



The Bacterial Replicative Helicase DnaB Evolved from a RecA Duplication

Detlef D. Leipe, L. Aravind, Nick V. Grishin, et al.

Genome Res. 2000 10: 5-16

Access the most recent version at doi:[10.1101/gr.10.1.5](https://doi.org/10.1101/gr.10.1.5)

References This article cites 70 articles, 22 of which can be accessed free at:
<http://genome.cshlp.org/content/10/1/5.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

A horizontal banner advertisement with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." in white. In the center, there is a white rectangular button with the text "LEARN MORE". On the right, there is a photograph of a woman wearing a red and white superhero cape and mask, with a green molecular structure logo above the word "CELLECTA" in white.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

The Bacterial Replicative Helicase DnaB Evolved from a RecA Duplication

Detlef D. Leipe,¹ L. Aravind,^{2,3} Nick V. Grishin,^{1,4} and Eugene V. Koonin^{1,5}

¹National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health, Bethesda Maryland 20894 USA; ²Department of Biology, Texas A&M University, College Station, Texas 70843 USA

The RecA/Rad51/DCMI family of ATP-dependent recombinases plays a crucial role in genetic recombination and double-stranded DNA break repair in Archaea, Bacteria, and Eukaryota. DnaB is the replication fork helicase in all Bacteria. We show here that DnaB shares significant sequence similarity with RecA and Rad51/DMC1 and two other related families of ATPases, Sms and KaiC. The conserved region spans the entire ATP- and DNA-binding domain that consists of about 250 amino acid residues and includes 7 distinct motifs. Comparison with the three-dimensional structure of *Escherichia coli* RecA and phage T7 DnaB (gp4) reveals that the area of sequence conservation includes the central parallel β -sheet and most of the connecting helices and loops as well as a smaller domain that consists of an amino-terminal helix and a carboxy-terminal β -meander. Additionally, we show that animals, plants, and the malarial *Plasmodium* but not *Saccharomyces cerevisiae* encode a previously undetected DnaB homolog that might function in the mitochondria. The DnaB homolog from *Arabidopsis* also contains a DnaG-primase domain and the DnaB homolog from the nematode seems to contain an inactivated version of the primase. This domain organization is reminiscent of bacteriophage primases-helicases and suggests that DnaB might have been horizontally introduced into the nuclear eukaryotic genome via a phage vector. We hypothesize that DnaB originated from a duplication of a RecA-like ancestor after the divergence of the bacteria from Archaea and eukaryotes, which indicates that the replication fork helicases in Bacteria and Archaea/Eukaryota have evolved independently.

Genetic recombination is an essential process for both recombinational repair and sexual reproduction. In Bacteria, the central role in recombination is played by the RecA recombinase enzyme (Radding 1989; Kowalczykowski and Eggleston 1994; Seitz et al. 1998). RecA is a DNA-dependent ATPase that promotes homologous pairing and strand exchange between different double-stranded (ds) DNA molecules and is therefore necessary for homologous recombination and DNA repair (Kowalczykowski et al. 1994). The biochemical activities of RecA include the ability to form regular helical filaments, bind single-stranded (ss) and dsDNA, and bind and hydrolyze nucleoside triphosphates (Kowalczykowski et al. 1994). In addition to its direct role in recombination, RecA functions as a cofactor in the cleavage reaction for LexA, the repressor of the SOS regulon (Little and Mount 1982; Witkin 1991). There are two types of RecA-like proteins in many eukaryotes, namely Rad51 and DMC1/Lim15. Rad51 is expressed in both meiotic and mitotic cells and mainly participates in recombinational repair of double-strand breaks (Shinohara et al. 1992; Doutriaux et al. 1998). DMC1 is expressed in meiotic cells, its null mutants

show a meiotic arrest phenotype, and it probably functions in the formation of synaptonemal complexes and also in double-strand break repair (Bishop et al. 1992; Dresser et al. 1997; Yoshida et al. 1998). Thus, there is functional overlap between Rad51 and DMC1 (Shinohara et al. 1997) and *Caenorhabditis elegans* seems to have only a single Rad51/DMC1 homolog (Takanami et al. 1998). A Rad51/DMC1 homolog (termed RadA) that catalyzes DNA pairing and strand exchange (Seitz et al. 1998) is also found in the Archaea (Sandler et al. 1996).

The RecA/RadA/DMC1 recombinases are closely related to three other groups of ATPases, namely bacterial Sms (also called RadA), bacterial DnaB, and archaeal and bacterial KaiC. The Sms protein is a poorly characterized bacterial homolog of RecA in which the RecA ATPase domain is fused to a Zn ribbon and a predicted serine protease domain (Koonin et al. 1996; Aravind et al. 1999) (hereafter we use the designation Sms to avoid confusion with the archaeal RadA). *Escherichia coli* sms mutants show increased sensitivity to X rays, UV radiation, and methyl methanesulfonate, suggesting a role in repair for the Sms protein (Neuwald et al. 1992; Song and Sargentini 1996).

The cyanobacterial KaiABC gene cluster constitutes the circadian clock in the cyanobacterium *Synechococcus* (Ishiura et al. 1998; Iwasaki et al. 1999). The KaiC protein generates a circadian oscillation by negative feedback control on its own expression (Ishiura et

³Present address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894 USA.

⁴Present address: Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75235 USA.

⁵Corresponding author.

E-MAIL koonin@ncbi.nlm.nih.gov; FAX (301) 435-7794.

al. 1998). The *Synechococcus* KaiC protein is composed of two RecA-like domains joined head to tail. Highly conserved homologs of KaiC are found in the cyanobacterium *Synechocystis*, the bacterium *Thermotoga*, and in all Archaea but absent from other bacteria and eukaryotes (Makarova et al. 1999).

The DnaB helicase is a crucial protein in bacterial DNA replication. It unwinds the DNA duplex ahead of the replication fork and is also responsible for attracting the DnaG primase to the replication fork (Tougu et al. 1994; Lu et al. 1996). The active form of the protein is a hexamer of identical 52.3-kD subunits that can form rings with threefold (C3) and sixfold (C6) symmetry (Yu et al. 1996) and it has been hypothesized that the amino-terminal ATPase domains of two adjacent protomers dimerizes to make the C6–C3 conversion (Fass et al. 1999). The crystal structure of the helicase domain of phage T7 helicase–primase (gp4) has recently been solved (Sawaya et al. 1999) and it has been found that the structure of the T7 helicase domain and its interactions with neighboring subunits in the crystal resemble those of the RecA and F₁ ATPase (Sawaya et al. 1999). In addition to the ATPase domain, *E. coli* DnaB comprises a globular amino-terminal domain (proteolytic fragment III) that is essential for interaction with other proteins involved in DNA replication like DnaA, DnaC, and the DnaG primase (Nakayama et al. 1984; Biswas et al. 1994; Sutton et al. 1998). The domain consists of six α helices (Weigelt et al. 1998; Fass et al. 1999; Weigelt et al. 1999) that are attached to the carboxy-terminal ATPase domain by a flexible hinge (Miles et al. 1997).

In addition to RecA, DMC1/Rad51/RadA, DnaB, Sms, and KaiC, there is a large number of proteins with more limited phylogenetic distribution that contain the core RecA ATPase domain. These include, among others, Rad51-interacting proteins Rad55 and Rad57 in

yeast (Game 1993), XRCC2 (Tambini et al. 1997), R5H2 and R5H3 (Cartwright et al. 1998), and TRAD (Kawabata and Sacki 1998) in mammals, and several other distinct RecA homologs found in Archaea and some bacteria (Aravind et al. 1999). Some of these orphan RecA homologs appear to contain an inactivated ATPase domain (Aravind et al. 1999). Additional domains associated with the RecA core include a modified amino-terminal helix–hairpin–helix (HhH) domain in the archaeoeukaryotic RadA/DMC1, a amino-terminal zinc finger and a carboxy-terminal Lon-type protease domain in Sms, and a GTPase in one of the archaeal RecA homologs (Aravind et al. 1999).

Here, using a combination of sequence database searches, sequence alignments, phylogenetic analysis, and structural comparison, we show that (1) DnaB, RecA, DMC1/RadA, Sms, and KaiC share significant sequence similarity along a region of 250 amino acids that includes both the ATP-binding domain and the DNA-binding site; (2) DnaB likely evolved from RecA by a gene duplication event at the onset of the evolution of the Bacteria; (3) RecA and DnaB are likely to perform their function by a similar mechanism of conformational change; (4) eukaryotes encode diverged homologs of DnaB, some of which also contain a DnaG-type primase domain; these genes might have been introduced into the eukaryotic genome by a horizontal transfer event involving a bacteriophage. We hypothesize that the common ancestor of the RecA/DnaB superfamily functioned as a recombinase in the last common ancestor (LCA) of all extant cells and that a RecA homolog (DnaB) was recruited for the helicase function at the replication fork once DNA replication evolved in bacteria. This interpretation lends further support to the hypothesis that the DNA replication machinery evolved independently in bacteria and archaea/eukaryotes (Leipe et al. 1999).

Figure 1 (See pages 7–9.) Multiple alignment of the core domain of the RecA/DnaB superfamily of ATPases. From top to bottom (separated by horizontal lines) the alignment contains sequences from bacterial and chloroplast DnaB, DnaB proteins and primase–helicase proteins from bacteriophages and eukaryotes, bacterial Sms proteins, KaiC from Archaea and Bacteria, RecA recombinase from Bacteria and phage T4, and RadA and Rad51/DMC1 recombinases from Archaea and Eukaryota. The 80% consensus for these proteins is shown below the aligned sequences. Numbers indicate the distance to the amino-terminal methionine and the carboxyl terminus of each protein and residues omitted within the alignment. (&) The position of inteins that have not been included in the alignment. The secondary structure elements derived from the X-ray structures of phage T7 gp4 and *E. coli* RecA are shown above the respective sequence. Helices are represented as cylinders, strands as arrows, and the unordered or mobile loops 1 and 2 as lines. Key residues that are discussed in the text are marked by arrowheads; the numbers identify the position of the residue in gp4 and RecA according to the original publications (Story et al. 1993; Sawaya et al. 1999). Highly conserved residues are color coded and indicated in the consensus line for the following groups. (Purple) Negatively charged (D,E); (red) positively charged (H,K,R), charged (c = D,E,H,K,R); (green) tiny (u = G,A,S); (yellow) hydrophobic (h = A,C,F,I,L,M,V,W,Y) or aliphatic (l = I,L,V); (pale yellow) alcohol (o = S, T, Y); (light blue) polar (p = D,E,H,K,N,Q,R,S,T), (reddish-brown) small (s = A,C,D,G,N,P,S,T,V); (gray) big (b = not small). Also colored are residues conserved only within the DnaB family. Where applicable, source organisms are identified by four-letter abbreviations. (Aepe) *Aeropyrum pernix*; (Aqae) *A. aeolicus*; (Arfu) *Archaeoglobus fulgidus*; (Arth) *A. thaliana*; (Basu) *Bacillus subtilis*; (Bobu) *Borrelia burgdorferi*; (T7) bacteriophage T7; (T4) bacteriophage T4; (Cael) *C. elegans*; (CDnaB_Odsi) *Odontella sinensis* chloroplast; (CDnaB_Popu) *Porphyra purpurea* chloroplast; (Chtr) *Chlamydia trachomatis*; (Ecol) *E. coli*; (Glma) *Glycine max*; (Hain) *Haemophilus influenzae*; (Hepy) *H. pylori*; (Hosa) *Homo sapiens*; (Lema) *Leishmania major*; (Meja) *Methanococcus jannaschii*; (Meth) *Methanobacterium thermoautotrophicum*; (Mumu) *Mus musculus*; (Myge) *Mycoplasma genitalium*; (Mytu) *Mycobacterium tuberculosis*; (Plch) *P. chabaudi*; (Rhma) *Rhodothermus marinus*; (Sace) *Saccharomyces cerevisiae*; (SPP1) *Bacillus subtilis* bacteriophage SPP1; (Suso) *Sulfolobus solfataricus*; (Sy68) *Synechocystis* PCC6803; (Teth) *Tetrahymena thermophila*; (Thma) *T. maritima*; (Trpa) *Treponema pallidum*.

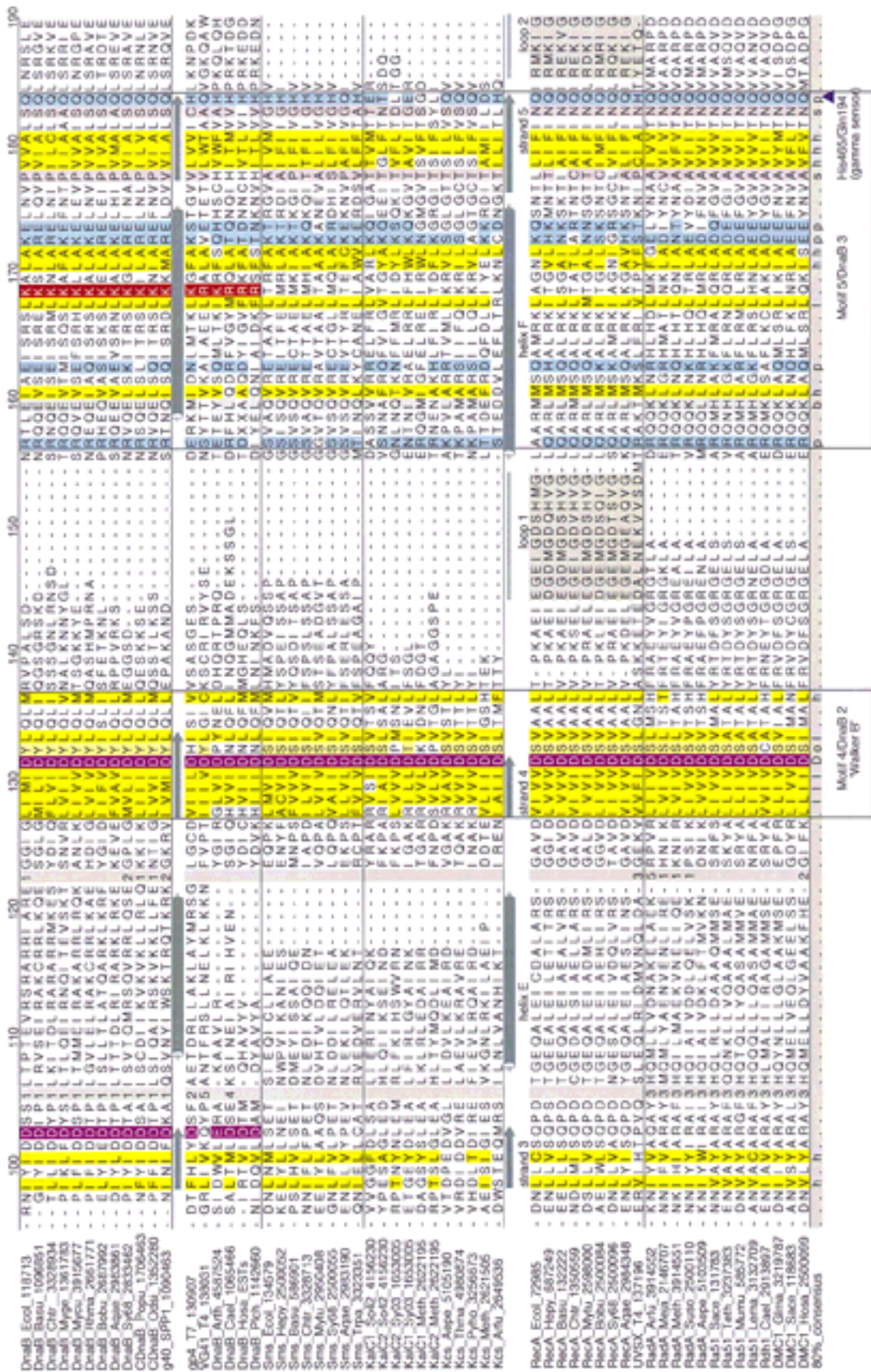


Figure 1 (Continued)

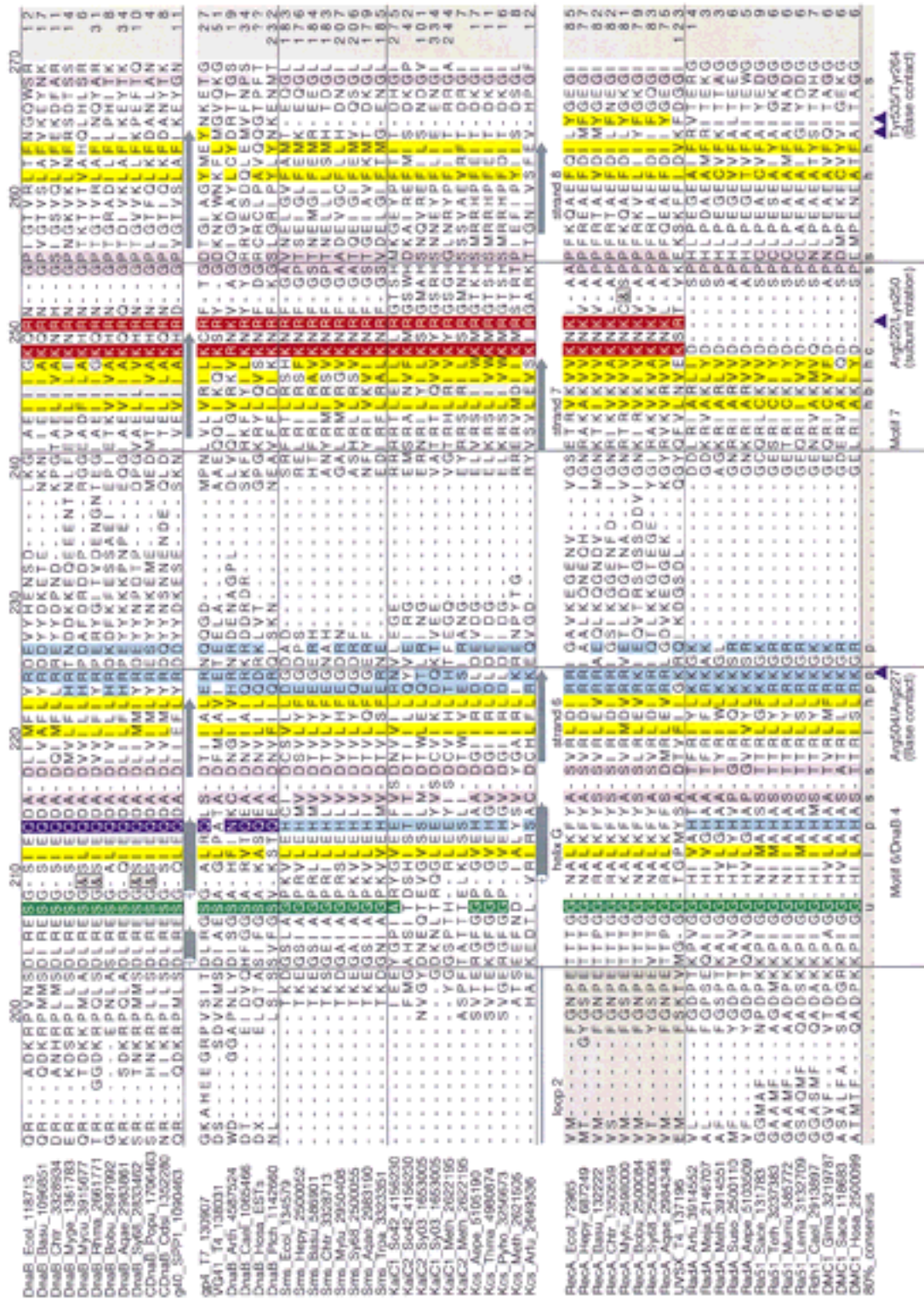


Figure 1 (Continued)

RESULTS AND DISCUSSION

The Core ATPase Domains of RecA and DnaB Are Specifically Related

BLAST searches seeded with the *E. coli* DnaB sequence retrieve the replicative helicases from a wide range of Bacteria and several bacteriophages with highly significant *E* values ($<10^{-40}$) and the helicase–primase proteins from bacteriophages T3/T7 and T4 with less significant *E* values (between 10^{-5} and 10^{-3}). The first iteration of the PSI-BLAST search unexpectedly retrieved, with highly significant *E* values, a number of members of the RecA superfamily, namely bacterial Sms proteins and archaeal and eukaryotic Rada/Rad51 proteins. For example, the sequence of the Sms protein from the bacterium *Aquifex aeolicus* was detected with an *E* value of 10^{-9} and the sequence of the murine Trad protein with an *E* value of 7×10^{-9} . In addition, previously undetected eukaryotic homologs of DnaB from *C. elegans*, *Arabidopsis thaliana*, and *Plasmodium chabaudi* were retrieved with *E* values between 10^{-7} and 10^{-5} ; a human homolog of these proteins was detected among EST products by searching the database of expressed sequence tags (dbEST) database (see discussion below). Subsequent search iterations retrieve the entire RecA family. Conversely, searches seeded with *E. coli* RecA retrieve members of the DnaB family starting with an *E* value of 0.001 for *Helicobacter pylori* DnaB in the first PSI-BLAST iteration, with all the other members of the DnaB family retrieved in subsequent iterations. In all of these searches, RecA family members and DnaB family members, respectively, were consistently retrieved from the database before any other ATPases. This suggests that within the class of P-loop ATPases, there is a specific structural and, by inference, evolutionary, relationship between the RecA, DMC1/RadA, Sms, KaiC, and DnaB families; hereafter, we refer to them collectively as the RecA/DnaB superfamily.

Sequence and Structure Conservation in the RecA/DnaB Superfamily

The structure of the *E. coli* RecA protein consists of a major central domain flanked by two smaller domains at the amino and carboxy termini (Story et al. 1992; see also Fig. 2, below). The central domain can be subdivided into a large subdomain encompassing strands 1–5 and the connecting helices and loops and a small subdomain that represents two noncontiguous regions of the sequence including helix B and strands 6–8 (Figs.

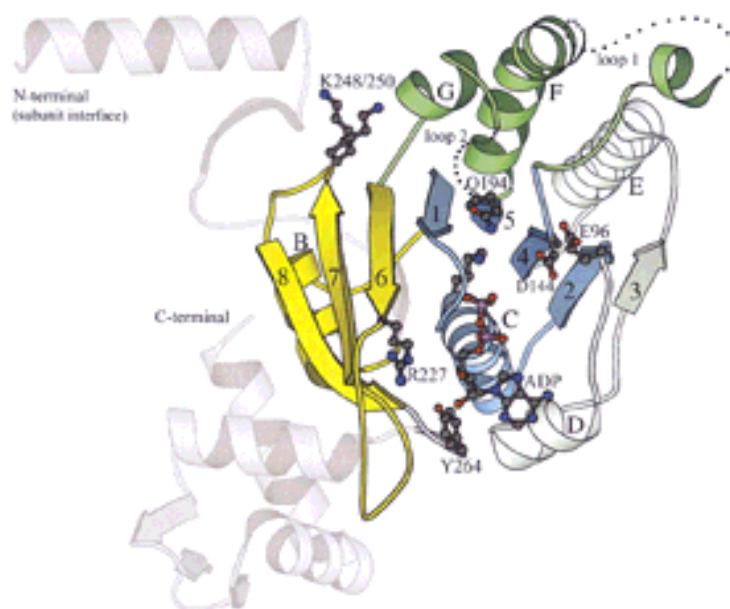


Figure 2 A molscript diagram of *E. coli* RecA structure. Areas with sequence conservation between DnaB and RecA/DMC1/RadA are highlighted, the non-conserved carboxy- and amino-terminal domain are shown in light gray. The central parallel μ -sheet is blue and the elements that are involved in coordinating loop 1 (between strand 4 and helix F) and loop 2 (between strand 5 and helix G) are green. Areas within the core domain that show no obvious sequence similarity between RecA and DnaB (helix D, strand 3 and helix E) are shown in light blue. The subdomain composed of helix B and strands 6–8 is shown in ball-and-stick representation. Conserved amino acid residues that are discussed in the text and indicated in the alignment (Fig. 1) are shown in ball-and-stick representation: Lys-72 (in the P-loop), Glu-96, Asp-144 (Walker B), Gln-194, Arg-227, Lys-248, Lys-250, and Tyr-264 are at or near the carboxyl terminus of strands 2, 4, 5, 6, and 7, respectively. Amino acid coordinates are from PDB file 2REB, location of ADP is from PDB file 1REA. The orientation of the monomer, labels of strands, helices, loops, and residue enumeration are in accordance with the original publications (Story et al. 1992; Story et al. 1993).

1 and 2). As detailed above, we found that the sequence of this 250-amino-acid central domain is specifically conserved between DnaB, RecA, DMC1/RadA, KaiC, and Sms protein families. The importance of this core for RecA function is underscored by the fact that Pk-REC, a truncated, 210-amino-acid DMC1/RadA homolog from *Pyrococcus*, which consists of the core domain alone, can complement UV-sensitive RecA mutants in *E. coli* (Rashid et al. 1996).

A multiple sequence alignment of the RecA and DnaB sequences was constructed on the basis of the PSI-BLAST output and refined manually using structural information on RecA and DnaB (Fig. 1). The region of sequence conservation between RecA, RadA/DMC1, Sms, KaiC, and DnaB extends for ~250 amino acids and includes the P-loop and the Mg^{2+} -binding site (Walker A and B motifs, respectively), which are involved in NTP binding and hydrolysis. Although the Walker A motif shows the typical G . . GKT pattern conserved in a vast variety of ATPase and GTPases (Saraste et al. 1990), it is noteworthy that the second carbox-

ylate typically found in the Walker B motif of several large groups of ATPases, for example, the AAA+ class of chaperone-like ATPases (Neuwald et al. 1999) and superfamily I and II helicases (Gorbalenya and Koonin 1993), is replaced by an alcohol residue in the RecA/DnaB superfamily (Fig. 1).

Motif 3 corresponds to *E. coli* RecA strand 2 and the following loop and is characterized by a completely conserved glutamate (hhh[SD].E) that has earlier been described as a conserved feature of the DnaB family (Ilyina et al. 1992). The conserved glutamate is assumed to activate the nucleophilic water molecule for an in-line attack of the ATP γ -phosphate (Story and Steitz 1992), and a E96D mutation in *E. coli* RecA results in a 100-fold reduction in the ATP hydrolysis rate (Campbell and Davis 1999a,b). The catalytic glutamate is highly conserved not only in the entire RecA/DnaB superfamily, but it is found in the same location (carboxy-terminal of the strand that follows the P-loop) in a large number of Walker-type ATPases, for example, F⁰/F¹ ATPases and Rho helicase (Yoshida and Amano 1995). Interestingly, however, this motif is not detectable in NTPases, for example, the AAA+ class and the superfamily 1 and 2 helicases, where the conserved aspartate in the Walker B motif (motif 4) is followed by another negatively charged residue (so-called DEXX box). As the conserved aspartate in motif 4 is followed by noncharged residue in the RecA/DnaB superfamily, it has been suggested that the second charged residue of the Walker B motif is functionally replaced by the conserved glutamate in motif 3 in the RecA/DnaB superfamily (Sawaya et al. 1999).

In addition to the catalytic glutamate in motif 3 and the Walker A and B motifs (motifs 2 and 4) that are found in a wide variety of ATPases, there are four other motifs (1, 5, 6, and 7 in Fig. 1) that show significant sequence conservation among the members of the RecA/DnaB superfamily and that can be correlated with elements known from the crystal structure of RecA and T7 gp4 (Story and Steitz 1992; Story et al. 1992; Sawaya et al. 1999) (Figs. 1 and 2).

Motif 1 is amino-terminal of the P-loop and corresponds to helix B and a glycine-rich loop containing a conserved negative charge with the consensus pattern h.[ST]G. .h[DE]. .G (where h stands for a hydrophobic residue, residues in square brackets are alternatives, and a dot stands for any residue). In *E. coli* RecA, the tight turn completed by helix B and the neighboring carboxy- and amino-terminal sequences is stabilized by hydrogen bonds between Thr-42 and Asp-48 side chains and Asp-48 and Gly-54 backbone atoms (Story et al. 1993); all four residues involved in these interactions are highly conserved within the entire RecA/DnaB superfamily (Fig. 1). No function has yet been assigned to motif 1, but it has been noted that this regions points towards the outside of the RecA

polymer and is thus distant from the (presumed) ATP and DNA binding sites (Story et al. 1993).

The most conserved RecA residue in motif 5 (Gln-194) is found at the carboxy-terminal end of strand 5. In the structure, this residue is adjacent to the ATP γ -phosphate and it has been proposed to mediate a structural change on binding of ATP that stabilizes a conformation in the following loop 2 and/or helix G with high affinity for DNA (Story and Steitz 1992). Similarly, the corresponding residue of phage T7 gp4 (His-465) is in a position to act as γ -phosphate sensor or conformational switch by forming a hydrogen bond with the ATP γ -phosphate (Sawaya et al. 1999). In addition to the conservation of the putative γ -phosphate sensor itself (glutamine in all bacterial DnaBs and histidine in the eukaryotic DnaB homologs, phage T7 gp4, phage T4 UvsX, and the Sms family), considerable sequence conservation is also found in the preceding helix F and strand 5 in all members of the RecA/DnaB superfamily (Fig. 1). This suggests that the general mode of ATP-binding/hydrolysis-mediated conformational change is conserved at least between RecA, RadA/DMC1, and DnaB. Whether that holds true for the entire superfamily is doubtful because the putative γ sensor (His-465/Gln-194) is not conserved in the double-domain KaiC proteins and because the loop between motifs 5 and 6 (loop 2) seems to be missing in KaiC and Sms (Fig. 1).

In addition to mediating a conformational change within a subunit, binding and hydrolysis of ATP is likely to induce the rotation of subunits within the T7 gp4 hexamer (Sawaya et al. 1999). It has been suggested that T7 gp4 residue Arg-522, which is close to the γ -phosphate of a bound ATP in a neighboring subunit, is responsible for coupling ATP hydrolysis to subunit rotation (Sawaya et al. 1999). The importance of the residue is underscored by the fact that Arg-522 is the third residue of a [KR].[KR] motif located between strands 7 and 8 that is completely conserved in the DnaB, RecA, Sms, and KaiC families (Fig. 1). Surprisingly, the [KR].KR motif appears to be missing in the archaeoeukaryotic RadA/DMC1 family (Fig. 1) although RadA/DMC1 shares the strand exchange function with RecA and shares the highest overall sequence similarity with RecA within the RecA/DnaB superfamily. There is a conserved positively charged residue nearby in the predicted strand 7 of the RadA/DMC1 family proteins (Fig. 1), but whether or not this residue is functionally equivalent to Arg-522 will have to await the first structure of a member of this family.

In T7 gp4, the base of the bound nucleotide is sandwiched between Arg-504 and Tyr-535 (Sawaya et al. 1999). Arg-504, at the carboxy-terminal end of strand 6 in motif 6 (Fig. 1), is conserved as either Arg or Lys in DnaB and RecA but not in most KaiC and Sms proteins. T7 gp4 Tyr-535, at the carboxy-terminal end

of strand 8 in motif 7, seems conserved as an aromatic residue (Phe, Tyr, His) within the DnaB family although exact superposition would require a gap in the bacterial DnaB sequences (Fig. 1). In *E. coli* RecA, the base of the bound ADP stacks on Tyr-103 (Story and Steitz 1992), which is a residue carboxyl terminus of motif 3 that seems conserved only in RecA but not in any of the other member of the RecA/DnaB superfamily (Fig. 1). The other residues that are close to the adenine base in the *E. coli* RecA structure are Asp-100, Tyr-264, and Gly-265 (Story and Steitz 1992). Interestingly, *E. coli* RecA Tyr-264 is conserved as an aromatic residue in the RecA family and located at the carboxy-terminal end of strand 8 similar (but seemingly not identical) to the position of T7 gp4 Tyr-535. A conserved aromatic residue close to the carboxy-terminal end of strand 8 is also present in the RadA/DMC1, Sms, and KaiC families, but they do not seem to align exactly with the aromatic residues in either RecA or gp4/DnaB (Fig. 1). The lack of exact superposition could be caused by a suboptimal alignment or, alternatively, might indicate that the spatial orientation of the nucleoside with respect to the phosphate moiety differs between the various members of the RecA/DnaB superfamily.

Similarities between DnaB and RecA can also be found in the subunit interface. Hexamer formation in T7 gp4 depends on helix A that is located at the amino terminus of the helicase domain (Sawaya et al. 1999). It protrudes from the rest of the molecule and completes a three-helix bundle (helices D1, D2, and D3) on a neighboring subunit (Sawaya et al. 1999). Similarly, in the RecA polymer, large parts of the subunit interface are formed by a protruding amino-terminal helix A (Fig. 2) and strand 0 of one subunit packing against strand 3 and helix E in a neighboring subunit (Story et al. 1992). Thus, although no sequence similarity has been detected in either the protruding amino-terminal helix A or the other interface half around helix D, the structural similarities suggest that the subunit interface is homologous and was already present in the common ancestor of DnaB and RecA. In contrast, the amino terminus of KaiC is located immediately before motif 1 (Fig. 1) and a protruding helix is likely absent. It is therefore unlikely that the KaiC proteins have the ability to hexamerize and the head-to-tail fusion of two RecA-like ATPase domain in the two-domain KaiC genes suggests that they might function as dimers. Similarly, the amino-terminal region of Sms proteins is taken up by the Zn-binding module, which might be an alternative means of dimerization but also could be a DNA-binding domain.

Evolution of the KaiC Family

Among the proteins considered here, the evolution of

the KaiC family is the most difficult to interpret. The gene seemingly has undergone multiple gene duplications and lateral transfers. The typical KaiC protein composed of two RecA-like domains joined head to tail is found in the Cyanobacteria and the Archaea *Archeoglobus*, *Pyrococcus*, and *Methanobacterium* (Fig. 3), whereas it is absent from *Methanococcus* and *Aeropyrum*. As an additional complication, the *Methanobacterium* KaiC is more closely related to one of the *Synechocystis* KaiC paralogs than to the double-domain KaiC found in other Archaea like *Archeoglobus* and *Pyrococcus* KaiC (Fig. 3). In addition to the double-domain KaiC proteins, there is a large number of single-domain KaiC homologs that are all archaeal with the exception of an apparent recent transfer into the hyperthermophilic bacterium *Thermotoga maritima* (Fig. 3). Indeed, whole-genome analysis has shown that almost a quarter of all *T. maritima* genes are likely acquired by lateral transfer from the Archaea (Logsdon and Fanny 1999; Nelson et al. 1999). The KaiC family as a whole seems to originate from the bacterial side of the RecA/DnaB superfamily and is identified as a sister group to the Sms family with varying statistical support in most phylogenetic analyses (results not shown). We hypothesize that the ancestral KaiC was a single-domain protein that has been laterally transferred from the Bacteria into the Archaea and that the two-domain KaiC originated by gene duplication and fusion within the Archaea. In this model, the occurrence of the double-domain KaiC in the Cyanobacteria and its lack in other Bacteria is interpreted as a secondary lateral transfer from the Archaea after the main bacterial lineages had been established.

Evolution of the Eukaryotic DnaB Proteins

There are two types of DnaB proteins in the Eukaryota. The DnaB sequences found in chloroplast genomes are highly similar to the bacterial sequences and the chloroplast DnaB of the red alga *Porphyra* also shares the intein position with Cyanobacteria and a few other bacteria (Pietrokovski 1996) (Fig. 1). There is therefore little doubt that these proteins are vertically inherited from the bacterial endosymbiont that gave rise to the plastids and that they are likely the functional helicases in chloroplast DNA replication. In contrast, the previously undetected nuclear eukaryotic DnaB homologs tend to group with the T-odd bacteriophage proteins (gp4) in which the DnaB helicase domain is fused to a DnaG-type primase domain, although there is no strong statistical support for this clade (Fig. 3). Also, when the nuclear eukaryotic DnaB sequences are used as queries for database searches, they typically show the greatest similarity to the bacteriophage DnaB homologs (data not shown). Furthermore, the DnaB homolog from *Arabidopsis* has the same domain architecture as the phage homologs, with the primase do-

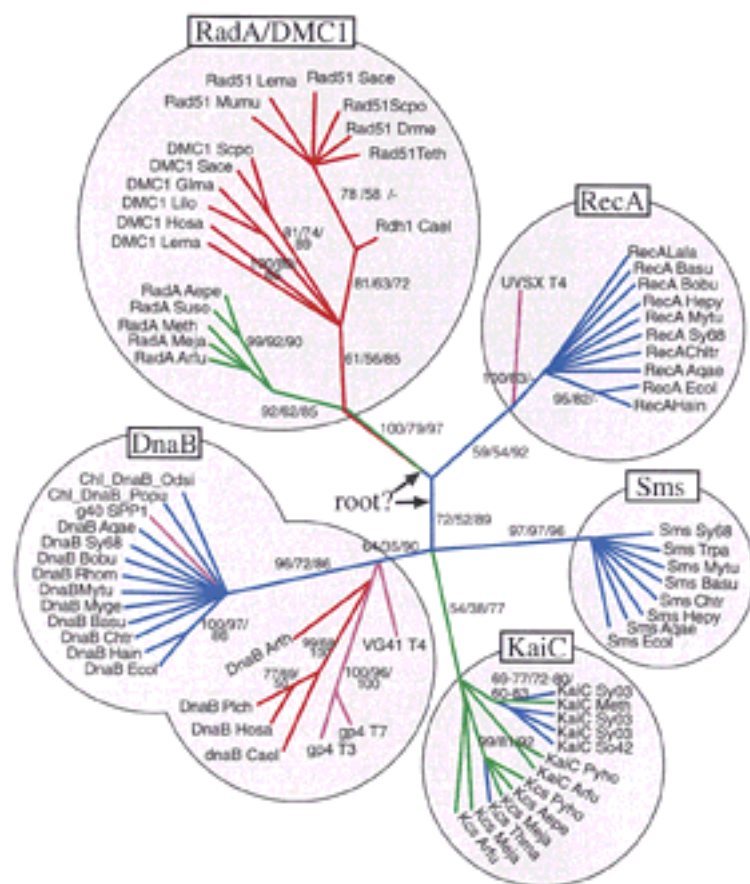


Figure 3 Unrooted phylogeny of the RecA/DnaB superfamily. The analysis is based on an alignment of the RecA/DnaB core domain shown in Fig. 1. The data matrix contains 221 residues seven of which are invariant or parsimony uninformative. Support for individual branches is indicated by bootstrap values for 1000 resampling of PAUP maximum parsimony (first number), PHYLIP distance analysis (second number), and the reliability value computed by the PUZZLE software (third number). Bootstrap values <50% are not recorded and branches without bootstrap numbers are derived from a distance tree computed with the PHYLIP programs protdist and fitch. Branch lengths are arbitrary and do not represent evolutionary distances. The two possible positions of the root as discussed in the text are indicated by black arrows. (Red) Eukaryota; (green) Archaea; (blue) Bacteria; (pink) Bacteriophages. Names in boxes identify the individual protein families. The sequence identifiers are the same as for Fig. 1 except that the GenBank identifier was omitted.

main located upstream of the DnaB domain and containing all the diagnostic sequence motifs of the Toprim domains of the DnaG-type primases (Ilyina et al. 1992; Aravind et al. 1998) (data not shown). The DnaB homolog from the nematode *C. elegans* seems to contain a diverged counterpart of the DnaG domain with disrupted catalytic motifs, and no trace of the DnaG domain could be detected in the homolog from *Plasmodium* (the human coding sequence is incomplete and it remains unclear whether or not the protein contains a DnaG domain). This conservation of a unique domain architecture between nuclear eukaryotic and bacteriophage DnaB homologs, together with the apparent absence of DnaB homologs in Archaea, suggests that the gene coding for the DnaB homolog

probably has been horizontally transferred into eukaryotes via a bacteriophage. Subsequent evolution of this gene in eukaryotes seemed to have involved degradation of the primase domain, at least in some lineages, whereas the helicase domain remained intact. The unexpected tree topology for the eukaryotic DnaB homologs, namely the strongly supported grouping of the *Plasmodium* protein with the human one and the lack of statistically significant grouping of the plant protein with the rest of the eukaryotes, suggest a complex evolutionary history of this gene, perhaps involving additional horizontal transfer events. The functions of the nuclear eukaryotic DnaB homologs remain unclear. The plant and animal DnaB homologs contain a amonia-terminal extension that is likely to function as an organellar import peptide; thus, a role in mitochondrial DNA replication or repair seems a possibility. This possible use of the phage DnaB for organellar function is reminiscent of a similar adaptation of a T-odd phage RNA polymerase in organellar transcription in plants (Hedtke et al. 1997).

Evolution of the RecA/DnaB Superfamily

The sequence similarity between DnaB and RecA and their shared ability to form hexameric rings or helices of similar quaternary structure (Ogawa et al. 1993; Yu and Egelman 1993, 1977; Yu et al. 1996; Seitz et al. 1998) raise the question of whether the RecA/DnaB superfamily is related to other hexameric P-loop NTPases. There is no evidence of a specific relationship with the hexameric/dodecameric branch-migration helicase RuvB (Mitchell and West 1994) or SV40 large T antigen helicase (Mastrangelo et al. 1989; Weisshart et al. 1999) both of which belong to the AAA+ class, a distinct division of P-loop NTPases (Neuwald et al. 1999; L. Aravind and E.V. Koonin, unpubl.). In contrast, there are distinct similarities between the RecA/DnaB superfamily and the family of ATPases that includes transcription termination factor Rho and F_1 -ATPase (Dombroski and Platt 1988; Gorbalenya and Koonin 1993; Miwa et al. 1995; Washington et al. 1996). Within the core domain of RecA and F_1 -ATPase (corresponding to strands 1–8 of RecA and the associated helices and loops), ~130 residues can be superimposed with a Rmsd of <2.0 Å (Abrahams et al. 1994) and secondary structure elements also are largely congruent (Washington et al. 1996). Although this leaves little doubt that the RecA/DnaB superfamily and the Rho/ F_1 family share a com-

mon ancestor that already had a hexameric quaternary structure, it also indicates that hexameric NTPases as a whole (including RecA/DnaB, Rho/F₁, and the AAA+ class) are not a monophyletic group.

Phylogenetic analysis based on the multiple alignment of the core RecA/DnaB domain (~250 residues) strongly supports the monophyly of six major groups, namely bacterial and chloroplast DnaB, eukaryotic DnaB homologs (with the exception of the plant one), bacterial Sms, KaiC, bacterial RecA, and the archaeal/eukaryotic Rad51/DMC1/RadA (Fig. 3). The most critical factor in interpreting this tree is the placement of the root. Unambiguous rooting is possible only when a reliable tree can be produced for two paralogous families resulting from a duplication known to be present in the last common ancestor (Gogarten et al. 1989; Iwabe et al. 1989; Brown and Doolittle 1995). To that end, we have used the Rho/F₁ ATPase family as the paralogous group for the entire RecA/DnaB superfamily. However, the information contained in the overall alignment was insufficient to obtain a reliable rooting (data not shown). Thus, the topology of the tree allows for two principal, competing interpretations (Fig. 3). Placing the root between the RecA/Rad51/DMC1/RadA recombinases and the predominately bacterial assemblage of Sms, DnaB, and KaiC suggests an evolutionary scenario in which a gene duplication in the LCA produced the ancestor of DnaB/Sms/KaiC on the one hand and the RecA/Rad51/RadA recombinases on the other hand, and a later gene duplication in the bacterial lineage gave rise to DnaB and Sms. Consequently, the model has to assume that the ancestor of DnaB/Sms/KaiC has been secondarily lost from the archaeo-eukaryotic lineage. Alternatively, the root can be placed between the archaeo-eukaryotic proteins (Rad51/DMC1/RadA) and the bacterial families (RecA/Sms/DnaB/KaiC) (Fig. 3). In this scenario, the RecA/DnaB superfamily evolved from a single gene in the LCA and the bacterial subfamilies, namely RecA, DnaB, Sms (and possibly KaiC), are derived from successive gene duplication events within the bacterial lineage. The data available do not allow us to distinguish with certainty between these two scenarios, but we favor the rooting between Rad51/DMC1/RadA and RecA because it is the more parsimonious alternative that does not invoke a secondary gene loss.

Conclusions

We show here that the DnaB and RecA/DMC1/RadA proteins form a distinct superfamily of structurally and evolutionarily related ATPases. Additionally, we describe previously undetected DnaB homologs from phylogenetically divergent eukaryotes. The eukaryotic DnaB homolog that shares a common domain organization with T-odd bacteriophage primases–helicases

might have been horizontally transferred into the eukaryotic lineage and is unlikely to play a critical role in eukaryotic nuclear DNA replication given its absence in yeast. Instead, the eukaryotic DnaB homologs are likely to function in organelles. These findings have consequences for our understanding of the evolution of DNA replication. Given the involvement of RecA/DMC1/RadA in recombinational processes in all domains of life, it seems likely that this particular family was already represented in the LCA of all extant cellular organisms. In contrast, DnaB, which is the principal helicase involved in bacterial DNA replication, has apparently been recruited for this function after the divergence of bacteria from the archaeal/eukaryotic lineage. Given that any replicative helicase has to be a highly processive enzyme, the ability of RecA to form hexameric rings (with the right diameter to encircle DNA) offers an explanation why a RecA derivative was a suitable candidate to be selected as the principal helicase for bacterial DNA replication. Conversely, eukaryotic replicative helicases might have been independently recruited from other classes of ATPases, such as the AAA+ class or the superfamily II helicases. The notion that the replicative DNA helicase of the Bacteria is not an ortholog of the corresponding replicative helicases in Archaea and Eukaryota is compatible with the recently discussed hypothesis that the modern-type system for the replication of ds DNA has evolved independently in the bacterial and archaeal/eukaryotic lineages (Leipe et al. 1999).

Methods

The nonredundant database of protein sequences at the NCBI (NR) was searched using the gapped BLASTP and PSI-BLAST programs (Altschul et al. 1997). Briefly, the PSI-BLAST program constructs a position-dependent weight matrix (profile) using multiple alignments generated from the BLAST hits above a certain expectation value (*E* value) and carries out iterative database searches using the information derived from the profile. The statistical evaluation of the PSI-BLAST results is based on the extreme value distribution statistics originally developed by Karlin and Altschul (1990) for local alignments without gaps and subsequently shown by extensive computer simulations to apply also to gapped alignments and to alignments obtained by using profiles (Altschul and Gish 1996; Altschul et al. 1997). It has been emphasized that *E* values reported for each retrieved sequence at the point when its alignment with the query sequence passes the cutoff for the first time are robust estimates of statistical significance. Once a sequence gets included in the profile, *E* values reported for it and its close homologs at subsequent iterations become inflated and do not represent the statistical significance (Altschul and Koonin 1998). Here we only report *E* values for the first appearance of the given sequence above the cutoff. The dbEST was searched using the gapped TBLASTN program (Altschul et al. 1997).

Multiple sequence alignments were constructed using the PSI-BLAST output and modified manually on the basis of structural considerations. The alignments were formatted us-

ing the SEAVIEW (Galtier et al. 1996) and ALSCRIPT programs (Barton 1993). Protein databank (PDB) files were visualized and manipulated using the MOLSCRIPT program (Kraulis 1991). Phylogenetic trees were constructed using distance (neighbor joining), maximum likelihood, and maximum parsimony methods as implemented in the PHYLIP (Felsenstein 1993), PUZZLE (Strimmer and von Haessler 1996), and PAUP (Strimmer and von Haessler 1996) (Swofford 1999) programs, respectively. To measure support for individual tree branches, the reliability values for the quartet puzzling method and bootstrap values for distance and parsimony trees have been recorded (Strimmer and von Haessler 1996; Swofford 1999).

ACKNOWLEDGMENTS

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

REFERENCES

- Abrahams, J.P., A.G. Leslie, R. Lutter, and J.E. Walker. 1994. Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature* **370**: 621–628.
- Altschul, S.F. and W. Gish. 1996. Local alignment statistics. *Methods Enzymol.* **266**: 460–480.
- Altschul, S.F. and E.V. Koonin. 1998. Iterated profile searches with PSI-BLAST—A tool for discovery in protein databases. *Trends Biochem. Sci.* **23**: 444–447.
- Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Aravind, L., D.D. Leipe, and E.V. Koonin. 1998. Toprim—A conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **26**: 4205–4213.
- Aravind, L., D.R. Walker, and E.V. Koonin. 1999. Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res.* **27**: 1223–1242.
- Barton, G.J. 1993. ALSCRIPT: A tool to format multiple sequence alignments. *Protein Eng.* **6**: 37–40.
- Bishop, D.K., D. Park, L. Xu, and N. Kleckner. 1992. DMC1: A meiosis-specific yeast homolog of *E. coli* recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* **69**: 439–456.
- Biswas, S.B., P.H. Chen, and E.E. Biswas. 1994. Structure and function of *Escherichia coli* DnaB protein: Role of the N-terminal domain in helicase activity. *Biochemistry* **33**: 11307–11314.
- Brown, J.R. and W.F. Doolittle. 1995. Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc. Natl. Acad. Sci.* **92**: 2441–2445.
- Campbell, M.J. and R.W. Davis. 1999a. Toxic mutations in the recA gene of *E. coli* prevent proper chromosome segregation. *J. Mol. Biol.* **286**: 417–435.
- . 1999b. On the in vivo function of the RecA ATPase. *J. Mol. Biol.* **286**: 437–445.
- Cartwright, R., A.M. Dunn, P.J. Simpson, C.E. Tambini, and J. Thacker. 1998. Isolation of novel human and mouse genes of the recA/RAD51 recombination-repair gene family. *Nucleic Acids Res.* **26**: 1653–16539.
- Dombroski, A.J. and T. Platt. 1988. Structure of rho factor: An RNA-binding domain and a separate region with strong similarity to proven ATP-binding domains. *Proc. Natl. Acad. Sci.* **85**: 2538–2542.
- Doutriaux, M.P., F. Couteau, C. Bergounioux, and C. White. 1998. Isolation and characterisation of the RAD51 and DMC1 homologs from *Arabidopsis thaliana*. *Mol. 131 Gen. Genet.* **257**: 283–291.
- Dresser, M.E., D.J. Ewing, M.N. Conrad, A.M. Dominguez, R. Barstead, H. Jiang, and T. Kodadek. 1997. DMC1 functions in a *Saccharomyces cerevisiae* meiotic pathway that is largely independent of the RAD51 pathway. *Genetics* **147**: 533–544.
- Fass, D., C.E. Bogden, and J.M. Berger. 1999. Crystal structure of the N-terminal domain of the DnaB hexameric helicase. *Struct. Fold Des.* **7**: 691–698.
- Felsenstein, J. 1993. PHYLIP 3.5c—Computer package distributed by the author, Department of Genetics SK-50, University of Washington, Seattle, Washington 98195. <http://evolution.genetics.washington.edu/phylip.html>
- Galtier, N., M. Gouy, and C. Gautier. 1996. SEAVIEW and PHYLO_WIN: Two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* **12**: 543–548.
- Game, J.C. 1993. DNA double-strand breaks and the RAD50-RAD57 genes in *Saccharomyces*. *Semin. Cancer Biol.* **4**: 73–83.
- Gogarten, J.P., H. Kibak, P. Dittrich, L. Taiz, E.J. Bowman, B.J. Bowman, M.F. Manolson, R.J. Poole, T. Date, T. Oshima, et al. 1989. Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. *Proc. Natl. Acad. Sci.* **86**: 6661–6665.
- Gorbalenya, A.E. and E.V. Koonin. 1993. Helicases: Amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Bio.* **3**: 419–429.
- Hedtke, B., T. Borner, and A. Weihe. 1997. Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*. *Science* **277**: 809–811.
- Ilyina, T.V., A.E. Gorbalenya, and E.V. Koonin. 1992. Organization and evolution of bacterial and bacteriophage primase-helicase systems. *J. Mol. Evol.* **34**: 351–357.
- Ishiura, M., S. Kutsuna, S. Aoki, H. Iwasaki, C.R. Andersson, A. Tanabe, S.S. Golden, C.H. Johnson, and T. Kondo. 1998. Expression of a gene cluster kaiABC as a circadian feedback process in cyanobacteria. *Science* **281**: 1519–1523.
- Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa, and T. Miyata. 1989. Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc. Natl. Acad. Sci.* **86**: 9355–9359.
- Iwasaki, H., Y. Taniguchi, M. Ishiura, and T. Kondo. 1999. Physical interactions among circadian clock proteins KaiA, KaiB and KaiC in cyanobacteria. *EMBO J.* **18**: 1137–1145.
- Karlin, S. and S.F. Altschul. 1990. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci.* **87**: 2264–2268.
- Kawabata, M. and K. Saeki. 1998. Sequence analysis and expression of a novel mouse homolog of *Escherichia coli* recA gene. *Biochim. Biophys. Acta* **1398**: 353–358.
- Koonin, E.V., R.L. Tatusov, and K.E. Rudd. 1996. *Escherichia coli* protein sequences: Functional and evolutionary implications. In *Escherichia coli and Salmonella. Cellular and molecular biology* (ed. F.C. Neidhardt et al.), pp. 2203–2217. Washington, D.C.
- Kowalczykowski, S.C. and A.K. Eggleston. 1994. Homologous pairing and DNA strand-exchange proteins. *Annu. Rev. Biochem.* **63**: 991–1043.
- Kowalczykowski, S.C., D.A. Dixon, A.K. Eggleston, S.D. Lauder, and W.M. Rehrauer. 1994. Biochemistry of homologous recombination in *Escherichia coli*. *Microbiol Rev.* **58**: 401–465.
- Kraulis, P.J. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**: 946–950.
- Leipe, D., L. Aravind, and E. Koonin. 1999. Did DNA replication evolve twice independently? *Nucleic Acids Res.* **27**: 3389–3401.
- Little, J.W. and D.W. Mount. 1982. The SOS regulatory system of *Escherichia coli*. *Cell* **29**: 11–22.
- Logsdon, J.M. and D.M. Faguy. 1999. Thermotoga heats up lateral gene transfer. *Curr. Biol.* **9**: R747–R751.
- Lu, Y.B., P.V. Ratnakar, B.K. Mohanty, and D. Bastia. 1996. Direct physical interaction between DnaG primase and DnaB helicase of *Escherichia coli* is necessary for optimal synthesis of primer RNA. *Proc. Natl. Acad. Sci.* **93**: 12902–12907.
- Makarova, K.S., L. Aravind, M.Y. Galperin, N.V. Grishin, R.L. Tatusov, Y.I. Wolf, and E.V. Koonin. 1999. Comparative

- genomics of the Archaea (Euryarchaeota): Evolution of conserved protein families, the stable core, and the variable shell. *Genome Res.* **9**: 608–628.
- Mastrangelo, I.A., P.V. Hough, J.S. Wall, M. Dodson, F.B. Dean, and J. Hurwitz. 1989. ATP-dependent assembly of double hexamers of SV40 T antigen at the viral origin of DNA replication. *Nature* **338**: 658–662.
- Miles, C.S., J. Weigelt, N.P. Stamford, N. Dammerova, G. Otting, and N.E. Dixon. 1997. Precise limits of the N-terminal domain of DnaB helicase determined by NMR spectroscopy. *Biochem. Biophys. Res. Commun.* **231**: 126–130.
- Mitchell, A.H. and S.C. West. 1994. Hexameric rings of Escherichia coli RuvB protein. Cooperative assembly, processivity and ATPase activity. *J. Mol. Biol.* **243**: 208–215.
- Miwa, Y., T. Horiguchi, and K. Shigesada. 1995. Structural and functional dissections of transcription termination factor rho by random mutagenesis. *J. Mol. Biol.* **254**: 815–837.
- Nakayama, N., N. Arai, Y. Kaziro, and K. Arai. 1984. Structural and functional studies of the dnaB protein using limited proteolysis. Characterization of domains for DNA-dependent ATP hydrolysis and for protein association in the primosome. *J. Biol. Chem.* **259**: 88–96.
- Nelson, K.E., R.A. Clayton, S.R. Gill, M.L. Gwinn, R.J. Dodson, D.H. Haft, E.K. Hickey, J.D. Peterson, W.C. Nelson, K.A. Ketchum et al. 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature* **399**: 323–329.
- Neuwald, A.F., D.E. Berg, and G.V. Stauffer. 1992. Mutational analysis of the Escherichia coli serB promoter region reveals transcriptional linkage to a downstream gene. *Gene* **120**: 1–9.
- Neuwald, A.F., L. Aravind, J.L. Spouge, and E.V. Koonin. 1999. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res.* **9**: 27–43.
- Ogawa, T., X. Yu, A. Shinohara, and E.H. Egelman. 1993. Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science* **259**: 1896–1899.
- Pietrokovski, S. 1996. A new intein in cyanobacteria and its significance for the spread of inteins. *Trends Genet.* **12**: 287–288.
- Radding, C.M. 1989. Helical RecA nucleoprotein filaments mediate homologous pairing and strand exchange. *Biochim. Biophys. Acta.* **1008**: 131–145.
- Rashid, N., M. Morikawa, and T. Imanaka. 1996. A RecA/RAD51 homologue from a hyperthermophilic archaeon retains the major RecA domain only. *Mol. Gen. 131 Genet.* **253**: 397–400.
- Sandler, S.J., L.H. Satin, H.S. Samra, and A.J. Clark. 1996. recA-like genes from three archaean species with putative protein products similar to Rad51 and Dmc1 proteins of the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **24**: 2125–2132.
- Saraste, M., P.R. Sibbald, and A. Wittinghofer. 1990. The P-loop—A common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* **15**: 430–434.
- Sawaya, M.R., S. Guo, S. Tabor, C.C. Richardson, and T. Ellenberger. 1999. Crystal structure of the helicase domain from the replicative helicase-primase of bacteriophage T7. *Cell* **99**: 167–177.
- Seitz, E.M., J.P. Brockman, S.J. Sandler, A.J. Clark, and S.C. Kowalczykowski. 1998. RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. *Genes & Dev.* **12**: 1248–1253.
- Shinohara, A., H. Ogawa, and T. Ogawa. 1992. Rad51 protein involved in repair and recombination in *S. cerevisiae* is a RecA-like protein. *Cell* **69**: 457–470.
- Shinohara, A., S. Gasiar, T. Ogawa, N. Kleckner, and D. K. Bishop. 1997. *Saccharomyces cerevisiae* recA homologues RAD51 and DMC1 have both distinct and overlapping roles in meiotic recombination. *Genes Cells* **2**: 615–629.
- Song, Y. and N.J. Sargentini. 1996. Escherichia coli DNA repair genes radA and sms are the same gene. *J. Bacteriol.* **178**: 5045–5048.
- Story, R.M. and T.A. Steitz. 1992. Structure of the recA protein-ADP complex. *Nature* **355**: 374–376.
- Story, R.M., I.T. Weber, and T.A. Steitz. 1992. The structure of the E. coli recA protein monomer and polymer. *Nature* **355**: 318–325.
- Story, R.M., D.K. Bishop, N. Kleckner, and T. A. Steitz. 1993. Structural relationship of bacterial RecA proteins to recombination proteins from bacteriophage T4 and yeast. *Science* **259**: 1892–1896.
- Strimmer, K. and A. von Haeseler. 1996. Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13**: 964–969.
- Sutton, M.D., K.M. Carr, M. Vicente, and J.M. Kaguni. 1998. Escherichia coli DnaA protein. The N-terminal domain and loading of DnaB helicase at the E. coli chromosomal origin. *J. Biol. Chem.* **273**: 34255–34262.
- Swofford, D.L. 1999. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.
- Takanami, T., S. Sato, T. Ishihara, I. Katsura, H. Takahashi, and A. Higashitani. 1998. Characterization of a Caenorhabditis elegans recA-like gene Ce-rdh-1 involved in meiotic recombination. *DNA Res.* **5**: 373–377.
- Tambini, C.E., A.M. George, J.M. Rommens, L.C. Tsui, S.W. Scherer, and J. Thacker. 1997. The XRCC2 DNA repair gene: identification of a positional candidate. *Genomics* **41**: 84–92.
- Tougu, K., H. Peng, and K.J. Marians. 1994. Identification of a domain of Escherichia coli primase required for functional interaction with the DnaB helicase at the replication fork. *J. Biol. Chem.* **269**: 4675–4682.
- Washington, M.T., A.H. Rosenberg, K. Griffin, F.W. Studier, and S.S. Patel. 1996. Biochemical analysis of mutant T7 primase/helicase proteins defective in DNA binding, nucleotide hydrolysis, and the coupling of hydrolysis with DNA unwinding. *J. Biol. Chem.* **271**: 26825–26834.
- Weigelt, J., C.S. Miles, N.E. Dixon, and G. Otting. 1998. Backbone NMR assignments and secondary structure of the N-terminal domain of DnaB helicase from E. coli. *J. Biomol. NMR* **11**: 233–234.
- Weigelt, J., S.E. Brown, C.S. Miles, N.E. Dixon, and G. Otting. 1999. NMR structure of the N-terminal domain of E. coli DnaB helicase: Implications for structure rearrangements in the helicase hexamer. *Struct. Fold Des.* **7**: 681–690.
- Weisshart, K., P. Taneja, A. Jenne, U. Herbig, D.T. Simmons, and E. Fanning. 1999. Two regions of simian virus 40 T antigen determine cooperativity of double-hexamer assembly on the viral origin of DNA replication and promote hexamer interactions during bidirectional origin DNA unwinding. *J. Virol.* **73**: 2201–2211.
- Witkin, E. M. 1991. RecA protein in the SOS response: Milestones and mysteries. *Biochimie* **73**: 133–141.
- Yoshida, K., G. Kondoh, Y. Matsuda, T. Habu, Y. Nishimune, and T. Morita. 1998. The mouse RecA-like gene Dmc1 is required for homologous chromosome synapsis during meiosis. *Mol. Cell* **1**: 707–718.
- Yoshida, M. and T. Amano. 1995. A common topology of proteins catalyzing ATP-triggered reactions. *FEBS Lett.* **359**: 1–5.
- Yu, X. and E. H. Egelman. 1993. DNA conformation induced by the bacteriophage T4 UvsX protein appears identical to the conformation induced by the Escherichia coli RecA protein. *J. Mol. Biol.* **232**: 1–4.
- . 1997. The RecA hexamer is a structural homologue of ring helicases. *Nat. Struct. Biol.* **4**: 101–104.
- Yu, X., M.J. Jezewska, W. Bujalowski, and E.H. Egelman. 1996. The hexameric E. coli DnaB helicase can exist in different Quaternary states. *J. Mol. Biol.* **259**: 7–14.

Received July 26, 1999; accepted in revised form November 16, 1999.