RecA Protein: Structure, Function, and Role in Recombinational DNA Repair[1]

ALBERTO I. ROCA AND MICHAEL M. COX

Department of Biochemistry
College of Agriculture and Life Sciences
University of Wisconsin
Madison, Wisconsin 53706

I. On the Function of Homologous Genetic Recombination in Bacteria .. 130
A. The Function of Homologous Genetic Recombination in Bacteria Is Recombinational DNA Repair 132
B. The Initiation of Recombination in an *Escherichia coli* Cell Normally Requires DNA Damage 136
II. The Structure of RecA Protein 138
A. General Properties 138
B. Sequence Alignments 138
C. Expanded Discussion of Structure–Function Relationships 162
III. RecA Protein Interactions with Its Ligands *in Vitro*; Biochemical Approaches 171
A. DNA Binding 171
B. Polar Assembly and Disassembly of RecA Filaments, and the Importance of 3′ DNA Ends 176
C. ATP Hydrolysis 179
IV. RecA Protein-mediated DNA Strand Exchange 184
A. Overview of the Reaction 184
B. DNA Pairing 186
C. Unidirectional Extension of the Hybrid DNA and the Role of ATP Hydrolysis 195
D. Exchange Reactions with Four DNA Strands 200
V. Interaction of RecA Protein with Other Proteins 200
A. The Single-Strand DNA Binding Protein 201
B. The RecF, RecO, and RecR Proteins 203
C. The RuvA and RuvB Proteins 206
D. Other Proteins Affecting RecA Activities 208
VI. Other Functions of RecA Protein *in Vivo* 208
A. The RecA Coprotease Function 208
B. SOS Mutagenesis 209
C. Chromosome Partitioning 210
D. Induced Stable DNA Replication 210

[1] Abbreviations: ATPγS, adenosine-5′-*O*-(3-thiotriphosphate); dsDNA, double-stranded DNA; pI, isoelectric point; pK, negative logarithm of dissociation constant; ssDNA, single-stranded DNA; UV, ultraviolet light.

Progress in Nucleic Acid Research and Molecular Biology, Vol. 56

Copyright © 1997 by Academic Press.
All rights of reproduction in any form reserved
0079-6603/97 $25.00

VII. Epilogue: Relating RecA Biochemistry to DNA Repair 210
References 213

The RecA protein plays a central role in recombinational processes in *Escherichia coli*. In this capacity, it binds to two DNA molecules and aligns homologous sequences within them. It then promotes a DNA strand-exchange reaction that creates branched DNA recombination intermediates. The protein also has a regulatory function, serving as part of the system that induces the SOS response to DNA damage. As a regulatory function, RecA protein facilitates the autocatalytic cleavage of the LexA repressor, the λ repressor, and a few other proteins. Because RecA does not participate as a classical protease in this reaction, the effect of RecA on repressor cleavage is usually referred to as the coprotease function. RecA also plays a direct role in a process called SOS mutagenesis. This review focuses largely on the recombinational activities of RecA protein.

The RecA protein has been found in all bacteria in which it has been carefully sought, including *Mycoplasma* with its minimal genome (*1*). The *recA* gene has been sequenced in over 60 bacterial species. The wide distribution of bacterial *recA* genes, in classes of bacteria that took separate evolutionary paths during precambrian times, indicates that the protein evolved very early. Structural homologs of RecA protein have been found in Archaeans[2] (*2*) and in eukaryotes from yeast to humans (*3–5*). The RecA protein of *E. coli* is a 352-aminoacid polypeptide of M_r 37,842. This modestly sized polypeptide has separate binding sites for at least three stands of DNA, ATP, the LexA and λ repressors, and other RecA monomers. The active species of RecA protein implicated in most of its activities is a structure in which RecA protein monomers are assembled into a right-handed helical filament on DNA. Its multitude of functions is reflected in a complex enzymology that still has significant chapters unwritten.

I. On the Function of Homologous Genetic Recombination in Bacteria

In the summer of 1984 at the Cold Spring Harbor Laboratory, there was a symposium entitled "Recombination at the DNA Level." It focused broadly on recombination mechanisms and covered what was then known about

[2] Our description of molecular taxonomy follows the recommendations of Woese *et al.* (*2a*).

the mechanisms of homologous genetic recombination, site-specific recombination, and transposition in a marathon of over 90 presentations. One of the most memorable presentations was the final one, in which Alan Campbell was asked to summarize the entire meeting. A paper outlining his remarks was later published in the symposium volume (6). Following an inexorable tide of papers on recombination mechanisms, the following comments on general recombination stood out:

> The function of general recombination has hardly surfaced in this Symposium. Perhaps there was nothing constructive to say on the subject. I am left uncertain as to how many investigators in that area consider that the function is too obvious to require discussion, how many think that general recombination serves no useful function, and how many consider the question uninteresting or intractable.
>
> But there is a real question that needs to be answered sometime. Our immediate interest . . . [is] . . . in recombinational mechanisms. But we all know that the immediate selective value of those [recombination] genes and products does not depend on their role in reshuffling genes in natural populations.

The passage reflects a persistent temptation to study the mechanism of recombination without adequately considering its biological functions. The interest in mechanism is easily explained. Homologous genetic recombination is the cellular process underlying most classical and many modern methods of molecular genetics. A sexual union of two individuals of any species is accompanied by genetic recombination, resulting in new combinations of genetic alleles in the offspring. In the laboratory, introduction of DNA into a cell is also often accompanied by recombination, leading to an insertion of the DNA into the cellular genome. The recombination associated with genetic crosses allowed Gregor Mendel to ask and answer new questions about inheritance. Today it allows gene hunters to isolate the genes responsible for human genetic diseases. Recombination generates genetic diversity and the generation of diversity allows us to study genomes. Without the inherent capacity of homologous chromosomes to come together and exchange genetic information via recombination, it is hard to imagine where our understanding of chromosomes, genes, and their functions might be. As we learn more about genomes, we seek to change them by genetic engineering and human gene therapy. Changing a genome requires recombination. Requiring further pretext or context for pursuing an understanding of recombination mechanisms seems almost gratuitous.

However, Campbell's point is not irrelevant or even esoteric. A real understanding of the recombination mechanism is impossible without an appreciation for recombination function. The function most important to a cell is not necessarily the function most important to geneticists. The absence of talks on recombination function at the 1984 symposium (6) did not reflect a

dearth of ideas on the subject. Key ideas about recombination function were simply not represented.

A. The Function of Homologous Genetic Recombination in Bacteria Is Recombinational DNA Repair

In addition to its role in the generation of genetic diversity, homologous genetic recombination is required for recombinational DNA repair and is involved in the successful segregation of chromosomes at cell division in all organisms. Which of these functions can account for the maintenance of recombination systems in all cells? One place to look for a discussion of this question is in the literature addressing one of the most intractable questions in evolutionary biology: the evolution of sex. "Sex" is commonly defined as either genital union, gender, or mixis. Mixis is more or less synonymous with recombination, and this is the context for our discussion. In higher eukaryotes, recombination is closely linked to reproduction.

A number of researchers have argued that the creation of genetic diversity is the primary function of recombination in eukaryotes, providing the selection pressure for the maintenance of recombination systems (*7–10*). Viewed in this context, there are several major advantages to recombination (*7, 10*). Recombination allows rapid adjustment of phenotypes to meet a changing environment. Beneficial mutations from different lineages can be readily combined, and deleterious mutations can be eliminated. All of these factors may help account for the limited distribution of asexual eukaryotes, in which genomic change is largely limited by the rate of mutation. Without sex, the combination of two beneficial new mutations would require two independent mutational events. Most mutations are bad, and an inability to eliminate deleterious mutations by recombination may be especially important in restricting the distribution of asexual species. A small sexual population lacking a mutant-free phenotype can always generate one by recombination. A similar asexual population can generate a mutant-free phenotype only by back mutation. A detailed theory of the consequences of deleterious mutations on asexual populations was presented by Muller (*11*), and has been given the name "Muller's ratchet" (*7, 10, 12*). Mutation loads may even limit the size of the genome of asexual organisms (*10*).

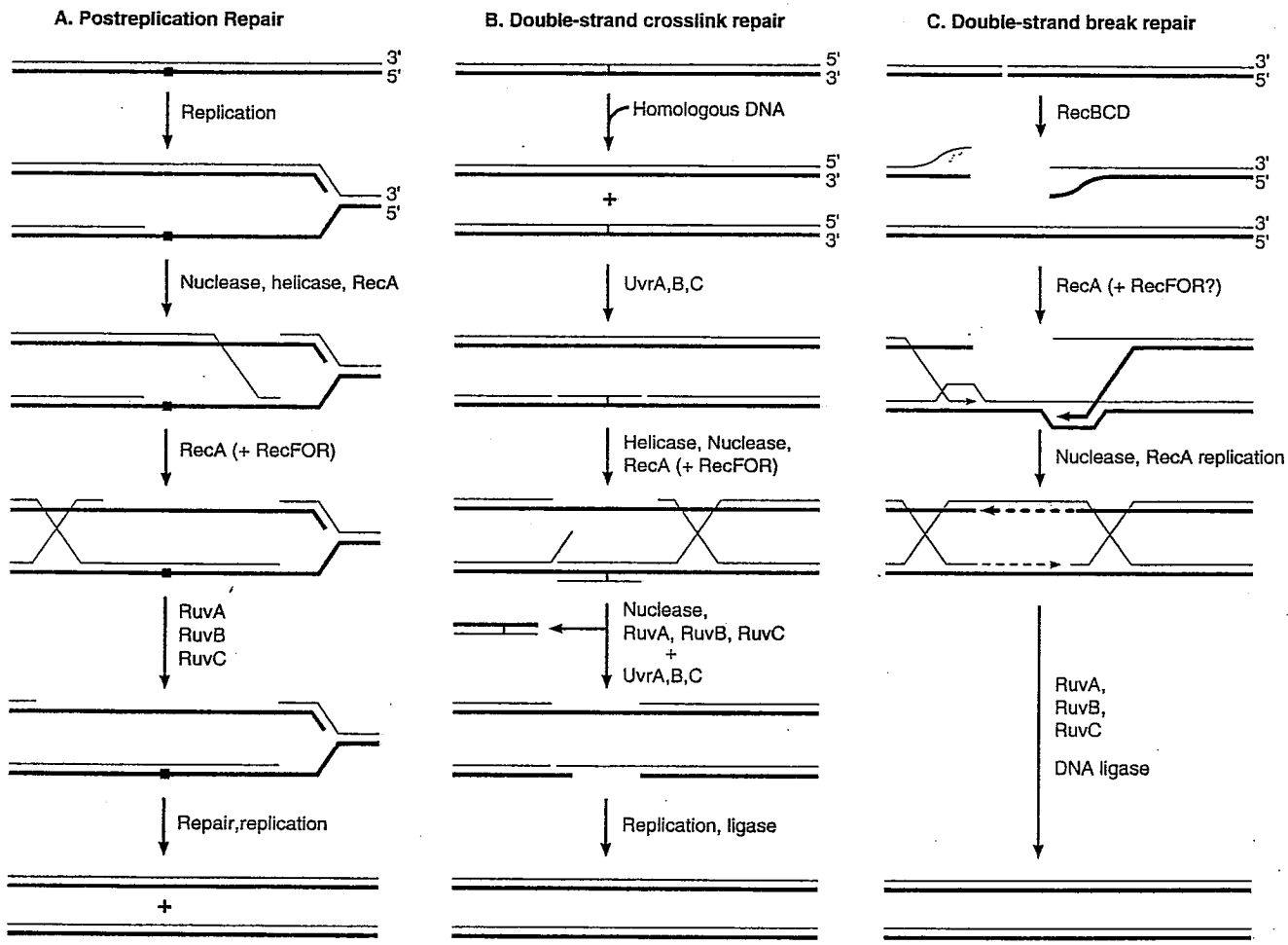
However, recombination comes at a price. For every 100 offspring, asexual organisms will contribute 100 individuals that can reproduce in the next generation. Sexual females will contribute only 50. The maintenance of males, along with the processes of meiosis and syngamy, represents a substantial energetic cost that must be borne by sexual species. Recombination may break up the same favorable combinations of mutant alleles that it creates. Recombination also does not provide a means for two or more rare genes, individually deleterious but beneficial when grouped, to spread

through a population together. The costs of recombination can lead to a short-term advantage for asexual populations over sexual ones. For these and other reasons, there has been a search for additional functions of recombination. One view holds that mechanisms for sexual genetic exchange arose at least in part from parasitic genetic elements (e.g., transposons) (*13*). A number of other researchers have argued that the primary function of eukaryotic recombination is to repair DNA (*14–18*). A significant advantage of this idea is that it provides a clear selective pressure for the maintenance of recombination systems at the level of individual organisms. The recent demonstration that the gene responsible for Bloom's syndrome in humans is a homolog of the bacterial *recQ* gene (*19*) [encoding a helicase involved in homologous recombination (*20, 21*)] provides an interesting link between recombination and DNA repair in eukaryotes. The gene for Werner's Syndrome also exhibits significant similarity to *recQ* (*22*).

The function of recombination and the selective advantage of sex in eukaryotes remain controversial. The major theses outlined above are not mutually exclusive, and several or all of the factors cited may have played a significant role in the evolution of eukaryotic sex.

The situation in bacteria is quite different. There is a broad consensus that recombination originated in bacteria or their progenitors as a DNA repair process (*7–10, 14, 16, 18, 23–27*). The requirement for DNA repair probably remains the major selection pressure for the universal maintenance of homologous recombination systems in bacteria (*26*). Sex is not linked to reproduction in bacteria. Sexual genetic exchanges are generally quite rare (*28–30*), and it is harder to invoke the creation of genetic diversity as a factor in the origin or maintenance of bacterial recombination systems. On the other hand, potentially mutagenic DNA lesions are virtually omnipresent (*25, 26*). Each bacterial cell growing aerobically in a rich laboratory broth suffers on average several thousand DNA lesions of all types in every generation (*26, 31*). When a bacterial cell lacks functional copies of key genes involved in recombination, the prototypical phenotype is an extreme sensitivity to DNA-damaging agents. Many cells mutant for recombination functions are very sick or dead (*26, 32*).

Recombinational DNA repair is a specialized process directed at DNA damage only when it occurs in certain contexts. Much DNA repair is made possible because DNA is double stranded. If one strand contains a lesion, a segment of the strand containing the lesion can be removed. The second strand can then be used as a template for replacement of the damaged strand segment. No recombination is required. However, there are important situations in which a second DNA strand is not available to direct repair. In some cases, cross-links or double-strand breaks produce damage in both strands. Alternatively, a DNA lesion can be located in a single-stranded region of the


A. Postreplication Repair
3'
5'
Replication
3'
5'
Nuclease, helicase, RecA
RecA (+ RecFOR)
RuvA
RuvB
RuvC
Repair,replication
+
B. Double-strand crosslink repair
5'
3'
Homologous DNA
5'
3'
+
5'
3'
UvrA,B,C
Helicase, Nuclease,
RecA (+ RecFOR)
Nuclease,
RuvA, RuvB, RuvC
+
UvrA,B,C
Replication, ligase
C. Double-strand break repair
3'
5'
RecBCD
3'
5'
3'
5'
RecA (+ RecFOR?)
Nuclease, RecA replication
RuvA,
RuvB,
RuvC
DNA ligase

DNA. In these situations, the information for faithful DNA repair must come from a different homologous DNA molecule via recombination.

Some current models for recombinational DNA repair are presented in Fig. 1. Each of the pathways is directed at a different type of damage, but there are a number of common themes. First, the pairing of two DNA molecules involves one single-stranded DNA and one duplex DNA. If single-stranded DNA is not already present, early stages of recombination are dominated by enzymes that produce it (helicases and nucleases). The pairing is likely to involve the transient formation of a novel triplex DNA intermediate. Second, pairing is followed by a strand exchange in which the single strand is paired with its complement in the duplex, and the like strand in the duplex is displaced. The product duplex is referred to as heteroduplex DNA if it contains one or more mismatches or lesions, or hybrid DNA if it does not. Third, the strand exchange usually results in the formation of branched DNA intermediates. Because they are formed from homologous DNA molecules, the branches can move in a process known as branch migration. Branch migration can continue into regions where both DNA molecules are duplex, forming the branched recombination intermediate first proposed by Holliday (*33*), and thus called the Holliday intermediate. Finally, repair is completed by enzymes that process the branched DNA intermediates, including specialized helicases and nucleolytic activities, DNA polymerases, and DNA ligase.[3]

Consider in more detail what is likely to happen when a replication fork encounters an unrepaired DNA lesion (Fig. 1A). DNA polymerases will not insert nucleotides opposite the lesion, so the replication halts. Whether or not replication resumes at some point further along the template, the lesion is left in a single-stranded DNA gap and must be repaired in a process known as postreplication DNA repair. The repair pathway shown is based on that proposed by Howard-Flanders and colleagues (*34*). The single-stranded DNA in the gap is paired with homologous duplex DNA derived from the other side of the replication fork. Nucleolytic processing of the donor duplex leads to formation of the branched recombination intermediate, and branch migration across the DNA lesion provides an undamaged complementary strand at that location. By rendering the DNA surrounding the lesion double-stranded, the lesion can be repaired by the excision repair pathways. Although the

[3] In Chapter 1 of his recent book (*33a*), Kuzminov has provided a detailed and lucid description of the evolution of thought about models for recombinational repair.

FIG. 1. Models for recombinational DNA repair. The postreplication repair model is adapted from West *et al.* (*34*). The double-strand cross-link repair pathway is based on a model by Cheng *et al.* (*414*). The double-strand break repair model shown is based on models presented by Szostak *et al.* (*35*) and Smith (*29*). See text for details.

details differ for the repair of double-strand breaks and cross-links, the same general sequence of steps is evident. The double-strand break-repair pathway (Fig. 1B) is also similar to the pathways proposed for most meiotic recombination in eukaryotes (*35*) and conjugational recombination in bacteria (*29*).

The central recombination activity in bacteria is the RecA protein. RecA is found in all bacteria, and clear structural and functional homologs have been identified in eukaryotes, from yeast to humans. RecA is broadly responsible for facilitating recombination steps involving DNA pairing and branch migration. To initiate, regulate, and complete recombinational exchanges, RecA is augmented by over a dozen additional proteins. The overall system is complex, sophisticated, adaptable to different repair and recombination scenarios, and critical for cell survival. *Escherichia coli* cells lacking a functional *recA* gene grow poorly. About 50% of the cells in a typical culture are dead and 10% contain no DNA (*32*). Cells containing both a *recA* null mutation and an additional mutation tending to increase DNA strand breaks (*lig, dam, xth, ung,* and *polA*) are inviable (*36–38*). Recombination and recombinational repair are all but eliminated in a *recA* cell.

A dramatic example of the relationship between recombination and repair can be seen in recent work on the bacterium *Deinococcus radiodurans,* which can survive a dose of radiation several thousand times that required to kill other organisms. This extraordinary radiation resistance depends completely on an intact *recA* gene (*39–41*). Homologous chromosomes in these bacteria may be arranged in pairs, with each pair permanently linked by many Holliday junctions to facilitate rapid repair of DNA breaks (*41*). This same mechanism may explain the resistance of *D. radiodurans* to DNA strand breaks caused by prolonged desiccation (*42*).

B. The Initiation of Recombination in an *Escherichia coli* Cell Normally Requires DNA Damage

Intracellular recombination in bacteria exhibits a demonstrable dependence on DNA damage. Some of the literature on which this statement is based has been reviewed (*25, 26*). New work continues to reinforce the theme. During infection of homoimmune *E. coli* lysogens, undamaged nonreplicating λ phage DNA circles undergo very little recombination. Prior ultraviolet (UV) irradiation of phages dramatically elevates recombinant frequencies (*43–45*). The initiation of λ recombination is dependent to a large extent on the exonuclease functions associated with methyl-directed mismatch repair, suggesting that it is the generation of single-strand gaps in the DNA that leads to the initiation of recombination (*45*).

Other observations correlate a requirement for RecA protein with oxidative DNA damage. Oxidative damage accounts for the major portion of the

thousands of genomic DNA lesions suffered by every *E. coli* cell in every generation when grown aerobically (*26, 31*). Introduction of a mutation in the *fur* gene produces an iron overload, leading to much increased oxidative stress and DNA damage, including lethal and mutagenic lesions (*46*). Double *fur recA* mutants die after a shift from anaerobic to aerobic conditions. Functional recombinational repair is therefore necessary for protection from the killing effects of aerobic environments, but SOS induction is not (*46*).

The prevalence of oxidative damage can also be correlated directly with the frequency of intracellular recombination. A duplication strain carrying a lac^+ allele between two direct-order flanking sequence repeats generates lac^- clones by recombination. Among 40 mutants isolated by J. Roth and colleagues that did not produce lac^- recombinants, 3 had mutations in the *recA* gene as expected. Unexpected was the finding that many mutations that eliminate recombination also cause phenotypes suggesting a block in oxidative metabolism (e.g., the citrate cycle and/or electron transport). These results suggest that intermediates or by-products (e.g., oxygen radicals) might be important in the generation of initiating substrates (nicks, breaks), which are a prerequisite for recombination (*46a*).

It is not the DNA damage per se that leads to the initiation of recombination, but the double-stranded DNA ends and single-stranded gaps that appear as a result of the damage. For example, the RecBCD enzyme is a key activity in bacterial double-strand break repair. RecBCD does not bind to DNA lesions; it binds to DNA ends created by DNA damage. Similarly, RecA filament formation is not stimulated by DNA lesions. Instead, RecA filaments form on single-stranded DNA gaps made available by other processes. The importance of the DNA gaps and breaks associated with damage as opposed to the damage itself can be seen in another RecA activity, the induction of the SOS response. Filamentous phages defective in minus-strand synthesis cause an induction of the SOS response when used to infect *E. coli* cells, whereas the wild-type phage did not (*47*). The single-stranded phage DNA, in the absence of DNA damage, was sufficient to induce the SOS response *in vivo*.

Recombination that occurs during the *rare* sexual exchange can be viewed as a by-product of the inherent recombinational repair processes. However, initiation of recombination during conjugation or transduction does not require DNA damage. During conjugation or transduction, the double-strand break-repair path (RecBCD) is simply appropriated to bring about an efficient genetic exchange. This appropriation comes about because the processes of conjugation and transduction present the cell with the substrates, DNA ends, and single-stranded DNA gaps, required to initiate recombination efficiently. The RecF pathway may be appropriated the same way in certain conjugational contexts. The process of conjugational recom-

bination is as efficient as it is because its evolution was shaped by the biological imperative to repair DNA.

II. The Structure of RecA Protein

A. General Properties

Bacterial strains engineered to produce the RecA protein at very high levels are widely available. The RecA protein is easily purified in large quantities (*48–53*), although care must be taken to remove persistent contamination with exonuclease I (*54*). The protein is quite soluble in aqueous buffers and is stable for many months if stored at −70°C. The isoelectric point of the purified protein has been measured in at least six studies, giving p*I* values from 5.0 (*55*) to 6.2 (*56*). Averaging all the measurements (*53, 55–59*) gives a p*I* of 5.6. In solution, the protein has a tendency to aggregate into oligomers, filaments, and bundles of filaments (*60–63*). The aggregation state of a given solution of the protein is affected by pH, ionic strength, and temperature. Concentrated solutions of the protein can appear almost opalescent.

The structure of RecA protein has been determined at 2.3 Å resolution (*64, 65*). There is a major central domain flanked by two smaller subdomains at the N and C termini (Fig. 2). Monomers in the crystal are packed to form a continuous spiral filament, with six monomers per right-handed helical turn (Fig. 3). These results are consistent with other studies showing that RecA protein forms a helical nucleoprotein filament on DNA. The filament exhibits a deep helical groove. A variety of physical studies (detailed below) indicate that the groove can accommodate up to three DNA strands. The structure was solved in two forms, one with bound ADP (*65*) and the other with no nucleotide (*64*). Both are highly informative, although some work indicates that they represent an inactive (ADP-bound) form of the protein (*3, 66–68*).

B. Sequence Alignments

Sequence alignments of related proteins from different species are commonly used to identify functionally important segments of a protein's primary structure (*69*). In a previous review, we published a multiple sequence alignment of 16 bacterial RecA proteins (*70*). This and other alignments helped to identify and elucidate the functions of several conserved regions of the RecA protein. For example, the Walker A motif defined by the sequence analysis of nucleotide binding proteins (including RecA) (*71*) has subsequently been shown to bind the phosphates of NTPs in the crystal structures of proteins such as Ras p21 (*72*), Ef-Tu (*73*), and RecA (*65*). RecA alignments have also helped identify RecA homologs in other classes of organisms (*4*).

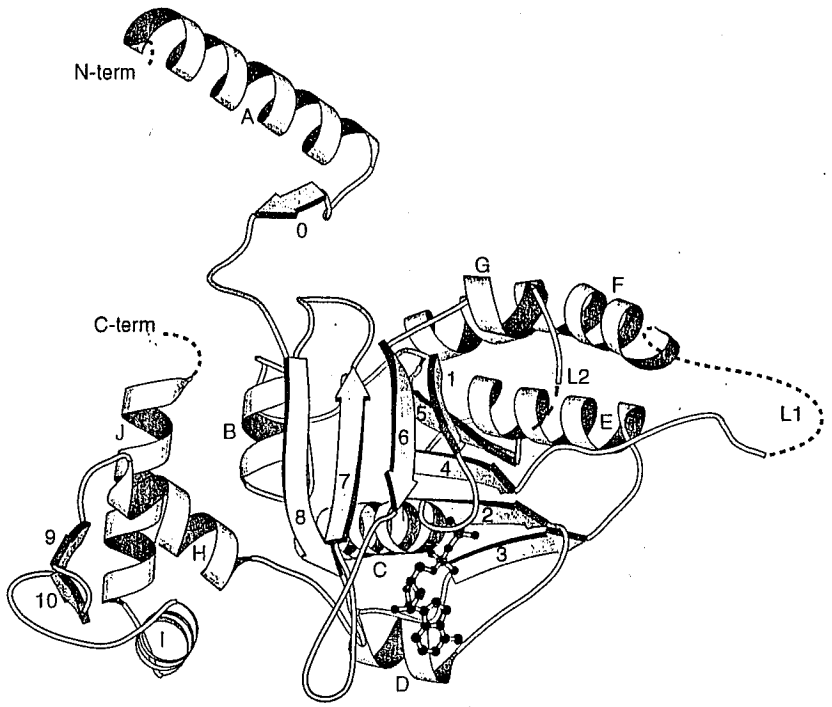


FIG. 2. A RecA monomer, based on the structure determined by Story *et al.* (*65*). The α-helical regions are labeled A through J; the β strands are labeled 0 through 10. Dashed lines denote regions that are disordered in the crystal structure.

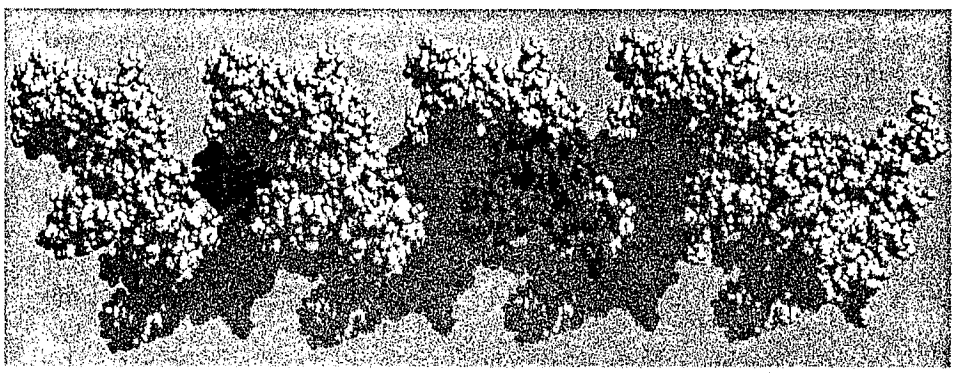


FIG. 3. A RecA filament, based on the structure published by Steitz and colleagues (*64, 65*). Four turns of a helical filament are shown in a space-filling representation. There are 6 RecA monomers per turn, or 24 altogether in the structure presented. Two monomers are shaded dark.

Several recent advances recommend a renewed examination of RecA sequences as an avenue for generating insights into structure–function relationships in this protein. First, the number of available bacterial RecA sequences has increased more than fourfold since the original alignment was published. The data set is now sufficient to use for phylogenetic analysis (*74, 75*). Second, the availability of the structure of the RecA protein, including an ADP-bound form (*64*), provides an enhanced context for sequence analysis. Third, the Rad51 protein of yeast has been identified as a structural (*3*) and functional (*76*) homolog of RecA. Finally, several additional eukaryotic RecA homologs have been advanced based on sequence identities, notably Dmc1 of yeast (*77*) and Rec2 of *Ustilago* (*78*). Together, these sequences have the potential to eliminate noise in the alignments and focus attention on structural motifs that are essential and perhaps unique to RecA function.

An expanded set of alignments is presented here, along with an analysis of some of the information to be derived from it. The results focus attention on a segment of the RecA primary structure that has been little explored to date, and that is likely to play a role in coupling ATP hydrolysis to a protein conformational change. The results also suggest the existence of a RecA homologue in *E. coli*.

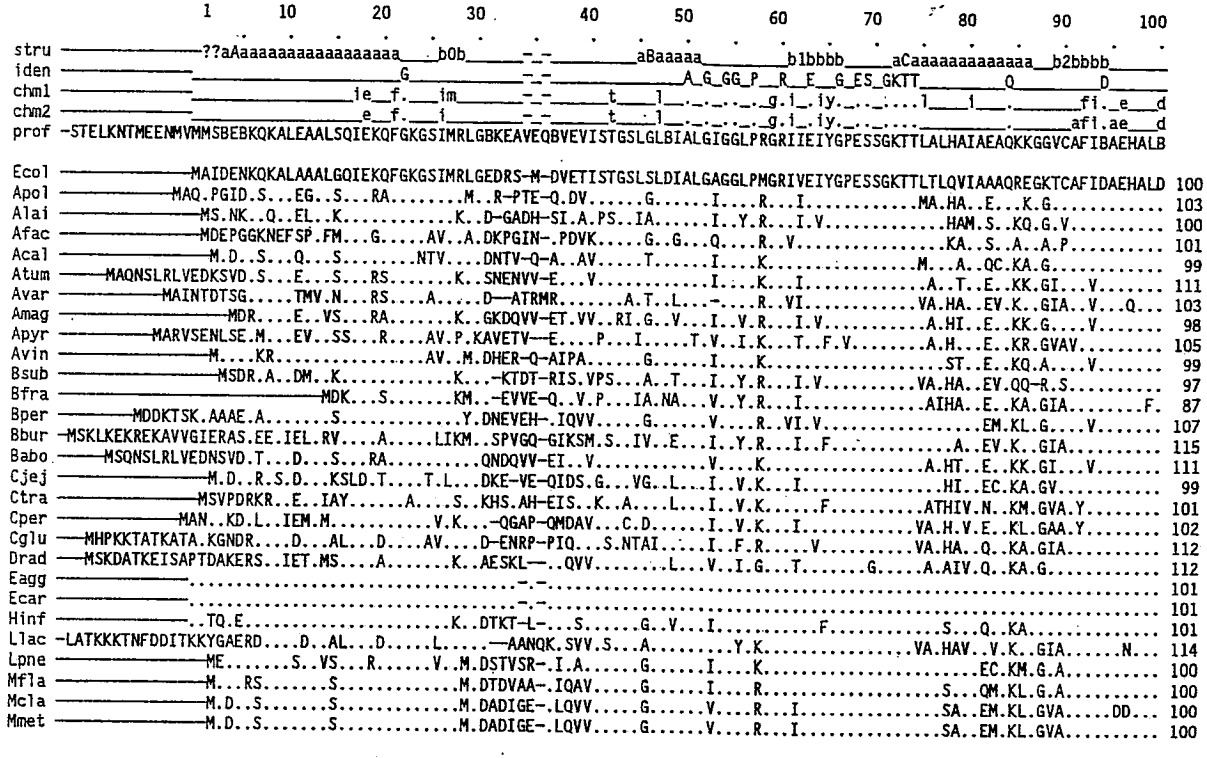
1. Bacterial RecA Alignment

The alignment of 64 full-length bacterial RecA homologs is depicted in Fig. 4. The sequences are from the GenBank and EMBL data bases and the respective accession numbers are given at the end of each sequence. Four of the sequences, indicated by dates, are personal communications. There were 27 sequence errors corrected during the preparation of this alignment (A. I. Roca, unpublished data). The summary lines at the top of Fig. 4 provide an overview of the conserved features of the alignment. For example, the Walker A box (involved in ATP binding) from residues 66 to 73 is highly conserved with six invariant residues (pointed out in the "iden" summary line) and two positions where nonconservative amino acid substitutions occur. This region is flanked by positions where chemically conservative changes occur, such as aromatic side chains at position 65 and hydrophobic residues at position 75, shown in the "chm1" and "chm2" summary lines. The alignment is placed within the context of the secondary structural elements ("stru" summary line) defined by the RecA crystal structure (*64*). For instance, the Walker A box is a loop that occurs between β-strand 1 and α-helix C. Finally, the fifth summary line ("prof") is a consensus sequence for bacterial RecA proteins calculated as the highest scoring residue at that position of the weighted alignment in the RecA profile (see Fig. 4 legend).

In general, the RecA protein is very well conserved. The percent identity

with respect to the *E. coli* sequence ranges from 49% for *Mycoplasma pulmonis* to 100% for *Shigella flexneri*. Proteins with identity as low as 30% can be structurally related (*79*), so it is likely that the structure of the bacterial homologs will resemble that of the *E. coli* RecA crystal structure (*64*). Indeed, electron microscopy shows that the RecA nucleoprotein filament from *Thermus aquaticus* is very similar in structure to that produced by the *E. coli* RecA protein (*80*). There are 59 residues that are invariant among the 64 bacterial sequences. Based on two slightly different sets of assumptions (chm1 versus chm2 summary lines), either 100 or 106 residues were identified wherein changes are restricted to chemically conservative substitutions.

FIG. 4. Multiple sequence alignment of 64 bacterial RecA proteins. In the alignment, a dash (-) represents a gap introduced in a sequence to optimize the alignment with the other RecA proteins. A period (.) indicates a residue identical to the amino acid found in the *E. coli* RecA sequence at the top of the alignment. An asterisk (*) marks the position of a protein intron found in the RecA proteins of *Mycobacterium tuberculosis* and *Mycobacterium leprae* (*91*). The GenBank accession numbers for the different published sequences are given at the end of the alignment (GenBank release 86 and EMBL release 41). Sequences with dates listed are personal communications. Above the alignment are several summary lines, calculated with a C program (frmtMSA) written by A. I. Roca. Here stru provides information on the secondary structural elements from the *E. coli* RecA crystal structure, where "a" denotes α-helices, "b" denotes β-strands, "1" denotes functional loops, and "?" denotes disordered regions (*64*). Each secondary structural element is uniquely designated by a letter or number placed at the second character in the string of characters defining the element. The α helices are labeled A to J, the β strands are labeled 0 to 10, and the loops are labeled 1 and 2 (*64*). Next, iden represents the 58 residues identical in all the sequences of the alignment; chm1 lists 105 chemically conservative residue substitutions based on the following classification: acidic (DE), small aliphatic (AG), amides (NQ), aromatic (FWY), basic (HKR), cysteine (C), hydrophilic (ST), hydrophobic (ILMV), and proline (P) (M. Gribskov, personal communication). Then chm2 enumerates 99 chemically conservative residue substitutions based on a different classification: acidic (DE), aliphatic (AGILV), amides (NQ), aromatic (FWY), basic (HKR), sulfur (CM), hydroxyl (ST), and imino (P) (*415*). In the chm1 and chm2 summary lines, the corresponding residue of the *E. coli* RecA protein sequence is the one listed. Also, invariant residues in these same summary lines are indicated by a period (.) for clarity. Finally, prof is a weighted consensus sequence of bacterial RecA proteins, generated by the PROFILEMAKE program of the GCG package using the BLOSUM62 symbol comparison table for scoring different amino acid substitutions (*416*). PROFILEMAKE is an implementation of profile analysis (*95*). The alignment of bacterial RecA sequences was generated with the PILEUP program of the GCG Wisconsin Sequence Analysis Package (Release 8.0, Genetics Computer Group, Inc.) (*417*) using default parameter values. PILEUP produces a progressive, pairwise alignment of sequences (*418*). A few manual adjustments of small gaps were performed after visual inspection of the alignment. The numbering of residues begins at alanine—the N-terminal residue found in the mature *E. coli* K-12 RecA protein (*419*). Amino acid positions cited throughout the text refer to the sequence of *E. coli* RecA unless stated otherwise. The bacterial alignment was weighted using an unpublished program (M. Gribskov, personal communication) that is based on a modified version of the Felsenstein algorithm (*420*).


1 10 20 30 40 50 60 70 80 90 100
stru ??aAaaaaaaaaaaaaaaaa bOb aBaaaaa b1bbbb aCaaaaaaaaaaaa b2bbbb
iden G A_G_GG_P__R__E__G_ES_GKTT Q D
chm1 ie__f.___im t___l g.i_.iy. l___i fi._e___d
chm2 e__f.___i t___l g.i_.iy. afi.ae___d
prof -STELKNTMEENMVMMSBEBKQKALEAALSQIEKQFGKGSIMRLGBKEAVEQBVEVISTGSLGLBIALGIGGLPRGRIIEIYGPESSGKTTLALHAIAEAQKKGGVCAFIBAEHALB
Ecol MAIDENKQKALAAALGQIEKQFGKGSIMRLGEDRS-M-DVETISTGSLSLDIALGAGGLPMGRIVEIYGPESSGKTTLTLQVIAAAQREGKTCAFIDAEHALD 100
Apol MAQ.PGID.S...EG..S...RA.....M..R-PTE-Q.DV.....G.....I....R...I.....MA.HA..E...K.G..... 103
Alai MS.NK..Q..EL..K.....K..D-GADH-SI.A.PS..IA.....I..Y.R...I.V.....HAM.S..KQ.G.V..... 100
Afac MDEPGGKNEFSP.FM...G.....AV..A.DKPGIN-.PDVK.....G..G...Q....R..V.....KA..S..A..A.P..... 101
Acal M.D..S...Q...S.....NTV....DNTV-Q-A..AV.....T.....I....K.....M...A..QC.KA.G..... 99
Atum MAQNSLRLVEDKSVD.S...E...S...RS.....K..SNENVV-E...V.....I....K...I.....A..T..E..KK.GI...V..... 111
Avar MAINTDTSG.....TMV.N...RS....A.....D-ATRMR.....A.T..L...-....R..VI.....VA.HA..EV.K..GIA..V...Q... 103
Amag MDR....E..VS...RA.....K..GKDQVV-ET.VV..RI.G..V...I..V.R...I.V.....A.HI..E..KK.G....V..... 98
Apyr MARVSENLSE.M...EV..SS...R....AV.P.KAVETV-E....P...I.....T.V..I.K...T..F.V.....A.H...E..KR.GVAV..... 105
Avin M....KR.....AV..M.DHER-Q-AIPA.....G.....I....K.....ST..E..KQ.A....V..... 99
Bsub MSDR.A..DM..K.....K...-KTDT-RIS.VPS...A..T...I..Y.R...I.V.....VA.HA..EV.QQ-R.S..... 97
Bfra MDK...S.....KM..-EVVE-Q..V.P...IA.NA...V..Y.R...I.....AIHA..E..KA.GIA.....F. 87
Bper MDDKTSK.AAAE.A.....S.....Y.DNEVEH-.IQVV.....G.....V....R..VI.V.....EM.KL.G....V..... 107
Bbur -MSKLKEKREKAVVGIERAS.EE.IEL.RV....A.....LIKM..SPVGQ-GIKSM.S..IV..E...I..Y.R...I..F.....A..EV.K..GIA..... 115
Babo MSQNSLRLVEDNSVD.T...D...S...RA.....QNDQVV-EI..V.....V....K.....A.HT..E..KK.GI...V..... 111
Cjej M.D..R.S.D...KSLD.T....T.L...DKE-VE-QIDS.G...VG..L....I..V.K...I.....HI..EC.KA.GV..... 99
Ctra MSVPDRKR..E..IAY.....A....S..KHS.AH-EIS..K..A....L....I..V.K.....F.....ATHIV.N..KM.GVA.Y..... 101
Cper MAN..KD.L..IEM.M.....V.K...-QGAP-QMDAV...C.D.....I..V.K...I.....VA.H.V.E..KL.GAA.Y..... 102
Cglu MHPKKTATKATA.KGNDR....D...AL...D....AV....D-ENRP-PIQ...S.NTAI.....I..F.R.....V.....VA.HA..Q..KA.GIA..... 112
Drad MSKDATKEISAPTDAKERS..IET.MS....A.....K..AESKL-..QVV.....L...V..I.G...T.....G.....A.AIV.Q..KA.G..... 112
Eagg 101
Ecar 101
Hinf ..TQ.E.....K..DTKT-L-...S.....G..V...I.....F.....S...Q..KA..... 101
Llac -LATKKKTNFDDITKKYGAERD....D..AL...D.....L.....-AANQK.SVV.S...A.....Y.K.....VA.HAV..V.K..GIA.....N... 114
Lpne ME.....S..VS...R.....V..M.DSTVSR-.I.A.....G.....I....K.....EC.KM.G.A..... 100
Mfla M...RS.....S.....M.DTDVAA-.IQAV.....G.....I....R.....S...QM.KL.G.A..... 100
Mcla M.D..S.....S.....M.DADIGE-.LQVV.....G.....V....R...I.....SA..EM.KL.GVA.....DD... 100
Mmet M.D..S.....S.....M.DADIGE-.LQVV.....G.....V....R...I.....SA..EM.KL.GVA..... 100

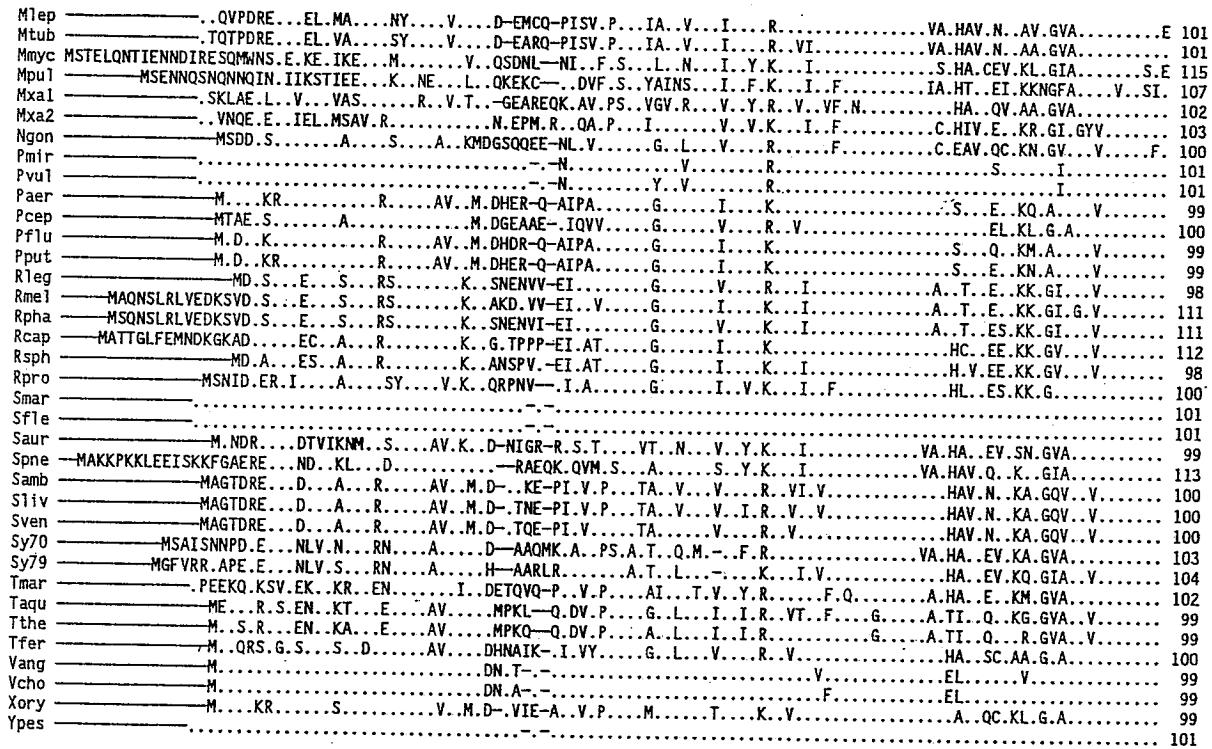
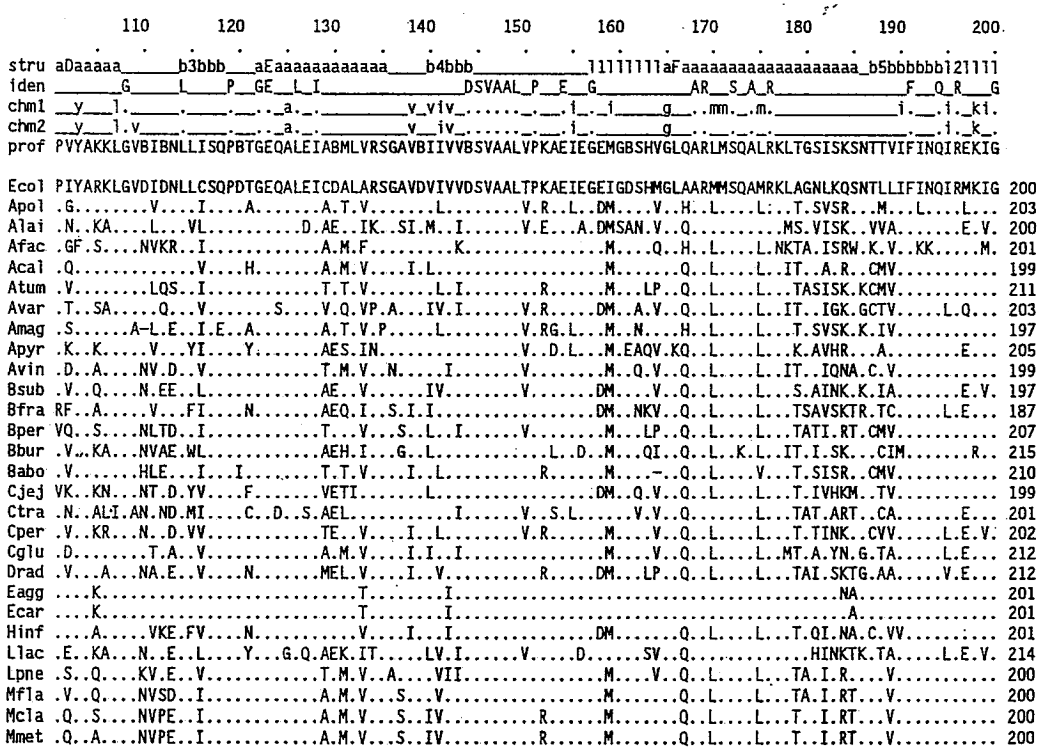
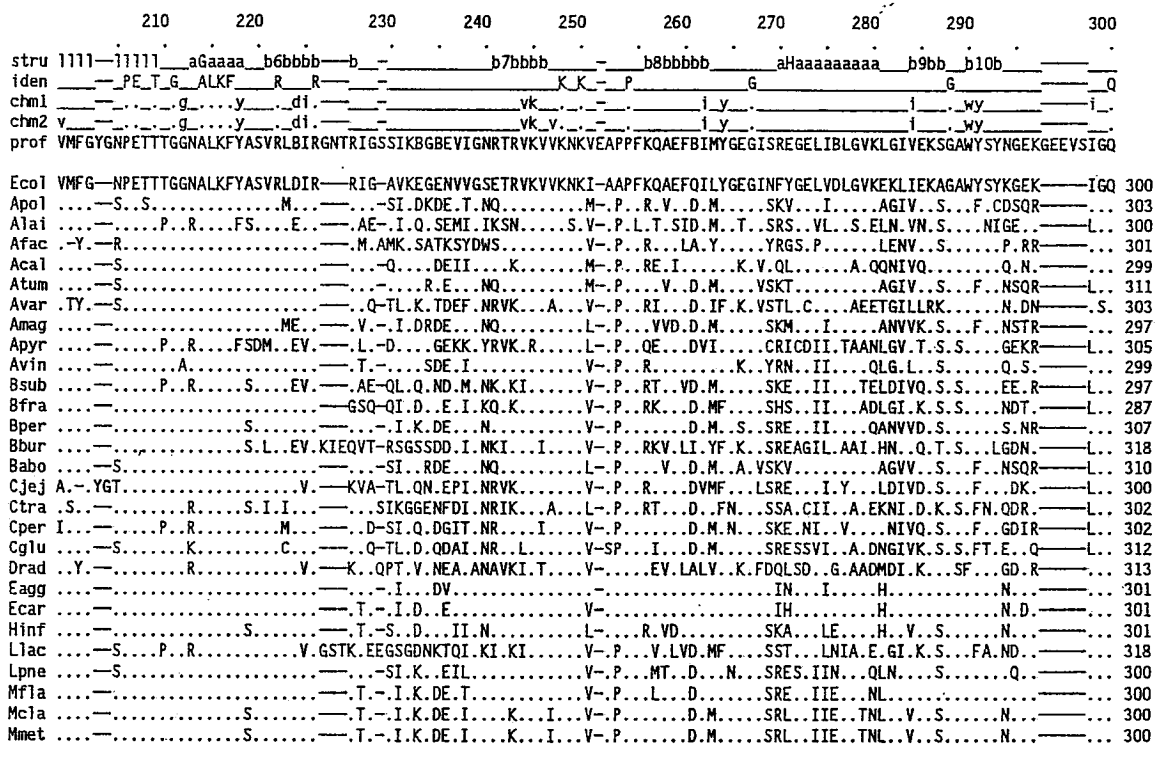

Mlep
Mtub
Mmyc MSTELQNTIENNDIRESQMWNS.E.KE.IKE...M.....V..QSDNL—NI..F.S...L..N...I..Y.K...I.....S.HA.CEV.KL.GIA.....S.E 115
Mpul
Mxal
Mxa2
Ngon
Pmir
Pvul
Paer
Pcep
Pflu
Pput
Rleg
Rmel
Rpha
Rcap
Rsph
Rpro
Smar
Sfle
Saur
Spne
Samb
Sliv
Sven
Sy70
Sy79
Tmar
Taqu
Tthe
Tfer
Vang
Vcho
Xory
Ypes

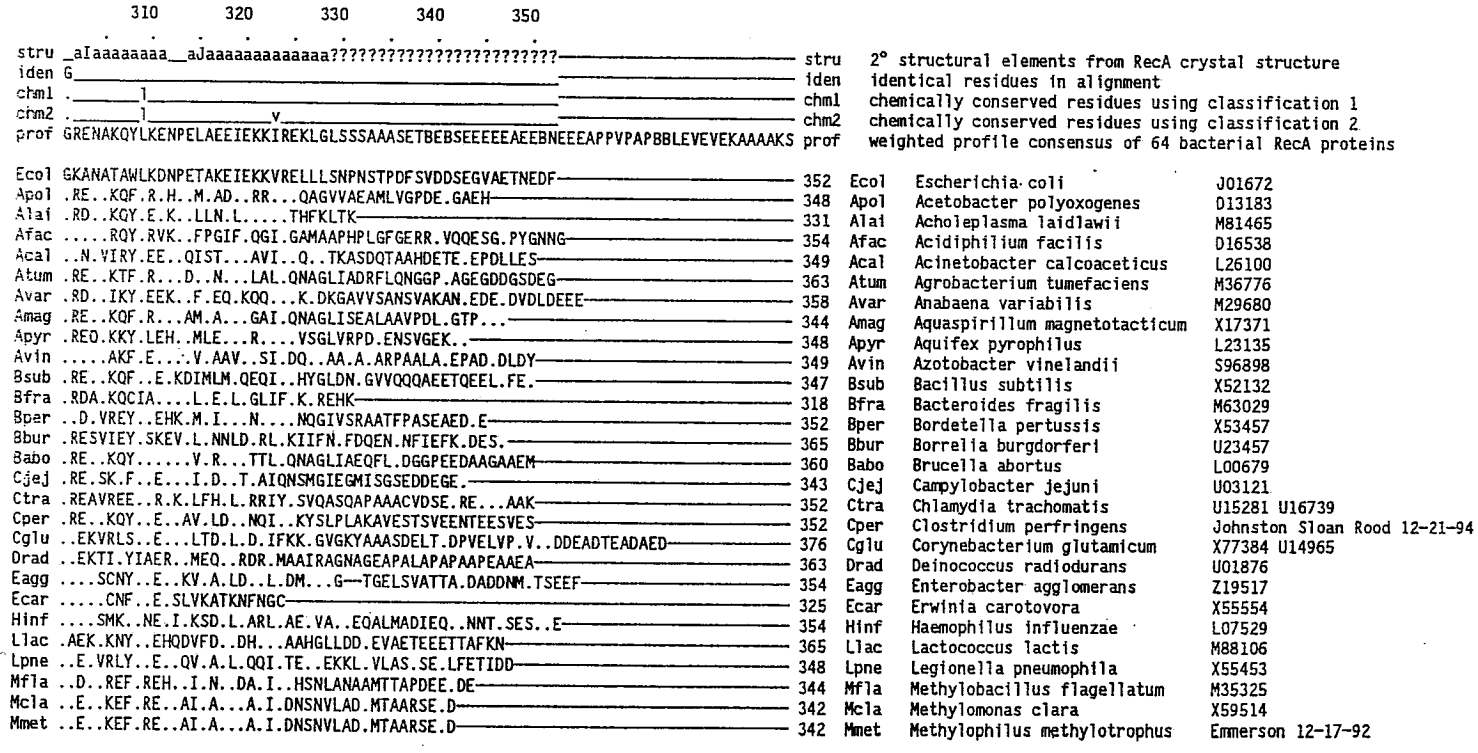
FIG. 4.


110 120 130 140 150 160 170 180 190 200
stru aDaaaaa_____b3bbb___aEaaaaaaaaaaa____b4bbb_____1111111aFaaaaaaaaaaaaaaaaaaa_b5bbbbbb121111
iden _____G_____L____P__GE__L_I_____DSVAAL_P__E__G_____AR__S_A_R_____F__Q_R___G
chml __y___l._____.____.__..._a._._____v_viv_.....__.__.i._.i_____g__..mm._.m._____i.__.i._ki.
chm2 __y___l.v____.____.__._a._._____v__iv_.....__.__.i_._____g__.__._._._____.__.i._k_.
prof PVYAKKLGVBIBNLLISQPBTGEQALEIABMLVRSGAVBIIVVBSVAALVPKAEIEGEMGBSHVGLQARLMSQALRKLTGSISKSNTTVIFINQIREKIG
Ecol PIYARKLGVDIDNLLCSQPDTGEQALEICDALARSGAVDVIVVDSVAALTPKAEIEGEIGDSHMGLAARMMSQAMRKLAGNLKQSNTLLIFINQIRMKIG 200
Apol .G.....V....I....A.....A.T.V.....L.....V.R..L..DM....V..H..L....L...T.SVSR...M...L....L... 203
Alai .N..KA....L...VL.....D.AE..IK..SI.M..I.....V.E...A.DMSAN.V..Q.....MS.VISK..VVA.....E.V. 200
Afac .GF.S....NVKR..I.....A.M.F.....K.....M....Q..H..L....L.NKTA.ISRW.K.V..KK.....M. 201
Acal .Q.....V....H.....A.M.V....I.L.....M.....Q..L....L..IT..A.R..CMV..... 199
Atum .V.....LQS..I.....T.T.V.....L.I.....R.....M...LP..Q..L....L...TASISK.KCMV..... 211
Avar .T..SA.....Q...V.....S....V.Q.VP.A...IV.I.....V.R.....DM..A.V..Q..L....L..IT..IGK.GCTV.....L.Q... 203
Amag .S.....A-L.E..I.E..A.....A.T.V.P.....L.....V.RG.L...M..N....H..L....L...T.SVSK.K.IV..... 197
Apyr .K..K.....V...YI....Y.....AES.IN.....V..D.L...M.EAQV.KQ..L....L...K.AVHR...A.....E... 205
Avin .D..A....NV.D..V.....T.M.V..N.....I.....V.....M..Q.V..Q..L....L..IT..IQNA.C.V..... 199
Bsub .V..Q....N.EE..L.....AE..V.....IV.....V.....DM....V..Q..L....L...S.AINK.K.IA.....E.V. 197
Bfra RF..A.....V...FI....N.....AEQ.I..S.I.I.....DM..NKV..Q..L....L...TSAVSKTR.TC.....L.E... 187
Bper VQ..S....NLTD..I.....T...V...S..L..I.....V.....M...LP..Q..L....L...TATI.RT.CMV..... 207
Bbur .V..KA...NVAE.WL.....AEH.I...G..L.....L..D..M...QI..Q..L..K.L..IT.I.SK...CIM.....R.. 215
Babo .V.....HLE...I...I.....T.T.V....I..L.....R.....M....-..Q..L....V...T.SISR..CMV..... 210
Cjej VK..KN...NT.D.YV....F.....VETI.....L.....DM..Q.V..Q..L....L...T.IVHKM..TV..... 199
Ctra .N..ALI.AN.ND.MI....C..D..S.AEL.....I.....V..S.L.....V.V..Q.....L...TAT.ART..CA.....E... 201
Cper .V..KR...N..D.VV.....TE..V....I..L.....V.R.....M....V..Q..L....L...T.TINK..CVV.....L.E.V. 202
Cglu .D.....T.A..V.....A.M.V....I.I..I.....M....V..Q..L....L..MT.A.YN.G.TA.....L.E... 212
Drad .V...A...NA.E..V....N.....MEL.V....I..V.....R.....DM...LP..Q..L....L...TAI.SKTG.AA.....V.E... 212
Eagg ....K.....T.....I.....NA..... 201
Ecar ....K.....T.....I.....A..... 201
Hinf ....A.....VKE.FV....N.....V....I...I.....DM.....Q..L....L...T.QI.NA.C.VV..... 201
Llac .E..KA...N..E..L....Y...G.Q.AEK.IT.....LV.I.....V.....D.....SV..Q.....HINKTK.TA.....L.E.V. 214
Lpne .S..Q....KV.E..V.....T.M.V..A....VII.....M....V..Q..L....L...TA.I.R....V..... 200
Mfla .V..Q....NVSD..I.....A.M.V...S...V.....M.....Q..L....L...TA.I.RT...V..... 200
Mcla .Q..S....NVPE..I.....A.M.V...S..IV.....R.....M.....Q..L....L...T..I.RT...V..... 200
Mmet .Q..A....NVPE..I.....A.M.V...S..IV.....R.....M.....Q..L....L...T..I.RT...V..... 200

```
Mlep .E..K.....T.S..V.............A.M.I....L.IV.I......V.R..L...M...YV..Q..L....L..MT.A.SN.G.TA.....L.E... 201
Mtub .D..K.....T.S..V.............A.M.I....L.IV.I......V.R..L...M....V..Q..L....L..MT.A.NN.G.TA.....L.D... 201
Mmyc .K.CQT..I.TNK..V....N.....D.LEM.IN.NSI.L.........V..T.LD..MS.QSI..Q....K.L...NWLIAK...TV.....L.E... 215
Mpul .V..KN..I.....I....S.......V.I..K..SI.L.........V.E..LN..MK.QSI..Q..L..K.L..IT.S.SKNK.SV.....V.E... 207
Mxal VS.......RVEE..V.............TEH.V......L.........V.R......M..A...VQ..L....L...T.AVSR.G.CI.......... 202
Mxa2 VG.......RT.D..L.............AEM.V....I..L........V....L...M..A...VQ..L....L...T.TIAK.Q.CV.......... 203
Ngon .V.......KVEE.YL.............T.V...GI.MV........V.......DM....V..Q..L....L...T.HI.KT...VV.......... 200
Pmir ....Q.............................S.....................V.....................N..................... 201
Pvul ....Q.............................S.....................V.....................N..................... 201
Paer .D..G....NV.D..V.............T.M.V..N.....I.......V........M..A.V..Q..L....L..IT..I.NA.C.V.......... 199
Pcep VQ..A....NVPE..I.............T...V...SI.M..I......V........M...LP..Q..L....L...T.TI.RT.C.V.......... 200
Pflu .E..G....NV.D..V.............T.M.V..N.I...........V........M..M.V..Q..L....L..IT..I.NA.C.V.......... 199
Pput .E..G....NV.D..V.............T.M.V..N.....I.......V........M..M.V..Q..L....L..IT..I.NA.C.V.......... 199
Rleg .V........LQ...I.............T.T.V...R...L.........R......M..TVP..Q..L....L...TASISK...MV.......... 198
Rmel .V........LE...I.............T.T.V....I.IL.I......V.R......M....LP.MQ..L....L...TASISK..CMV.......... 211
Rpha .V........LQ...I.............T.T.V.......L.........R......M....LP..Q..L....L...TASISK...MV.......... 211
Rcap .Y..K....SLED..I.............V.T.V.....SLV...............DM..ATV.AQ..L........TASIGR..CMV.......... 212
Rsph .Q..K....NL.E..I.............V.T.V.....NL............S....DM..MQ..SQ..L........TASIGR..CMV.......... 198
Rpro .A..K....N..E.II.............A.T.I...GI.M.II......V..S.....M..AQ.ASQ..L....L...TASINRT.CITV.......... 200
Smar ....K.............................T........I.....................................NA.................. 201
Sfle ..................................................................................................... 201
Saur .E..QA........YL....H...G...AE.FV......IV.................M..T.V..Q..L....L...S.AISK...TA.......E.V. 199
Spne .A..AA...N..E..L....S...G...AGK.ID.....LV........V.R...D.D.....V..Q...........GASINKTK.IA.....L.E.V. 213
Samb .E..Q.........IL....N.......V.M.V....L.L..I......V.R......M....V..Q..L....L..ITSA.N..K.TA.....L.E... 200
Sliv .E..K.........IL....N.......V.M.V....L.L..I......V.R......M....V..Q..L....L..ITSA.N..K.TA.....L.E... 200
Sven .E..K.........IL....N.......V.M.V....L.L..I......V.R......M....V..Q..L....L..ITSA.N..K.TA.....L.E... 200
Sy70 .T.SAA.....E...VA...N..S....A.Q.V..A...L..I......V.R......M..VQV..Q..L..K.L..I...MGR.GCTV..L..L.Q... 203
Sy79 .V..TSV........IC......M....V.Q.V..A...IV.I......V.R......M..AQV..Q..L.......IT..IGK.GCTV..L..L.Q... 204
Tmar .V..KN....LKS..I....H.......V.E.V...V..L.........V.R.....AM..MQV..Q..L....L..I..SVNK.KAVV..T........ 202
Taqu .L..K.....VQE..V.............VEL.................V........M..Q.V..Q..L....L...TAV.SK...AA.....V.E.V. 199
Tthe .L..QR...QVED..V.............VEL.................V.R......M..Q.V..Q..L....L...TAV.AK...AA.....V.E.V. 199
Tfer .G..H.....LE...I.............A.M.V......L..I..............M....V..Q..L....L.N.TA.ISR....V.......... 200
Vang ....K....N..E..V.....................I....I..............M.........Q...L.......T.......CMC.......... 199
Vcho .V..K....N..E..V.....................................M.........Q...L.......T.......CMC.......... 199
Xory .V..A....NV.D..L.............A.M.V..SS..IV.I..............M..QLP..Q..L....L...T..I.R....VV....L..... 199
Ypes ....K.............................T........I.....................................NA.................. 201
```

FIG. 4. (*continued*)


210 220 230 240 250 260 270 280 290 300
stru
iden
chm1
chm2
prof VMFGYGNPETTTGGNALKFYASVRLBIRGNTRIGSSIKBGBEVIGNRTRVKVVKNKVEAPPFKQAEFBIMYGEGISREGELIBLGVKLGIVEKSGAWYSYNGEKGEEVSIGQ
Ecol VMFG—NPETTTGGNALKFYASVRLDIR——RIG—AVKEGENVVGSETRVKVVKNKI—AAPFKQAEFQILYGEGINFYGELVDLGVKEKLIEKAGAWYSYKGEK———IGQ 300
Apol ....—S..S.....M...——...-SI..DKDE.T.NQ.....M-.P..R.V..D.M.....SKV...I.....AGIV..S...F.CDSQR———... 303
Alai ....—.....P..R....FS....E..——.AE-.I.Q.SEMI.IKSN.....S.V-.P.L.T.SID.M..T..SRS..VL..S.ELN.VN.S....NIGE..———L.. 300
Afac .-Y.—R.....——.M.AMK.SATKSYDWS.....V-.P..R...LA.Y.....YRGS.P.....LENV..S.....P.RR———... 301
Acal ....—S.....——...-Q....DEII....K.....M-.P..RE.I.....K.V.QL.....A.QQNIVQ.....Q.N.———... 299
Atum ....—S.....——...-....R.E...NQ.....M-.P....V..D.M....VSKT.....AGIV..S...F..NSQR———L.. 311
Avar .TY.—S.....——..Q-TL.K.TDEF.NRVK...A...V-.P..RI...D.IF.K.VSTL.C....AEETGILLRK.....N.DN———.S. 303
Amag ....—.....ME..——.V.-.I.DRDE...NQ.....L-.P...VVD.D.M.....SKM...I.....ANVVK.S...F..NSTR———... 297
Apyr ....—.....P..R....FSDM..EV.——.L.-D....GEKK.YRVK.R.....L-.P..QE...DVI.....CRICDII.TAANLGV.T.S.S....GEKR———L.. 305
Avin ....—.....A.....——.T.-....SDE.I.....V-.P..R.....K..YRN..II....QLG.L..S.....Q.S.———... 299
Bsub ....—.....P..R.....S....EV.——.AE-QL.Q.ND.M.NK.KI.....V-.P..RT..VD.M.....SKE..II...TELDIVQ.S.S....EE.R———L.. 297
Bfra ....—.....——GSQ-QI.D..E.I.KQ.K.....V-.P..RK...D.MF....SHS..II...ADLGI.K.S.S....NDT.———L.. 287
Bper ....—.....S.....——...-.I.K.DE...N.....V-.P.....D.M..S..SRE..II....QANVVD.S.....S.NR———... 307
Bbur ....—.....S.L..EV.KIEQVT-RSGSSDD.I.NKI...I....V-.P..RKV.LI.YF.K..SREAGIL.AAI.HN..Q.T.S...LGDN.———L.. 318
Babo ....—S.....——...-SI..RDE...NQ.....L-.P....V..D.M..A.VSKV.....AGVV..S...F..NSQR———L.. 310
Cjej A.-.YGT.....V.——KVA-TL.QN.EPI.NRVK.....V-.P..R....DVMF...LSRE...I.Y...LDIVD.S...F...DK.———L.. 300
Ctra .S..—.....R.....S.I.I...——...SIKGGENFDI.NRIK...A...L-.P..RT...D..FN...SSA.CII..A.EKNI.D.K.S.FN.QDR.———L.. 302
Cper I...—.....P..R.....M...——..D-SI.Q.DGIT.NR....I....V-.P.....D.M.N...SKE.NI..V....NIVQ.S...F..GDIR———L.. 302
Cglu ....—S.....K.....C...——..Q-TL.D.QDAI.NR..L.....V-SP...I...D.M.....SRESSVI..A.DNGIVK.S.S.FT.E..Q———L.. 312
Drad ..Y.—.....R.....V.——K..QPT.V.NEA.ANAVKI.T....V-.....EV.LALV..K.FDQLSD..G.AADMDI.K...SF...GD.R———... 313
Eagg ....—.....——...-.I...DV.....-.....IN...I.....H.....N...———... 301
Ecar ....—.....——.T.-.I.D..E.....V-.....IH.....H.....N.D.———... 301
Hinf ....—.....S.....——.T.-S..D...II.N.....L-....R.VD.....SKA...LE....H..V..S.....N...———... 301
Llac ....—S....P..R.....V.GSTK.EEGSGDNKTQI.KI.KI.....V-.P...V.LVD.MF....SST...LNIA.E.GI.K.S...FA.ND..———... 318
Lpne ....—S.....——...-SI.K..EIL.....V-.P...MT..D...N...SRES.IIN...QLN....S.....Q..———... 300
Mfla ....—.....——.T.-.I.K.DE.T.....V-.P...L...D.....SRE..IIE...NL.....———... 300
Mcla ....—.....S.....——.T.-.I.K.DE.I....K...I...V-.P.....D.M.....SRL..IIE..TNL..V..S.....N...———... 300
Mmet ....—.....S.....——.T.-.I.K.DE.I....K...I...V-.P.....D.M.....SRL..IIE..TNL..V..S.....N...———... 300


310 320 330 340 350
stru _aIaaaaaaaa__aJaaaaaaaaaaaaa???????????????????????????
iden G
chm1 .
chm2 .
prof GRENAKQYLKENPELAEEIEKKIREKLGLSSSAAASETBEBSEEEEEAEEBNEEEAPPVPAPBBLEVEVEKAAAAKS
Ecol GKANATAWLKDNPETAKEIEKKVRELLLSNPNSTPDFSVDDSEGVAETNEDF 352 Ecol Escherichia coli J01672
Apol .RE..KQF.R.H..M.AD..RR...QAGVVAEAMLVGPDE.GAEH 348 Apol Acetobacter polyoxogenes D13183
Alai .RD..KQY.E.K..LLN.L.....THFKLTK 331 Alai Acholeplasma laidlawii M81465
Afac .....RQY.RVK..FPGIF.QGI.GAMAAPHPLGFGERR.VQQESG.PYGNNG 354 Afac Acidiphilium facilis D16538
Acal ..N.YIRY.EE..QIST...AVI..Q..TKASDQTAAHDETE.EPDLLES 349 Acal Acinetobacter calcoaceticus L26100
Atum .RE..KTF.R...D..N...LAL.QNAGLIADRFLQNGGP.AGEGDDGSDEG 363 Atum Agrobacterium tumefaciens M36776
Avar .RD..IKY.EEK..F.EQ.KQQ...K.DKGAVVSANSVAKAN.EDE.DVDLDEEE 358 Avar Anabaena variabilis M29680
Amag .RE..KQF.R...AM.A...GAI.QNAGLISEALAAVPDL.GTP... 344 Amag Aquaspirillum magnetotacticum X17371
Apyr .REQ.KKY.LEH..MLE...R....VSGLVRPD.ENSVGEK.. 348 Apyr Aquifex pyrophilus L23135
Avin .....AKF.E...V.AAV..SI.DQ..AA.A.ARPAALA.EPAD.DLDY 349 Avin Azotobacter vinelandii S96898
Bsub .RE..KQF..E.KDIMLM.QEQI..HYGLDN.GVVQQQAEETQEEL.FE. 347 Bsub Bacillus subtilis X52132
Bfra .RDA.KQCIA....L.E.L.GLIF.K.REHK 318 Bfra Bacteroides fragilis M63029
Bper ..D.VREY..EHK.M.I...N....NQGIVSRAATFPASEAED.E 352 Bper Bordetella pertussis X53457
Bbur .RESVIEY.SKEV.L.NNLD.RL.KIIFN.FDQEN.NFIEFK.DES. 365 Bbur Borrelia burgdorferi U23457
Babo .RE..KQY.....V.R...TTL.QNAGLIAEQFL.DGGPEEDAAGAAEM 360 Babo Brucella abortus L00679
Cjej .RE.SK.F..E...I.D..T.AIQNSMGIEGMISGSEDDEGE. 343 Cjej Campylobacter jejuni U03121
Ctra .REAVREE..R.K.LFH.L.RRIY.SVQASQAPAAACVDSE.RE...AAK 352 Ctra Chlamydia trachomatis U15281 U16739
Cper .RE..KQY..E..AV.LD..NQI..KYSLPLAKAVESTSVEENTEESVES 352 Cper Clostridium perfringens Johnston Sloan Rood 12-21-94
Cglu ..EKVRLS..E...LTD.L.D.IFKK.GVGKYAAASDELT.DPVELVP.V..DDEADTEADAED 376 Cglu Corynebacterium glutamicum X77384 U14965
Drad ..EKTI.YIAER..MEQ..RDR.MAAIRAGNAGEAPALAPAPAAPEAAEA 363 Drad Deinococcus radiodurans U01876
Eagg ....SCNY..E..KV.A.LD..L.DM...G--TGELSVATTA.DADDNM.TSEEF 354 Eagg Enterobacter agglomerans Z19517
Ecar .....CNF..E.SLVKATKNFNGC 325 Ecar Erwinia carotovora X55554
Hinf ....SMK..NE.I.KSD.L.ARL.AE.VA..EQALMADIEQ..NNT.SES..E 354 Hinf Haemophilus influenzae L07529
Llac .AEK.KNY..EHQDVFD..DH...AAHGLLDD.EVAETEEETTAFKN 365 Llac Lactococcus lactis M88106
Lpne ..E.VRLY..E..QV.A.L.QQI.TE..EKKL.VLAS.SE.LFETIDD 348 Lpne Legionella pneumophila X55453
Mfla ..D..REF.REH..I.N..DA.I..HSNLANAAMTTAPDEE.DE 344 Mfla Methylobacillus flagellatum M35325
Mcla ..E..KEF.RE..AI.A...A.I.DNSNVLAD.MTAARSE.D 342 Mcla Methylomonas clara X59514
Mmet ..E..KEF.RE..AI.A...A.I.DNSNVLAD.MTAARSE.D 342 Mmet Methylophilus methylotrophus Emmerson 12-17-92
stru 2° structural elements from RecA crystal structure
iden identical residues in alignment
chm1 chemically conserved residues using classification 1
chm2 chemically conserved residues using classification 2
prof weighted profile consensus of 64 bacterial RecA proteins

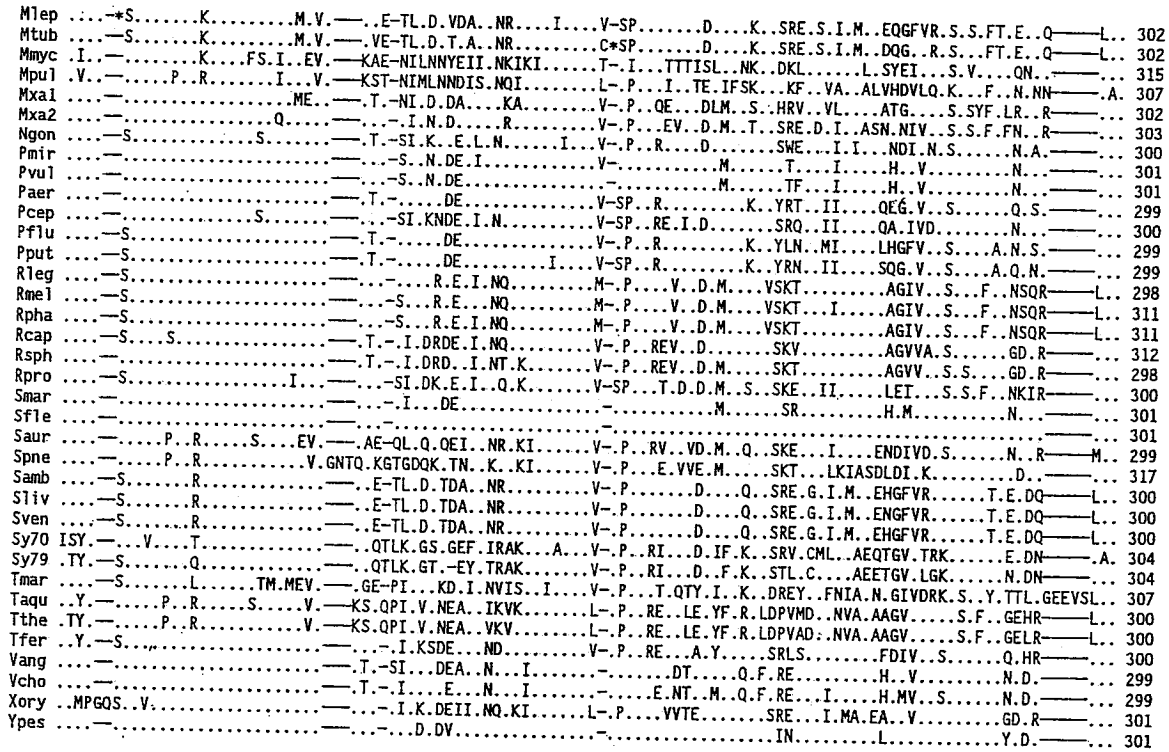


FIG. 4. (*continued*)

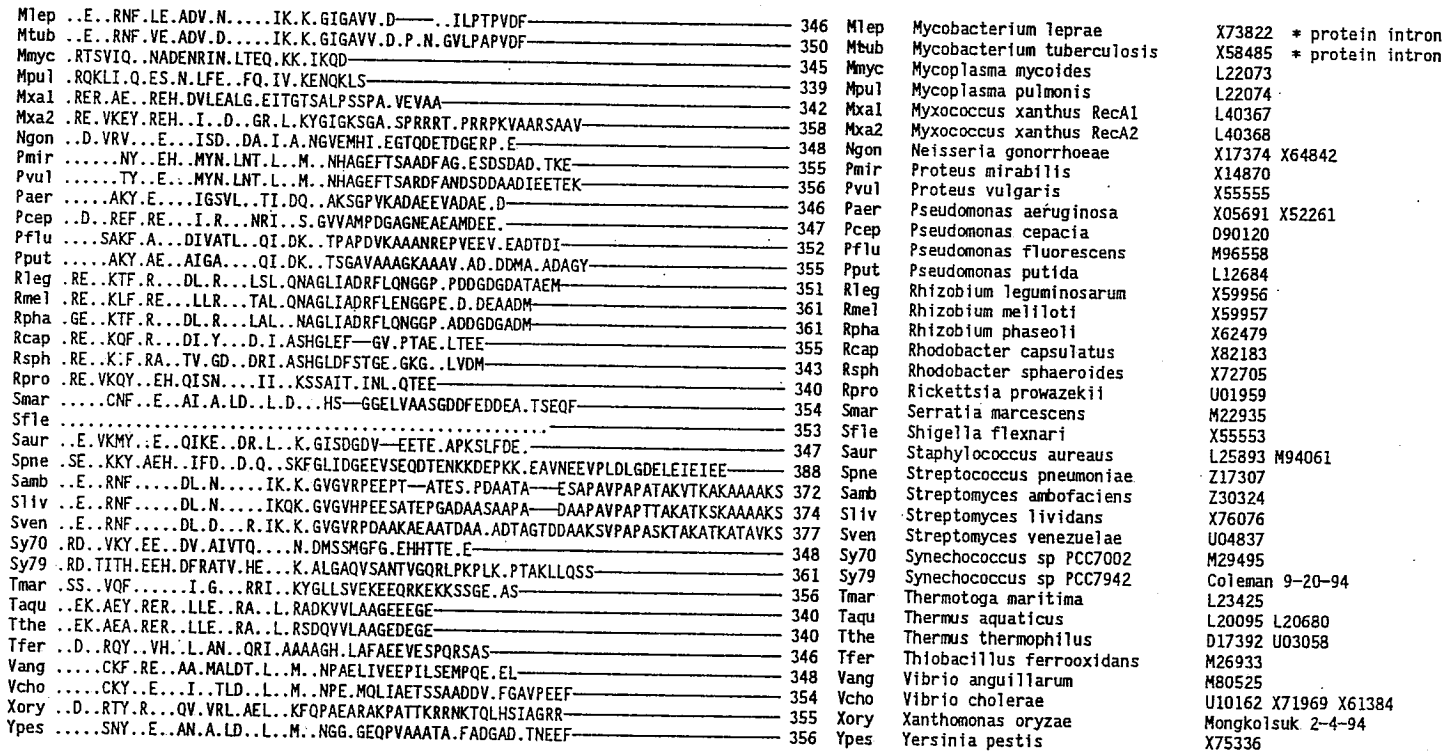


FIG. 4. (*continued*)

Figure 5 shows the invariant and chemically conserved residues mapped onto the crystal structure of the RecA protein with ADP bound (*64*). The conserved residues cluster in three areas corresponding to the major central domain and flanking two smaller subdomains (*64*). In Fig. 5, the N-terminal subdomain corresponds to the gray residues near the top left corner of monomer *a*, whereas the C-terminal subdomain is in the bottom left corner of the same monomer. Complementary information is presented in Fig. 6. Figure 6A outlines the major regions of secondary structure in RecA, along with two loop regions (L1 and L2) that are undefined in the crystal structure. A similarity curve that highlights regions of sequence conservation among bacterial RecA proteins is presented in Fig. 6B. Figure 6C shows the trends in the charge conservation in the RecA family and may provide insight into those regions of the crystal structure that are disordered and hence hinder the calculation of electrostatic potential surfaces. As expected, wherever there is low sequence conservation (Fig. 6B) there is high charge variability, e.g., the carboxyl-terminus region (Fig. 6C). Interestingly, there are a few places where there is strong sequence conservation yet also some charge variability, such as around position 290.

Finally, Fig. 6D is a difference plot showing the change in molecular surface area per residue in the RecA monomer crystal structure (*64*) in the presence of different ligands. These values are defined by the calculated interactions of residues with an imaginary spherical probe having a radius of 1 Å. These calculations can provide a general sense of the ligand interactions with respect to the RecA protein sequence. If a ligand blocks or impedes the probe's interaction with a given amino acid residue, there is a decrease in the molecular surface area that registers as a peak in Fig. 6D. There is a limitation to this approach. These calculations can overestimate interactions, i.e., a peak does not necessarily represent a van der Waals contact. For example, Tyr-264 does not make substantial contacts with ADP in the crystal structure. However, the cavity between Tyr-264 and ADP cannot accommodate the probe and therefore a signal is generated in the plot. Nevertheless, most of the peaks in Fig. 4D reflect identifiable interactions seen in the crystal structure.

Because key regions in the interior of the RecA filament are disordered (e.g., loops L1 and L2), it is impossible to calculate an electrostatic potential surface (*81*) in much of the region where DNA binding is known to occur. Keeping this caveat in mind, it was still interesting to visualize the distribution of positive and negative surfaces of the RecA filament as seen in the crystal structure. Figure 7 shows that there is an asymmetric distribution of charge with respect to the long axis of the filament. Namely, the 5′ end of the filament with respect to bound single-stranded DNA (ssDNA) (smooth face) is predominantly positive, whereas the 3′ end (lobes) is predominantly neg-

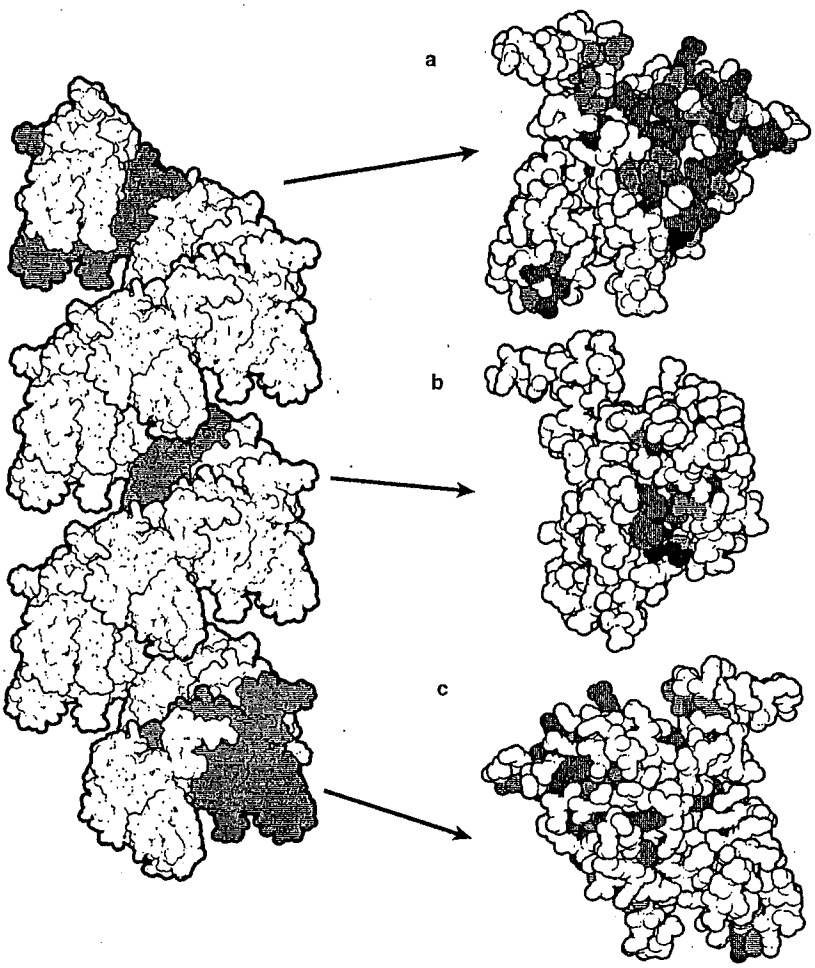


FIG. 5. Conserved residues in the bacterial alignment mapped onto the RecA crystal structure (Brookhaven entry 2REB). An 18-subunit right-handed RecA filament (left) has three symmetry-related subunits in gray. Each of these monomers (labeled *a*, *b*, and *c*) is enlarged on the right. If one is looking down from the top of the filament, monomer *b* is rotated 60° in a clockwise direction from monomer *a*. Monomer *c* is rotated 180° from monomer *a*. Hence, monomer *a* is the inside face of RecA with respect to the filament axis. The rodlike protrusion at the top left part of monomer *a* is α-helix A (see Figs. 2, 4, and 6A). Note that ≈15% of the RecA protein is disordered in the crystal structure and therefore is not visible. The identical residues in the alignment are in dark gray, the chemically conserved residues are in light gray, and the ADP molecule is black. Monomers *a* and *c* show all of the identical and conserved residues that are visible on the protein faces shown. Monomer *b* shows only the conserved residues in class I (ADP binding) (*4*). Conserved residues in this figure are based on an analysis of 58 bacterial RecA sequences, excluding Bbur, Ctra, Cper, Rcap, Samb, and Sy79, which were not available when the figure was generated. This representation of the RecA proteins was generated with the help of David Goodsell (Scripps Research Institute) using a black and white rendering technique that he developed (*421*). The filament is oriented with the end nearest the 5′ end of a bound single DNA strand oriented toward the top.

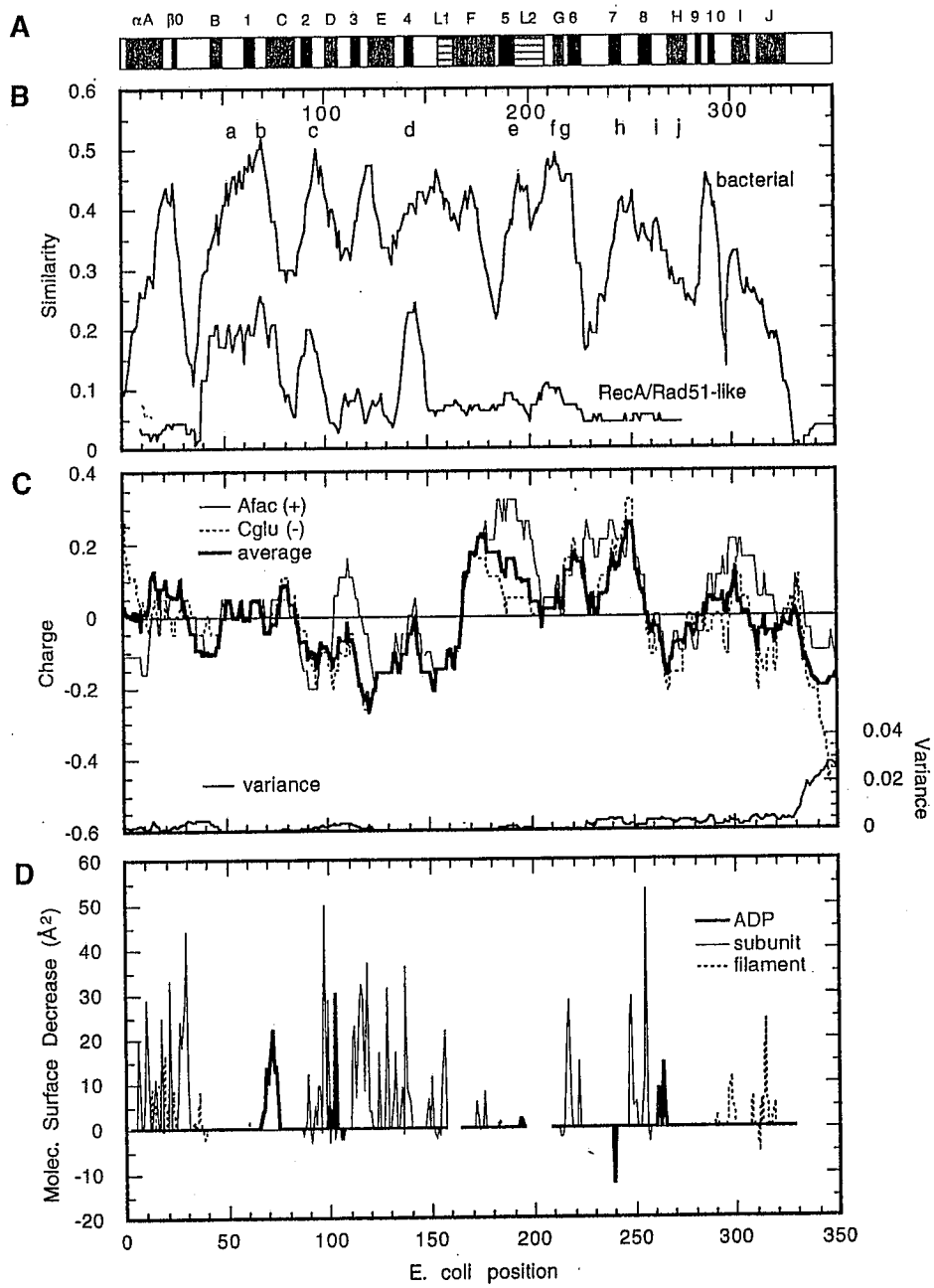


FIG. 6. Plots of similarity, charge density, and molecular surface with respect to the bacterial RecA alignment. (A) Secondary structural elements from the RecA.crystal structure (*64*) are schematically shown with α helices in gray and β strands in black. The L1 and L2 loops are de-

ative. Support for this calculation comes from the measurement of a large permanent dipole in a RecA–dsDNA complex (*82*). The positively charged amino terminus (see Fig. 6C) contributes to the basic surface seen in Fig. 7. The primary role of the amino terminus is in forming part of the

---

picted as horizontal lines. (B) A similarity plot showing the conserved regions in the bacterial RecA (Fig. 4) and the RecA/Rad51-like alignments (Fig. 9B). Interesting regions of the RecA protein are marked as follows: (a) MAW motif; (b) Walker A box (P-loop); (c) Glu-96; (d) Walker B box; (e) Asn-193; (f) Gly-211 and Gly-212; (g) RecA signature; (h) proposed DNA-binding wing; (i) Tyr-264; (j) C-terminal end of Rad51 and Dmc1. The similarity plots were generated using the PLOTSIMILARITY program of GCG. For the unweighted alignment of bacterial RecA proteins in Fig. 4, a window size of 10 residues was used. This is approximately the average length of secondary structural elements in the RecA crystal structure. The BLOSUM62 table was used to calculate sequence similarities. Values were scaled from 0 to 1, where 1 is the maximum possible score. Only that part of the curve corresponding to the *E. coli* RecA sequence is shown, *i.e.*, 350 residues out of the approximately 400 residues (columns) in the multiple sequence alignment. The PLOTSIMILARITY program is not sensitive to weights such that the entire RecA/Rad51-like alignment could not be used to generate a similarity curve. To alleviate the problem caused by the overrepresentation of bacterial RecA sequences, a subset of the alignment, using only the following sequences, was used: Ecol, Rad51, and Dmc1 of *S. cerevisiae,* and Sms, Rad57, Rad55, Mei3, Rec2, and UvsX. The same window size, symbol comparison table, and scaling were used as described above. Also, only the values corresponding to the *E. coli* RecA sequence are shown, *i.e.*, the values corresponding to the gap insertions in the *E. coli* RecA protein (Fig. 9B) were omitted in the plot. (C) Charge density plot of the RecA alignment. The average of all 64 bacterial sequences is graphed. The most basic (Afac, full length calculated p*I* = 9.5) and the most acidic (Cglu, p*I* = 4.9) RecA proteins are also portrayed. Instead of plotting the data from the other 62 RecA sequences, the variance in the data is shown below the charge density curves. As before, only those parts of the curve corresponding to the *E. coli* RecA protein are shown. The Protean program (Version 3.04, DNAStar, Inc.) was used to calculate charge. This method sums charge at pH 7 over a defined window size using standard p*K* tables (*422*). A window size of 19 residues was used. This larger window size was used to reduce noise in the data. (D) Molecular surface area changes due to ADP, adjacent subunits, or a neighboring filament interacting with a RecA monomer in the crystal structure. A positive peak reflects a decrease in molecular surface area, and therefore potential contacts between the RecA monomer and the respective ligand. Note that there is no signal for the Walker B box, because the γ phosphate of ATP is not present in the RecA crystal structure (*65*). The solvent accessible molecular surface (*423*) was calculated for one RecA monomer in the crystal structure (*65*) using the DMS program of the MIDAS 2.0 system (*424*). In short, an imaginary spherical probe is rolled over the entire surface of a macromolecule, thereby generating a molecular surface. This molecule surface is smoother than a van der Waals surface because small crevices cannot accommodate the rolling probe. The smallest probe possible (radius = 1 Å) was used to generate a surface revealing the most features. The default values for atomic radii were used. Values for the surface area (in $Å^2$) for each nondisordered residue were obtained. Calculations were repeated in the presence of each of the following interacting ligands: ADP bound by the RecA monomer, subunits abutting the RecA monomer within the filament, and monomers in a neighboring filament making interfilament contacts. The decrease in molecular surface area due to the presence of the ligand is plotted.

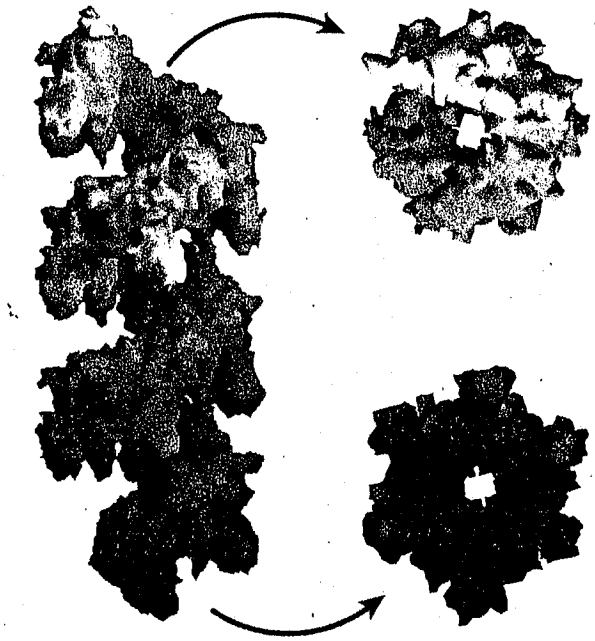


FIG. 7. Electrostatic potential surface of an 18-monomer RecA filament calculated using the program GRASP with default settings (*425*). The gradient of electrostatic potential values ranges from -4 to $+4$ $k_B t$ and is represented from black to white, respectively. The orientation of the filament is the same as that in Fig. 5. On the right are two views looking down the long axis of the filament from the 5′ end (*top*) and 3′ end (*bottom*). A color version of this figure is available using the following URL: http://atahualpa.biochem.wisc.edu/.

monomer–monomer interface within a RecA filament (see Fig. 6D and *83*), but part of it is exposed in the filament groove. The overall dipole is a significant feature of the helical entrance to the central DNA binding pocket of the RecA filament. It is interesting to speculate that the dipole may play an indirect role in DNA binding or in the mechanism by which secondary DNA molecules enter and exit a nucleoprotein filament formed on ssDNA.

Conserved regions may be best described in the context of the crystal structure. Four classes of conserved residues have been defined (*4*). Class I residues are involved in making ADP contacts in the crystal structure (Fig. 6D), and possibly catalyzing ATP hydrolysis and mediating the induced conformational switch induced by ATP hydrolysis (*4*). Class II residues are

involved in the hydrophobic packing of the RecA monomer. Many of the conserved but not necessarily invariant residues are evident in the central domain of monomers *a* and *c* in Fig. 5. Class III residues lie at the subunit–subunit interface (Fig. 6D). The Class IV residues constitute a "structurally unusual" part of the RecA protein (*4*), described in more detail below. No conserved residues have been uniquely associated with DNA binding.

The amino-terminal subdomain has conserved residues surrounding β-strand 0 (Figs. 4 and 6B) that fall into class III. The carboxyl-terminal subdomain has conserved residues near β-strands 9 and 10 and the beginning of α-helix I. Figure 6D shows that this region is involved in interfilament contacts in the crystal structure (*64*). Most of the conserved residues are in the major central domain. This domain comprises the inside face of the filament, which is where important RecA activities such as DNA strand exchange are thought to take place. Monomer *a* in Fig. 5 shows the inside face of the RecA protein. Many of the conserved residues on the inside face are involved in ADP binding, and monomer *b* in Fig. 5 shows only the class I residues. They surround the bound ADP molecule (shown in black). Monomer *c* in Fig. 5 shows that the outside face of the RecA monomer has conserved residues also. There are other conserved residues, but they are not visible in these illustrations because they are buried in the crystal structure. Additional information about many of the conserved residues can be found in several recent reviews (*70, 84, 85*).

In some cases, sequence similarities are found in clusters (Fig. 6B) that correspond to known motifs such as the Walker A box (*71*), now called the P-loop motif (*86*). In the RecA alignment, the motif is 66-GpESsGKT-73 (invariant residues are capitalized), which most resembles the motifs found in Ras p21 proteins, elongation factors, myosins, and thymidine kinases (GxxxxGK[TS], where x represents any residue and the residues in brackets refer to mutually exclusive choices) (*86*). This loop binds the phosphates of nucleotides.

Another ATP-binding motif that is less well defined among different protein families is the Walker B box (*71*). The consensus sequence is four hydrophobic residues followed by an aspartate, which in the crystal structures of Ras p21 (*72*) and Ef-Tu (*73*) is a conserved aspartate at the end of a β-strand. In the RecA alignment this occurs at β-strand 4, whose sequence is 140-vivvD-144. It has been proposed that the conserved aspartate interacts with the γ-phosphate of ATP, although this moiety is not present in the crystal structure of the RecA protein (*65*).

The PROSITE data base includes a unique RecA signature sequence based on a smaller RecA alignment (*87*). The signature occurs at 214-ALK-

FyasvR-222 and the new signature based on the expanded alignment in Fig. 4 is 21-A-L-K-F-[FY]-[AS]-[DST]-[ILMQV]-R-222. From the crystal structure, this site makes contacts with an adjacent monomer in the RecA filament (*64*), as seen in Fig. 6D. Another substantial region of conservation is found to the N-terminal side of the Walker A box. This region, spanning residues 42–65, is described in more detail below.

Regions of substantial divergence in the alignment occur in loops (residues 30–40 and 230–240), at the ends of secondary structural elements (residues 85–90, 110–113, 130–135, 180–190, 280–285 and 295–300), and at the amino and carboxyl termini (Fig. 6B). Two of these divergent regions (residues 110–113 and 130–135) correspond to parts of the monomer–monomer interface in the RecA filament (*64*). Four of these areas (residues 30–40, 180–190, 280–295, and 295–300) form part of the interfilament contact regions seen in the crystal structure (Fig. 6D) (*64, 88*). Divergence at these sites indicates that the interfilament contacts are either serendipitous interactions unique to the crystal, or that there is species specialization in forming such interfaces.

Unique insertions relative to the *E. coli* RecA protein occur in loops such as those made up of residues 295–300 and 195–209 (loop L2) as well as at the end of β-stand 6. Especially notable insertions are the unrelated protein introns [called "inteins" (*89*)] of *M. tuberculosis* and *M. leprae* found in the turn between β-strands 7–8 and in loop L2, respectively (Figure 4) (*90, 91*). Interestingly, it has been found that an unprocessed *Mycobacterium tuberculsosi* RecA protein (containing a 440-residue insertion) will negatively complement processed, i.e., functional, *M. tuberculosis* RecA protein. This suggests that the unprocessed *M. tuberculosis* RecA protein folds properly so that it can form mixed filaments with wild-type RecA protein and then inactivate RecA function (*90, 92*).

The first two residues of the amino terminus and the last 24 residues of the carboxyl terminus are disordered in the crystal structure and presumably are solvent-exposed. The amino terminus varies in length and is not well conserved (Fig. 6B), although almost all of the amino termini of the bacterial proteins are long enough to span α-helix A of the *E. coli* RecA crystal structure. The sequence of the amino terminus of a mature RecA protein is not always certain due to the presence of alternative start codons at the 5′ end of *recA* genes (*93*). The carboxyl terminus is also poorly conserved in the alignment (Fig. 6B). Though the sequences may not be similar, there is a general pattern of negatively charged residues (Fig. 6C), with some interesting exceptions. Figure 8 shows a plot of the length of the carboxyl termini versus their calculated isoelectric point. In general, the carboxyl termini average 50 residues in length and have an acidic character ($pI \approx 5$).

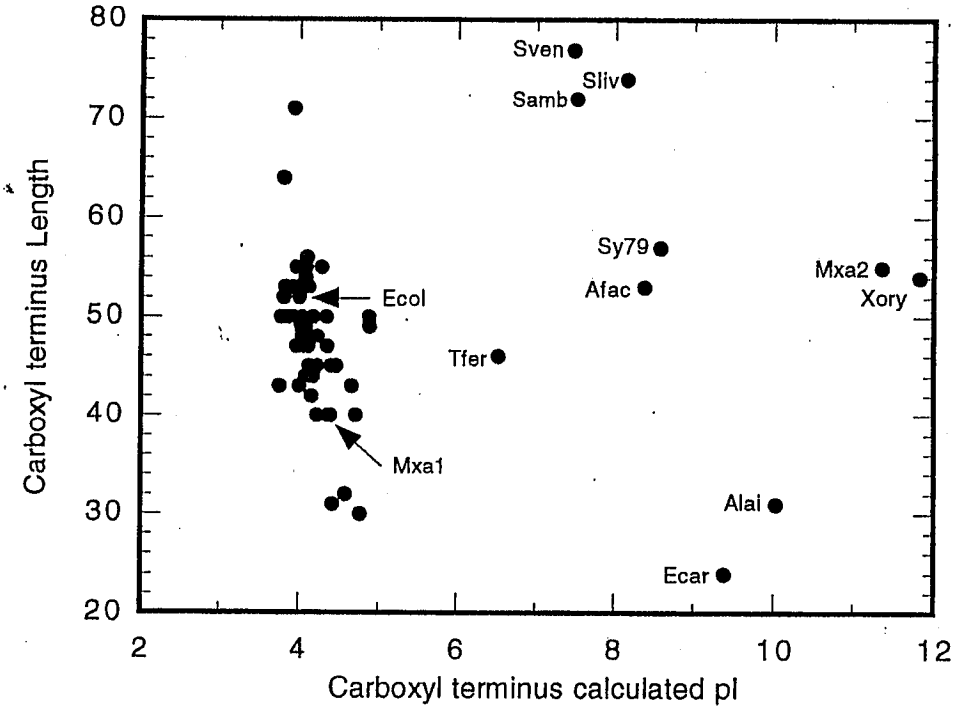


FIG. 8. Distribution of bacterial RecA carboxyl-terminal length versus calculated p*I*. For this analysis, the carboxyl terminus of the RecA protein is defined as starting at the invariant glycine, which occurs at position 301. Isoelectric (p*I*) points were calculated using the GCG ISOELECTRIC program with default parameters. The average carboxyl-terminal length and calculated p*I* are 49 ± 9.5 and 4.9 ± 1.9, respectively. The p*I* values reported are for comparison purposes only and are not intended to reflect an experimentally determined p*I* for an equivalent polypeptide.

2. RecA Alignments Including Bacteriophage and Eukaryotic RecA Homologs

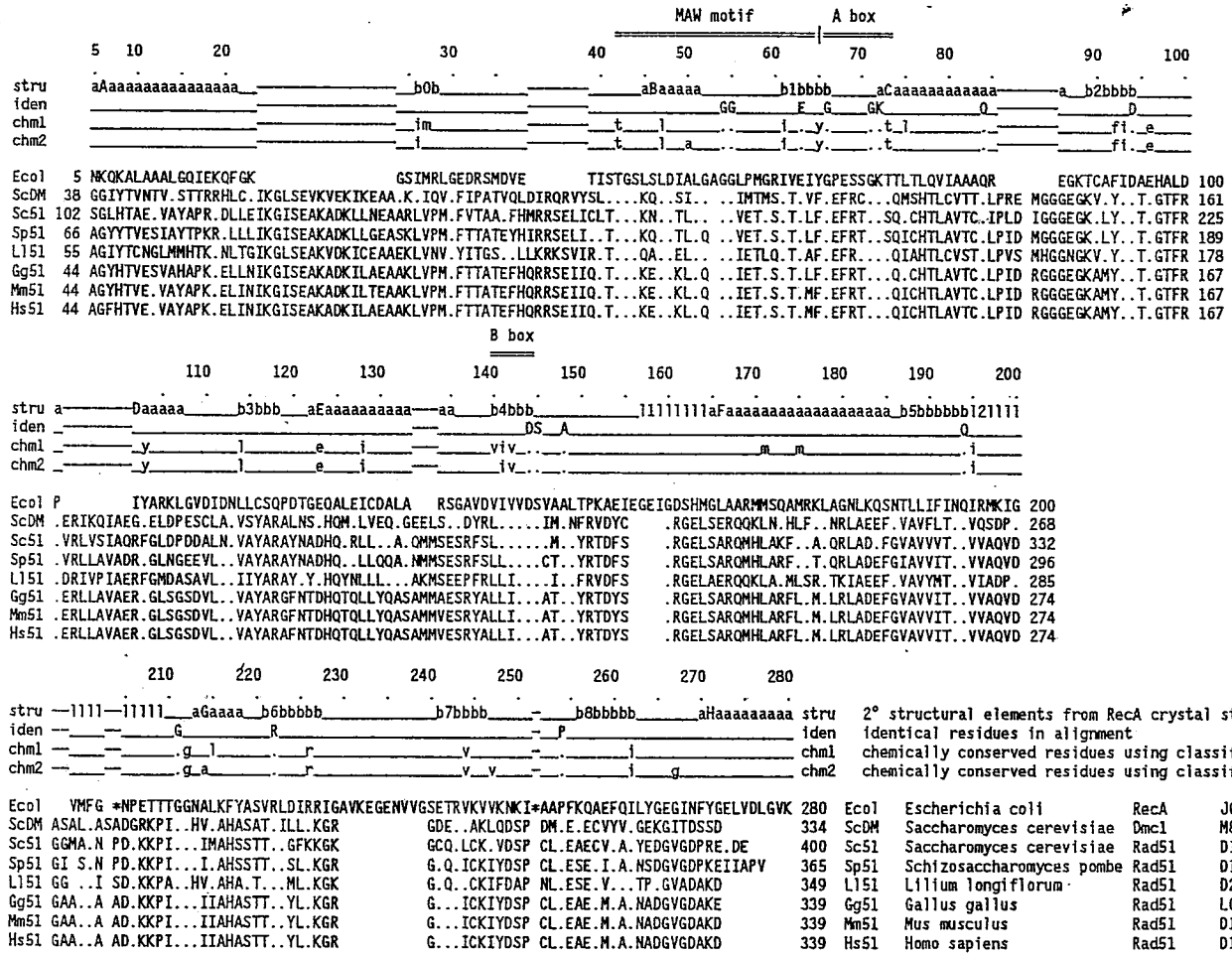
An alignment of bacterial RecA proteins and known structural (*3, 4*) and functional (*76, 94*) homologs is presented in Fig. 9A. This expanded alignment focuses attention on a subset of the residues conserved in the bacterial alignment. In general, these residues are involved in ATP interactions, monomer–monomer contacts, and the hydrophobic core of RecA (*64*). Note that the summary lines represent conserved residues from *all* of the RecA/Rad51-like sequences. Bacterial RecA protein sequences other than *E. coli* RecA are not shown in the listing of sequences, to conserve space.

The bacterial RecA alignment was used as a basis to search for additional RecA homologs in the data base. As described in the Fig. 9 legend, proteins exhibiting sequence similarities confined to the ATP binding motifs were discarded in the analysis. Proteins found with the Profile (*95*) and BLAST (*96*) searches are aligned with the bacterial RecA and eukaryotic Rad51 proteins in Fig. 9B. The similarity curve for this alignment is shown

---

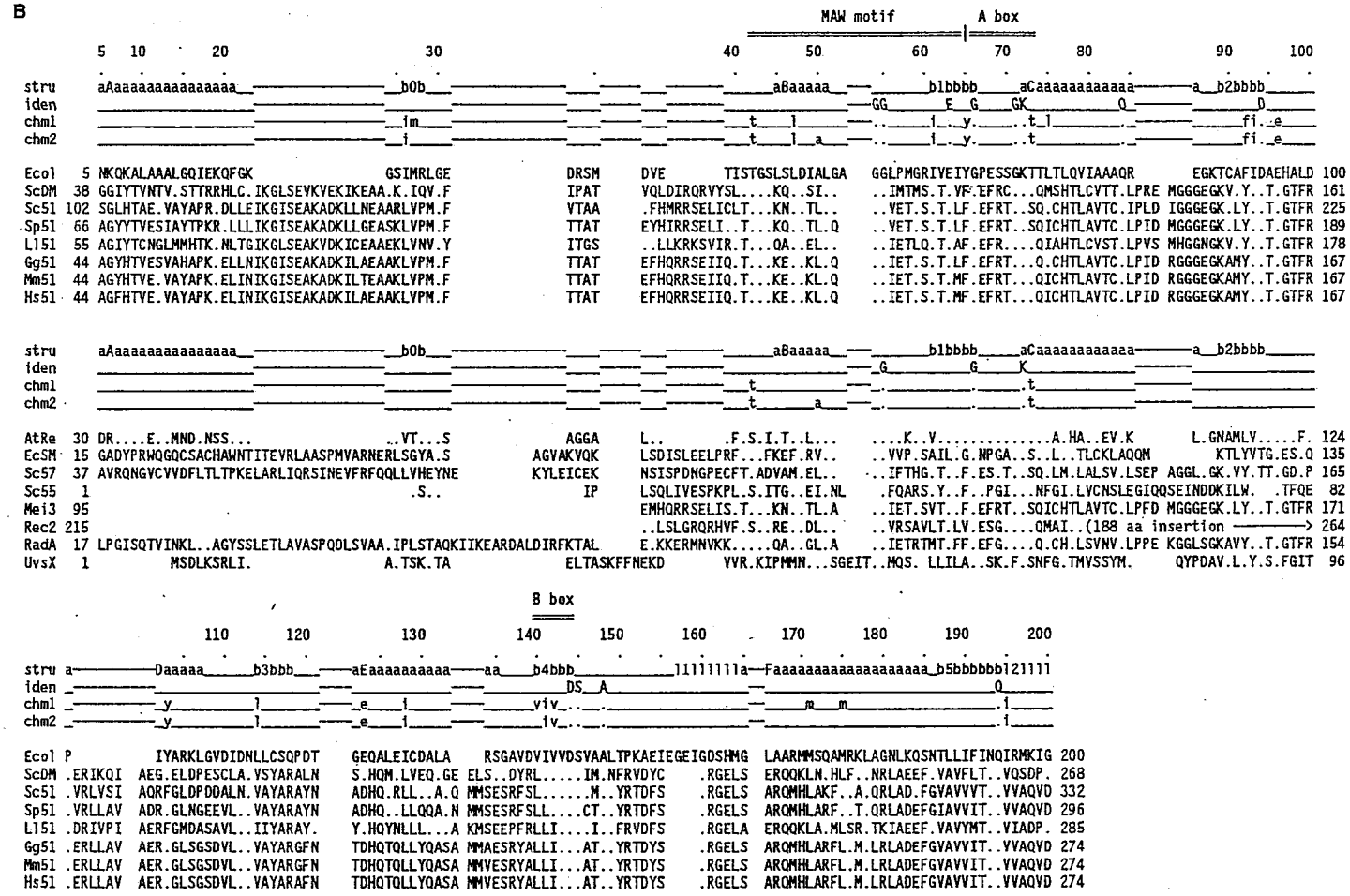
FIG. 9. Alignment of 64 bacterial RecA proteins with RecA/Rad51-like homologs. (A) Bacterial RecA aligned with Dmc1 and Rad51 homologs. (B) RecA, Dmc1, and Rad51 proteins aligned with RecA/Rad51-like candidate homologs. See Fig. 4 legend for explanation of alignment format. Due to the larger number of gaps in these alignments, spaces represent gaps for clarity. The dash (-) identifies gaps in the summary lines with respect to the *E. coli* RecA sequence. The locations of the MAW motif and Walker A and B boxes are indicated. Note that the other 63 bacterial RecA proteins are not shown due to space constraints. However, the summary lines were calculated using those sequences. Therefore, some residues do not appear in the summary lines, *e.g.*, Leu-51 is not conserved due to Thr (Apyr, Tmar) and Met (Sy70) substitutions in sequences not shown. Also, nowhere in the Rec2 insertion could a reasonable catalytic glutamate (corresponding to *E. coli* position 96) be found that is in the same context as the other Rad51-like proteins. A weighted RecA profile was first generated with the PROFILEMAKE program using the BLOSUM62 table. The RecA profile was used to search the SwissProt data base (Release 30) for RecA-like sequences using the PROFILESEARCH program and the following optimal parameters: no averaging, gap penalty = 17, and gap length weight penalty = 1.7. The strength of the structural similarity is reported as a Z score (standard normal deviate, where unrelated sequences have a mean of 0 and a standard deviation of 1). The Z scores above 6 are an indication of significant structural similarity (*112*). In an analysis of small structural features such as the helix–turn–helix motif, it was found that significant Z scores were around 10 (*426*). Using the RecA profile, it was found that full-length "normal" bacterial RecA sequences, *i.e.*, nonpartial sequences and nonprotein intron-containing sequences, had standardized Z scores greater than 120. The Dmc1 and Rad51 homologs had Z scores around 40. Except for Rad55, all the search hits with Z scores between 6 and 9.5 were ATP-binding proteins with the highly conserved A-box (P-loop) motif (*86*). Those sequences were ignored in the following analysis. Search hits with Z scores above 9.5 (except Rad55, Z = 8.2) were used to generate an alignment of RecA/Rad51-like proteins. Many of these proteins were known to have similarity to RecA (*5, 78, 109*). Pairwise comparisons of the profile consensus and profile search data base hits were generated with the PROFILEGAP program. BLAST searches (*96*) were also performed with the profile consensus and various individual RecA or Rad51 sequences, but this did not yield any new hits. It is difficult to automatically align sequences with low sequence identity (*79*). For example, two different alignments of bacteriophage T4 UvsX protein with the *E. coli* RecA have been reported (*70, 427*). Therefore, to generate the alignments of the Dmc1 and Rad51 proteins, these sequences were first automatically aligned using PILEUP. This alignment was manually merged with the previously generated RecA alignment. The structural alignment of Dmc1 with the RecA crystal structure was used as a guide during this process (*4*). The alignment in A differs slightly from the structural alignment in the L2 loop region. This region is disordered in the crystal structure and poorly conserved among RecA/Rad51-like proteins. Each of the remaining RecA/Rad51-like sequences were manually added using the BLAST and the pairwise PROFILEGAP alignments as guides. The Rec2 protein (*78*) has a very large insertion between the Walker A and B boxes that is not found in the other RecA/Rad51-like proteins. This insertion is not shown here for the sake of clarity.

A



2° structural elements from RecA crystal structure
identical residues in alignment
chemically conserved residues using classification 1
chemically conserved residues using classification 2

Ecol	Escherichia coli	RecA	J01672	280
ScDM	Saccharomyces cerevisiae	Dmc1	M87549	334
Sc51	Saccharomyces cerevisiae	Rad51	D10023	400
Sp51	Schizosaccharomyces pombe	Rad51	D13805	365
Ll51	Lilium longiflorum	Rad51	D21821	349
Gg51	Gallus gallus	Rad51	L09655	339
Mm51	Mus musculus	Rad51	D13803	339
Hs51	Homo sapiens	Rad51	D14134	339


B
MAW motif
A box
B box
stru
iden
chm1
chm2
Ecol
ScDM
Sc51
Sp51
L151
Gg51
Mm51
Hs51
AtRe
EcSM
Sc57
Sc55
Mei3
Rec2
RadA
UvsX

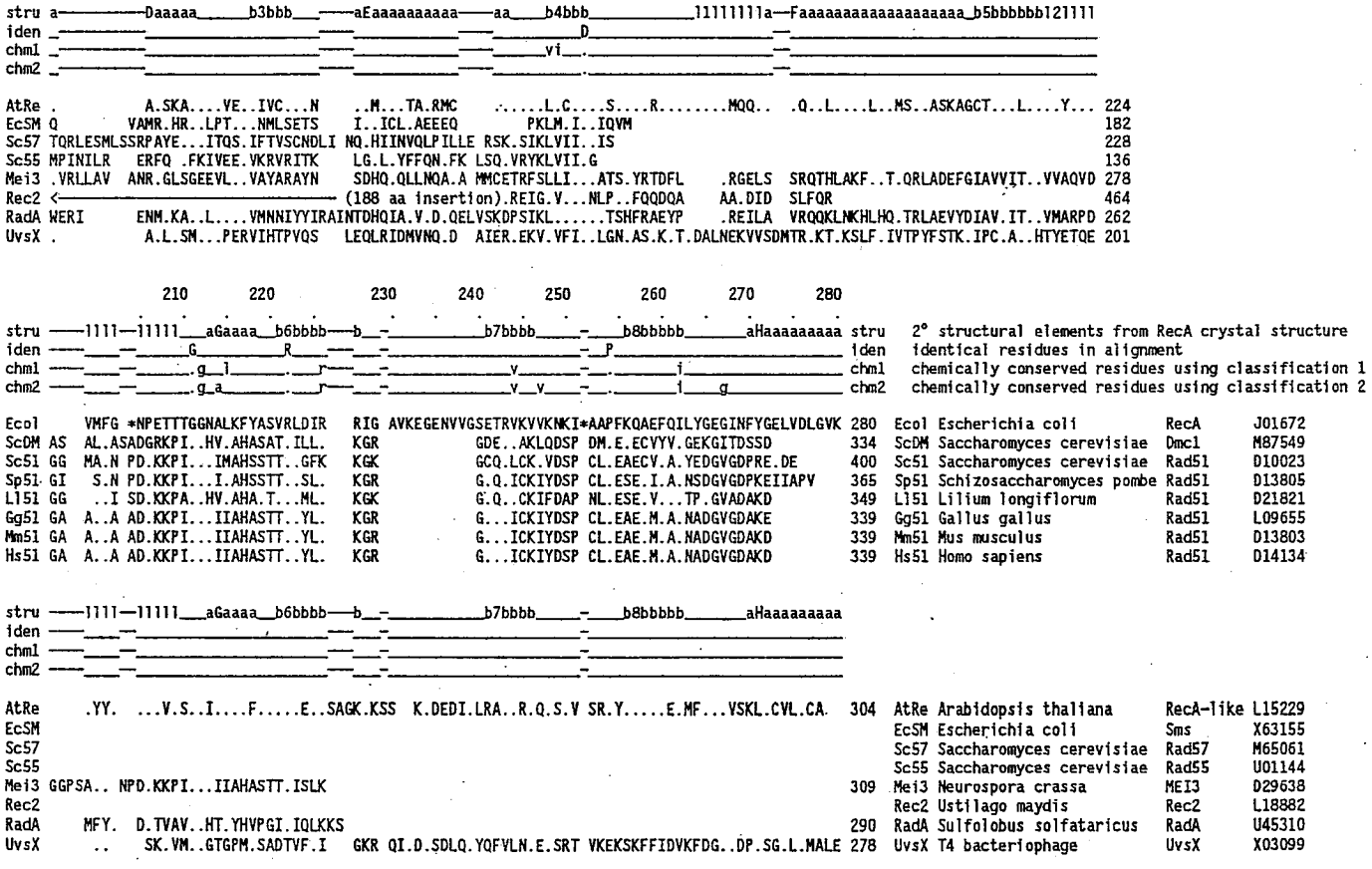


FIG. 9. (*continued*)

in Fig. 6B. Four conserved structural elements are highlighted by the similarity curve for this alignment. Three of these areas are involved in ATP interactions: the A box, Glu-96, and the B box. The fourth region involves residues 42–65, discussed further below.

Proteins identified in the search for RecA homologs include a number of proteins previously shown to have some similarity to RecA and to have some role in recombination processes. One is the UvsX protein of bacteriophage T4, a protein with a well-studied structural (*4, 97, 98*) and functional (*99–101*) relationship to RecA. The RecA-like protein of *Arabidopsis thaliana* chloroplasts is very similar to bacterial RecA proteins (over 50% identity) (*102*). This reflects the widely held view that chloroplasts originated from cyanobacterial ancestors (*103*). Mei3 and Rec2 are believed to be Rad51 homologs from *Neurospora crassa* (*104*) and *Ustilago maydis* (*78*), respectively. In *Saccharomyces cerevisiae,* Rad55 and Rad57 are in the same genetic pathway for radiation repair as Rad51 (*105, 106*). Both proteins act downstream of Rad51 in the *RAD50–RAD57* epistasis group (*107*). Recently, it has been shown that Rad51 and Dmc1 colocalize in prophase nuclei, and it has been suggested that both proteins may interact to form components of recombination nodules (*108*). Last, mutations in the *sms* gene of *E. coli* result in sensitivity to the DNA-damaging agent, methylmethane sulfonate (*109*).

We also analyzed the RadA protein sequence of *Sulfolobus solfataricus* (*110*), a member of the Archaea domain (*111*). This protein is the first demonstration of a RecA homolog in this domain (*110*), whereas homologs have been characterized in both the Bacteria and Eucarya domains. The major ATP interaction motifs of RecA protein are present in the RadA protein.

The only region with substantial sequence conservation in Fig. 9B that is not clearly involved with ATP binding and/or hydrolysis occurs on the N-terminal side of the Walker A box. This region is largely buried in the crystal structure (Fig. 6D). Some class II residues and all of the class IV residues fall into this region, spanning residues 42–65 and encompassing α-helix B and most of β-strand 1. The sequence in this region is not similar to any other known motifs found in data bases such as PROSITE (*87*) or validated Profiles (*112*). Sequence conservation in this region has been noted previously in virtually all studies making use of RecA sequence alignments, but no function has been assigned.

C. Expanded Discussion of Structure–Function Relationships

Structure–function analysis with respect to the primary structure is incomplete, and some conflicts exist with respect to reported functional as-

signments for individual amino acid residues. The overall effort to relate structure to function is bolstered by an array of over 400 available mutants, over 35 of which have been purified and studied *in vitro* (*70, 85*). The discussion here is limited to activities with a role in the recombinogenic function of RecA.

1. ATP Binding

The single binding site for ATP or ADP within the central domain of the RecA structure represents the most highly conserved and clearly identified substructure in the RecA protein. At least three of the structural elements making up this binding site are generally present in bacterial RecA proteins and in the eukaryotic homologs. These elements include the P-loop (*86*), spanning residues Gly-66 through Thr-73 in *E. coli* RecA, with the RecA consensus sequence being GpESsGKT; the Walker B box, spanning residues 140-vivvD-144; and the single residue Glu-96. The interaction of these three structural elements with ATP and ADP is also generally well-supported by the published structure (*65*). The structure provides additional information. The following class-I residues mediate additional interactions with ADP. Asp-94 is believed to stabilize the loop containing Glu-96 by a hydrogen bond to the backbone NH of residues 96 and 97. Asp-144 is thought to interact with the γ-phosphate of ATP and the bound Mg^{2+} ion via a water molecule. This residue has an unusual cis peptide bond between it and Ser-145. Gln-194 has also been proposed to interact with the γ-phosphate of ATP and mediate a conformation change on ATP hydrolysis between ATP and α-helix G and loop L2. Additional contacts between RecA protein and bound ADP occur at Asp-100, Tyr-103, and Gly-265. There is limited amino acid residue conservation at these positions.

The P-loop has been analyzed extensively by mutagenesis. Essentially no amino acid substitutions are permitted at positions 66 and 68, and only a few are permitted at position 69, whereas permitted changes at position 70 are restricted to amino acids with small side chains (*113*). At positions 71–73, limited mutagenesis indicates that there is little tolerance for changes, because all changes introduced to date produce a *rec*$^-$ phenotype.

A notable mutation is K72R, which creates a RecA protein that binds but does not hydrolyze ATP (*114*). *In vitro,* this mutant RecA protein forms nucleoprotein filaments on DNA, carries out a limited DNA strand-exchange reaction, and facilitates cleavage of the LexA protein (*114, 115*). The protein is therefore at least minimally functional in standard RecA activities *in vitro.* However, the mutant gene confers a *rec*$^-$ phenotype equivalent to a *recA* deletion (*113*). The properties of this mutant *in vivo* and *in vitro* have become

important in efforts to sort out the function of RecA-mediated ATP hydrolysis.

Other interesting P-loop mutations occur at position 67, because some of the changes at this position permit a mutagenic separation of RecA functions (*116*). A Pro to Trp change at position 67 creates a RecA protein that has a high constitutive coprotease activity, but affects recombination functions only modestly. Glu and Asp substitutions dramatically reduce recombination functions but leave a low-level constitutive coprotease activity. Changes to Lys or Arg, in contrast, eliminate the coprotease activity, but generally have minimal effects on recombination functions.

2. DNA Binding

To organize the discussion, we point out at the outset that RecA filaments can bind up to three DNA strands within the filament groove, with three distinct DNA strand binding sites (*117–120*). Ample data suggest that four DNA strands cannot be bound in the center of the filament (*121*). A variety of evidence serving to define the three sites will be presented in a number of sections below. The sites are usually called simply I, II, and III (*117–120*). Site I has the highest affinity for a DNA strand and is generally the site at which ssDNA is bound to RecA proteins. Site II is the site generally occupied by a strand complementary to that in site I, although heterologous sequences are tolerated. Site III would then be the site occupied by a DNA strand displaced during DNA strand exchange. The numbering of these sites varies from group to group. When strands of identical ssDNAs are added successfully to a RecA filament, site III actually binds to the DNA with the second highest affinity. In some schemes, site III is therefore called site II, and *vice versa* (*117–119*).

Because there is no RecA–DNA cocrystal, the DNA binding sites of RecA protein have been difficult to define structurally. On the basis of the crystallographic studies, Steitz and co-workers proposed that two loops that are undefined in the crystal structure, L1 (residues 157–164) and L2 (residues 195–209), are involved in DNA binding (*64*). They further suggested that L2 might be part of the single-strand DNA binding site (site I), whereas L1 might be involved in the binding of duplex DNA (site II) (*64*). There is some sequence conservation in these regions among the bacterial RecA proteins. The conservation breaks down when the alignments are extended beyond bacteria. Loop L1 displays a fairly well-conserved net negative charge among the bacterial proteins (Fig. 6C). The charge density in loop 2 is less conserved. It has been suggested that the polypeptide backbone of these loops might contact DNA, obviating a need for amino acid conservation (*4, 122*). If the L1 and L2 loops are involved in DNA binding, the crystallographic studies pro-

vide no information to indicate whether they are involved in DNA interactions at sites I, II, or III, or in more than one of these sites.

The loop L1 region has also been targeted for mutagenesis, and has exhibited significant mutational flexibility (*122*). Pro-151 and Glu-154 are upstream of the L1 loop. Pro-151 can accept a Ser substitution and maintain activity. Glu-154 is a critical residue and all changes except a conservative aspartate substitution result in a *recA*$^-$ phenotype. A number of substitutions can be made at Gly-157 (at the beginning of the L1 loop) and display partial wild-type activity.

Camerini-Otero and co-workers have carried out studies that tend to support the idea that the loop L2 region is involved in DNA binding at DNA strand-binding site I. These include proteolysis mapping (*123*) and cross-linking studies (*124*), as well as studies of the weak but measurable DNA binding and DNA pairing properties of a 20-amino acid peptide that includes L2 (*125*).

The recent protein–DNA cross-linking studies have been intriguing, but have not yet clarified the picture. One study demonstrated cross-linking between bound DNA containing the photoreactive analog 5-iodouracil (IdU) and amino acid residues in the L1 and L2 loops (*124*), generally supporting the proposal of Story *et al.* (*64*). A more detailed examination by Wang and Adzuma indicated that the L1 and L2 loops were involved in binding to different DNA strands at sites I and II, respectively (*126*). Binding of IdU-containing ssDNA to RecA protein, followed by irradiation, resulted in cross-linking to Met-164, in loop L1. When RecA was first bound to a nonphotoreactive oligonucleotide, then chased with an IdU-containing oligo to place it in site II, cross-linking occurred at Met-202 or Phe-203, in loop L2. This work supports the general proposal of Steitz and co-workers (*64*) for DNA binding motifs, but it conflicts with the idea that loop L2 is part of site I (*125*).

Two studies involving direct UV cross-linking of RecA protein to bound DNA yielded different results. In one, Tyr-103 was cross-linked to DNA in the absence of nucleotide cofactor (*127*). When ATPγS was added to the reaction mixture, Tyr-103 was again cross-linked, along with another residue in the region between residues 178 and 183 (*127*). Neither of these cross-links occurs within loops L1 or L2. In another study, cross-links were found in the regions of residues 61–72, 178–183 (probably Lys-183), and 233–243 (*128*). The cross-link at residue 183 is intriguing because it is located near the exterior of the RecA filament, far from the filament axis region where DNA is presumed to bind. In evaluating the various results, it must be kept in mind that RecA protein must have binding sites for three DNA strands, and may have additional DNA interaction sites on the filament exterior to facilitate

the movement of DNA into and out of the filament core. In effect, the various DNA binding studies and proposals need not be mutually exclusive, and all of them may reflect a protein–DNA interaction important in some context of RecA action. Even if they are all informative, we may still have an incomplete outline of the protein–DNA interaction scheme in a RecA filament.

Roles in DNA binding have also been proposed for the N and C termini (*129–131*). These proposals have not yet been substantiated by further studies, and are presently difficult to evaluate. Circular dichroism spectroscopy was used to show that a synthetic peptide representing the first 24 residues of the RecA protein amino terminus does interact electrostatically with ssDNA (*129*). However, the lack of sequence or charge conservation (Fig. 6) in the amino terminus suggests that this is not a general DNA binding mechanism used by RecA proteins. The carboxyl terminus has been implicated in modulating dsDNA binding through the hypothesized mechanism of electrostatic repulsion with the double helix (*130, 131*). However, the existence of basic carboxyl termini in some RecA proteins challenges the universality of this model. It is interesting to note that *Myxococcus xanthus* has two RecA proteins—one with an acidic C terminus (Mxa1) and one with a basic C terminus (Mxa2). Also, all of the *Streptomyces* RecA proteins have long basic carboxyl termini (Fig. 8).

Finally, a DNA binding role has been assigned to various aromatic and charged amino acids between residues 240 and 310 on the basis of sequence similarities among ssDNA-binding proteins, such as *E. coli* SSB and GVP of bacteriophage IKe (*132*). Based on NMR studies (*133*), a structural model for a "DNA-binding wing" of GVP has been proposed, which consists of an antiparallel β sheet made of two β strand separated by a turn. It has been proposed that an analogous DNA-binding wing may exist from residue 243 to 257 of the RecA protein (*133*). Although this region of the bacterial RecA proteins is conserved and positively charged (Fig. 6, B and C), the eukaryotic RecA-like proteins do not retain this motif. This motif was dismissed as having a role in DNA binding because of its location in the structure relative to the filament axis and its involvement in the monomer–monomer interface (Fig. 6D) (*64*).

An aspect of the DNA binding problem not addressed by any of these proposals is the extension and underwinding of DNA that accompanies binding (*134–136*). Intercalators promote the binding of RecA to DNA (*137–139*), and intercalation of aromatic amino acid side chains could provide a mechanism for disrupting base-stacking interactions. It has been suggested that some ssDNA-binding proteins use aromatic residues for intercalation (*140*). However, neither the tryptophans nor tyrosines of RecA protein are involved in DNA

intercalation (*141, 142*). It remains a formal possibility that phenylalanine residues are involved in DNA binding. Phe residues at positions 191 and 217 are invariant in the bacterial alignment, but the former is buried in the monomer crystal structure and the latter is involved in monomer–monomer contacts in the filament (Fig. 3D) (*64*). Other phenylalanines are chemically well conserved in the bacterial alignment (positions 21, 92, 203, and 255), but only position 92 is conserved among all of the eukaryotic RecA-like homologs. This residue is buried in the crystal structure (*64*). The phenylalanine at position 203 is part of loop L2, but it is not intercalated into the DNA (*143*).

Binding to multiple DNA strands is the essence of RecA function, and the lack of conserved residues that can be associated with DNA binding is somewhat surprising. Any discussion of the molecular mechanism of RecA binding to DNA currently entails a high degree of uncertainty. In general, the DNA binding proposals made to date are not well-supported by the alignments presented above. The best proposals so far presented are likely to define only a portion of the overall DNA binding problem.

3. The Monomer–Monomer Interface

Residues Leu-114, Ile-128, and Ala-148 are involved in monomer–monomer contacts (see Figures 2 and 6D). Specifically, Leu-114 belongs to the class III residues as defined by Story *et al.* (*4*). Residues Ser-172, Arg-176, Lys-250, and Pro-254 are the invariant residues making monomer–monomer contacts at the amino-terminal side of the monomer in the RecA filament. Residue 248 also forms part of the polymerization interface at the amino terminus of the RecA monomer. A RecA mutant (K248A) (*144*) shows no activity for ssDNA binding, ssDNA-dependent ATPase, filament formation, and strand exchange. *In vivo*, this mutant shows a UV-sensitive phenotype. Other data have demonstrated the involvement of parts of the N-terminal 33 amino acids in monomer–monomer interactions (*83*).

Skiba and Knight have reported on an extensive mutagenic analysis of part of the monomer–monomer interface within the RecA filament (*145*). The region they chose is coincident with the RecA signature described above. Weak partial activity is maintained with a number of substitutions at Lys-216, Phe-217, and Arg-222. Arg-222 is in the amino-terminal polymerization interface of the filament, and is hydrogen bonded to Glu-63. Lys-216 is hydrogen bonded to the main chain carbonyl of Ala-95 on the adjacent monomer (*145*). Phe-217 makes van der Waals contacts with Thr-150 and Ile-155 of the adjacent monomer. Arg-222 is hydrogen-bonded to His-97 of the adjacent monomer. They proposed that the effect of ATP is propagated through the filament by the interaction of Glu-63 and Arg-222.

4. The MAW Motif

The structural motif between residues 42 and 65 in the *E. coli* RecA sequence is conserved among all RecA homologs. Only two mutants have been isolated in this region: RecA13 (L51F) and RecA56 (R60C). *In vivo*, both mutants display null phenotypes. The *E. coli* harboring *recA13* are sensitive to UV light, are recombination defective (*146–149*), and cannot induce λ prophage (*147*) or participate in the autocatalytic cleavage of the LexA repressor (*150*).

The *recA56* strains have similar phenotypes with respect to recombination deficiency, UV sensitivity, and prophage induction (*148, 151, 152*), along with the inability to induce the SOS response (*153, 154*) or promote SOS mutagenesis (*155*). *In vitro*, both mutants can bind ssDNA in the absence or presence of ADP. However, neither can bind ssDNA in the presence of ATP. In addition, neither mutant protein can hydrolyze ATP. These mutants do not display the allosteric effect enabled by ATP (or ATPγS) binding, i.e., they cannot adopt an active conformation for strand exchange in the presence of ATP (*156*).

The "structurally unusual" class IV residues (*4*) occur within this region at positions 42, 48, 54, and 55. Thr-42 makes a hydrogen bond with Asp-48, which is completely buried (Fig. 6D). In the alignment, both are conservatively substituted by Ser and Asn, respectively. The invariant Gly-54 and Gly-55 have unusual ϕ and ψ torsion angles and stack on top of Asp-48. Consequently, there is no room for side chains at these two positions, explaining the lack of residue substitutions in the alignment. Finally, the amide backbone of Gly-54 makes a hydrogen bond with the carbonyl of Asp-48.

Additional residues conserved in this region include Ala-50, Gly-52, Pro-57, and Glu-63, which are invariant. Glu-63 is in a tract of conserved hydrophobic residues (Ile-61, Val-62, Ile-64, and Tyr-65). The Leu residues at positions 47 and 51 are also notable. Leu-47 is chemically conserved in the analysis at the top of Fig. 4. Leu-51 is found in 61 of 64 bacterial RecA sequences, and does not appear to be chemically conserved only because of one Met and two Thr substitutions in the alignment. Curiously, both of the Thr substitutions are in the RecA proteins of thermophilic bacteria, where an overall decrease in the content of polar residues is observed (*157*). A leucine residue corresponding to Leu-51 occurs in all of the RecA homologs examined to date (Fig. 9B). Based on the crystal structure (*4*), residues Leu-47, Leu-51, Ile-61, and Ile-64 were assigned to class II (part of the hydrophobic core of the RecA protein). Leu-47 and Ile-64 make van der Waals contacts with both Leu-75 and Ile-225 (*158*). Thus, residues in this region interact with different parts of the RecA protein on either side of the ATP binding site, at least in the conformation illustrated by the crystal structure.

This area of the RecA protein is therefore likely to be involved in an ATP-induced conformational switch, consistent with one of several functional suggestions advanced by Kowalczykowski (*84*). We suggest that the region encompassing residues 42 to 65 be defined as a unique RecA structural motif with the acronym MAW, for *M*akes *A*TP *W*ork. The MAW motif is summarized in Fig. 10 (42-tgxxxldxalxxGGlxxgxivEiy-65) and its structure is depicted in Fig. 11.

Conservation of a protein sequence motif across all bacterial and eukaryotic domains often correlates with a ligand-binding function, as seen with the RecA motifs associated with ATP binding. No ligand-binding function has been suggested for the MAW motif, in part because most of the motif is buried in the RecA structure.

5. Miscellaneous Conserved Residues in the RecA Protein

To our knowledge, no information exists for residues Gly-22, Ala-50, Gly-52, Pro-57, Glu-63, Gly-108, Glu-123, Ala-147, Leu-149, Ala-168, Ala-174, Pro-206, Arg-226, Gly-267, and Gly-288. With respect to the functional significance of other conserved amino acid residues in the RecA structure, we focus primarily on work carried out since our previous review (*70*).

The crystallographic studies shed light on the function of a number of other residues (*4, 64, 65, 141, 142*). Gly-122 is at the beginning of α-helix E. Val-146 is a class II residue involved in the hydrophobic core of the RecA monomer. All of these residues are excellent candidates for site-directed mutagenesis studies.

6. Incorporation of Tryptophan Reporter Groups into the RecA Protein

This allows the use of fluorescence assays to monitor the conformational changes that occur on nucleotide and/or DNA binding. Morimatsu *et al.* (*159*) replaced Tyr-103 and Tyr-264 with tryptophan. Previous studies (*65, 160, 161*) showed that these residues interact with ATP and/or ADP. *In vivo,* both mutants retain wild-type levels of UV resistance. *In vitro,* the mutant proteins exhibit a modest reduction in DNA strand exchange activity. The fluorescence of each tryptophan is quenched in the presence of ADP or ATP. However, the quenching of Y103W in the presence of ADP is greater than that of Y264W. This can be rationalized by noting that the Tyr side chain stacks next to the adenosine base of ADP in the crystal structure (*65*). Interestingly, complete quenching was not observed when the Y103W was bound to ATP, suggesting that there might be different binding modes for ADP and ATP. A quenching of the fluorescence signals was observed for both mutants

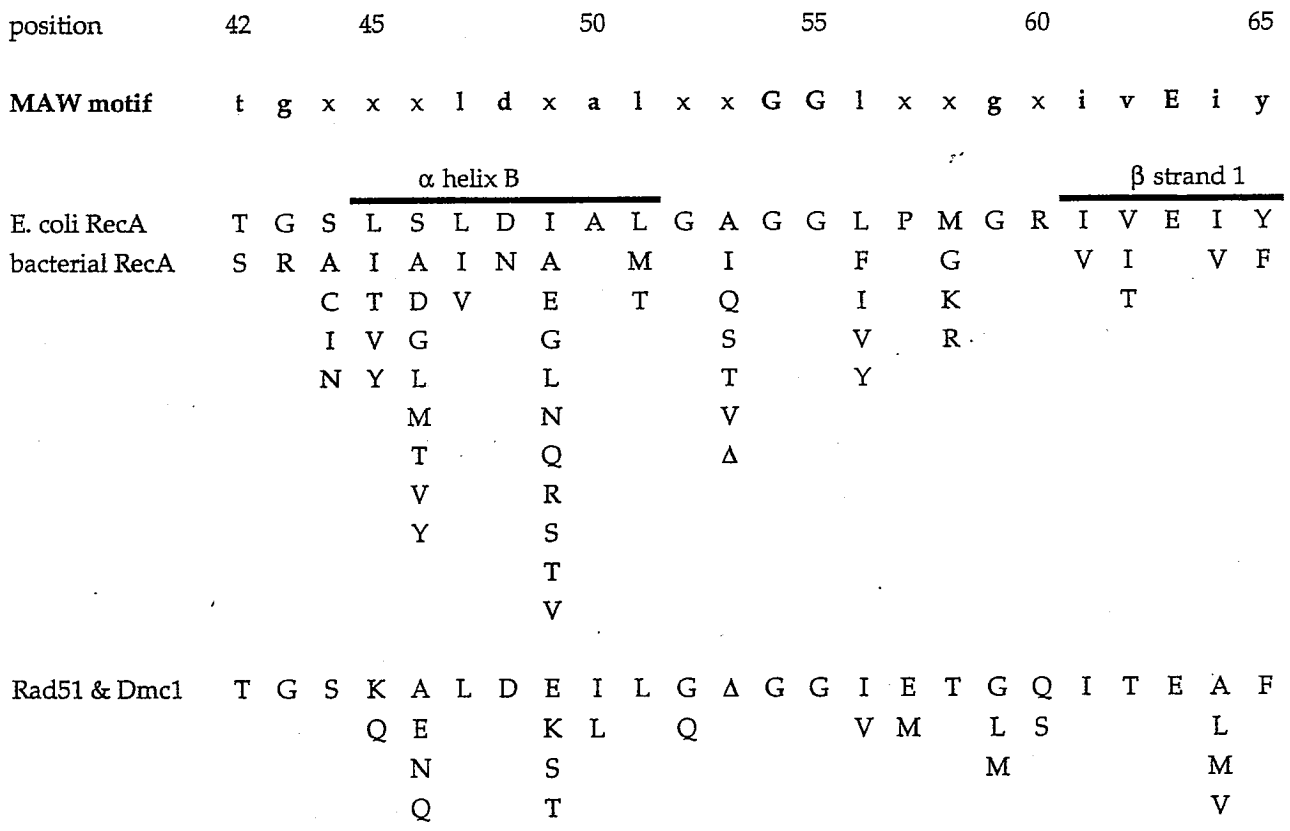


FIG. 10. Sequence of the MAW motif and its distribution in the bacterial and eukaryotic RecA homologs. Uppercase letters represent invariant residues. Lowercase letters represent semiconservative changes. In some instances, lowercase letters are used when the frequency of nonconservative substitutions is low, *e.g.*, at position 43, Arg occurs only once (Amag). A triangle (Δ) represents a one-residue deletion. The positions of the secondary structural elements are indicated.

in the presence of different DNA substrates. This suggests that these residues form part of the binding site or are involved in a conformational change associated with DNA binding. As controls, alanine substitutions at both positions were also studied. *In vivo,* these mutants were UV sensitive whereas *in vitro* they were not competent for strand-exchange activity.

In a similar study, His-163 was changed to Trp; this mutant could be used as a diagnostic tool to monitor the conformational state of the RecA protein (*162*). This work is considered in the following discussion of the RecA ATPase activity. Finally, Phe-203 of the L2 loop was replaced by a tryptophan (*143*).

7. The Sms Protein

All of the universally conserved RecA sequence motifs occur in this protein, advancing the interesting possibility that a second protein with RecA-like function exists in *E. coli.* This notion is supported by the sensitivity of *sms* null mutants to alkylating agents, suggesting that Sms has some role in DNA repair. The *sms* gene is cotranscribed with the *serB* gene, which encodes a component of the serine biosynthesis pathway. It has been suggested that the Sms protein may be involved in the repair of endogenous alkylation damage (*109*). The C-terminal region of the Sms protein, not shown in Fig. 9B, also has sequences similar to the Lon protease of *E. coli.* The *sms* gene is identical to *radA,* mutations in which confer sensitivity to UV- and X-irradiation when logarithmic phase cells are grown in rich media (*163, 163a*). Because Sms lacks significant sequence similarity to RecA protein in regions beyond RecA residue 150, an *in vitro* analysis of Sms activities could help elucidate the function of the MAW motif, as well as the function of structural motifs in RecA beyond residue 150, such as the L1 and L2 loops. The sequence similarity between Sms and RecA is noted in another review (*163b*).

III. RecA Protein Interactions with Its Ligands *in Vitro;* Biochemical Approaches

A. DNA Binding

In the context of recombinational DNA repair, RecA protein brings two homologous DNA molecules together and promotes a DNA strand-exchange reaction between them. Binding of the first DNA can be thought of as primary binding, involving either ssDNA to strand-binding site I or dsDNA to sites I and II. Binding of the second DNA can be called secondary binding. When ssDNA is bound at site I, secondary binding can involve dsDNA binding at strand binding sites II and III or the binding of a second ssDNA at site

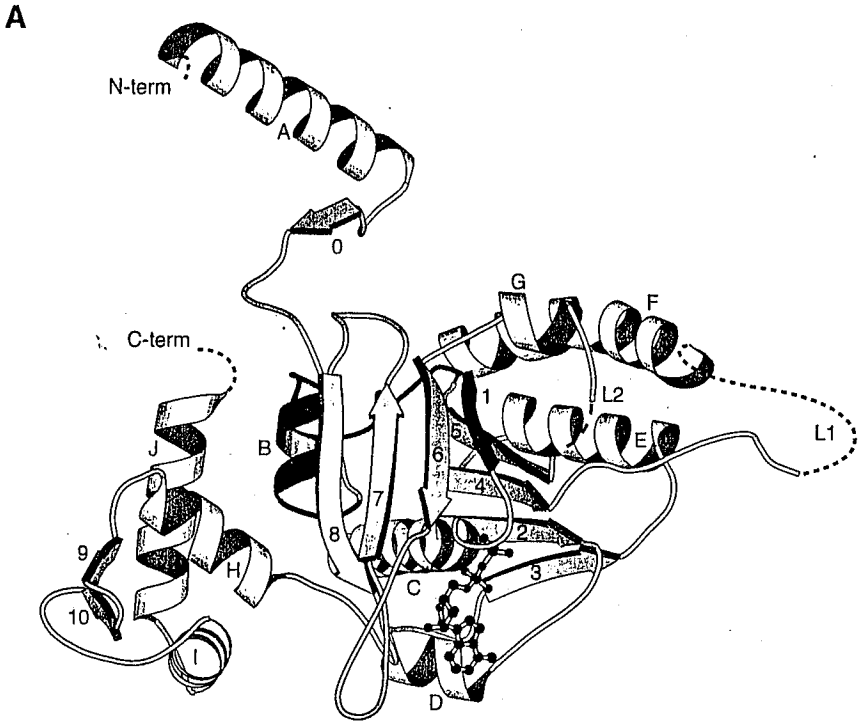


FIG. 11. (A) Schematic representation of the crystal structural of the RecA monomer in the ADP-bound conformation. This view shows the inside face of the RecA monomer. The α helices and β strands are indicated by coils and arrows, respectively. The disordered loops L1 and L2 are indicated by dashes. The ADP molecule is displayed as a ball-and-stick figure. The MAW motif (black) is in the center of the structure. The Walker A box is at the turn that immediately follows β-strand 1. The Walker B box is located at β-strand 4, which is behind the Walker A box. (B) A close-up of the structure of the MAW motif. The conserved residues from Fig. 10 are shown. The hydrogen bond between Thr-42 and Asp-48 is represented by a dashed line. The figures were generated using the program MOLSCRIPT (*428*).

II (*117–120*). This discussion covers only the primary binding of RecA protein to single-stranded, gapped, and double-stranded DNA substrates.

1. Binding to Single-Stranded DNA

Early studies detected direct binding of RecA protein only to single-stranded DNA (ssDNA) (*164, 165*). In the absence of nucleotide cofactor or with ADP, RecA protein binds to ssDNA to form a "collapsed" filament. The resulting structure has a helical pitch of 64 Å and an axial rise of 2.1 Å per nucleotide (*166, 167*). In the presence of ATP, dATP, ATPγS, or ADP·AlF_4^-,

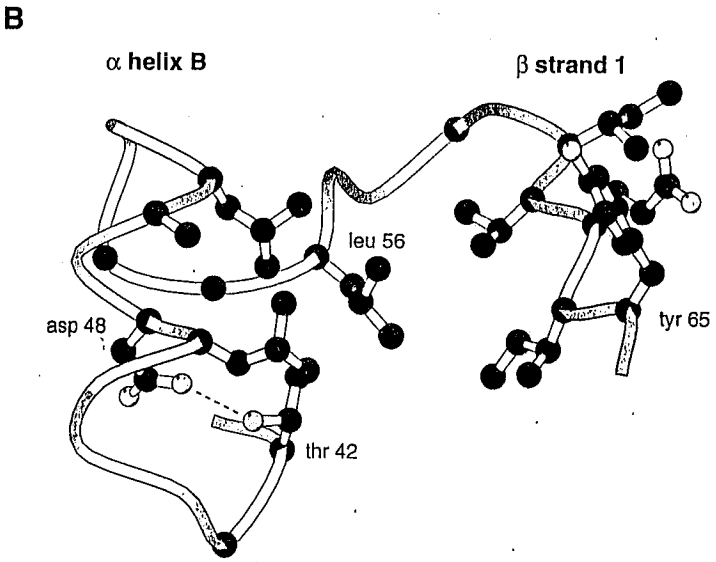


FIG. 11. (*continued*)

an extended filament is formed. This is a right-handed helical form with six RecA monomers per turn, a pitch of 95 Å, and a diameter of 100 Å (*66, 168–170*). There are three nucleotides of ssDNA per RecA monomer in this filament. The DNA is bound along the ribose-phosphate backbone, with the DNA bases displayed in the filament groove (*137, 171, 172*). This single strand of DNA binds to a specific site within the filament, deep within the groove and near the filament axis (*173*), which has been called strand-binding site I (*117, 118, 120*). RecA binding to ssDNA is generally nonspecific with respect to sequence, although the protein exhibits an enhanced affinity to certain homopolymers such as poly(dT) (*174, 175a*). Optimal binding of wild-type RecA protein to poly(dT) generally occurs only with polymers more than 50 nucleotides long (*176*).

Although most measurements indicate a RecA to ssDNA stoichiometry of one monomer per three nucleotides, some titration experiments provided a stoichiometry twice as high (*177–180*). These results could be explained by an intrinsic site size of six nucleotides per RecA monomer, with the lower site size resulting from the binding of a second set of RecA monomers to the first. Alternatively the intrinsic site size could be three nucleotides per monomer, and the higher site size could reflect the binding of a second DNA strand to site II within the filament. The larger site size is observed only when the titration involves a DNA-based signal, such as a fluorescence change arising from

RecA binding to a suitably modified ssDNA. The issue has been resolved in favor of the latter interpretation by Brenner and colleagues (*181*). Their results are also in broad agreement with a range of observations from electron microscopy (*3, 173, 182, 183*).

2. Binding to Double-Stranded DNA

The barrier for RecA binding to double-stranded DNA (dsDNA) turned out to be kinetic rather than thermodynamic. At pH values above neutrality, the rate-limiting step in dsDNA binding is the nucleation of filament formation (*184–186*). Once nucleated, filament extension is rapid and the resulting filaments are stable as long as ATP is regenerated. The nucleation step is highly pH dependent. It occurs rapidly at pH 6.0, making this the preferred pH for many studies of RecA interaction with dsDNA.

Kinetic analysis has divided the nucleation step into two parts (*185*). A preequilibrium has been detected in which RecA protein binds weakly to native DNA. This phase involves the net release of one proton, and exhibits little dependence on temperature or DNA length. The rate-limiting nucleation then occurs, accompanied by partial DNA underwinding (*185*). This phase is accompanied by the uptake of three protons (or a net uptake of two protons for the overall process), explaining the sharp decrease in nucleation rates as pH is increased (*55, 185*).

The nucleation phase is also very temperature dependent, exhibiting an Arrhenius activation energy of 39 kcal mol^{-1} (*185*). Larger DNAs provide bigger targets for nucleation, and thus stimulate the overall binding process. Also, nucleation is accompanied by DNA underwinding, so any alteration in DNA structure that tends to unwind the DNA [negative superhelicity, single-strand gaps, DNA lesions, increased temperature (*70*)] enhances the rate of nucleation (*185*). Reports of enhanced RecA binding to DNAs containing DNA lesions (*187–189a*) thus do not reflect a special RecA affinity for DNA lesions. Instead, they are readily accommodated into a general model for RecA binding to dsDNA in which structural perturbations of the DNA by the lesions enhance the rate of the nucleation step in the binding pathway (*185*). Reports that RecA protein binds more readily to Z form DNA (*190, 191*) have been explained by a similar kinetic perturbation in the general DNA binding pathway (*191, 192*). The filaments formed after RecA binds to Z DNA are still right-handed in their helicity (*191, 192*).

Unlike the case with ssDNA, binding to dsDNA occurs only in the presence of ATP or ATP analogs such as ATPγS (*184, 185*). Therefore, only the extended form of the filament is observed when RecA is bound to dsDNA. Within the RecA filament, the dsDNA is also extended and underwound. The approximately 1.5-fold extension gives a pitch of 100 Å (the same as for bind-

ing to ssDNA), and an axial rise per base pair of 5.1 Å. the DNA is underwound by 43% in the presence of ATPγS (*135*), with a small reduction to 39.6% underwinding when ATP is used and is being hydrolyzed (*136*). This level of underwinding corresponds to about 18 bp per turn of the nucleoprotein filament helix. The degree of dsDNA underwinding observed within filaments actively hydrolyzing ATP varies very little from one filament to another, implying a broad structural homogeneity in these complexes (*136*). Notably, the structural limit of DNA extension is 70% (*193, 194*).

Like ssDNA, dsDNA is bound along the phosphate-ribose backbone and the minor groove (*137, 141, 171, 172*). The major groove is evidently displayed in the filament groove. The binding does not involve the intercalation of Trp or Tyr residues (*141, 195*). The mechanism by which the DNA is extended within RecA filaments has not been elucidated. The bases of the DNA are oriented perpendicular to the filament axis (*195, 196*). It has been argued, on the basis of a variety of experiments, that the two strands are still base-paired to each other (*172, 195–198*), i.e., underwinding does not translate into strand separation within the filament.

Two studies using oligonucleotides modified with the fluorescent chromophore (+)*anti*-benzo(*a*)pyrene-7,8-dihydrodiol-9,10-epoxide (BPDE) provided evidence that all three strands that can bind to a RecA filament interact (*199, 200*). There is a clear homology dependence seen in the signal changes evoked when additional DNA strands are added to filaments with a DNA strand already bound at site I. However, the work also indicates that the base pairing in duplex DNA is not as tight as in B-form DNA (*199, 200*). The dsDNA is located deep within the filament groove (*201*), at a site that should include DNA strand binding sites I and II.

3. Binding to dsDNA with ssDNA Gaps

A single-strand gap of 50 nucleotides or more is the best way to circumvent the slow nucleation step in RecA binding to dsDNA at neutral pH values and above. Nucleation occurs in the gap, and the filament is rapidly extended to encompass the adjacent dsDNA. The entire filament exhibits a structural polarity that reflects the orientation of the strand in the gap (*183*). The two strands of the dsDNA are also bound asymmetrically. The strand in the gap apparently is bound in the site normally occupied by ssDNA, and thus can be considered bound at site I. The strand bound at site I is better protected from nuclease digestion than its complement (bound in site II) by factors of 2–3 (*202, 203*), providing one experimental distinction helping to define sites I and II. This reflects the different fates of each strand. During strand exchange reactions, the strand in site I remains in the filament throughout the process, while its complement is exchanged out of the filament.

B. Polar Assembly and Disassembly of RecA Filaments, and the Importance of 3′ DNA Ends

Distinct nucleation and extension phases of RecA filament assembly on DNA can be discerned when either ssDNA or dsDNA is the lattice on which assembly occurs. On ssDNA, both phases are rapid and the kinetics have not been analyzed in sufficient detail to provide rates for individual steps. On ssDNA circles over 8000 nucleotides long, filament formation is completed in less than 2 min at 37°C. If nucleation is sufficiently rate limiting as to restrict the process to only one nucleation event per circle, we can infer a rate of filament extension of over 1100 RecA monomers per minute. The active binding unit or protomer (monomer, dimer, or hexamer?) has not been defined. Rods of RecA protein are a predominant structural form under conditions commonly used to monitor RecA reactions *in vitro* (*60*).

The single-stranded DNA binding protein (SSB) of *E. coli* plays an important role in the assembly process. RecA protein does not readily bind to regions of secondary structure in ssDNA. When RecA binds to ssDNA molecules with significant secondary structure, such as the DNA derived from bacteriophage M13, binding does not lead to full coating of the DNA circle. Added SSB greatly stimulates binding and facilitates the creation of a contiguous and stable filament that coats the entire DNA molecule (*204–209*). The effects of SSB are best explained by a mechanism in which SSB binds to and melts out the secondary structure in the ssDNA, and is then displaced by RecA protein (*210, 211*). In effect, the RecA protein and SSB compete for binding sites on the ssDNA. When bound to ATP, RecA protein can displace SSB from the DNA. However, SSB affects the steps of filament formation differently. SSB strongly inhibits the nucleation of RecA filament formation, while it is readily displaced from the ssDNA during the filament extension phase. This leads to an order-of-addition effect. When RecA is added first to a solution containing ssDNA, it binds to regions lacking secondary structure. The best stimulation of RecA filament formation occurs when SSB is added after RecA protein, where it removes secondary structure in the DNA and is displaced by extension of the RecA filaments already bound (*207, 210–212*). However, when SSB is added before RecA protein it completely covers the ssDNA and there is a substantial lag in RecA protein filament formation (*207, 212*) explained by the inhibition of filament nucleation.

The properties of the RecA filament assembly process have played a prominent role in the development of models to explain the function of RecA and RecBCD proteins in the initiation of recombinational processes. Many recent efforts to reconstitute portions of homologous recombination reactions *in vitro* have stressed the importance of prepared regions of single-

stranded DNA with free 3′ ends in the initiation of genetic exchanges (*213–215*). In bacteria, the *in vivo* evidence that 3′ ends are more recombinogenic or invasive is limited. The observation that *sbcB* mutations, which affect the 3′ to 5′ exonuclease I, restore recombination levels in *recBC* mutants (*27, 216–218*) has often been cited as evidence that the generation and protection of 3′ ends is important. However, some mutations or deletions that eliminate exoI function do not restore recombination in *recBC* mutants (*219*), calling the importance of 3′ ends into question. The key *in vitro* observation relevant to the role of 3′ ends in recombination has been the polarity of the RecA filament assembly reaction.

On ssDNA, filament extension after nucleation is polar and proceeds 5′ and 3′ (*220*). On circular ssDNA, the process continues until the DNA is completely encompassed. On linear ssDNA, nucleation is presumed to occur virtually anywhere and to proceed to the 3′ end. The 3′ end is therefore much more likely, than the 5′ end to be coated with RecA protein. The observation is readily incorporated into models in which 3′ ends exposed by the action of nucleases or helicases are more recombinogenic or invasive than 5′ ends. A greater propensity for RecA filaments on ssDNA to promote DNA pairing at the 3′ end of a linear ssDNA has been directly demonstrated *in vitro* (*213, 221–224*). However, as detailed below, the final chapter on the molecular origin of these effects and their implications for recombination *in vivo* has not yet been written.

Filament assembly is also polar on dsDNA. With linear dsDNA, the slow nucleation step is readily circumvented with 5′ single-strand extensions but not by 3′ extensions (*203, 225, 226*). The single-stranded DNA in a 5′ single-strand extension or a single-strand gap becomes the initiating strand in filament formation. The filament extension then encompasses the adjacent dsDNA, proceeding 5′ to 3′ along the initiating strand.

The major pathway for disassembly of RecA filaments is also unidirectional and end dependent. On dsDNA, RecA monomers dissociate from the end opposite to that at which assembly occurs (*226*). Filament disassembly therefore proceeds with the same polarity as assembly, 5′ to 3′ relative to the DNA strand in site I. The rate of disassembly is highly pH dependent. Below pH 6.75, linear RecA filaments exhibit little or no net disassembly. Above pH 6.75, end-dependent disassembly is observed, increasing in rate with increasing pH until a maximum is observed above pH 8.1 (*203*). The maximum rate may approach 200 monomers min^{-1} (*203*).

Filament assembly must have the equivalent equilibrium constants for subunit addition at either end, unless assembly or disassembly at one end is coupled in some manner to an input of chemical energy. Inasmuch as assembly and disassembly of RecA filaments can occur at different filament ends in one test tube under a single set of conditions, one process or both

must be facilitated by ATP hydrolysis. Assembly occurs readily with ATP analogs that are not hydrolyzed, but disassembly does not, indicating that end-dependent filament disassembly is coupled to ATP hydrolysis. The disassembly process accounts for only a small fraction of the ATP hydrolyzed by a filament. In the filament interior, ATP is hydrolyzed by each monomer but protein–protein interactions on both sides of the monomer apparently prevent dissociation most of the time. On the disassembly end, the probability that the monomer will dissociate as a result of ATP hydrolysis is a function of pH.

More recently, a 5′ to 3′ disassembly of RecA filaments from linear ssDNA has also been documented (227). Net disassembly is observed only in the presence of SSB. As RecA monomers or protomers dissociate, they are replaced by SSB, which serves to inhibit renucleation of new filaments. Because the filament extension process displaces SSB but proceeds 5′ to 3′, the SSB bound to growing segments of the DNA at the disassembly end is not readily displaced by RecA (227). This disassembly process must contribute to the observed bias for RecA-mediated DNA pairing at the 3′ end of the linear ssDNA. Not only can filaments nucleate at points away from the 5′ end, but the disassembly process will actively remove RecA protein from the 5′ end. The rate of disassembly from ssDNA and its pH dependence have not yet been analyzed in detail, but at pH 7.5 the rates are comparable to those observed from dsDNA under the same conditions. It should be noted that disassembly is prevented if dATP replaces ATP as nucleotide cofactor, or if the *E. coli* RecO and RecR proteins (described in Section V,B) are present (227). In the presence of the RecO and RecR proteins, RecA filaments stably coat linear ssDNAs contiguously from the 5′ to the 3′ ends. It can be argued, therefore, that free 5′ DNA ends should participate in RecA protein-mediated DNA pairing processes as readily as 3′ ends in any cell that contains the RecO and RecR proteins.

RecA monomers in the interior of filaments formed on ssDNA exhibit only a very modest exchange reaction with free RecA monomers, and no exchange at all when dATP replaces ATP (228). Some direct transfer of RecA monomers or protomers between adjacent RecA nucleoprotein filaments is observed (229), but ATP hydrolysis rarely results in simple dissociation of a RecA monomer from the filament interior.

When filaments are formed on dsDNA, however, a relatively robust exchange between bound and free RecA monomers in the filament interior is observed (228). This indicates that there is a change in the state or conformation of the filament when a second DNA strand is added at site II. Experiments probing the interaction of DNA strands within the RecA filament with BPDE-labeled oligonucleotides also suggest a change in filament state when strand-binding sites II and/or III are occupied (*199, 200*).

The mechanism of the exchange observed for monomers in the RecA filament interior is not yet known. It could involve simple dissociation and replacement by other monomers. Alternatively, it could represent an exchange between RecA nucleoprotein filaments and unbound aggregates or filaments of RecA protein with which they come into transient contact. The latter process could be mechanistically related to the direct transfer of RecA protomers between RecA–ssDNA filaments (*229*), or from nucleoprotein filaments to unbound DNA (*177, 230*). Conformation changes observed within bundles of RecA filaments associated with slow hydrolysis of ATPγS might be interpreted as a side-by-side exchange of RecA monomers between filaments (*231*).

C. ATP Hydrolysis

The RecA protein is a DNA-dependent ATPase. Rates of ATP hydrolysis are modest. The maximum reported value for k_{cat}/K_M at 37°C is about 2×10^4 sec^{-1} M^{-1}, some 10^{-4} that of the diffusion-controlled limit. The activity is affected by the structure of the bound DNA and by monomer–monomer interactions within the filament. The kinetics are complicated by cooperativity, and the K_M is more accurately reported as $S_{0.5}$.

Two methods are commonly used to monitor ATP hydrolysis by RecA. One employs labeled ATP, and thin-layer chromatography to separate ATP, ADP, and P_i (*232, 233*). The other is a spectrophotometric assay in which ATP hydrolysis is enzymatically coupled to an NAD–NADH interconversion (*203, 212, 233–235*).

1. DNA-independent ATP Hydrolysis

RecA-mediated ATP hydrolysis in the absence of DNA is slow but detectable (*232*). Under conditions used to monitor many RecA-mediated processes *in vitro*, with the aggregate concentration of salt and buffers at less than 150 m*M*, the DNA-independent ATPase activity peaks at about pH 6.0 with a k_{cat} of 0.1 min^{-1}. At pH 7.5, the k_{cat} drops to 0.015 min^{-1}. The rate of ATP hydrolysis increases as a third- to fourth-order function of salt concentration. At very high (1.5–2 *M*) concentrations of a wide variety of salts, the rates of ATP hydrolysis become comparable to rates observed in the presence of DNA (*236, 237*). The nature of the salt dependence and observed changes in circular dichroism spectra at high salt concentrations (*238, 238a*) suggest that salt can facilitate a conformation change that to some degree mimics that induced by the binding of DNA to RecA protein.

2. DNA-dependent ATP Hydrolysis

Accurate measurement of RecA-mediated ATP hydrolysis on ssDNA depends either on the use of DNA cofactors that lack secondary structure

[such as poly(dT)] or on the presence of SSB, which serves to eliminate secondary structure and permit uniform DNA binding by RecA (as already described). When bound to ssDNA, RecA protein hydrolyzes ATP with a k_{cat} of 28–30 min^{-1} at 37°C. The ATPase activity exhibits no dependence on pH between 5.5 and 9.0 (*232*). The effects of temperature between 25 and 45°C produce a linear Arrhenius plot, from which an Arrhenius activation energy of 11.8 ± 0.3 kcal mol^{-1} can be derived (*233*) for the reaction.

The ATP hydrolysis occurs uniformly throughout the RecA filament, with no enhancement at filament ends (*176, 212, 239*). As ssDNA is titrated into a reaction mixture, the observed rate of ATP hydrolysis generally increases linearly until sufficient DNA is present (three nucleotides per RecA monomer) to bind all of the RecA protein present. This direct relationship between ATP hydrolysis and DNA binding is observed under most, but not all, conditions. The rate of ATP hydrolysis can be a convenient method to monitor DNA binding if appropriate controls and corroborating methods are included in the data set.

ATP is hydrolyzed to ADP and P_i. ADP acts as a competitive inhibitor of ATP hydrolysis. Another common inhibitor is the ATP analog ATPγS [adenosine 5′-*O*-(3-thio)triphosphate), which is bound but not appreciably hydrolyzed by RecA protein and acts as a potent competitive inhibitor. The inhibition patterns in both cases are greatly complicated by additional effects of these two nucleotides (*240, 241*). As already described, each tends to stabilize cooperatively different forms of a RecA filament.

A variety of rNTPs and dNTPs are hydrolyzed by RecA protein (*232, 242, 243*). Among them are dATP, UTP, PTP (purine ribonucleoside triphosphate), and dUTP, which serve as cofactors in at least some RecA-mediated reactions (*178, 232, 243*). Others, including ITP, CTP, dCTP, and GTP, are hydrolyzed with a nearly normal V_{max}, but do not support DNA strand exchange. Bryant and colleagues have pointed out that although a number of nucleotide triphosphates are hydrolyzed, only those with a measured $S_{0.5}$ below 100 μ*M* serve as cofactors in the DNA strand-exchange reaction (*243, 244*). The reactions with dATP are particularly notable. With dATP, hydrolytic rates are increased about 20% relative to ATP hydrolysis. RecA filament stability and the rates of DNA strand exchange are also enhanced (*115, 178, 228*). The RecA mutant, RecA K72R, which binds but does not hydrolyze ATP, will promote a limited DNA strand-exchange reaction only if dATP is used as the nucleotide cofactor (*114, 115*).

When RecA protein is bound to dsDNA, the k_{cat} for ATP hydrolysis is reduced to 20–22 min^{-1}. The rates again exhibit no dependence on pH between 5.5 and 9, as long as RecA filaments are fully formed on the DNA (*184*).

3. Mechanistic Aspects of ATP Hydrolysis: Conformation Changes and Cooperativity

As already described, three different forms of RecA filaments have been defined structurally, principally by electron microscopy. When bound to ATP or to certain ATP analogs, the resulting nucleoprotein filament on ssDNA is extended. With either ADP or no nucleotide cofactor, the filament forms are described as collapsed, although they are distinct. The structure derived from crystallographic studies probably most closely approximates the ADP-bound form. The same three filament forms can also be defined on the basis of DNA binding studies. The affinity of RecA protein for ssDNA is increased by ATP and decreased by ADP, with a form exhibiting intermediate affinity present in the absence of nucleotide cofactor (*245, 246*). The ADP-bound and no-nucleotide filament forms are inactive in DNA pairing and strand-exchange activities.

Based on the extrapolation that these RecA species represent conformations occurring during the normal ATP hydrolytic cycle, these observations have inspired a number of models for ATP hydrolysis. One proposal suggested that ATP hydrolysis would be coordinated in the monomers along the filament, leading to an accordion-like extension and retraction of the filament structure (*166, 167*). However, a detailed examination of conformations observed during the slow hydrolysis of ATPγS (*66*) indicated that hydrolysis resulted in no collapse to anything resembling the ADP-bound filament form. The degree of DNA underwinding within a filament actively hydrolyzing ATP is quite uniform from one filament to the next, contributing to a picture in which there is a broad structural homogeneity between filaments (*136, 203, 227*). Complementary studies on the RecA ATPase inhibition patterns observed with ADP and ATPγS demonstrated that conversion of the ADP filament form to the ATP form proceeded with a transient phase during which the ADP filaments were disassembled and reassembled in the ATP form. The two forms are mutually exclusive. The observed ATP and ADP filament forms are therefore not directly interconvertible in the sense of the entire filament alternating between them, and the accordion model has been eliminated from consideration.

The possibility remains that within a filament actively hydrolyzing ATP, a subset of the RecA monomers might exist at any moment in a conformation equivalent to the ADP filament conformation defined by the studies carried out with ADP alone. ATP hydrolysis in individual monomers could result in cycling between the ATP and ADP filament forms, with the filament as a whole representing a dynamic mixture of these and perhaps other conformations. As suggested (*5, 85*), this could lead to a cycling between conformations with high and low affinities for ssDNA, effectively coupling the

ATP hydrolysis to an association/dissociation cycle. This would solve the molecular dilemma posed by the competing requirements of a protein like RecA to bind tightly to DNA to promote some reaction, yet dissociate when the reaction was completed (*5, 85*). In support, studies on the end-dependent disassembly of RecA filaments provide evidence that ATP hydrolysis is involved in the dissociation of RecA monomers (or larger protomers) from one filament end (*5, 85, 226, 227*).

However, there is no obligate coupling of ATP hydrolysis to RecA dissociation, and dissociative processes rarely account for more than a minute fraction of the ATP hydrolyzed. When RecA filaments are bound to ssDNA circles, there is very little exchange between free and bound RecA, and none at all if dATP is used as the nucleotide cofactor (*228*). On linear ssDNA, ATP-mediated dissociation is observed at the filament end nearest the 5′ end of the DNA, but ATP hydrolysis in the filament interior does not result in dissociation. Even the end-dependent dissociation is suppressed on ssDNA when dATP is used (*227*). A picture emerges in which dissociation in the filament interior is limited by interactions with RecA monomers on either side, while dissociation at the "disassembly" end would occur with a probability influenced by the nucleotide cofactor used and other factors (*203, 227*).

As of this writing, it is unclear whether the inactive RecA filament conformation observed in the presence of ADP alone represents a conformation that occurs even transiently in individual monomers during the normal ATP hydrolytic cycle. New probes of RecA conformation changes are clearly needed. A particularly promising approach involves the RecA H163W mutant protein, with the change located in loop L1 (*162, 247*). The substitution does not appear to affect appreciably any aspect of RecA function. However, the new tryptophan residue provides a fluorescence signal that is quenched on the binding of nucleotide cofactors. ADP addition to a RecA–ssDNA complex leads to a small quenching of the fluorescence signals, providing new evidence that the ADP filament form is conformationally distinct from the form present in the absence of nucleotide (*162*). ATP addition leads to a much greater quenching of the fluorescence signal (*162, 247*). The change in signal reflects a first-order process that occurs with a rate constant equivalent to the k_{cat} for ATP hydrolysis (E. Stole and F. R. Bryant, personal communication). The signal is also closely associated with the conformation change to an extended filament form that accompanies ATP binding. ITP is hydrolyzed by RecA protein with a k_{cat} equivalent to that for ATP, but it does not support DNA strand exchange, it does not promote the conformation change to the extended filament, and its binding does not cause the quenching of the fluorescence signal brought about by ATP (*247*). The signal change observed with ATP is similar under at least some conditions to that brought about by ATPγS binding (E. Stole and F. R. Bryant, personal communication).

These observations are leading to a dissection of steps in the ATP hydrolytic cycle. The rate-limiting step is a first-order process and thus follows ATP binding. The absence of a burst of ATP (or ATPγS) hydrolysis indicates that the rate-limiting step precedes the hydrolytic step. Filament extension occurs with a first-order rate constant identical to the k_{cat} for ATP hydrolysis. However, the rate-limiting step and filament extension are not equivalent, because ITP hydrolysis appears to share the same rate-limiting step with ATP hydrolysis, yet does not support filament extension. The result is a scheme in which NTP binding is followed by a slow transition, possibly a conformation change that does not affect the fluorescence properties of the H163W mutant protein. If ATP is the nucleotide cofactor, the slow transition is accompanied by or rapidly followed by filament extension. A relatively fast hydrolytic step then follows (F. R. Bryant, personal communication). The steps involved in dissociation of ADP and P_i are not yet defined, but do not limit the rate of the reaction. RecA protein does not promote a detectable exchange of [^{3}H]ADP ⇌ ATP, $HPO_4^- \rightleftharpoons H_2{}^{18}O$, or $HP^{18}O_4 \rightleftharpoons H_2O$ (*248*). Thus, ATP hydrolysis is neither macroscopically nor microscopically reversible on the enzyme.

Once extended, the filament remains in this state as long as ATP is regenerated. Filament collapse and disassembly occur only when the ADP/ATP ratio exceeds 1.0 (*240, 241, 246, 248*). As previously indicated, the degree of extension and associated DNA underwinding is essentially the same whether the extension is brought about by ATPγS, or within a filament actively hydrolyzing ATP. The RecA monomer conformation observed in all monomers in filaments bound to ADP must be populated for a very small fraction of the time required for the entire ATP hydrolytic cycle. Alternatively, this conformation may not occur at all during the hydrolytic cycle. Monomer–monomer interactions within the filament may constrain each monomer to maintain an extended conformation even when ADP is transiently bound. Individual monomers might go through a series of conformation changes during the hydrolytic cycle, but the conformation present when ADP was transiently bound would be distinct from that observed when the entire filament was formed in the presence of ADP. Inhibition studies show that ATPγS and ADP are antagonistic inhibitors of ATP hydrolysis, and that the ATP and ADP filament forms are structurally incompatible (*240*). Whereas filaments form on ssDNA in the presence of either ATP or ADP, the incompatibility of the two filament forms results in nearly total dissociation when a 50/50 mixture of ATP and ADP is present, with each nucleotide antagonizing the filament form stabilized by the other (*241*).

A reasonable hypothesis is that the binding of ATP to a sufficiently large subfraction of the monomers in a filament results in the formation of a static and extended structural core conformation maintained by certain seg-

ments of the RecA monomer structure and reinforced cooperatively through monomer–monomer interactions. Outside of this core structure, other elements of the structure of each monomer may undergo conformation changes in concert with the ATP hydrolytic cycle to effect the late stages of DNA strand exchange. When ATP hydrolysis occurs in a monomer or a segment of monomers at the disassembly end of a filament, a momentary collapse of the core structure to the ADP form could lead to dissociation at that point. These ideas incorporate some aspects of the active cluster hypothesis proposed by Kowalczykowski (*242, 246*).

These ideas and observations imply a high degree of cooperativity between monomers within a RecA filament. The cooperativity is manifested by Hill coefficients greater than 1.0 for ATP hydrolysis observed under many conditions (*249*), and a cooperativity parameter, ω, of 50 ± 10 for binding to ssDNA (*245*). It is also manifested by unusual kinetic effects. For example, RecA-mediated GTP hydrolysis is greatly stimulated by low concentrations of ATP (*250*). Also, very low concentrations of ATPγS actually stimulate the hydrolysis of ATP under some conditions (*240*).

Cooperativity must play a role in the filament disassembly reaction. Within RecA filaments formed on dsDNA, individual monomers hydrolyze ATP with a turnover of about 20 min^{-1}, or about 1 ATP every 3 sec. If the ATP hydrolytic cycles of adjacent RecA monomers were entirely independent, ATP hydrolysis would occur in the monomer at the disassembly end about 1.5 sec (on average) after the previous monomer dissociated. If ATP hydrolysis and end-dependent dissociation are coupled, the maximum rate of dissociation would be one every 1.5 sec, or 40 monomers per minute. However, the maximum observed rates of filament disassembly are at least three times this. This implies that the ATP hydrolytic cycle of adjacent monomers are coupled and temporally offset from one another, leading to waves of hydrolysis proceeding down the filament from the disassembly to the assembly ends (*70, 226*). To reconcile the observed rates of filament disassembly with the ATPase turnover rates, the waves must be coordinated and spaced every 6–12 monomers (*70, 226*). In the filament interior, a new wave must reach a given monomer every 3 sec. More accurate measurements of end-dependent filament disassembly rates are needed to refine this picture.

IV. RecA Protein-mediated DNA Strand Exchange

A. Overview of the Reaction

Typical DNA strand-exchange reactions used to study RecA function *in vitro* are illustrated in Fig. 12. The reactions shown are arranged for convenience, because it is easy to distinguish products from substrates with a wide

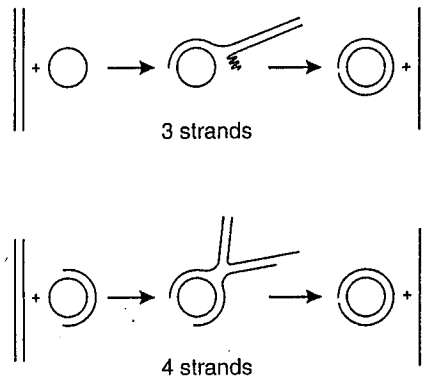


FIG. 12. RecA protein-mediated DNA strand-exchange reactions.

range of assays. DNA substrates are usually derived from bacteriophages, often M13 or its derivatives. The reaction can involve either three or four DNA strands, as shown, and it nicely mimics several of the putative steps in recombinational DNA repair. A Holliday intermediate is formed transiently in the four-strand reaction. The discussion remains organized around the concept of three DNA strand-binding sites in the RecA filament, in spite of the capacity of RecA filaments to promote the four-strand exchange.

There are at least four experimentally or at least conceptually distinguishable phases of a RecA protein-mediated DNA stand-exchange reaction. The first is the formation of a RecA nucleoprotein filament that completely coats the single-stranded or gapped DNA substrate. In the discussion below, this DNA substrate is called DNA1, and it binds at the DNA strand-binding site defined as site I. The resulting nucleoprotein filament is effectively a sequence-specific DNA-binding entity, with the sequence specificity determined by the bound DNA. Phase 1 is also called presynapsis. The bound single strand is aligned with a linear duplex DNA (DNA2) in the second phase of the reaction. DNA2 is brought into the filament at DNA strand-binding sites II and III. In the third phase, there is a rapid strand switch to create 1000 bp or more of hybrid DNA. This often produces the branched molecules shown in Fig. 12. Phase 3 requires ATP but not its hydrolysis. In reactions involving four DNA strands, the DNA pairing and strand switching in phases 2 and 3 are limited to the single-strand gap in the gapped duplex DNA (*115*, *251*). In other words, the fundamental DNA pairing process in RecA filaments will accommodate only three DNA strands. Phases 1–3 can result in the formation of thousands of base-pairs of hybrid DNA without ATP hydrolysis. However, under most conditions, the completion of DNA strand exchange requires a fourth phase in which the nascent hybrid DNA is extend-

ed in a facilitated branch migration reaction that proceeds until products are formed. This final phase requires ATP hydrolysis and will accommodate four DNA strands. For a variety of reasons, phases 2 and 3 are difficult to distinguish experimentally. Phases 1 and 4 are readily distinguished and experimentally defined relative to the DNA pairing in phases 2 and 3.

The RecA filament formed in phase 1 hydrolyzes ATP with a k_{cat} of 30 min^{-1}. On addition of a homologous duplex DNA, the rate of ATP hydrolysis decreases rapidly to a turnover of 20–22 min^{-1} (*239*). The decline is dependent on the length of homology in the duplex DNA, and therefore reflects a change brought on by DNA pairing (*239*). The apparent efficiency of the strand-exchange reaction is low, with about 100 ATPs hydrolyzed for every base pair of hybrid DNA formed under typical reaction conditions.

The branch migration in phase 4 is unidirectional (5′ to 3′ with respect to the single strand initially bound), and proceeds at a rate of 380 bp min^{-1} at 37°C (*233*). This polarity affects the efficiency with which DNA pairing occurs at the two ends of the duplex DNA substrate. The end of the duplex at which a complete strand exchange would normally be initiated (the 5′ end with respect to the displaced strand) is therefore called the proximal end, and the other end is called the distal end. Phase 4 proceeds readily past mismatches, lesions, and even heterologous inserts (up to a few hundred base pairs in length) in one or both DNA substrates, a capability that may be critical to its function in recombinational DNA repair (*115, 252–255*).

The DNA binding and filament formation that occurs in phase 1, as well as the properties of the resulting filaments, has already been described in Section III,A. The discussion below focuses on the DNA pairing (phases 2 and 3), and the hybrid DNA extension that occurs in phase 4.

B. DNA Pairing

DNA pairing occurs within the filament, as originally proposed by Paul Howard–Flanders (*256*) and confirmed over the past decade in a wide range of studies. However, the four-stranded DNA pairing intermediate suggested is yielding to models in which DNA pairing within the RecA filament is restricted to three strands. The initial pairing interaction in these reactions always involves alignment of a single strand with a homologous duplex DNA. Holliday structures involving reciprocal strand exchange among four DNA strands are created via branch migration following a three-stranded DNA pairing event. Although four-stranded DNA structures have often been promoted as potential recombination intermediates (*5, 256–260*), there is no unambiguous evidence for such an intermediate in RecA reactions and much evidence against it (*121*).

Consider the DNA binding properties of RecA protein reviewed above. RecA binds readily to ssDNA, but very slowly to dsDNA. Many of the en-

zymes identified as functioning early in recombinational processes are helicases and nucleases (RecJ, RecQ, RecBCD) that help generate a single-stranded DNA substrate for RecA (*29, 85, 217*). A wide range of biochemical and biophysical studies highlight the ease with which three DNA strands are incorporated into the major groove of a RecA filament, and the general absence of evidence for the incorporation of a fourth DNA strand (*117, 119, 121, 199, 261–264*). The DNA strand binding site III can be distinguished by some of the same experimental approaches used to distinguish site I from site II. For example, a DNA strand bound at site II is more sensitive to nuclease degradation than strands bound at site I or III (*196*). When combined with *in vivo* work demonstrating the importance of single-stranded DNA in initiating recombination, a simple picture emerges. In most situations requiring homologous recombination, the major DNA pairing event involves a single-stranded DNA bound by RecA protein and a homologous duplex DNA.

The mechanism of the DNA pairing process in the DNA strand-exchange reaction has not been worked out in detail. However, several experimental facts seem clear and consistent over a range of studies in different laboratories. First, the DNA pairing process does not require ATP hydrolysis. Second, DNA pairing does require the extended RecA filament conformation brought about by binding to ATP or ATPγS. Finally, the product of the reaction is a complex in which the hybrid DNA product of a strand switch is stabilized in the filament, probably with the displaced strand remaining within the filament, wound around the hybrid DNA but not necessarily interacting with it.

1. Kinetics

The kinetic mechanism of DNA pairing is surprisingly poorly understood. DNA pairing can, in principle, be broken down into a number of steps. The RecA nucleoprotein filaments on ssDNA will bind at least weakly to duplex DNA whether it is homologous or not (*118, 119*). Nonhomologous interactions with the duplex DNA substrate would be a presumed first step. These interactions lead to the formation of extensive aggregated networks of DNA when RecA protein is mixed with heterologous duplex DNA (*265, 266*). Rapid DNA pairing can occur within these networks when the DNA is homologous, although incubation of RecA nucleoprotein filaments with completely heterologous DNA can block pairing of homologous DNA added later (*266*).

The presence of heterologous DNA sequences flanking a homologous sequence in the duplex DNA did not enhance the rate of DNA pairing in one study using contiguous RecA filaments formed on ssDNA in the presence of SSB (*267*). However, another study demonstrated an enhancing effect of flanking heterologous DNA when the more fragmented filaments formed in the absence of SSB were used (*265*). Pairing, in the sense of an initial ho-

mologous alignment, can occur anywhere along the length of either DNA substrate, and is generally completed within a few minutes (*239, 268, 269*). Formation of a paired DNA species retaining significant stability when RecA protein is removed can take somewhat longer and generally requires the presence of a DNA end allowing a net exchange of DNA strands (*268, 269*).

There is still much to be learned about this process by careful kinetic studies. One effort in this direction has recently been reported (*270*). The use of very short single-strand oligonucleotides (suboptimal for RecA binding under most conditions) and an unusual mixture of ATPγS and ADP in this study produced rather slow rates of pairing, making the results difficult to compare with other RecA studies. However, this and most other studies done to date contribute to the present consensus that the homologous alignment of the two DNA substrates (roughly the process described above as phase 2) does not limit the overall rate at which stably paired complexes are formed. This suggests that some aspect of the strand switch in phase 3 is rate limiting.

2. DNA Structures and Proposed Intermediates

A more vigorous effort has been directed at the DNA structural aspects of the pairing reaction. The central idea guiding much of this effort is the possible involvement of a novel DNA triplex structure, sometimes referred to as R-form DNA (*271*). Perhaps the first mention of a recombination triplex DNA intermediate was put forward by Lacks (*272*), and the structures of DNA triplex intermediates suggested by a variety of studies are remarkably consistent with the original proposal by Lacks. A consideration of possible paths of DNA pairing reveals two reaction stages during which a DNA triplex could occur (Fig. 13). In one path, the duplex DNA would enter the RecA filament groove and align via its major groove with the ssDNA. This would form what has been called a prestrand switch triplex (*273*). The strands of the duplex substrate would retain their Watson–Crick base pairing in this intermediate. Rotation of the DNA bases to pair the ssDNA with its complement to make a hybrid duplex would subsequently place the displaced DNA strand into the minor groove of the hybrid. In the second path, the duplex DNA would enter the filament and interact with the ssDNA initially through its minor groove. Rotation of the bases would generate a hybrid DNA and leave the displaced DNA strand in the major groove to form a poststrand switch triplex. In a poststrand switch triplex, the hybrid DNA strands would be paired in a Watson–Crick conformation. A variety of models have been proposed for the location of the third strand in either triplex, with the most probable involving secondary hydrogen bonds to groups in the major groove (*271, 274–277*). A current proposal for base triplets in an R-form triplex is presented in Fig. 14.

The first pathway in Fig. 13 is particularly attractive, because it provides

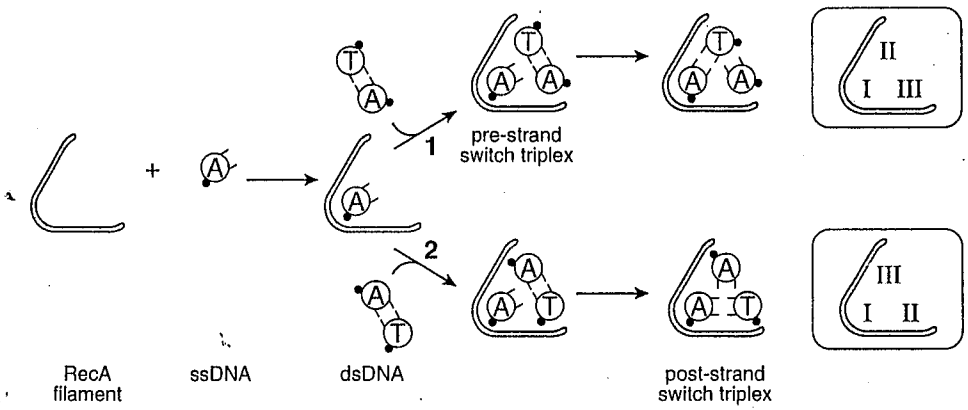


FIG. 13. Possible DNA pairing pathways during DNA strand exchange in the RecA protein filament. Nucleotide bases are represented by circles; the small attached filled circles denote location of the covalent bonds to deoxyribose. Pathway 1 is one in which the major groove of the dsDNA substrate is presented first in the filament groove. Rotation of the bases then results in the displaced strand positioned in the minor groove of the hybrid DNA product. Pathway 2 is a "minor groove first" path. In this case, the displaced DNA strand is positioned in the major groove of the hybrid DNA. DNA strand binding sites would be numbered as shown in the insets for the two pathways. Potential DNA triplex structures in which the extra DNA strand is arrayed in the major groove of the Watson–Crick duplex is indicated.

a clear structural mechanism whereby the two DNA molecules can be homologously aligned. However, there is no evidence for a prestrand switch triplex in RecA reactions. Most studies have detected only the hybrid DNA present after the strand switch has taken place. The RecA filament appears to stabilize the products of DNA strand exchange, using binding energy to promote the strand switch. The existence of a triplex DNA intermediate on the reaction pathway remains controversial. Experiments probing the structure of paired DNA species formed within RecA filaments in the presence of ATPγS suggest a Watson–Crick pairing relationship between the two strands of the hybrid DNA product (*198*). When distal joints are formed with ATPγS, and cross-linked with psoralen derivatives prior to deproteinization, only the two strands corresponding to the product hybrid duplex are cross-linked (*278*), consistent with the result of Adzuma (*198*). These results do not provide information about the location of the displaced DNA strand.

A series of recent chemical probing and cross-linking studies (279–281) has provided strong support for the second pathway in Fig. 13. The duplex DNA appears to enter the groove of the filament and align via its minor groove with the ssDNA. A rotation of the DNA bases produces hybrid DNA. The results of one of these studies suggest that the displaced strand is with-

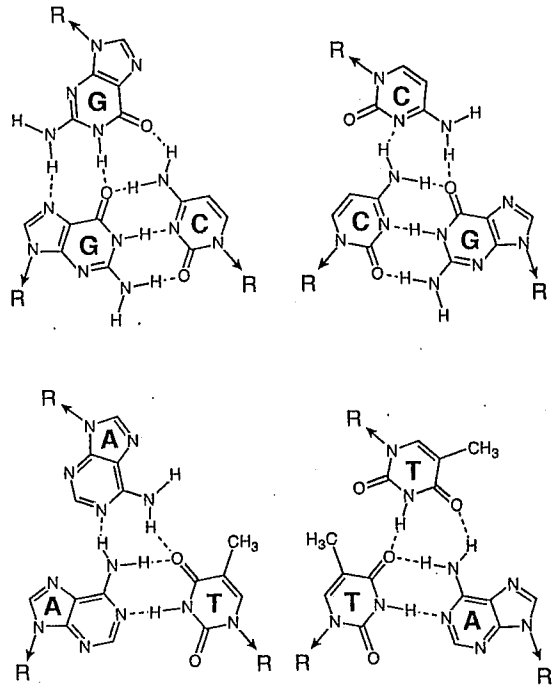


FIG. 14. Proposed DNA triplets for the R-form DNA. Reprinted from Cox (*121*), with permission.

in the RecA filament, but somewhat displaced from the major groove of the hybrid duplex and not hydrogen-bonded to it (*281*). However, some direct evidence that all three strands that can bind to a RecA filament are close enough to interact has been provided by studies using BPDE-labeled oligonucleotides (*199, 200*). An important implication of this work is that there is no homologous recognition between the two DNA molecules prior to the strand switch.

All three strands are cross-linked only in distal joints formed within a RecA filament while ATP was being hydrolyzed (*278, 282*). This structure may represent a triplex DNA species, but the cross-linking patterns are also consistent with a dynamic structure in which the two alternative duplexes are locally interconverted as ATP is hydrolyzed, or a structure in which regions of hybrid DNA alternate with regions that have not undergone a strand switch (*278, 282*). When distal joints were formed with ATP, but cross-linking was carried out after the RecA protein was removed, cross-links were limited to the two strands corresponding to the hybrid duplex (*278*). At a mini-

mum, this indicates that the DNA structure changes when RecA protein is removed.

In distal joints formed within the RecA filament, the third strand appears to be at least transiently displaced when ATP is hydrolyzed. This strand is susceptible to degradation by exonuclease I (*54*). When pairing is restricted to the distal end of the duplex DNA, the addition of exonuclease I results in a reverse strand-exchange reaction, in which the hybrid DNA is stabilized by the progressive enzymatic removal of the displaced stand (*54*).

Much of the evidence for a poststrand switch triplex involves the characterization of a putative stable triplex species that persists after the RecA filament is removed (*276, 277, 283–285*). The stable triplex is detected only when DNA pairing is restricted to the distal end of the duplex DNA substrate. The putative triplexes remaining after the removal of RecA protein cannot be extended as they must be within the RecA filament, and their potential relationship to the species present in the filament can be debated. The triplex structure itself, however, could represent a novel DNA conformation with interesting evolutionary significance. The triplexes have been defined by their thermal stability (distal joints appear to be more stable than proximal joints), and by the resistance of all three strands to various nucleolytic and chemical probes (*277, 283–285*). To function in recombination, an R-form triplex must have like strands arranged in parallel, and the structure must form in a sequence-independent manner. These features set it apart from the stable triplex DNA species characterized to date (*286, 287*).

Early models for R-form DNA placed the third strand in the major groove, stabilized in part by hydrogen bonds to the N-7 guanine (which is not involved in Watson–Crick base pairing) and secondary hydrogen bonds to groups involved in Watson–Crick pairing. However, substituting 7-deazaguanine for guanine in both strands of the duplex DNA substrate had no effect on the rate or efficiency of the DNA strand-exchange reaction (*275*). This was also true when three of four strands were substituted in a four-strand reaction. This result has been extended to DNA substituted with 7-deazaadenine, and DNA with both purine N-7 substitutions (S. K. Jain and M. M. Cox, unpublished results). These results argue against triplex models that rely on hydrogen bonding to purine N-7 for significant stabilization, and have been reinforced by methylation protection experiments done on the structures remaining after RecA deproteinization (*277*). Recent computer modeling predicts that stable triplex structures can be assembled without involving purine N-7 (*271*).

The formation of stable triplex DNA species has proved difficult to confirm by electron microscopy (EM). The triplexes have a reported thermal stability at least as great as duplex DNA (*277, 283, 284*), and the methods used for spreading DNA do not denature duplexes. Nevertheless, no triplexes have

been observed in many EM trials, using conditions inspired by the various published reports (*278;* R. B. Inman and M. M. Cox, unpublished results). Any triplexes present in these reactions are disrupted by surface tension or some other factor associated with DNA spreading.

There has been only one report of results suggesting the spontaneous formation of an R-form DNA triplex (*288*). However, the DNA species studied have a T_m of only about 30°C. Many of the studies of stable DNA species formed after removal of RecA protein have been carried out at higher temperatures. This further suggests that any triplexes left behind after RecA removal exhibit some hysteresis in their formation.

Within the "minor groove first" pathway (Fig. 13), the formation of a DNA triplex, even transiently, remains attractive as a structural device to facilitate homologous recognition between the two DNA substrates. The one viable alternative is a pathway in which the RecA filament catalyzes the conversion of the duplex substrate to a duplex hybrid DNA without a triplex intermediate of any kind. In this last scenario, homologous recognition would be mediated by Watson–Crick base-pairing alone, with the homology search involving iterative rotation of bases in the duplex substrate followed by dissociation until homologous alignment was established. This part of the RecA-mediated DNA strand-exchange mechanism is still in need of further definition.

3. Energetics

Whatever the detailed pathway for the pairing reaction, it is clear that it results in an efficient formation of hybrid DNA and does not require ATP hydrolysis. The strand switch and stabilization of the hybrid DNA product are facilitated by binding energy within the major RecA filament groove. This has been demonstrated in several studies. First, a limited DNA strand-exchange reaction (producing 1–2 kbp of hybrid DNA) can occur under some conditions in the presence of ATPγS (*251, 255, 289–291*). A similar DNA strand-exchange reaction is promoted by the RecA K72R mutant protein, with an alteration in the consensus ATP binding site that allows it to bind but not hydrolyze ATP (*114, 115*). It has been further demonstrated that it is merely the filament conformation conferred by ATP that is required for DNA strand exchange, because an NTP cofactor is not needed if this conformation can be established (*292*). These results are in line with the fact that DNA strand exchange between homologous DNA molecules is not an energetically demanding reaction. It occurs under some conditions in the complete absence of proteins (*293*), and is promoted by a variety of eukaryotic proteins that do not hydrolyze ATP and (in many cases) have no role in recombination (*5, 294*).

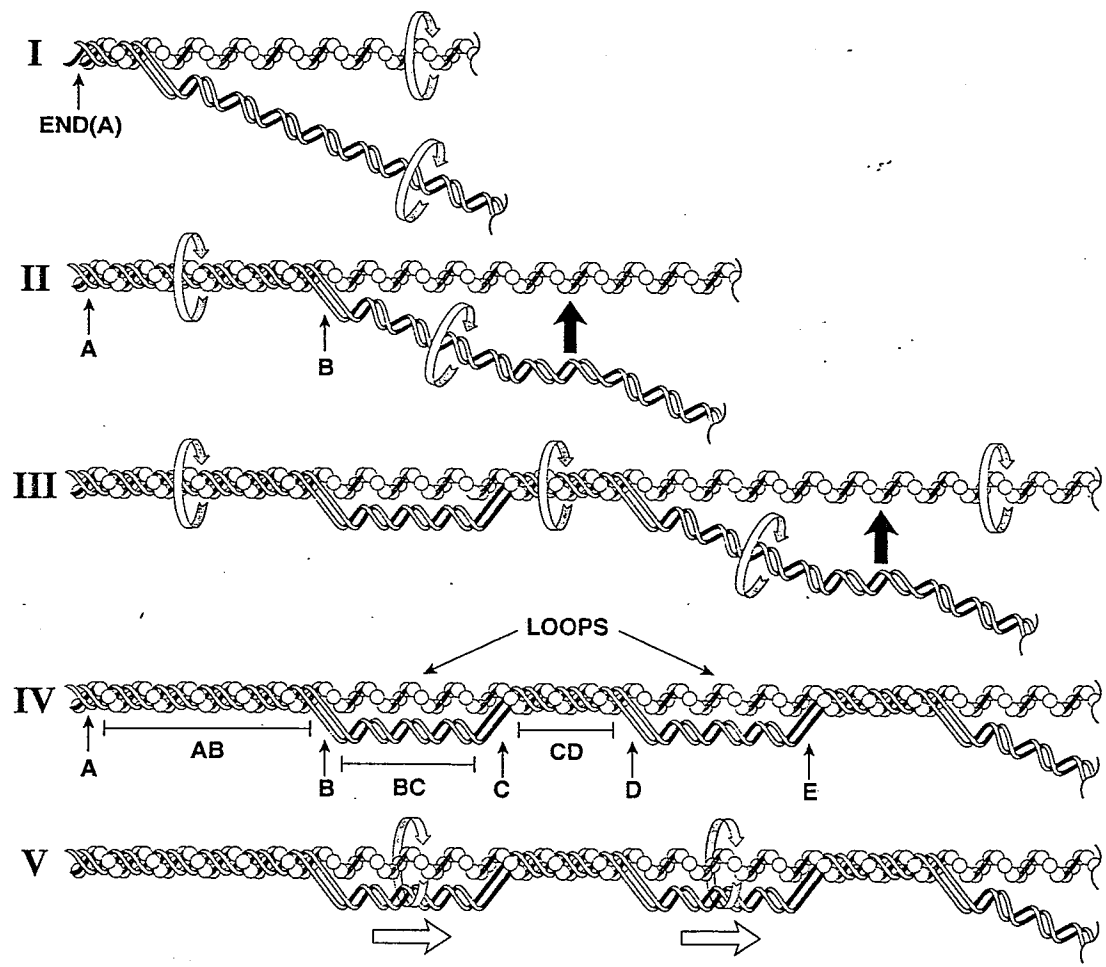
The reaction observed in the absence of ATP hydrolysis is generally lim-

ited to formation of 1–2 kbp of hybrid DNA (*289, 291*). The reason for this limitation is not clear. It has been proposed that the limitation reflects discontinuities in the RecA filaments (*114, 298, 292*). Addition of excess RecA protein does not overcome the limitation (*115*), but small discontinuities might still be present. An alternative proposal is presented in Fig. 15. The discontinuities could be not in the filaments but in the DNA pairing intermediate.

Efficient DNA pairing in the absence of ATP hydrolysis requires the addition of Mg^{2+} in significant excess to that required to saturate the available NTP cofactor as Mg-NTP (*115*). Initiation of pairing at one point in the filament would lead to propagation as the filament binding energy drew in additional DNA. The propagation would be accompanied by rotation of the filament and the duplex DNA, as a by-product of the binding energy-facilitated incorporation of more DNA. As propagation continued, DNA pairing at other locations within the filament would be an intramolecular and potentially favorable process. Pairing at some distance from the point where the initial pairing interaction was being propagated would create a topologically trapped external loop of substrate duplex DNA (Fig. 15). Multiple loops of this type might be created. Once RecA protein was removed, the stable hybrid DNA would be defined by the DNA between the end of the substrate duplex and the first external loop. Because moving these loops would require disruption of paired regions to extend others, and would have to involve rotation about the RecA filament, such loops might well be immobile without an active mechanism to produce the required rotation or disassembly of RecA filaments in these locations.

The effects of Mg^{2+} on the DNA strand-exchange reaction in the absence of ATP hydrolysis may be instructive. When Mg^{2+} concentrations are similar to the NTP concentration so that the concentration of free Mg^{2+} is minimal, the formation of paired DNA is reduced considerably. The limited pairing that does occur, however, can slowly generate completely exchanged hybrid DNA products over 7 kbp in length at significant yields (*115*). If the probability of pairing is reduced, the probability of the secondary pairing events needed to form the loops must also be reduced. Hence, the pairing that does occur generates longer hybrid DNAs. The capacity of the filaments to generate the longer hybrid DNAs under these conditions also tends to argue against the idea that limitations to the length of the hybrid DNA at higher Mg^{2+} concentrations are brought about by filament discontinuities. The effects of high Mg^{2+} concentrations in blocking DNA strand exchange are not relieved if RecA K72R filaments are formed at low Mg^{2+} concentration, followed by raising the latter (*115*).

The DNA strand-exchange reaction through phase 3 is probably sufficient to explain the role of RecA protein in genetic recombination during con-


I
END(A)
II
A
B
III
LOOPS
IV
A
AB
B
BC
C
CD
D
E
V

jugation or transduction. However, the limitations of the pairing reactions described above are not only manifested in the length of the hybrid DNA generated. Without ATP hydrolysis, DNA strand exchange also is bidirectional (*223, 291*), will not bypass structural barriers in the DNA (*115, 254, 255*), and will not accommodate four DNA strands (*115, 251*). A unidirectional DNA strand exchange capable of bypassing structural barriers in the DNA could be useful in the context of recombinational DNA repair, and herein lies a function for ATP hydrolysis.

C. Unidirectional Extension of the Hybrid DNA and the Role of ATP Hydrolysis

NTPases can generally be classified according to one of three biological functions: motor proteins, molecular timing devices, or recycling functions (*295*). The ATPase activity of RecA is most commonly portrayed as a recycling function (*5, 85, 257, 295*), causing the dissociation of RecA monomers from the filament after DNA strand exchange has occurred. As already noted, there is evidence that the end-dependent dissociation of RecA monomers from RecA filaments is at least one of the functions of RecA-mediated ATP hydrolysis (*226*).

However, we have also noted that disassembly of RecA filaments rarely accounts for more than a minute fraction of the ATP hydrolyzed in RecA filaments. ATP is hydrolyzed uniformly by RecA monomers throughout the filament (*176, 212, 239*), and under some conditions ATP or dATP hydrolysis proceeds in the absence of measurable RecA dissociation (*228, 229*).

Defining the role of the RecA ATPase can be divided into two parts: biochemical function and mechanism. The observation that DNA strand exchange proceeds in the absence of ATP hydrolysis provides an opportunity to further define function. In effect, the limitations of the reaction in the absence of ATP hydrolysis elucidate biochemical functions provided by the ATPase activity. When ATP is hydrolyzed, the DNA strand-exchange reac-

---

FIG. 15. A model to explain the cessation of DNA strand exchange in the absence of ATP hydrolysis after formation of 1–2 kbp of hybrid DNA. Formation of a discontinuous DNA pairing intermediate is illustrated in five steps. (I) Pairing is initiated at one end of a duplex DNA substrate. Extension of the paired region requires the rotation of both the filament and the duplex DNA, as shown by circular arrows. (II) As the paired region lengthens, some probability exists for an intramolecular pairing interaction elsewhere in the filament (black arrow). (III) Pairing at the new location creates a new point for continued spooling of the duplex DNA into the filament. However, a segment of DNA is left outside of the filament as an external loop by the second pairing initiation. (IV). Multiple loops can form (e.g., segments B–C and D–E), with paired segments (e.g., C–D) between them. (V) Resolution of the loops requires their rotation about the axis of the RecA nucleoprotein filament. Reprinted from Shan *et al.* (*115*), with permission.

tion becomes unidirectional (*223, 291*), will bypass substantial structural barriers in the DNA substrates (*115, 254, 255*), and will accommodate four DNA strands (*115, 251*). All of these become properties defining a fourth phase of DNA strand exchange, a phase involving the active and ATP-dependent conversion of the DNA pairing intermediates generated in phase 3 to fully exchanged DNA products.

The molecular mechanism by which the hydrolysis of ATP confers these properties on the strand-exchange reaction is more difficult to resolve. At least three ideas have been proposed for how ATP hydrolysis might effect DNA strand exchange at the molecular level. Konforti *et al.* (*223*) proposed that ATP hydrolysis caused dissociation of RecA from the hybrid DNA, with the overall reaction rendered unidirectional by unidirectional assembly of a new filament on the displaced strand. This mechanism provides no explanation for the RecA-mediated bypass of short heterologous insertions during strand exchange. In addition, in reactions carried out in the presence of SSB, the RecA protein involved can be quantitatively accounted for in a complex with the hybrid dsDNA long after DNA strand exchange is complete (*296*).

Other results have shown that the SSB is bound to the displaced DNA strand (*297*). No net disassembly of the RecA filament is required during unidirectional DNA strand exchange, nor is assembly of RecA filaments on the displaced strand. The results eliminate the mechanistic proposal that DNA strand exchange is driven by dissociation of RecA monomers from the initial filament followed by a unidirectional reassembly of the filament on the displaced strand. However, they do not rule out a localized redistribution of RecA monomers within the filament.

The two remaining models are both consistent with existing data. The first is derived from the proposal that ATP hydrolysis is coupled to dissociation of RecA monomers from filaments at the point where strand exchange is taking place (*256*). In its original form, this proposal did not provide an explanation for RecA-mediated bypass of DNA structural barriers during strand exchange. This idea has been updated with the suggestion (*289*) that discontinuities in the RecA filament limit the progress of ATP-independent DNA strand exchange. ATP hydrolysis would then be required simply to redistribute the RecA protein (*114, 289, 292*). Continued extension of the hybrid DNA would depend on available contiguous RecA filament, remaining largely ATP independent.

The second model (*298*) is derived from the fact that if DNA strand exchange is to take place, the two DNA substrates must rotate (Fig. 16). The most straightforward way to facilitate strand exchange is to couple ATP hydrolysis directly to this rotation. This could involve external loops of DNA with the RecA protein facilitating the rotation shown in panel V of Figure 15. If one of the two DNA substrates binds to the outside of the filament, even

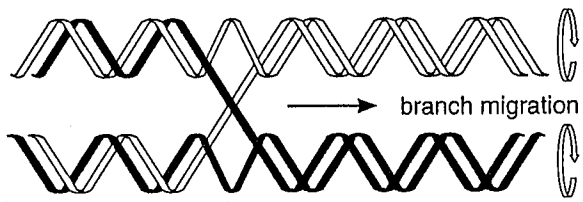


FIG. 16. A DNA strand exchange between two duplex DNA molecules. To move the DNA branch, the molecules must be rotated.

over a short region, the rotation can be effected by passing this external segment between monomers in a ratchetlike action coupled to ATP hydrolysis (*298*). There are six RecA monomers per turn in the filament. Each 360° rotation would occur in six steps, with one-sixth of the RecA monomers in the filament participating in each step. Because DNA within a RecA filament is underwound to 18 bp per turn, each complete rotation would move the DNA branch by 18 bp. A unidirectional rotation would necessarily result in unidirectional strand exchange. A rotation of this kind coupled to ATP hydrolysis would resolve any external loops of DNA such as those depicted in Fig. 15.

This model has been called the facilitated DNA rotation model (*298*). It requires that ATP hydrolysis in the filament be organized into waves that would be separated by perhaps six monomers and travel unidirectionally through the filament, much as was suggested in Section III,C, describing the ATP hydrolytic activity of RecA protein. The model has the problem that no DNA binding sites have been defined on the filament exterior, although little effort has been directed at this possibility. A different version of a facilitated DNA rotation model has recently been proposed (*298, 299*). In this altered version, the DNA is rotated, but all strands remain within the filament.

Both the RecA redistribution and facilitated DNA rotation models provide an explanation for how heterologous sequences are bypassed during DNA strand exchange in an ATP-dependent manner. Bypass requires that the heterologous DNA sequences be unwound. Radding and colleagues demonstrated that a nick in the heterologous region prevents bypass, indicating that the bypassed DNA is unwound indirectly through the application of torsional stress (*299*). The underwound state of DNA bound by RecA is the key to understanding how redistribution of RecA monomers might facilitate bypass (Fig. 17). RecA must be bound to a homologously paired three-stranded complex on both sides of the heterologous insertion. Disassembly of part of the RecA filament on one side of the insertion would lead to the release of underwound DNA. If free rotation of the remaining filament ends was precluded (as it would be in the circular nucleoprotein filaments commonly used

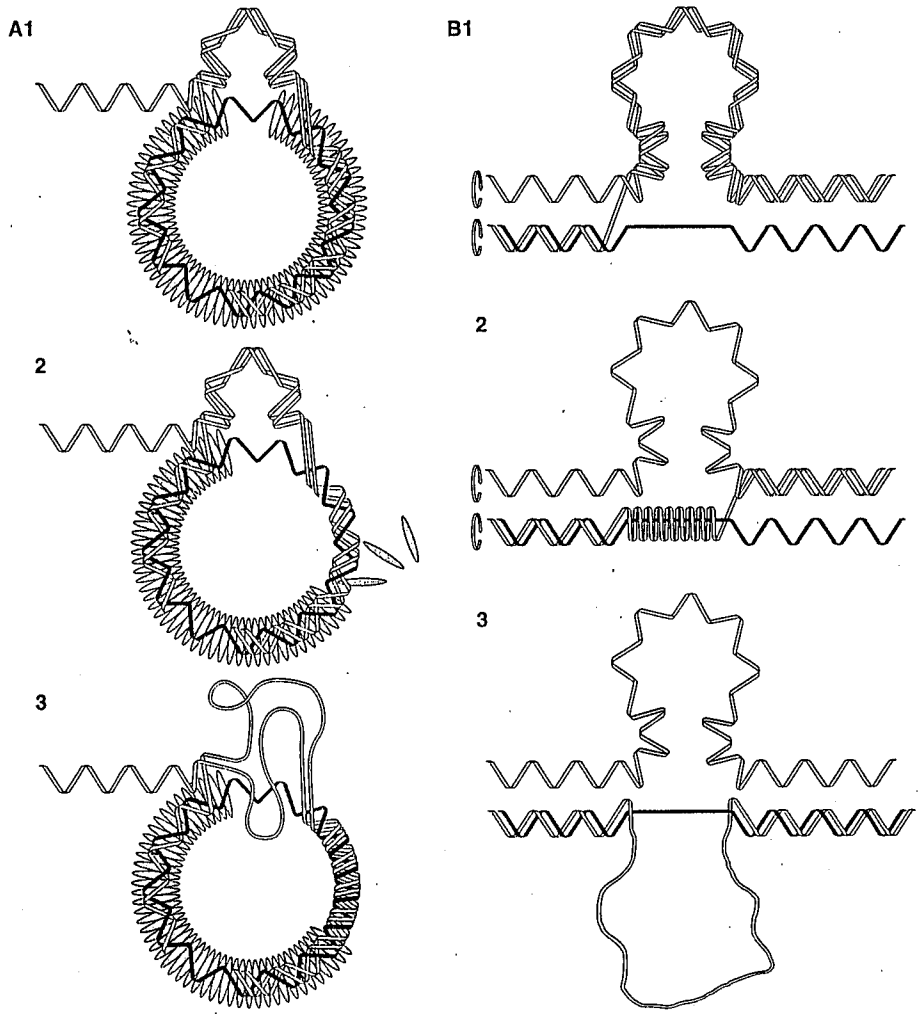


FIG. 17. Mechanisms for the bypass of heterologous DNA insertions in the duplex DNA substrate during DNA strand exchange (A) Bypass by filament disassembly. DNA strand exchange proceeds up to the homology–heterology boundary without the aid of ATP hydrolysis. DNA bound within the RecA filament is extended and underwound by about 40%. Dissociation of some RecA monomers (ovals) results in the release of some underwound DNA. Because the filament is formed on a DNA circle and has no ends free to rotate, the underwinding in the DNA can be translated into unwinding of the DNA in the heterologous insertion. DNA strand exchange can then be continued beyond the insertion without the aid of ATP hydrolysis. This model is based on ideas communicated to us by S. Kowalczykowski. (B) Bypass by facilitated DNA rotation. If RecA protein is bound to the DNA with the filled strand, and promotes the DNA rotation shown by a mechanism described elsewhere (298), the DNA within the heterologous insertion will be unwound as a result of the rotation.

in these experiments), then the underwinding in the released DNA could be translated into strand separation, with as many as 8 bp separated for every six RecA monomers dissociated. A modest degree of filament disassembly in a stalled strand-exchange complex could therefore explain the bypass. In the facilitated DNA rotation model, continued rotation of the exchanging DNAs stalled at the insertion would directly translate into unwinding of the DNA in the insertion, as shown in Fig. 17.

The RecA redistribution model predicts that the bypass of heterologous insertions requires dissociation of part of the adjacent filament, while maintaining the underwound state of the released DNA so that it can be translated into unwinding of the insertion. A linear filament of RecA protein, unconstrained and free to rotate at the ends, should be unable to promote bypass if the model is correct. However, RecA filaments formed on linear ssDNA promote an efficient bypass of heterologous insertions (K. MacFarland and M. M. Cox, unpublished results). In addition, the rate of filament assembly of a RecA filament is generally faster than disassembly at the other end, making it unlikely that a gap in the filament large enough to facilitate the unwinding of 50 to 100-bp heterologous inserts would appear on a circular DNA substrate with any reasonable probability. However, there is an exchange of RecA monomers between free and bound forms that is observed when the RecA is bound to duplex DNA or during DNA strand exchange (*228;* Q. Shan and M. M. Cox, unpublished results). This could signal the kind of RecA redistribution required by the model under strand-exchange conditions, and an evaluation of the role of this exchange process must await a definition of its mechanism.

The facilitated DNA rotation model has the unique virtue of providing a quantitative accounting of the ATP hydrolyzed by RecA protein during DNA strand exchange. There are two relevant calculations. First, to bring about a 360° rotation by the facilitated rotation mechanism (*298*), every RecA monomer in the filament would hydrolyze 1 ATP. In a typical filament containing 2000 RecA monomers, this translates into 2000 ATPs to move the DNA branch by 18 bp, or 100 ATPs per base pair of hybrid DNA created. This is the efficiency observed under optimal conditions *in vitro*. The evident inefficiency can be rationalized as an energetic surplus that permits the bypass of structural barriers during repair. Notably, DNA repair processes tend to be profligate consumers of chemical energy (*26*). Consider, for example, the chemical energy invested in the degradation and replacement of a strand of DNA, 1 kb or more in length, to repair one DNA mismatch (*300*).

Second, if ATP hydrolysis and DNA rotation are coupled, the rate of branch movement should have a predictable relationship to the rate of ATP hydrolysis. A given monomer will hydrolyze ATP every time the external DNA molecule passes, so that each hydrolytic event corresponds to one ro-

tation. During strand exchange, each monomer hydrolyzes about 20 ATPs per minute. The resulting 20 rotations should move the branch 20 × 18 bp, or 360 bp.

This predication has been tested in a detailed study of the temperature dependence of the DNA strand-exchange reaction and the ATP hydrolysis that accompanies it. Arrhenius plots were constructed for strand exchange and RecA-mediated ATP hydrolysis during strand exchange, based on experiments carried out from 25° to 45° (233). The plots were linear and parallel over the entire temperature range, providing another piece of evidence consistent with a direct coupling of ATP hydrolysis to the last phase of DNA strand exchange (233). Perhaps more significantly, the distance between the lines corresponded to a factor of about 18 bp.

The role of ATP hydrolysis in the DNA strand-exchange reaction remains a controversial topic. However, the potential exists for the redefinition of the RecA ATPase activity in terms of a novel motor function in recombinational DNA repair. Resolution of the various issues remaining in defining the function and mechanism of ATP hydrolysis will continue to be a challenge.

D. Exchange Reactions with Four DNA Strands

ATP hydrolysis is required for the promotion of a reciprocal DNA strand exchange involving four DNA strands. This is perhaps the observation most difficult to rationalize in terms of the RecA redistribution proposal. If a RecA filament can bind only three DNA strands within its major groove, and the filament cannot promote a four-strand DNA exchange reaction without ATP hydrolysis, redistribution of the monomers within the filament is unlikely to produce a different result. The facilitated DNA rotation model provides a mechanism for the promotion of a four-strand exchange reaction. Once pairings were initiated in the single-strand gap as a three-strand reaction, as required, homologous DNA in the flanking four-strand regions would effectively be aligned as well. DNA rotation as depicted in Fig. 16 would inevitably spool one strand of the duplex region and the gapped DNA substrate out of the filament ahead of the point where a strand was being brought into the filament from the second duplex substrate. By this mechanism, only two DNA strands would occupy the interior groove of a RecA filament at any point during the four-strand portion of an exchange reaction.

V. Interaction of RecA Protein with Other Proteins

The RecA protein promotes only a few stages of recombinational processes in which over a dozen other proteins may take part. The list below is admittedly selective and focuses on proteins with a demonstrated effect on

RecA activities *in vivo* and/or *in vitro*. The discussion is also limited to proteins involved in the recombinational functions of RecA protein. A particularly complete listing and description of bacterial recombination functions appears in another review (*85*).

A. The Single-Strand DNA Binding Protein

Some effects of the single-strand DNA binding protein (SSB) on RecA protein binding to ssDNA have already been noted. SSB and RecA compete for binding sites on the DNA. SSB inhibits the nucleation of RecA filaments, but not extension. When RecA protein is added before or concurrently with SSB, some RecA protein will bind to ssDNA regions lacking DNA secondary structure. The SSB will melt out regions of secondary structure, then will be displaced from the DNA by RecA filament extension (*210, 211*). A relatively nonspecific binding to ssDNA is the only biochemical activity of SSB.

The *ssb* gene was first detected by virtue of the effects of gene defects on DNA replication (*301*). Several mutant alleles have been studied. In addition to effects on DNA replication, *ssb* mutants also confer a sensitivity to DNA-damaging agents such as UV light, and a reduction in recombination proficiency (*302*). One allele, *ssb-3*, confers a UV sensitivity approaching that observed for *recA* deletions (*303*). The wild-type gene encodes a protein with 177 amino acids with a combined molecular weight of 18,843 (*304*).

The DNA-binding properties of SSB have been reviewed in detail (*305, 306*). SSB generally binds to DNA as a tetramer. At least four distinct binding modes have been detected, with transitions between them occurring as a function of solution conditions. The various SSB species have binding site sizes ranging from 35 to 65 nucleotides of ssDNA. The species predominating at low salt concentrations and relatively high SSB binding densities is SSB_{35}. The smaller site size in this DNA binding mode reflects an interaction of only two of the tetramer subunits with the DNA. This form of SSB is characterized by a high level of cooperativity in DNA binding (*305, 306*), giving rise to a fairly uniform coating of SSB on the DNA (*307*). The DNA contour length is reduced by about 60% when complexed with SSB_{35} (*307*). At very high salt concentrations or low SSB binding densities, the SSB_{65} species predominates. This form is characterized by a more limited cooperativity reflecting interactions between two tetramers to form octamers, but no higher order structures (*308*). In the electron microscope, the octamers give rise to a "beads on a string" appearance to the SSB–DNA complexes (*307*). The contour length of the ssDNA is reduced about 78% when complexed to SSB_{65} (*307*). A prominent intermediate DNA binding mode is SSB_{56}, with many properties similar to SSB_{65}, including a beaded appearance on ssDNA when viewed in the electron microscope. Another species detected in some studies, SSB_{42}, has not been characterized in detail. SSB_{56} is the predominant

species under the conditions used for most *in vitro* experiments with RecA protein (*309*), and is the species responsible for the enhancement of RecA protein filament formation (*307, 310, 311*). Bound RecA protein tends to be displaced by SSB under conditions favoring the formation of SSB_{35} (*311*).

The maintenance of contiguous extended RecA filaments (hydrolyzing ATP) on random sequence ssDNA requires a continuous presence of SSB (*311*). This could be explained by a cycle of partial RecA dissociation followed by SSB-facilitated rebinding. Alternatively, it could reflect a stabilizing interaction of SSB with RecA filaments. The former explanation is favored by the observation that other single-strand DNA-binding proteins that should not interact with RecA protein, including the gene-32 protein of bacteriophage T4 and even the yeast protein yRPA, will facilitate RecA filament formation. In addition, there is no indication that SSB is present in RecA filaments formed in the presence of SSB when visualized by electron microscopy (*307, 312;* R. B. Inman and M. M. Cox, unpublished results).

However, it has already been noted that RecA filaments on ssDNA exhibit little exchange between free and bound RecA monomers in the presence of SSB, as might be expected in the dissociation/rebinding model. A direct and apparently specific interaction between SSB and RecA protein is suggested by changes in the intrinsic fluorescence of SSB (*234*) and by immunoprecipitation studies (*310*). In addition, the SSB that promotes the formation of RecA filaments in one reaction mixture is not immediately available to promote formation of new RecA filaments when combined with a second RecA–ssDNA reaction mixture incubated without SSB (M. M. Cox, unpublished results). The latter experiments suggest that SSB is sequestered in a weak interaction with RecA filaments.

We postulate that a relatively weak interaction between SSB and RecA filaments exists that is subject to disruption with the methods used to spread RecA protein for electron microscopy. The functional significance of such an interaction is not clear. It must be inconsequential for the formation of RecA filaments *in vitro,* because other single-strand binding proteins can substitute for SSB in this function. Many studies indicate that the elimination of secondary structure in the ssDNA is the unique function of SSB in presynapsis. Once RecA filaments were formed, an SSB interaction with them might have a modest effect in preventing the dissociation of RecA monomers in the interior of RecA filaments. Indeed, these filaments undergo a gradual partial disassembly when SSB is removed (*311*). However, the most likely role of an SSB–RecA interaction would come in the later stages of DNA strand-exchange reactions. SSB binds to the displaced single strand as it is generated and tends to prevent secondary DNA pairing reactions involving it (297). *In vitro,* an interaction of SSB with the RecA filament might not provide an

enhancement of SSB binding to the displaced DNA strand. However, an advantage might be realized in the more complex *in vivo* environment if the same RecA–SSB interaction helped sequester SSB and position it for binding to the displaced DNA strand. Over a billion years of coevolution has provided ample opportunity for the development of both weak and strong interactions between proteins involved in recombinational processes, to better organize events in recombinational DNA repair.

B. The RecF, RecO, and RecR Proteins

RecA protein may never act by itself as a uniform filament of RecA monomers in the cell. Evidence is accumulating that a number of proteins not only affect the assembly and disassembly of RecA filaments, they also remain intimately associated with the filament and may affect its function. Primary candidates at present are the RecF, RecO, and RecR proteins. The RecO and RecR proteins have already been shown to interact with RecA filaments in a way that facilitates assembly and blocks disassembly.

The *recF* gene was first detected in 1973 (*152*) as a UV-sensitive, recombination-deficient mutant in a *recBC sbcBC* background. The gene is contained in an operon that also includes the *dnaN* gene, an arrangement that seems to suggest some link with replication. It is speculated that RecF is part of a system that diverts ssDNA from being a template for replication to being a substrate for recombination (*27*). The RecO and RecR proteins described below are included in this hypothesis. The sequenced *recF* gene encodes a 357-amino acid polypeptide (40.5 kDa). The RecF protein has been purified by at least four research groups and characterized *in vitro* (*313–318*). The protein contains a consensus nucleotide-binding fold (Walker A box). RecF protein alone has no stimulatory effect on any RecA activity tested to date. At sufficiently high concentrations, it inhibits RecA binding to DNA and RecA protein-mediated DNA strand exchange (*314, 316*). The RecF protein possesses a very weak ATP hydrolytic activity (k_{cat} about 0.3 min^{-1}) (*318*).

In vivo, *recF* mutant bacteria exhibit a delayed activation of the SOS response that might reflect slow formation of the RecA filaments required to facilitate LexA cleavage (*319*). The *E. coli* strains in which SSB is overexpressed exhibit a *recF*-like phenotype (*320*). Mutations in *recF* are suppressed by certain *recA* mutations, such as *recA441* and *recA803* (*319, 321*). The RecA441 (previously *tif*) and RecA803 proteins exhibit an enhanced capacity to displace SSB and bind ssDNA *in vitro* (*322, 323*). This suggests some sort of *in vivo* interaction between RecF protein and RecA protein, although the function of RecF has not yet been elucidated.

The *recO* gene was identified in 1985 (*324*). It is situated in a operon with the *rnc* gene, which encodes ribonuclease III, and the *era* gene, which en-

codes a GTP-binding protein with sequence similarities to the yeast RAS proteins (*325*). The sequenced *recO* gene encodes a protein with 242 amino acids (26,000 Da) (*326, 327*). A predicated nucleotide-binding fold in the RecO protein (*326*) is not conserved among the limited number of other RecO sequences available (A. I. Roca, unpublished results). No RecO-mediated ATP hydrolysis has been detected.

The purified RecO protein has been characterized in several studies (*227, 316, 317, 328, 329*). It binds to both ssDNA and dsDNA and behaves as a monomer in solution. It promotes an ATP-independent renaturation of complementary DNA strands (*328*). A weak assimilation of ssDNA fragments into homologous supercoiled duplexes to form D-loops is catalyzed by RecO protein (*329*).

The *recR* gene was identified in 1989 (*330, 331*). It is located on the *E. coli* chromosome near the *dnaX* gene and shares an operon with a small open reading frame of unknown function called *orf-12*. The sequenced gene encodes a 201-amino acid protein with a predicted molecular mass of 21,965 Da. A RecR sequence alignment with the two other available RecR sequences (A. I. Roca, unpublished results) does not show a canonical Walker A box, contrary to the reported existence of a nucleotide-binding fold (*332*). There is a highly conserved Cys cluster ($CxxCx_8CxxC$), which may chelate zinc as in other proteins (*332a*). Zinc fingers are required for DNA damage recognition by UvrA (*333*). The RecR protein migrates in SDS–PAGE as a 26-kDa protein.

The purified protein has been studied in concert with the RecF and RecO proteins (*227, 316–318*). The RecR protein from *Bacillus subtilis* has also been purified (*332*). The *B. subtilis* RecR protein binds to dsDNA, and the binding is stimulated by DNA damage, ATP, and divalent cations. The *B. subtilis* and *E. coli* genes exhibit a 43% sequence identity.

Several lines of evidence indicate that the RecF, RecO, and RecR proteins function at the same stage of recombination, possibly as part of a single complex. The phenotypes of mutations in the three genes are very similar, defining them as an epistatic group (*27, 334*). Mutations in all three genes are suppressed by *recA441* and *recA803* mutations (*335*). In addition, a gene in bacteriophage λ, called *orf*, which can replace *recF, recO,* and *recR* functions in lambda recombination has been identified (*336*).

An interaction has been detected between the RecF and RecR proteins *in vitro*. In the presence of ATP, RecF protein binds to dsDNA very weakly, and RecR protein does not bind to dsDNA by itself at all. When both proteins are added to a reaction together, RecF protein binds to dsDNA much more readily and appears to coat the DNA if enough protein is present (*318*). RecR protein is present in these complexes. The ATPase activity of RecF pro-

tein is also stimulated two- to threefold by RecR protein under some conditions (*318*). The functional significance of these reactions has not yet been worked out. Alone, RecF binds preferentially to gapped DNA (*336a*).

The RecF, RecO, and RecR proteins have been examined together *in vitro* in two studies (*316, 317*). The RecO and RecR proteins, either with or without RecF protein, stimulated RecA protein binding to ssDNA coated with SSB. None of the proteins was active on its own. Although RecF protein was not required for the stimulatory effect, the presence of RecO and RecR seemed to nullify the inhibitory effect seen when RecF was used alone. RecO and RecR proteins remain bound to the DNA after formation of the RecA filament. The resulting filament has only 70% of the RecA of a normal filament, but it can still promote DNA strand exchange (*317*). Significant amounts of RecO, RecR, and SSB remained associated with RecA nucleoprotein filaments formed on ssDNA (*317*).

The effects of RecO and RecR proteins do not end once the RecA filament is assembled. As already noted, the RecO and RecR proteins completely block the end-dependent disassembly of RecA filament from ssDNA (227). The proteins may remain bound as part of a joint filament with the RecA protein. Both proteins are required in amounts equivalent to about 1 monomer of each for every 50 monomers of RecA protein (227). This may reflect a kind of regulatory function for RecO and RecR. At a stalled replication fork where recombinational DNA repair is required, the ssDNA will generally be bound with SSB.

The RecO and RecR proteins greatly facilitate the binding of RecA to an SSB-bound DNA substrate, and they prevent disassembly once bound. This may give the filament a sufficient lifetime to carry out an efficient DNA strand-exchange reaction to effect repair (see Fig. 1). The filament must eventually be disassembled and the replication fork restarted. Recent genetic studies indicating overlapping functions for the PriA protein (involved in the assembly of the replication primosome) and RecF protein are consistent with a role for RecF in the reassembly of an active replication fork (*337*). Additional work on *priA* has served to highlight the important role of DNA replication in many recombinational processes (*338*).

A DNA repair perspective may be especially useful in constructing hypotheses for the function of the RecF, RecO, and RecR proteins that can be addressed experimentally. Mutations of the *recF, recO,* and *recR* genes exhibit a substantial defect in recombination only when they appear in *recBC sbc* genetic backgrounds (*27, 85, 217*). Single mutations in some of these genes do affect plasmid recombination, but nearly normal levels of conjugational DNA recombination can be achieved without them. However, cells that contain mutations in any of these genes in an otherwise wild-type background are sensitive to DNA-damaging agents such as ionizing radiation,

mitomycin C, or UV light (*152, 331, 339–343*). All of these gene products appear to have a fundamental role in DNA repair.

C. The RuvA and RuvB Proteins

Some mechanism must exist for the disassembly of RecA filaments *in vivo*, especially if they are stabilized by the RecO and RecR proteins. RecA filaments may be displaced from the DNA by the combined action of two other proteins, the RuvA and RuvB proteins.

The first *ruv* mutants were isolated in 1974 (*344*). Two separate genes, *ruvA* and *ruvB*, were subsequently found at the *ruv* locus in a single, LexA-regulated operon (*345*). The *ruvC* gene is nearby. The *ruvA* gene encodes a 203-amino acid polypeptide with a molecular mass of 22 kDa. The protein migrates in SDS–PAGE at the position expected for a protein of 27 kDa. The *ruvB* gene encodes a 306-amino acid polypeptide of 37 kDa; the sequence includes a consensus nucleotide-binding fold. As with the RuvA protein, RuvB migrates in SDS–PAGE at a position expected for a slightly larger protein (41 kDa). The fact that these proteins are induced during SOS provides another suggestive link with DNA repair. The RuvA protein structure has been determined (*345a*).

The RuvA and RuvB proteins were first purified by the Shinagawa and West groups (*346, 347*), and some of their *in vitro* activities have been described. The proteins act together to process branched DNA strand-exchange intermediates created by the RecA protein.

RuvA protein binds to DNA, binding most tightly to Holliday intermediates and cruciform structures (*348*). One proposed role is that of a molecular matchmaker (*348, 349*), targeting the RuvB protein to the DNA. RuvB protein exhibits a DNA-dependent ATPase activity and catalyzes the branch migration of Holliday intermediates (*346, 348, 350, 351*). Both activities are stimulated by RuvA protein. The requirement for RuvB protein decreases when RuvA protein is present, and the optimal conditions for ATP hydrolysis or branch migration change. Reported rates of ATP hydrolysis are greater when RuvB is bound to circular dsDNA, suggesting that it is coupled to a processive movement along the DNA (*352, 353*). RuvB protein binds weakly to ssDNA or dsDNA, but the binding can readily be measured in the presence of high (>10 m*M*) levels of Mg^{2+}.

RuvB protein is also a helicase (*351, 354*), although detectable helicase activity is observed only in the presence of RuvA protein. RuvB protein moves 5′ to 3′ along one DNA strand that is partially single stranded, displacing a short complementary strand. Complementary strands nearly 600 nucleotides long can be displaced, although the efficiency falls off more or less linearly as the length of the strand increases from 100 to 600 nucleotides (*354*). No unwinding of fully duplex DNA has been observed, even for duplexes as short

as 50 bp. The helicase is thought to have an unspecific mechanistic role in the branch migration activity. The helicase activity is enhanced at DNA branch points (*351*).

Under some conditions, the RuvB protein binds to dsDNA as a dodecamer (440 kDa), forming back-to-back hexameric rings with D_6 symmetry (*355*). The DNA takes a path through the center of the fused rings, much like the situation with the β subunit of DNA polymerase III (*356*). The openings at either end of this donutlike structure have a diameter of about 20 Å. More recent results indicate that the active form of RuvB protein is hexamer (*352*). The functional structure may be similar to that of some other helicases, such as the SV40 T-antigen (*357, 358*), the rho transcriptional terminator (*359–364*), and the DnaB protein (*356*). An analysis of the RuvB ATPase activity indicates the presence of nonequivalent ATP hydrolytic sites within a RuvB hexamer (*353*). ATP hydrolysis in individual subunits is likely to be coupled to conformation changes that modulate the interaction of each subunit with DNA so as to effect movement along the DNA.

Overall, the RuvA protein is thought to bind to branched recombination intermediates and facilitate the binding of RuvB. The RuvB protein (or a RuvAB complex) then promotes a branch migration reaction that has been advanced as the major RuvAB function *in vivo*. EM observations and other work have led West and colleagues to propose a model in which a RuvA protein is bound to a Holliday junction, sandwiched between two RuvB hexamers (*366, 367*). In the model, two branches of the Holliday structure are propelled outward through the hexameric RuvB rings to promote branch migration (*366, 367*). These proteins may also modulate in some way the binding and cleavage of Holliday intermediates by RuvC protein (*348, 368*).

When added to RecA protein-mediated DNA strand-exchange reactions, the RuvA and RuvB proteins greatly enhance the bypass of heterologous insertions in the duplex DNA substrate. The combined action of these proteins permits significant bypass of inserts that are 1 kbp in length (*369*). When barriers that cannot be bypassed are encountered, the RuvA and RuvB proteins promote an efficient reversal of the DNA strand-exchange reaction carried out by RecA protein (*370*). Based on *in vitro* results, RuvA and RuvB could act as antirecombinase activities under some conditions, reversing potentially deleterious intrachromosomal recombinational events (*370*). Although the existence of a protective antirecombination pathway has been postulated (*371*), the idea that RuvA and RuvB might be involved has not been tested *in vivo*.

It has been directly demonstrated that the RuvA and RuvB proteins will displace RecA filaments from dsDNA under some conditions (*372*). This may be a general pathway for RecA filament disassembly in the cell, but many questions remain about where RecA protein leaves off and RuvA and RuvB take over in a recombinational process. The issue is complicated by the pres-

ence of another protein in *E. coli*, RecG protein, with functions that at least partially overlap those of RuvA and RuvB (*368, 373, 374*).

D. Other Proteins Affecting RecA Activities

A number of other proteins can affect RecA protein-mediated DNA strand-exchange reactions. Kowalczykowski and colleagues have reconstituted a DNA pairing reaction that is dependent on the concerted action of the RecBCD enzyme, SSB, and RecA proteins (*213, 214, 375, 376*). The reaction is also dependent on the recombinational hotspot sequence, *chi*. The system provides an important model system for the presumed early steps in the double-strand break recombinational repair pathway.

RecA function *in vitro* is also affected by some nucleases. RecJ is a 63-kDa protein with a 5′ to 3′ exonuclease activity (*377, 378*). It is involved in the RecF and RecE pathways for genetic recombination (*379–381*). Under some conditions, the RecJ exonuclease enhances the RecA protein-mediated strand exchange by degrading the displaced DNA strand (*382*). Exonuclease I is a 3′ to 5′ exonuclease specific for ssDNA (*383*). Some mutant alleles of the gene encoding exonuclease I (*sbcB*) suppress the recombination-deficient phenotype of *recBC* mutations (*384*). In concert with the RecA protein, exonuclease I will facilitate a DNA strand-exchange reaction with a polarity opposite to that normally promoted by RecA protein (*54*). As with RecJ, the reaction is accompanied by digestion of the displaced DNA strand. Exonuclease I copurifies with RecA protein through a wide range of chromatographic resins, suggesting a rather strong association of the two proteins (*54*). However, there has been no demonstration that the interaction is specific, and no function has been suggested or demonstrated.

VI. Other Functions of RecA Protein *in Vivo*

The role of RecA protein in DNA metabolism is remarkably diverse. In its various capacities, RecA protein exhibits enzymatic activities not described above and interacts with a number of proteins not linked with recombination. These other functions are summarized briefly below.

A. The RecA Coprotease Function

In the early 1960s and early 1970s, a series of observations had suggested the existence of an elaborate pathway for DNA repair and mutagenesis that was induced by DNA damage (*385, 386*). The observations and ideas were unified under the SOS hypothesis in 1974 by Radman (*387*). The first activity RecA protein was associated with *in vitro* was the cleavage of the bacteriophage λ repressor as a biochemical step in the induction of the bacteri-

al SOS response (*388*). Later, it became clear that the LexA repressor was inactivated by RecA-mediated proteolytic cleavage as well (*389*). Repressor cleavage was also the first functional assay used to guide the purification of RecA protein (*388*).

The SOS induction pathway has been worked out in some detail (*390–393*). High levels of DNA damage cause an interruption of replication. The single-stranded DNA created as one result of the interruption is the primary signal for SOS induction (*47*). RecA protein binds to the ssDNA, perhaps aided by the RecF, RecO, and RecR proteins. The bound RecA protein assumes the extended filament conformation in the presence of ATP, and this filament form facilitates an autocatalytic cleavage of the LexA repressor, the bacteriophage λ repressor, and a few other proteins. The result is a general induction of over 20 genes repressed by LexA and induction of any bacteriophage λ lysogens present. The induced gene products include a variety of enzymes with functions in DNA repair, replication, recombination, and a specialized pathway for mutagenesis. RecA protein is induced as an SOS function.

Facilitating the autocatalytic cleavage of LexA protein involves the binding of LexA deep within the groove of a RecA filament (*394*). The binding site generally spans two adjacent RecA monomers (*394*). The form of RecA protein bound on ssDNA has often been referred to as an activated species, or RecA* (*390–393*). It is likely that this RecA species is identical to the extended RecA filament involved in recombinational processes (*394*). The use of ssDNA as the molecular signal for SOS induction may be a simple manifestation of the fundamental DNA binding properties of RecA protein. The propensity of RecA to bind ssDNA almost exclusively under physiological conditions serves to target RecA protein to ssDNA gaps created by the interruption of replication. There is evidence that RecA protein mutants that bind more readily to dsDNA exhibit a phenotype in which SOS is constitutively induced (*131, 395*), presumably because RecA protein is constantly and nonselectively bound to DNA as an extended filament. Binding of LexA in the groove of a RecA filament blocks DNA strand exchange by excluding the dsDNA substrate (*395a,b*).

B. SOS Mutagenesis

Heavy DNA damage leads to the induction of a specialized system for mutagenic DNA lesion bypass as a part of the SOS system (*392, 396–398*). The existence of the system can be rationalized as a pathway permitting replication restart in an environment where replication would otherwise be impossible (*398*). The deleterious mutagenic effects of replicative bypass of lesions would be balanced by the survival of a few cells that would otherwise perish. The pathway may also permit a resistance to DNA lesion for which no effective repair pathways exist (*393, 398*). RecA protein has at least three roles

in SOS mutagenesis. First, some of the required activities are induced as part of the SOS response. Second, at least one of the required proteins is proteolytically processed in a reaction facilitated by RecA protein. Finally, RecA protein participates directly in the lesion bypass.

Replicative lesion bypass in SOS mutagenesis requires the combined activities of proteins, including the UmuC, UmuD′, and RecA proteins, along with DNA polymerase III (*398, 399*). UmuD′ is a processed form of the UmuD protein brought about by a RecA facilitated proteolytic cleavage (*400–402*). Several studies have provided evidence that RecA protein has a direct role in this process (*403–405*), and an *in vitro* replication bypass requiring all of these proteins has been demonstrated (*399*). However, the precise molecular role of RecA in this process remains obscure.

C. Chromosome Partitioning

The abnormal chromosome complement in cells lacking *recA* function has led to a proposal that RecA protein has a role in the proper partitioning of chromosomes at cell division (*406*). Part of the role of RecA in partitioning may be indirect, mediated by the protection of chromosomes from the nuclease digestion that occurs in the absence of RecA (*407*). RecA could also be required to produce in concatemers of daughter chromosomes the tension needed for partitioning (*406*), mirroring the need for genetic recombination for the proper partitioning of chromosomes in cell division in eukaryotes.

D. Induced Stable DNA Replication

DNA damage brings about elevated levels of DNA replication that are not dependent on initiation at the bacterial replication origin, *oriC*. This phenomenon is called induced stable DNA replication, or iSDR (*338, 408–410*). The replication is closely linked to recombination functions and recombinational DNA repair (*338, 408–410*). To some extent, it may reflect replication that must accompany many of the kinds of recombinational DNA repair processes outlined in Fig. 1. It has been suggested that single-strand 3′ ends involved in an early stage of recombinational processes might, after being paired with a new complementary strand, act as primers for this DNA synthesis (*411*).

VII. Epilogue: Relating RecA Biochemistry to DNA Repair

Many aspects of RecA enzymology remain controversial. The DNA pairing pathway is still in need of molecular definition, and the role of ATP hydrolysis continues to be debated. Another key challenge for the future is the

elucidation of the collaborative roles of the various recombination proteins. In particular, a better understanding is needed of where one activity leaves off and another takes over.

These issues are interrelated. For example, it can be argued that a hybrid DNA extension reaction coupled to ATP hydrolysis would in any case be of limited physiological significance (*85, 348, 412*). The RuvA and RuvB proteins promote the extension of hybrid DNA (a DNA branch migration) significantly faster than RecA protein, and they promote the bypass of long heterologous insertions that cannot be bypassed by RecA alone (*369*). There seems little need for RecA in this capacity.

At a minimum, it is clear that the ATPase activity of RecA protein is biologically important, in spite of the existence of a range of proteins that promote DNA strand exchange without ATP (*5*). The ATP binding site in RecA protein is the most conserved part of the protein structure. Also, the RecA K72R mutant protein, which binds but does not hydrolyze ATP and promotes a significant strand-exchange reaction *in vitro*, produces a *recA* null phenotype *in vivo* (*113;* R. Devoret, personal communication).

We have postulated that RecA-mediated ATP hydrolysis provides a molecular motor that augments an intrinsic ATP-independent DNA strand-exchange process, rendering it unidirectional and capable of barrier bypass in the context of efficient recombinational DNA repair. The prediction arises that it should be possible to find *recA* mutants that confer a repair-deficient, recombination-proficient phenotype. Such mutants would have an intact strand-exchange capability within the filament, but the motor required to direct this process past DNA lesions during recombinational DNA repair would be compromised. Searching for such mutants is complicated by the fact that RecA protein is required to induce the SOS response to DNA damage. A repair deficiency can therefore result from a defect in the coprotease activity rather than a defect in recombinational repair.

The mutant in question should therefore confer a recombination-proficient, repair-defective, and coprotease-constitutive phenotype (or the mutant should retain the repair deficiency when the SOS response is induced by other genetic devices). One mutant of this kind has been found and characterized (*413*). It is designated RecA423, or R169H (in α-helix F). Cells containing *recA423* are nearly as proficient in conjugational recombination, transductional recombination, and recombination of λ *red*$^-$ gam$^-$ phage as are wild-type cells. At the same time, the mutant cells are deficient in intrachromosomal recombination and nearly as sensitive to UV irradiation as are *recA* deletion strains.

RecA423 is coprotease constitutive, so the repair defect does not reflect a lack of SOS repair functions. *In vitro*, the mutant protein exhibits a reduced capacity to bind DNA, and promotes DNA strand exchange and ATP hy-

drolysis at much reduced rates. The strand exchange is blocked by short heterologous insertions in one of the DNA substrates (*413*). The results demonstrate that a RecA mutant protein with a significantly compromised DNA strand-exchange capacity can suffice completely in conjugational and transductional recombination. However, the structural and energetic requirements for recombinational DNA repair are clearly more stringent, because the mutant cells are almost completely devoid of recombinational DNA repair function. The work provides experimental support for the idea that many of the biochemical features of the RecA protein evolved to address the requirements of DNA repair. It also argues that RuvA and RuvB do not compensate for a reduced RecA strand-exchange function in DNA repair.

The models for recombinational DNA repair presented in Fig. 1 are quite similar to current models drawn for homologous recombination carried out to generate genetic diversity. A focus on the mechanisms of conjugational and transductional recombination can then be viewed as a means of elucidating the mechanisms of homologous genetic recombination in all contexts. However, the biological context within which research is carried out can greatly affect the outcome of the research, and a focus on recombination as a means to generate genetic diversity has not always been fruitful. A decades-long and fruitless search for four-stranded DNA pairing intermediates in recombination was undertaken. Work on RecBCD enzyme and RecA protein proceeded at a fast pace in the late 1970s and 1980s, while many other interesting bacterial recombination functions were left in relative obscurity over this same period, perhaps due in part to their perceived involvement in secondary recombination pathways. The ATPase activity of RecA protein is still evaluated in the context of the minimal energetic requirements for homologous DNA strand exchange evident in the conjugational recombination pathway.

In thinking about the function of proteins involved in recombination, and their mechanisms of action, a recombinational DNA repair paradigm is likely to have a predictive value superior to that provided by a paradigm centered on conjugational and transductional recombination. An understanding of protein functions in recombinational DNA repair should facilitate a more rapid advance in the understanding of all recombination processes.

Acknowledgments

The authors thank Randy Bryant, John Clark, Howard Gamper, Richard Michod, and John Roth for helpful discussions; and Randy Bryant, Robert Schleif, William Reznikoff, Kenji Adzuma, Mikael Kubista, Howard Gamper, Masayuki Takahashi, and Brian Cali for reading and commenting on parts of the manuscript. We appreciate the help of John Clark, Steven Sandler,

Stephen Kowalczykowski, Howard Gamper, Randy Bryant, John Roth, Dan Camerini-Otero, and Stephen West in communicating results prior to publication. The model in Fig. 17A was based on ideas communicated to MMC by Stephen Kowalczykowski during a meeting in Seillac, France, during the summer of 1994.

We also thank the following people for communicating unpublished sequences: John Coleman, Peter Emmerson, Sumiko and Masayori Inouye, Skorn Mongkolsuk, Wai Mun Huang, Joanne L. Johnston, Joan Sloan, and Julian Rood. We thank A. John Clark and Steve Sandler for communication of the Archaea RadA sequence prior to publication. Irv Edelman and Mike Gribskov helped with technical aspects of the sequence analysis; Mike Hogan and Don Thomson provided computer programming assistance; and David Goodsell, Jean-Yves Sgro, Adam Steinberg, and Gary Wesenberg helped with computer graphics. Finally, we gratefully acknowledge the researchers who published RecA protein sequences and, in particular, those who verified their data for this work. This work was supported by NIH grant GM32335.

References

1. C. M. Fraser, J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, et al., *Science* **270,** 397 (1995).
2. S. Sandler, L. H. Satin, H. S. Samara and A. J. Clark, *NARes* **24,** 2125(1996).
2a. C. R. Woese, O. Kandler and M. L. Whellis, *PNAS* **87,** 4576 (1990).
3. T. Ogawa, X. Yu, A. Shinohara and E. H. Egelman, *Science* **259,** 1896 (1993).
4. R. M. Story, D. K. Bishop, N. Kleckner and T. A. Steitz, *Science* **259,** 1892 (1993).
5. S. C. Kowalczykowski and A. K. Eggleston, *ARB* **63,** 991 (1994).
6. A. Campbell, *CSHSQB* **49,** 839 (1984).
7. J. F. Crow, *in* "The Evolution of Sex: An Examination of Current Ideas" (R. E. Michod and B. R. Levin, eds.), p. 56. Sinauer Assoc., Sunderland, Massachusetts, 1988.
8. M. T. Ghiselin, *in* "The Evolution of Sex: An Examination of Current Ideas" (R. E. Michod and B. R. Levin, eds.), p. 7. Sinauer Assoc., Sunderland, Massachusetts, 1988.
9. J. Maynard Smith, "The Evolution of Sex." Cambridge Univ. Press, London and New York, 1978.
10. J. Maynard Smith, *in* "The Evolution of Sex: An Examination of Current Ideas" (R. E. Michod and B. R. Levin, eds.), p. 106. Sinauer, Assoc., Sunderland, Massachusetts, 1988.
11. H. J. Muller, *Mutat. Res.* **1,** 2 (1964).
12. J. Felsenstein, *Genetics* **78,** 737 (1974).
13. D. A. Hickey and M. R. Rose, *in* "The Evolution of Sex: An Examination of Current Ideas" (R. B. Michod and B. R. Levin, eds.), p. 161. Sinauer Assoc., Sunderland, Massachusetts, 1988.
14. E. C. Dougherty, *Syst. Zool.* **4,** 145 (1955).
15. H. Bernstein, *BioScience* **33,** 326 (1983).
16. H. Bernstein, F. A. Hopf and R. E. Michod, *in* "The Evolution of Sex: An Examination of Current Ideas" (R. E. Michod and B. R. Levin, eds.), p. 139. Sinauer Assoc., Sunderland, Massachusetts, 1988.
17. B. R. Levin, *in* "The Evolution of Sex: An Examination of Current Ideas" (R. B. Michod and B. R. Levin, eds.), p. 194. Sinauer Assoc., Sunderland, Massachusetts, 1988.
18. R. E. Michod, "Eros and Evolution: A Natural Philosophy of Sex." Addison Wesley, Menlo Park, California, 1995.
19. N. A. Ellis, J. Groden, T.-Z. Te, J. Straughen, D. J. Lennon, S. Ciocci, M. Proytcheva and J. German, *Cell* **83,** 655 (1995).

233. W. A. Bedale and M. M. Cox, *JBC* **271,** 5725 (1996).
234. S. W. Morrical, J. Lee and M. M. Cox, *Bchem* **25,** 1482 (1986).
235. L. J. Roman and S. C. Kowalczykowski, *Bchem* **25,** 7375 (1986).
236. B. F. Pugh and M. M. Cox, *in* "Protein Structure, Folding, and Design 2" (D. L. Oxender, ed.), p. 275. Alan R. Liss, New York, 1987.
237. B. F. Pugh and M. M. Cox, *JBC* **263,** 76 (1988).
238. K. A. Kumar, S. Mahalakshmi and K. Muniyappa, *JBC* **268,** 26162 (1993).
238a. P. Wittung, B. Nordén and M. Takahashi, *EJB* **228,** 149 (1995).
239. B. C. Schutte and M. M. Cox, *Bchem* **26,** 5616 (1987).
240. J. W. Lee and M. M. Cox, *Bchem* **29,** 7666 (1990).
241. J. W. Lee and M. M. Cox, *Bchem* **29,** 7677 (1990).
242. S. C. Kowalczykowski, *Bchem* **25,** 5872 (1986).
243. K. L. Menge and F. R. Bryant, *Bchem* **31,** 5151 (1992).
244. K. L. Menge and F. R. Bryant, *Bchem* **31,** 5158 (1992).
245. J. P. Menetski and S. C. Kowalczykowski, *JMB* **181,** 281 (1985).
246. J. P. Menetski, A. Varghese and S. C. Kowalczykowski, *Bchem* **27,** 1205 (1988).
247. E. Stole and F. R. Bryant, *JBC* **270,** 20322 (1995).
248. M. M. Cox, D. A. Soltis, I. R. Lehman, C. DeBrosse and S. J. Benkovic, *JBC* **258,** 2586 (1983).
249. G. M. Weinstock, K. McEntee and I. R. Lehman, *JBC* **256,** 8845 (1981).
250. K. L. Menge and F. R. Bryant, *Bchem* **27,** 2635 (1988).
251. J. I. Kim, M. M. Cox and R. B. Inman, *JBC* **267,** 16444 (1992).
252. M. E. Bianchi and C. E. Radding, *Cell* **35,** 511 (1983).
253. Z. Livneh and I. R. Lehman, *PNAS* **79,** 3171 (1982).
254. W. Rosselli and A. Stasiak, *EMBO J.* **10,** 4391 (1991).
255. J. I. Kim, M. M. Cox and R. B. Inman, *JBC* **267,** 16438 (1992).
256. P. Howard-Flanders, S. C. West and A. Stasiak, *Nature (London)* **309,** 215 (1984).
257. S. C. West, *ARB* **61,** 603 (1992).
258. S. McGavin, *JMB* **55,** 293 (1971).
259. J. H. Wilson, *PNAS* **76,** 3641 (1979).
260. R. A. Fishel and A. Rich, *in* "Mechanisms and Consequences of DNA Damage Processing" (E. C. Friedberg and P. C. Hanawalt, eds.), p. 23. Alan R. Liss, New York, 1988.
261. J. E. Lindsley and M. M. Cox, *JBC* **265,** 10164 (1990).
262. B. Müller, T. Koller and A. Stasiak, *JMB* **212,** 97 (1990).
263. S. A. Chow, S. K. Chiu and B. C. Wong, *JMB* **223,** 79 (1992).
264. E. C. Conley and S. C. West, *JBC* **265,** 10156 (1990).
265. S. A. Chow and C. M. Radding, *PNAS* **82,** 5646 (1985).
266. D. K. Gonda and C. M. Radding, *JBC* **261,** 13087 (1986).
267. D. A. Julin, P. W. Riddles and I. R. Lehman, *JBC* **261,** 1025 (1986).
268. M. Bianchi, C. Das Gupta and C. M. Radding, *Cell* **34,** 931 (1983).
269. P. W. Riddles and I. R. Lehman, *JBC* **260,** 165 (1985).
270. J. E. Yancey-Wrona and R. D. Camerini-Otero, *Curr. Biol.* **5,** 1149 (1995).
271. V. B. Zhurkin, G. Raghunathan, N. B. Ulyanov, O. R. Camerini and R. L. Jernigan, *JMB* **239,** 181 (1994).
272. S. Lacks, *Genetics* **53,** 207 (1966).
273. B. Burnett, B. J. Rao, B. Jwang, G. Reddy and C. M. Radding, *JMB* **238,** 540 (1994).
274. S. Umlauf, "Unusual DNA Structure in Site-Specific and Homologous Recombination." Ph.D. Thesis, University of Wisconsin-Madison, 1990.
275. S. K. Jain, R. B. Inman and M. M. Cox, *JBC* **267,** 4215 (1992).
276. A. Stasiak, *Mol. Microbiol.* **6,** 3267 (1992).

20. N. Irino, K. Nakayama and H. Nakayama, *MGG* **205,** 298 (1986).
21. K. Umezu and H. Nakayama, *JMB* **230,** 1145 (1993).
22. C.-E. Yu, J. Oshima, Y.-H. Fu, E. Wijsman, F. Hisama, R. Alisch, S. Matthews, J. Nakura, T. Miki, S. Ouais, G. M. Martin, J. Mulligan and G. D. Schellenberg, *Science* **272,** 258 (1996).
23. H. Bernstein, H. C. Byerly, F. A. Hopf and R. E. Michod, *Science* **229,** 1277 (1985).
24. H. Potter and D. Dressler, *in* "The Recombination of Genetic Material" (K. B. Low, ed.), p. 217. Academic Press, San Diego, 1988.
25. M. M. Cox, *Mol. Microbiol* **5,** 1295 (1991).
26. M. M. Cox, *BioEssays* **15,** 617 (1993).
27. A. J. Clark and S. J. Sandler, *Crit. Rev. Microbiol.* **20,** 125 (1994).
28. R. K. Selander and B. R. Levin, *Science* **210,** 545 (1980).
29. G. R. Smith, *Cell* **64,** 19 (1991).
30. J. M. Smith, *J. Hered.* **84,** 326 (1993).
31. E. M. Park, M. K. Shigenaga, P. Degan, T. S. Korn, J. W. Kitzler, C. M. Wehr, P. Kolachana and B. N. Ames, *PNAS* **89,** 3375 (1992).
32. F. N. Capaldo, G. Ramsey and S. D. Barbour, *J. Bact.* **118,** 242 (1974).
33. R. Holliday, *Genet. Res.* **5,** 282 (1964).
33a. A. Kuzminor, "Recombinational Repair of DNA Damage." R. G. Landes Co., Austin, Texas, 1996.
34. S. C. West, E. Cassuto and P. Howard-Flanders, *Nature (London)* **294,** 659 (1981).
35. J. W. Szostak, W. T. L. Orr, R. J. Rothstein and F. W. Stahl, *Cell* **33,** 25 (1983).
36. L. S. Morse and C. Pauling, *PNAS* **72,** 4645 (1975).
37. A. G. Miguel and R. M. Tyrrell, *Biophys. J.* **49,** 485 (1986).
38. T. C. Wang and K. C. Smith, *J. Bact.* **165,** 1023 (1986).
39. M. J. Daly, L. Ouyang, P. Fuchs and K. W. Minton, *J. Bact.* **176,** 3508 (1994).
40. K. W. Minton, *Mol. Microbiol.* **13,** 9 (1994).
41. K. W. Minton and M. J. Daly, *BioEssays* **17,** 457 (1995).
42. V. Mattimore and J. R. Battista, *J. Bact.* **178,** 633 (1996).
43. W. Y. Feng, E. H. Lee and J. B. Hays, *Genetics* **129,** 1007 (1991).
44. J. B. Hays and J. G. Hays, *Biopolymers* **31,** 1565 (1991).
45. W. Y. Feng and J. B. Hays, *Genetics* **140,** 1175 (1995).
46. D. Touati, M. Jacques, B. Tardat, L. Bouchard and S. Despied, *J. Bact.* **177,** 2305 (1995).
46a. T. Galitski, Ph.D. Thesis, University of Utah, 1996.
47. N. Higashitani, A. Higashitani, A. Roth and K. Horiuchi, *J. Bact.* **174,** 1612 (1992).
48. M. M. Cox, K. McEntee and I. R. Lehman, *JBC* **256,** 4676 (1981).
49. T. Shibata, R. P. Cunningham and C. M. Radding, *JBC* **256,** 7557 (1981).
50. S. M. Cotterill, A. C. Satterthwait and A. R. Fersht, *Bchem* **21,** 4332 (1982).
51. J. Griffith and C. G. Shores, *Bchem* **24,** 158 (1985).
52. S. Kuramitsu, K. Hamaguchi, T. Ogawa and H. Ogawa, *J. Biochem. (Tokyo)* **90,** 1033 (1981).
53. T. Ogawa, H. Wabiko, T. Tsurimoto, T. Horii, H. Masukata and H. Ogawa, *CSHSQB* **2,** 909 (1979).
54. W. A. Bedale, R. B. Inman and M. M. Cox, *JBC* **268,** 15004 (1993).
55. T. Simonson, M. Kubista, R. Sjoback, H. Ryberg and M. Takahashi, *J. Mol. Recog.* **7,** 199 (1994).
56. L. J. Gudas and D. W. Mount, *PNAS* **74,** 5280 (1977).
57. J. H. Krueger and G. C. Walker, *PNAS* **81,** 1499 (1984).
58. N. Garvey, A. C. St. John and E. M. Witkin, *J. Bact.* **163,** 870 (1985).
59. C. Cazaux, F. Larminat, G. Villani, N. P. Johnson, M. Schnarr and M. Defais, *JBC* **269,** 8246 (1994).
60. S. L. Brenner, A. Zlotnick and J. D. Griffith, *JMB* **204,** 959 (1988).

61. M. Takahashi, *JBC* **264,** 288 (1989).
62. D. H. Wilson and A. S. Benight, *JBC* **265,** 7351 (1990).
63. S. L. Brenner, A. Zlotnick and W. F. Stafford, *JMB* **216,** 949 (1990).
64. R. M. Story, I. T. Weber and T. A. Steitz, *Nature (London)* **355,** 318 (1992).
65. R. M. Story and T. A. Steitz, *Nature (London)* **355,** 374 (1992).
66. X. Yu and E. H. Egelman, *JMB* **227,** 334 (1992).
67. E. H. Egelman and A. Stasiak, *Micron* **24,** 309 (1993).
68. E. H. Egelman, *Current Opin. Struct. Biol.* **3,** 189 (1993).
68a. C. Ellouze, M. Takahashi, P. Wittung, K. Mortensen, M. Schnarr and B. Nordén, *EJB* **233,** 579 (1995)
69. M. Kimura and T. Ohta, *PNAS* **71,** 2848 (1974).
70. A. I. Roca and M. M. Cox, *CRC Crit. Rev. Biochem. Molec. Biol.* **25,** 415 (1990).
71. J. E. Walker, M. Saraste, M. J. Runswick and N. J. Gay, *EMBO J* **1,** 945 (1982).
72. E. F. Pai, U. Krengel, G. A. Petsko, R. S. Goody, W. Kabsch and A. Wittinghofer, *EMBO J* **9,** 2351 (1990).
73. T. F. M. la Cour, J. Nyborg, S. Thirup and B. F. C. Clark, *EMBO J* **4,** 2385 (1995).
74. J. A. Eisen, *J. Mol. Evol.* **41,** 1105 (1995).
75. S. Karlin, G. M. Weinstock and V. Brendel, *J. Bact.* **177,** 6881 (1995).
76. P. Sung, *Science* **265,** 1241 (1994).
77. D. K. Bishop, D. Park, L. Xu and N. Kleckner, *Cell* **69,** 439 (1992).
78. B. P. Rubin, D. O. Ferguson and W. K. Holloman, *McBiol* **14,** 6287 (1994).
79. T. Flores, C. A. Orengo and D. Moss, *Protein Sci.* **2,** 1811 (1993).
80. X. Yu, E. Angov, R. D. Camerini-Otero and E. Egelman, *Biophys. J.* **69,** 2728 (1995).
81. B. Honig and A. Nicholls, *Science* **268,** 1144 (1995).
82. M. Jonsson, U. Jacobsson, M. Takahashi and B. Nordén, *JCS Faraday Trans.* **89,** 2791 (1993).
83. T. Mikawa, R. Masui, T. Ogawa, H. Ogawa and S. Kuramitsu, *JMB* **250,** 471 (1995).
84. S. C. Kowalczykowski, *Biochimie* **73,** 289 (1991).
85. S. C. Kowalczykowski, D. A. Dixon, A. K. Eggleston, S. D. Lauder and W. M. Rehrauer, *Microbiol. Rev.* **58,** 401 (1994).
86. M. Saraste, P. R. Sibbald and A. Wittinghofer, *Trends Biochem. Sci.* **15,** 430 (1990).
87. A. Bairoch, *NARes* **19,** 2241 (1991).
88. S. K. Liu, J. A. Eisen, P. C. Hanawalt and I. Tessman, *J. Bact.* **175,** 6518 (1993).
89. F. B. Perler, E. O. Davis, G. E. Dean, F. S. Gimble, W. E. Jack, N. Neff, C. J. Noren, J. Thorner and M. Belfort, *NARes* **22,** 1125 (1994).
90. E. O. Davis, P. J. Jenner, P. C. Brooks, M. J. Colston and S. G. Sedgwick, *Cell* **71,** 201 (1992).
91. E. O. Davis, H. S. Thangaraj, P. C. Brooks and M. J. Colston, *EMBO J* **13,** 699 (1994).
92. R. A. Kumar, M. B. Vaze, N. R. Chandra, M. Vijayan and K. Muniyappa, *Bchem* **35,** 1793 (1996).
93. W. Selbitschka, W. Arnold, U. B. Priefer, T. Rottschafer, M. Schmidt, R. Simon and A. Pühler, *MGG* **229,** 86 (1991).
94. A. Shinohara, H. Ogawa and T. Ogawa, *Cell* **69,** 457 (1992).
95. M. Gribskov, A. D. McLachlan and D. Eisenberg, *PNAS* **84,** 4355 (1987).
96. S. F. Altschul, W. Gish, W. Miller, E. W. Myers and D. J. Lipman, *JMB* **215,** 403 (1990).
97. J. Griffith and T. Formosa, *JBC* **260,** 4484 (1985).
98. X. Yu and E. H. Egelman, *JMB* **232,** 1 (1993).
99. T. Formosa and B. M. Alberts, *JBC* **261,** 6107 (1986).
100. T. Yonesaki and T. Minagawa, *EMBO J* **4,** 3321 (1985).
101. T. Kodadek, M. L. Wong and B. M. Alberts, *JBC* **263,** 9427 (1988).
102. H. Cerutti, M. Osman, P. Grandoni and A. T. Jagendorf, *PNAS* **89,** 8068 (1992).

103. R. Cedergren, M. W. Gray, Y. Abel and D. Sankoff, *J. Mol. Evol.* **28,** 98 (1988).
104. R. Cheng, T. I. Baker, C. E. Cords and R. J. Radloff, *Mutat. Res.* **294,** 223 (1993).
105. J. C. Game, *in* "Yeast Genetics: Fundamental and Applied Aspects" (J. F. T. Spencer, D. Spencer and A. R. W. Smith, eds.), p. 109. Springer Publ., New York, 1983.
106. M. A. Resnick, *in* "Meiosis" (P. B. Moens, ed.), p. 157. Academic Press, New York, 1987.
107. A. J. Rattray and L. S. Symington, *Genetics* **139,** 45 (1995).
108. D. K. Bishop, *Cell* **79,** 1081 (1994).
109. A. F. Neuwald, D. E. Berg and G. V. Stauffer, *Gene* **120,** 1 (1992).
110. S. J. Sandler, L. H. Satin, H. S. Samra and A. J. Clark, *NARes* **24,** 2125 (1996).
111. C. R. Woese, O. Kandler and M. L. Whellis, *PNAS* **87,** 4576 (1990).
112. M. Gribskov, *in* "Computer Analysis of Sequence Data, Part II" (A. M. Griffin and H. G. Griffin, eds.), p. 247. Humana Press, Totowa, New Jersey, 1994.
113. J. T. Konola, K. M. Logan and K. L. Knight, *JMB* **237,** 20 (1994).
114. W. M. Rehrauer and S. C. Kowalczykowski, *JBC* **268,** 1292 (1993).
115. Q. Shan, M. M. Cox and R. B. Inman, *JBC* **271,** 5712 (1996).
116. J. T. Konola, H. G. Nastri, K. M. Logan and K. L. Knight, *JBC* **270,** 8411 (1995).
117. M. Takahashi, M. Kubista and B. Nordén, *Biochimie* **73,** 219 (1991).
118. M. Takahashi and B. Nordén, *Adv. Biophys.* **30,** 1 (1994).
119. M. Kubista, T. Simonson, R. Sjöback, H. Widlund and A. Johansson, eds., "*Towards an Understanding of the Mechanism of DNA Strand Exchange Promoted by RecA Protein.* Biological Structure and Function." Proceedings of the Ninth Conversation, The State University of New York (R. H. Sarma and M. H. Sarma, eds.). Adenine Press, New York, 1995.
120. H. Kurumizaka, B. J. Rao, T. Ogawa, C. M. Radding and T. Shibata, *NARes* **22,** 3387 (1994).
121. M. M. Cox, *JBC* **270,** 26021 (1995).
122. H. G. Nastri and K. L. Knight, *JBC* **269,** 26311 (1994).
123. R. V. Gardner, O. N. Voloshin and R. D. Camerini-Otero, *EJB* **233,** 419 (1995).
124. V. A. Malkov and R. D. Camerini-Otero, *JBC* **270,** 30230 (1995).
125. O. N. Voloshin, L. Wang and R. D. Camerini-Otero, *Science* **272,** 868 (1996).
126. Y. Wang and K. Adzuma, *Bchem* **35,** 3563 (1996).
127. K. Morimatsu and T. Horii, *EJB* **228,** 772 (1995).
128. W. M. Rehrauer and S. C. Kowalczykowski, *JBC* **271,** 11996 (1996).
129. A. Zlotnick and S. L. Brenner, *JMB* **209,** 447 (1989).
130. R. C. Benedict and S. C. Kowalczykowski, *JBC* **263,** 15513 (1988).
131. S. Tateishi, T. Horii, T. Ogawa and H. Ogawa, *JMB* **223,** 115 (1992).
132. B. V. Prasad and W. Chiu, *JMB* **193,** 579 (1987).
133. E. A. de Jong, D. J. van, B. J. Harmsen, G. I. Tesser, R. N. Konings and C. W. Hilbers, *JMB* **206,** 133 (1989).
134. A. Stasiak, E. Di Capua and T. Koller, *JMB* **151,** 557 (1981).
135. A. Stasiak and E. Di Capua, *Nature (London)* **299,** 185 (1982).
136. B. F. Pugh, B. C. Schutte and M. M. Cox, *JMB* **205,** 487 (1989).
137. D. F. Dombroski, D. G. Scraba, R. D. Bradley and A. R. Morgan, *NARes* **11,** 7487 (1983).
138. R. J. Thresher and J. D. Griffith, *PNAS* **87,** 5056 (1990).
139. S. K. Kim, B. Nordén and M. Takahashi, *JBC* **268,** 14799 (1993).
140. R. V. Prigodich, J. Casas-Finet, K. R. Williams, W. Konigsberg and J. E. Coleman, *Bchem* **23,** 522 (1984).
141. M. I. Khamis, J. R. Casas-Finet and A. H. Maki, *BBA* **950,** 132 (1988).
142. S. Eriksson, B. Nordén and M. Takahashi, *JBC* **268,** 1805 (1993).
143. F. Maraboeuf, O. Voloshin, O. R. Camerini and M. Takahashi, *JBC* **270,** 30927 (1995).
144. T. T. Nguyen, K. A. Muench and F. R. Bryant, *JBC* **268,** 3107 (1993).
145. M. C. Skiba and K. L. Knight, *JBC* **269,** 3823 (1994).

146. P. Howard-Flanders and L. Theriot, *Genetics* **53,** 1137 (1966).
147. N. S. Willetts and A. J. Clark, *J. Bact.* **100,** 231 (1969).
148. R. G. Lloyd and B. Low, *Genetics* **84,** 675 (1976).
149. L. N. Csonka and A. J. Clark, *Genetics* **93,** 321 (1979).
150. M. Sassanfar and J. W. Roberts, *JMB* **212,** 79 (1990).
151. A. J. Clark, *J. Cell. Physiol.* **70,** 165 (1967).
152. Z. Horii and A. J. Clark, *JMB* **80,** 327 (1973).
153. A. Bagg, C. J. Kenyon and G. C. Walker, *PNAS* **78,** 5749 (1981).
154. J. M. Weisemann, C. Funk and G. M. Weinstock, *J. Bact.* **160,** 112 (1984).
155. D. G. Ennis, N. Ossanna and D. W. Mount, *J. Bact.* **171,** 2533 (1989).
156. S. D. Lauder and S. C. Kowalczykowski, *JMB* **234,** 72 (1993).
157. J. G. Wetmur, D. M. Wong, B. Ortiz, J. Tong, F. Reichert and D. H. Gelfand, *JBC* **269,** 25928 (1994).
158. K. M. Logan and K. L. Knight, *JMB* **232,** 1048 (1993).
159. K. Morimatsu, T. Horii and M. Takahashi, *EJB* **228,** 779 (1995).
160. K. L. Knight and K. McEntee, *JBC* **260,** 867 (1985).
161. S. Eriksson, B. Nordén, K. Morimatsu, T. Horii and M. Takahashi, *JBC* **268,** 1811 (1993).
162. E. Stole and F. R. Bryant, *JBC* **269,** 7919 (1994).
163. W. P. Diver, N. J. Sargentini and K. C. Smith, *Int. J. Rad. Biol. Rel. Stud. Physics. Chem. Med.* **42,** 339 (1982).
163a. Y. Song and N. J. Sargentini, *J. Bact.* **178,** 5045 (1996).
163b. E. V. Koonin, R. L. Tatusov and K. E. Rudd, *in* "*Escherichia coli* and *Salmonella* (F. C. Neidhardt, ed.), p. 2203. ASM Press, Washington, D.C., 1996.
164. M. M. Cox and I. R. Lehman, *ARB* **56,** 229 (1987).
165. C. M. Radding, *in* "Genetic Recombination" (R. Kucherlapati and G. R. Smith, eds.), p. 193. American Society for Microbiology, Washington, D.C., 1988.
166. A. Stasiak and E. H. Egelman, *in* "Genetic Recombination" (R. Kucherlapati and G. R. Smith, eds.), p. 265. American Society for Microbiology, Washington, D.C., 1988.
167. J. Heuser and J. Griffith, *JMB* **210,** 473 (1989).
168. E. H. Egelman and A. Stasiak, *JMB* **191,** 677 (1986).
169. E. Di Capua, A. Engel, A. Stasiak and T. Koller, *JMB* **157,** 87 (1982).
170. B. Nordén, C. Elvingson, M. Kubista, B. Sjoberg, H. Ryberg, M. Ryberg, K. Mortensen and M. Takahashi, *JMB* **226,** 1175 (1992).
171. M. C. Leahy and C. M. Radding, *JBC* **261,** 6954 (1986).
172. E. Di Capua and B. Müller, *EMBO J.* **6,** 2493 (1987).
173. E. H. Egelman and X. Yu, *Science* **245,** 404 (1989).
174. K. McEntee, G. M. Weinstock and I. R. Lehman, *JBC* **256,** 8835 (1981).
175. M. Amaratunga and A. S. Benight, *BBRC* **157,** 127 (1988).
175a. C. Cazenave, M. Chabbert and J. J. Toulme, *BBA* **781,** 7 (1984).
176. S. L. Brenner, R. S. Mitchell, S. W. Morrical, S. K. Neuendorf, B. C. Schutte and M. M. Cox, *JBC* **262,** 4011 (1987).
177. J. P. Menetski and S. C. Kowalczykowski, *JBC* **262,** 2085 (1987).
178. J. P. Menetski and S. C. Kowalczykowski, *Bchem* **28,** 5871 (1989).
179. R. S. Mitchell, A. Zlotnick and S. L. Brenner, *Biophys. J.* **53,** 220 (1988).
180. S. D. Lauder and S. C. Kowalczykowski, *JBC* **266,** 5450 (1991).
181. A. Zlotnick, R. S. Mitchell, R. K. Steed and S. L. Brenner, *JBC* **268,** 22525 (1993).
182. E. H. Egelman and A. Stasiak, *JMB* **200,** 329 (1988).
183. A. Stasiak, E. H. Egelman and P. Howard-Flanders, *JMB* **202,** 659 (1988).
184. B. F. Pugh and M. M. Cox, *JBC* **262,** 1326 (1987).
185. B. F. Pugh and M. M. Cox, *JMB* **203,** 479 (1988).

186. S. C. Kowalczykowski, J. Clow and R. A. Krupp, *PNAS* **84,** 3127 (1987).
187. C. Lu, R. H. Scheuermann and H. Echols, *PNAS* **83,** 619 (1986).
188. M. Kojima, M. Suzuki, T. Morita, T. Ogawa, H. Ogawa and M. Tada, *NARes* **18,** 2707 (1990).
189. Y. H. Wang, C. D. Bortner and J. Griffith, *JBC* **268,** 17571 (1993).
189a. S. K. Kim, M. Takahashi, B. Jernstrom and B. Nordén, *Carcinogenesis* **14,** 311 (1993).
190. J. A. Blaho and R. D. Wells, *JBC* **262,** 6082 (1987).
191. J. A. Blaho and R. D. Wells, *This Series* **37,** 107 (1989).
192. J. Kim, J. Heuser and M. M. Cox, *JBC* **264,** 21848 (1989).
193. P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay and F. Caron, *Science* **271,** 792 (1996).
194. S. B. Smith, Y. Cui and C. Bustamante, *Science* **271,** 795 (1996).
195. M. Takahashi, M. Kubista and B. Nordén, *JMB* **205,** 137 (1989).
196. M. Kubista, M. Takahashi and B. Nordén, *JBC* **265,** 18891 (1990).
197. B. Nordén, C. Elvingson, T. Eriksson, M. Kubista, B. Sjoberg, M. Takahashi and K. Mortensen, *JMB* **216,** 223 (1990).
198. K. Adzuma, *Genes Dev.* **6,** 1679 (1992).
199. P. Wittung, B. Nordén, S. K. Kim and M. Takahashi, *JBC* **269,** 5799 (1994).
200. P. Wittung, M. Funk, B. Jernström, B. Nordén and M. Takahashi, *FEBS Lett.* **368,** 64 (1995).
201. E. Di Capua, M. Schnarr and P. A. Timmins, *Bchem* **28,** 3287 (1989).
202. S. A. Chow, S. M. Honigberg and C. M. Radding, *JBC* **263,** 3335 (1988).
203. J. E. Lindsley and M. M. Cox, *JMB* **205,** 695 (1989).
204. T. Shibata, C. Das Gupta, R. P. Cunningham and C. M. Radding, *PNAS* **77,** 2606 (1980).
205. K. McEntee, G. M. Weinstock and I. R. Lehman, *PNAS* **77,** 857 (1980).
206. M. M. Cox and I. R. Lehman, *PNAS* **78,** 3433 (1981).
207. M. M. Cox and I. R. Lehman, *JBC* **257,** 8523 (1982).
208. M. M. Cox, D. A. Soltis, Z. Livneh and I. R. Lehman, *CSHSQB* **47,** 803 (1983).
209. M. M. Cox, D. A. Soltis, Z. Livneh and I. R. Lehman, *JBC* **258,** 2577 (1983).
210. K. Muniyappa, S. L. Shaner, S. S. Tsang and C. M. Radding, *PNAS* **81,** 2757 (1984).
211. S. C. Kowalczykowski, J. Clow, R. Somani and A. Varghese, *JMB* **193,** 81 (1987).
212. S. C. Kowalczykowski and R. A. Krupp, *JMB* **193,** 97 (1987).
213. D. A. Dixon and S. C. Kowalczykowski, *Cell* **66,** 361 (1991).
214. S. C. Kowalczykowski, *Experientia* **50,** 204 (1994).
215. G. R. Smith, *Experientia* **50,** 234 (1994).
216. G. J. Phillips, D. C. Prasher and S. R. Kushner, *J. Bact.* **170,** 2089 (1988).
217. G. R. Smith, *Cell* **58,** 807 (1989).
218. A. J. Clark, *Annu. Rev. Microbiol.* **25,** 437 (1971).
219. H. Razavy, S. K. Szigety and S. M. Rosenberg, *Genetics* **142,** 133 (1996).
220. J. C. Register III and J. Griffith, *JBC* **260,** 12308 (1985).
221. B. B. Konforti and R. W. Davis, *JBC* **265,** 6916 (1990).
222. B. B. Konforti and R. W. Davis, *JBC* **266,** 10112 (1991).
223. B. B. Konforti and R. W. Davis, *JMB* **227,** 38 (1992).
224. M. Dutreix, B. J. Rao and C. M. Radding, *JMB* **219,** 645 (1991).
225. S. L. Shaner and C. M. Radding, *JBC* **262,** 9211 (1987).
226. J. E. Lindsley and M. M. Cox, *JBC* **265,** 9043 (1990).
227. Q. Shan, J. M. Bork, B. L. Webb, R. B. Inman and M. M. Cox, *unpublished* (1996).
228. Q. Shan and M. M. Cox, *JMB* **257,** 756 (1996).
229. S. K. Neuendorf and M. M. Cox, *JBC* **261,** 8276 (1986).
230. J. P. Menetski and S. C. Kowalczykowski, *JBC* **262,** 2093 (1987).
231. X. Yu and E. H. Egelman, *JMB* **225,** 193 (1992).
232. G. M. Weinstock, K. McEntee and I. R. Lehman, *JBC* **256,** 8829 (1981).

233. W. A. Bedale and M. M. Cox, *JBC* **271,** 5725 (1996).
234. S. W. Morrical, J. Lee and M. M. Cox, *Bchem* **25,** 1482 (1986).
235. L. J. Roman and S. C. Kowalczykowski, *Bchem* **25,** 7375 (1986).
236. B. F. Pugh and M. M. Cox, *in* "Protein Structure, Folding, and Design 2" (D. L. Oxender, ed.), p. 275. Alan R. Liss, New York, 1987.
237. B. F. Pugh and M. M. Cox, *JBC* **263,** 76 (1988).
238. K. A. Kumar, S. Mahalakshmi and K. Muniyappa, *JBC* **268,** 26162 (1993).
238a. P. Wittung, B. Nordén and M. Takahashi, *EJB* **228,** 149 (1995).
239. B. C. Schutte and M. M. Cox, *Bchem* **26,** 5616 (1987).
240. J. W. Lee and M. M. Cox, *Bchem* **29,** 7666 (1990).
241. J. W. Lee and M. M. Cox, *Bchem* **29,** 7677 (1990).
242. S. C. Kowalczykowski, *Bchem* **25,** 5872 (1986).
243. K. L. Menge and F. R. Bryant, *Bchem* **31,** 5151 (1992).
244. K. L. Menge and F. R. Bryant, *Bchem* **31,** 5158 (1992).
245. J. P. Menetski and S. C. Kowalczykowski, *JMB* **181,** 281 (1985).
246. J. P. Menetski, A. Varghese and S. C. Kowalczykowski, *Bchem* **27,** 1205 (1988).
247. E. Stole and F. R. Bryant, *JBC* **270,** 20322 (1995).
248. M. M. Cox, D. A. Soltis, I. R. Lehman, C. DeBrosse and S. J. Benkovic, *JBC* **258,** 2586 (1983).
249. G. M. Weinstock, K. McEntee and I. R. Lehman, *JBC* **256,** 8845 (1981).
250. K. L. Menge and F. R. Bryant, *Bchem* **27,** 2635 (1988).
251. J. I. Kim, M. M. Cox and R. B. Inman, *JBC* **267,** 16444 (1992).
252. M. E. Bianchi and C. E. Radding, *Cell* **35,** 511 (1983).
253. Z. Livneh and I. R. Lehman, *PNAS* **79,** 3171 (1982).
254. W. Rosselli and A. Stasiak, *EMBO J.* **10,** 4391 (1991).
255. J. I. Kim, M. M. Cox and R. B. Inman, *JBC* **267,** 16438 (1992).
256. P. Howard-Flanders, S. C. West and A. Stasiak, *Nature (London)* **309,** 215 (1984).
257. S. C. West, *ARB* **61,** 603 (1992).
258. S. McGavin, *JMB* **55,** 293 (1971).
259. J. H. Wilson, *PNAS* **76,** 3641 (1979).
260. R. A. Fishel and A. Rich, *in* "Mechanisms and Consequences of DNA Damage Processing" (E. C. Friedberg and P. C. Hanawalt, eds.), p. 23. Alan R. Liss, New York, 1988.
261. J. E. Lindsley and M. M. Cox, *JBC* **265,** 10164 (1990).
262. B. Müller, T. Koller and A. Stasiak, *JMB* **212,** 97 (1990).
263. S. A. Chow, S. K. Chiu and B. C. Wong, *JMB* **223,** 79 (1992).
264. E. C. Conley and S. C. West, *JBC* **265,** 10156 (1990).
265. S. A. Chow and C. M. Radding, *PNAS* **82,** 5646 (1985).
266. D. K. Gonda and C. M. Radding, *JBC* **261,** 13087 (1986).
267. D. A. Julin, P. W. Riddles and I. R. Lehman, *JBC* **261,** 1025 (1986).
268. M. Bianchi, C. Das Gupta and C. M. Radding, *Cell* **34,** 931 (1983).
269. P. W. Riddles and I. R. Lehman, *JBC* **260,** 165 (1985).
270. J. E. Yancey-Wrona and R. D. Camerini-Otero, *Curr. Biol.* **5,** 1149 (1995).
271. V. B. Zhurkin, G. Raghunathan, N. B. Ulyanov, O. R. Camerini and R. L. Jernigan, *JMB* **239,** 181 (1994).
272. S. Lacks, *Genetics* **53,** 207 (1966).
273. B. Burnett, B. J. Rao, B. Jwang, G. Reddy and C. M. Radding, *JMB* **238,** 540 (1994).
274. S. Umlauf, "Unusual DNA Structure in Site-Specific and Homologous Recombination." Ph.D. Thesis, University of Wisconsin-Madison, 1990.
275. S. K. Jain, R. B. Inman and M. M. Cox, *JBC* **267,** 4215 (1992).
276. A. Stasiak, *Mol. Microbiol.* **6,** 3267 (1992).

277. B. J. Rao, S. K. Chiu and C. M. Radding, *JMB* **229,** 328 (1993).
278. S. K. Jain, M. M. Cox and R. B. Inman, *JBC* **270,** 4943 (1995).
279. R. Baliga, J. W. Singleton and P. B. Dervan, *PNAS* **92,** 10393 (1995).
280. M. A. Podyminogin, R. B. Meyer and H. B. Gamper, *Bchem* **34,** 13098 (1995).
281. M. A. Podyminogen, R. B. Meyer and H. B. Gamper, *Bchem* **35,** 7267 (1996).
282. S. W. Umlauf, M. M. Cox and R. B. Inman, *JBC* **265,** 16898 (1990).
283. P. Hsieh, O. C. Camerini and O. R. Camerini, *Genes Dev.* **4,** 1951 (1990).
284. B. J. Rao, M. Dutreix and C. M. Radding, *PNAS* **88,** 2984 (1991).
285. S. K. Chiu, B. J. Rao, R. M. Story and C. M. Radding, *Bchem* **32,** 13146 (1993).
286. H. E. Moser and P. B. Dervan, *Science* **238,** 645 (1987).
287. R. D. Wells, D. A. Collier, J. C. Hanvey, M. Shimizu and F. Wohlrab, *FASEB J.* **2,** 2939 (1988).
288. A. K. Shchyolkina, E. N. Timofeev, O. F. Borisova, I. A. Il'icheva, E. E. Minyat, E. B. Khomyakova and V. L. Florentiev, *FEBS Lett.* **339,** 113 (1994).
289. J. P. Menetski, D. G. Bear and S. C. Kowalczykowski, *PNAS* **87,** 21 (1990).
290. W. Rosselli and A. Stasiak, *JMB* **216,** 335 (1990).
291. S. K. Jain, M. M. Cox and R. B. Inman, *JBC* **269,** 20653 (1994).
292. S. C. Kowalczykowski and R. A. Krupp, *PNAS* **92,** 3478 (1995).
293. J. L. Sikorav and G. M. Church, *JMB* **222,** 1085 (1991).
294. E. Kaslan and W. D. Heyer, *JBC* **269,** 14103 (1994).
295. B. Alberts and R. Miake-Lye, *Cell* **68,** 415 (1992).
296. C. J. Ullsperger and M. M. Cox, *Bchem* **34,** 10859 (1995).
297. P. E. Lavery and S. C. Kowalczykowski, *JBC* **267,** 9315 (1992).
298. M. M. Cox, *Trends Biochem. Sci.* **19,** 217 (1994).
299. B. Jwang and C. M. Radding, *PNAS* **89,** 7596 (1992).
300. P. Modrich, *ARGen* **25,** 229 (1991).
301. R. R. Meyer, J. Glassberg and A. Kornberg, *PNAS* **76,** 1702 (1979).
302. R. R. Meyer and P. S. Laine, *Microbiol. Rev.* **54,** 342 (1990).
303. S. C. Schmellik and E. S. Tessman, *J. Bact.* **172,** 4378 (1990).
304. A. Sancar, K. R. Williams, J. W. Chase and W. D. Rupp, *PNAS* **78,** 4274 (1981).
305. T. M. Lohman, W. Bujalowski and L. B. Overman, *Trends Biochem. Sci.* **13,** 250 (1988).
306. T. M. Lohman and M. E. Ferrari, *ARB* **63,** 527 (1994).
307. J. D. Griffith, L. D. Harris and J. Register III, *CSHSQB* **49,** 553 (1984).
308. W. Bujalowski and T. M. Lohman, *JMB* **195,** 897 (1987).
309. W. Bujalowski and T. M. Lohman, *Bchem* **25,** 7799 (1986).
310. K. Muniyappa, K. Williams, J. W. Chase and C. M. Radding, *NARes* **18,** 3967 (1990).
311. S. W. Morrical and M. M. Cox, *Bchem* **29,** 837 (1990).
312. C. Bortner and J. Griffith, *JMB* **215,** 623 (1990).
313. T. Griffin IV and R. D. Kolodner, *J. Bact.* **172,** 6291 (1990).
314. M. V. V. S. Madiraju and A. J. Clark, *NARes* **19,** 6295 (1991).
315. M. V. Madiraju and A. J. Clark, *J. Bact.* **174,** 7705 (1992).
316. K. Umezu, N. W. Chi and R. D. Kolodner, *PNAS* **90,** 3875 (1993).
317. K. Umezu and R. D. Kolodner, *JBC* **269,** 30005 (1994).
318. B. L. Webb, M. M. Cox and R. B. Inman, *JBC* **270,** 31397 (1995).
319. M. V. Madiraju, A. Templin and A. J. Clark, *PNAS* **85,** 6592 (1988).
320. P. L. Moreau, *J. Bact.* **170,** 2493 (1988).
321. M. Sassanfar and J. Roberts, *J. Bact.* **173,** 5869 (1991).
322. P. E. Lavery and S. C. Kowalczykowski, *JMB* **203,** 861 (1988).
323. M. V. Madiraju, P. E. Lavery, S. C. Kowalczykowski and A. J. Clark, *Bchem* **31,** 10529 (1992).

324. R. Kolodner, R. A. Fishel and M. Howard, *J. Bact.* **163,** 1060 (1985).
325. J. Ahnn, P. E. March, H. E. Takiff and M. Inouye, *PNAS* **83,** 8849 (1986).
326. P. T. Morrison, S. T. Lovett, L. E. Gilson and R. Kolodner, *J. Bact.* **171,** 3641 (1989).
327. H. E. Takiff, S. M. Chen and D. L. Court, *J. Bact.* **171,** 2581 (1989).
328. C. Luisi-DeLuca and R. Kolodner, *JMB* **236,** 124 (1994).
329. C. Luisi-DeLuca, *J. Bact.* **177,** 566 (1995).
330. A. A. Mahdi and R. G. Lloyd, *NARes* **17,** 6781 (1989).
331. A. A. Mahdi and R. G. Lloyd, *MGG* **216,** 503 (1989).
332. J. C. Alonso, A. C. Stiege, B. Dobrinski and R. Lurz, *JBC* **268,** 1424 (1993).
332a. J. M. Berg and Y. Shi, *Science* **271,** 1081 (1996).
333. B. Van Houten and A. Snowden, *BioEssays* **15,** 51 (1993).
334. G. R. Smith, *Genome* **31,** 520 (1989).
335. T. C. Wang, H. Y. Chang and J. L. Hung, *Mutat. Res.* **294,** 157 (1993).
336. J. A. Sawitzke and F. W. Stahl, *Genetics* **130,** 7 (1992).
336a. S. P. Hegde, M. Rajagopalan and M. V. Madiraju, *J. Bact.* **178,** 184 (1996).
337. S. J. Sandler, *Mol. Microbiol.* **19,** 871 (1996).
338. T. Kogoma, G. W. Cadwell, K. G. Barnard and T. Asai, *J. Bact.* **178,** 1258 (1996).
339. N. J. Sargentini and K. C. Smith, *Radiat. Res.* **107,** 58 (1986).
340. R. G. Lloyd, M. C. Porton and C. Buckman, *MGG* **212,** 317 (1988).
341. Y. C. Tseng, J. L. Hung and T. C. Wang, *Mutat. Res.* **315,** 1 (1994).
342. S. J. Sandler and A. J. Clark, *J. Bact.* **176,** 3661 (1994).
343. T. V. Wang and K. C. Smith, *J. Bact.* **158,** 727 (1984).
344. N. Otsuji, H. Ilyehara and Y. Hideshima, *J. Bact.* **117,** 337 (1974).
345. F. E. Benson, G. T. Illing, G. J. Sharples and R. G. Lloyd, *NARes* **16,** 1541 (1988).
345a. J. B. Rafferty, S. E. Sedelnikova, D. Hargreaves, P. J. Artymiuk, P. J. Baker, G. J. Sharples, A. A. Mahdi, R. G. Lloyd and D. W. Rice, *Science* **274,** 415 (1996).
346. I. R. Tsaneva, B. Muller and S. C. West, *Cell* **69,** 1171 (1992).
347. T. Shiba, H. Iwasaki, A. Nakata and H. Shinagawa, *MGG* **237,** 395 (1993).
348. S. C. West, *Cell* **76,** 9 (1994).
349. A. Sancar and J. E. Hearst, *Science* **259,** 1415 (1993).
350. B. Müller, I. R. Tsaneva and S. C. West, *JBC* **268,** 17179 (1993).
351. I. R. Tsaneva and S. C. West, *JBC* **269,** 26552 (1994).
352. A. H. Mitchell and S. C. West, *JMB* **243,** 208 (1994).
353. P. E. Marrione and M. M. Cox, *Bchem* **34,** 9809 (1995).
354. I. R. Tsaneva, B. Muller and S. C. West, *PNAS* **90,** 1315 (1993).
355. A. Stasiak, I. R. Tsaneva, S. C. West, C. J. Benson, X. Yu and E. H. Egelman, *PNAS* **91,** 7618 (1994).
356. J. Kuriyan and M. O'Donnell, *JMB* **234,** 915 (1993).
357. F. B. Dean and J. Hurwitz, *JBC* **266,** 5062 (1991).
358. F. B. Dean, J. A. Borowiec, T. Eki and J. Hurwitz, *JBC* **267,** 14129 (1992).
359. Y. Wang and P. H. von Hippel, *JBC* **268,** 13947 (1993).
360. J. Geiselmann, T. D. Yager and P. H. von Hippel, *Protein Sci.* **1,** 861 (1992).
361. J. Geiselmann, S. E. Seifried, T. D. Yager, C. Liang and P. H. von Hippel, *Bchem* **31,** 121 (1992).
362. J. Geiselmann, T. D. Yager, S. C. Gill, P. Calmettes and P. H. von Hippel, *Bchem* **31,** 121 (1992)
363. J. Geiselmann, Y. Wang, S. E. Seifried and P. H. von Hippel, *PNAS* **90,** 7754 (1993).
364. S. E. Seifried, J. B. Easton, and P. H. von Hippel, *PNAS* **89,** 10454 (1992).
365. W. Bujalowski, M. M. Klonowska, and M. J. Jezewska, *JBC* **269,** 31350 (1994).
366. K. Hiom and S. C. West, *Cell* **80,** 787 (1995).

367. C. A. Parsons, A. Stasiak, R. J. Bennett and S. C. West, *Nature (London)* **374,** 375 (1995).
368. B. Müller and S. C. West, *Experientia* **50,** 216 (1994).
369. L. E. Iype, E. A. Wood, R. B. Inman and M. M. Cox, *JBC* **269,** 24967 (1994).
370. L. E. Iype, R. B. Inman and M. M. Cox, *JBC* **270,** 19473 (1995).
371. A. Segall, M. J. Mahan and J. R. Roth, *Science* **241,** 1314 (1988).
372. D. E. Adams, I. R. Tsaneva and S. C. West, *PNAS* **91,** 9901 (1994).
373. R. G. Lloyd and G. J. Sharples, *NARes* **21,** 1719 (1993).
374. M. C. Whitby, S. D. Vincent and R. G. Lloyd, *EMBO J.* **13,** 5220 (1994).
375. L. J. Roman and S. C. Kowalczykowski, *JBC* **264,** 18340 (1989).
376. L. J. Roman, D. A. Dixon and S. C. Kowalczykowski, *PNAS* **88,** 3367 (1991).
377. S. T. Lovett and R. D. Kolodner, *PNAS* **86,** 2627 (1989).
378. S. T. Lovett and R. D. Kolodner, *J. Bact.* **173,** 353 (1991).
379. S. T. Lovett and A. J. Clark, *J. Bact.* **157,** 190 (1984).
380. S. T. Lovett and A. J. Clark, *J. Bact.* **162,** 280 (1985).
381. S. T. Lovett, D. C. Luisi and R. D. Kolodner, *Genetics* **120,** 37 (1988).
382. S. E. Corrette-Bennett and S. T. Lovett, *JBC* **270,** 6881 (1995).
383. I. R. Lehman and A. L. Nussbaum, *JBC* **239,** 2628 (1964).
384. S. R. Kushner, H. Nagaishi, A. Templin and A. J. Clark, *PNAS* **68,** 824 (1971).
385. M. Defais, P. Fauquet, M. Radman and M. Errera, *Virology* **43,** 495 (1971).
386. D. W. Mount, K. B. Low and S. J. Edmiston, *J. Bact.* **112,** 886 (1972).
387. M. Radman, *in* "Molecular and Environmental Aspects of Mutagenesis" (L. Prakash, F. Sherman, M. Miller, C. Lawrence and H. W. Tabor, eds.), p. 128. Charles C. Thomas Publ., Springfield, Illinois, 1974.
388. J. W. Roberts, C. W. Roberts and N. L. Craig, *PNAS* **75,** 4714 (1978).
389. J. W. Little, S. H. Edmiston, L. Z. Pacelli and D. W. Mount, *PNAS* **77,** 3225 (1980).
390. G. C. Walker, *ARB* **54,** 425 (1985).
391. G. C. Walker, *in "Escherichia coli* and *Salmonella typhimurium:* Cellular and Molecular Biology" (F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter and H. E. Umbarger, eds.), p. 1346. American Society for Microbiology, Washington, D.C., 1987.
392. R. Devoret, *Ann. l'Inst. Pasteur Actual.* **1,** 11 (1992).
393. E. C. Friedberg, G. C. Walker and W. Siede, "DNA Repair and Mutagenesis." ASM Press, Washington, D.C., 1995.
394. X. Yu and E. H. Egelman, *JMB* **231,** 29 (1993).
395. C. Lu and H. Echols, *JMB* **196,** 497 (1987).
395a. W. M. Rehrauer, P. E. Lavery, E. L. Palmer, R. N. Singh and S. C. Kowalczykowski, *JBC* **271,** 23865 (1996).
395b. F. G. Harmon, W. M. Rehrauer and S. C. Kowalczykowski, *JBC* **271,** 23874 (1996).
396. S. Murli and G. C. Walker, *Curr. Opin. Genet. Dev.* **3,** 719 (1993).
397. Z. Livneh, F. O. Cohen, R. Skaliter and T. Elizur, *Crit. Rev. Biochem. Mol. Biol.* **28,** 465 (1993).
398. G. C. Walker, *Trends Biochem. Sci.* **20,** 416 (1995).
399. M. Rajagopalan, C. Lu, R. Woodgate, M. O'Donnell, M. F. Goodman and H. Echols, *PNAS* **89,** 10777 (1992).
400. H. Shinagawa, H. Iwasaki, T. Kato and A. Nakata, *PNAS* **85,** 1806 (1988).
401. S. E. Burckhardt, R. Woodgate, R. H. Scheuermann and H. Echols, *PNAS* **85,** 1811 (1988).
402. R. Woodgate, M. Rajagopalan, C. Lu and H. Echols, *PNAS* **86,** 7301 (1989).
403. M. Dutreix, B. Burnett, A. Bailone, C. M. Radding and R. Devoret, *MGG* **232,** 489 (1992).
404. J. B. Sweasy, E. M. Witkin, N. Sinha and V. Roegner-Maniscalco, *J. Bact.* **172,** 3030 (1990).
405. E. G. Frank, J. Hauser, A. S. Levine and R. Woodgate, *PNAS* **90,** 8169 (1993).
406. J. W. Zyskind, A. L. Svitil, W. B. Stine, M. C. Biery and D. W. Smith, *Mol. Microbiol.* **6,** 2525 (1992).

407. K. Skarstad and E. Boye, *J. Bact.* **175,** 5505 (1993).
408. T. R. Magee, T. Asai, D. Malka and T. Kogoma, *EMBO J.* **11,** 4219 (1992).
409. T. Asai, S. Sommer, A. Bailone and T. Kogoma, *EMBO J.* **12,** 3287 (1993).
410. T. Asai and T. Kogoma, *J. Bact.* **176,** 7113 (1994).
411. T. Asai and T. Kogoma, *J. Bact.* **176,** 1807 (1994).
412. S. C. West and B. Connolly, *Mol. Microbiol.* **6,** 2755 (1992).
413. K. Ishimori, S. Sommer, A. Bailone, M. Takahashi, M. M. Cox and R. Devoret, *JMB* (1996). Submitted.
414. S. Cheng, A. Sancar and J. E. Hearst, *NARes* **19,** 657 (1991).
415. S. Karlin and G. Ghandour, *PNAS* **82,** 8597 (1985).
416. S. Henikoff and J. G. Henikoff, *PNAS* **89,** 10915 (1992).
417. J. Devereux, P. H. Haeberli and O. S. Smithies, *NARes* **12,** 387 (1984).
418. D. F. Feng and R. F. Dolittle, *J. Mol. Evol.* **25,** 351 (1987).
419. T. Horii, T. Ogawa and H. Ogawa, *PNAS* **77,** 313 (1980).
420. J. Felsenstein, *Am. J. Human Genet.* **25,** 471 (1973).
421. D. S. Goodsell and A. J. Olson, *J. Mol. Graphics* **10,** 235 (1992).
422. A. White, P. Handler and E. L. Smith, "Principles of Biochemistry," 3rd Ed., p. 124. McGraw-Hill, New York, 1964.
423. F. M. Richards, *Annu. Rev. Biophys. Bioeng.* **61,** 151 (1977).
424. T. E. Ferrin, C. C. Huang, L. E. Jarvis and R. Langridge, *J. Mol. Graphics* **6,** 13 (1988).
425. A. Nicholls, K. A. Sharp and B. Honig, *Proteins Struct. Funct. Genet.* **11,** 281 (1991).
426. M. Gribskov and D. Eisenberg, *in* "Techniques in Protein Chemistry" (T. E. Hugli, ed.), p. 108. Academic Press, New York, 1989.
427. H. Fujisawa, T. Yonesaki and T. Minagawa, *NARes* **13,** 7473 (1985).
428. P. J. Kraulis, *J. Appl. Crystallog.* **24,** 946 (1991).