3762–3769 *Nucleic Acids Research, 1999, Vol. 27, No. 18*

Drosophila and human RecQ5 exist in different isoforms generated by alternative splicing

Jeff J. Sekelsky[1,2,*], Michael H. Brodsky[3], Gerald M. Rubin[3] and R. Scott Hawley[1]

[1]Section of MCB, University of California, Davis, CA 95616, USA, [2]Department of Biology, University of North Carolina, Chapel Hill, NC 27599, USA and [3]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA

Received March 23, 1999; Revised and Accepted July 26, 1999

DDBJ/EMBL/GenBank accession nos AF134239, AF135183

ABSTRACT

Members of the RecQ helicase superfamily have been implicated in DNA repair, recombination and replication. Although the genome of the budding yeast *Saccharomyces cerevisiae* encodes only a single member of this family, there are at least five human RecQ-related genes: *RecQL*, *BLM*, *WRN*, *RecQ4* and *RecQ5*. Mutations in at least three of these are associated with diseases involving a predisposition to malignancies and a cellular phenotype that includes increased chromosome instability. Metazoan RecQ helicases are defined by a core region with characteristic helicase motifs and sequence similarity to *Escherichia coli* RecQ protein. This core region is typically flanked by extensive, highly charged regions, of largely unknown function. The recently reported human RecQ5, however, has only the core RecQ-homologous region. We describe here the identification of the *Drosophila RecQ5* gene. We recovered cDNAs corresponding to three alternative splice forms of the *RecQ5* transcript. Two of these generate nearly identical 54 kDa proteins that, like human RecQ5, consist of the helicase core only. The third splice variant encodes a 121 kDa isoform that, like other family members, has a C-terminal extension rich in charged residues. A combination of RACE and cDNA analysis of human *RECQ5* demonstrates extensive alternative splicing for this gene also, including some forms lacking helicase motifs and other conserved regions.

INTRODUCTION

Genome instability is a significant factor in the development of many cancers. Mutations in genes that function in DNA metabolism frequently result in genome instability and, in humans, cancer-associated syndromes (reviewed in 1). One such class of genes is that encoding proteins related to *Escherichia coli* RecQ helicase. Three of the five known human members of this class have been shown to be associated with hereditary diseases that include predisposition to cancers. Mutations in *BLM* result in Bloom syndrome (BS), mutations in *WRN* result in Werner syndrome (WS), and mutations in *RECQ4* result in Rothmund–Thomson syndrome (RTS) (2–4). Each of these syndromes is a rare, autosomal recessive disorder with unique clinical features. BS is characterized by dwarfism and immune deficiency (reviewed in 5); WS is characterized by the premature onset of a number of conditions associated with aging (reviewed in 6,7); and RTS is characterized by skin and skeleton abnormalities, but also includes features associated with premature aging (reviewed in 8). Each of these syndromes involves a greatly increased risk for many types of malignancy, and each exhibits a cellular phenotype that includes genome instability. In BS cells, genome instability is manifested as the presence of chromosome aberrations and increased levels of exchange between homologous chromosomes and between sister chromatids (9,10). Chromosome instability in WS cells is revealed as an increased rate of chromosome translocations and deletions (11,12). RTS patients have been observed to acquire somatic mosaicism for chromosome rearrangements and changes in chromosome number (13,14).

Insights into the cellular functions of RecQ helicases have come from studies in model organisms, especially microorganisms. The protein family is named for *E.coli* RecQ, a 3' to 5' DNA helicase that is required either to promote or to inhibit recombination, depending upon the pathway (15–17). A single RecQ helicase-encoding gene has been found in each of the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* (18–20). Mutations in *S.cerevisiae SGS1* result in hyper-recombination and defects in both mitotic and meiotic chromosome segregation (18,21). Sgs1p interacts with topoisomerases II and III in a two-hybrid assay, and *sgs1* mutations suppress the slow growth phenotype of *top3* mutations (22). These phenotypes suggest that Sgs1p is involved in both replication and recombination.

One issue that studies in microorganisms cannot address is the multiplication and divergence of the RecQ family in animals. In addition to the *BLM*, *WRN* and *RECQ4* loci described above, two other human genes encoding RecQ-related proteins, *RECQL* and *RECQ5*, have been reported (23–25). The genome of the nematode *Caenorhabditis elegans* encodes at least four RecQ-related proteins, and it is possible to define orthologous relationships between these proteins and human WRN, BLM,

*To whom correspondence should be addressed at: Department of Biology, CB #3280, Fordham Hall, University of North Carolina, Chapel Hill, NC 27599-3280, USA. Tel: +1 919 843 9400; Fax: +1 919 962 8472; Email: sekelsky@email.unc.edu

Table 1. Human *RECQ5* cDNAs

GenBank	I.M.A.G.E.	Library tissue source	Class[a]	Insert[b]
AF135183	–	brain	A1	1–1356; 1823–2178
AA155835	590051	umbilical vein endothelium	A1	985–1356; 1823–2178
AI686448	2267876	serous papillary carcinoma	A1 or B	<1928–2176
AI684041	2267435	serous papillary carcinoma	A2 or B	<1961–2397
AI537883	2189520	serous papillary carcinoma	A3	1–1356; 1823–2828
N53539	284217	multiple sclerosis lesions	A3 or B	2198–2828
AI218469	1845920	mixed fetal lung, testis, B-cell	C1	(500)–1113; +332 bp
AI363275	2016056	glioblastoma	C2	311–1113; +396 bp
AI310229	1914385	kidney clear cell tumor	F1	278–1034; +191 bp
AI671940	2316345	kidney	F2	<866–1225; +47 bp
AI567000	2191126	serous papillary carcinoma	F3	846–1356; +780 bp

[a]See text for definitions of transcript classes. Some cDNAs could not be assigned unambiguously to a single class due to insufficient length or sequence information.

[b]Coordinates are relative to GenBank accession number AF135183. Numbers preceded by + denote sequences (of the indicated length) not present in AF135183. Numbers in parentheses are approximations. Some cDNAs were unavailable for analysis; these are marked with < to indicate that the insert continues in the 5' direction. Coordinates for these cDNAs are from the sequence in the corresponding GenBank record.

RecQL and RecQ5, based on sequence similarities within the region conserved among family members. Thus, it appears that this gene family underwent substantial divergence early in metazoan evolution, with different family members presumably acquiring different functions (26). Genetic characterization of RecQ-related genes in model metazoans is likely to provide important new insights into these functions.

We report here the identification of the *Drosophila melanogaster RecQ5* gene. Characterization of cDNAs encoding *Drosophila* RECQ5 show that alternative splicing produces drastically different isoforms. We also found evidence that human *RECQ5* is alternatively spliced to produce different isoforms.

MATERIALS AND METHODS

GenBank accession numbers

AF134239, *Drosophila RecQ5* (genomic sequence); AF135183, human *RECQ5* (for transcripts A and B).

DNA clones and sequences

We screened 10^5 clones from a 0–4 h *Drosophila* embryo cDNA library (27) and recovered 25 *RecQ5* cDNAs. We initially characterized four of these further. Two contained the entire coding region and two were truncated at the 5' end. Comparison to genomic sequence showed that three of these had the larger (327 nt) intron removed, and therefore encode the 121 kDa isoform. One of the truncated cDNAs had the smaller intron (75 nt) removed, and therefore encodes the 54 kDa isoform. We analyzed the remaining cDNAs by PCR and found that 19 had the larger intron removed and two had the smaller intron removed. Two *RecQ5* EST sequences were subsequently deposited into GenBank (28). One of these, clone LD21474 (GenBank accession no. AA735537), had the larger intron removed. The other, clone GH01404 (GenBank accession no. AI062257), had no intron removed at the alternative splice site, although it carries a polyadenylation tract at the 3' end and other introns were removed.

The original human *RECQ5* cDNA in the EST database is I.M.A.G.E. Consortium CloneID 590051 (GenBank accession nos AA155835 and A155882) (29). Using this insert as a probe, we screened 5×10^5 plaques from each of two human brain cDNA libraries and obtained a single positive clone. The sequence of this cDNA and our RACE sequences has been deposited into GenBank under accession number AF135183.

During the course of this work, several I.M.A.G.E. Consortium cDNA clone EST sequences were deposited into GenBank. We obtained and characterized the cDNAs that were available (Table 1 and Fig. 3). CloneID 2191126 carries a 1292 bp insert. The library from which this clone is derived has inserts cloned directionally into the *Sal*I (5') and *Not*I (3') sites of pCMV-SPORT6. The *Sal*I end appears to be intact. However, 162 bp near this end matches the end of a 5' EST read from CloneID 380349, but in the opposite orientation. The last 512 bp of the insert match *RECQ5* sequence in a 3' to 5' direction. There is no poly(A) tract at this end, and there is a fragment of rearranged polylinker sequence between the *Not*I site and the insert. We conclude that this clone carries a cDNA fragment with its 5' end toward the *Not*I site and its 3' end (which does not include the terminus) at the *Sal*I end.

Sequence analysis

Sequence comparisons were performed with the Wisconsin Package, v.9.1 (Genetics Computing Group). Multiple alignments of the RecQ conserved amino acid sequences (corresponding to human RecQ5 residues 1 through 406) were generated with the PILEUP program. Protein sequence distances were determined with the DISTANCES program (Kimura scoring method), and

a phylogenetic tree was constructed with the GROWTREE program (neighbor-joining method). Other methods gave slightly different trees, especially with respect to the placement of human RecQ4, but all methods generated essentially similar orthologous relationships between the metazoan sequences analyzed.

Searches for nuclear localization signals were done on the PSORT server at http://psort.nibb.ac.jp

PCR and 3' RACE

Human *RECQ5* introns were confirmed by PCR from genomic DNA (Promega), used at a concentration of 1 ng/µl. Primers used were Q12 (GCGCTTTGGCTTCTCTGA) and Q205a (CAAACCCAAAGCCTTCTT), which amplify a 666 bp fragment, including a 472 bp intron at 113; Q974 (GGGTGTGAACGCCAAGGCTT) and Q1061a (CAGGGACCTTCTCCTCCCAT), which amplify a 626 bp fragment that includes a 538 bp intron at 1011; Q1090 (GGGAGTGGATAAAGCCAAT) and QC326 (GGGCACAGCACTAGGCAAT), which amplify a 90 bp fragment across the 5' boundary of the intron between motifs V and VI; Q1318 (GGCCTTTGATGCCCTGGTGACC) and Q2081a (CCTCGATCTACCATGAGCTT), which amplify a 765 bp fragment that spans the 467 nt intron position; Q1773 (GGGGAGTCATGTGCTTTGAA) and Q2111a (CTGAAAATCAGGAGACGGG), which amplify a 339 bp fragment; and Q2458 (GGCCAAGTGTTCCTGTTCAT) and Q3157a (GCCCCGCCTCATTAGTTA), which amplify a 700 bp fragment.

To determine whether there are alternative 3' ends on *RECQ5* transcripts, we used two gene-specific primers from the 1.7 kb cDNA: Q1173 (GGGCTGGCAGGATGGGAAGCCTT) and Q1318. Amplification was done with the Marathon kit (Clontech), on human placenta second-strand cDNA ligated to adapters, according to the instructions. We obtained bands of 2.7 and 1.7 kb with primer 1173, and 2.5 and 1.5 kb with primer 1318. Several smaller bands were also detected by probing a blot of the RACE products, but these were not characterized. Bands were gel-isolated and reamplified with the same or internal primers for sequencing.

Radiation hybrid mapping

Radiation hybrid mapping was performed using the Stanford G3 Radiation Hybrid Panel (20) obtained from Research Genetics, according to the instructions. Primers used were 1852 (CTGAGGGCTGCTTGGTGTAGTCAGGTT; 3' end of coding) and 2081a (above; 3' UTR). Raw scoring was submitted to the RH server at http://www-shgc.stanford.edu/RH/index.html . The result returned indicated linkage to AFMb054zf9 on chromosome 17, with a LOD score of 21.7, and a distance of 2.17 cR_{10000}.

Northern blot hybridization

A blot of human mRNA was obtained from OriGene Technologies (Rockville, MD). The 229 bp PCR amplimer from the radiation hybrid mapping was ^{32}P-labeled using the Ready-to-Go DNA labeling kit (Pharmacia), and the blot probed and washed according to the manufacturer's instructions. We also probed a similar blot (Clontech) containing mRNA from leukocytes, colon, small intestine, ovary, testis, prostate, thymus and spleen, and obtained similar results to those in Figure 4 (data not shown).

Immunolocalization of *Drosophila* RECQ5

The protein-coding region of a *RecQ5* cDNA encoding the 121 kDa isoform was fused in-frame downstream of a segment encoding the FLAG epitope DYKDDDDK. This was placed under the control of the *Ubiquitin* promotor through the initiating methionine codon. 5×10^6 Schneider S2 cells were transiently transfected with 5 µg DNA and 10 µl Superfect (Qiagen). Two days after transfection, these cells were stained with a 1:1000 dilution of anti-FLAG antibody M5 (Sigma) and 0.5 µg/ml DAPI.

RESULTS

RecQ5 is conserved between *Drosophila* and humans

We discovered a *D.melanogaster* gene encoding a novel member of the RecQ helicase superfamily while cloning an unrelated cell cycle checkpoint gene. Database searches using the deduced protein sequence revealed putative orthologues in *C.elegans* (E03A3.2) and humans. A sequence of the human ortholog was recently published with the name *RECQ5* (25), so we designate the *Drosophila* gene *RecQ5*. The predicted human RecQ5 and *Drosophila* RECQ5 proteins are 49% identical to one another in the RecQ-conserved region. Sequence comparisons between family members using this region place human RecQ5, *Drosophila* RECQ5 and *C.elegans* E03A3.2 onto a distinct branch (Fig. 1), suggesting that these genes share a conserved function in DNA metabolism.

We mapped the *Drosophila RecQ5* gene to interval 70E1-4 by *in situ* hybridization to polytene chromosomes. We determined the genomic position of human *RECQ5* by radiation hybrid mapping using the Stanford G3 panel (Materials and Methods). Our results place the gene in 17q23–25 near the marker AFMb054zf9.

Drosophila RECQ5 isoforms generated by alternative splicing

Previously described members of the eukaryotic RecQ helicase family range in size from 75 kDa for human RecQL to 162 kDa for human WRN. In most family members, the highly conserved helicase region is flanked by extensive sequences that show similarity to one another primarily in amino acid composition, being rich in charged residues. The reported human *RECQ5* sequence encodes a 46 kDa protein that consists of the core region conserved among RecQ family members. Analysis of *Drosophila RecQ5* cDNAs reveals predicted isoforms of 54 and 121 kDa (Fig. 2). The smaller isoforms are very similar in structure to human RecQ5, consisting entirely of the region encompassing the seven motifs common to known helicases (30), together with flanking sequences conserved in all RecQ family members. The larger isoform has in addition an extensive C-terminal region abundant in charged residues. Among the 588 residues in this region, 98 (16.7%) are acidic, and 126 (21.4%) are basic. The charged residues are primarily glutamic acid (10.9%) and lysine (14.1%). No sequence similarity is apparent between this region and other RecQ family members, other than in composition.

The different *Drosophila* RECQ5 isoforms are encoded by transcripts that differ by alternative splicing (Fig. 2). The region encoding the helicase motifs is contained entirely on the second exon, which can be joined to either of two alternative third exon start sites. Exon 3a, following removal of a 327 nt

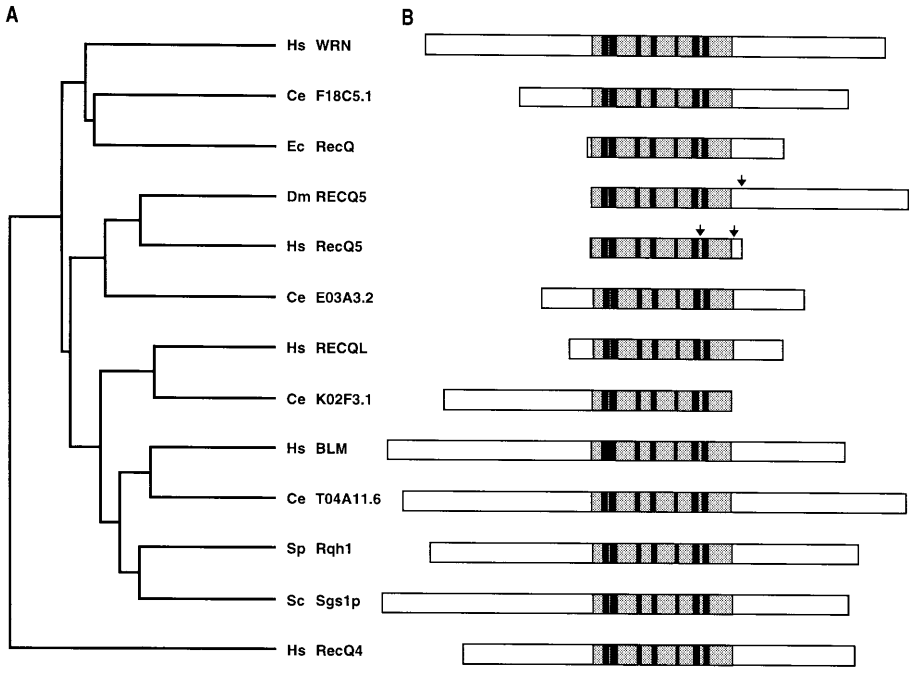

Figure 1. The eukaryotic RecQ helicase family. (**A**) Dendrogram depicting relationships between RecQ helicase family members from *E.coli* (Ec), *S.cerevisiae* (Sc), *S.pombe* (Sp), *C.elegans* (Ce) and humans (Hs), along with *Drosophila* RECQ5. The relationship of human RecQ4 to the other sequences is unclear (Materials and Methods), but the other metazoan proteins fall into four branches. Another *Drosophila* RecQ family member, DmBLM, the ortholog of human BLM, was recently reported (26), but is not indicated here. (**B**) Schematics of the corresponding protein structures. The shaded region represents the RecQ helicase-related conserved region that was used in sequence comparisons. Black boxes indicate the seven conserved helicase motifs. Arrows in *Drosophila* and human RecQ5 indicate termination sites of alternative isoforms (see Figs 2 and 3).

intron, maintains an open reading frame into subsequent exons, resulting in a 121 kDa isoform (RECQ5a). Exon 3b, following removal of a 75 nt intron, has an immediate stop codon at the 5' end, resulting in the truncated 54 kDa isoform (RECQ5b). We also found a cDNA in which no intron was removed at this position. Other introns had been removed and there was a poly-adenylation tract at the 3' end, indicating that this cDNA is derived from a processed transcript. This cDNA encodes a 54 kDa isoform (RECQ5c) almost identical to RECQ5b.

Human RECQ5 isoforms generated by alternative splicing

We wanted to know whether alternative splicing is a feature conserved in human *RECQ5*. Human *RECQ5* was originally defined by EST sequences from a single cDNA (Table 1, line 2). We obtained and sequenced this cDNA and found it to be incomplete at the 5' end. Using this clone as a probe, we screened human cDNA libraries, and recovered one *RECQ5* clone from a brain cDNA library. The 1710 bp sequence of this cDNA is identical to the sequence of the EST clone, except that it is longer at the 5' end and it contains a complete open reading frame. Both cDNAs have a polyadenylation tract 275 bp 3' to the termination codon. The corresponding mRNA encodes a 49 kDa polypeptide, slightly longer than the 46 kDa polypeptide predicted by the published *RECQ5* coding sequence (25). Comparison of the published nucleotide sequence to our sequences showed that they are identical except at the extreme 3' end of the open reading frame, at which point the published sequence has a stop codon not found in our sequences.

To investigate the differences between our cDNAs and the published sequence, we performed 3' RACE on human placenta mRNA, using two gene-specific primers to sequences near the end of the protein-coding region (Materials and Methods). With each primer we amplified a product of the size predicted by our cDNA sequences (data not shown). In addition, we obtained two larger products for each primer, which we refer to as RACE products 1 and 2.

Sequencing of these larger RACE products showed that each contains an insertion of 467 bp near the end of the coding region (Fig. 3). The insertion begins with a GT and ends with an AG, so we considered that it might correspond to an intron, and that RACE products 1 and 2 represent amplification of transcripts in which this intron had not been removed. We designed primer pairs that span either the 5' or 3' boundary of the insertion, and that span the entire insertion (Materials and Methods). When we used these primers in PCR reactions using human genomic DNA as a template, products of the predicted sizes were specifically amplified (data not shown), confirming that this insertion is indeed derived from an unspliced intron.

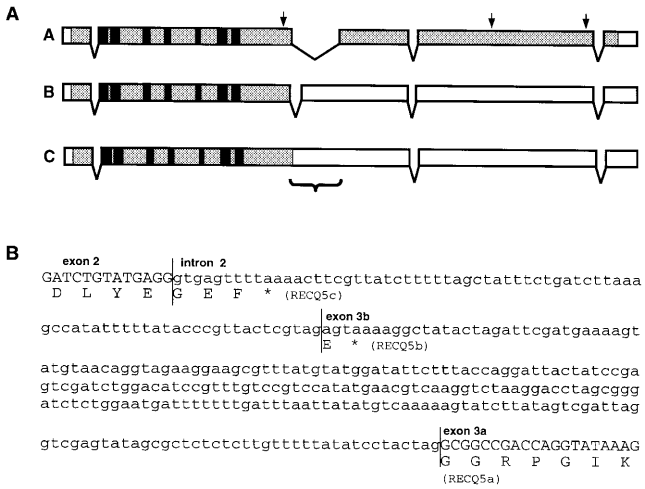

Figure 2. Alternative splicing at *Drosophila RecQ5*. (**A**) Schematic of three alternative *RecQ5* transcripts. Each box represents one exon. Protein-coding regions are shaded; black boxes indicate the seven helicase motifs. Transcript A encodes a 121 kDa isoform (RECQ5a); transcripts B and C encode nearly identical 54 kDa isoforms (RECQ5b and RECQ5c). Positions of potential nuclear localization signals (KRPKK, HRRKR and PYYKRKI) are indicated with arrows. (**B**) DNA sequence corresponding to the region of the alternative splice (indicated by a bracket in 3a), determined by sequencing cDNAs and genomic DNA. The sequence begins near the end of the common second exon. The sequence corresponding to the larger version of the intron is in lower case. Predicted amino acid sequences are shown, with asterisks indicating stops.

One possible explanation is that RACE products 1 and 2 were from unprocessed primary transcripts. Several observations suggest that this is not the case. First, the template in the RACE reactions was double-stranded cDNA made from polyadenylated mRNA, and both products carry a polyadenylation tract at their 3' ends. Second, the 3' end of product 1 is identical to that of two 3' EST sequences in the database (see below). Third, the 5' end of the unspliced intron has an in-frame stop codon, resulting in a sequence that is identical to the published *RECQ5* protein-coding sequence (25). We conclude that this 467 nt intron is removed from some transcripts but not from others. As is the case in *Drosophila*, failure to remove this intron results in a truncation of the polypeptide, in this case removing 25 residues to yield a 46 kDa isoform.

RACE products 1 and 2 also differ at their 3' ends from our cDNA. Both RACE products contain the entire 3' end of transcript A, followed by additional sequences. Product 1 has an additional 1118 bp, and product 2 has this same sequence followed by an additional 718 bp, ending in an Alu sequence. There is a putative polyadenylation signal near the 3' end of the Product 1 sequence, and database searches revealed two ESTs from cDNAs with identical 3' ends (Table 1 and Fig. 3), confirming that this is a bona fide polyadenylation site.

During the course of this work, additional human *RECQ5* EST sequences were deposited into GenBank (Table 1). Analysis of these cDNAs reveals additional polyadenylation sites, splicing patterns and predicted protein isoforms (Fig. 3). We have designated each transcript class according to predicted protein product (distinguished by different letters) and polyadenylation site (distinguished by different numbers). Transcripts in class A encode the 49 kDa isoform. There are at least three different polyadenylation sites for class A transcripts. Transcripts in class B encode the 46 kDa isoform, due to failure to remove the 467 bp intron. This class is presently represented only by our RACE products and the coding sequence reported by Kitao *et al.* (25).

A third class, designated C, is represented by I.M.A.G.E. Consortium CloneIDs 2016056 and 1845920. The inserts in these cDNAs match our *RECQ5* cDNA sequence from their 5'

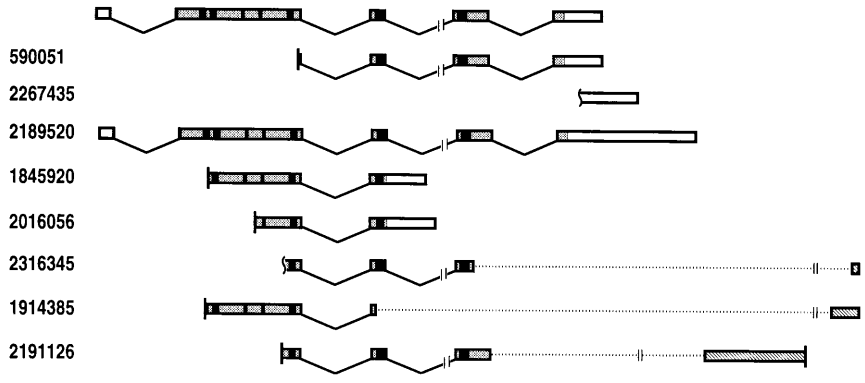

Figure 3. Transcripts and partial genomic structure of human *RECQ5*. (**A**) The structures of representative cDNAs carrying *RECQ5* sequences are shown. Protein-coding regions are shaded; black boxes indicate the seven helicase motifs. Known intron positions are indicated, but other intron positions were not determined. The third intron depicted is ~2.1 kb. The fourth intron depicted is the one not removed in class B transcripts. The uppermost schematic represents our 1710 bp cDNA, and others represent I.M.A.G.E. Consortium cDNAs (Table 1). A vertical line indicates that the cDNA was probably incomplete at that end. A sigmoid line indicates that the cDNA continues for an undetermined distance (these clones were not yet available for analysis). The lower three cDNAs appear to be fusions between *RECQ5* at and the 3' end of another gene (hatched boxes).

ends to bp 1113 (of our sequence). At this junction, novel sequence begins, extending 332 and 396 bp, respectively, to a polyadenylation tract. The novel sequence begins with GT, suggesting that these cDNAs might correspond to transcripts in which termination and polyadenylation occurred within an intron. PCR from human genomic DNA using primer pairs spanning the junction point yielded the fragment predicted from these two EST sequences, indicating that the sequences across the junction point are contiguous in the genome. Hence, there is an intron of ~2 kb between the regions encoding helicase motifs V and VI; transcripts in class C terminate within this intron. These transcripts encode a 36 kDa protein that lacks motif VI as well as additional sequences conserved among RecQ family members.

Transcript class D is represented by I.M.A.G.E. Consortium CloneIDs 1914385, 2316345 and 2191126 (Table 1 and Fig. 3). These transcripts appear to be hybrids between *RECQ5* and another gene. They each have *RECQ5* sequences at their 5' ends and another sequence at their 3' ends. CloneID 1914385 has *RECQ5* sequence through 1034 (of *RECQ5*), followed by 191 bp of sequence that is identical to the 3' ends of a group of 19 other EST sequences. These other EST sequence reads, which are up to 601 bp in length, are identical to one another along their entire lengths, and none of this sequence is from *RECQ5*, even in cases where 5' EST reads are available. Thus, these cDNAs are probably derived from an independent gene, which we will refer to as gene *D*.

CloneID 2316345 is similar to 1914385, except that the junction point is at 1229 in *RECQ5*, and there are only 51 bp of the 3' end of gene *D*. The insert in 2191126 also appears to be a fusion between *RECQ5* and gene *D*. This insert is in a rearranged orientation (Materials and Methods) and is lacking a polyadenylation tract at the 3' end, so it likely represents an internal fragment of a cDNA. The 3'-most sequences from this insert overlap the 5' end of CloneID 380349, whose 3' end identifies it as a cDNA from gene *D*. Thus, 2191126 appears to identify a fusion point more 5' in gene *D*.

RECQ5 is expressed in many tissues

To determine the relative abundances and tissue distribution of the different *RECQ5* transcripts, we probed blots of human mRNA. We detected a 1.7 kilonucleotide (knt) transcript at a low level in all tissues examined (Fig. 4). If our RACE products correspond to transcripts identical at the 5' end to the 1710 bp cDNA we sequenced, as is suggested by the published *RECQ5* sequence, these transcripts would be 2828 and 3646 nt (excluding polyadenylation tracts). However, we were unable to detect these larger transcripts using either probes within the region in common to all transcripts or probes specific to the larger transcripts. This was true even for placenta mRNA, which was the source of our RACE products. In contrast, Kitao *et al.* (25) detected transcript sizes of 3.6 and 3.8 knt for *RECQ5*, also present in all tissues examined. However, these authors did not report any signals corresponding to smaller transcripts.

Drosophila RECQ5 localizes to the nucleus

Saccharomyces cerevisiae Sgs1p and human WRN are both found to be concentrated in the nucleolus (31,32). To determine whether *Drosophila* RECQ5 shows a similar pattern of subnuclear localization, we transiently expressed the 121 kDa isoform, tagged with an N-terminal FLAG epitope, in *Drosophila*

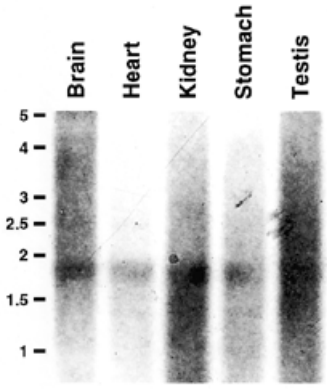

Figure 4. Expression of human *RECQ5*. A blot containing human mRNA (OriGene Technologies) was probed for *RECQ5* using sequences near the 3' end of transcript A. The probe used should detect transcripts A and B, but not C and D. RNA size markers are indicated to the left. The sources of mRNA are indicated above each lane.

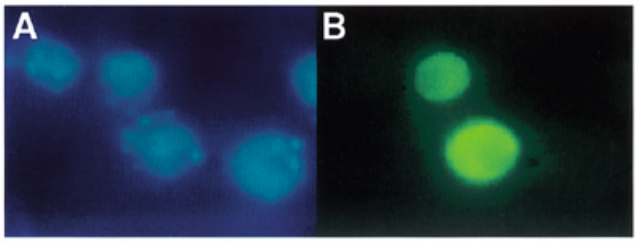

Figure 5. Immunolocalization of *Drosophila* RECQ5 to the nucleus. (**A**) DAPI staining of a field of cells transfected with the *Ubiq::FLAG-RecQ5* construct. Four nuclei can be seen in their entirety. (**B**) The same field stained with anti-FLAG antibody. Two cells can be seen to express *FLAG-RecQ5*. In both cells the protein is strongly localized to the nucleus.

Schneider2 cells. When cells were stained with anti-FLAG antibody to detect FLAG-RECQ5, we detected localization to the nucleus, but no sub-nuclear localization was apparent (Fig. 5).

DISCUSSION

We have described a new *D.melanogaster* RecQ helicase family member that shares a high degree of amino acid identity with human RecQ5. A unique feature of the *Drosophila* gene is the existence of isoforms that differ by the presence or absence of a highly charged C-terminus. Most other eukaryotic family members have regions of a similar size on one or both sides of the helicase region, but the functions of these regions are largely unknown. Perhaps the best understood case is WRN, which has a recognizable exonuclease motif at its N-terminus (33). This region has been demonstrated to possess a 3' to 5' exonuclease activity *in vitro*, and this activity is separable from the helicase activity (34). The region N-terminal to the helicase region of Sgs1p has been shown to interact with the C-terminal

portion of topoisomerase II in a yeast two-hybrid assay (18). It is possible that the highly charged, C-terminal region specific to the 121 kDa isoform of *Drosophila* RECQ5 interacts with other proteins, and that the difference between the isoforms is a way of regulating these interactions. The existence of different *Drosophila* RecQ5 isoforms provides a unique opportunity to study the *in vivo* functions of these sequences outside the core RecQ-related region. Such dramatically different isoforms have not been reported for any other family members. It is interesting to note that most family members have a dibasic (RR) sequence at the C-terminus of the RecQ-conserved region, suggesting the possibility of proteolytic processing at this location.

Alternative splicing also generates different isoforms of human RecQ5. In one case, the isoform produced depends on whether or not an intron is removed, as was observed for the *Drosophila* gene. The size difference between the 49 and 46 kDa RecQ5 isoforms, however, is much less than the difference between the *Drosophila* isoforms. These two human RecQ5 isoforms carry all of the sequences conserved among RecQ family members, and therefore both may have helicase activity. Nonetheless, it is possible that these different isoforms possess different biochemical activities or different capacities to interact with other cellular proteins that may regulate activity.

Another possibility is that the different C-termini regulate protein stability or sub-cellular localization. The gene encoding murine RecQL is alternatively spliced in the testis to remove a short region believed to contain a nuclear localization signal (35). The only recognizable nuclear localization signal in human RecQ5 is the sequence PERRVRS, which begins at residue 12. All predicted isoforms of *Drosophila* RECQ5 have at least one potential nuclear localization sequence (Fig. 2A, arrows). Consistent with this observation is our finding that at least the larger isoform localizes to the nucleus when expressed in a *Drosophila* cell line. It seems unlikely that the function of *Drosophila* RECQ5 alternative splicing is to regulate nuclear localization, but it is possible that this is the case for human RecQ5.

A third human RecQ5 isoform is produced by transcripts in class C, in which polyadenylation occurs 5' to the sequences encoding the last helicase motif. This 36 kDa protein lacks both helicase motif VI and additional sequences conserved among RecQ helicases, and likely has no catalytic function. Transcripts in this class may be important for regulating *RECQ5* expression.

Class D transcripts are somewhat perplexing. They appear to be fusions between *RECQ5* and another gene. Such fusions could have occurred in the DNA (e.g. by chromosome rearrangement), the RNA (e.g. by aberrant splicing), or the cDNA (e.g. by cloning artifacts). The three fusion cDNAs came from different libraries, and they have different fusion points in both genes, so it seems unlikely that these represent cloning artifacts. It is possible that these transcripts result from aberrant splicing. The fusion point in 2191126 is at a *RECQ5* exon boundary. However, the fusion point of 1914385 is not at an exon boundary. Either chromosome rearrangements or aberrant splicing could therefore account for these transcripts. Conceptual translation of available sequences from this second gene does not reveal any similarities to other proteins in the databases.

We used radiation hybrid mapping to map human *RECQ5* to the tip of 17q, in agreement with the *in situ* hybridization mapping of Kitao *et al.* Loss of heterozygosity in this region has been associated with ovarian cancer and familial breast carcinomas (36–40). Given the phenotypes associated with mutations in other RecQ helicase family members, human *RECQ5* may be a good candidate for a tumor suppressor gene.

We mapped *Drosophila RecQ5* to polytene region 70E1–4. We were unable to identify probable candidates for mutations in *RecQ5* among known mutations in the region. Efforts are now underway to generate mutations in this gene. Preliminary results suggest that *Drosophila RecQ5* is not an essential gene. Analysis of these mutations will provide valuable insights into the function of this gene in *Drosophila*, and perhaps other organisms, and into the functional significance of the different isoforms.

ACKNOWLEDGEMENTS

We thank Ken Burtis and Sarah Wayson for helpful suggestions, Elaine Kwan for assistance with cell culture experiments, and Todd Laverty for the *in situ* hybridization to polytene chromosomes. J.J.S. was supported in part by a fellowship from the Cancer Research Fund of the Damon Runyon, Walter Winchell Foundation, DRG 1355.

REFERENCES

1. Friedberg,E.C., Walker,G.C. and Siede,W. (1995) *DNA Repair and Mutagenesis.* American Society for Microbiology, Washington, DC.
2. Ellis,N.A., Groden,J., Ye,T.-Y., Straughen,J., Lennon,D.J., Ciocci,S., Proytcheva,M. and German,J. (1995) *Cell*, **83**, 655–666.
3. Yu,C.-E., Oshima,J., Fu,Y.-H., Wijsman,E.M., Hisama,F., Alisch,R., Metthews,S., Nakura,J., Miki,T., Ouais,S., Martin,G.M., Mulligan,J. and Schellenberg,G.D. (1996) *Science*, **272**, 258–262.
4. Kitao,S., Shimamoto,A., Goto,M., Miller,R.W., Smithson,W.A., Lindor,N.M. and Furuichi,Y. (1999) *Nature Genet.*, **22**, 82–84.
5. German,J. (1993) *Medicine*, **72**, 393–406.
6. Goto,M., Horiuchi,Y., Tanimoto,K., Ishii,T. and Nakashima,H. (1978) *J. Am. Geriatr. Soc.*, **26**, 341–347.
7. Goto,M., Tanimoto,K., Horiuchi,Y. and Sasazuki,T. (1981) *Clin. Genet.*, **19**, 8–15.
8. Vennos,E.M., Collins,M. and James,W.D. (1992) *J. Am. Acad. Dermatol.*, **27**, 750–762.
9. Chaganti,R.S., Schonberg,S. and German,J. (1974) *Proc. Natl Acad. Sci. USA*, **71**, 4508–4512.
10. German,J., Schonberg,S., Louie,E. and Chaganti,R.S. (1977) *Am. J. Hum. Genet.*, **29**, 248–255.
11. Schonberg,S., Niermeijer,M.F., Bootsma,D., Henderson,E. and German,J. (1984) *Am. J. Hum. Genet.*, **36**, 387–397.
12. Scappaticci,S., Forabosco,A., Borroni,G., Orecchia,G. and Fraccaro,M. (1990) *Ann. Genet.*, **33**, 5–8.
13. Lindor,N.M., Devries,E.M., Michels,V.V., Schad,C.R., Jalal,S.M., Donovan,K.M., Smithson,W.A., Kvols,L.K., Thibodeau,S.N. and Dewald,G.W. (1996) *Clin. Genet.*, **49**, 124–129.
14. Miozzo,M., Castorina,P., Riva,P., Dalpra,L., Fuhrman Conti,A.M., Volpi,L., Hoe,T.S., Khoo,A., Wiegant,J., Rosenberg,C. and Larizza,L. (1998) *Int. J. Cancer*, **77**, 504–510.
15. Hanada,K., Ukita,T., Kohno,Y., Saito,K., Kato,J.-I. and Ikeda,H. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 3860–3865.
16. Harmon,F.G. and Kowalczykowski,S.C. (1998) *Genes Dev.*, **12**, 1134–1144.
17. Lovett,S.T. and Sutera,V.A.J. (1995) *Genetics*, **140**, 27–45.
18. Watt,P.M., Louis,E.J., Borts,R.H. and Hickson,I.D. (1995) *Cell*, **81**, 253–260.
19. Murray,J.M., Lindsay,H.D., Munday,C.A. and Carr,A.M. (1997) *Mol. Cell. Biol.*, **17**, 6868–6875.
20. Stewart,E., Chapman,C.R., Al-Khodairy,F., Carr,A.M. and Enoch,T. (1997) *EMBO J.*, **16**, 2682–2692.

21. Watt,P.M., Hickson,I.D., Borts,R.H. and Louis,E.J. (1996) *Genetics*, **144**, 935–945.
22. Gangloff,S., McDonald,J.P., Bendixen,C., Arthur,L. and Rothstein,R. (1994) *Mol. Cell. Biol.*, **14**, 8391–8398.
23. Seki,M., Miyazawa,H., Tada,S., Yanagisawa,J., Yamaoka,T., Hoshino,S.-i., Ozawa,K., Eki,T., Nogami,M., Okumura,K., Taguchi,H., Hanaoka,F. and Enomoto,T. (1994) *Nucleic Acids Res.*, **22**, 4566–4573.
24. Puranam,K.L. and Blackshear,P.J. (1994) *J. Biol. Chem.*, **269**, 29838–29845.
25. Kitao,S., Ohsugi,I., Ichikawa,K., Goto,M., Furuichi,Y. and Shimamoto,A. (1998) *Genomics*, **54**, 443–452.
26. Kusano,K., Berres,M.E. and Engels,W.R. (1999) *Genetics*, **151**, 1027–1039.
27. Brown,N.H. and Kafatos,F.C. (1988) *J. Mol. Biol.*, **203**, 425–437.
28. Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein,P., Lewis,S. and Rubin,G.M. (1997) The BDGP/HHMI *Drosophila* EST Project.
29. Lennon,G.G., Auffray,C., Polymeropoulos,M. and Soares,M.B. (1996) *Genomics*, **33**, 151–152.
30. Gorbalenya,A.E., Koonin,E.V., Donchenko,A.P. and Blinov,V.M. (1989) *Nucleic Acids Res.*, **17**, 4713–4730.
31. Sinclair,D.A., Mills,K. and Guarente,L. (1997) *Science*, **277**, 1313–1316.
32. Marciniak,R.A., Lombard,D.B., Johnson,F.B. and Guarente,L. (1998) *Proc. Natl Acad. Sci. USA*, **95**, 6887–6892.
33. Mushegian,A.R., Bassett,D.E., Boguski,M.S., Bork,P. and Koonin,E.V. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 5831–5836.
34. Huang,S., Li,B., Gray,M.D., Oshima,J., Mian,I.S. and Campisi,J. (1998) *Nature Genet.*, **20**, 114–116.
35. Wang,W.S., Seki,M., Yamaoka,T., Seki,T., Tada,S., Katada,T., Fujimoto,H. and Enomoto,T. (1998) *Biochim. Biophys. Acta*, **1443**, 198–202.
36. Eccles,D.M., Russell,S.E.H., Haites,N.E., Atkinson,R., Bell,D.W., Gruber,L., Hickey,I., Kelly,K., Kitchener,H. and Leonard,R. (1992) *Oncogene*, **7**, 2069–2072.
37. Lindblom,A., Skoog,L., Andersen,T.I., Rotstein,S., Nordenskjold,M. and Larsson,C. (1993) *Hum. Genet.*, **91**, 6–12.
38. Cornelis,R.S., Devilee,P., van Vliet,M., Kuipers-Deijkshoorn,N., Kersenmaeker,A., Bardoel,A., Meera Khan,P. and Cornelisse,C.J. (1993) *Oncogene*, **8**, 781–785.
39. Cropp,C.S., Champeme,M.-H., Lidereau,R. and Callahan,R. (1993) *Cancer Res.*, **53**, 5617–5619.
40. Plummer,S.J., Paris,M.J., Myles,J., Tubbs,R., Crowe,J. and Casey,G. (1997) *Cancer Genet. Cytogenet.*, **20**, 354–362.