The recA Gene from the Thermophile Thermus aquaticus YT-1: Cloning, Expression, and Characterization

EVELINA ANGOV AND R. DANIEL CAMERINI-OTERO*

Genetics and Biochemistry Branch, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20892

Received 16 August 1993/Accepted 28 December 1993

We have cloned, expressed, and purified the RecA analog from the thermophilic eubacterium Thermus aquaticus YT-1. Analysis of the deduced amino acid sequence indicates that the T. aquaticus RecA is structurally similar to the Escherichia coli RecA and suggests that RecA-like function has been conserved in thermophilic organisms. Preliminary biochemical analysis indicates that the protein has an ATP-dependent single-stranded DNA binding activity and can pair and carry out strand exchange to form a heteroduplex DNA under reaction conditions previously described for E. coli RecA, but at 55 to 65°C. Further characterization of a thermophilically derived RecA protein should yield important information concerning DNA-protein interactions at high temperatures. In addition, a thermostable RecA protein may have some general applicability in stabilizing DNA-protein interactions in reactions which occur at high temperatures by increasing the specificity (stringency) of annealing reactions.

The Escherichia coli RecA protein plays an essential role in genetic recombination via a multistep pathway involving the formation of a single-stranded nucleoprotein filament, DNA pairing, and subsequent strand exchange to form heteroduplex DNA. RecA is directly involved in DNA repair and the induction of the SOS response by cleavage of the LexA repressor (6, 28, 34). The E. coli recA gene and protein are well characterized (16). To date, a large number of other bacterial genes with sequence homology to the E. coli recA gene have been described, both in very closely related enteric bacteria (e.g., Proteus mirabilis) as well as in more distant genera (e.g., cyanobacteria). The encoded proteins have amino acid sequence identities to E. coli RecA that range from 56 to 100% (30). Significantly, among each of the 23 bacterial RecA proteins described, more than 55% of the amino acids are either invariant or conservative replacements. Even so, in vitro biochemical activities such as single-stranded DNA binding and strand exchange have been described for only two other bacterial RecA proteins, those from P. mirabilis and Bacillus subtilis (19, 39).

The more distantly related bacteriophage T4 UvsX protein promotes homologous recombination in a similar manner. This protein, however, does not cross-react antigenically with antibodies raised to E. coli RecA and reveals only 23% identity at the amino acid level (8). Recently, two eukaryotic RecA analogs from Saccharomyces cerevisiae, RAD51 and DMC1 (2, 33), have been isolated and described. The deduced amino acid sequences of the two yeast proteins are homologous to that of E. coli RecA protein. In addition, on the basis of amino acid sequence analysis of DMC1, RAD51, and T4 UvsX and crystallographic data from E. coli RecA, the yeast proteins and T4 UvsX protein are structurally similar to the E. coli RecA (35, 37). Even though the UvsX protein has been demonstrated to carry out strand exchange and single-stranded DNA binding in vitro (15) and the RAD51 protein has been shown to form distinctive RecA-like nucleoprotein filaments on double-stranded DNA (27), in vitro strand exchange activity or single-stranded DNA binding has not yet been demonstrated for the eukaryotic RecA homologs.

We present data for the first known thermophilic eubacterial RecA protein and, to our knowledge, the first gene encoding a thermostable homologous recombination protein of any kind. This is of particular importance because homologous recombination involves nucleic acid hybridizations and exchange reactions such as triplex DNA formation and branch migration (4), nucleic acid transactions that are all very sensitive to temperature. Analysis of this protein is also of significant interest, since few other thermostable enzymes of nucleic acid metabolism have been described. In the present study, we have cloned the recA gene from the thermophilic eubacterium Thermus aquaticus YT-1 and overproduced the thermostable protein in E. coli. Analysis of the deduced amino acid sequence indicates that the T. aquaticus RecA is highly conserved relative to the E. coli protein. The protein has been purified, and preliminary biochemical data indicate that the T. aquaticus RecA has ATP-dependent single-stranded DNA binding activity and can pair homologous DNAs to form stable joint molecules under well-defined experimental conditions. A thermostable RecA protein may be suitable for biochemical reactions which occur at higher temperatures by enhancing the specificity (stringency) of the pairing reaction between homologous DNAs.

MATERIALS AND METHODS

Bacterial strains. T. aquaticus YT-1 (ATCC 25105) was purchased from the American Type Culture Collection (Rockville, Md.). E. coli XL1-Blue (recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB+ lacIq lacZAM15 Tn10(Tet) ]) was used as the host for screening the T. aquaticus genomic DNA library by plaque hybridization (Stratagene). E. coli JM109 (recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 Δ(lac-pro-AB)[F' proAB+ lacP+ lacZAM15 traD36]) was the host strain used for T. aquaticus RecA protein expression.

Enzymes and chemicals. Restriction enzymes were purchased from New England Biolabs and from Promega. T4 DNA ligase, T4 DNA polymerase, and T4 polynucleotide kinase were purchased from New England Biolabs. [α-32P]
dCTP, [γ-32P]ATP, and [α-35S]dATP were obtained from Dupont, NEN.

Cloning the T. aquaticus recA gene. Standard molecular cloning techniques were used to clone the T. aquaticus recA gene (20). T. aquaticus cell cultures were grown at 70°C in Thermus medium (4 g of yeast extract per liter, 8 g of Polypeptone [BBL 11910] per liter, 2 g of NaCl per liter, adjusted to pH 7.5). T. aquaticus genomic DNA was prepared as described previously (18). Degenerate and nondegenerate oligonucleotides for PCR and DNA sequencing were synthesized by the phosphoramidite method on an Applied Biosystems model 380B DNA synthesizer. Double-stranded DNA fragments were isolated by PCR (Perkin-Elmer Cetus) and were random prime labeled with the Ambion Decaprime DNA labeling kit for probing Southern blots and plaque hybridization. The primers used to amplify the 210-bp fragment from T. aquaticus genomic DNA were 5'-GGGGAATTCGG(T/C/G)(A/G)T(T/G)GA(T/C)TC(T/G/A) (SacI) and 5'-GGGGGATCCGT(G/A)CT(G/T)GA(T/C)TC(T/G/A) (Glycol). Amplified DNA products were subcloned into pBluescript vectors and sequenced with T3 and T7 primers. A SacI genomic DNA library was constructed in the λZAPII expression system and subsequently probed with the 210-bp fragment mentioned above. Positive plaques were purified through four rounds, and clones were excised by in vivo excision (Stratagene) to generate the cloned fragments in the pBluescript vector. A 1,417-bp fragment which carries the T. aquaticus recA gene was isolated.

DNA sequencing. DNA sequencing was performed by the Sanger dideoxy chain termination method with the Sequenase version 2.0 DNA sequencing kit (U.S. Biochemical) (32). Computer analysis of nucleotide and amino acid primary sequences was done by the University of Wisconsin Genetics Computer Group sequence analysis package (BESTFIT program).

Construction of a recombinant expression vector for T. aquaticus RecA. To overproduce T. aquaticus RecA from E. coli, we subcloned the T. aquaticus recA gene, which is part of the 1.4-kb fragment isolated by in vivo excision of a λZAPII library construct, into the expression vector pTrc99a (Pharmacia LKB Biotechnology), which carries a strong tac promoter (trp [−35] region and the lacUV5 [−10] region), a ribosome-binding site, an ampicillin resistance gene, lacIq, and strong transcription termination signals in cis. Plasmid pTrc99a was digested with NcoI and treated with T4 DNA polymerase and then was treated with calf intestinal phosphatase. The insert DNA was prepared by digesting the pBluescript plasmid with SacI to generate the 1.4-kb fragment and treated with T4 DNA polymerase to generate blunt ends. The DNA was ligated with T4 DNA ligase at 14°C. Recombinant clones were sequenced on both strands to determine the orientation of inserts.

Expression and purification of the T. aquaticus RecA in E. coli. E. coli JM109 cells transformed with the recombinant plasmid were grown at 37°C in Luria-Bertani medium (10 g of tryptone per liter, 10 g of NaCl per liter, 5 g of yeast extract per liter) containing 100 μg of ampicillin (Sigma Chemical Co.) per ml to mid-log phase (optical density at 600 nm, 0.4) and induced with isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 0.4 mM (Gold Biotechnology, Inc.). The cells were harvested after 6 h by centrifugation. The purification scheme is based on an E. coli RecA purification described by Steve Kowalczykowski (15a) with some modifications. The harvest cells were resuspended in lysozyme buffer (2 ml/g of wet cells) containing 100 mM NaCl, 40 mM Tris-HCl (pH 8.0), 10 mM MgCl2, 1 mM EDTA, and 1 mM β-mercaptoethanol. Cells were disrupted by sonication and treated with DNase I (20 μg/ml) and 3 mM phenylmethylsulfonyl fluoride at 23°C for 1 h. A crude cell extract was obtained by centrifugation (30 min at 10,000 × g). Most endogenous E. coli proteins were denatured by heat treatment at 70°C for 45 min and removed from the cell extract by centrifugation at 16,000 × g for 1 h. The supernatant fluid was brought to 65% saturation with (NH4)2SO4 precipitated, and collected by centrifugation (16,000 × g for 1 h). The pellet was resuspended in phosphate buffer (20 mM potassium phosphate [pH 6.5], 10% [vol/vol] glycerol, 0.1 mM dithiothreitol, 0.1 mM EDTA). The protein sample was dialyzed extensively to remove excess (NH4)2SO4 and was applied to a 20-ml DEAE-Sephal column (Pharmacia Fine Chemicals) equilibrated in phosphate buffer. (Subsequent protein purifications of T. aquaticus RecA omitted the ammonium sulfate precipitation since it did not contribute to the overall purification scheme.) The T. aquaticus RecA was eluted at 180 mM NaCl with a linear gradient (300-ml total volume) of NaCl from 0 to 400 mM in phosphate buffer. (Even though the Coomassie-stained gel of Fig. 3A does not demonstrate significant purification at this step, the DEAE fractionation was not omitted since its exclusion yields a less pure protein fraction from single-stranded DNA-agarose.) Fractions containing T. aquaticus RecA were pooled, dialyzed against phosphate buffer, and applied to a 25-ml packed single-stranded DNA-agarose-affinity column (Bethesda Research Laboratories) equilibrated in phosphate buffer. Non-specifically bound proteins were eluted with phosphate buffer containing 50 mM NaCl. Bound T. aquaticus RecA protein was eluted from the single-stranded DNA-agarose column with 1 mM ADP in phosphate buffer containing 50 mM NaCl. The final pooled fraction was concentrated with Ficoll 400 (Pharmacia LKB Biotechnology) and dialyzed into 20 mM Tris-HCl (pH 7.5)-50% (vol/vol) glycerol-0.1 mM dithiothreitol-0.1 mM EDTA-100 mM KCl for storage at 4°C. Twenty grams of wet cells yielded approximately 2 mg of T. aquaticus RecA protein. The relative protein concentration was determined by the Bradford method (Bio-Rad). The absolute concentration of T. aquaticus RecA was determined from the calculated difference in its relative response to Coomassie brilliant blue G-250 (5) compared with that of E. coli RecA and a calibration curve determined for the E. coli protein from its known extinction coefficient at 280 nm (17). Protein staining of mini-12% Tris-glycine—sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels (Novex) was with Coomassie brilliant blue R250.

Immunochromatic detection of T. aquaticus RecA. Protein samples were run on 12% Tris-glycine—SDS-polyacrylamide gels and transferred to nitrocellulose membranes. Nonspecific binding sites were blocked in 10% (wt/vol) nonfat dry milk-0.1% (vol/vol) Tween 20–1× phosphate-buffered saline (pH 7.4) at room temperature. Blots were incubated with primary polyclonal anti-E. coli RecA antibody, which was purified with MAbTrap G (Pharmacia LKB) (1:6,000), followed by incubation with anti-rabbit immunoglobulin G-conjugated alkaline phosphatase (Sigma) (1:1,200). Primary and secondary antibodies were diluted in 5% (wt/vol) nonfat dry milk-0.1% (vol/vol) Tween 20–1× phosphate-buffered saline. Color was developed with substrates provided by the Promega Promega Western blot (immunoblot) alkaline phosphatase system.

Filter binding assay. Millipore HAWP 025 (0.45-μm pore size) nitrocellulose membranes were prepared by treatment with 0.3 N NaOH for 30 min in the dark before being rinsed extensively with distilled H2O. Membranes were stored at room temperature in 50 mM Tris-HCl (pH 7.5)-12.5 mM MgCl2 until use. A 30-mer oligonucleotide (5'-CTG CAATGATA CCG GAC ACC CAC GTC CAC-3') was 5' end labeled with [γ-32P]ATP and T4 polynucleotide ligase.
kinase as described previously (20). Reaction mixtures (conditions optimized for formation of synaptic complex with E. coli RecA [9]) containing 20 mM Tris-HCl (pH 7.5), 0.4 mM dithiothreitol, 12.5 mM MgCl₂, 1.1 mM ADP, 0.3 mM ATP-P-S, 15 ng of 32P-labeled oligonucleotide, and 1.5 μg of T. aquaticus RecA were incubated for 30 min at the designated temperatures. Each reaction mixture was applied to a filter under vacuum at room temperature and washed with 3 ml of buffer (room temperature) containing 20 mM Tris-HCl (pH 7.5) and 12.5 mM MgCl₂. Filters were air dried and counted in a Hewlett-Packard 2200CA scintillation counter.

**Strand exchange assay.** The strand exchange conditions were described previously (10). M13mp18 viral DNA (50 ng) (New England Biolabs) and BamHI-linearized M13mp18 replicative-form DNA (50 ng) (New England Biolabs) were incubated with T. aquaticus RecA in the presence of 20 mM Tris-HCl (pH 7.5)–0.4 mM dithiothreitol–12.5 mM MgCl₂–1 mM ATP and an ATP-regenerating system. The concentration of T. aquaticus RecA protein was as indicated in the legend to Fig. 5. T. aquaticus RecA protein was incubated with single-stranded DNA for 2 min at 65°C, followed by an additional 2-min incubation at 65°C with 1 μg of E. coli single-stranded DNA binding protein (Promega). Double-stranded DNA was added to the reaction mixture, and the reaction mixture was further incubated at 65°C for 15 min unless otherwise indicated. Reaction mixtures were deproteinated with 1% SDS and 10 mM EDTA in 0.04% bromophenol blue–10 mM Tris-acetate (pH 8.0)–10% glycerol and electrophoresed at room temperature on 0.8% agarose gels in buffer (40 mM Tris-acetate [pH 8.0], 1 mM EDTA) containing 0.75 μg of ethidium bromide per ml for 14 to 18 h at 0.6 V/cm.

**Nucleotide sequence accession number.** The nucleotide sequence of the T. aquaticus recA gene is available in GenBank under accession no. L20680.

**RESULTS**

Molecular cloning of the T. aquaticus recA gene. There are several highly conserved amino acid sequences among prokaryotic RecA proteins (30). Degenerate oligonucleotides homologous to regions of the E. coli RecA protein that are either highly conserved or functionally significant were synthesized. These oligonucleotides were used as primers in PCR to amplify DNA products from T. aquaticus genomic DNA. A 210-bp DNA product was isolated, sequenced, and shown to have 68% homology at the DNA level and 73% homology at the amino acid level to the respective regions in E. coli. This 210-bp fragment was random primer labeled and used as a probe for Southern hybridization of T. aquaticus genomic DNA digested with several different restriction enzymes. The probe hybridized to a 1.4-kb SacI fragment on Southern blots (data not shown). Subsequently, a SacI T. aquaticus genomic DNA library was constructed in λZAPII and probed with the 210-bp fragment described above. Primary plaque hybridization yielded seven positive plaques. Four of the seven isolates were plaque purified through four rounds. The λZAPII inserts were excised to form pBluescript-derived plasmids by in vivo excision. Restriction digestion of the cloned inserts in the respective pBluescript plasmids yielded an insert size of 1.4 kb, as expected from Southern hybridization.

Both strands of the 1.4-kb SacI insert were sequenced. (Subsequent to submission of the manuscript, a second T. aquaticus recA sequence became available from GenBank [J. G. Wetmur, D. M. Wong, B. Ortiz, F. Reichert, and D. H. Gelfand; accession no. L20095]. Comparison of the two sequences indicated two sites of discrepancy. Reanalysis and sequencing indicated that the second T. aquaticus recA sequence is correct (amino acid positions 132, 164, and 165 and a frameshift at position 336); thus, all further analysis and characterization follow from the corrected sequence.) The sequence indicates the presence of an open reading frame of 1,023 bp with 61 bp upstream of a putative ATG and 334 bp downstream of a putative TAG termination codon (Fig. 1). The open reading frame encodes a putative 341-amino-acid protein with a predicted molecular weight of 36,451. The ATG translation initiation codon is preceded by a consensus Shine-Dalgarno sequence which is complementary to the 3' end of E. coli and Thermus thermophilus 16S rRNA (25). The nucleotide sequence downstream from the TAG translational termination codon does not resemble features of E. coli rho-independent transcription terminators (23, 31).

![Image of the nucleotide sequence of the T. aquaticus recA gene](image_url)
TABLE 1. Codon usage in *T. aquaticus* and *E. coli* recA genes

| Amino acid | Codon | *T. aquaticus* | *E. coli* | Amino acid | Codon | *T. aquaticus* | *E. coli* |
|------------|-------|---------------|-----------|------------|-------|---------------|-----------|
| Phe        | TTT   | 2             | 4         | Tyr        | TAT   | 0             | 0         |
|            | TTC   | 6             | 6         | His        | CAT   | 0             | 0         |
|            | TTA   | 0             | 0         |            | CAC   | 3             | 2         |
|            | TTG   | 4             | 2         |            |       |               |           |
| Leu        | CTT   | 4             | 3         |            |       |               |           |
|            | CTC   | 7             | 2         |            |       |               |           |
|            | CTA   | 1             | 0         |            |       |               |           |
|            | CTG   | 20            | 24        |            |       |               |           |
| Met        | ATG   | 7             | 10        |            |       |               |           |
|            | GTG   | 0             | 0         |            |       |               |           |
| Val        | GTT   | 1             | 4         |            |       |               |           |
|            | GTC   | 11            | 3         |            |       |               |           |
|            | GTA   | 1             | 5         |            |       |               |           |
|            | GTG   | 24            | 10        |            |       |               |           |
| Ser        | TCT   | 0             | 6         |            |       |               |           |
|            | TCC   | 6             | 6         |            |       |               |           |
|            | TCA   | 0             | 2         |            |       |               |           |
|            | TCG   | 4             | 1         |            |       |               |           |
|            | AGT   | 4             | 5         |            |       |               |           |
| Pro        | CTT   | 0             | 0         |            |       |               |           |
|            | CCC   | 9             | 0         |            |       |               |           |
|            | CCA   | 0             | 1         |            |       |               |           |
|            | CGG   | 5             | 9         |            |       |               |           |
| Thr        | ACT   | 1             | 3         |            |       |               |           |
|            | ACC   | 8             | 9         |            |       |               |           |
|            | ACA   | 0             | 0         |            |       |               |           |
|            | AGC   | 2             | 5         |            |       |               |           |
| Ala        | GCT   | 2             | 4         |            |       |               |           |
|            | GCC   | 26            | 4         |            |       |               |           |
|            | GCA   | 0             | 11        |            |       |               |           |
|            | GCG   | 8             | 19        |            |       |               |           |

*Numbers reflect the number of times a particular codon is used by *T. aquaticus* and *E. coli.*

Amino acid composition and codon usage. The amino acid composition of *T. aquaticus* RecA reflects the inherent differences in amino acid preferences between thermophilic and mesophilic proteins. Similar to other *Thermus* proteins analyzed to date, *T. aquaticus* RecA has no cysteine residues. The number of proline residues, however, is significantly greater than that seen for *E. coli* (14 versus 10) (Table 1). The number of polar amino acids, e.g., serine and threonine, is lower than that found in *E. coli* RecA (26% polar amino acids in *T. aquaticus* versus 31% in *E. coli*), while 46% of the amino acids in *T. aquaticus* are hydrophobic or nonpolar compared with 43% in *E. coli*. Substitution of polar amino acids by hydrophobic amino acids would tend to yield a more compact structure. These types of substitutions reduce the overall hydrophilicity and protein flexibility and have been implicated in stabilizing proteins at higher temperatures (3).

Sequence analysis at the nucleotide level of the open reading frame reflects a higher G/C content than in *E. coli* and a G/C content similar to that reported for other *Thermus* genes (3, 7, 13, 18, 26). The high G/C composition at the nucleotide level may help to stabilize the DNA of thermophiles at high temperatures. The overall G/C content for *T. aquaticus* recA is 67% compared with 55% for *E. coli* recA. Additionally, the third letters in degenerate codons are highly G/C rich (13) and are reflected in the amino acid bias, i.e., the preference for amino acid codons that are G/C rich (Table 1). Deduced nucleotide sequence alignments of *T. aquaticus* recA with *E. coli* recA show 60% identity, while the amino acid identity is 59%, with 78% similarity (Fig. 2).

Expression of *T. aquaticus* RecA. The *T. aquaticus* recA gene, on the 1.4-kb SacI fragment, was subcloned into the prokaryotic expression vector pTrc99a. The gene was expressed under inducible IPTG control from the strong trc promoter (trp - 35 region, lacUV5 – 10 region). The pTrc99a vector includes these features: a ribosome-binding site, an ATG translation initiation codon in the multiple cloning region, strong transcription termination signals downstream of the multiple cloning region, and lacP. Sequence analysis of the cloning site suggests that translational initiation from this construct was from the *T. aquaticus* recA ATG, since the vector ATG site (NcoI site in multiple cloning region) was not returned for initiation because of a fortuitous deletion, presumably during the T4 DNA polymerase reaction, of the G nucleotide in the initiation codon. The resulting *T. aquaticus* RecA protein is IPTG inducible and migrates on SDS-PAGE at the molecular weight expected from the predicted amino acid sequence.

The *T. aquaticus* RecA protein purification scheme is described in detail in Materials and Methods. Typical samples of protein induction and purification are shown in Fig. 3. Of particular interest is the ability to partially purify thermostable proteins from other endogenous *E. coli* proteins by a brief 70°C incubation. This technique is widely used for purifying thermostable proteins (18, 26). Coomassie-stained gels were used to monitor the purification scheme (Fig. 3A). After heat
purification, protein was concentrated by precipitation with ammonium sulfate. Samples were further purified by elution from a DEAE-sephacel column. Pooled fractions were affinity purified on a single-stranded DNA agarose column. An approximately 40-fold purification was achieved in the final purification fraction with a yield of approximately 15%. The relative migration of the protein is in agreement with the expected molecular weight determination from the amino acid sequence. Immunodetection with antibodies against E. coli RecA suggests that these antibodies recognize and cross-react with the major protein present, T. aquaticus RecA (Fig. 3B). The lower levels of antibody binding to T. aquaticus RecA presumably reflect the partial amino acid identity (59%) between the two proteins.

**Filter binding assay.** The similarity between the E. coli RecA and the T. aquaticus RecA proteins, simply on the basis of amino acid comparisons, suggests that the T. aquaticus RecA protein may be functionally similar to E. coli RecA and thus may have similar biochemical activities at 65°C. First, the ability to bind single-stranded DNA by the T. aquaticus RecA protein was measured by retention on nitrocellulose membranes. Reactions were performed in the presence of the cofactors ADP and the nonhydrolyzable ATP analog, ATPγS, since, under these conditions, a stable protein–three-stranded DNA complex (synaptic complex) is formed by E. coli RecA (9). Figure 4A shows the titration of T. aquaticus RecA with 15 ng of an oligonucleotide DNA (30-mer) at 65°C. Maximum retention under these conditions occurs with 5 μg of T. aquaticus RecA. To determine the optimal temperature of T. aquaticus RecA binding to the 32P-labeled oligonucleotide, reaction mixtures were incubated at various temperatures from 25 to 95°C and then filtered under vacuum at room temperature (Fig. 4B). The filter binding assays indicate maximal amounts of complex formation at approximately 55°C. Unlike

**FIG. 3.** (A) T. aquaticus RecA protein purification. Samples were run on a 12% Tris-glycine-SDS gel and stained with Coomassie blue R-250. Protein molecular mass markers (in kilodaltons) are indicated to the left (21.5, soybean trypsin inhibitor; 30, carbonic anhydrase; 43, ovalbumin; 67, bovine serum albumin; 97.4, phosphorylase B). Lanes: 1, 50 ng of E. coli RecA; 2, total cell extract; 3, post-heat purification at 70°C; 4, ammonium sulfate precipitate; 5, fraction eluted from DEAE-sephacel; 6, single-stranded DNA agarose fraction. The arrow indicates the T. aquaticus RecA protein. (B) Western blot analysis of T. aquaticus RecA with polyclonal antibody raised against E. coli RecA. E. coli RecA and purified T. aquaticus RecA samples were analyzed on SDS-PAGE gels and subjected to Western blot analysis. Molecular mass markers are as described for panel A. Lanes: 1, 50 ng of E. coli RecA; 2, 250 ng of T. aquaticus RecA. An arrow indicates the location of the purified T. aquaticus RecA.

**FIG. 4.** Single-stranded DNA filter binding assays. (A) Measurement of retention of T. aquaticus RecA protein–single-stranded DNA complex formation on nitrocellulose filters at 65°C with increasing concentrations of T. aquaticus RecA from 0.1 to 15 μg. The binding reaction mixtures were incubated at 65°C for 5 min and were then applied to a vacuum filtering apparatus at room temperature (see Materials and Methods). The concentration of the single-stranded DNA oligonucleotide (30-mer) is constant at 15 ng. The x axis is T. aquaticus RecA measured in micrograms, while the y axis is the percent retention of complex to nitrocellulose membranes. (B) Measurement of protein–single-stranded DNA complex retention to nitrocellulose membranes at different temperatures (25, 37, 45, 55, 65, 70, 75, 85, and 95°C) for E. coli RecA (○) and T. aquaticus RecA (●). The amount of E. coli and T. aquaticus RecA was 1.5 μg. Standard error of the mean values are indicated for each point in both panels. Background binding to membranes was subtracted from each value. Since some experiments at high temperatures (>65°C) indicate low levels of association with single-stranded DNA by E. coli RecA, control experiments were performed in the absence and presence of ATPγS and ADP to determine whether the binding was real or artificial. These experiments indicated that the binding of the E. coli RecA to single-stranded DNA above 65°C is not dependent on nucleotide cofactors and that the 1 to 15% retention is probably due to the inherent variability of filter binding assays.

E. coli RecA, T. aquaticus RecA has a single-stranded DNA binding activity over a range of 45 to 65°C. Single-stranded DNA binding activity is reduced above 70°C, with less than 10% activity above 75°C (Fig. 4B). In contrast, single-stranded DNA binding assays with E. coli RecA (1.5 μg) show approximately 75% retention at 37°C and a reduction in retention above 45°C and at 25°C. The amount of RecA protein used in the assay was within the linear range for binding to single-stranded DNA by E. coli RecA (data not shown). Similar to E. coli RecA, single-stranded DNA binding to T. aquaticus RecA protein is an ATP-dependent process. Filter binding reactions performed in the absence of ADP and ATPγS gave retention of less than 1% for both RecA proteins when measured at 37°C for E. coli and 65°C for T. aquaticus (data not shown).

**Strand exchange.** Homologous DNA recombination promoted by E. coli RecA is a multistep process in which RecA binds to single-stranded DNA in an ATP-dependent fashion to form an active nucleoprotein filament prior to the exchange of homologous strands. Filter binding assays demonstrated that T. aquaticus RecA bound to single-stranded DNA, forming a presynaptic complex. Since the recombination reaction promoted by E. coli RecA ultimately involves the exchange of DNA strands, we investigated the ability of T. aquaticus RecA to pair homologous DNAs and go on through to the exchange of strands by assaying for the formation of heteroduplex joint molecules. The substrate DNAs in this reaction were linearized double-stranded M13mp18 and the homologous single-stranded circular M13mp18 DNA. In this assay, the strand
FIG. 5. Formation of joint molecules by T. aquaticus RecA at 65°C. The results of a strand exchange assay with T. aquaticus RecA containing 50 ng of M13mp18 viral DNA and 50 ng of M13mp18 linear duplex DNA are shown. Assay conditions are described in Materials and Methods. Lanes: 1, M13mp18 linear duplex alone; 2, M13mp18 viral DNA alone; 3, control reaction in the absence of T. aquaticus RecA; 4, 5, and 6, addition of 5, 7, and 10 μg of T. aquaticus RecA protein, respectively. Abbreviations: T. aq., T. aquaticus; ss, single stranded; ds, double stranded; JM, joint molecule.

exchange reaction occurs by pairing a linear duplex DNA with homologous regions of the single-stranded circular DNA. Branch migration then forms a heteroduplex DNA product or joint molecule (38). Strand exchange products are visualized on ethidium bromide-stained agarose gels. The substrate DNAs migrate differently from joint molecules, as indicated in Fig. 5. The joint molecules formed represent intermediates in the exchange process. If the reaction goes to completion, the product formed is nicked circular DNA (form II), which migrates just above the joint molecule intermediate in ethidium bromide-stained agarose gels (not seen in Fig. 5). This reaction also requires Mg²⁺ ions and is greatly enhanced by the E. coli single-stranded DNA binding protein (16).

We show evidence for strand exchange by T. aquaticus RecA under reaction conditions similar to those described for E. coli RecA (19). The results of the strand exchange assay with T. aquaticus RecA are shown in Fig. 5. The data show that T. aquaticus RecA formed joint molecules at 65°C. Densitometric scanning of the Polaroid negative indicated 20, 24, and 19% joint molecule formation for lanes 4, 5, and 6, respectively, in Fig. 5. The formation of joint molecules by T. aquaticus RecA at 65°C was not as efficient as that by E. coli RecA (at 37°C) under similar reaction conditions but represented a measurable strand exchange activity. E. coli RecA converts two to four times more duplex DNA to joint molecule in the same amount of time as T. aquaticus RecA. Furthermore, we have not been able to convert the joint molecules to form II by increasing the time of incubation or increasing the amount of protein. However, the differences in conversion may be due in part to our use of the mesophilic analog E. coli single-stranded DNA binding protein in our reaction mixtures, which acts to eliminate secondary structure barriers in single-stranded DNA (29).

In our hands, E. coli single-stranded DNA binding protein stimulates the conversion of duplex to joint molecule by E. coli RecA by a factor of at least 2. In the future, it may be possible to increase the conversion to joint molecule by T. aquaticus RecA in the presence of a thermostable single-stranded DNA binding protein. Similar strand exchange experiments performed in the absence of ATP did not demonstrate any joint molecule formation (data not shown). An experiment to determine the optimal temperature for strand exchange by E. coli RecA confirmed that optimal strand exchange occurs at 37°C, with an equivalent amount of activity at 25°C, near background levels at 45°C, and no activity at 55°C and above, as measured by densitometric scanning (data not shown). In a similar experiment with T. aquaticus RecA, we observed a broader range of strand exchange activity from 37 to 70°C, with peak activity at 65°C (data not shown). These data were in relative agreement with those from the single-stranded DNA filter binding studies described in the preceding section in which peak single-stranded DNA binding occurred near 55°C under these conditions (Fig. 4B).

DISCUSSION

A RecA analog from a thermophilic organism was cloned and expressed in E. coli. The T. aquaticus RecA protein is 341 amino acids in length, compared with 353 amino acids for E. coli RecA. The relative molecular weight, as estimated by denaturing SDS gel electrophoresis, of the T. aquaticus RecA protein agrees with the deduced amino acid sequence (Fig. 3). Also, as predicted by the sequence, the molecular weight of T. aquaticus RecA is lower than that of E. coli RecA. A polyclonal antibody raised against E. coli RecA cross-reacts with purified T. aquaticus RecA. This finding suggests that the thermophilic analog is significantly similar not only in primary sequence but also in conformation to the E. coli RecA. The purification scheme described is similar to that used for E. coli RecA purification by our laboratory but with some modifications. We have taken advantage of the thermostability of T. aquaticus RecA and utilized a heat treatment step early in the purification scheme to remove a large percentage of the endogenous E. coli proteins. Subsequent steps include anion exchange and a single-stranded DNA binding column.

Retention of T. aquaticus RecA and single-stranded DNA complexes on nitrocellulose filters suggests that the T. aquaticus RecA binds to single-stranded DNA. Our assay conditions are those optimized for E. coli RecA binding to single-stranded DNA for the formation of synaptic complexes (9) and may not represent the best conditions for the binding of T. aquaticus RecA to single-stranded DNA. However, we have demonstrated that T. aquaticus RecA can bind single-stranded DNA under these conditions over a wide range of temperatures, with optimal binding at a temperature significantly above that of E. coli RecA (Fig. 4) and that our protein is a true thermostable RecA and not a contaminating mesophilic activity. Strand exchange by T. aquaticus RecA occurs under experimental conditions optimized for E. coli RecA strand exchange. This suggests that T. aquaticus RecA may promote strand exchange in a manner similar to E. coli RecA. The reaction conditions for T. aquaticus RecA remain to be optimized.

It is interesting to speculate on the nature of the interaction of proteins with DNA at high temperatures. The temperatures at which these reactions occur in thermophiles would be sufficient to alter the secondary structure of DNA and to denature many proteins. Therefore, the interaction of proteins with DNA at these temperatures must be stabilized in such a way as to allow their association. On the basis of analysis of the amino acid composition of proteins derived from thermophiles, a more compact and rigid secondary and tertiary structure is predicted. The amino acid replacements or substitutions from mesophilic to the thermophilic analogs described tend to generate compact, hydrophobic structures. The elimination of thiol groups by either lowering the number of cysteine residues or by reducing their accessibility to solvent has been implicated in promoting thermal stability (24). It has been suggested that the configurational entropy differences in transfer from the unfolded to the folded state are due to rotational flexibility about the β-carbon (21). Proline residues restrict rotational flexibility of the backbone because of their pyrrolidine ring structure. Mutagenesis studies to engineer
thermal stable proteins have shown that Gly→Pro substitutions may restrict backbone conformation and thus alter the configurational entropy of protein unfolding (22). *T. aquaticus* RecA demonstrates some of the typical preferences for amino acids which are implicated in playing a role in stabilizing proteins at higher temperatures.

By substituting *T. aquaticus* RecA amino acids directly onto the induced RecA crystal structure with Sybyl molecular modeling software on an Evans and Sutherland workstation, we made several interesting observations. As described in Results, *T. aquaticus* RecA has no cysteine residues. These residues have been substituted with either Ala or Val, thereby increasing the overall hydrophobicity of the protein. In addition, the availability of the deduced *E. coli* RecA crystal structure (37) has allowed us to make the following speculations. First, most of the nonconservative amino acid replacements (polar, hydrophobic, or bulky groups) occur in the carboxy-terminal domain or in amino acids that are exposed to solvent, while those portions of the RecA protein implicated in nucleotide binding and putative single- and double-stranded DNA binding sites (36, 37) are highly conserved. Second, a step of amino acid replacements that can stabilize mesophilic proteins at high temperatures implicates substitutions which stabilize α-helices (22). These replacements decrease flexibility and increase hydrophobicity (e.g., Gly→Ala, Ser→Ala, Lys→Arg, and Lys→Ala) and often occur in external α-helical regions. Of these types of replacements, four occur in putative α-helical regions of *T. aquaticus* RecA, and two occur in exposed looped regions on the basis of the *E. coli* RecA crystal structure. Finally, the amino acid backbone may be significantly more rigid because of the increased number of proline residues. Substitution of *T. aquaticus* RecA proline residues onto the *E. coli* RecA crystal structure yields proline residues predominantly located in putative loop regions of the thermophilic RecA. Therefore, a higher energy input than that necessary for looped regions without proline residues might be required to alter peptide conformations. To our knowledge, there are only a few known mesophilic nucleic acid binding proteins with both cloned thermostable analogs and structures as determined by X-ray crystallography: i.e., the DNA polymerase I Klenow fragment of *E. coli* and its cloned thermostable analog *T. aquaticus* DNA polymerase (1) and the *E. coli* RNase HI with its thermostable analog from *T. thermophilus* HB8, RNase H (12, 14). Recently, the *T. thermophilus* HB8 RNase H protein was crystallized and analyzed to determine how the protein may be thermally stabilized (11). As predicted from amino acid analysis and studies on predicting thermally stabilizing amino acid substitutions (22), amino acid replacements which decrease flexibility and increase hydrophobicity in α-helical regions contribute significantly to thermal stability. However, analysis of the mesophilic RNase H and thermostable RNase H crystals indicates that simple amino acid comparisons cannot sufficiently explain differences in protein thermostability.

The interest in proteins derived from thermophiles or hyperthermophiles has increased in recent years because of their potential commercial applicability. Because of the nature of the *E. coli* RecA protein and its essential role in pairing homologous DNAs, a thermostable RecA protein could potentially be utilized to increase the specificity and efficiency of pairing DNAs in reactions that occur at higher temperatures, e.g., PCR. Any reaction that requires annealing of homologous primers to DNA to template DNA could be enhanced in the presence of a thermostable RecA protein. The evidence thus far generated suggests that the thermostable RecA is functionally similar to *E. coli* RecA. However, there still remain a vast number of questions to be answered concerning the nature of these reactions at higher temperatures. It is evident that proteins from these organisms have evolved to accommodate the extreme environments in which they flourish. We are interested in determining how homologous recombination has evolved to accommodate these conditions.

ACKNOWLEDGMENTS

We thank Peggy Hsieh, Moon G. Kim, Sushma Pati, and Janet Yancey-Wrona for helpful discussions and critical review of the manuscript.

REFERENCES

1. Beese, L. S., V. Derbyshire, and T. A. Steitz. 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA. Science 260:352–355.
2. Bishop, D., 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 69:439–450.
3. Bowen, D., J. A. Littlechild, J. E. Fothergill, and H. C. Watson. 1988. Nucleotide sequence of the phosphoglycerate kinase gene from the extreme thermophile *Thermus thermophilus*. Biochemistry 25:509–517.
4. Camerini-Otero, R. D., and P. Hsieh. 1993. Parallel DNA triplexes, homologous recombination, and other homology-dependent DNA interactions. Cell 72:217–223.
5. Compton, J. S., and C. G. Jones. 1985. Mechanism of dye response and interference in the Bradford protein assay. Anal. Biochem. 151:369–374.
6. Cox, M. M., and I. R. Lehman. 1987. Enzymes of general recombination. Annu. Rev. Biochem. 56:229–262.
7. Dekker, K., H. Yamagata, K. Sakaguchi, and S. Ueda. 1991. Xylose (glucose) isomerase gene from the thermophile *Thermus thermophilus*: cloning, sequencing, and comparison with other thermostable xylose isomerases. J. Bacteriol. 173:3078–3083.
8. Fujisawa, H., T. Yonesaki, and T. Minagawa. 1985. Sequence of the T4 recombination gene, uvsX, and its comparison with that of the recA gene of *Escherichia coli*. Nucleic Acids Res. 13:7473–7481.
9. Hsieh, P., C. S. Camerini-Otero, and R. D. Camerini-Otero. 1992. The synopsis event in the homologous pairing of DNA: RecA recognizes and pairs less than one helical repeat of DNA. Proc. Natl. Acad. Sci. USA 89:6492–6496.
10. Hsieh, P., and R. D. Camerini-Otero. 1989. Formation of joint DNA molecules by two eukaryotic strand exchange proteins does not require melting of a DNA duplex. J. Biol. Chem. 264:5089–5097.
11. Ishikawa, K., M. Okumura, K. Katayanagi, S. Kimura, S. Kanaya, H. Nakamura, and K. Morikawa. 1993. Crystal structure of ribonuclease H from *Thermus thermophilus* HB8 refined at 2.8 Å resolution. J. Mol. Biol. 230:529–542.
12. Itaya, M., and K. Kondo. 1991. Molecular cloning of a ribonuclease H (RNase HI) gene from an extreme thermophile *Thermus thermophilus* HB8: a thermostable RNase H can functionally replace the *Escherichia coli* enzyme in vivo. Nucleic Acids Res. 19:4443–4449.
13. Kagawa, Y., H. Nojima, N. Nukiwa, M. Ishizuka, T. Nakajima, T. Yasuhara, T. Tanaka, and T. Oshima. 1984. High guanine plus cytosine content in the third letter of codons of an extreme thermophile: DNA sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. J. Biol. Chem. 259:2956–2960.
14. Kanaya, S., and M. Itaya. 1992. Expression, purification, and characterization of a recombinant ribonuclease H from *Thermus thermophilus* HB8. J. Biol. Chem. 267:10184–10192.
15. Kodadek, T., M. L. Wong, and B. Alberts. 1988. The mechanism of homologous DNA strand exchange catalyzed by the bacteriophage T4 uvsX and gene 32 proteins. J. Biol. Chem. 263:9427–9436.
16. Kowalczykowski, S. C. 1991. Biochemistry of genetic recombina-
27. Ogawa, T., X. Yu, A. Shinozara, and E. H. Egelman. 1993. Similarity of the yeast RAD51 filament to the bacterial RecA filament. Science 259:1896–1898.

28. Radding, C. M. 1982. Homologous pairing and strand exchange in genetic recombination. Annu. Rev. Genet. 16:405–437.

29. Radding, C. M. 1988. Homologous pairing and strand exchange promoted by *Escherichia coli* RecA protein. p. 193–229. In R. Kucherlapati and G. R. Smith (ed.), Genetic recombination. American Society for Microbiology, Washington, D.C.

30. Roca, A. L. and M. M. Cox. 1990. The RecA protein: structure and function. Crit. Rev. Biochem. Mol. Biol. 25:415–456.

31. Rosenberg, M., and D. Court. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. 13:319–353.

32. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463–5467.

33. Shinozara, 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.

34. Smith, G. R. 1988. Homologous recombination in procaryotes. Microbiol. Rev. 52:1–28.

35. 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.

36. Story, R. M., and T. A. Steitz. 1992. Structure of the recA protein-ADP complex: implications for ATPase coupling to DNA binding. Nature (London) 358:374–376.

37. Story, R. M., I. T. Weber, and T. A. Steitz. 1992. The structure of the *E. coli* recA protein monomer and polymer. Nature (London) 355:318–325.

38. West, S. C. 1992. Enzymes and molecular mechanisms of genetic recombination. Annu. Rev. Biochem. 61:503–640.

39. West, S. C., J. K. Countryman, and P. Howard-Flanders. 1983. Purification and properties of the recA protein of *Proteus mirabilis*. J. Biol. Chem. 258:4648–4654.