Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 12681–12685, December 1994
Biochemistry

A *Drosophila* protein homologous to the human p70 Ku autoimmune antigen interacts with the P transposable element inverted repeats

(fruit fly/DNA repair/transposition/double-strand break)

EILEEN L. BEALL*, ARIE ADMON†‡, AND DONALD C. RIO*§

*Department of Molecular and Cell Biology and †Howard Hughes Medical Institute, 401 Barker Hall, University of California, Berkeley, CA 94720-3204

Communicated by Allan C. Spradling, September 13, 1994 (received for review July 29, 1994)

ABSTRACT **P transposable elements in *Drosophila* are mobilized via a cut-and-paste mechanism. This mode of transposition requires repair of both a double-strand break at the donor DNA site and gapped DNA at the target site. Biochemical studies have identified a cellular non-P element-encoded DNA binding protein, termed the inverted repeat binding protein (IRBP), that specifically interacts with the outer half of the 31-bp terminal inverted repeats. Protein sequence information was used to isolate cDNA clones encoding IRBP. Sequence analysis shows that IRBP is related to the 70-kDa subunit of the human Ku autoimmune antigen. The mammalian Ku antigen binds free DNA termini and has been implicated in immunoglobulin VDJ recombination, DNA repair, and transcription. In addition, Ku is the DNA binding subunit of the double-strand DNA-dependent protein kinase. Cytogenetic mapping indicates that the IRBP gene maps to chromosomal position 86E on the right arm of the third chromosome.**

Transposable elements are mobile segments of DNA found in many prokaryotic and eukaryotic organisms (for review, see ref. 1). Highly organized nucleoprotein complexes between transposable element-encoded proteins and the transposon termini carry out the catalytic events of transposition: cleavage of the donor transposon DNA and covalent joining of the transposon to the target DNA in a strand transfer reaction involving a single step transesterification reaction (for review, see ref. 2).

Because of its widespread use as genetic tools and as vectors for gene transfer, the P element family of transposable elements in the fruit fly, *Drosophila melanogaster*, is one of the most studied eukaryotic transposons (3). P elements are found in natural populations and can range in size from about 0.5 to 2.9 kb. When these elements transpose, a DNA intermediate is formed in which an 8-bp target site duplication is created upon insertion. The P element termini are required for transposition, which include the 31-bp terminal inverted repeats, the 11-bp internal inverted repeats, and unique DNA sequences encompassing ≈150 bp at each end (3). Complete full-length P elements encode an 87-kDa sequence-specific DNA-binding transposase protein that recognizes internal sites at each end (4). Recently, a faithful *in vitro* P element transposition system was developed which showed that GTP and Mg^{2+} serve as cofactors for the reaction (5). Consistent with these biochemical studies, genetic experiments suggested that P elements transpose via a cut-and-paste mechanism similar to the bacterial transposons Tn*10* and Tn*7* (6–9).

The model for P element transposition suggests that a double-strand break is left at the donor site after an excision event (5, 6). This double-strand break must be repaired to prevent chromosomal loss. When a homologous chromosome is present, this repair is hypothesized to occur by a gene conversion type of mechanism similar to double-strand gap repair (6). This repair process must involve non-P element-encoded factors. Indeed, it is known that many bacterial transposition reactions are often aided by host-encoded proteins. In the case of P elements, genetic studies have shown that a gene called *mus101* (for mutagen sensitive) can affect the recovery of chromosomes that have undergone P element excision (10). In addition, several other *mus* mutants are known to be involved in DNA excision or postreplication repair. Biochemical assays provide another approach to identify host cell components involved in P element transposition. Using the P element inverted repeats as a probe, a *Drosophila* DNA binding protein was identified that bound specifically to the outer half of the P element inverted repeats (11). Because the inverted repeats are absolutely required for transposition and are not bound by the P element-encoded transposase, the inverted repeat binding protein (IRBP) is a good candidate for a host factor involved in transposition.

To begin an investigation of the potential role of IRBP in P element transposition, the protein was purified and subjected to microprotein sequence analysis. This sequence information was used to isolate cDNA clones encoding IRBP.¶ The protein bears significant primary sequence homology to the 70-kDa subunit of the human Ku autoimmune antigen. This mammalian heterodimeric DNA binding protein complex was identified by using human patient sera from autoimmune syndrome disease patients (12–14). The human Ku antigen binds both at DNA termini and at internal sites *in vitro* (14–16) and has been implicated in DNA repair, transposition, transcription, and DNA replication (13, 17, 18). Recently, the human Ku antigen was shown to be a subunit of the double-strand DNA-dependent protein kinase (dsDNA PK) (19, 20), which can phosphorylate a number of nuclear proteins *in vitro* (21–23). However, the physiological substrate(s) of the dsDNA PK is still unknown. Furthermore, recent studies with mutant x-ray-sensitive mammalian cells that are defective for double-strand break repair and immunoglobulin VDJ (variable–diversity–joining) recombination (24, 25) have shown that the Ku antigen is either absent or defective in these mutant cell lines. Reversion to x-ray resistance restores Ku immunoreactivity and terminal DNA binding activity (26–28). Our analysis suggests that the IRBP cDNA encodes a DNA binding protein that recognizes the P element inverted repeats. Cytogenetic mapping localized the IRBP gene to region 86E2-3 on the third chromosome. Interestingly, this region contains a mutation, *mus309*, that is

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IRBP, inverted repeat binding protein; dsDNA PK, double-strand DNA-dependent protein kinase; V, variable; D, diversity; J, joining.

‡Present address: Department of Biology, Technion Institute, Haifa, 32000, Israel.
§To whom reprint requests should be addressed.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. U15004).

```
          10          20          30          40          50          60          70          80          90
CCC CCC CCC CCC GAA TCG ATA ACA ACA GCG ACG ATA AGC GCG CCC GCC TTC TTG TGA AAT TAG CGG AAA GTC AGA TTT TTA GCA ATC TAA AGA AGC CGC

100         110         120         130         140         150         160         170         180         190
ATA ATG AGC ACC TGG AAT CCG GAG AAC GAT GTG GAC CTG CTG TCT GGG TCC GAG GAC GAG GAG GAT GTG TCC ATG AAG CGG GAC TAC CAT GGG CGC GAG
     M   S   T   W   N   P   E   N   D   V   D   L   L   S   G   S   E   D   E   E   D   V   S   M   K   R   D   Y   H   G   R   E    32

200         210         220         230         240         250         260         270         280         290
GCC ATT CTG TTC GTG GTA GAC GCC AAT CTT CAG ACA GCC GGC GTG GAG CGC CTG TTG GAG GCA CTG AAC ATC ATC CGG ACG GCC TTT ATA TCC GGA CTT
 A   I   L   F   V   V   D   A   N   L   Q   T   A   G   V   E   R   L   L   E   A   L   N   I   I   R   T   A   F   I   S   G   L    65

300         310         320         330         340         350         360         370         380         390
CTG GTT AAC GAC AAG GAC CTC ATC GGA CTC ATC TTC GCC AAC ACC AAG CAC AGT CCG CCG CCG CTG GAA GCC AGT GCA TTG GAC AAC ATC GTA ATG CCG
 L   V   N   D   K   D   L   I   G   L   I   F   A   N   T   K   H   S   P   P   P   L   E   A   S   A   L   D   N   I   V   M   P    98

 400        410         420         430         440         450         460         470         480         490
GAT AAC TGC GCA GTG TTC TTG CCC CTT CGC CAA CTA ACC AAA CCC ATT GTG GAG CAC TAT CTG GAA TTC ATG GGC GGG GTG GAG ACG CAG TTC GCC GAT
 D   N   C   A   V   F   L   P   L   R   Q   L   T   K   P   I   V   E   H   Y   L   E   F   M   G   G   V   E   T   Q   F   A   D   131

 500        510         520         530         540         550         560         570         580         590
GTG TAT GGC CTG GCG GAA CCC GAT GGT CGC GGC AGG TTT GAC CTT ATG ATC CGG CTC TGC ATC GAG ATG CTG GAA AAG TGC GGC AAG AAG CTA AAC AAC
 V   Y   G   L   A   E   P   D   G   R   G   R   F   D   L   M   I   R   L   C   I   E   M   L   E   K   C   G   K   K   L   N   N   164

  600       610         620         630         640         650         660         670         680         690
GCC AAG ATC GCC TAT GTC ACG GAC GTC AGG GAA CCT CAT CCA TCG AAC AGC AAT CAC TTC CAG GCT GCC CTG CAA AAG GCC AGC GAT CTG GAG GGC AAG
 A   K   I   A   Y   V   T   D   V   R   E   P   H   P   S   N   S   N   H   F   Q   A   A   L   Q   K   A   S   D   L   E   G   K   197

   700      710         720         730         740         750         760         770         780         790
GAG TTC GAG TTT CAT GTC ATT CCC ATG GTC GAT GAC TTT GAC TAC GAG CCG TTT TAC AAG GAG TTC ATC ACG TTG TCA AGA GCT ATC GAA CTG GAC GCC
 E   F   E   F   H  _V___I___P___M___V___D___D___F___D___Y_  E   P   F   Y   K   E  _F___I___T___L___S_  R  _A___I___E___L___D___A_  230

   800      810         820         830         840         850         860         870         880         890
TTC CAG GTG CCA GAT GCC CAG ATG CTG CGC GAA ATC CTG TCC GAT CGT AAG TTG AAG CAG GAT TTC CTT CGC CGA TGC CTG GGC CAC TTC AGT TTT TAT
_F___Q___V_  P   D   A   Q   M   L   R   E   I   L   S   D   R   K   L   K   Q   D   F   L   R   R   C   L   G   H   F   S   F   Y   263

    900     910         920         930         940         950         960         970         980         990
CTG GGC CCC AAC CTA TCC ATG TCC GTA CAG TAC TAC AAT TAC TTT CAG CGA CGC GCC TAT CCG CGC AAA GTG CAA ATC CTG CGC AGG GAC AAC AGT GTG
 L   G   P   N   L   S   M   S   V   Q   Y   Y   N   Y   F   Q   R   R   A   Y   P   R   K   V   Q   I   L   R   R   D   N   S   V   296

        1000        1010        1020        1030        1040        1050        1060        1070        1080
GTG CGT ACT AAG CGA GTG ATT ACG GTG CAA AAA CAA AAG GAC GAT GGC TCG CAG GAT ATC GAG CAC GAG TAT CAG ATT AAG GTG ACG GGC GGT TGG TAC
 V   R   T   K   R   V   I   T   V   Q   K   Q   K   D   D   G   S   Q   D   I   E   H   E   Y   Q   I   K   V  _T___G___G___W___Y_  329

1090    1100        1110        1120        1130        1140        1150        1160        1170        1180
ACT TGC AAC GTG GGC GAA AGG GAT CTG CGC ATC AGC ATG GAT CAG TTG AAC AGG GTG CGC AAT CTG CAC AAG CCG CAA ATG ATG CTG CTG GGC TTC AAG
_T___C___N___V___G___E_  R   D   L   R   I   S   M   D   Q   L   N   R   V   R   N   L   H   K   P   Q   M   M   L   L   G   F   K   362

1190    1200        1210        1220        1230        1240        1250        1260        1270        1280
CAT CGA TCC TCT CTT CCC GAA GTT AGC TAC ATC AAG CCA GCG AAT TTC ATG TAC CCC GAT GAT CAG AGC ATC ATC GGA TCG AAG CGC TTG TTC CGC GCA
 H   R   S   S   L   P   E   V   S   Y   I   K   P   A   N   F   M   Y   P   D   D   Q   S   I   I   G   S   K   R   L   F   R   A   395

1290     1300        1310        1320        1330        1340        1350        1360        1370        1380
TTG TGG GAA CGA TGC TTG GTG CGC GAC AAG ATT GCC ATT TGC CTG TTC ATG TGC AAG CGC AAG TCG ATA CCT CGC TAT GTG GCA CTT GTG CCA GTA GAG
 L   W   E   R   C   L   V   R   D   K   I   A   I   C   L   F   M   C   K   R   K   S   I   P   R  _Y___V___A___L___V___P___V___E_  428

1390     1400        1410        1420        1430        1440        1450        1460        1470        1480
GCC CCA GAT AAT GGG GAA GAT AAG AAC TAT CGC TCT CTG CTC TGC GGT GAC GGA TTC AAG ATT GTC TAC TTG CCG GAG GCC AAG CAC ATC CGC CAC CTA
_A___P___D___N___G___E___D___K___N___Y___R_  S   L   L   C   G   D   G   F   K  _I___V___Y___L___P___E___A___K_  H   I   R  _H___L_  461

 1490      1500        1510        1520        1530        1540        1550        1560        1570        1580
GAC TTG CAG GAC TGG AAC AAT ACG GAA AAC ACT GCT GAC GAA CAG AAA GTC GAG TTT TTC CAA AAG ATC ATC AAG AAG CTG CGC GTT GAC TAT CAG CCG
_D___L___Q___D___W___N___N___T___E___N___T___A___D___E___Q___K___V___E___F___F___Q_  K   I   I   K   K   L   R   V   D   Y   Q   P   494

 1590      1600        1610        1620        1630        1640        1650        1660        1670        1680
AAT CTC ATC AAC GAT CCA AGT CTG GAC GCC CTG CAG GCG AAT CTT CTG GCC CTC TCC CTG GAC TTT TCG ACA GAT ACT AAA GGA CTC GAT AAT CTG CTG
 N   L   I   N   D   P   S   L   D   A   L   Q   A   N   L   L   A   L   S   L   D   F   S   T   D   T   K   G   L   D   N   L   L   527

   1690      1700        1710        1720        1730        1740        1750        1760        1770        1780
GAC ACT TCG CAA CAG GAC AAG CGC ATA GAA AAG CTG CTG CCG GAC TAT GAA ATG TTC GCT CCG GAA GCA GAA CCC CCT AAG AAG CGA GCA GCC AAG TCC
 D   T   S   Q   Q   D   K   R   I   E   K   L   L   P   D   Y   E   M   F   A   P   E   A   E   P   P   K   K   R   A   A   K   S   560

   1790      1800        1810        1820        1830        1840        1850        1860        1870        1880
ACC ACA GCG GGA GCG AGC GGT CCC AAA ATG GCC AAG ATC GAT GAT GAT CAA CTC AAG GAA TTC GAA TTC GTA AAG AGC CTG AAC AAG GAT GAG GCT CTG
 T   T   A   G   A   S   G   P   K   M   A   K   I   D   D   D   Q   L   K   E   F   E   F   V   K  _S___L___N___K___D___E___A___L_  593

   1890      1900        1910        1920        1930        1940        1950        1960        1970        1980
ACG AGT TGC ACG GCC GCC CAG TTG CAT TTT ATT CTG CAG CAT CAC TTC GAT GTT ACA ATG CCC AAG TCG TCA AAG AAG GCA AAA CTG GTC GCC AAA ATC
_T___S___C___T___A___A___Q___L_  H   F   I   L   Q   H   H   F   D   V   T   M   P   K   S   S   K   K   A   K   L   V   A   K   I   626

        1990        2000        2010        2020        2030        2040        2050        2060        2070
GAG GAA CTA CAT AAG TAA TTA AAA TAT TCC ATA CTT TGA GTG ATA AAT CTG TAC CAA TCT AGC TTG TAT TAA CCA TTG GTA AAC TTA GGA TTA GAA AAT
 E   E   L   H   K   *                                                                                                               631

2080    2090        2100        2110        2120        2130        2140        2150
AAT GTT TTG TTG TCC GAA TTG TTA ATT AAA TAC ATA CCA ACA CAT AAT TAC TTG TAA AAT AAA AAA AAA AAA A
```

FIG. 1. DNA and deduced amino acid sequences of IRBP. The sequence of one full-length cDNA is shown as well as the deduced amino acid coding sequence of 631 amino acids. Peptides derived from trypsin and V-8 protease digestion followed by amino acid sequence analysis are underlined. The three peptides used to make anti-peptide antibodies are shown in italics.

involved in DNA repair (29, 30). Finally, the divergence between IRBP and human Ku 70 may explain why P elements have not been found to transpose in mammalian cells.

MATERIALS AND METHODS

Protein Purification, Microsequencing, and Peptide Antibody Production. The IRBP was purified by heparin agarose and several cycles of DNA affinity chromatography as described (11). These fractions were concentrated and subjected to SDS/polyacrylamide gel electrophoresis. Electroblotting, protease treatment, and HPLC were as described (31, 32). Several of these peptide peaks were sequenced on a model 477A sequenator (Applied Biosystems). Three sequences were chosen (IRBP1, VIPMDVAFIYC; IRBP2, AIELDAFQVC; IRBP3, SIVHPSYNC) for peptide antibody generation. Polyclonal antiserum was raised in New Zealand White female rabbits by injection of 100–200 μg of KLH–peptide conjugate. Polyclonal mouse anti-IRBP antibodies were prepared by nitrocellulose implant of immunoblotted protein (33). Antibody was affinity-purified (33).

cDNA Library Screening, Cloning, DNA Sequencing, and RNA. A random-primed cDNA library in bacteriophage λgt11 was made from *Drosophila* Kc cell poly(A)$^{+}$ mRNA as described (34). This library was screened with the IRBP2 anti-peptide antibodies and positives were rescreened with

polyclonal mouse anti-IRBP or rabbit anti-IRBP1 peptide antibodies (34). One insert was sequenced (K_c1.1) and then used in DNA hybridization screening of a *Drosophila* 4- to 8-hr embryonic plasmid cDNA library (35). One new cDNA (pNB409-1) was subjected to DNA sequence analysis. RNA blot hybridization was carried out by standard methods at 65°C with single-stranded antisense RNA probes (34).

***In Situ* Hybridization to Polytene Chromosomes.** Biotinylated DNA probes were prepared by nick-translation (34) and hybridized to polytene chromosomes as described (36).

UV Crosslinking and Immunoprecipitation. Antibodies were prepared against IRBP expressed in *Escherichia coli* as described (33). UV photochemical crosslinking was carried out as described (11) with ^{32}P-labeled, BrdUrd-substituted DNA probes prepared as described (11).

RESULTS

Isolation, Sequence Analysis, and Comparison of IRBP cDNA Clones. IRBP was identified as a DNA-binding polypeptide of 65–70 kDa and can be purified to near homogeneity by a combination of heparin-agarose and DNA affinity chromatography (11). Purified IRBP was subjected to gel electrophoresis, electroblotting, enzymatic digestion, HPLC separation of peptide fragments, and amino acid sequencing. Three peptide sequences were used to generate synthetic peptides and polyclonal antisera as well as electroblotted IRBP protein purified from K_c cells (11). These antibodies were affinity-purified and shown to specifically crossreact with IRBP in partially purified IRBP preparations on protein immunoblots (data not shown). The IRBP anti-peptide antibodies were then used to screen a random-primed *Drosophila* cDNA library made from K_c tissue culture cells. Positive clones were retested with IRBP1 anti-peptide and anti-IRBP antibodies. One of the largest inserts was subcloned into a plasmid vector and sequenced. This cDNA fragment was used as a probe to isolate full-length clones from a 4- to 8-hr embryonic cDNA plasmid library (35). Restriction endonuclease mapping as well as *in vitro* transcription and translation of isolated clones [using the SP6 promoter in the library plasmid vector pNB40 (35)] were used to identify full-length clones (data not shown) and one of these was sequenced in its entirety (Fig. 1). This open reading frame encodes a predicted protein of 631 amino acids and M_r 72,353. The original two peptide sequences as well as numerous additional peptides from subsequent peptide sequencing experiments were all identified in this sequence (Fig. 1).

Sequence data base searches revealed significant matches with the 70-kDa subunit of the human Ku autoimmune antigen (Fig. 2). The best region of homology is between amino acids 354 and 459 of the IRBP sequence and amino acids 342 and 443 of the human Ku p70 protein (37) (Fig. 2). However, by inclusion of gaps in the alignment it is clear that there are other regions of identity throughout the entirety of the two sequences (27% identity and 34% similarity with conservative amino acid substitutions) (Fig. 2). There is also significant identity with a recently isolated yeast Ku p70 homolog (Fig. 2; ref. 38). The human 70-kDa Ku protein has a leucine zipper motif located between amino acids 187 and 248 (37, 39) that may be involved in heterodimeric DNA binding complex formation with the p86 Ku subunit. Although this region is similar, it is not absolutely conserved in the *Drosophila* IRBP cDNA sequence (Figs. 1 and 2). Recently, we have found that IRBP is associated with another protein of 80 to 86 kDa. This protein may correspond to the larger Ku p86 subunit that was presumably fractionated away during the IRBP purification (11). It is possible that association of IRBP with a p86 subunit as a heterodimer might alter its DNA recognition properties. In addition, the purified IRBP appears to be proteolyzed, missing ≈6 kDa of terminal peptide sequence relative to the cloned IRBP cDNA (E.L.B. and D.C.R., unpublished observations). These findings may explain the instability of IRBP DNA binding activity in chromatography experiments and the clearly preferred binding of IRBP to the P element 31-bp inverted repeats located at internal sites in DNA binding experiments (11) (E.L.B. and D.C.R., unpublished observations). However, it should be noted that mammalian Ku antigen binds both terminal and internal DNA sites, but the Ku and IRBP DNA binding activities cannot currently be compared (13–18).

Expression of the IRBP Gene. The expression profile of the gene encoding IRBP was examined by RNA blot hybridization with poly(A)$^+$ mRNA from *Drosophila* K_c tissue culture cells and tissues from a variety of *Drosophila* developmental stages. Using antisense RNA derived from the IRBP cDNA as a probe, a single 2.4-kb RNA species was identified in all RNA samples (Fig. 3). This size of mRNA is consistent with the cDNA sequence analysis (Fig. 1). The 2.4-kb IRBP mRNA was expressed at a low level throughout development but was particularly abundant in ovaries, 0- to 2-h embryos, and adult females (Fig. 3). In addition, a low level of the 2.4-kb mRNA was observed in adult males (data not shown). This expression profile is consistent with the transcript being made at the highest levels in the female germ line during oogenesis.

Cytogenetic Mapping of the IRBP Gene. The use of polytene chromosome *in situ* hybridization in *Drosophila* allows a direct link to be made between a cloned DNA segment and the genetic map of the organism. Often this information is useful to aid in gene identification and physical mapping and to identify potential candidate genes corresponding to a particular cDNA. Using a biotin-labeled IRBP cDNA as a probe, a single hybridization signal was detected at cytological position 86E2-3 on the right arm of the third chromosome

```
HP70  MSGWESYYKT EGDEEAEEEQ EENLEASGDY KYSGRDSLIF LVDASKAMFE   50
DP70  MSTW...NP. ENDVDLLSGS EDEEDVSMKR DYHGREAILF VVDANLQ...   43
YP70  M......RPV TNAFGNSGEL NDQVDEKGYR KFDIHEGILF CIELSETMFK   44

HP70  SQSEDEL.TP FDMSIQCIQS VYISKIISSD RDLLAVVFYG TEKDKNSVN.   98
DP70  ...TAGV.ER LLEALNIIRT AFISGLLVND KDLIGLIFAN TKHSPPPLEA   89
YP70  ESSDLDVKSP LLEILESLDE LMSQLVITRP GTAIGCYFYY CNREDAKEGI   94

HP70  ..FKNIYV.. ........LQE LDNPGAKRIL ELDQFKGQQG QKRFQDMM..  135
DP70  SALDNIVMPD NCAVFLPLRQ LTKPIVEHYL EFMGGVETQF ADVY.GLA..  136
YP70  YEL....... .....FPLRD INATFMKKLN DLLEDLSSGR ISLYDYFMFQ  132

HP70  ..GHGSDYSL SEVLWVCANL FSD...VQFK MSHKRIMLFT NEDNPHGNDS  180
DP70  ..EPDGRGRF DLMIRLCIEM .LEK..CGKK LNNAKIAYVT DVREPHPSNS  181
YP70  QTGSEKQVRL SVLFTFMLDT FLEEIPGQKQ LSNKRLFLFT DIDKPQEAQD  182

HP70  AKASRARTKA G.DLRDTGIF LDLMHLKKPG GFDISLFYRD IISIA....E  225
DP70  NHFQAALQKA S.DLEGKEFE FHVIPM..VD DFDYEPFYKE FITLS....R  224
YP70  IDERARLRRL TIDLFDNKVN FATFFIGYAD KPFDNEFYSD ILQLGSHTNE  232

HP70  DEDLR.VHFE ESSKLED..L LRKVRAKET. RKRALSRLKL KLNKDIVISV  271
DP70  AIELDAFQVP DAQMLRE..I LSDRKLKQDF LRRCLGHFSF YLGPNLSMSV  272
YP70  NTGLDSEFDG PSTKPIDAKY IKSRILRKKE VKRIMFQCPL ILDEKTNFIV  282

HP70  GIYNLVQKAL KPPPIKLYRE TNEPVKTK.. .......... ....TRTFNT  305
DP70  QYYNYFQRRA YPRKVQILRR DNSVVRTKRV ITVQKQKDDG SQDIEHEYQI  322
YP70  GVKGY..... .....TMYTH EKAGVRYKLV YEHEDIRQEA Y..SKRKFLN  320

HP70  STGGLLLPSD TKRSQIYGSR QIILEKEETE ELKRF...DD PGLMLMGFKP  352
DP70  KVTGGWYTCN V......GER DLRISMDQLN RVRNL...HK PQMMLLGFKH  363
YP70  PITGEDVTGK TVKVYPYGDL DINLSDSQDQ IVMEAYTQKD AFLKIIGFRS  370

HP70  LV.LLKKHHY LRPSLFVYPE ESLVIGSSTL FSALLIKCLE KEVAALCRYT  401
DP70  RS.SLPEVSY IKPANFMYPD DQSIIGSKRL FRALWERCLV RDKIAICLFM  412
YP70  SSKSIHYFNN IDKSSFIVPD EAKYEGSIRT LASLLKILRK KDKIAILWGK  420

HP70  PRRNIPPYFV ALVPQEEELD DQ....KIQV TPPGFQLVFL PFADDKRKMP  447
DP70  CKRKSIPRYV ALVPVEAPDN GEDKNYRSLL CGDGFKIVYL PEAKHIRHLD  462
YP70  LKSNSHPSLY TLSPSSVKDY NE........ ...GFYLYRL PFLDEIRKFP  459

HP70  FTE....KIM ATPEQVGKMK AIVE...... KLRFTYRSDS FENPVLQQHF  487
DP70  LQDWNNTENT ADEQKVEFFQ KIIK...... KLRVDYQPNL INDPSLDALQ  506
YP70  .SLLSYDDGS EHKLDYDNMK KVTQSIMGYF NLRDGYNPSD FKNPLLQKHY  508

HP70  RNLEALALDL MEPEQAVDLT LPKVEAMNKR LGSLVDEF.. KELVYPPDYN  535
DP70  ANLLALSLDF STDTKGLD.N LLDTSQQDKR IEKLLPDY.. .E.MFAPEAE  551
YP70  KVLHDYLLQI ETTFDENETP NTKKDRMMRE DDSLRKLYYI RNKILESEKS  558

HP70  PEGKVTKRKH DNEGSGSKRP KVEYSE.... EELKTHISKG TLGKFTVPML  581
DP70  PPKKRAAKST TAGASGPKMA KIDDDQLKEF EFVKSLNKDE ALTSCTAAQL  601
YP70  EDPTIQRLNK YVKIWNMFYK KFNDDNISIK EEKKPFDKKP KFNI        602

HP70  K...EACRAY GLKSGLKKQE LLEALTKHFQ D                      609
DP70  HFILQHHFDV TMPKSSKKAK LVAKIEELHK                        631
```

FIG. 2. Amino acid sequence comparison of IRBP (DP70) and the human Ku p70 (HP70) (37) and the yeast (YP70) (38) proteins. Identical amino acids are indicated in white with black background, conservative amino acid changes (D/E, R/K, L/I/V, G/A, Q/N, S/T) are indicated in white with a stippled gray background, and nonidentical amino acids are indicated in black letters. Gaps are indicated by black dots and amino acid numbers are indicated on the right.

(3R) (Fig. 4). Examination of the known genes that lie in this interval (40) indicated that one gene, *mus309*, lies in the cytogenetic interval 86E2–87B5. *mus309* mutants are members of a class of mutants in *Drosophila* that are hypersensitive to mutagenic chemicals. Interestingly, the *mus309* mutants also exhibit sterility in females or both sexes depending on the mutant allele (29, 40). This observation is provocative because IRBP mRNA levels were found to be elevated in the female germ line (Fig. 3).

Immunoprecipitation of IRBP–DNA Adducts. IRBP was identified by UV-induced DNA–protein crosslinking and DNase I footprinting (11). In highly purified IRBP preparations, a single polypeptide of ≈65 kDa was crosslinked to the P element inverted repeat DNA (11). Therefore, we prepared antibody to recombinant IRBP protein expressed from the IRBP cDNA in *E. coli*. We used this affinity-purified antibody to immunoprecipitate the ^{32}P-labeled DNA–protein adduct from highly purified IRBP preparations. This antibody crossreacted with both crude and purified *Drosophila* IRBP on protein immunoblots (Fig. 5*A*). Indeed, this antibody was capable of retrieving the IRBP protein–DNA complex (Fig. 5*B*, lane 1), whereas with control antibody or no antibody, no immunoprecipitation was observed (Fig. 5*B*, lanes 2 and 3). Note that the protein–DNA adduct has slower electrophoretic mobility than the free protein due to the covalently attached ^{32}P-labeled DNA (11). Therefore, the cDNA we isolated corresponds to the 65-kDa IRBP polypeptide originally purified from *Drosophila* K_c cells.

The mammalian Ku p70 subunit is often associated with a second Ku p86 subunit as a heterodimer (13, 37). However, we purified IRBP as a single polypeptide from *Drosophila* K_c cells that is homologous to the mammalian Ku p70 (11). Previous studies have shown that mammalian p70 is capable of binding DNA in the absence of p86 (41), suggesting that there may be a free pool of the 70-kDa subunit in cells and that association with p86 might modulate DNA binding by p70. Expression studies and isolation of the *Drosophila* p86 gene are needed to clarify whether IRBP p70 alone or a p70-p86 heterodimer allows specific binding to P element DNA sequences as observed with endogenous purified IRBP protein (11). It is possible that an IRBP p70-p86 heterodimer might have altered DNA recognition properties.

DISCUSSION

Previous studies have shown that P element transposase binds to internal sites on P element DNA but does not recognize the terminal 31-bp inverted repeats (4). A non-P element-encoded protein was identified by DNase I protection analysis that recognized the outer half of the inverted repeats and was termed the IRBP (11). Here, we have isolated the gene encoding IRBP and determined its deduced protein sequence. Interestingly, *Drosophila* IRBP shares significant sequence homology to the 70-kDa subunit of the mammalian Ku autoimmune antigen. Chromosome *in situ* hybridization analysis mapped the IRBP gene to cytological position 86E2-3. This genomic region contains a gene called *mus309* that is a member of a class of genes that are hypersensitive to mutagenic chemicals (29, 30). Many of these genes have been shown to be involved in DNA repair.

Is IRBP Involved in P Element Transposition? Several observations suggest that IRBP may be involved in P element transposition. First, *in vitro* DNA binding studies demonstrated that IRBP interacts with the P element 31-bp inverted repeats adjacent to the transposition cleavage site. This is significant because the inverted repeats are absolutely required for P element transposition *in vivo* (3) and transposase does not interact with these sequences *in vitro* (4). Second,

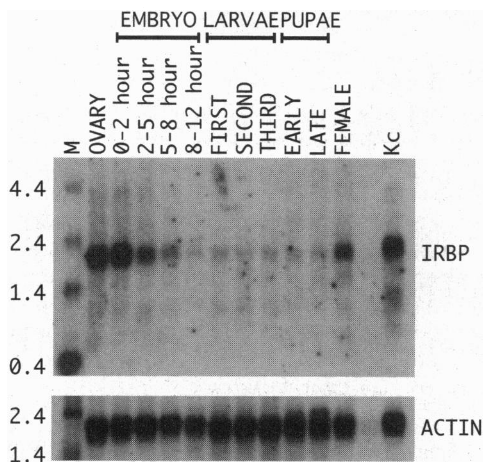

FIG. 3. (*Upper*) Developmental profile of *Drosophila* IRBP transcripts. RNA blot hybridization was carried out with poly(A)$^+$ RNA from the indicated developmental stage. (*Lower*) Same filter probed with a *Drosophila* 5C actin gene probe used as a control. Lane M, size markers (kb).

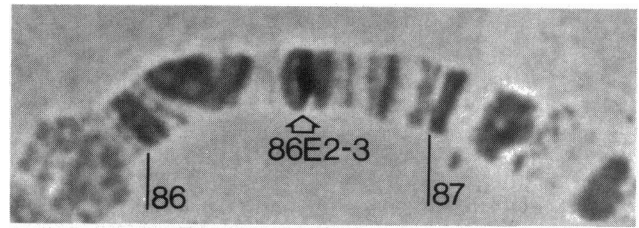

FIG. 4. Cytogenetic mapping of *Drosophila* IRBP gene. Chromosome *in situ* hybridization was carried out with the biotinylated IRBP cDNA and detected with avidin-alkaline phosphatase (36). The cytogenetic region of chromosome 3R 86-87 is shown and the signal at 86E2-3 is indicated by an open arrow.

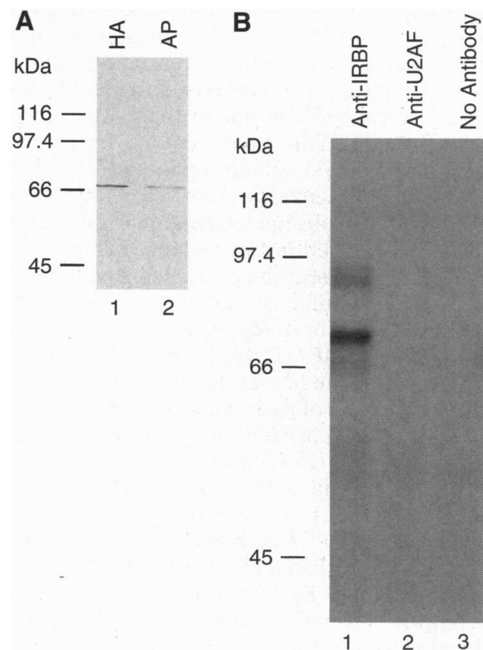

FIG. 5. Immunoprecipitation of IRBP–DNA complexes. (*A*) Protein immunoblot analysis of IRBP in crude heparin agarose (lane 1) and purified DNA affinity (lane 2) fractions (11) using affinity-purified antibodies to recombinant IRBP expressed in *E. coli*. (*B*) Antibodies raised to recombinant IRBP were used to immunoprecipitate radiolabeled IRBP–DNA complexes produced by UV photochemical crosslinking with ^{32}P-labeled BrdUrd-substituted DNA (11).

the IRBP protein sequence suggests that it is related to the human Ku autoimmune antigen. This heterodimeric DNA binding complex binds to both DNA termini and internal sites (14–16) and is known to be involved in DNA double-strand break repair and immunoglobulin VDJ recombination in mammalian cells (26–28). Third, the IRBP gene was cytogenetically mapped to interval 86E2-3 on the right arm of the third chromosome. This region contains a gene called *mus309* that may be involved in DNA repair and P element transposition.

The mutagen-sensitive (*mus*) mutants in *Drosophila* were isolated in the early 1970s (29, 30). One member of this class of mutants, *mus101*, has been demonstrated genetically to effect P element transposition (10). These experiments showed that there was failure to recover P element carrying chromosomes in a *mus101* mutant background after transposase-induced P element mobilization (10). This observation suggests that *mus101* is involved in repair of the double-strand break induced at the donor chromosomal site after P element mobilization (5, 6). It is likely that other *mus* genes will be involved in P element transposition and that *mus309* is a possible candidate.

The Mammalian Ku Antigen Is Involved in DNA Repair. Based on the biochemical observation that the purified mammalian Ku antigen recognizes free DNA termini (14–16), it was proposed that Ku might be involved in DNA repair or transposition (13). A number of studies have led to the purification of Ku as a DNA binding protein involved in transcription. However, these studies have not been definitive (17, 20, 21). More recently, the Ku antigen was purified as a subunit of the dsDNA PK (19, 20). These experiments showed that the 350-kDa kinase polypeptide required the heterodimeric Ku subunits for specific interaction with DNA and for catalytic activity (19, 20). The kinase is known to phosphorylate a variety of nuclear proteins, including simian virus 40 T antigen, p53, and other DNA binding proteins *in vitro* (21–23). However, the physiological significance of phosphorylation of these proteins is not known. Interestingly, the sequence specificity for phosphorylation by the dsDNA PK was investigated for p53 and was shown to be a serine or threonine residue always flanked by a glutamine residue (23). There are seven of these potential dipeptide phosphorylation sites in the first 140 amino acids of P element transposase. It is conceivable that free DNA ends might act as a signal to trigger phosphorylation of transposase or other target proteins during transposition or at other times in the cell cycle—for instance, after DNA replication. Interestingly, transcription of the genes encoding the mammalian p70 and p86 Ku subunits is induced when cells proliferate after G_0 cell cycle arrest (42).

It has recently been found that x-ray-sensitive mammalian cell lines are defective for both double-strand break repair and immunoglobulin VDJ recombination (24, 25). Interestingly, these mutant cell lines are defective in a DNA end binding activity, which was recently demonstrated to be antigenically related to the Ku antigen (26–28). More importantly, reversion of the x-ray sensitivity was accompanied by a restoration of the Ku-like DNA binding activity (27, 28). Thus, it appears likely that the Ku antigen is involved either directly or indirectly in double-strand break repair and VDJ recombination.

We gratefully acknowledge the contributions of Rhonda Doll and Amanda Hannaford to the early stages of this project involving protein purification and anti-peptide antibody generation. We also acknowledge T. Laverty, G. M. Rubin, and the *Drosophila* genome project for assistance with polytene chromosome *in situ* hybridization and localization. This work was funded by the National Institutes of Health (Grant R01-GM48862-03).

1. Berg, D. E. & Howe, M. M., eds. (1989) *Mobile DNA* (Am. Soc. Microbiol., Washington, DC).
2. Mizuuchi, K. (1992) *Annu. Rev. Biochem.* **61,** 1011–1051.
3. Engels, W. R. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC), pp. 437–484.
4. Kaufman, P. D., Doll, R. F. & Rio, D. C. (1989) *Cell* **59,** 359–371.
5. Kaufman, P. D. & Rio, D. C. (1992) *Cell* **69,** 27–69.
6. Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B. & Sved, J. (1990) *Cell* **62,** 515–525.
7. Bainton, R., Gamas, P. & Craig, N. L. (1991) *Cell* **65,** 805–816.
8. Benjamin, H. W. & Kleckner, N. (1989) *Cell* **59,** 373–383.
9. Haniford, D. B., Benjamin, H. W. & Kleckner, N. (1991) *Cell* **64,** 171–179.
10. Banga, S. S., Velazquez, A. & Boyd, J. B. (1991) *Mutat. Res.* **255,** 79–88.
11. Rio, D. C. & Rubin, G. M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8929–8933.
12. Francoeur, A.-M., Peebles, C. L., Gomper, P. T. & Tan, E. M. (1986) *J. Immunol.* **136,** 1648–1653.
13. Mimori, T., Hardin, J. A. & Steitz, J. A. (1986) *J. Biol. Chem.* **261,** 2274–2278.
14. Griffith, A. J., Blier, P. R., T., M. & Hardin, J. A. (1992) *J. Biol. Chem.* **267,** 331–338.
15. Paillard, S. & Strauss, F. (1991) *Nucleic Acids Res.* **19,** 5619–5624.
16. Zhang, W.-W. & Yaneva, M. (1992) *Biochem. Biophys. Res. Commun.* **186,** 574–579.
17. Knuth, M. W., Gunderson, S. I., Thompson, N. E., Strasheim, L. A. & Burgess, R. R. (1990) *J. Biol. Chem.* **265,** 17911–17920.
18. Stuiver, M. H., Coenjaerts, F. E. J. & van der Vliet, P. C. (1990) *J. Exp. Med.* **172,** 1049–1054.
19. Gottlieb, T. M. & Jackson, S. P. (1993) *Cell* **72,** 1–20.
20. Dvir, A., Peterson, S. R., Knuth, M. W., Lu, H. & Dynan, W. S. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 11920–11924.
21. Jackson, S. P., MacDonald, J. J., Lees-Miller, S. & Tjian, R. (1990) *Cell* **63,** 155–165.
22. Lees-Miller, S. P., Chen, Y.-R. & Anderson, C. W. (1990) *Mol. Cell. Biol.* **10,** 6472–6481.
23. Lees-Miller, S. P., Sakaguchi, K., Ullrich, S. J., Apella, E. & Anderson, C. W. (1992) *Mol. Cell. Biol.* **12,** 5041–5049.
24. Taccioli, G., Rathbun, E., Oltz, E., Stamato, T., Jeggo, P. & Alt, F. (1993) *Science* **260,** 207–210.
25. Pergola, F., Zdzienick, M. & Lieber, M. (1993) *Mol. Cell. Biol.* **13,** 3464–3471.
26. Rathmell, W. K. & Chu, G. (1994) *Mol. Cell. Biol.* **14,** 4741–4748.
27. Rathmell, W. & Chu, G. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 7623–7627.
28. Getts, R. C. & Stamato, T. D. (1994) *J. Biol. Chem.* **269,** 15981–15984.
29. Boyd, J. B., Golino, M. D., Shaw, K. E. S., Osgood, C. J. & Green, M. M. (1981) *Genetics* **97,** 607–623.
30. Boyd, J. B. & Setlow, R. B. (1976) *Genetics* **84,** 507–526.
31. Kennedy, T. E., Gawinowicz, M. A., Barzilai, A., Kandel, E. R. & Sweatt, J. D. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7008–7012.
32. Fernandez, J., DeMott, M., Atherton, D. & Mische, S. M. (1992) *Anal. Biochem.* **201,** 255–264.
33. Harlow, E. & Lane, D., eds. (1989) *Antibodies: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
34. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
35. Brown, N. H. & Kafatos, F. C. (1988) *J. Mol. Biol.* **203,** 425–437.
36. Ashburner, M. (1989) *Drosophila: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
37. Reeves, W. H. & Sthoeger, Z. M. (1989) *J. Biol. Chem.* **264,** 5047–5052.
38. Feldmann, H. & Winnacker, E. (1993) *J. Biol. Chem.* **268,** 12895–12900.
39. Landschultz, W. H., Johnson, P. F. & McKnight, S. L. (1988) *Science* **240,** 1759–1764.
40. Lindsley, D. L. & Zimm, G. G. (1992) *The Genome of Drosophila melanogaster* (Academic, San Diego).
41. Messier, H., Fuller, T., Mangal, S., Brickner, H., Igarashi, S., Gaikwad, J., Fotedar, R. & Foredar, A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 2685–2689.
42. Roberts, M., Han, Y., Fienberg, A., Hunihan, L. & Ruddle, F. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6354–6358.