In vitro selection of preferred DNA pairing sequences by the *Escherichia coli* RecA protein

Robert B. Tracy and Stephen C. Kowalczykowski

Division of Biological Sciences, Sections of Microbiology and of Molecular and Cellular Biology, Graduate Group in Microbiology, University of California at Davis, Davis, California 95616 USA

GENES & DEVELOPMENT 10:1890–1903 ISSN 0890-9369/96 $5.00

In vitro selection of preferred DNA pairing sequences by the *Escherichia coli* RecA protein

Robert B. Tracy and Stephen C. Kowalczykowski[1]

Division of Biological Sciences, Sections of Microbiology and of Molecular and Cellular Biology, Graduate Group in Microbiology, University of California at Davis, Davis, California 95616 USA

The RecA protein and other DNA strand exchange proteins are characterized by their ability to bind and pair DNA in a sequence-independent manner. In vitro selection experiments demonstrate, unexpectedly, that RecA protein has a preferential affinity for DNA sequences rich in GT composition. Such GT-rich sequences are present in loci that display increased recombinational activity in both eukaryotes and prokaryotes, including the *Escherichia coli* recombination hotspot, χ (5′-GCTGGTGG-3′). Interestingly, these selected sequences, or χ-containing substrates, display both an enhanced rate and extent of homologous pairing in RecA protein-dependent homologous pairing reactions. Thus, the binding and pairing of DNA by RecA protein is composition-dependent, suggesting that a component of the elevated recombinational activity of χ and increased genomic rearrangements at certain DNA sequences in eukaryotes is contributed by enhanced DNA pairing activity.

[*Key Words*: Homologous recombination; RecA protein; in vitro selection; χ; GT-rich sequences]

Received May 21, 1996; revised version accepted June 26, 1996.

Genetic and biochemical studies show that the RecA protein is indispensable to the process of homologous recombination. Mutations in the *recA* gene reduce both conjugation (Clark 1973) and transduction (Clark et al. 1984) by as much as 5×10^4-fold relative to wild-type strains. The biochemical functions of RecA protein in recombination have been well characterized in vitro (for review, see Radding 1988; Kowalczykowski 1991; West 1992; Cox 1993; Kowalczykowski et al. 1994; Kowalczykowski and Eggleston 1994). RecA protein promotes the ATP-stimulated renaturation of complementary single-stranded DNA (ssDNA) and the ATP-dependent pairing and exchange of DNA strands between a variety of ssDNA and double-stranded DNA (dsDNA) substrates. DNA strand exchange occurs via a series of kinetically distinct steps: presynapsis, synapsis, and DNA heteroduplex extension. It is during the initial stages of DNA strand exchange that the RecA protein forms a presynaptic (or nucleoprotein) filament that is functional in the search for homologous DNA sequences. Though DNA strand exchange has been carefully examined in vitro, certain features remain unknown. For instance, it is not known whether the RecA protein preferentially binds to DNA structures [e.g., triplex structures (Griffith et al. 1989; Hsieh et al. 1990; Menetski et al. 1990; Rao et al. 1991; Stasiak 1992; Camerini-Otero and Hsieh 1993)] that are putative intermediates during the DNA strand exchange process or whether it has a heightened affinity for specific DNA sequences or DNA compositions which could ultimately enhance the function of RecA protein during DNA strand exchange.

The formation of a presynaptic filament requires the RecA protein to bind nonspecifically to DNA. This is necessary so that RecA protein can homologously pair any DNA sequence to its partner. Therefore, is it plausible to suggest that the RecA protein can display a bias for a specific DNA sequence or a particular set of DNA sequences? While an increased affinity has been demonstrated for the polynucleotides, polydeoxycytidylic acid and polydeoxythymidylic acid [poly(dC) and poly(dT), respectively] (McEntee et al. 1981), there is no direct evidence that RecA protein has a preferential affinity for a sequence (or sequences) with a mixed base composition. However, recent indirect evidence suggests that RecA protein could contribute to the stimulatory effect of χ in the initiation phase of homologous recombination, potentially via an interaction with χ sites (Dixon and Kowalczykowski 1995). χ sites are DNA sequences that en-

[1]Corresponding author.

Table 1. *Oligodeoxyribonucleotides*

Designation and description		Length	Sequence
SKBT0	random pool of 106-mers	106	AGGCAAGCTTGGCTGCAG(N_{70})TAGGATCCGAGTCCTCCT
SKBT1	random pool of 54-mers	54	AGGCAAGCTTGGCTGCAG(N_{18})TAGGATCCGAGTCCTCCT
SKBT13	χ at the 3′ end	54	GGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCGGCTAGGCTGGTGGGGTTG
SKBT15	complement of SKBT13	54	CAACCCCACCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACC
SKBT16	contains selected sequence 1	54	AGCTTGCATGCCTGCAGGGGCGTGTGTGGTGGTGTGCTAGGATCCCCGGGTACC
SKBT17	complement of SKBT16	54	GGTACCCGGGGATCCTAGCACACCACCACACACGCCCCTGCAGGCATGCAAGCT
SKBT18	random pool of 54-mers	54	AGGCAAGCTTGGCTGCAG(N_{18})TAGGATCCGAGTCCTCCT
SKBT19	homologous to pBT54CN1	54	TGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTA
SKBT20	homologous to pBT54CN1	54	TCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAG
SKBT21	contains selected sequence 8	54	AAGCTTGCATGCCTGCAGGGCGTTATGGTTTATGTGTAGGATCCCCGGGTACCG
SKBT22	complement of SKBT21	54	CGGTACCCGGGGATCCTACACATAAACCATAATGCCCTGCAGGCATGCAAGCTT
SKBT23	contains selected sequence 9	54	GCCTAGTGTGCCCTGCAGGCTGCGAGTGGTTTTGGCTAGGATCCCCGGGTACCG
SKBT24	complement of SKBT23	54	CGGTACCCGGGGATCCTAGCCAAAATCACTCGCAGCCTGCAGGGCACACTAGGC
SKBT25	selected sequence 1	18	GCGTGTGTGGTGGTGTGC
SKBT26	complement of SKBT25	18	GCACACCACCACACACGC
SKBT33	pared-down selected sequence 1	9	GTGTGGTGG
SKBT34	pared-down selected sequence 1	12	GTGTGGTGGTGT
SKBT35	pared-down selected sequence 1	15	GTGTGTGGTGGTGTG
SKBT36	complement of SKBT35	15	CACACCACCACACAC
SKJC6	contains three χ sites	59	GTCATTAATGCTGGTGGGCGCAGACTCGCTGGTGGTCACATGGCGGCTGGTGGCTGCAG
SKJC7	complement of SKJC6	59	GTCCTGCAGCCACCAGCCGCCATGTGACCACCAGCGAGTCTGCGCCCACCAGCATTAAT

hance homologous recombination in their vicinity 5- to 10-fold (Lam et al. 1974). Both genetic and physical analysis determined the sequence of χ to be 5′-GCTGGTGG-3′, its complement, or both strands (Smith et al. 1981). In vitro, χ stimulates recombination by acting as a regulatory element, which attenuates the 3′ strand-specific nuclease activity, but not the helicase activity, of the RecBCD enzyme (Dixon and Kowalczykowski 1991, 1993). As a consequence, an intact strand of DNA containing χ at its 3′ terminus, is available for homologous pairing by the RecA protein. This strand is more invasive than any other strand of ssDNA produced by the action of RecBCD enzyme and thus raises the question of whether χ itself is responsible for this heightened pairing activity (Dixon and Kowalczykowski 1995).

In vitro selection, which has also been termed SELEX (Tuerk and Gold 1990; Gold 1995), in vitro genetics (Szostak 1992), SAAB (Blackwell and Weintraub 1990), or CASTing (Wright and Funk 1993), is a procedure that can capitalize on small differences in protein-binding affinity to select for a rare nucleic acid sequence or structure from a very large pool of random sequences. In this paper we use in vitro selection (or SELEX) to show that RecA protein has a preference for DNA sequences that have a high GT content and as a consequence, some contain χ-like sequences. The physical basis for this bias appears to be an increased apparent ssDNA-binding affinity. Furthermore, and most significantly, when the selected sequences or χ-containing substrates were used in RecA protein-dependent joint molecule formation assays, they

A

Clone 1: 5'-AGCAGACTGTTCGCTCTTGGTCACTACCTATGTGTTAGGGTCTAATGGTGACTTAGGCGATTGTGCCTGT-3'
Clone 2: 5'-CATGCAGAATCCAGGTAGTAACTGGGAGGGAAGCAGCTTAAGTTGTGTGAGTGGCGGGCAGGATTGGCGT-3'
Clone 3: 5'-GTTGGTGTGAACGGTTTGTTGTTAGTATCGGTGGTTTTCTGGGATTAGTCTTATGTGGGGGTACTGGTAG-3'
Clone 4: 5'-TATGTGCAGATCGTTGACCTCAGAGTCTCGCAGAGCTGGTCTGTAAGTGTTGTGTGTGTTTAAGTCTGGT-3'
Clone 5: 5'-CACTGGTGTTAAGTGTGGGTTATCATGATACGTGCGGGGTGTCTTGGCGCGTTCAGACTTGAGAGCTGGG-3'
Clone 6: 5'-TACATGTCTGTCTGTCACGCATGTGTATCTCGTTCTGTTTCTATATTTTTGTTGCGGCATTTGTATAAGG-3'
Clone 7: 5'-ACGGGTGTTCGTCGTGTCGCATTCTGCTTCGGACTCCGGTGATACGTTGCGCATGTATCTGTTCCGGCCG-3'
Clone 8: 5'-TAGGACGGATTAATACACGGTAGCGTTCAGCTATGTTCTAGTGTTGGTTCGAAAAGGTGTGGCTGCTATG-3'
Clone 9: 5'-GAACATGGTCAGGTCCTAGCCAAGGTTGAGACGCATGGCCGGGACCCGTTCTCTTATGGGAGTTGGGCTC-3'
Clone 10: 5'-GGGGATCAATAGGTCAAAACTCATCGTCGAATATGCTCTTGTTAAGCTTTCTGGTTTTTGCCGTGGTGTC-3'
Clone 11: 5'-GGCAACGGCTCAGAAGGGTGGTATGACAGGATGACATAGCTGTTGTGGGTGTGTGGCGGGTGGTTGACTA-3'
Clone 12: 5'-GTACTCGTGCTTAAGCTCTTGTTTCCCGTATACTTGGTATTGGGTCCTTTATCACGCTCGTGTGATGGTC-3'
Clone 13: 5'-GTGGCTAAGGTAACATTTCACGTGTAAACCATCTCGAAGATTTAGGTTATGCGGTCGTATAGGGCCGGTC-3'
Clone 14: 5'-TAGGGATTGGTTTAGCTGGTCGGGATGTTAATCGTAACTTGCGTCGTCACTGCGCTGTCACGTCCAGGTC-3'
Clone 15: 5'-GGGAAACGGGCTAGTGTGTTGGATCTTTGATTCGCGTAGTATAAGTCTAGTGAGTGACTTAGACATCTGC-3'
Clone 16: 5'-CGTGTAGGCAGATTGCTGGGTGCGCGCGTGATGGTGTATGGTAGGACTCTCTGCTCTGTGATGCGCAGTG-3'
Clone 17: 5'-GGCTGGCTGGGTGTATAGGCCCGTTGGATGAACTAAAGTGATGGGGGGATTAATGAGACGTTGGATGGTT-3'
Clone 18: 5'-GACAATACGTGGATGGGGGTAATAGCCCAAGCTATGTTGGTGGGTAAATTCTGCAAGGATTCGCGGCGTG-3'
Clone 19: 5'-GCTGCGTAGACCAGGTGTCAGTGTCTTTGCATTTGATGCGTGTAGGTATGGTAGCTAGCGGGTTAGGCTG-3'
Clone 20: 5'-GTGGGCAGTGTCGGATAACGTAATGGGCGGGTATGAGAGAGCTATGCGCATGCTGCTGGGCCCTAGGCTG-3'
Clone 21: 5'-TGGGTCCGGCGTAATTCCCACTCTGAGCGGCATATCAGACGAGCCCATATGCCGTCGACTGTTGCCTGCC-3'
Clone 22: 5'-TATGCGGACTCCTTAGATGGTGTGTCTTGTATGGCGCCTGACAGTGTTGGTGGTGACCGTTGGTTATGCC-3'

Average %G: 33.5; %A: 17.9; %T: 30.7; %C: 17.9

B

		#G	#A	#T	#C
Clone 1:	5'-GCGTGTGTGGTGGTGTGC-3'	10 (56%)	0 (0%)	6 (33%)	2 (11%)
Clone 2:	5'-GTTCAGTGATACTGGTTG-3'	6 (33%)	3 (17%)	7 (39%)	2 (11%)
Clone 3:	5'-GATGCTGAGGGGTGTGTT-3'	9 (50%)	2 (11%)	6 (33%)	1 (5%)
Clone 4:	5'-TGTGTCTGCTATGGCTGG-3'	7 (39%)	1 (5%)	7 (39%)	3 (17%)
Clone 5:	5'-GGGATGTTGATGGGTTGG-3'	10 (56%)	2 (11%)	6 (33%)	0 (0%)
Clone 6:	5'-GCTGAGTCAGCTTTGGTC-3'	6 (33%)	2 (11%)	6 (33%)	4 (22%)
Clone 7:	5'-GGGTAAGTGTTAGTGGGC-3'	9 (50%)	3 (17%)	4 (28%)	1 (5%)
Clone 8:	5'-GGCGTTATGGTTTATGTG-3'	7 (39%)	2 (11%)	8 (44%)	1 (5%)
Clone 9:	5'-GCTGCGAGTGGTTTTGGC-3'	8 (44%)	1 (5%)	6 (33%)	3 (17%)
Clone 10:	5'-GTTAAGGCTCTGCTTTGG-3'	6 (33%)	2 (11%)	7 (39%)	3 (17%)
Clone 11:	5'-GTTCAGTTCTTTTTGTTG-3'	4 (22%)	1 (5%)	11 (61%)	2 (11%)
Clone 12:	5'-GCGTGTGTGGTAGTGTGC-3'	9 (50%)	1 (5%)	6 (33%)	2 (11%)
Clone 13:	5'-GGGCGTTATAGTTTGGTC-3'	7 (39%)	2 (11%)	7 (39%)	2 (11%)
Clone 14:	5'-TGTAGGTTGGTGTTAGTC-3'	7 (39%)	2 (11%)	8 (44%)	1 (5%)
Clone 15:	5'-CTGTGTGTGGTAGTGTGG-3'	9 (50%)	1 (5%)	7 (39%)	1 (5%)
Clone 16:	5'-TATAGCGTCTATTTGGTT-3'	4 (22%)	3 (17%)	9 (50%)	2 (11%)
Clone 17:	5'-TATGACGGTGTCTAGTTG-3'	6 (33%)	3 (17%)	7 (39%)	2 (11%)
Clone 18:	5'-ACACCTCTGGTAGCCTGG-3'	5 (28%)	3 (17%)	4 (22%)	6 (33%)
Clone 19:	5'-GTATGAGCAGGCTGGTTT-3'	7 (39%)	3 (17%)	6 (33%)	2 (11%)
Clone 20:	5'-GTAGTGATTGGAAGGCCT-3'	7 (39%)	4 (22%)	5 (28%)	2 (11%)
Clone 21:	5'-GTATTTGGTTACGTCGAC-3'	6 (33%)	3 (17%)	7 (39%)	2 (11%)
Clone 22:	5'-GAGTAGGTCTGGGTTGGC-3'	9 (50%)	2 (11%)	5 (28%)	2 (11%)
Clone 23:	5'-CTGGTATATGTTATGTGA-3'	5 (28%)	4 (22%)	8 (44%)	1 (5%)
Clone 24:	5'-GGTTGATTGATCCATTGG-3'	6 (33%)	3 (17%)	7 (39%)	2 (11%)
	Average:	38.3%	12.7%	37.3%	11.2%

Figure 1. RecA protein selects for GT-rich DNA sequences from random pools of oligonucleotides. In vitro selection was performed on a random pool of 70-mers or 18-mers as described in Materials and methods. (*A*) The sequences of the 70-mers that were selected by RecA protein after eight cycles of selection and amplification from two independent in vitro selection experiments. The bold and underlined part of each clone corresponds to a χ-like sequence (at least 5 of the 8 nucleotides found in χ). The average base composition (*bottom*) is for all of the sequences. (*B*) The sequences of the 18-mers that were selected by RecA protein after five cycles of selection and amplification from two independent in vitro selection experiments. The occurrence of each base in each sequence is shown. Additionally, the *bottom* of *B* shows the average base composition for all of the sequences.

exhibited an enhanced rate and extent of homologous pairing activity relative to their complements and to several control substrates. Therefore, we suggest that the enhanced pairing demonstrated by the selected sequences may contribute to the increased pairing activity of χ sequences and may provide an explanation for the stimulation of recombination by GT-rich sequences observed in eukaryotic organisms.

Results

In vitro selection of GT-rich DNA sequences

In vitro selection was initially performed with a pool of $\sim 10^{14}$ 106-mers (SKBT0; Table 1). The 106-mers consisted of a random internal region of 70 nucleotides and two defined regions of 18 nucleotides on the 5′ and 3′ sides of the random region. Although we were aware that we could not sample the total population of DNA molecules ($\sim 10^{42}$ different molecules) with such a large random region, we chose this length because our original intent was to determine whether the RecA protein could select for a unique DNA structure.

Using limiting concentrations of RecA protein to ensure substantial competition between available 106-mers, eight cycles of selection and PCR amplification were performed to obtain an enriched pool of oligonucleotides. Figure 1A shows the sequences of the internal 70-mer regions that were selected by RecA protein from two independent in vitro selection experiments; 22 individual clones were randomly selected for sequencing. The sequences were found to be over-represented in GT content. In addition, surprisingly, χ-like sequences (i.e., containing at least 5 of the 8 nucleotides in the χ sequence) were present in all of the clones. When the χ-like sequences from all 22 clones are aligned, the derived consensus sequence matches the known χ sequence, with a fairly high base frequency at each position (64%–82%); (data not shown). Because of the large random region involved and the GT bias present in the selected pool, it was not clear whether the RecA protein had selected these sequences on the basis of GT-richness or because of the presence of χ or χ-like sequences.

Consequently, the selection experiments were repeated using an oligonucleotide (SKBT1 or SKBT18) that

consisted of a random internal region of only 18 nucleotides and two defined regions of 18 nucleotides on the 5′ and 3′ ends (Table 1). The oligonucleotide SKBT18 is identical to SKBT1 except that it was synthesized by an independent source to verify that SKBT1 did not have a sequence bias associated with it. An 18-nucleotide random region was used for several reasons: (1) We could determine more accurately whether the RecA protein has a preference for GT-rich sequences, for χ or χ-like sequences, or for another specific DNA sequence; (2) all of the possible sequence variants could be sampled ($\sim 7 \times 10^{10}$ different sequences); and (3) 18 nucleotides represents the theoretical length bound by one complete turn of the RecA protein helical filament, which may be a more accurate representation of the binding properties of a functional RecA protein (Hsieh et al. 1992). Again, using limiting concentrations of RecA protein, five cycles of selection and amplification were performed. Figure 1B shows the RecA protein-selected 18-mers from two independent in vitro selection experiments, using either SKBT1 or SKBT18. As with the 70-mers, we sequenced a sample of all of the available clones. In addition to the 24 different sequences displayed, the sequence shown for clone 1 (selected sequence 1) was obtained 16 times. Clone 12 is identical to clone 1 except for a single-base change (G → A) at position 12, and clone 15 is similar to clone 1 except for changes at four base positions. There is no unique sequence present among all of the clones. However, like the results obtained for the prior selection experiment, these sequences also have a very high content of G and T residues (38% and 37%, respectively; Fig. 1B). In addition, 15 of the 24 clones contain segments that consist of 5 or 6 of the nucleotides present within χ.

Control reactions of the in vitro selection procedure, with the random pool of 18-mers, show no bias towards a high GT content (Table 2). Substitution of the *Escherichia coli* single-stranded DNA-binding (SSB) protein for RecA protein verified that selection of GT-rich sequences is specific to RecA protein and not a general property of nonspecific, cooperative DNA-binding proteins. Conducting the PCR reactions in the absence of RecA protein confirmed that the PCR process itself was not generating a sequence bias. Finally, the filter-binding control was performed to verify that the nitrocellulose filters themselves do not preferentially bind GT-rich sequences, which are then eluted by the procedures we used. Table 2 shows that for all of the controls, the sequences are essentially random with no over-representation of G or T residues. This demonstrates that there were no other circumstances contributing to the GT-rich sequence bias observed for RecA protein.

Table 2. *Control reactions for in vitro selection experiments*

Control reactions[a]	Average (%)[b]			
	G	A	T	C
SSB protein-selected sequences	24.4	27.2	27.7	20.7
PCR control sequences	24.2	29.7	19.8	26.3
Background (filter-binding) control sequences	17.9	29.4	21.6	31.1

[a]In vitro selection control reactions were performed using a random pool of 18-mers as described in Materials and methods.
[b]For each control reaction, the average percent of each base was determined from 10 sequences. Average percent was calculated by dividing the number of occurrences of a particular base by the total number of occurrences of all of the bases.

These results demonstrate that the RecA protein has a preference for DNA sequences that have a GT-rich composition. Because the χ sequence is GT-rich, it is not surprising that χ-like sequences are present in all of the selected 70-mers and in many of the selected 18-mers.

Over-representation of dinucleotides and trinucleotides in the selected DNA sequences

The initial characterization of the selected sequences showed that there is a preponderance of G and T residues. These sequences were examined more closely for distinguishing features. Because RecA protein binds three nucleotides per protein monomer within the functional presynaptic complex (Kowalczykowski 1991), we focused on the frequency of trinucleotide occurrence in the RecA protein-selected 18-mers (Table 3). As shown in Table 3, a particular group of trinucleotides (TGG, GTG, GTT, GGT, and TGT) are highly over-represented relative to all other trinucleotides assuming a random outcome from the selection process. In addition, they are also over-represented given the GT-richness of the selected clones. To verify that the observed frequencies are not a consequence of random chance, a χ^2 test was performed using all of the trinucleotides. The χ^2 test yielded a value of 529.60 (data not shown). Because there are 64 possible outcomes, there are 63 degrees of freedom. A χ^2 test with 63 degrees of freedom has a 0.05% chance of exceeding 106.6 (Ostle and Malone 1988). Therefore, the over-representation of exclusively GT-containing trinucleotides and the under-representation of mostly AC-containing triplets, is highly significant. In general, the same pattern of occurrence was observed with the selected 70-mers (data not shown). We also determined the frequency of occurrence of dinucleotides within the selected 18-mers (Table 3). Table 3 shows that TG and GT are highly over-represented compared with the rest of the dinucleotides and relative to the expected frequency of 6.3%. Again, the χ^2 test showed that this is not a consequence of random chance, and that it is highly significant (data not shown). When considering the GT-richness of the selected clones, these two dinucleotides are still over-represented.

RecA protein displays a heightened affinity for a selected sequence

The sequences selected by RecA protein in vitro were identified ostensibly by virtue of their preferential affinity. To verify that this was the case, competitive filter

Table 3. *Frequency of dinucleotides and trinucleotides present within the 18-mer sequences selected by RecA protein*

Trinucleotides	Frequency (%)[a]	Trinucleotides	Frequency (%)[a]
TGG	7.6	TAC	0.5
GTG	7.6	ACA	0.3
GTT	6.3	ACC	0.3
GGT	6.0	ACT	0.3
TGT	6.0	ATC	0.3
TTG	4.7	CAC	0.3
AGT	3.4	CAT	0.3
CTG	3.4	CCA	0.3
TAT	3.1	CGA	0.3
TTT	3.1	CGG	0.3
TAG	2.9	GAA	0.3
ATG	2.6	GCA	0.3
GTA	2.6	TCC	0.3
TGA	2.6	AAA	0.0
GCT	2.3	AAC	0.0
GGC	2.3	AAT	0.7
GGG	2.1	AGA	0.0
GTC	2.1	CAA	0.0
TTA	2.1	CCC	0.0
GAT	1.8	CCG	0.0
TCT	1.8	CGC	0.0
AGG	1.6	TCG	0.0
CGT	1.6	**Dinucleotides**	**Frequency (%)[b]**
GCG	1.6	TG	17.6
TGC	1.6	GT	17.4
ATT	1.3	GG	12.0
GAG	1.3	TT	10.8
AGC	1.0	TA	6.6
ATA	1.0	AG	5.6
CAG	1.0	GC	5.6
AAG	0.8	CT	5.4
CCT	0.8	AT	4.9
CTA	0.8	GA	3.9
CTT	0.8	TC	3.4
GGA	0.8	CG	2.0
TCA	0.8	AC	1.5
TTT	0.8	CA	1.5
ACG	0.5	CC	1.0
CTC	0.5	AA	0.7
GAC	0.5		
GCC	0.5		
TAA	0.5		

[a]Frequency was calculated by determining the number of times a particular trinucleotide occurred in all 24 sequences and dividing this number by the total number of trinucleotide occurrences (384). The expected frequency for a trinucleotide is 1.6% ($1/4^3 = 1/64$).

[b]Frequency was calculated by determining the number of times a particular dinucleotide occurred in all 24 sequences and dividing this number by the total number of dinucleotide occurrences (408). The expected frequency for a dinucleotide is 6.3% ($1/4^2 = 1/16$).

binding assays were performed. Clone 1 from the selected 18-mers (Fig. 1B) was synthesized (SKBT25; Table 1) and used in these assays. The complement of this selected sequence, which is AC-rich, was used as a control (SKBT26; Table 1). Figure 2 depicts the dissociation kinetics of RecA protein–oligonucleotide complexes in the presence of increasing concentrations of a competitor, $(dT)_{32}$. The complex formed with the selected sequence

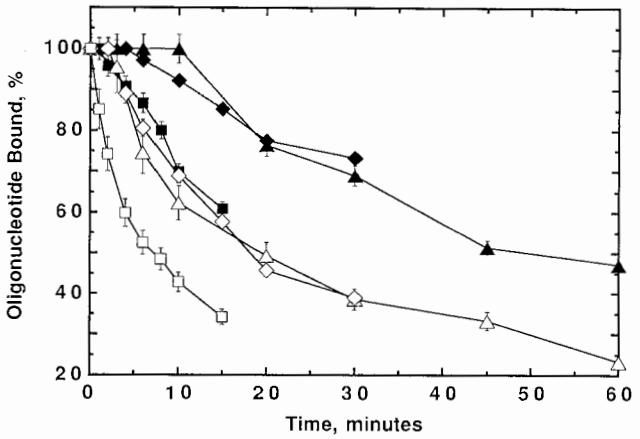

Figure 2. Preferred binding of a selected sequence by RecA protein. All reactions were conducted as described in Materials and methods. Displayed is a time course for dissociation of RecA protein-oligonucleotide complexes in the presence of increasing concentrations of the competitor $(dT)_{32}$: 2-fold excess (▲,△), 6-fold excess (◆,◇), and 10-fold excess (■,□). (▲,◆,■) Reactions performed with SKBT25 (selected sequence 1); (△,◇,□) reactions carried out with SKBT26 (complement of SKBT25). The values reported are the average of three experiments; error bars indicate the standard deviation.

is clearly more stable than that formed with the complement: RecA protein dissociates with a more than twofold faster rate from the complement relative to the selected sequence at all $(dT)_{32}$ concentrations. Furthermore, this data is substantiated by experiments conducted at other $(dT)_{32}$ concentrations (data not shown).

The slower rate of RecA protein dissociation from the selected sequence could result from a reduced rate of ATP hydrolysis when bound to this sequence, thereby making the complexes longer lived (Menetski and Kowalczykowski 1985; Kowalczykowski 1991). However, ATP hydrolysis assays demonstrated no difference in the rate of ATP hydrolysis for RecA protein with either SKBT25 or SKBT26 (data not shown), suggesting that an enhancement in the intrinsic binding affinity of RecA protein is responsible for the observed stabilization of the complex formed with the selected sequence.

The selected sequences display enhanced RecA protein-dependent joint molecule formation

The finding that RecA protein displays a preferential affinity for the selected DNA sequences raises the question of whether a similar bias is also manifest in its DNA-pairing activity. To test this possibility, the ability of RecA protein to promote invasion of supercoiled DNA by these oligonucleotides, to form joint molecules (D-loops), was measured. Figure 3 shows the pairing positions for several oligonucleotides on the plasmid pBT54CN1. Each oligonucleotide was paired with pBT54CN1 in a joint molecule formation (pairing) assay. The same 18-mer sequence (clone 1 in Fig. 1B) used in

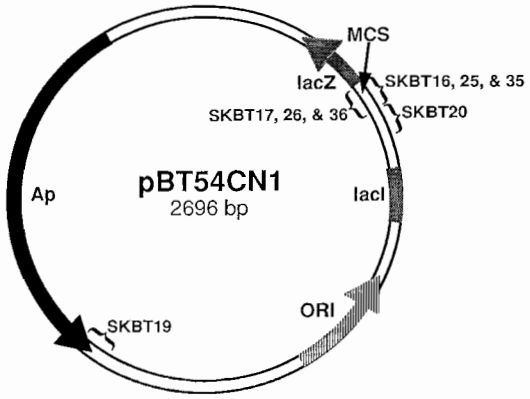

Figure 3. Location of homologous pairing sites for oligonucleotides on the plasmid pBT54CN1. pBT54CN1 is a derivative of pUC19, which contains selected sequence 1 in the multiple cloning site between the *Bam*HI and *Pst*I restriction sites. Construction of this plasmid and verification of the presence of selected sequence 1 is described in Materials and methods. The sequence of each designated oligonucleotide is shown in Table 1. (SKBT16, 25, 35) Selected sequence and its variants; (SKBT17, 26, 36) complements of selected sequence and variants; (SKBT19) distal control oligonucleotide; (SKBT20) proximal control oligonucleotide.

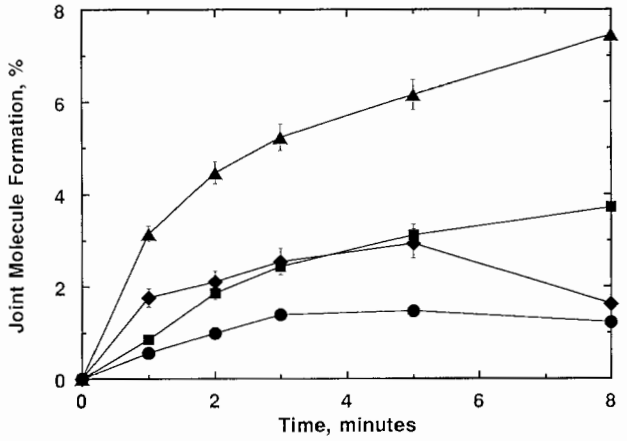

Figure 4. RecA protein-dependent joint molecule formation with oligonucleotides containing selected sequences. All joint molecule assays were performed as described in Materials and methods with pBT54CN1 and either SKBT16 (▲, contains selected sequence 1); SKBT17 (◆, complement of SKBT16); SKBT19 (●, opposite side of plasmid); or SKBT20 (■, adjacent to SKBT16) in the presence of 2 mM magnesium acetate. Data for SKBT21, SKBT22, SKBT23, and SKBT24 are not shown. Error bars indicate the standard deviation of three experiments.

the DNA competition experiments was also used in these assays. Initially, however, the 18-mer sequence was examined within the context of the flanking cloning sequences, SKBT16 (i.e., a 54-mer oligonucleotide comprised of selected sequence 1 plus 18 nucleotide cloning regions on both sides). The rationale for using a 54-mer (rather than the 18-mer) was to ensure that pairing would occur, as a minimum of 26 nucleotides is required for stable pairing once RecA protein is removed (Hsieh et al. 1992). Three different oligonucleotides were used as controls. The first is the complement of the selected sequence (SKBT17). The complement was used because it pairs in exactly the same region of the plasmid as the selected sequence and thus removes the possibility of plasmid position-dependent effects. Two other controls, SKBT19 and SKBT20, were used to verify that the region of the plasmid containing the selected sequence was not an unusual location for homologous pairing.

Figure 4 depicts RecA protein-dependent formation of joint molecules at a magnesium ion concentration of 2 mM. The maximal rate and extent of joint molecule formation by SKBT16 are increased relative to the three control 54-mers—increasing on average ~3.5- and 3.4-fold, respectively, compared to the controls. At 4 mM Mg^{2+}, the maximal rate and extent of joint molecule formation were increased ~2.9- and 1.9-fold, respectively, compared with the controls (data not shown). Studies conducted at optimal in vitro reaction conditions (10 mM Mg^{2+}) showed no differences in the pairing activity of these oligonucleotides (data not shown). This bias in pairing activity is not unique to selected sequence 1: Two other 54-mers containing selected sequences 8 and 9 (SKBT21 and SKBT23, respectively; Table 1) also demonstrated an increased rate and extent of joint molecule formation compared with their complements, SKBT22 and SKBT24, respectively (data not shown).

Because the initial oligonucleotide contained nonselected DNA sequences in the flanking regions, we wondered whether the differences might be more pronounced if only the selected 18-nucleotide sequences were examined. Consequently, selected sequence 1 (SKBT25; Table 1) was synthesized and tested. Surprisingly, despite its short length (18 nucleotides), joint molecules were formed (Fig. 5). A somewhat greater rate and extent of joint molecule formation was observed at 4 and 10 mM Mg^{2+} (data not shown). In contrast, Figure 5 also shows that when the complement of SKBT25 (SKBT26) was used, there was no joint molecule formation at 2 mM Mg^{2+}; this was also true at 4 and 10 mM Mg^{2+} (data not shown). To determine the minimal size of selected sequence 1 that could form a joint molecule, pared down versions of the sequence were tested [a 3-mer (TGG), a 6-mer (TGGTGG), a 9-mer (SKBT33), a 12-mer (SKBT34), and a 15-mer (SKBT35)]. Figure 5 shows that joint molecule formation also occurs with the 15-mer (SKBT35), although to a lesser extent than with the 18-mer. Fifteen nucleotides appears to be at or close to the minimum size for pairing in the presence of ATP, as none of the smaller oligonucleotides demonstrated joint molecule formation (data not shown). As observed for the complement of SKBT25, the complement of SKBT35 (SKBT36) also did not form joint molecules (Fig. 5). When other selected sequences (8 and 9) were tested in the joint molecule assay as 18-mers, they displayed no RecA protein-dependent joint molecule formation (data not shown). This finding suggests that there is a hierarchy of homologous pairing sequences which may also parallel RecA protein binding preferences.

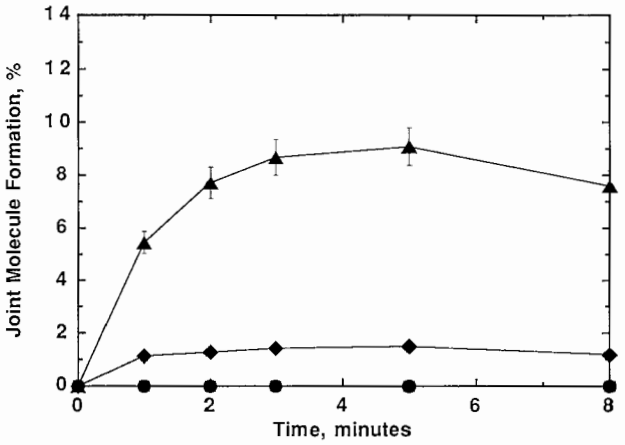

Figure 5. RecA protein-dependent joint molecule formation using a selected sequence and its complement. In all reactions, joint molecules, were formed as described in Materials and methods with pBT54CN1 and either SKBT25 [▲, selected sequence 1 (18-mer)]; SKBT26 (■, complement of SKBT25); SKBT35 [◆, selected sequence 1 (15–mer)]; or SKBT36 (●, complement of SKBT35) in the presence of 2 mM magnesium acetate. Data for SKBT33 and SKBT34 are not shown. Error bars indicate the standard deviation of three experiments.

Substrates containing the χ sequence demonstrate enhanced RecA protein-promoted joint molecule formation

Because χ-containing DNA fragments are paired preferentially in the coupled RecABCD reaction (Dixon and Kowalczykowski 1991, 1993, 1995) and because RecA protein appears to have a preference for sequences that are a subset of those within the χ sequence, we wanted to establish whether the presence of the χ sequence results in an enhancement of RecA protein-promoted joint molecule formation. This was initially examined by performing joint molecule assays with the plasmid pBR322χ$^+$3F3H and either one of two oligonucleotides, SKJC6 (Table 1), which contains three χ sites, or SKJC7 (Table 1), which is the complement of SKJC6. Figure 6A demonstrates that both the rate and extent of joint molecule formation are increased for SKJC6. The maximal rate and extent of joint molecule formation are increased ~2- and 2.1-fold, respectively, relative to its complement SKJC7. To determine whether the presence of just one χ site also enhanced joint molecule formation, SKBT13 (Table 1), which contains one χ site five nucleotides from the 3′ end, and SKBT15 (Table 1), which is the exact complement of SKBT13, were tested. SKBT13 and SKBT15 are homologous to the plasmid pBR322χ$^+$F225. SKBT13 displayed an increased rate and extent of joint molecule formation compared with its complement; however, the difference was not as pronounced as when three χ sites were present (data not shown).

In addition to studies on the effect of χ placed within oligonucleotides, we also sought to establish whether joint molecule formation could be enhanced when χ is present within substrates that have a size that is more representative of substrates encountered in vivo by RecA protein. This was investigated by carrying out joint molecule assays with the plasmid pBR322χ$^+$FH and linear single-stranded pBR322χ$^+$H containing a χ site five nucleotides from the 3′ end, or with the plasmid pBR322χ° and linear single-stranded pBR322χ°, which does not contain χ. In the presence of 2 mM Mg^{2+}, the maximal rate and extent of joint molecule formation are increased ~2.9- and 2.8-fold, respectively, when χ is at the 3′ end relative to the substrate lacking the χ sequence (Fig. 6B). When the concentration of magnesium was increased to 4 mM, the substrate containing χ at the 3′ end still displayed an enhanced rate and extent of joint molecule formation (2.0- and 2.3-fold, respectively) (data not shown).

These results demonstrate that χ sites enhance RecA protein-dependent homologous pairing. As stated earlier, however, stimulation is likely not attributable to the χ sites per se but rather, stems from the fact that RecA protein has a preference for GT-rich sequences.

Discussion

Although the RecA protein-mediated process of DNA strand exchange has been studied intensively in vitro, there are many features of the process that remain unknown. For instance, it was not known whether the RecA protein has a preference for particular DNA structures or DNA sequences that might serve to facilitate the function of RecA protein during DNA strand exchange. Using in vitro selection, we determined that RecA protein has a preference for DNA sequences that have a rich GT composition. In addition, we showed that RecA protein has a heightened affinity for at least one of the selected sequences and, most surprisingly, that these sequences display an enhanced rate and extent of joint molecule formation.

RecA protein not only displays a preference in binding to these selected sequences, but it also exhibits a bias in its DNA-pairing activity. There is a hierarchy of enhancement of homologous pairing that exists among the selected sequences. When selected sequences 1, 8, and 9 are contained within 54-mers, they all show enhanced pairing; however, when the selected sequences themselves (18-mers) were used, only selected sequence 1 displayed joint molecule formation. The fact that selected sequence 1 by itself exhibits joint molecule formation is significant in three respects: (1) It demonstrates that the enhanced pairing activity is attributable solely to the selected sequence itself; (2) it shows that nonselected sequences provide a lower background level of pairing capacity and their removal makes homologous pairing potential more conspicuous; and (3) it identifies this sequence as being unique, as it is the smallest oligonucleotide capable of participating in joint molecule formation in the presence of ATP.

These intrinsic differences in pairing activity between the various substrates became readily apparent at suboptimal reaction conditions. This observation is not com-

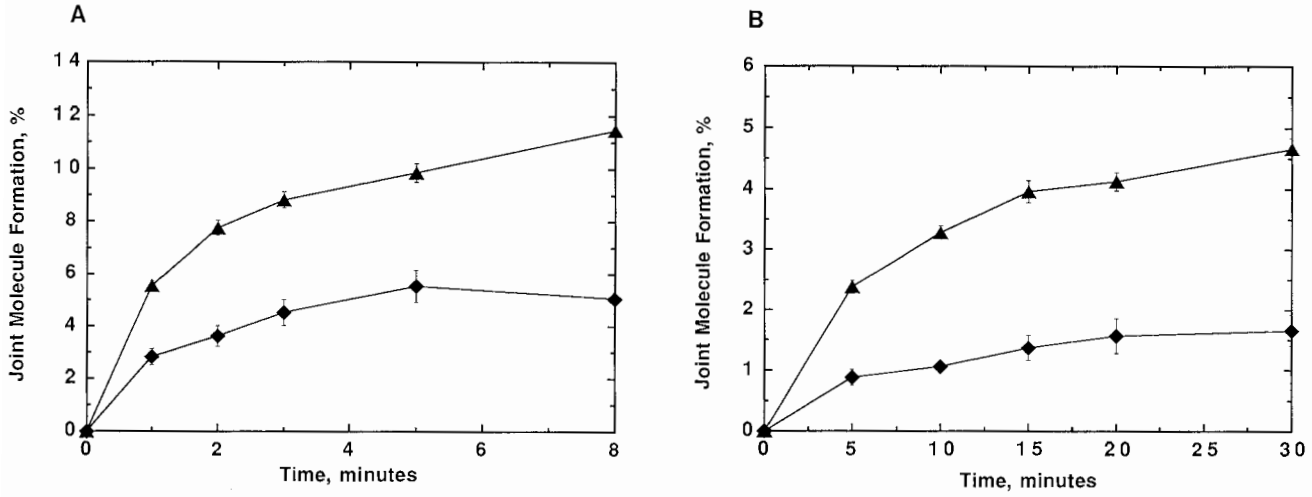

Figure 6. RecA protein-promoted joint molecule formation using χ-containing substrates. Joint molecule formation was conducted as described in Materials and methods. (*A*) Joint molecule formation with pBR322χ⁺3F3H and either SKJC6 (▲, contains 3 tandem χ sites) or SKJC7 (◆, complement of SKJC6) in the presence of 2 mM magnesium acetate. Data for SKBT13 and SKBT15 are not shown. (*B*) Joint molecule formation with either supercoiled pBR322χ⁺FH and single-stranded pBR322χ⁺H (▲, χ at the 3′ end), or supercoiled pBR322χ° and single-stranded pBR322χ° (◆, does not contain χ) in the presence of 2 mM magnesium acetate. Error bars indicate the standard deviation of two experiments.

pletely unexpected given that biological systems must often operate under suboptimal conditions to be able to both up- and down-regulate their processes. Consequently, it is very likely that the effect the selected sequences and χ have on joint molecule formation at the lower Mg^{2+} ion conditions is biologically important. Supporting this contention are two facts: (1) The intracellular free Mg^{2+} ion concentration in *E. coli* is 1–2 mM (Alatossava et al. 1985), which is the concentration of magnesium ion in vitro where we detect the most dramatic effects on joint molecule formation; and (2) the effect of χ on homologous recombination in vivo is ~5- to 10-fold, whereas our in vitro studies demonstrate an effect of 2- to 3.5-fold on joint molecule formation, potentially accounting for 20%–70% of the physiological effect. Although RecA protein-promoted DNA strand exchange in vitro occurs optimally at magnesium ion concentrations of 6–10 mM (Cox and Lehman 1982; Roman and Kowalczykowski 1986), in vivo conditions are suboptimal (1–2 mM Mg^{2+}). This view is consistent with observations that a mutant RecA protein (RecA441) has enhanced recombination activity in vivo (Lloyd 1978), but at optimal conditions in vitro (8 mM Mg^{2+}), it is no more proficient at joint molecule formation than wild-type RecA protein. However, under suboptimal conditions (2–4 mM Mg^{2+}), RecA441 protein displays increased DNA strand exchange activity relative to the wild-type protein (Lavery and Kowalczykowski 1990). Such results demonstrate that optimal pairing conditions may not be sufficiently restrictive to detect differences between the selected or the χ-containing sequences and the control substrates.

A cursory survey of the selected sequences showed that they are GT-rich. A more careful examination, however, revealed that these sequences have features in common with the genomic sequence of *E. coli*, and with the genomic sequences of several bacteriophages and many eukaryotic organisms. To determine whether the genomes of organisms display a characteristic fingerprint, Cardon et al. (1993) determined the frequency of short oligonucleotides (di-, tri-, and tetranucleotides) in two relatively long (~100 kb) contigs of *E. coli* and in the EcoSeq2 data base (>1.6 Mb) of *E. coli*. They found that the two most over-represented trinucleotides are TGG/CCA and CTG/CAG. This is interesting in light of the fact that the most frequent triplet in the selected 18-mers is TGG and the seventh most frequent is CTG (Table 3). Furthermore, the χ sequence is highly over-represented in the 68% of the *E. coli* genome now sequenced [1 in every 4.4 kbp as determined using the MOTIF computer program (Cockwell and Giles 1989) to scan the *E. coli* data base collection (Wahl and Kroger 1995)], and χ contains five different trinucleotides, four of which are among the top seven trinucleotides found within the selected 18-mers (Table 3). Finally, Burge et al. (1992) determined that TGG/CCA is the most over-represented triplet in many *E. coli*-specific bacteriophages (λ, P1, T4, and T7), as well as in many eukaryotic organisms (*Saccharomyces cerevisiae, Drosophila melanogaster, Xenopus laevis*, chicken, and human).

There also appears to be a relationship between our selected sequences and certain genomic sequences of eukaryotic organisms. When the selected 18-mers (Fig. 1B) were compared to the GenBank data base, several of the sequences (clones 1, 3, 9, 12, 15, and 22) were found to be similar to DNA sequences that tend to be recombinogenic and GT-rich: microsatellite DNA from humans, the constant region of the heavy chains of immuno-

globulins from several mammalian species, and Alu repetitive elements from humans; furthermore, many of these sequences contain χ-like sequences. Microsatellite DNA sequences, also called variable tandem repeats (VTRs), are short tandemly repeated sequences, such as GTGTGT, that are present throughout the genome of humans and other mammals (Alberts et al. 1989). They are distinguished from minisatellites in that they have a shorter repeat unit (Wright 1994). The variability in the number of repeats is believed to be attributable to unequal mitotic or meiotic exchanges between tandem arrays (Jeffreys et al. 1985), or because of DNA slippage during replication (Levinson and Gutman 1987). Although no function has been assigned to microsatellite sequences, they have been implicated as hot spots for homologous recombination in mammalian cells (Jeffreys et al. 1985; Steinmetz et al. 1986; Wahls et al. 1990). Interestingly, many human minisatellites contain a χ-like sequence, GC[A/T]GG[A/T]GG (Krowczynska et al. 1990). Similarly, immunoglobulin switch sequences contain regions that resemble χ (Kenter and Birshtein 1981). Switching of the heavy chains of immunoglobulins occurs via homologous recombination between DNA sequences (switch sequences) within the introns of the constant region. The switch sequences consist of multiple copies of tandemly repeated sequences that can range anywhere from as short as 4–6 nucleotides to as long as 40–80 nucleotides in length (Lutzker and Alt 1989). Finally, Rudiger and co-workers (1995) discovered that particular Alu elements contain a well-conserved 26-bp core sequence, within which is a portion of the *E. coli* χ sequence (5′-GCTGG-3′). Alu repetitive elements, also called short interspersed repetitive DNA elements (SINEs) or retroposon elements, are short, highly repeated DNA sequences found in introns within the satellite DNA of humans (Deininger 1989). In humans, it was demonstrated that homologous recombination between Alu elements results in genomic rearrangements (Vanin et al. 1983; Henthorn et al. 1986; Hobbs et al. 1986; Nicholls et al. 1987; Rouyer et al. 1987; Rudiger et al. 1991). Since we show that the selected sequences are enhanced for joint molecule formation and it appears that these eukaryotic sequences are putative hot spots for recombination, GT-rich sequences may stimulate recombination in general. The presence of χ-like sequences in these regions likely reflects their GT-rich composition and, hence, offers an explanation for the coincidence of χ-like sequences in recombinationally hot regions, both in prokaryotes and eukaryotes. Because eukaryotic homologs of RecA protein have been identified (Ogawa et al. 1993; Kowalczykowski and Eggleston 1994), it is now possible to ascertain whether they display a similar sequence bias.

As noted above, a potential link exists between the selected sequences and the *E. coli* recombination hotspot χ. It is necessary to point out, however, that this link likely reflects the preference of RecA protein for GT-rich sequences, rather than a specific affinity for χ sequences. Because χ is GT-rich and is highly over-represented, and because four of the five triplets found within χ are among the top seven trinucleotides found within the selected 18-mers, it is likely that χ is a more preferred site within the genome for targeted binding and subsequent homologous pairing by RecA protein. Dixon and Kowalczykowski (1991, 1995) showed, in vitro, that χ-dependent pairing products are more prominent relative to χ-independent products in RecABCD-dependent joint molecule assays. One possible contribution to this prominence is the preferential interaction of RecA protein with χ-containing ssDNA fragments. Processing by the RecBCD enzyme results in a ssDNA end containing χ at the 3′ end (Dixon and Kowalczykowski 1991; Kowalczykowski et al. 1994). Because RecA protein polymerizes onto ssDNA in a 5′ → 3′ polar fashion, the 3′ end of ssDNA is more invasive than the 5′ end (Konforti and Davis 1992). Therefore, the ultimate disposition of χ at the 3′ end of RecBCD enzyme-processed DNA provides RecA protein with the opportunity to readily bind and subsequently utilize χ for joint molecule formation. The preferred interaction of RecA protein with χ likely contributes to, but does not entirely explain, the prominence of the χ-dependent pairing products in vitro. It is also possible that the RecBCD enzyme "presents" the χ-specific ssDNA to RecA protein in such a manner that it facilitates its loading onto the DNA (Dixon and Kowalczykowski 1995). Such an enhanced loading hypothetically would allow RecA protein to compete more effectively with SSB protein for DNA-binding sites and thus coat the DNA faster (D.G. Anderson and S.C. Kowalczykowski, unpubl.). Figure 6 demonstrates that at least for the substrates used here, enhancement of pairing by χ does not require RecBCD enzyme. Thus, the enhanced recombinogenic behavior of χ-containing ssDNA in coupled RecABCD reactions contains a contribution from the greater intrinsic pairing capability of this sequence.

The fact that TGG is the most frequent triplet in the selected 18-mers, as well as being one of the most over-represented trinucleotides in the genome of *E. coli* and of many other organisms, warrants some speculation. First, perhaps RecA protein has adapted to the sequence composition of the *E. coli* genome and, therefore, particular regions of the *E. coli* genome might be more attractive for RecA protein binding and subsequent RecA protein function (i.e., homologous pairing). Second, because many eukaryotic organisms have TGG as one of the most frequently occurring triplets, it is interesting to speculate that the recombination enzymes of these organisms may also demonstrate composition-dependent DNA binding and, hence, possibly explain why particular regions of their genomes exhibit recombinational hotspot activity. Finally, although it is possible that particular regions of prokaryotic and eukaryotic genomes display recombinational hot spot activity because of the preferred binding and subsequent action of recombination enzymes, it is also possible that hot spot activity is inherent to the DNA itself. In this case, the over-representation of TGG or of other GT-rich compositions in certain regions of the genome might still help recruit recombination enzymes, yet the observed stimulation of

recombination would ultimately be intrinsic to properties of the DNA.

The molecular basis for the enhanced homologous pairing activity of the selected sequences and the χ-containing sequences is unknown. Two mechanisms could account for the enhanced pairing activity of these sequences, however. One possibility is that the initial pairing between the DNA substrates is longer lived, because the inherent stability of the complexes formed between RecA protein and the selected sequences. Therefore, the homologously paired complexes that form do not dissociate as rapidly, allowing stable joint molecules to form. As mentioned above, another possible mechanism is that the enhanced pairing is intrinsic to the DNA sequence itself and may reflect a unique conformation of the DNA (i.e., a non-B-form DNA structure) that allows a faster rate of pairing. This hypothetical structure must be asymmetric to permit the strand bias that we observe in homologous pairing. Preliminary results demonstrate that SKBT16 (contains selected sequence 1) displays enhanced joint molecule formation even in the absence of RecA protein (R.B. Tracy and S.C. Kowalczykowski, unpubl.), providing some support for this unexpected possibility. Both scenarios promote the idea that beyond the ability of RecA protein to generally increase the rate of pairing of all DNA sequences, the unique composition of the selected and the χ-containing sequences contribute to a stimulation of DNA pairing for these important sequences.

Materials and methods

Chemicals

Chemicals were reagent-grade, and solutions were prepared using Barnsted NANOpure water. ATP and ATPγS were from Pharmacia LKB Biotechnology, Inc., and Boehringer Mannheim, respectively, and were dissolved as concentrated stock solutions at pH 7.5. Concentrations of adenine nucleotides were determined spectrophotometrically by using an extinction coefficient of 1.54×10^4 M^{-1} cm^{-1} at 260 nm.

Proteins

RecA protein was purified from *E. coli* strain JC12772 (Uhlin and Clark 1981) using a modified procedure (S.C. Kowalczykowski, unpubl.) based on spermidine precipitation (Griffith and Shores 1985). Protein concentration was determined using an extinction coefficient of 2.7×10^4 M^{-1} cm^{-1} at 280 nm.

SSB protein was isolated from strain RLM727 using a protocol provided by Dr. Roger McMacken of John Hopkins University (as described in LeBowitz 1985). Protein concentration was determined using an extinction coefficient of 3.0×10^4 M^{-1} cm^{-1} at 280 nm (Ruyechan and Wetmur 1975).

DNA ligase, Klenow fragement, restriction endonucleases (*Ahd*I, *Bam*HI, *Nhe*I, *Pst*I, and *Xba*I), and T4 polynucleotide kinase were purchased from New England BioLabs. *Taq* DNA polymerase was purchased from Stratagene. T7 sequenase, version 2.0, was purchased from U.S. Biochemical Corp. The enzymes were used according to Sambrook et al. (1989) or as directed by the specific vendor.

DNA substrates

The sequences of the oligodeoxyribonucleotides used are shown in Table 1. $(dT)_{32}$ was provided by Alex Mazin from our laboratory; SKJC6 and SKJC7 were provided by Jason Churchill from our laboratory. All oligonucleotides were synthesized on a MilliGen/Biosearch 8400 Cyclone Plus DNA Synthesizer (Millipore Corp., Milford, MA) using standard solid-phase chemistry. All DNA synthesizer chemicals, including β-cyanoethyl phosphoramidites, were purchased from Millipore. Oligonucleotides were cleaved from their respective columns and then deprotected using DNA MATE reagent 1030 (Barrskogen, Inc., Bowie, MD). Oligonucleotides were cleaved from the column for 5 min at room temperature and were then deprotected at 70°C for 3 hr. After deprotection, the oligonucleotides were precipitated with DNA MATE reagent 1040 (Barrskogen, Inc., Bowie, MD). After precipitation, the oligonucleotides were resuspended in 1 ml of TE (10 mM Tris-HCl, 1 mM EDTA at pH 7.5). Oligonucleotides were gel purified using a 10% denaturing polyacrylamide gel containing 89 mM Tris base, 89 mM boric acid, and 2 mM EDTA at pH 8.0. Bands were visualized by UV shadowing, cut from the gel, sliced up, and eluted by soaking in 0.3 M sodium-acetate at pH 7.5, at 37°C for 15 hr. Oligonucleotides were extracted with one volume of phenol/chloroform/isoamyl alcohol (25:24:1), ethanol precipitated, and resuspended in TE at pH 7.5. The concentration of each oligonucleotide was determined spectrophotometrically using its respective extinction coefficient at 260 nm. The extinction coefficients, ε (M^{-1} cm^{-1}), were determined using the method described in Fasman (1975). Oligonucleotides were 5′-end labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP (6000 Ci/mmole) (New England Nuclear Research Products, Boston, MA).

The plasmids pBT54CN1, pBT54CN8, and pBT54CN9 were constructed during the cloning step of the in vitro selection protocol. Their construction is described below. The plasmids pBR322χ° and pBR322χ$^+$F225 (Smith et al. 1981) were prepared from strains S819 and S818, respectively, provided by G.R. Smith and A.F. Taylor (Fred Hutchinson Cancer Research Center, Seattle, WA). pBR322χ° does not contain a χ site, whereas pBR322χ$^+$F225 contains a χ site at position 1493–1500. The plasmid pBR322χ$^+$FH, which was constructed by D.A. Dixon (Dixon and Kowalczykowski 1993), contains χ sequences at positions 1493–1500 and 3549–3556. The plasmid pBR322χ$^+$3F3H was constructed by Jason Churchill in our laboratory. Briefly, a dsDNA oligomer containing three tandem χ sequences (5′-GCTGGTGG-3′), which were each separated by 10 nucleotides, was inserted into wild-type plasmid pBR322 between the *Ase*I (3537) and *Pst*I (3607) sites. In addition, another dsDNA oligomer containing three tandem χ sequences (also separated by 10 nucleotides) was inserted into the same plasmid at the *Ppu*MI site (1438 and 1480). This resulted in the formation of the plasmid pBR322χ$^+$3F3H. The oligonucleotides SKJC6 and SKJC7 are homologous to this plasmid at the *Ppu*MI site. The plasmids were purified by using the Qiagen Maxi Kit (Qiagen, Inc., Chatsworth, CA).

In vitro selection

The oligodeoxyribonucleotides used for in vitro selection (SKBT0, SKBT1, and SKBT18) are shown in Table 1. SKBT1 and SKBT18 are identical to each other except that SKBT18 was synthesized by an independent source, GIBCO-BRL Custom Primers (Life Technologies, Inc., Gaithersburg, MD). All three oligonucleotides were constructed such that they contained a random internal region of 70 nucleotides (SKBT0) or 18 nucleotides (SKBT1 and SKBT18), and two defined regions on the 5′

and 3′ ends that are 18 nucleotides in length. The random region contained an equimolar mixture of the 4 bases. The defined regions were used for PCR and cloning. In vitro selection with each of the oligonucleotides began with $\sim 10^{14}$ molecules. Each pool of oligonucleotides (SKBT0, SKBT1, or SKBT18) was 5′-end labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP. An aliquot of the radiolabeled oligonucleotides (200 μM) was mixed with selection buffer (25 mM Tris-acetate at pH 7.5, 4 mM Mg-acetate, 1 mM DTT) and 1 mM ATP. In vitro selection was then initiated with the addition of 2 μM RecA protein (1:100 ratio of RecA protein to nucleotides). The reaction mixture was incubated at 37°C for 10 min. After 10 min, 1 mM ATPγS [adenosine 5′-γ-(thiotriphosphate)] was added, followed by an additional incubation of 5 min. The reaction mixture (100 μl) was then applied to a nitrocellulose filter (HAWP 0025 from Millipore Corp., Milford, MA) and subsequently washed three times with 1 ml of selection buffer. The nitrocellulose filters were pretreated with 0.4 M KOH for 40 min, washed with double-distilled water for 20 min, and soaked in selection buffer for at least 30 min prior to use (Lin and Riggs 1972). The filters were washed, dried, and counted using a Betagen Betascope 603 radioisotopic analyzer; they were then counted to quantitate the amount of oligonucleotides selected by the RecA protein. For quantitation purposes, a background control (absence of RecA protein) and a 100% retention control were carried out. The retained oligonucleotides (ranged from 3% to 60%, depending on the number of cycles) were then eluted with elution buffer (selection buffer and 0.5% SDS). The eluted material was extracted once with phenol/chloroform/isoamyl alcohol (25:24:1), ethanol precipitated, and resuspended in TE at pH 7.5. The selected DNA was then amplified via asymmetric PCR reamplification (Ausbel et al. 1994) using the following primers: 5′-AGGAGGACTCGGATCCTA-3′ (primer 1) and 5′-AGGCAAGCTTGGCTGCAG-3′ (primer 2). Asymmetric PCR reamplification consists of two steps: (1) Equimolar amounts of both primers are used to generate a large quantity of dsDNA; and (2) a small amount (~1 μl) of this dsDNA is then used in a subsequent PCR reaction with only one primer to produce ssDNA. For the first step of the reaction, 100 ng of the selected oligonucleotides was mixed with *Taq* polymerase buffer, 300 μM of each dNTP, 50 pmoles of primer 1, and 50 pmoles of primer 2. The reaction mix was incubated at 94°C for 5 min (hot start). Five units of *Taq* polymerase was then added to initiate the reaction. The reaction mixture was cycled through 20 cycles at 94°C for 30 sec, 52°C for 1 min, and 72°C for 1 min. This was followed by a final extension at 72°C for 7 min. The dsDNA product from this step was extracted once with phenol/chloroform/isoamyl alcohol (25:24:1) and ethanol precipitated. About 1 μl of the dsDNA was then used in the second or reamplification step. The same reaction conditions were used in the second step, except that primer 2 (50 pmoles) was used exclusively. The ssDNA that was generated in this second step was used in a subsequent cycle of selection with RecA protein and amplification via PCR. A total of eight cycles of selection and amplification were carried out on the random pool of 106-mers (SKBT0), whereas five cycles were carried out on the random pool of 54-mers (SKBT1 or SKBT18). It should be noted that the concentration of selected oligonucleotides to begin each cycle of selection varied according to the amount of DNA amplified from the PCR reaction. The concentration ranged anywhere from 50 to 200 μM nucleotides. The RecA protein concentration was then adjusted accordingly so that it was kept at a 1:100 ratio with respect to the DNA.

After eight cycles of selection and amplification with the 106-mers or five cycles with the 54-mers, the selected sequences were then cloned into the plasmid pUC19. The selected oligonucleotides and pUC19 were cut with *Bam*HI and *Pst*I. The cut DNA fragments were then purified via agarose gel electrophoresis followed by the use of SpinBind DNA extraction units (FMC BioProducts, Rockland, ME). The purified pUC19 and selected sequences were then ligated together with T4 DNA ligase at 16°C for 4 hr. The ligated products were transformed into competent *E. coli* DH5α cells (10–25 ng was transformed). Transformants were selected owing to the presence of an ampicillin resistance gene on pUC19. A blue/white screening was employed to distinguish between those plasmids that were presumed to contain an insert and those that were not. Standard alkaline lysis minipreps were carried out on the putative clones, followed by precipitation with polyethylene glycol (PEG)-8000 to remove all of the RNA. Putative clones were then sequenced by the dideoxy-chain termination method using [α-^{32}P]dATP as the isotope. Alkaline denaturation was used to prepare the DNA for sequencing. The plasmids, designated pBT54CN1, pBT54CN8, and pBT54CN9, were constructed during the cloning step.

The three in vitro selection control reactions with the random pool of 54-mers, were performed as described for the RecA protein except for the following differences: (1) The first control used SSB protein (1:50 ratio of SSB protein to DNA) rather than RecA protein; (2) the PCR control involved reiterating the PCR reaction for five cycles and did not involve any selection; and (3) the filter-binding control was carried out in the absence of any protein.

Determination of dinucleotide and trinucleotide composition within the selected sequences

A computer program called MOTIF was employed to determine the composition of di- or trinucleotides within the RecA protein-selected sequences (Cockwell and Giles 1989). MOTIF scans sequences (either DNA or protein) for matches to user-defined motifs based on identity. Once MOTIF had tabulated the occurrence of each di- or trinucleotide, we determined the frequency (%) of each di- or trinucleotide.

Competitive filter-binding assays

Competitive filter binding assays were performed as follows: Alkaline-treated Millipore nitrocellulose filters (described in the in vitro selection protocol) were soaked in filter-binding buffer (25 mM Tris-acetate at pH 7.5, 4 mM Mg-acetate, 1 mM DTT) for at least 30 min prior to use. For the time-course measurements, reaction mixtures (800 μl) contained filter-binding buffer, 1 mM ATP, 3 μM (nucleotides) 5′-end labeled SKBT25 or SKBT26, 1 μM RecA protein, and either 6, 18, or 30 μM (nucleotides) of the competitor $(dT)_{32}$. All components of the reaction except the competitor were preincubated at 37°C for 3 min; after 3 min, a 100 μl aliquot was removed that corresponded to 0 min; $(dT)_{32}$ was then added to a concentration indicated above; 100 μl aliquots were then removed at indicated times and immediately loaded onto nitrocellulose filters under suction; after the liquid passed through the filters, the filters were washed three times with 1 ml of filter-binding buffer; finally, the filters were dried with a heat lamp. The filters were then quantitated using a Betagen Betascope 603 radioisotopic analyzer. A background control was performed in the absence of RecA protein.

Joint molecule formation assays with oligonucleotides

Each oligonucleotide was 5′-end labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase. Joint molecule formation was carried

out in a reaction mixture (250 μl) consisting of 25 mM Tris-acetate (pH 7.5), Mg-acetate concentration as indicated, 1 mM DTT, 1 mM ATP, 1 μM (nucleotides) oligonucleotides, 0.333 μM RecA protein, and 18 μM (nucleotides) plasmid DNA. The reactions were initiated by the addition of plasmid DNA, after preincubation of the other components for 1 min at 37°C. At 0, 1, 2, 3, 5, and 8 min 40-μl aliquots were removed, and added to 5 μl of a 10% SDS/0.5 M EDTA mixture, to quench the samples, and 5 μl of DNA loading buffer (0.25% bromphenol blue, 0.25% xylene cyanole, and 25% Ficoll). Samples were electrophoresed through a 1% agarose gel at 50 V for 4 hr. The gels were dried and then quantitated using a Betagen Betascope 603 radioisotopic analyzer. Control reactions were performed in the absence of either plasmid DNA or RecA protein. The percentage of joint molecules formed was determined relative to the total amount of plasmid DNA (which was limiting) present in a reaction.

Joint molecule formation assays with pBR322-sized substrates

The linear ssDNA substrate containing χ at the 3′ end was generated by cleaving 4 μg of supercoiled pBR322χ$^+$FH with *Xba*I (this puts χ one nucleotide from the 3′ end); filling-in and radiolabeling the 3′-recessed ends with Klenow fragment, cold dNTPs (dCTP, dGTP, and dTTP), and [α-^{32}P]dATP (this puts χ five nucleotides from the 3′ end on one strand); the radiolabel on the strand opposite from the χ-containing strand was cleaved off with *Ahd*I; the unincorporated radiolabeled nucleotides were removed by using an S-200 column from Pharmacia; finally, the ^{32}P-labeled DNA was denatured by heating to 95°C for 5 min followed by quick-chilling in an isopropanol–dry-ice bath. The linear ssDNA substrate lacking χ was produced in essentially the same manner except that pBR322χ° was cleaved with *Bam*HI and *Nhe*I. Joint molecule formation was performed at 37°C in a reaction mixture (250 μl) consisting of 25 mM Tris-acetate (pH 7.5), magnesium–acetate concentration as indicated, 1 mM DTT, 1 mM ATP, an ATP regenerating system (2 mM PEP and 1 U/μl of pyruvate kinase), 10 μM (nucleotides) of linear ^{32}P-labeled DNA, 6 μM RecA protein, and 20 μM (nucleotides) supercoiled plasmid DNA. The reactions were initiated by the addition of plasmid DNA, after preincubation of the other components for 2 min at 37°C. At 0, 5, 10, 15, 20, and 30 min, 40 μl aliquots were removed and added to 5 μl of a 10% SDS/0.5 M EDTA mixture and 5 μl of DNA loading buffer. Samples were electrophoresed through a 1% agarose gel at 35 V for 12 hr. The gels were dried and quantitated using a Betagen Betascope 603 analyzer. The 100% control was the amount of radiolabeled ssDNA that was present at time 0.

Acknowledgments

We thank the members of our laboratory, Dan Anderson, Piero Bianco, Joel Brockman, Deana Haddox, Frank Harmon, David Hickleton, Alex Mazin, and Bill Rehrauer, for their helpful comments and critical reading of this manuscript. This work was supported by a grant (AI-18987) from the National Institutes of Health.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Alatossava, T., H. Jutte, A. Kuhn, and E. Kellenberger. 1985. Manipulation of intracellular magnesium content in polymyxin B nonapeptide-sensitized Escherichia coli by ionophore A23187. *J. Bacteriol.* **162:** 413-419.

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J.D. Watson. 1989. *Molecular biology of the cell.* Garland Publishing, New York, NY.

Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. 1994. *Current protocols in molecular biology.* John Wiley/Greene, New York, NY.

Blackwell, T.K. and H. Weintraub. 1990. Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. *Science* **250:** 1104–1110.

Burge, C., A.M. Campbell, and S. Karlin. 1992. Over- and under-representation of short oligonucleoties in DNA sequences. *Proc. Natl. Acad. Sci.* **89:** 1358–1362.

Camerini-Otero, R.D. and P. Hsieh. 1993. Parallel DNA triplexes, homologous recombination, and other homology-dependent DNA interactions. *Cell* **73:** 217–223.

Cardon, L.R., C. Burge, G.A. Schachtel, B.E. Blaisdell, and S. Karlin. 1993. Comparative DNA sequence features in two long Escherichia coli contigs. *Nucleic Acids Res.* **21:** 3875–3884.

Clark, A.J. 1973. Recombination deficient mutants of E. coli and other bacteria. *Annu. Rev. Genet.* **7:** 67–86.

Clark, A.J., S.J. Sandler, D.K. Willis, C.C. Chu, M.A. Blanar, and S.T. Lovett. 1984. Genes of the RecE and RecF pathways of conjugational recombination in *Escherichia coli. Cold Spring Harbor Symp. Quant. Biol.* **49:** 453–462.

Cockwell, K.Y. and I.G. Giles. 1989. Software tools for motif and pattern scanning: Program descriptions including a universal sequence reading algorithm. *Comput. Appl. Biosci.* **5:** 227–232.

Cox, M.M. 1993. Relating biochemistry to biology: How the recombinational repair function of RecA protein is manifested in its molecular properties. *BioEssays* **15:** 617–623.

Cox, M.M. and I.R. Lehman. 1982. RecA protein-promoted DNA strand exchange: Stable complexes of recA protein and single-stranded DNA formed in the presence of ATP and single-stranded DNA binding protein. *J. Biol. Chem.* **257:** 8523–8532.

Deininger, P.L. 1989. SINEs: Short interspersed repeated DNA elements in higher eucaryotes. In *Mobile DNA* (ed. D.E. Berg and M.M. Howe), pp. 619–636. American Society for Micorbiology, Washington, DC.

Dixon, D.A. and S.C. Kowalczykowski. 1991. Homologous pairing in vitor stimulated by the recombination hotspot, Chi. *Cell* **66:** 361–371.

———. 1993. The recombination hotspot, Chi, is a regulatory sequence that acts by attenuating the nuclease activity of the Escherichia coli recBCD enzyme. *Cell* **72:** 87-96.

———. 1995. Role of the Escherichia coli recombination hotspot, Chi, in RecABCD-dependent homologous pairing. *J. Biol. Chem.* **270:** 16360–16370.

Fasman, G. 1975. CRC handbook of biochemistry and molecular biology, Vol 1. CRC Press, Cleveland, OH.

Gold, L. 1995. Oligonucleotides as research, diagnostic, and therapeutic agents. *J. Biol. Chem.* **270:** 13581–13584.

Griffith, J. and C.G. Shores. 1985. RecA protein rapidy crystallizes in the presence of spermidine: A valuable step in its purification and physcial characterization. *Biochemistry* **24:** 158–162.

Griffith, J., C. Bortner, G. Christeiansen, J. Register, and R. Thresher. 1989. The structure of three standed joints catalyzed by the recA protein. In *Molecular mechanisms in DNA replication and recombination* (ed. I.R. Lehman and C.C. Richardson), pp. 105–114. Wiley–Liss, New York, NY.

Henthorn, P.S., D.L. Mager, T.H. Huisman, and O. Smithies. 1986. A gene deletion ending within a complex array of repeated sequences 3′ to the human β-globin gene cluster. *Proc. Natl. Acad. Sci.* **83:** 5194–5198.

Hobbs, H.H., M.S. Brown, J.L. Goldstein, and D.W. Russell. 1986. Deletion of exon encoding cysteine-rich repeat of low density lipoprotein receptor alters its binding specificity in a subject with familial hypercholesterolemia. *J. Biol. Chem.* **261:** 13114–13120.

Hsieh, P., C.S. Camerini-Otero, and R.D. Camerini-Otero. 1990. Pairing of homologous DNA sequences by proteins: Evidence for three stranded DNA. *Genes & Dev.* **4:** 1951–1963.

———. 1992. The synapsis event in the homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of DNA. *Proc. Natl. Acad. Sci.* **89:** 6492–6496.

Jeffreys, A.J., V. Wilson, and S.L. Thein. 1985. Hypervariable "minisatellite" regions in human DNA. *Nature* **314:** 67–73.

Kenter, A.L. and B.K. Birshtein. 1981. Chi, a promoter of generalized recombination in lambda phage, is present in immunoglobulin genes. *Nature* **293:** 402–404.

Konforti, B.B. and R.W. Davis. 1992. ATP hydrolysis and the displaced strand are two factors that determine the polarity of RecA-promoted DNA strand exchange. *J. Mol. Biol.* **227:** 38–53.

Kowalczykowski, S.C. 1991. Biochemistry of genetic recombination: Energetics and mechanism of DNA strand exchange. *Annu. Rev. Biophys. Biophys. Chem.* **20:** 539–575.

Kowalczykowski, S.C. and A.K. Eggleston. 1994. Homologous pairing and DNA strand exchange proteins. *Annu. Rev. Biochem.* **63:** 991–1043.

Kowalczykowski, S.C., D.A. Dixon, A.K. Eggleston, S.D. Lauder, and W.M. Rehrauer. 1994. Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol. Rev.* **58:** 401–465.

Krowczynska, A.M., R.A. Rudders, and T.G. Krontiris. 1990. The human minisatellite consensus at breakpoints of oncogene translocations. *Nucleic Acids Res.* **18:** 1121–1127.

Lam, S.T., M.M. Stahl, K.D. McMilin, and F.W. Stahl. 1974. Rec-mediated recombinational hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity. *Genetics* 77: 425–433.

Lavery, P.E. and S.C. Kowalczykowski. 1990. Properties of recA441 protein-catalyzed DNA strand exchange can be attributed to an enhanced ability to compete with SSB protein. *J. Biol. Chem.* **265:** 4004–4010.

LeBowitz, J. 1985. "Biochemical mechanism of stand initiation in bacteriophage lambda DNA replication". Ph.D. thesis, The Johns Hopkins University, Baltimore, MD.

Levinson, G. and G.A. Gutman. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Bio. Evol.* **4:** 203–221.

Lin, S.Y. and A.D. Riggs. 1972. Lac repressor binding to non-operator DNA: Detailed studies and a comparison of equilibrium and rate competition methods. *J. Mol. Biol.* **72:** 671–690.

Lloyd, R.G. 1978. Hyper-recombination in Escherichia coli K-12 mutants constitutive for protein X sysnthesis. *J. Bacteriol.* **134:** 929–935.

Lutzker, S.G. and F.W. Alt. 1989. Immunoglobulin heavy-chain class switching. In *Mobile DNA* (ed. D.E. Berg and M.M. Howe), pp. 693–714. American Society for Microbiology, Washington, D.C.

McEntee, K., G.M. Weinstock, and I.R. Lehman. 1981. Binding of the recA protein of Escherichia coil to single- and double-stranded DNA. *J. Biol. Chem.* **256:** 8835–8844.

Menetski, J.P. and S.C. Kowalczykowski. 1985. Interaction of recA protein with single-stranded DNA. Quantitative aspects of binding affinity modulations by nucleotide cofactors. *J. Mol. Biol.* **181:** 281–295.

Menetski, J.P., D.G. Bear, and S.C. Kowalczykowski. 1990. Stable DNA heteroduplex formation catalyzed by the Escherichia coli RecA protein in the absence of ATP hydrolysis. *Proc. Natl. Acad. Sci.* **87:** 21–25.

Nicholls, R.D., N. Fischel-Ghodsian, and D.R. Higgs. 1987. Recombination at the human α-globin gene cluster: Sequence features and topological constraints. *Cell* **49:** 369–378.

Ogawa, T., A. Shinohara, A. Nabetani, T. Ikeya, X. Yu, E.H. Egelman, and H. Ogawa. 1993. RecA-like recombination proteins in eukaryotes: Function and structures of *RAD51* genes. *Cold Spring Harbor Symp. Quant. Biol.* **58:** 567–576.

Ostle, B. and L.C. Malone. 1988. *Statistics in research.* Iowa State University Press, Ames, IA.

Radding, C.M. 1988. Homologous pairing and strand exchange promoted by Escherichia recA protein. In *Genetic recombination* (ed. R. Kucherlapati and G.R. Smith), pp. 193–229. American Society for Microbiology, Washington, DC.

Rao, B.J., M. Dutreix, and C.M. Radding. 1991. Stable three-stranded DNA made by RecA protein. *Proc. Natl. Acad. Sci.* **88:** 2984–2988.

Roman, L.J. and S.C. Kowalczykowski. 1986. Relationship of the physical and enzymatic properties of Escherichia coli recA protein to its strand exchange activity. *Biochemistry* **25:** 7375–7385.

Rouyer, F., M.C. Simmler, D.C. Page, and J. Weissenbach. 1987. A sex chromosome rearrangement in a human XX male caused by Alu-Alu recombination. *Cell* **51:** 417–425.

Rudiger, N.S., P.S. Hansen, M. Jorgensen, O. Faergeman, L. Bolund, and N. Gregersen. 1991. Repetitive sequences involved in the recombination leading to deletion of exon 5 of the low-density-lipoprotein receptor gene in a patient with familial hypercholesterolemia. *Eur. J. Biochem.* **198:**107–111.

Rudiger, N.S., N. Gregersen, and M.C. Kielland-Brandt. 1995. One short well conserved region of Alu-sequences is involved in human gene rearrangements and has homology ith prokaryotic chi. *Nucleic Acids Res.* **23:** 256–260.

Ruyechan, W.T. and J.G. Wetmur. 1975. Studies on the cooperative binding of the Escherichia coli DNA unwinding protein to single-stranded DNA. *Biochemistry* **14:** 5529–5534.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Smith G.R., S.M. Kunes, D.W. Schultz, A. Taylor, and K.L. Triman. 1981. Structure of Chi hotspots of generalized recombination. *Cell* **24:** 429–436.

Stasiak, A. 1992. Three-stranded DNA structure; is this the secret of DNA homologous recognition? *Mol. Microbiol.* **6:** 3267–3276.

Steinmetz, M., D. Stephan, and K. Fischer Lindahl. 1986. Gene organization and recombinational hotspots in the murine major histocompatibility complex. *Cell* **44:** 895–904.

Szostak, J.W. 1992. In vitro genetics. *Trends Biochem. Sci.* **17:** 89–93.

Tuerk, C. and L. Gold. 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249:** 505–510.

Uhlin, B.E. and A.J. Clark. 1981. Overproduction of the Escherichia coli recA protein without stimulation of its proteolytic activity. *J. Bacteriol.* **148:** 386–390.

Vanin, E.F., P.S. Henthorn, D. Kioussis, F. Grosveld, and O. Smithies. 1983. Unexpected relationships between four large deletions in the human β-globin gene cluster. *Cell* **35:** 701–709.

Wahl, R. and M. Kroger. 1995. ECDC–a totally integrated and interactively usable genetic map of Escherichia coli K12. *Microbiol. Res.* **150:** 7–61.

Wahls, W.P., L.J. Wallace, and P.D. Moore. 1990. Hypervariable minisatellite DNA is a hotspot for homologous recombination in human cells. *Cell* **60:** 95–103.

West, S.C. 1992. Enzymes and molecular mechanisms of genetic recombination. *Annu. Rev. Biochem.* **61:** 603–640.

Wright, J. 1994. Mutation at VNTRs: Are minisatellites the evolutionary progeny of microsatellites? *Genome* **37:** 345–347.

Wright, W.E. and W.D Funk. 1993. CASTing for multicomponent DNA-binding complexes. *Trends Biochem. Sci.* **18:** 77-80.