DNA Recombination and Repair in the Archaea

ERICA M. SEITZ, CYNTHIA A. HASELTINE, AND
STEPHEN C. KOWALCZYKOWSKI
Sections of Microbiology and of Molecular and Cellular Biology
Center for Genetics and Development
University of California, Davis
Davis, California 95616-8665

I. Introduction

DNA is subjected daily to considerable environmental and endogenous damage, which challenges both the integrity of the essential information that it contains and its ability to be transferred to future generations. All cells, however, are prepared to handle damage to the genome through an extensive DNA repair system, thus underscoring the importance of this process in cell survival. The Archaea represent a rather diverse group of organisms, including many members who thrive under conditions that would be lethal for most bacteria and eukaryotes. These conditions, such as extreme temperatures, also present a new challenge to the Archaea and to their genomes, reinforcing the need to possess an efficient DNA repair system (DiRuggiero *et al.*, 1999; Grogan, 2000). This, and the fact that the Archaea are a largely unexplored domain of life prompted

ADVANCES IN APPLIED MICROBIOLOGY, VOLUME 50

0065-2164/01 $35.00

interest in the types of DNA repair mechanisms that operate within this domain.

Studies carried out in bacteria, especially in *Escherichia coli,* or in eukaryotes, particularly in the yeast *Saccharomyces cerevisiae,* revealed much of what is known about these processes. These studies showed that DNA repair occurs by several pathways (Lindahl and Wood, 1999); these include reversal of DNA damage, excision of damaged nucleotides (nucleotide excision repair or NER) or bases (base excision repair or BER), excision of misincorporated nucleotides (mismatch repair or MMR), and recombinational repair (Friedberg *et al.,* 1995). Although relatively little was known about DNA repair in Archaea, the recent sequencing of several archaeal genomes permitted the identification of structural homologues of many proteins involved in these different pathways. In this article, we review the most important features of DNA repair learned from studies of organisms such as *E. coli* and *S. cerevisiae.* In particular, we emphasize the elements which have been conserved throughout evolution, either at the level of global mechanisms or at the level of the protein effectors. We apply this knowledge to the third domain of life, the Archaea, and review what is known about DNA repair in this domain of life, with a specific emphasis on recombinational repair.

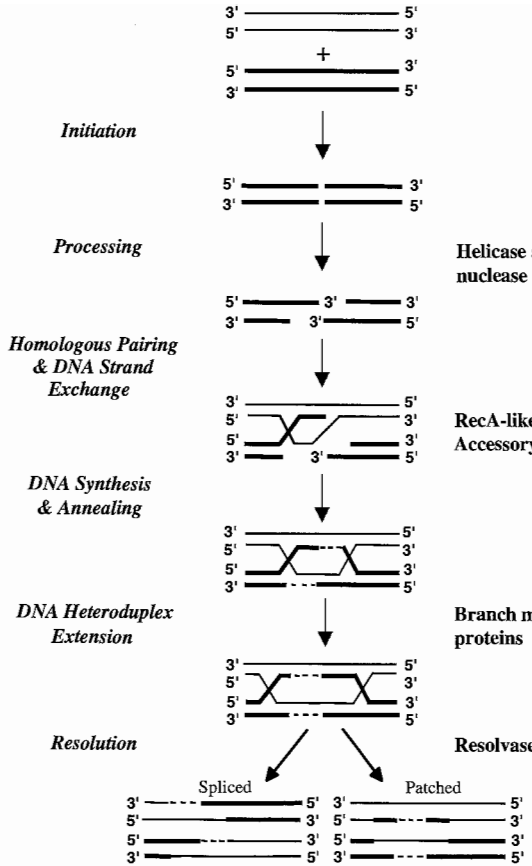
II. Recombinational Repair

One of the most serious types of damage that can be inflicted on the genome is a DNA break either in a single strand or in both strands of DNA [a double-stranded DNA (dsDNA) break; DSB]. DNA breaks of any type pose a particularly significant problem to the cell because they challenge the integrity of the DNA molecule and can lead, if not repaired, to loss of information, gross chromosomal rearrangements, and chromosome missegregation. Because of these potentially lethal consequences, both bacterial and eukaryal organisms have mechanisms for repairing this type of DNA lesion, although the manner by which each repairs the lesion differs. In the Bacteria, this type of damage is remedied primarily by the process of homologous DNA recombination (Kowalczykowski *et al.,* 1994; Kuzminov, 1999), whereas in the Eukarya, the DSB is repaired by either homologous recombination or nonhomologous end joining (NHEJ) (Pâques and Haber, 1999; Sung *et al.,* 2000). Recombination involves pairing of the damaged DNA with a homologous partner to copy any lost information from the homologue, thereby accurately repairing the DSB, whereas NHEJ involves ligation of the DSB without the need for significant homology, thus being inherently error-prone. Here we focus on DSB repair by homologous recombination, as NHEJ appears to be a uniquely eukaryal process.

A. An Overview of Homologous Recombination

Homologous DNA recombination is a primary means for the repair of DSBs. Although the general mechanism is similar in bacteria and eukaryotes, the proteins that are involved in this process differ (Fig. 1). Figure 1 depicts the DSB repair model (Resnick, 1976; Szostak *et al.*, 1983) and the likely proteins that act at each step. After DSB formation, both ends of the break are resected to create single-stranded DNA (ssDNA), which then invades a homologous dsDNA molecule. After DNA strand invasion occurs, the 3′ ends of the invading strands serve as primers for the initiation of nascent DNA synthesis, which leads to the formation of two Holliday junctions that are cleaved in one of two orientations to generate two types of recombinant molecules (Fig. 1).

Biochemical studies have revealed the function of many enzymes that participate in the process of homologous recombination. In *E. coli*, it was determined that the process of homologous recombination involves the action of more than 25 different proteins (Kowalczykowski *et al.*, 1994). Figure 1 shows some of the enzymes from *E. coli* and *S. cerevisiae* that act at each step in this process (Kowalczykowski *et al.*, 1994; Pâques and Haber, 1999) and for which there are, or may be, either structural or functional homologues in the Archaea. The first step in the homologous DNA recombination pathway is an initiation or processing step, which involves processing of the broken DNA molecule so that a region with a partially ssDNA character is generated. This processing can be accomplished through the action of DNA helicases, nucleases, or both. The next step corresponds to the search for the homologous target DNA molecule, which is immediately followed by the exchange of their DNA strands. This step is accomplished by DNA strand exchange proteins, which bind to the ssDNA that was generated previously. The resultant nucleoprotein filament is the active form of these proteins, which acts both in the homology search process and in the invasion of the recipient DNA molecule. The consequence of this initial pairing event is a region of newly paired or heteroduplex DNA, which is also known as a joint molecule (Kowalczykowski and Eggleston, 1994). The third step involves the reciprocal exchange of the two DNA strands, creating a four-stranded structure known as a Holliday junction. The regions of heteroduplex DNA are extended by protein-promoted branch migration, which involves the action of either the DNA strand exchange protein or a specialized DNA helicase. The final step involves symmetric cleavage of the Holliday junction in one of two orientations by a Holliday junction-specific endonuclease to produce one of two alternative recombinant products (Kowalczykowski *et al.*, 1994; West, 1994a,b; White *et al.*, 1997; Lilley and White, 2000). Despite differences between the



	E. coli	*S. cerevisiae*	Archaea
	---	Spo11	Spo11[1]
Helicase and/or nuclease	RecBCD	---	---
	SbcCD	Mre11/Rad50/Xrs2	Mre11/Rad50
	RecQ	Sgs1[2]	Sgs1[2,3]
RecA-like protein and Accessory proteins	RecA	Rad51	RadA
	SSB	RPA	RPA/SSB
	---	Rad55/57	RadB
	---	Rad54	Rad54[3]
	RecO(R)	Rad52	---
	RecF(R)	---	---
Branch migration proteins	RuvAB	Rad54[4]	Rad54[4]
	RecG		
Resolvase	RuvC	?[5] Cce1 (mitochondrial)	Hjc/ Hje[6]

well-studied bacterial (namely, *E. coli*) and eukaryal systems (namely, *S. cerevisiae*), these basic steps remain mostly conserved.

1. Bacterial Homologous DNA Recombination

E. coli possesses two pathways for the repair of DNA strand breaks (Kowalczykowski *et al.*, 1994; Kuzminov, 1999): the RecBCD pathway, which repairs DSBs; and the RecF pathway, which repairs primarily single-strand gaps but can repair DSBs as well. Both of these pathways for recombinational repair depend on the action of the RecA protein. In the RecBCD pathway, the RecBCD helicase/nuclease both processes the DSB to create ssDNA and loads RecA protein onto this ssDNA in anticipation of DNA strand exchange. In the RecF pathway, RecQ helicase processes the broken DNA molecule to produce ssDNA, and the RecO and RecR proteins aid in loading RecA protein onto the ssDNA by mediating the removal of ssDNA binding (SSB) protein (Umezu *et al.*, 1993; Harmon and Kowalczykowski, 1998; Kuzminov, 1999).

2. Eukaryal Homologous DNA Recombination

Homologous DNA recombination is studied in the Eukarya most extensively with the yeast, *S. cerevisiae,* but recent studies in mammals demonstrate the commonality of this eukaryotic process (Pâques and Haber, 1999). As discussed later, some parallels can be drawn between the yeast and the bacterial systems, but for the most part, the system in yeast exists as a more complex process. The repair of DSBs by homologous recombination requires members of the yeast *RAD52* epistasis group, which consists of *RAD50, RAD51, RAD52, RAD54, RAD55, RAD57, RAD59, MRE11, XRS2,* and *RDH54/TID1* genes (Game, 1993; Pâques and Haber, 1999). The function of the proteins encoded by these genes has been studied both genetically and biochemically, but the precise function of some proteins is not yet fully understood (Fig. 1).

3. Archaeal Homologous DNA Recombination

The genome sequences of several archaeons has made it possible to identify structural homologues of many proteins involved in the process of homologous DNA recombination. In addition, some of these proteins

FIG. 1. Mechanism for double-stranded DNA break repair by homologous recombination, and the proteins involved. Shown are the proteins that are either known or proposed to act in each step of this process in *E. coli, S. cerevisiae,* and the Archaea. Notes: [1]The archaeal Spo11 protein is a subunit of TopoVI, and a direct role in DSB formation is not clearly defined. [2]A role for Sgs1 in initiation is unclear. [3]Assignment is based only on sequence homology. [4]The Rad54 protein is not a structural homologue of either the RuvAB or the RecG protein; however, it will promote DNA heteroduplex extension (J. Solinger *et al,* in press). [5]? refers to the fact that an activity has been found in human cells but the responsible protein is unknown. [6]Hje refers to an activity only; the protein has not been identified.

have been studied biochemically, and there is some genetic evidence supporting the role of these genes in archaeal homologous DNA recombination. Evidence for stimulation of chromosomal marker exchange in the hyperthermophilic archaeon *Sulfolobus acidocaldarius* provides evidence for DNA repair, conjugation, and homologous recombination processes in these organisms (Schmidt *et al.*, 1999). Figure 1 and Table I present mainly the proteins involved in this process for which homologues have been found in the Eukarya and Archaea. For the most part, the proteins identified in the Archaea show greater structural and, in some cases, functional, similarity to eukaryal proteins than to their bacterial counterparts (Fig. 1).

B. Generation of DNA Breaks

DNA breaks can occur either in a single DNA strand, creating ssDNA gaps, or in both strands, DSBs. There are many routes for production of ssDNA gaps or DSBs, but DNA replication is a major mechanism for converting ssDNA lesions into larger gaps or DSBs (Kogoma, 1997; Kuzminov, 1999; Kowalczykowski, 2000; Michel, 2000). As illustrated in Fig. 2, ssDNA gaps can be created if a blocking lesion is not removed by repair processes prior to the arrival of the DNA replication machinery. If the lesion is on the lagging strand template, then Okazaki fragments cannot be joined; if the lesion is on the leading strand, then the replication fork halts and may initiate farther downstream. In either case, a region of single-stranded, unreplicated DNA is created. Lesions having the ability to halt the progression of replicative DNA polymerases are numerous and include the well-studied 6–4 thymine photoproducts and cyclobutane pyrimidine dimers caused by ultraviolet (UV) light (Edenberg, 1976).

DSBs can arise from several sources. Exogenously, DSBs are caused by ionizing radiation such as X-rays or γ rays or by various radiomimetic chemicals. Endogenously, DSBs can be created directly by reactive oxygen species and can also arise as a consequence of replicating a nicked DNA template (Fig. 2). Indeed, if a DNA replication fork encounters an interruption (nick or ssDNA gap) in one of the two DNA strands, this interruption will be converted to a DSB (Kuzminov, 1999; Pâques and Haber, 1999; Kowalczykowski, 2000). Nicks in DNA can result from numerous sources, including unsealed Okazaki fragments on the lagging strand and incision of a damaged DNA strand by another repair system, such as either nucleotide or base excision repair. DSBs can also be created as a consequence of the replication apparatus stalling or halting. Stalling can occur, for example, due to the presence of a chemical imperfection in the DNA or a protein complex tightly bound to DNA, either of which can block the progression of the fork. The stalled DNA replication

TABLE I

ARCHAEAL RECOMBINATION PROTEIN HOMOLOGUES[a]

	RadA paralogue											
	Mre11	Rad50	Xrs2	RPA/SSB	RadA	RadB	Other	Rad52	Rad54[b]	Rad59	Spo11	Hjc[c]
A. fulgidus	+	+	−	R	+	+	−	−	?	−	+	+
Halobacterium	+	+	−	R	+	+	−	−	?	−	+	?
M. jannaschii	+	+	−	R	+	+	?[d]	−	−	−	+	+
M. thermoautotrophicum	+	+	−	R	+	+	−	−	−	−	+	+
P. abyssii	+	+	−	R	+	+	−	−	?	−	+	?
P. furiosus	+	+	−	R	+	+	−	−	?	−	+	+
P. horikoshii	+	+	−	R	+	+	−	−	?	−	+	+
A. pernix	+	+	−	S	+	+	−	−	?	−	+	+
S. solfataricus	+	+	−	S	+	+	−	−	+[e]	−	+	+

[a] The potential recombination protein homologues from nine fully sequenced archaeal genomes are represented. A + indicates the presence of a single homologous protein sequence, while a − represents the failure to detect a homologue. A ? is shown for single protein sequences where a homologue may be present, but sufficiently high levels of homology to permit confident assignment are not apparent. An R represents the presence of an RPA-like structural protein homologue, while an S represents the presence of an SSB-like structural protein homologue.

[b] A single protein sequence with limited homology to Rad54 was identified in each of the organisms indicated by a ?.

[c] A Holliday-junction endonuclease activity distinct from Hjc was found *S. shibatae* and *S. solfataricus* and is called Hje.

[d] A single protein sequence was identified with homology to Rad55.

[e] This homologue is based on sequence similarity only.

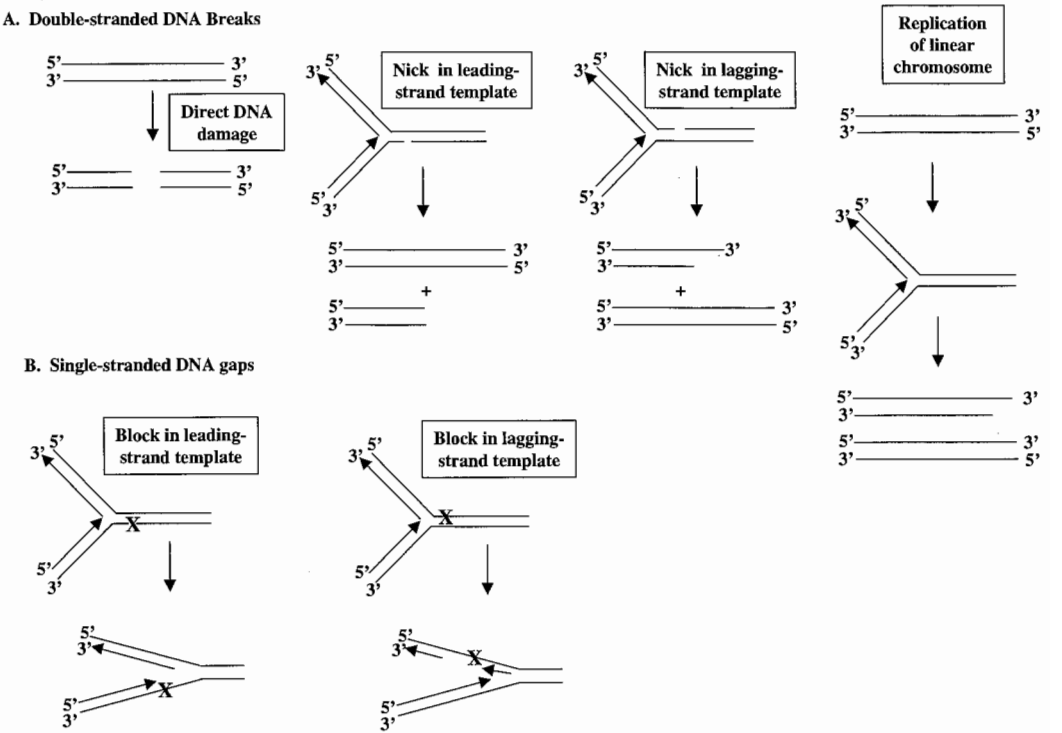


FIG. 2. Single-stranded DNA gaps or double-stranded DNA breaks are formed by DNA damaging agents and by DNA replication through the lesion. Depicted is the production of (A) a double-stranded DNA break, formed either by DNA damaging agents directly or by replication through a nicked template, and (B) a single-stranded DNA gap, formed by replication stopping at the lesion. Both ssDNA gaps and dsDNA breaks can be repaired by homologous recombination. [Adapted from Kowalczykowski (2000).]

forks must be restarted for the replication of the genome to be completed. This restart can be achieved through the introduction of a DSB at the regressed replication fork, followed by recombination-dependent replication (Kogoma, 1997; Michel *et al.*, 1997; Kuzminov, 1999; Pâques and Haber, 1999; Kowalczykowski, 2000; Marians, 2000; Michel, 2000).

In addition to these general mechanisms for DSB formation, DSBs in the Eukarya are also produced in a programmed and specific manner. For example, in meiotic cells, DSBs are enzymatically introduced during the initiation phase of meiosis, to ensure the crossing-over of homologues needed for their faithful segregation (Keeney *et al.*, 1997; Haber, 2000a,b).

1. DSBs in the Bacteria

In *E. coli,* DNA replication initiates at the chromosomal origin, OriC, and progresses bidirectionally along the two arms of the circular chromosome toward the replication terminus. The majority of these replication forks encounters an obstacle to their progression, leading to their stalling (Kogoma, 1997; Michel *et al.*, 1997; Kuzminov, 1999). These obstacles can be chemical lesions, DNA-bound protein complexes, or secondary DNA structure. Regardless of the obstacle, complete replication of the chromosome requires the origin-independent restart of the stalled replication fork. DNA recombination is responsible for this restart (Kogoma, 1996). Recent studies indicate that the first step in this process involves regression of the replication fork by reannealing of the two newly synthesized DNA strands after replication fork arrest. This creates an X-shaped Holliday junction that contains one accessible dsDNA end (Postow *et al.*, 2000; Flores *et al.*, 2001). The RecG protein, a DNA helicase involved in homologous recombination, can catalyze such Holliday junction formation by replication fork reversal (McGlynn and Lloyd, 2000). At this stage, this intermediate can be processed in either of two ways. The RecBCD enzyme, an enzyme involved in the initiation of DNA recombination in bacteria (see below), is a dsDNA nuclease that acts on the DSB created at the Holliday junction (which was formed by replication fork reversal) and starts degrading the DNA. This nucleolytic action effectively shortens the two newly synthesized strands and allows the replication fork to move back from the point where it initially stalled, giving it another opportunity to progress past the previous block after it reinitiates. Alternatively, the regressed replication fork/Holliday junction can be recognized and cleaved by the RuvABC complex to produce a DSB (Michel *et al.*, 1997; Seigneur *et al.*, 1998). The RuvAB complex is involved in the branch migration of Holliday junctions, and RuvC is an endonuclease that specifically cleaves these junctions, as discussed in more detail below. The DSB is then repaired by homologous recombination and is used to restart

replication through the action of the PriA protein, which links recombination and replication restart (Kogoma, 1996, 1997; Kowalczykowski, 2000; Marians, 2000; Michel, 2000; Sandler and Marians, 2000).

2. DSBs in the Eukarya

The importance of the above findings is underlined by the fact that sites which are known to block DNA replication in mitotic eukaryal cells promote chromosomal instability due to an increased frequency of homologous recombination, suggesting that the relationship between replication blockage and recombination-dependent replication fork restart is universal (Rothstein *et al.,* 2000). In yeast cells undergoing meiosis, DSBs have long been observed to coincide with known meiotic recombination hot spots (Nicolas *et al.,* 1989; Sun *et al.,* 1989; Debrauwere *et al.,* 1999). These meiotic DSBs were mapped at nucleotide resolution along the entire length of chromosome III and were found to cluster in intergenic promoter-containing intervals, but their occurrence did not require transcription (Baudat and Nicolas, 1997; Borde *et al.,* 1999). Because some breaks were found to have the Spo11 protein covalently linked to the 5′ ends of the break sites (Liu *et al.,* 1995; Keeney *et al.,* 1997), it was hypothesized that this protein is the endonuclease responsible for the formation of the meiotic DSB. Mutation of a conserved tyrosine residue in this protein (the residue that attacks the phosphodiester bond and results in a transient covalent DNA–protein complex) eliminated the DSBs and meiotic recombination (Bergerat *et al.,* 1997). Following this discovery, Spo11 homologues were discovered in *Schizosaccharomyces pombe, Drosophila melanogaster, Caenorhabditis elegans,* and *Mus musculus* and were found to be essential for meiotic recombination (Dernburg *et al.,* 1998; McKim and Hayashi-Hagihara, 1998; Celerin *et al.,* 2000; Cervantes *et al.,* 2000). In mice, knockouts of the *Spo11* gene result in drastic gonadal abnormalities due to defective meiosis, and this gene is additionally required for meiotic synapsis (Baudat *et al.,* 2000; Romanienko and Camerini-Otero, 2000). Overall, these studies demonstrate that homologous DNA recombination during meiosis is initiated by the formation of specific DSBs. Recent results demonstrate that the formation of these breaks in yeast is carefully controlled by the cell and is coupled to the last round of meiotic DNA replication (Borde *et al.,* 2000).

3. Spo11 in the Archaea

An archaeal type II topoisomerase from the hyperthermophile *Sulfolobus shibatae* that showed homology to the *S. cerevisiae* Spo11 protein was discovered and is referred to as topoisomerase VI (TopoVI) (Bergerat *et al.,* 1994, 1997). TopoVI is a type II topoisomerase, and these enzymes help regulate DNA topology during transcription, replication,

and recombination by catalyzing DNA strand transfer through transient DSBs. This particular topoisomerase is composed of two subunits, A and B, and defines a new family of topoisomerases. The A subunit showed significant homology to the Spo11 protein in *S. cerevisiae* and to the Spo11 homologue in *S. pombe,* the Rec12 protein. Upon inspection of the nine fully sequenced archaeal genomes, we identified several additional homologues, and Fig. 3 shows an alignment of these proteins from eight archaeal organisms. A Spo11 protein homologue was not found in *P. furiosus.* Overall, these proteins share 28–35% similarity to the *S. cerevisiae* Spo11 protein, and each has five conserved DNA gyrase motifs, labeled I–V (Figs. 3 and 4). The *S. shibatae* TopoVI can relax both positive and negative supercoils and has a strong decatenase activity, implying a function in the maintenance of chromosome topology (Bergerat *et al.,* 1997).

C. Initiation of Homologous DNA Recombination: DSB End Processing

After the formation of a DSB, processing of the DNA ends must occur to create a suitable substrate for the next step in homologous recombination, which is catalyzed by a DNA strand exchange protein (Fig. 1). In *E. coli,* the RecBCD enzyme is responsible for this end-processing event (for review, see Kowalczykowski *et al.,* 1994; Kuzminov, 1999; Arnold and Kowalczykowski, 1999), but in the Eukarya and Archaea the mechanism by which this initial processing event occurs is largely unknown. There are, however, enzymes involved in some aspect of DNA end processing that are homologous between the Eukarya and the Archaea; these are the Rad50 and Mre11 proteins (Pâques and Haber, 1999; Sung *et al.,* 2000), which, interestingly, also share homology with a DNA nuclease in *E. coli,* comprised of the SbcC and SbcD proteins (Sharples and Leach, 1995).

The RecBCD enzyme is not the only protein capable of initiating recombination in *E. coli.* In a *recBC*$^-$ *sbcBC*$^-$ background, recombination proceeds by an alternate pathway known as the RecF pathway. In the absence of the RecBCD enzyme, another helicase, RecQ, processes the DSB (Clark and Sandler, 1994; Mendonca *et al.,* 1995). Interestingly, the Eukarya also have structural homologues of the RecQ helicase; in *S. cerevisiae* it is the Sgs1 protein, and it also affects recombination, but its precise function is unclear (Gangloff *et al.,* 1994; Watt *et al.,* 1995). In humans, there are five proteins that in their conserved helicase domains show significant amino acid similarity to the *E. coli* RecQ helicase: Blm, Wrn, RecQL, RecQ4, and RecQ5 (Puranam and Blackshear, 1994; Seki *et al.,* 1994; Ellis *et al.,* 1995; Yu *et al.,* 1996; Kitao *et al.,* 1998; Shen and Loeb, 2000). Mutations at the *BLM, WRN,* and *RECQ4* loci lead to Bloom's, Werner's, or Rothmund–Thomson

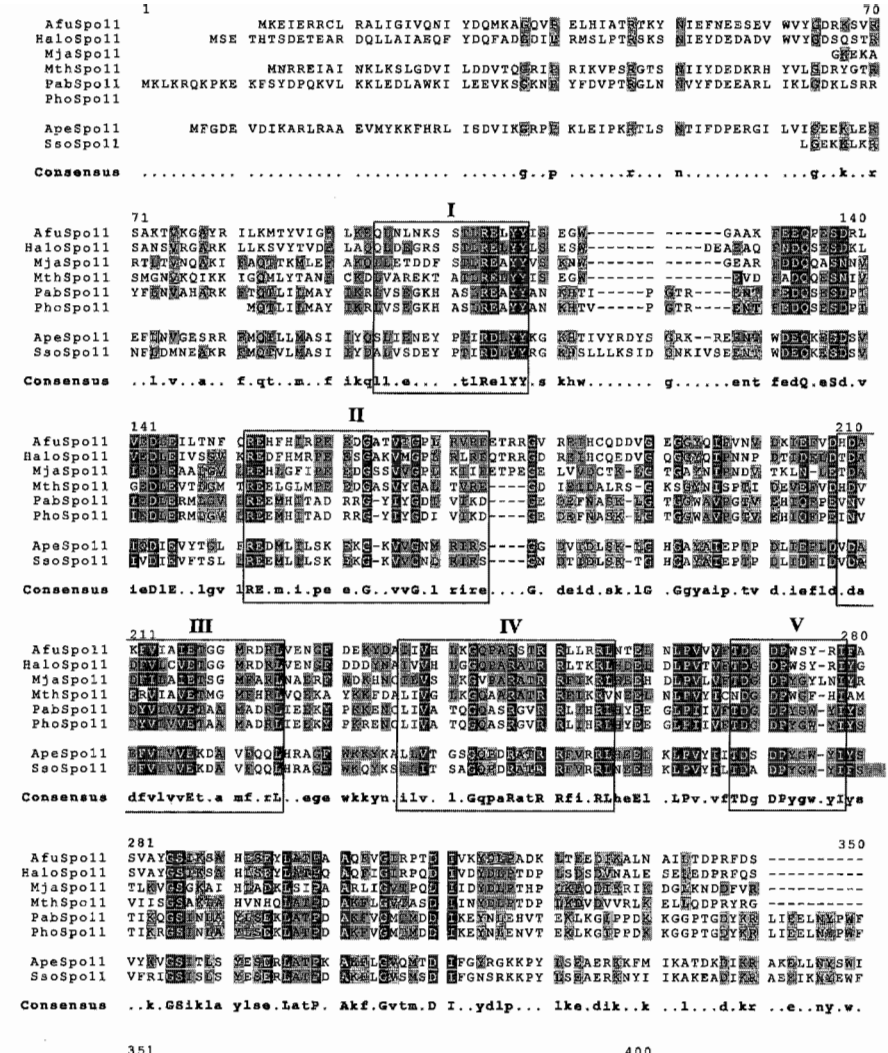


FIG. 3. Multiple alignment of archaeal Spo11 protein homologues. Sequences were as follows: *A. fulgidus* (Afu), gi2649657; *Halobacterium* sp. NRC-1 (Halo), gi10580448; *M. jannaschii* (Mja), mj0369; *M. thermoautotrophicum* (Mth), gi2622109; *P. abyssii* (Pab), gi5458027; *P. horikoshii* (Pho), ph1563; *A. pernix* (Ape), gi5104364; and *S. solfataricus* (Sso), bac04_042. The sequences were aligned using MULTALIN at http://www.toulouse.inra.fr/multalin.html. Highly conserved residues are shaded in dark gray, while moderately conserved residues are shaded in light gray. DNA gyrase motifs I–V are indicated.

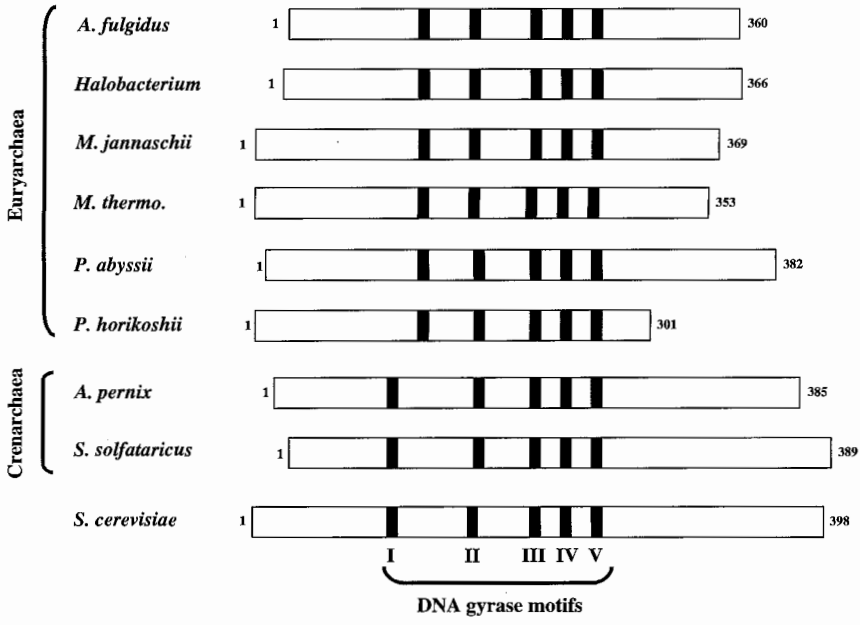


FIG. 4. Schematic representation of archaeal Spo11 protein homologues. Also shown, for comparison, is the *S. cerevisiae* Spo11 protein. DNA gyrase motifs I–V are indicated.

syndromes, respectively, which are rare, inherited diseases that result in DNA replication abnormalities and genomic instability (Kitao *et al.*, 1999a,b; Chakraverty and Hickson, 1999; Shen and Loeb, 2000). Interestingly, a member of the RecQ helicase family was identified in the crenarchaeote *A. pernix* (Kawarabayasi *et al.*, 1999).

1. Bacterial RecBCD-like Enzymes

DNA processing in wild-type *E. coli* is carried out by the RecBCD enzyme, a heterotrimeric protein complex that possesses DNA helicase activity, as well as dsDNA and ssDNA exonuclease activities (Kowalczykowski *et al.*, 1994; Arnold and Kowalczykowski, 1999; Kuzminov, 1999). The exonuclease activity of the RecBCD enzyme initially degrades DNA in a preferential 3′-to-5′ direction (Fig. 5). This destructive activity is regulated by the interaction of the RecBCD enzyme with an eight-nucleotide DNA hot-spot sequence called χ (Lam *et al.*, 1974; Smith *et al.*, 1980; Dixon and Kowalczykowski, 1993; Anderson and Kowalczykowski, 1997a; Bianco and Kowalczykowski, 1997). When the RecBCD enzyme encounters a properly oriented χ site, the 3′-to-5′ exonuclease activity is attenuated, while a weaker 5′-to-3′ exonuclease is activated (Fig. 5). Since the helicase activity

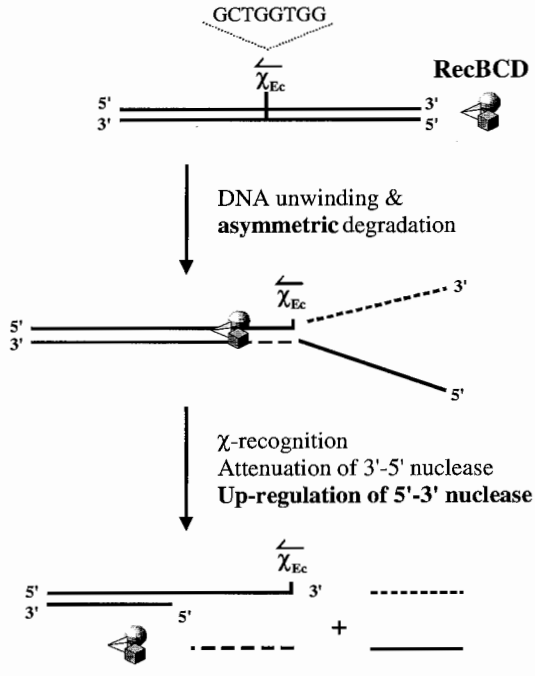


FIG. 5. RecBCD helicase/exonuclease activity is regulated by the recombination hot spot, χ. The RecBCD enzyme enters the DSB and both unwinds and degrades the DNA (the 3′-strand is degraded more extensively than the 5′-strand). Recognition of χ (5′-GCTGGTGG-3′) is followed by both attenuation of the 3′–5′ nuclease activity and a switch in the polarity of nuclease degradation (to 5′–3′), resulting in degradation of the opposite DNA strand. Also (not shown), the RecA protein is loaded by the RecBCD enzyme onto the χ-containing strand. [Adapted from Anderson and Kowalczykowski (1997a).]

is unaffected, these changes result in a switch in polarity of DNA strand degradation: before χ, the RecBCD enzyme preferentially degrades the 3′-ending strand, whereas after encountering a χ site, the RecBCD enzyme degrades the 5′-ending strand (Fig. 5) (Dixon and Kowalczykowski, 1993; Anderson and Kowalczykowski, 1997a). This processing results in a DNA molecule containing a 3′-ssDNA overhang, onto which the RecBCD enzyme also facilitates the loading of the RecA protein. The RecA nucleoprotein filament then promotes homologous pairing and DNA strand exchange (Anderson and Kowalczykowski, 1997b). Indeed, this facilitated loading of the RecA protein by the RecBCD enzyme is essential to the RecBCD-mediated recombination pathway (Arnold and Kowalczykowski, 2000). Functional homologues of the RecBCD enzyme exist in other bacteria, and although their mechanism of action differs somewhat, the net effect is to process DSBs into 3′-tailed ssDNA (Chédin *et al.*, 2000).

There is no known homologue of the RecBCD enzyme in either the Eukarya or the Archaea at this time, but a structural homologue of the SbcCD enzyme of *E. coli* exists in both of these phylogenetic domains (Fig. 1) (Connelly *et al.*, 1999). The SbcC and SbcD proteins form a complex that possesses ATP-independent ssDNA endonuclease and ATP-dependent dsDNA exonuclease activities (Connelly and Leach, 1996; Connelly *et al.*, 1997). The SbcC protein contains an ATP-binding motif, and the SbcD protein contains a nuclease domain. This complex can also recognize and cleave DNA hairpins (Connelly *et al.*, 1998, 1999; Cromie *et al.*, 2000).

2. E. coli RecQ Helicase

The RecQ helicase is responsible for processing DSBs in the absence of a functional RecBCD enzyme, and it functions in the RecF pathway of recombination. Null mutations in *recQ*, in combination with other mutations, result in a 100-fold reduction in homologous recombination proficiency and cause an increase in sensitivity to UV irradiation (Nakayama *et al.*, 1984, 1985). RecQ is a 3′-to-5′ DNA helicase that can initiate homologous recombination either at a DSB or at ssDNA regions (Lanzov *et al.*, 1991; Lloyd and Buckman, 1995) and can unwind a variety of DNA substrates, including intermediates formed by homologous pairing events (Harmon and Kowalczykowski, 1998). RecQ helicase, in the presence of RecA and SSB proteins, can also initiate homologous recombination *in vitro* (Harmon and Kowalczykowski, 1998). Another function for RecQ helicase comes from evidence that it acts together with topoisomerase III to control recombination (Harmon *et al.*, 1999).

3. Eukaryal Sgs1 Helicase

The *S. cerevisiae* Sgs1 helicase is a member of the RecQ helicase family that is involved in the segregation of chromosomes, control of aging, and regulation of recombination. Mutation of *SGS1* results in premature aging in yeast cells and the accumulation of extrachromosomal rDNA circles (Gangloff *et al.*, 1994; Watt *et al.*, 1995, 1996; Sinclair and Guarente, 1997; Saffi *et al.*, 2000). The Sgs1 protein is also a 3′-to-5′ helicase (Bennett *et al.*, 1998). Additionally, like the *E. coli* system, the Sgs1 protein interacts with *S. cerevisiae* TopoIII to control recombination events (Gangloff *et al.*, 1994; Bennett *et al.*, 2000; Duno *et al.*, 2000; Fricke *et al.*, 2000).

Five additional members of the RecQ helicase family exist in humans, and three are responsible for causing diseases, known as Werner's, Bloom's, and Rothmund–Thomson syndromes (Ellis *et al.*, 1995; Yu *et al.*, 1996; Kitao *et al.*, 1998,1999a,b). These diseases are characterized by the premature onset of aging and an increased incidence of chromosomal abnormalities (Epstein and Motulsky, 1996; Lindor *et al.*, 2000).

4. Archaeal Sgs1 Helicase

A putative Sgs1 protein homologue exists in the crenarchaeote *A. pernix* (gi5105033) (Kawarabayasi *et al.,* 1999). Searching the rest of the fully sequenced archaeal genomes has not yet resulted in convincing Sgs1 protein homologues. The *A. pernix* Sgs1 protein homologue is similar in size to the S. cerevisiae Sgs1 protein and shows 42% similarity to the S. cerevisiae Sgs1 protein and 47% similarity to the *E. coli* RecQ protein in the region containing the helicase domains.

5. Eukaryal MRE11/RAD50/ XRS2 (NBS1) Proteins

The genes involved in DNA end processing in *S. cerevisiae* are called *RAD50, MRE11,* and *XRS2,* and their gene products form a complex. This complex is involved in many DNA repair processes, which include homologous recombination, nonhomologous end joining, telomere maintenance, and the generation of DSBs in meiosis (Pâques and Haber, 1999; Sung *et al.,* 2000). The Rad50 protein shows homology to the *E. coli* SbcC protein, while the Mre11 protein shows homology to the *E. coli* SbcD protein (Sharples and Leach, 1995). The Rad50 protein is a member of a family of proteins called the *s*tructural *m*aintenance of *c*hromosomes (SMC) family (Hirano, 1999). This protein has ATP-dependent DNA binding and partial DNA unwinding activities (Raymond and Kleckner, 1993). Several mutations near the nucleotide binding site additionally cause defects in meiotic but not in mitotic DSB repair (Alani *et al.,* 1990). The Mre11 protein is homologous to a family of phosphodiesterases (Ogawa *et al.,* 1995). In accordance with this fact, both the *S. cerevisiae* and the human Mre11 proteins have ssDNA endonuclease activity and a 3′-to-5′ exonuclease activity (Furuse *et al.,* 1998; Paull and Gellert, 1998; Usui *et al.,* 1998). The Mre11 and Rad50 proteins from humans and yeast form a complex, which results in enhanced exonuclease activity. These proteins, like the bacterial SbcD protein, specifically require manganese for activation of nuclease activity (Furuse *et al.,* 1998). Processing of DSBs during meiotic recombination is dependent on the nuclease activity of Mre11, which is proposed to remove the DSB-promoting protein, Spo11, from the 5′ terminus of the DSB to which it is covalently attached (Sung *et al.,* 2000). The Rad50/Mre11 complex interacts with a third protein called Xrs2. This interaction takes place via the Mre11 subunit (Johzuka and Ogawa, 1995), although the role of Xrs2 in changing the function of the Mre11/Rad50 complex remains undefined.

In humans, the Rad50/Mre11 complex interacts with a third protein, called p95 or NBS1 (named due to its involvement in Nijemegen breakage syndrome) (Dolganov *et al.,* 1996). Although this third subunit appears to be analogous to the yeast Xrs2 protein, there is essentially

no sequence homology between these two proteins (Petrini, 1999). This third protein confers upon the complex the ability to open DNA hairpins efficiently, as well as an ATP-dependent endonuclease activity that acts on 3′-ssDNA tails adjacent to a duplex region (Paull and Gellert, 1999). This complex can also unwind duplex DNA to a limited extent, causing strand separation that is stimulated by ATP (Paull and Gellert, 1999).

6. Archaeal RAD50/MRE11 Proteins

Rad50 and Mre11 protein homologues exist in at least nine archaeons to date (Table I and Figs. 6–8). The archaeal Rad50 proteins share 30–38% similarity with the *S. cerevisiae* Rad50 protein and 5–13% similarity with *E. coli* SbcC protein and have conserved Walker-A and -B domains (Fig. 6). We also identified archaeal Mre11 protein homologues in each of the fully sequenced genomes available; these share 20–25% similarity with the *S. cerevisiae* Mre11 protein and 8–20% similarity with *E. coli* SbcD protein. The archaeal Mre11 proteins all contain the four domains that were proposed to be essential for nuclease activity (I–IV in Figs. 7 and 8). A homologue of either the Xrs2 or the NBS1 subunit has not yet been detected, raising the possibility that the Archaea lack this third subunit.

Mre11 (*pfMre11*) and *Rad50* (*pfRad50*) from the euryarchaeote *Pyrococcus furiosus* were recently cloned, and their gene products purified (Hopfner *et al.*, 2000a). This Mre11 homologue, the pfMre11 protein, showed sequence similarity with other members of the Mre11 protein family and had 29% identity and 42% similarity with the human Mre11 protein in the conserved N-terminal domains of the two proteins. The pfMre11 protein, alone, digests ssDNA in a Mn^{2+}-dependent manner. The *pfRad50* gene is located next to the *pfMre11* gene in the *P. furiosus* genome, which is similar to the genetic organization of the *E. coli sbcC* and *sbcD* genes. The pfRad50 protein displays only 19% homology to the human Rad50 protein, although the key residues of the Walker-A and -B ATP binding motifs are conserved between the pfRad50 protein and other members of this protein family (Hopfner *et al.*, 2000a).

The pfMre11 and pfRad50 proteins form a stable complex (pfMRE11/Rad50), which can digest linear plasmid DNA in an ATP-dependent manner. pfMRE11/Rad50 shows 3′-to-5′ ssDNA exonuclease activity, and this activity is ATP dependent, like the bacterial SbcCD complex and the eukaryal Mre11/Rad50 complex. These activities were observed at elevated temperatures of 50°C (Hopfner *et al.*, 2000a). The high-resolution X-ray crystal structures of the ATP-bound and ATP-free Rad50 catalytic domains were determined for *pfRad50*. The two Rad50 catalytic domains associate in an ATP-dependent manner and form a putative DNA binding groove at the interface of this interaction

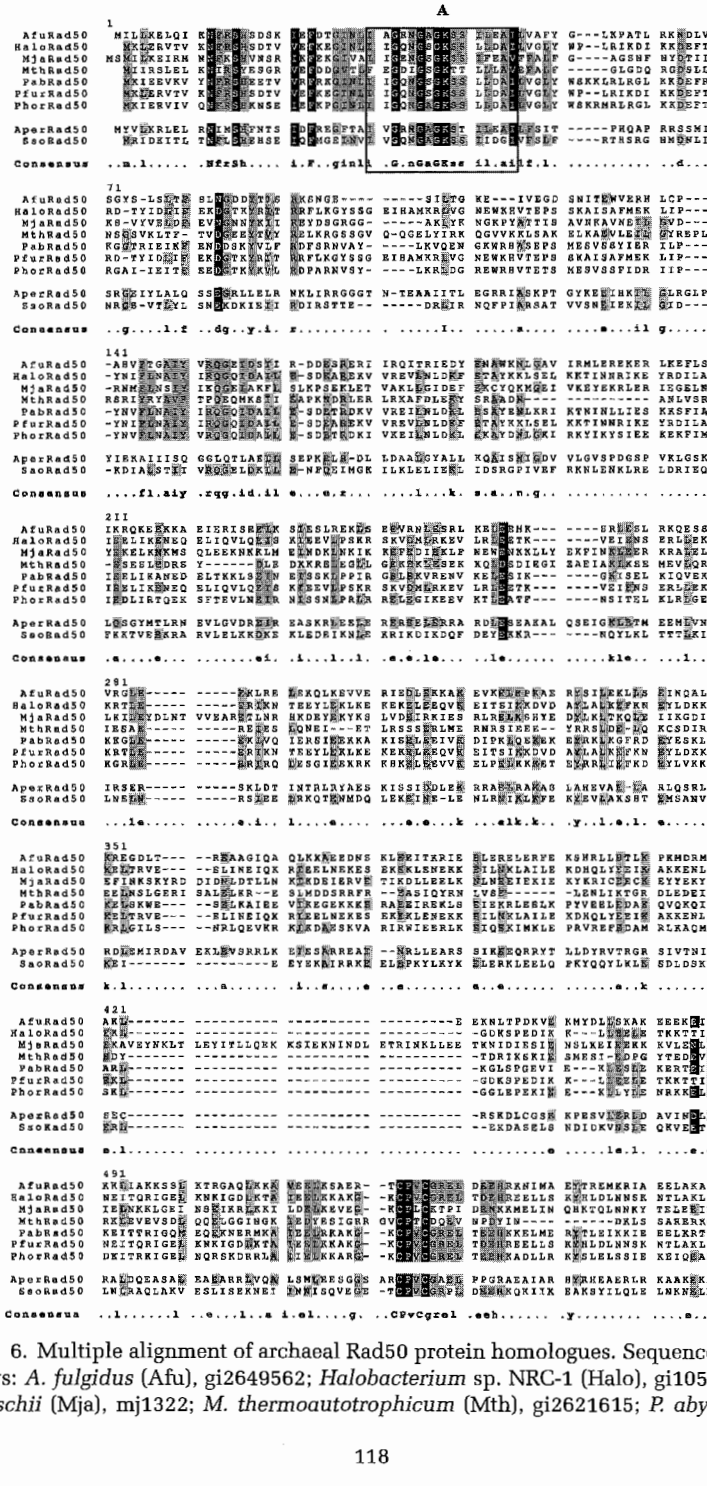
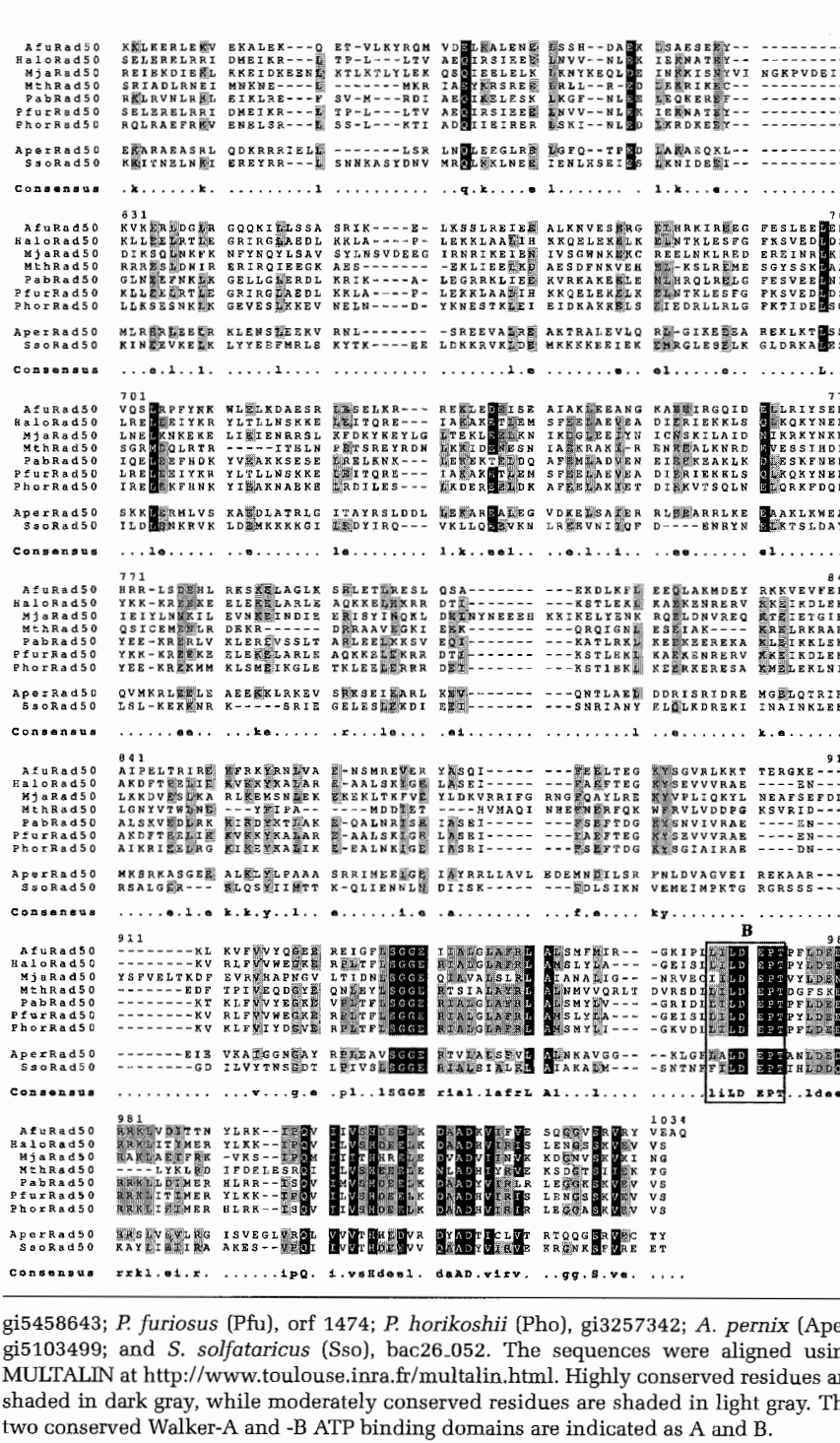


FIG. 6. Multiple alignment of archaeal Rad50 protein homologues. Sequences were as follows: *A. fulgidus* (Afu), gi2649562; *Halobacterium* sp. NRC-1 (Halo), gi10580117; *M. jannaschii* (Mja), mj1322; *M. thermoautotrophicum* (Mth), gi2621615; *P. abyssii* (Pab),



gi5458643; *P. furiosus* (Pfu), orf 1474; *P. horikoshii* (Pho), gi3257342; *A. pernix* (Ape), gi5103499; and *S. solfataricus* (Sso), bac26_052. The sequences were aligned using MULTALIN at http://www.toulouse.inra.fr/multalin.html. Highly conserved residues are shaded in dark gray, while moderately conserved residues are shaded in light gray. The two conserved Walker-A and -B ATP binding domains are indicated as A and B.

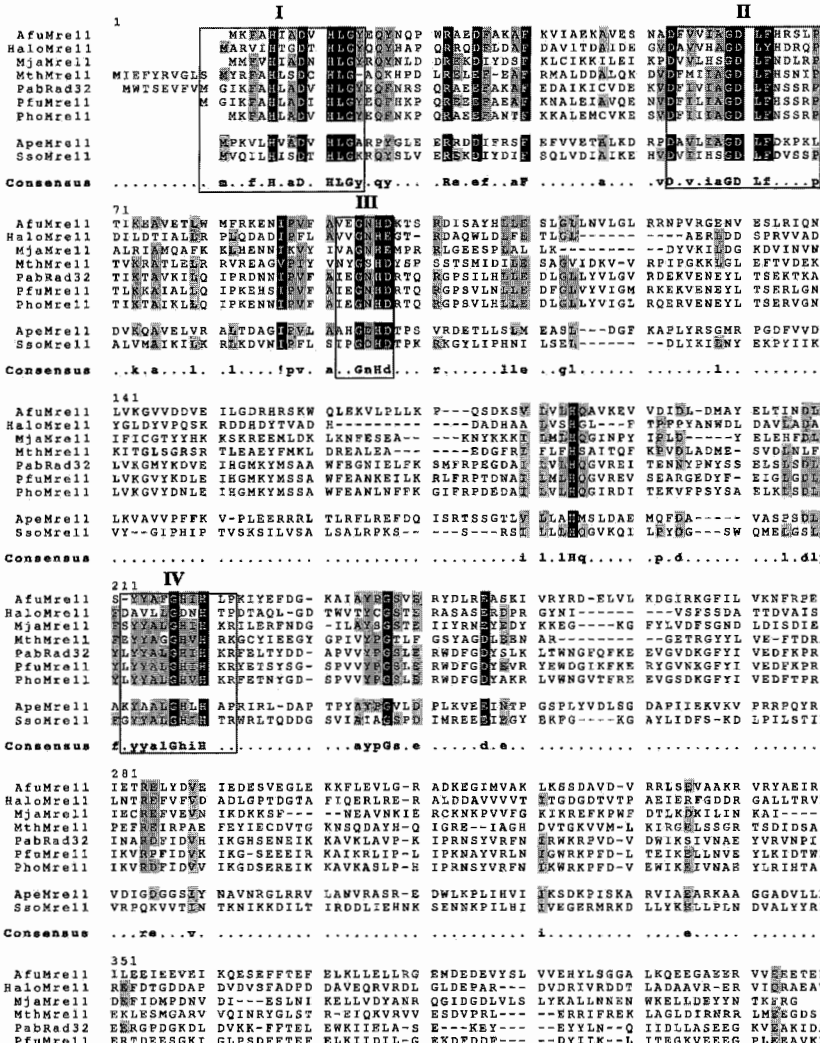


FIG. 7. Multiple alignment of archaeal Mre11 protein homologues. Sequences were as follows: *A. fulgidus* (Afu), G69378; *Halobacterium* sp. NRC-1 (Halo), gi10580116; *M. jannaschii* (Mja), B64465; *M. thermoautotrophicum* (Mth), E69171; *P. abyssii* (Pab), E75103; *P. furiosus* (Pfu), orf1475; *P. horikoshii* (Pho), D71083; *A. pernix* (Ape), E72765; and *S. solfataricus* (Sso), bac26_053. The sequences were aligned using MULTALIN at http://www.toulouse.inra.fr/multalin.html. Highly conserved residues are shaded in dark gray, while moderately conserved residues are shaded in light gray. Conserved nuclease domains I–IV as described for the Mre11 family are indicated.

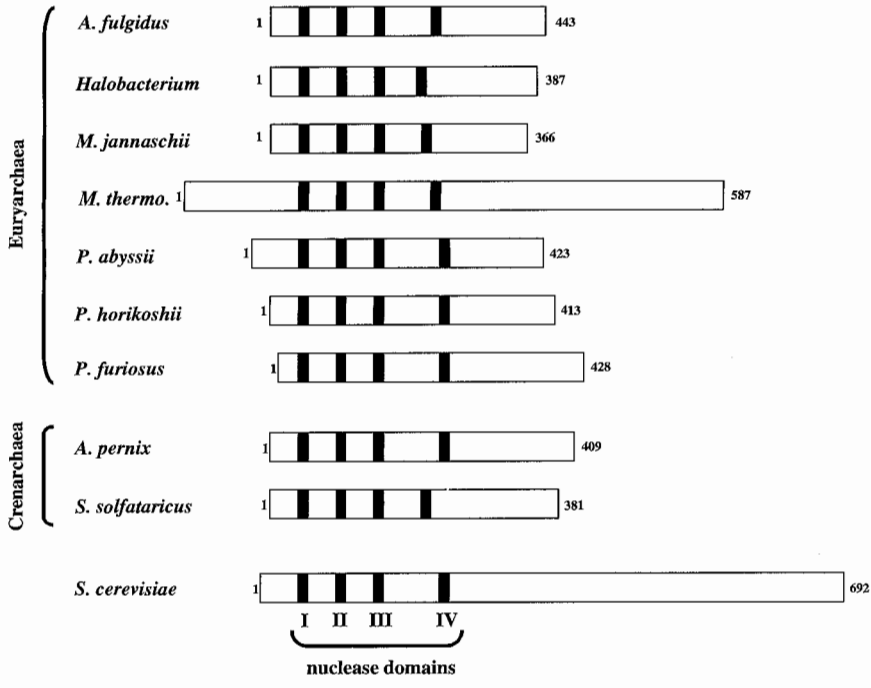


FIG. 8. Schematic representation of archaeal Mre11 protein homologues. Also, shown, for comparison, is the *S. cerevisiae* Mre11 protein. Conserved nuclease domains I–IV are indicated.

(Fig. 9). This suggests that the Rad50 protein may regulate DNA binding and release after DNA end processing through its association with the Mre11 protein (Hopfner *et al.*, 2000b). The fact that the Archaea possess both a Mre11/Rad50 protein homologue and a Spo11 protein homologue suggests that this group of organisms may both form and process DSBs more similarly to the Eukarya than to the Bacteria.

D. DNA PAIRING AND STRAND EXCHANGE

Perhaps the most crucial step in homologous recombination is that of homologous pairing and DNA strand exchange (Fig. 10) (Kowalczykowski and Eggleston, 1994; Bianco *et al.*, 1998; Kuzminov, 1999). The first archaeal recombination protein identified was a DNA strand exchange protein. This protein was discovered based upon its homology to both the bacterial and the eukaryal DNA strand exchange proteins, although it displayed more homology to the eukaryal DNA strand exchange protein (Sandler *et al.*, 1996). In the Bacteria, the role of

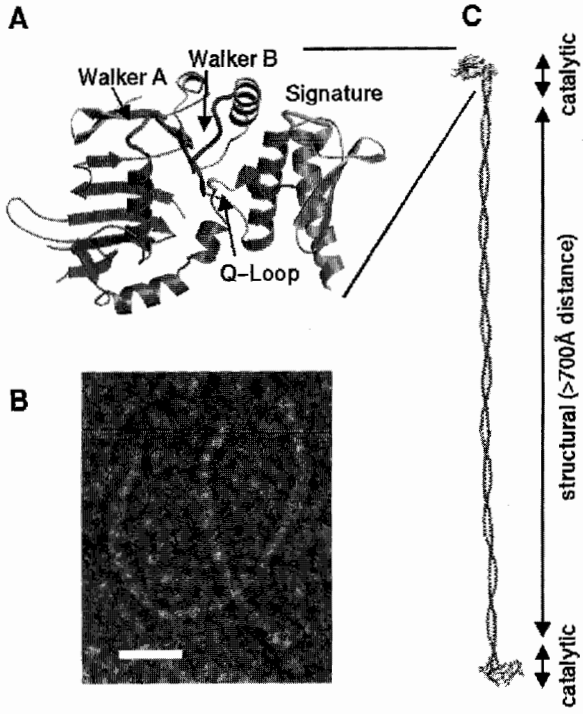


FIG. 9. Structure of the catalytic domain of the *P. furiosus* Rad50 protein. (A) The bilobal ABC type ATPase fold of the Rad50 protein catalytic domain, which is created by association of the N-terminal and C-terminal ATPase segments of Rad50 protein. The Walker-A and -B motifs, as well as other important catalytic domains, are indicated. (B) Electron micrograph of the elongated rods of the 600-residue coiled-coil domain of the Rad50 protein homodimer. The scale bar is 10 nm. (C) Proposed structure of a Rad50 homodimer. (Courtesy of J. A. Tainer, Scripps Research Institute.)

homologous pairing and DNA strand exchange is fulfilled by the RecA protein (Bianco and Kowalczykowski, 1999). In the Eukarya, the Rad51 protein, which is homologous to the RecA protein, assumes this role (Ogawa *et al.*, 1993), and in the Archaea, this DNA strand exchange step is mediated by the RadA protein (Seitz *et al.*, 1998).

1. Bacterial DNA Strand Exchange: The RecA Protein

Pioneering work on the *E. coli* RecA protein helped to define its role as the prototypical DNA strand exchange protein. The *recA* gene was originally isolated in *E. coli* over 30 years ago as a mutation responsible for a dramatic reduction in recombination levels, and its involvement was eventually established for almost all pathways of bacterial recombination (Clark and Margulies, 1965). Subsequently, the

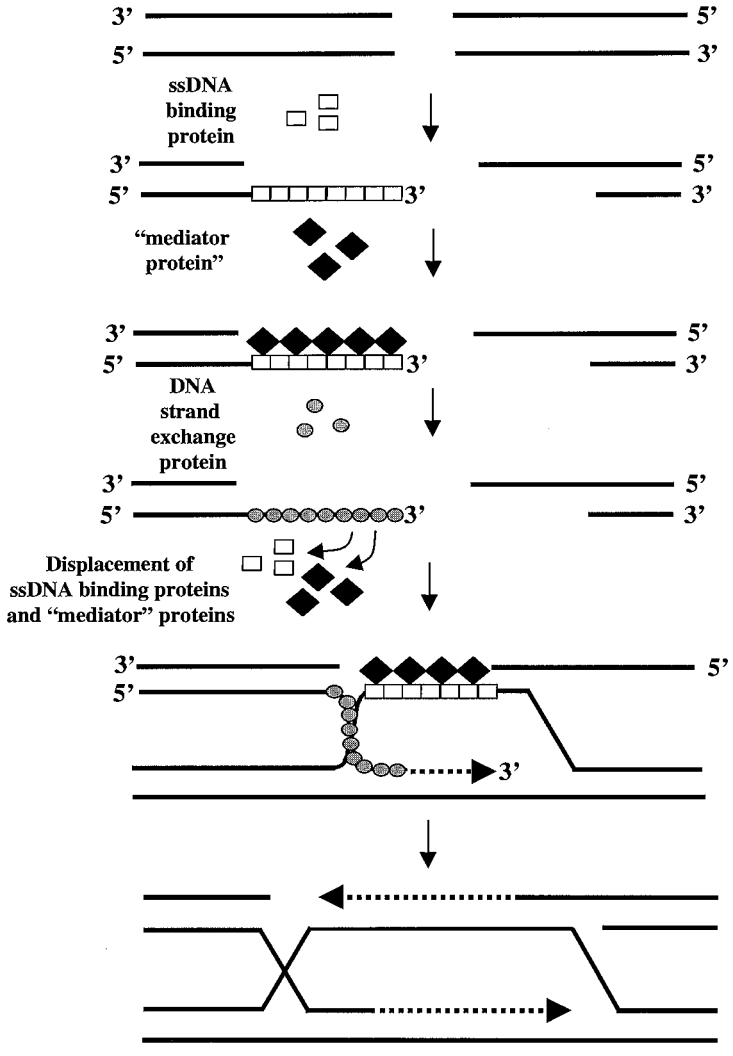


FIG. 10. Biochemical mechanism for the homologous pairing and DNA strand exchange step of homologous recombination. Shown is the DNA strand exchange protein-mediated homologous pairing event between a dsDNA molecule with a DSB and an intact target DNA molecule. After processing of the DSB, ssDNA tails are created, to which a ssDNA binding protein binds. To bind the ssDNA, the DNA strand exchange protein must then displace the ssDNA binding protein; this replacement is aided by mediator or exchange proteins. Next the DNA strand exchange protein catalyzes a homology search and pairs the two DNA molecules. The opposite end of the DSB, after processing, pairs either by the same process or by annealing of the displaced ssDNA in the joint molecule with the repair of ssDNA in the DSB. After DNA strand invasion, the 3′ end serves as a primer for DNA replication (dashed line).

RecA protein was found to possess many biochemical activities: ss- and dsDNA-dependent ATPase, DNA- and ATP-dependent coprotease, ATP-stimulated DNA annealing and ATP-dependent DNA strand exchange activities (Radding, 1989; Cox, 1999; Bianco and Kowalczykowski, 1999). After initial processing of the DSB ends by the RecBCD or RecQ enzymes (Anderson and Kowalczykowski, 1997b; Harmon and Kowalczykowski, 1998), the RecA protein begins a search for homology and catalyzes the pairing and exchange of a DNA strand between each of the two DNA molecules (Fig. 10). RecA protein-mediated homologous pairing and DNA strand exchange occur through a series of distinct steps: presynapsis, synapsis, and DNA heteroduplex extension. During presynapsis, the RecA protein binds to ssDNA in a stoichiometric fashion, with one RecA monomer bound per three nucleotides of ssDNA. The RecA protein interacts with ssDNA in a nonspecific, cooperative manner but does display a preference for binding and pairing DNA sequences rich in G and T residues (Tracy and Kowalczykowski, 1996). RecA protein assembly on ssDNA is polar and occurs in a 5′-to-3′ direction to yield a continuous right-handed helical nucleoprotein filament of RecA protein termed the "presynaptic complex" (Stasiak *et al.*, 1984; Egelman and Stasiak, 1986; Stasiak and Egelman, 1986, 1994). Formation of this presynaptic complex occurs much more readily in the presence of a single-stranded DNA binding protein, the SSB protein. Because the RecA protein binds poorly to dsDNA, the presence of secondary structure in ssDNA impedes the formation of a contiguous RecA protein filament. The SSB protein removes this block by disrupting the secondary structure and is subsequently displaced by the RecA protein. Removal of this ssDNA secondary structure permits contiguous filament formation by the RecA protein (Kowalczykowski and Krupp, 1987). The formation of the active RecA nucleoprotein filament typically depends on the presence of a cofactor such as ATP or dATP, and in this ATP-bound form, the RecA protein is in a state that has a high affinity for binding to DNA. The RecA protein hydrolyzes ATP at a rate (k_{cat}) of 25–30 min^{-1}. Although this ATP hydrolysis is not required for the homologous pairing and DNA strand exchange step, it is important in converting the RecA protein from a high-affinity ATP-bound form to an ADP-bound form that has a low affinity for DNA (Kowalczykowski, 1991). This allows the RecA protein both to bind tightly to DNA and to dissociate readily from DNA. Within the filament lies the ssDNA molecule, which has been extended by binding of the RecA protein to 1.5 times the axial spacing of regular B-form DNA (Stasiak *et al.*, 1981; Egelman and Stasiak, 1986, 1998; Stasiak and Egelman, 1986, 1984; Egelman and Yu, 1989).

During the synaptic step of this process, the RecA nucleoprotein filament catalyzes the search for homology within another dsDNA

molecule and exchanges DNA strands between the two molecules. First, the RecA filament makes a series of random, nonhomologous contacts with the target duplex DNA molecule before finding the homologous sequence. Next, the RecA protein catalyzes the exchange of DNA strands, producing a joint molecule. Subsequent to the formation of this joint molecule, the heteroduplex DNA can be extended by the RecA protein through a branch migration step that occurs in only one direction (5′ to 3′ relative to the displaced ssDNA) (Cox and Lehman, 1981); however, *in vivo,* the RuvAB proteins likely assume this function (West, 1997). The SSB protein also plays a second function in DNA strand exchange at this postsynaptic step, by binding to the displaced ssDNA strand and preventing RecA protein-dependent reinvasion of the duplex DNA molecule by the displaced strand (Kowalczykowski *et al.,* 1994).

2. Eukaryal DNA Strand Exchange: The Rad51 Protein

The existence of a RecA protein homologue in the Eukarya was discovered almost 10 years ago (Shinohara *et al.,* 1992). Mutants of *S. cerevisiae* were isolated on the basis of their sensitivity to ionizing radiation and their inability to undergo meiosis. Of the corresponding genes, studies showed that a *rad51* null mutant is defective in both mitotic and meiotic recombination and is impaired in DSB repair (Game, 1993). Additionally, it was found that the Rad51 protein showed a strong amino acid similarity to the RecA protein (Shinohara *et al.,* 1992). The Rad51 protein possesses many of the same biochemical activities as the RecA protein: stoichiometric binding to DNA (one Rad51 protein monomer per three nucleotides of DNA), ssDNA-dependent ATPase activity, and catalysis of DNA strand exchange (Sung, 1994). The Rad51 protein also forms a right-handed helical nucleoprotein filament on DNA, similar to that of the RecA protein (Ogawa *et al.,* 1993). Interesting differences do exist between these two homologues, however: the Rad51 protein hydrolyzes ATP at a much slower rate (0.7 min^{-1}), has a greater affinity for dsDNA binding, and catalyzes DNA strand exchange much less efficiently, even in the presence of the eukaryotic SSB protein, replication protein-A (RPA), than the RecA protein. Rad51 protein-promoted DNA strand exchange is almost entirely dependent on the presence of a ssDNA binding protein, in contrast to the RecA protein-promoted reaction (Sung and Robberson, 1995; Sugiyama *et al.,* 1997). The ready binding of Rad51 protein to dsDNA poses a unique problem, in that it blocks DNA strand exchange *in vitro* (Sung and Robberson, 1995). Interestingly, the Rad51 protein also shows a pairing bias that is opposite to that of the RecA protein (Mazin *et al.,* 2000b), suggesting that the biochemical properties of the two nucleoprotein filaments may be different. Additionally, the Rad51 protein interacts

with other members of the *RAD52* epistasis group, some of which stimulate activities of the Rad51 protein (Sung *et al.*, 2000) (Fig. 10; see below).

3. Archaeal DNA Strand Exchange: The RadA Protein

A role for the RadA protein (Sandler *et al.*, 1996) in DNA repair via homologous recombination came from genetic analysis showing that deletion of the *radA* gene in *Haloferax volcanii* (Woods and Dyall-Smith, 1997) resulted in an archaeon that exhibited a decreased growth rate and an increased sensitivity to DNA damaging agents such as UV irradiation and ethylmethane sulfonate (EMS). The RecA protein homologue from the hyperthermophilic crenarchaeote *Sulfolobus solfataricus* was the first to be purified and studied biochemically (Seitz *et al.*, 1998). It shares many of the same biochemical characteristics of the RecA and Rad51 proteins: the RadA protein is a DNA-dependent ATPase, forms a helical nucleoprotein filament on DNA (Fig. 11), and catalyzes DNA strand exchange. The RadA protein also binds ssDNA with the same stoichiometry as do the RecA and Rad51 proteins, one RadA monomer per three nucleotides of DNA, and it shows a preference for binding to and pairing DNA sequences that are rich in G and T residues (Seitz and Kowalczykowski, 2000). These biochemical

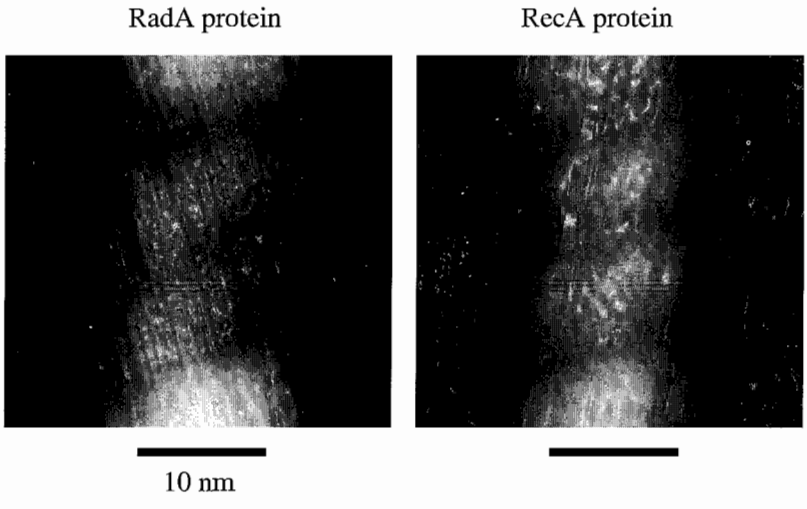


FIG. 11. Nucleoprotein filaments of RecA and RadA proteins imaged by atomic force microscopy. Shown are complexes of the RadA and RecA proteins assembled on pBR322 dsDNA in the presence of the ATP analogue, ADP · Al · F_4. As shown, the RadA protein forms a right-handed helical structure that is similar to the structure formed by the RecA protein. [Adapted from Seitz *et al.* (1998).]

activities were seen only at elevated temperatures, close to those at which *S. solfataricus* thrives. The nucleoprotein filament formed by the archaeal RadA protein is the same right-handed helical structure formed by the *E. coli* RecA and the *S. cerevisiae* Rad51 proteins (Egelman and Stasiak, 1986; Ogawa *et al.*, 1993; Seitz *et al.*, 1998) (Fig. 11). The RadA protein's biochemical activities seem more akin to those of the Rad51 protein, however, in that the rate of ATPase activity is rather low (k_{cat} = 0.2 min^{-1}), and the efficiency of DNA strand exchange is also rather poor (Seitz *et al.*, 1998).

The RadA proteins from other hyperthermophilic archaeons, *Desulfurococcus amylolyticus, Pyrobaculum islandicum,* and *P. furiosus,* possess similar biochemical activities, also at elevated temperatures (Kil *et al.*, 2000; Komori *et al.*, 2000b; Spies *et al.*, 2000). Figure 12 shows an alignment of nine archaeal RadA protein sequences, demonstrating the extensive sequence conservation; the well-conserved Walker-A and -B nucleoside triphosphate binding motifs are indicated. In accord with its biochemical similarity to the eukaryal Rad51 protein, the amino acid sequences show that the archaeal RadA proteins are structurally more closely related to the eukaryal Rad51 protein (34–42% identical and 53–63% similar) than to their bacterial counterpart (14–17% identical and 25–31% similar). Domain analysis of the RadA protein from *P. furiosus* demonstrates that the C-terminal portion of the protein, which contains the central core domain (Domain II), possesses DNA-dependent ATPase activity and DNA strand exchange activity, although much reduced in comparison to those of the native RadA protein. Addition of the missing N-terminal peptide to the C-terminal portion restored RadA protein activity to 60% of the wild-type level as measured by ATPase and DNA strand exchange activities, which suggests that the N terminus is needed for the protein to achieve the proper structure for optimal activity (Komori *et al.*, 2000a).

E. Single-Stranded DNA Binding Proteins

As stated previously, DNA strand exchange takes place in essentially three stages. During the steps of presynapsis and postsynapsis, ssDNA binding proteins help to alleviate ssDNA secondary structure and to prevent reinvasion of the displaced single strand of DNA after synapsis, respectively (Kowalczykowski *et al.*, 1994). These functions are fulfilled in bacteria by the ssDNA binding (SSB) protein and in eukaryotes by replication protein-A (RPA) (Fig. 13). Several ssDNA binding proteins have also been identified in the Archaea. Although single-stranded DNA binding proteins are conserved throughout the Archaea, Bacteria, and Eukarya, their protein architectures are quite different.

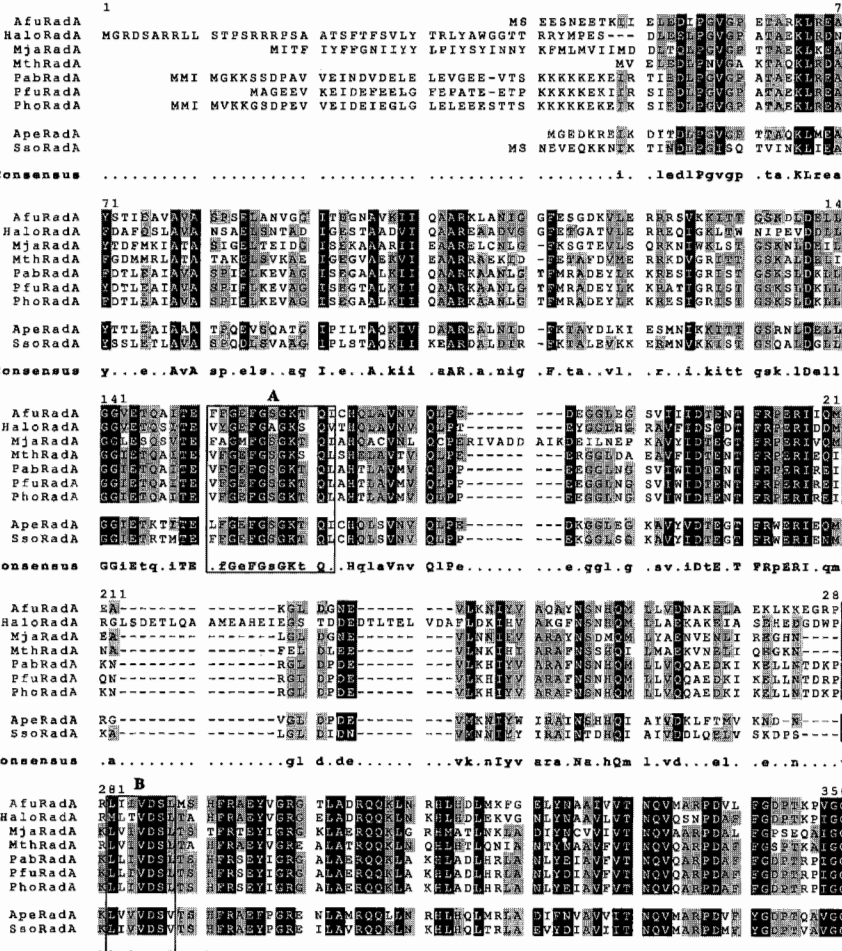


FIG. 12. Multiple alignment of archaeal RadA protein homologues. Sequences were as follows: *A. fulgidus* (Afu), gi2649602; *Halobacterium* sp. NRC-1 (Halo), gi10581871; *M. jannaschii* (Mja), gi2146708; *M. thermoautotrophicum* (Mth), gi2622493; *P. abyssii* (Pab), gi7448305; *P. furiosus* (Pfu), gi3560537; *P. horikoshii* (Pho), gi3256652; *A. pernix* (Ape), gi5103509; and *S. solfataricus* (Sso), gi2129447. The sequences were aligned using MULTALIN at http://www.toulouse.inra.fr/multalin.html. Highly conserved residues are shaded in dark gray, while moderately conserved residues are shaded in light gray. The two conserved Walker-A and -B domains are indicated as A and B.

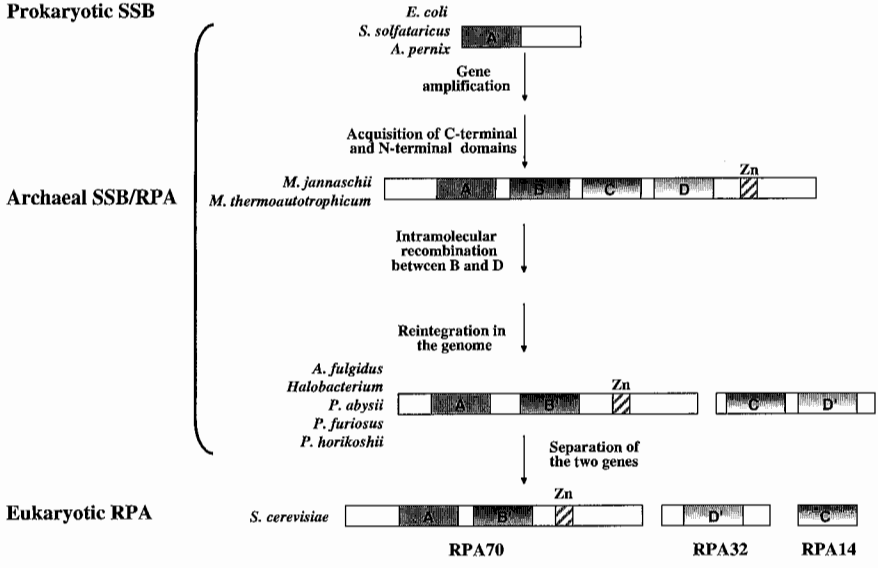


FIG. 13. A model for the evolutionary relationship between the single-stranded DNA binding proteins. Shown is a possible scheme for the evolution of the heterotrimeric eukaryal RPA protein from the single subunit of the bacterial and archaeal SSB proteins. The path illustrated is the simplest and does not necessarily imply the actual evolutionary mechanism. [Adapted from Chédin *et al.* (1998a).]

1. The Bacterial SSB Protein

The *E. coli* SSB protein is important in the processes of replication, recombination, mutagenesis, transposition, repair, and response to DNA damage (Meyer and Laine, 1990). This protein binds preferentially and cooperatively to ssDNA (Lohman and Ferrari, 1994). The *E. coli* SSB protein is encoded by a single gene, while the active form of the protein is a homotetramer in which each monomer contains one ssDNA binding domain (Lohman and Ferrari, 1994). During the process of homologous recombination, the SSB protein is involved in stimulation of RecA protein-mediated DNA strand exchange and in protecting ssDNA from nucleolytic degradation (Kowalczykowski *et al.*, 1994; Anderson and Kowalczykowski, 1998).

2. The Eukaryal RPA

The eukaryal RPA complex is composed of three distinct subunits (Gomes and Wold, 1995, 1996; Wold, 1997). The large subunit of this protein, RPA70, has several domains. The N terminus mediates interactions between RPA and many cellular proteins, while the middle

region contains two functional and homologous ssDNA binding sites. The C terminus is involved in interactions with the other subunits of this heterotrimeric complex and also contains a zinc-finger domain, which is important for RPA function (Wold, 1997). RPA32 carries a third functional ssDNA binding site and is phosphorylated in a cell cycle-dependent manner (Bochkareva *et al.*, 1998). Finally, the smallest subunit, RPA14, has an additional ssDNA binding domain. Although the bacterial and eukaryal proteins have completely different protein architectures and share little homology overall, a significant amount of homology is found between their ssDNA binding domain motifs. For example, the ssDNA binding domain A of the RPA70 subunit shows similarity to the *E. coli* SSB protein. This homology also extends to phage-encoded SSB's and, now, to the archaeal ssDNA binding proteins (Philipova *et al.*, 1996; Chédin *et al.*, 1998b; Kelly *et al.*, 1998; Haseltine, 2001; Wadsworth and White, 2001).

3. Archaeal ssDNA Binding Proteins

A ssDNA binding protein was initially found by sequence analysis in each of three archaeons: *Methanococcus jannaschii, Methanobacterium thermoautotrophicum,* and *Archaeoglobus fulgidus* (Fig. 13 and Table I) (Chédin *et al.*, 1998b). These proteins are homologous to the eukaryal *RFA1* gene, which corresponds to the RPA70 subunit, the largest subunit of the RPA heterotrimeric complex. Interestingly, the ssDNA binding proteins discovered in these three archaeons possessed completely different architectures from either the SSB protein or RPA (Chédin *et al.*, 1998b; Kelly *et al.*, 1998). The euryarchaeal *M. jannaschii* and *M. thermoautotrophicum* proteins exist as a single polypeptide chain and encompass four ssDNA binding domains in tandem, all of which show homology to each other (Fig. 13). Additionally, these ssDNA binding domains contain amino acids that are conserved in the eukaryal RPA70 subunit and are known to make contacts with DNA. Furthermore, a strongly conserved zinc-finger domain was also found within these proteins. This finding implies that these proteins function as a single subunit that does not require multimerization, as in the case of the SSB protein, or association with other subunits, as in the case of the eukaryal RPA.

Investigation into other members of the Archaea, however, revealed ssDNA binding proteins with varied architectures (Chédin *et al.*, 1998b). For example, in *A. fulgidus,* a protein containing two subunits with two DNA binding domains in each was discovered. The second subunit also contained a putative zinc-finger motif. This organization proved to be true for *Pyrococcus abysii, Pyrococcus horikoshii, P. furiosus,* and *Halobacterium* sp. NRC-1 as well (Fig. 13). Finally, the genomes

of *Aeropyrum pernix* and *S. solfataricus*, two members of the Crenarchaeota, possess proteins with a completely different architecture (Haseltine, 2001; Wadsworth and White, 2001). These proteins contain a single subunit with a single ssDNA binding domain and an acidic C terminus, which are hallmarks of an *E. coli* SSB protein-like structure. This suggests that the ssDNA binding proteins from members of the Crenarchaeota and Euryarchaeota must have diverged early in evolution and that representatives of each type of ssDNA binding protein still exist in members of the Archaea.

The ssDNA binding proteins from *M. jannaschii* (Kelly *et al.*, 1998; E. M. Seitz and S. C. Kowalczykowski, unpublished observation) and, most recently, *S. solfataricus* (Haseltine, 2001; Wadsworth and White, 2001) were purified. Both proteins show ssDNA binding activity at elevated temperatures, but neither stimulate the ATPase activity or DNA strand exchange activities of the RadA protein. Since secondary structure is not stable in ssDNA at elevated (75–80°C) temperatures, there may be little need for an SSB protein in the presynaptic step of archaeal recombination. Consequently, these ssDNA binding proteins might be needed only for postsynaptic steps.

F. Additional Proteins Involved in DNA Strand Exchange

During the process of DNA strand exchange, the RecA, Rad51, and RadA proteins may encounter obstacles that prevent them from binding to ssDNA or from efficiently completing the DNA strand exchange or DNA heteroduplex extension step. In some instances, ssDNA binding proteins can actually serve as competitors to binding of the DNA strand exchange proteins to ssDNA. This competition is overcome by "mediator" proteins that can facilitate the binding of the DNA strand exchange protein to ssDNA (Fig. 10). In *E. coli*, the RecF, RecO, and RecR proteins serve this function by facilitating binding of the RecA protein to a SSB protein-coated ssDNA gap (Umezu *et al.*, 1993; Webb *et al.*, 1997; Kuzminov, 1999). While there is no structural homologue of either the RecF, the RecO, or the RecR protein in the Eukarya, two factors, the Rad52 protein and Rad55/57 proteins, help the Rad51 protein to overcome the competition imposed by the binding of RPA to ssDNA (Pâques and Haber, 1999; Sung *et al.*, 2000). The Rad55/57 proteins share homology to the Rad51 protein and are, therefore, referred to as Rad51 protein paralogues. Homologues of the RecF, RecO, RecR or Rad52 protein have not been identified in the Archaea. However, there exists a RadA protein paralogue, the RadB protein (Komori *et al.*, 2000b), whose function is unclear, but it may also serve a "mediator" role during DNA strand exchange.

1. Recombination Mediator/DNA Annealing Proteins

a. Bacterial RecFOR Proteins. In both the Bacteria and the Eukarya, there exist proteins that aid the DNA strand exchange protein. In wild-type *E. coli,* the need for these "accessory" proteins is revealed when the DNA lesion is a daughter strand gap, whose repair occurs via the RecF pathway of recombinational repair (Horii and Clark, 1973; Kuzminov, 1999). In this pathway, three proteins facilitate aspects of RecA nucleoprotein filament formation: RecF, RecO, and RecR (Fig. 1) (Kolodner *et al.,* 1985). In the course of daughter strand gap repair, SSB protein is the first protein to bind to the ssDNA within the gap. To facilitate the exchange of RecA protein for SSB protein, the RecOR protein complex binds to the SSB protein–ssDNA complex and facilitates the polymerization of the RecA protein filament at the expense of the SSB protein-coated ssDNA. The RecA protein can now pair the ssDNA gap with a homologous sequence to permit repair of the ssDNA gap. In this capacity, the RecO and -R proteins help both to direct the RecA protein to the gap and to displace the SSB protein that is coating the ssDNA. The RecF protein forms a complex with the RecR protein, and this complex binds randomly to dsDNA to stop RecA nucleoprotein filament extension (Webb *et al.,* 1997). The RecO protein can also anneal complementary ssDNA (Luisi-DeLuca and Kolodner, 1994) and, in fact, can anneal ssDNA that is complexed with the SSB protein (N. Kantake, M. V. V. M. Madiraju, T. Sugiyama, and S. Kowalczykowski, in preparation). To date, no structural homologues of RecF, RecO, or RecR have been uncovered in eukaryal or archaeal organisms, although these proteins are conserved throughout the Bacteria; however, functional homologues exist.

b. The Eukaryal Rad52 Protein. The importance of *S. cerevisiae RAD52* in recombination is underscored by the fact that null mutations in *RAD52* eliminate the cell's ability to carry out all homologous recombination events (Game, 1993; Rattray and Symington, 1994). *RAD52* has therefore been implicated in multiple recombination pathways: homologous recombination, ssDNA annealing (SSA), and break-induced replication (BIR) (Pâques and Haber, 1999; Sung *et al.,* 2000). The Rad52 protein bears no structural homology to any known recombination factors in the Bacteria; however, it appears to be a functional homologue of the RecO(R) protein. Additionally, no Rad52 protein homologues have been identified in the Archaea.

The Rad52 protein binds ssDNA and mediates DNA strand annealing between two homologous DNA molecules; this activity is stimulated by the presence of RPA bound to the DNA (Mortensen *et al.,* 1996; Shinohara *et al.,* 1998; Sugiyama *et al.,* 1998). The Rad52 protein binds to DNA by forming ring-shaped multimers (Shinohara *et al.,* 1998; Van

Dyck *et al.*, 1999) and binds to ssDNA with a higher affinity than to dsDNA (Mortensen *et al.*, 1996; Van Dyck *et al.*, 1999). The Rad52 protein forms a complex with the Rad51 protein, as shown by immunoprecipitation (Sung, 1997b). The Rad52 protein is also able to form a complex with RPA or with RPA–ssDNA complexes (Shinohara *et al.*, 1998; Sugiyama *et al.*, 1998). During DNA strand exchange, the Rad52 protein is able to overcome the inhibition to Rad51 protein posed by the binding of RPA to ssDNA (New *et al.*, 1998; Shinohara *et al.*, 1998). While the Rad52 protein can bind ssDNA, it does not displace RPA from ssDNA; rather it mediates an efficient exchange between the Rad51 protein and RPA (Sung, 1997b; New *et al.*, 1998; Shinohara and Ogawa, 1998). The mechanism by which the Rad52 protein carries out this role as "mediator" may be through its ability to target the Rad51 protein to ssDNA, although presently the exact mechanism is not entirely clear.

2. Rad51 and RadA Protein Paralogues

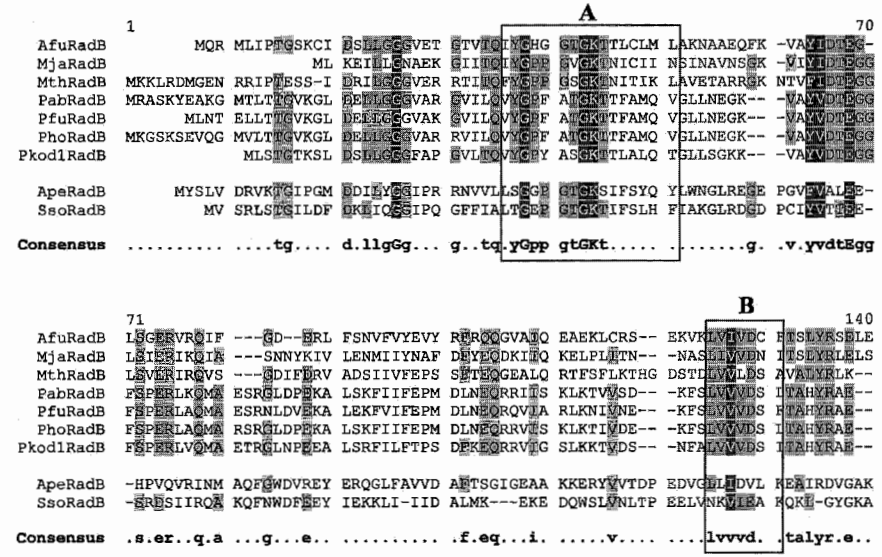
a. Eukaryal Rad55/57 Proteins (Rad51 Protein Paralogues). Additional members of the yeast *RAD52* epistasis group function in conjunction with the Rad51 protein, and some of these members exist in archaeal genomes. Two proteins in *S. cerevisiae* show limited homology to both the RecA and the Rad51 proteins and are called the Rad55 and Rad57 proteins (Sung *et al.*, 2000). The homology between these proteins and either the RecA or the Rad51 protein resides mainly in the sequence motifs that are involved in nucleoside triphosphate binding. In yeast, mutations in these genes result in cells that are cold-sensitive for both recombination and sensitivity to ionizing radiation. The recombination defect of a *rad55 rad57* double mutant is no greater than that of either single mutation alone, which suggests an epistatic relationship between the two genes (Lovett and Mortimer, 1987). The two proteins interact with one another, as evidenced by yeast two-hybrid experiments and coimmunoprecipitation (Johnson and Symington, 1995). The Rad55/57 complex aids the Rad51 protein in forming a more continuous filament on ssDNA that is complexed with RPA during the presynaptic step of DNA strand exchange (Sung, 1997a).

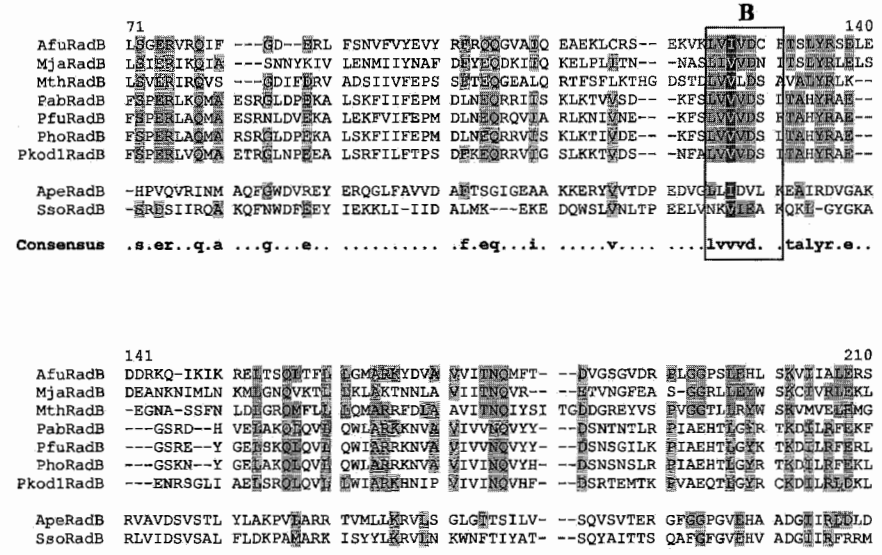
Human cells contain five Rad51 paralogues of unknown function, known as XRCC2, XRCC3, Rad51B, Rad51C, and Rad51D. These human Rad51 paralogues are all mitotically expressed (Albala *et al.*, 1997; Rice *et al.*, 1997; Cartwright *et al.*, 1998a,b; Dosanjh *et al.*, 1998; Liu *et al.*, 1998), and share 20–30% amino acid homology with the human Rad51 protein and with each other. The *XRCC2* and *XRCC3* genes are important for chromosome stability in mammalian cells (Fuller and Painter, 1988; Tucker *et al.*, 1991; Cui *et al.*, 1999), and *XRCC2* and *XRCC3* are

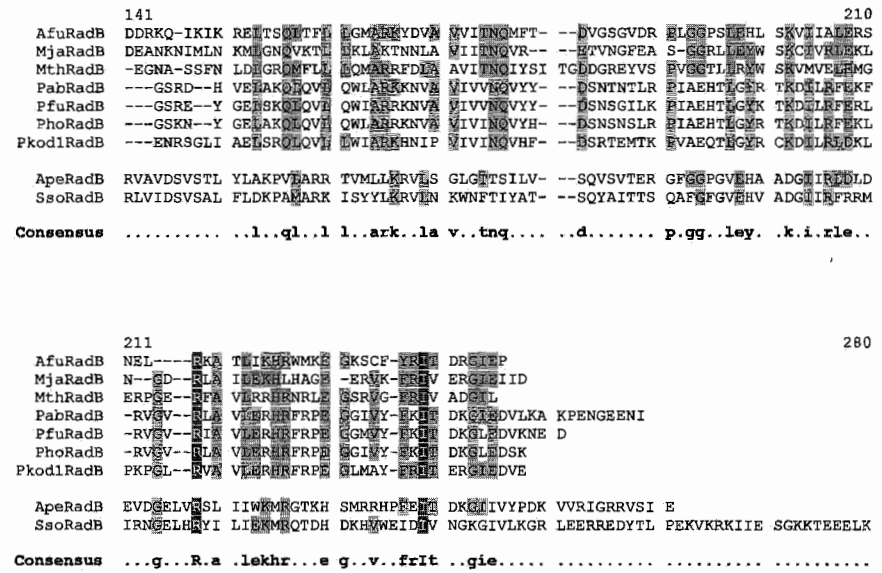
important for efficient repair of DSBs by homologous recombination (Cui *et al.*, 1999; Pierce *et al.*, 1999). Additionally, these five human Rad51 paralogues interact with one other (Schild *et al.*, 2000).

b. Archaeal RadA Protein Paralogues. The Archaea possess proteins homologous to RadA protein as well, and they may serve the same sort of presynaptic role in homologous recombination as demonstrated for Rad55/57 (Fig. 1). The RadA protein paralogue in the Archaea is referred to as RadB. Figure 14 shows an alignment of nine RadB proteins and the conserved Walker-A and -B motifs. These RadB proteins differ from RadA in two ways: first, the RadB proteins are smaller than the RadA protein, lacking both an N- and a C-terminal extension (Fig. 15). Second, while the sequences are homologous, they share only about 30–40% similarity with the RadA protein. In addition, there is a difference between the euryarchaeal and the crenarchaeal RadB protein sequences. The crenarchaeal RadB proteins show more sequence similarity to the *E. coli* RecA protein and, in fact, cannot be identified through a Blast search with the *S. cerevisiae* Rad51 protein sequence. Figure 16a shows an alignment between the *E. coli* RecA protein and the RadB proteins from the crenarchaeotes *S. solfataricus* and *A. pernix*. The crenarchaeal RadB protein is truncated on both the N and the C termini in comparison to the RecA protein but shows 25–27% amino acid similarity over the entire protein. Conversely, RadB proteins from euryarchaeotes show more sequence similarity to the *S. cerevisiae* Rad51 protein (Bult *et al.*, 1996; Klenk *et al.*, 1997; Smith *et al.*, 1997; Kawarabayasi *et al.*, 1998; Komori *et al.*, 2000b). Figure 16b shows an alignment of euryarchaeal RadB proteins with the *S. cerevisiae* Rad51 protein. These euryarchaeal RadB proteins share 38–54% amino acid similarity, across the entire protein, to the Rad51 protein.

The *radB* gene from *P. furiosus* was cloned and its gene product purified (Komori *et al.*, 2000b). This protein possesses a weak DNA-independent ATPase activity, and, interestingly, a higher affinity for binding to ssDNA than does the RadA protein. The RadB protein inhibits RadA protein-promoted D-loop formation under all conditions examined. Inhibition is also seen in RadA protein-promoted DNA strand exchange unless the RadB protein is added after the RadA protein is allowed to bind the ssDNA. Electron microscopy reveals that the RadB protein forms a filamentous structure on ssDNA. The RadB protein did not show any interaction with the RadA protein, which differs from the situation with Rad51 and Rad55/57. Interestingly, this protein coimmunoprecipitates with the Hjc enzyme from *P. furiosus*, a Holliday junction-resolving enzyme (see below), and the RadB protein inhibited Holliday junction cleavage by the Hjc protein. The fact that the RadB







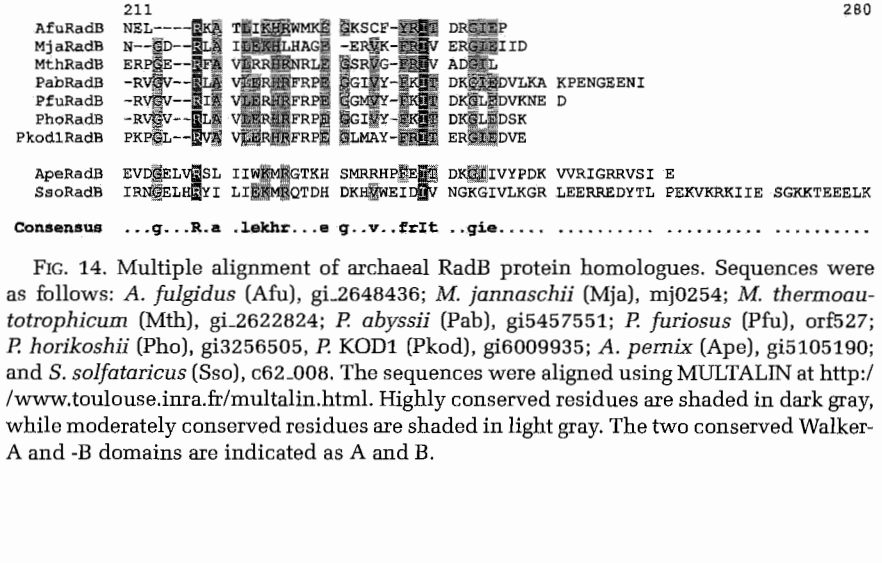


FIG. 14. Multiple alignment of archaeal RadB protein homologues. Sequences were as follows: *A. fulgidus* (Afu), gi_2648436; *M. jannaschii* (Mja), mj0254; *M. thermoautotrophicum* (Mth), gi_2622824; *P. abyssii* (Pab), gi5457551; *P. furiosus* (Pfu), orf527; *P. horikoshii* (Pho), gi3256505, *P.* KOD1 (Pkod), gi6009935; *A. pernix* (Ape), gi5105190; and *S. solfataricus* (Sso), c62_008. The sequences were aligned using MULTALIN at http://www.toulouse.inra.fr/multalin.html. Highly conserved residues are shaded in dark gray, while moderately conserved residues are shaded in light gray. The two conserved Walker-A and -B domains are indicated as A and B.

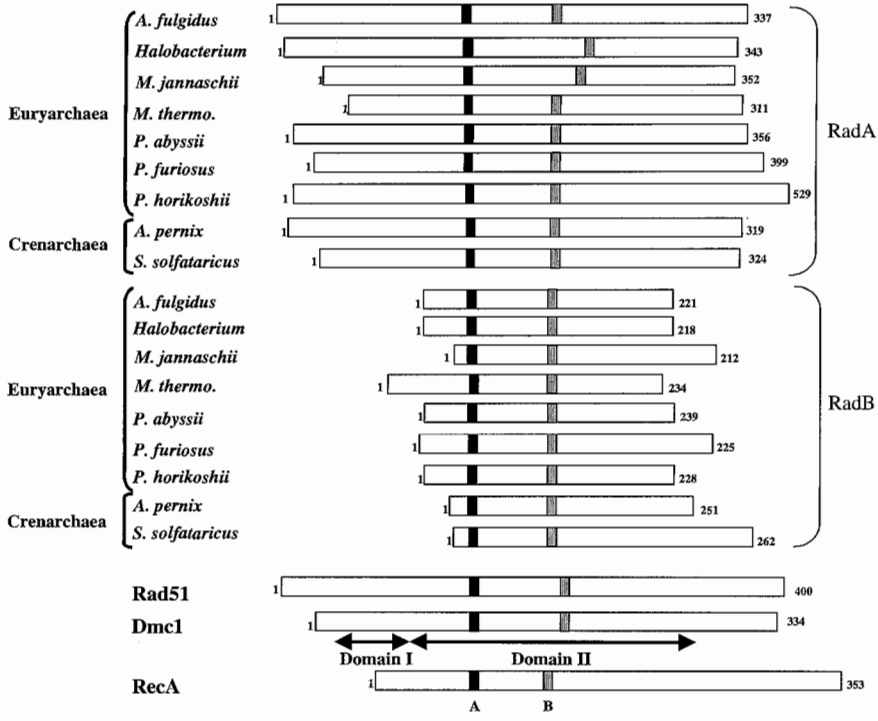


FIG. 15. Schematic representation of the archaeal RadB proteins compared to the RadA proteins and to the RecA/Rad51 proteins. Shown for comparison are the *S. cerevisiae* proteins, Rad51 and Dmc1, and the *E. coli* protein, RecA. RadA proteins are approximately 100 amino acids longer than RadB proteins at the N terminus (Domain I). RadB proteins consist primarily of a central core domain (Domain II). The two conserved Walker-A and -B domains are indicated as A and B.

protein did not stimulate any RadA protein activity could be due to the fact that, to function properly, it must form a heterodimer with another unknown protein, like the *S. cerevisiae* Rad55/57 protein complex (Komori *et al.*, 2000b).

3. Rad54 Proteins

a. The Yeast Rad54 Protein. Another member of the *RAD52* epistasis group, the Rad54 protein, was shown in *S. cerevisiae* to enhance Rad51 protein function during the synaptic phase of DNA strand exchange (Petukhova *et al.*, 1999; Mazin *et al.*, 2000a; Van Komen *et al.*, 2000). This protein belongs to a group of proteins known as the Swi2/Snf2 family, which are involved in a variety of chromosomal processes (Eisen *et al.*, 1995). The Rad54 protein has dsDNA-dependent ATPase activity,

and it can induce a conformational change in dsDNA, which is manifest as a change in the linking number of covalently closed dsDNA (Petukhova *et al.,* 1999; Tan *et al.,* 1999). The Rad54 protein interacts with the Rad51 protein in both yeast two-hybrid and *in vitro* analyses (Petukhova *et al.,* 1998), and the Rad54 protein stimulates, by more than 10-fold, Rad51 protein-dependent homologous DNA pairing (Petukhova *et al.,* 1999; Mazin *et al.,* 2000a; Van Komen *et al.,* 2000).

b. Archaeal Rad54 Protein Homologues. A putative Rad54 protein homologue exists in the crenarchaeote *S. solfataricus* (Table I and Figs. 16 and 17). The *S. solfataricus* Rad54 homologue shows conservation of the seven helicase motifs that are found in the yeast Rad54 protein, and it is about 30 amino acids longer than the yeast protein. Figure 17 shows an alignment of the *S. solfataricus* Rad54 protein with the *S. cerevisiae* Rad54 protein, and the conserved helicase motifs are labeled. Also indicated are conserved leucine residues that may constitue a leucine zipper motif. Figure 18 is a schematic comparison of these two proteins. The *S. solfataricus* Rad54 protein lacks the nuclear localization signal (NLS) of the *S. cerevisiae* Rad54 protein but has 47 and 25% amino acid similarity and identity, respectively, to the first 200 amino acids immediately following the yeast Rad54 NLS. This 200-amino acid region makes the Rad54 protein family distinct from other Swi2/Snf2 DNA-dependent ATPases (Kanaar *et al.,* 1996). Additionally, the *S. solfataricus* Rad54 protein has a conserved leucine zipper motif that is found in the *S. cerevisiae* Rad54 protein. Homologues of the Rad54 protein cannot be identified unequivocally in other archaeons due to weak sequence conservation, and currently there is no biochemistry available for any putative archaeal Rad54 protein.

G. Holliday-Junction Cleaving Enzymes

When first proposed, the Holliday model for recombination envisioned that exchange of both single strands of dsDNA with a homologous duplex DNA would produce a four-way junction, termed the Holliday (1964) junction. This four-way Holliday-junction is central to many models of homologous recombination, and physical evidence for this junction in meiotic recombination was demonstrated (Schwacha and Kleckner, 1995). The formation of this four-way junction is followed by branch migration, which includes the progressive exchange of base-pairing between the homologous duplex DNA molecules (West, 1992; White *et al.,* 1997). Cleavage of this junction by the introduction of two symmetric phosphodiester cleavages (Fig. 19) in one of two possible orientations results in two possible recombinant

(a)

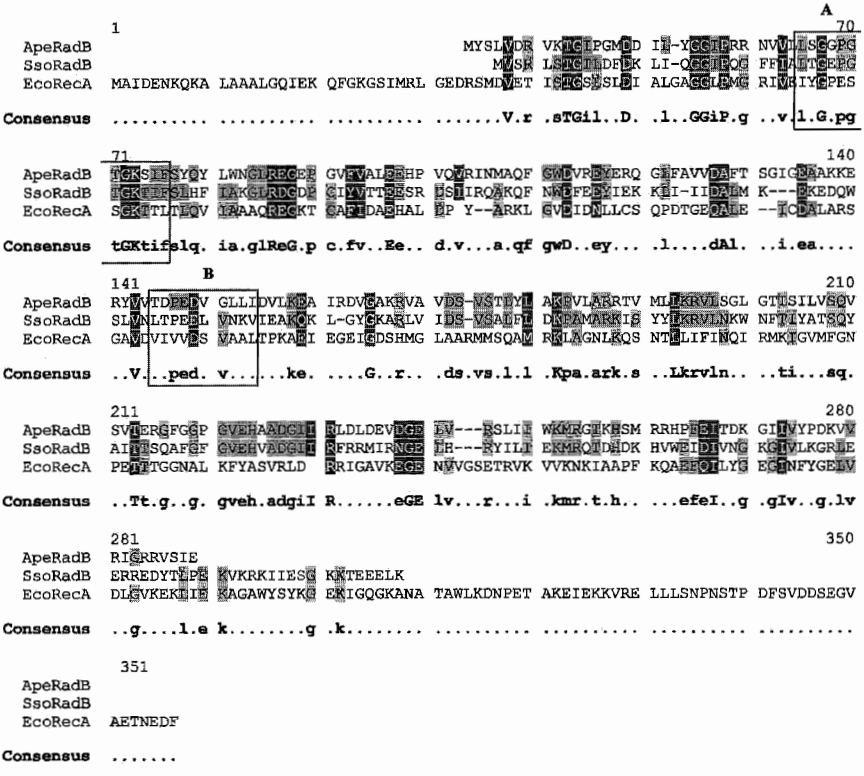


FIG. 16. Multiple alignment of RadB protein homologues. (a) Alignment of crenarchaeal RadB proteins with *E. coli* RecA protein. Sequences were as follows: *A. pernix* (Ape), gi5105190; *S. solfataricus* (Sso), c62_008; and *E. coli* (Eco), gi1789051. (b) Alignment of euryarchaeal RadB proteins with *S. cerevisiae* Rad51 protein. Sequences were as follows: *A. fulgidus* (Afu), gi_2648436; *M. jannaschii* (Mja), mj0254; *M. thermoautotrophicum* (Mth), gi_2622824; *P. abyssii* (Pab), gi5457551; *P. furiosus* (Pfu), orf527; *P. horikoshii* (Pho), gi3256505, P. KOD1 (Pkod), gi6009935; and *S. cerevisiae* (Sce), gi603333. The sequences were aligned using MULTALIN at http://www.toulouse.inra.fr/multalin.html. Highly conserved residues are shaded in dark gray, while moderately conserved residues are shaded in light gray. The two conserved Walker-A and -B domains are indicated as A and B.

DNA products: spliced, which results in exchange of genetic markers; and patched, which results in heteroduplex DNA but no exchange of the flanking genetic markers.

The branch migration step (Fig. 1) can be catalyzed by a DNA strand exchange protein; however, in *E. coli* two proteins, RuvA and RuvB, which form the heterodimer called RuvAB, promote particularly

(b)

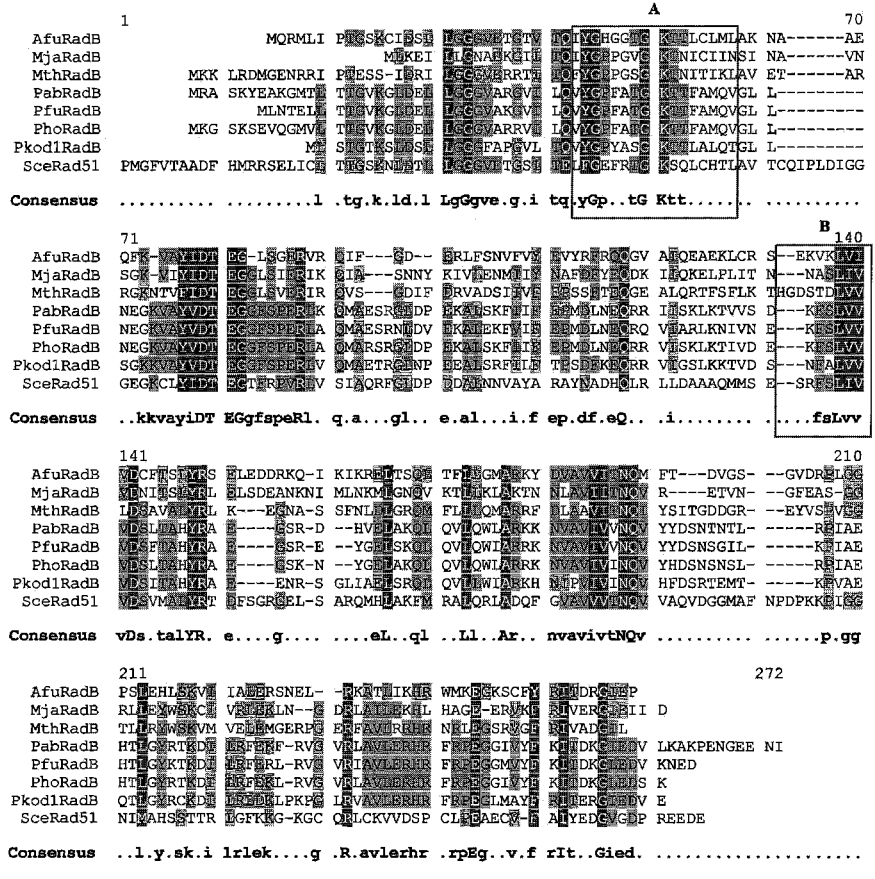


FIG. 16. (*Contd.*)

efficient branch migration (Iwasaki *et al.*, 1992; West, 1997). In addition, the RecG protein has DNA unwinding activity that can promote branch migration (Lloyd and Sharples, 1993; Whitby and Lloyd, 1998).

Holliday-junction cleaving or resolving enzymes are found throughout all three domains of life (Aravind *et al.*, 2000) and are also present in bacteriophages (White *et al.*, 1997). These nucleases are specific for DNA molecules that contain branch points and, in particular, four-way junctions. Holliday-junction resolving enzymes can be divided into three types. Type 1 enzymes cleave Holliday junctions at specific dinucleotide sequences, and members include *E. coli* RuvC, yeast mitochondrial Cce1, *E. coli* RusA (White *et al.*, 1997), and perhaps archaeal

```
SceRad54:   189 DNKEEESKKMIKSTQEKDNINKEKNSQEERPTQRIGRHPALMTNGVRNKP--LRE-LL 244
                D  EEE  K++   +    +        +E+  Q+I      + +   +K   LRE LL
SsoRad54:   356 DISEEEFMKLVSENRTIVELGGNLVEIDEKSLQKIKDLLYKIKSKKIDKIDILRESLL 414

                                                                      I
SceRad54:   245 GDSENSAE--NKKKFASVPVVIDPKLAKI-LRPHQVEG---VRFLNRGAYGCIMADEM 297
                GD E + E  ++ +      +++P   K  LRP+Q++G   +RF+N+  +G  +AD+M
SsoRad54:   415 GDIEINDELLDRLRGNKSFQLLEPYNIKANLRPYQIKGFSWMRFMNKLGFGICLADDM 473

                                          IA
SceRad54:   298 GLGKTLQCIALMWTLLRQGPQGKRLIDKCIIVCPSSLVNNWANELIKWLGPNTLTPLA 356
                GLGKTLQ IA+     ++       +  +++CP S++ NW  EL K+
SsoRad54:   474 GLGKTLQTIAVFSDAKKENE-----LTPSLVICPLSVLKNWEEELSKFAPHLRFAVFH 527

                                                                     II
SceRad54:   357 VDGKKSSMGGGNTTVSQAIHAWAQAQGRNIVKPVIIISYETLRRNVDQLKNCNVGLML 415
                 D  K  +   +                          +++ +Y  L R+  +LK      ++
SsoRad54:   528 EDRSKIKLEDYD----------------------IILTTYAVLLRDT-RLKLVEWKYIV 565

                                          III
SceRad54:   416 ADEGHRLKNGDSLTFTALDSISCHRRVILSGTPIQNDLSEYFALLSFSNPGLLGSRAE 474
                 DE   +KN  +  F A+  +    R+ L+GTPI+N + + ++++++F NPGLLGS +E
SsoRad54:   566 IDEAQNIKNPQTKIFKAVKELKSKYRIALTGTPIENKVDDLWSIMTFLNPGLLGSYSE 624

SceRad54:   475 FRKNFENPILRGRDADATDKEITKGEAQLQKLSTIVSKFIIRRT--NDILAKYLPCKY 531
                F++ F  PI +G   D   KE          +L  I+S FI+RRT     +  +LP K
SsoRad54:   625 FKSKFATPIKKG---DNMAKE---------ELKAIISPFILRRTKYDKAIINDLPDKI 671

SceRad54:   532 EHVIFVNLKPLQNELYNKLIKSREVKKVVKGVGGSQPLRAIGI-------LKKLCNHP 583
                E ++  NL P Q  +Y       EV+ +   +     ++  G+       LK++ +HP
SsoRad54:   672 ETNVYCNLTPEQAAMYKA-----EVENLFNNIDSVTGIKRKGMILSTLLKLKQIVDHP 725

SceRad54:   584 NLLNFEDEFDDEDD-----LELPDDYNMPGSKARDVQTKYSAKFSILERFLHKIKTES 637
                 LL   ++           +E+ ++    G K   + T++     I+   + K   E
SsoRad54:   726 ALLKGGEQSVRRSGKMIRTMEIIEEALDEGDKIA-IFTQFVDMGKIIRNIIEK---EL 780

                IV
SceRad54:   638 DLKIVLISNYTQTLDLIEKMCRYKHYSAVRLDGTMSINKRQKLVDRFNDPEGQEFIFL 696
                + ++  +       +  ++ C +    AV L   +       ++ +F +    +FI +
SsoRad54:   781 NTEVPFLYGELSKKERDDRECSH----AVILFDIIMRTLPDDIISKFQNNPSVKFI-V 834

                 V                    VI
SceRad54:   697 LSSKAGGCGINLIGANRLILMDPDWNPAADQQALARVWRDGQKKDCFIYRFISTGTIE 755
                LS KAGG GINL  ANR+I  D  WNPA + QA  RV+R GQ ++  +++ IS GT+E
SsoRad54:   835 LSVKAGGFGINLTSANRVIHFDRWWNPAVEDQATDRVYRIGQTRNVIVHKLISVGTLE 893

SceRad54:   756 EKIFQRQSMKMSLSSCVVDAKEDVERLFSSDNLRQLFQ 794
                EKI Q  + K SL   ++ + +      S++ LR++ +
SsoRad54:   894 EKIDQLLAFKRSLFKDIISSGDSWITELSTEELRKVIE 932
```

FIG. 17. Comparison of the *S. solfataricus* Rad54 protein homologue with *S. cerevisiae* Rad54. Protein sequences (Sso Rad54 homologue sh13a0224_002&004 and Sce Rad54 protein gi6321275) were aligned using BLAST at http://www.ncbi.nlm.nih.gov/BLAST/ . The seven helicase domains characteristic of Swi2/Snf2 DNA-dependent ATPases are indicated, although the homology in motif IV is weak. Identical residues are represented by the single-letter amino acid code, while highly conserved residues are indicated by a +. Residues that may constitute a leucine zipper motif are circled.

Hjc (Kvaratskhelia and White, 2000a). This sequence requirement is probably important to limit cleavage to the Holliday junction. Type 2 enzymes, on the other hand, which include the bacteriophage enzymes T4 endo VII and T7 endo I, have little or no substrate specificity. These endonucleases can cleave a wide variety of other DNA structures, such

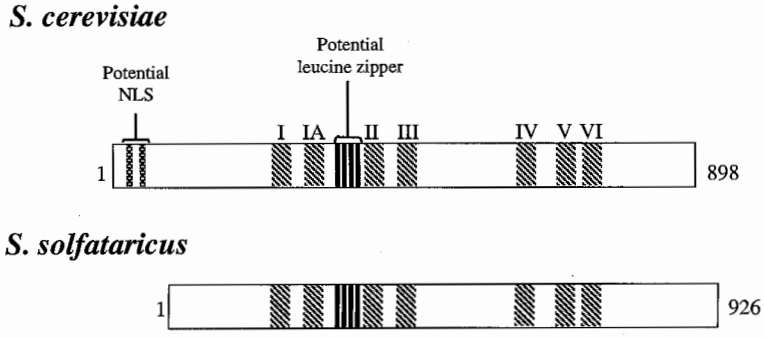


FIG. 18. Schematic representation of the *S. solfataricus* Rad54 protein homologue. Potential nuclear localization signal (NLS) and potential leucine zipper regions are indicated. The seven helicase domains characteristic of Swi2/Snf2 DNA-dependent ATPases are represented by cross-hatched boxes, although the homology in motif IV is weak.

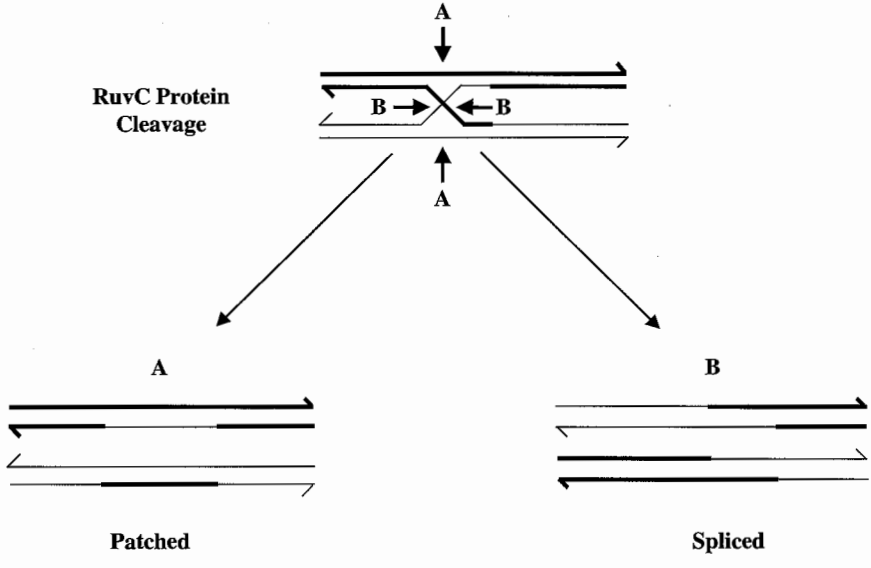


FIG. 19. Holliday-junction cleaving enzymes are responsible for resolution of Holliday junctions in one of two possible orientations. Shown are the products of the endonucleolytic cleavage by the RuvC protein of a Holliday junction in either of two possible orientations, A or B. Cleavage in the A orientation results in a patched recombinant product, while cleavage in the B orientation results in a spliced recombinant product.

as three-way junctions, bulged duplexes, mismatches, and cisplatin adducts (White *et al.,* 1997). The third type of Holliday-junction resolvases is defined by a newly discovered archaeal Hje enzyme. Like type 1, this enzyme shows substrate specificity, but like type 2, it does not exhibit sequence specificity for cleavage (Kvaratskhelia and White, 2000b). Although these Holliday-junction resolving enzymes show the same type of specificity for binding to and cleaving four-way junctions, at the amino acid level these proteins show little or no conservation. Indeed, while a Holliday-junction cleaving activity is detected in yeast nuclei and mammalian extracts, no proteins have been assigned to this activity as of yet (Constantinou *et al.,* 2001).

1. The Bacterial RuvC Protein

The *E. coli* RuvC protein is the prototypic Holliday-junction cleaving enzyme (Bennett and West, 1996; Shah *et al.,* 1997, West, 1997; Eggleston and West, 2000). The crystal structure of RuvC was determined at atomic resolution and demonstrates that the catalytic center, comprising four acidic residues, lies at the bottom of a cleft that fits a DNA duplex (Ariyoshi *et al.,* 1994a,b). The RuvC protein specifically binds four-way Holliday junctions as a dimer and cleaves the strands in a magnesium- and homology-dependent manner. The ssDNA nicks made by RuvC are symmetric; they are found in strands of similar polarity, exclusively on the 3′ site of thymine residues. Strand cleavage by the RuvC dimer occurs in a sequence-specific manner, and the optimal sequence for cleavage is (A~T)TT↓(C >G~A) (Fogg *et al.,* 1999).

2. Archaeal Holliday-Junction Cleaving Enzymes

The first archaeal Holliday-junction cleaving activity was detected in the hyperthermophilic archaeon, *P. furiosus;* the gene was cloned, and the protein was subsequently purified (Komori *et al.,* 1999). This protein, named Hjc (for Holliday-junction cleavage), introduces symmetrically related nicks into two DNA strands of similar polarity, as observed with the *E. coli* RuvC enzyme and other known resolvases. This *P. furiosus* Hjc enzyme resolves Holliday junctions by introducing paired cuts, 3′ to the point of strand exchange, without discernible sequence specificity. The *P. furiosus* Hjc protein does not share any sequence similarity with any of the other known resolvases, although this sequence is highly conserved in the genomes of other archaeons (Table I and Fig. 20). *P. furiosus* Hjc protein cleaves the recombination intermediates that are formed by the *E. coli* RecA protein as efficiently as does the *E. coli* RuvC enzyme (Komori *et al.,* 1999).

FIG. 20. Multiple alignment of archaeal Holliday-junction cleavage protein homologues. Sequences were as follows: *A. fulgidus* (Afu), gi2648580; *M. jannaschii* (Mja), gi2496010; *M. thermoautotrophicum* (Mth), gi2622382; *P. furiosus* (Pfu), gi5689160; *P. horikoshii* (Pho), gi5689160; *A. pernix* (Ape), gi5104108; and *S. solfataricus* (Sso), gi6015898. The sequences were aligned using MULTALIN at http://www.toulouse.inra.fr/multalin.html. Highly conserved residues are shaded in dark gray, while moderately conserved residues are shaded in light gray. The *P. aerophilum* homologue is not shown because the genome sequence has not been publicly released.

The *S. solfataricus* Hjc protein was identified based on homology to the *P. furiosus* Hjc protein and showed 34% amino acid sequence identity to this protein. Additional homologues of the Hjc enzyme were identified in the archaea shown in Fig. 20, plus *Pyrobaculum aerophilum.* These proteins show 35% amino acid identity between them, including 13 totally conserved residues that may function in binding the catalytic metal ions (Fig. 20). This conserved catalytic metal ion binding domain was identified previously in several restriction enzymes and is part of the active site of the type II restriction enzyme *Eco*RV (Kvaratskhelia *et al.,* 2000). Domain analysis of the *P. furiosus* Hjc enzyme also revealed the importance of several residues that confer enzymatic activity to this protein, three of which were found to be conserved in the motif found in type II restriction endonuclease family proteins (Komori *et al.,* 2000a). The *S. solfataricus* Hjc enzyme binds specifically to four-way DNA junctions in a Mg^{2+}-dependent manner, cleaves the junction 3′ to the center of the junction, and may show some sequence specificity for cleavage (Kvaratskhelia and White, 2000a).

Another archaeal Holliday-junction resolving enzyme, Hje (for Holliday-junction endonuclease), was found in two members of the crenarchaeota, *S. solfataricus* and *S. shibatae* (Table I) (Kvaratskhelia and White, 2000b). The partial purification of these enzymes showed

that these endonucleases resolve Holliday junctions in a Mg^{2+}-dependent manner by introducing paired nicks in opposing strands, thereby releasing nicked duplex DNA products. Further experiments showed that the Hje protein does not show sequence specificity for junction cleavage, suggesting that Hje does not belong to the type 1 class of sequence-specific junction resolving enzymes, such as the *E. coli* RuvC and yeast mitochondrial Cce1 proteins. The Hje proteins do not cleave three-way junctions as does the T4 endonucleaseVII enzyme but do discriminate between the continuous and the exchanging strands of the four-way DNA junction to a greater extent than any other known Holliday-junction cleavage enzyme (Kvaratskhelia and White, 2000b). The archaeal Hje enzyme may therefore use this type of discrimination for recognition and resolution of Holliday junctions to achieve specificity without having to rely on the local nucleotide sequence, like the RuvC enzyme. The Hje enzyme introduces a new class of Holliday-junction resolving enzymes that is unlike any of the previously studied enzymes (Kvaratskhelia and White, 2000a). The *S. solfataricus* Hje enzyme produces a cleavage pattern completely different from that of the Hjc enzyme, which suggests that there are two Holliday-junction resolving enzymes in this archaeon (Kvaratskhelia and White, 2000a).

H. Summary: Archaeal Recombinational Repair

The process of homologous DNA recombination in the Archaea has only just begun to be explored. This nascent analysis has been greatly facilitated by the relatively recent sequencing of several different archaeal genomes, since the ability to perform genetic screens in these organisms is still rather difficult due to unusual growth requirements, as well as the inability to transform genetically many members of this group.

The picture emerging for this process in the Archaea is one that shows much more similarity to the pathway of eukaryal homologous DNA recombination than to that of bacterial recombination. Homologues of the eukaryal Spo11 protein, which is involved in the creation of DSBs in meiosis, exist in nearly all members of the Archaea, although it is unclear at this point whether this protein plays a direct role in the initiation of homologous recombination in the Archaea, since it is a subunit of topoisomerase VI. The lack of a bacterial RecBCD enzyme homologue to process the DSB suggests that there is a different initiation or DNA end processing mechanism in the Archaea. Homologues of another eukaryal/bacterial nuclease complex that can process DNA ends are, however, found in the Archaea: the Rad50 and Mre11 proteins (Fig. 1 and Table I). Although their precise role in recombination is uknown, perhaps in conjunction with a DNA helicase, appropriate DSB processing

can be effected. Interestingly, there also exists at least one example of an archaeal homologue of the RecQ/Sgs1 helicase family. Therefore, related mechanisms of DSB processing are likely for the Archaea and Eukarya.

The archaeal homologous DNA strand exchange protein, RadA, clearly shows more homology to the eukaryal Rad51 protein rather than to the bacterial RecA protein, both structurally and functionally. The fact that RadA protein homologues exist in more than 14 archaeons illustrates the importance of this protein in archaeal cellular function, and given the ubiquity of the Rad51 and RecA proteins, all Archaea are expected to have a RadA homologue.

The Archaea also possess an interesting family of ssDNA binding proteins, which likely serves an important function in the processes of DNA replication, recombination, and repair. These proteins are also more similar at the sequence level to the eukaryal RPA, but they display very diverse structural forms. The euryarchaeal proteins closely resemble RPA in that they also incorporate a zinc binding domain within the protein; however, these proteins exist in one- or two-subunit structural variants, rather than the three-subunit quaternary structure of RPA (Fig. 13). In contrast, however, the crenarchaeal protein resembles the bacterial SSB protein in structural form (a single ssDNA binding domain with an acidic tail, which assembles into a tetramer) while retaining sequence similarity to the binding domains of eukaryal RPA.

The existence of *RAD52* epistasis group homologues in the Archaea also substantiates this similarity to the eukaryal process. These homologues include members, known as RadB proteins, that bear similarity to RecA or Rad51 proteins but are distinct from RadA proteins. RadB proteins, which are RadA protein paralogues, may be homologues of Rad55 or Rad57 proteins. A putative Rad54 protein homologue is also present.

Finally, Holliday-junction resolvases exist in the Archaea. While these enzymes do not show homology to any known resolvases, they are able to bind to four-way Holliday junctions and promote their cleavage in a Mg^{2+}-dependent manner, as shown for all other Holliday-junction cleaving enzymes. The Hjc enzyme, present in most archaeons, is a Holliday-junction resolving enzyme, which may show some sequence specificity for cleavage. The Hje enzymes seem to define their own class of Holliday-junction resolvases, in that they do not display any sequence specificity for cleavage of the Holliday junction but do discriminate between stacked four-way junctions that contain continuous and those that contain exchanging strands, which is different from any Holliday-junction resolving enzyme known to date. Until the identification of the eukaryal Holliday-junction resolvases responsible for this step in homologous recombination, it is impossible to say whether the archaeal resolvases resemble eukaryal resolvases.

Thus, the archaeal system does seem to represent a "simpler" version of the complex eukaryal process, but with unique features, and with some features that bear resemblance to those of Bacteria.

III. DNA Repair Pathways

All living cells have many mechanisms for repairing the various types of DNA damage encountered (Lindahl and Wood, 1999). The multiple pathways employed can be divided into several distinct groups: direct reversal of DNA damage, which chemically reverses DNA damage; base excision repair (BER), which removes the damaged base; nucleotide excision repair (NER), which removes lesions in oligonucleotide form; mismatch repair (MMR), which corrects mispaired bases in DNA; and bypass pathways, which involve specialized DNA polymerases that can insert residues opposite damaged sites so that DNA replication can continue. In this article, we focus mainly on the pathways where homologues have been identified or studied in the Archaea. These processes include direct reversal of DNA damage, NER, and BER (Fig. 21). Toward the end of the article we discuss what is known in the other

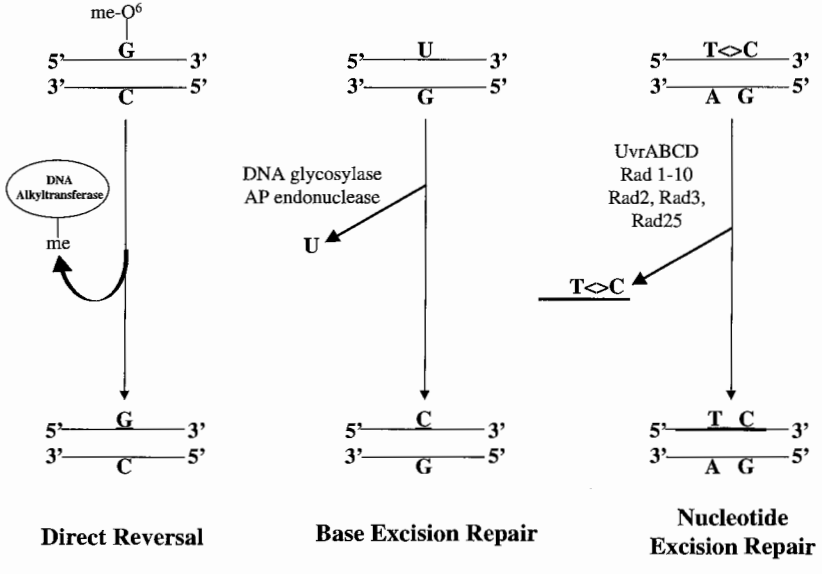


FIG. 21. Three DNA repair pathways common to all phylogenetic domains. Direct reversal chemically reverses the modification and includes the removal of a methyl group from O^6-methylguanine. Base excision repair corrects modifications, such as the incorporation of a uracil residue, by removing a single base. Nucleotide excision repair involves the removal of intact nucleotides, such as a T–C pyrimidine dimer; the lesion is excised as an oligonucleotide, whose length differs for bacterial and eukaryal NER systems.

pathways of MMR and error-prone DNA repair in this phylogenetic domain.

A. Direct DNA Damage Reversal

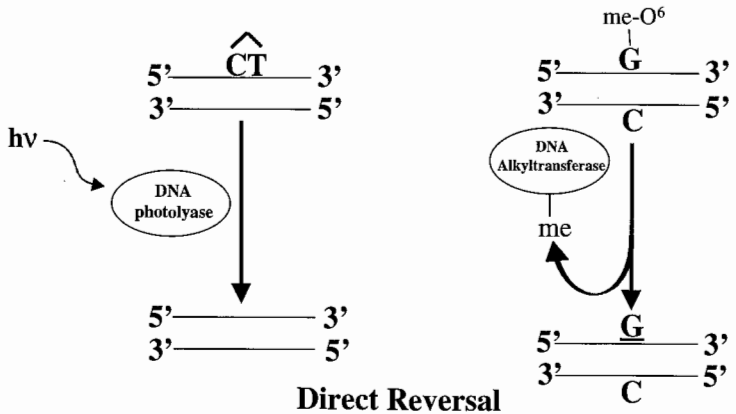
The first DNA repair mode to be discovered was photoreactivation of DNA (Friedberg *et al.*, 1995). Photoproducts in DNA are created by exposure to UV radiation at wavelengths near the absorption maximum of DNA. To repair the major photoproduct formed, a pyrimidine dimer, organisms have a photoreactivation system to reverse the base damage directly. Photoreactivation is a light-dependent process involving the enzyme-catalyzed monomerization of *cis-syn*-cyclobutyl pyrimidine dimers (Fig. 22), and the enzymes that catalyze the photoreactivation of pyrimidine dimers in DNA are referred to as DNA photolyases or photoreactivating enzymes (Friedberg *et al.*, 1995). This activity is widely distributed in nature and exists in the Bacteria, Eukarya, and Archaea (Friedberg *et al.*, 1995; DiRuggiero *et al.*, 1999; Grogan, 2000).

1. Photolyase

Photolyase is able to split dimers using visible light as the source of energy. This enzyme is able to absorb visible or near-UV light because it contains a photochemically active chromophore (reduced FAD) as well as another chromophore which transduces the absorbed energy to the FAD cofactor. In bacteria, such as *E. coli*, the *phrB* gene encodes the DNA photolyase; in lower eukaryotes, such as *S. cerevisiae*, this gene is referred to as *PHR1*. The *E. coli* and *S. cerevisiae* photolyases contain 5,10-methenyltetrahydrofolate (MTHF) as the second chromophore and have an absorption maximum at 380 nm (Sancar *et al.*, 1987; Johnson *et al.*, 1988). However, the Gram-positive bacterium *Streptomyces griseus* and the cyanobacterium *Anacystis nidulans* contain 8-hydroxy-5-deazaflavin as a second chromophore, which has an absorption maximum at 440 nm (Eker *et al.*, 1981, 1990; Yasui *et al.*, 1988; Sack *et al.*, 1998). Photoreactivation activity has been detected in four archaeons *in vivo: H. halobium, M. thermoautotrophicum, S. solfataricus,* and *S. acidocaldarius* (Fig. 22) (Grogan, 2000). The DNA photolyase from *M. thermoautotrophicum* was purified and characterized and was found to have an absorption maximum at 440 nm (Kiener *et al.*, 1989).

2. DNA Alkyltransferases

Another mechanism of DNA damage repair occurs in response to certain mutagenic alkylating agents, which react with DNA to produce both O-alkylated and N-alkylated products. O^6-alkylguanine and O^4-alkylthymine are potentially mutagenic lesions because they can



Direct Reversal

Protein Function	*E. coli*	*S. cerevisiae*	Euryarchaeota	Crenarchaeota
DNA Photolyase	PhrB	Phr1	Activity: *H. halobium, M. thermoautotrophicum* Homologue: *M. thermoautotrophicum* (gi2507184)	Activity: *S. acidocaldarius, S. solfataricus*
DNA alkyl-transferase	Ogt	Mgt1	Activity: *T. litoralis, P. furiosus* Ogt Homologues: *A. aeolicus* (gi2983880), *A. fulgidus* (gi2648205), *M. thermoautotrophicum* (gi2621699), *P. abyssii* (gi5457822), *P. horikoshii* (gi3258272) Mgt1 homologue: *Pyrococcus* KOD1 (gbD86335), *M. jannaschii* (mj1529)	Activity: *P. islandicum, S. acidocaldarius* Ogt Homologues: *A. pernix* (gi5104628), *S. solfataricus* (bac03_008)

FIG. 22. Proteins involved in direct reversal DNA repair that are common to all phylogenetic domains. The table compares proteins involved in the photoreactivation and DNA alkyl transfer processes for the Bacteria, Eukarya, and Archaea.

mispair during semiconservative DNA synthesis. The DNA repair protein, O^6-alkylguanine DNA alkyltransferase (ATase), functions by transferring the problematic alkyl groups from the O^6 position of guanine and the O^4 position of thymine to a cysteine residue at the active site of the protein (Foote *et al.*, 1980; Olsson and Lindahl, 1980). This irreversible process results in the stoichiometric inactivation of the protein.

The *E. coli* enzyme that is responsible for transferring methyl groups from the O^6 position of O^6-methylguanine was originally called O^6-methylguanine DNA methyltransferase, but it is also known as Ada due to its importance in the adaptive response to alkylation damage (Friedberg *et al.*, 1995). This protein is able to recognize methyl groups and larger alkyl groups as substrates. *E. coli* possesses an additional protein, however, called Ogt (a DNA alkyltransferase encoded by the *ogt* gene), which transfers the alkyl groups from O^4-methylthymine and O^6-methylguanine to a cysteine residue in the ATase (Goodtzova *et al.*, 1997). The protein responsible for O^6-alkylguanine DNA alkyltransferase activity in *S. cerevisiae* is the product of the *MGT1* gene and is known as Mgt1 protein. This protein shows conservation with the *E. coli* Ada and Ogt proteins and with the human and mammalian Mgt1 proteins as well (Xiao and Samson, 1992).

In the Archaea, DNA alkyltransferases and DNA methyltransferases were found in several members. The protein MGMT (for O^6- methylguanine DNA methyltransferase) was isolated from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1 and possesses methyltransferase activity at temperatures as high as 90°C (Leclere *et al.*, 1998). Additionally, alkyltransferase activity was detected in cell extracts from two euryarchaeotes, *Thermococcus litoralis* and *P. furiosus*, and two crenarchaeotes, *S. acidocaldarius* and *P. islandicum*. The principal activity of these extracts resembled that of the *E. coli* Ogt protein (Skorvaga *et al.*, 1998). Subsequent analysis of sequenced archaeal genomes revealed Ogt homologues also in *A. aeolicus, A. fulgidus, A. pernix, M. thermoautotrophicum, M. jannaschii, P. abysii, P. horikoshii*, and *S. solfataricus* (Fig. 23) (Grogan, 2000). Figure 23 shows an alignment of eight archaeal Ogt protein homologues with the bacterial Ogt protein from *T. maritima*. These proteins all have a conserved methyl-acceptor cysteine residue. The conservation of these alkyltransferases throughout evolution suggests a strong need for this function, which is most likely due to the toxic and mutagenic consequences of this type of DNA damage.

B. Base Excision Repair

Base excision repair (BER) involves the removal of nonbulky DNA lesions such as uracil, thymine glycols and hydrates, and 8-oxo-guanine in essentially two steps (Fig. 24). First, a DNA glycosylase releases the

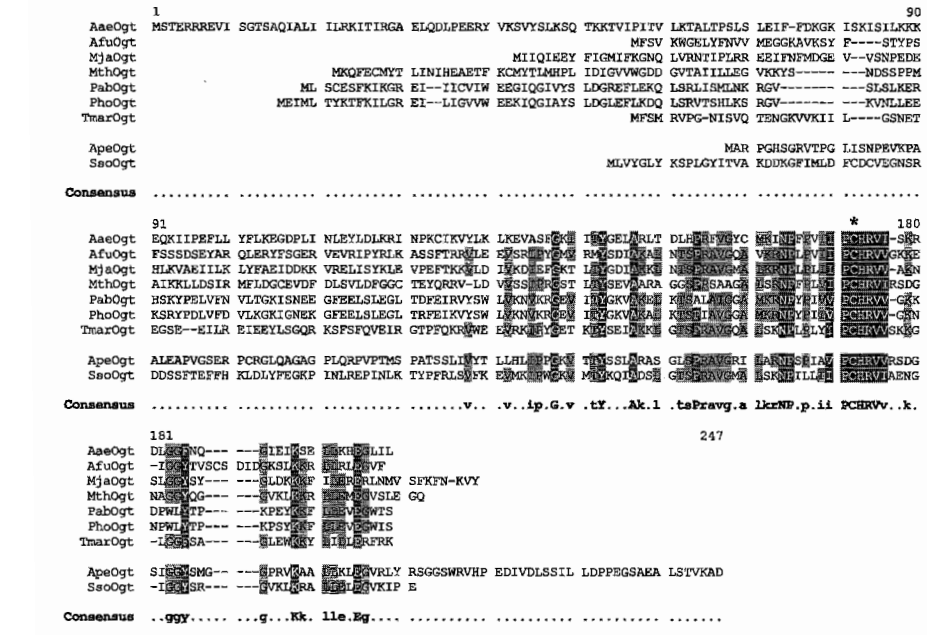


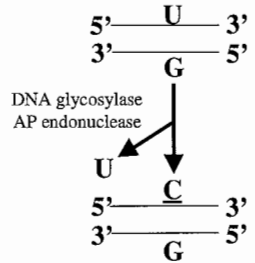
FIG. 23. Multiple alignment of Ogt protein homologues. Sequences were as follows: *Aquifex aeolicus* (Aae), gi2983880; *A. fulgidus* (Afu), gi2648205; *M. jannaschii* (Mja), mj1529; *M. thermoautotrophicum* (Mth), gi2621699; *P. abyssii* (Pab), gi5457822; *P. horikoshii* (Pho), gi3258272; *A. pernix* (Ape), gi5104628; *S. solfataricus* (Sso), bac03–008; and *Thermotoga maritima* (Tmar), gi4981422. *T. maritima* is a member of the Bacteria. The sequences were aligned using MULTALIN at http://www.toulouse.inra.fr/multalin.html. Highly conserved residues are shaded in dark gray, while moderately conserved residues are shaded in light gray. The methyl acceptor cysteine is marked by an asterisk.

base by cleaving the glycosidic bond that connects the base to the deoxyribose. Next, the abasic sugar (apurinic/apyrimidinic (AP) site) is released by the combined actions of AP lyase and AP endonucleases (Friedberg *et al.*, 1995; Sancar, 1996; Wood, 1996).

1. DNA Glycosylases

DNA glycosylases recognize only a certain form of base damage, such as a specific inappropriate base (*e.g.*, uracil) or a specific base mispairing.

FIG. 24. Proteins involved in base excision repair (BER) that are common to all phylogenetic domains. The table compares proteins involved in BER for the Bacteria, Eukarya, and Archaea, showing conserved homologues of a uracil DNA glycosylase, a mismatch glycosylase, an 8-oxoguanine DNA glycosylase, and an apurinic nuclease. Notes: [1]Although reported as UDG homologues (Sandigursky and Franklin, 2000), these sequences are annotated in their respective genomes as DNA polymerase homologues. [2]This protein has also been suggested to be a novel mismatch glycosylase (Horst and Fritz, 1996; Begley *et al.*, 1999) and has been categorized here as a MutY homologue for simplicity.



Base Excision Repair

Protein Function	Bacteria	Eucarya	Euryarchaeota	Crenarchaeota
Uracil DNA Glycosylase	UDG	UDG	Activity: *A. fulgidus, P. islandicum, P. furiosus, T. litoralis* Homologues: *A. fulgidus* (gi2648243)[1], *P. abyssii* (gi3257896)[1], *P. horikoshii* (gi5458117)[1]	Activity: *S. shibatae, S. solfataricus* Homologue: *A. pernix* (gi5104069)[1]
Mismatch Glycosylase	Nth	---	Homologues: *A. fulgidus* (gi2648861), *Halobacterium* (gi10580185), *M. jannaschii* (mj1434), *M. thermoformicicum* (gi232205), *P. aerophilum*, *P. abyssii* (gi5458097), *P. furiosus* (orf1411), *P. horikoshii* (gi3257923)	Homologue: *S. solfataricus* (c04_006)
	MutY	---	Homologue: *Halobacterium* (gi10581009), *M. thermoautotrophicum* MIG (gi2621835)[2]	Homologue: *A. pernix* (gi5104542)
8-oxoguanine DNA Glycosylase	oxoG	Ogg1	Activity: *M. jannaschii* (mjOgg) Homologue: *M. jannaschii* (gi2833558)	---
Apurinic Endonuclease	Endo IV	Apn	Endo IV Homologues: *M. jannaschii* (mj1614), *M. thermoautotrophicum* (gi8928109)	---
	Endo V	---	Homologue: *M. thermoautotrophicum* (gi2622612)	

DNA glycosylases were first identified in *E. coli* but are ubiquitous in nature. Generally speaking, DNA glycosylases are small, single-subunit proteins that have no cofactor requirement. These enzymes recognize the presence of damaged or mismatched bases and catalyze the breakage of the glycosyl bond between the base and the DNA sugar–phosphate backbone. Some of these enzymes have an associated AP lyase activity that produces 3′-α,β-unsaturated aldehyde and 5′-phosphate products (McCullough *et al.*, 1999). Glycosylase action, or the loss of purines or pyrimidines, results in the production of a common intermediate, the AP site. These sites are further processed by the AP endonucleases or AP lyases that cleave the phosphodiester bond either 5′ or 3′ to the AP site, respectively. This site is then processed further to yield a 3′-OH suitable for polymerization and ligation (Sancar, 1996).

a. Uracil DNA Glycosylases. Deamination of cytosine results in the formation of a uracil base. Since uracil will base-pair with adenine, cytosine deamination results in a transition mutation from G–C to A–T, if the uracil-containing strand is used as a replication template (Friedberg *et al.*, 1995). DNA glycosylases that excise uracil or thymine at the *N*-glycosidic bond can be classified into two major types according to amino acid sequence and function. The first type is uracil DNA glycosylase (UDG), which excises uracil from both ss- and dsDNA (U/G and U/A mispairs). This type of enzyme does not, however, excise thymine from T/G mismatches. UDG is found in all organisms, and there is 56% amino acid sequence identity between *E. coli* UDG and human UDG (Olsen *et al.*, 1989; Krokan *et al.*, 1997). The second type of DNA glycosylase includes a mismatch-specific uracil DNA glycosylase (MUG), found in *E. coli* and *Serratia marcescens*, and thymine DNA glycosylase (TDG) from humans (Neddermann *et al.*, 1996). MUG and TDG recognize the mismatched basepairs in dsDNA and remove both mismatched uracil and thymine. TDG recognizes and repairs U/G and T/G mispairs equally, while MUG is mostly U/G mispair specific. MUG has 32% amino acid identity with the central part of human TDG.

A uracil DNA glycosylase (UDG) was first described based on protein activity in the archaea *S. shibatae, S. solfataricus, P. islandicum, P. furiosus,* and *T. litoralis* (Fig. 24) (Koulis *et al.*, 1996). Subsequent to this discovery, a UDG from the archaeon *A. fulgidus* was isolated (Sandigursky and Franklin, 2000). These enzymes showed biochemical characteristics similar to those of the *E. coli* enzyme, as well as the same enzyme from the thermophilic bacterium *T. maritima* (Sandigursky and Franklin, 1999). This archaeal UDG enzyme can remove uracil opposite guanine, as would occur in DNA after cytosine deamination. However, this glycosylase was not able to remove thymine from a similar substrate containing a T–G base pair, which is similar to the

activity of the *T. maritima* UDG (Sandigursky and Franklin, 1999). Additional homologues of this protein exist in *P. horikoshii, P. abysii,* and *A. pernix* and were identified based on amino acid sequence homology (Fig. 24) (Sandigursky and Franklin, 2000).

b. Mismatch Glycosylases. A mismatch glycosylase (Mth-MIG) that shows functional similarity to MUG/TDG glycosylases was discovered encoded on the cryptic plasmid pV1 of *M. thermoautotrophicum* (Fig. 24). Mth-MIG processes U/G and T/G but not U on a single strand of DNA (Horst and Fritz, 1996; Begley *et al.,* 1999). Mth-MIG shows little amino acid similarity to MUG/TDG and UDG but shows significant sequence similarity to the [4Fe–4S]-containing Nth/MutY DNA glycosylase family, which catalyzes *N*-glycosylic reactions on DNA substrates other than U/G and T/G mispairs and which are conserved in both the Bacteria and the Eukarya. These types of DNA glycosylases include DNA endonuclease III (Nth; thymine glycol DNA glycosylase), MutY DNA glycosylase (A/G-specific adenine glycosylase), UV endonuclease (UV endo), and methylpurine DNA glycosylase II (MpgII). The unique structural and functional characteristics of Mth-MIG suggest that it is a new type of U/G and T/G mismatch-specific glycosylase. Another putative homologue of this protein was identified in the archaeon *M. jannaschii* based upon sequence homology to endonuclease III (Fig. 24) (Begley *et al.,* 1999).

An additional DNA glycosylase with significant sequence homology to [4Fe–4S]-containing Nth/MutY DNA glycosylases was discovered in the hyperthermophilic archaeon *P. aerophilum* (Fig. 24) (Yang *et al.,* 2000). This protein, Pa-MIG, shows 34% amino acid identity to the *M. thermoformicicum* Mth-MIG protein and 30% amino acid identity to the *E. coli* MutY protein. This protein also has amino acid residues that are generally conserved in the [4Fe–4S]-containing Nth/MutY DNA glycosylase family (Lu and Fawcett, 1998; Yang *et al.,* 2000). The Pa-MIG protein also has a conserved tyrosine residue that is conserved among all Nth proteins and is critical for associated AP lyase activity. Biochemically, the Pa-MIG protein processes both U/G and T/G mismatches and may have a weak AP lyase activity associated with the enzyme, as does the *E. coli* MutY enzyme. This protein could also process T/7,8-dihydro-8-oxoguanine (GO) and U/GO substrates but could not process A/G and A/GO mispairs, which are substrates for the MutY protein, or G/G and G/GO mispairs. Members of this Nth/MutY/MIG/MpgII/UV endoglycosylase superfamily can also be found in *A. pernix, A. fulgidus, M. jannaschii,* and *P. horikoshii* (Yang *et al.,* 2000). Figure 25 shows an alignment of nine archaeal members of this DNA glycosylase family. The conserved lysine residue within the Nth protein family is indicated, and the cysteine residues involved in the [4Fe–4S] binding cluster are also indicated. *M. thermoformicicum* Mth-MIG is not indicated due

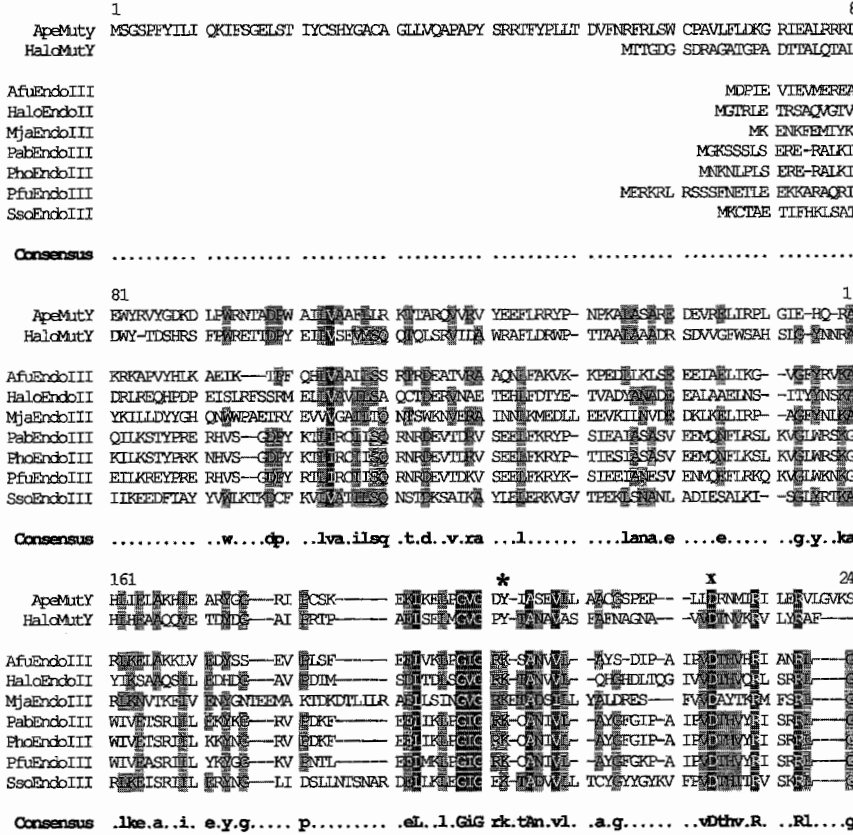


FIG. 25. Multiple alignment of archaeal MutY and endonuclease III protein homologues. Sequences were as follows: *A. pernix* (Ape), gi5104542; *Halobacterium* (HaloMutY), gi10581009; *A. fulgidus* (Afu), gi2648861; *Halobacterium* (HaloEndoIII), gi10580185; *M. jannaschii* (Mja), mj1434; *P. abyssii* (Pab), gi5458097; *P. furiosus* (Pfu), orf1411; *P. horikoshii* (Pho); gi3257923; and *S. solfataricus* (Sso), gi3257923. The sequences were aligned using MULTALIN at http://www.toulouse.inra.fr/multalin.html. Highly conserved residues are shaded in dark gray, while moderately conserved are residues are shaded in light gray. The conserved lysine residue within the Nth family is marked with an asterisk. The strictly conserved aspartic acid residue is indicated by an "x." The cysteine residues involved in binding the [4Fe–4S] cluster are marked with dots.

to the incompletion of this genome sequencing project at this date, and *P. aerophilum* is not indicated due to restrictions on obtaining the sequences. The archaeal MIG family is remotely related to the human MBD4 thymine glycosylase (Pa-MIG shows 21% amino acid identity in the glycosylase domain to human MBD4 protein), which also repairs T/G and U/G mismatches in dsDNA. The C-terminal catalytic domain of the human MBD4 protein shows homology to *E. coli* endonuclease III and MutY proteins (Petronzelli *et al.,* 2000).

2. 8-Oxoguanine DNA Glycosylases

Another member of the DNA glycosylase family that has a homologue in the Archaea is 8-oxoguanine DNA glycosylase (Gogos and Clarke, 1999). 8-Oxoguanine (oxoG) is caused by oxidizing agents or ionizing radiation and can be highly mutagenic if not repaired properly. DNA glycosylases that are specific for this oxoG type of lesion were discovered throughout the Bacteria and Eukarya, although they do not appear to belong to the same family. The eukaryal oxoG DNA glycosylases of yeast and mammals (Ogg 1 protein in *S. cerevisiae* and humans) belong to a protein sequence-related family of DNA glycosylases whose members have a wide range of specificities. The bacterial enzymes, however, such as the *E. coli* MutM enzyme (or Fpg), make up their own distinct family that share sequence conservation, require zinc for activity, and have a strong δ-elimination activity (Girard *et al.,* 1997). An oxoG DNA glycosylase was identified, based on sequence homology to the DNA glycosylase superfamily, in the euryarchaeote *M. jannaschii,* and its gene product purified (Fig. 24). This protein, called mjOgg, is distantly related to other known oxoG-specific enzymes belonging to the same glycosylase superfamily and shows no greater sequence homology with the eukaryal Ogg1 protein than other members. mjOgg shows DNA glycosylase activity and a specificity for oxoG. This enzyme also has an associated DNA lyase activity (Gogos and Clarke, 1999).

3. AP Endonucleases

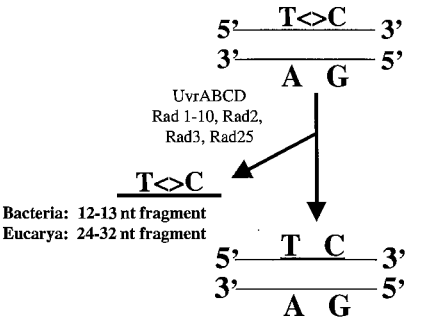
The AP endo/endonuclease IV family is another class of enzymes involved in BER that have putative representatives in the Archaea, based on sequence analysis. Homologues have been found in *M. jannaschii* and *M. thermoautotrophicum* (Fig. 24). Following the release of free, damaged, or inappropriate bases by DNA glycosylases, AP sites are produced. The repair of these lesions is initiated by AP endonucleases, which catalyze the incision of DNA exclusively at AP sites, and this prepares the DNA for subsequent excision, repair synthesis, and DNA ligation. Endonuclease IV, encoded by the *nfo* gene in *E. coli,* catalyzes the formation of ssDNA breaks at sites of base loss in duplex DNA.

Endo IV attacks phosphodiester bonds 5′ to the sites of base loss in DNA, leaving 3′-OH groups. The bacterial endo IV protein is a homologue of eukaryotic apurinic endonucleases (Aravind *et al.*, 1999). Additionally, a homologue of *E. coli* Nfi, or endonuclease V, was tentatively identified, based on sequence homology, in *M. thermoautotrophicum* (Fig. 24) (Aravind *et al.*, 1999). These putative protein homologues have yet to be studied biochemically.

C. Nucleotide Excision Repair

Another ubiquitous repair pathway is the nucleotide excision repair (NER) pathway (Friedberg *et al.*, 1995; Sancar, 1996). During NER, damaged bases such as pyrimidine dimers and (6–4) photoproducts are enzymatically excised from DNA as intact nucleotides that are a part of an oligonucleotide fragment (Fig. 26). There are two excision mechanisms. One is via an endonuclease–exonuclease mechanism, where an endonuclease makes an incision at a phosphodiester bond either 5′ or 3′ to the lesion, and then an exonuclease digests the damaged strand past the lesion. The second mechanism involves the action of an excision nuclease (excinuclease), which incises the phosphodiester bonds on either side of the lesion and at some distance away from the lesion, to excise the lesion in a nucleotide fragment of a unique length. The fragment and UvrC protein are then released by the action of a DNA helicase (UvrD protein, or helicase II, in *E. coli*) (Fig. 26) (Friedberg *et al.*, 1995; Sancar, 1996).

NER has been characterized in detail in both the Bacteria and the Eukarya, where the damage to the DNA is excised by the combined actions of several proteins in an ATP-dependent manner. The multisubunit complex that comprises the excinuclease in *E. coli* is made up of the UvrA, UvrB, and UvrC proteins (Sancar, 1996). The UvrA protein functions in recognizing the site of DNA damage, while the UvrB and UvrC proteins catalyze the excision reaction, hydrolyzing the eighth phosphodiester bond on the 5′ side of the damaged base or bases and the fourth or fifth phosphodiester bond on the 3′ side of the damaged base or bases. This leads to the excision of the lesion in the form of a 12- to 13-nucleotide fragment (Sancar, 1996). The UvrD protein (helicase II) then releases the oligonucleotide fragment as well as the DNA-bound UvrC protein. The eukaryal excinuclease incises the 20th–25th phosphodiester bond 5′ and the 3rd–8th phosphodiester bond 3′ to the lesion to generate 24- to 32-nucleotide fragments (Fig. 26). This NER system, however, involves the action of many more proteins than the bacterial process and is, thus, much more complex. None of the protein subunits that make up the eukaryal excinuclease show any significant homology



Nucleotide Excision Repair

Protein Function	Bacteria	Eucarya	Euryarchaeota	Crenarchaeota
Excinuclease	UvrABCD	---	Activity: *M. thermoautotrophicum*	---
	UvrA	---	Homologue: *M. thermoautotrophicum* (MT443)	
	UvrB	---	Homologue: *M. thermoautotrophicum* (MT442)	
	---	Rad1	Homologues: *A. fulgidus* (AF0264), *M. jannaschii* (MJ1505), *M. thermoautotrophicum* (MT1415)	
	---	Rad2	Homologues: *A. fulgidus* (AF0264), *M. thermoautotrophicum* (MT1633), *P. abyssii* (PAB1877)	
	---	Rad3	Homologue: *P. abyssii* (PAB2385)	

FIG. 26. Proteins involved in nucleotide excision repair (NER) that are common to all phylogenetic domains. The table compares proteins involved in NER for the Bacteria, Eukarya, and Archaea, showing the conserved excinucleases involved in this process.

to the bacterial enzyme. The eukaryal system, however, is conserved throughout the Eukarya (Wood, 1996).

When the Archaea were explored for the presence of NER activity, the activity was found to be more similar to that of the bacterial system. The first experiments, using a cell extract from *M. thermoautotrophicum,* demonstrated the release of an oligomer containing the lesion that was 10–11 nucleotides in length (Ogrunc *et al.,* 1998). This finding paralleled the results with the purified *E. coli* excinuclease, which released a 12-mer fragment, whereas the mammalian excinuclease released a 27-mer fragment. The archaeal reaction was ATP dependent, in accordance with the behavior of both the bacterial and the eukaryal excinucleases. This archaeon also has UvrA and UvrB homologues, based on sequence homology (Fig. 26) (Grogan, 2000).

The mechanism of NER seems to differ, however, for other members of the Archaea, and homologues of the eucaryal NER system were detected. These include homologues of Rad1, Rad2, Rad3, Rad25, and Rad27, as well as mouse ERCC1 and human XP-F proteins (Fig. 26) (Aravind *et al.,* 1999; Grogan, 2000). In the Eukarya, two nucleases are used to create the dual incisions during NER. In *S. cerevisiae,* the nucleases are the Rad2 protein and the Rad1–10 protein complex (Game, 1993, 2000; Prakash and Prakash, 2000). The Rad1 and Rad10 proteins form a complex that has a ssDNA endonuclease activity which cleaves 3′-ended ssDNA at the junction with duplex DNA (Rad1–10). The Rad2 protein also has ssDNA endonuclease activity. Homologues of the yeast Rad1 protein were uncovered in the archaea *M. jannaschii, A. fulgidus,* and *M. thermoautotrophicum,* although none was found in bacteria (Aravind *et al.,* 1999). All of the nucleases from this Rad1 family of proteins contain a conserved $ERKX_2SD$ motif and a conserved aspartate residue. The archaeal homologues predict, interestingly, an N-terminal helicase domain that is normally inactive in eukaryotes (Aravind *et al.,* 1999). Putative homologues of Rad2 were identified in *P. abysii, A. fulgidus,* and *M. thermoautotrophicum* (Fig. 26) (Aravind *et al.,* 1999; Grogan, 2000). Two helicases in *S. cerevisiae,* Rad3 and Rad25, are also involved in NER. These helicases are responsible for creating a bubble structure during NER (Prakash and Prakash, 2000), and a homologue of the Rad3 helicase was identified in *P. abysii* (Fig. 25) (Grogan, 2000).

D. Mismatch Repair

Both bacterial and eukaryal organisms can repair mismatched DNA base pairs. Mismatches arise by several mechanisms, including errors generated during the process of DNA replication, the formation of

heteroduplex DNA as part of the recombination process, and through the deamination of 5-methylcytosine. This type of modified base can be found in the DNA of many organisms, from bacteria to eukaryotes. Deamination causes the conversion of a G–5-mC base pair to a G–T base pair (Friedberg *et al.,* 1995; Yang, 2000).

The basic enzymology of the major MMR processes is very similar in bacteria and eukaryotes. MMR in *E. coli* has been studied extensively and occurs via a methyl-directed MutHLS system. MutS protein initiates this process by binding, as a homodimer, to base–base mismatches and loop insertions–deletions that may have arisen due to polymerase misincorporation and slippage errors, respectively. This MutS repair complex then recruits a MutL protein homodimer, which activates the endonuclease activity of MutH. The ATP binding and hydrolysis activities of MutS and MutL proteins may cause conformational changes to regulate binding to mismatches and subsequent interactions with other factors such as MutH. Once MutH is activated, its endonuclease activity is directed to incise the newly replicated DNA strand at hemimethylated sites formed after the passage of the replication fork. The nicked strand is then unwound by the activity of helicase II and degraded back past the mismatch, either by 5′-to-3′ or by 3′-to-5′ exonucleases, and repair synthesis fills in the resulting gap (Modrich, 1991; Yang, 2000).

Unlike the system in *E. coli, S. cerevisiae* has six MutS protein homologues, which are referred to as MutS homologue (MSH) proteins (Kolodner and Marsischky, 1999). In yeast, MMR begins with a MSH2 protein recognizing the mismatch and forming a heterodimer with either a MSH3 or a MSH6 protein to bind the mismatches; each of the latter provides specificity for the type of error that is recognized (Eisen, 1998; Kolodner and Marsischky, 1999). The roles of the other MutS homologues in yeast are not as well understood. MSH1 protein is involved in MMR in mitochondrial DNA, although the function of this protein has not yet been completely characterized (Chi and Kolodner, 1994). The MSH4 and MSH5 proteins are not involved in MMR but, instead, function during meiotic crossing-over and chromosome segregation (Pochart *et al.,* 1997). Mismatch recognition and repair mechanisms in humans and other higher eukaryotes show similarity to those that exist in yeast (Fishel and Wilson, 1997; Kolodner and Marsischky, 1999).

The Archaea, so far, have been shown to possess only a single MutS protein homologue (Eisen, 1998; Aravind *et al.,* 1999). The putative MutS protein homologue was detected in only one member of the Archaea, *M. thermoautotrophicum* (Eisen, 1998), based on sequence homology to the *E. coli* MutS protein; however, this MutS protein homologue was shown to group closer to a subgroup of MutS protein homologues that includes MSH4 and MSH5, which are chromosome

crossover and segregation proteins (Eisen, 1998). There is no biochemical characterization of this protein as of yet.

E. Flap Endonuclease Protein Homologues

DNA structures containing single-stranded branches or "flaps" are found as intermediates of DNA replication, recombination, or repair (DeMott *et al.*, 1996; Bambara *et al.*, 1997). Degradation of these flap structures during these processes is carried out by a protein known as FEN-1 (flap endonuclease-1). This protein possesses 5′-to-3′ exonuclease activity and can act as an endonuclease for 5′ ssDNA flaps. FEN-1 protein homologues were discovered in several members of the Archaea: *A. fulgidus, P. furiosus, M. jannaschii,* and *P. horikoshii* (Hosfield *et al.*, 1998a; Rao *et al.*, 1998; Matsui *et al.*, 1999). These proteins show a high level of sequence homology with the human FEN-1 protein; the *M. jannaschii* FEN-1 homologue shows 76% amino sequence similarity, and the homologues from *A. fulgidus* and *P. furiosus* show 72 and 74% amino sequence similarity, respectively. The *A. fulgidus, P. furiosus, M. jannaschii,* and *P. horikoshii* FEN-1 protein homologues were purified, and they show specificity for flap DNA structures (Hosfield *et al.*, 1998a; Rao *et al.*, 1998; Matsui *et al.*, 1999). The FEN-1 protein from *P. furiosus* was crystallized, and the structure was determined (Hosfield *et al.*, 1998b).

F. Translesion DNA Synthesis and Mutagenesis

In the bacterium *E. coli,* mutagenesis that occurs after exposure to DNA-damaging agents requires a distinct system (the SOS-induced mutagenesis system), which processes DNA damage in an error-prone manner. Several genes in *E. coli* are regulated by the SOS system, and two of these are error-prone DNA polymerases: $UmuD'_2C$, which is also referred to as PolV (Tang *et al.*, 1999; Goodman, 2000); and DinB, which is referred to as PolIV (Wagner *et al.*, 1999). Homologues of the *E. coli* DinB protein were discovered in *S. cerevisiae, C. elegans, M. musculus,* and *H. sapiens* (Gerlach *et al.*, 1999; Woodgate, 1999). In yeast, the Rad30 protein is homologous to both the UmuC and the DinB proteins and is a DNA polymerase (DNA pol η) that can replicate thymine dimers in template DNA (Johnson *et al.*, 1999). Additionally, a human homologue of yeast Rad30 (*Xeroderma pigmentosum* variant; XPV) shows activities similar to those of the yeast pol η (Masutani *et al.*, 1999a; Masutani *et al.*, 1999b).

A DinB/UmuC protein homologue was identified by sequence analysis in the archaeon, *S. solfataricus* (Kulaeva *et al.*, 1996). This protein

homologue shows 32% sequence similarity to the DinB protein and 22% sequence similarity to the UmuC protein. Additionally, DNA mutagenesis induced by exposure to UV radiation was detected in the *Pyrococcus* species of archaea (Watrin and Prieur, 1996). Biochemical characterization of this archaeal protein homologue is yet to be reported.

G. Summary: DNA Repair Mechanisms in the Archaea

As discussed above, recombinational repair in the Archaea shares more orthologous protein components with the eukaryal system than with the bacterial system, based on the similarities with many components of the yeast *RAD52* epistasis group.

However, the comparison of other DNA repair pathways has not produced a simple conclusion. Proteins involved in the direct reversal of DNA damage are similar in both bacteria and eukaryotes, and the archaeal protein homologues show similarities to both as well. The archaeal DNA alkyltransferases, however, show homology to the bacterial Ogt protein.

The archaeal DNA glycosylases involved in BER show homology to both bacterial and eukaryal enzymes, a consequence of the fact that many bacterial DNA glycosylases are also conserved in the Eukarya. The archaeal UDG protein displays both biochemical and sequence similarity to bacterial UDG proteins. The Archaea have a mismatch glycosylase with homology to the Nth/MutY/MIG/MpgII/UV endoglycosylase superfamily, which is also conserved in both the Bacteria and the Eukarya. An archaeal 8-oxoguanine DNA glycosylase exists in *M. jannaschii,* but the sequence of this enzyme differs greatly from those of both its eukaryal and its bacterial counterparts. Finally, the members of the AP endo/endonuclease IV family in the Archaea are similar in sequence to the bacterial proteins.

In the case of NER, the archaeal proteins show similarities in some cases to the bacterial proteins and in other species to the eukaryal proteins. An activity was identified in *M. thermoautotrophicum* that mimics the action of the UvrABCD proteins, and UvrA and B protein homologues exist, based on sequence similarity, in this archaeon. However, in other archaea, protein homologues of the eukaryal NER machinery were detected.

Less information is available on the processes of mismatch repair and error-prone DNA repair in this third domain of life. So far, only one MutS homologue was found; although this homologue was discovered based on sequence homology to the *E. coli* MutS protein, it groups closer to a subgroup that includes eukaryal MutS protein homologues. Another protein involved in DNA replication, recombination,

and repair, FEN-1 protein, has homologues in several archaea, and these show a high degree of sequence homology with the human FEN-1 protein. Finally, a homologue of a bacterial protein involved in error-prone DNA replication, DinB/UmuC, was found in just one member of the Archaea.

In conclusion, it appears that the Archaea possess proteins involved in DNA repair that are similar to both bacterial and eukaryal components and some proteins that are only distantly related to either. For this reason, it is difficult to classify the entire archaeal domain as being "more" bacterial or eukaryal in its means for repairing damage to its DNA. Further investigation into the processes by which the Archaea are able to repair DNA damage will reveal mechanisms by which this unique domain of life deals with the classic problem of DNA damage and should lend insight into the evolution of DNA repair processes.

ACKNOWLEDGMENTS

We thank the following members of the Kowalczykowski lab for providing comments on the manuscript: Piero Bianco, Carole Bornarth, Joel Brockman, Frédéric Chédin, Mark Dillingham, Naofumi Handa, Alex Mazin, Jim New, and Yun Wu. We also thank Dr. John A. Tainer for providing the Rad50 protein structural figure. This work was supported by NIH Training Grant GM07377 to E.M.S., NSF Postdoctoral Fellowship in Microbial Biology 0074380 to C.A.H., and NIH Grants GM62653 and GM 41347 and Human Frontiers Science Program Grant HFSP-RG63 to S.C.K.

REFERENCES

Alani, E., Padmore, R., and Kleckner, N. (1990). *Cell* **61,** 419–436.

Albala, J. S., Thelen, M. P., Prange, C., Fan, W., Christensen, M., Thompson, L. H., and Lennon, G. G. (1997). *Genomics* **46,** 476–479.

Anderson, D. G., and Kowalczykowski, S. C. (1997a). *Genes Dev.* **11,** 571–581.

Anderson, D. G., and Kowalczykowski, S. C. (1997b). *Cell* **90,** 77–86.

Anderson, D. G., and Kowalczykowski, S. C. (1998). *J. Mol. Biol.* **282,** 275–285.

Aravind, L., Walker, D. R., and Koonin, E. V. (1999). *Nucleic Acids Res.* **27,** 1223–1242.

Aravind, L., Makarova, K. S., and Koonin, E. V. (2000). *Nucleic Acids Res.* **28,** 3417–3432.

Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Fujishima, A., Shinagawa, H., and Morikawa, K. (1994a). *J. Mol. Biol.* **241,** 281–282.

Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994b). *Cell* **78,** 1063–1072.

Arnold, D. A., and Kowalczykowski, S. C. (1999). *In* "Encyclopedia of Life Sciences." Nature Publishing Group, London, www.els.net.

Arnold, D. A., and Kowalczykowski, S. C. (2000). *J. Biol. Chem.* **275,** 12261–12265.

Bambara, R. A., Murante, R. S., and Henricksen, L. A. (1997). *J. Biol. Chem.* **272,** 4647–4650.

Baudat, F., and Nicolas, A. (1997). *Proc. Natl. Acad. Sci. USA* **94,** 5213–5218.

Baudat, F., Manova, K., Yuen, J. P., Jasin, M., and Keeney, S. (2000). *Mol. Cell* **6,** 989–998.

Begley, T. J., Haas, B. J., Noel, J., Shekhtman, A., Williams, W. A., and Cunningham, R. P. (1999). *Curr. Biol.* **9,** 653–656.

Bennett, R. J., and West, S. C. (1996). *Proc. Natl. Acad. Sci. USA* **93,** 12217–12222.
Bennett, R. J., Sharp, J. A., and Wang, J. C. (1998). *J. Biol. Chem.* **273,** 9644–9650.
Bennett, R. J., Noirot-Gros, M. F., and Wang, J. C. (2000). *J. Biol. Chem.* **275,** 26898–26905.
Bergerat, A., Gadelle, D., and Forterre, P. (1994). *J. Biol. Chem.* **269,** 27663–27669.
Bergerat, A., de Massy, B., Gadelle, D., Varoutas, P. C., Nicolas, A., and Forterre, P. (1997). *Nature* **386,** 414–417.
Bianco, P. R., and Kowalczykowski, S. C. (1997). *Proc. Natl. Acad. Sci. USA* **94,** 6706–6711.
Bianco, P. R., and Kowalczykowski, S. C. (1999). *In* "Encyclopedia of Life Sciences." Nature Publishing Group, London, www.els.net.
Bianco, P. R., Tracy, R. B., and Kowalczykowski, S. C. (1998). *Front. Biosci.* **3,** D570–D603.
Bochkareva, E., Frappier, L., Edwards, A. M., and Bochkarev, A. (1998). *J. Biol. Chem.* **273,** 3932–3936.
Borde, V., Wu, T. C., and Lichten, M. (1999). *Mol. Cell. Biol.* **19,** 4832–4842.
Borde, V., Goldman, A. S., and Lichten, M. (2000). *Science* **290,** 806–809.
Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., Fitzgerald, L. M., Clayton, R. A., and Gocayne, J. D. (1996). *Science* **273,** 1058–1073.
Cartwright, R., Dunn, A. M., Simpson, P. J., Tambini, C. E., and Thacker, J. (1998a). *Nucleic Acids Res.* **26,** 1653–1659.
Cartwright, R., Tambini, C. E., Simpson, P. J., and Thacker, J. (1998b). *Nucleic Acids Res.* **26,** 3084–3089.
Celerin, M., Merino, S. T., Stone, J. E., Menzie, A. M., and Zolan, M. E. (2000). *EMBO J.* **19,** 2739–2750.
Cervantes, M. D., Farah, J. A., and Smith, G. R. (2000). *Mol. Cell* **5,** 883–888.
Chakraverty, R. K., and Hickson, I. D. (1999). *Bioessays* **21,** 286–294.
Chédin, F., Noirot, P., Biaudet, V., and Ehrlich, S. D. (1998a). *Mol. Microbiol.* **29,** 1369–1377.
Chédin, F., Seitz, E. M., and Kowalczykowski, S. C. (1998b). *Trends Biochem. Sci.* **23,** 273–277.
Chédin, F., Ehrlich, S. D., and Kowalczykowski, S. C. (2000). *J. Mol. Biol.* **298,** 7–20.
Chi, N. W., and Kolodner, R. D. (1994). *J. Biol. Chem.* **269,** 29984–29992.
Clark, A. J., and Margulies, A. D. (1965). *Proc. Natl. Acad. Sci. USA* **53,** 451–459.
Clark, A. J., and Sandler, S. J. (1994). *Crit. Rev. Microbiol.* **20,** 125–142.
Connelly, J. C., and Leach, D. R. (1996). *Genes Cells* **1,** 285–291.
Connelly, J. C., de Leau, E. S., Okely, E. A., and Leach, D. R. (1997). *J. Biol. Chem.* **272,** 19819–19826.
Connelly, J. C., Kirkham, L. A., and Leach, D. R. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 7969–7974.
Connelly, J. C., de Leau, E. S., and Leach, D. R. (1999). *Nucleic Acids Res.* **27,** 1039–1046.
Constantinou, A., Davies, A. A., and West, S. C. (2001). *Cell* **104,** 259–268.
Cox, M. M. (1999). *Prog. Nucleic Acid Res. Mol. Biol.* **63,** 311–366.
Cox, M. M., and Lehman, I. R. (1981). *Proc. Natl. Acad. Sci. USA* **78,** 6018–6022.
Cromie, G. A., Millar, C. B., Schmidt, K. H., and Leach, D. R. (2000). *Genetics* **154,** 513–522.
Cui, X., Brenneman, M., Meyne, J., Oshimura, M., Goodwin, E. H., and Chen, D. J. (1999). *Mutat. Res.* **434,** 75–88.
Debrauwere, H., Buard, J., Tessier, J., Aubert, D., Vergnaud, G., and Nicolas, A. (1999). *Nature Genet.* **23,** 367–371.
DeMott, M. S., Shen, B., Park, M. S., Bambara, R. A., and Zigman, S. (1996). *J. Biol. Chem.* **271,** 30068–30076.
Dernburg, A. F., McDonald, K., Moulder, G., Barstead, R., Dresser, M., and Villeneuve, A. M. (1998). *Cell* **94,** 387–398.
DiRuggiero, J., Brown, J. R., Bogert, A. P., and Robb, F. T. (1999). *J. Mol. Evol.* **49,** 474–484.

Dixon, D. A., and Kowalczykowski, S. C. (1993). *Cell* **73,** 87–96.
Dolganov, G. M., Maser, R. S., Novikov, A., Tosto, L., Chong, S., Bressan, D. A., and Petrini, J. H. (1996). *Mol. Cell. Biol.* **16,** 4832–4841.
Dosanjh, M. K., Collins, D. W., Fan, W., Lennon, G. G., Albala, J. S., Shen, Z., and Schild, D. (1998). *Nucleic Acids Res.* **26,** 1179–1184.
Duno, M., Thomsen, B., Westergaard, O., Krejci, L., and Bendixen, C. (2000). *Mol. Gen. Genet.* **264,** 89–97.
Edenberg, H. J. (1976). *Biophys. J.* **16,** 849–860.
Egelman, E. H., and Stasiak, A. (1986). *J. Mol. Biol.* **191,** 677–697.
Egelman, E. H., and Stasiak, A. (1988). *J. Mol. Biol.* **200,** 329–349.
Egelman, E. H., and Yu, X. (1989). *Science* **245,** 404–407.
Eggleston, A. K., and West, S. C. (2000). *J. Biol. Chem.* **275,** 26467–26476.
Eisen, J. A. (1998). *Nucleic Acids Res.* **26,** 4291–4300.
Eisen, J. A., Sweder, K. S., and Hanawalt, P. C. (1995). *Nucleic Acids Res.* **23,** 2715–2723.
Eker, A. P., Dekker, R. H., and Berends, W. (1981). *Photochem. Photobiol.* **33,** 65–72.
Eker, A. P., Kooiman, P., Hessels, J. K., and Yasui, A. (1990). *J. Biol. Chem.* **265,** 8009–8015.
Ellis, N. A., Groden, J., Ye, T. Z., Straughen, J., Lennon, D. J., Ciocci, S., Proytcheva, M., and German, J. (1995). *Cell* **83,** 655–666.
Epstein, C. J., and Motulsky, A. G. (1996). *Bioessays* **18,** 1025–1027.
Fishel, R., and Wilson, T. (1997). *Curr. Opin. Genet. Dev.* **7,** 105–113.
Flores, M. J., Bierne, H., Ehrlich, S. D., and Michel, B. (2001). *EMBO J.* **20,** 619–629.
Fogg, J. M., Schofield, M. J., White, M. F., and Lilley, D. M. (1999). *Biochemistry* **38,** 11349–11358.
Foote, R. S., Mitra, S., and Pal, B. C. (1980). *Biochem. Biophys. Res. Commun.* **97,** 654–659.
Fricke, W. M., Kaliraman, V., and Brill, S. J. (2000). *J. Biol. Chem.* (in press).
Friedberg, E. C., Walker, G. C., and Siede, W. (1995). "DNA Repair and Mutagenesis." ASM Press, Washington DC.
Fuller, L. F., and Painter, R. B. (1988). *Mutat. Res.* **193,** 109–121.
Furuse, M., Nagase, Y., Tsubouchi, H., Murakami-Murofushi, K., Shibata, T., and Ohta, K. (1998). *EMBO J.* **17,** 6412–6425.
Game, J. C. (1993). *Semin. Cancer Biol.* **4,** 73–83.
Game, J. C. (2000). *Mutat. Res.* **451,** 277–293.
Gangloff, S., McDonald, J. P., Bendixen, C., Arthur, L., and Rothstein, R. (1994). *Mol. Cell. Biol.* **14,** 8391–8398.
Gerlach, V. L., Aravind, L., Gotway, G., Schultz, R. A., Koonin, E. V., and Friedberg, E. C. (1999). *Proc. Natl. Acad. Sci. USA* **96,** 11922–11927.
Girard, P. M., Guibourt, N., and Boiteux, S. (1997). *Nucleic Acids Res.* **25,** 3204–3211.
Gogos, A., and Clarke, N. D. (1999). *J. Biol. Chem.* **274,** 30447–30450.
Gomes, X. V., and Wold, M. S. (1995). *J. Biol. Chem.* **270,** 4534–4543.
Gomes, X. V., and Wold, M. S. (1996). *Biochemistry* **35,** 10558–10568.
Goodman, M. F. (2000). *Trends Biochem. Sci.* **25,** 189–195.
Goodtzova, K., Kanugula, S., Edara, S., Pauly, G. T., Moschel, R. C., and Pegg, A. E. (1997). *J. Biol. Chem.* **272,** 8332–8339.
Grogan, D. W. (2000). *Trends Microbiol.* **8,** 180–185.
Haber, J. E. (2000a). *Mutat. Res.* **451,** 53–69.
Haber, J. E. (2000b). *Curr. Opin. Cell. Biol.* **12,** 286–292.
Harmon, F. G., and Kowalczykowski, S. C. (1998). *Genes Dev.* **12,** 1134–1144.
Harmon, F. G., DiGate, R. J., and Kowalczykowski, S. C. (1999). *Mol. Cell* **3,** 611–620.
Haseltine, C. A., and Kowalczykowski, S. C. (2001). Submitted for publication.
Hirano, T. (1999). *Genes Dev.* **13,** 11–19.

Holliday, R. (1964). *Genet. Res.* **5,** 282–304.
Hopfner, K. P., Karcher, A., Shin, D., Fairley, C., Tainer, J. A., and Carney, J. P. (2000a). *J. Bacteriol.* **182,** 6036–6041.
Hopfner, K. P., Karcher, A., Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P., and Tainer, J. A. (2000b). *Cell* **101,** 789–800.
Horii, Z.-I., and Clark, A. J. (1973). *J. Mol. Biol.* **80,** 327–344.
Horst, J. P., and Fritz, H. J. (1996). *EMBO J.* **15,** 5459–5469.
Hosfield, D. J., Frank, G., Weng, Y., Tainer, J. A., and Shen, B. (1998a). *J. Biol. Chem.* **273,** 27154–27161.
Hosfield, D. J., Mol, C. D., Shen, B., and Tainer, J. A. (1998b). *Cell* **95,** 135–146.
Iwasaki, H., Takahagi, M., Nakata, A., and Shinagawa, H. (1992). *Genes Dev.* **6,** 2214–2220.
Johnson, J. L., Hamm-Alvarez, S., Payne, G., Sancar, G. B., Rajagopalan, K. V., and Sancar, A. (1988). *Proc. Natl. Acad. Sci. USA* **85,** 2046–2050.
Johnson, R. D., and Symington, L. S. (1995). *Mol. Cell. Biol.* **15,** 4843–4850.
Johnson, R. E., Prakash, S., and Prakash, L. (1999). *Science* **283,** 1001–1004.
Johzuka, K., and Ogawa, H. (1995). *Genetics* **139,** 1521–1532.
Kanaar, R., Troelstra, C., Swagemakers, S. M., Essers, J., Smit, B., Franssen, J. H., Pastink, A., Bezzubova, O. Y., Buerstedde, J. M., Clever, B., Heyer, W. D., and Hoeijmakers, J. H. (1996). *Curr. Biol.* **6,** 828–838.
Kawarabayasi, Y., Sawada, M., and Horikawa, H. *et al.* (1998). *DNA Res.* **5,** 55–76.
Kawarabayasi, Y., Hino, Y., and Horikawa, H., *et al.* (1999). *DNA Res.* **6,** 83–101, 145–152.
Keeney, S., Giroux, C. N., and Kleckner, N. (1997). *Cell* **88,** 375–384.
Kelly, T. J., Simancek, P., and Brush, G. S. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 14634–14639.
Kiener, A., Husain, I., Sancar, A., and Walsh, C. (1989). *J. Biol. Chem.* **264,** 13880–13887.
Kil, Y. V., Baitin, D. M., Masui, R., Bonch-Osmolovskaya, E. A., Kuramitsu, S., and Lanzov, V. A. (2000). *J. Bacteriol.* **182,** 130–134.
Kitao, S., Ohsugi, I., Ichikawa, K., Goto, M., Furuichi, Y., and Shimamoto, A. (1998). *Genomics* **54,** 443–452.
Kitao, S., Lindor, N. M., Shiratori, M., Furuichi, Y., and Shimamoto, A. (1999a). *Genomics* **61,** 268–276.
Kitao, S., Shimamoto, A., Goto, M., Miller, R. W., Smithson, W. A., Lindor, N. M., and Furuichi, Y. (1999b). *Nature Genet.* **22,** 82–84.
Klenk, H. P., Clayton, R. A., and Tomb, J. F., *et al.* (1997). *Nature* **390,** 364–370.
Kogoma, T. (1996). *Cell* **85,** 625–627.
Kogoma, T. (1997). *Microbiol. Mol. Biol. Rev.* **61,** 212–238.
Kolodner, R. D., and Marsischky, G. T. (1999). *Curr. Opin. Genet. Dev.* **9,** 89–96.
Kolodner, R., Fishel, R. A., and Howard, M. (1985). *J. Bacteriol.* **163,** 1060–1066.
Komori, K., Sakae, S., Shinagawa, H., Morikawa, K., and Ishino, Y. (1999). *Proc. Natl. Acad. Sci. USA* **96,** 8873–8878.
Komori, K., Miyata, T., Daiyasu, H., Toh, H., Shinagawa, H., and Ishino, Y. (2000a). *J. Biol. Chem.* **275,** 33791–33797.
Komori, K., Miyata, T., DiRuggiero, J., Holley-Shanks, R., Hayashi, I., Cann, I. K., Mayanagi, K., Shinagawa, H., and Ishino, Y. (2000b). *J. Biol. Chem.* **275,** 33782–33790.
Koulis, A., Cowan, D. A., Pearl, L. H., and Savva, R. (1996). *FEMS Microbiol. Lett.* **143,** 267–271.
Kowalczykowski, S. C. (1991). *Annu. Rev. Biophys. Biophys. Chem.* **20,** 539–575.
Kowalczykowski, S. C. (2000). *Trends Biochem. Sci.* **25,** 156–165.
Kowalczykowski, S. C., and Eggleston, A. K. (1994). *Annu. Rev. Biochem.* **63,** 991–1043.
Kowalczykowski, S. C., and Krupp, R. A. (1987). *J. Mol. Biol.* **193,** 97–113.

Kowalczykowski, S. C., Dixon, D. A., Eggleston, A.K., Lauder, S.D., and Rehrauer, W. M. (1994). *Microbiol. Rev.* **58,** 401–465.
Krokan, H. E., Standal, R., and Slupphaug, G. (1997). *Biochem. J.* **325,** 1–16.
Kulaeva, O. I., Koonin, E. V., McDonald, J. P., Randall, S. K., Rabinovich, N., Connaughton, J. F., Levine, A. S., and Woodgate, R. (1996). *Mutat. Res.* **357,** 245–253.
Kuzminov, A. (1999). *Microbiol. Mol. Biol. Rev.* **63,** 751–813.
Kvaratskhelia, M., and White, M. F. (2000a). *J. Mol. Biol.* **297,** 923–932.
Kvaratskhelia, M., and White, M. F. (2000b). *J. Mol. Biol.* **295,** 193–202.
Kvaratskhelia, M., Wardleworth, B. N., Norman, D. G., and White, M. F. (2000). *J. Biol. Chem.* **275,** 25540–25546.
Lam, S. T., Stahl, M. M., McMilin, K. D., and Stahl, F. W. (1974). *Genetics* **77,** 425–433.
Lanzov, V., Stepanova, I., and Vinogradskaja, G. (1991). *Biochimie* **73,** 305–312.
Leclere, M. M., Nishioka, M., Yuasa, T., Fujiwara, S., Takagi, M., and Imanaka, T. (1998). *Mol. Gen. Genet.* **258,** 69–77.
Lilley, D. M., and White, M. F. (2000). *Proc. Natl. Acad. Sci. USA* **97,** 9351–9353.
Lindahl, T., and Wood, R. D. (1999). *Science* **286,** 1897–1905.
Lindor, N. M., Furuichi, Y., Kitao, S., Shimamoto, A., Arndt, C., and Jalal, S. (2000). *Am. J. Med. Genet.* **90,** 223–228.
Liu, J., Wu, T. C., and Lichten, M. (1995). *EMBO J.* **14,** 4599–4608.
Liu, N., Lamerdin, J. E., Tebbs, R. S., Schild, D., Tucker, J. D., Shen, M. R., Brookman, K. W., Siciliano, M. J., Walter, C. A., Fan, W., Narayana, L. S., Zhou, Z. Q., Adamson, A. W., Sorensen, K. J., Chen, D. J., Jones, N. J., and Thompson, L. H. (1998). *Mol. Cell* **1,** 783–793.
Lloyd, R. G., and Buckman, C. (1995). *Genetics* **139,** 1123–1148.
Lloyd, R. G., and Sharples, G. J. (1993). *EMBO J.* **12,** 17–22.
Lohman, T. M., and Ferrari, M. E. (1994). *Annu. Rev. Biochem.* **63,** 527–570.
Lovett, S. T., and Mortimer, R. K. (1987). *Genetics* **116,** 547–553.
Lu, A. L., and Fawcett, W. P. (1998). *J. Biol. Chem.* **273,** 25098–25105.
Luisi-DeLuca, C., and Kolodner, R. D. (1994). *J. Mol. Biol.* **236,** 124–138.
Marians, K. J. (2000). *Trends Biochem. Sci.* **25,** 185–189.
Masutani, C., Araki, M., Yamada, A., Kusumoto, R., Nogimori, T., Maekawa, T., Iwai, S., and Hanaoka, F. (1999a). *EMBO J.* **18,** 3491–3501.
Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999b). *Nature* **399,** 700–704.
Matsui, E., Kawasaki, S., Ishida, H., Ishikawa, K., Kosugi, Y., Kikuchi, H., Kawarabayashi, Y., and Matsui, I. (1999). *J. Biol. Chem.* **274,** 18297–18309.
Mazin, A. V., Bornarth, C. J., Solinger, J. A., Heyer, W. D., and Kowalczykowski, S. C. (2000a). *Mol. Cell* **6,** 583–592.
Mazin, A. V., Zaitseva, E., Sung, P., and Kowalczykowski, S. C. (2000b). *EMBO J.* **19,** 1148–1156.
McCullough, A. K., Dodson, M. L., and Lloyd, R. S. (1999). *Annu. Rev. Biochem.* **68,** 255–285.
McGlynn, P., and Lloyd, R. G. (2000). *Cell* **101,** 35–45.
McKim, K. S., and Hayashi-Hagihara, A. (1998). *Genes Dev.* **12,** 2932–2942.
Mendonca, V. M., Klepin, H. D., and Matson, S. W. (1995). *J. Bacteriol.* **177,** 1326–1335.
Meyer, R. R., and Laine, P. S. (1990). *Microbiol. Rev.* **54,** 342–380.
Michel, B. (2000). *Trends Biochem. Sci.* **25,** 173–178.
Michel, B., Ehrlich, S. D., and Uzest, M. (1997). *EMBO J.* **16,** 430–438.
Modrich, P. (1991). *Annu. Rev. Genet.* **25,** 229–253.
Mortensen, U. H., Bendixen, C., Sunjevaric, I., and Rothstein, R. (1996). *Proc. Natl. Acad. Sci. USA* **93,** 10729–10734.

Nakayama, H., Nakayama, K., Nakayama, R., Irino, N., Nakayama, Y., and Hanawalt, P. C. (1984). *Mol. Gen. Genet.* **195,** 474–480.
Nakayama, K., Irino, N., and Nakayama, H. (1985). *Mol. Gen. Genet.* **200,** 266–271.
Neddermann, P., Gallinari, P., Lettieri, T., Schmid, D., Truong, O., Hsuan, J. J., Wiebauer, K., and Jiricny, J. (1996). *J. Biol. Chem.* **271,** 12767–12774.
New, J. H., Sugiyama, T., Zaitseva, E., and Kowalczykowski, S. C. (1998). *Nature* **391,** 407–410.
Nicolas, A., Treco, D., Schultes, N. P., and Szostak, J. W. (1989). *Nature* **338,** 35–39.
Ogawa, H., Johzuka, K., Nakagawa, T., Leem, S. H., and Hagihara, A. H. (1995). *Adv. Biophys.* **31,** 67–76.
Ogawa, T., Yu, X., Shinohara, A., and Egelman, E. H. (1993). *Science* **259,** 1896–1899.
Ogrunc, M., Becker, D. F., Ragsdale, S. W., and Sancar, A. (1998). *J. Bacteriol.* **180,** 5796–5798.
Olsen, L. C., Aasland, R., Wittwer, C. U., Krokan, H. E., and Helland, D. E. (1989). *EMBO J.* **8,** 3121–3125.
Olsson, M., and Lindahl, T. (1980). *J. Biol. Chem.* **255,** 10569–10571.
Pâques, F., and Haber, J. E. (1999). *Microbiol. Mol. Biol. Rev.* **63,** 349–404.
Paull, T. T., and Gellert, M. (1998). *Mol. Cell* **1,** 969–979.
Paull, T. T., and Gellert, M. (1999). *Genes Dev.* **13,** 1276–1288.
Petrini, J. H. (1999). *Am. J. Hum. Genet.* **64,** 1264–1269.
Petronzelli, F., Riccio, A., Markham, G. D., Seeholzer, S. H., Stoerker, J., Genuardi, M., Yeung, A. T., Matsumoto, Y., and Bellacosa, A. (2000). *J. Biol. Chem.* **275,** 32422–32429.
Petukhova, G., Stratton, S., and Sung, P. (1998). *Nature* **393,** 91–94.
Petukhova, G., Van Komen, S., Vergano, S., Klein, H., and Sung, P. (1999). *J. Biol. Chem.* **274,** 29453–29462.
Philipova, D., Mullen, J. R., Maniar, H. S., Lu, J., Gu, C., and Brill, S. J. (1996). *Genes Dev.* **10,** 2222–2233.
Pierce, A. J., Johnson, R. D., Thompson, L. H., and Jasin, M. (1999). *Genes Dev.* **13,** 2633–2638.
Pochart, P., Woltering, D., and Hollingsworth, N. M. (1997). *J. Biol. Chem.* **272,** 30345–30349.
Postow, L., Ullsperger, C., Keller, R. W., Bustamante, C., Vologodskii, A. V., and Cozzarelli, N. R. (2001). *J. Biol. Chem.* **276,** 2790–2796.
Prakash, S., and Prakash, L. (2000). *Mutat. Res.* **451,** 13–24.
Puranam, K. L., and Blackshear, P. J. (1994). *J. Biol. Chem.* **269,** 29838–29845.
Radding, C. M. (1989). *Biochim. Biophys. Acta* **1008,** 131–145.
Rao, H. G., Rosenfeld, A., and Wetmur, J. G. (1998). *J. Bacteriol.* **180,** 5406–5412.
Rattray, A. J., and Symington, L. S. (1994). *Genetics* **138,** 587–595.
Raymond, W. E., and Kleckner, N. (1993). *Nucleic Acids Res.* **21,** 3851–3856.
Resnick, M. A. (1976). *J. Theor. Biol.* **59,** 97–106.
Rice, M. C., Smith, S. T., Bullrich, F., Havre, P., and Kmiec, E. B. (1997). *Proc. Natl. Acad. Sci. USA* **94,** 7417–7422.
Romanienko, P. J., and Camerini-Otero, R. D. (2000). *Mol. Cell* **6,** 975–987.
Rothstein, R., Michel, B., and Gangloff, S. (2000). *Genes Dev.* **14,** 1–10.
Sack, S. Z., Liu, Y., German, J., and Green, N. S. (1998). *Clin. Exp. Immunol.* **112,** 248–254.
Saffi, J., Pereira, V. R., and Henriques, J. A. (2000). *Curr. Genet.* **37,** 75–78.
Sancar, A. (1996). *Annu. Rev. Biochem.* **65,** 43–81.
Sancar, G. B., Smith, F. W., and Heelis, P. F. (1987). *J. Biol. Chem.* **262,** 15457–15465.
Sandigursky, M., and Franklin, W. A. (1999). *Curr. Biol.* **9,** 531–534.
Sandigursky, M., and Franklin, W. A. (2000). *J. Biol. Chem.* **275,** 19146–19149.

Sandler, S. J., and Marians, K. J. (2000). *J. Bacteriol.* **182,** 9–13.
Sandler, S. J., Satin, L. H., Samra, H. S., and Clark, A. J. (1996). *Nucleic Acids Res.* **24,** 2125–2132.
Schild, D., Lio, Y., Collins, D. W., Tsomondo, T., and Chen, D. J. (2000). *J. Biol. Chem.* **275,** 16443–16449.
Schmidt, K. J., Beck, K. E., and Grogan, D. W. (1999). *Genetics* **152,** 1407–1415.
Schwacha, A., and Kleckner, N. (1995). *Cell* **83,** 783–791.
Seigneur, M., Bidnenko, V., Ehrlich, S. D., and Michel, B. (1998). *Cell* **95,** 419–430.
Seitz, E. M., and Kowalczykowski, S. C. (2000). *Mol. Microbiol.* **37,** 555–560.
Seitz, E. M., Brockman, J. P., Sandler, S. J., Clark, A. J., and Kowalczykowski, S. C. (1998). *Genes Dev.* **12,** 1248–1253.
Seki, M., Miyazawa, H., Tada, S., *et al.* (1994). *Nucleic Acids Res.* **22,** 4566–4573.
Shah, R., Cosstick, R., and West, S. C. (1997). *EMBO J.* **16,** 1464–1472.
Sharples, G. J., and Leach, D. R. (1995). *Mol. Microbiol.* **17,** 1215–1217.
Shen, J. C., and Loeb, L. A. (2000). *Trends Genet.* **16,** 213–220.
Shinohara, A., and Ogawa, T. (1998). *Nature* **391,** 404–407.
Shinohara, A., Ogawa, H., and Ogawa, T. (1992). *Cell* **69,** 457–470.
Shinohara, A., Shinohara, M., Ohta, T., Matsuda, S., and Ogawa, T. (1998). *Genes Cells* **3,** 145–156.
Sinclair, D. A., and Guarente, L. (1997). *Cell* **91,** 1033–1042.
Skorvaga, M., Raven, N. D., and Margison, G. P. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 6711–6715.
Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., *et al.* (1997). *J. Bacteriol.* **179,** 7135–7155.
Smith, G. R., Kunes, S. M., Schultz, D. W., Taylor, A., and Triman, K. L. (1980). *In* "Mechanistic Studies of DNA Replication and Genetic Recombination" B. Alberts, ed. pp. 927–931. Academic Press, New York.
Spies, M., Kil, Y., Masui, R., Kato, R., Kujo, C., Ohshima, T., Kuramitsu, S., and Lanzov, V. (2000). *Eur. J. Biochem.* **267,** 1125–1137.
Stasiak, A., and Egelman, E. H. (1986). *Biochem. Soc. Trans.* **14,** 218–220.
Stasiak, A., and Egelman, E. H. (1994). *Experientia* **50,** 192–203.
Stasiak, A., Di Capua, E., and Koller, T. (1981). *J. Mol. Biol.* **151,** 557–564.
Stasiak, A., Stasiak, A. Z., and Koller, T. (1984). *Cold Spring Harbor Symp. Quant. Biol.* **49,** 561–570.
Sugiyama, T., Zaitseva, E. M., and Kowalczykowski, S. C. (1997). *J. Biol. Chem.* **272,** 7940–7945.
Sugiyama, T., New, J. H., and Kowalczykowski, S. C. (1998). *Proc. Natl. Acad. Sci. USA* **95,** 6049–6054.
Sun, H., Treco, D., Schultes, N. P., and Szostak, J. W. (1989). *Nature* **338,** 87–90.
Sung, P. (1994). *Science* **265,** 1241–1243.
Sung, P. (1997a). *Genes Dev.* **11,** 1111–1121.
Sung, P. (1997b). *J. Biol. Chem.* **272,** 28194–28197.
Sung, P., and Robberson, D. L. (1995). *Cell* **82,** 453–461.
Sung, P., Trujillo, K. M., and Van Komen, S. (2000). *Mutat. Res.* **451,** 257–275.
Szostak, J. W., Orr-Weaver, T. L., Rothstein, R. J., and Stahl, F. W. (1983). *Cell* **33,** 25–35.
Tan, T. L., Essers, J., Citterio, E., Swagemakers, S. M., de Wit, J., Benson, F. E., Hoeijmakers, J. H., and Kanaar, R. (1999). *Curr. Biol.* **9,** 325–328.
Tang, M., Shen, X., Frank, E. G., O'Donnell, M., Woodgate, R., and Goodman, M. F. (1999). *Proc. Natl. Acad. Sci. USA* **96,** 8919–8924.
Tracy, R. B., and Kowalczykowski, S. C. (1996). *Genes Dev.* **10,** 1890–1903.

Tucker, J. D., Jones, N. J., Allen, N. A., Minkler, J. L., Thompson, L. H., and Carrano, A. V. (1991). *Mutat. Res.* **254,** 143–152.
Umezu, K., Chi, N. W., and Kolodner, R. D. (1993). *Proc. Natl. Acad. Sci. USA* **90,** 3875–3879.
Usui, T., Ohta, T., Oshiumi, H., Tomizawa, J., Ogawa, H., and Ogawa, T. (1998). *Cell* **95,** 705–716.
Van Dyck, E., Stasiak, A. Z., Stasiak, A., and West, S. C. (1999). *Nature* **398,** 728–731.
Van Komen, S., Petukhova, G., Sigurdsson, S., Stratton, S., and Sung, P. (2000). *Mol. Cell* **6,** 563–572.
Wadsworth, R. I., and White, M. F. (2001). *Nucleic Acids Res.* **29,** 914–920.
Wagner, J., Gruz, P., Kim, S. R., Yamada, M., Matsui, K., Fuchs, R. P., and Nohmi, T. (1999). *Mol. Cell* **4,** 281–286.
Watrin, L., and Prieur, D. (1996). *Curr. Microbiol.* **33,** 377–382.
Watt, P. M., Louis, E. J., Borts, R. H., and Hickson, I. D. (1995). *Cell* **81,** 253–260.
Watt, P. M., Hickson, I. D., Borts, R. H., and Louis, E. J. (1996). *Genetics* **144,** 935–945.
Webb, B. L., Cox, M. M., and Inman, R. B. (1997). *Cell* **91,** 347–356.
West, S. C. (1992). *Annu. Rev. Biochem.* **61,** 603–640.
West, S. C. (1994a). *Cell* **76,** 9–15.
West, S. C. (1994b). *Ann. N.Y. Acad. Sci.* **726,** 156–163; discussion, 163–154.
West, S. C. (1997). *Annu. Rev. Genet.* **31,** 213–244.
Whitby, M. C., and Lloyd, R. G. (1998). *J. Biol. Chem.* **273,** 19729–19739.
White, M. F., Giraud-Panis, M. J., Pohler, J. R., and Lilley, D. M. (1997). *J. Mol. Biol.* **269,** 647–664.
Wold, M. S. (1997). *Annu. Rev. Biochem.* **66,** 61–92.
Wood, R. D. (1996). *Annu. Rev. Biochem.* **65,** 135–167.
Woodgate, R. (1999). *Genes Dev.* **13,** 2191–2195.
Woods, W. G., and Dyall-Smith, M. L. (1997). *Mol. Microbiol.* **23,** 791–797.
Xiao, W., and Samson, L. (1992). *Nucleic Acids Res.* **20,** 3599–3606.
Yang, H., Fitz-Gibbon, S., Marcotte, E. M., Tai, J. H., Hyman, E. C., and Miller, J. H. (2000). *J. Bacteriol.* **182,** 1272–1279.
Yang, W. (2000). *Mutat. Res.* **460,** 245–256.
Yasui, A., Takao, M., Oikawa, A., Kiener, A., Walsh, C. T., and Eker, A. P. (1988). *Nucleic Acids Res.* **16,** 4447–4463.
Yu, C. E., Oshima, J., Fu, Y. H., Wijsman, E. M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., Martin, G. M., Mulligan, J., and Schellenberg, G. D. (1996). *Science* **272,** 258–262.