doi:10.1006/jmbi.2000.3861 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2000) 300, 469–479

JMB

A Novel, 11 Nucleotide Variant of χ, χ*: One of a Class of Sequences Defining the *Escherichia coli* Recombination Hotspot χ

Deana A. Arnold[2], Naofumi Handa[3], Ichizo Kobayashi[3] and Stephen C. Kowalczykowski[1,2]*

[1]*Sections of Microbiology and of Molecular and Cellular Biology*

[2]*Genetics Graduate Group University of California at Davis, Davis, CA 95616, USA*

[3]*Department of Molecular Biology, Institute of Medical Science, University of Tokyo Shiroganedai, Minato-ku Tokyo 108-8639, Japan*

**Corresponding author*

In wild-type *Escherichia coli*, recognition of the recombination hotspot, χ (5′-GCTGGTGG-3′), by the RecBCD enzyme is central to homologous recombination. However, in the *recC** class of RecBCD mutants, stimulation of recombination by the canonical χ sequence is not detectable, but the levels of homologous recombination are nearly wild-type. *In vivo* studies demonstrate that a member of this class of mutants, the *recC1004* allele, encodes an enzyme that responds to a novel variant of χ, termed χ* (5′-GCTGGTGCTCG-3′). Here, we establish that, *in vitro*, the χ* sequence is recognized more efficiently by the RecBC^{1004}D enzyme than is the wild-type χ. This is manifest by both a greater modification of nuclease activity and a higher stimulation of RecA protein-mediated joint molecule formation at χ* than at χ. Sequencing of the *recC1004* gene revealed that it contains a frameshift mutation, which results in a replacement of nine of the wild-type amino acid residues by eight in the mutant protein, and defines a locus that is important for the specificity of χ-recognition. In addition, we show that this novel, 11 nucleotide χ* sequence also regulates the wild-type RecBCD enzyme, supporting the notion that variants of the canonical χ constitute a class of sequences that regulate the recombination function of RecBCD enzyme.

© 2000 Academic Press

Keywords: RecBCD enzyme; χ (chi); RecA protein; genetic recombination; recombination hotspot

Introduction

The main pathway of homologous recombination in *Escherichia coli* is the RecBCD-pathway. Recombination by this pathway requires the concerted action of the DNA strand-exchange protein, RecA, and the RecBCD enzyme, which processes double-stranded DNA (dsDNA) to produce a RecA protein-accessible single-stranded DNA (ssDNA) substrate (for reviews, see Eggleston & West, 1997; Kowalczykowski *et al*., 1994; and see Myers & Stahl, 1994). Specific recombination hotspots, known as χ sites (χ = 5′-GCTGGTGG-3′), stimulate recombination *via* the RecBCD pathway both by regulating the nuclease activities of the RecBCD enzyme, and by coordinating the activities of the RecA and RecBCD proteins (Anderson & Kowalczykowski, 1997a; Dixon & Kowalczykowski, 1993; Kobayashi *et al*., 1982; Lam *et al*., 1974; Ponticelli *et al*., 1985; Smith *et al*., 1981; Stahl *et al*., 1975; Stahl & Stahl, 1977; Taylor *et al*., 1985).

The first step in the initiation of recombination at χ involves a modification of the nuclease activities of RecBCD enzyme (Figure 1(a)). The RecBCD enzyme unwinds duplex DNA from blunt, or nearly blunt, dsDNA ends. Prior to χ-recognition, the enzyme displays a vigorous 3′ to 5′ nuclease activity, which preferentially degrades the ssDNA that is 3′-terminal at the entry site of the enzyme, defined as the top-strand (Figure 1(a)) (Dixon & Kowalczykowski, 1991, 1993; Kowalczykowski *et al*., 1994). Upon recognition of a χ site in the correct orientation (i.e. from the 3′-side of the sequence 5′-GCTGGTGG-3′) (Taylor *et al*., 1985),

Abbreviations used: ssDNA, single-stranded DNA; dsDNA, double-stranded DNA; SSB protein, single-stranded DNA-binding protein; HMW, high molecular weight.

E-mail address of the corresponding author: sckowalczykowski@ucdavis.edu

0022-2836/00/030469–11 $35.00/0

© 2000 Academic Press

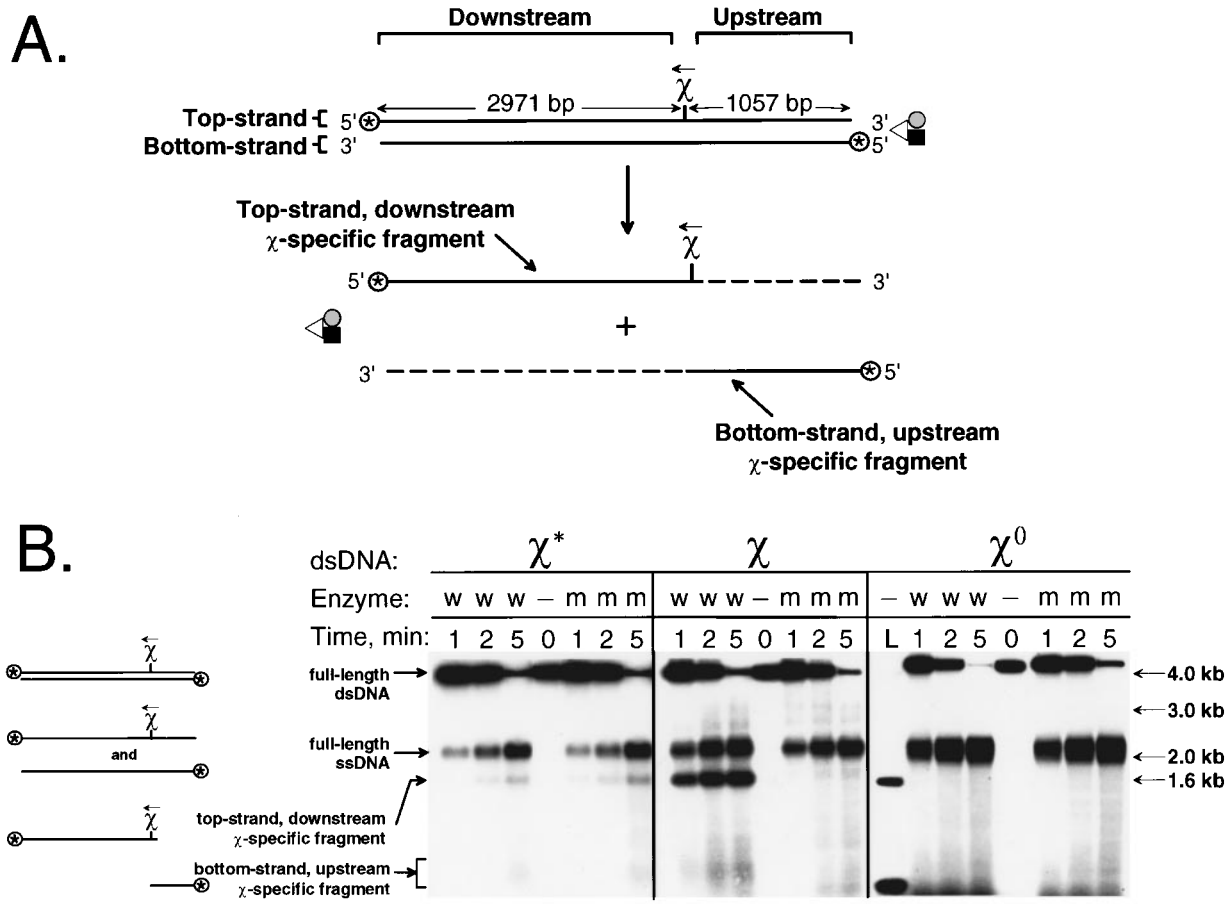


Figure 1. χ* regulates the nuclease activities of RecBCD and RecBC^{1004}D enzymes to a similar level. (a) A schematic for the production of both top-strand, downstream and bottom-strand, upstream χ-specific fragment production by wild-type RecBCD enzyme. (b) A linear substrate, containing χ* or χ , or devoid of χ (χ^0), was 5′-end labeled and added to the reaction mix as described in Materials and Methods. The reactions were initiated by the addition of a subsaturating amount (0.1 nM, functional) of RecBCD (w) or RecBC^{1004}D enzyme (m), and aliquots of these reactions were subjected to 1% (w/v) agarose gel electrophoresis. A typical gel is shown with the relevant bands identified to the left. The lane representing the zero time-point contains one-half the reaction volume of the other lanes. The lane designated L contains a 5′-end labeled dsDNA molecular mass marker (Gibco BRL), and the relative sizes of each fragment are identified to the right.

there is both an attenuation of the 3′ to 5′ nuclease activity and an activation of a weaker 5′ to 3′ nuclease activity on the 5′-terminal strand relative to the entry site of the enzyme, designated the bottom-strand (Anderson & Kowalczykowski, 1997a; Dixon & Kowalczykowski, 1991, 1993). Further, we define the dsDNA region that is processed by the enzyme prior to χ-recognition (to the right in Figure 1(a)) as the upstream region, and the region processed after χ (to the left in Figure 1(a)) as the downstream region. Thus, for small, plasmid-sized substrates, the continued unwinding and degradation after χ-recognition results in the production of both a top-strand, downstream χ-specific fragment containing χ at the 3′ terminus, as well as a bottom-strand, upstream χ-specific fragment (Figure 1(a)). The RecBCD enzyme also facilitates the loading of RecA protein onto the top-strand, downstream χ-specific fragment, thus explaining the preferred use of this substrate in RecABCD protein-mediated joint molecule formation assays (Anderson & Kowalczykowski, 1997b; Dixon & Kowalczykowski, 1991). Indeed, conjugal and transductional recombination occur preferentially in the vicinity of χ sites in the *E. coli* genome, and this evidence together with the *in vitro* work described above suggest that nearly all recombination events involving RecBCD enzyme are mediated by χ sequences (Dower & Stahl, 1981).

There is, however, a basal level of RecBCD-dependent recombination in the complete absence of the canonical χ (Ennis *et al.*, 1987; Lam *et al.*, 1974; Holbeck & Smith, 1992). This phenomenon is particularly apparent in one specific class of *recBCD* mutants, the *recC** class. The *recC** strains (alleles 1001 through 1004) were isolated as pseudorevertants of a recombination-deficient strain, *recC73*, and exhibit varying degrees of recombination proficiency and resistance to DNA damage, without enhancement of recombination in response

to χ (Schultz *et al*., 1983). The way in which these mutants, as well as the wild-type RecBCD enzyme, promote recombination in the absence of χ is not clearly understood, but interactions with novel variants of the canonical χ sequence are likely to play a role. In support of this hypothesis, the wild-type RecBCD enzyme is able to recognize some single base-pair mutations of χ, and to stimulate recombination at levels up to 40 % of the canonical χ-activity (Cheng & Smith, 1984, 1987; Schultz *et al*., 1981). Thus, the basal levels of RecBCD-dependent, χ-independent recombination in phage λ crosses can be attributed to the presence of variants of canonical χ, suggesting that these sequences may also be contributing to the overall levels of generalized recombination in *E. coli*.

The high levels of recombination displayed by the *recC** class suggested that it could represent a set of mutants that possess altered χ-recognition, responding to one or more non-canonical χ sequences at levels comparable to the RecBCD-canonical χ interaction. To investigate this possibility, a screen was designed to search for sequences that regulate the nuclease activity possessed by one of the RecBC*D enzymes, RecBC^{1004}D (Handa *et al*., 1997). Based on an *in vivo* assay that measures protection of DNA from degradation by RecBCD enzyme in the presence of χ (Dabert *et al*., 1992), it was demonstrated that RecBC^{1004}D enzyme does not efficiently recognize the canonical χ site. However, it does display a significant response to a variant of χ (χ* = 5′-GCTGGTGCTCG-3′) (Handa *et al*., 1997). The presence of this sequence in a rolling circle replication plasmid permitted the accumulation of high molecular weight (HMW) plasmid DNA, which is indicative of an attenuation of nuclease activity; further, this sequence promoted an increase in *red gam* λ phage burst size, suggesting that it could stimulate recombination in *recC1004* cells as well. However, although nuclease attenuation was clearly elicited by the χ* sequence, it was not clear from that study whether this sequence could stimulate recombination, and thus function in all respects as a χ site.

Previous *in vitro* analyses of purified RecBC^{1004}D enzyme revealed that it possesses wild-type levels of DNA helicase, dsDNA nuclease and ATPase activities, as well as modestly enhanced ssDNA exo- and endonuclease activities (Arnold *et al*., 1998). Further, we found that the mutant enzyme displays only a very weak recognition of canonical χ sites. However, the significant reduction in canonical χ-recognition, as compared to the wild-type enzyme, is not accompanied by an equivalent reduction in recombination proficiency, suggesting that the recognition of non-canonical sites must compensate for the reduced recognition of canonical χ in the *recC1004* strain (Arnold *et al*., 1998; Schultz *et al*., 1983).

In this study, we investigate the behavior of the χ* sequence *in vitro*, focusing on two fundamental characteristics that define the wild-type RecBCD enzyme-χ interaction. First, we examined whether χ* modifies the nuclease activities of both the RecBCD and RecBC^{1004}D enzymes. Second, an *in vitro* recombination reaction was utilized to measure the stimulation of RecA protein-mediated joint molecule formation at χ*. We find that the χ* site does indeed function in all respects like the canonical χ sequence and is able to stimulate recombination *in vitro* in a manner analogous to, although less efficient than, that of the canonical χ sequence. In addition, we examined the activity of single base-pair mutations of χ* and found that they also function as weak χ sites, although the mutations do reduce their activity as compared to the χ* site. These findings contribute to the growing appreciation that the *E. coli* χ is not a unique sequence, but rather a class of DNA sequences of varying effectiveness, and that the members of this class are characterized by their ability to regulate genetic recombination by stimulating biochemical modifications within the RecBCD enzyme.

Results

The χ* sequence modifies the nuclease of both the RecBCD and RecBC^{1004}D enzymes

Previously, we showed that the RecBC^{1004}D enzyme undergoes χ-specific modification of nuclease activity at a frequency approximately 0.05 that of wild-type RecBCD enzyme (Arnold *et al*., 1998). To determine biochemically whether χ* mediates the same nuclease modifications of the RecBCD and RecBC^{1004}D enzymes as does χ, we compared χ and χ*-specific fragment production (Figure 1(a)). For this, a linear dsDNA substrate containing χ, χ*, or no χ site (χ^0) was 5′-end labeled and treated with a subsaturating amount of RecBCD or RecBC^{1004}D enzyme.

Since the nuclease activities of the wild-type and mutant enzymes are highly sensitive to the concentration of free magnesium ions *in vitro*, a range of conditions was tested to determine those optimal for χ*-specific fragment production, as described (Arnold *et al*., 1998). Maximum χ*-specific fragment production occurs at 2 mM magnesium acetate and 1 mM ATP, the same conditions that promote optimal χ-specific fragment production for both enzymes (Arnold *et al*., 1998) (our unpublished results). Thus, all fragment production assays quantified in this work were performed at these conditions.

Figure 1(b) shows that, for the substrate containing χ*, both top-strand, downstream and bottom-strand, upstream χ*-specific fragments are produced by the wild-type and mutant enzymes. The average amounts of top-strand, downstream χ*-specific fragment produced in ten independent experiments were 5.3 % and 5.8 % for the wild-type and mutant enzymes, respectively. These values represent a sevenfold decrease for wild-type enzyme, and a fourfold increase for mutant enzyme, when compared to the χ-specific recog-

nition efficiencies reported previously for both of these enzymes (Figure 1(b); and see Arnold *et al.*, 1998). The wild-type and mutant enzymes also produced similar amounts of bottom-strand, upstream χ*-specific fragments (4% and 3%, respectively). The production of both downstream and upstream χ*-specific fragments by the RecBCD and RecBC^{1004}D enzymes indicates that interaction with χ* promotes both an attenuation of the 3′ to 5′ nuclease activity and an activation of the 5′ to 3′ nuclease activity. In addition, RecBC^{1004}D enzyme is reversibly inactivated by the canonical χ (Arnold *et al.*, 1998) as well as by χ* (our unpublished results). Consequently, at least with respect to nuclease activities, the response of both the wild-type and mutant enzymes to χ* is analogous to that elicited by the canonical χ sequence.

These assays were also performed with dsDNA either containing a canonical χ site, or devoid of χ (χ^0) (Figure 1(b)). The RecBC^{1004}D enzyme displays only a weak recognition of χ, as evidenced by the meager production of χ-specific fragments (1.5%), as previously reported (Arnold *et al.*, 1998). In contrast, the RecBCD enzyme exhibits χ-dependent modification of nuclease activity at much higher levels (35.3%). Quantification of the amounts of top-strand, downstream χ or χ*-specific fragment produced in reactions containing either the χ or χ* substrate, respectively, are shown in Figure 2(b). There is no specific fragment produced when the χ^0 substrate is processed by the wild-type or mutant enzyme, demonstrating that the fragments produced from the χ and χ*-containing dsDNA are specific to the χ and χ* sequences, respectively.

The four additional nucleotides of χ* are essential for regulation of RecBCD and RecBC^{1004}D nuclease activity

Previously, it was shown that variants of χ* with base-pair changes at each of the last four positions of the sequence (Table 1) are unable to increase λ phage plaque size or burst size in the *recC1004* strain (Handa *et al.*, 1997). To determine whether these base-pairs are required for recognition *in vitro*, we measured downstream χ*-specific fragment production by both wild-type and mutant enzymes (Figure 2(a)). The data show that all four of the novel nucleotides are essential for maximal χ* activity. For the RecBC^{1004}D enzyme, the production of downstream χ*-specific fragments is reduced by three- to tenfold relative to χ* for the base-pair changes occurring at positions 9, 10 and 11, and there is no detectable recognition of the variant with an A in the eighth position. For the RecBCD enzyme, these same variants of χ* displayed similar reductions in recognition efficiency (two- to eightfold), and the lowest level of recognition is observed when A is in the eighth position. These results demonstrate that the four bases on the 3′-end of the χ* sequence are essential for optimal recognition of this sequence by both the RecBCD and RecBC^{1004}D enzymes.

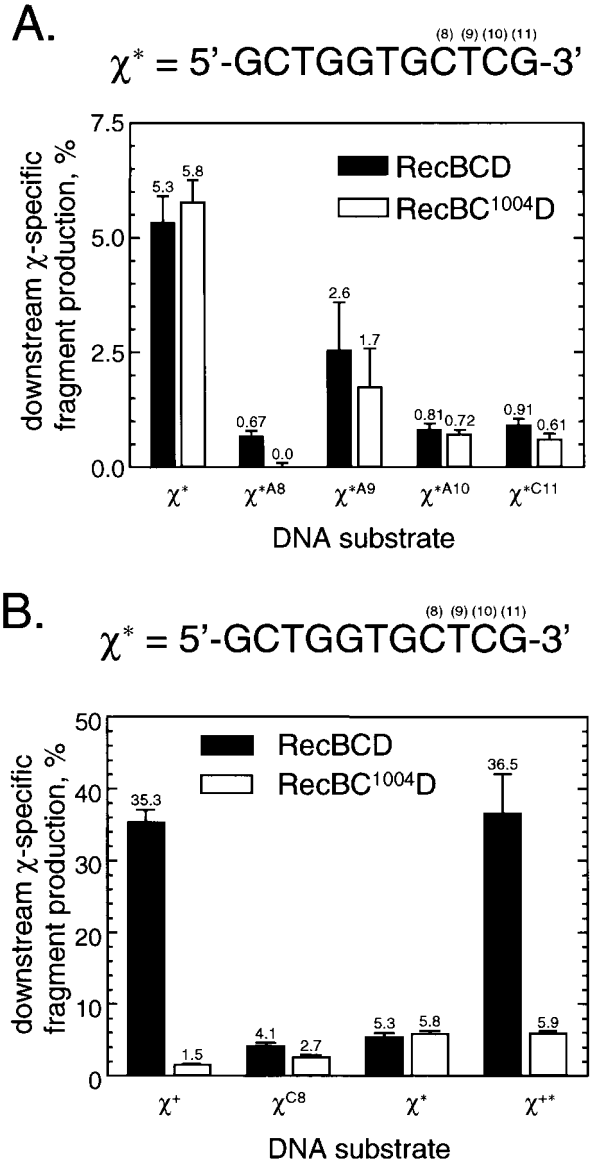


Figure 2. The base-pairs immediately upstream of χ affect recognition by the wild-type and mutant enzymes. Assays were performed to measure specific fragment production in response to variants of χ, and to compare this response to that observed at (a) χ* and (b) χ. The average amounts of specific downstream fragments produced by RecBCD (filled bars) and RecBC^{1004}D (open bars) enzymes are shown in the bar graph. Each bar represents at least three independent experiments. The values for χ are averages of our previous work (Arnold *et al.*, 1998) and current experiments performed in parallel with the other DNA substrates.

Since RecBCD enzyme recognizes an octameric χ sequence, the possibility existed that the mutant and wild-type enzymes were recognizing two overlapping octamers, thus creating an apparent 11-mer variant. If this were true, then mutations at positions that are not shared between the two octamers would affect recognition of only one of them. Since mutations at positions 9, 10 and 11 are only

Table 1. Nucleotide sequences of χ and variants of χ

Variants of χ	Nucleotide sequences
χ^{+}	3′-AGCTT**GCTGGTGG**CAGCTG-5′
χ^{*}	3′-AGCTT**GCTGGTGCTCG**CAGCTG-5′
χ^{*A8}	3′-AGCTT**GCTGGTGATCG**CAGCTG-5′
χ^{*A9}	3′-AGCTT**GCTGGTGCACG**CAGCTG-5′
χ^{*A10}	3′-AGCTT**GCTGGTGCTAG**CAGCTG-5′
χ^{*C11}	3′-AGCTT**GCTGGTGCTCC**CAGCTG-5′
χ^{+*}	3′-AGCTT**GCTGGTGGTCG**CAGCTG-5′
χ^{C8}	3′-AGCTT**GCTGGTGC**CAGCTG-5′

The variants of χ and flanking sequences are shown in bold and non-bold print, respectively. The χ* positions that are underlined represent variations as compared to the canonical χ. The underlined positions for the other variants of χ represent variations as compared to χ*.

within the ''second'' putative octamer, then the reduced recognition of these variants of χ*, as compared to χ* (Figure 2(a)), would have to be solely due to a decrease in the recognition of the second octamer. The variant of χ* with the lowest level of recognition by the mutant enzyme is χ^{*C11} (0.6 %). Therefore, if recognition of χ* were due to two overlapping octamers, the putative second octamer must possess a recognition efficiency of at least 5.2 %, and the first potential octameric sequence would contribute 0.6 % or less to the total 5.8 % recognition efficiency of χ* by RecBC^{1004}D enzyme. To test whether this prediction was true, we measured recognition of the first eight nucleotides of χ*, which is a χ site with a cytosine (C) residue in position 8 (χ^{C8}). We found that although the recognition of χ^{C8} by the RecBC^{1004}D enzyme is reduced (2.7 %) as compared to recognition of the χ* sequence (5.8 %), it is still significantly higher than the 0.6 % value predicted if χ* were comprised of two overlapping and independent octamers (Table 1; Figure 2(b)). The same line of reasoning holds true for the wild-type enzyme. These results confirm the original characterization of the χ* sequence as an 11-mer.

We also tested a variant of χ* with G in position 8, which restores the canonical χ sequence, to see if recognition of the canonical sequence plus the extended three nucleotides of χ* (''χ^{+*}'') would surpass recognition of χ by the RecBCD and RecBC^{1004}D enzymes (Table 1). Production of χ-specific fragments occurred at levels similar to recognition at χ for the wild-type RecBCD enzyme; the mutant enzyme, however, showed a fourfold enhancement of recognition over that observed at χ, a level that was comparable to that seen at χ* (Figure 2(b)). Therefore, these experiments show that, although the mutant enzyme exhibits no strong sequence preference for C over G in position 8 of χ*, positions 9-11 of χ* do enhance recognition of χ by the mutant RecBC^{1004}D enzyme. On the other hand, recognition of χ by RecBCD enzyme is not enhanced, indicating that the eight canonical nucleotides of χ are sufficient for maximal recognition by the wild-type enzyme.

The RecBC^{1004}D enzyme promotes the formation of RecA protein-dependent joint molecules in response to the χ* sequence

We have shown that the χ* sequence can modify the nuclease activity of the RecBCD and RecBC^{1004}D enzymes *in vitro*, a modification that represents the first step of recombination initiation. However, while this initial processing of the duplex DNA substrate is necessary, it is not sufficient for stimulation of recombination at χ (Anderson & Kowalczykowski, 1997b). The question remained as to whether the χ* sequence could also direct the loading of RecA protein by RecBCD or RecBC^{1004}D enzyme onto the χ*-containing ssDNA. This loading results in the enhancement of RecA protein-mediated strand invasion of supercoiled DNA by homologous, linear ssDNA *in vitro*. Since RecBCD enzyme facilitates loading of the RecA protein specifically onto the downstream χ-specific fragments, this ssDNA product is preferentially incorporated into the RecA protein-promoted joint molecules. To determine if this crucial step in the initiation of recombination could occur, the formation of χ*-dependent joint molecules was examined.

Typical joint molecule formation assays with the wild-type and mutant enzymes are shown in Figure 3(b). The production of both χ*-dependent and χ*-independent joint molecules is clearly visible for both wild-type and mutant enzymes. As observed with DNA containing wild-type χ (Arnold *et al.*, 1998), the χ*-dependent joint molecules appear earlier and predominate over χ*-independent joint molecules. To determine whether there is a preferential incorporation of χ*-specific ssDNA over full-length ssDNA into joint molecules, the yield of joint molecule formation was determined and normalized to the total amount of each ssDNA species produced. As shown in Figure 3(c), the maximum yields of χ*-dependent joint molecules for RecBC^{1004}D and RecBCD enzymes were found to be 49 % and 45 %, respectively, with the RecA protein-mediated incorporation of χ*-specific fragments into joint molecules approximately threefold higher than incorporation of full-length ssDNA. This preferential utilization of the χ*-specific fragments is comparable to that reported previously for χ-specific fragments, although the overall yield of χ*-dependent joint molecules, as compared to that of χ-dependent joint molecules, is reduced for the wild-type enzyme and increased for the mutant enzyme, reflecting the relative differences in the χ and χ*-recognition efficiencies of these enzymes (Anderson & Kowalczykowski, 1997b; Arnold *et al.*, 1998). We also found that the χ^{*A9}, χ^{*A10}, and χ^{*C11} sequences stimulate χ-dependent joint molecule formation in the presence of either enzyme, whereas DNA lacking a variant of χ* does not (our unpublished results; Dixon & Kowalczykowski, 1991). As expected, based on the low recognition efficiencies for these variants of χ*, the absolute

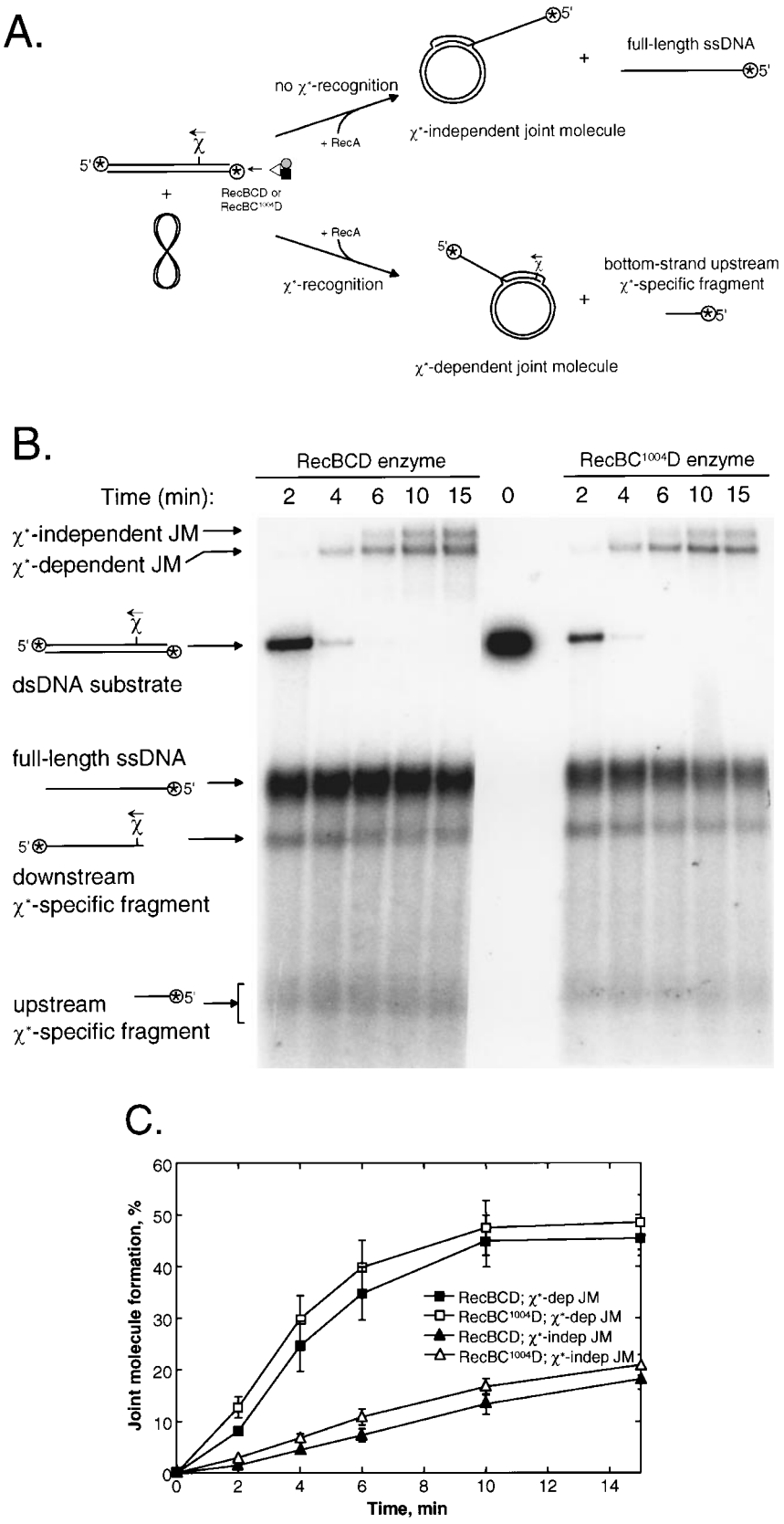


Figure 3. The incorporation of χ*-containing ssDNA into joint molecules by RecA protein is threefold more efficient than incorporation of full-length ssDNA for both the wild-type and mutant enzymes. (a) A linear, 5′-end labeled χ* dsDNA substrate is processed by RecBCD or RecBC^{1004}D for use in homology-dependent strand invasion of a supercoiled DNA molecule by RecA protein. (b) Aliquots taken from ongoing reactions are subjected to 1% (w/v) agarose gel electrophoresis. (c) By normalizing joint molecule formation to the amount of appropriate ssDNA species present at that time-point (in that lane), we show that the stimulation of χ*-dependent over χ*-independent joint molecule formation is similar for both the wild-type and mutant enzymes.

yields were reduced. Since the mutant enzyme is unable to generate detectable levels of χ^{*A8}-specific fragments, only wild-type enzyme allowed the production of χ^{*A8}-dependent joint molecules.

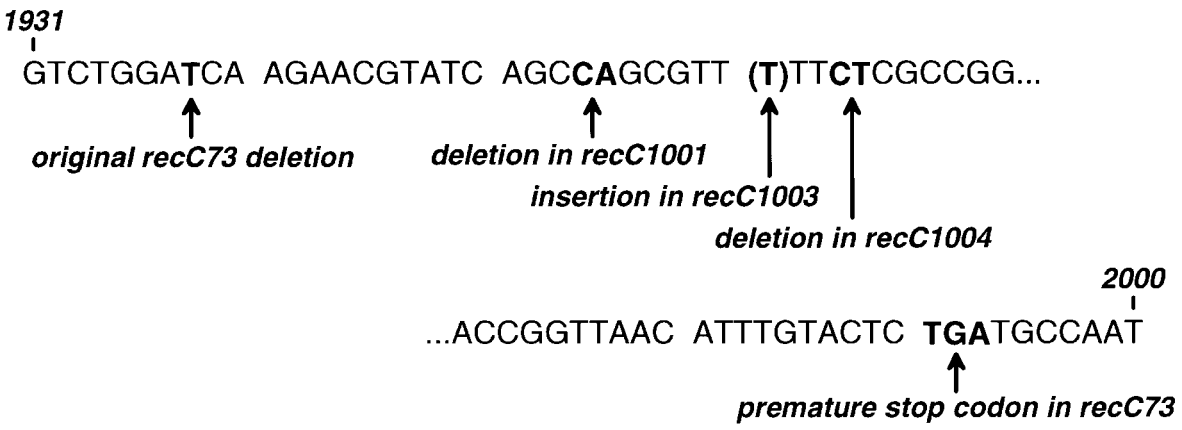
Molecular basis for the altered recognition specificity of the RecC* mutants

To ascertain the nature of the suppressing mutation(s) present in the $RecC^{1004}$ subunit that altered the specificity of χ-recognition, the *recC1004* allele, and its parent, the *recC73* allele, were sequenced (Figure 4(a)). Previously classified as a presumed missense mutant (Templin *et al.*, 1978), we found that *recC73* is actually a frameshift mutation, resulting in a truncated polypeptide of only 663 amino acid residues. A 1 bp deletion occurs at position 1938 and is followed by a region of frameshift that terminates at a stop codon (TGA) 53 bp downstream. Hence, the predicted RecC73 polypeptide lacks amino acids from the C-terminal region of RecC protein (Figure 4(b)). Since the first codon alteration (coding for aspartic acid) is a silent mutation (GAT → GAC), the resulting 663 residue polypeptide contains the first 646 residues of the wild-type *recC* gene followed by a 17 residue region that differs from the wild-type sequence.

The *recC1004* gene contains the *recC73* deletion, as well as a second deletion that removes two additional base-pairs (positions 1963 and 1964, with respect to the wild-type sequence) (Figure 4(a)). This second deletion restores the reading frame for the C-terminal portion of the consequent polypeptide, while maintaining the initial eight amino acid residues of the *recC73* frameshift (Figure 4(b)). Thus, we conclude that a change of nine residues in the wild-type protein to the eight residues in the $RecC^{1004}$ protein is responsible for the altered sequence-specificity, as well as the modest increase in ssDNA nuclease activities of the $RecBC^{1004}D$ enzyme (Arnold *et al.*, 1998).

We also examined the sequence of two other *recC** alleles, *recC1001* and *recC1003*, to determine whether their pseudoreversion mutations are similar to that of the *recC1004* allele, as might be expected based on the phenotypic similarities of the RecC* mutant class. The original single base-

A.



B.

Protein	Amino acid sequence (position)										
	645	647	649	651	653	655	657	659	661	663	
RecC	L D Q E R I S Q R F L A G P V N I C T L										
RecC73	L D K N V S A S V F S P D R L T F V L -										
$RecC^{1001}$	L D K N V S A * R F L A G P V N I C T L										
$RecC^{1003}$	L D K N V S A S V F L A G P V N I C T L										
$RecC^{1004}$	L D K N V S A S V F * A G P V N I C T L										

Figure 4. Sequences of the truncated, frameshift *recC73* allele and the *recC** pseudorevertants. (a) The sequences of the wild-type *recC*, *recC73*, and *recC** genes. (b) The corresponding amino acid sequences of the RecC, RecC73, and RecC* proteins, with a dash denoting the termination of the *recC73* translation product, and asterisks used to maintain alignment of the C-terminal end of sequences that have a net loss of one amino acid.

pair deletion of *recC73* and suppressor mutations were found in both genes (Figure 4(a)). Like *recC1004*, *recC1001* contains a 2 bp deletion downstream of the *recC73* mutation. Since the *recC1001* pseudoreversion occurs closer to the *recC73* deletion than the *recC1004* pseudoreversion (16 residues downstream), there are five altered amino acid residues that replace six of the wild-type RecC protein (Figure 4(b)). The *recC1003* allele possesses a 1 bp insertion that restores the proper frame 23 bp downstream of the *recC73* deletion. This results in a predicted protein sequence in which seven residues of the wild-type polypeptide are replaced by seven in the $RecC^{1003}$ protein (Figure 4(b)).

Discussion

Historically, the *E. coli* recombination hotspot, χ, has been defined as the sequence 5′-GCTGGTGG-3′. However, a significant level of RecBCD enzyme-dependent recombination is observed in the absence of χ (Signer & Weil, 1968). The fact that certain single base-pair mutations of χ can also stimulate recombination *via* the RecBCD pathway, although less efficiently than the canonical χ sequence, provided a mechanism for χ-independent recombination (Cheng & Smith, 1984, 1987). A degeneracy of sequence-recognition is not limited to *E. coli*; the *Haemophilus influenzae* χ site (Chi_{Hi}) was recently identified as, not one, but several sequences related to the *E. coli* χ site, all of which exhibit ''Chi activity'' as measured by the accumulation of HMW plasmid DNA by rolling circle replication (Sourice *et al*., 1998). Furthermore, the fact that RecA protein is also required for HMW DNA formation indicates that recombination, in addition to inactivation of nuclease activity, is involved in this process. The moderate to wild-type levels of recombination exhibited by the *E. coli recC** mutants, which are recombination-proficient, yet exhibit no stimulation of recombination in the presence of canonical χ (Schultz *et al*., 1983), suggest that recognition and response to variants of the canonical χ site can substitute for canonical χ-recognition. Biochemical characterization of the $RecBC^{1004}D$ protein verifies that the only significant difference between this and the wild-type enzyme is the loss of specificity for canonical χ (this work; and see Arnold *et al*., 1998).

An 11 nucleotide variant of χ, χ*, was identified by virtue of its ability to specifically interact with the $RecBC^{1004}D$ enzyme (Handa *et al*., 1997). All genomic fragments that allowed large plaque formation by *red gam* double mutants of phage λ in the *recC1004* strain contained this 11-mer sequence, and single base-pair mutations at positions 8, 9, 10 and 11 destroyed this effect (Handa *et al*., 1997). These results indicated that the χ* sequence conferred protection from $RecBC^{1004}D$ enzyme nuclease, and further implied that the resulting replication products were actively participating in RecA-dependent recombination. This finding also suggested that the *recC1004* mutation had changed the recognition-specificity, and had extended the χ-recognition domain of the RecBCD enzyme to encompass 11 rather than eight nucleotides. The data we present here support the original identification of χ* as an 11 nucleotide sequence that specifically interacts with the $RecBC^{1004}D$ enzyme. The single-nucleotide mutations of χ*, with the exception of χ^{+*}, decrease production of χ*-specific fragments by $RecBC^{1004}D$ enzyme *in vitro* (Figure 2(a)). We further demonstrate that the wild-type RecBCD enzyme can also respond to this variant of χ, and that the efficiency of this response is impaired by mutations at positions 8-11. The fact that the wild-type RecBCD enzyme is able to specifically recognize an 11 nucleotide sequence establishes that this extended recognition capability is not unique to the *recC1004* mutation. In addition, the effect of χ* on lambda plaque size *in vivo* was observed for wild-type and *recC1004* strains, but not for several other RecBCD mutants with deficiencies in canonical χ-recognition, including *recC1001*, *recC1002*, *recC1003*, *recB2154*, *recB2155*, *recC2145*, *recB344*, and *recC343* (our unpublished results). Recognition of χ* by both wild-type and mutant enzymes is accompanied by all of the enzymatic modifications associated with the wild-type response to χ. Thus, our findings show that the interaction of the RecBCD and $RecBC^{1004}D$ enzymes with the 11 nucleotide χ* sequence is analogous to, though less efficient than, interaction with the canonical χ sequence.

Since the last three base-pairs are required for maximal recognition of χ*, it was conceivable that the addition of these nucleotides to the canonical χ sequence would augment its recognition as well. However, addition of positions 9-11 of χ* to the canonical χ site did not enhance recognition by the wild-type enzyme, although recognition by the mutant enzyme was bolstered (Figure 2(b)). Therefore, even though positions 9, 10 and 11 are essential for maximal χ*-recognition by the wild-type enzyme, only the eight nucleotides of the canonical χ sequence are necessary and sufficient for full χ activity. We propose that the sequence flanking χ becomes a factor only when recognition efficiency is compromised either by a mutation in the enzyme, as in the case of $RecBC^{1004}D$ enzyme, or by the presence of a non-canonical χ such as χ*. Extensive mutagenesis of χ, and the sequences flanking χ, would be necessary to fully characterize these effects.

Our data show that χ* elicits the same biochemical changes in the mutant enzyme that χ elicits in the wild-type enzyme. However, since this site occurs only seven times in the *E. coli* genome and not once in the lambda genome, it cannot alone account for the significant levels of conjugal and transductional recombination detected in the *recC1004* strain (Schultz *et al*., 1983). It is more likely that the recombination proficiency of the *recC1004* strain is the additive result of interactions

with χ, χ* and other non-canonical χ sites sufficient to produce the significant levels of recombination detected *in vivo*. The RecBC^{1004}D enzyme does not appear to be a gain-of-function mutant, however, since the RecBCD enzyme recognizes all variants of χ tested at levels comparable to, or better than, the mutant enzyme. Rather, although the specificity for χ is absent, it appears that the RecBC^{1004}D enzyme, and perhaps the other RecBC*D enzymes as well, simply retain the wild-type ability to stimulate recombination in response to variants of χ. The activity of non-canonical χ sites is obscured in wild-type cells, however, due to both the overabundance and high activity of canonical χ sequences.

We suggest that any sequence that stimulates recombination by a specific interaction with RecBCD enzyme be called a χ site, and that a departure from the canonical sequence should not reclassify the sequence as ''χ-like''. Thus, ''χ'' represents a class of sequences that stimulate recombination through a specific interaction with RecBCD enzyme, with the most efficient being the canonical χ sequence, 5′-GCTGGTGG-3′. In order to rank these variants of χ with respect to their relative activities, a means to accurately measure recognition by RecBCD enzyme is essential. Traditional quantification of χ activity in phage λ crosses reveals a five- to tenfold stimulation of recombination due to χ, relative to the recombination frequency seen for the rest of the λ genome (Lam *et al*., 1974; McMilin *et al*., 1974; Stahl *et al*., 1975). Therefore, it follows that χ stimulates recombination five- to tenfold better than the totality of the other octamers in the 48.5 kb λ genome. Since there are ~48,500 octamers, the χ site is 243,000 to 485,000-fold more active than the average octamer. Presented in this manner, the magnitude of the χ effect appears great; however, the measurable activity of χ in these crosses is still just tenfold. Thus, any variant of χ whose recognition efficiency is reduced by tenfold will not be detectable in these assays, even though this level of recognition would represent a 24,000 to 48,500-fold higher activity than the average octamer in λ. In other words, the detection of χ-variant activity is a matter of sensitivity over background, and because the background is high in λ crosses, the sensitivity to χ-variant activity is compromised. However, measuring the production of downstream χ-specific fragments *in vitro* allows detection and quantification of recognition at levels that are only 0.05 the level for wild-type enzyme and, hence, is a more sensitive method to rank the strength of non-canonical χ sites. For example, the *recC1004* strain has no χ-activity as measured by phage λ crosses *in vivo* (Schultz *et al*., 1983), yet RecBC^{1004}C enzyme produced detectable levels of χ-specific fragments *in vitro* (Arnold *et al*., 1998). It is worthwhile to note that the sensitivity of these assays can be even further increased by using tandem arrangements of the non-canonical χ sequence in question and/or a higher concentration of shorter dsDNA substrates (Arnold *et al*., 1998).

The sequence of the entire *recC1004* gene revealed two mutations: a single base-pair deletion at nucleotide position 1938, and a 2 bp deletion corresponding to positions 1963 and 1964. The single base-pair deletion was identified as the parental mutation by sequencing the *recC73* allele; sequences obtained from the *recC1001* and *recC1003* alleles also contain the *recC73* deletion, as well as mutations that correct the reading frame of the *recC* gene. Since the *recC73*, *recC1001*, and *recC1003* alleles were sequenced over a 557 bp region, we cannot exclude the possibility of other mutations outside of this region. However, based on the P1-mediated transduction experiments reported in the original characterization, we feel that it is reasonable to assume that there are no other mutations in the *recC73*, *recC1001*, and *recC1003* alleles, and we propose that the variation in the size of the frameshifted region for each allele is responsible for the moderate phenotypic differences (Schultz *et al*., 1983). Interestingly, the phenotype of the *recC1001* mutation is hyperrecombinogenic (Schultz *et al*., 1983), suggesting that it may result from a reduced specificity of recognition of χ-class sequences.

The RecBC^{1004}D enzyme contains a nine residue change that leaves essentially all the activities of the enzyme intact, except for the specificity of sequence recognition. The fact that amino acid changes in the RecC subunit disrupt the specificity of χ-recognition clearly demonstrates that this region is involved in sequence recognition. Although these data suggest that the χ-recognition domain resides within the RecC subunit, the possibility remains that this region interacts with, and allows recognition of χ by, another subunit in the RecBCD holoenzyme. Further mutagenesis of this region would be invaluable in further defining the χ-recognition domain.

Materials and Methods

Chemicals and buffers

All solutions were made using Barnstead NANOpure water and reagent-grade chemicals. Taq DNA polymerase was purchased from Pharmacia Biotech. All other chemicals, DNA-modifying enzymes, and radioisotopes, were purchased from vendors as listed, and used as described (Arnold *et al*., 1998).

Proteins

Wild-type RecBCD, RecBC^{1004}D, SSB, and RecA proteins were purified as described (Arnold *et al*., 1998).

DNA substrates

Plasmid substrates were purified, linearized, and radioactively end-labeled as described (Arnold *et al*., 1998). pNH92 (4033 bp) and pNH94 (4036 bp) are pBR322 derivatives containing either a χ or χ* sequence,

respectively, flanked by the *Bam*HI and *Hin*dIII sites. pNH95-98, pDAχ^{+*}, and pDAχ^{C8} are similar constructs, bearing instead mutations of the χ or χ* sequence (Table 1). pBR322 served as the χ*-free (χ^0) control. All concentrations of DNA are given in nucleotides, unless otherwise indicated. The strain containing the plasmid from which the *recC1004* gene was sequenced is also the strain that was used to purify the RecBC^{1004}D enzyme; V194, Δ(*argA-thyA*)323, IN(*rrnD-rrnE*)1, pDWS2 derivative; *recC1004* (Arnold *et al*., 1998). The strain that was used to determine the *recC73* mutation was V79, *recC73 argA recF143 his-4 met rpsL31* λ$^-$ (F′15 *thyA*$^+$ *recC73 argA*$^+$) (Chaudhury & Smith, 1984; Schultz *et al*., 1983). The strain used to determine the *recC1001* mutation was V211, Δ(*argA-thyA*)323, IN(*rrnD-rrnE*)1, pDWS2 derivative; *recC1001*; and the strain used to determine the *recC1003* mutation was V193, Δ(*argA-thyA*)323, IN(*rrnD-rrnE*)1, pDWS2 derivative; *recC1003*. The *recC1004*, *recC73*, *recC1001*, and *recC1003* strains were all generously provided by A. F. Tayor, S. K. Amundsen, and G. R. Smith (Fred Hutchinson Cancer Research Center, Seattle, WA). The plasmid template for sequencing the *recC1004* gene was purified using a Qiagen midi-prep kit (Qiagen, Inc.). The primers for sequencing were purchased from Operon Technologies, Inc., and their sequences are available upon request.

χ-Specific fragment production

All reactions were carried out as described (Arnold *et al*., 1998). The χ and χ*-containing plasmids, as well as the plasmid lacking χ (χ^0), were linearized with *Ava*I restriction endonuclease; the plasmid containing the three tandem χ sites was linearized with *Sca*I restriction endonuclease. The reaction buffer contained 25 mM Tris-acetate (pH 7.5), 2 mM magnesium acetate, 1 mM ATP, 1 mM DTT, 10 μM nucleotides of linearized plasmid DNA, and 2 μM SSB protein. The reactions were initiated by addition of 0.10 nM functional RecBCD or RecBC^{1004}D enzyme, corresponding to one functional enzyme per 25 dsDNA ends. The products of these reactions were separated on a 1% (w/v) agarose gel and quantified using a Molecular Dynamics STORM 840 PhosphorImager and ImageQuaNT software. The amount of χ-specific fragment produced in each reaction was normalized to the amount of dsDNA unwound at each time-point; this value was further corrected for the orientation-dependence of interaction with χ and the amount of label per molecule of DNA. Each value represents the average of at least three independent assays performed on different days.

Joint molecule formation assays

Reactions were performed as described (Arnold *et al*., 1998).

Sequencing of the recC* and recC73 alleles

The *recC1004* allele was sequenced by primer walking on both strands to ensure that even a single-nucleotide change would be detected, and as a means to distinguish sequencing artifacts. The *recC73*, *recC1001*, and *recC1003* alleles were sequenced over a 557 bp region, corresponding to nucleotides 1653-2210, which flanks the *recC1004* mutation. All samples were submitted for sequencing to the Division of Biological Science Automated DNA Sequencing Facility on the U. C. Davis campus. The sequences of the primers are available upon request. These sequence data have been submitted to the GenBank database (http://www.ncbi.nlm.nih.gov/) under accession numbers AF176618 (RecC1004), AF176619 (RecC73), AF176620 (RecC1001), and AF176621 (RecC1003).

Acknowledgments

Many thanks to our colleagues in the Kowalczykowski laboratory, Daniel Anderson, Richard Ando, Carole Bornarth, Piero Bianco, Frederic Chédin, Frank Harmon, Noriko Kantake, Julie Kleiman, James New, Erica Seitz, and Susan Shetterley, for careful reading and contributive discussion of this manuscript. We thank Dr Frank Stahl for convincing us of the true magnitude of the χ effect. D. A. was supported by funds from National Institutes of Health grant GM-41347; N. H. is supported by JSPS Research Fellowships for Young Scientists. The work at the Institute of Medical Science is supported by grants to I. K. from the Ministry of ESSC (International collaboration, DNA recombination, DNA repair, Molecular evolution, Class B, genome) and from NEDO.

References

Anderson, D. G. & Kowalczykowski, S. C. (1997a). The recombination hot spot χ is a regulatory element that switches the polarity of DNA degradation by the RecBCD enzyme. *Genes Dev.* **11**, 571-581.

Anderson, D. G. & Kowalczykowski, S. C. (1997b). The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ-regulated manner. *Cell,* **90**, 77-86.

Arnold, D. A., Bianco, P. R. & Kowalczykowski, S. C. (1998). The reduced levels of χ recognition exhibited by the RecBC^{1004}D enzyme reflect its recombination defect *in vivo*. *J. Biol. Chem.* **273**, 16476-16486.

Chaudhury, A. M. & Smith, G. R. (1984). *Escherichia coli recBC* deletion mutants. *J. Bacteriol.* **160**, 788-791.

Cheng, K. C. & Smith, G. R. (1984). Recombinational hotspot activity of Chi-like sequences. *J. Mol. Biol.* **180**, 371-377.

Cheng, K. C. & Smith, G. R. (1987). Cutting of Chi-like sequences by the recBCD enzyme of *E. coli*. *J. Mol. Biol.* **194**, 747-750.

Dabert, P., Ehrlich, S. D. & Gruss, A. (1992). High-molecular-weight linear multimer formation by single-stranded DNA plasmids in *Escherichia coli*. *J. Bacteriol.* **174**, 173-178.

Dixon, D. A. & Kowalczykowski, S. C. (1991). Homologous pairing *in vitro* stimulated by the recombination hotspot, Chi. *Cell,* **66**, 361-371.

Dixon, D. A. & Kowalczykowski, S. C. (1993). The recombination hotspot χ is a regulatory sequence that acts by attenuating the nuclease activity of the *E. coli* RecBCD enzyme. *Cell,* **73**, 87-96.

Dower, N. A. & Stahl, F. W. (1981). Chi activity during transduction-associated recombination. *Proc. Natl Acad. Sci. USA,* **78**, 7033-7037.

Eggleston, A. K. & West, S. C. (1997). Recombination initiation: easy as A, B, C, D... χ? *Curr. Biol.* **7**, R745-R749.

Ennis, D. G., Amundsen, S. K. & Smith, G. R. (1987). Genetic functions promoting homologous recombi-

nation in *Escherichia coli*: a study of inversions in phage lambda. *Genetics,* **115**, 11-24.

Handa, N., Ohashi, S., Kusano, K. & Kobayashi, I. (1997). Chi-star, a Chi-related 11-mer sequence partially active in an *E. coli recC1004* strain. *Genes Cells,* **2**, 525-536.

Holbeck, S. L. & Smith, G. R. (1992). Chi enhances heteroduplex DNA levels during recombination. *Genetics,* **132**, 879-891.

Kobayashi, I., Murialdo, H., Crasemann, J. M., Stahl, M. M. & Stahl, F. W. (1982). Orientation of cohesive end site cos determines the active orientation of Chi sequence in stimulating *recA- recBC*-mediated recombination in phage lambda lytic infections. *Proc. Natl Acad. Sci. USA,* **79**, 5981-5985.

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

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

McMilin, K. D., Stahl, M. M. & Stahl, F. W. (1974). Rec-mediated recombinational hot spot activity in bacteriophage lambda. I. Hot spot activity associated with *spi*$^-$ deletions and *bio* substitutions. *Genetics,* **77**, 409-423.

Myers, R. S. & Stahl, F. W. (1994). Chi and the RecBC D enzyme of *Escherichia coli*. *Annu. Rev. Genet.* **28**, 49-70.

Ponticelli, A. S., Schultz, D. W., Taylor, A. F. & Smith, G. R. (1985). Chi-dependent DNA strand cleavage by recBC enzyme. *Cell,* **41**, 145-151.

Schultz, D. W., Swindle, J. & Smith, G. R. (1981). Clustering of mutations inactivating a Chi recombinational hotspot. *J. Mol. Biol.* **146**, 275-286.

Schultz, D. W., Taylor, A. F. & Smith, G. R. (1983). *Escherichia coli recBC* pseudorevertants lacking Chi recombinational hotspot activity. *J. Bacteriol.* **155**, 664-680.

Signer, E. R. & Weil, J. (1968). Recombination in bacteriophage lambda. I. Mutants deficient in general recombination. *J. Mol. Biol.* **34**, 261-271.

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

Sourice, S., Biaudet, V., El Karoui, M., Ehrlich, S. D. & Gruss, A. (1998). Identification of the Chi site of *Haemophilus influenzae* as several sequences related to the *Escherichia coli* Chi site. *Mol. Microbiol.* **27**, 1021-1029.

Stahl, F. W. & Stahl, M. M. (1977). Recombination pathway specificity of Chi. *Genetics,* **86**, 715-725.

Stahl, F. W., Crasemann, J. M. & Stahl, M. M. (1975). *Rec*-mediated recombinational hot spot activity in bacteriophage lambda. III. Chi mutations are site-mutations stimulating *rec*-mediated recombination. *J. Mol. Biol.* **94**, 203-212.

Taylor, A. F., Schultz, D. W., Ponticelli, A. S. & Smith, G. R. (1985). RecBC enzyme nicking at Chi sites during DNA unwinding: location and orientation-dependence of the cutting. *Cell,* **41**, 153-163.

Templin, A., Margossian, L. & Clark, A. J. (1978). Suppressibility of *recA, recB*, and *recC* mutations by nonsense suppressors. *J. Bacteriol.* **134**, 590-596.

Edited by M. Gottesman

(Received 1 February 2000; received in revised form 2 May 2000; accepted 5 May 2000)