Isolation and characterization of a gene encoding DNA topoisomerase I in *Drosophila melanogaster*

Tao-shih Hsieh, Sheryl D.Brown, Patti Huang and Jennifer Fostel
Department of Biochemistry, Duke University Medical Center, Durham, NC 27710, USA

ABSTRACT

We synthesized a DNA probe specific for the gene encoding eucaryotic DNA topoisomerase I by the polymerase chain reaction. The sequences of the primers for this reaction were deduced from the regions with extensive homology among the enzymes from the fission and budding yeasts, and the human. From the clones isolated by screening a Drosophila cDNA library with this DNA probe, two cDNA clones of 3.8 and 5.2 kb were characterized and completely sequenced. Both cDNA sequences contain an identical open reading frame for 972 amino acid residues, The 3.8 kb messenger RNA is likely generated by using a polyadenylation site 5' upstream to that used in generating the 5.2 kb mRNA. The predicted amino acid sequence shows that a segment of 420 amino acid residues at the amino terminus is hydrophilic, similar to the amino terminal 200 residues in the yeast and human enzymes. Furthermore, the Drosophila enzyme is unique in that the amino terminal 200 residues are enriched in serine and histidine residues; most of them are present in clusters. The rest of the Drosophila sequence is highly homologous to those from yeast and human enzymes. The evolutionarily conserved residues are identified and are likely the critical elements for the structure and function of this enzyme. A plasmid vector containing the cloned cDNA was constructed for the expression of Drosophila protein in *Escherichia coli*. The enzymatic and immuno-chemical analysis of the polypeptide produced in this heterologous expression system demonstrated that the expressed protein shares similar enzymatic properties and antigenic epitopes with DNA topoisomerase I purified from Drosophila embryos or tissue culture cells, thus establishing the bacterial expression system being useful for the future structure/function analysis of the Drosophila enzyme.

INTRODUCTION

DNA topoisomerases can affect many, if not all aspects of the biological functions of DNA by their enzymatic activities in modulating the structures of DNA (see recent reviews, 1—3). These enzymes are grouped into two types according to their mechanism of action: type I enzymes work by reversibly breaking one DNA strand at a time, and type II enzymes can break both strands concertedly and pass a segment of DNA through this transient double strand break. During the breakage/rejoining process, DNA topoisomerases can form covalent protein-DNA adducts through phosphodiester bonds between specific tyrosine residues in the enzyme and the phosphate group at the DNA break. Both eucaryotic and procaryotic type II DNA topoisomerases share significant similarity in their structure and mechanism. In contrast, the type I enzymes are more diverse in their structure and mechanism. The ω protein from *Escherichia coli*, representing the majority of topoisomerase I activity in bacterial cells, can only relax highly negatively supercoiled DNA efficiently (4) and the enzyme can form a covalent linkage to the 5'-phosphoryl end at the site of enzyme-mediated DNA cleavage (5). The most abundant nuclear type I DNA topoisomerases like those isolated from mammalian and yeast cells, are proteins of about 100 kD with no apparent sequence homology to the bacterial ω protein. They can readily relax both positive and negative DNA supercoils, and form a covalent linkage to 3' phosphoryl end at the DNA cleavage site (reviewed in 6). In addition to the swivelase function during the replication and transcription, one of the noted biological functions of yeast topoisomerase I, and possibly topoisomerase II, is the suppression of homologous recombination within the rDNA repeats (7,8). Another yeast genetic locus *TOP3*, originally discovered because of the effect of its mutant allele in enhancing recombination between δ sequence repeats of *Ty1* transposons, shares significant homology in its coding sequence with the *E. coli* protein (9) and is enzymatically similar to bacterial topoisomerase III (10).

Type I DNA topoisomerase activity has been isolated from Drosophila embryos and it has the hallmarks of the eucaryotic topoisomerase I activities (11—13). However, the size of the Drosophila enzyme is 135 kD (12,13), significantly larger than 100 kD type I topoisomerases isolated from yeast and mammalian cells. The topoisomerase I isolated from Xenopus oocytes has a molecular weight of 165 kD (14). It is unclear yet whether the Drosophila or Xenopus enzymes are structurally related to the yeast and mammalian enzymes. In order to address this question and to establish a molecular genetic system to investigate the functions of DNA topoisomerase I in Drosophila cells, we have isolated the *top1* cDNA from Drosophila embryos based on its homology to a highly conserved region among the eucaryotic topoisomerases I. The sequence information of the cDNA clone and the heterologous expression data indicate that the cloned gene encodes a eucaryotic DNA topoisomerase I, corresponding to the 135 kD enzyme isolated from Drosophila embryos.

MATERIALS AND METHODS

Synthesis of a *top1* DNA probe by the polymerase chain reaction

The sequences for synthetic DNA primers were generated by reverse translation of peptide sequences in the region of extensive homology among eucaryotic topoisomerases I (Fig. 1). The sequences for oligo 1 and oligo 2, each consisting of 256 and 384 permutations respectively, to reflect all possible codon degeneracies, are 5'TT(T/C)(C/A)G(A/G/C/T)AC(A/G/C|T)-TA(T/C)AA(T/C)GC3' oligo 1, and 5'(T/C)TG(A/G)TG(A/G)-TT(A/G)CA(A/G/C/T)A(A/G)(A/G/T)AT3' oligo 2. The Drosophila genomic DNA sequence between the two primers was amplified via the polymerase chain reaction using standard reaction conditions (15) except the primer concentrations were 17 μM. The annealing of the primers to the Drosophila genomic DNA was carried out at 43°C for 2 min followed by a 3 min ramp from 43°C to 72°C. The DNA synthesis was extended at 72°C for 1 min and the cycle was repeated 25 times. The DNA products were routinely analyzed by electrophoresis on a 5% polyacrylamide gel.

Cloning and sequence analysis of *top1* cDNA

A radioactive DNA probe for eucaryotic topoisomerase I gene was prepared to a specific activity of about 10^9 cpm/μg by the polymerase chain reaction. A Drosophila cDNA library, cloned in the λgt11 vector (16), was screened by the plaque hybridization method (17). DNA sequences were determined by the chain termination method (18) using T7 DNA polymerase (US Biochemical, Cleveland, OH). The *top1* cDNA fragments were subcloned in the Bluescript plasmid vector (Strategene, La Jolla, CA) and their nested deletions were generated by unidirectional exonuclease III digestion (19). The nucleotide sequence data were assembled and processed by the Pustell computer program (International Biotechnologies, Inc., New Haven, CT).

Construction of a plasmid vector for expression of the cloned Drosophila topoisomerase I gene

We removed most of the 5' untranslated region in the cDNA clone by EspI digestion at nucleotide 241. The resulting *top1* cDNA fragment, including the polyA tail, was processed for adding BamHI linkers to the ends and inserted at the BamHI cloning site in the pET3b expression vector (20). The sequence surrounding the 5' end of the cloned segment was confirmed by DNA sequencing. The construction results in an additional 23 amino acids before the initiating methionine in the Drosophila *top1* cDNA. For expressing the cloned *top1* gene, the pET3b recombinant plasmid was transformed into BL21(DE3,LysS) *E.coli* cells and induction of the cloned segment by isopropyl-β-D-thiogalactopyranoside was carried out according to the published procedure (20). *E.coli* cells were harvested and stored at −70°C.

The bacterial cell paste was thawed and lysis ensued due to the presence of endogenous T7 lysozyme. The cell lysate was either processed for analysis by SDS polyacrylamide gel electrophoresis (21) or was centrifuged to remove the cell debris. The cleared lysate was assayed for DNA supercoil relaxation activity in a reaction mixture containing 10mM Tris HCl (pH 7.9), 1mM Na$_3$EDTA, 100mM NaCl, and 25 μg/ml plasmid DNA. The reaction was incubated at 30°C and the time course was monitored. To assay the relaxation activity for positive DNA supercoils, ethidium bromide at a final concentration of 3 μg/ml


```
H    MEGLTAKVFRTYNASITLQQQLKELTAPDENIPAKILSYNRANRAVAILCNHQR    581 - 634

SC   MPGLTAKVFRTYNASKTMQDQLDLIPNKGSVAE KILKYNAANRTVAILCNHQR   508 - 560

SP   MDGLSAKVFRTYNASYTMAEELKKMPKNLTLAD KILFYNRANRTVAILCNHQR   553 - 605

DM           FRTYNASKTLQSQLDLLTDPSATVPEKLLAYNRANRAVAILCNHQ
```

Figure 1. Amino acid sequences in a region of extensive homology among eucaryotic type I DNA topoisomerases. A segment of amino acid sequence for the enzymes from human (H), *Saccharomyces cerevisiae* (SC), *Schizosaccharomycespombe* (SP), and *Drosophila melanogaster* (DM) are shown with the identical residues in bold type. The human and yeasts sequences are from published data (24–26) and numbers on the right indicate the positions in the protein sequences. The Drosophila sequence was deduced from the DNA sequence of the segment amplified between two opposing primers as indicated by the arrows. The primers oligo 1 and oligo 2 correspond to the regions covered by the arrows pointing rightward and leftward, respectively.

was added to a reaction mixture in which the plasmid DNA was already relaxed. Aliquots collected during the reaction time course were extracted with phenol then ether. The DNA products were analyzed by agarose gel electrophoresis as described before (22).

Other methods

Immunoblot analysis of the polypeptides separated by SDS polyacrylamide gel electrophoresis was performed using the published procedure (23). The Drosophila topoisomerase I was purified to nearly homogeneity from either embryos or tissue culture cells using the conventional chromotographic procedures (13). The anti-topoisomerase I antibody was generated by immunizing a rabbit with purified 13S kD Drosophila topoisomerase I (13).

RESULTS

Synthesis of a DNA probe for the gene encoding eucaryotic DNA topoisomerase I

Analysis of the deduced amino acid sequence of the DNA topoisomerase I from *Saccharomyces cerevisiae* (24), *Schizosaccharomyces pombe* (25), and human (26) revealed regions of extensive sequence homology. In order to generate a DNA probe using the polymerase chain reaction (15), two neighboring blocks of protein homology were reverse translated to give sequences for two opposing DNA primers (Fig. 1). These regions were selected because they appear to have the least number of codon degeneracies in the DNA sequences, and the expected separation of the two DNA probes is about 130 nucleotides which is well within the optimal range of DNA synthesis by the polymerase chain reaction. These two synthetic DNA primers, each of which is a mixture of sequences accounting for all possible degeneracies, were used to amplify the DNA sequence between the primers in the Drosophila genomic DNA. Under conditions where annealing of the primers with the Drosophila genomic DNA was at a temperature between 41 and 50°C, a DNA fragment of 135 bp was the predominant amplification product (data not shown). When the annealing temperature was below 37°C, a number of DNA fragments were amplified, presumably due to the lowered stringency in the polymerase reaction cycles. At an annealing temperature above 53°C, no significant amplification product was observed.

To further examine the DNA sequences amplified from the Drosophila genomic DNA, the primary amplification product was purified by polyacrylamide gel electrophoresis and subcloned in
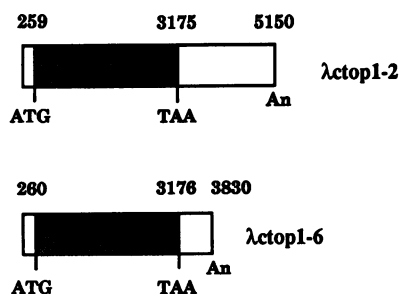
Figure 2. A schematic diagram of the cloned *top1* cDNA sequences. The *top1* cDNA inserts from two recombinant λ phages (λctop1-2 and λctsp1-6) were sequenced and their features are diagramatically summarized. Both sequences have polyA (An) at their 3'-ends and have an identical open reading frame (shaded box) with initiation and termination codons marked. The nucleotide positions for the first nucleotides of these codons and the lengths of the inserts are indicated at the top. The complete nucleotide sequences are deposited in GenBank (M74557).

a plasmid vector. The nucleotide sequence analysis of the cloned insert revealed in addition to the primer DNA sequences used in the amplification reaction, a stretch of deduced sequence of 33 amino acids. The deduced Drosophila sequence shares homology with the other three eucaryotic topoisomerses I in this region (Fig. 1). This result suggests that by amplifying a region between two DNA primers corresponding to highly homologous blocks we have isolated a DNA segment encoding part of the Drosophila DNA topoisomerase I.

Isolation and nucleotide sequence analysis of cDNA clones for Drosophila DNA topoisomerase I

We prepared a radioactively labeled DNA probe using the polymerase chain reaction to amplify the Drosophila genomic DNA sequences within the DNA topoisomerase I coding region (shown in Fig. 1). Using this labeled probe for the topoisomerase I gene, we screened about 1×10^6 plaques from a Drosophila cDNA library cloned into the λgt11 vector. Five independent recombinant plaques gave the strongest hybridization signals throughout several cycles of plaque purification and screening procedures. The phage DNA containing these cDNA clones were prepared. The analysis by restriction enzyme digestion indicated that four of the clones contained a DNA insert of about 5.2 kb while one contained a 3.8 kb DNA insert.

The cDNA inserts of 5.2 kb and 3.8 kb from recombinant phages λctop1-2 and λctop1-6, respectively, were subcloned into a plasmid vector and both strands of the cloned inserts were sequenced by the chain termination method (18). The nucleotide sequence of the 5.2 kb cDNA insert shows that it contains 5150 nucleotides including a stretch of polyA at its end (Fig. 2). The nucleotide sequence of the 3.8 kb cDNA insert is identical to the 5.2 kb cDNA except it terminates at nucleotide 3789 (or 3790/3791) due to the presence of a run of two adenosine residues at these positions and it also contains a stretch of polyA sequence at its end (Fig. 2). The presence of the polyA sequence in both cDNA inserts allows us to assign the polarity of the gene. It therefore appears that 3.8 kb message was probably generated by a cleavage/polyadenylation event occuring 5' upstream to that of 5.2 kb transcript. The 5' end of the 3.8 kb cDNA also contains an additional thymine residue, which was probably derived from the mRNA based on our sequencing information from the genomic clones of *top1*. The complete nucleotide sequence of

A

```
MSGDVAAENS  IHIQNGGSCE  VVQSNGVTTN  GHGHHHHHHS  SSSSSSKHKS  SSKDKHRDRE   60
REHKSSNSSS  SSKEHKSSSR  DKDRHKSSSS  SSKHRDKDKE  RDGSSNSHRS  GSSSSHKDKD  120
GSSSSKHKSS  SGHHKRSSKD  KERRDKDKDR  GSSSSSRHKS  SSSSRDKERS  SSSHKSSSSS  180
SSSKSKHSSS  RHSSSSSSKD  HPSYDGVFVK  PEPVSQQLMH  SGSVDAFQMQ  QLGSYEAAAA  240
GTNFNGNGNV  AGANYKNGYE  ESIVDIKKEE  ESFNNLSQAS  SCDYSMSQFR  ADEPPFVVKH  300
EQSYAEEDST  MNYNDHDDDA  DEMNDDEEDV  PLAMRKRKQE  ATDRPDGGMD  NDDDDDIPLL  360
ARKKVKKEKI  KKESKEKSKK  RVKEEPSDDY  GNVKPKKKKM  KKEPEPAVSP  GKRQKAKAKV  420
EEEEVWRHHE  EEKRADGVKH  STLEHKGPVF  APRYERVPRN  VRFYYDGKPL  ELSEETEEAA  480
TFYAKMLNHD  YCTKEVFNNN  FFKDFRKSMT  PREREIIKDF  RKCNFQEHFN  YFQAESKRK   540
AASKEEKLIK  KNENEALMKE  FGFCMIDGHK  EKIGNFRLEP  PGLFRGRGEH  PKMGMIKRRI  600
QASDVSINCG  KDSKVPSPPP  GSRWKEVRHD  NTYTWLASWI  ENVQGQVKYI  MLNPSSKLKG  660
EKDHIKYETA  BRLDKVIDKI  RATYRDEWKS  KEMRVRQRAV  ALYFIDKLAL  RAGNEKDEDQ  720
ADTVGCCSLR  VEHVQLHKEL  NGKENVVVFD  FPGKDSIRYY  NEVEVEKRVF  KNLELFMEHK  780
KEGDDLFDRL  NTQVLNEHLK  ELHEGLTAKV  FRTYNASKTL  QSQLDLLTDP  SATVPEKLLA  840
YNRANRAVAI  LCNHQRSVPK  SHEKSMENLK  EKIKAKREAI  EKCESEYHSR  DEKKGKQLER  900
LRDQLKKLEL  QETDRDENKT  IALGTSKLNY  LDPRISVAWC  KKHDVPIEKI  FNKTQRTKFL  960
HAVHMADENY  RF                                                        972
```

B

Key: ▨ Serine/Histidine Rich Region
☒ Hydrophilic Region
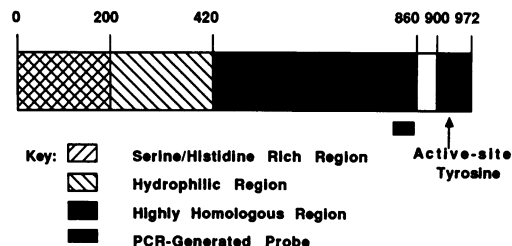■ Highly Homologous Region
■ PCR-Generated Probe

Figure 3. A. Deduced amino acid sequence of Drosophila topoisomerase I. The underlined residues indicate those that are conserved in identity comparing the sequences of yeasts and human enzymes. The homologous substitutions which tend to be distributed around these conserved residues, are not indicated here. The conserved tyrosine at position 930 was shown to be involved in the protein/DNA linkage for the homologous yeast enzyme (29,30). **B.** A schematic diagram of the structural features of the Drosophila enzyme. The homologous regions were mapped with the help of the Pustell sequence comparison program (33) by pairwise searches between the Drosophila sequence and sequences of other eucaryotic nuclear enymes. Notice the amino terminal 200 residues are both hydrophilic and rich in Ser/His.

top1 cDNA was deposited in the GenBank database with an accession number M74557.

In both cDNA sequences, the first ATG codon from their 5' ends is located at nucleotide 259 (or nucleotide 260 for the 3.8 kb cDNA insert) and it initiates a contiguous open reading frame with 972 amino acid residues (Fig. 2). The sequence context of this initiation codon, CAAA<u>ATG</u>A, is in general agreement with the consensus start codon sequence, (C/A)AA(A/C)<u>ATG</u>(A/G), in Drosophila (27). The termination codon TAA is located at nucleotide position 3175 and thus leaves a long untranslated region of 1951 nucleotides before the polyA end of the 5.2 kb clone. The difference in the structures between these two cDNA clones is primarily in the length of the untranslated region at the 3' ends; for the 3.8 kb cDNA the untranslated region at the 3' end is 614 nucleotides long. The 3' end untranslated region is rich in adenine and thymine (A/T composition 64%). There is no perfect match to the polyA signal motif AATAAA upsteam to the polyadenylation sites in both cDNA species. However, sequences that are closely related to this motif and can function as polyadenylation signals (28) can be located upstream to both poly A addition sites.

The expression of *top1* and the sizes of mRNA species were analyzed by RNA blot hybridization experiments using cloned cDNA sequences as probes. The *top1* messages in the RNA prepared from Drosophila embryos consist of two major species
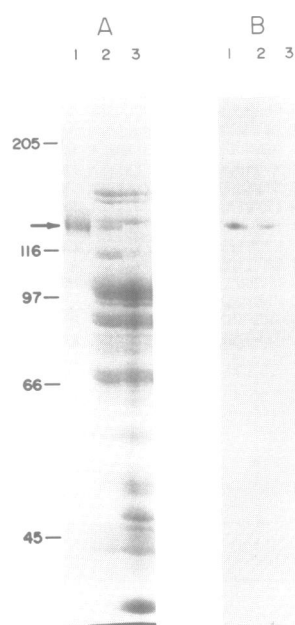
Figure 4. Analysis of the polypeptides induced in cells harboring an expression vector for Drosophila *top1* cDNA. Duplicate portions of an SDS polyacrylamide gel were either stained with Coomassie blue dye (Panel A) or processed for immunoblotting with anti-topoisomersae I antibody (Panel B). Lane 1 shows the purified 135 kD topoisomerase I from Drosophila cells. Also shown are *E. coli* cell extracts prepared under conditions where the cloned *top1* cDNA was induced for expression by the addition of IPTG (lane 2) or without induction (lane 3). The lines and numbers on the left mark the positions and sizes (in kD) of the protein size markers used in this experiment. The arrow highlights the position of 135 kD polypeptide.
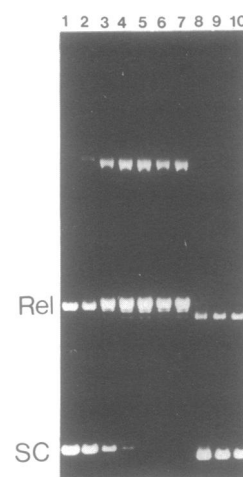

Figure 5. DNA supercoil relaxation assay of the *E. coli* extract containing the polypeptide from the expressed Drosophila *top1* cDNA. The plasmid DNA substrate was incubated with the *E. coli* cell extract and aliquots were taken at timepoints 10, 20 and 30 min. (lanes 2−4). After a 30 min. incubation the reaction mixture was split into two parts. Incubation was continued for one half of the sample with the timepoints taken at another 10, 20 and 330 min. (lanes 5−7) and a saturating amount of ethidium bromide was added to a concentration of 3 μg/ml to the other half with identical timepoints taken (lanes 8−10). The substrate plasmid DNA is in lane 1 and the positions of supercoiled and relaxed DNA are marked by SC and Rel, respectively. The DNA species migrating slower than the relaxed monomer plasmid (Rel), evident in lanes 2−7, are the relaxed dimer DNA. The supercoiled dimer comigrates with the relaxed monomer under these gel electrophoretic conditions.

with sizes of 5.2 kb and 4.0 kb (data not shown). The 5.2 kb message appears to be homogeneous in size. It comigrates with the 5.2 kb RNA size marker generated by *in vitro* transcription of the cloned cDNA segment and therefore, the 5.2 kb cDNA cloned is likely to be full-length or nearly so. There is also a population of *top1* messages with an average size of 4.0 kb. Based on the results of 3.8 kb cDNA sequences, it would be interesting to establish in future experiments if the population of 4.0 kb RNA are all generated by using different cleavage/polyadenylation sites or by other mechanisms as well.

Homology and structural comparison of Drosophila topoisomerase I to other eucaryotic topoisomerases I

The deduced structure of Drosophila topoisomerase I contains 972 amino acid residues (Fig. 3A), which is significantly longer than the yeast and human : topoisomerases I, ranging between 765 and 812 amino acid residues. The Drosophila DNA topoisomerase I is clearly evolutionarily related to other eucaryotic topoisomerases I. We carried out pairwise sequence comparisons between the Drosophila enzyme and the enzymes from fission and budding yeasts, and from man (29). These comparison searches indicate that there is a high level of identity, approximately 40%, among the amino acid sequences of these eucaryotic DNA topoisomerses I. The conserved amino acid residues that are identical in the sequences of these four nuclear topoisomerases I are marked in Drosophila sequence (Fig. 3A). Using Drosophila topoisomerase I sequence as a reference, the regions of homolgy are concentrated in two stretches, residues

420 to 860 and 900 to 972 (Fig. 3, A and B). The sequences used in generating the screening probe are located toward the carboxyl terminal in the first homology block, and the tyrosine residue shown to be involved in the protein/DNA linkage for the yeast topoisomerase I (30,31) is located in a region of extensive homology near the center of the second block (Fig. 3B). The type I DNA topoisomerases from yeast and human are also characterized by a hydrophilic domain of about 200 amino acid residues at the amino terminals (24−26). The first 420 residues for the Drosophila enzyme are enriched in charged amino acids (Fig. 3 B). However, the Drosophila enzyme is uniquely characterized by clusters of serines and histidines in a stretch of about 200 residues at its amino terminus (see Fig, 3A). In this region, nearly half of the amino acids are either serines (39.5%) or histidines (11.5%). It therefore appears that in comparison with other eucaryotic enzymes, Drosophila topoisomerase I has an additional Ser/His rich hydrophilic domain of 200 amino acid residues at its amino terminus. This extra stretch of amino acids could also account for the size difference between the 135 kD Drosophila enzyme and the 100 kD enzymes from yeast and man.

Expression of cloned Drosophila *top1* in *E. coli* cells

We have established heterologous expression systems to further confirm that the gene cloned corresponds to a Drosophila topoisomerase I gene. The *top1* cDNA insert was engineered under the control of phage 17 promoter in a pET3b expression vector (20). The construction of the recombinant *E. coli* expression vector is described in MATERIALS AND METHODS. As a result of adding linker sequences and the amino terminal peptide from T7 gene 0 protein during the subcloning

steps, there are an additional 23 amino acid residues preceding the initiating methionine of topoisomerase I. The expected size of the fusion protein produced in *E.coli* is 995 amino acid residues.

Total protein from *E.coli* cells harboring the recombinant pET3b plasmids was analyzed by polyacrylamide gel electrophoresis (Fig. 4). Part of the gel was stained with Coomassie blue dye (Panel A, Fig. 4), while an identical half was processed for immunoblotting (Panel B, Fig. 4). The antibody used for the immunoblot was obtained by immunizing a rabbit with a sample of the 135 kD topoisomerase I purified from Drosophila cells (13). Under inducing conditions, a polypeptide with a size of about 135 kD was synthesized (lane 2 in Fig. 4). This polypeptide was not present in *E.coli* cells before induction (lane 3 in Fig. 4), nor in the cells without the cloned *top1* cDNA (data not shown). The induced polypeptide comigrates with the 135 kD topoisomerase I isolated from Drosophila cells (lane 1 in Fig. 4). Furthermore, the anti-topoisomerase I antibody cross-reacts with the polypeptide induced in *E. coli* cells as well as with the isolated Drosophila enzyme (lanes 1 and 2, Fig. 4B). Therefore the protein encoded by the cloned Drosophila gene must be closely related to the Drosophila topoisomerase I. We have also constructed a yeast expression system in which the cloned gene is under the control of *GAL1* promoter in a yeast-*E. coli* shuttle vector described earlier (32). A novel polypeptide of 135 kD, recognized by the antitopoisomerase I antibody, can be induced in yeast cells growing in galactose-containing media (data not shown). This result suggests that the translational initiation signal in the cloned cDNA sequence can be recognized by another eucaryotic expression system.

The eucaryotic topoisomerases I, including the Drosophila enzyme, are catalytically active in the absence of a divalent cation and can relax both positively and negatively supercoiled DNA (reviewed in 6). These properties distinguish the eucaryotic topoisomerases I from the bacterial enzymes. For instance, the most potent topoisomerase I activity from *E. coli*, the ω protein, can only relax highly negatively supercoiled DNA and this relaxation activity depends on the presence of Mg^{++} (4). We have taken advantage of these differences in eucaryotic and bacterial topoisomerases I to identify whether the polypeptide encoded by the cloned Drosophila gene possesses any eucaryotic topoisomerase I activity. The cleared lysates from *E. coli* cells grown under inducing conditions contained an activity capable of relaxing negatively supercoiled DNA in the presence of 1mM EDTA and without any added divalent cations (Lanes 2−7, Fig. 5). Under identical conditions, the cell lysates prepared from uninduced cells contained no such relaxation activity (data not shown). At the timepoint when most of the negatively supercoiled DNA was relaxed, the DNA sample was split into two aliquots. Incubation was continued under identical conditions for half of the sample (lanes 5 to 7, Fig. 5), while a saturating amount of ethidium was added to the other half of the sample (lanes 8 to 10, Fig. 5). The presence of a DNA intercalator e ethidium bromide can convert the relaxed DNA into positively supercoiled DNA, relaxation of which can again lead to the formation of highly negatively supercoiled DNA after the removal of bound ethidium. Analysis of the DNA sample after the additon of ethidium bromide demonstrated the generation of tightly supertwisted DNA from the relaxed DNA molecules (lanes 8−10, Fig. 5), thus suggesting the presence of an activity in these crude extracts capable of removing positive DNA supertwists.

These activities are hallmarks for eucaryotic DNA topoisomerase I and therefore, the gene cloned here indeed encodes Drosophila topoisomerase I. Furthermore, the efffciency in the expression of the cloned *top1* suggests the potential use of of this system in the structure/function analysis of the Drosophila enzyme.

DISCUSSION

We designed a set of DNA primers to synthesize a DNA probe specific for the eucaryotic topoisomerase I gene. The sequences of these primers were selected based upon the amino acid sequences in the region with extensive identity among the yeasts and human DNA topoisomerases I. A specific DNA fragment can be efficiently amplified from the Drosophila genomic DNA by using these primers under typical conditions for polymerase chain reactions. It is possible that the same set of primers can be used to amplify a similar region in the DNA topoisomerase I gene from other eucaryotic sources. We have isolated several Drosophila cDNA clones using the DNA probes generated from these primers. The cloned cDNA segment in heterologous expression systems, can direct the synthesis of a novel 135 kD polypeptide that shares both enzymatic properties and antigenic determinants with the 135 kD DNA topoisomerase I isolated from Drosophila cells, thus confirming the identity of the cloned DNA sequence. The nucleotide sequences of the cDNAs reveal an open reading frame of 972 amino acid residues with extensive homology to the yeasts and human enzymes. The amino terminal 420 residues in the Drosophila enzyme are hydrophilic, a property also shared by 200 residues at the amino termini of the yeasts and human enzymes. We have identified the amino acid residues that are conserved in the four sequences of nuclear enzymes from human, fruit fly, and yeasts (Fig. 3A). These highly conserved residues may correspond to regions critical for topoisomerase I functions and as such, they will be interesting targets for future structure/function analysis.

In addition to the regions of extensive homology, the Drosophila protein has a unique region at its amino terminus. The first 200 amino acid residues of the Drosophila enzyme are not only hydrophilic but also are characterized by a high content of serine and histidine residues, some of which are present in runs of identical residues. The biological functions of these clustered amino acids are unknown. It is plausible that the phosphorylation of serines may provide a mechanism for regulating topoisomerase I functions, and metal chelation by histidines may modulate the interactions of these residues with nucleic acids.

The analysis of genomic clones, isolated by screening a Drosophila DNA library with *top1* cDNA probes, is currently underway. Both *in situ* polytene chromosome hybridization experiments and genomic Southern hybridization results indicate that the topoisomerase I is encoded by a single copy gene in Drosophila (unpublished data). This information should facilitate our efforts in isolating mutants with deficiencies at the *top1* locus and allow further analysis of the biological functions of topoisomerase I in a multicellular organism.

ACKNOWLEDGEMENTS

REFERENCES

1. Maxwell, A. and Gellert, M. (1986). Adv. Protein Chem. *38*, 69-109.
2. Cozzarelli, N. R. and Wang, J. C. (1990). DNA topology and its biological effects. Cold Spring Harbor Laboratory Press.
3. Wang, J. C. (1991). J. Biol. Chem. *266*, 6659−6662.
4. Wang, J. C. (1971). J. Mol. Biol. *55*, 523−533.
5. Depew, R. E., Liu, L. F. and Wang, J. C. (1978). J. Biol. Chem. *253*, 511−518.
6. Champoux, J. J. (1990). *Mechanistic aspects of type I topoisomerases.* In: *DNA Topology and its Biological Effects.*, pp. 217−242., N. R. Cozzarelli and J. C. Wang, eds. Cold Spring Harbor Lab Press, NY.
7. Christman, M. F., Dietrich, F. S. and Fink, G. R. (1988). Cell *55*, 413−425.
8. Kim, R. A. and Wang, J. C. (1989). Cell *57*, 975−985.
9. Wallis, J. W., Chrebet, G., Brodsky, G., Rolfe, M. and Rothstein, R, (1989). Cell *58*, 409−419.
10. Kim, R. A. and Wang, J. C. (1992). J. Biol. Chem. *267*, 17178−17185.
11. Baase, W. A. and Wang, J. C. (1974). Biochem. *13*, 4299−4303.
12. Javaherian, K., Tse, Y. -C. and Vega, J. (1982). Nucleic Acids Res. *10*, 6945−6955.
13. Fostel, J. M. (1990). Isolation and characterization of type I DNA topoisomerase from *Drosophila melanogaster*. Ph.D. thesis, Duke University.
14. Richard, R. E. and Bogenhagen, D. F. (1989). J. Biol. Chem. *264*, 4704−4709.
15. Saiki, R. F., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Science *239*, 487−491.
16. Young, R. A. and Davis, R. W. (1983). Proc. Natl. Acad. Sci. USA *80*, 1194−1198.
17. Benton, W. D. and Davis, R. W. (1977). Science *196*, 180−182.
18. Sanger, F., Nicklen, S. and Coulson, A. R. (1977). Proc. Natl. Acad. Sci. USA *74*, 5463-5467.
19. Henikoff, S. (1984). Gene *28*, 351-359.
20. Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorf, J. W. (1990). Methods in Enzymology *185*, 60−89.
21. Laemmli, U. K. (1970). Nature *227*, 680-685.
22. Sander, M. and Hsieh, T. (1983). J. Biol. Chem. *258*, 8421−8428.
23. Towbin, H., Staehlin, T. and Gordon, J. (1979). Proc. Natl. Acad. Sci. USA *82*, 4374−4378.
24. Thrash, C., Bankier, A. T., Barrell, B. G. and Sternglanz, R. (1985). Proc. Natl. Acad. Sci. USA *82*, 4374−4378.
25. Uemura, T., Morino, K., Uzawa, S., Shiozake, K. and Yanagida, M. (1987) Nucleic Acids Res. 15, 9727−9739.
26. D'Arpa, P., Machlin, P. S., Ratrie III, H., Rothfield, N. F., Cleveland, D. W. and Earnshaw, W. C. (1988). Proc. Natl. Acad. Sci. USA 85, 2543−2547.
27. Cavener, D. R. (1987). Nucleic Acids Res. 15, 1353−1361.
28. Sheets, M. D., Ogg, S. C., and Wickens, M. P. (1990). Nucleic Acids Res. 18, 5799−5805.
29. Pearson, W. R. and Lipman, D. H. (1988). Proc. Natl. Acad. Sci. *USA* 85, 2444−2448.
30. Lynn, R. M., Bjornsti, M. -A., Caron, P. R. and Wang, J. C. (1989). Proc. Natl. Acad. Sci. USA 86, 3559−3563.
31. Eng, W. -k., Pandit, S. D. and Sternglanz, R. (1989). J. Biol. Chem. *264*, 13373−13376.
32. Wyckoff, E. and Hsieh, T. (1988). Proc. Natl. Acad. Sci. USA 85, 6272−6276.
33. Pustell, J. and Kafatos, F.C. (1984). Nucleic Acids Res. 12, 643−655.