Nucleic Acids Research, Vol. 19, No. 16 4523–4529

Cloning and characterization of Rrp1, the gene encoding *Drosophila* strand transferase: carboxy-terminal homology to DNA repair endo/exonucleases

Miriam Sander, Ky Lowenhaupt[1], William S.Lane[2] and Alexander Rich[1]
Laboratory of Genetics D3-04, National Institute of Environmental Health Sciences, POB 12233, Research Triangle Park, NC 27709, [1]Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139 and [2]Harvard Microchemistry Facility, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA

Received March 28, 1991; Revised and Accepted July 17, 1991 GenBank accession no. M62472

ABSTRACT

We previously reported the purification of a protein from *Drosophila* embryo extracts that carries out the strand transfer step in homologous recombination (Lowenhaupt, K., Sander, M., Hauser, C. and A. Rich, 1989, J. Biol. Chem. 264, 20568). We report here the isolation of the gene encoding this protein. Partial amino acid sequence from a tryptic digest of gel purified strand transfer protein was used to design a pair of degenerate oligonucleotide primers which amplified a 635 bp region of *Drosophila* genomic DNA. Recombinant bacteriophage were isolated from genomic and embryo cDNA libraries by screening with the amplified DNA fragment. These bacteriophage clones identify a single copy gene that expresses a single mRNA transcript in early embryos and in embryo-derived tissue culture cells. The cDNA nucleotide sequence contains an open reading frame of 679 amino acids within which are found 5 tryptic peptides from the strand transfer protein. Expression of this cDNA in *E. coli* produces a polypeptide with the same electrophoretic mobility as the purified protein. The deduced protein sequence has two distinct regions. The first 427 residues are basic, rich in glutamic acid and lysine residues and unrelated to known proteins. The carboxy-terminal 252 residues are average in amino acid composition and are homologous to the DNA repair proteins, *Escherichia coli* exonuclease III and *Streptococcus pneumoniae* exonuclease A. This protein, which we name Rrp1 (Recombination Repair Protein 1), may facilitate recombinational repair of DNA damage.

INTRODUCTION

The molecular analysis of homologous recombination events in prokaryotes has relied largely on a genetic approach. In *Escherichia coli*, the importance of the strand transfer activity of recA protein for all conjugative recombination pathways was established by the properties of *E. coli* strains deficient in this activity (for reviews see refs.(1,2)). In eukaryotes, no regulatory protein which is functionally homologous to the recA protein has been identified, although strand transfer proteins have been identified by biochemical assay (3–9), and mutants deficient in recombination functions have been isolated from species such as *Saccharomyces cerevisiae* (10–14), *Ustilago* (15) and *Drosophila* (16).

RecA protein is known to be essential for the regulation of a complex DNA damage response system known as SOS (for review see ref. (17)). Inducible DNA damage responses have been demonstrated in yeast (18,19) and in eukaryotic cells (20,21), but the mechanism(s) controlling this type of response remains unclear.

We previously reported the purification of a strand transfer activity from *Drosophila* embryos (22). This activity copurifies with a polypeptide whose apparent molecular mass is 105 kDa based on electrophoretic mobility. The activity requires Mg^{++}, homologous DNA, is ATP-independent, and proceeds by a strand displacement mechanism. To further understand both the biological function and the biochemical characteristics of this protein, which we have named Rrp1 (Recombination Repair Protein 1), we have isolated the gene that encodes it. We present here the characterization of the cDNA sequence of the *Rrp1* gene.

MATERIALS AND METHODS

Nucleic acids and Enzymes

Restriction enzymes were purchased from either Life Technologies, Inc. or New England Biolabs and used according to the suppliers' protocols. AmpliTAQ DNA polymerase, T7 RNA polymerase and Klenow DNA polymerase were purchased from Perkin-Elmer Cetus, Stratagene and New England Biolabs, respectively. Plasmid DNA was purified as described by Birnboim (23) followed by ion exchange chromatography on Qiagen (Qiagen, Inc.). Phage nucleic acids were purified using chromatography on Qiagen (Qiagen, Inc.). Drosophila genomic DNA was prepared from 6–18 h embryos (24). The expression

vector pET3a and the *E. coli* host strains for expression were purchased from Novagen.

Protein microsequencing

Approximately 20 μg fraction IV strand transferase protein (22) was purified by SDS-PAGE (25), and prepared for *in situ* tryptic digestion essentially as described (26). Tryptic peptides were purified using narrowbore reverse phase HPLC and sequenced on an Applied Biosystems 477A protein sequencer.

Isolation and characterization of recombinant bacteriophage

Mixed base oligonucleotides were synthesized on a Milligen Biosearch DNA synthesizer. PCR amplification of genomic DNA was carried out as described by Compton (27). The 635 bp amplified fragment was treated with Klenow DNA polymerase in the presence of deoxynucleosidetriphosphates and ligated into the EcoRV site of pBluescript SK (−) (Stratagene). The insert of this plasmid (pMS111) was removed by restriction digestion, purified by agarose gel electrophoresis and used to generate hybridization probes.

Screening of genomic (28) or cDNA libraries (29) was carried out as described (30). The radiolabeled probe for the genomic library screen was generated by the random priming method (31) using the DNA fragment amplified by PCR (Fig. 1c). Screening of the cDNA library was carried out with an RNA runoff transcript probe generated from pMS111 corresponding to the PCR amplified region of genomic DNA.

The insert region of λc5 was removed by digestion of the phage with KpnI and SstI and then ligated to KpnI/SstI double digested pBluescript SK (−) to create the plasmid pMS215. The nucleotide sequence of the cDNA portion of pMS215 was determined using dideoxynucleotide sequencing protocols for double-stranded plasmid DNA (32) and Sequenase 2.0 from United States Biochemical Corporation. Oligonucleotide primers for sequencing were synthesized by either Oligos Etc. Inc. or Research Genetics. The DNA sequence of both DNA strands was determined in its entirety. Ambiguous regions were resolved using dITP sequencing. Selected regions were also sequenced from plasmid subclones of λc1 and the *Rrp1* genomic DNA.

Hybridization to DNA, RNA and polytene chromosomes

Hybridization probes were prepared using DNA restriction fragments of plasmid DNA which were purified by agarose gel electrophoresis. DNA fragments were labeled by the random priming method (31).

Drosophila genomic DNA was digested with restriction enzymes, separated on agarose gels and transferred to nitrocellulose membranes (Hybond C extra, Amersham). High stringency hybridization was in 5×SSC at 68°C, washed to 0.2×SSC, 65°C. Low stringency hybridization was in 5×SSC, 60°C, washed to 2×SSC, 60°C. This stringency allows the visualization of members of the tubulin gene family (A. Ayme-Southgate, personal communication).

RNA was prepared from 0−3 h embryos and Schneider 2L cells as described by Ayme and Tissieres (32). mRNA was selected using oligo(dT)-cellulose (Type III, Collaborative Research), separated on formaldehyde gels (32) and transfered to Hybond N (Amersham). Hybridizations were as above (high stringency). Size markers were purchased from Life Technologies, Inc. and visualized by the addition of ^{32}P-labeled λ DNA to the hybridization mix.

Polytene chromosomes were prepared and *in situ* hybridization performed as described by Pardue *et al.* (33).

Construction and characterization of expression plasmid

To construct pRrp1-E1 two oligonucleotide primers with the sequences 5′-GGGC**GGATCC**ATGCCGCGTGTCAAGGCC-G-3′ and 5′-GGGCGTCGACTCCATTCCCATTGGCATTT-3′ were used to amplify a 95 bp DNA fragment of pMS215 that contains the nucleotide sequence from the amino-terminal methionine codon of the Rrp1 ORF to the SalI site in the cDNA sequence. The first oligonucleotide introduces a BamHI site (in bold typeface in the oligonucleotide sequence) immediately 5′ to the amino-terminal Rrp1 ATG codon (underlined in the oligonucleotide sequence) such that an in frame fusion of the Rrp1 ORF to the 12 amino acid leader peptide of the pET3a vector was created by ligation into BamHI restricted pET3a. To construct pRrp1-E1, dephosphorylated SalI digested pMS215 was ligated to the SalI digested 95 bp PCR fragment and the ligation product was then cut with SstI. The SstI end was converted to a blunt end with T4 DNA polymerase in the presence of the four deoxynucleosidetriphosphates. The sample was then digested with BamHI and the 3200 bp fragment containing the Rrp1 cDNA sequence was purified by agarose gel electrophoresis. This fragment was ligated to BamHI cut pET3a vector in a two-step reaction. A first ligation reaction was carried out, then T4 DNA polymerase was added in the presence of the four deoxynucleosidetriphosphates to fill in the unligated BamHI ends, and then a second ligation was performed. Two recovered plasmids, pRrp1-E1 and pRrp1-E9 have BamHI recognition sites at both cloning junctions due to the presence of a terminal C:G base pair at the SstI end of the insert fragment. pRrp1-E1 and pRrp1-E9 carry the cDNA insert in the forward and reverse orientations, respectively (Fig. 4). The nucleotide sequence of the cloning junction and the PCR amplified region of pRrp1-E1 was determined. No nucleotide sequence changes were introduced into this region through the subcloning process.

To analyze expression of the plasmids pRrp1-E1 or pRrp1-E9, appropriate host cells carrying the plasmid were incubated at 37°C in LB medium until the culture reached mid- to late-log growth. IPTG was added to a final concentration of 0.4 mM and incubation was continued. At appropriate time points, 1 ml of cells were centrifuged in a microfuge. The cell pellets were resuspended in 50 μl H_2O, followed by addition of 50 μl of a solution containing 10% glycerol, 75 mM Tris-HCl pH 6.8, 2% sodiumdodecylsulfate, 100 mM dithiothreitol and 0.1% bromophenol blue. 20 μl of each sample was analyzed by SDS-PAGE (25) on a 9% polyacrylamide gel.

Nucleotide and protein sequence analysis

Sequence analyses were done using programs from the University of Wisconsin Genetics Computer Group (UWGCG) software package (34).

RESULTS

Isolation of genomic and cDNA bacteriophage lambda clones encoding the strand transferase protein

The gene for the strand transfer protein was isolated using a reverse genetics approach. A protein fraction purified through three chromatographic steps (fraction lV strand transferase) was

further purified by preparative gel electrophoresis. The appropriate gel slice was transfered to nitrocellulose membrane and then subjected to tryptic digestion *in situ*. After HPLC purification, the amino acid sequences of several peptides were determined. The sequences of five tryptic peptides are shown in Fig. 1a. This protein sequence information was used to design oligonucleotide probes which could hybridize to exon regions of *Drosophila* genomic DNA.

For each of the two tryptic peptides 1 and 2, a pair of sense and antisense primers are shown in Fig. 1b. These primers were used in a PCR reaction with *Drosophila* genomic DNA as a template. The primer pair consisting of oligonucleotides 2 and 3 amplified a 635 bp segment very efficiently, whereas the other primer pair, oligonucleotides 1 and 4, did not amplify any specific sequence (data not shown). The amplified segment was used to screen a *Drosophila* genomic library (28) and an embryo cDNA library (29). The results of these experiments are summarized in Fig. 1c, which shows a restriction map of the genomic region and the segments present in three cDNA clones.

The gene for the strand transfer protein lies within a 10 kb BamHI fragment of genomic DNA (Fig. 1c). Hybridization to genomic DNA using the 10 kb BamHI DNA fragment showed that there is no repetitive DNA within this region. Even at low stringency hybridization, no cross hybridization has been detected, indicating that the *Rrp1* gene is not part of a multigene family. *Rrp1* maps to the 23 BC region on chromosome arm 2L by *in situ* hybridization to polytene chromosomes (data not shown). This is a relatively uncharacterized region of the *Drosophila* genetic map.

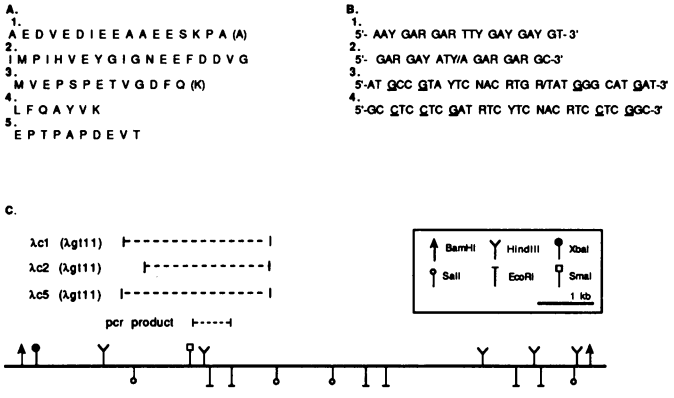

Fig.1. Isolation of cDNA clones for *Drosophila Rrp1* using tryptic amino acid sequences and PCR amplification. A. The amino acid sequences of 5 tryptic fragments of the Rrp1 protein are shown. The two terminal amino acids shown in parentheses were assigned with a lower level of confidence than the remaining amino acids. B. Four pools of DNA oligonucleotide sequences were synthesized which code for parts of the amino acid sequences 1 (pools 1 and 3) and 2 (pools 2 and 4) shown in A. Pools 1 and 2 are sense DNA sequences and pools 3 and 4 are antisense DNA sequences. The underlined nucleotides are positions where some of the possible amino acid codons in the genomic DNA would form mispairs with the oligonucleotide. PCR amplification using *Drosophila* genomic DNA as substrate was carried out under standard conditions (27) and resulted in amplification of a 635 bp genomic region using oligonucleotide pools 2 and 3, but in no amplification with oligonucleotide pools 1 and 4 (data not shown). C. A restriction map of the genomic DNA region surrounding the *Rrp1* gene is shown. The regions contained in 3 cDNA phage are indicated. The fragment amplified by PCR and used as probe to isolate the cDNA clones is indicated. Three small introns (less than 80 bp in length) which are spliced out of the precursor mRNA are not indicated on this diagram.

Northern analysis of mRNA transcription from *Rrp1*

RNA was isolated from Drosophila embryos and tissue culture cells and selected using oligo(dT)-cellulose. Hybridization to this RNA using a probe corresponding to the PCR amplified region of *Rrp1* identified a unique transcript of about 3 kb (Fig. 2). This transcript has been localized primarily to a 2.6 kb SalI fragment on the genomic map (Fig 1c; data not shown). This SalI fragment also hybridizes to the cDNA phage isolated by screening with the PCR amplified fragment. Three representative recombinant phage are shown in Fig 1c. The longest phage isolates, λc5 and λc1, have inserts 2.4 kb in length, which is close to the size of the mRNA detected on Northern blots (Fig. 2). Nucleotide sequence analysis of the cDNA as well as expression of the cDNA in *E. coli* demonstrate that the entire open reading frame (ORF) encoding Rrp1 protein is contained within the 2.4 kb inserts of λc5 and λc1(see below).

Figure 2 shows that Rrp1 mRNA is present in 0−3 h embryos. Since these embryos are not actively undergoing transcription, the mRNAs represented are maternally contributed. A maternal role for Rrp1 is supported by the fact that this transcript, although present in all stages tested, is most abundant in embryos less than 6 h old and in adult females (K.L., unpublished results). This expression pattern may reflect a role in events surrounding fertilization, including meiosis. In addition, the presence of Rrp1 mRNA and protein (not shown) in tissue culture cells suggests that there may be a function in mitotic tissue.

Nucleotide and predicted protein sequence of the *Rrp1* cDNA

The *Rrp1* cDNA sequence, as determined from λc5, contains a 679 codon ORF (Fig. 3). The putative amino-terminal methionine of the Rrp1 protein is located at nucleotide 133 of the cloned cDNA. The 133 nucleotide leader sequence contains multiple stop codons which occur in all 3 reading frames. (The

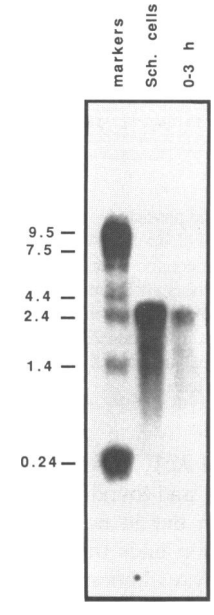

Fig. 2. Northern blot analysis of transcription from *Rrp1*. Approximately 10 μg of each mRNA sample was fractionated on a formaldehyde gel together with a size marker standard. The gel was transfered to Hybond-N membrane and probed with a random prime labeled probe of the PCR amplified region of the *Rrp1* gene. Lanes are (from left to right): size marker, Schneider cell mRNA, 0−3 h embryo mRNA. Marker sizes are indicated in kb.

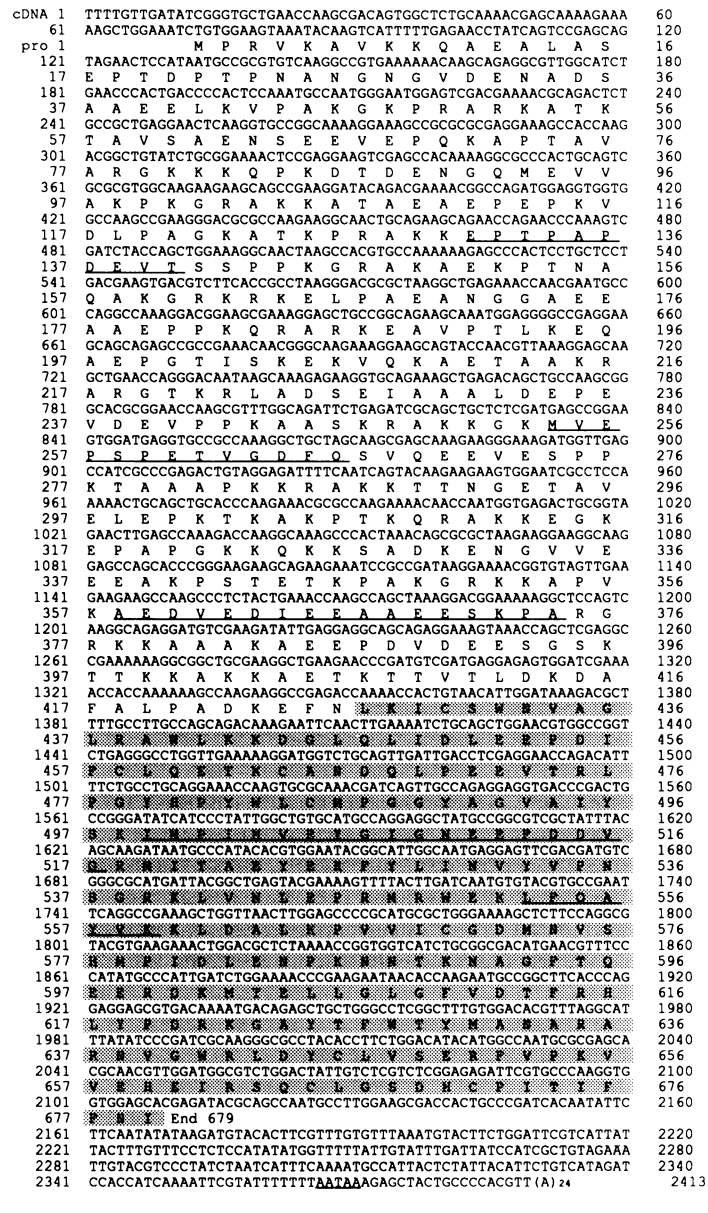

Fig.3. Nucleotide sequence and predicted amino acid sequence of the *Rrp1* cDNA. The DNA sequence of the 2413 bp cDNA insert is shown. The predicted Rrp1 protein sequence begins at nucleotide 133 and is shown above the coding sequence. Underlined amino acid sequences are identical with the tryptic peptide fragments of the strand transferase protein shown in Fig. 1a. The polyA addition signal sequence, AATAA, occurs at nucleotide 2366 and is underlined in the sequence. The shaded amino acid sequence indicates the region of Rrp1 which is homologous to DNA repair endo/exonucleases.

nucleotide sequence of λc1, an independently isolated cDNA, was determined in part, and found to differ in several nucleotides at the cloning junction but in no nucleotides of the 5'-leader sequence.) A stop codon ends the ORF at nucleotide position 2170, followed by a polyA addition signal, AATAA, at nucleotide position 2366, producing a 3'-untranslated region 220 nucleotides in length.

The protein predicted by this ORF contains sequences that match each of the tryptic peptide sequences from the strand transfer protein (Fig. 3, underlined amino acid sequences and Fig. 1a). This strongly suggests that this gene encodes the major polypeptide in the strand transferase preparation, whose apparent

Table I. Amino acid composition of Rrp1 protein

Rrp1 amino acid coordinates	1–426		427–679		1–679	
	Number of residues	%	Number of residues	%	Number of residues	%
Nonpolar						
ala	73	17.1	13	5.1	86	12.8
val	26	6.1	18	7.1	44	6.5
leu	10	2.3	24	9.5	34	5.0
ile	3	0.7	15	5.9	18	2.6
met	3	0.7	8	3.2	11	1.6
pro	43	10.1	14	5.5	57	8.2
phe	3	0.7	11	4.3	14	2.1
trp	0	0.0	6	2.4	6	0.9
Polar						
gly	22	5.2	17	6.7	39	5.7
ser	17	4.0	7	2.7	24	3.5
thr	30	7.0	10	3.9	40	5.9
cys	0	0.0	8	3.2	8	1.2
tyr	0	0.0	13	5.1	13	1.9
asn	11	2.6	14	5.5	25	3.7
gln	12	2.8	6	2.4	18	2.6
Basic						
arg	20	4.7	14	5.5	34	5.0
his	0	0.0	6	2.4	6	0.9
lys	76	17.8	16	6.3	92	13.4
Acidic						
asp	19	4.5	14	5.5	33	4.9
glu	58	13.6	19	7.5	77	11.3

molecular mass based on electrophoretic mobility is 105 kDa. However, the protein encoded by the ORF present in the Rrp1 cDNA has a predicted molecular weight of 75 kDa. This exact reason for the discrepancy between the predicted size and the apparent size is not known; however, the discrepancy is due to some characteristic of the polypeptide sequence *per se* (see below), and this may be the density of charged residues in the amino-terminal region of the protein.

The Rrp1 protein sequence is rich in lysine residues (13.4%) and glutamic acid residues (11.3%) (Table I). The density of charged residues is highest in the amino terminal two-thirds of the protein: 40% of the residues from 1–426 are charged (17.8% lysine and 13.6% glutamic acid), while 25% of the residues from 427–679 are charged (6.3% lysine and 7.5% glutamic acid). The protein segment from 94–254 contains 13 net positive charges and has a predicted pI of 10.7. It has been shown previously that highly charged protein regions can result in lowered electrophoretic mobility during SDS-PAGE (35).

To confirm the identity of the cloned gene, an expression plasmid, pRrp1-E1, was constructed in which the 679 amino acid ORF is fused to a 14 amino acid amino-terminal peptide and placed behind the T7 RNA polymerase promoter of the plasmid vector pET3a (Fig. 4). In this expression system the *E. coli* host strain, BL21(DE3)pLysS, has a chromosomal copy of the gene encoding T7 RNA polymerase that is under the control of an isopropyl-β-thio-galactopyranoside (IPTG) inducible promoter (36). Fig. 5 shows SDS-PAGE analysis of the polypeptides present in extracts of strains carrying pRrp1-E1 or pRrp1-E9, a control plasmid that has the BamHI fragment containing the coding sequence for Rrp1 in the reverse orientation to that shown in Fig. 4. A polypeptide with the expected electrophoretic mobility (102 kDa) accumulates with time of incubation in the presence of IPTG in the extracts of *E. coli* cells that carry both the expression construct pRrp1-E1 and a source of T7 RNA

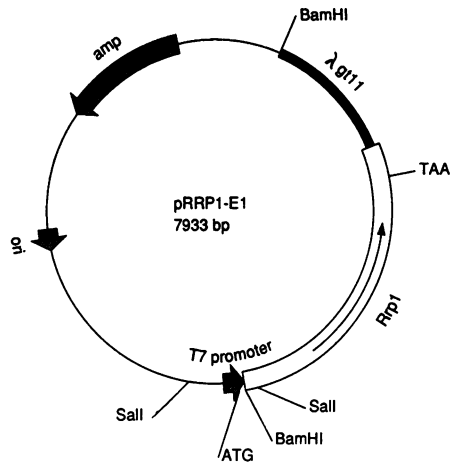

Fig. 4. Map of the plasmid pRrp1-E1. The vector plasmid pET3a is a derivative of pBR322 into which the T7 promoter has been inserted (36). In the plasmid pRrp1-E1 a fragment containing the Rrp1 cDNA nucleotide sequence and adjacent λ DNA is inserted at the BamHI site of pET3a. The protein synthesized from this construct consists of the 14 amino acid amino-terminal peptide MASMTGGQQMGRGS fused to the 679 amino acid sequence predicted from the Rrp1 cDNA and shown in Fig.3. In the plasmid pRrp1-E9 the orientation of the insert is in the reverse orientation to that shown above. The λ ΔNA adjacent to the T7 promoter in pRrp1-E9 encodes a portion of the β-galactosidase gene from λgt11.

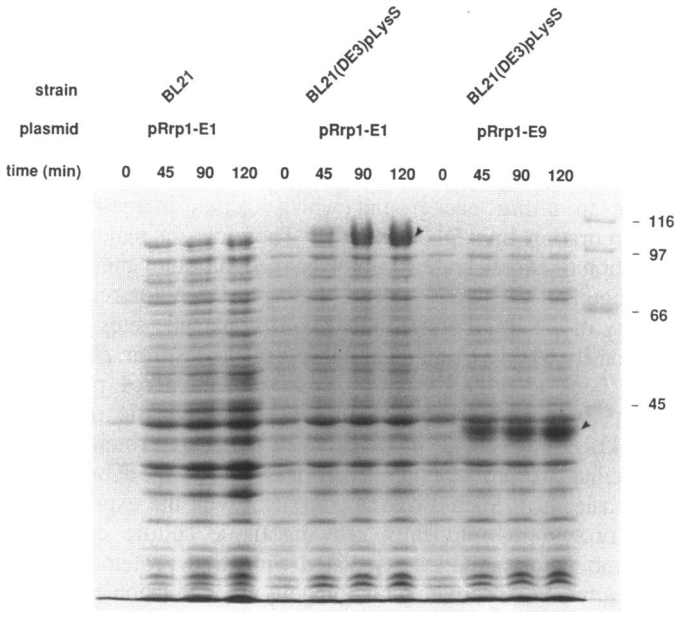

Fig.5. SDS-PAGE analysis of polypeptides in *E. coli* cells expressing Rrp1 protein. Kinetic analyses of the IPTG inducible protein synthesis in three *E. coli* cultures are shown. The host strain used and plasmid carried is indicated for each culture. IPTG was added at zero time to a concentration of 0.4 mM, and samples were prepared for SDS-PAGE as described under Materials and Methods. The IPTG induced polypeptides are indicated by arrowheads. The mobilities of marker proteins and their molecular weights in kDa are indicated to the right side of the figure.

polymerase (Fig. 5; central four lanes; arrowhead indicates the induced polypeptide). In the control culture that carries the plasmid pRrp1-E1 but lacks a source of T7 RNA polymerase (host strain BL21; Fig. 5, four leftmost lanes) the 102 kDa induced polypeptide is not present. In cells which have an

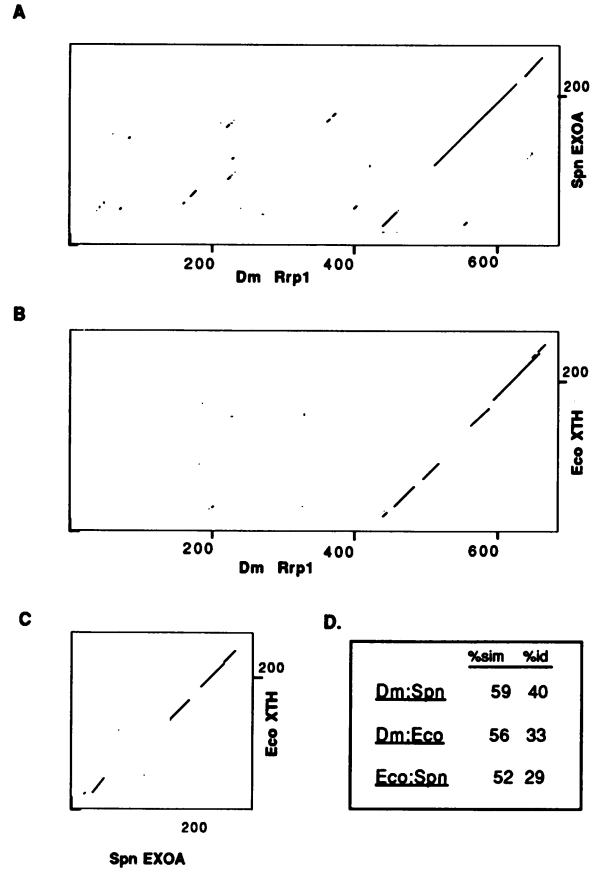

Fig.6. Homology relationships between the Rrp1 protein and two bacterial DNA repair proteins. A–C. Three pairwise comparisons were carried out on the three protein sequences *Drosophila* Rrp1(Dm Rrp1), *S. pneumoniae* exonuclease A (Spn EXOA) and *E. coli* exonuclease III (Eco XTH) using the COMPARE program from the UWGCG sequence analysis software (34). Regions of ≥50% identity in a window of 30 residues is indicated as a line on the diagonal. D. The % similarity and % identity for each pairwise comparison are shown. Only residues 427–679 of Rrp1 are considered in the calculation.

inducible T7 RNA polymerase gene but carry the plasmid pRrp1-E9 (Fig. 5, four rightmost lanes) the 102 kDa polypeptide is not present, but an induced protein with an $M_r \approx 40$ kDa is present (indicated by arrowhead). This species is encoded by the λgt11 DNA that is adjacent to the T7 prometer in pRrp1-E9.

Since the protein expressed from the plasmid pRrp1-E1 in *E. coli* has an electrophoretic mobility very close to that of the strand transfer protein purified from *Drosophila* embryos, several conclusions can be made. First, the cDNA sequence shown in Fig. 4 contains the complete open reading frame for the Rrp1 protein. Second, posttranslational modification of native Rrp1 protein by phosphorylation or glycosylation is not a likely cause for its aberrant electrophoretic mobility, since the native modification systems are not present in *E. coli* cells. The aberrant mobility must therefore be due to some characteristic of the protein sequence *per se*. Third, since the protein sequence predicted from the Rrp1 cDNA contains amino acid sequences that match all known tryptic peptides of the strand transfer protein, and since this protein, when synthesized in *E. coli*, has an electrophoretic mobility very close to that of the protein purified by strand transfer assay, the identity of the cloned gene as coding for the Drosophila strand transfer protein is confirmed.

A preliminary analysis of the enzymatic activity of the

expressed Rrp1 protein has been carried out. ATP-independent strand transfer activity is present in extracts which contain the induced 102 kDa protein, but is undetectable in control extracts that lack this polypeptide (M.S. unpublished results). This result suggests that the induced Rrp1 protein may be sufficient for the strand transfer function.

The Rrp1 protein sequence was examined for its relationship to other proteins using the University of Wisconsin Genetics Computer Group (UWGCG) software (34) to search the protein databases (GenEMBL, NBRF, Swissprot). The Rrp1 protein contains a region highly homologous to two bacterial DNA repair proteins, *S. pneumoniae* exonuclease A and *E. coli* exonuclease III, that were shown previously to be related to each other (37). The homology with Rrp1 extends throughout the entire length of the two bacterial proteins, but is limited to the carboxy-terminal one-third of the Rrp1 protein. This result is depicted graphically in Figs. 6a–c, and summarized in Fig.6d. In Figs. 6a–c, protein regions with a ≥50% identity over any window of 30 amino acids are displayed as a solid line on the diagonal between the paired protein sequences. *E. coli* exonuclease III and Rrp1 are 33% identical through a region of 278 amino acids, and *S. pneumoniae* exonuclease A and Rrp1 are 40% identical through the same region.

Enzymatic characterization of exonuclease III and exonuclease A has shown that both proteins are active as both AP endonucleases and 3′-exonucleases (37,38). The relatively high level of identity between these two proteins and *Drosophila* Rrp1 suggests that a functional similiarity between them could exist (discussed below).

The amino-terminal two-thirds of the Rrp1 protein is not related to any protein in the database. Its function is still uncharacterized; however, strand transfer, single-strand DNA (ssDNA) binding and ssDNA renaturation activities are likely possibilities. If these activities all map to the amino-terminal region of Rrp1, this may help explain the novel enzymatic properties of the Rrp1 protein; the bacterial members of this protein family do not demonstrate ssDNA renaturation or strand transfer activity (data not shown).

DISCUSSION

To facilitate further analysis of the biochemical properties and biological function of the Drosophila strand transferase protein, the gene encoding this protein has been isolated and its cDNA sequence determined. A single copy gene was identified that expresses a unique 3 kb mRNA during early embryogenesis and in mitotically growing embryo tissue culture cells. The cDNA sequence predicts a 679 amino acid protein that is unusually rich in lysine and glutamic acid residues. The protein sequence has two distinct regions. The amino-terminal 427 amino acids of the predicted protein sequence are not related to any known protein sequences. The carboxy-terminal 252 amino acids of the predicted protein sequence are 40% and 33% homologous to the DNA repair proteins *S. pneumoniae* exonuclease A and *E. coli* exonuclease III, respectively (Fig.6). Based on predicted roles in homologous recombination and DNA repair, this gene has been named *Rrp1*.

The gene identified is thought to be the gene encoding the protein purified by strand transfer assay for the following reasons. First, five peptide sequences determined by microsequencing the gel purified strand transferase protein are found within the protein sequence predicted from the cDNA (Fig. 3). Second, overexpression of the protein encoded by this cDNA in *E. coli* demonstrates the synthesis of a protein with the expected electrophoretic mobitily (Fig. 5). Third, a preliminary analysis of the enzymatic characteristics of the expressed Rrp1 protein demonstrates the presence ATP-independent strand transfer activity in both crude extracts and in a partially purified fraction of *E. coli* cells which contain the induced polypeptide, while control extracts lacking this polypeptide lack this activity (M.S. unpublished results).

It has not been ruled out that a protein component present as a minor contaminant in the purified *Drosophila* strand transfer fraction is required for some of its enzymatic functions. However, expression of the cloned cDNA reported here suggests that the Rrp1 protein may be sufficient for strand transfer activity (M.S. unpublished results). Further enzymatic analysis of the bacterially expressed protein and of truncated or altered forms of the protein will now allow the function(s) of the protein's amino-terminal and carboxy-terminal regions to be more easily tested.

Previous characterization of the purified strand transfer fraction indicated the presence of a 3′-exonuclease activity (22), which is consistent with the protein sequence homology to exonucleases described above. Furthermore, this exonuclease has a specificity identical to that of exonuclease III and copurifies with Rrp1 strand transferase activity (39). The exonuclease activity acts in a limited manner on the dsDNA substrate in strand transfer reactions, and may be required prior to initiation of the strand displacement reaction which has been demonstrated biochemically (22). An apurinic (AP) endonuclease with properties similar to that of exonuclease III is also associated with the Rrp1 protein (39).

The protein sequence deduced from the *Rrp1* cDNA suggests a possible role for the protein in DNA repair. In *E. coli*, exonuclease III is the major AP endonuclease. Mutants in the gene for exo III (*xth*) are slightly sensitive to ionizing radiation, oxidative damage and alkylating agents (40,41), and are inviable when in a *dut* background which causes increased uracil incorporation into DNA (42,43). Therefore, a major *in vivo* function of exonuclease III involves recognition of sites of DNA damage and preparation of the damaged DNA for repair. The ability of the Rrp1 protein to recognize apurinic sites has been demonstrated (39). Additional AP endonucleases from *Drosophila* have been identified, which differ from the Rrp1 protein in polypeptide size, chromatographic and enzymatic properties (44,45). However, an *in vivo* role of the Rrp1 protein may be to facilitate recombinational repair of DNA damage at or adjacent to abasic sites. Such a role is consistent with the expression of this protein in mitotically growing tissue culture cells. The recombination activities associated with Rrp1 protein may also be important in promoting homologous recombination events that are not initiated or induced by DNA damage.

The physical and enzymatic properties of Rrp1 protein suggest that DNA repair and homologous recombination activities reside in a single polypeptide. Further study of this unusual protein will hopefully aid our understanding of the mechanisms by which these two processes are coordinated *in vivo*.

ACKNOWLEDGEMENTS

We gratefully acknowledge Agnes Ayme-Southgate and Katrin Valgeirsdottir for assistance with both library screening and Northern blot analysis. Maxwell Ping-Lee provided the Drosophila embryo cDNA library. We also thank Dr. Burke H.Judd and Dr. Thomas A.Kunkel for their careful readings of

this manuscript. This work was supported by grants to A.R. from NIH, NSF, the Office of Naval Research and ACS. M.S. was supported in part by a postdoctoral fellowship from ACS.

REFERENCES

1. Smith,G.R. (1987) Ann. Rev. Genet., **21**, 179–201.
2. Clark,A.J. (1973) Ann. Rev. Genet., **7**, 67–86.
3. Kolodner,R., Evans,D.H. and Morrison,P.T. (1987) Proc. Natl. Acad. Sci. USA, **84**, 5560–5564.
4. Sugino,A., Nitiss,J. and Resnick,M.R. (1988) Proc. Natl. Acad. Sci. USA, **85**, 3683–3687.
5. Hsieh,P., Meyn,M.S. and Camerini-Otero,R.D. (1986) Cell, **44**, 885–894.
6. Eisen,A. and Camerini-Otero,R.D. (1988) Proc. Natl. Acad. Sci. USA, **85**, 7481–7485.
7. Moore,S.P. and Fishel,R. (1990) J. Biol. Chem., **265**, 11108–11117.
8. Hotta,Y., Tabata,S., Bouchard,R.A., Pinon,R. and Stern,H. (1985) Chromosoma, **93**, 140–151.
9. Ganea,D., Moore,P., Chekuri,L. and Kucherlapati,R. (1987) Mol. Cell Biol., **7**, 3124–3130.
10. Dykstra,C.C., Kitada,K., Clark,A.B., Hamatake,R.K. and Sugino,A. (1991) Mol. Cell Biol., **in press**
11. Esposito,M.S. and Brown,J.T. (1990) Curr. Genet., **17**, 7–12.
12. Alani,E., Padmore,R. and Kleckner,N. (1990) Cell, **61**, 419–436.
13. Prakash,S., Prakash,L., Burke,W. and Montelone,B.A. (1980) Genetics, **94**, 31–50.
14. Engebrecht,J., Hirsch,J. and Roeder,G.S. (1990) Cell, **62**, 927–937.
15. Holliday,R., Halliwell,R.E., Evans,M.W. and Rowell,V. (1976) Genet. Res., **27**, 413–453.
16. Hawley,R.S. (1988) In Kucherlapati,R. and Smith,G. (eds.), Genetic Recombination. American Society for Microbiology, Washington,DC, pp. 497–525.
17. Walker,G.C. (1985) Ann. Rev. Biochem., **54**, 425–457.
18. Sebastian,J., Kraus,B. and Sancar,G.B. (1990) Mol. Cell Biol., **10**, 4630–4637.
19. Eckhardt,F., Moustacchi,E. and Haynes,R.H. (1978) In Hanawalt,P., Friedberg,E. and Fox,C. (eds.), DNA repair mechanisms. Academic Press, Inc., New York, NY, pp. 421–423.
20. Fornace,A.J.Jr., Alamo,I.Jr. and Hollander,M.C. (1988) Proc. Natl. Acad. Sci. USA, **85**, 8800–8804.
21. Akaboshi,E. and Howard-Flanders,P. (1989) Mutation Res., **227**, 1–6.
22. Lowenhaupt,K., Sander,M., Hauser,C. and Rich,A. (1989) J. Biol. Chem., **264**, 20568–20575.
23. Birnboim,H.C. (1983) Methods Enz., **100**, 243–255.
24. Blin,N. and Stafford,D.W. (1976) Nucleic Acids Res., **3**, 2303–2305.
25. Laemmli,U.K. (1970) Nature, **227**, 680–685.
26. Aebersold,R., Leavitt,J., Saavedra,R., Hood,L.E. and Kent,S.B.H. (1987) Proc. Natl. Acad. Sci. USA, **84**, 6970–6974.
27. Compton,T. (1990) In Innis,M.A., Gelfand,D.H., Sninsky,J.J. and White,T.J. (eds.), PCR protocols. Academic Press, Inc., San Diego, CA, pp. 39–45.
28. Maniatis,T., Hardison,R.C., Lacy,E., Lauer,J., O'Connel,C., Quon,D., Sim,G.K. and Efstratiadis,A. (1978) Cell, **15**, 687–701.
29. Nolan,J.M., Lee,M.P., Wyckoff,E. and Hsieh,T. (1986) Proc. Natl. Acad. Sci. USA, **83**, 3664–3668.
30. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
31. Feinberg,A.P. and Vogelstein,B. (1983) Anal. Biochem., **132**, 6–13.
32. Ayme,A. and Tissieres,A. (1985) EMBO J., **4**, 2949–2954.
33. Pardue,M.L., Lowenhaupt,K., Rich,A. and Nordheim,A. (1987) EMBO J., **6**, 1781–1789.
34. Devereux,J., Haeberli,P. and Smithies,O. (1984) Nucleic Acids Res., **12**, 387–395.
35. Query,C.C., Bentley,R.C. and Keene,J.D. (1989) Cell, **57**, 89–101.
36. Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Methods Enz., **185**, 60–89.
37. Puyet,A., Greenberg,B. and Lacks,S.A. (1989) J. Bacteriology, **171**, 2278–2286.
38. Rogers,S.G. and Weiss,B. (1980) Methods Enz., **65**, 201–211.
39. Sander,M., Lowenhaupt,K. and Rich,A. (1991) Proc. Natl. Acad. Sci. USA, (in press)
40. Demple,B., Halbrook,J. and Linn,S. (1983) J. Bacteriology, **153**, 1079–1082.
41. Sammartino,L.J. and Tuveson,R.W. (1983) J. Bacteriology, **156**, 904–906.
42. el-Hajj,H.H., Zhang,H. and Weiss,B. (1988) J. Bacteriol., **170**, 1069–1075.
43. Taylor,A.F. and Weiss,B. (1982) J. Bacteriol., **151**, 351–357.
44. Spiering,A.L. and Deutsch,W.A. (1986) J. Biol. Chem., **261**, 3222–3228.
45. Kelley,M.R., Venugopal,S., Harless,J. and Deutsch,W.A. (1989) Mol. Cell Biol., **9**, 965–973.