2120–2126 Nucleic Acids Research, 1995, Vol. 23, No. 12

Human helicase gene *SKI2W* in the HLA class III region exhibits striking structural similarities to the yeast antiviral gene *SKI2* and to the human gene *KIAA0052*: emergence of a new gene family

Andrew W. Dangel[1], Liming Shen[1,2,3], Anna R. Mendoza[1], Lai-chu Wu[2,4] and C. Yung Yu[1,2,3,5,*]

[1]Wexner Institute for Pediatric Research, Children's Hospital Research Foundation, 700 Children's Drive, Columbus, OH 43205, USA, [2]The Ohio State Biochemistry Program, [3]Department of Pediatrics, [4]Department of Medical Biochemistry and [5]Department of Medical Microbiology and Immunology, The Ohio State University, Columbus, OH 43210, USA

Received March 29, 1995; Revised and Accepted May 11, 1995

EMBL accession no. Z48796

ABSTRACT

Helicases are essential enzymes for life because DNA replication, DNA repair, recombination, transcription, RNA splicing and translation all involve more than one helicase to unwind DNA or RNA. We have discovered, cloned and partially characterized a novel human helicase gene, *SKI2W*. The human *SKI2W* is located between the *RD* and *RP1* genes in the class III region of the major histocompatibility complex (MHC) on chromosome 6, a genomic region associated with many malignant, genetic and autoimmune diseases. Derived amino acid sequence of human *SKI2W* showed an open reading frame for 1246 residues. It contains consensus sequences for structural motifs of an RNA helicase with a DEVH box. It has a leucine zipper motif that may be important for protein dimerization, and an RGD motif close to the N-terminus that might serve as a ligand for integrin or cell adhesion molecules. SKI2W shares a striking and extensive similarity to the yeast Ski2p that is involved in the inhibition of translation of poly(A) negative [poly(A)$^-$] RNA, and plays an important role in antiviral activities. Human SKI2W fusion protein expressed in insect cells using a baculovirus vector has ATPase activity. The human SKI2W protein and the yeast Ski2p share extensive sequence similarities to another putative human protein KIAA0052, suggesting the presence of a new gene family that may be involved in translational regulation of cellular and viral RNA.

INTRODUCTION

The major histocompatibility complex (MHC) is an important genomic region involved in the recognition of self and non-self. In humans it is also known as the HLA and is located on the short arm of chromosome 6. It contains three classes of genes and spans ~3500 kb in size (1). The structure, function and location of some HLA genes appear to be correlated. The class I and class II genes are likely formed by gene duplications and their protein products are involved in antigen presentation processes of the cell-mediated immune response. The class III region is located between the class I and class II genes. The genes for the complement components forming the C3 convertases, C2, factor B, C4A and C4B are present in tandem in the class III region. More than 40 diseases have been associated with the HLA. Population genetic studies in humans have inferred the presence of susceptibility genes in the HLA for nasopharyngeal carcinoma (2,3) and Hodgkin's lymphoma (4–6), IgA deficiency and common variable immunodeficiency (7). It has been shown that some congenic mouse strains are more vulnerable to chemical carcinogenesis or to viral infections and the susceptibility genes were mapped to the MHC (8–10). Breeding experiments in rats have demonstrated the presence of a growth and reproduction complex (*grc*) in (or linked to) the MHC (11). A cancer gene *rcc* has been mapped in the *grc* because *grc*$^-$ rats are highly susceptible to chemically induced carcinogenesis. These observations motivate our search for cancer and disease susceptibility genes in the MHC.

We have cloned and characterized a novel gene *RP1* located 611 bp upstream of C4. *RP1* encodes a 55 kDa protein that contains a bipartite nuclear localization signal and also leucine-rich sequences (12). It shares limited sequence similarity with the yeast DNA repair protein RAD7 (13). Located ~12 kb upstream of the *RP1* gene is the *RD* gene that contains 24 copies of repeating Arg–Asp or Arg–Glu amino acids. Although the function of RD is unknown, the presence of RNA binding motifs suggests that it may be involved in nucleic acid metabolism (14–16). The *RP1* and *RD* genes are separated by an intergenic distance of 11 kb. Based on the fact that the MHC class III region

* To whom correspondence should be addressed at: Wexner Institute for Pediatric Research, 700 Children's Drive, Columbus, OH 43205-2696, USA

is tightly packed with genes and some neighboring genes have intergenic regions <1 kb in size (12,17), we investigated the possibility for the presence of another gene at this 11 kb region. A genomic DNA probe derived from an 1.3 kb *Bam*HI fragment located upstream of the *RP1* gene was employed for Northern blot analysis. Transcripts ~3.9 kb in size were detected from multiple RNA samples isolated from different tissues. Thus, cDNA clones for this new gene, *W*, were isolated and sequenced. Derived amino acid sequence for *W* reveals seven structural motifs of helicases (18) and extensive sequence similarities to an undefined human protein KIAA0052 (19), and to the yeast protein Ski2p that plays an important role in the inhibition of translation of poly(A)$^-$ and/or uncapped cellular RNA and viral RNA (20,21). Thus, *Gene W* is also named *SKI2W*. Glutathione S-transferase (GST) fusion protein for human SKI2W were produced in insect cells. Purified GST–SKI2W fusion proteins were used to demonstrate the presence of ATPase activity that was enhanced by the addition of small nucleic acid molecules.

MATERIALS AND METHODS

Cell lines, isolation of RNA and Northern blot analysis

The cell lines used for extraction of total RNA were previously described (12,22). They included HT29 (human colon carcinoma), IMR32 (human neuroblastoma), BJAB (human B lymphoma), Raji (human B lymphoma), CNE (human nasopharyngeal carcinoma) and NPC-LC (cotton-top tamarin, EBV transformed B lymphoblast). Total RNA was isolated by guanidinium isothiocyanate lysis method following standard protocols (23). Northern blot analysis was performed using 50 μg of total RNA resolved by electrophoresis with a 0.8% formaldehyde–agarose gel (23).

Isolation of cDNA clones and DNA probes

cDNA clones were isolated from λgt10 cDNA libraries from a monocytic cell line U937 (Clontech, Palo Alto, CA), a T lymphocytic cell line RPMI 8402 (a gift of Dr Terry Rabbitts, MRC Laboratory of Molecular Biology, Cambridge, UK) and a B lymphocytic cell line DAUDI (Clontech, Palo Alto, CA). DNA probes used for library screenings included a 1.3 kb *Bam*HI–*Sal*I genomic DNA probe from cos 3A3 which is located 2 kb upstream of the *RP1* gene (12), a 1.6 kb *Gene W* 3′ cDNA probe obtained from a clone isolated from the U937 cDNA library, a 1.4 kb *Bam*HI–*Pst*I fragment and a 0.8 kb *Bam*HI–*Pst*I fragment from cosmids containing human *Gene W* genomic DNA (L. -M. Shen, A. R. Mendoza, A. W. Dangel and C. Y. Yu, manuscript in preparation).

DNA sequencing and sequence analysis

DNA sequences were determined by Sanger's dideoxy method using Sequenase (US Biochemicals, Cleveland, OH) and [^{35}S]dATP after the manufacturer's protocol. DNA templates were generated by: (i) *Bam*HI, *Bgl*II, *Pst*I, *Sma*I and *Sst*I restriction fragments from W1.6, W2.2, W0.8 and W1.9 cDNA clones isolated from U937, RPMI 8402 and DAUDI cDNA libraries, respectively; (ii) exonuclease III treatments of W2.2 DNA and W1.9 DNA using the 'Erase-a-Base' kit (Promega, Madison, WI), after appropriate linearization of plasmids by *Kpn*I and *Xho*I digests to generate a 3′ overhanging end and a 5′ overhanging end. Gel readings were compiled by PC/GENE software (Intelligenetics, Mountain View, CA). Gaps in sequence melds were filled by synthesizing new oligonucleotide primers for sequencing. A motif search was carried out on PROSITE (24) using the PC/GENE software.

Comparisons of the derived amino acid sequence with GenBank, PIR and SWISSPROT databases were achieved by GCG program FASTA (25,26) through the Pittsburgh Supercomputing Center. The email implementation of the Blast algorithm was employed to scan a non-redundant database built from SWISSPROT, PIR and GenèPept (27). Alignments of protein sequences were performed by GCG programs BESTFIT, PILEUP and PRETTY (26).

Determination of the 5′ sequence of *SKI2W* cDNA

The 5′ RACE (rapid amplification of cDNA ends) method was employed to clone the 5′ end sequence of *Gene W* cDNA, using a kit provided by Clontech (Palo Alto, CA). Briefly, 2 μg human poly(A)$^+$ RNA (from CNE cell line) was reverse transcribed using a *Gene W* antisense primer WR100 (5′-AGG GAT CCG TCT TCC TCT GCC AT-3′) located 100 bp downstream of the 5′ end of the compiled 3.6 kb *Gene W* cDNA sequence. The RACE products were used as templates for a nested PCR using an antisense *Gene W* primer WR50 (5′-AGC CAG GCT GGG GAT GAC A-3′) and an anchor primer provided by the RACE kit. The PCR products were cloned directly into the pCR II vector (Invitrogen, San Diego, CA) and used for sequencing.

Production of GST–SKI2W fusion protein in insect cells

A nearly full length 3.6 kb *SKI2W* cDNA (W3.6K) was constructed by joining a 1.5 kb *Sac*I fragment to a 1.2 kb *Sac*I–*Bam*HI fragment and a 0.9 kb *Bam*HI–*Eco*RI fragment derived from three overlapping *SKI2W* cDNA clones, W1.9, W2.2 and W1.6, respectively, in correct order and reading frame. This W3.6K cDNA encodes 1197 amino acids of the SKI2W protein without the first 49 residues at the N-terminal region. The 3.6 kb *SKI2W* cDNA fragment was released by an *Eco*RI digestion and ligated into baculovirus transfer vector pAcG2T (Pharmingen, San Diego, CA). The recombinant SKI2W product was expressed as a fusion protein with the GST protein under the control of the strong AcNPV polyhedrin promoter in the transfer vector, after the manufacturer's protocol (28). Mutagenized baculovirus DNA containing a lethal deletion was co-transfected with the pAcG2T/W3.6 construct into the *Sf*9 insect cells. Homologous recombinations between the mutagenized virus and the fusion protein constructs reconstituted viable viruses and produced a 160 kDa GST–SKI2W fusion protein in the *Sf*9 cells. Four days after co-transfection, the culture media were collected and the recombinant viruses amplified by infection of *Sf*9 cells. A plaque-purification procedure was performed. A SDS–PAGE analysis was performed on the infected *Sf*9 cell lysates to identify high expressors of SKI2W fusion proteins.

Purification of GST–SKI2W fusion protein and assay of ATPase activity

The GST–SKI2W fusion protein was purified from *Sf*9 cell lysates by sequential precipitation, centrifugation and solubilization of the inclusion bodies and insoluble pellets with 1, 2 and 4 M urea supplemented with 0.5–2% of the detergent NP-40 and

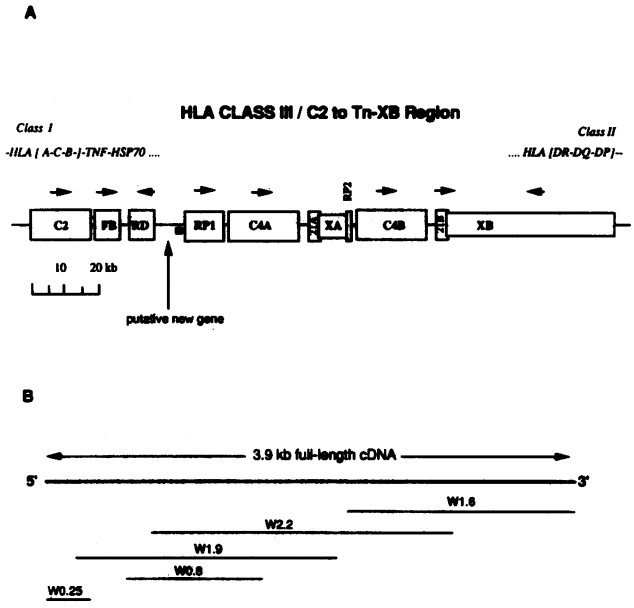

Figure 1. (A) Gene organization of the HLA class III region from complement C2 to tenascin Tn-XB (1,12,30,56). Horizontal arrows indicate gene orientation and the filled box indicates the location of a 1.3 kb *Bam*HI–*Sal*I genomic fragment used for Northern blot analysis. (B) A diagram indicating the position and length of five cDNA clones that constitute the complete open reading frame of *Gene W*.

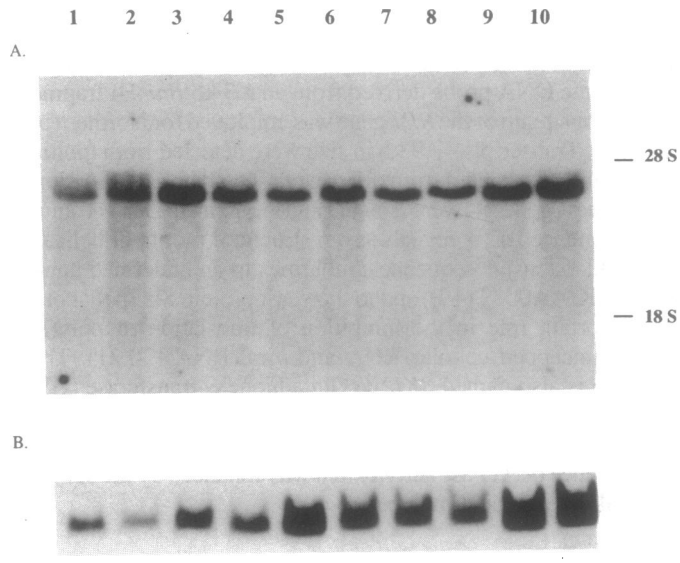

Figure 2. Northern blot analysis of human RNAs hybridized to (A) *SKI2W* probe and (B) human glyceraldehyde-3-dehydrogenase probe (57). RNA samples present in the gel are: lanes 1 and 2, HT29; lanes 3 and 4, IMR32; lanes 5 and 6, BJAB; lanes 7 and 8, NPC/LC; lanes 9 and 10, Raji.

0.25–1% β-mercaptoethanol. Protein molecules solubilized in 4 M urea were resolved by 10% SDS–PAGE with molecular weight markers and electroblotted onto a PVDF membrane. The area of the membrane with 160 kDa proteins was excised; membrane-immobilized fusion proteins were denatured and renatured by five sequential treatments of guanidinium thiocyanate from concentrations of 8, 6, 3, 1.5–0.75 M, and followed by ATPase reaction buffer.

ATPase assays were performed at 37°C for 1 h in 25 mM Hepes, pH 7.6, 5 mM $MgCl_2$, 1 mM dithiothreitol, 33 nM [γ-^{32}P]ATP and using ~100 pmol of purified SKI2W fusion protein (29). The reaction was stopped by the addition of an equal volume of formamide dye, and resolved on a 20% polyacrylamide sequencing gel or with a thin layer chromatography plate. The ATPase activities were compared by adding oligo-dT, a single-stranded oligonucleotide W (5′-GCC CTT GTC CTC AGT GCA-3′) and tRNA.

RESULTS

Discovery of a new gene located upstream of *RP1*

Restriction maps and sequence data suggest that *RD* and *RP1* (12,30) genes are ~11 kb apart (Fig. 1A). An investigation was initiated for the possibility of an unidentified gene between *RD* and *RP1*. Figure 2 shows the result of a Northern blot analysis using RNA samples isolated from human cell lines (colon carcinoma, neuroblastoma and B lymphomas) and hybridized to a DNA probe corresponding to the intergenic region between *RD* and *RP1*. A single major transcript ~3.9 kb in size was detected in all samples tested, which confirmed the prediction of a new gene between *RD* and *RP1*. This new gene was tentatively named *W*.

Three different human cDNA libraries (prepared from lymphocytic RNAs) were screened using a 1.3 kb *Bam*HI fragment, located upstream of the *RP1* gene, as a probe, *W* cDNA probes derived afterwards, and *W* 5′ genomic DNA probes derived from cosmid clones. These libraries yielded four overlapping cDNA clones spanning ~3.6 kb (Fig. 1B) including a poly(A) tail. The RACE method was employed to determine the transcriptional start site, which yielded an additional 198 bp of cDNA sequences. We consider that the 5′ end of the *Gene W* cDNA was complete for two reasons: (i) the total size of the *W* cDNA clone matches that revealed by Northern blot analysis; (ii) the *RD* and *SKI2W* genes are arranged in head-to-head configuration, the putative transcriptional start site of *Gene W* is only 142 bp away from the transcriptional start site of the *RD* gene (L. M. Shen, A. R. Mendoza, A. W. Dangel and C. Y. Yu, manuscript in preparation).

The entire *W* cDNA sequence was determined and submitted to the EMBL data base (accession number Z48796). The compiled *Gene W* cDNA spans 3874 bp that includes a 52 bp 5′ untranslated region and a 33 bp 3′ untranslated region. The polyadenylation signal is an unusual variant GATAAA that is only five nucleotides upstream of the poly(A) tail. The cDNA sequence contains an open reading frame of 1246 amino acid residues. The predicted molecular weight of the putative protein is 137 kDa before post-translational processing/modification.

The most striking feature of the derived amino acid sequence for W is the presence of a helicase domain located between residues 332 and 706. This helicase domain contains all the seven boxes of helicase motifs, of which Box 1 (residue 332–339), Box 1a (residues 352–362) and Box 2 (residues 419–428) may be involved in nucleotide binding, ATPase activity and/or Mg^{2+} binding, Box 3 (residues 448–460) and Box 6 (residues 699–706) may be involved in DNA or RNA unwinding (helicase) activity. The roles for Box 4 (residues 419– 428) and Box 5 (residues 651–669) are not clear. The presence of DEVH at Box 2, SAT in

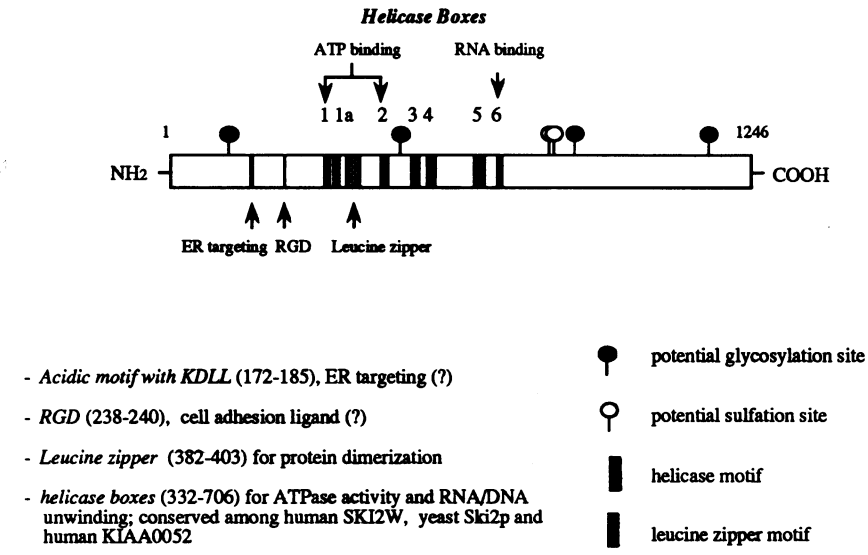

Figure 3. Structural motifs of human SKI2W and their possible functions.

Box 3 and QXXGRAGR in Box 6 would suggest that *Gene W* codes for an RNA helicase (31–33).

A leucine zipper motif is present between Box 1a and Box 2 of W that may be important in the dimerization of protein molecules. Leucine zipper motifs have been found N-terminal to the helicase domain in yeast RAD5 (a DNA repair protein) (34) and C-terminal to the helicase domain in the bacteria *mfd* (a transcription-repair coupling protein) (35). This is probably the first example noting the presence of a leucine zipper motif within a helicase domain.

Located between residues 172 and 185 is an acidic region with six E (Glu) residues followed by KDLL. This motif resembles the endoplasmic reticulum targeting sequence that is characterized by a KDEL motif preceded by a stretch of acidic residues (36). An RGD motif is present at residues 238–240. The RGD motif is a ligand to integrins important in cell-adhesion (37,38). The presence of such a motif for adhesive recognition and receptor interaction in an RNA helicase protein is puzzling. However, such a motif is also present in the HIV transactivator protein Tat that is an RNA binding protein regulating lentivirus transcription (39). A diagram depicting the structural features of human SKI2W is shown in Figure 3.

Human *Gene W* encodes a protein homologous to the yeast viral RNA translation inhibitor SKI2

A database search revealed that the derived W amino acid sequence is extensively similar to the yeast antiviral protein Ski2p. The putative human W protein and the yeast Ski2p contain 1246 and 1285 amino acid residues, respectively. Overall, these two proteins share 39.1% sequence identity, and 58.8% sequence similarity when conservative amino acid changes were taken into account. There are three regions where >200 residues of the two proteins are 65.8–85.9% similar to each other. These highly homologous regions are located at the helicase domain and at the C-terminus. The two proteins both contain a leucine zipper motif between the helicase Box 1a and Box 2. The C-termini of the yeast Ski2p and human W are hydrophobic and almost identical, except that human W has one extra residue. Based on these striking similarities between yeast Ski2p and the product of human *Gene W*, it is presumed that human W is a homolog of the yeast Ski2p. Therefore, *Gene W* is renamed *SKI2W*. However, the human SKI2W and the yeast Ski2p contain distinct features. For examples, the human SKI2W does not contain an RGG motif between helicase Box 4 and Box 5 that may be involved in RNA binding for many nucleolus proteins (20,40); the RGD motif located close to the N-terminal region of human SKI2W is absent in the yeast Ski2p.

A putative human protein KIAA0052 with marked similarity to the human SKI2W and yeast Ski2p: emergence of a new helicase family

Further search for SKI2W related sequences with the NIH databases using the BLASTMAIL program revealed another protein sequence KIAA0052 (data base accession number D29641) with marked sequence similarity to SKI2W and to Ski2p. KIAA0052 is a putative protein whose amino acid sequence was derived from a 3353 bp clone isolated from a human immature myeloid cell line KG1 cDNA library (19,41). Presumably this cDNA clone is incomplete with a truncated 5′ end. The assignment of the open reading frame for this clone was 837 amino acids with a 3′ untranslated region of 841 bp (nucleotides 2512–3353). An alignment of this reading frame with the yeast Ski2p and the human SKI2W sequences revealed a 58% sequence similarity (36–37% sequence identities) of KIAA0052 to each protein. Extensive sequence similarities among human SKI2W, human KIAA0052 and yeast Ski2p suggested the presence of a new gene family that may be involved in translational regulation of cellular and/or viral poly(A)$^-$ RNA molecules.

Expression of human SKI2W in insect cells and purification of SKI2W fusion proteins

To characterize the biochemical and functional properties of SKI2W, it is imperative to obtain large quantities of SKI2W

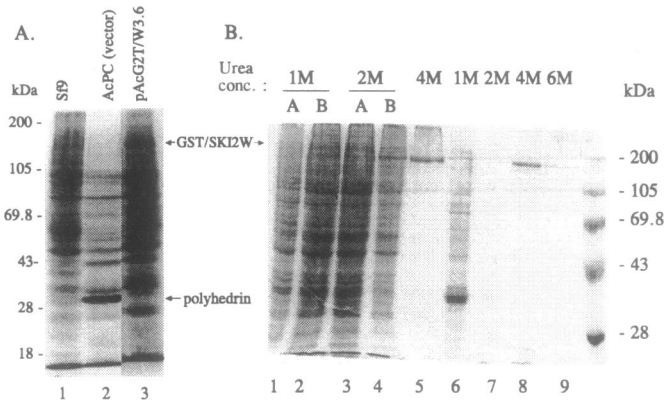

Figure 4. (**A**) Expression of GST–SKI2W fusion protein. *Sf9* insect cell proteins were metabolically labeled by [^{35}S]Met and [^{35}S]Cys for 3 h at 27°C, resolved by SDS–PAGE and subjected to autoradiography. (**B**) Solubilization and purification of GST–SKI2W fusion proteins. *Sf9* cell lysates after 1 day (lanes 1–5) and 3 days (lanes 6–9) of pAcG2T/W3.6 infection were subject to sequential solubilization in 1, 2, 4 M (and 6 M) urea and centrifugation. Solubilization of pellets with the same concentration of urea buffer (1 and 2 M) was repeated for lanes 2 and 4. The soluble proteins in the supernatant were resolved by SDS–PAGE and stained by Coomassie blue.

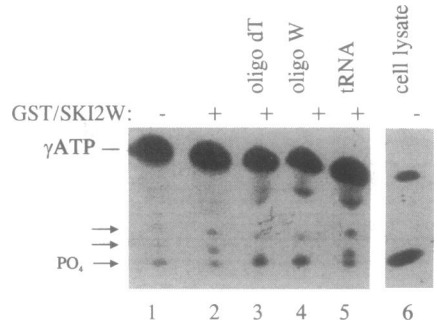

Figure 5. An assay of ATPase activity of renatured GST–SKI2W immobilized on PVDF membrane. Equal amounts of γ-ATP were used as substrate for each reaction. The reaction products were resolved by SDS–PAGE with 20% polyacrylamide gel and subject to autoradiography. Arrows indicate reaction products which contained radioactive phosphate groups.

protein. A near full length 3.6 kb *SKI2W* cDNA (W3.6K) was constructed from three different restriction fragments derived from three overlapping *SKI2W* cDNA clones. This W3.6K cDNA encodes 1197 amino acids of the SKI2W protein without the first 49 residues at the N-terminal region.

The 3.6 kb *SKI2W* cDNA fragment was ligated into the baculovirus transfer vector pAcG2T. The recombinant SKI2W products were expressed as a fusion protein with GST protein under the control of the strong AcNPV polyhedrin promoter in the transfer vector. Figure 4A shows an autoradiograph of a SDS–PAGE analysis of *Sf9* total proteins metabolically labeled by [^{35}S]methionine and [^{35}S]cysteine for 3 h. Lane 1 is a negative control using uninfected *Sf9* cells, lane 2 is a positive control using *Sf9* cells infected by wild-type baculovirus AcNPC (a major 29 kDa polyhedrin protein is present), lane 3 is the *Sf9* cells infected with the pAcG2T/W3.6 recombinant virus. A 160 kDa GST–SKI2W fusion protein is detectable in lane 3.

The GST–SKI2W fusion protein was present in the inclusion bodies of the insect cells. Trials using glutathione–Sepharose to bind the GST–SKI2W fusion protein from the soluble fraction were unsuccessful. Subsequently, purification of SKI2W fusion proteins was performed by sequential treatment of insect cell lysates with 1–4 M urea solutions and centrifugation to pellet insoluble proteins. By increasing the concentration of urea in the presence of the detergent NP-40, most of the cellular proteins except the GST–SKI2W were solubilized in 2 M urea (lanes 1–4, Fig. 4B). Although the 160 kDa polypeptide was detectable in 2 M urea solution, the majority of GST–SKI2W in inclusion bodies could be dissolved only in 4 M urea (lanes 5 and 8). Thus, the GST–SKI2W fusion protein was purified by sequential precipitation, centrifugation and solubilization in 1, 2 and 4 M urea. The proteins solubilized in 4 M urea were further resolved by SDS–PAGE with 10% polyacrylamide, electroblotted onto a PVDF membrane and the area of the membrane with the 160 kDa protein was excised. The immobilized proteins were subject to a denaturation and renaturation step for refolding and reactivation. These procedures were adopted: (i) to circumvent the difficulties of protein precipitation experienced during dialysis to reduce the urea concentration; (ii) to minimize the possible contamination of the GST–SKI2W protein preparation by other insoluble proteins of different molecular weights. Under the same conditions, *Sf9* insect cells infected only with the baculovirus vector did not yield inclusion bodies and no proteins were recovered in the 4 M urea solution.

Demonstration of ATPase activity of SKI2W

Figure 5 shows an assay of ATPase activity by the highly purified and partially renatured GST–SKI2W fusion protein immobilized on a PVDF membrane. The negative control was PVDF membrane without GST–SKI2W fusion protein (lane 1). The positive control was a reaction of total HeLa cell extract with the γ-ATP that almost completely hydrolyzed the radioactive substrate to yield radioactive, inorganic phosphate (lane 6). Slight enhancement of ATPase activity was observed when GST–SKI2W alone was used (lane 2) or supplemented with tRNA (lane 5). About 3–5-fold increase in ATPase activities was detected when the reaction substrates were supplemented with oligo-dT (lane 3) or a pyrimidine-rich oligonucleotide (lane 4). Similar results were obtained from more than three independent experiments.

DISCUSSION

Helicases were initially recognized as essential components of DNA replication through their activity to unwind the double helix (42). The energy required for this process is provided by the hydrolysis of nucleoside triphosphates (ATP or GTP). Recent advances have shown that helicases are involved in a wide spectra of activities involving DNA, RNA and DNA–RNA hybrids. Besides being an essential component of DNA repair (43), helicase domains are present in the transcription factor TFIIH (in human) and in the transcription-DNA repair coupling protein MutS (in *Escherichia coli*) (35). They are also found in protein complexes involved in DNA recombination such as the Ku-80 in the immunoglobulin gene rearrangement (44), in RNA splicing factors (45–48), and in the translation elongation factor eIF-4A (49). Interestingly, the endonuclease subunit responsible for the restriction–modification of DNA in *E.coli* also contains the

conserved helicase domain (50). Sequence alignment of >30 helicases from a wide range of organisms revealed seven consensus structural motifs among all helicases which are named Box 1, Box 1a and Boxes 2–6. The helicases are further subdivided into superfamily I (SFI) that contains Asp–Glu–Ala–Asp in Box 2 (DEAD box) and His–X–X–Gly–Arg–Ala–Gly–Arg (HXXGRAGR) in Box 6, and superfamily II (SFII) that contains Asp–Glu–X–His (DEXH box) and Gln–X–X–Gly–Arg–Ala–Gly–Arg (QXXGRAGR) at the analogous positions (18,31,50,51). The presence of DEVH in Box 2, SAT in Box 3, and QMAGRAGR in Box 6 infers that human SKI2W is a member of the SF-II helicase family.

In the MHC there are multiple precedents for genes of related functions to be closely located together, such as the genes coding for subunits for the complement C3 convertases (BF, C2, C4A and C4B) in the class III region, and for the antigen processing proteins TAP1 and TAP2 in the class II region (52,53). It may be more than a coincidence for *RD*, *SKI2W* and *RP1* genes to be tightly linked. We speculate that their protein products may have related functions. All three genes are constitutively expressed and appear to encode proteins involved in nucleic acid metabolism (12,15,16,30), making them candidate genes for MHC-associated diseases and cancers.

The yeast Ski2p is involved in an important mechanism that defends the microbe against virus propagation (54). Molecular cloning of the yeast *SKI2* gene and subsequent elegant genetic and biochemical experiments revealed that Ski2p exerts its antiviral action by translational inhibition of poly(A)$^-$ RNA (20,21). A substantial proportion of double-stranded and single-stranded RNA viruses in yeast and also in humans does not contain the poly(A) tail and/or is not capped. Thus, Ski2p may be regarded as a defence protein at the nucleic acid level, an approach also adopted by the restriction–modification enzyme in *E.coli* (50). Ski2p may also inhibit or regulate the translation of the cellular poly(A)$^-$ or uncapped RNAs such as histone mRNA, degraded transcripts or structural RNAs (ribosomal RNA and tRNAs), which may be important in maintaining normal cellular function.

The extensive amino acid sequence identities of 36–39% among yeast Ski2p, human SKI2W and human KIAA0052 protein suggests that these three proteins are evolutionarily related. The three genes probably diverged from each other almost at the same time since their sequence identities are very similar among themselves. The most conserved regions among the three proteins are located at the helicase domains and the C-terminal regions. All three proteins contain a leucine zipper motif between Box 1a and Box 2 of the helicase domain, suggesting that these proteins may interact with themselves or with other protein molecules through dimerization or multimerizations.

The demonstration of the ATPase activity for GST–SKI2W fusion protein is an important step towards understanding the functional properties of SKI2W. The low solubilities of GST–SKI2W fusion protein caused considerable difficulties in the purification of active protein for functional studies. Sequential solubilization and centrifugation of the fusion proteins by increasing the concentration of urea enabled purification of denatured protein to 80–90% purity, as judged by SDS–PAGE. Further purification by polyacrylamide gel electrophoresis and electroblotting to PVDF membrane enabled renaturation of the fusion proteins without gross re-precipitation. This method enabled the demonstration of ATPase activity for SKI2W. Further work is necessary to determine if the ATPase and possibly the helicase activities of SKI2W are induced by single-stranded or double-stranded DNA or RNA molecules, and by poly(A)$^+$ or poly(A)$^-$ RNA molecules.

While this paper was being prepared, a partial human cDNA sequence *170A* with a reading frame showing striking similarity to yeast Ski2p was published (55). This clone was obtained by screening an expression library with a monoclonal antibody (mAb) 170A1. Indirect immunofluorescence studies showed that the antigen for mAb 170A1 is expressed in the nucleolus and also in the endoplasmic reticulum. The human *170A* cDNA sequence has a continuous reading frame of 729 amino acids in its 3094 bp sequence. The *170A* cDNA sequence is almost identical to the 3′ cDNA sequence of human *SKI2W*. However, two stop codons were present at positions 2188 and 2293 of the 170A cDNA that resulted in premature truncation of the protein (55). These stop codons are not present in the *SKI2W* cDNA sequence. Compared with the full length amino acid sequence of human SKI2W, it became clear that the 729 amino acid reading frame from *170A* has missed the first 239 amino acid residues at the N-terminal region and the last 300 amino acid residues at the C-terminal region of SKI2W. Moreover, the first 10 and the last 14 amino acids of 170A did not match to human SKI2W. The close sequence identities between SKI2W and 170A would suggest that they are from the same gene. The discrepancies described above could be due to sequencing errors or mutations.

ACKNOWLEDGEMENTS

We are indebted to Dr Smita Patel (Department of Biochemistry, The Ohio State University) for help and advice on assays of ATPase activities, Shangxiang Zhang for trial experiments to express SKI2W in bacteria, Dr Sue O'Dorisio and Bradley Baker for reviewing the manuscript, Dr Reed Wickner (National Institute of Health) for sharing a preprint before publication, and Dr Nobuo Nomura (Kazusa DNA Research Institute, Japan) for informing us the gene name of the data base entry D29641. This work was aided in part by a Basil O'Conner Starter Scholar Award from the March of Dimes Foundation for Birth Defects (5-FY94-0774), the Children's Hospital Research Foundation, Columbus, Ohio (020-313 and 020-832), an Eastern Star Award for Cancer Research, and the Pittsburgh Supercomputing Center through the NIH Center for Research Resources Cooperative agreement (1P41 RR06009) (C.Y.Y.). L.C.W. is supported by NIH grant GM48798-02.

REFERENCES

1 Trowsdale,J., Ragoussis,J. and Campbell,R.D. (1991) *Immunol. Today* **12**, 443–476.
2 Lu,S., Day,N.E., Degos,L., Lepage,V., Wang,P.C., Chan,S.H., Simons,M., McKnight,B., Easton,D., Zeng,Y. and de-The,G. (1990) *Nature* **346**, 470–471.
3 Coffin,C.M., Rich,S.S. and Dehner,L.P. (1991) *Cancer* **68**, 1323–1328.
4 Torres,A., Martinez,F., Gomez,P., Comez,C., Garcia,J. and Nunez–Roldan,A. (1980) *Cancer* **46**, 838–843.
5 Bowers,T.K., Moldow,C.F., Bloomfield,C.D. and Yunis,E.J. (1977) *Vox Sang.* **33**, 273–277.
6 Kalidi,I., Masset,M., Gony,J., Marcelli,A., Jeannet,M., Irle,C., Bauters,F., Tchernia,G., Turpin,F., Gisselbrecht,C. and Hors,J. (1989) *Nouv. Rev. Fr. Hematol.* **31**, 149–152.

7 Schaffer,F.M., Palermos,J., Zhu,Z.B., Barger,B.O., Cooper,M.D. and Volanakis,J.E. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8015–8019.
8 Demant,P., Oomen,L. and Oudshoorn–Snoek,M. (1989) *Adv. Cancer Res.* **53**, 117–179.
9 Oomen,L., van der Valk,M., Hart,A., Demant,P. and Emmelot,P. (1988) *Cancer Res.* **48**, 6634–6641.
10 Malo,D. and Skamene,E. (1994) *Trends Genet.* **10**, 365–371.
11 Melhem,M.F., Kunz,H.W. and Gill,T.J.III (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1967–1971.
12 Shen,L.M., Wu,L.C., Sanlioglu,S., Chen,R., Mendoza,A.R., Dangel,A., Carroll,M.C., Zipf,W. and Yu,C.Y. (1994) *J. Biol. Chem.* **269**, 8466–8476.
13 Perozzi,G. and Prakash,S. (1986) *Mol. Cell. Biol.* **6**, 1497–1507.
14 Levi-Strauss,M., Carroll,M.C., Steinmetz,M. and Meo,T. (1988) *Science* **240**, 201–204.
15 Surowy,C.S., Hoganson,G., Gosink,J., Stunk,K. and Spritz,R.A. (1990) *Gene* **90**, 299–302.
16 Cheng,J., Macon,K.J. and Volanakis,J.E. (1993) *Biochem. J.* **294**, 589–593.
17 Wu,L.-C., Morley,B.J. and Campbell,R.D. (1987) *Cell* **48**, 331–342.
18 Gorbalenya,A.E., Koonine,E.V., Donchenko,A.P. and Blinov,V.M. (1989) *Nucleic Acids Res.* **17**, 4713–4730.
19 Normura,N., Miyajima,N., Szuka,T., Tanaka,A.,, Kawarabayashi,Y., Nagase,T., Isshikawa,K., Seki,T. and Tabata,S. (1994) GenBank Accession no. D29641.
20 Widner,W.R. and Wickner,R.B. (1993) *Mol. Cell. Biol.* **13**, 4331–4341.
21 Masison,D., Blanc,A., Ribas,J.C., Carroll,K., Sonenberg,N. and Wickner,R.B. (1995) *Mol. Cell. Biol.* **15**, 2763–2771.
22 Dangel,A.W., Mendoza,A.R., Baker,B.J., Daniel,C.M., Carroll,M.C., Wu,L–C. and Yu,C.Y. (1994) *Immunogenetics* **40**, 425–436.
23 Ausubel, F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1987). *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, NY.
24 Bairoch,A. and Bucher,P. (1994) *Nucleic Acids Res.* **22**, 3583–3589.
25 Pearson,W.R. and Lipman,D.J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
26 Genetics Computer Group, Inc. (1991). *GCG Sequence Analysis Software Package: Program Manual, Version 7*, Genetics Computer Group, Inc., Madison.
27 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.* **215**, 403–410.
28 O'Reilly,D., Miller,L.K., and Luckow,V.A. (1992) *Baculovirus Expression Vectors, A Laboratory Manual*, W.H. Freeman and Company, New York, NY.
29 Patel,S.S., Rosenberg,A.H., Studier,F.W. and Johnson,K.A. (1992) *J. Biol. Chem.* **267**, 15013–15021.
30 Speiser,P.W. and White,P.C. (1989) *DNA* **8**, 745–751.
31 Bork,P. and Koonin,E.V. (1993) *Nucleic Acids Res.* **21**, 754–752.
32 Plumpton,M., McGarvey,M. and Beggs,J. (1994) *EMBO J.* **13**, 879–887.
33 Fuller-Pace,F.V. (1994) *Trends Cell Biol.* **4**, 271–274.
34 Johnson,R.E., Henderson,S.T., Petes,T.D., Prakash,S., Bankmann,M. and Prakash,L. (1992) *Mol. Cell. Biol.* **12**, 3807–3817.
35 Selby,C.P. and Sancar,A. (1993) *Science* **260**, 53–58.
36 Munro,S. and Pelham,H.R.B. (1987) *Cell* **48**, 899–907.
37 Smith, J.W. (1994) In *Integrins: Molecular and Biological Responses to the Extracellular Matrix* (Cheresh, D.A. and Mecham, R.P., eds) pp. 1–32. Acedemic Press, Inc., San Diego.
38 Kuhn,K. and Eble,J. (1994) *Trends Cell Biol.* **4**, 256–261.
39 Bayer,P., Kratt,M., Ejchart,A., Westendrop,M., Frank,R. and Rosch,P. (1995) *J. Mol. Biol.* **247**, 529–535.
40 Kiledjian,M. and Dreyfuss,G. (1992) *EMBO J.* **11**, 2655–2664.
41 Nomura,N., Miyjima,N., Sazuka,T., Tanaka,A., Kawarabayasi,Y., Sato,S., Nagase,T., Seki,N., Ishiawa,K. and Tabata,S. (1994) *DNA Res.* **1**, 27–35.
42 Kornberg, A. and Baker, T.A. (1992) In *DNA Replication* W. H. Freeman & Co., NY, pp. 355–378.
43 Friedber,E.C. (1992) *Cell* **71**, 887–889.
44 Troelstra,C. and Jaspers,N. (1994) *Curr. Biol.* **4**, 1149–1154.
45 Kim,S.-H., Smith,J., Claude,A. and Lin,R.-J. (1992) *EMBO J.* **11**, 2319–2326.
46 Schwer,B. and Guthrie,C. (1992) *EMBO J.* **11**, 5033–5039.
47 Company,M., Arenas,J. and Abelson,J. (1991) *Nature* **349**, 487–493.
48 Strauss,E.J. and Guthrie,C. (1994) *Nucleic Acids Res.* **22**, 3187–3193.
49 Pause,A., Methot,N., Svitkin,Y., Merrick,W. and Sonneberg,N. (1994) *EMBO J.* **13**, 1205–1215.
50 Gorbalenya,A.E. and Koonin,E.V. (1991) *FEBS Lett.* **291**, 277–281
51 Koonin,E.V. (1991) *Nature* **352**, 290.
52 DeMars,R. and Spies,T. (1993) *Trends Cell Biol.* **2**, 81–86.
53 Hill,A. and Ploegh,H. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 341–343.
54 Wickner, R.B. (1991) In *The Molecular and Cellular Biology of the Yeast Saccharomyces* (Broach, J.R., Pringle, J.R. and Jones, E.W., eds) pp. 263–296. CSHL Press, NY.
55 Lee,S.–G., Lee,I., Park,S.H., Kang,C. and Song,K. (1995) *Genomics* **25**, 660–666.
56 Bristow,J., Tee,M., Gitelman,S., Mellon,S. and Miller,W. (1993) *J. Cell. Biol.* **122**, 265–278.
57 Tso,J.Y., Sun,X.H., Kao,T.H. and Wu,R. (1985) *Nucleic Acids Res.* **13**, 2485–2502.