INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Cloning, Sequencing and Biochemical Characterization of RecA and RuvB from Divergent Thermophilic Bacteria

by

Jie Tong

A dissertation submitted to the Graduate Faculty in Biomedical Sciences in partial fulfillment of the requirement for the degree of Doctor of Philosophy, The City University of New York

1997

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

5/19/97
Date

Lu-Hai Wang, Ph.D.
Chair of Examining Committee

5/21/97
Date

Terry A. Krulwich, Ph.D.
Executive Officer

James G. Wetmur, Ph.D.

Jerome L. Schulman, M.D.

John A. Blaho, Ph.D.

William K. Holloman, Ph.D.

Examining Committee

The Mount Sinai School of Medicine of the City University of New York

ABSTRACT

Cloning, Sequencing and Biochemical Characterization of RecA and RuvB from Divergent Thermophilic Bacteria

by

Jie Tong

Adviser: Professor James G. Wetmur

The *recA* gene of *Thermus thermophilus* was cloned, sequenced and expressed in *Escherichia coli*. Three other thermostable RecA proteins were cloned by other members of our laboratory group from the highly-divergent thermophilic eubacteria *Thermus aquaticus, Thermotoga maritima* and *Aquifex pyrophilus*. In contrast to the *E. coli* RecA protein, all the purified thermophilic RecA proteins exhibited single-stranded DNA-dependent ATPase activity optima above 70°C. In spite of substantial sequence divergence, interesting characteristics of the thermostable RecA proteins included increased valine content, common amino acid replacements at two highly conserved sites, and an increase in the calculated isoelectric point of approximately a full pH unit.

The *ruvB* genes of the highly divergent thermophilic eubacteria *T. thermophilus* and *T. maritima* were cloned, sequenced and expressed in *Escherichia coli*. Both thermostable RuvB proteins were purified to homogeneity. Like *E. coli* RuvB protein, both purified thermostable RuvB proteins showed strong double-stranded DNA-dependent ATPase activity at their temperature optima (≥70°C). In the absence of ATP, *T. thermophilus* RuvB protein bound to linear double-stranded DNA with a preference for the ends. Addition of ATP or ATPγS

destabilized the *T. thermophilus* RuvB-DNA complexes. Both thermostable RuvB proteins displayed helicase activity on supercoiled DNA. Expression of thermostable *T. thermophilus* RuvB protein in the *E. coli ruvBrecG* mutant strain N3395 partially complemented the UV sensitive phenotype suggesting that *T. thermophilus* RuvB protein has a function similar to *E. coli* RuvB *in vivo*.

To solve the crystal structure, *T. maritima* and *T. thermophilus* RuvB proteins were expressed and purified from the *E. coli* strain BL21(DE3) pLysS. Crystallization of the *T. maritima* and *T. thermophilus* RuvB proteins was achieved at Scripps Research Institute by both the hanging-drop and sitting-drop vapor diffusion techniques. *T. maritima* and *T. thermophilus* RuvB crystals gave diffraction data to at least 2.2 Å and 3.6 Å, respectively. Two *T. maritima* RuvB cysteine mutants, A308C and N32C, were generated for derivatization by mercurials. Both mutant proteins were expressed, purified and crystallized. Selenomethionine-substituted RuvB protein was generated for multiwavelength anomalous diffraction phasing. Based on mass spectra data, the level of selenomethionine substitution was almost complete.

FORMAT OF THESIS

This thesis was prepared according to the guidelines of the City University of New York. The first chapter contains a general introduction. The second chapter contains my data from Wetmur et al., *Journal of Biological Chemistry* (1994). The third chapter is Tong & Wetmur, *Journal of Bacteriology* (1996). The fourth chapter contains unpublished data resulting from a collaboration with Christopher D. Putnam and John A. Tainer in The Scripps Research Institute. The final chapter contains an overall discussion.

ACKNOWLEDGEMENTS

I wish to express my gratitude to my mentor Dr. James G. Wetmur for his encouragement, support and guidance through out my thesis research. Without him this thesis would not have been possible.

I would like to thank Dr. John A. Blaho for his helpful advice on topological assay and two-dimensional gel electrophoresis.

I would like to express my gratitude to my advisory committee members, Dr. Lu-Hai Wang and Dr. Jerome L. Schulman for their advice throughout my graduate study.

I want to thank Ms. Daphne Szeto, Mr. H. G. V. Rao and Ms. Amy Rosenfeld for their help and friendship.

I would also like to thank our collaborator Dr. John A. Tainer and Mr. Christopher D. Putnam in The Scripps Research Institute.

TABLE OF CONTENTS

	Page
Approval Page	ii
Abstract	iii
Thesis Format	v
Acknowledgements	vi
Table of Contents	vii
List of Publications	xiii
List of Tables	xiv
List of Figures	xv
I. Introduction	1
A. RecA protein	1
Presynapsis	3
Synapsis	4
DNA heteroduplex formation	4
B. Applications of RecA protein	5
C. Eukaryotic homologs of RecA protein	6
Ustilago maydis REC2	6
Saccharomyces cerevisiae RAD51	7
Saccharomyces cerevisiae DMC1	7
Arabidopsis thaliana RecA homolog	8
D. Processing of recombination intermediates	8
Ruv locus	8

RuvA and RuvB 9

RecG 12

RuvC 13

Rus 14

E. Research goals 15

F. Helicase function 16

G. Crystal structure of helicases 18

II. Cloning, Sequencing and Expression of RecA from *Thermus thermophilus* 20

A. Introduction 21

B. Experimental procedures 23

Genomic DNA, plasmids, nucleotides and enzymes 23

Oligodeoxynucleotides 23

DNA amplification 24

Cloning, sequencing and Southern hybridization 25

Expression 25

Purification 26

ATPase assay 27

Computer analysis 27

C. Results and discussion 28

Cloning *Thermus thermophilus recA* 28

Expression of *T. Thermophilus* RecA in *E. coli* 31

Purification of recombinant *T. Thermophilus* RecA protein 33

Phylogenetic tree based on RecA protein sequences 36

Common characteristics of thermostable RecA proteins 42

III. Cloning, Sequencing and Expression of *ruvB* and Characterization of RuvB Protein from Two Distantly Related Thermophilic Eubacteria 45

A. Introduction 46

B. Materials and methods 48

Bacterial strains and plasmids 48

Genomic DNA, nucleotides and enzymes 48

Oligodeoxynucleotides 49

DNA amplification, cloning and sequence analysis 50

Expression 50

purification 51

RuvB ATPase assay 51

RuvB protein-DNA complex formation 51

Topological assay for DNA unwinding by *T. thermophilus* and *T. maritima* RuvB 52

RuvB complementation assay 52

Genbank 53

C. Results 53

Cloning and sequencing of *T. Thermophilus* and *T. maritima* *ruvB* genes 53

Expression and purification of *T. thermophilus* and 54

T. maritima RuvB protein

ATPase activity of thermostable *T. thermophilus* and *T. maritima* RuvB 57

Binding of *T. thermophilus* and *T. maritima* RuvB to linear dsDNA 58

Unwinding supercoiled DNA by thermostable *T. thermophilus* and *T. maritima* RuvB 60

T. thermophilus RuvB can complement *E. coli* RuvB *in vivo* 63

D. Discussion 64

E. Acknowledgments 67

IV. Crystal Structure Determination of Thermostable RuvB Proteins 69

A. Introduction 70

B. Materials and methods 71

Bacterial strains and plasmids 71

Site-directed mutagenesis of pET11c-TmaRuvB 71

Sequence analysis 72

Expression and purification of wild-type *T. thermophilus* and *T. maritima* RuvB and cysteine-containing *T. maritima* RuvB mutants 72

RuvB ATPase assay 72

Crystallization of *T. thermophilus* and *T. maritima* RuvB protein 73

Expression and purification of selenomethionine-substituted 73

T. maritima RuvB protein

Mass spectrometry 73

C. Preliminary results and discussion 74

Overexpression and purification of *T. thermophilus* and *T. maritima* RuvB proteins 74

Crystallization and diffraction studies of the *T. maritima* RuvB 74

Crystallization and diffraction studies of the *T. thermophilus* RuvB 74

Designing cysteine mutants for phasing *T. maritima* RuvB diffraction data 77

Generating N32C and A308C *T. maritima* RuvB mutants 77

N32C and A308C *T. maritima* RuvB mutant proteins are thermostable 77

N32C and A308C *T. maritima* RuvB mutant sequences 79

Expression and purification of N32C and A308C *T. maritima* RuvB mutant proteins 79

Generating selenomethionine-substituted RuvB 79

V. Discussion 85

A. Cloning and analysis of genes encoding thermostable recombination proteins 85

B. Analysis of thermostable RecA and RuvB sequences 89

C. Attempts to clone thermostable RuvA proteins 94

D. Structure studies of thermostable RuvB helicases 96

VI. Bibliography 98

PUBLICATIONS

Wetmur, J.G., Wang, D.M., Ortiz, B., Tong, J., Reichert, F. & Gelfand D.H. (1994) Cloning, Sequencing, and Expression of RecA Proteins from Three Distantly Related Thermophilic Eubacteria. *J. Biol. Chem.* **269**:25928-25935

Chen, J.M., Tong, J., Sukegawa, I.T. & Wang, L-H. (1995) Cloning and Functional Characterization of the Chicken *c-ros* Promoter. *Cell Growth & Diff.* **6**: 1523-1530

Tong, J. & Wetmur, J.G. (1996) Cloning, Sequencing, and Expression of *ruvB* and Characterization of RuvB Proteins from Two Distantly Related Thermophilic Eubacteria, *J. Bacteriol.* **178**: 2695-2700

1. Calculated isoelectric pc
2. Calculated isoelectric pc

LIST OF TABLES

Page

1. Calculated isoelectric points of RecA proteins 44

2. Calculated isoelectric points of RuvB proteins 95

LIST OF FIGURES

		Page
1.	Roles of recombination proteins in *E. coli* recombination process	2
2.	Tripartite structure of RuvAB complex	11
3.	Degenerate primers for cloning *Thermus thermophilus* RecA	29
4.	Analysis of PCR product with degenerate primers and inverse PCR primers	30
5.	Southern Blot hybridized with inverse PCR primer	32
6.	SDS-PAGE depicting purification of *T. thermophilus* RecA protein	34
7.	*T. thermophilus* RecA coding sequence	35
8.	Similarity of *E. coli* and *T. thermophilus* RecA Sequences	37
9.	Temperature dependence of the ATPase activity of RecA proteins	38
10.	Thermophilic RecA protein sequences	39
11.	Unrooted evolutionary tree for RecA proteins	41
12.	Unrooted evolutionary tree for RuvB proteins	55
13.	SDS-PAGE analysis of *T. thermophilus* and *T. maritima* RuvB proteins at each step of purification	56
14.	Temperature dependence of ATPase activity of *T. thermophilus* RuvB and *T. maritima* RuvB proteins	58
15.	Binding of *T. thermophilus* RuvB to linear duplex DNA	59
16.	Competitive binding of two different sizes of linear dsDNA with *T. thermophilus* RuvB protein	61
17.	Detection of topoisomers of supercoiled pUC19 DNA following *T.*	62

thermophilus and *T. maritima* RuvB mediated unwinding and topoisomerase I treatment on one- and two-dimensional agarose gels

18. UV survival of *E. coli recG ruvB* mutant strain N3395 containing pUC19, pTth expressing *T. thermophilus* RuvB protein, pEc expressing *E. coli* RuvB protein and wild type strain XAO 65
19. *T. maritima* and *T. thermophilus* RuvB crystals 75
20. Diffraction data from the *T. maritima* and *T. thermophilus* RuvB crystals 76
21. Screening for N32C and A308C *T. maritima* RuvB mutants 78
22. SDS-PAGE analysis of thermostable N32C and A308C *T. maritima* RuvB expression 80
23. SDS-PAGE depicting purification of A308C and N32C *T. maritima* RuvB mutant proteins 81
24. SDS-PAGE depicting purification of selenomethionine-substituted *T. maritima* RuvB protein 82
25. Mass Spectra data of wild-type and selenomethionine-substituted *T. maritima* RuvB proteins 84
26. *T. thermophilus ruvB* sequence 86
27. *T. maritima ruvB* sequence 87
28. Pileup of RuvB sequences 90

I. Introduction

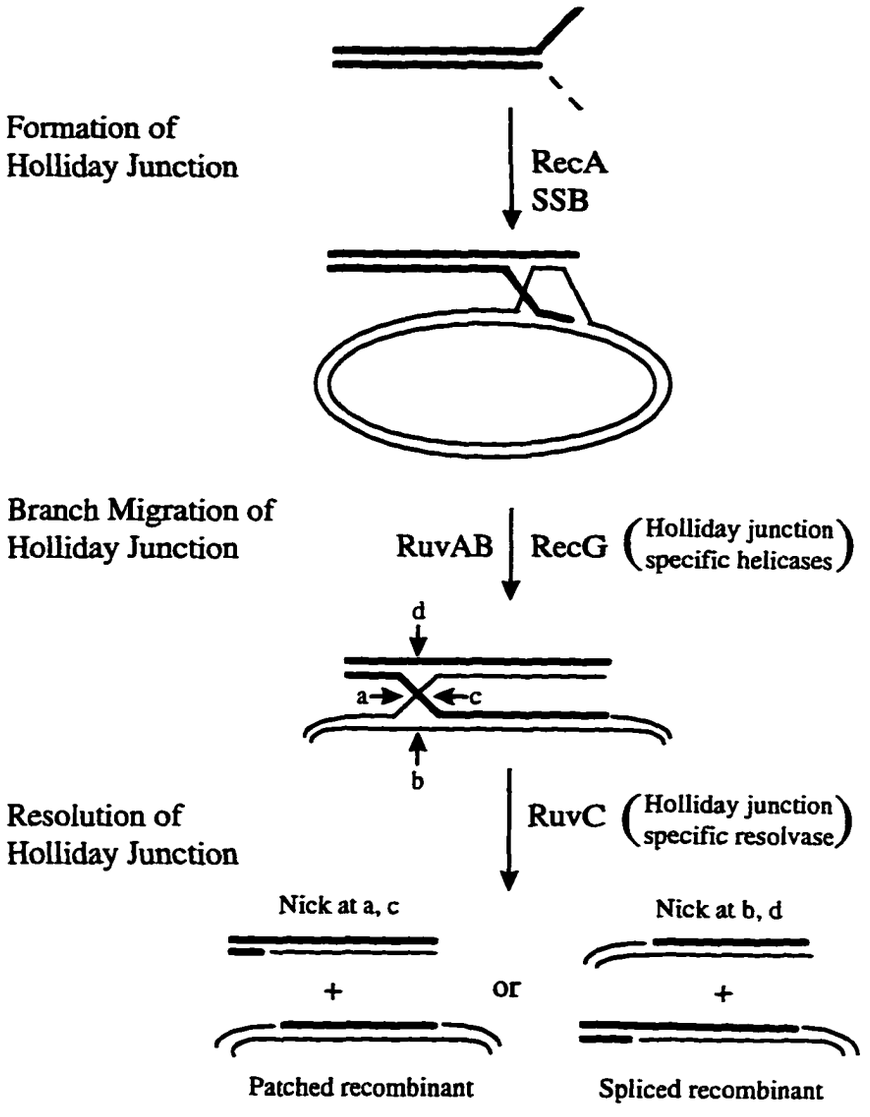
Genetic recombination plays an essential role in both the maintenance of genetic information and the generation of biological diversity. Genetic recombination involves the concerted action of binding proteins, helicases, ligases and many other classes of enzymes. Much of the current understanding of the mechanism of this complicated process comes from studies of homologous recombination in *Escherichia coli.*

In 1965, Clark and Margulies reported the identification of the first gene required for homologous genetic recombination in *E. coli.* Mutations in this gene (*recA*) created a recombination-negative phenotype (*rec*$^-$) in cells acting as recipients in a conjugation assay. To date, more than 16 other recombination genes have been identified in *E. coli.* Genetic recombination in *E. coli* may utilize products of the *recA, recB, recC, recD, recE, recF, recG, recJ, recN, recQ, recR, ruvA, ruvB* and *ruvC* genes as well as enzymes involved in general DNA metabolism such as DNA polymerases, topoisomerases and ligases (reviewed by West, 1994; Shinagawa and Iwasaki, 1995). The functions of some of these proteins in recombination are illustrated in Figure 1.

A. RecA protein

The *E. coli* RecA protein (M_r 37,842) is a multifunctional protein which plays important roles in genetic recombination, DNA repair and UV-induced

Figure 1. Roles of recombination proteins in *E. coli* recombination process



mutagenesis. RecA protein is the central mediator of the SOS response to DNA damage, and, when complexed with single-stranded DNA, catalyzes the autocleavage of LexA repressor (Little, 1984), UmuD protein (Nohmi, et al., 1988; Shinagawa, et al., 1988a), and λ repressor (Craig and Roberts, 1980; Little, 1984). Cleavage of LexA derepresses transcription from several promoters leading to the synthesis of more than 20 SOS repair proteins, including LexA, RecA, RecN, RecQ, RuvA, RuvB, UvrA, UvrB, UvrD, UmuC, and UmuD (reviewed by West, 1992).

RecA plays a central role in recombination both as a structural protein and as a catalyst. Much of our current understanding of the mechanism of homologous pairing is derived from *in vitro* studies of *E. coli* RecA protein. Reactions of circular viral single strands with homologous linear duplex DNA have provided a convenient model for elucidating the three distinct phases by which RecA protein can carry out recombination: presynapsis, synapsis, and heteroduplex DNA formation (Kowalczykowski and Eggleston, 1994)

Presynapsis. RecA protein first polymerizes on single stranded DNA (ssDNA) in the 5′ → 3′ direction (Register and Griffith,1985; Lindsley and Cox, 1989) to form a right-handed helical structure known as the presynaptic complex. One unusual characteristic of this complex is that the DNA is extended ~ 50% relative to B-form duplex DNA, with increased axial spacing between base pairs of 5.1 Å versus 3.4 Å in B form and with unwinding to 18.6 bp per turn versus 10.4 bp per turn in B form. Each RecA monomer binds 3 nucleotides (Henser and Griffith, 1989; Dicapua, et al., 1990; Kowalczykowski, et al., 1994).

Synapsis. The presynaptic filament is a prerequisite for searching for DNA sequence homology on duplex DNA. Although the exact mechanism of the homology search remains unclear, it is believed that the first step involves the formation of contacts of limited homology, perhaps independent of strand orientation. These contacts result in the formation of networks within which the DNA concentration is high. The search continues and random collisions ultimately lead to productive homologous alignment (Kowalczykowski and Eggleston, 1994).

DNA heteroduplex formation. The region of DNA heteroduplex formed in the synaptic phase can enlarge progressively creating heteroduplex DNA within the filament while displacing a single strand. DNA heteroduplex formation is unidirectional with 5′ → 3′ polarity relative to the displaced ssDNA (Kahn, et al., 1981; West, et al., 1981; West, et al., 1982). Under typical reaction conditions, RecA protein-promoted DNA heteroduplex extension occurs at a rate of 2-10 bp/s (Kahn, et al., 1981; West, et al., 1981). Although extension is slow, the reaction is processive and can traverse kilobase pairs of DNA. Moreover, heteroduplex formation can traverse not only limited mismatches and frameshifts but even heterologous insertions as long as 50-100 base pairs (Bianchi and Radding, 1983). The formation of such mismatched heteroduplexes requires ATP hydrolysis.

Thus, the RecA protein plays a central role in the recombination process by catalyzing homologous pairing and strand exchange, leading to the formation of recombination intermediates in which DNA molecules are linked by Holliday junctions.

Since the sequencing of *E. coli recA* (Sancar, et al., 1980), genes encoding proteins with extensive sequence identify with *E. coli* RecA have been cloned and sequenced from many other bacterial species.

A few years ago, Story *et al.* (Story, et al., 1992; Story and Steitz, 1992) solved the crystal structure of the *E. coli* RecA protein at 2.3 Å resolution. The structure revealed a RecA polymer. This helical polymer closely resembled the low-resolution structure of RecA/DNA filaments seen in the electron microscope, and contained an extensive interface between subunits. The amino acid residues that are highly conserved among RecA-like proteins from various bacterial species were located toward the center of the helical filament.

B. Applications of RecA protein

E. coli RecA-ssDNA complexes have been employed as reagents. In one application, RecA-coated biotinylated probes were used for plasmid library screening with 10-20% recovery and enrichment of 10^4-10^5-fold (Rigas, et al., 1986). In a second application, RecA-oligodeoxynucleotide complexes were used to protect a specific site on duplex DNA from methylation, making this site uniquely susceptible to subsequent cleavage by a restriction endonuclease (Ferrin and Camerini-Otero, 1991; Koob, et al., 1992). The protected site may be less than one helical turn (Hsieh, et al., 1992). In a third application, RecA-oligodeoxynucleotide complexes were used to block transcription initiation and elongation *in vitro*, suggesting a potential for antisense application *in vivo* (Golub,

et al., 1992). The demonstration that a RecA fusion protein with a nuclear localization signal moved to the nucleus and bound chromosomal DNA suggests a potential for both antisense or gene therapy applications (Kido, et al., 1992). The availability of thermophilic RecA proteins may increase the effectiveness of current *in vitro* applications and may open new possibilities both *in vitro* and *in vivo* (Kato and Kuramitsu, 1993; Angov and Camerini-Otero, 1994; Wetmur, et al., 1994).

C. Eukaryotic homologs of RecA protein

***Ustilago maydis* REC2.** In 1982 Kmiec and Holloman purified the first eukaryotic RecA-like homologous pairing protein from mitotic cells of *Ustilago maydis*. This protein was originally named REC1 because its activity was not detectable in the *rec1* mutant of *Ustilago*. It was later found that the *REC1* gene of *Ustilago maydis* encoded a different protein, a 3′- 5′ exonuclease without pairing activity (Thelen, et al., 1994), and REC1 was renamed REC2. The 70 kD protein was shown to have single stranded DNA-dependent ATPase activity, to promote renaturation of complementary single stranded DNA, and to catalyze homologous pairing and strand exchange in a model assay in which single stranded circular DNA and homologous linear duplex DNA were used as substrates (Kmiec and Holloman, 1982).

The 70 kD pairing protein of *Ustilago maydis* was found to be a truncated form of a previously identified *REC2* gene product (Kmiec, et al., 1994). Like the

70 kD pairing protein, the purified 84 kD recombinant REC2 protein demonstrated DNA-dependent ATPase activity, catalyzed DNA renaturation, and promoted homologous pairing activity. Amino acid sequence comparison revealed that a 47-amino acid fragment surrounding the nucleotide triphosphate binding Walker A motif in REC2 had 39% identity with *E. coli* RecA, 42% identity with DMC1, and 40% identity with RAD51 (Rubin, et al., 1994).

***Saccharomyces cerevisiae* RAD51.** The primary sequence of the *Saccharomyces cerevisiae RAD51* gene showed that the 221 amino acid central portion of the encoded protein was 30% identical to that of the *E. coli* RecA. Yeast RAD51 has been implicated in mitotic recombination, double-strand break repair and in meiosis (Shinohara, et al., 1992). Yeast RAD51 protein has also been shown to form RecA-like filaments on DNA and to promote strand exchange *in vitro* (Benson, et al., 1992; Ogawa, et al., 1993; Sung, 1994), indicating the functional similarity between RAD51 and *E. coli* RecA. Since the discovery of yeast RAD51, mouse and human RAD51 homologs have been identified. Both proteins are 83% identical to yeast RAD51 (Shinohara, 1993).

***Saccharomyces cerevisiae* DMC1.** Another protein from *Saccharomyces cerevisiae*, DMC1, shares 26% identity with the central portion of *E. coli* RecA protein. Interestingly, DMC1 is highly identical to yeast RAD51, with 45% identity at the protein core region. Unlike RAD51, DMC1 is meiosis specific. It has been shown to be involved in recombination, synaptonemal complex formation and cell cycle progression (Bishop, et al., 1992). The mouse and human homologs of yeast DMC1 have also been identified and have been shown to have 54%

sequence identity with the yeast protein (Habu, et al., 1996).

***Arabidopsis thaliana* RecA homolog.** By using a fragment of a cyanobacterial *recA* gene derived from *Synechococcus* as a probe to screen a cDNA library of *Arabidopsis*, a *recA* gene homolog was identified and cloned. This gene was 60% identical to *Synechococcus recA* in the central core region. Immunological assay with polyclonal antibody to *E. coli* RecA revealed that the RecA homolog was located in the chloroplast. Hybridization assays demonstrated that the gene was located in the nucleus (Cerutti, et al., 1992).

The finding of RecA homologs in eukaryotic cells suggests that the mechanisms of homologous pairing and strand exchange may have been conserved.

D. Processing of recombination intermediates

***Ruv* locus.** Cells carrying *ruv* mutations were originally detected by their sensitivity to mitomycin C (Otsuji, et al., 1974) and by their reduced ability to promote recombination between duplicated *gal* genes (Stacey and Lloyd, 1976). They were also sensitive to UV light and ionizing irradiation, indicating a role in the recombinational repair of DNA damage. Further studies showed that the *ruv* locus on the *E. coli* chromosome encoded three genes, designated *ruvA*, *ruvB*, and *ruvC* (Sharples, et al., 1990). The *ruvA* and *ruvB* genes form an operon that is regulated by lexA (Benson, et al., 1991; Shinagawa, et al., 1988b), and is part of the SOS-inducible DNA repair response (Walker, 1985). The *ruvC* gene forms a

separate operon with *orf-26*, which could potentially encode a 26 kD protein of unknown function (Sharples, et al., 1990; Sharples and Lloyd, 1991; Takahagi, et al., 1991).

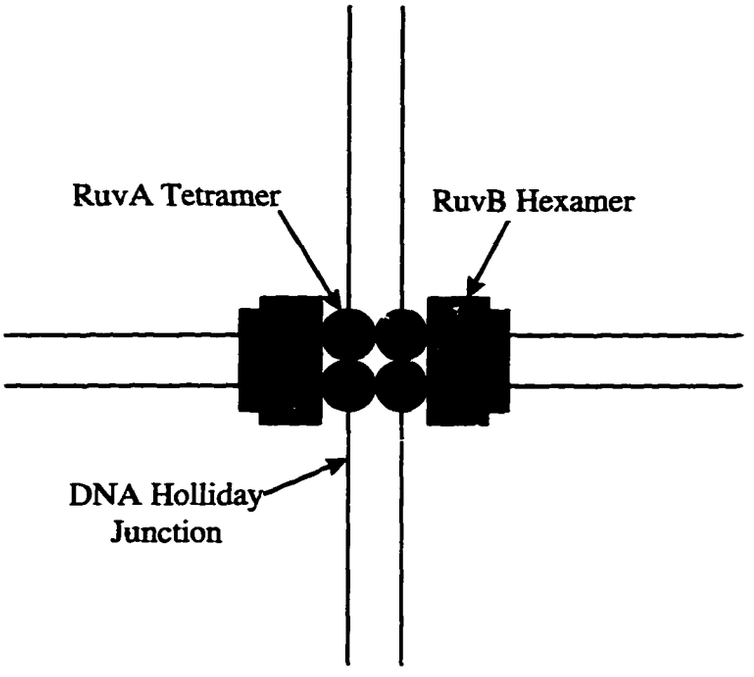
RuvA and RuvB. In order to generate heteroduplex DNA following formation of a Holliday junction, the junction needs to move along the DNA. This process is called branch migration. Studies have showed that spontaneous branch migration is slow, especially in the presence of magnesium, and can be severely inhibited by non-homologous sequence (Panyutin and Hsieh, 1993), indicating that this process requires catalysis. Recently, the RuvA and RuvB proteins have been found to promote ATP-dependent branch migration.

RuvA encodes a 22 kD DNA-binding protein, while *ruvB* encodes a 37 kD protein with ATPase activity (Iwasaki, et al., 1989; Shiba, et al., 1991). RuvA was shown to stimulate the ATPase activity of RuvB in the presence of DNA, indicating the potential for these two proteins to interact with one another (Shiba, et al., 1991). Subsequently, RuvA and RuvB were shown to act together to mediate the branch migration of recombination intermediates (Holliday junctions) made by RecA. At high RuvB concentrations, RuvB promoted branch migration in the absence of RuvA, indicating that the RuvB ATPase provided the driving force for branch migration. RuvA protein acts as a specificity factor by binding to a Holliday junction and directing RuvB to the junction. Interestingly, the rate of branch migration catalyzed by RuvA and RuvB is significantly faster than that catalyzed by RecA protein. Moreover, RuvAB-mediated branch migration can effectively overcome UV-induced DNA damage that severely inhibits RecA-mediated strand

exchange (Tsaneva, et al., 1992). RuvA and RuvB can even facilitate the bypass of a heterologous DNA insertion of at least 1 Kb during RecA protein-mediated DNA strand exchange (Iype, et al., 1994) and promote branch migration through at least 1.8 Kb of heterologous DNA in a RecA-free X-structure (Parsons, et al., 1995).

Sequence analysis has identified RuvB as a member of a superfamily of helicases. Several human diseases involving DNA recombination and/or repair, such as Xeroderma pigmentosum, Cockayne's syndrome, Bloom's syndrome and Werner's syndrome, may be caused by defects in DNA helicases (West, 1996). An electron microscopic study revealed that RuvB proteins formed double hexameric rings around DNA. These hexameric rings had an outer diameter of ~120 Å with a 20-25 Å diameter hole through the center of the rings (Stasiak, et al., 1994). Electron microscopy also revealed the RuvAB tripartite complex in which RuvA bound a Holliday junction and was sandwiched between two diametrically opposed hexameric rings of RuvB (Figure 2). The Holliday junction within this complex adopts a square-planar structure. Branch migration reactions are 1000 fold faster for this square-planar form of the Holliday junction seen in the absence of magnesium than the folded junction in the presence of magnesium (Panyutin and Hsieh, 1994). The role of RuvA is not restricted to RuvB targeting. It is required continually during branch migration (Michell and West, 1996). The movement of DNA through the RuvAB complex is proposed to be driven by the exertion of equal and opposite forces generated by the two rings of RuvB protein (Parsons, et al., 1995). Using a topological assay that measures the unwinding

Figure 2. Tripartite structure of RuvAB complex



of covalently closed duplex DNA, RuvA and RuvB were found to promote the transient unwinding of relaxed or supercoiled DNA, suggesting that branch migration and strand exchange might be coupled to transient DNA unwinding when the two DNA molecules pass through the RuvAB complex (Adam and West, 1995).

The crystal structure of RuvA has been determined at a resolution of 1.9 Å. RuvA crystals contain a monomer in the asymmetric unit. Each of the monomers of the RuvA tetramer are related by a fourfold symmetry in a manner similar to a four-petaled flower. A model of RuvA with a synthetic Holliday junction complex was constructed in which the four DNA duplex arms were in a square planar configuration, and each arm was docked into a groove on the concave surface of a monomer unit with the connecting single strands surrounding a central pin (Rafferty, et al., 1996). This model is consistent with the electron microscopy observations.

RecG. Mutants in *ruvA*, *ruvB* or *ruvC* genes all show defects in DNA repair and recombination (Lloyd, et al., 1984). However, unlike the repair defect, the recombination-defective phenotype is weak. The finding that *ruv recG* double mutants exhibit a strong recombination defective phenotype (Lloyd, 1991) suggests a functional overlap between a Ruv protein function and RecG.

Further studies showed that the 76 kD RecG protein of *E. coli* was a junction-specific DNA helicase similar as RuvAB. It catalyzed branch migration of Holliday junctions made by RecA in both directions (Whitby, et al., 1993). It could also catalyze branch migration of three-stranded recombination intermediates

made by RecA with a preference for the direction of RecA catalyzed strand exchange (Whitby and Lloyd, 1995). The overlap observed between RuvAB and RecG functions suggests at least two alternative pathways for branch migration.

However, *in vitro* evidence differentiates the RecG function from the RuvAB function. RecG may inhibit strand exchange mediated by RecA by promoting branch migration in the reverse direction, while RuvAB promotes strand exchange mediated by RecA by catalyzing branch migration in both directions (Whitby, et al., 1993; Al-Deib, et al., 1996). Although RecG does not have a resolvase activity, it has been proposed to provide an alternative way of resolving a Holliday junction by preferentially promoting branch migration back toward the starting substrates or ends of molecules (Whitby, et al., 1993). However, it is still possible that RecG functions together with some resolvase other than RuvC.

RuvC. Following branch migration, genetic recombination requires the resolution of the Holliday junction. This process is achieved by RuvC resolvase. In 1990, Connolly and West reported a resolution activity in cell-free extracts of *E. coli*. Subsequent experiments showed that this activity was absent from extracts made from the recombination, repair defective mutant *ruvC* (Connolly, et al., 1991), providing the key to the identification of the RuvC resolvase.

RuvC is a 19 kD protein. The role of the RuvC protein in the resolution of Holliday junctions was established by *in vitro* experiments that demonstrated the specific cleavage of, first, recombination intermediates made by RecA protein (Connolly, et al., 1991; Dunderdale, et al., 1991), second, synthetic Holliday junctions produced by annealing four oligonucleotides (Dunderdale, et al., 1991);

and third, palindromic sequences that extrude cruciform structures from supercoiled plasmid DNA (Iwasaki, et al., 1991).

By using small synthetic junctions, S. C. West's group showed that recognition of a junction by RuvC is structure-specific and occurs in the absence of metal cofactors. The cleavage of the junction, which is dependent upon Mg^{2+}, occurs by the introduction of symmetrically related nicks in strands of like polarity. RuvC demonstrates sequence specificity during cleavage. The nicks introduced by RuvC occur exclusively at the 3′ side of thymine residues (Bennett, et al., 1993).

The crystal structure of the *E. coli* RuvC was determined at a resolution of 2.5 Å. The RuvC crystal contained two dimeric molecules of RuvC in an asymmetric unit. RuvC formed a dimer related by a dyad axis of symmetry. This crystallographic study, coupled with extensive mutational analyses, revealed that the catalytic center of RuvC contained four acidic residues, all localized at the bottom of the putative DNA-binding cleft within the monomeric RuvC structure (Ariyoshi, et al., 1994). It is likely that RuvC dimer acts directly on the opposing arms of an intact RuvA-Holliday junction complex.

Rus. Recently, the suppressor mutation *rus-1* was found to restore UV resistance to *ruv* strains (Mandal, et al., 1993). DNA cloning and sequencing revealed that *rus-1* is an insertion of a 1.3 Kb IS2 element upstream of an open reading frame encoding a 14 kD polypeptide. The insertion probably enhanced its expression. This 14 kD peptide was named Rus. Rus is a resolvase that cleaves Holliday junctions in a manner remarkably similar to RuvC (Sharples, et

al., 1994). Multicopy plasmids expressing Rus can suppress *ruvA*, *ruvB* and *ruvC* mutants. Suppression depends on RecG (Manhi, et al., 1996). This evidence suggested that Rus protein might have been the resolvase that functioned together with RecG. However, it was later found that the *rusA* gene was located within a defective lambdoid prophage, DLP12 (Manhi, et al., 1996). In normal *E. coli* cells, the expression level of Rus is very low because of the poorly functional or tightly regulated *rusA* promoter (Manhi, et al., 1996). The exact role for this protein in recombination in *E. coli* is not clear.

E. Research goals

RecA, RuvA, RuvB, RuvC proteins are the major players in genetic recombination in *E. coli*. RecA mediates the formation of recombination intermediates (Holliday junctions). RuvAB processes the recombination intermediates by promoting branch migration. RuvC resolves the Holliday junctions to form mature recombination products.

The rapid progress in understanding the mechanism of bacterial recombination coupled with modern cloning and protein purification technology make it a realistic goal to develop a complete *in vitro* recombination system.

Recently RecA proteins from the thermophilic eubacteria *Thermus aquaticus*, *Thermus thermophilus*, *Thermotoga maritima* and *Aquifex pyrophilus* have been cloned and expressed in *E. coli*. Purified thermostable recombinant RecA proteins showed DNA-dependent ATPase activity at high temperatures

(Kato and Kuramitsu, 1993; Angov and Camerini-Otero, 1994; Wetmur, et al., 1994).

This thesis describes important steps toward developing a thermostable *in vitro* recombination system. Chapter II describes the cloning, expression and characterization of *T. thermophilus* RecA. Chapters III and IV describe the cloning, expression and characterization of *T. thermophilus* and *T. maritima* RuvB. There are several theoretical and practical advantages to a thermostable *in vitro* recombination system. First, at 70-80°C where thermostable proteins have their highest activity, most secondary structure has been removed from naked ssDNA. The study of RecA-mediated presynaptic complex formation may be simplified. Secondary structure can interfere with its formation (Muniyappa, et al., 1984; Tsang, et al., 1985; Kowalczykowski and Krupp, 1987), often leading to the additional requirement of a single-strand binding protein at 37°C. Similarly, the contribution of RecA to single-strand branch migration or RuvB to Holliday junction branch migration equilibria and kinetics may be investigated by direct comparison of reactions in the presence and absence of RecA or RuvB proteins. Secondly, it is easy to purify thermostable recombinant proteins free of protein and nucleic acid contamination from *E. coli*. Trace amount of nucleases, helix-destabilizing proteins and helicases can lead to artifacts (Kowalczykowski and Eggleston, 1994).

F. Helicase function

DNA helicases share the common property of being able to unwind duplex DNA. They use energy from the hydrolysis of nucleoside triphosphates to catalyze the breakage of hydrogen bonds between two DNA strands and to translocate along DNA. The first helicase identified was *E. coli* helicase I, which is required during conjugation (Abdel-Monem, et al., 1976). Since then a large number of DNA helicases have been identified in prokaryotes and eukaryotes as well as in bacteriophages and viruses (Lohman and Bjornson, 1996). Helicases play an essential role in almost all cellular processes involving nucleic acids such as DNA replication, recombination, repair and transcription. Despite the importance of these enzymes in DNA metabolism, the exact mechanism of helicase function remains unclear (Lohman and Bjornson, 1996). Sequence comparison has revealed that many helicases contain seven highly conserved motifs, including the Walker A and B motifs involved in nucleoside triphosphate binding (West, 1996). In general, DNA unwinding occurs with a unique directionality, either 5′-3′ or 3′-5′ polarity relative to the strand of DNA that is bound by the enzyme. The active forms of most helicases are oligomeric, predominantly dimeric and hexameric. The oligomeric nature of helicases is consistent with the requirement for two or more DNA binding sites in a functional helicase (Lohman and Bjornson, 1996). Interestingly, RuvB and other hexameric helicases such as *E. coli* Rho and bacteriophage T7 gp4 have structural homology to RecA and F_1-ATPase (Abrahams, et al., 1994; Washington, et al., 1996). RuvB forms a hexameric ring around DNA; six RecA monomers form one helical turn with the DNA lying within the deep groove of the nucleoprotein filament; three α- and three β- subunits of

F_1-ATPase form a hexameric ring with the γ- subunit positioned within the central cavity of the ring. Structurally, the R-loop in the F1-ATPase which contacts the γ-subunit corresponds to the L2-loop of RecA that interacts with DNA (Washington, et al., 1996). Functionally, the hexameric helicases and the F_1-ATPase are both implicated in rotational motion driven by ATP hydrolysis. They may share similar mechanisms for coupling ATP hydrolysis to rotation through conformational changes in their subunits. The recent direct observation of the rotation of the γ-subunit in the $\alpha_3\beta_3$ hexameric ring of F_1-ATPase may provide a model for studying helicase function (Noji, et al., 1997).

G. Crystal structure of helicases

Our current understanding of the structures of most helicases comes from low resolution electron microscopy. One helicase crystal structure has been determined, the PcrA helicase from *Bacillus stearothermophilus*. The structures of a monomeric form of the protein were determined alone and in a complex with ADP at 2.5 and 2.9 Å resolution, respectively (Subramanya, et al., 1996). Unfortunately, the biochemical properties of PcrA are not well characterized. PcrA shares 41% identity with *E. coli* Rep protein, which functions as a dimer, suggesting that PcrA probably belongs to the dimeric class of helicases. Surprisingly, the structure of the A subdomain of the PcrA helicase was similar to the ATPase domain of *E. coli* RecA protein. Comparison of the ATPase domain of the PcrA-ADP complex and the RecA-ADP complex revealed that the relative

coordinates of all of the key residues of the RecA active site were conserved in the helicase active site (Subramanya, et al., 1996).

Understanding the mechanism of helicase function will require detailed information about the structure of helicases, which can only be obtained by x-ray diffraction. No crystal structure of a hexameric helicase has been reported. In chapter IV, we will describe collaborative efforts toward solving the crystal structure of the hexameric DNA helicase RuvB.

II. Cloning, Sequencing and Expression of RecA from *Thermus thermophilus*

Jie Tong and James G. Wetmur

Data originally presented in Wetmur, J. G., Wang, D. M., Ortiz, B., Tong, J., Reichert, F. & Gelfand D. H. (1994) Cloning, Sequencing, and Expression of RecA Proteins from Three Distantly Related Thermophilic Eubacteria. *J. Biol. Chem.* **269**:25928-25935

A. Introduction

Eubacterial RecA proteins carry out several reactions essential for homologous recombination, including homologous pairing and strand exchange (Radding, 1991). *Escherichia coli* RecA is able to promote the search for rare sequences in duplex DNA (Honigberg, et al., 1986). It will form a complex with single-stranded DNA in the presence of ATP or the non-hydrolyzable substrate ATPγS. This complex will specifically bind homologous duplex DNA to form another complex with the three DNA strands in an unknown configuration (Camerini-Otero and Hsieh, 1993). This unusual structure appears to be transiently stable following deproteinization (Rao, et al., 1991; Hsieh, et al., 1990). The structure cannot be that of a classical triple helix (Jain, et al.,1992) and may be stabilized by differences in winding numbers between adjacent D-loops (Jwang and Radding, 1992). The availability of thermophilic RecA proteins with temperature optima nearer to the melting temperature of duplex DNA, where ssDNA lacks secondary structure, should facilitate the study of the formation and function of RecA-DNA complexes.

The crystal structure of *E. coli* RecA protein (Story, et al., 1992) and its ADP complex (Story and Steitz, 1992) have been reported. The conservation of prokaryotic RecA amino acid sequences is quite impressive, extending even to plant chloroplast RecA (Cerutti, et al., 1992). This conservation has permitted use of PCR with degenerate primers to amplify fragments of *recA* genes (Dybvig, et al., 1992; Duwat, et al., 1992). Although electron microscope image analysis has

suggested that the RAD51 protein of the eukaryote *Saccharomyces cerevisiae* adopts the same helical polymeric structure as RecA in the presence of DNA (Ogawa, et al., 1993) and may be part of a family of similar proteins including *S. cerevisiae* DMC1 protein and bacteriophage T4 UvsX protein (Story, et al., 1993), all of these proteins are markedly different from prokaryotic RecA proteins.

Determination of the structural features contributing to the thermostability of many proteins has been approached both through comparison of the amino acid sequences and the crystal structures of homologous mesophilic and thermophilic proteins and through protein engineering (Perutz and Raidt, 1975; Zuber, 1988; Menéndez-Arias and Argos, 1989; Zülli, et al., 1991; Wozniak, et al., 1990; Kelly, et al., 1993; Matthews, 1993). Although several recurring structural features have been identified, the data set and the theory are both inadequate for assessing the effects of amino acid substitution on protein thermophilicity and thermostability. The availability of the *E. coli* RecA protein crystal structure and the extensive amino acid sequence library of over 40 complete mesophilic RecA proteins suggested that additional insight into protein thermophilicity and thermostability might be extracted from the study of several RecA proteins from diverse thermophilic organisms, as well as from subsequent site-directed mutagenesis. In this work, we discuss several common features with the potential for contributing to RecA thermophilicity and thermostability.

B. Experimental procedures

Genomic DNA, plasmids, nucleotides and enzymes. All DNA manipulations used standard techniques and procedures (Sambrook, et al., 1989). Genomic DNA of *T. thermophilus* was kindly provided by Dr. David H. Gelfand, Roche Molecular Systems, Inc. Plasmids employed for cloning and expression were pUC19 and pDG160 (Lawyer, et al., 1993), which were grown in *E. coli* DH5α cells and DG116 cells (Lawyer, et al., 1993), respectively. All absorbance spectra were determined using a Hewlett-Packard diode array spectrophotometer equipped with a peltier temperature controller. Concentrations of DNA and primers were determined by using 50 and 36 μg ml^{-1} A_{260}^{-1}, respectively, as conversion factors. Deoxynucleoside triphosphates were obtained from Perkin-Elmer or purchased from Boehringer-Mannheim. [α-^{35}S]dATP and [γ-^{32}P] ATP were purchased from NEN/DuPont. Amplitaq DNA Polymerase, purchased from Perkin-Elmer, and native Taq polymerase, purchased from several suppliers, were used in the buffer supplied by the manufacturer. Restriction endonucleases, T4 polynucleotide kinase and T4 DNA ligase were purchased from New England Biolabs and used as recommended by the manufacturer. Simultaneous reactions with two or more restriction endonucleases were carried out in New England Biolabs NEB3 buffer. Simultaneous reactions with restriction endonucleases and T4 DNA ligase were carried out in the same buffer supplemented with 1 mM ATP.

Oligodeoxynucleotides. All synthetic oligodeoxynucleotide primers for PCR and sequencing were synthesized on automated instruments using standard

phosphoramidite chemistry and alcohol precipitated. The degenerate sense primer 5′ GCGGAATTCCAGCC(G/C)GACAC(G/C)GG(G/C)GA(A/G)CA(A/G)GC 3′ and antisense primer 5′ GCGAGATCTGGGTT(G/C)CC(A/G)AACAT(G/A/T/C)-AC(G/A/T/C)CC 3′, corresponding to amino acids 119-126 and 201-207 in *E. coli* RecA (see Figure 3, p 29), were used for cloning a fragment of the *T. thermophilus recA* gene [*Eco*RI (GAATTC) and *Bgl*II (AGATCT) recognition sequences are underlined]. The initial inverse PCR primers for *T. thermophilus* were 5′ GCGAGATCTGGTGCGGGAGAAGGTGGGCGT 3′ plus 5′ GCGGAATTCCCAC-CGCCCCCGAGCGGGC 3′. Additional inverse PCR (cloning) primers of the same type were synthesized as required for use with circularized *T. thermophilus* DNA templates. Additional sequencing primers lacking the GCG cap and restriction endonuclease sites were synthesized as required. PCR primers for cloning *T. thermophilus recA* genes into pDG160 were 5′ GCGAAGCTT***ATG***GACGAGAG-CAAGCGCAAGG 3′ plus 5′ GCGGGATCCAAAGCACCAGGTCCAGAAGGG 3′, respectively, where the *Bam*HI and *Hind*III sites are underlined and the initiation codon in the forward primer is shown in bold italics.

DNA amplification. PCR amplifications were carried out in a Perkin-Elmer/Cetus DNA Thermal Cycler or a USA/Scientific Gene Machine II with DNA templates in 100 µl containing 1 µM of each primer, 10 mM Tris-HCl buffer, pH 8.3, 50 mM KCl, 2.5-5 units Taq DNA polymerase, and 200 µM of each dNTP (Saiki, et al., 1988). Typically, simultaneous reactions were initiated by addition of a $MgCl_2$ solution to Mg^{2+}-free PCR mixtures at >80°C to yield final concentrations of 0.8-2 mM followed by denaturation for 30 sec at 96°C. When using degenerate

primers and 50 ng genomic DNA template, the first 5 cycles employed a 30 sec annealing step at 45°C followed by a 2 min ramp to 72°C before denaturation. An additional 30-35 cycles were carried out with a 55°C annealing temperature. For inverse PCR (Ochman, et al., 1990), genomic DNA was digested to completion with a restriction endonuclease leaving a 3′ or 5′ 4-base overhang, phenol extracted, and ligated overnight at a DNA concentration of less than 50 µg/ml. When using unique primers and 50 ng genomic DNA or circularized genomic DNA for direct or inverse PCR, respectively, the first 5 cycles were omitted.

Cloning, sequencing and Southern hybridization: Products of PCR amplifications were phenol extracted to remove Taq polymerase and filtered on Millipore Ultrafree-MC 30,000 NMWL filter units to remove primers. PCR products with *Bgl*II cloning sites were cloned into pUC19 by simultaneous digestion of vector and insert with *Bgl*II, *Bam*HI, and *Eco*RI, optional heat inactivation, ligation, and redigestion with *Bam*HI to destroy religated vectors without inserts. Inserts in pUC19 and pDG160 were sequenced in both orientations using insert-specific and vector-specific oligodeoxynucleotide primers with the Sequenase DNA Sequencing Kit (U.S. Biochemicals, Inc.). Cycle sequencing was carried out with Taq DNA polymerase using either ^{32}P-labeled primers. Southern hybridizations of restriction endonuclease-cleaved genomic DNAs were carried out with oligodeoxynucleotides labeled with ^{32}P using T4 polynucleotide kinase.

Expression. PCR products resulting from use of the expression primers on the *T. thermophilus* genomic DNAs were ligated into the *Bam*HI and *Hin*dIII sites in pDG160 for transformation of *E. coli* DG116 cells and growth at 30°C on LB-

AMP plates. Colonies derived from independent amplification reactions were grown overnight at 30°C in LB-AMP, diluted 1/40 to 100 ml into the same medium and grown to A_{600} > 0.5, induced at 42°C for 15 min, grown for an additional 3-4 hrs at 39°C, collected by centrifugation for 5 min at 10,000 rpm in a Beckman preparative ultracentrifuge in 3 tubes of an SW27 rotor and frozen as cell pellets in these tubes at -70°C for subsequent processing. The pellets from the 3 tubes were resuspended in 1 ml 0.05 M Tris-HCl, 0.005 M EDTA, pH 8 in a 1.5 ml microcentrifuge tube, disrupted on ice with the microtip of a Heat Systems Sonifier Cell Disrupter at power level 7 for pulses totaling 3 min, cleared of cell debris by microcentrifugation (10,000 g) for 5 min at 4°C, transferred to a new tube, made 0.3 M $(NH_4)_2SO_4$ by addition of 1 M stock, heated to 75°C for 15 min to denature thermolabile proteins, placed on ice for 30 min to aggregate the denatured proteins, cleared of DNA and denatured proteins by microcentrifugation for 15 min at 4°C, transferred to a new tube and made 10% glycerol for storage at -20°C. The partially purified products were assayed for the presence of a thermostable protein of the correct size by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Purification. The initial purification of the *T. thermophilus* RecA proteins was accomplished using the procedure described above. The product was loaded onto a 1 ml HiTrap Blue column (Pharmacia) column and extensively washed with 2 M NaCl, 10 mM Tris-HCl, pH 7.8, 1 mM EDTA to remove nucleic acids and low molecular weight thermostable polypeptides. The buffer was changed to 0.1 M NaOAc, 10 mM Tris-HCl, pH 7.8, 1 mM EDTA before elution with 2.0 M guanidine

HCl in that same buffer. The final products were analyzed by SDS-PAGE, protein concentrations were determined using the by the Bio-Rad Protein Assay kit (Bradford), and complete absorbance spectra were determined to ensure removal of nucleic acids.

ATPase assay. In addition to RecA protein, complete ATPase reaction mixtures contained 20 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 1 mM ATP and 3 μM (nucleotides) M13 ssDNA. A standard colorimetric assay was used to determine released P_i (LeBel, et al., 1978). Samples were taken as a function of time, stopped and assayed as described. Single-stranded DNA-dependent ATPase activity was proportional to the difference between the activity in the complete reaction mixture and activity in the reaction mixture lacking M13 ssDNA.

Computer analysis. Nucleic acid and protein sequence analyses were carried out using programs in GCG (Devereux, et al.,1984). Mesophilic RecA protein sequences were obtained from Genbank 80.0 (Benson, et al., 1993). The RecA protein sequences from 22 Gram-negative mesophilic bacteria, 4 Gram-positive mesophilic bacteria (including one mycoplasma species) and 2 cyanobacteria were aligned with the 4 newly-determined thermophilic RecA sequences using PILEUP. Also included was the *Arabidopsis thaliana* chloroplast RecA sequence. The multiple alignments were truncated to include only amino acids corresponding to 6 (NKQKA...) - 327 (...KVREL) of *E. coli* RecA prior to analysis using PHYLIP (Phylogeny Inference Package) version 3.5c (Felsenstein, et al., 1989). Pairwise distances between amino acids in the RecA sequences were calculated using PROTDIST with either the Dayhoff PAM matrix or the

Categories distance. Unrooted trees were calculated using FITCH with global rearrangement and jumbling before plotting with DRAWTREE.

C. Results and discussion

Cloning *Thermus thermophilus recA*. RecA protein is ubiquitous in prokaryotes. The amino acid sequences of RecA proteins among different bacterial species are quite conserved. This conservation permitted us to use PCR with degenerate primers to amplify a fragment of the *T. thermophilus recA* gene.

Two conserved regions were chosen by comparing two very unrelated bacterial RecA sequences: *E. coli* and *B. subtilis*. Two partially degenerate PCR primers (Figure 3) were designed based on the two conserved regions. In order to reduce the degeneracy of the primers, codon usage data for *T. aquaticus* (Lawyer, et al., 1989) were used to construct the 5′ end. Complete degeneracy was incorporated at the 3′ end, where mismatching most affects PCR. A 285 bp *T. thermophilus recA* gene fragment (Figure 4A) was amplified by PCR with the pair of degenerate primers. This PCR product was then cloned into pUC19 and sequenced.

In order to clone the whole *T. thermophilus recA* gene, the inverse PCR method was used instead of screening a genomic library. A pair of inverse PCR primers, which have a GCG cap and restriction endonuclease sites, 5′ GCGAGATCTGGTGCGGGAGAAGGTGGGCGT 3′ plus 5′ GCGGAATTCCCAC-CGCCCCCGAGCGGGC 3′, were synthesized based on the actual sequence of

Figure 3. Degenerate primers for cloning *Thermus thermophilus recA*

119	120	121	122	123	124	125	126
Q	P	D	T	G	E	Q	A
CAR	CCN	GAY	ACN	GGN	GAR	CAR	GC

Forward primer with *EcoRI* site (underlined):

GCGGAATTCCAGCC(G/C)GACAC(G/C)GG(G/C)GA(G/A)CA(G/A)GC

201	202	203	204	205	206	207
G	V	M	F	G	N	P
GGN	GTN	ATG	TTY	GGN	AAY	CC

Reverse primer with *BglII* site (underlined):

GCGAGATCTGGGTT(G/C)CC(G/A)AACAT(G/A/T/C)AC(G/A/T/C)CC

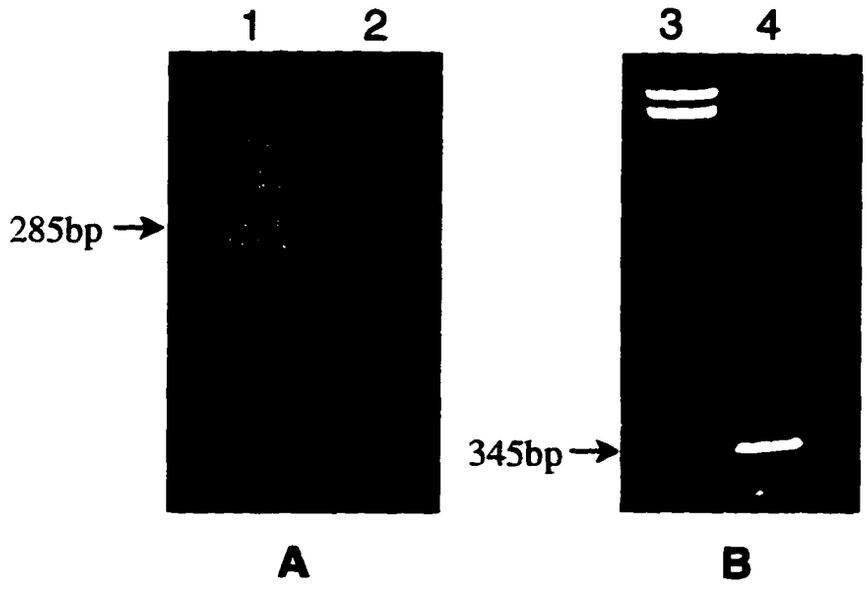


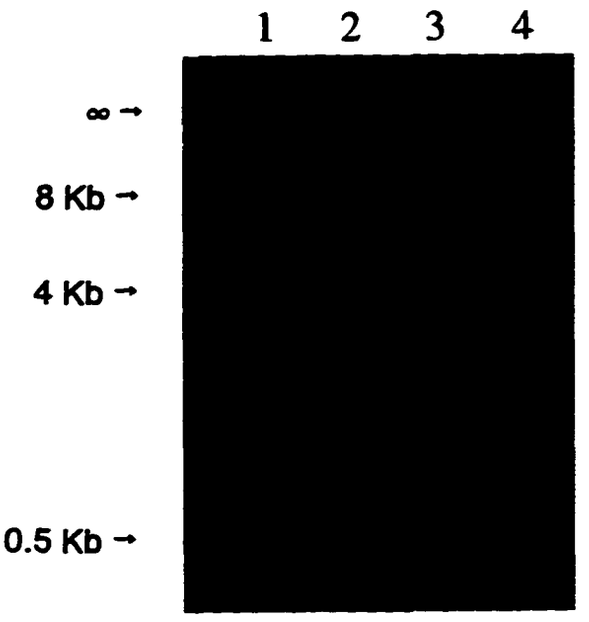
Figure 4. Analysis of PCR product with degenerate primers (A) and inverse PCR primers (B). Lane 1 and 3, markers; Lane 2, PCR product from degenerate priming; Lane 4, PCR product with inverse primers.

the T. thermophilus recA gene fragment between the degenerate primers. One inverse primer was used to probe a southern blot (Figure 5). The blot showed that the cloned sequence was derived from *T. thermophilus*. In addition, the blot demonstrated that SacI would be a useful restriction endonuclease for constructing circularized PCR templates. The 345 bp inverse PCR product (Figure 4B) was cloned into pUC19 and sequenced.

Inverse PCR procedures were repeated by using *Sau*3AI and *cfo*I digested and reclosed circles until the 5′ATG initiation codon and 3′ stop codon were identified. The 5′ end of *T. thermophilus recA* sequence was verified by direct cycle sequencing of the PCR product.

Expression of *T. thermophilus* RecA in *E. coli*. The expression primers were synthesized based on 5′ end sequence and 3′ untranslated region of the circularly permuted *T. thermophilus* DNA sequence. The sense primer was 5′ GCGAAGCTT***ATG***GACGAGAGCAAGCGCAAGG 3′, and the antisense primer was 5′ GCGGGATCCAAAGCACCAGGTCCAGAAGGG 3′, where the *Bam*HI and *Hind*III are underlined and the initiation codon in the forward primer is shown in bold italics. The full length *T. thermophilus recA* gene was amplified by PCR using expression primers. The PCR product was digested with *Bam*HI and *Hind*III, ligated into the expression vector pDG160 and cloned into DG116. Expression was regulated by a temperature-sensitive λ repressor. Cells are grown overnight at 30°C in LB-AMP, diluted 1/40 in 100 ml aliquots of the same medium, grown at 30°C to $A_{600} \approx 0.5$, induced at 42°C for 15 min, propagated for 3-4 hrs at 39°C, and pelleted at 5000 g for 15 min. Cell pellets were stored at -70°C before

Figure 5. Southern Blot hybridized with inverse PCR primer



Lane 1: *T. thermophilus* genomic DNA cut with *Sac*I.

Lane 2: *T. thermophilus* genomic DNA cut with *Pst*I.

Lane 3: *T. thermophilus* genomic DNA cut with *Bgl*II.

Lane 4: *T. thermophilus* genomic DNA cut with *Bam*HI.

purification.

Purification of recombinant *T. thermophilus* RecA protein. For initial purification, cell pellet from a 100 ml culture was resuspended in 1 ml 0.05 M Tris, 5 mM EDTA, pH 8.0, sonicated 3 min at 4°C to break the cells, and pelleted in a microfuge for 5 min at 4°C. The supernatant was transferred to a new tube, and $(NH_4)_2SO_4$ solution was added to the final concentration of 0.3 M. Cell lysates were incubated at 75°C to denature *E. coli* proteins, which were then precipitated on ice and removed by sedimentation. The supernatant contained a crude *T. thermophilus* RecA protein extract with nucleic acids and greatly reduced levels of *E. coli* proteins. For further purification, crude *T. thermophilus* RecA protein extract was bound to a 1 ml cibacron blue 3G-A agarose column which was previously equilibrated with 0.1 M NaOAc, 10 mM Tris-HCl, pH 7.8, 1 mM EDTA. The column was washed extensively with 2 M NaCl, 10 mM Tris-HCl, pH 7.8, 1 mM EDTA to remove nucleic acids and low molecular weight thermostable polypeptides. The buffer was changed to 0.1 M NaOAc, 10 mM Tris-HCl, pH 7.8, 1 mM EDTA before elution with 2 M guanidine HCl in the same buffer. *T. thermophilus* RecA protein in the elution buffer was later dialyzed into 0.2 M NaCl, 10 mM TrisHCl, pH 7.8, 1 mM EDTA. The recombinant *T. thermophilus* RecA protein purified by this method was free of *E. coli* proteins and nucleic acids contamination as verified by SDS-PAGE (Figure 6) and absorbance spectroscopy.

Figure 7 shows the *T. thermophilus* RecA coding sequence (Genbank accession number U03058). The *T. thermophilus recA* gene sequence was verified to be authentic by sequencing three independently-derived clones

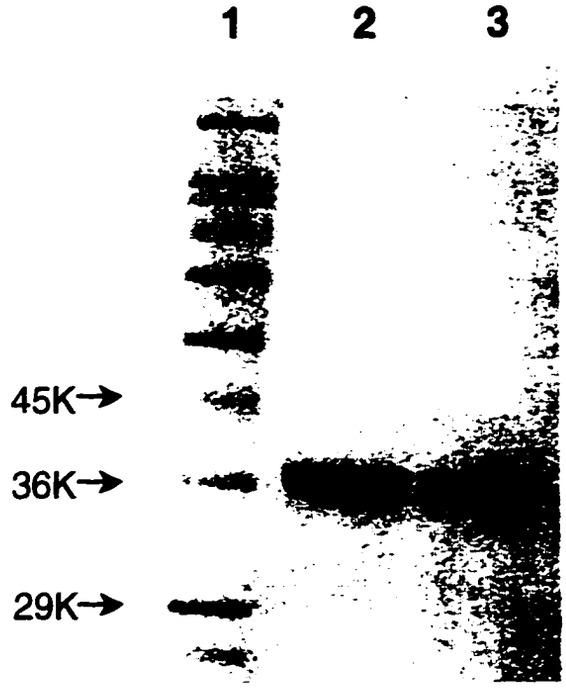


Figure 6. SDS-PAGE depicting purification of *T. thermophilus* RecA protein. Lane 1, molecular weight marker; Lane 2, fraction after heating at 70°C; Lane 3, fraction after HiTrap Blue chromatography. The SDS-PAGE was 10% and was stained with Coomassie Brilliant Blue R.

Figure 7. *T. thermophilus* RecA coding sequence

```
   1  ATGGACGAGA GCAAGCGCAA GGCCCTGGAG AACGCCCTGA AGGCGATTGA
  51  GAAGGAGTTC GGCAAGGGGG CGGTGATGCG GCTGGGCGAG ATGCCCAAGC
 101  AGCAGGTGGA CGTGATCCCC ACCGGCTCCC TCGCCCTAGA CCTCGCCCTG
 151  GGGATCGGCG GCATCCCCCG GGGGCGGATC GTGGAGATCT ATGGCCCCGA
 201  GTCCGGGGGC AAGACCACCC TCGCCCTCAC CATCATCGCC CAGGCCCAGA
 251  GGCGGGGCGG GGTGGCCGCC TTCGTGGATG CGGAGCACGC CCTGGACCCC
 301  CTGTACGCCC AGCGCCTCGG CGTCCAGGTG GAGGACCTCC TGGTCTCCCA
 351  GCCCGACACG GGCGAGCAGG CCCTGGAGAT CGTGGAGCTC CTCGCCCGCT
 401  CGGGGGCGGT GGACGTGATC GTGGTGGACT CGGTGGCCGC CTTGGTCCCC
 451  CGGGCGGAGA TTGAGGGGGA GATGGGGGAC CAGCACGTGG GCCTCCAGGC
 501  CCGGCTCATG AGCCAGGCCC TCCGCAAGCT CACCGCGGTG CTCGCCAAGA
 551  GCAACACCGC CGCCATCTTC ATCAACCAGG TGCGGGAGAA GGTGGGGGTC
 601  ACGTACGGCA ACCCCGAGAC CACCCCGGGG GGGAGGGCGC TGAAGTTCTA
 651  CGCCAGCGTG CGCCTGGACG TGCGCAAAAG CGGCCAGCCC ATCAAGGTGG
 701  GGAACGAGGC CGTGGGCGTC AAGGTGCGGG TCAAGGTGGT GAAGAACAAG
 751  CTCGCCCCCC CCTTCCGCGA GGCGGAGCTG GAGATCTACT TCGGCCGGGG
 801  CCTGGACCCG GTGGCCGACC TGGTGAACGT GGCCGTGGCC GCGGGGGTCA
 851  TTGAGAAGGC CGGGTCCTGG TTCTCCTACG GGGAGCTCCG CCTGGGCCAG
 901  GGGAAGGAGA AGGCGGCCGA GGCCCTGCGG GAGCGGCCCG AGCTTTTGGA
 951  GGAGATCCGC GCCAAGGTCT TGGAGCGCTC GGACCAGGTG GTCCTGGCCG
1001  CGGGCGAGGA CGAGGGGGAG TAG
```

G + C CONTENT: 70%

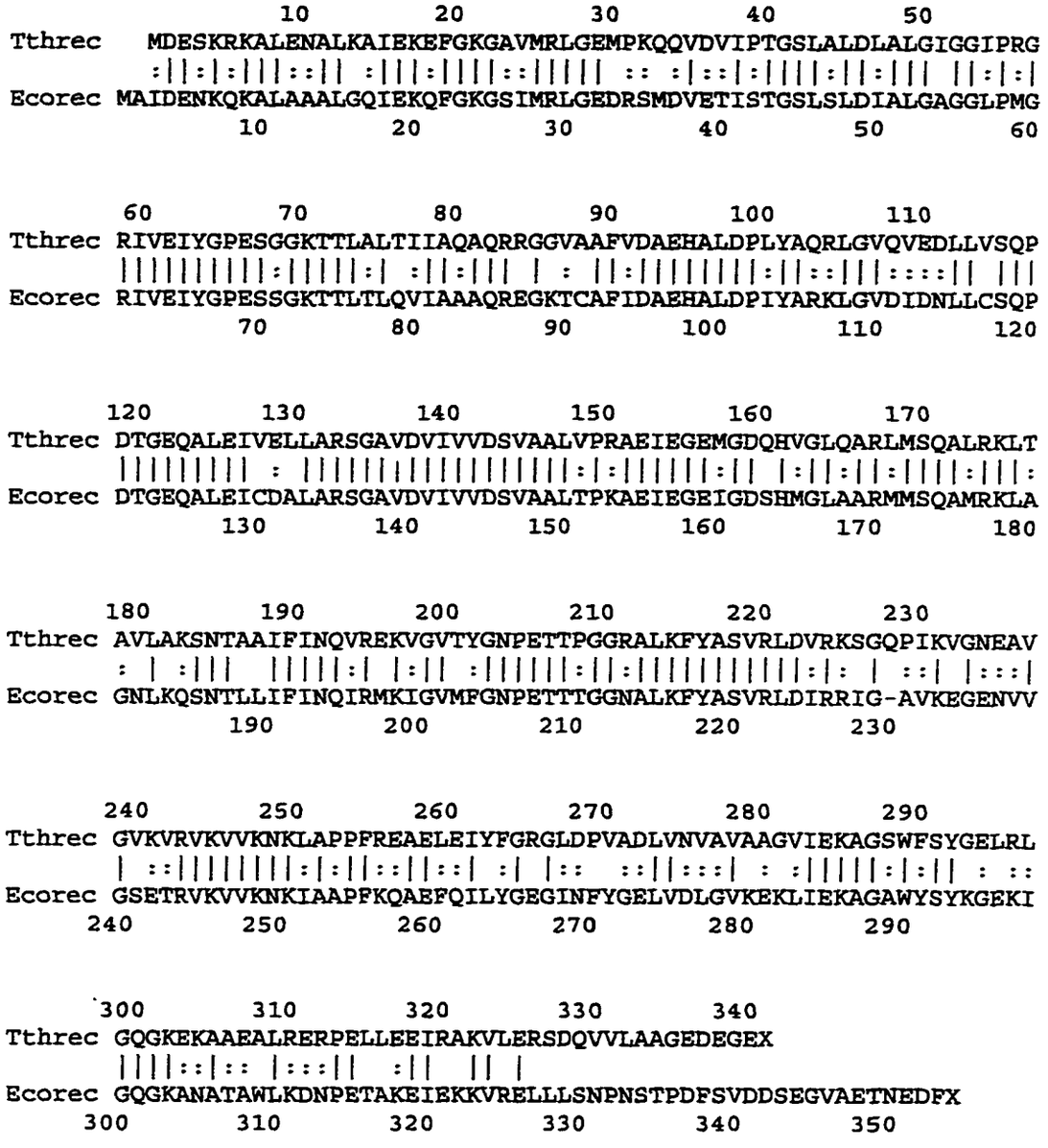
expressing thermostable *T. thermophilus* RecA proteins. The G+C content of the *T. thermophilus* RecA coding sequence was 70%, which coincides well with the G+C content of *T. thermophilus* genomic DNA (69%) (Oshima and Imahori, 1974). Since deposit of this sequence in Genbank, the same *T. thermophilus* RecA sequence has been reported by another group (Kato and Kuramitsu, 1993).

As shown in Figure 8, the amino acid sequence of *T. thermophilus* RecA is very similar with that of the *E. coli* RecA (62.5% identity).

Figure 9 shows the temperature-dependence of the DNA-dependent ATPase activity of thermostable *T. thermophilus* RecA protein compared to the temperature dependence of *E. coli* RecA. *T. thermophilus* RecA has a ssDNA-dependent ATPase optimum at 70°C versus *E. coli* RecA at 37°C.

Phylogenetic tree based on RecA protein sequences. The other three thermostable RecA proteins were cloned by other members of our laboratory group from the highly-divergent thermophilic eubacteria *T. aquaticus, T. maritima* and *A. pyrophilus* by using the same method. The amino acid sequences of the thermostable RecA proteins are compared in Figure 10, demonstrating their extensive degree of similarity. *A. pyrophilus*, *T. aquaticus*, *T. maritima* and *T. thermophilus* are 60.7, 61.5, 61.3 and 61.8% identical to *B. subtilis* RecA and 61.8, 60.6, 62.4 and 59.9% identical to *E. coli* RecA. Figure 11 depicts the unrooted evolutionary tree determined using the categories distance in PROTDIST for RecA protein sequences from the 4 Gram-negative thermophilic bacteria as well as 22 Gram-negative mesophilic bacteria, 4 Gram-positive mesophilic bacteria (including one mycoplasma species), and two cyanobacteria plus the plant *Arabidopsis*

Figure 8. Similarity of *E. coli* and *T. thermophilus* RecA Sequences



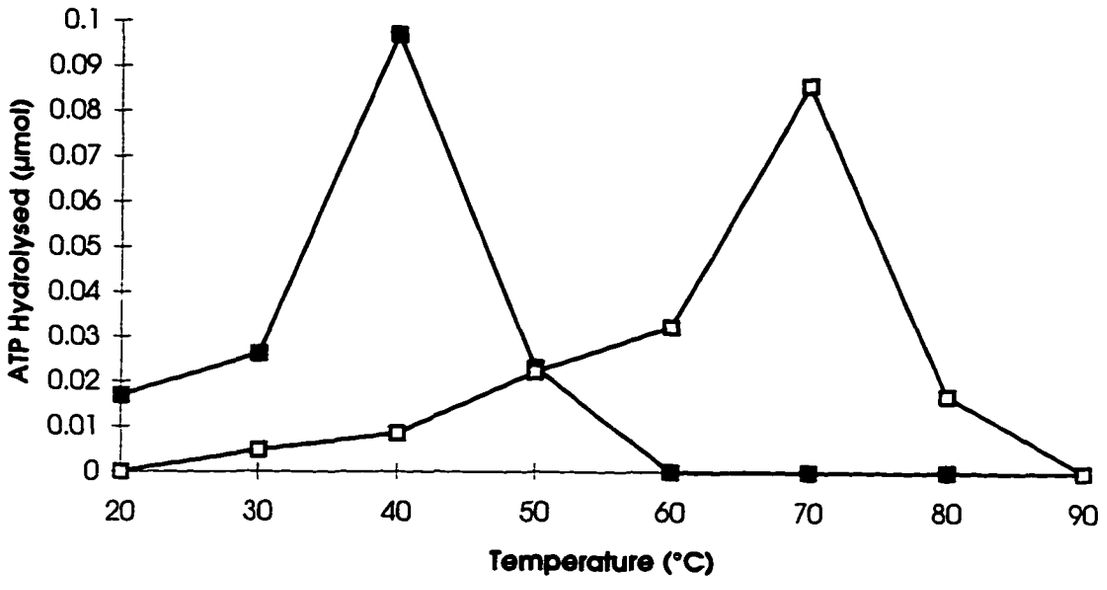


Figure 9. Temperature dependence of the ATPase activity of RecA proteins. Filled squares: *E. coli*; open squares: Tth.

Figure 10. Thermophilic RecA protein sequences

```
          1                                                  50
Apyreca   MARVSENLSE KMKALEVALS SIEKRFGKGA VMPLKAVETV E.VETIPTGS
Taqreca   ......MEEN KRKSLENALK TIEKEFGKGA VMRLGEMPKL Q.VDVIPTGS
Tthreca   ......MDES KRKALENALK AIEKEFGKGA VMRLGEMPKQ Q.VDVIPTGS
Tmareca   ....MPEEKQ KKSVLEKALK RIEENFGKGS IMILGDETQV QPVEVIPTGS

          51                                                 100
Apyreca   ISLDIATGVG GIPKGRITEI FGVESSGKTT LALHVIAEAQ KRGGVAVFID
Taqreca   LGLDLALGIG GIPRGRVTEI FGPESGGKTT LALTIIAQAQ KGGGVAAFVD
Tthreca   LALDLALGIG GIPRGRIVEI YGPESGGKTT LALTIIAQAQ RRGGVAAFVD
Tmareca   LAIDIATGVG GYPRGRIVEI FGQESSGKTT LALHAIAEAQ KMGGVAAFID

          101                                                150
Apyreca   AEHALDPKYA KKLGVDVDNL YISQPDYGEQ ALEIAESLIN SGAVDVIVVD
Taqreca   AEHALDPLYA KKLGVDVQEL LVSQPDTGEQ ALEIVELLAR SGAVDVIVVD
Tthreca   AEHALDPLYA QRLGVQVEDL LVSQPDTGEQ ALEIVELLAR SGAVDVIVVD
Tmareca   AEHALDPVYA KNLGVDLKSL LISQPDHGEQ ALEIVDELVR SGVVDLIVVD

          151                                                200
Apyreca   SVAALVPKDE LEGEMGEAQV GKQARLMSQA LRKLKGAVHR SNTALIFINQ
Taqreca   SVAALVPKAE IEGEMGDQHV GLQARLMSQA LRKLTAVLSK SNTAAIFINQ
Tthreca   SVAALVPRAE IEGEMGDQHV GLQARLMSQA LRKLTAVLAK SNTAAIFINQ
Tmareca   SVAALVPRAE IEGAMGDMQV GLQARLMSQA LRKIAGSVNK SKAVVIFTNQ
```

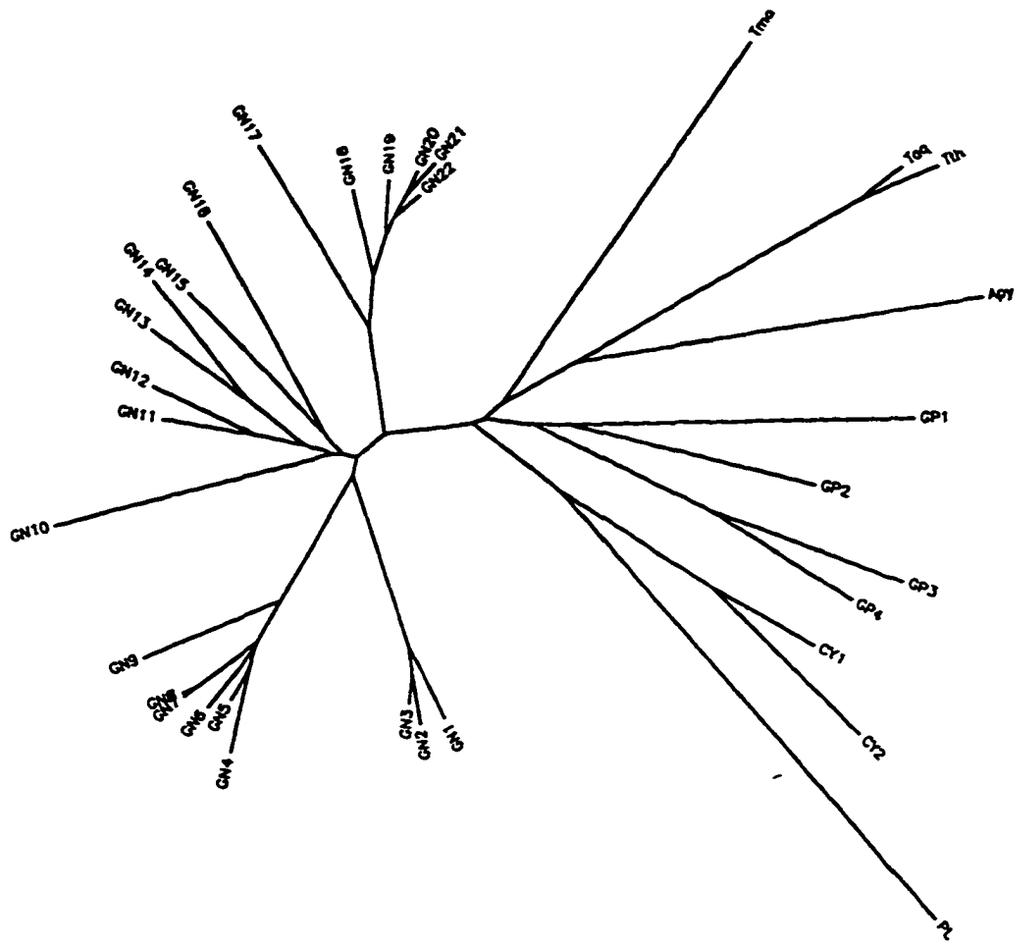
```
          201                                                  250
Apyreca   IREKIGVMFG NPETTPGGRA LKFFSDMRLE VRRLGD.VKE GGEKKGYRVK
Taqreca   VREKVGVMYG NPETTPGGRA LKFYSSVRLD VRKSGQPIKV GNEAVGIKVK
Tthreca   VREKVGVTYG NPETTPGGRA LKFYASVRLD VRKSGQPIKV GNEAVGVKVR
Tmareca   IRMKIGVMFG SPETTTGGLA LKFYATMRME VRR.GEPIKE GKDVIGNVIS

          251                                                  300
Apyreca   VRVVKNKLAP PFQEAEFDVI YGEGICRICD IIDTAANLGV ITKSGSWYSY
Taqreca   VKVVKNKLAP PFREAELEIY FGRGLDPVMD LVNVAVAAGV IEKAGSWFSY
Tthreca   VKVVKNKLAP PFREAELEIY FGRGLDPVAD LVNVAVAAGV IEKAGSWFSY
Tmareca   VKIVKNKVAP PFKTAQTYII YGKGIDREYE LFNIAVNEGI VDRKGSWYYY

          301                                                  350
Apyreca   .....GEKRL GQGREQAKKY LLEHPEMLEE IERKVREVSG LVRPDTENSV
Taqreca   .....GEHRL GQGKEKAAEY LRERPELLEE IRAKVLERAD KVVLAAGEEE
Tthreca   .....GELRL GQGKEKAAEA LRERPELLEE IRAKVLERSD QVVLAAGEDE
Tmareca   TTLKGEEVSL GQGSSNAVQF LKDNPEIAGE IERRIREKYG LLSVEKEEQR

          351       362
Apyreca   GEKSE*.... ..
Taqreca   GE*....... ..
Tthreca   GE*
```

Figure 11. Unrooted evolutionary tree for RecA proteins



GN1: Pseudomonas fluorescens, GN2: Azotobacter vinelandii, GN3: Pseudomonas aeruginosa, GN4: Proteus vulgaris, GN5: Serratia marcescens, GN6: Escherichia coli, GN7: Erwinia carotovora, GN8: Proteus mirabilis, GN9: Vibrio anguillarum, GN10: Neisseria gonorrhoeae, GN11: Methylobacillus flagellatum, GN12: Methylomonas clara, GN13: Pseudomonas cepacia, GN14: Bordetella pertussis, GN15: Legionella pneumophila, GN16: Thiobacillus ferrooxidans, GN17: Aquaspirillum magnetotacticum, GN18: Brucella abortus, GN19: Rhizobium meliloti, GN20: Rhizobium leguminosarum viciae, GN21: Rhizobium leguminosarum phaseoli, GN22: Agrobacterium tumefaciens, Tma: Thermotoga maritima, Taq: Thermus aquaticus, Tth: Thermus thermophilus, Apy: Aquifex pyrophilus, GP1: Acholeplasma laidlawii, GB2: Bacillus subtilis, GP3: Lactococcus lactis, GP4: Streptococcus pneumoniae, CY1: Anabaena variabilis, CY2: Synechococcus sp., PL: Arabidopsis thaliana (transported to chloroplast)

thaliana, where the *recA* gene is in the nucleus but the RecA protein is directed to the chloroplasts. The same four arms are found using the Dayhoff PAM matrix in PROTDIST (data not shown). The tree structure indicates that the sequences of the thermophilic bacteria are as different from the mesophilic Gram-negative bacteria as are these sequences from Gram-positive bacteria or cyanobacteria.

Common characteristics of thermostable RecA proteins. In spite of substantial sequence divergence, interesting common characteristics could be distinguished among the four thermostable RecA amino acid sequences. Overall, after eliminating the effects of codon bias, the major trends were loss of C, substitution of E for D, increased positively-charged amino acid content, decreased polar N, S and T, and increased hydrophobic P and especially V content. Of particular interest are substitutions common to thermostable proteins. A Y, found in 94% of all mesophilic RecA proteins at *E. coli* position 66, is replaced by F in *A. pyrophilus*, *T. aquaticus* and *T. maritima*. This site is near the end of β-sheet 1 and immediately preceding the consensus GXXXXGKT/S ATP-binding site leading into helix C. This substitution accounts for all of the Y to F. Similarly, an A found in 94% of all mesophilic RecA proteins at *E. coli* position 290, in the turn between β-sheet elements 9 and 10 involved in RecA-RecA interactions, is replaced by S in *A. pyrophilus*, *T. aquaticus*, *T. maritima* and *T. thermophilus*, again accounting for all of the A to S substitutions. These two positions would be good candidates for site-directed mutagenesis aimed at analyzing the temperature-dependence of the thermostable RecA proteins.

Another global effect of the substitutions found in the thermophilic

organisms was an increase in the calculated isoelectric point (GCG-ISOELECTRIC) compared to that of mesophilic RecA proteins by greater than a full pH unit. Table 1 lists representative values for mesophilic and thermophilic organisms at 37°C. It is likely that the apparent increase in isoelectric point would be different if the calculations were made at the temperatures of maximum enzymatic activity.

Table 1

Calculated isoelectric points of RecA proteins

	Mesophilic species	**pI**	
Gram-negative:	*Agrobacterium tumefaciens*	5.1	
	Escherichia coli	4.9	
	Neisseria gonorrhoeae	5.0	Average = 5.0
Gram-positive:	*Bacillus subtilis*	4.9	
Cyanobacteria:	*Anabaena variabilis*	4.9	
	Thermophilic species		
Gram-negative:	*Aquifex pyrophilus*	6.2	
	Thermus aquaticus	6.0	Average = 6.0
	Thermotoga maritima	6.4	
	Thermus thermophilus	5.5	

III. Cloning, Sequencing and Expression of *ruvB* and Characterization of RuvB Proteins from Two Distantly Related Thermophilic Eubacteria

JIE TONG AND JAMES G. WETMUR

Department of Microbiology, Mount Sinai School of Medicine, New York, NY 10029

Originally published in *J. Bacteriol.* **178**: 2695-2700 (1996). Reprinted with permission from the publisher.

A. Introduction

Formation and resolution of homologous recombination intermediates in bacteria depend on RecA and RuvA, B and C proteins. *E. coli* RuvA and RuvB proteins function together to catalyze the branch migration of Holliday junctions (reviewed by Kowalczykowski, et al., 1994; Müller and West, 1994; West, 1994; West, 1995). Although RecA can catalyze branch migration through mismatched sequences, RuvA and RuvB are required both to accelerate branch migration and to bypass insertions greater than 100 bp (Iype, et al., 1994). Individual roles have been assigned to RuvA and RuvB proteins. RuvB is a DNA-dependent ATPase (Iwasaki, et al., 1989; Marrione and Cox, 1995) which forms hexameric rings on duplex DNA (Stasiak, et al., 1994; Mitchell and West, 1994). At high concentrations, RuvB alone can promote branch migration in the presence of ATP and high concentrations of Mg^{2+} (≥15 mM) (Tsaneva, et al., 1992; müller, et al., 1993a), indicating that RuvB is the engine which drives branch migration of Holliday junctions. RuvA is a DNA binding protein with specificity for Holliday junctions (Parsons, et al., 1992). RuvA will form a stable complex with RuvB in solution (Shiba, et al., 1993). At low concentrations of Mg^{2+} and RuvB, RuvA is required for branch migration (müller, et al., 1993a, müller, et al., 1993b). RuvA directs the RuvB helicase to its Holliday junction substrate (Hiom and West, 1995; Parsons, et al., 1995).

E. coli RuvA and RuvB not only function together but are expressed from the same operon. There are two operons in the *E. coli ruv* locus. The *ruvA* and *ruvB* genes form an operon that is regulated by LexA (Benson, et al., 1988;

Shinagawa, et al., 1988b). The *ruvC* gene is expressed from a separate operon which is located within 2 Kb upstream of the *ruvAB* operon.

To date, only two sequences similar to *E. coli ruvAB* have been identified, one from *Mycobacterium leprae* (Genbank No. U00011, K. Robison, Harvard), an organism phylogenetically related to Gram-positive bacteria, and the other from *Haemophilus influenzae* (Genbank No. U32716) (Benson, et al., 1993; Fleischman, et al., 1995), a Gram-negative bacterium. In both cases, the *ruvA* and *ruvB* genes are contiguous and apparently part of one operon. The presence of *ruvAB* in both Gram-negative and Gram-positive bacteria suggests conservation of RuvAB function in highly divergent prokaryotes.

RecA proteins have been cloned and expressed from several thermophilic bacteria including *T. thermophilus*, *T. aquaticus*, *T. maritima* and *A. pyrophilus* (Kato and Kuramitsu, 1993; Angov and Camerini-Otero, 1994; Wetmur, et al., 1994). In this paper, we report the cloning, sequencing, expression and purification of RuvB proteins from two distantly related thermophilic eubacteria, *T. thermophilus* and *T. maritima*. Surprisingly, both *T. thermophilus* and *T. maritima ruvB* genes were not part of an operon with *ruvA* and did not follow *ruvC*. Nevertheless, these two thermostable RuvB proteins share high amino acid identity with *E. coli* RuvB (55% for both *T. thermophilus* and *T. maritima*). In this work, we demonstrate that the thermostable RuvB proteins have similar functions to *E. coli* RuvB *in vivo* and *in vitro*. Expression of *T. thermophilus* RuvB partially complemented a mutation in *E. coli ruvB*. Both *T. thermophilus* and *T. maritima* RuvB formed stable complexes with duplex DNA and demonstrated strong double-

stranded DNA-dependent ATPase activity. Both *T. thermophilus* and *T. maritima* RuvB unwound supercoiled DNA, with *T. maritima* RuvB showing a greater degree of unwinding within the limits of the assay employed.

B. Materials and methods

Bacterial strains and plasmids. *E. coli* strains DH5α, with plasmid pUC19, and NovaBlue(DE3)pLysS, with plasmid pET11c (Novagen) were used for plasmid manipulation and RuvB expression, respectively. *E. coli* strains N3395 (*recG ruvB*) (Lloyd, 1991) and XAO (*recG*$^+$ *ruvB*$^+$, *E. coli* Genetic Stock Center) were used for the complementation assay. Supercoiled (Form I) pUC19 DNA was purified to homogeneity by equilibrium centrifugation in CsCl-ethidium bromide.

Genomic DNA, nucleotides and enzymes. All DNA manipulations used standard techniques and procedures (Sambrook et al., 1989). Genomic DNAs of *T. thermophilus* and *T. maritima* were kindly provided by Dr. David H. Gelfand, Roche Molecular Systems, Inc. All absorbance spectra were determined with a Hewlett-Packard diode array spectrophotometer. Protein Concentrations were measured by using a protein assay kit (Bio-Rad) with bovine serum albumin as the standard. Deoxynucleoside triphosphates and ATP were purchased from Boehringer-Mannheim. ATPγS was purchased from Sigma. [α-^{35}S]dATP and [γ-^{32}P] ATP were purchased from NEN/DuPont. Amplitaq DNA Polymerase, purchased from Perkin-Elmer, and native Taq polymerase, purchased from several suppliers, were used in the buffers supplied by the manufacturers. Calf thymus topoisomerase I (topo I) was purchased from GibcoBRL, Inc. Restriction

endonucleases, T4 polynucleotide kinase and T4 DNA ligase were purchased from New England Biolabs and used as recommended by the manufacturer.

Oligodeoxynucleotides. All synthetic oligodeoxynucleotide primers for PCR and sequencing were ordered from GENSET. The degenerate sense primer 5′ GCGGAATTCCC(G/C)GG(G/C)CT(G/C)GG(G/C)AA(G/A)AC 3′ and antisense primer 5′ GCGAGATCTTGGATCTC(G/A)TC(G/A/T)AT(G/A)AA 3′ corresponding to amino acids 64-69 (PGLGKT) and 111-116 (FIDEIH) in *E. coli* RuvB, were used for cloning a fragment of the *T. thermophilus ruvB* gene. *Eco*RI (GAATTC) and *Bgl*II (AGATCT) recognition sequences are underlined. The degenerate sense primer 5′ GCGGAATTCAG(A/G)GA(T/C)AG(A/G)TT(T/C)GG 3′ and antisense primer 5′ GCGAGATCTAAGTC(T/C)CT(A/C)AC(T/C)CT(T/C)CT 3′, corresponding to amino acids 172-176 (RDRFG) and 228-233 (RRVRDF) in *E. coli* RuvB, were used for cloning a fragment of the *T. maritima ruvB* gene. Additional inverse PCR and sequencing primers were synthesized as required. PCR primers for cloning *T. thermophilus* and *T. maritima ruvB* genes into pET11c were 5′ GCGCATA***TG***GAAGACCTCGCCCTTAGG 3′ plus 5′ GCGGGATCCAGGTCCAG-CTCGGAAAGCGTG 3′, 5′ GCGCATA***TG***AGTGAATTTCTCACACCT 3′ plus 5′ GATGGATCCCTGTTTGCTCG 3′, respectively, where the *Nde*I and *Bam*HI sites are underlined and the initiation codon in the forward primer is shown in bold italics. Forward *T. thermophilus* and *E. coli ruvB* primers for cloning into pUC19 were 5′ GCGAAGCTTGGAGGAATTTTAACGC***ATG*** followed by additional *ruvB* sequences; reverse primers were 5′ GCGGAATTC followed by antisense sequences following the translation terminator. The *Hind*III and *Eco*RI sites are

underlined.

DNA amplification, cloning and sequence analysis. PCR amplifications with degenerate and unique primers and inverse PCR on circularized templates were carried out in a Ericomp Powerblock thermocycler and ligated into pUC19 or into the *Nde*I and *Bam*HI sites of pET11c using conditions previously described (Wetmur, et al., 1994) and transformed in the appropriate strains. Inserts in pET11c were sequenced in both orientations using insert-specific and vector-specific oligodeoxynucleotide primers with the Sequenase DNA Sequencing Kit (U.S. Biochemicals, Inc.). Cycle sequencing of PCR products was carried out with Taq DNA polymerase using ^{32}P-labeled primers. Nucleic acid and protein sequence analyses were carried out as previously described (Wetmur, et al., 1994) using Genbank sequences (Devereux, et al., 1984) and the GCG (Benson et al., 1993) and PHYLIP (Felsenstein, 1989) programs.

Expression. NovaBlue(DE3)pLysS cells expressing *T. thermophilus* or *T. maritima* RuvB from a pET11c construct were propagated at 37°C in LB containing 50 µg/ml ampicillin (AMP) and 25 µg/ml chloramphenicol (CAM). Overnight cultures were diluted 1/100 into the same media, grown to $A_{600} > 0.5$, induced by adding IPTG to the final concentration of 1 mM, and grown for an additional 4-5 hrs. Cells were collected, frozen, disrupted and clarified as previously described (Wetmur, et al., 1994). The resulting solutions were heated at 70°C for 15 min to denature thermolabile proteins, placed on ice for 30 min to aggregate the denatured proteins, and cleared of denatured proteins by microcentrifugation for 15 min at 4°C.

Purification. Crude RuvB protein, approximately 1 ml per 250 ml culture, was loaded onto a 1 ml HiTrap Q anion exchange column (Pharmacia), repeatedly washed with 20 mM Tris-HCl, pH 8.0 or 9.0 for *T. thermophilus* or *T. maritima* RuvB, respectively, and eluted with 0.2 M NaCl in the same buffer. The eluate was loaded onto a 1 ml HiTrap Blue affinity column (Pharmacia). Columns containing *T. thermophilus* or *T. maritima* RuvB were washed extensively with 2 M NaCl, 10 mM Tris-HCl, pH 7.8, 1 mM EDTA, or with 0.1 M NaOAc in the same buffer, respectively, and eluted in 2.0 M guanidine HCl in that same buffer. After dialysis against 0.2 M KCl, 0.1 M Tris-HCl, pH 8.0, 1 mM EDTA and concentration using Centricon-30 (Amicon), protein concentrations were determined and compared with complete absorbance spectra to determine an extinction coefficient and to verify removal of nucleic acids. Purification from other proteins was verified by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) where overloaded gels were stained with Coomassie Brilliant Blue R.

RuvB ATPase assay. The ATPase activity of *T. thermophilus* and *T. maritima* RuvB proteins was measured by colorimetric detection of released P_i as described (LeBel, et al., 1978). Reaction mixtures (0.5 ml) contained 10 µg *T. thermophilus* RuvB or 20 µg *T. maritima* RuvB, 20 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$, 1 mM ATP. Some reactions contained 6 µM (nucleotides) M13 single-stranded DNA (ssDNA), pUC19 linearized by *Eco*RI, or supercoiled pUC19.

RuvB protein-DNA complex formation. PCR-derived linear 337 bp double-stranded DNA (dsDNA) at 20 µg/ml was incubated with various concentrations of RuvB protein in binding buffer (10 mM TrisHCl, pH 8.0, 0.1 mM

EDTA, 50 mM NaCl, 10 mM $MgCl_2$) at 60°C for 1 hr prior to electrophoretic separation on a 1% agarose gel, staining with ethidium bromide and fluorography. Some reactions contained 1 mM ATP or ATPγS or 20 µg/ml competing PCR-derived linear 1079 bp dsDNA.

Topological assay for DNA unwinding by *T. thermophilus* and *T. maritima* RuvB. A topological assay for DNA unwinding was performed as described (Adams and West, 1995). Supercoiled pUC19 (30 µM in nucleotides) was incubated with 4.0 µM *T. thermophilus* or *T. maritima* RuvB in a 100 µl reaction containing 30 mM Tris-HCl (pH 8.0), 30 mM $MgCl_2$, 0.1 mM EDTA, 30 mM NaCl, 0.5 mM dithiothreitol, 30 µg bovine serum albumin/ml at 60°C for 15 min. Some samples contained 0.5 mM ATP and/or 0.5 mM ATPγS. After 5 min temperature equilibration at 37°C, each incubation was continued for 1 hr with 6 units of calf thymus topo I. The reaction was stopped by adding SDS, EDTA and proteinase K to 0.4% (w/v), 40 mM and 20 µg/ml, respectively. DNA was extracted with phenol/chloroform, ethanol precipitated, resuspended and analyzed by electrophoresis in 1.5% agarose in TAE (0.04 M Tris-acetate, 0.001 M EDTA, pH 8.0) at ~2 V/cm for 14 hr. Two-dimensional gel analyses were performed as described (Kilpatrick, et al, 1984; Blaho, et al., 1988). The first dimension was the same as described above. The gel was soaked in the same buffer plus 1 µM chloroquine for 1 hr prior to electrophoresis. The second dimension was carried out for the same time in the chloroquine-containing TAE buffer. Gels were stained with ethidium bromide and analyzed by fluorography.

RuvB complementation assay. Polycistronic pUC19 plasmids pTth

and pEc, expressing *T. thermophilus* and *E. coli ruvB* genes, respectively, were constructed such that a ribosome binding site and a stop codon were introduced into the pUC19 *lac* gene at the *Hin*dIII site and followed immediately by *ruvB* initiation codons. Single colonies of wild type *E. coli* XAO or N3395 *ruvBrecG* N3395 containing pUC19 or one of the expression plasmids were grown to late log phase in LB-AMP. Cells were diluted to ~1×10^4 colony-forming units (CFU) per 0.1 ml, plated in the same media and exposed to a calibrated 254-nm UV source for various times. After overnight incubation at 37°C, the surviving fraction was determined as the CFU exposed plates divided by CFU on unexposed plates.

Genbank. The nucleotide sequences of *T. maritima* and *T. thermophilus* RuvB have been deposited in Genbank (Accession Numbers: *T. thermophilus* RuvB, U22817; *T. maritima* RuvB, U38840)

C. Results

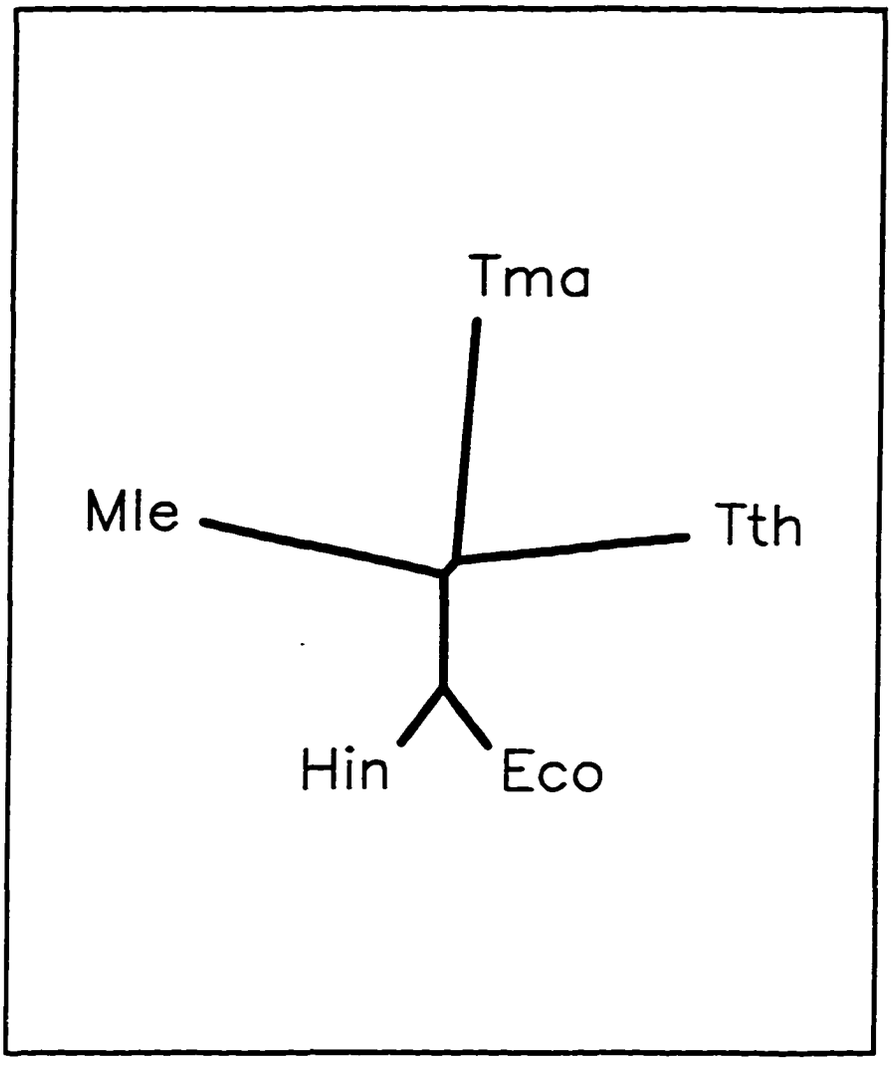
Cloning and sequencing of *T. thermophilus* and *T. maritima ruvB* genes. Degenerate primers corresponding to conserved amino acid sequences in *E. coli* and *Mycobacterium leprae* were used to amplify, clone and sequence fragments of the *T. thermophilus* and *T. maritima ruvB* genes. The complete sequences were determined by inverse PCR, and amplified full-length genes were cloned into the expression plasmid pET11c. The *T. thermophilus* and *T. maritima ruvB* sequences were confirmed by sequence identity of at least two independently derived clones expressing a thermostable RuvB protein. The

guanine plus cytosine content of *T. thermophilus* and *T. maritima* RuvB were 70 and 50, respectively.

The *T. thermophilus* and *T. maritima* translated amino acids sequences were 54% identical and both 55% identical to *E. coli* RuvB. Figure 12 shows an unrooted evolutionary tree of all RuvB sequences known to date. *T. thermophilus* and *T. maritima* RuvB are as divergent from one another and from mesophilic Gram negative organisms as are mesophilic Gram-negative organisms from mesophilic Gram-positive organisms.

Expression and purification of *T. thermophilus* and *T. maritima* RuvB protein. The pET11c plasmids expressing *T. thermophilus* or *T. maritima* RuvB were transformed into NovaBlue(DE3)pLys which is an *endA* mutant rather than BL21(DE3)pLys because BL21 expresses endonuclease I, which was determined to be not only thermostable but thermoactive. *T. thermophilus* and *T. maritima* RuvB protein were expressed at 10-20% of total cellular protein (Figure 13A and 13B, lane 2). Figure. 13A and 13B, lane 3 depict the extensively purified crude fractions subsequent to heat-treatment step. In the absence of expression of thermostable RuvB (vector without insert), no protein band the size of RuvB was observed (data not shown). Figure 13A and 13B, lanes 4 and 5 show only one protein band following anion and affinity chromatography. The HiTrap Blue affinity chromatography was necessary for removing residual nucleic acids. After chromatography, the absorbance spectrum showed OD_{280} / OD_{260} = 1.3. Both *T. thermophilus* and *T. maritima* RuvB proteins were purified to homogeneity as determined both by SDS-PAGE (Figure 13A and 13B, lane 5). All subsequent

Figure 12. Unrooted evolutionary tree for RuvB proteins



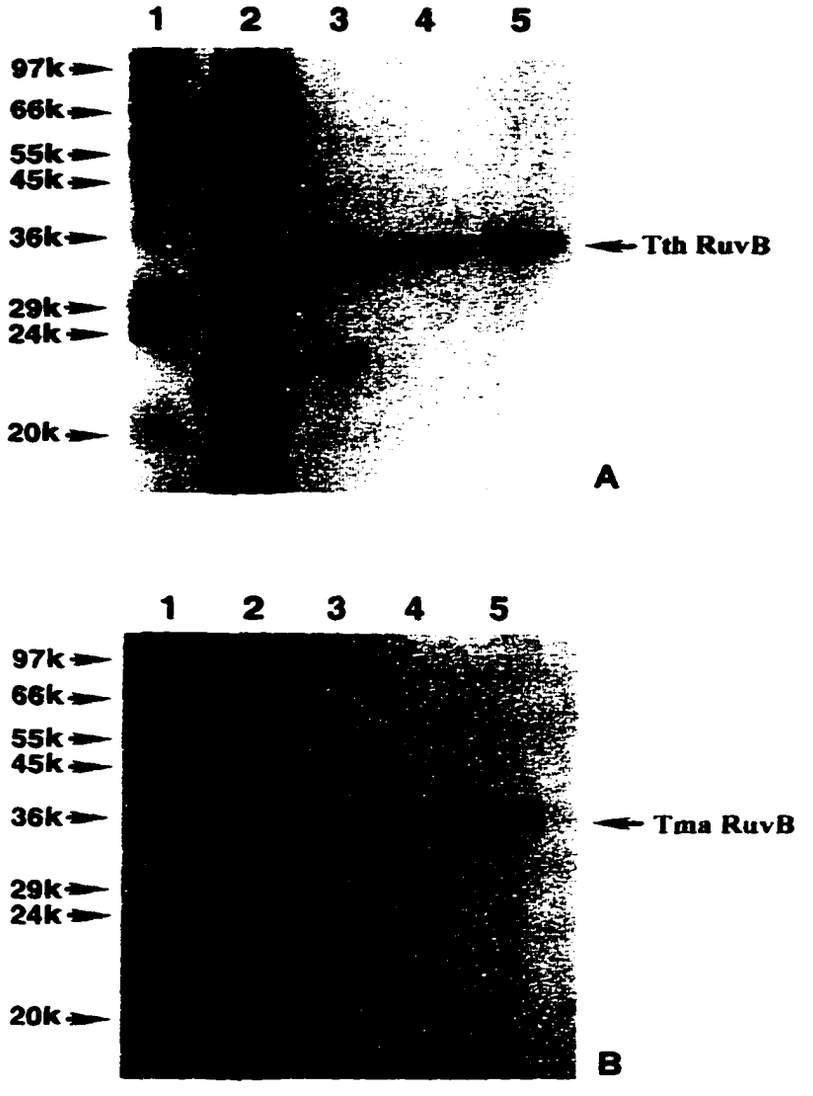


Figure 13. SDS-PAGE analysis of *T. thermophilus* and *T. maritima* RuvB proteins at each step of purification. Panel A and B: Lane 1, molecular weight marker; Lane 2, induced cell lysate; Lane 3, fraction after heating at 70°C; Lane 4, fraction eluted from HiTrap Q column; Lane 5, fraction eluted from HiTrap Blue column. The SDS-PAGE was 12.5% and was stained with Coomassie Brilliant Blue R.

experiments were carried out using these purified RuvB proteins.

ATPase activity of thermostable *T. thermophilus* and *T. maritima* RuvB. Figure 14A and 14B depict the temperature dependence of ATPase activity of *T. thermophilus* RuvB (A) and *T. maritima* RuvB (B) proteins and their dependence on various DNA substrates. *T. thermophilus* and *T. maritima* RuvB have activity optima at 70°C and 80°C, respectively. At 70°C, *T. thermophilus* RuvB rates of ATP hydrolysis were 1.2 (no DNA), 1.2 (ssDNA), 17 (linear pUC19 dsDNA) and 20 (supercoiled pUC19 DNA) mol ATP hydrolyzed/min per mol of RuvB. At 80°C, *T. maritima* RuvB rates of ATP hydrolysis were 1.2 (no DNA), 1.9 (ssDNA), 13 (linear dsDNA) and 11 (supercoiled DNA) mol ATP hydrolyzed/minute per mol of RuvB. The equivalent stimulation of *T. thermophilus* and *T. maritima* RuvB ATPase activity by linear or supercoiled dsDNA sharply contrasts with the greater stimulation of *E. coli* RuvB by supercoiled DNA (Mitchell and West, 1994; Marrione and Cox, 1995).

Binding of *T. thermophilus* and *T. maritima* RuvB to linear dsDNA. Figure 15 depicts a gel retardation assay for *T. thermophilus* RuvB binding to 337 bp linear DNA at 60°C. Increasing *T. thermophilus* RuvB concentration led to increasing numbers of well-spaced bands (lanes b-f), indicating complexes with increasing numbers of *T. thermophilus* RuvB molecules. The binding was apparently non-cooperative. In contrast, complexes of *T. maritima* RuvB and DNA did not produce well-spaced bands (data not shown), either indicating heterogeneity of structure or instability of the complexes. Addition of ATP (1 mM) to *T. thermophilus* RuvB-DNA complexes led to release of bound *T. thermophilus*

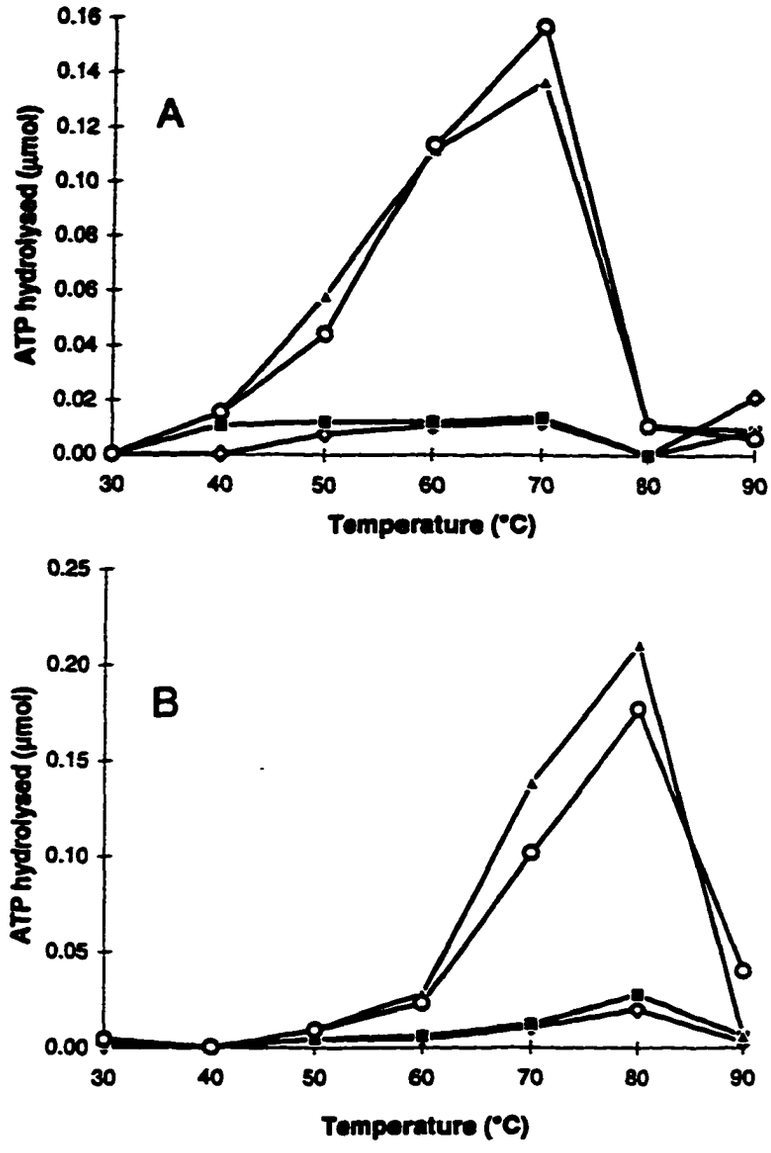


Figure 14. Temperature dependence of ATPase activity of *T. thermophilus* RuvB (A) and *T. maritima* RuvB (B) proteins in the presence of circular M13 ssDNA (filled squares), linear pUC19 (filled triangles), supercoiled pUC19 (open circles) or no DNA (open diamonds) in a 30 min reaction.

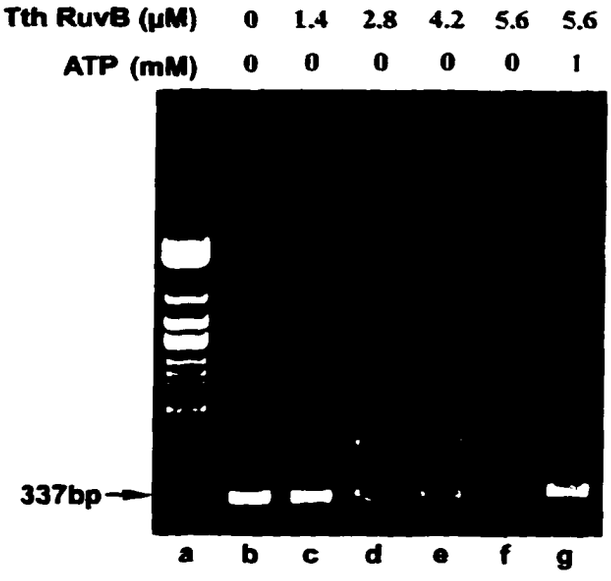


Figure 15. Binding of *T. thermophilus* RuvB to linear duplex DNA. Lane a: DNA marker. Lane b-g: 0.2 μg of 337 bp PCR product incubated with the indicated concentrations of RuvB protein in binding buffer for 1 hour at 60° in the absence (Lane b-f) or presence (Lane g) of ATP (1 mM). Reaction products were analyzed by electrophoresis on a 1% agarose gel.

RuvB (Figure 15, lane g). ATPγS (1 mM) had the same effect (data not shown).

Figure 16 depicts a competition assay for *T. thermophilus* RuvB binding to dsDNA of equal masses of two different lengths (337 bp and 1079 bp) under the same conditions as Figure 15. The extent of formation of complexes with one or more RuvB molecules was independent of the length of the dsDNA molecules, indicating a preference for binding at the ends of the molecules (see especially lanes e-h). The preference for ends is not exclusive. *T. thermophilus* RuvB will also bind to supercoiled DNA (see Figure 17). The strong binding of *T. thermophilus* RuvB to the ends of linear dsDNA is consistent with the strong stimulation of *T. thermophilus* (and *T. maritima*) RuvB ATPase activity by linear dsDNA.

Unwinding supercoiled DNA by thermostable *T. thermophilus* and *T. maritima* RuvB. Figure 17 demonstrates the effect of thermostable RuvB proteins on topoisomer distribution in relaxed supercoiled dsDNA as determined using one-dimensional (A-B) and two-dimensional (C-E) gels. One- and two-dimensional gels of supercoiled pUC19 DNA after treatment with topo I only are depicted in Figure 17A and 17B, lane a, and Figure 17C, respectively. One positive and one negative topoisomer with less than integral writhing number contributed to the top band (relaxed) on the one-dimensional gels. The additional bands were positively supercoiled with increasing mobility corresponding to increasing writhing number. Addition of *T. thermophilus* or *T. maritima* RuvB without ATP or ATPγS resulted in little change in the topoisomer distribution (Figure 17A and 17B, lane e).

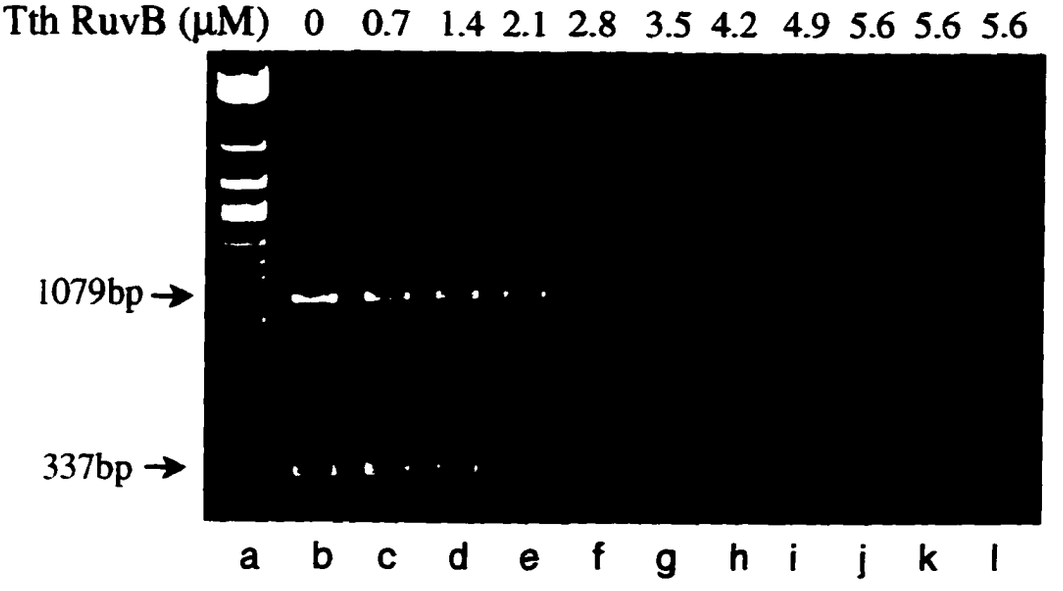


Figure 16. Competitive binding of two different sizes of linear dsDNA with *T. thermophilus* RuvB protein. Lane a, molecular weight markers; 0.2 μg of 337 bp and 0.2 μg of 1079 bp PCR product (lane b-j) or 0.2 μg of 337 bp PCR product alone (lane k) or 0.2 μg of 1079 bp PCR product alone (lane l) were incubated with indicated concentrations of *T. thermophilus* RuvB protein in binding buffer for 1 hour at 60°. Reaction products were analyzed by electrophoresis on a 1% agarose gel.

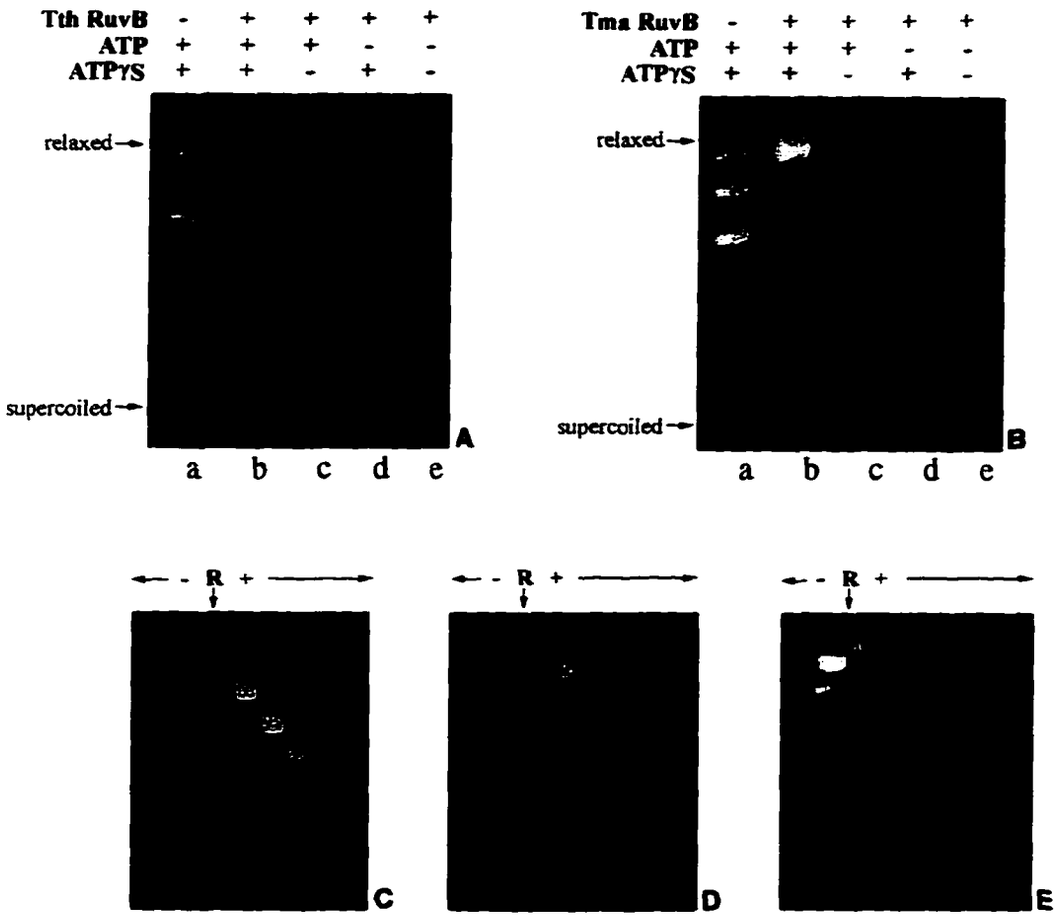


Figure 17. Detection of topoisomers of supercoiled pUC19 DNA following *T. thermophilus* and *T. maritima* RuvB mediated unwinding and topoisomerase I treatment on one- (A, B) and two- (C-E) dimensional agarose gels. Relaxed: writhing number = 0; supercoiled: highly negatively supercoiled. Two-dimensional gels followed incubation with ATP (0.5 mM) and ATPγS (0.5 mM) at 60° for 15 min before topo I treatment. C: no RuvB; D: *T. thermophilus* RuvB; E: *T. maritima* RuvB. First dimension: top to bottom; R = relaxed; sign (+, -) = direction of supercoiling. Second dimension: left to right.

Addition of *T. thermophilus* RuvB and both ATP and ATPγS resulted in a small shift toward negative superhelicity (Figure 17D) and the appearance of numerous additional faint bands representing topoisomers of increasing negative superhelicity (Figure 17A, lane b). Addition of ATP without ATPγS eliminated the small shift toward negative superhelicity but retained the additional faint negative topoisomers (Figure 17A, lane c). Addition of ATPγS without ATP resulted in the same small shift toward negative superhelicity but increased the fraction of highly negatively-supercoiled species to the extent that they represented a major fraction (supercoiled) of topoisomers (Figure 17A, lane d).

Addition of *T. maritima* RuvB and both ATP and ATPγS resulted in a large shift toward negative superhelicity (Figure 17E) and the appearance of numerous additional very faint bands representing topoisomers of increasing negative superhelicity (Figure 17B, lane b). In contrast to *T. thermophilus* RuvB, addition of *T. maritima* RuvB and ATP without ATPγS resulted in little change compared to *T. maritima* RuvB alone (lane c). Addition of ATPγS without ATP resulted in the same large shift toward negative superhelicity and an increased the fraction of more negatively-supercoiled species (lane d). Unlike differences in the effects of cofactors, the apparent differences in the extent of unwinding of superhelical DNA by *T. thermophilus* and *T. maritima* RuvB may be a function of the assay, where the helicase and topo I have different temperature optima.

***T. thermophilus* RuvB can complement *E. coli* RuvB *in vivo*.** In order to determine whether *T. thermophilus* RuvB could substitute for *E. coli* RuvB helicase *in vivo*, the *E. coli ruvBrecG* double mutant strain N3395 was used as a

host for control pUC19 and for pUC19-based plasmids pTth or pEc expressing *T. thermophilus* and *E. coli* RuvB, respectively. Both expression plasmids produced RuvB proteins when induced, although pEc was more efficient (data not shown). Figure 18 depicts the UV sensitivity of these strains. The UV resistance was 5- to 10-fold greater for the pTth strain than for the control pUC19 strain for the UV doses tested, although the strain was still UV-sensitive compared to the pEc strain or to wild type (*ruvb*$^{+}$) *E. coli.*

D. Discussion

In this study, the *ruvB* genes of the highly divergent thermophilic eubacteria *T. thermophilus* and *T. maritima* were cloned, sequenced and expressed in *Escherichia coli*. *T. thermophilus*, *T. maritima*, mesophilic Gram-negative bacteria and mesophilic Gram-positive bacteria are equally divergent. RuvA and RuvB are expressed from a single operon in both Gram-negative (*E. coli* and *Haemophilus influenzae*) and Gram-positive (*M. leprae*) mesophilic bacteria. Surprisingly, the sequence 1000 bp or 470 bp immediately upstream from *T. thermophilus* and *T. maritima ruvB*, respectively, contained no sequences similar to *ruvA*. In fact, a search of Genbank revealed that the translation of the *T. thermophilus* upstream sequence encoded, in the opposite orientation, a protein similar to *E. coli* PanB. The *T. maritima* upstream open reading frame did not show significant identity to any bacterial sequences in Genbank. The absence of linkage of *ruvA* to *ruvB* in the two thermophilic genera may be (i) fortuitous or (ii)

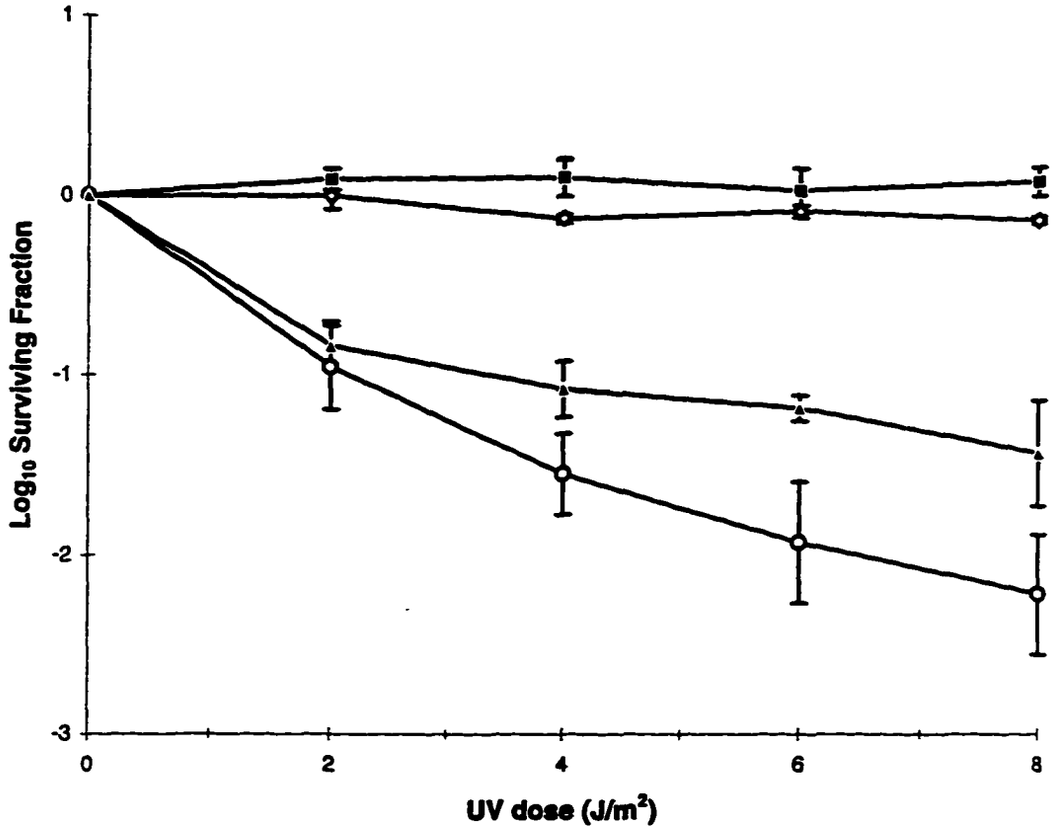


Figure 18. UV survival of *E. coli recG ruvB* mutant strain N3395 containing pUC19 (open circles), pTth expressing *T. thermophilus* RuvB protein (filled triangles), pEc expressing *E. coli* RuvB protein (filled squares) and wild type strain XAO (open diamonds). Each data point is the average of four experiments. The error bars indicate the average ± standard deviation.

may suggest that thermophilic RuvB binding to Holliday junctions is independent of RuvA or (iii) depends on another protein. Given the partial complementation of a *ruvBrecG E. coli* by *T. thermophilus* RuvB, suggesting that *T. thermophilus* RuvB and *E. coli* RuvB have a similar function *in vivo*, the fortuitous option is the most likely possibility. Incomplete complementation was to be expected because of the low helicase activity of *T. thermophilus* RuvB at 37°C. Initial attempts to clone genomic sequences encoding a protein similar to RuvA from several thermophilic bacterial species have been unsuccessful, but may reflect the poorer conservation of RuvA amino acid sequences between highly-divergent genera compared to RuvB amino acid sequences. Additional future efforts will be directed toward cloning genes which express thermophilic RuvA proteins.

T. thermophilus, *T. maritima* and mesophilic RuvB proteins contain conserved regions including GxGKT and DExH motifs, which are shared by several ATP-dependent helicases in *E. coli* (Tsaneva, et al., 1993), consistent with the conservation of the function of these thermostable RuvB proteins. Although *T. thermophilus* and *T. maritima* RuvB showed functional similarity to *E. coli* RuvB in their dsDNA affinity, dsDNA-dependent ATPase activity and ability to unwind supercoiled DNA, *T. thermophilus* and *T. maritima* RuvB showed some unique properties. (i) *E. coli* RuvB ATPase activity is preferentially stimulated by supercoiled DNA (Mitchell and West, 1994; Marrione and Cox, 1995). However, *T. thermophilus* and *T. maritima* RuvB ATPase activities were stimulated as well by linear dsDNA. This result is consistent with our additional observation that *T. thermophilus* RuvB protein has a preference for binding to the ends of linear

dsDNA. Furthermore, increasing the assay temperature increases the pitch of the DNA helix, decreasing the negative superhelicity of the supercoiled DNA and making it more like relaxed circular and linear dsDNA. (ii) *E. coli* RuvB requires ATP or ATPγS for the binding of DNA (Müller, et al., 1993b). However, *T. thermophilus* and *T. maritima* RuvB form complexes with linear DNA which are stable to electrophoresis in the absence of nucleotide cofactors. In fact, adding ATP or ATPγS destabilized *T. thermophilus* RuvB-dsDNA complexes, suggesting that the hydrolysis of ATP might be involved in the dissociation of *T. thermophilus* RuvB from DNA. At 70°C, ATPγS was slowly hydrolyzed by *T. thermophilus* RuvB (data not shown). (iii) With *E. coli* RuvB, transiently-unwound forms of supercoiled DNA could only be detected after incubation and topoisomerase I treatment in the presence of both ATP and ATPγS (Adams and West, 1995). However, with *T. thermophilus* and *T. maritima* RuvB, the greatest unwinding of supercoiled DNA was detected in the presence of ATPγS alone, followed by the unwinding observed in the presence of a mixture of ATP and ATPγS. Studies with additional substrates may help to elucidate the significance of these differences.

E. Acknowledgments

We are grateful to Dr. Barry Rosenstein of this institution for calibration the UV source, to Dr. David Gelfand of Roche Molecular Systems for genomic DNA from *T. aquaticus* and *T. maritima*, to Dr. Jen-i Mao of Genome Therapeutics Corp. for providing the *M. leprae* cosmid containing the *ruv* genes, to Dr. Garry J.

Sharples of the University of Nottingham (U.K.) for providing the *ruvBrecG E. coli* strain N3395, and to Dr. John A. Blaho of this institution for helpful advice on the topological assay and two-dimensional gel electrophoresis and for critical reading of the manuscript. This research was supported by a grant from the National Institutes of Health (HG00446).

IV. Crystal Structure Determination of Thermostable RuvB Proteins

This chapter contains progress to date for a collaboration between our laboratory and the laboratory of Dr. John A. Tainer at the Scripps Research Institute, La Jolla, CA.

A. Introduction

Helicases play an essential role in almost all cellular processes involving nucleic acids such as DNA replication, recombination, repair and transcription. These enzymes act as molecular motors to couple the hydrolysis of nucleoside triphosphate with unwinding of and translocation along DNA. Despite the importance of these enzymes in DNA metabolism, the mechanistic details ranging from nucleotide-binding to helicases to hydrolysis-driven conformational changes are poorly understood (Lohman and Bjornson, 1996). Our best understanding of the structures of most helicases comes from low resolution electron microscopy studies. Understanding the mechanism of helicase function will require detailed information about helicase structure that can only be obtained by x-ray crystallography. The only known atomic structural data for a helicase is the crystal structure of a monomeric form of PcrA from *Bacillus stearothermophilus*, determined both alone and in a complex with ADP (Subramanya, et al., 1996). Little is known about the biochemical properties of this protein. In contrast, *E. coli* RuvB is a well characterized, hexameric DNA helicase. Together with the Holliday junction specific DNA binding protein RuvA, RuvB will catalyze the branch migration of Holliday junctions in a reaction that is coupled to ATP hydrolysis (West, 1996). *E. coli* RuvB is a good candidate for crystal structure studies. Unfortunately, *E. coli* RuvB crystals do not give diffraction data to sufficient resolution to solve the crystal structure (Patnum and Tainer, unpublished data). As reported in the previous chapter, two thermostable RuvB proteins have been

cloned and expressed from two distantly related thermophilic eubacteria, *T. thermophilus* and *T. maritima*. They were shown to be DNA-dependent ATPases like the *E. coli* homolog. *T. thermophilus* RuvB can complement *E. coli* RuvB *in vivo* suggesting it has a similar function in recombination *in vivo* (Tong and Wetmur, 1996). Thermostable proteins often have more compact structure than mesophilic proteins, making thermostable RuvB proteins better candidates for crystal structure studies. In this chapter, we will describe our contributions to an effort to solve the crystal structure of these hexameric DNA helicases. This structure should provide the basis for elucidation of the structural biochemistry of helicases and may contribute toward understanding many genetic diseases which result from DNA helicase defects, including Xeroderma pigmentosum, Cockayne's syndrome, Bloom's syndrome and Werner's syndrome (West, 1996).

B. Materials and methods

Bacterial strains and plasmids. *E. coli* strains BL21(DE3)pLysS and B834(DE3)pLysS (Novagen), with plasmid pET11c-TthRuvB and pET11c-Tma RuvB were used for RuvB expression. PET11c-Tma RuvB was used as template for PCR based mutagenesis. XL1-Blue supercompetent cells were used for transformation of mutated plasmids.

Site-directed mutagenesis of pET11c-Tma RuvB. The PCR based QuickChange™ Site-Directed Mutagenesis Kit (Stratagene) was used to introduce cysteines into *T. maritima* RuvB. Sense primer 5′ CATTGGTCAGGAA**TGC**GT-

GAAAAAGAAACTC 3′and complementary primer 5′ GAGTTTCTTTTTCAC**GCA**-TTCCTGACCAATG 3′were used to generate the *T. maritima* RuvB N32C mutant, and sense primer 5′ CAGGCAGGATTCCTC**TGC**AGAACTCCCAGAGG 3′ and complementary primer 5′ CCTCTGGGAGTTCT**GCA**GAGGAATCCTGCCTG 3′were used to generate the *T. maritima* RuvB A308C mutant. Mutated bases are shown in bold. PCR was performed in an Ericomp Powerblock thermocycler using two complementary primers, 125 ng each, with 10 ng template plasmid pET11c-Tma RuvB and 2.5U *Pfu* DNA polymerase. A denaturation step of 95°C, 30 sec was followed by 16 cycles of 95°C, 30 sec; 55°C, 1 min; 68 °C, 14 min. *Dpn*I digestion and transformation of XL-1 Blue cells were performed as instructed by the manufacturer.

Sequence analysis. Inserts screened for the presence of new *Bsm*I or *Pst*I sites, expected for the cysteine mutations A308C or N32C, respectively, were sequenced using insert-specific and pET11c vector-specific oligodeoxynucleotide primers and the Sequenase DNA Sequencing Kit (U.S. Biochemicals, Inc.).

Expression and purification of wild-type *T. thermophilus* and *T. maritima* RuvB and cysteine-containing *T. maritima* RuvB mutants. Expression and purification to homogeneity of *T. thermophilus* RuvB and both wild-type and mutant *T. maritima* RuvB basically followed the procedures previously described (Chapter III; Tong and Wetmur, 1996). In order to obtain the higher yields necessary for crystallography, BL21(DE3)pLysS cells were substituted for Novablue (DE3)pLysS cells.

RuvB ATPase assay. The ATPase activities of *T. maritima* RuvB

cysteine-containing proteins were measured by colorimetric detection of released P_i as described (LeBel, et al., 1978). Reaction mixtures (0.5 ml) contained 20 µg *T. maritima* RuvB N32C or A308C mutant proteins, 20 mM Tris-HCl, pH 7.8, 10 mM $MgCl_2$ and 1 mM ATP. Some reactions contained 6 µM (nucleotides) ϕX174 single-stranded DNA (ssDNA), or supercoiled pUC19 DNA.

Crystallization of *T. thermophilus* and *T. maritima* RuvB proteins. Crystallization of *T. thermophilus* and *T. maritima* RuvB was performed by our collaborators, Christopher O. Putnam and Dr. John A. Tainer at the Scripps Research Institute, La Jolla, CA.

Expression and purification of selenomethionine-substituted *T. maritima* RuvB protein. The plasmid pET11c-Tma RuvB was transformed into the methionine auxotrophic *E. coli* strain B834(DE3)pLysS [the parent strain of BL21(DE3)]. Cells were grown overnight in M9 medium containing 1 mg/ml thiamine, 40 µg/ml selenomethionine, 50 µg/ml ampicillin, 25 µg/ml chloramphenicol and 5% LB medium. This culture was diluted 1/200 into the same medium without LB and shaken for 16 hours at 37°C before inducing with IPTG at a final concentration of 1 mM. Cells were harvested after an additional 9 hours of growth. Selenomethionine-substituted RuvB protein was purified using the same method employed for wild type *T. maritima* RuvB protein except that the HiTrap Q column step was omitted.

Mass spectrometry. MALDI mass spectra were determined on wild-type and selenomethionine-substituted *T. maritima* RuvB at the Mass Spectrometry Core Facility at the Scripps Research Institute, La Jolla, CA.

C. Preliminary results and discussion

Over expression and purification of *T. thermophilus* and *T. maritima* RuvB proteins. *T. thermophilus* and *T. maritima* RuvB proteins were each expressed in BL21(DE3)pLysS in forty half-liter cultures. In both cases, 20 mg of protein were obtained from 20 liters of culture after purification as previously described (Chapter 3; Tong and Wetmur, 1996).

Crystallization and diffraction studies of the *T. maritima* RuvB. Crystallization of the *T. maritima* RuvB protein was achieved by both the hanging-drop and sitting-drop vapor diffusion techniques. The *T. maritima* RuvB crystallized in several crystal forms depending on the solvents employed. The orthorhombic crystal pictured in Figure 19A was 0.3 mm x 0.3 mm x 0.2 mm. These crystals were in space group $P2_12_12_1$, with unit cell parameters a=117.8 Å, b=134.3 Å, c=139.3 Å, $\alpha = \beta = \gamma = 90°$, and contain a hexamer in the asymmetric unit with an estimated solvent content of 48%. A diffraction data set from 20-2.5 Å was collected by using Siemens SRA anode generator (Figure 20A).

Crystallization and diffraction studies of the *T. thermophilus* RuvB. Two crystal forms were obtained for *T. thermophilus* RuvB. The tetragonal crystal was in the space group $P4_12_12$ (Figure 19B) with unit cell parameters a = b = 86.7 Å, c=362.5 Å, $\alpha = \beta = \gamma = 90°$ and contained three molecules in the asymmetric unit with overall solvent content of 62%. A diffraction data set from 40-3.6 Å was collected at the Cornell High Energy Synchrotron Source (Figure 20B).

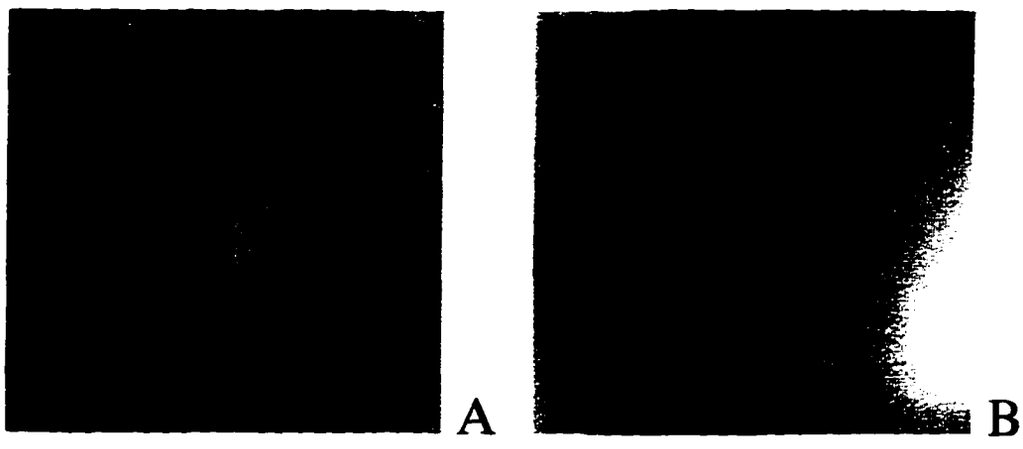


Figure 19. *T. maritima* (A) and *T. thermophilus* (B) RuvB crystals.

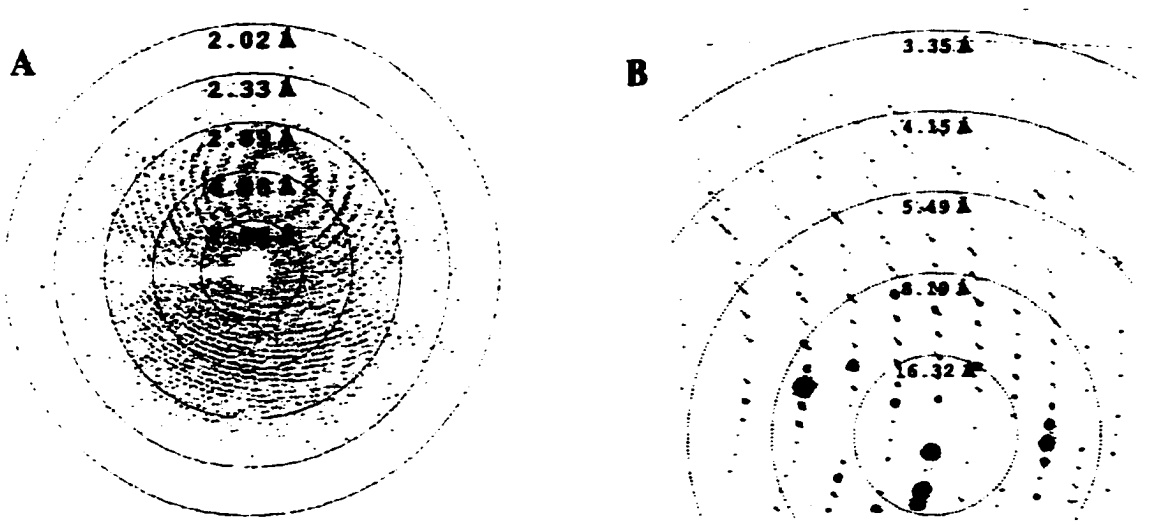


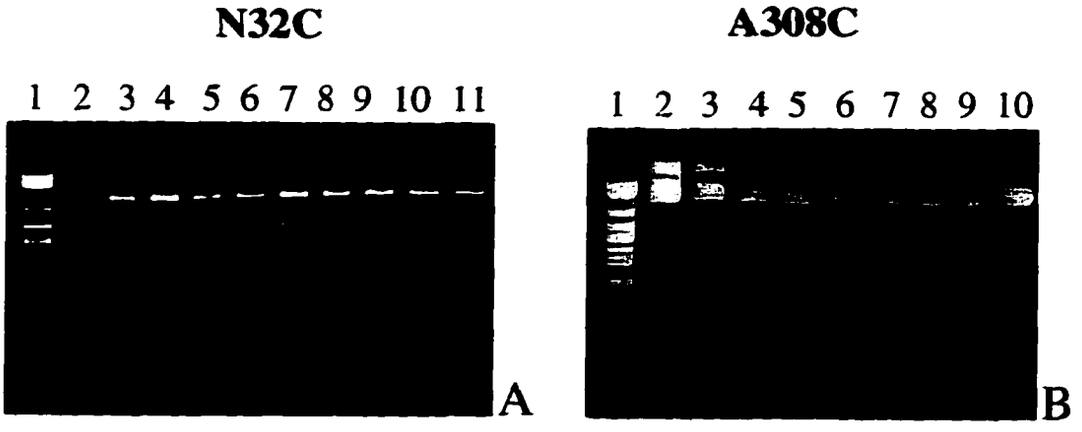
Figure 20. Diffraction data from the *T. maritima* (A) and *T. thermophilus* (B) RuvB crystals.

Designing cysteine mutants for phasing *T. maritima* RuvB diffraction data. Since *T. maritima* RuvB has no cysteines, two *T. maritima* RuvB mutants, N32C and A308C, were designed for derivatization by mercurials by our collaborators in Scripps based on three criteria: 1) secondary structure and solvent accessibility predictions with the program PHD (Rost, et al., 1994); 2) alignment with exposed residues on potentially conserved regions of RecA with the program CKWHENCE (Burns, Scripps, unpublished); 3) sites of residues variation among bacterial RuvB sequence.

Generating N32C and A308C *T. maritima* RuvB mutants. The N32C and A308C mutants were generated by using the PCR based QuickChange™ Site-Directed Mutagenesis Kit (Stratagene). Plasmid pET11c-Tma RuvB only has one *Bsm*I site. Introducing the N32C mutation created a second *Bsm*I site. Thus, *Bsm*I digestion was screen for N32C mutant plasmids. Similarly, pET11c-Tma RuvB only has one *Pst*I site. The A308C mutation created a second *Pst*I site, setting up the same type of screen for A308C mutant plasmids. In the screen for the N32C mutation, all ten independent isolates analyzed by *Bsm*I digestion had the second cleavage site (Figure 21A). Similarly, In the screen for the A308C mutation, all nine independent isolates analyzed by *Pst*I digestion had the second cleavage site (Figure 21B).

N32C and A308C *T. maritima* RuvB mutant proteins are thermostable. Three N32C clones and four A308C clones were transformed into BL21(DE3)pLysS cells. Protein expression and heating was performed as described (Tong and Wetmur, 1996). All the N32C and A308C mutant RuvB

Figure 21. Screening for N32C and A308C *T. maritima* RuvB mutants



*Bsm*I digestion

Lane 1: Marker
Lane 2-11: N32C clones

*Pst*I digestion

Lane 1: Marker
Lane 2-9: A308C clones
Lane 10: pET11c-Tma RuvB

proteins were thermostable (Figure 22).

N32C and A308C *T. maritima* RuvB mutant sequences. Two isolates were selected for each mutant, and the complete inserts were sequenced. Of the two potential N32C mutant plasmids sequenced, both had the expected mutation and neither had a PCR error. Of the two potential A308C mutant plasmids sequenced, both had the expected mutation. One clone had no PCR errors, and the second had a 10 base pair inversion. A thermostable protein had been detected in spite of this inversion.

Expression and purification of N32C and A308C *T. maritima* RuvB mutant proteins. Twenty mg of each mutant protein was purified to homogeneity for crystallization (Figure 23).

Generating selenomethionine-substituted RuvB. In order to improve the initial phases and to identify the positions of the methionines when fitting the structure, selenomethionine-substituted RuvB was generated. Because the B834(DE3)pLysS (pET11c-Tma RuvB) strain cannot recover from stationary phase if selenomethionine is the only source of methionine in the medium, 5% LB medium was added to the initial overnight culture (Yang, et al., 1990). After dilution, subsequent culture involved IPTG induction of *T. maritima* RuvB and incorporation of selenomethionine. The cells grew very poorly in this medium with a doubling time of about 3 hours. Following lysis and heat treatment, the crude protein fraction showed only the RuvB band on SDS-PAGE (Figure 24, lane 2). Therefore, the HiTrap Q step could be omitted from the purification protocol. HiTrap Blue affinity chromatography was performed to remove residual nucleic

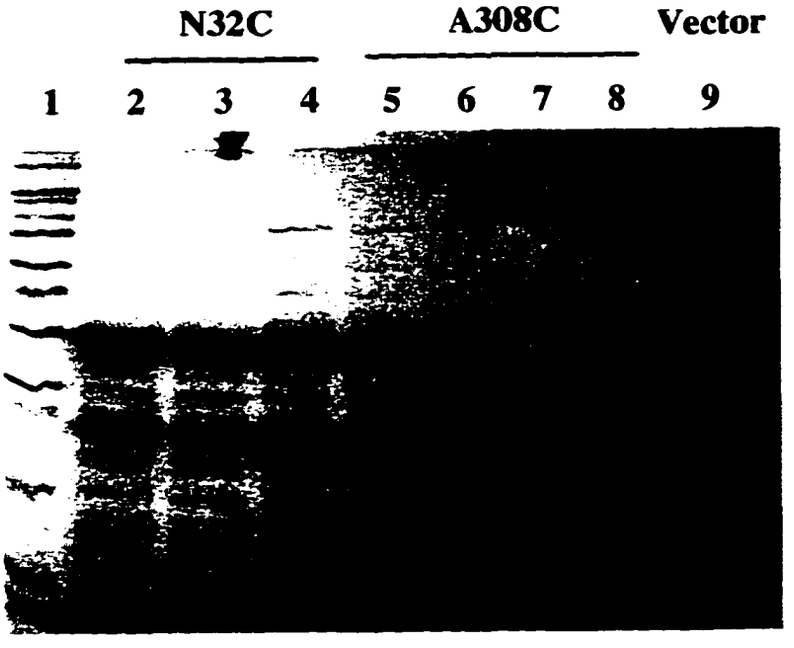


Figure 22. SDS-PAGE analysis of thermostable N32C and A308C *T. maritima* RuvB expression. Lane 1, molecular weight marker; Lane 2-4, initial purification of independently derived expression clones for N32C; Lane 5-8, initial purification of independently derived expression clones for A308C; Lane 9, vector clone.

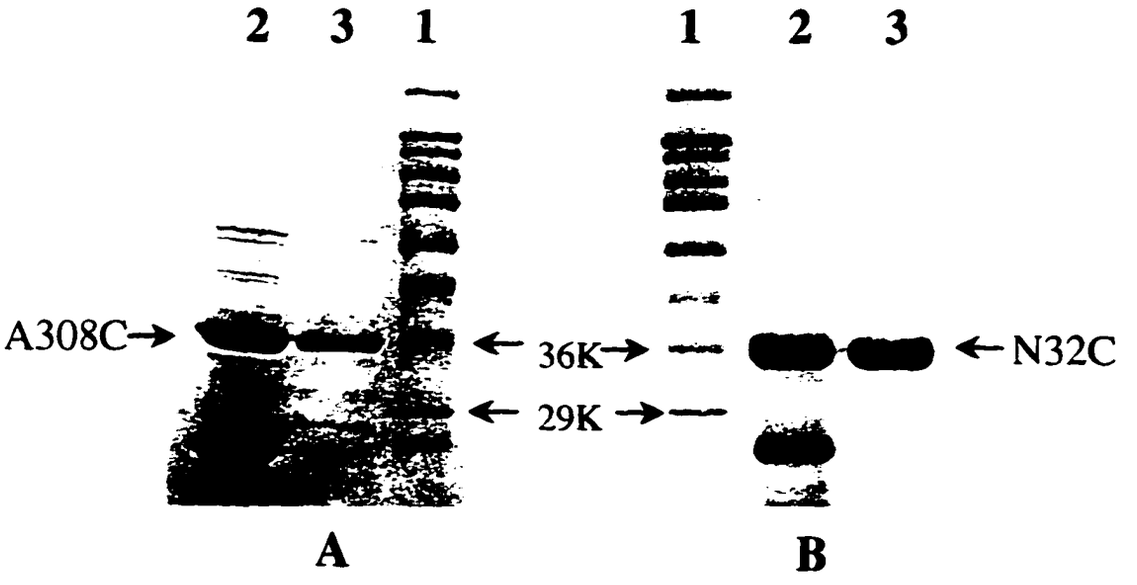


Figure 23. SDS-PAGE depicting purification of A308C (A) and N32C (B) *T. maritima* RuvB mutant proteins. Lane 1, molecular weight marker; Lane 2, fraction after heating at 70°C; Lane 3, fraction after HiTrap Q and HiTrap Blue chromatography. The SDS-PAGE was 12.5% and was stained with Coomassie Brilliant Blue R.

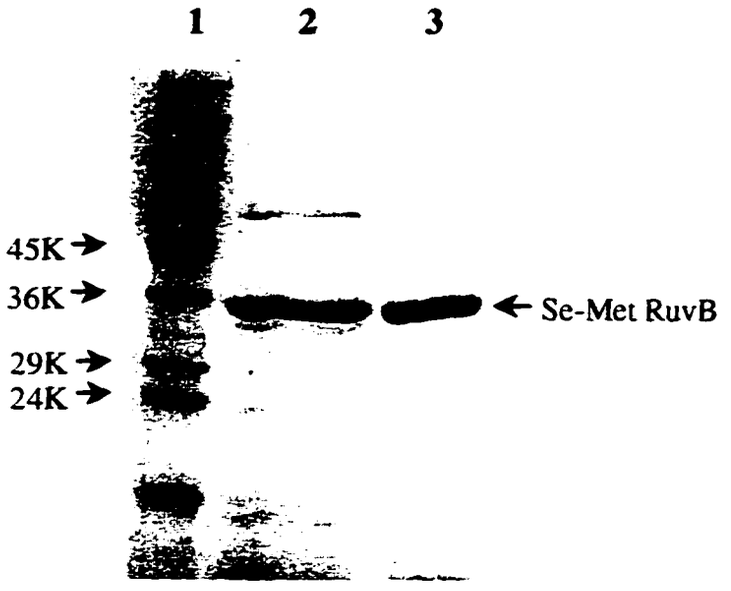


Figure 24. SDS-PAGE depicting purification of selenomethionine-substituted *T. maritima* RuvB protein. Lane 1, molecular weight marker; Lane 2, fraction after heating at 70°C; Lane 3, fraction after HiTrap Blue chromatography. The SDS-PAGE was 12.5% and was stained with Coomassie Brilliant Blue R.

acids. Four mg of selenomethionine-substituted RuvB was purified to homogeneity for crystallization (Figure 24, lane 3).

T. maritima RuvB contains eight methionine residues per monomer. The predicted change in mass of a protein containing eight selenium atoms replacing eight sulfur atoms is 376 daltons (47 daltons/atom). In a MALDI mass spectrum, the wildtype *T. maritima* RuvB protein gave a peak at 36,994 Daltons (Figure 25A). The selenomethionine-substituted RuvB gave a peak at 37,366 Daltons (Figure 25B). The difference is 372 daltons, Therefore, the level of selenomethionine substitution appears to be almost complete. The crystallization of the selenomethionine-substituted *T. maritima* RuvB is underway in the laboratory of our collaborator in Scripps.

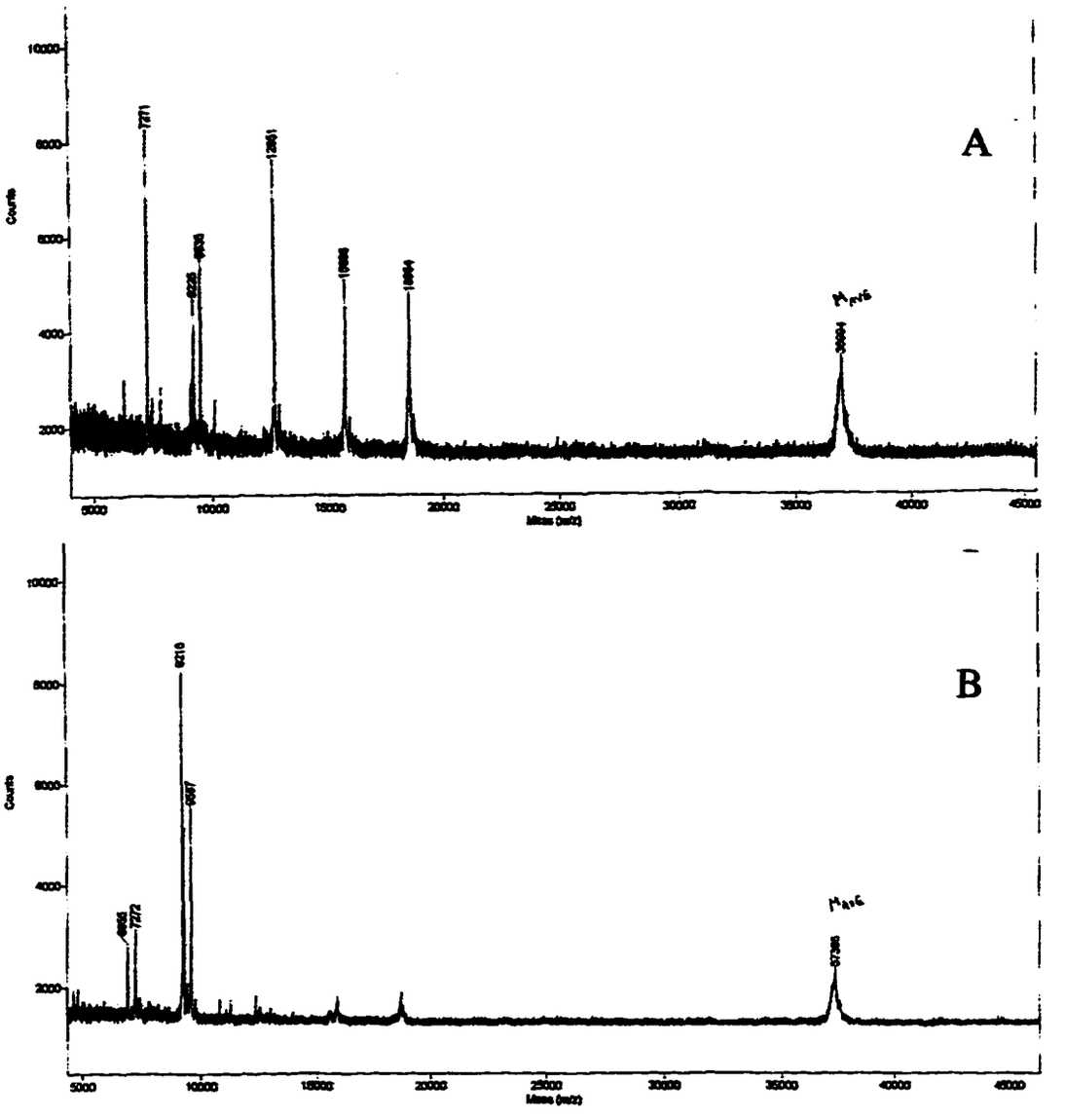


Figure 25. Mass Spectra data of wild-type (A) and selenomethionine-substituted (B) *T. maritima* RuvB proteins. Arrow indicates the location of the peak of interest.

V. Discussion

A. Cloning and analysis of genes encoding thermostable recombination proteins

The same approach was used for cloning both *recA* and *ruvB* genes from thermophiles. The *recA* gene from the thermophilic eubacterium *T. thermophilus* was cloned, sequenced and expressed in *E. coli* . Three other recA genes from distantly-related thermophilic eubacteria (*A. pyrophilus*, *T. aquaticus*, *T. maritima*) have also been cloned by other members of our laboratory (Wetmur, et al., 1994). *T. thermophilus* is closely related to *T. aquaticus*. The *ruvB* genes of the highly divergent thermophilic eubacteria *T. thermophilus* and *T. maritima* also were cloned, sequenced and expressed in *E. coli.* The coding sequences of *T. thermophilus* and *T. maritima* RuvB are shown in Figure 26 and Figure 27, respectively. All cloning was carried out using PCR technology without the need for library construction. Inverse PCR is a rapid method for obtaining sequence data for the 5′- and 3′-flanking regions of bacterial genes, the prerequisite for generation of primers for PCR cloning into an expression vector. Because of the inherent error frequency of *in vitro* DNA replication, care was taken to demonstrate that sequences of independently-derived expression clones were identical.

Attempts to clone *ruvB* from *A. pyrophilus* were unsuccessful. Recently, the complete sequence of the related species *Aquifex VF5* has been completed by Recombinant Biocatalysis, Inc. At our behest, this sequence was searched for

Figure 26. *T. thermophilus ruvB* sequence

```
   1  gggtgccttg gtgtggtgga gcatctcctc caaggtcacg ggcaccgtgg
  51  aggggtagcc cagaccacca tcccgaagga gtccccgacc aaaatcggtc
 101  cacccgcgcc tgcccaggcg ggccgtgggg tagtcgtagg cggtgaggta
 151  gacgaccgct ggcccttggc gttgcggaag tccttcaccg tgcgccgcac
 201  gggaataagg ctatcatgga ggagcGTGGA AGACCTCGCC CTTAGGCCCA
 251  AGACCCTGGA CGAGTACATC GGCCAGGAGC GCCTGAAGCA AAAGCTCCGG
 301  GTCTACCTCG AGGCGGCCAA GGCCCGAAAA GAGCCCTTGG AGCACCTCCT
 351  CCTCTTCGGC CCCCCGGGCC TGGGCAAGAC CACCCTGGCC CACGTCATCG
 401  CCCACGAGCT TGGGGTCAAC CTCCGGGTCA CCTCCGGGCC CGCCATAGAG
 451  AAGCCCGGGG ACCTCGCCGC CATCCTGGCC AACTCCCTGG AGGAAGGGGA
 501  CATCCTCTTC ATTGACGAGA TCCACCGCCT GAGCCGCCAG GCCGAGGAGC
 551  ACCTCTACCC CGCCATGGAG GACTTCGTCA TGGACATCGT CATCGGCCAA
 601  GGTCCGGCGG CGAGGACCAT CCGGCTGGAG CTTCCCCGCT TCGCCCTGAT
 651  CGGGGCCACC ACCCGGCCCG GCCTCATCAC CGCGCCCCTC CTGAGCCGCT
 701  TCGGCATCGT GGAGCACCTG GAGTACTACA CCCCTGAGGA GCTGGCCCAA
 751  GGGGTGATGC GGGACGCCCG CCTCCTCGGG GTGAGGATCA CGGAGGAGGC
 801  CGCCCTGGAG ATCGGTAGGC GGAGCCGGGG CACCATGCGC GTCGCCAAGC
 851  GCCTCTTCCG GAGGGTGCGG GACTTCGCCC AGGTGGAGGG GGAGGAGGTC
 901  ATCACCCGGG AGCGGGCCCT GGAGGCCCTT GCCGCCCTGG GCCTGGACGA
 951  GCTGGGCTTG GAGAAGCGGG ACCGGGAGAT CCTGGAGGTC CTCATCCTCC
1001  GCTTCGGGGC CGGCCCCGTG GGCCTCGCCA CCCTGGCCAC CGCCCTCTCC
1051  GAGGACCCGG GCACCTTGGA AGAGGTGCAC GAGCCCTACC TCATCCGCCA
1101  GGGCCTTTTG AAGCGCACTC CCCGGGGCCG GGTGGCCACG GAACTGGCCT
1151  ACCGCCACCT GGGCTACCCG CCCCCGGTGG GCCCCCTCTT GGAGCCGTGA
1201  tccgggccac gctttccgag ctggacctgg gggagctc
```

G + C CONTENT: 70%

Lowercase characters indicate noncoding sequence; uppercase characters indicate coding sequence; bold characters indicate start codon or stop codon.

Figure 27. *T. maritima ruvB* sequence

```
   1  tcggttttgt gatgggaagc ggtaagtgat cttctttttc aggaaagaaa
  51  agggggcgca agcccccttt tttgttcgag tggtataatt ttttcgagaa
 101  aaaggaaatc agaagaggtg gaaagGTGAG TGAATTTCTC ACACCTGAAA
 151  GGACCGTTTA CGACTCTGGT GTACAGTTTC TAAGGCCCAA AAGCCTCGAT
 201  GAATTCATTG GTCAGGAAAA CGTGAAAAAG AAACTCTCCC TCGCTCTCGA
 251  AGCCGCGAAG ATGAGGGGAG AAGTGCTCGA TCATGTCCTC CTCGCAGGAC
 301  CACCGGGACT CGGAAAGACG ACCCTTGCAC ACATAATCGC CAGCGAACTC
 351  CAGACGAACA TCCACGTTAC GAGCGGACCG GTTCTTGTGA AACAGGGAGA
 401  TATGGCCGCT ATCCTCACAA GTCTGGAACG GGGAGACGTT CTTTTCATAG
 451  ACGAAATACA CCGATTGAAC AAGGCAGTGG AAGAGCTTCT TTACTCTGCC
 501  ATCGAAGACT TCCAGATAGA CATCATGATC GGAAAGGGCC CGAGTGCAAA
 551  GTCCATTAGG ATAGACATCC AGCCTTTTAC GCTCGTTGGA GCCACGACGA
 601  GAAGTGGTCT TTTGAGTTCT CCTCTCAGAA GCAGGTTTGG TATCATCCTC
 651  GAACTGGACT TCTACACTGT GAAAGAACTG AAGGAAATCA TAAAAAGAGC
 701  GGCCAGCTTG ATGGACGTTG AGATAGAAGA CGCAGCAGCA GAGATGATCG
 751  CGAAAAGATC GAGAGGCACA CCGAGGATCG CTATAAGACT CACGAAGAGA
 801  GTGAGGGACA TGCTCACGGT GGTAAAGGCA GACAGAATCA ATACCGATAT
 851  CGTTTTGAAG ACCATGGAAG TTCTGAACAT AGACGACGAG GGACTTGATG
 901  AGTTCGACAG GAAGATCCTG AAGACGATCA TAGAGATTTA CAGGGGAGGC
 951  CCCGTCGGAT TGAACGCCCT CGCCGCTTCA CTCGGTGTAG AAGCGGACAC
1001  CCTGAGCGAA GTTTATGAAC CTTACCTCCT CCAGGCAGGA TTCCTCGCCA
1051  GAACTCCCAG AGGAAGGATC GTCACTGAAA AGGCTTACAA ACACCTGAAG
1101  TACGAAGTCC CGGAAAACCG TCTTTTCTGA ggtgatctcc catgggattg
1151  aaagaaaacc tcgaaagggt tctcaacaga atgaaaaacg ctgctcttcg
1201  agcaaacagg gatccatccg aagtgaggct cgtagttgct tcaaaatacg
```

G + C CONTENT: 50%

Lowercase characters indicate noncoding sequence; uppercase characters indicate coding sequence; bold characters indicate start codon or stop codon.

*ruv*genes. Surprisingly, none of the *ruv* genes were found. Perhaps *Aquifex* uses a different system, such as the recG system, for migration and resolution of Holliday structures.

RuvA and RuvB are expressed from a single operon in both Gram-negative (*E. coli*, *Haemophilus influenzae*, *Pseudomonas aeruginosa*, and *Borrelia burgdorferi*) and Gram-positive type (*M. leprae* and *Mycoplasma pneumoniae*) mesophilic bacteria. The sequences 1.2 Kb or 1.0 Kb immediately upstream from *T. thermophilus* or *T. maritima ruvB*, respectively, contained no sequences similar to *ruvA*. In fact, a search of Genbank revealed that the translation of the *T. thermophilus* upstream sequence encoded, in the opposite orientation, a protein similar to *E. coli* PanB. The *T. maritima* upstream open reading frame did not show significant identity to any bacterial sequences in Genbank. The absence of linkage of *ruvA* to *ruvB* in the two thermophilic genera (i) may be fortuitous or (ii) may suggest that thermophilic RuvB binding to Holliday junctions is independent of RuvA or depends on another protein. Given the partial complementation of a *ruvBrecG E. coli* by *T. thermophilus* RuvB, suggesting that *T. thermophilus* RuvB and *E. coli* RuvB have a similar function *in vivo*, the fortuitous option is the most likely possibility. The fact that the absence of linkage of *ruvA* to *ruvB* was also observed in the cyanobacterium *Synechocystis sp.* further supports the fortuitous option.

B. Analysis of thermostable RecA and RuvB sequences

The relationship of the sequence of *T. thermophilus* RecA to other RecA sequences was illustrated in Figures 8 and 10. Figure 28 shows the pileup of RuvB amino acid sequences of *T. thermophilus*, *T. maritima*, a mesophilic Gram-negative bacterium (*E. coli*) and a mesophilic Gram-positive type bacterium (*M. leprae*). They are equally divergent.

A phylogenetic analysis of 16S rRNA from many species including *A. pyrophilus* and *T. maritima* has indicated that *A. pyrophilus* was the deepest rooted in the eubacterial tree (Burggraf, et al., 1992). To compensate for G+C% bias, especially in certain thermostable species such as *T. aquaticus* and *T. thermophilus*, an additional tree was constructed using only transversions because the purine content of 16S rRNA is nearly constant. The result was the same. The phylogenetic analysis of the thermostable RecA and RuvB proteins depicted in Figures 11 and 12, respectively, indicated a midpoint somewhere in the vicinity of the branchpoints of the (i) Gram-positive and (ii) Gram-negative mesophilic eubacteria, (iii) cyanobacteria, and (iv) thermophilic bacteria. This analysis, indicating a single branch for the thermophiles, may be complicated by the impossibility of separating drift in the RecA or RuvB amino acid sequences from selection for amino acid substitutions conferring thermostability. Nevertheless, among the thermophilic bacteria, *T. maritima* and not *A. pyrophilus* was the most deeply rooted in the RecA tree. Because *T. maritima* has a G+C% of 47%, no correction for codon bias was needed.

Figure 28. Pileup of RuvB sequences

```
          1                                                  50
Ruvbeco   ..MIEADRLI SAGTTLPEDV ADRA...IRP KLLEEYVGQP QVRSQMEIFI
Ruvbmle   MSEDYLDRDV SPALTVGEAD IDVS...LRP RSLREFIGQP RVREQLQLVI
Ruvbtth   .......... .........V EDLA...LRP KTLDEYIGQE RLKQKLRVYL
Ruvbtma   .........V SEPLTPERTV YDSGVQFLRP KSLDEFIGQE NVKKKLSLAL

          51                                                 100
Ruvbeco   KAAKLRGDAL DHLLIFGPPG LGKTTLANIV ANEMGVNLRT TSGPVLEKAG
Ruvbmle   EGAKNRGATP DHILLSGPPG LGKTSLAMII AAELGSSLRM TSGPALERAG
Ruvbtth   EAAKARKEPL EHLLLFGPPG LGKTTLAHVI AHELGVNLRV TSGPAIEKPG
Ruvbtma   EAAKMRGEVL DHVLLAGPPG LGKTTLAHII ASELQTNIHV TSGPVLVKQG

          101                                                150
Ruvbeco   DLAAMLTN.L EPHDVLFIDE IHRLSPVVEE VLYPAMEDYQ LDIMIGEGPA
Ruvbmle   DLAVMLSN.L VEHDVLFIDE IHRIARPAEE MLYLAMEDFR VDVIVGKGPG
Ruvbtth   DLAAILANSL EEGDILFIDE IHRLSRQAEE HLYPAMEDFV MDIVIGQGPA
Ruvbtma   DMAAILT.SL ERGDVLFIDE IHRLNKAVEE LLYSAIEDFQ IDIMIGKGPS

          151                                                200
Ruvbeco   ARSIKIDLPP FTLIGATTRA GSLTSPLRDR FGIVQRLEFY QVPDLQYIVS
Ruvbmle   ATSIPLEVAP FTLVGATTRS GALTGPLRDR FGFTAHMDFY EPTELEGVLA
Ruvbtth   ARTIRLELPR FALIGATTRP GLITAPLLSR FGIVEHLEYY TPEELAQGVM
Ruvbtma   AKSIRIDIQP FTLVGATTRS GLLSSPLRSR FGIILELDFY TVKELKEIIK
```

```
          201                                                  250
Ruvbeco   RSARFMGLEM SDDGALEVAR RARGTPRIAN RLLRRVRDFA EVKHDGTISA
Ruvbmle   RAAGILGIEL GVEAGAEIAR RSRGTPRIAN RLLRRVRDFA EVRADGVITR
Ruvbtth   RDARLLGVRI TEEAALEIGR RSRGTMRVAK RLFRRVRDFA QVEGEEVITR
Ruvbtma   RAASLMDVEI EDAAAEMIAK RSRGTPRIAI RLTKRVRDML TVVKADRINT

          251                                                  300
Ruvbeco   DIAAQALDML NVDAEGFDYM DRKLLLAVID KFFGGPVGLD NLAAAIGEER
Ruvbmle   DVAKAALAVY DVDELGLDRL DRAVLSALTR SFGGGPVGVS TLAVAVGEEA
Ruvbtth   ERALEALAAL GLDELGLEKR DREILEVLIL RFGAGPVGLA TLATALSEDP
Ruvbtma   DIVLKTMEVL NIDDEGLDEF DRKILKTIIE IYRGGPVGLN ALAASLGVEA

          301                                                  350
Ruvbeco   ETIEDVLEPY LIQQGFLQRT PRGRMATTRA WNHFGITPPE MP*.......
Ruvbmle   TTVEEVCEPF LVRAGMVART PRGRVATAQA WTYLCMTPPV GVTGLSQPGL
Ruvbtth   GTLEEVHEPY LIRQGLLKRT PRGRVATELA YRHLGYPPPV GPLLEP*...
Ruvbtma   DTLSEVYEPY LLQAGFLART PRGRIVTEKA YKHLKYEVPE NRLF*.....

          351
Ruvbeco   ....
Ruvbmle   FES*
Ruvbtth   ....
Ruvbtma   ....
```

The larger number of known mesophilic RecA sequences compared to RuvB sequences permitted more detailed analysis of RecA. In spite of substantial sequence divergence, interesting common characteristics could be distinguished among the four thermostable RecA amino acid sequences. Of these proteins, only *A. pyrophilus* contained C, and the two C residues were not located at a site of C residues in mesophilic RecA proteins. The common absence of C residues in proteins of the genus *Thermus* has been observed in DNA polymerase I and DNA ligase. Selection against C may result from the lability of C residue at high temperature (Olcott, 1951). A comparison of the amino acid composition of *A. pyrophilus* and *T. maritima* RecA proteins with those of a representative set of 27 complete mesophilic RecA sequences in Swissprot revealed three types of significant changes. First, among the charged amino acids, a decrease in D was compensated by an increase in E, and K was increased. The increase in charged residues correlates with the observation that the formation of additional surface ion pairs is important for thermotability (Perutz and Raidt, 1975; Walker, et al., 1980; Kelly, et al.,1993). The substitution of E for D may reflect increased potential for salt bridge formation related to the longer side chain in E. These changes are opposite to those obtained from a comparison of many mesophilic and thermophilic enzymes (Zuber, 1988), suggesting that care must be taken in interpreting substitutions in one enzyme (Kelly, et al., 1993). Second, among the polar amino acids, N and the hydroxy-amino acids S and T were decreased. Third, among the hydrophobic amino acids, there was a decrease in A and a large increase in V. A large increase in V has been described in several analyses

(Zuber, 1988; Menéndez-Arias and Argos,1989). Substitution of valine for alanine has been found to increase the thermostability of enzymes such as α-amylase from *Bacillus licheniformi* (Declerck, et al., 1995). This stabilization may result from the increased hydrophobilicity of V compared to A and the larger side chains which could increase intra- and interhelical compactness. *A. pyrophilus* and *T. maritima* are encoded by sequences with G+C% 48 and 47, respectively. The codons for D, E, V, S and T show no G+C% bias, whereas the codons for K, N and A begin with AA, AA and GC, respectively. *T. aquaticus* and *T. thermophilus* are encoded by sequences with G+C% 67 and 70, respectively. A comparison of their amino acid composition with the same mesophilic RecA proteins again revealed the decrease in D compensated by an increase in E, the decrease in N, S and T, and the large increase in V. The increase in positively charged amino acids occurred through a statistically significant increase in R, with a preference for high %G+C, rather than K, especially in *T. thermophilus* RecA. The decrease in A, encoded by GCN, disappeared. *T. aquaticus* and *T. thermophilus* showed an increase in P and L, which have large and small %G+C codon biases, respectively, and large decreases in M and I, which are encoded by ATN. Overall, after eliminating the effects of codon bias, the major trends were loss of C, substitution of E for D, increased positively-charged amino acid content, decreased polar N, S and T, and increased hydrophobic P and especially V content.

Another global effect of the substitutions found in the thermophilic organisms was an increase in the calculated isoelectric point compared to that of

mesophilic RecA and RuvB proteins by greater than a full pH unit. Table 1 lists representative values for RecA isoelectric points of mesophilic and thermophilic organisms at 37°C. Table 2 shows comparable results for RuvB isoelectric points. It is likely that the apparent increase in isoelectric point would be different if the calculations were made at the temperatures of maximum enzymatic activity.

C. Attempts to clone thermostable RuvA proteins

RuvA is a relatively small protein. *E. coli* (Gram negative) and *M. Leprae* (Gram positive type organism) RuvA exhibit 32% amino acid identity in 205 amino acids of overlap compared to 55% identity over 309 amino acids of RuvB. Degenerate PCR primers were synthesized based on the conserved amino acid sequences (1) VVRED (forward), (2) PGIGK (reverse) and (3) GVGPK (reverse) at *E. coli* RuvA positions 52, 114, and 80, respectively. The primer design was identical to that we have used previously for cloning RecA and RuvB from diverse genera of thermophiles. Oligonucleotides 1 and 2 primed larger PCR products from *T. thermophilus* and *T. maritima* inconsistent with the expected size. Oligonucleotides 1 and 3 primed a band of the anticipated size from *T. maritima*. These products were cloned and several were sequenced. None of these sequences was consistent with RuvA. Other approaches that could be employed to clone *T. maritima* or *T. thermophilus ruvA* are (i) screening a genomic DNA libraries using the *E. coli ruvA* gene as a probe at low-stringency hybridization and (ii) complementation, using plasmid genomic expression libraries to transform the

Table 2

Calculated isoelectric points of RuvB proteins

	Mesophilic species	**pI**	
Gram-negative:	*Escherichia coli*	4.9	
	Hemophilus influenza	4.9	Average = 5.0
	Pseudomonas aeruginosa	5.3	
Gram-positive:	*Mycobacterium leprae*	5.0	
	Thermophilic species		
Gram-negative:	*Thermotoga maritima*	6.4	Average = 6.3
	Thermus thermophilus	6.2	

UV sensitive *recGruvA E. coli* strain N3397 (Lloyd, 1991), with a screen for UV resistant clones. The latter method will only work if there is has residual RuvA activity at 37 °C. *T. thermophilus* RuvB retained activity at 37 °C, but *T. maritima* RuvB did not.

D. Structural studies of thermostable RuvB helicases

The Stratagene PCR-based QuickChange site-directed mutagenesis kit was chosen to generate *T. maritima* RuvB cysteine mutants N32C and A308C. Traditional site-directed mutagenesis methods generally require single-stranded DNA as templates, thus necessitating recloning of the gene into M13-based bacteriophage vectors. The QuickChange site-directed mutagenesis protocol allows the introduction of site-specific mutations directly into any double-stranded plasmid of up to 10 Kb. PCR is performed using higher fidelity Pfu DNA polymerase which does not displace the mutant oligonucleotide primers. Newly synthesized PCR products are selected by GATC methylation-dependent *Dpn*I digestion that cleaves parental DNA templates but not the unmethylated PCR products. This protocol generated N32C and A308C mutants with greater than 90% efficiency. After sequencing the *ruvB* gene inserts of two N32C clones and two A308C clones, we did not observe any PCR errors in three of the four clones, but one clone has a 10 base pair inversion. The flanking region around the 10 base pair fragment had the potential to form a stable hairpin structure, which might have caused the PCR error through a strand-switching mechanism. Based on our

experience, the method appears robust and should be useful for generating additional mutants for studying the active sites of RuvB helicases and the sites involved in hexamer formation.

The *T. maritima* RuvB cysteine mutant N32C protein was thermostable and had dsDNA-dependent ATPase activity. The *T. maritima* RuvB cysteine mutant A308C protein was also thermostable, but lacked dsDNA-dependent ATPase activity. Both mutant proteins have been crystallized under conditions similar to those used for as the wild-type protein. Preliminary experiments suggest that the A308C mutant is proficient for hexamer assembly (Putnam and Tainer, unpublished data). Thus, the loss of dsDNA-dependent ATPase activity is more likely due to an altered active site conformation than to a gross alteration in protein conformation or folding. Additional studies showed that the N32C mutant, but not the A308C mutant, accelerated the cleavage of Ellman's reagent, placing the N32C cysteine on the exterior of the protein and suggesting that the A308C cysteine was buried.

The fact that *T. thermophilus* RuvB partially complemented the UV sensitivity of *E. coli ruvBrecG* strain N3395 *in vivo*, but *T. maritima* RuvB did not (data not shown), correlates with the observation that *T. thermophilus* had residual ATPase activity at 37°C (Figure 14A) while *T. maritima* RuvB did not (Figure 14B). The absence of a low-temperature ATPase activity in *T. maritima* RuvB may be useful for freezing reaction intermediates in *T. maritima* RuvB cocrystals with substrates and/or cofactors.

VI. Bibliography

Abdel-Monem, M., Durwald, H., and Hoffmann-Berling, H. (1976): DNA unwinding enzyme II of *Escherichia coli*.2. Characterization of the DNA unwinding activity. Eur. J. Biochem. 65, 441-449.

Abrahams, J. P., Leslie, A. G. W., Lutter, R., and Walker, J. E. (1994): Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. Nature 370, 621-628.

Adams, D. E., and West, S. C. (1995): Unwinding of closed circular DNA by the *Escherichia coli* RuvA and RuvB recombination/repair proteins. J. Mol. Biol. 247, 404-417.

Al-Deib, A. A., Mahdi, A. A., and Lloyd, R. G. (1996): Modulation of recombination and DNA repair by the RecG and PriA helicases of *Escherichia coli* K-12. J. Bacteriol. 178, 6782-6789.

Angov, E., and Camerini-Otero, R. D. (1994): The *recA* gene from the thermophile *Thermus aquaticus* YT-1: cloning, expression, and characterization. J. Bacteriol. 176, 1405-1412.

Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994): Atomic structure of the RuvC resolvase: A Holliday junction-specific Endonulease from E. coli. Cell 78: 1063-1072.

Bennett, R. J., Dunderdale, H. J., and West, S. C. (1993): Resolution of Holliday junctions by RuvC resolvase: cleavage specificity and DNA distortion. Cell 74, 1021-1031.

Benson, D., Lipman, D.J., and Ostell, J. (1993): Genbank. Nucleic Acids Res. 21, 2963-2965.

Benson, F. E., Illing, G.T., Sharples, G. J., and Lloyd, R.G. (1988): Nucleotide sequencing of the *ruv* region of *Escherichia coli* K-12 reveals a LexA regulated operon encoding two genes. Nucleic Acids Res. 16, 1541-1549.

Benson, F. E., Collier, S., and Lloyd, R. G. (1991): Evidence of abortive recombination in ruv mutants of Escherichia coli K12. Mol. Gen. Genet. 225, 266-272.

Benson, F. E., Stasiak, A., and West, S. C. (1992): Purification and characterization of the human Rad51 protein, an analogue of *E. coli* RecA. EMBO J. 13, 5764-5771.

Bianchi, M. E., and Radding, C. M. (1983): Insertions, deletions and mismatches in heteroduplex DNA made by recA protein. Cell 35, 511-520.

Bishop, D. K., Park, D., Xu, L., and Kleckner, N. (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.

Blaho, J. A., Larson, J. E., McLean, M. J., and Wells, R. D. (1988): Multiple DNA secondary structures in perfect inverted repeat inserts in plasmids. J. Biol. Chem. 263, 14446-14455.

Burggraf, S., Olsen, G. J., Stetter, K.O., and Woese, C. R. (1992): A phylogenetic analysis of *Aquifex pyrophilus*. System. Appl. Microbiol. 15, 352-356.

Camerini-Otero, R. D., and Hsieh, P. (1993): Parallel DNA triplexes, homologous

recombination, and other homology-dependent DNA interactions. Cell 73, 217-223.

Cerutti, H., Osman, M., Grandoni, P., and Jagendorf, A. T. (1992): A homolog of *Escherichia coli* RecA protein in plastids of higher plants. Proc. Natl. Acad. Sci. U. S. A. 89, 8068-8072.

Clark, A. J., Margulies, A. D. (1965): Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12. Proc. Natl. Acad. Sci. U. S. A. 53, 451-459.

Connolly, B., Parson, C. A., Benson, F. E., Dunderdale, H. J., Sharples, G. J., Lloyd, R. G., and West, S. C. (1991): Resolution of Holliday junctions in vitro requires the Escherichia coli ruvC gene product. Proc. Natl. Acad. Sci. U. S. A. 88, 6063-6067.

Craig, N., and Roberts, J. W. (1980): *E. coli* recA protein-directed cleavage of phage lambda repressor requires polynucleotide. Nature 283, 26-30.

Declerck, N., Joyet, P., Trosset, J. Y., Garnier, J., and Gaillardin, C. (1995): Hyperthermostable mutants of *Bacillus licheniformis* alpha-amylase: multiple amino acid replacements and molecular modelling. Protein Engineering 8, 1029-1037.

Devereux, J., Haelberti, P., and Smithies, O. (1984): A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12, 387-395.

Dicapua, E., Schnarr, M., Ruigrok, R. W., Lindner, P., Timmins, P. A. (1990): Complexes of RecA protein in solution. A study by small angle neutron

scattering. J. Mol. Biol. 214, 557-70.

Dunderdale, H. J., Benson, F. E., Parsons, C. A., Sharples, G. J., Lloyd, R. G., and West, S. C. (1991): Formation and resolution of recombination intermediates by E. coli RecA and RuvC proteins. Nature 354, 506-510.

Duwat, P., Ehrlich, S. D., and Gruss, A. (1992): A general method for cloning recA genes of gram-positive bacteria by polymerase chain reaction. J. Bacteriol. 174, 5171-5175.

Dybvig, K., Hollingshead, S. K., Heath, D.G., Clewell, D. B., Sun, F., and Woodard, A. (1992): Degenerate oligonucleotide primers for enzymatic amplification of *recA* sequences from gram-positive bacteria and mycoplasmas. J. Bacteriol. 174, 2729-2732.

Felsenstein, J. (1989): PHYLIP -- Phylogeny Inference Package (Version 3.2). Cladistics 5, 164-166.

Ferrin, L. J., and Camerini-Otero, R. D. (1991): Selective cleavage of human DNA: RecA-assisted restriction endonuclease (RARE) cleavage. Science 254, 1494-1497.

Fleischman, R. D., Adams, M. D., White, O., Clayton, R.A., Kirkness, E. F., Kerlavage, A. R., Bult, C.J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L. I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L.,

M[illegible],
J. [illegible]
Hae[illegible]s infl[illegible]

Golub, E. I., W[illegible]
protein [illegible]
elongatio[illegible]

Habu, T., Taki, T., We[illegible]
human homo[illegible]
recombination [illegible]
transcript in meio[illegible]

Heuser, J., and Griffith, J. (1[illegible] Visualiz[illegible]
with DNA by Quick-[illegible]deep[illegible]
210, 473-484.

Hiom, K., and West, S. C. ([illegible] Bra[illegible]
recombination: assembly [illegible]AB-H[illegible]
Cell 80, 787-793.

[illegible]onigberg, S. M., Rao, B. J., and Ra[illegible] M[illegible]
to promote a search for rare se[illegible]e in d[illegible]
Sci. U. S. A. 83, 9586-9590.

P., Camerini-Otero, C. S., and Ca[illegible]Otero [illegible]
homologous DNA sequences by p[illegible]
[illegible]NA. Genes Dev. 4, 1951-1963.

[illegible]amerini-Otero, C. S., and Camerini-[illegible]. (1992)

McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., and Venter, J. C. (1995): Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269, 496-512.

Golub, E. I., Ward, D. C., and Radding, C. M. (1992): Joints formed by RecA protein from oligonucleotides and duplex DNA block initiation and elongation of transcription. Nucleic Acids Res. 20, 3121-3125.

Habu, T., Taki, T., West, A., Nishimune, Y., and Morita, T. (1996): The mouse and human homologs of DMC1, the yeast meiosis-specific homologous recombination gene, have a common unique form of exon-skipped transcript in meiosis. Nucleic Acid Res. 24, 470-477.

Heuser, J., and Griffith, J. (1989): Visualization of RecA protein and its complexes with DNA by Quick-freeze/deep-etch electron microscopy. J. Mol. Biol. 210, 473-484.

Hiom, K., and West, S. C. (1995): Branch migration during homologous recombination: assembly of a RuvAB-Holliday junction complex in vitro. Cell 80, 787-793.

Honigberg, S. M., Rao, B. J., and Radding, C. M. (1986): Ability of RecA protein to promote a search for rare sequences in duplex DNA. Proc. Natl. Acad. Sci. U. S. A. 83, 9586-9590.

Hsieh, P., Camerini-Otero, C. S., and Camerini-Otero, R. D. (1990): Pairing of homologous DNA sequences by proteins: evidence for three-stranded DNA. Genes Dev. 4, 1951-1963.

Hsieh, P., Camerini-Otero, C. S., and Camerini-Otero, R. D. (1992): The synapsis

event in the homologous pairing of DNAs: RecA recognizes and pairs less than one helical repeat of DNA. Proc. Natl. Acad. Sci. U. S. A. 89, 6492-6496.

Iwasaki, H., Shiba, T., Makino, K., Nakata, A., and Shinagawa, H. (1989): Overproduction, purification, and ATPase activity of the *Escherichia coli* RuvB protein involved in DNA repair. J. Bacteriol. 171, 5276-5280.

Iwasaki, H., Takahagi, M., Shiba, T., Nakata, A., and Shinagawa, H. (1991): Escherichia coli RuvC protein is an endonuclease that resolves the Holliday structure. EMBO. J. 10, 4381-4389.

Iype, L. E., Wood, E. A., Inman, R. B., and Cox, M. M. (1994): RuvA and RuvB proteins facilitate the bypass of heterologous DNA insertions during RecA protein-mediated DNA strand exchange. J. Biol. Chem. 269, 24967-24978.

Jain, S. K., Inman, R. B., and Cox, M. M. (1992): Putative three-stranded DNA pairing intermediate in RecA protein-mediated DNA strand exchange: no role for guanine N-7. J. Biol. Chem. 267, 4215-4222.

Jwang, B., and Radding, C. M. (1992): Torsional stress generated by RecA protein during DNA strand exchange separates strands of a heterologous insert. Proc. Natl. Acad. Sci. U. S. A. 89, 7596-7600.

Kahn, R., Cunningham, R. P., DasGupta, C., Radding, C. M. (1981): Polarity of heteroduplex formation promoted by Escherichia coli recA protein. Proc. Natl. Acad. Sci. U. S. A. 78, 4786-4790.

Kato, R., and Kuramitsu, S. (1993): RecA protein from the extremely thermophilic

bacterium, *Thermus thermophilus* HB8. J. Biochem. (Tokyo). 114, 926-929.

Kelly, C. A., Nishiyama, M., Ohnishi, Y., Beppu, T., and Birktoft, J. J. (1993): Determinants of protein stability observed in the 1.9 Angstrom crystal structure of malate dehydrogenase from the thermophilic bacterium *Thermus flavus*. Biochemistry 32, 3913-3922.

Kido, M., Yoneda, Y., Nakanishi, M., Uchida, T., and Okada, Y. (1992): *Escherichia coli* RecA protein modified with a nuclear location signal binds to chromosomes in living mammalian cells. Exp. Cell Res. 198, 107-114.

Kilpatrick, M. W., Klysik, J., Singleton, C. K., Zarling, D. A., Jovin, T. M., Hanau, L. H., Erlanger, B. F., and Wells, R. D. (1984): Intervening sequences in human fetal globin genes adopt left-handed Z helices. J. Biol. Chem. 259, 7268-7274.

Kmiec, E., and Holloman, W. K. (1982): Homologous pairing of DNA molecules promoted by a protein from Ustilago. Cell 29, 367-374.

Kmiec, E.B., Cole, A., and Holloman, W. K. (1994): The REC2 gene encodes the homologous pairing protein of Ustilago maydis. Mol. Cell Biol. 14, 7163-7172.

Kmiec, E. B., and Holloman, W. K. (1994): DNA strand exchange in the absence of homologous pairing. J. Biol. Chem. 269, 10163-10168.

Koob, M., Burkiewicz, A., Kur, J., and Szybalski, W. (1992): RecA-AC: single-site cleavage of plasmids and chromosomes at any predetermined restriction

site. Nucleic Acids Res. 20, 5831-5836.

Kowalczykowski, S. C., Krupp, R. A. (1987): Effect of Escherichia coli SSB protein on the single-stranded DNA-dependent ATPase activity of Escherichia coli RecA protein. Evidence that SSB protein facilitate the binding of RecA protein to regions of secondary structure within single-stranded DNA. *J. Mol. Biol.* 193, 97-113.

Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D., and Rehrauer, W. M. (1994): Biochemistry of homologous recombination in *Escherichia coli.* Microbiol. Rev. 58, 401-465.

Kowalczykowski, S. C., Eggleston, A. K. (1994): Homologous pairing and DNA strand exchange proteins. Annu. Rev. Biochem. 63, 991-1043.

Lawyer, F. C., Stoffel, S., Saiki, R. K., Chang, S-Y., Landre, P. A., Abramson, R. D., and Gelfand, D. H. (1993): High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. PCR. Methods. Appl. 2, 275-287.

Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H. (1989): Isolation, characterization and expression in *Escherichia coli* of the DNA-polymerase gene from *Thermus aquaticus*. J. Biol. Chem. 264, 6427-6437.

LeBel, D., Poirier, G. G., and Beaudoin, A. R. (1978): A convenient method for the ATPase assay. Anal. Biochem. 85, 86-89.

Lindsley, J. E., Cox, M. M. (1989): Dissociation pathway for RecA nucleoprotein

filaments formed on linear duplex DNA. J. Mol Biol. 205, 695-711.

Little, J. W. (1984): Autodigestion of LexA and phage lamda repressors. Proc. Natl. Acad. Sci. U. S. A. 81, 1375-1379.

Lloyd, R. G., Benson, F. E., and Shurvinton, C. E. (1984): Effect of ruv mutations on recombination and DNA repair in Escherichia coli K12. Mol. Gen. Genet. 194, 303-309.

Lloyd, R. G. (1991): Conjugational recombination in resolvase-deficient *ruvC* mutants of *Escherichia coli* K-12 depends on *recG*. J. Bacteriol. 173, 5414-5418.

Lohman, T. M., and Bjornson, K. P. (1996): Mechanisms of Helicase-catalyzed DNA unwinding. Annu. Rev. Biochem. 65, 169-214.

Mahdi, A. A., Sharples, G .J., Mandal, T. N., and Lloyd, R. G. (1996): Holliday junction resolvases encoded by homologous *rusA* genes in Escherichia coli K-12 and phage 82. J. Mol. Biol. 257, 561-573.

Mandal, T. N., Mahdi, A. A., Sharples, G. J., and Lloyd, R. G. (1993): Resolution of Holliday intermediates in recombination and DNA repair: indirect suppression of *ruvA*, *ruvB*, and *ruvC* mutations. J. Bacteriol. 175, 4325-34.

Marrione, P. E., Cox, M. M. (1995): RuvB protein-mediated ATP hydrolysis: Functional asymmetry in the RuvB hexamer. Biochemistry 34, 9809-9818.

Matthews, B. W. (1993): Structural and genetic analysis of protein stability. Annu. Rev. Biochem. 62, 139-160.

Menéndez-Arias, L., and Argos, P. (1989): Engineering protein thermal stability: sequence statistics point to residue substitutions in α-helices. J. Mol. Biol. 206, 397-406.

Mitchell, A. H., and West, S. C. (1994): Hexameric rings of *Escherichia coli* RuvB protein. Cooperative assembly, processivity and ATPase activity. J. Mol. Biol. 243, 208-215.

Müller, B, Tsaneva, I. R., and West, S. C. (1993a): Branch migration of Holliday junctions promoted by the *Escherichia coli* RuvA and RuvB proteins. I. Comparison of RuvAB- and RuvB-mediated reactions. J. Biol. Chem. 268, 17179-17184.

Müller, B., Tsaneva, I. R., and West, S. C. (1993b): Branch migration of Holliday junctions promoted by the *Escherichia coli* RuvA and RuvB proteins. II. Interaction of RuvB with DNA. J. Biol. Chem. 268, 17185-17189.

Müller, B., and West, S. C. (1994): Processing of Holliday junctions by the *Escherichia coli* RuvA, RuvB, RuvC and RecG proteins. Experientia. 50, 216-222.

Muniyappa, K., Shaner, S. L., Tsang, S. S., Radding, C. M. (1984): Mechanism of the concerted action of recA protein and helix-destabilizing proteins in homologous recombination. Proc. Natl. Acad. Sci. U. S. A. 81, 2757-2761.

Nohmi, T., Battista, J. R., Dodson, L. A., Walker, G. C. (1988): RecA-mediated cleavage activates UmuD for mutagenesis: mechamistic telationship between transcriptional derepression and posttranslational activation.

Proc. Natl. Acad. Sci. U. S. A. 85, 1819-1820.

Noji, H., Yasuda, R., Yoshida, M., and Kinosita Jr, K. (1997): Direct observation of the rotation of F_1-ATPase. Nature 386, 299-302.

Ochman, H., Medhora, M. E., Garza, D., and Hartl, D. L. (1990): Amplification of flanking sequences by inverse PCR. In: PCR Protocols. A Guide to Methods and Applications. (Eds: Innis,MA; Gelfand,DH; Sninsky,JJ; White,TJ) Academic Press, Inc, San Diego, 219-227.

Ogawa, T., Yu, X., Shinohara, A., and Egelman, E. H. (1993): Similarity of the yeast RAD51 filament to the bacterial RecA filament. Science 259, 1896-1899.

Olcott, H. S. (1951): Methods for the determination of amino acids. In: Greenberg, D. M. (ed) Amino acids and proteins - theory, methods, application. 56-114

Oshima, T., and Imahori, K. (1974): Physiochemical properties of deoxyribonucleic acid from an extreme thermophile. J. Biochem. (Tokyo). 75, 179-183.

Otsuji, N., Iyehara, H., Hideshima, Y. (1974): Isolation and characterization of an Escherichia coli ruv mutant which forms nonseptate filaments after low doses of ultraviolet light irradiation. J. Bacteriol. 117, 337-344.

Panyutin, I. G., and Hsieh, P., (1993): Formation of a single base mismatch impedes spontaneous DNA branch migration. J. Molec. Biol. 230, 413-424.

Panyutin, I. G., and Hsien, P. (1994): The kinetics of spontameous DNA branch migration. Proc. Natl. Acad. Sci. USA. 91, 2021-2025.

Parsons, C. A., Stasiak, A., Bennett, R. J., and West, S. C. (1995): Structure of a multisubunit complex that promotes DNA branch migration. Nature 374, 375-378.

Parsons, C. A., Tsaneva, I., Lloyd, R. G., and West, S. C. (1992): Interaction of Escherichia coli RuvA and RuvB proteins with synthetic Holliday junctions. Proc. Natl. Acad. Sci. U. S. A. 89, 5452-5456.

Perutz, M. F., and Raidt, H. (1975): Stereochemical basis of heat stability in bacterial ferrodoxins and in haemoglobin A2. Nature 255, 256-259.

Radding, C. M. (1991): Helical interactions in homologous pairing and strand exchange driven by RecA protein. J. Biol. Chem. 266, 5355-5358.

Rafferty, J. B., Sedelnikova, S. E., Hargreaves, D., Artymiuk, P. J., Baker, P. J., Sharples, G. J., Mahdi, A. A., Lloyd, R. G., and Rice, D. W. (1996): Crystal structure of DNA recombination protein RuvA and a model for its binding to the Holliday junction. Science 274: 415-421.

Rao, B. J., Dutreix, M., and Radding, C. M. (1991): Stable three-stranded DNA made by RecA protein. Proc. Natl. Acad. Sci. U. S. A. 88, 2984-2988.

Register, J. C. 111, Griffith, J. (1985): The direction of RecA protein assembly onto single strand DNA is the same as the direction of strand assimilation during strand exchange. J. Biol. Chem. 260, 12308-12312.

Rigas, B., Welcher, A. A., Ward, D. C., Weissman, S. M. (1986): Rapid plasmid library screening using RecA-coated biotinylated probes. Proc. Natl. Acad. Sci. U. S. A. 83, 9591-9595.

Rost, B., Sander, C., and Schneider R. (1994): PHD-an automatic mail server for

protein secondary structure prediction. CABIOS 10, 53-60.

Rubin, B. P., Ferguson, D. O., and Holloman, W. K. (1994): Structure of REC2, a recombinational repair gene of Ustilago maydis, and its function in homologous recombination between plasmid and chromosomal sequences. Mol. Cell Biol. 14, 6287-6296.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988): Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487-491.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989): Molecular cloning, a laboratory manual. 2nd ed. Cold Spring Harbor University Press, Cold Spring Harbor.

Sancar, A., Stachelek, C., Konigsberg, W., and Rupp, W. D. (1980): Sequences of the recA gene and protein. Proc. Natl. Acad. Sci. U.S.A.77, 2611-2615.

Sharples, G. J., Benson, F. E., Illing, G. T., and Lloyd, R. G. (1990): Molecular and functional analysis of the ruv region of Escherichia coli K-12 reveals three genes involved in DNA repair and recombination. Mol. Gen. Genet. 221, 219-226.

Sharples, G. J., and Lloyd, R. G. (1991): Resolution of Holliday junctions in Escherichia coli: identification of the *ruvC* gene product as a 19-kilodalton protein. J. Bacteriol. 173, 7711-7715.

Sharples, G. J., Chan, S. N., Mahdi, A. A., Whitby, M. C., and Lloyd, R. G. (1994):

Processing of intermediates in recombination and DNA repair: identification of a new endonuclease that specifically cleaves Holliday junction. EMBO J. 13, 6133-6142.

Shiba, T., Iwasaki, H., Nakata, A., and Shinagawa, H., (1991): SOS-inducible DNA repair proteins, RuvA and RuvB, of Escherichia coli: functional interactions between RuvA and RuvB for ATP hydrolysis and renaturation of the cruciform structure in supercoiled DNA. Proc. Natl. Acad. Sci. U. S. A. 88, 8445-8449.

Shiba, T., Iwasaki, H., Nakata, A., and Shinagawa, H. (1993): *Escherichia coli* RuvA and RuvB proteins involved in recombination repair: Physical properties and interactions with DNA. Mol. Gen. Genet. 237, 395-399.

Shinagawa, H., Iwasaki, H., Kato, T., Nakata, A. (1988a): RecA protein-dependent cleavage of UmuD protein and SOS mutagenesis. Proc. Natl. Acad., Sci. U. S. A. 85, 1806-1810.

Shinagawa, H., Makino, K., Amemura, M., Kimura, S., Iwasaki, H., and Nakata, A. (1988b): Structure and regulation of the *Escherichia coli ruv* operon involved in DNA repair and recombination. J. Bacteriol. 170, 4322-4329.

Shinagawa, H., and Iwasaki, H. (1995): Molecular mechanisms of Holliday junction processing in Escherichia coli. Adv. Biophys., 31, 49-65.

Shinohara, A., Ogawa, H., and Ogawa, T. (1992): Rad51 protein involved in repair and recombination in S. cerevisiae is a RecA-like protein. Cell 69, 457-470.

Stasiak, A., Tsaneva, I. R., West, S. C., Benson, C. J. B., Yu, X., and Egelman, E.

H (1994): The *Escherichia coli* RuvB branch migration protein forms double hexameric rings around DNA. Proc. Natl. Acad. Sci. U. S. A. 91, 7618-7622.

Stacey, K. A., Lloyd, R. G., (1976): Isolation of *rec*-mutants from an F-prime merodiploid strain of *Escherichia coli* K-12. Mol. Gen. Genet. 143, 223-232.

Story, R. M., Bishop, D. K., Kleckner, N., and Steitz, T. A. (1993): Structural relationship of bacterial RecA proteins to recombination proteins from bacteriophage T4 and yeast. Science 259, 1892-1896.

Story, R. M., and Steitz, T. A. (1992): Structure of the RecA protein-ADP complex. Nature 355, 374-376.

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

Subramanya, H. S., Bird, L. E., Brannigan, J. A., and Wigley, D. B. (1996): Crystal structure of a DEXX box DNA helicase. Nature 384, 379-383.

Sung, P. (1994): Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast Rad51 protein. Science 265, 1241-1243.

Takahagi, M., Iwassaki, H., Nakata, A., and Shinagawa, H. (1991): Molecular analysis of the *Escherichia coli ruvC* gene, which encodes a Holliday junction-specific endonuclease. J. Bacteriol. 173, 5747-5753.

Thelen, M. P., Onel, K., and Holloman, W. K. (1994): The REC1 gene of Ustilago maydis involved in the cellular response to DNA damage encodes an exonuclease. J. Biol. Chem. 269, 747-754.

Tong, J., and Wetmur, J. G. (1996): Cloning, sequencing, and expression of *ruvB* and characterization of RuvB proteins from two distantly related thermophilic eubacteria. J. Bacteriol. 178, 2695-2700.

Tsaneva, I. R., Muller, B., and West, S. C. (1992): ATP-dependent branch migration of Holliday junctions promoted by the RuvA and RuvB proteins of E. coli. Cell 69, 1171-1180.

Tsaneva, I. R., Muller, B., and West, S. C. (1993): RuvA and RuvB proteins of *Escherichia coli* exhibit DNA helicase activity in vitro. Proc. Natl. Acad., Sci. U. S. A. 90, 1315-1319.

Tsang, S. S., Muniyappa, K., Azhderian, E., Gonda, D. K., Radding, C. M., Flory, J., and Chase, J. W. (1985): Intermediates in homologous pairing promoted by RecA protein. Isolation and characterization of active presynaptic complexes. J. Mol. Biol. 185, 295-309.

Walker, G. C. (1985): Inducible DNA repair system. Annu. Rev. Biochem. 54, 425-458.

Walker, J. E., Wonacott, A. J., and Harris, J. I. (1980): Heat stability of a tetrameric enzyme, D-glyceraldehyde-3-phosphate dehydrogenase. Eur. J. Biochem. 108, 581-586.

Washington, M. T., Rosenberg, A. H., Griffin, K., abd Studier, F. W. (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.

West, S. C., Cassuto, E., Howard-Flanders, P. Heteroduplex formation by recA

protein: polarity of strand exchange (1981): Proc. Natl. Acad. Sci. U. S. A. 78, 6149-6153.

West, S. C., Cassuto, E., Howard-Flanders, P. (1982): Postreplication repair in *E. coli*: strand exchange reactions of gapped DNA by RecA protein. Mol.Gen.Genet. 187, 209-217.

West, S. C. (1992): Enzymes and molecular mechanisms of genetic recombination. Annu. Rev. Biochem. 61, 603-640.

West, S. C. (1994): The processing of recombination intermediates: Mechanistic insights from studies of bacterial proteins. Cell 76, 9-15.

West, S. C. (1995): Formation, translocation and resolution of Holliday junctions during homologous genetic recombination. Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences 347, 21-25.

West, S. C. (1996): DNA helicases: New breeds of translocating motors and molecular pump. Cell 86, 177-180.

Wetmur, J. G., Wong, D. M., Ortiz, B., Tong, J., Reichert, F., and Gelfand, D. H. (1994): Cloning, sequencing, and expression of RecA proteins from three distantly related thermophilic eubacteria. J. Biol. Chem. 269, 25928-25935.

Whitby, M. C., Ryder, L., and Lloyd R. G. (1993): Reverse branch migration of Holliday junctions by RecG protein: a new mechanism for resolution of intermediates in recombination and DNA repair. Cell 75, 341-350.

Whitby, M. C., and Lloyd, R. G. (1995): Branch migration of three-strand

recombination intermediates by RecG, a possible pathway for securing exchanges initiated by 3′-tailed duplex DNA. EMBO J. 14, 3302-3310.

Wozniak, J. A., Zhang, X.-J., Wilson, K., Weaver, L. H., Tronrud, D. E., Pjura, P. E., Nicholson, H., Matsumura, M., Karpusas, M., Jacobson, R., Faber, R., Dao-pin, S., Alber, T., and Matthews, B. W. (1990): Crystallographic and genetic approaches toward design of proteins of enhanced thermostability. In: Crystallographic and Modeling Methods in Molecular Design. (Eds: Bugg,CE; Ealick,SE) Springer-Verlag, New York, 80-94.

Yang, W., Hendrickson, W. A., Kalman, E. T., and Crouch, R. J. (1990): Expression, purification, and crystallization of natural and selenomethionyl recombinant ribonuclease H from *Escherichia coli.* J. Biol. Chem. 265, 13553-13559.

Zuber, H. (1988): Temperature adaptation of lactate dehydrogenase. Structural, functional and genetic aspects. Biophys. Chem. 29, 171-179.

Zülli, F., Schneiter, R., Urfer, R., and Zuber,H (1991): Engineering thermostability and activity of lactate dehydrogenases from bacilli. Biol. Chem. Hoppe-Seyler 372, 363-372.