reasons, the actual data points are omitted and only the fitted curves are shown in panel A. In order to view the clear concentration-dependent kinetics, the total amplitude changes of the five traces in panel A were normalized to be the same. Panel B shows the original data points overlapped with the simulated reaction progress curves (black solid curves) using a sequential four steps reaction scheme as shown in Figure 3.12. For clarity, each data is offsetted 0.02 signal level to separate from each others.
Figure 3.8 Concentration-dependence of the total amplitude changes (A) and the reciprocal of relaxation times (B, C and D) of Reaction I (filled squares), C (open circles) and O (open squares). The concentrations used were as following: in Reaction I, the RecA(S1) filament was at 0.1 μM, while the dsDNA D0 was at 0.5, 0.7, 1.0, 1.5, and 2.5 μM. In Reaction C, the dsDNA D2 was at 0.1 μM and the RecA(S0) was at 0.5, 0.7, 1.0, 1.5 and 2.0 μM. In Reaction O, the dsDNA D0 was at 0.1 μM while the RecA(S0) filament was at 0.5, 0.7, 1.0 and 1.5 μM. As shown in A, the total amplitudes changes for all three reactions are not dependent on the concentration range used. In the contrast, the relaxation times of the three reactions exhibit strong concentration-dependence. In B, the reciprocal of the relaxation times of the first phases of Reaction I and O exhibit similar linear concentration dependence; In C, the reciprocal of the relaxation times of the second phase of the Reaction I and first phase of Reaction C are plotted together because of the similar concentration-dependent behaviors; In D the reciprocal of the relaxation times of the third phase of reaction I and the second phases of Reaction C and O are plotted.
Figure 3.8
(Figure 3.8B). A pseudo-first order concentration-dependence analysis revealed bimolecular association rate constants for Reaction $I$ and $O$ of $6.5 \pm 0.5 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$ and $5.9 \pm 0.6 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$, respectively, and unimolecular dissociation rate constants of $1.2 \pm 0.5$ and $0.7 \pm 0.4 \text{ s}^{-1}$, respectively (Figure 3.8B). This implies that the first phase in Reaction $I$ and $O$ likely represents the same initial bimolecular association step. When the reciprocal of the relaxation times of the later phases of the three reactions were analyzed, we found that the first phase of Reaction $C$ exhibit similar concentration-dependent behavior as the second phase of Reaction $I$ (Figure 3.8C), while the concentration-dependence of the second phase of Reaction $C$ is more similar to that of Reaction $O$ than to the third phase of Reaction $I$ (Figure 3.8D). The agreement and disagreement among the relaxation times of the three reactions indicate that though we only observed three apparent phases for the RecA-mediated strand exchange reaction, there may be as many as four discrete reaction steps involved.

**Heterologous contact between RecA(I) filament and dsDNA substrate**

To probe the nonsequence-specific interaction between the RecA(ssDNA) filament with its dsDNA substrate, which is a prerequisite for the ensuing homology alignment and pairing steps, we investigated the impact of increasing heterology on the kinetics of the strand exchange reaction. Figure 3.9A shows the time course collected for a homologous reaction (Reaction $I$: RecA(S1) + D0, solid curve), a partially homologous
Figure 3.9 Overview of concentration dependence of heterologous reactions. In A, Reaction I" (RecA(S1)+H0) is shown as open diamonds, Reaction Het1 (RecA(S1)+H1) as filled triangles and reaction Het1 (RecA(s1)+H2) as open triangles. For comparison reasons, the signal collected by mixing the RecA(S1) filament with reaction buffer (gray circles) and the fitted exponential curve of homologous reaction I (black curve) were also plotted. For clarity, only 20% data of each trace was shown. All the data were collected under standard reaction conditions. In B, the RecA(S1) filament concentration and the homologous dsDNA D0 were maintained at 0.1 and 0.5 μM, respectively, while the concentration of heterologous dsDNA H1 was increased from 0.0 to 2.0 μM as indicated by the arrow. In C, the concentration of the dsDNA D1 and the RecA(S0) filament were maintained at 0.1 and 0.1 μM respectively, while the concentration of heterologous dsDNA H1 was increased from 0.0 to 2.0 μM as indicated by the arrow.
reaction (Reaction I": RecA(S1) + H0, open diamond), and two completely heterologous reactions (Reaction Het1: RecA(S1) + H1, filled triangles; and Reaction Het2: RecA(S1) + H2, open triangles). The time course of a control reaction involving the mixing of the RecA(S1) filament with reaction buffer (gray circles) is also shown. In Reaction I", the dsDNA H0 contains the same sequence as D0 except that two adjacent AT base pairs in D0 were substituted with two TA base pairs. This substitution did not substantially alter the stability of the dsDNA substrate, while the C strand in H0 now contains a two-base mismatch relative to the sequence of the I strand (S1). For the two completely heterologous reactions, the dsDNA substrates H1 and H2 contain sequence homologies no greater than three consecutive basepairs with D0, and only differ from each other in their base compositions: one strand of H1 has a base composition identical to that of the I strand S1, while H2 has no base composition bias in either strand.

As shown in Figure 3.9A, the partially homologous Reaction I" is apparently slower than the completely homologous Reaction I, but the total amplitude change did not show a significant difference. Interestingly, both the two heterologous reactions indicated that the RecA(I) filament bound the dsDNA substrates in spite of their lack of sequence homology, and the fluorescence changes and/or the extents of the heterologous "reactions" appear dependent on the sequence of the duplex DNA (Figure 3.9A, filled and open triangles). Most importantly, the two heterologous reactions are obviously different from the completely or partially homologous Reaction I and I", indicating that
the heterologous dsDNA is incompetent to carry out the homologous strand exchange reaction. Nevertheless, these results demonstrate the presence of sequence-nonspecific contacts between the RecA(I) filament and the heterologous dsDNA substrates. This nonspecific contact is a prerequisite for the occurrence of homologous DNA strand exchange.

Because we observed the sequence-nonspecific contact between the RecA(I) filament with both heterologous dsDNAs H1 and H2, and the fluorescence change observed on H1 is greater than that on H2, we conclude that the sequence-nonspecific interaction is general to the RecA-mediated strand exchange reaction under our solution conditions. However, the observation of this contact using fluorescence emission changes is greatly enhanced during the reaction with heterologous duplex H1. In the following experiments, heterologous duplex H1 was used to investigate the sequence-nonspecific contact in order to achieve a better visible signal level.

We first tested the possibility of inhibition of homologous strand exchange in Reaction I and O by increasing the concentration of heterologous DNA. In the inhibition test for Reaction I, 0.1 μM RecA(S1) filament was mixed with a combination of 0.5 μM homologous D0 and different concentrations of heterologous H1 (0 to 1.5 μM). In the inhibition test for Reaction O, 0.1 μM RecA(S0) filament was mixed with a combination of 0.1 μM D1 and different concentrations of H1 (0 to 2.0 μM). As shown in Figure
3.9B and C, the total amplitude changes for Reactions I and O decreased when the concentration of H1 increased. This suggests that the presence of heterologous dsDNA inhibits the homologous strand exchange reaction, probably by transiently associating with the RecA(I) filament thus leaving less filament available for productive interaction with the homologous dsDNA.

We then tested the concentration dependence for the partially homologous Reaction I\(^n\) (RecA(S1) + H0) and the completely heterologous Reaction Het1 (RecA(S1) + H1). A three-exponential function was required to describe adequately the time course of Reaction I\(^n\), while a double-exponential function was sufficient for the description of Reaction Het1. The fitted relaxation times and amplitude changes at different substrate concentration are summarized in Table 3.3 and plotted in Figure 3.10. The concentration-independent total amplitude change for Reaction I\(^n\) (Figure 3.10A, diamond) is similar to that of Reaction I. This indicates that two consecutive mismatches between the RecA-bound ssDNA and the complementary strand of the dsDNA substrate does not substantially influence the extent of the strand exchange reaction. In contrast, the total amplitude change for Reaction Het1 (Figure 3.10A, triangle) is strongly concentration-dependent and is always much smaller in magnitude than those of Reactions I\(^n\) and I. The difference in the total amplitude change between Reactions I (or I\(^n\)) and Het1 demonstrated the RecA(I) filament is capable of discriminating a partially homologous and a completely heterologous substrate.
<table>
<thead>
<tr>
<th>[S] (\mu\text{M})</th>
<th>(\tau_1) (s)</th>
<th>(\Delta F_1)</th>
<th>(\tau_2) (s)</th>
<th>(\Delta F_2)</th>
<th>(\tau_3) (s)</th>
<th>(\Delta F_3)</th>
<th>(\Delta F_{\text{Total}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction I* R(S1) + H0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.1 ± 0.16</td>
<td>-0.012 ± 0.0017</td>
<td>4.6 ± 0.47</td>
<td>-0.019 ± 0.0014</td>
<td>27 ± 2.9</td>
<td>-0.0071± 0.00065</td>
<td>-0.038</td>
</tr>
<tr>
<td>0.7</td>
<td>0.72 ± 0.076</td>
<td>-0.015 ± 0.0012</td>
<td>3.5 ± 0.29</td>
<td>-0.016 ± 0.0011</td>
<td>26 ± 2.9</td>
<td>-0.0038± 0.00034</td>
<td>-0.035</td>
</tr>
<tr>
<td>1.0</td>
<td>0.58 ± 0.048</td>
<td>-0.020 ± 0.0012</td>
<td>2.9 ± 0.24</td>
<td>-0.014 ± 0.0011</td>
<td>31 ± 4.3</td>
<td>-0.0024± 0.00018</td>
<td>-0.036</td>
</tr>
<tr>
<td>1.5</td>
<td>0.39 ± 0.035</td>
<td>-0.021 ± 0.0014</td>
<td>1.8 ± 0.18</td>
<td>-0.012 ± 0.0014</td>
<td>29 ± 3.7</td>
<td>-0.0021± 0.00013</td>
<td>-0.036</td>
</tr>
<tr>
<td>2.5</td>
<td>0.24 ± 0.020</td>
<td>-0.026 ± 0.0015</td>
<td>1.2 ± 0.15</td>
<td>-0.0098± 0.0015</td>
<td>33 ± 4.1</td>
<td>-0.0020± 0.00009</td>
<td>-0.038</td>
</tr>
<tr>
<td>Reaction Het1 R(S1) + H1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.5 ± 0.08</td>
<td>-0.0092± 0.00022</td>
<td>21 ± 2.5</td>
<td>-0.0024± 0.00014</td>
<td></td>
<td></td>
<td>-0.012</td>
</tr>
<tr>
<td>0.7</td>
<td>1.2 ± 0.06</td>
<td>-0.013 ± 0.0003</td>
<td>12 ± 0.8</td>
<td>-0.0052± 0.00021</td>
<td></td>
<td></td>
<td>-0.018</td>
</tr>
<tr>
<td>1.0</td>
<td>1.1 ± 0.04</td>
<td>-0.016 ± 0.0002</td>
<td>16 ± 0.6</td>
<td>-0.0071± 0.00016</td>
<td></td>
<td></td>
<td>-0.023</td>
</tr>
</tbody>
</table>

*Values reported are the parameters that describe the exponential curve fit of the raw data generated by KaleidaGraph. Amplitude changes (\(\Delta F\)) are the true values reported by KaleidaGraph for the curve fits. *In both reactions, the duplex DNA concentration is varied as indicated in the [S] column while the concentration of the preformed R(S1) filament is maintained at 0.1 \(\mu\text{M}\); All concentrations are expressed in molecules.
Table 3.4: Exponential Curve Fit Parameters for Reaction I at different temperatures *^a^

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>$\tau_1$ (s)</th>
<th>$\Delta F_1$</th>
<th>$\tau_2$ (s)</th>
<th>$\Delta F_2$</th>
<th>$\tau_3$ (s)</th>
<th>$\Delta F_3$</th>
<th>$\Delta F_{Total}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>1.1 ± 0.21</td>
<td>-0.0091 ± 0.0014</td>
<td>5.8 ± 1.5</td>
<td>-0.0094 ± 0.0010</td>
<td>35 ± 7.7</td>
<td>-0.0080 ± 0.0010</td>
<td>-0.026</td>
</tr>
<tr>
<td>25</td>
<td>0.51 ± 0.067</td>
<td>-0.015 ± 0.0011</td>
<td>3.6 ± 0.60</td>
<td>-0.013 ± 0.0010</td>
<td>22 ± 3.2</td>
<td>-0.0084 ± 0.0010</td>
<td>-0.035</td>
</tr>
<tr>
<td>30</td>
<td>0.41 ± 0.037</td>
<td>-0.018 ± 0.0008</td>
<td>3.4 ± 0.37</td>
<td>-0.012 ± 0.0007</td>
<td>25 ± 3.5</td>
<td>-0.0046 ± 0.00047</td>
<td>-0.035</td>
</tr>
<tr>
<td>34</td>
<td>0.26 ± 0.023</td>
<td>-0.021 ± 0.0009</td>
<td>2.4 ± 0.19</td>
<td>-0.017 ± 0.0007</td>
<td>20 ± 2.5</td>
<td>-0.0044 ± 0.00042</td>
<td>-0.043</td>
</tr>
<tr>
<td>37</td>
<td>0.17 ± 0.016</td>
<td>-0.023 ± 0.0011</td>
<td>1.8 ± 0.12</td>
<td>-0.018 ± 0.0008</td>
<td>25 ± 4.0</td>
<td>-0.0030 ± 0.00023</td>
<td>-0.044</td>
</tr>
</tbody>
</table>

*Values reported are the parameters that describe the 3-exponential curve fit of the raw data generated by KaleidaGraph. Amplitude changes ($\Delta F$) are the true values reported by KaleidaGraph for the curve fits. Figure 1 and 2 shows the components of the reaction I indicated here.*
When a pseudo-first order concentration dependence analysis was performed on
the reciprocal relaxation times of the first phases of Reactions I" and Het1, bimolecular
association rate constants of $1.7 \pm 0.2 \times 10^6$ and $0.4 \pm 0.1 \times 10^6$ M$^{-1}$s$^{-1}$, respectively, and
unimolecular dissociation rate constants of $0.2 \pm 0.2$ and $0.5 \pm 0.1$ s$^{-1}$, respectively, were
determined (Figure 3.10B). A comparison of the equilibrium constants ($K_1$ ) for
formation of the first detectable intermediates in Reactions I/O, I", and Het1, we found
that the $K_1$ of Reaction Het1 ($1.0 \times 10^6$ M$^{-1}$) is significantly smaller that that of Reaction
I/O (average value of $7.2 \times 10^6$ M$^{-1}$) or I" ($8.3 \times 10^6$ M$^{-1}$).

**Temperature dependence of the strand exchange reaction**

We investigated the temperature dependence of Reaction I in order to study the
thermodynamic nature of the RecA-mediated strand exchange reaction. In this
experiment, 0.1 μM RecA(S1) filament was mixed with 0.5 μM dsDNA D0 at different
temperatures and the fluorescence change during the reaction were recorded as a function
of time, and analyzed using a three-exponential function as described above. The fitted
relaxation times and amplitude changes are summarized in Table 3.4. Figure 3.11A
demonstrates that the relaxation times decreased while the total amplitude change
increased at higher temperature. The half-time of the slowest reaction, observed at 20 °C,
suggests that it is completed within the 100 s time window of data collection. These
Figure 3.10 Concentration-dependence of the total amplitude changes (A) and the reciprocal of relaxation times (B) of Reaction Het1 (RecA(S1)+H1, filled triangles) and Reaction II’ (RecA(S1)+H0, open diamonds). For both of the reactions, the concentration of RecA(S1) filament was maintained at 0.1 μM while the concentration of dsDNA varied. For H1, the concentration was 0.5, 0.7, and 1.0 μM, respectively. For H0, the concentration was at 0.5, 0.7, 1.0, 1.5 and 2.5 μM respectively. As shown in A, the total amplitudes changes for Reaction Het1 is strongly dependent on the dsDNA concentration while Reaction II’ is not. In B, the reciprocal of the relaxation times of the first phases of Reaction II’ exhibit a greater linear-dependence on the dsDNA concentration than the Reaction Het1.
Figure 3.11 Overview of the temperature-dependence of Reaction I. The concentration of the RecA(S1) filament and dsDNA D0 were maintained at 0.1 and 0.5 μM, respectively. The temperatures under which the reaction was performed were 20, 25, 30, 34 and 37°C from bottom to top, respectively. For clarity reasons, the actual data points are omitted and only the fitted curves are shown in panel A. In order to view the clear temperature-dependent kinetics, the starting points of the five traces in panel A were also offsetted to be the same. Panel B shows the original data points overlapped with the simulated reaction progress curves (black solid curves) using a sequential four steps reaction scheme as shown in Figure 3.11. For clarity, each data is offsetted 0.02 signal level to separate from each others.
observations suggest that both the rate and the overall equilibrium of the reaction were influenced by the temperature.

**Kinetic simulation of the concentration-dependent data**

In order to resolve a robust kinetic scheme and true rate constants for molecular events of the RecA-mediated strand exchange reaction, we employed DynaFit (267) to simulate the experimental data. We used the following two procedures during the simulations. First, in order to alleviate the large noise inherent in the early time data, we used the exponentially-fitted and 95% confidence interval weighted curves as the input data for the simulations. Thus, all data points are equally weighted during the simulation. In addition, in order to introduce less systematic error into the simulation, we normalized the fitted curves of the concentration-dependent experiments to have the same extent of fluorescence change based on the observation that the total amplitude changes of the reactions are effectively independent of substrate concentration. For the same reason, we also off-set the starting fluorescence signals of the temperature dependent experiments to the same level.

Second, in order to have as few variables as possible in the simulation, we calculated the response factors of the initial fluorescent substrate by dividing the fluorescence signal of the reaction at time 0 by the concentration of the fluorescent substrate. For example, the initial fluorescence signal was 0.096 for Reaction I and the
response factor of the fluorescent RecA(S1) filament is $0.096/0.1 \mu M = 0.96 \mu M^{-1}$. To further simplify the simulation, we also estimated the response factor of the product, namely, the RecA(tsDNA) complex using the following procedure. We first assumed that the reaction proceeds to 100% product formation. This is a reasonable assumption based on the small apparent dissociation constant ($K_d \leq 0.01 \mu M$, greater than 98% product conversion at the standard reaction condition). We then calculated the corresponding response factor of the product by simply dividing the fluorescence signal at the end of the reaction by 0.1 \mu M. Next, in the simulation, we manually adjusted the calculated response factor of the product within a 10% range to have the simulated $K_d$ (calculated based on the simulated rate constants) of the reaction match the upper limit of the $K_d$ we measured using the steady-state fluorescence titration ($\leq 0.01 \mu M$). This adjustment only resulted in small changes of the rate constants of the last two steps while the rate constants of the first one or two steps were not affected at all.

The criterion we used in selecting the best kinetic scheme is summarized as follows. First, we selected the mechanism with the smallest $\chi^2$ given by the simulation. Second, we used multiexponential functions as described in Materials and Methods to fit the simulated curves and extract phase relaxation times and amplitudes from the simulated data sets. Finally, we selected the mechanism for which the concentration-
dependent phase relaxation times and amplitudes most closely agreed with those derived directly from the experimental data.

We first simulated the concentration-dependent data sets for Reactions I, C and O using a common kinetic scheme with globally varied rate constants but locally varied response factors of the intermediates for each reaction. As many as thirteen concentration-dependent traces for the three reactions were simulated simultaneously. We tried a variety of mechanistic models that involved between three and five discrete steps. The following sequential four-step mechanism gave the best approximation of experimental data:

\[
\begin{align*}
R + D & \overset{k_1}{\underset{k_{-1}}{\rightleftharpoons}} I_1 & \overset{k_2}{\underset{k_{-2}}{\rightleftharpoons}} I_2 & \overset{k_3}{\underset{k_{-3}}{\rightleftharpoons}} I_3 & \overset{k_4}{\underset{k_{-4}}{\rightleftharpoons}} P
\end{align*}
\]

In Scheme 1, R is the RecA(I) filament, D is the dsDNA substrate, I₁ to I₃ are the intermediates, and P is the RecA(tsDNA) complex. The first rate constant \(k_{+1}\) is a bimolecular association constant while the others are all first order rate constants in unit of \(s^{-1}\). The best simulated rate constants and response factors are summarized in Table 3.5. This mechanism with this set of parameters accounts very well for all the concentration–dependent data of the three reactions, as can be visualized by the excellent superimposition of the simulated traces (black lines) and the experimental traces (grey circles) of Reaction I plotted in Figure 3.7B as an example. Interestingly, we found that
Table 3.5. Rate constants of Reaction I, C and O at different substrate concentrations solved by global simulation based on the sequential four steps mechanism

<table>
<thead>
<tr>
<th>Data Set</th>
<th>$k_1$ (μM$^{-1}$s$^{-1}$)</th>
<th>$k_2$ (s$^{-1}$)</th>
<th>$k_3$ (s$^{-1}$)</th>
<th>$k_4$ (s$^{-1}$)</th>
<th>$k_5$ (s$^{-1}$)</th>
<th>$k_6$ (s$^{-1}$)</th>
<th>$k_7$ (s$^{-1}$)</th>
<th>$k_8$ (s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global*</td>
<td>6.4 ± 0.02</td>
<td>0.84 ± 0.12</td>
<td>0.12 ± 0.020</td>
<td>0.13 ± 0.042</td>
<td>0.42 ± 0.051</td>
<td>0.015 ± 0.011</td>
<td>0.053 ± 0.009</td>
<td>≤ 0.06</td>
</tr>
<tr>
<td>Reaction I</td>
<td>0.96</td>
<td>0.63 ± 0.070</td>
<td>0.25 ± 0.32</td>
<td>0.60 ± 0.010</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction C</td>
<td>0.15</td>
<td>0.19 ± 0.013</td>
<td>0.80 ± 0.13</td>
<td>0.56 ± 0.052</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction O</td>
<td>0.20</td>
<td>0.81 ± 0.045</td>
<td>0.97 ± 0.066</td>
<td>0.84 ± 0.018</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The value reported here are the parameters obtained from a global simulation on the concentration dependence data of reaction I, C and O using a sequential four steps mechanism as described in Figure 7, and are presented as the best simulated parameters plus or minus formal standard error as defined by the simulation program Dynafit. The response factors for R or D are calculated by dividing the starting fluorescence signal of each reaction by the concentration of the fluorescent species (0.1 μM). The response factors for I$_1$, I$_2$ and I$_3$ were obtained from the global simulation as described above, and presented as the averaged values of the five or four concentration points for each reaction plus or minus one standard deviation. The response factors for P were calculated by dividing the final fluorescence signal by the product concentration, which is assumed to be 0.1 μM according to steady-state $K_d$ measurement (Figure 3.5).
the first forward and reverse rate constants ($k_{+1} = 6.4 \pm 0.02 \mu\text{M}^{-1}\text{s}^{-1}$ and $k_{-1} = 0.84 \pm 0.12 \text{s}^{-1}$) are in excellent quantitative agreement with those obtained using the pseudo-first order concentration dependence analyses for Reactions I and O. This indicates that the first step is less coupled with the following steps and the pseudo-first order analysis is a good approximation. The latter three steps have similar and much slower rate constants, suggesting that they correspond to the rate-limiting steps of the reaction. Simulation of the concentration-dependent data for Reaction I showed similar results with slightly decreased values of $k_{+1}$ and $k_{-1}$, recapitulating the pseudo-first order analysis (Figure 3.10B).

We also examined the response factors of the reaction species. These fluorescence response factors characterize the intrinsic fluorescence properties of each reaction species and reflect, in part, the local environment of the fluorophore in each species. Because the three sets of concentration-dependent experiments used in the simulation were performed on different dates with different photomultiplier settings, the response factors given by the simulation are all arbitrary units and can only be compared within but not among reactions. Therefore, we performed the following normalization in order to compare the three reactions directly. First, we normalized the response factor of P in Reaction C to be the same as that of P in Reaction I. This normalization is based on the observation in the steady- and transient-state measurements that both Reaction C and I have similar fluorescence intensity at the end of the reaction (see Figure 3.6A, D and
Table 3.1). Then we normalized the response factors of the other species (D, I₁ to I₃) of Reaction C in relative to the normalized response factor of P. Second, we normalized the response factor of the most fluorescent intermediate I₂ in Reaction O to be the same as that of the RecA(I) filament (R) of Reaction I. This normalization is based on the observation in steady- and transient-state measurements that the RecA(I) filament has the highest fluorescence among all the RecA(DNA) complexes (Table 3.1 and Figure 3.6A), and the O strand is likely to be a RecA-bound ssDNA near the end of the reaction (261). Thus, we used the fluorescence of the RecA(I) filament as an upper limit for the fluorescence that the O strand could possibly achieve during the reaction. Then we normalized the response factors of other species (R, I₁, I₃ and P) of Reaction O in relative to the normalized response factor of I₂. The final values of the normalized response factors of the species in the three reactions are reported in Table 3.1. Interestingly, the normalized response factors of D in Reaction C and O are close to each other (0.15 for Reaction C and 0.20 for Reaction O), implying that the assumption we used for the normalization are good approximations.

These normalized response factors of different species of the three reactions were plotted along the reaction progress (Figure 3.12), and an interesting pattern emerged. In general, the changes of the fluorescence properties of each species follow the same trends as we expected. In Reaction I, the 6MI-labeled I strand goes from a RecA-bound ssDNA to a RecA-bound dsDNA, thus the fluorescence of the intermediates and product in
Figure 3.12 Response factors of each species in Reaction I (blue), C (red) and O (green) resulting from the global simulation using a sequential four steps mechanism as shown at the bottom of the plot. The fitted parameters are listed in Table 5. The global simulation takes the three sets of concentration-dependence data of the three reactions and simultaneously simulate the reaction progress curves using the same mechanism with locally varied response factors. The response factors of species with fitting errors plotted here are normalized as described in text. R: RecA(ssDNA) filament; D: dsDNA; I₁, I₂ and I₃ are intermediates and P is the product. Please note that the response factors of the reactants (R and D) and product (P) are well defined while intermediate I₂ has the largest fitting uncertainties.
Reaction I is highly quenched. In Reaction C, the 6MI-labeled C strand goes from a RecA-free dsDNA to a RecA-bound dsDNA, thus the fluorescence of the intermediates and product in Reaction C is enhanced. In Reaction O, the 6MI-labeled O strand goes from a RecA-free dsDNA to a RecA-bound ssDNA, thus the fluorescence of the intermediates and product in Reaction O is highly enhanced. Most intriguingly, we found that the fluorescence of the first intermediate I₁ in Reaction I and O changed dramatically from the starting substrate R or D, while the fluorescence of the intermediates in Reaction C remained essentially the same until the reaction proceeds to intermediate I₂. This pattern of fluorescence change of the intermediates may explain why the observations of first and second phases are mutually exclusive for Reactions C and O. In addition, the similar fluorescence response factors of I₁ in Reaction I and O implies that in the first intermediate, the local environment of the I and O strand may be very similar to each other, while the close contact between the I and C strand, namely, base pairing, did not evolve until the reaction proceeds to the last intermediate I₃.

**Kinetic simulation of temperature-dependent data**

We then simulated the temperature-dependent data for Reaction I using the same sequential four-step mechanism (Scheme 1). We first calculated the response factor of the substrate R as described previously. Then we calculated the response factors of the three intermediates and product according to the relative ratios of the response factors in
Reaction I given by the global simulation of the concentration dependence experiments. In order to have the simulated \( K_d \) of the reaction at 37 °C match the experimentally measured \( K_d \), we also manually adjusted the response factor of P slightly. This adjustment results in about 2% difference (0.693) from the calculated response factor of P (0.707) using the relative ratio. The fluorescence properties of 6MI are largely insensitive to temperature over the range of temperatures employed herein (270). Therefore, the response factors calculated above were fixed as constants for the simulations of the reactions at different temperatures. Finally, each time trace for Reaction I at a unique temperature was simulated individually using Scheme 1 with the rate constants variable. The best-fitted rate constants are summarized in Table 3.6. The fitting goodness can be visualized by the superimposition of the simulated curves (solid lines) and the experimentally-obtained data (grey circles) as shown in Figure 3.11B.

**Arrhenius and van’t Hoff analysis of temperature-dependent constants**

By analyzing the rate constants of the reaction at different temperatures, we found that the temperature exerts its largest influence on the first step of the reaction, especially in the forward, associative direction. The later three steps exhibit rate and equilibrium constants with substantially smaller temperature-dependence. Hence, we further investigated the thermodynamic nature of the first step of the reaction. Arrhenius (Figure 3.13A) and van’t Hoff plots (Figure 3.13B) were constructed to analyze the first
Table 3.6. Rate constants and equilibrium constants of Reaction I at different temperatures solved by simulation based on the sequential four steps mechanism.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>$k_1$ (μM$^{-1}$s$^{-1}$)</th>
<th>$k_3$ (s$^{-1}$)</th>
<th>$K_{eq1}$ (μM$^{-1}$)</th>
<th>$k_2$ (s$^{-1}$)</th>
<th>$k_{eq2}$</th>
<th>$k_3$ (s$^{-1}$)</th>
<th>$K_{eq3}$</th>
<th>$k_4$ (s$^{-1}$)</th>
<th>$K_{eq4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.5</td>
<td>0.9</td>
<td>0.6</td>
<td>0.20</td>
<td>0.25</td>
<td>0.8</td>
<td>0.12</td>
<td>0.08</td>
<td>1.4</td>
</tr>
<tr>
<td>25</td>
<td>1.6</td>
<td>1.2</td>
<td>1.4</td>
<td>0.16</td>
<td>0.22</td>
<td>0.7</td>
<td>0.13</td>
<td>0.10</td>
<td>1.2</td>
</tr>
<tr>
<td>30</td>
<td>3.4</td>
<td>2.6</td>
<td>1.3</td>
<td>0.53</td>
<td>0.63</td>
<td>0.8</td>
<td>0.34</td>
<td>0.17</td>
<td>2.0</td>
</tr>
<tr>
<td>34</td>
<td>6.8</td>
<td>3.8</td>
<td>1.8</td>
<td>0.63</td>
<td>0.52</td>
<td>1.2</td>
<td>0.42</td>
<td>0.16</td>
<td>2.6</td>
</tr>
<tr>
<td>37</td>
<td>9.5</td>
<td>2.8</td>
<td>3.3</td>
<td>0.49</td>
<td>0.48</td>
<td>1.0</td>
<td>0.46</td>
<td>0.29</td>
<td>1.6</td>
</tr>
</tbody>
</table>

$^a$ Values reported are the parameters simulated on the temperature dependence data of reaction I (Figure 3.11) using a sequential four steps mechanism as described in Figure 8. The response factors for each reaction species are R = 1.18, I1 = 0.78, I2 = 0.39, I3 = 0.73, P = 0.69. They were obtained from the global simulation of the concentration dependence data of reaction I and maintained constant through different temperatures.
**Figure 3.13** Thermodynamic analysis of strand exchange. The data plotted here are generated from the best simulated kinetic parameters for the temperature-dependent data of Reaction I using the same sequential four steps mechanism. The best simulated parameters are listed in Table 6. A: Arrhenius plot of the forward rate constants of the first step revealed a large activation energy at 135 kJ·mol\(^{-1}\). B: van’t Hoff plot of the equilibrium constants of the first step shows positive enthalpy and entropy changes at 65 kJ·mol\(^{-1}\) and 334 J·mol\(^{-1}\)·K\(^{-1}\). C: free energy (\(\Delta G\), diamonds), enthalpy (\(\Delta H\), right triangles) and the product of temperature and entropy (\(\Delta S\cdot T\), inverted triangles) of Reaction I at 37 °C calculated from the best fitted rate constants. Error bars are plotted for each data point but some of them are too small to be visible with the scale used. It is apparent that the largest negative free energy change occurs at the first step while the following three steps are almost isoenergetic. Interestingly, for each step, the product of temperature and entropy changes covers the positive enthalpy changes.
Figure 3.13
bimolecular association rate constant \( k_{+1} \) and first equilibrium constant \( (K_i) \). The Arrhenius plot revealed a large, positive activation energy, \( E_a = 135 \pm 1 \text{ kJ} \cdot \text{mol}^{-1} \), while the van't Hoff plot revealed positive enthalpy and entropy changes for the first step, \( \Delta H^\circ = 65 \pm 1 \text{ kJ} \cdot \text{mol}^{-1} \) and \( \Delta S^\circ = 334 \pm 2 \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1} \).

**DISCUSSION**

We have presented data describing the characteristics of the RecA-mediated DNA strand exchange reaction using ODN substrates containing a fluorescent base analog substituted for a single guanine at a selected position on one of the three substrate DNA strands. The reaction occurs between a RecA(ssDNA) filament and naked dsDNA to produce a RecA(tsDNA) product comprised of a heteroduplex formed between the incoming ssDNA and the complementary strand of the dsDNA and a displaced (but RecA-bound) outgoing strand. Mechanistically, the reaction proceeds via at least four discrete steps, the first of which is a rapid bimolecular association and the last three of which are slower, unimolecular events. As the reaction proceeds, the total emission of the single fluorophore changes in a way that is characteristic of each of the three strand’s fate during the reaction. Nevertheless, the real-time fluorescence changes for the three strands are simultaneously described by the four-step mechanism, indicating that the three strands are processed in a concerted, or parallel, manner. Further analysis of the fluorescence and energetic changes associated with each step can reveal clues as to the
nature of the molecular events involved, and a discussion of our inferences are discussed below.

**RecA-mediated strand exchange using oligonucleotides with single fluorescent base analogs**

Historically, the strand exchange reaction mediated by the RecA protein has been investigated using long (3-10 kb) natural DNA substrates such as the single-stranded or replicative forms of bacteriophages (for examples, see (123,154,191)). While the long DNA substrates have the advantage of mimicking the *in vivo* substrates, they also have disadvantages associated with topological constraints (131,271), coaggregation (190,191), and nonuniform distribution of long-range interactions (69,145,272). Recently the use of short (< 100 bp) oligonucleotides (ODNs) has demonstrated the above problems may be overcome by performing the strand exchange reaction with relatively short ODN substrates (180,200-202,234,261,273,274).

It is important to emphasize that while the use of short ODNs undoubtedly leads to fewer observable events, such as fewer nucleation events and the exclusion of late-stage branch migration processes, it also allows a focus on the early events in homologous recombination. Importantly, we demonstrated that our reaction system, with the combination of short ODNs and intrinsic DNA fluorescence, is able to not only carry the RecA-mediated strand exchange reaction efficiently, but also recapitulates many aspects of the RecA-mediated strand exchange reaction. Specifically, these aspects
include the rapid formation of sequence-nonspecific contact between the RecA(I) filament and dsDNA substrate (Figure 3.9), the sequence-specific formation of a synaptic complex (Figure 3.6), the discrimination of the three strands inside the RecA filament during and after strand exchange (Figure 3.4 and 6), and characteristic energetic requirements for each reaction step (Figure 3.11 and 12).

In our previous work using a pair of external fluorophores (261), we have demonstrated that in a post-exchange intermediate, the three DNA strands inside the RecA filament adopt similar extended and unwound conformations. Herein we showed in great detail that although the three DNA helical geometries are similar to each other, the DNA-DNA and protein-DNA interactions among the three strands are not equivalent. This difference can only be ascertained by the use of 6MI, a sensitive fluorescent guanine analog, which greatly facilitates the direct observation of the DNA base during the reaction with the least structural perturbations (for review, see 275). Historically, the use of fluorescent ODNs has involved fluorescent base analogs such as 2-aminopurine (2AP) (239), etheno-modified dA(εA) (151, 276-278), or externally conjugated fluorophores such as the fluorescein, rhodamine (201, 202, 261), hexachlorofluorescein (200), and cyanine dyes (239). The advantage of fluorescent base analogs 2AP and εA over external bulky fluorophores are apparent, in that they cause less perturbation of structure and interaction. However, the ultraviolet colors and low quantum yields in DNA (≈ 0.02)
of 2AP and εA make them less desirable for studying the RecA protein. This situation was because the RecA protein is not an enzyme with respect to its DNA binding activity. Instead, stoichiometric amounts of the protein coat the DNA molecule. Such high tryptophan levels lead to a large background fluorescence problem. The interfering background is exacerbated by the fact that large RecA complexes also produce a substantial light scattering signal (279). Pioneering work using the εA monomers (276,280,281) and multiple 2AP fluorophores (239) allowed RecA-DNA interactions to be observed, but the 6MI moiety presents significant advantages. First, the visible color of 6MI fluorescence greatly reduces the effect of background tryptophan signal. Second, the enhanced 6MI photophysical properties (e.g., unique excitation and larger emission quantum yield (∼ 0.3, derived from Hawkins et al. (268) ) allows the labeling of ODNs at specific single bases. Thus, the DNA signal will be more readily interpretable from a structural perspective. By contrast, the fluorescence from εA in chemically treated DNA represents the average of a population of heterogeneously labeled molecules with multiple fluorophores. Third, the 6MI base is a functional and structural analog of guanine and therefore has least perturbation on the DNA structure. In comparison, the εA base can not basepair with thymine, and it locally disrupts the DNA helix. Lastly, 6MI is extremely sensitive to local environmental changes such as protein-DNA interaction, base-pairing and stacking (275). In summary, these features make our
reaction system a simple yet clean and powerful *in vitro* model system for the study of *in vivo* homologous DNA recombination.

**The RecA protein distinguishes the three DNA substrate strands.**

In the steady state, the combination of fluorescence intensity, emission maximum, and anisotropy of the 6MI-containing ODNs after strand exchange are characteristic of each strand, suggesting that the local environments of the three DNA strands inside the RecA filament may be unique. This finding is consistent with the presence of two or three distinct DNA binding sites inside the RecA filament, as suggested by previous work (102,147,282,283). The ability of the RecA protein to rapidly distinguish the three strands is reflected in the transient-state fluorescence measurements: the 6MI contained in each strand exhibits a characteristic kinetic trace during the reaction as shown in Figure 3.6. In addition, the fluorescence property of each strand solved by the computational simulation also indicates that the three strands are treated differently by the RecA protein during the reaction time course. However, a robust simulation of the unique kinetic traces can be adequately achieved using kinetic rate constants that are common for the strand exchange reaction and do not reflect differences in the strands. Hence, we conclude that while RecA mediates the rapid sensing of unique features of the strands’ sequences, the strands are undergoing changes in a relatively concerted manner.
The influence of heterologous DNA on strand exchange

In contrast to the clear distinction among the three DNA strands in the homologous reactions discussed above, we found that the interaction between a RecA(I) filament and a heterologous dsDNA is quite different. The fluorescence intensity and emission maximum of Reactions Het1 and Het2, as shown in Figure 3.4, appear largely similar to those of the starting substrates, reflecting the inability of these heterologous substrates to complete the strand exchange reaction. Nevertheless, we observed increased anisotropy values, as well as slightly altered fluorescence intensities in the heterologous reactions. This observation suggests the presence of sequence-nonspecific contact between the filament and the dsDNA substrates, and possibly the formation of a heterologous RecA(tsDNA) complex. The putative existence of such a complex is further substantiated by the effect of heterologous duplex DNA on the kinetics of homologous reactions. The observed inhibition of the homologous reaction by the presence of heterologous duplex DNA is likely caused by competitive, nonspecific binding of heterologous dsDNA to the RecA(I) filament whereby less filament is available for productive strand exchange.

Previous studies on the effect of heterologous DNA on homologous reactions produced varied results. The Camerini-Otero laboratory reported that heterologous DNA did not affect the rate of the synaptic complex formation (199), while the Radding
laboratory reported that heterologous DNA paradoxically accelerates the formation of joint molecule (192). The first conclusion was inferred from measurements of the protection from restriction enzyme digestion sites afforded by the formation of synaptic complex over an eight-minute time window. This assay method requires that the nonspecific heterologous complex have a stability and/or concentration comparable to that of the homologous synaptic complex in order to be detected. According to our measurements, the heterologous complex has a smaller equilibrium constant. Hence, the two requirements were not met in the studies by Yancey-Wrona et al. (1995), and they failed to detect the inhibition effect of heterologous DNA. The second study used long DNA substrates under conditions in which RecA-DNA networks or coaggregates likely form. The increased rate of synaptic complex formation by heterologous DNA may be due to the increased concentration of local DNA by the formation of a sequence nonspecific network. This situation is not represented in our reaction conditions.

It is important to note the previous work demonstrating that the RecA protein has different binding affinities for ssDNA substrates with different base compositions (103,242,284,285). During our heterologous reaction experiments using a conserved fluorophore placement, we observed different fluorescence changes whose magnitudes depended on the sequence of the heterologous dsDNA. We speculate that the RecA-ssDNA filament may also demonstrate sequence-dependent association that is independent of homology, and experiments are underway to test related hypotheses.
Indeed, several previous work reported that when the heterologous dsDNA was composed of scrambled sequences (mimicking our Reaction RecA(S1)+H2), no significant signal changes were observed (201,202). In the experiments described herein, we maximized the visible signal changes characteristic of the weak heterologous interactions by using duplex H1 instead of H2. This represents an "exaggerated" case in which we simply amplified the magnitudes of the apparent interaction between nucleoprotein filament and dsDNA while maintaining the heterologous nature of the interaction.

**A plausible kinetic scheme for RecA-mediated DNA strand exchange**

Based on our real time fluorescence data coupled with numerical simulation, we have proposed a sequential four-step mechanism to describe the RecA-mediated strand exchange reaction. This mechanism is the conceptually simplest model (e.g., no branched or circular pathways required) that can adequately account for the multiexponential kinetics observed.

Previous studies have also proposed multi-steps mechanisms, in which as many as three steps were involved (152,199-201,203). Our work is the first time that the reaction has been illustrated in much greater detail, a significant improvement resulting from both the high time resolution generated by the rapid-mixing stopped-flow fluorimetry and the high sensitivity of the 6MI fluorescence to local environment.
The equilibrium constant for formation of the first intermediate was determined to be $K_1 \approx 5 \mu M^{-1}$ (averaged value from Table 3.5 and 6), a value which agrees well with two previous measurements suggesting an equilibrium constant near $10 \mu M^{-1}$ (199,201). However, the first bimolecular association rate constant, $k_{+1} \approx 5 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$, as determined both by the pseudo-first order analysis (Figure 3.8B) and simulation (Table 3.5 and 6), is more than two orders of magnitude faster than the value reported by Yancey-Wrona et al. ($10^5$ to $10^6 \text{ M}^{-1}\text{s}^{-1}$) (199), while it is close to the value measured by Bazemore et al. ($\approx 1 \times 10^6 \text{ M}^{-1}\text{s}^{-1}$) (201). The apparent second-order rate constant reported by Yancey-Wrona et al. (1995) was inferred from a steady-state kinetic analysis, in which the initial rate of the formation of the synaptic complex over an eight-minute time window was monitored. Thus, these authors likely did not profit from an experimental setup affording a time resolution as high as that of stopped-flow fluorometry. Moreover, the results of Yancey-Wrona et al. (1995) were obtained in the presence of high concentrations ADP, the product of ATP hydrolysis and an inhibitor of RecA-DNA binding and strand exchange. Bazemore et al. (1997) reported a bimolecular association rate constant of about $10^6 \text{ M}^{-1}\text{s}^{-1}$ (201,202), which is about five-fold slower than what we measured in this study. Although this difference is not dramatic, it is likely significant. Thus we discuss these possible reasons for the difference. First, we used up to three, rather than two, exponential functions to analyze our data, which produce a better fitting
goodness as judged both visually and by reduced $\chi^2$. Second, our reaction is carried in the presence of ATPγS instead of ATP. Because ATPγS is well known to enhance the stability of recA-ssDNA complexes, it is also possible that the affinity of the RecA(ssDNA) filament to a homologous dsDNA may be different in these two conditions. Third, we used intrinsic DNA fluorescence as a sensitive local environmental change reporter, while Bazemore et al. (1997) used FRET between an extrinsic dye-label pair (fluorescein and rhodamine). If our proposal that the first step is the local association of the partially-unwound dsDNA with the RecA(ssDNA) filament (see schematic drawing for the structure of $I_1$ in Figure 3.16 and the following discussion) is correct, then we anticipate that the FRET efficiency between fluorescein and rhodamine, conjugated to the termini of 83mer ODNs, may not be sensitive to formation of the putative $I_1$. That is, they may not observe formation of the first intermediate as we saw using the intrinsic DNA fluorophore 6MI. In supporting of this idea, a pseudo-first order analysis assuming that the first intermediate $I_1$ has the same response factor as the substrates R+D (thus the formation of $I_1$ is “invisible”), we found that the apparent rate constant $k_{s1}$ for the first step (R+D $\rightarrow$ $I_2$) decreased about 10-fold (Figure 3.14). This may reconcile the difference between our measured $k_{s1}$ value and that reported by Bazemore et al. (1997)
Figure 3.14. A pseudo first order analysis comparison between the first phase of our study (filled squares, Figure 3.8B, Reaction I) and a proposed first phase based on the conditions employed by Bazemore et al. (1997) (open squares). Because the latter work reported a slower association rate constant than what we detected, we argue that the difference may be due to the inability of their reaction system to detect the first intermediate I₁. Therefore, we intended to view the kinetics by assuming the first intermediate I₁ is not detected using our reaction system. We first calculated the concentration change over time of each reaction species (R, D, I₁, I₂, I₃, and P) using our sequential four steps mechanism (Table 3.5 and 6) at different reaction concentrations of the dsDNA (varied from 0.5 to 2.5 μM) and RecA(ssDNA) filament (0.1 μM). Second, we assumed that the response factor of I₁ equals to that of R, which will render the
formation of $I_1$ "invisible" in the time traces. Third, we calculated the fluorescence change of the reaction by using the altered response factors and the concentration change profiles of each species. Then we used double-exponential functions to fit the reconstructed time-resolved fluorescence changes of the reactions and isolated the relaxation times of the first phase. Finally, we did the pseudo first order analysis by plotting the reciprocal of the first relaxation time against different concentrations of dsDNA. The apparent association constant for the altered analysis is $0.4 \pm 0.2 \times 10^6 \mu M^{-1}s^{-1}$, and the dissociation constant is $0.4 \pm 0.3 \times 10^6 s^{-1}$. 
In comparison with the first bimolecular association step, which is well characterized in part because it is well resolved from the subsequent events, the later three steps are less well defined, as indicated by the relatively large parameter uncertainties encountered in the simulation (Table 3.5). Nevertheless, the rate constants of the last three steps are significantly smaller than those of the first steps, indicating that they are the rate-limiting steps of the reaction. In fact, an analysis of the free energy changes for each step (Figure 3.13C) reveals only the first step of the reaction is associated with a large negative free energy change such that the first step dominates the overall thermodynamics of the reaction. Kowalczykowski and coworkers have suggested that the driving force for the strand exchange reaction is the different binding affinities of sites I and II for the respective DNA substrates and products of the reaction (103). Although we anticipated that any steps reflecting the crucial change in site occupancy (i.e., ssDNA → heteroduplex in site I and dsDNA → ssDNA in site II) would be reflected by a signature in the thermodynamic profile, we don't observe any energetic changes that can be unambiguously ascribed to such an event.

**Rapid kinetic discrimination of heterology during formation of the first detectable intermediate**

Because the rate constants of the bimolecular association between the filament and the dsDNA are reasonably well defined by both the pseudo-first order analysis and the results of simulation (compare the apparent $k_{+1}$ and $k_{-1}$ generated in Figure 3.8B and
the simulated results in Table 3.5), we compared the association and dissociation rate constants for the first steps of fully homologous Reaction I, partially homologous Reaction I\textsuperscript{m}, and completely heterologous Reaction Het1 based on the pseudo-first order analysis (Figure 3.10B). Using these rate constants, we calculated the activation free energies (\(\Delta G^a\)) and the free energy change for formation of the first intermediate I\textsubscript{1} (\(\Delta G^r\)) of the three reactions, and plotted the differences in the free energy changes (\(\Delta \Delta G^a\) and \(\Delta \Delta G^r\)) between the homologous Reaction I and Reaction I\textsuperscript{m} or Reaction Het1 in Figure 3.15.

Within experimental uncertainty, we noticed that for the partially heterologous Reaction I\textsuperscript{m}, the energy status of the first transition state is higher than that of the completely homologous Reaction I (\(\Delta \Delta G^a = 3.5 \pm 0.3\ \text{kJ}\cdot\text{mol}^{-1}\)), while that of the first intermediate I\textsubscript{1} is not significantly different from the homologous Reaction I (\(\Delta \Delta G^r = -1.1 \pm 2.4\ \text{kJ}\cdot\text{mol}^{-1}\)). In contrast, for Reaction Het1, the energy status of the first transition state and intermediate I\textsubscript{1} are both significantly higher than those of the homologous Reaction I and partially heterologous Reaction I\textsuperscript{m} (\(\Delta \Delta G^a = 6.9 \pm 0.5\ \text{kJ}\cdot\text{mol}^{-1}\), \(\Delta \Delta G^r = 4.5 \pm 1.4\ \text{kJ}\cdot\text{mol}^{-1}\)). The fact that the energy barrier between the starting substrates and the transition state increases along the presence of heterology in the duplex DNA substrate suggests that the kinetic discrimination against a heterologous sequence is made by the RecA(I) filament in the first step. In addition to this kinetic discrimination, the decreased
Figure 3.15 The difference in the free energy changes for formation of the first transition state ($\Delta \Delta G^\ddagger$) and the first intermediate $I_1$ ($\Delta \Delta G^\circ$) between Reaction I and $I^m$ (open bars), or between Reactions I and Het1 (RecA(S1)+H1, grey bars) calculated using equation 9 and 10. The rate constants ($k_{i,i}$) and equilibrium constants ($K_1$) were determined from the pseudo-first order analysis as shown in Figure 3.8B and 3.10B. For the partially heterologous Reaction $I^m$, the energy status of the transition state is higher than that of the completely homologous Reaction I ($\Delta \Delta G^\ddagger = 3.5 \pm 0.3$ kJ·mol$^{-1}$), while that of the first intermediate $I_1$ is not significantly different from the homologous Reaction I ($\Delta \Delta G^\circ = -1.1 \pm 2.4$ kJ·mol$^{-1}$). In contrast, for Reaction Het1, the energy status of the transition state and intermediate $I_1$ are both significantly higher than those of the homologous Reaction I and partially heterologous Reaction $I^m$ ($\Delta \Delta G^\ddagger = 6.9 \pm 0.5$ kJ·mol$^{-1}$, $\Delta \Delta G^\circ = 4.5 \pm 1.4$ kJ·mol$^{-1}$).
thermostability of I₁ in Reaction Het1 indicates that thermodynamic discrimination also plays a role in heterology screening.

The kinetic, but not thermodynamic distinction, between the perfectly homologous duplex D0 and the 2-bp mismatched duplex H0 by the RecA(I) filament demonstrates that disruptions in homology influence the energy level of the first transition state. It seems unlikely that such discrimination would occur during the formation of an initial collision complex, and we infer the presence of sequence-nonspecific collision complex that is not directly detected by our experimental system. Presumably, then, the rapid reorganization of the collision complex to the subsequent intermediate I₁, in which the three DNA strands are in close proximity, is the process which allows heterology discrimination.

We also note that the increasing heterology of the duplex DNA decreased the association between the filament and the duplex DNA more dramatically than accelerating the dissociation of the first RecA(tsDNA) complex. Within experimental uncertainty, we found that the dissociation rate constants of the first step for the fully homologous Reaction I (k₁ = 1.2 ± 0.5 s⁻¹), the partially homologous Reaction I₉ (0.2 ± 0.2 s⁻¹), and the completely heterologous Reaction Het1 (0.5 ± 0.1 s⁻¹) are not significantly different. In contrast, the association rate constants changed more than 10-fold from 0.4 x 10⁶ M⁻¹s⁻¹ to 6.5 x 10⁶ M⁻¹s⁻¹ (compare Figures 3.8B and 10B). Previous studies reported that increasing heterology leads to both slower association and faster
dissociation (152,202). We suspect that the heterology-independent dissociation rate constants may be the result of a high intrinsic affinity of the RecA filament towards the sequences employed herein (242). This may also be related to the observation of a greater fluorescence change in Reaction Het1 using H1 instead of H2.

**Energetic aspects of the reaction**

The influence of temperature on the strand exchange reaction can be interpreted in the context of an Arrhenius analysis of the rate constants and a van't Hoff analysis of the equilibrium constants. In the former case, we only observe a substantial influence of temperature on the rate constant, \( k_{+1} \), for the first forward step of the reaction (Figure 3.13A). The large, positive activation energy \( E_a \) is inconsistent with a first phase in which helix formation – either double- or triple-stranded – is occurring since the latter is characterized by \( \Delta H^\ddagger < 0 \). This conclusion is corroborated by the fact that the apparent entropy of activation is positive (\( \Delta S^\ddagger > 0 \); derived from an Eyring analysis, result not shown) while that of helix formation is accompanied by a negative activation entropy. The conclusion that the first step is not canonical helix formation is further corroborated by the observation that \( \Delta H^o \) derived from the van’t Hoff analysis of \( K_1 \) (Figure 3.13B) is also positive rather than negative.

An alternative mechanistic possibility for the first step is that it involves filament-induced melting of the substrate dsDNA. While the signs of the activation and
thermodynamic parameters are consistent with this idea, the magnitudes of $E_a$ (135 kJ·mol$^{-1}$) and $\Delta H^\circ$ (65 kJ·mol$^{-1}$) are clearly different from those required for melting a 30mer dsDNA calculated using nearest neighbor parameters (286). It remains a formal possibility that step one involves only local and partial disruption of the dsDNA.

A related possibility is that step one involves the association of dsDNA in a pre-stretched conformation. The stretched dsDNA can naturally occur during thermal fluctuation and the binding of the RecA protein could stabilize the high-energy state of the stretched, naked dsDNA. The apparent activation energy would then simply reflect the thermal energy necessary to stretch a local segment of dsDNA. Interestingly, a recent work demonstrated the preference for RecA binding to mechanically stretched dsDNA rather than canonical B-DNA (204). The demonstration that mechanical and thermal denaturation of dsDNA are processes related by common physics suggests that the models involving RecA binding to locally “melted” and locally “stretched” dsDNA are conceptually indistinguishable.

Whatever the exact mechanism of the association represented by the first step, it is clear from Figure 3.13C that the sum of the new RecA-DNA and DNA-DNA interactions formed in the first intermediate, $I_1$, do not compensate for the loss of stacking in the incoming dsDNA. Indeed, it is the influence of $\Delta S$ that favors the forward reaction at the temperatures investigated here. This observation leads to the question as to how the association between biopolymers can proceed with a positive entropy. One possible
explanation is that water molecules and/or ions are released from the interacting substrates upon association. By combining this notion with that regarding the importance of thermal dsDNA stretching, we speculate that the initial RecA nucleoprotein filament contains a pre-formed ideal binding site for extended dsDNA wherein thermally stretched DNA is bound with concomitant release of water and/or ions previously organized in the filament. The concept of an idealized, pre-organized binding site is supported by the observation that even completely heterologous dsDNA associates transiently with the filament to form I₁.

A final interesting feature of the thermodynamic profile diagrammed in Figure 3.13C is the absence, following the first step, any large changes in ΔH. Such changes would be expected to accompany both site changes associated with the putative thermodynamic driving force for strand exchange as well as cooperative DNA helix melting or formation. The absence of thermodynamic signatures revealing such molecular events suggests that they must be revealed by other observable or that they do not occur during the discrete kinetic steps.

**Structural information of intermediates inferred from the kinetics**

Based on the discussion above, we proposed a molecular model for the RecA-mediated strand exchange reaction (Figure 3.16). In this model, we speculate that the intermediate I₁ represents a pre-strand exchange intermediate, because the similar
Figure 3.16 Proposed molecular model depicting the RecA-mediated DNA strand exchange in the absence of ATP hydrolysis. The upper panel shows the cross section of the RecA filament, with the different shades indicating different binding sites. Four kinetically distinct steps based on this study are proposed. The first step, followed by three slow, isomerization steps, is a fast association event between the nucleoprotein filament and the dsDNA, in which a true collision complex may be not detected under our reaction condition. The occupation of each binding site by the DNA strands during this process is also shown. The possible structures of each intermediate are drawn in the bottom panel.
stabilities of the respective Iᵢ's formed in Reactions Iᵢ⁺ and I (Figure 3.15) suggest that substantive base-pairing between the I and C strands have not occurred yet. This conclusion is also consistent with the changes in the response factors derived from the simulations. The fluorescence of the C and I strands do not become similar until the reaction proceeds to the third intermediate I₃ (Table 3.5; Figure 3.12). In addition, judging by the magnitude of the activation energy required in the formation of Iᵢ, we speculate that, during the formation of Iᵢ from the collision complex, a sequence comparison between the I and the C⋅O strands is likely mediated by a few, transient noncovalent interactions. Given a late, Iᵢ-like transition state structure, better homologous alignment would lead to stronger transient contacts, a more stable Iᵢ (lower energy level), and a lower barrier for conversion to the first stable and detectable intermediate Iᵢ. Based on these considerations, we schematically draw the structure of Iᵢ as a RecA(tsDNA) complex in which the dsDNA makes cross contact with the RecA(I) filament only using a few base pairs (Figure 3.15). This cross contact is in consistent of the previously proposed interaction between the RecA(I) filament and its homologous dsDNA, in which three inclined base pairs of the dsDNA register with the bases of the RecA-bound ssDNA (142), and our work formally supports this possibility.

We were unable to discuss the structure of intermediate I₂ in a great detail due to the large uncertainty in the kinetic simulation (Figure 3.12 and Table 3.5). However, we speculate that intermediate I₂ should be another RecA(tsDNA) complex, in which now
all the three DNA strands are brought into registry with each other given the homology has been identified during the formation of I₁.

With respect to I₃ and P, we must consider what is known about the structure of P. Our previous work has reported that the final product in our reaction, a RecA(tsDNA) complex, is a post-exchange triple-stranded DNA with the outgoing strand shifted toward its 5’ terminus 3 or 4 nts (261). This structure is in agreement with Malkov et al. (184), and it can also explain previous failed attempts to probe a post-exchange triplex DNA or a R-DNA complex in the presence of ATPγS (158,159,185,287). In the presence of ATP, however, we suspect that the reaction may proceed directly to the final separation of heteroduplex DNA and outgoing ssDNA, either from this trapped RecA(tsDNA) complex or from an earlier intermediate, for example, I₁. This reasoning is based on the following three considerations:

First, previous kinetic work by Gumbs and Shaner showed that among the three kinetic phases they observed for the strand exchange reaction, the first two phases observed in the presence of ATP or ATPγS are nearly identical, while only the last phase shows a four-fold difference (200). This indicates the early intermediates in the reaction should be the similar to each other in the presence of ATP and ATPγS. Second, previous work demonstrated that the strand exchange reaction is reversible in the presence of ATP (201,288), but the joint molecules are easily trapped in the presence of ATPγS (288,289)
(as well as indicated by the small $K_d$ value measured in this study). In our reaction scheme, we noticed that the middle two steps (from $I_1$ to $I_2$) have comparable forward and reverse rate constants, while the reverse rate constant of the last step (from $I_2$ to $P$) is approaching zero. This indicates that the conversion from $I_2$ to $P$ is the step where the tsDNA is trapped. This should be the consequence of the arrangement of the outgoing strand in $P$ sitting off plane of the heteroduplex DNA, which naturally prohibit the reverse reaction. Lastly, in the presence of ATP hydrolysis, Jain et al. (173) detected cross-linking of all three DNA strands, while in the presence of ATP$_\gamma$S, only the heteroduplex DNA is cross-linked. This further suggests that the RecA(tsDNA) complex we detected in the absence of ATP hydrolysis may be different from that in the presence of ATP hydrolysis.

Taking all these considerations together, it is possible that when the reaction proceeds to $I_3$, the outgoing strand will dissociate from the complex upon ATP hydrolysis, while in the absence of ATP hydrolysis, the outgoing strand is trapped on its way to dissociation. Thus, it is possible that the intermediate $I_3$ is a triplex DNA with all three corresponding bases on the same base pair plane, which can only be detected in the presence of ATP hydrolysis. It may adopt the proposed post-exchange R-DNA (141) structure. Experiments are under way to test these hypothesis.
During the RecA-mediated DNA strand exchange between single-stranded (ssDNA) and homologous double-stranded DNA (dsDNA) molecules, a key intermediate comprised of three DNA strands simultaneously bound to a RecA filament (RecA·tsDNA complex) forms. In the absence of ATP hydrolysis, we have probed the helical structures of the three DNA strands of the RecA·tsDNA complex, evaluated the kinetics for the formation of this key intermediate, and presented a molecular model recapitulating the key aspects of the RecA-mediated DNA strand exchange (Figure 3.16).

We first measured the helical geometry of the three DNA strands of the RecA(tsDNA) complex using fluorescence resonance energy transfer (FRET) under physiologically relevant solution conditions. FRET donor and acceptor dyes were used to label different DNA strands, and the interfluorophore distances were inferred from energy transfer efficiencies measured as a function of the base-pair separation between the two dyes. The energy transfer efficiencies were first measured on a control RecA·dsDNA complex, and the calculated helical parameters \( h = 5 \, \text{Å}, \, \Omega_h = 20^\circ \) were consistent with structural conclusions derived from EM and other classic biochemical methods. Measurements of the helical parameters for the RecA·tsDNA complex revealed that all three DNA strands adopt extended and unwound conformations similar to those of RecA-bound dsDNA. The structural data are consistent with the hypothesis that this complex is a late, post-strand-exchange intermediate with the outgoing strand shifted by about three base pairs with respect to its registry with the incoming and complementary
strands. Furthermore, the bases of the incoming and complementary strands are displaced away from the helix axis toward the minor groove of the heteroduplex, and the bases of the outgoing strand lie in the major groove of the heteroduplex. These evidence strongly suggest that homologous contacts proceeding the strand exchange arise in the minor groove of the substrate dsDNA.

We then monitored the formation of this RecA-tsDNA complex in real time using a fluorescent base analog 6MI. Although the three DNA strands in the RecA filament after strand exchange adopt similar helical structures, the usage of this sensitive intrinsic fluorophore allowed us to characterize the DNA-DNA and protein-DNA interactions in a great detail. We obtained evidence showing that each strand is treated differently by the RecA protein according to its sequence relative to the incoming ssDNA strand, and the respective interactions among the three DNA strands and the RecA protein are unique to each other. By monitoring the time-dependent changes of polarized emission from the fluorophore, we observed at least three phases during the formation of the RecA-tsDNA complex. The first phase is strongly dependent on substrate concentration and reaction temperature. Kinetic simulation revealed a sequential four-steps mechanism as the best description for the reaction. The first step is the fast association of the nucleoprotein filament and the dsDNA substrates, followed by three slow, reorganization steps. Thermodynamic analysis indicates the reaction is entropy favorable and a large activation energy is necessary for the association step. Comparison of the energy diagrams between
a homologous, partially homologous and a completely heterologous dsDNA substrates suggests that the RecA protein discriminates homology both kinetically and thermodynamically.

Based on these studies, a structural and mechanistic model for RecA-mediated strand exchange was constructed (Figure 3.16). In this model, the first productive intermediate is likely to be the association of the RecA·ssDNA filament with a partially stretched and unwound dsDNA, with a few base pairs registering for homology identification. Homology of the dsDNA is selected both kinetically and thermodynamically against heterology during this step. In the following steps, both of the two strands of the dsDNA is uptaken into the RecA·ssDNA filament once homology is identified, and the three DNA strands are reorganized to form a product-like three-stranded DNA complex. In this complex, the complementary and incoming strands are likely to be base-paired, while the outgoing strand may still remain in the same base pair plane with the heteroduplex DNA. In the absence of ATP hydrolysis, the outgoing strand may diffuse away from the heteroduplex DNA and form the stable final post-exchange complex with the outgoing strand displaced in the major groove of the heteroduplex DNA.
REFERENCE


by 11.5 degrees/base pair in the presence of adenosine 5'-O-(3-thiotriphosphate). *J Biol Chem.* **258**, 12624-12631


134. Yu, X., and Egelman, E. H. (1992) Structural data suggest that the active and inactive forms of the RecA filament are not simply interconvertible. *J Mol Biol.* 227, 334-346


iodine 125: evidence for recognition of homology in the major groove of the target duplex. *J Mol Biol.* **299**, 629-640


