Molecular determinants responsible for recognition of the single-stranded DNA regulatory sequence, χ, by RecBCD enzyme

Naofumi Handa[a,b], Liang Yang[a], Mark S. Dillingham[c], Ichizo Kobayashi[b], Dale B. Wigley[d], and Stephen C. Kowalczykowski[a,1]

[a]Department of Microbiology and Department of Molecular and Cellular Biology, University of California, Davis, CA 95616; [b]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108-8639, Japan; [c]DNA-Protein Interactions Unit, School of Biochemistry, University of Bristol, Bristol BS8 1TD, United Kingdom; and [d]Division of Structural Biology, Chester Beatty Laboratories, Institute of Cancer Research, London SW3 6JB, United Kingdom

Contributed by Stephen C. Kowalczykowski, April 11, 2012 (sent for review February 21, 2012)

The RecBCD enzyme is important for both restriction of foreign DNA and recombinational DNA repair. Switching enzyme function from the destructive antiviral state to the productive recombinational state is regulated by the recombination hotspot, χ (5′-GCTGGTGG-3′). Recognition of χ is unique in that it is recognized as a specific sequence within single-stranded DNA (ssDNA) during DNA translocation and unwinding by RecBCD. The molecular determinants of χ recognition and the subsequent alteration in function are unknown. Consequently, we mutated residues within the RecC subunit that comprise a channel where ssDNA is thought to be scanned for a χ sequence. These mutants were characterized in vivo with regard to χ recognition, UV-sensitivity, phage degradation, and recombination proficiency. Of 38 residues mutated, 11 were previously undescribed mutations that altered χ recognition. The mutants fell into two classes: five that failed to respond to χ, and six that suggested a relaxed specificity for χ recognition. The location of the first set of mutations defines a recognition structure responsible for sequence-specific binding of ssDNA. The second set defines a highly conserved structure, linked to the recognition structure, which we hypothesize regulates conversion of RecBCD from a molecular machine that destroys DNA to one that repairs it. These findings offer insight into the evolution of enzymes with alternate χ recognition specificities.

helicase | nuclease | protein–DNA interactions

Homologous recombination in *Escherichia coli* is initiated by the heterotrimeric helicase/nuclease, RecBCD (1, 2). RecBCD comprises two motor subunits (RecB and RecD) with opposite translocation polarities that travel along the complementary DNA strands (2–4). The enzyme unwinds and degrades double-stranded DNA (dsDNA) from an end until it encounters an 8-nt sequence, called Chi [crossover hotspot instigator (χ): 5′-GCTGGTGG-3′] (5, 6). When the enzyme recognizes this single-stranded DNA (ssDNA) sequence, its nucleolytic activities are altered: the vigorous 3′→5′ nuclease is down-regulated and the 5′→3′ nuclease is up-regulated (7, 8). In addition to these changes, χ-activated RecBCD displays the capacity to load RecA protein, an essential factor for DNA strand exchange, onto the χ-containing ssDNA (9–11). Thus, by switching in response to χ recognition, RecBCD changes from an enzyme that provides an antiviral function to cells by degrading foreign DNA and bacteriophages, to one that provides a recombinational dsDNA-break repair function to maintain chromosome integrity (2, 12). As a reflection of this important function, the χ sequence is over-represented in the *E. coli* genome by ∼sixfold and is found on average every 4.5 kb (13).

To date, a few RecBCD mutants with a deficiency in χ recognition have been reported (14–20). Because recognition of χ results from a number of interdependent steps, defects in χ recognition can result from a number of molecular events (2). Lack of recognition can arise from: (*i*) a defective RecB motor, which is needed to pump the χ-containing ssDNA through a channel in RecBCD to a χ-scanning site in RecC (15–17); (*ii*) loss of the RecD subunit, which results in an enzyme, RecBC, that is constitutive for recombination functions (21–23); (*iii*) the inability to recognize χ (18–20); or (*iv*) the inability to couple the χ recognition to the requisite conformational change.

The crystal structure of the RecBCD enzyme showed that the dsDNA is separated into two strands at a pin structure located in RecC, such that the 5′- and 3′-DNA strands are positioned to bind the motor domains of the RecD and RecB subunits, respectively (4). The structure also reveals the presence of two channels: one each for the 5′- and 3′-terminated DNA strands. Moreover, the exit for the 3′-ssDNA channel is facing the nuclease domain of RecBCD, which is located at the C terminus of the RecB subunit; consequently, unless the conformation of RecBCD is changed, the 3′-terminated DNA strand has a high probability of being degraded. The RecC subunit was proposed to scan the ssDNA for the χ sequence as the unwound DNA is pumped through the internal channel (4). Pseudorevertants (*recC** mutants) of a *recC*-null strain were found that did not respond to the canonical χ but were inferred to recognize an altered sequence (14). All of these mutants arose from insertions or deletion downstream of the original frame-shift mutation to restore the reading frame and to change 7–9 amino acid residues of RecC (19). One mutant (*recC1004*) did not respond to normal χ but recognized an altered χ sequence, χ*, 5′-GCTGGTG*CT-CG*-3′ (19, 20, 24). By virtue of displaying altered specificity, rather than ambiguous loss of recognition, this is the only mutant that directly implicates a subunit, RecC, in the χ recognition process.

Because the χ sequence is essentially used as a bar-code for DNA self-recognition, its sequence varies in divergent bacteria (25). *Bacillus subtilis* uses 5′-AGCGG-3′ as its cognate χ sequence to regulate its RecBCD homolog, AddAB (2, 26), and *Lactococcus lactis* uses 5′-GCGCGTG-3′ as χ to regulate its RecBCD homolog, RexAB enzyme, which is an AddAB ortholog (27). The χ sequences of the Gram-negative bacterium, *Haemophilus influenzae*, are the χ variants 5′-GNTGGTGG-3′ and 5′-G[C/G]TGGAGG-3′ (28). Finally the χ sequence of *E. coli* was found to attenuate the endogenous ATP-dependent dsDNA exonuclease in a variety of related enteric bacteria, including *Shigella Sonnei*, *Citrobacter freundii*, *Klebsiella pneumoniae*, *Serratia marcescens*, *Salmonella typhimurium*, and *Proteus mirabilis*, suggesting a common or degenerate variant in these closely

Author contributions: N.H., M.S.D., D.B.W., and S.C.K. designed research; N.H. and L.Y. performed research; N.H., L.Y., and I.K. contributed new reagents/analytic tools; N.H., M.S.D., I.K., D.B.W., and S.C.K. analyzed data; and N.H., M.S.D., D.B.W., and S.C.K. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: sckowalczykowski@ucdavis.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1206076109/-/DCSupplemental.

related species (29–31). Soley from statistical analysis of genome sequences, it has become possible to identify overrepresented sequences with the genomic characteristics of χ sequences (25). This strategy led to the identification of 5′-GAAGCGG-3′ in *Staphylococcus aureus* and 5′-GCGCGTG-3′ in several other *Streptococci* as χ-like sequences. Thus, there is great variety in the sequences of χ-like elements, and the strategy of using a DNA sequence-regulated helicase/nuclease is broadly used within the Bacteria.

Although the structure of RecBCD provided considerable insight into RecBCD function, it did not identify the χ recognition motifs because the ssDNA had not penetrated into the RecC subunit. Predicting the domain responsible for recognition is not possible because, despite the multitude of structures that reveal the nature of sequence-specific recognition of dsDNA, there is relatively little known regarding the molecular determinants of sequence-specific recognition of ssDNA. Consequently, we identified and mutated candidate residues on the surface of the channel within the RecC subunit that could be involved in χ recognition. We found 11 different mutations that result in an altered response to χ. The mutations fell into two categories: one category behaved in vivo as though it has lost the capacity to recognize χ; the second behaved as although the sequence determinant for χ recognition had become relaxed. These mutations define the χ interaction region and offer insights into an allosteric relay system that transmits recognition of a specific regulatory sequence within ssDNA: recognition that results in a switch in enzymatic and functional personality of RecBCD.

Results

Identification of Amino Acids Within the RecC Channel That Potentially Interact with χ. The crystal structure was used to identify candidate residues that are surface-accessible along the putative χ recognition channel (Fig. 1*A*) (4). An initial group of 22 residues was selected and, based on the subsequent identification of positive candidates, another group of more focused residues was mutagenized. In the end, 35 amino acids were individually changed to alanine (Fig. 1*B*). An additional three mutations (E399G, V450G, and A766V) were generated as a consequence of PCR mutagenesis errors and did not alter phenotype. The selected residues are widely spread on the RecC polypeptide but cluster in three regions of primary structure (Fig. 1*B*): Q38 to R186, D414 to F442, and Q652 to R708. The last region is close to the position of the *recC*1004* alteration (20).

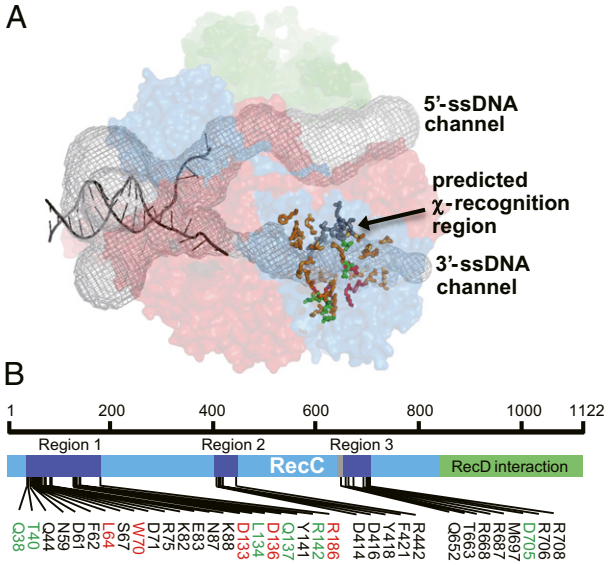

Fig. 1. The RecC subunit of the RecBCD complex. (*A*) Structure of RecBCD showing the location of DNA and channels (4). RecB, RecC, and RecD are red, blue, and green, respectively. The two internal channels (black mesh) were calculated using MOLE (47). Type 1 or type 2 residues are red or green, respectively; other residues mutated in this study are bronze-orange sticks; and those altered in *recC1004* (19) are gray sticks. (*B*) Primary structure of RecC showing residues mutated here. Type 1 or type 2 mutations are in red or green, respectively. The region altered in *recC** mutants is gray (19).

RecC-Channel Mutants Retain the Capacity to Nucleolytically Destroy T4 Phage in Vivo. The individual *recC* mutations were inserted into a plasmid, transformed into a *recC*-null strain, and then assessed for functions in vivo. All of the mutants at least partially complemented the null strain with regard to UV survival, showing that at least partially functional proteins were being produced (Table S1). Of the mutants examined, all displayed RecBCD-dependent nuclease activity, as judged by the inability of mutant T4 *gene 2*$^-$ bacteriophage to plate on these cells (Table S1; some mutants were not fully characterized because, based on initial screening, their phenotypes did not pass the criteria defined below). The *gene 2* product of bacteriophage T4 is a DNA end-binding protein that normally protects the linear dsDNA of the phage from degradation by RecBCD enzyme; T4 phage lacking *gene 2* function are restricted for growth because of RecBCD nuclease function (15, 32). Consequently, each of these RecBCD mutants is proficient for dsDNA exonuclease activity. Note that T4 *gene 2*-defective phage grow well on cells lacking *recD* (22); this is because the RecBC enzyme is a helicase but not a nuclease, and the nuclease activity of RecBCD is needed to block survival of T4 *gene 2*$^-$ phage (21). Thus, the failure of this phage to grow on our *recC* mutant strains shows that none of the mutations resulted in the loss of the RecD subunit from RecBCD to produce a *recD*$^-$-like phenotype. This finding is significant because the RecBC enzyme does not recognize χ (23): had any of our mutations resulted in the loss of RecD (15), then they may have appeared as false-negatives with respect to loss of χ recognition.

RecC-Channel Mutants Display Two Distinctive Classes of Defects with Regard to Their Response to χ. The recognition of χ was examined by plating bacteriophage λ, without or with a χ sequence, on bacteria expressing the mutant RecBCD (20, 33). Phage λ propagates by two routes: one is recombination and the other is rolling circle replication to make a tandemly repeated genome for packaging into phage particles. Both routes are stimulated by the interaction between RecBCD and χ, when the phage's recombination functions (*red* and *gam* genes) are inactive: χ protects the λ genome from degradation by RecBCD, and χ also activates RecBCD to enhance recombination frequency. Therefore, a λ phage with a χ sequence (χ^+) produces more progeny than wild-type λ (which does not have a χ sequence, χ^0) in wild-type cells (6). As a result, χ^+ phage make larger plaques compared with isogenic χ^0 phage. Large plaques can also appear if the nuclease activity of the RecBCD enzyme is reduced; however, because all of our mutant RecBCD enzymes (except one) display nuclease activity as measured by the T4 *gene 2*$^-$ plating assay, then such a possibility can be excluded.

The results of plating phage λ on all 38 mutant *recBCD* cells are summarized in Table S1 (columns four and five). Wild-type bacteria show the typical response that results in plaques for χ^+ phage that are about two- to threefold larger than for χ^0 phage. Because the RecBCD nuclease restricts λ phage growth, phage without a χ sequence forms large plaques on a *recC*$^-$ strain. Therefore, to identify RecC-channel mutants that potentially lacked χ recognition, we applied the following phenotypic criteria: (*i*) the mutant displayed nuclease activity [i.e., blocked plating of T4 *gene 2*$^-$ mutant phage (Table S1, column three)]; (*ii*) the mutant cells produced plaques for χ^0 phage λ that are the same size as wild-type cells (Table S1, column four); and (*iii*) the mutant cells produced plaques for χ^+ phage λ that are the same size as for the χ^0 phage (Table S1, column five). There were five mutants that met these criteria: L64A, W70A, D133A, D136A, and R186A (Fig. 1*A*); they are summarized in Table S2 and classified as "type 1." For each mutant, the growth of either χ^0 or

χ^+ phage λ results in the same plaque size, and the size is the same as for χ^0 phage λ grown on wild-type cells (Fig. 2*A*).

A second possible type of χ recognition mutant is an enzyme that either recognized a new χ sequence that was present on both χ^0 and χ^+ phages, or that displayed a relaxed recognition of sequences that were a subset of the canonical χ sequence. To identify such candidates, the following phenotypic criteria were applied: (*i*) the mutant displayed nuclease activity [blocked plating of T4 *gene 2*$^-$ mutant phage (Table S1, column three)]; (*ii*) the mutant cells produced a large plaque size for χ^0 phage λ that was at least as large as the plaque for χ^+ phage λ on wild-type cells (Table S1, column four); and (*iii*) the mutant cells produced plaque sizes for χ^+ phage λ that are the same as for χ^0 phage (Table S1, column five). There were six mutants that met these criteria: Q38A, T40A, L134A, Q137A, R142A, and D705A (Fig. 1*A*); they are also summarized in Table S2 and classified as "type 2." For each mutant, the growth of χ^0 or χ^+ phage λ results in a large plaque; in fact, the plaque size is as large, or larger, than that for χ^+ phage growing on wild-type cells (Fig. 2*A*). Interestingly, for the D705A mutation, the plaque sizes approach the size observed for the *recBCD*-null strain, but clearly neither the D705A mutation nor the other type 1 and 2 mutations result in a null phenotype for UV sensitivity (Fig. 2*B*) or for T4 *gene 2*$^-$ growth, showing that they are not simply inactive. Finally, when we combined some of the type 2 mutations into a single RecC polypeptide for reasons that will become apparent below, the phenotype of Q137A-D705A, R142A-D705A, and Q137A-R142A-D705A was the same as for the single mutations (Table S1), showing that they were not synergistic.

Both Type 1 and Type 2 Mutants Reveal Alterations in χ-Mediated Recombination. To verify that these newly isolated RecBCD mutants are altered for χ recognition, recombination between *red*$^-$ *gam*$^-$ λ phages (which are defective for phage-encoded recombination function) with or without a χ sequence was measured in vivo (20). The cross is designed to determine whether a sequence inserted into the *int* gene region (between *J* and *cI*) behaves as a recombination hotspot by measuring the ratio of clear and turbid plaques formed by the recombinant progeny (Fig. 3*A*) (20). Recombinants possessing the S^+ and *Jh* markers are assayed to determine whether their phenotype with regard to the intervening immunity region is turbid (cI^+) or clear (*cI*). If spontaneous recombination occurred evenly along the DNA, then the recombination frequency on both the left and right sides of the immunity region should be the same (1:1); if a recombination hotspot is inserted at the *int* locus, then turbid plaques will be prominent (20). All of the type 1 and 2 mutants were proficient for basal χ-independent recombination (Table S1) that occurs because of a low frequency of spontaneous activation of RecBCD (18).

For the wild-type protein, the response to χ is evident in both the recombination frequency and the hotspot activity: both measurements are increased [2.6 (± 0.5)-fold and 3.8 (± 1.2)-fold, respectively] by χ (Fig. 3 *B* and *C* and Table S2). We note that in our assay, the mutant *recC* genes are expressed from a multicopy plasmid under the control of the P_{tac} promoter. It was previously reported that overexpression of the RecC subunit reduces χ hotspot activity (34). In agreement, even though the expression in our work was in the absence of isopropyl-β-D-thiogalactopyranoside (IPTG), we also observed that the increased gene dosage of RecC diminished the measured χ response (Fig. S1). The control with wild-type RecC expressed from the chromosome showed a 1.8-fold higher χ hotspot activity when compared to the same assay with RecC expressed from the plasmid, which is the same fold-difference previously noted (34). All of the type 1 mutants (L64A, W70A, D133A, D136A, and R186A) displayed recombination frequencies similar to wild-type for the χ^0 phage and, significantly, none showed an increased frequency in response to χ (Fig. 3*B*). In addition, none of the type 1 mutants showed higher hotspot activity in the χ^+ cross (Fig. 3*C*). We also produced a histidine variant at position R186 (R186H) because this mutation is present in the previously characterized *recC2145* allele, which does not recognize χ in vivo (15). The *recC2145* allele actually has two substitutions in RecC (R186H and G304S) (2). Given that R186H

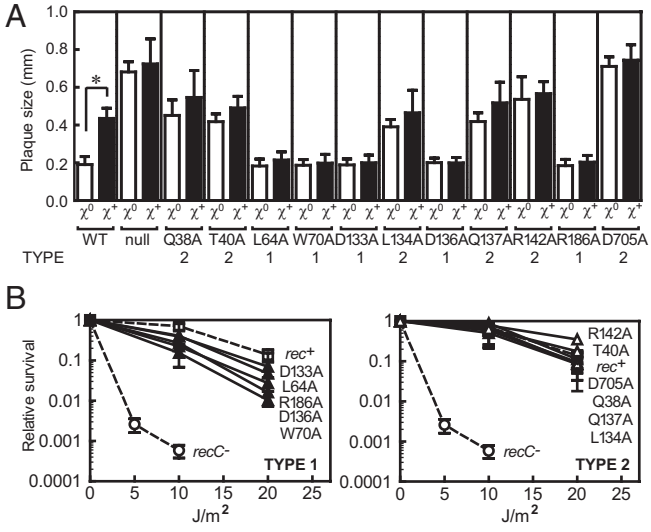

Fig. 2. Phenotypic assays reveal defects in recognition of χ by the RecC-channel mutants. (*A*) The sizes of phage λ plaques distinguish χ recognition for the *recC* mutants. Each bar represents average with SD of 10 individual plaques; *$P < 0.005$ in *t* test. (*B*) UV resistance of *recC* mutants. (*Left*) Type 1 mutants, ▲; (*Right*) type 2 mutants, △; *recC*$^+$, □; and *recC*$^-$ (without plasmid), ○. Averages with SD from at least two independent experiments.

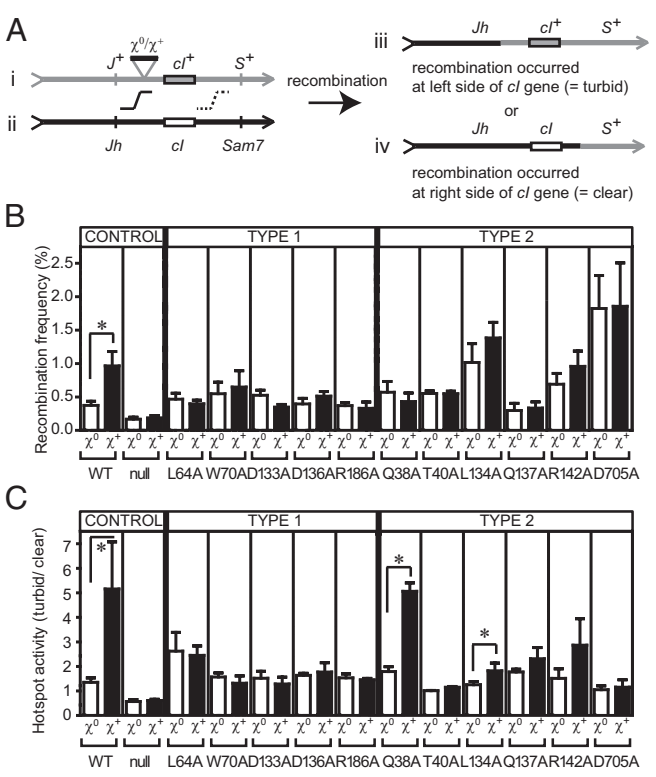

Fig. 3. In vivo λ recombination assays. (*A*) Schematic of λ recombination assay. (*B*) Recombination frequency (%). (*C*) Hotspot activity (ratio of turbid/clear plaques). Averages with SDs from at least three assays. Recombination frequency of χ^0 and χ^+ λ for wild-type strain (chromosomal *recC*$^+$ gene) was 0.63 ± 0.16 and 2.8 ± 0.5, and hotspot activity was 1.8 ± 0.2 and 7.2 ± 0.6, respectively. *$P < 0.005$ in *t* test.

and R186A have the same χ recognition deficiencies, we conclude the R186 position is responsible for the phenotype of the *recC2145* allele. Thus, all of the type 1 mutants fail to recognize χ, and they define residues that are likely to be directly involved in binding the recombination hotspot.

The type 2 mutants, however, revealed a complex recombination behavior. The L134A, R142A, and D705A mutations showed higher recombination frequency than wild-type for χ^0 and χ^+ phage (Fig. 3*B*) but apparently no hotspot activity (Fig. 3*C*). On the other hand, the Q38A mutant showed approximately wild-type levels of recombination frequency for χ^0 phage (Fig. 3*B*) and an increased hotspot activity (Fig. 3*C*). However, if these mutations caused an altered or relaxed recognition specificity, then interpretation of the phage recombination measurements can be complicated if the alternative χ sequence precedes the recombination interval being studied. If these mutant RecBCD enzymes recognize novel (perhaps shorter) χ-like sequences (see *Discussion*), then they could respond to the putative χ-like sequence first, which would preclude recognition of the distal wild-type χ sequence (35). Phage λ does indeed possess at least two χ-like sequences (cCTGGTGG at 541 bp and GtTGGTGG at 2,763 bp from the right end) upstream of the selected S^+ marker. If any of the type 2 mutants are activated by an upstream χ variant, then recombination of both χ^0 and χ^+ phage will be increased (Fig. 3*B*, D705A, for example), resulting in a ratio of hotspot activity of only 1.0 (Fig. 3*C*, D705A, for example).

To determine whether the two classes of mutants represent a gain or loss of function, we measured their genetic behavior in cells containing wild-type RecBCD. One allele (R186A or D705A) from each class was investigated (Fig. S2). The D705A allele (type 2) is completely dominant over the wild-type, resulting in large plaques for both χ^0 and χ^+ phage in the wild-type background. In contrast, the R186A allele (type 1) has no effect on χ^0 phage but reduced the plaque size for χ^+ phage; therefore, the R186A mutant is codominant to the wild-type RecBCD.

Residues Identified by Mutagenesis Are Highly Conserved in all RecC Homologs. To determine whether the residues identified by mutational analysis in this study are generally important to RecBCD function in other bacteria, we examined their conservation in *recC* orthologs. The sequences of 112 unique bacterial *recC* genes were aligned, showing clear islands of conservation (Fig. S3). The residues identified here fall into five regions of conservation (Fig. 4). Many of these χ recognition residues are highly conserved, with several showing extraordinary preservation. For example, at position 142, arginine is found in 97% of the 112 orthologs, and at position 705, aspartate is nearly always (90%) found and the residue is always acidic (aspartic or glutamic acid). Interestingly, although some amino acids (e.g., R706) are highly conserved, they have no effect on χ recognition, at least when mutated to an alanine (Table S1). In the RecBCD structure (Fig. 5), R142 and D705 form a salt bridge that, based on the conservation, must be a crucial and universal structural element of the χ response. The significance of the locations of the other residues involved in χ recognition, and their conservation, will be made apparent in the *Discussion*.

Discussion

The crystal structure of RecBCD revealed that each DNA strand passes through a different interior channel and engages with a motor domain in either the RecB or RecD subunit (4). The 3′-ssDNA tail is translocated by the RecB motor into a proposed χ-scanning channel in RecC (17), eventually exiting the complex at the nuclease domain, which is at the C terminus of RecB. Several crystal structures of RecBCD have revealed the details of the interaction between the 3′-tail and the RecB motor, but this tail was always too short to cross the RecB/RecC interface (36). However, the channel through RecC in the protein complex is a clear structural feature and, together with several previous studies, suggested that χ sequence recognition occurs in the RecC subunit (14, 19, 20). Indeed, here we established that this region

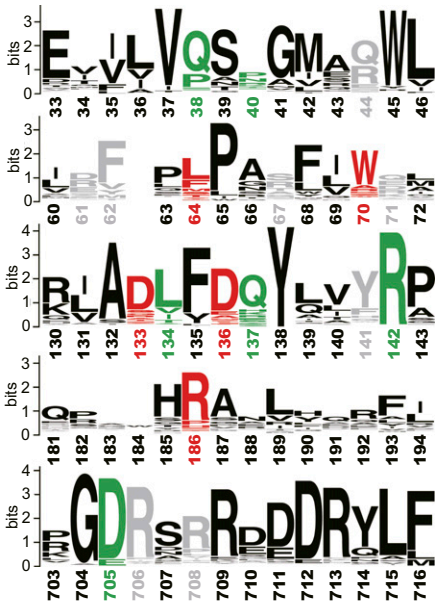

Fig. 4. Conservation of the amino acid residues involved in χ recognition among RecC orthologs. Five regions of highly conserved sequence implicated in χ recognition are shown in WebLogo format (48, 49). Type 1 residues are in red, and type 2 in green. Residues in gray were mutated but did not affect χ recognition. Gap in line 2 reflects a very small number of RecC sequences (not including *E. coli*) having a single amino acid insert there (Fig. S3).

serves as the χ-scanning site of RecBCD, and we identified residues that are crucial to the recognition and regulation process.

Among the 35 individual mutations to alanine that we intentionally constructed, 11 mutants showed a phenotype characteristic of altered χ recognition (Fig. 5). Because of PCR mutagenesis errors, sequencing uncovered several other substitutions. The accidental channel mutations, R186C and D705H, and the intentional substitution, R186H, have the same phenotypes as their respective alanine substitution, showing that at least for these positions, the phenotype is not unique to alanine substitution. Therefore, in all, we analyzed 41 different mutations at 38 amino acid positions, of which 11 unique loci resulted in altered χ recognition (Fig. 1, Table S1, and Table S2). Based on their in vivo characteristics, these mutations fell into two groups: type 1 mutations, which resulted in a loss of χ recognition or response,

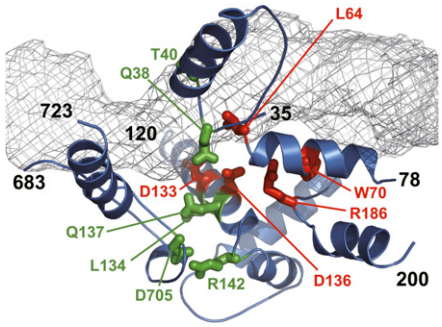

Fig. 5. Locations of residues in RecC structure implicated in χ recognition. Channel for 3′-ssDNA is gray mesh calculated using MOLE (47). Residues important for χ recognition are in stick format, and red for type 1 or green for type 2. Secondary structural elements are from regions of conservation in Fig. 4; black residue numbers indicate the starts and ends. Note that the first two elements, as well as the third and fourth, are continuous; the fifth is separate and comprises a loop joined by the ionic interaction between the highly conserved R142 and D705 residues.

and type 2 mutations, which resulted in, we infer, an altered or relaxed recognition of χ.

The type 1 residues (L64, W70, D133, D136, and R186) seem to define regions that are directly responsible for either recognition of, or response to, χ. These RecBCD mutant proteins either fail to recognize the canonical χ sequence or they fail to switch to the χ-activated form of the enzyme. These mutants could have completely lost the capacity to recognize any χ sequence or they might require a sequence that is not present in bacteriophage λ. Interestingly, cells expressing the type 1 mutations were not drastically UV sensitive (Fig. 2*B*). A likely explanation is that these mutant enzymes show a low but meaningful basal level of repair mediated by χ-like or nonspecific sequences, or a low level of spontaneous activation (18), which is augmented by the RecF-pathway components to effect repair via a "hybrid" pathway (37). Biochemical analysis of the mutant RecBCD enzymes, which is described in the companion article by Yang et al. (38), shows that they are unable to recognize χ.

The RecC structure itself resembles a Superfamily 1 (SF1) helicase and may interact with ssDNA using an equivalent region of the structure (4). The most striking structural feature of the type 1 residues is that they form an almost continuous patch of surface on subdomain 1B near the rear exit of the RecC tunnel (Fig. 5 and Fig. S4), defining a part of the χ-scanning site. Although not contiguous within the primary structure, D133-D136-R186-W70 form an approximately linear array that comprises the ssDNA-recognition motif that could read the sequence of the incoming ssDNA in the 3′ to 5′ direction, offering a possible mechanism for χ recognition. Because alteration of any one of these residues prevents response of RecBCD to χ, we imagine that both the acidic aspartate and the basic arginine are forming specific hydrogen bonds with the nucleotides of the χ sequence, and that both the aromatic tryptophan and hydrophobic leucine (L64) are making hydrophobic interactions with the bases in the χ sequence. Interestingly, arginine interacts predominantly with G and secondarily with T, and aspartate interacts with C (39): these are the nucleotides that constitute the χ sequence.

Although much is known about sequence-specific binding to dsDNA (see, for example, ref. 39), there are relatively few examples of sequence-specific recognition of ssDNA. The RNA polymerase of N4 bacteriophage interacts with ssDNA as a template, but requires a DNA stem-loop structure (40); the structure of this complex shows that an arginine and lysine hydrogen-bond with guanine bases, and a tryptophan stacks on the guanine in the ssDNA loop; in addition, a specificity loop makes base-specific contacts with the stem via aspartate and arginine residues and salt bridges with lysine residues (41). The σ70-ssDNA complex is another example: it uses an amphipathic helix with exposed residues to directly contact bases in the ssDNA formed upon melting (42). The hnRNP K protein binds ssDNA by a network of hydrogen bonds between isoleucine residues and two specific bases (43). Finally, the telomeric end-binding protein of *Oxytricha* binds $(T_4G_4)_2$ specifically via stacking interactions of tyrosine on bases, and H-bonding of tyrosine, glutamine, and arginine residues with the bases (44). Characterization of the χ recognition locus of RecBCD adds to this limited set of sequence-specific ssDNA interactions.

The type 2 mutations define a particularly interesting and previously undescribed class of altered RecBCD enzymes. These mutants show high levels of recombination even in the absence of χ, yet they possess nuclease activity, suggesting that they are either constitutively activated to promote recombination or they recognize a DNA sequence that is much more frequent than the canonical 8-nt χ sequence. If these mutant enzymes were to be activated for recombination by a 6- or 7-nt sequence, then they would initiate recombination in phage λ crosses in an interval before the one being measured (downstream of the canonical χ). Thus, their recombination activity would be high but they would appear to be unresponsive to the canonical χ because the mutant RecBCD would have been activated to promote a crossover before the wild-type χ sequence; in addition, this gain of function—that is, the capacity to recognize many more χ-like sequences—would be consistent with the dominant phenotype (Fig. S2). The type 2 residues (Q38, T40, L134, Q137, R142, and D705) are found in the subdomains 1A, 1B, and 2A of RecC and, based on the crystal structure of RecBCD, it is evident that all of these residues cannot interact directly with χ (Fig. 5). However, we identified a polypeptide loop lining the channel that may have a role in this process (4, 36). The structure of RecC is related to SF1 helicases but it also has an additional loop inserted in the 2A domain. One of the χ-recognition mutants (D705) is located in this loop and it makes an ion pair with another χ-recognition mutant (R142) situated in the 1B domain. Although these residues are remote from the channel, R142 is at the far end of an α-helix—the χ-recognition helix in the 1B domain—that contains a cluster of other χ-recognition mutations that do line the channel (D133, L134, D136, and Q137), suggesting a network of interactions that constitute a structure important for transmission of an allosteric signal from the χ-recognition helix to the ion-pair structure. This signal could mediate the conformational changes needed to alter the functionality of RecBCD (Fig. S5).

Interestingly, some of the amino acid residues identified here as being involved in χ recognition show a remarkable degree of conservation with 112 other bacterial RecC proteins (Fig. 4). In particular, the nearly absolute maintenance of the two residues that comprise the ion pair (R142 and D/E705) uncovers the universal importance of this structural element to RecBCD function; in agreement, this structure, although different in detail, is conserved in AddAB enzyme (45). In addition, most of the other residues important to χ recognition are also very well conserved and are found in conserved regions.

The χ sequence is used as a bar code to permit bacteria to identify DNA that has a high probability of being its own chromosomal DNA, which is in need of recombinational DNA repair (1, 2): if RecBCD enzyme does not encounter a cognate χ sequence while unwinding DNA, then it will nucleolytically destroy it. For this reason, the cognate χ sequences are the most abundant sequences in their respective organisms (25, 46); in *E. coli* they are found every ~4.5 kb in the genome [1,009 sites for the 4,639,675-bp genome (13)]. Our identification of the molecular determinants for χ recognition provides insight into the specificity of χ sequence recognition by RecBCD homologs. It is known that many enteric bacteria use the same sequence, 5′-GCTGGTGG, as χ (29, 30). In fact, even though *Salmonella*'s RecC is somewhat diverged, all of the χ recognition residues identified here are conserved. Interestingly, *Haemophilus* RecC has T40P and L64M substitutions (Table S3), and here we showed that mutations at these positions in *E. coli* RecC change its χ recognition. In fact, the *Haemophilus* χ sequence is 5′-GNTGGTGG, or 5′-G[C/G]TGGAGG (28). Given that T40 is also mutated to a proline in both *P. mirabilis* and *Serratia proteamaculans*, which recognize the same χ as *E. coli* or a degenerate version, this fact is consistent with our interpretation herein that T40 confers a possible relaxed specificity of χ recognition. Thus, T40 and L64 may well contribute to the recognition of the second or sixth letter of the χ sequence, and it is easy to imagine how the recognition of divergent χ sequences could evolve via slight modification of these or the other residues identified herein. Detailed in vitro analysis (38) and a crystal structure of RecBCD in a χ recognition complex will provide further insight into the mechanism of ssDNA sequence-specific χ recognition and its effects on the biochemical properties of RecBCD.

Materials and Methods

In Vivo Assays: UV Sensitivity, Bacteriophage λ Plaque Size, T4 *Gene 2*⁻ Bacteriophage Plating, and Bacteriophage λ Recombination. *E. coli,* bacteriophage, plasmid, construction of RecC-channel mutants and other details are provided in *SI Materials and Methods* and Tables S4 and S5. For the phenotypic analyses, cells expressing *recC* from a relatively low-copy plasmid were grown in the absence of IPTG induction. UV sensitivity was measured using expo-

nentially growing cultures diluted and spread on L agar plates. Plates were irradiated with UV light (254 nm). Bacteriophage λ plaque sizes were measured as described previously (33). RecBCD nuclease activity in vivo was examined using the T4 *gene 2*$^{-}$ bacteriophage plating assay (32). Bacteriophage λ recombination crosses were performed as described previously (20, 24), except that IPTG induction was not necessary. The adsorption frequency of parental phages was 96.0 ± 3.1% (n = 42).

Sequence Alignments. The sequences of RecC polypeptide were manually selected at KEGG GENES database (http://www.genome.jp/kegg/genes.html) to ensure that there were no redundant entries (e.g., fragments or mutant strains). These sequences were aligned to the *E. coli* protein using MAFFT (http://www.genome.jp/tools/mafft/). For the alignment, the 112 unique bacterial genes listed in the *SI Materials and Methods* were used.

ACKNOWLEDGMENTS. This work was supported by the following institutions: grants from the Japan Society for the Promotion of Science Postdoctoral Fellowship for Research Abroad, The Naito Foundation, the Kato Memorial Bioscience Foundation, the Takeda Foundation, and the Sumitomo Foundation, and Grants-in-Aid for Scientific Research from both the Japan Society for the Promotion of Science and the Ministry of Education, Culture, Sports, Science and Technology (to N.H.); a Royal Society and Wellcome Trust grant (to M.S.D.); Global COE Genome Information Big Bang, Grants-in-Aid for Scientific Research, and Grant-in-Aid for Scientific Research on Innovative Areas (to I.K.); and National Institutes of Health Grant GM41347 (to S.C.K.).

1. Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD, Rehrauer WM (1994) Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol Rev* 58: 401–465.
2. Dillingham MS, Kowalczykowski SC (2008) RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol Rev* 72:642–671.
3. Dillingham MS, Spies M, Kowalczykowski SC (2003) RecBCD enzyme is a bipolar DNA helicase. *Nature* 423:893–897.
4. Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB (2004) Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* 432:187–193.
5. Ponticelli AS, Schultz DW, Taylor AF, Smith GR (1985) Chi-dependent DNA strand cleavage by RecBC enzyme. *Cell* 41:145–151.
6. Lam ST, Stahl MM, McMilin KD, Stahl FW (1974) Rec-mediated recombinational hot spot activity in bacteriophage lambda. II. A mutation which causes hot spot activity. *Genetics* 77:425–433.
7. Dixon DA, Kowalczykowski SC (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.
8. Anderson DG, Kowalczykowski SC (1997) 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.
9. Spies M, Kowalczykowski SC (2006) The RecA binding locus of RecBCD is a general domain for recruitment of DNA strand exchange proteins. *Mol Cell* 21:573–580.
10. Churchill JJ, Kowalczykowski SC (2000) Identification of the RecA protein-loading domain of RecBCD enzyme. *J Mol Biol* 297:537–542.
11. Anderson DG, Kowalczykowski SC (1997) The translocating RecBCD enzyme stimulates recombination by directing RecA protein onto ssDNA in a χ-regulated manner. *Cell* 90:77–86.
12. Handa N, Ichige A, Kusano K, Kobayashi I (2000) Cellular responses to postsegregational killing by restriction-modification genes. *J Bacteriol* 182:2218–2229.
13. Blattner FR, et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–1462.
14. Schultz DW, Taylor AF, Smith GR (1983) *Escherichia coli* RecBC pseudorevertants lacking chi recombinational hotspot activity. *J Bacteriol* 155:664–680.
15. Amundsen SK, Neiman AM, Thibodeaux SM, Smith GR (1990) Genetic dissection of the biochemical activities of RecBCD enzyme. *Genetics* 126:25–40.
16. Holbeck SL, Smith GR (1992) Chi enhances heteroduplex DNA levels during recombination. *Genetics* 132:879–891.
17. Spies M, Dillingham MS, Kowalczykowski SC (2005) Translocation by the RecB motor is an absolute requirement for χ-recognition and RecA protein loading by RecBCD enzyme. *J Biol Chem* 280:37078–37087.
18. Arnold DA, Bianco PR, Kowalczykowski SC (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.
19. Arnold DA, Handa N, Kobayashi I, Kowalczykowski SC (2000) A novel, 11 nucleotide variant of χ, χ*: One of a class of sequences defining the *Escherichia coli* recombination hotspot χ. *J Mol Biol* 300:469–479.
20. 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.
21. Chaudhury AM, Smith GR (1984) A new class of *Escherichia coli recBC* mutants: Implications for the role of RecBC enzyme in homologous recombination. *Proc Natl Acad Sci USA* 81:7850–7854.
22. Amundsen SK, Taylor AF, Chaudhury AM, Smith GR (1986) *recD*: The gene for an essential third subunit of exonuclease V. *Proc Natl Acad Sci USA* 83:5558–5562.
23. Churchill JJ, Anderson DG, Kowalczykowski SC (1999) The RecBC enzyme loads RecA protein onto ssDNA asymmetrically and independently of chi, resulting in constitutive recombination activation. *Genes Dev* 13:901–911.
24. Handa N, Kowalczykowski SC (2007) A RecA mutant, RecA(730), suppresses the recombination deficiency of the RecBC(1004)D-chi* interaction in vitro and in vivo. *J Mol Biol* 365:1314–1325.
25. Halpern D, et al. (2007) Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling. *PLoS Genet* 3:1614–1621.
26. Chédin F, Noirot P, Biaudet V, Ehrlich SD (1998) A five-nucleotide sequence protects DNA from exonucleolytic degradation by AddAB, the RecBCD analogue of *Bacillus subtilis*. *Mol Microbiol* 29:1369–1377.
27. el Karoui M, Ehrlich D, Gruss A (1998) Identification of the lactococcal exonuclease/recombinase and its modulation by the putative Chi sequence. *Proc Natl Acad Sci USA* 95:626–631.
28. Source S, Biaudet V, El Karoui M, Ehrlich SD, 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.
29. McKittrick NH, Smith GR (1989) Activation of Chi recombinational hotspots by RecBCD-like enzymes from enteric bacteria. *J Mol Biol* 210:485–495.
30. Rinken R, de Vries J, Weichenhan D, Wackernagel W (1991) The recA-recBCD dependent recombination pathways of *Serratia marcescens* and *Proteus mirabilis* in *Escherichia coli*: Functions of hybrid enzymes and hybrid pathways. *Biochimie* 73: 375–384.
31. Smith GR, Roberts CM, Schultz DW (1986) Activity of Chi recombinational hotspots in *Salmonella typhimurium*. *Genetics* 112:429–439.
32. Oliver DB, Goldberg EB (1977) Protection of parental T4 DNA from a restriction exonuclease by the product of gene 2. *J Mol Biol* 116:877–881.
33. Handa N, Ichige A, Kobayashi I (2009) Contribution of RecFOR machinery of homologous recombination to cell survival after loss of a restriction-modification gene complex. *Microbiology* 155:2320–2332.
34. Rinken R, Wackernagel W (1992) Inhibition of the recBCD-dependent activation of Chi recombinational hot spots in SOS-induced cells of *Escherichia coli*. *J Bacteriol* 174: 1172–1178.
35. Stahl FW, Thomason LC, Siddiqi I, Stahl MM (1990) Further tests of a recombination model in which chi removes the RecD subunit from the RecBCD enzyme of *Escherichia coli*. *Genetics* 126:519–533.
36. Saikrishnan K, Griffiths SP, Cook N, Court R, Wigley DB (2008) DNA binding to RecD: Role of the 1B domain in SF1B helicase activity. *EMBO J* 27:2222–2229.
37. Ivančić-Baće I, et al. (2003) RecFOR function is required for DNA repair and recombination in a RecA loading-deficient *recB* mutant of *Escherichia coli*. *Genetics* 163: 485–494.
38. Yang L, et al. (2012) Alteration of χ recognition by RecBCD reveals a regulated molecular latch and suggests a channel-bypass mechanism for biological control. *Proc Natl Acad Sci USA* 109:8907–8912.
39. Luscombe NM, Laskowski RA, Thornton JM (2001) Amino acid-base interactions: A three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res* 29:2860–2874.
40. Davydova EK, Santangelo TJ, Rothman-Denes LB (2007) Bacteriophage N4 virion RNA polymerase interaction with its promoter DNA hairpin. *Proc Natl Acad Sci USA* 104: 7033–7038.
41. Gleghorn ML, Davydova EK, Rothman-Denes LB, Murakami KS (2008) Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol Cell* 32:707–717.
42. Malhotra A, Severinova E, Darst SA (1996) Crystal structure of a sigma 70 subunit fragment from *E. coli* RNA polymerase. *Cell* 87:127–136.
43. Braddock DT, Baber JL, Levens D, Clore GM (2002) Molecular basis of sequence-specific single-stranded DNA recognition by KH domains: Solution structure of a complex between hnRNP K KH3 and single-stranded DNA. *EMBO J* 21:3476–3485.
44. Peersen OB, Ruggles JA, Schultz SC (2002) Dimeric structure of the *Oxytricha nova* telomere end-binding protein alpha-subunit bound to ssDNA. *Nat Struct Biol* 9: 182–187.
45. Saikrishnan K, et al. (2012) Insights into Chi recognition from the structure of an AddAB-type helicase-nuclease complex. *EMBO J* 31:1568–1578.
46. Touzain F, Petit MA, Schbath S, El Karoui M (2011) DNA motifs that sculpt the bacterial chromosome. *Nat Rev Microbiol* 9:15–26.
47. Petrek M, Kosinová P, Koca J, Otyepka M (2007) MOLE: A Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure* 15:1357–1363.
48. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14:1188–1190.
49. Schneider TD, Stephens RM (1990) Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res* 18:6097–6100.

Supporting Information

Handa et al. 10.1073/pnas.1206076109

SI Materials and Methods

Bacterial Strains, Bacteriophages, and Plasmids. The bacteria, phage, and plasmids used in this study are listed in Table S4. The *Escherichia coli* K-12 derivatives used here are DH5 (S.C.K. laboratory collection); BNH963 [*recC*::Tn*10* derivative of AB1157 (1)], which was mainly used for in vivo characterization; and V330 [Δ(*recC-argA*)*234* su^0] (2), which was used for in vivo nuclease assay. A *gene 2*$^-$ mutant of T4 bacteriophage (S.C.K. laboratory collection) was used for in vivo nuclease assay. For λ recombination assay, BIK808 and JM1 were used as indicator strains after recombination (3, 4). Bacteriophage λ with or without a χ sequence [LIK950 and LIK916 (3)] were used for χ recognition assays, and these two and LIK1068 (3) were used for λ recombination assays in vivo. pPB700 (5), pNH336 (present work), pWKS6 (6), and pMS421 (7) were expression vectors for *recB*, *recC*, *recD* (5), and *lacI*^{q} (7) genes, respectively. These plasmids were introduced into V330 for the T4 *gene 2*$^-$ assays, Plasmids that express the mutant RecC proteins are described in Table S4.

Media. *E. coli* strains were grown in L broth [1.0% (wt/vol) Bacto-tryptone, 0.5% (wt/vol) yeast extract, and 0.5% (wt/vol) sodium chloride]. Antibiotics were added at the following concentrations when required: ampicillin (amp) at 100 μg/mL, chloramphenicol (cam) at 25 μg/mL, kanamycin (kan) at 50 μg/mL, and spectinomycin (spc) at 30 μg/mL.

Construction of Plasmids Expressing the RecC Channel Mutants. Construction of all single amino acid substitutions was carried out as follows. We first constructed a wild-type *recC* expression vector from a larger plasmid that was used previously [pPB520 (5)] to facilitate site-directed mutagenesis. The *recC* gene with a *tac* promoter was amplified using forward (2106–2125: 5′-AACGACAGGAGCACGATCAT) and reverse (7358–7377: 5′-GGCTGATCAAGCAGATTGTACTGAGAGTGC) primers from pPB520. The amplified DNA fragment was digested with BamHI (3204: between the forward primer and the promoter) and NdeI (7350: between the reverse primer and end point of the *recC* gene) and ligated into pACYC184 digested with BamHI and PshBI. The resulting plasmid, pNH336, was sequenced to confirm that no mutations were incorporated. The single alanine-substitution mutations were constructed by site-directed mutagenesis from pNH336 by using a mutated primer set (Table S5) and PCR. The entire *recC* gene for all mutant constructs was confirmed by sequencing. All constructs have no additional mutations unless stated otherwise.

UV Sensitivity Measurements. A plasmid with wild-type or mutated *recC* was introduced into BNH963 (*recC*::Tn*10*). Exponential cultures in L broth with appropriate antibiotics were diluted in broth and spread on L agar plates. The plates were irradiated with UV light (254 nm) for various doses (times). Colonies were counted after incubation at 37 °C for 20 h in dark. Survival relative to the absence of UV irradiation was plotted.

Bacteriophage λ Plaque Size Assay. Details were described previously (8). *E. coli* host cells (BNH963), with a plasmid expressing a *recC* allele were freshly grown in 1.0% (wt/vol) Bacto-tryptone 0.5% (wt/vol) NaCl, 0.2% (wt/vol) maltose, 10 mM $MgSO_4$, and 10 μg/mL vitamin B1 at 37 °C until they reached mid-log phase. An aliquot of the culture was mixed with melted agar [0.6% (wt/vol)] and prewarmed media, and then poured onto an agar plate [1.2% (wt/vol)] containing 1.0% (wt/vol) polypeptone and 0.5% (wt/vol) NaCl. After the top agar cooled, diluted phage LIK916 (χ^0) or LIK950 (χ^+) were spotted onto the plate and incubated at 37 °C for overnight. After incubation, plates were scanned by a computer-connected scanner (Canon) and the image was analyzed with Illustrator software (Adobe). The diameter of 10 independent plaques for each strain and phage were measured using the pointing tool, and values were calibrated using a ruler scanned by the same manner.

T4 *Gene 2*$^-$ Bacteriophage in Vivo Nuclease Assay. Nuclease activity in vivo was examined using the T4 *gene 2*$^-$ bacteriophage plating assay (9). *Gene 2* encodes a DNA end-binding protein for T4 phage that protects the phage from nucleolytic degradation by RecBCD; consequently, *gene 2*$^-$ T4 mutant phage cannot grow in the presence of RecBCD enzyme. Because the *gene 2*$^-$ mutation is an amber mutation, strain V330 (su^0) containing plasmids for expression of the *recB*, *recC*, and *recD* genes was used instead of AB1157 (*supE*).

In Vivo Recombination Using a Bacteriophage λ Cross Assay. *E. coli* host cells (BNH963) with a plasmid expressing a *recC* allele were freshly grown in 1.0% (wt/vol) Bacto-tryptone 0.5% (wt/vol) NaCl, 0.2% (wt/vol) maltose, 10 mM $MgSO_4$, and 10 μg/mL vitamin B1 at 37 °C to a density of 4×10^8 cells/mL isopropyl-β-D-thiogalactopyranoside induction was not necessary and consequently omitted. An aliquot of the culture was mixed with a half volume each of LIK916 (χ^0) or LIK950 (χ^+) (at 4×10^9 phages/mL) and LIK1068 (at 4×10^9 phages/mL); the multiplicity of infection was 5 for both parental phages. The mixture was incubated at 37 °C without shaking for 30 min, then the mixture was transferred into prewarmed liquid media (100-times dilution), and then incubated further for 90 min with shaking at 37 °C. A few drops of $CHCl_3$ were added, and the phage lysate was mixed with the BIK808 indicator strain to measure *Jh* S^+ recombinant phage; total phage was measured by mixing with strain JM1. After adequate dilution, these mixtures were mixed with melted and prewarmed top agar [0.6% (wt/vol)] and poured onto plates (see above). The recombination frequency was calculated as the recombinant phage titer measured using BIK808 divided by the total phage titer measured using JM1, multiplied by 100. Hotspot activity was calculated as the ratio of turbid plaques divided by clear plaques measured using BIK808. The adsorption frequency of the parental phages was 96.0 ± 3.1% (n = 42).

Sequence Alignments. The sequences of RecC polypeptide that were annotated as "exonuclease V gamma subunit" or "recC" were manually selected at KEGG GENES database (http://www.genome.jp/kegg/genes.html) to ensure that they were not redundant entries (e.g., fragments or sequence mutants). These sequences were aligned to the *E. coli* protein using MAFFT (http://www.genome.jp/tools/mafft/). For the alignment, 112 unique bacterial genes used were: *Escherichia coli* K-12 MG1655, *Salmonella enterica* subsp. *enterica serovar Typhi* CT18, *Yersinia pestis* CO92, *Shigella sonnei*, *Erwinia carotovora*, *Erwinia tasmaniensis*, *Photorhabdus luminescens*, *Buchnera aphidicola* APS, *Wigglesworthia glossinidia*, *Sodalis glossinidius*, *Enterobacter* sp. 638, *Enterobacter sakazakii*, *Klebsiella pneumoniae*, *Citrobacter koseri* ATCC BAA-895, *Serratia proteamaculans*, *Proteus mirabilis*, *Edwardsiella ictaluri*, *Candidatus Blochmannia floridanus*, *Candidatus Hamiltonella defense*, *Dickeya dadantii* Ech703, *Xenorhabdus bovienii*, *Pantoea*

ananatis, Haemophilus influenzae Rd KW20, *Haemophilus somnus* 129PT, *Pasteurella multocida*, *Mannheimia succiniciproducens*, *Actinobacillus pleuropneumoniae* L20, *Aggregatibacter aphrophilus*, *Xylella fastidiosa* 9a5c, *Xanthomonas campestris* pv. *campestris* ATCC 33913, *Stenotrophomonas maltophilia* K279a, *Vibrio cholerae* O1, *Vibrio fischeri*, *Photobacterium profundum*, *Pseudomonas aeruginosa* PAO1, *Azotobacter vinelandii*, *Psychrobacter arcticum*, *Acinetobacter* sp. ADP1, *Moraxella catarrhalis*, *Shewanella oneidensis*, *Idiomarina loihiensis*, *Colwellia psychrerythraea*, *Pseudoalteromonas haloplanktis*, *Marinobacter aquaeolei*, *Alteromonas macleodii*, *Psychromonas ingrahamii*, *Methylococcus capsulatus*, *Francisella tularensis* subsp. *tularensis* SCHU S4, *Allochromatium vinosum*, *Alkalilimnicola ehrlichei*, *Halorhodospira halophila*, *Thioalkalivibrio* sp. K90mix, *Halothiobacillus neapolitanus*, *Hahella chejuensis*, *Chromohalobacter salexigens*, *Alcanivorax borkumensis*, *Marinomonas* sp. MWYL1, *Aeromonas hydrophila*, *Tolumonas auensis*, *Dichelobacter nodosus*, *Acidithiobacillus ferrooxidans* ATCC 53993, *Baumannia cicadellinicola*, *Gamma proteobacterium* HdN1, *Neisseria meningitidis* Z2491 (serogroup A), *Chromobacterium violaceum*, *Laribacter hongkongensis*, *Burkholderia mallei* ATCC 23344, *Delftia acidovorans*, *Variovorax paradoxus*, *Comamonas testosterone*, *Leptothrix cholodnii*, *Thiomonas intermedia*, *Nitrosomonas eutropha*, *Aromatoleum aromaticum* EbN1, *Azoarcus* sp. BH72, *Thauera* sp. MZ1T, *Accumulibacter phosphatis*, *Geobacter sulfurreducens*, *Pelobacter propionicus*, *Desulfomicrobium baculatum*, *Desulfotalea psychrophila*, *Desulfurivibrio alkaliphilus*, *Desulfococcus oleovorans*, *Desulfatibacillum alkenivorans*, *Desulfobacterium autotrophicum*, *Haliangium ochraceum*, *Syntrophus aciditrophicus*, *Mycobacterium tuberculosis* H37Rv, *Nocardia farcinica*, *Rhodococcus* sp. RHA1, *Gordonia bronchialis*, *Tsukamurella paurometabola*, *Cellulomonas flavigena*, *Nocardioides* sp. JS614, *Nakamurella multipartite*, *Kineococcus radiotolerans*, *Candidatus Protochlamydia amoebophila*, *Waddlia chondrophila*, *Borrelia burgdorferi* B31, *Leptospira interrogans serovar lai*, *Chitinophaga pinensis*, *Fibrobacter succinogenes*, *Coraliomargarita akajimensis*, *Gemmatimonas aurantiaca*, *Synechococcus* sp. WH8102, *Prochlorococcus marinus* SS120, *Chlorobaculum tepidum*, *Chlorobium chlorochromatii*, *Pelodictyon luteolum*, *Deferribacter desulfuricans* SSM1, and *Denitrovibrio acetiphilus*.

1. Bachmann BJ (1987) Derivation and genotypes of some mutant derivatives of *Escherichia coli* K-12. *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*, eds Neidhardt FC, et al. (American Society for Microbiology, Washington, DC), Vol 2, pp 1190–1219.
2. Amundsen SK, Taylor AF, Smith GR (2002) A domain of RecC required for assembly of the regulatory RecD subunit into the *Escherichia coli* RecBCD holoenzyme. *Genetics* 161:483–492.
3. 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.
4. Handa N, Kowalczykowski SC (2007) A RecA mutant, RecA(730), suppresses the recombination deficiency of the RecBC(1004)D-chi* interaction in vitro and in vivo. *J Mol Biol* 365:1314–1325.
5. Boehmer PE, Emmerson PT (1991) *Escherichia coli* RecBCD enzyme: Inducible overproduction and reconstitution of the ATP-dependent deoxyribonuclease from purified subunits. *Gene* 102:1–6.
6. Handa N, Bianco PR, Baskin RJ, Kowalczykowski SC (2005) Direct visualization of RecBCD movement reveals cotranslocation of the RecD motor after χ recognition. *Mol Cell* 17:745–750.
7. Heath JD, Weinstock GM (1991) Tandem duplications of the lac region of the *Escherichia coli* chromosome. *Biochimie* 73:343–352.
8. Handa N, Ichige A, Kobayashi I (2009) Contribution of RecFOR machinery of homologous recombination to cell survival after loss of a restriction-modification gene complex. *Microbiology* 155:2320–2332.
9. Oliver DB, Goldberg EB (1977) Protection of parental T4 DNA from a restriction exonuclease by the product of *gene 2*. *J Mol Biol* 116:877–881.

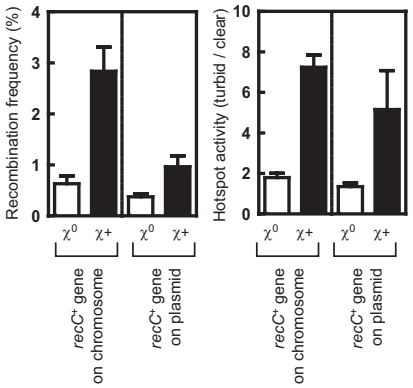

Fig. S1. Comparison of recombination frequency and hotspot activities for chromosomally expressed versus plasmid-expressed *recC*. The indicated assays were carried out for the *recC*$^+$ gene expressed from the chromosome and from plasmid pNH336. (*Left*) Recombination frequency (%). (*Right*) Hotspot activity (ratio of turbid to clear plaque morphology). Recombination and hotspot assays were carried out more than three times, and both the average and SD are presented.

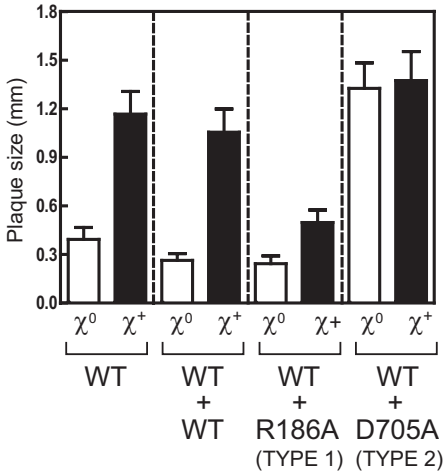

Fig. S2. RecBCD mutants displaying χ-recognition defects are at least partially dominant with respect to wild-type RecBCD. The plaque sizes of either χ^0 or χ^+ phage λ were measured as described. Plasmids pNH336 (WT), pNH386 (R186A), or pNH751 (D705A) were transformed into AB1157 (wild-type). Each bar represents the average with SD of 10 individual plaques.

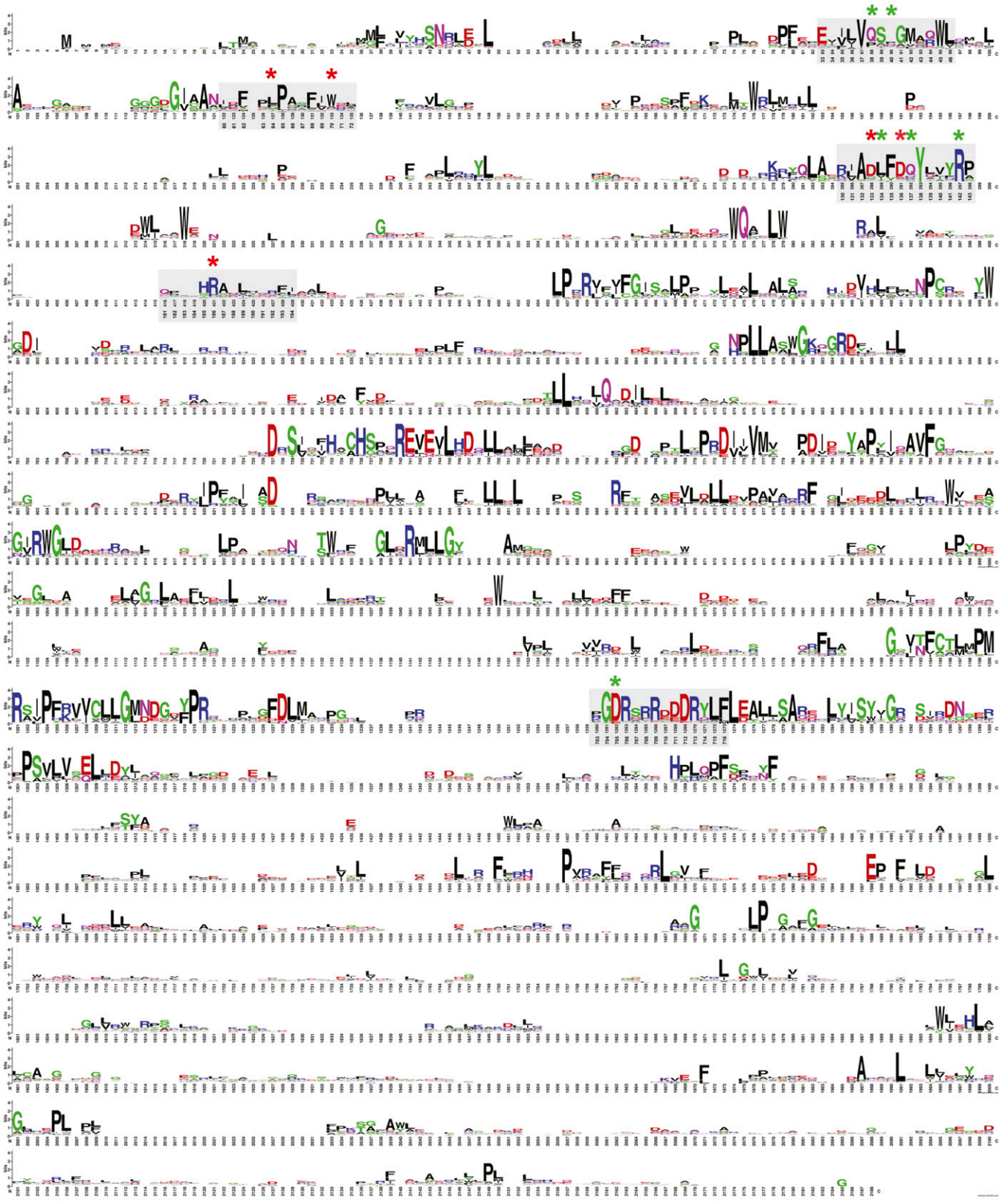

Fig. S3. Amino acid sequence conservation in the RecC polypeptide. The 112 unique RecC sequences were aligned as described in *SI Materials and Methods*. The multiple sequence alignment was then analyzed using WebLogo (1, 2) to produce a graphic representation of conservation in the sequence. Amino acids are color-coded according to their chemical properties and their height indicates the degree of conservation at each position. The red or green asterisks above residues denote type 1 and type 2 residues, respectively. The sequences are, in general, rather divergent but there are several regions of well-conserved residues, five of which (highlighted in gray; the residues numbers for *E. coli* are provided below those regions) contain residues implicated in the recognition of the χ sequence (Fig. 4).

1. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14:1188–1190.
2. Schneider TD, Stephens RM (1990) Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res* 18:6097–6100.

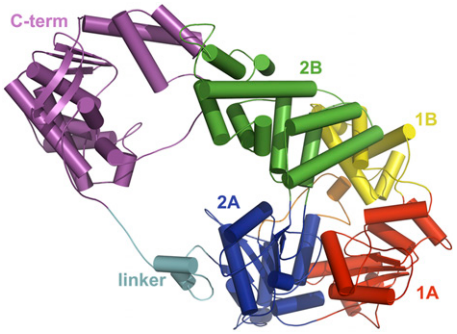

Fig. S4. Ribbon representation of the RecC polypeptide. The protein is colored according to the domain architecture. The 1A, 2A, 1B, and 2B subdomains of the N-terminal inactivated helicase domain are shown in red, blue, yellow, and green, respectively. A short insert in the 2A subdomain that forms part of the proposed latch mechanism is colored orange and this contacts the 1B domain (yellow) via a highly conserved ionic interaction. The C-terminal domain of RecC, which is involved in protein-protein interactions with RecD, is shown in magenta and a linker region between the N and C domains is colored cyan.

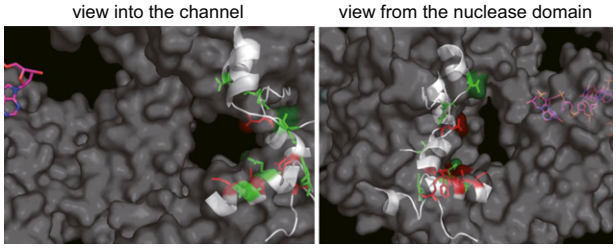

Fig. S5. Locations within the RecBCD structure of the residues in RecC implicated in χ recognition. Location of type 1 (red) and type 2 (green) mutations in the RecBCD structure. On the left is a view into the χ-recognition channel of RecC (for clarity, residues 688–691 are not shown), and on the right is a view from the back where the nuclease domain is located.

Table S1. Genetic analysis of RecBCD enzymes with RecC-channel mutations

Mutation	UV survival* (10 J/m^2)	T4 *gene 2*$^-$ plaque	λ Plaque size† χ^0	λ Plaque size† χ^+	λ Cross frequency χ^0	λ Cross frequency χ^+	λ Cross hotspot activity χ^+/χ^0
WT	0.57 ± 0.21	—	0.19 ± 0.04	0.44 ± 0.05	0.32 ± 0.02	0.77 ± 0.03	4.8 ± 0.6
Null	0.00055 ± 0.00019	+	0.68 ± 0.05	0.72 ± 0.13	0.17 ± 0.03	0.18 ± 0.04	1.1 ± 0.1
Q38A	0.45 ± 0.10	—	0.45 ± 0.08	0.55 ± 0.14	0.37 ± 0.19	0.55 ± 0.27	2.8 ± 0.2
T40A	0.51 ± 0.22	±	0.42 ± 0.04	0.49 ± 0.06	0.55 ± 0.04	0.55 ± 0.04	1.1 ± 0.04
Q44A	0.41 ± 0.22	—	0.29 ± 0.06	0.94 ± 0.10	N.T.	N.T.	N.T.
N59A	0.62 ± 0.35	N.T.	0.29 ± 0.05	0.83 ± 0.05	N.T.	N.T.	N.T.
D61A	0.59 ± 0.15	—	0.29 ± 0.08	0.98 ± 0.10	N.T.	N.T.	N.T.
F62A	0.24 ± 0.08	—	0.24 ± 0.04	0.79 ± 0.05	N.T.	N.T.	N.T.
L64A	0.41 ± 0.16	—	0.18 ± 0.04	0.22 ± 0.04	0.47 ± 0.09	0.40 ± 0.06	0.96 ± 0.18
S67A	0.49 ± 0.29	—	0.25 ± 0.04	0.77 ± 0.06	N.T.	N.T.	N.T.
W70A	0.15 ± 0.09	—	0.18 ± 0.03	0.20 ± 0.05	0.55 ± 0.17	0.65 ± 0.24	0.85 ± 0.22
D71A	0.58 ± 0.17	—	0.31 ± 0.10	0.81 ± 0.05	N.T.	N.T.	N.T.
R75A	0.70 ± 0.24	N.T.	0.28 ± 0.04	0.89 ± 0.10	N.T.	N.T.	N.T.
K82A	0.58 ± 0.15	N.T.	0.30 ± 0.06	0.97 ± 0.09	N.T.	N.T.	N.T.
E83A	0.54 ± 0.17	—	0.30 ± 0.05	0.92 ± 0.09	N.T.	N.T.	N.T.
N87A	0.49 ± 0.25	—	0.28 ± 0.07	0.98 ± 0.08	N.T.	N.T.	N.T.
K88A	0.47 ± 0.11	—	0.31 ± 0.05	0.86 ± 0.10	N.T.	N.T.	N.T.
D133A	0.32 ± 0.12	—	0.19 ± 0.03	0.20 ± 0.04	0.53 ± 0.07	0.35 ± 0.04	0.87 ± 0.23
L134A	0.41 ± 0.09	—	0.39 ± 0.04	0.47 ± 0.12	1.0 ± 0.3	1.4 ± 0.2	1.5 ± 0.3
D136A	0.27 ± 0.05	—	0.20 ± 0.02	0.20 ± 0.03	0.40 ± 0.08	0.51 ± 0.07	1.1 ± 0.12
Q137A	0.55 ± 0.22	—	0.42 ± 0.05	0.52 ± 0.11	0.31 ± 0.18	0.30 ± 0.15	1.3 ± 0.3
Y141A	0.46 ± 0.19	—	0.46 ± 0.09	1.0 ± 0.1	N.T.	N.T.	N.T.
R142A	0.79 ± 0.10	—	0.54 ± 0.12	0.57 ± 0.06	0.69 ± 0.16	0.96 ± 0.23	1.8 ± 0.3
R186A	0.23 ± 0.03	—	0.19 ± 0.03	0.20 ± 0.04	0.37 ± 0.05	0.33 ± 0.09	0.95 ± 0.07
R186H	0.055 ± 0.003	—	N.D. (S)	N.D. (S)	0.42 ± 0.10	0.34 ± 0.08	0.93 ± 0.21
R186C‡	0.099 ± 0.014	—	N.D. (S)	N.D. (S)	N.T.	N.T.	N.T.
E399G‡	0.41 ± 0.20	—	0.29 ± 0.08	1.0 ± 0.2	N.T.	N.T.	N.T.
D414A	0.55 ± 0.11	—	0.28 ± 0.05	0.94 ± 0.06	N.T.	N.T.	N.T.
D416A	0.49 ± 0.18	—	0.33 ± 0.07	1.1 ± 0.2	N.T.	N.T.	N.T.
Y418A	0.66 ± 0.34	—	0.41 ± 0.05	1.2 ± 0.1	N.T.	N.T.	N.T.
F421A	0.50 ± 0.25	—	0.28 ± 0.08	0.76 ± 0.09	N.T.	N.T.	N.T.
R442A	0.58 ± 0.17	—	0.28 ± 0.09	0.77 ± 0.09	N.T.	N.T.	N.T.
V450G‡	0.46 ± 0.36	—	0.43 ± 0.05	1.0 ± 0.11	N.T.	N.T.	N.T.
Q652A	0.59 ± 0.22	—	0.32 ± 0.06	0.79 ± 0.06	N.T.	N.T.	N.T.
T663A	0.62 ± 0.21	—	0.30 ± 0.06	0.65 ± 0.05	N.T.	N.T.	N.T.
R668A	0.044 ± 0.001	—	VS	VS	0.22 ± 0.09	0.41 ± 0.19	1.9 ± 0.6
R687A	0.42 ± 0.29	—	0.40 ± 0.04	0.84 ± 0.11	N.T.	N.T.	N.T.
M697A	0.61 ± 0.20	—	0.32 ± 0.07	1.1 ± 0.1	N.T.	N.T.	N.T.
D705A	0.89 ± 0.14	—	0.71 ± 0.05	0.74 ± 0.08	1.8 ± 0.5	1.9 ± 0.7	1.1 ± 0.4
D705H	0.49 ± 0.20	—	N.D. (L)	N.D. (L)	N.T.	N.T.	N.T.
R706A	0.44 ± 0.23	—	0.23 ± 0.05	0.63 ± 0.06	N.T.	N.T.	N.T.
R708A	0.56 ± 0.14	—	0.34 ± 0.06	0.81 ± 0.08	0.41 ± 0.03	0.62 ± 0.01	1.3 ± 0.1
A766V‡	0.50 ± 0.24	N.T.	0.41 ± 0.05	1.0 ± 0.2	N.T.	N.T.	N.T.
Q137A-R142A	0.40 ± 0.14	—	N.D. (L)	N.D. (L)	N.T.	N.T.	N.T.
Q137A-D705A	0.43 ± 0.16	—	N.D. (L)	N.D. (L)	N.T.	N.T.	N.T.
Q142A-D705A	0.46 ± 0.33	—	N.D. (L)	N.D. (L)	N.T.	N.T.	N.T.
Q137A-R142A-D705A	0.428 ± 0.071	—	N.D. (L)	N.D. (L)	N.T.	N.T.	N.T.

Values are given as an average ± SD. N.D., not determined; N.T., not tested. All strains except for null (*recC*::Tn*10*) are BNH963 with a plasmid harboring the mutant *recC*. Strains used for T4 *gene 2*$^-$ assay are in the Su^0 with the same plasmid.

*More than two measurements were carried out.

†Plaque sizes are presented as the measured diameter (mm) for most mutants; in some cases, the sizes were not measured precisely but rather the size is simply denoted as (L) large, (S) small, or (VS) very small.

‡Mutation was accompanied with a second mutation: R186C with D133A, E399G with D414A, V450G with R687A, and A766V with R75A.

Table S2. Mutations in *recC* that result in RecBCD enzymes with altered χ recognition phenotypes

Phenotype	Mutation*	T4 *gene 2*$^-$ plating	Plaque size[†], χ^0 *(mm)*	Plaque size[†], χ^+ *(mm)*	Hotspot activity[‡] χ^+/χ^0
Wild-type	—	—	0.19 ± 0.04	0.44 ± 0.05	3.8 ± 1.2
Null	deletion	+	0.68 ± 0.05	0.72 ± 0.13	1.1 ± 0.1
Type 1	L64A	—	0.18 ± 0.04	0.22 ± 0.04	0.96 ± 0.2
(Lost-recognition)	W70A	—	0.18 ± 0.03	0.20 ± 0.05	0.85 ± 0.2
	D133A	—	0.19 ± 0.03	0.20 ± 0.04	0.87 ± 0.2
	D136A	—	0.20 ± 0.02	0.20 ± 0.03	1.1 ± 0.2
	R186A	—	0.19 ± 0.03	0.20 ± 0.04	0.95 ± 0.1
Type 2	Q38A	—	0.45 ± 0.08	0.55 ± 0.14	2.8 ± 0.2
(Relaxed-specificity)	T40A	±[§]	0.42 ± 0.04	0.49 ± 0.06	1.1 ± 0.1
	L134A	—	0.39 ± 0.04	0.47 ± 0.12	1.5 ± 0.3
	Q137A	—	0.42 ± 0.05	0.52 ± 0.11	1.3 ± 0.3
	R142A	—	0.54 ± 0.12	0.57 ± 0.06	1.8 ± 0.3
	D705A	—	0.71 ± 0.05	0.74 ± 0.08	1.1 ± 0.4

Values are given as an average ± SD.
*All strains, except the null (*recC*::Tn*10*), were BNH963 with a plasmid harboring the indicated *recC* gene. The T4 *gene 2*$^-$ assays were performed in a Su^0 with the same plasmids.
[†]Size of bacteriophage λ plaque in millimeters.
[‡]Recombination hotspot activity as defined in *Materials and Methods*.
[§]The number of plaque-forming units was 0.014 ± 0.012, whereas the values for wild-type and other mutants were less than 0.00005 (relative to the *recC* null mutant, which was assigned a relative value of 1).

Table S3. Residues involved in χ recognition in *E. coli* and their equivalents in organism that recognize known χ sequences

Chi sequence	Mutation class / Residue	2 Q38	2 T40	1 L64	1 W70	1 D133	2 L134	1 D136	2 Q137	2 R142	1 R186	2 D705
5′-GCTGGTGG	*Escherichia coli*	—	—	—	—	—	—	—	—	—	—	—
	Shigella Sonnei	—	—	—	—	—	—	—	—	—	—	—
	Salmonella typhimurium	—	—	—	—	—	—	—	—	—	—	—
	Klebsiella pneumoniae	—	—	—	—	—	—	—	—	—	—	—
	Citrobacter koseri	—	—	—	—	—	—	—	—	—	—	—
	Proteus mirabilis	—	P	—	—	—	—	—	—	—	—	—
	Serratia proteamaculans	—	P	—	—	—	—	—	—	—	—	—
5′-GNTGGTGG	*Haemophilus influenzae*	—	P	M	—	—	—	—	—	—	—	—
Conservation* (%)		68	6	48	64	66	66	72	65	97	66	90

A "—" indicates the same sequence as in *E. coli* RecC.
*Identity is reported after alignment by MAFFT; only one sequence from each of 112 species was selected to avoid redundancy and bias.

Table S4. *E. coli* strains, bacteriophage, and plasmids used in this study

Strain	Genotype (alternate designation)	Source or reference
E. coli strains		
DH5	*deoR endA1 gyrA96 hsdR17 recA1 relA1 supE44 thi-1* F^- λ^-	S.C.K. lab collection
AB1157	*supE44 thr-1 ara-14 leuB6* Δ*(gpt-proA)62 lacY1 tsx-33 galK2 hisG4 rfbD1 mgl-51 rpsL kdgK51 xyl-5 mtl-1 argE3 thi-1* λ^- F^-	(1)
BNH963	AB1157 *recC*::Tn*10*	Present work
V330	(*recC-argA*)234 su^0	(2)
BNH1486	V330 [pMS421, pPB700, pWKS6]	Present work
JM1	*recB21 recC22 sbcA20 supF*	S.C.K. lab collection, (3)
BIK808	C600 λ^r *recB21 supE*	(3)
Bacteriophage strains		
T4 2^-	T4 *gene 2*	S.C.K. lab collection
LIK916	λ *Bam1*^{0} Δ*B int* Δ*(red-gam) imm21 nin5 shn6*^{0}	(3)
LIK950	λ *Bam1*^{0} Δ*B int::*χ^{+C157} Δ*(red-gam) imm21 nin5 shn6*^{0}	(3)
LIK1068	λ *Jh int4* Δ*(red-gam) cI26 sam7*	(3)
Plasmids:		
pACYC184	p15A ori Tet Cml	S.C.K. lab collection
pMS421	*lacI*^{q} expression vector (Spc)	(7)
pPB700	*recB* expression vector (Amp)	(5)
pPB520	*recC* expression vector (Cml)	(5)
pWKS6	*recD* expression vector (Kan)	(5)
pNH336	*recC*^{WT} expression vector	Present work
pNH504	pNH336 derivative, *recC*^{Q38A}	Present work
pNH738	pNH336 derivative, *recC*^{T40A}	Present work
pNH641	pNH336 derivative, *recC*^{Q44A}	Present work
pNH727	pNH336 derivative, *recC*^{N59A}	Present work
pNH740	pNH336 derivative, *recC*^{D61A}	Present work
pNH371	pNH336 derivative, *recC*^{F62A}	Present work
pNH629	pNH336 derivative, *recC*^{L64A}	Present work
pNH388	pNH336 derivative, *recC*^{S67A}	Present work
pNH361	pNH336 derivative, *recC*^{W70A}	Present work
pNH631	pNH336 derivative, *recC*^{D71A}	Present work
pNH742	pNH336 derivative, *recC*^{R75A}	Present work
pNH730	pNH336 derivative, *recC*^{K82A}	Present work
pNH667	pNH336 derivative, *recC*^{E83A}	Present work
pNH395	pNH336 derivative, *recC*^{N87A}	Present work
pNH365	pNH336 derivative, *recC*^{K88A}	Present work
pNH586	pNH336 derivative, *recC*^{D133A}	Present work
pNH1041	pNH336 derivative, *recC*^{L134A}	Present work
pNH634	pNH336 derivative, *recC*^{D136A}	Present work
pNH374	pNH336 derivative, *recC*^{Q137A}	Present work
pNH375	pNH336 derivative, *recC*^{Y141A}	Present work
pNH592	pNH336 derivative, *recC*^{R142A}	Present work
pNH386	pNH336 derivative, *recC*^{R186A}	Present work
pNH519	pNH336 derivative, *recC*^{R186H}	Present work
pNH585	pNH336 derivative, *recC*$^{R186C-D133A}$	Present work
pNH540	pNH336 derivative, *recC*$^{E399G-D414A}$	Present work
pNH537	pNH336 derivative, *recC*^{D414A}	Present work
pNH620	pNH336 derivative, *recC*^{D416A}	Present work
pNH379	pNH336 derivative, *recC*^{Y418A}	Present work
pNH553	pNH336 derivative, *recC*^{F421A}	Present work
pNH382	pNH336 derivative, *recC*^{R442A}	Present work
pNH569	pNH336 derivative, *recC*$^{V450G-R687A}$	Present work
pNH560	pNH336 derivative, *recC*^{Q652A}	Present work
pNH509	pNH336 derivative, *recC*^{T663A}	Present work
pNH383	pNH336 derivative, *recC*^{R668A}	Present work
pNH568	pNH336 derivative, *recC*^{R687A}	Present work
pNH1029	pNH336 derivative, *recC*^{M697A}	Present work
pNH751	pNH336 derivative, *recC*^{D705A}	Present work
pNH757	pNH336 derivative, *recC*^{D705H}	Present work
pNH596	pNH336 derivative, *recC*^{R706A}	Present work
pNH512	pNH336 derivative, *recC*^{R708A}	Present work
pNH743	pNH336 derivative, *recC*$^{A766V-R75A}$	Present work
pNH1045	pNH336 derivative, *recC*$^{Q137A-R142A}$	Present work

Table S4. Cont.

Strain	Genotype (alternate designation)	Source or reference
pNH1035	pNH336 derivative, *recC*$^{Q137A-D705A}$	Present work
pNH1038	pNH336 derivative, *recC*$^{Q142A-D705A}$	Present work
pNH1043	pNH336 derivative, *recC*$^{Q137A-R142A-D705A}$	Present work

Table S5. Primers used for site-directed mutagenesis

Mutation (in plasmid)	Primer name	Sequence
Q38A	Q38A	5′- AGATGATTCTGGTGGCAAGTACCGGTATGGC
(pNH504)	Q38A-R	5′- GCCATACCGGTACTTGCCACCAGAATCATCT
T40A	T40A	5′- CTGGTGCAAAGTGCCGGTATGGCACAG
(pNH738)	T40A-R	5′- CTGTGCCATACCGGCACTTTGCACCAG
Q44A	Q44A	5′- GTACCGGTATGGCAGCGTGGCTGCAAATGAC
(pNH641)	Q44A-R	5′- GTCATTTGCAGCCACGCTGCCATACCGGTAC
N59A	N59A	5′- GTTTGGTATTGCGGCAGCCATTGATTTTCCGCTGC
(pNH727)	N59A-R	5′- GCAGCGGAAAATCAATGGCTGCCGCAATACCAAAC
D61A	D61A	5′- GCGGCAAACATTGCTTTTCCGCTGCCAG
(pNH740)	D61A-R	5′- CTGGCAGCGGAAAAGCAATGTTTGCCGC
F62A	F62A	5′- GGCAAACATTGATGCTCCGCTGCCAGCGA
(pNH371)	F62A-R	5′- TCGCTGGCAGCGGAGCATCAATGTTTGCC
L64A	L64A	5′- CAAACATTGATTTTCCGGCGCCAGCGAGCTTTATCTG
(pNH629)	L64A-R	5′- CAGATAAAGCTCGCTGGCGCCGGAAAATCAATGTTTG
S67A	S67A	5′- CCGCTGCCAGCGGCCTTTATCTGGGAT
(pNH388)	S67A-R	5′- ATCCCAGATAAAGGCCGCTGGCAGCGG
W70A	W70A	5′- GCGAGCTTTATCGCGGATATGTTCGTC
(pNH361)	W70A-R	5′- GACGAACATATCCGCGATAAAGCTCGC
D71A	D71A	5′- GAGCTTTATCTGGGCTATGTTCGTCCGGG
(pNH631)	D71A-R	5′- CCCGGACGAACATAGCCCAGATAAAGCTC
R75A	R75A	5′- CTGGGATATGTTCGTCGCGGTGTTACCGGAAATC
(pNH742)	R75A-R	5′- GATTTCCGGTAACACCGCGACGAACATATCCCAG
K82A	K82A	5′- TACCGGAAATCCCCGCAGAGAGCGCCTTTAAC
(pNH730)	K82A-R	5′- GTTAAAGGCGCTCTCTGCGGGGATTTCCGGTA
E83A	E83A	5′- CGGAAATCCCCAAAGCGAGCGCCTTTAACA
(pNH667)	E83A-R	5′- TGTTAAAGGCGCTCGCTTTGGGGATTTCCG
N87A	N87A	5′- GAGAGCGCCTTTGCCAAACAGAGCATG
(pNH395)	N87A-R2	5′- CATGCTCTGTTTGGCAAAGGCGCTCTC
K88A	K88A	5′- AGCGCCTTTAACGCACAGAGCATGAGC
(pNH365)	K88A-R	5′- GCTCATGCTCTGTGCGTTAAAGGCGCT
D133A	D133A	5′- TCAAAAGCGGCGGCCCTGTTTGACCAG
(pNH586)	D133A-R	5′- CTGGTCAAACAGGGCCGCCGCTTTTGA
L134A	L134A	5′- AGCGGCGGACGCGTTTGACCAGTATCTGGTCTATCGT
(pNH1041)	L134A-R	5′-CTGGTCAAACGCGTCCGCCGCTTTTGAGGAAAGCTGG
D136A	D136A	5′- GCGGACCTGTTTGCCCAGTATCTGGTC
(pNH634)	D136A-R	5′- GACCAGATACTGGGCAAACAGGTCCGC
Q137A	Q137A	5′- CGGACCTGTTTGACGCGTATCTGGTCTATCG
(pNH374)	Q137A-R	5′- CGATAGACCAGATACGCGTCAAACAGGTCCG
Y141A	Y141A	5′- CAGTATCTGGTCGCTCGTCCGGACTGG
(pNH375)	Y141A-R	5′- CCAGTCCGGACGAGCGACCAGATACTG
R142A	R142A	5′- GTATCTGGTCTATGCTCCGGACTGGCTGG
(pNH592)	R142A-R	5′- CCAGCCAGTCCGGAGCATAGACCAGATAC
R186A	R186A	5′- CCGCGCTGGCACGCCGCCAATCTCTAT
(pNH386)	R186A-R	5′- ATAGAGATTGGCGGCGTGCCAGCGCGG
R186H	R186H	5′- CCGCGCTGGCACCACGCCAATCTCTAT
(pNH519)	R186H-R	5′- ATAGAGATTGGCGTGGTGCCAGCGCGG
D414A	D414A	5′- CGTGATGGTGGCTGCTATCGACAGCTACA
(pNH537)	D414A-R	5′- TGTAGCTGTCGATAGCAGCCACCATCACG
D416A	D416A2	5′- GTGGCTGATATCGCCAGCTACAGTCCG
(pNH620)	D416A-R2	5′- CGGACTGTAGCTGGCGATATCAGCCAC
Y418A	Y418A	5′- CTGATATCGACAGCGCCAGTCCGTTTATTCA
(pNH379)	Y418A-R	5′- TGAATAAACGGACTGGCGCTGTCGATATCAG
F421A	F421A	5′- ACAGCTACAGTCCGGCTATTCAGGCTGTGTT
(pNH553)	F421A-R	5′- AACACAGCCTGAATAGCCGGACTGTAGCTGT

Table S5. Cont.

Mutation (in plasmid)	Primer name	Sequence
R442A	R442A	5′- GCCATTTCCGACGCTCGTGCGCGGCAG
(pNH382)	R442A-R	5′- CTGCCGCGCACGAGCGTCGGAAATGGC
Q652A	Q652A	5′- GGCGAGAAAACGCGCGCTGATACGTTC
(pNH560)	Q652A-R	5′- GAACGTATCAGCGCGCGTTTTCTCGCC
T663A	T663A	5′- CATTGGCATCAGAGCACAAATGTTAAC
(pNH509)	T663A-R	5′- GTTAACATTTGTGCTCTGATGCCAATG
R668A	R668A	5′- TTGAACGGAATTGAAGCCATTGGCATCAGAG
(pNH383)	R668A-R	5′- CTCTGATGCCAATGGCTTCAATTCCGTTCAA
R687A	R687A	5′- TGGCGCAAGCTGAGCTGGATAAACGCC
(pNH568)	R687A-R	5′- GGCGTTTATCCAGCTCAGCTTGCGCCA
M697A	M697A	5′-CTTTGACCTGGCCAGCCAGAAACCGAAGCGTGGCGACC
(pNH1029)	M697A-R	5′-GTTTCTGGCTGGCCAGGTCAAAGCCCAATGGCGCAAGC
D705A	D705A	5′- CCGAAGCGTGGCGCCCGTAGCCGTCGC
(pNH751)	D705A-R	5′- GCGACGGCTACGGGCGCCACGCTTCGG
R706A	R706A	5′- ATCGCGACGGCTAGCGTCGCCACGCTT
(pNH596)	R706A-R	5′- AAGCGTGGCGACGCTAGCCGTCGCGAT
R708A	R708A	5′- GTCGTCATCGCGAGCGCTACGGTCGCC
(pNH512)	R708A-R	5′- GGCGACCGTAGCGCTCGCGATGACGAC

The directed mutation is underlined.[†]