J. Biochem. **125**, 939-946 (1999)

Molecular Cloning of Mouse p47, a Second Group Mammalian RuvB DNA Helicase-Like Protein: Homology with Those from Human and *Saccharomyces cerevisiae*[1]

Takashi Gohshi,* Midori Shimada,* Shigeru Kawahire,† Norio Imai,* Tohru Ichimura,* Saburo Omata,* and Tsuneyoshi Horigome*,[2]

* *Department of Biochemistry, Faculty of Science, and* †*Course of Biosystem Science, Graduate School of Science and Technology, Niigata University, 2-Igarashi, Niigata 950-2181*

Received January 25, 1999; accepted February 8, 1999

A 47k protein (p47) in a high-salt buffer extract of a rat liver nuclear matrix fraction was purified by means of a wheat germ agglutinin affinity column, reversed phase HPLC, and SDS-PAGE, and partial amino acid sequences were analyzed. Based on these sequences, the mouse cDNA of the protein was cloned and sequenced, and its amino acid sequence was deduced. Mouse p47 consists of 463 amino acid residues with a molecular weight of 51,112. The amino acid sequences of human and *Saccharomyces cerevisiae* p47s were also deduced from the nucleotide sequences of "expressed sequence tag" fragments and genomic DNA, respectively. These sequences contain helicase motifs and show homology to bacterial RuvB DNA helicases acting in homologous recombination. They also show homology with the putative mammalian helicases p50/TIP49 and RUVBL1. Comparison of the amino acid sequences of p47 group proteins and those of p50/TIP49 group proteins revealed the p47 group proteins to comprise a group distinct from the p50/TIP49 proteins. Ultracentrifugation and gel filtration analyses showed that p47 in the rat liver cytosol fraction exists as large complexes of 697k.

Key words: cDNA cloning, DNA helicase, p47, RuvB, RUVB-like protein.

Homologous genetic recombination promotes genetic diversity in individual organisms and plays an important role in the repair of various types of lethal DNA damage. The activities of many proteins are required for recombination. In bacteria, for instance, RecA, with the assistance of single-stranded DNA-binding protein, promotes strand exchange with a homologous duplex and creates a four-strand intermediate, the Holliday junction. This junction is then translocated by RuvA and RuvB through branch migration and resolved by RuvC. RuvA is a structure-specific DNA-binding protein that binds with the greatest affinity to Holliday junctions and facilitates the binding of RuvB to these junctions. RuvB is a DNA helicase that forms large hexameric rings. The rings of RuvB face each other across a junction and promote a novel dual helicase action that "pumps" DNA through the RuvAB complex, using the energy provided by ATP hydrolysis. A third protein, RuvC, resolves the Holliday junction by introducing nicks into the two DNA strands (*1, 2*).

Studies on eukaryotic recombination proteins are still in their infancy. However, the identification and characterization of the Rad51 proteins from yeast (*3, 4*), frog (*5*), and humans (*6, 7*), which are homologous to *Escherichia coli* RecA, indicated a remarkable structural and functional conservation throughout evolution. It is also likely that higher organisms have proteins equivalent to RuvA, RuvB, and RuvC.

Recently, putative eukaryotic homologues of RuvB were isolated by various methods. Kanemaki *et al.* isolated a 49k TATA-binding protein-interacting protein (TIP49) from a rat liver nuclear extract and cloned its cDNA (*8*). We isolated the same protein (rp50) from a wheat germ agglutinin (WGA) bound fraction of a rat liver nuclear matrix extract and cloned its cDNA (*9*). A human homologue (RUVBL1) cDNA clone was isolated with a yeast two-hybrid system using the 14k subunit of human replication protein A as the bait (*10*). All these proteins have DNA helicase motifs (*11, 12*), and show high amino acid sequence homology with bacterial RuvB (*8-10, 13*). The homology of the amino acid sequences of these eukaryotic proteins and prokaryotic RuvBs strongly suggests that these homologues act in genetic recombination and the repair of DNA damage in eukaryotes in a manner similar to bacterial RuvB. Furthermore, the interaction of the rp50/TIP49 protein with the TATA-binding protein (*8*), and the existence of RUVBL1 in the RNA polymerase II holoen-

[1] This work was supported by Grants-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan, and by a grant from the Biodesign Research Project.
[2] To whom correspondence should be addressed. Phone: +81-25-262-6160, Fax: +81-25-262-6165, E-mail: thori@sc.niigata-u.ac.jp
Abbreviations: bp, base pair(s); EST, expressed sequence tag; GlcNAc, *N*-acetylglucosamine; ORF, open reading frame; PCR, polymerase chain reaction; RT-PCR, reversed transcription-polymerase chain reaction; RUVBL1, RuvB-like protein 1; RUVBL2, RuvB-like protein 2; p47, 47k RuvB-like protein; p50, 50k RuvB-like protein; SDS-PAGE, SDS-polyacrylamide gel electrophoresis; TIP49, 49k TATA-binding protein-interacting protein; WGA, wheat germ agglutinin.

zyme complex (*10*), suggest that these proteins play some role in transcription.

In this paper, we report the identification of another RuvB-like protein in the wheat germ agglutinin (WGA)-Sepharose bound fraction of a rat liver nuclear matrix extract, and the molecular cloning of the mouse homologue cDNA. This protein has a DNA-helicase motif and shows amino acid sequence homology to bacterial RuvB. However, the protein is classified as a member of a group of proteins distinct from the above eukaryotic RuvB-like proteins. As in the case of bacterial RuvBs, the protein exists in the extract as a large complex.

EXPERIMENTAL PROCEDURES

Buffers and Solvents—Buffer A: 10 mM Tris-HCl buffer, pH 7.4, containing 0.2 mM $MgCl_2$; buffer B: 50 mM Tris-HCl buffer, pH 7.2, containing 0.1 mM $CaCl_2$, 500 mM NaCl, and 0.2 mM PMSF; buffer C: 50 mM Tris-HCl buffer, pH 7.2, containing 0.1 mM $CaCl_2$, 1% Triton X-100, and 0.2 mM PMSF; solvent A: 60% (v/v) formic acid; and solvent B: 33% (v/v) isopropanol in 60% (v/v) formic acid.

Preparation of the Rat Liver Nuclear Matrix—Rat liver nuclei were isolated from fasting rats by the method established by Blobel and Potter (*14*), with the exception that all buffers were supplemented with proteinase inhibitors: 1 mM phenylmethanesulfonyl fluoride, 2 mM benzamidine, 10 μg/ml of leupeptin, and 5 μg/ml each of antipain, chymostatin, and pepstatin A. The nuclear matrix was prepared from the nuclei as described by Kita *et al.* (*15*). Briefly, isolated nuclei were suspended in 50 mM Tris-HCl buffer, pH 7.4, containing 0.25 M sucrose, 5 mM $MgSO_4$, and proteinase inhibitors. The suspension was treated with DNase I and RNase A (final 125 μg/ml of each enzyme) at 4°C for 2 h, and then centrifuged at 800×g for 15 min. A "nuclear matrix fraction" was obtained in the form of a pellet. This fraction contained nuclear membrane proteins and insoluble nuclear matrix proteins.

Preparation of a WGA-Bound Fraction—The nuclear matrix fraction suspended in buffer A containing 500 mM NaCl, 4% 2-mercaptoethanol, and proteinase inhibitors was incubated at 4°C for 30 min, and then centrifuged at 10,000×g for 15 min. The supernatant was designated the "salt extract fraction" of the nuclear matrix. The fraction was dialyzed against buffer B and then applied to a WGA-Sepharose column (Seikagaku, Tokyo) equilibrated with buffer B. The column was washed with buffer B and then buffer C containing 1 M NaCl. Proteins bound to the column were eluted with buffer B containing 500 mM *N*-acetylglucosamine (GlcNAc). The eluate thus obtained was designated as the "WGA-bound fraction" of the nuclear matrix salt extract. For determination of the Stokes' radius and sedimentation coefficient of the rat p47 (rp47) monomer, rp47 was eluted from the WGA-Sepharose column with buffer C containing 1 M urea. The eluted fraction ("urea-eluted fraction") was used for the above analyses in the absence of $MgCl_2$ after concentration on a small DEAE-cellulose column. rp47 in the fraction obtained in the absence of $MgCl_2$ exists as a monomer.

Preparation of rp47 for Partial Amino Acid Sequencing—Reversed-phase HPLC in 60% formic acid was carried out as described previously (*16*). The WGA-bound fraction was mixed with a 1/9 volume of 10% SDS and a 1/20 volume of 2-mercaptoethanol, and then incubated at 4°C for 30 min. Then, the solution was mixed with 1.5 volumes of formic acid diluted with an equal volume of solvent A, and centrifuged at 6,500×g for 10 min. The supernatant was applied to a Poros 10R1 column (0.46×8 cm) (PerSeptive Biosystems, USA), and eluted with a 75-min linear gradient of 0 to 33% (v/v) isopropanol in solvent A at a flow rate 0.5 ml/min. The fractions containing rp47 were combined and then subjected to 8% SDS-polyacrylamide gel electrophoresis (SDS-PAGE). For internal amino acid sequences, the gel was stained with Coomassie Brilliant Blue R-250, and gel slices containing rp47 (about 28 μg) were subjected to lysylendopeptidase digestion as described previously (*17*). The slices containing peptides were centrifuged after homogenization. The supernatant was collected, concentrated by lyophilization, and then applied to a reversed-phase HPLC equipped with a Silica-base C8 column (4.6× 250 mm, Capcel Pak C8 column; Shiseido, Tokyo). Peptides were eluted with a linear gradient of 5-75% acetonitrile containing 0.1% trifluoroacetic acid at a flow rate of 0.5 ml/min. The sequences of the isolated peptides were determined with a Protein Sequencer 470A (Applied Biosystems, USA).

cDNA Cloning and Sequencing of Mouse p47 (mp47)—A search of expressed sequence tag (EST) databases revealed that the six partial amino acid sequences matched parts of the deduced amino acid sequences of seven mouse clones (GenBank accession numbers AA726368, AA547270, AA144822, D77893, AA117260, AA895357, and AA529352). These clones seem to encode regions of the mp47 cDNA. Therefore, on the basis of these nucleotide sequences, oligonucleotide primers were synthesized for the cloning of mp47 cDNA: N, 5′-TTCCGCTGAGGCATTGGCGG-3′ as a 5′ primer; and C, 5′-AAGGGCTGGGGTGTCAGGTC-3′ as a 3′ primer (C is complementary). Then the reverse transcription-polymerase chain reaction (RT-PCR) was carried out with mouse liver total RNA using these oligonucleotide primers. The 1.4 kbp PCR product was cloned into the Bluescript II SK(−) plasmid vector, and its sequence was determined by the dideoxy chain termination method (*18*).

Gel Filtration and Density-Gradient Centrifugation—The Stokes' radius of the rp47 complex was determined with the WGA-bound fraction on a Superose 6 HR 10/30 column (Pharmacia Biotech, Sweden), 1×30 cm, equilibrated with 50 mM Tris-HCl buffer, pH 7.2, containing 500 mM NaCl, 2.5 mM $MgCl_2$, and 0.2% Emulgen 109P (Kao, Tokyo). The column was eluted at 4°C at a flow rate of 0.4 ml/min. The Stokes' radii of the marker proteins used to calibrate the column were as follows: thyroglobulin, 8.6 nm; catalase, 5.23 nm; bovine serum albumin, 3.62 nm; and ovalbumin, 2.82 nm. The Stokes' radius of the rp47 monomer was determined on a Superdex 200 HR 10/30 column (Pharmacia Biotech, Sweden) with a urea-eluted fraction in 20 mM Tris-HCl buffer, pH 8.0, containing 500 mM NaCl and 0.2% Emulgen 109P.

The sedimentation coefficients of the rp47 complex and the monomer were determined by glycerol density-gradient centrifugation. The WGA-bound and urea-eluted fractions were loaded onto 4 ml 15-40% glycerol gradients in 50 mM Tris-HCl buffer, pH 7.2, containing 500 mM NaCl, 2.5 mM $MgCl_2$, and 0.2% Emulgen 109P for the complex, and 20 mM Tris-HCl buffer, pH 8.0, containing

500 mM NaCl and 0.2% Emulgen 109P for the monomer; centrifuged at 110,000 × *g* for 18 h; fractionated; and then analyzed by SDS-PAGE and silver staining. The sedimentation coefficients of the marker proteins used were as follows: thyroglobulin, 20 S; catalase, 11 S; bovine serum albumin, 4.4 S; ovalbumin, 3.7 S; and carbonic anhydrase, 3.0 S.

RESULTS

Purification and Partial Amino Acid Sequencing of rp47—A high salt extract fraction of rat liver nuclear matrix contains peripheral membrane proteins of nuclear envelopes, nuclear pore complex proteins, and nuclear matrix proteins. The GlcNAc-bearing and GlcNAc protein-interacting proteins in this fraction were bound to a WGA-affinity column and eluted with GlcNAc. For partial amino acid sequence analysis, rp47 was separated from the WGA-bound fraction by reversed phase HPLC in 60% formic acid, and then subjected to SDS-PAGE (Fig. 1). The rp47 protein band on the gel was excised and digested with lysylendopeptidase. The peptides generated were purified by C8 reversed-phase HPLC and their amino acid sequences were determined. The six partial amino acid sequences obtained are shown in Table I.

Molecular Cloning of the cDNA of mp47—Searches of nucleic acid databases revealed that the six partial amino acid sequences of rp47 match parts of the deduced amino acid sequences of seven clones of mouse-EST. Based on these EST sequences, two oligonucleotide primers for PCR were synthesized. Using these primers, a nucleotide fragment approximately 1.4 kbp in length was obtained by RT-PCR toward mouse liver total RNA. The nucleotide sequence was determined by the dideoxy chain termination

TABLE I. **Amino acid sequences of peptides obtained by lysylendopeptidase digestion of rp47.**

Peptide	Sequence
1	I AGRAVLTAGQPGTGKT
2	TTEMETI YDL
3	LGRSFTRARDYDAMG
4	FVQCPDG
5	RVYSLFLDESRSTQY
6	EYQDAFLFNELK

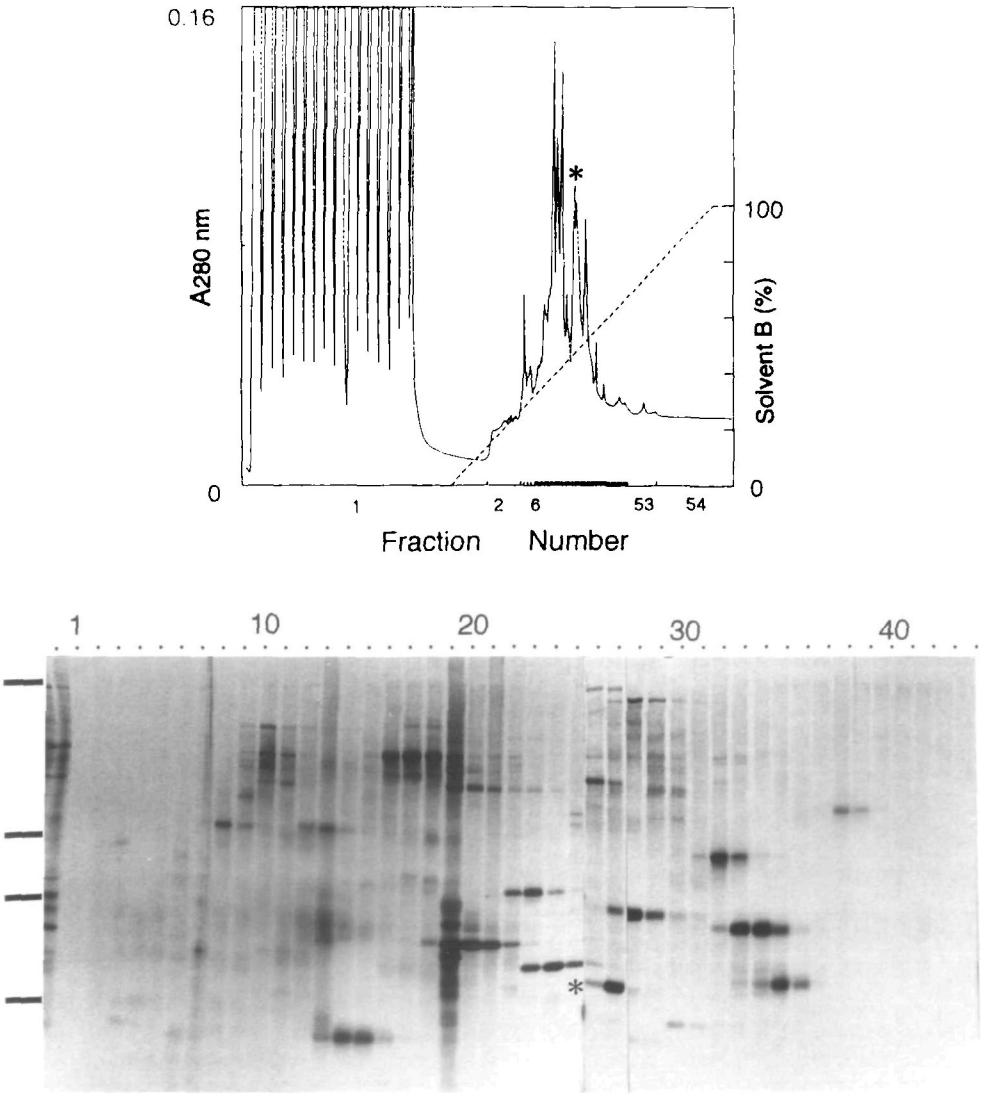


Fig. 1. **Purification of rp47 from the WGA-bound fraction of a rat liver nuclear matrix extract.** The WGA-bound fraction of a nuclear matrix salt extract, 1.5 mg, was treated with formic acid, and then applied to a Poros 10R1 column and eluted with a 75-min linear gradient of isopropanol in 60% formic acid. The eluate was fractionated as indicated in the elution profile. Fraction 1 corresponds to the flow through fraction. Aliquots of the fractions were analyzed by SDS-PAGE and silver staining. The numbers above the gel correspond to the fraction numbers. The bars to the left of the gel indicate the positions of marker proteins of 200k, 97k, 66k, and 43k, from top to bottom. The asterisks indicate the positions of rp47.

method after cloning to the Bluescript II SK(−) plasmid vector (Fig. 2). This clone contained an open reading frame (ORF) encoding a M_r 51,112 polypeptide comprising 463 amino acid residues. All amino acid sequences obtained for the rp47 peptides were found in the deduced amino acid sequence of the ORF (underlined region in Fig. 2). The sequence contained helicase superfamily motifs, GXXXX-GKT/S (amino acid residues 77-84), DEXH (amino acid residues 299-302), and others, which are boxed in Fig. 2 (*11, 12*).

mp47 Homologues in Other Organisms—To examine for the presence of mp47 homologues in other eukaryotes, we searched databases of *Saccharomyces cerevisiae* genome DNA sequences and EST. One clone, YPL235W (GenBank accession number Z73591), encodes an *S. cerevisiae* homologue, scp47 (Fig. 3A), whose sequence is the same as that of scRUVBL2 reported recently by Qiu *et al.* (*10*). The cDNA sequence of the complete ORF of human p47 (hp47) could be constructed with 28 human EST sequences (GenBank accession numbers T08989, H16424, AA120808, AA295369, W16757, AA442510, AA641138, AA554571, AA316505, AA313988, W74106, T35920, T32891, T35784, M86061, AA374580, R19091, T30926, H16341, F12454, R87393, AA311649, T35858, AA340214, AA121446, AA130812, AA464478, and AA832093) found in the EST database (Fig. 3A). The amino acid sequences of

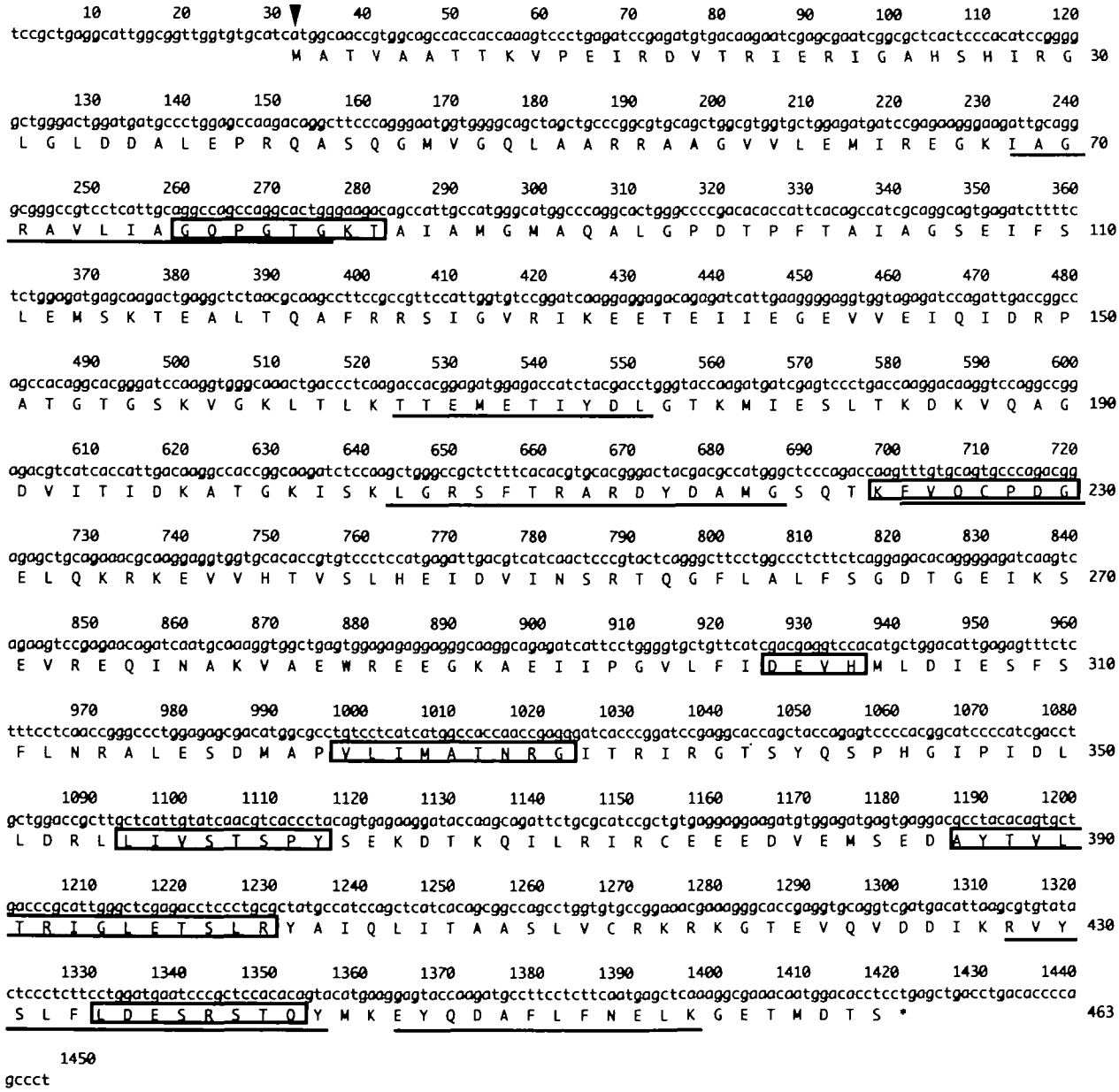
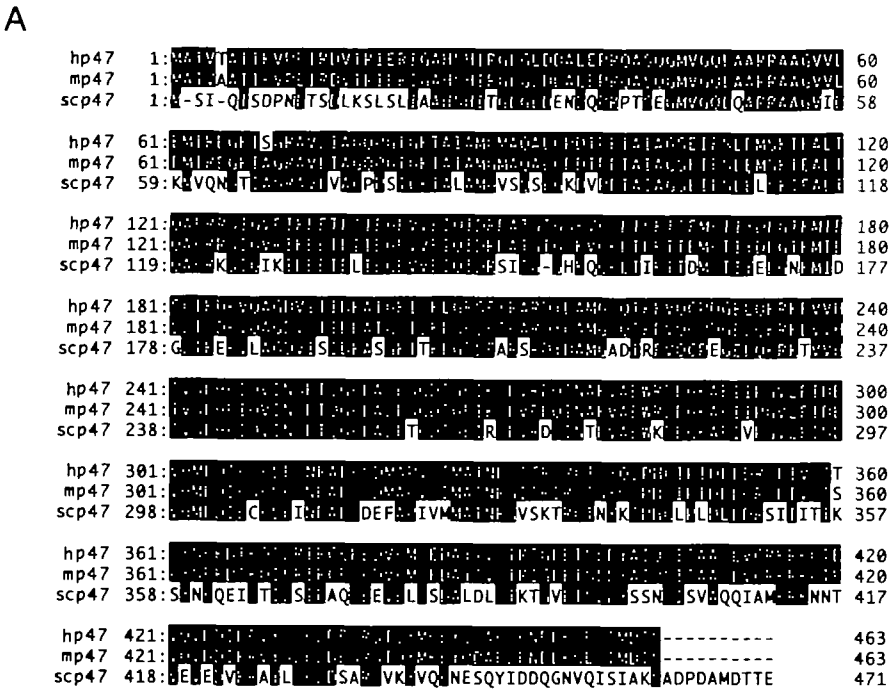


Fig. 2. **Nucleotide and amino acid sequences of mp47.** Nucleotides are numbered above each line. Amino acid residues are represented by their one letter codes and are numbered to the right. The arrowhead and asterisk indicate the start codon, and an in-frame stop codon, respectively. The sequences obtained for the lysylendopeptidase-digest of rp47 are underlined. Helicase motifs I, Ia, II, III, IV, V, and VI are boxed, from the amino terminal to the carboxyl terminal. The GenBank accession number for mp47 is AB013912.

mp47 and hp47 completely match each other with the exception of three amino acid replacements: A^5 to T^5, A^{69} to S^{69}, and T^{360} to S^{360}. The amino acid residues of mp47 and scp47 are 69% identical (Fig. 3B). As can be seen in Fig. 3B, these three proteins: *i.e.*, mp47, hp47, and scp47, should be classified into an RuvB-like protein group distinct from that of rp50 RuvB-like proteins (discussed below). The amino acid sequence of mp47 was then compared with those of bacterial RuvBs (Fig. 4). For example, *Thermotoga maritima* RuvB and mp47 show 33% (55%) and 38% (54%)



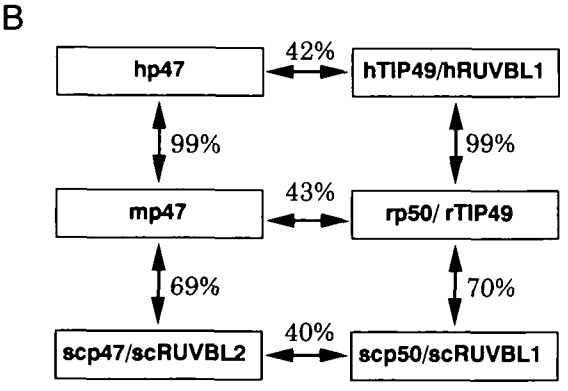


Fig. 3. **Comparison of the amino acid sequences of mouse and other eukaryotic p47s.** (A) Alignment of the amino acid sequences of mp47 and its homologues in man (hp47) and *S. cerevisiae* (scp47). The black boxes contain amino acid identity among more than two sequences. The amino acid sequences of scp47 and hp47 were deduced from nucleotide sequences found in the databases of *S. cerevisiae* genomic DNA sequences and EST. (B) Amino acid sequence identity among mp47 and rp50 and their eukaryotic homologues. Amino acid sequences of hp47, mp47, scRUVBL2 (*10*), hTIP49 (GenBank accession number AB002406), rTIP49 (GenBank accession number AB002406), and scRUVBL1 (*10*) were compared. The percentages of identical amino acids are indicated.

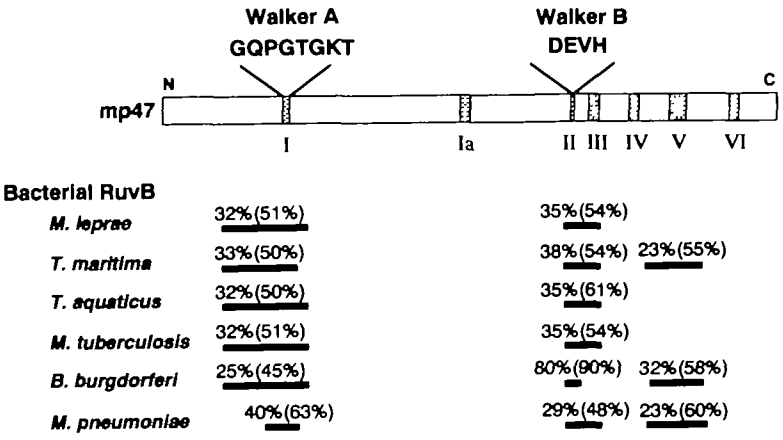


Fig. 4. **Amino acid sequence homology of mp47 and prokaryotic RuvB helicases.** The amino acid sequence of mp47 was aligned with the sequences of *Mycobacterium leprae* (Swiss-prot accession number P46833), *Thermotoga maritima* (Q56313), *Thermus aquaticus* (Q56214), *Mycobacterium tuberculosis* (Q50629), *Borrelia burgdorferi* (P70828), and *Mycoplasma pneumoniae* (P75242) RuvBs. Horizontal bars denote conserved regions in mp47 and prokaryotic RuvBs. The percentage identity (similarity) for each homologous region compared with mp47 is indicated. The Walker A motif (helicase motif I), Walker B motif (helicase motif II), and other helicase motifs Ia, III, IV, V, and VI in mp47 are indicated.

TABLE II. **The molecular masses and molecular shapes of the rp47 monomer and complex.** rp47 complexes were analyzed in medium containing 2.5 mM $MgCl_2$, whereas the monomer was analyzed in the same medium with $MgCl_2$ omitted.

	Molecular mass	Sedimentation coefficient (S)	Stokes' radius (nm)	Native molecular mass ($\times 10^{-3}$)	Frictional ratio (f/f_0)	Axial ratio
rp47						
Complex	–	13.9±0.9[a]	11.9±0.4[a]	697±53	2.03±0.06	21.1±1.6
Monomer	47,000[b]	3.2±0.1[a]	3.6±0.0[a]	52±1.1	1.6±0.0	11±0.0
rp50						
Complex	–	13.9±0.9[c]	11.9±0.4[c]	697±53[c]	2.03±0.06[c]	21.1±1.6[c]
Monomer	50,000[b,c]	3.3±0.0[c]	3.9±0.0[c]	54±0.3[c]	1.5±0.0[c]	10±0.0[c]

[a] $n = 3$. [b] These values were obtained by SDS-PAGE. [c] These values were reported previously in Ref. 9.

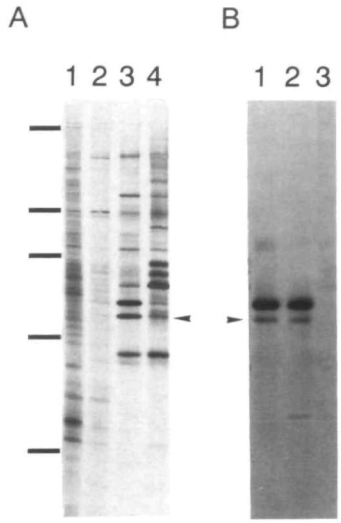


Fig. 5. **Binding of rp47 to WGA-Sepharose.** (A) A rat liver nuclear matrix salt extract (1) was applied to a WGA-Sepharose column, and the bound protein was successively eluted with buffer C containing 1 M NaCl (2), 1 M urea (3), and 0.5 M GlcNAc (4). (B) Partially purified rp47 (1) was applied to a WGA-Sepharose column and the unbound fraction was collected (2). Then the bound protein (3) was eluted immediately with buffer C containing 0.5 M GlcNAc. Aliquots were analyzed by SDS-PAGE followed by silver staining. Arrowheads indicate the positions of rp47. Bands just above p47 in lanes A3, B1, and B2 are rp50. The bars to the left of the gel indicate the positions of marker proteins of 200k, 97k, 66k, 43k, and 29k, from top to bottom.

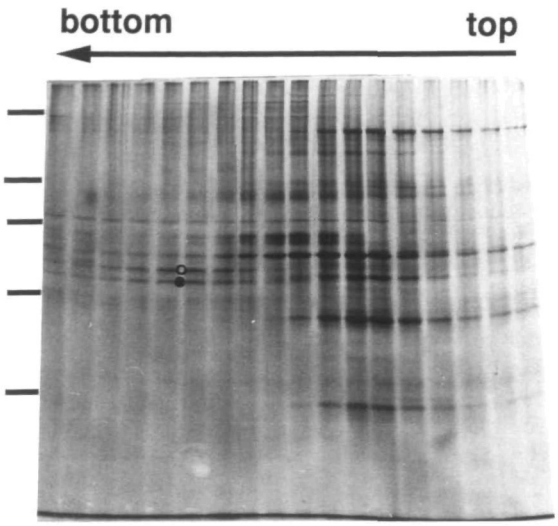


Fig. 6. **Co-migration of rp47 and rp50 in a glycerol density gradient.** The WGA-bound fraction of a rat liver nuclear matrix extract was applied to a glycerol gradient, centrifuged, fractionated, and then analyzed by SDS-PAGE followed by silver staining. The open circle and filled circle indicate the rp50 and rp47 bands, respectively. The bars to the left of the gel indicate the positions of marker proteins as in Fig. 5.

same (similar) amino acids around the Walker A and B motifs (helicase motifs I and II, respectively), and 23% (50%) same (or similar) amino acids around helicase motifs IV to V. These results suggest that mp47 is a mammalian homologue of RuvB DNA helicase.

Characterization of rp47 in the WGA-Bound Fraction

The rat liver nuclear matrix extract was applied to a WGA-Sepharose column, and bound proteins were eluted successively with 1 M NaCl, 1 M urea, and 0.5 M GlcNAc (Fig. 5). rp47 was eluted with 1 M urea (Fig. 5A, lane 3) but not 1 M NaCl (Fig. 5A, lane 2). Most WGA-bound proteins were eluted somewhat with 1 M urea, the majority remaining on the column (Fig. 5A, lane 4, and Ref. *15*). Moreover, rp47 did not react with WGA-horseradish peroxidase conjugate (data not shown). These results suggest that rp47 does not bear a GlcNAc residue but binds to WGA-Sepharose indirectly through a GlcNAc-bearing protein(s). To confirm the indirect binding of rp47 to WGA-Sepharose, we examined whether or not partially purified rp47, purified from the 1 M urea eluted fraction (Fig. 5A, lane 3) by gel filtration (*19*), binds to WGA-Sepharose. As expected, the partially purified p47 did not bind directly to WGA-Sepharose (Fig. 5B, lane 3). These results suggest that rp47 exists in the WGA-bound fraction as a complex containing GlcNAc-bearing protein(s).

Protein Complexes Containing rp47—At first, the sedimentation coefficient and Stokes' radius of rp47 were determined using the "urea-eluted fraction" in the absence of $MgCl_2$ (Table II). The native molecular mass and axial ratio were calculated from the S-value and Stokes' radius, respectively (Table II). These values indicated that rp47 exists as a monomer in the absence of $MgCl_2$. On the other hand, when crude rp47 in the WGA-bound fraction of rat liver nuclear matrix was analyzed by gel filtration in the presence of $MgCl_2$, rp47 eluted in the region of the large protein complexes (data not shown). It is known that the hexameric structure of RuvB is stabilized by Mg^{2+} ions (*20*). Therefore, we determined the sedimentation coefficient and Stokes' radius of the complex in the presence of $MgCl_2$ (Table II). The native molecular mass was calculated from the S-value and Stokes' radius. The results showed that rp47 exists as a large complex (697k) in the presence of $MgCl_2$. This molecular size is larger than that of the p47 hexamer (discussed below). The large frictional ratio, 2.03,

indicates that the shape of the complex deviates from a globular form. The axial ratio of the complex as a prolate, as estimated from the frictional ratio using the table of Schachman (*21*), was about 21 (Table II). When the sedimentation coefficients and Stokes' radii of rp47 and rp50 were compared, a small but obvious difference was observed in the Stokes' radii of the monomers. The monomers of these two proteins eluted from the gel filtration column in apparently different fractions (data not shown). However, the complexes containing rp47 and rp50 sedimented in a glycerol density gradient and eluted from a gel filtration column at exactly the same positions. So, exactly the same molecular masses were assumed for the complexes containing rp47 and rp50 (Table II). For example, the sedimentation patterns of rp47 and rp50 in their complex forms are shown in Fig. 6. Various proteins in the WGA-bound fraction were sedimented depending on their *S*-values. The complexes containing rp47 and rp50 seem to have exactly the same *S*-values. No other protein with the same sedimentation profile as rp47 and rp50 was observed as an obvious band. These results suggest that the major components of rp47- and rp50-containing complexes are rp47 and rp50 themselves. These two proteins may be components of the same complex (discussed below).

DISCUSSION

We cloned and sequenced the cDNA of mp47. We could not obtain a peptide containing the amino terminal residue of rp47 from its lysylendopeptidase digest. The amino acid sequence of the amino terminal region was also not obtained by direct sequencing of the intact rp47 protein. Therefore, we estimated the AUG start codon of mp47, as shown in Fig. 2, based on the following indirect data. First, the molecular weight, 51,112, calculated from the deduced amino acid sequence of mp47, was close to that of the rp47 protein, $(52 \pm 1.1) \times 10^3$, as estimated by hydrodynamic methods (Table II). Second, the DNA sequence around Met^1, ATCATGG, is very similar to the Kozak consensus sequence for translational initiation, ACCATGG (*22, 23*). Third, upstream of the coding region of mp47 cDNA, there is a stop codon at -90 to -88 in frame, but no Met between the stop codon and Met^1.

We could not obtain an rp47 cDNA. Therefore, the amino acid sequence of rp47 is not yet known other than the partial sequences shown in Fig. 2 and that deduced from a rat EST nucleotide sequence (corresponding to amino acid residues 344-462 of mp47; GenBank accession number, AA859838). All these sequences completely match the corresponding regions of mp47. Furthermore, the amino acid sequences of mouse and human p47s are identical except for two amino acids (Fig. 3). Therefore, the amino acid sequence of rp47 may be entirely same as that of mp47.

The deduced amino acid sequence of mp47 contains ATP/GTP binding Walker motifs A, GXXXXGKT, and B, DEXH (amino acid residues 77-84 and 299-302) (Fig. 2) (*24*). These two motifs constitute major parts of helicase motifs (*12, 25*). Moreover, the amino acid sequence from the amino terminal to the carboxyl terminal is similar to that of putative rat DNA-helicase rp50 (*9*): 43% (84%) of the amino acids are identical (or similar). The amino acid residues of three portions of mp47: 47-94, 295-325, and 355-392, are identical (or similar) to those of RuvB helicase from *T. maritima* (*26*), 33% (55%), 38% (54%), and 23% (50%), respectively (Fig. 4). These results suggest that p47 is a RuvB-like mammalian DNA helicase. The well conserved amino acid sequences of p47s in organisms ranging from yeast to mammals (Fig. 3) suggest that p47s are essential to the basic functions of eukaryotes. The similarity of the amino acid sequences of p47 and RuvB suggests that p47 plays essential roles in DNA replication, repair, and recombination (*1, 2, 27*). In association with RuvA, RuvB DNA helicase in prokaryotes promotes the branch migration of Holliday junctions during genetic recombination and DNA repair (*25, 28*). Similar functions are expected for p47, a putative DNA-helicase. The function of p47 may be regulated by phosphorylation because consensus sites for phosphorylation by cAMP- and cGMP-dependent protein kinases: R·X·X·S/T (T^{419}), casein kinase II: S/T·X·X·E·X and S/T·X·X·D·X (T^{17}, S^{114}, T^{166}, T^{170}, S^{243}, and S^{363}), and protein kinase C: S/T·X·K/R (T^7, T^{81}, T^{162}, T^{199}, T^{328}, S^{363}, and S^{398}) are found in mp47. The basic amino acid stretch of RKRK (amino acid residues 414-417) may be a nuclear localization signal of mp47 (*29*).

The p47s found in three eukaryotes should be classified into an RuvB-like protein group distinct from that of p50 RuvB-like proteins. This is because (i) the percentages of identical amino acids between members within each group are higher than those between members of different groups (Fig. 3B); and (ii) the percentages of identical amino acids between hp47 and hTIP49, mp47 and rp50, and scp47 and scp50 are very similar (Fig. 3B). This seems to mean that an ancestor helicase-like protein differentiated to the ancestor proteins of p47 and p50 at a very early time in the couse of eukaryotic evolution and then amino acid mutations have accumulated in each protein at almost the same rates. To examine which of p47 and p50 is more similar to RuvB, amino acid sequences from the amino terminal to the carboxyl terminal of RuvBs from *T. maritima* and *T. thermophilus* were compared with those of mp47 and rp50: 11 and 17% of the amino acids were identical, respectively (RuvBs from *T. maritima* and *T. thermophilus* were most similar to mp47 and rp50, respectively). These results suggest that p50 is more similar to RuvB than p47.

Both rp47 and rp50 were extracted from rat liver nuclei with same high salt buffer. Therefore, these proteins seem to bind to a nuclear matrix structure directly or indirectly through an ionic interaction. These proteins extracted with 500 mM NaCl remained in large complexes containing WGA-binding glycoprotein(s), because they bound to WGA-Sepharose (Fig. 5 and see Ref. *9*). It is known that many WGA-binding glycoproteins, such as the Sp1 (*30*), CTF (*30*), and HNF1 (*31*) transcription factors, the largest subunit of RNA polymerase II (*32*), and others, act in nuclei (*33, 34*). It is also known that p50 is included in an RNA polymerase II holoenzyme complex (*10*). Therefore, WGA-binding glycoprotein(s) in rp47 and rp50 complexes may be RNA polymerase II holoenzymes and/or transcription factors bearing *N*-acetylglucosamine residues. The 697k complex is stable in 1 M NaCl-1% Triton X-100 but not in 1 M urea, because rp47 bound to WGA-Sepharose indirectly could be eluted with 1 M urea (Fig. 5) and rp47 in the urea-eluted fraction existed as a monomer (Table II).

The subunit composition of the 697k complex containing rp47 is not yet clear. However, there are several suggestive

observations as to the composition: (i) RuvB type helicases usually exist as hexamers (*20*); (ii) the sedimentation of rp47 with rp50 in a glycerol density gradient may mean that these two kinds of protein are contained in the same complex (Fig. 6); (iii) complexes containing rp47 and rp50 eluted in the same fraction from gel filtration and showed the same Stokes' radii (Table II); and (iv) immunoprecipitation of rp50 from a salt extract fraction of rat liver nuclei (unpublished observation) and hp50/hRUVBL1 from 293T cell lysates (*10*) causes the co-precipitation of a protein of about 47k. These observations suggest that 697k complexes contain six rp47s and six rp50s, and also that there may be some minor component(s). However, the possibility that rp47 and rp50 are contained in distinct complexes can not be excluded.

We wish to thank Kazuko Hasegawa for her help in preparing the pictures.

REFERENCES

1. Shinagawa, H. and Iwasaki, H. (1996) Processing the Holliday junction in homologous recombination. *Trends Biochem. Sci.* **21**, 107-111
2. West, S.C. (1997) Processing of recombination intermediates by the RuvABC proteins. *Annu. Rev. Genet.* **31**, 213-244
3. Aboussekhara, A., Chanet, R., Adjira, A., and Fabre, F. (1992) Semidominant suppressors of *srs2* helicase mutations of *Saccharomyces cereviciae* map in the *RAD51* gene whose sequence predicts a protein with similarities to prokaryotic RecA proteins. *Mol. Cell. Biol.* **12**, 3224-3234
4. Sugiyama, T., Zaitseva, E.M., and Kowalczykowski, S.C. (1997) A single-stranded DNA-binding protein is needed for efficient complex formation by the *Saccharomyces cerevisiae* Rad51 protein. *J. Biol. Chem.* **272**, 7940-7945
5. Maeshima, K., Morimatsu, K., Shinohara, A., and Horii, T. (1995) Rad51 homologs in *Xenopus laevis*: Two distinct genes are highly expressed in ovary and testis. *Gene* **160**, 195-200
6. Yoshimura, Y., Morita, T., Yamamoto, A., and Matsushiro, A. (1993) Cloning and sequence of the human *recA*-like gene cDNA. *Nucleic Acids Res.* **21**, 1665
7. Baumann, P., Benson, F.E., and West, S.C. (1996) Human Rad51 protein promotes ATP-dependent homologous pairing and strand transfer reactions in vitro. *Cell* **87**, 757-766
8. Kanemaki, M., Makino, Y., Yoshida, T., Kishimoto, T., Koga, A., Yamamoto, K., Yamamoto, M., Moncollin, V., Egly, J.-M., Muramatsu, M., and Tamura, T. (1997) Molecular cloning of a rat 49-kDa TBP-interacting protein (TIP49) that is highly homologous to the bacterial RuvB. *Biochem. Biophys. Res. Commun.* **235**, 64-68
9. Kikuchi, N., Gohshi, T., Kawahire, S., Tachibana, T., Yoneda, Y., Isobe, T., Lim, C.R., Kohno, K., Ichimura, T., Omata, S., and Horigome, T. (1999) Molecular shape and ATP binding activity of rat p50, a putative mammalian homologue of RuvB DNA helicase. *J. Biochem.* **125**, 487-494
10. Qiu, X.-B., Lin, Y.-L., Thome, K.C., Pian, P., Schlegel, B.P., Weremowicz, S., Parvin, J.D., and Dutta, A. (1998) An eukaryotic RuvB-like protein (RUVBL1) essential for growth. *J. Biol. Chem.* **273**, 27786-27793
11. Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., and Blinov, V.M. (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17**, 4713-4730
12. Gorbalenya, A.E. and Koonin, E. V. (1993) Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Opin. Struct. Biol.* **3**, 419-429
13. Makino, Y., Mimori, T., Koike, C., Kanemaki, M., Kurokawa, Y., Inoue, S., Kishimoto, T., and Tamura, T. (1998) TIP49, homologous to the bacterial DNA helicase RuvB, acts as an autoantigen in human. *Biochem. Biophys. Res. Commun.* **245**, 819-823
14. Blobel, G. and Potter, V.R. (1966) Nuclei from rat liver: Isolation method that combines purity with high yield. *Science* **154**, 1662-1665
15. Kita, K., Omata, S., and Horigome, T. (1993) Purification and characterization of a nuclear pore glycoprotein complex containing p62. *J. Biochem.* **113**, 377-382
16. Kikuchi, N., Yamakawa, Y., Ichimura, T., Omata, S., and Horigome, T. (1997) Reversed-phase high-performance liquid chromatography of membrane proteins on perfusion-type polystyrene resin columns in 60% formic acid. *Chromatography* **18**, 175-184
17. Kawahire, S., Takeuchi, M., Gohshi, T., Sasagawa, S., Shimada, M., Takahashi, M., Abe, T.K., Ueda, T., Kuwano, R., Hikawa, A., Ichimura, T., Omata, S., and Horigome, T. (1997) cDNA cloning of nuclear localization signal binding protein NBP60, a rat homologue of lamin B receptor, and identification of binding sites of human lamin B receptor for nuclear localization signals and chromatin. *J. Biochem.* **121**, 881-889
18. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
19. Imai, N., Saito, M., Denta, K., Ichimura, T., Omata, S., and Horigome, T. (1995) Purification of 39 kDa and 50 kDa putative nucleoporins from rat liver nuclear envelopes. *Biochem. (Life Sci. Adv.)* **14**, 59-66
20. 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
21. Schachman, H.K. (1959) *Ultracentrifugation in Biochemistry*, p. 239, Table IX, Academic Press, New York and London
22. Kozak, M. (1989) The scanning model for translation: An update. *J. Cell Biol.* **108**, 229-241
23. Mcbratney, S. and Sarnow, P. (1996) Evidence for involvement of *trans*-acting factors in selection of the AUG start codon during eukaryotic translational initiation. *Mol. Cell. Biol.* **16**, 3523-3534
24. Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982) Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes, and a common nucleotide binding fold. *EMBO J.* **8**, 945-951
25. Tsaneva, I.R., Müller, B., and West, S.C. (1993) RuvA and RuvB proteins of *Escherichia coli* exhibit DNA helicase activity *in vitro*. *Proc. Natl. Acad. Sci. USA* **90**, 1315-1319
26. 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
27. Kadadek, T. (1998) Mechanistic parallels between DNA replication, recombination and transcription. *Trends Biochem. Sci.* **23**, 79-83
28. 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
29. Garcia-Bustos, J., Heitman, J., and Hall, M.N. (1991) Nuclear protein localization. *Biochim. Biophys. Acta* **1071**, 83-101
30. Jackson, S.P. and Tjian, R. (1989) Purification and analysis of RNA polymerase II transcription factors by using wheat germ agglutinin affinity chromatography. *Proc. Natl. Acad. Sci. USA* **86**, 1781-1785
31. Lichtsteiner, S. and Schibler, U. (1989) A glycosylated liver-specific transcription factor stimulates transcription of the albumin gene. *Cell* **57**, 1179-1187
32. Kelly, W.G., Dahmus, M.E., and Hart, G.W. (1993) RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *J. Biol. Chem.* **268**, 10416-10424
33. Haltiwanger, R.S., Busby, S., Grove, K., Li, S., Mason, D., Medina, L., Moloney, D., Philipsberg, G., and Scartozzi, R. (1997) O-Glycosylation of nuclear and cytoplasmic proteins: regulation analogous to phosphorylation? *Biochem. Biophys. Res. Commun.* **231**, 237-242
34. Hart, G.W., Kreppel, L.K., Comer, F.I., Arnold, C.S., Snow, D.M., Ye, Z., Cheng, X., DellaManna, D., Caine, D.S., Earles, B.J., Akimoto, Y., Cole, R.N., and Hayes, B.K. (1996) O-GlcNAcylation of key nuclear and cytoskeletal proteins: reciprocity with O-phosphorylation and putative roles in protein multimerization. *Glycobiology* **6**, 711-716