3762–3768 *Nucleic Acids Research, 1998, Vol. 26, No. 16*

Hex1: a new human Rad2 nuclease family member with homology to yeast exonuclease 1

David M. Wilson III*, James P. Carney[1], Matthew A. Coleman, Aaron W. Adamson, Mari Christensen and Jane E. Lamerdin

Biology and Biotechnology Research Program, L-452, Lawrence Livermore National Laboratory, Livermore, CA 94551, USA, [1]Life Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA and Department of Radiation Oncology, University of California, San Francisco, CA 94103, USA

Received April 13, 1998; Revised May 26, 1998; Accepted June 25, 1998 DDBJ/EMBL/GenBank accession nos AF042282 and AC004783

ABSTRACT

Nucleolytic processing of chromosomal DNA is required in operations such as DNA repair, recombination and replication. We have identified a human gene, named *HEX1* for human exonuclease 1, by searching the EST database for cDNAs that encode a homolog to the *Saccharomyces cerevisiae EXO1* gene product. Based on its homology to this and other DNA repair proteins of the Rad2 family, most notably *Schizosaccharomyces pombe* exonuclease 1 (Exo1), Hex1 presumably functions as a nuclease in aspects of recombination or mismatch repair. Similar to the yeast proteins, recombinant Hex1 exhibits a 5′→3′ exonuclease activity. Northern blot analysis revealed that *HEX1* expression is highest in fetal liver and adult bone marrow, suggesting that the encoded protein may operate prominently in processes specific to hemopoietic stem cell development. *HEX1* gene equivalents were found in all vertebrates examined. The human gene includes 14 exons and 13 introns that span ~42 kb of genomic DNA and maps to the chromosomal position 1q42–43, a region lost in some cases of acute leukemia and in several solid tumors.

INTRODUCTION

DNA repair has evolved to protect cells against the mutagenic and cytotoxic effects of spontaneous, oxidative, alkylation or compound-derived DNA damage (1). Most of these corrective systems require an exo- or endonuclease for processing of chromosomal imperfections. In mammalian mismatch repair, improperly paired bases are recognized by Msh2, which directs a cascade of events that presumably leads to the incision of DNA either 5′ or 3′ of the mismatch and removal of the target DNA strand, prior to re-synthesis and ligation (2–4). During nucleotide excision repair, XPG and the ERCC1/XPF complex incise 5′ and 3′ of the damage respectively and are required for the eventual removal of the damage-containing DNA strand (5). Distinct nucleases are essential for site-specific cleavage and termini processing in base excision repair (6). In addition, 5′→3′ exonucleolytic degradation is a requisite step for generating a 3′ single-stranded terminus necessary for strand invasion and the processing of DNA double-strand breaks during meiotic recombination, as demonstrated most clearly in the yeast *Saccharomyces cerevisiae* (7–9). Integration of exogenous DNA by non-homologous end-joining (i.e. illegitimate recombination) has also been shown to require exonuclease activity (10). While the major nucleases that participate in the excision repair pathways have been identified, the nucleases involved in mammalian mismatch and recombinational repair remain undefined. Instead, much of our current knowledge of these repair processes stems from studies performed in yeast and bacteria.

In search of a recombination-promoting exonuclease, Szankasi and Smith (11) found that *Schizosaccharomyces pombe* possesses a nuclease activity specifically induced during meiosis. This activity, termed exonuclease 1 (Exo1), was shown to act as a 5′→3′ exonuclease at DNA double-strand break ends and nicks. Deletion of the *exo1* gene resulted in elevated spontaneous reversion rates of various alleles of *ade6* (i.e. a mutator phenotype), as well as an altered recombination frequency (12). Detailed analysis of the mutational spectrum of the *exo1* cells suggested a role for the encoded protein in mismatch repair, most likely during homologous recombination.

In *Drosophila melanogaster*, the *EXO1* gene equivalent, *TOS*, was shown to be expressed specifically in the female germline during early embryogenesis, with the highest level of mRNA seen in the developing oocyte, a location of active meiotic recombination (13). These findings indicated a role for the *TOS* gene product, Tosca, in homologous recombination.

Tishkoff and colleagues (14), while searching for yeast factors that interact with the yeast Msh2 mismatch repair protein in two-hybrid screens, found that one interacting clone encoded a 5′→3′ exonuclease with significant homology to *S.pombe* Exo1 and *D.melanogaster* Tosca; this gene was termed *S.cerevisiae EXO1*. Simultaneously, Fiorentini *et al*. (15) demonstrated that extracts from Exo1-deficient *S.cerevisiae* strains were less effective at recombining linear DNA molecules possessing overlapping homologous ends and that *EXO1* encoded a previously characterized 5′→3′ exonuclease that generates 3′ single-stranded complementary tails which promote joint molecule formation (16). Furthermore, *exo1* mutants were found to display a 6-fold reduction in the rate of mitotic recombination between non-tandem duplications (15). Taken together, these findings indicate that the Exo1 homologs function as 5′→3′ exonucleases in aspects of both mismatch and recombinational repair.

Upon identification of *S.pombe* Exo1, Szankasi and Smith (12) noted that this protein displayed significant similarity to the

*To whom correspondence should be addressed. Tel: +1 925 423 0695; Fax: +1 925 422 2282; Email wilson61@llnl.gov

proteins of the Rad2 nuclease family (reviewed in 17). However, these authors also recognized that Exo1 had several unique features which suggested that this protein might comprise a third Rad2 classification.

The first Rad2 class includes the human XPG protein, which cleaves at the 3′-side of the damage-containing bubble structure formed during nucleotide excision repair (18). Class II is comprised of the eukaryotic Fen-1 homologs. The Fen-1 proteins are structure-specific nucleases that remove 5′-DNA flaps produced by polymerase strand displacement (19) and are also responsible for the resolution of Okazaki fragments during DNA replication (20,21). Class III is made up of the Exo1-like enzymes. In addition, phage and bacterial 5′→3′ exonucleases, namely T4 RNase H (22) and T5 D15 5′-exonuclease (23), as well as several proof-reading polymerases, share homology to the Rad2 family members. Much of the amino acid sequence conservation among the Rad2 family is concentrated within two domains, termed the N (N-terminal) and I (internal) nuclease domains (17). Furthermore, many active site catalytic residues, as predicted from the X-ray crystal structure of T4 RNase H (22), are also conserved throughout the Rad2 family and many of these residues have been shown to be essential for nuclease or binding activity of Fen1 (24).

We describe here the identification of a human cDNA and a genomic clone that encodes a protein, termed Hex1, which shares significant homology to *S.cerevisiae* Exo1, *S.pombe* Exo1 and *D.melanogaster* Tosca. We report the initial enzymatic characterization of Hex1, its mRNA expression pattern, the genomic organization and the chromosomal mapping position. The possible biological roles of Hex1 are discussed in light of these findings.

MATERIALS AND METHODS

Identification of *HEX1* cDNA

The dbEST database was searched using the TBLASTN program (25) to identify cDNA clones that contain an open reading frame homologous to the *S.cerevisiae* Exo1 protein. Only human EST clone 843301 (GenBank accesion no. AA489549; Stratagene HeLa library) was found to have significant homology, displaying 44% identity along the first 75 N-terminal amino acids of Exo1; the same clone was identified by searching the dbEST with *S.pombe* Exo1. The entire nucleotide sequence of clone 843301 (obtained from the IMAGE consortium at Lawrence Livermore National Laboratory) was determined using the Primer Island™ *in vitro* transposition system (Perkin Elmer, Foster City, CA). Thirteen clones, each containing a single random Ty1 transposon insertion, were sequenced with transposon-specific primers using Taq FS dye terminator chemistry. The resultant sequence ladders were collected on an ABI 377 automated DNA sequencer. The cDNA sequence was reconstructed using PHRAP (P.Green, University of Washington) and edited using CONSED (D.Gordon, University of Washington). The *HEX1* cDNA sequence has been assigned Genbank accession no. AF042282.

Expression and purification of recombinant Hex1 fusion protein

The *HEX1* coding region was PCR amplified using primers 5′HEXNCO (5′-GGCACCATGGGGATACAGGGAT-3′) and HEX1XTN (5′-CGGGATCCCCGAATTTTTTAAATCCAA-3′) and sub-cloned after partial digestion into the *Nco*I and *Bam*HI sites of pET28a (Novagen, Madison, WI) to generate pET28Hex1-His. This construct permits expression of a six histidine C-terminal-tagged Hex1 fusion protein under control of the T7 RNA polymerase promoter. Protein expression was performed by transforming pET28Hex1-His into BL21 (λDE3) (Novagen) and growing these bacteria in 4 l LB medium containing 20 μg/ml kanamycin. Cells were grown to an OD of 0.6 and isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 0.5 mM. Inductions were carried out for 3 h and the cells were harvested by centrifugation at 3000 *g* for 20 min. Cell pellets were washed once with 0.5× phosphate-buffered saline (PBS) and frozen at –80°C or processed immediately. Cells were thawed on ice, resuspended in cold buffer H (50 mM HEPES–NaOH, 200 mM NaCl, 20% glycerol) and sonicated for two 15 s bursts using a Misonix XL sonicator. All subsequent steps were performed at 4°C. Insoluble material was removed by centrifugation at 20 000 *g* for 20 min and the supernatant, termed 'crude cell extract', was assayed or further processed by application to a Q20 anion exchange column using the BioLogic Workstation (BioRad, Hercules, CA). The Q20 flow-through was collected and imidazole was added to a final concentration of 20 mM. Two millilters of packed Ni-NTA resin (Qiagen, Santa Clarita, CA) equilibrated in buffer H were added to the Q20 flow-through and this mixture was incubated for 1 h with rocking. The Ni-NTA material was then poured into a 5 ml glass column and washed with 10 ml buffer H containing 20 mM imidazole. Bound protein was eluted with a linear gradient of 20–500 mM imidazole in buffer H and the fractions were assayed for exonuclease activity (see below) and analyzed by SDS–PAGE and Coomassie blue staining or western blotting. Immunoblotting was performed with histidine-specific primary antibodies (G-18) and horseradish peroxidase-conjugated secondary antibodies from Santa Cruz Biotechnology (Santa Cruz, CA) and Chem Illuminesence (Amersham, Arlington Heights, IL).

Exonuclease assays

Exonuclease activity was measured by monitoring the release of trichloroacetic acid (TCA)-soluble mononucleotides from restriction enzyme-linearized [^{3}H]PM2 DNA. Briefly, [^{3}H]PM2 DNA was digested with *Hinc*II and used as substrate in 50 μl reactions containing 50 mM HEPES–KOH, pH 7.8, 5 mM $MgCl_2$, 1 mM dithiothreitol, 1 nmol (total nucleotide) [^{3}H]PM2 DNA and a crude cell extract or nickel-purified fraction. Reactions were performed for 30 min at 37°C and stopped by addition of 5 μl 4 mg/ml salmon sperm DNA and 75 μl 10% TCA. Nucleic acid was precipitated for 10 min on ice, the samples were centrifuged for 10 min at 14 000 *g* and the supernatants were assayed for the released [^{3}H]nucleotide monophosphate residues by liquid scintillation counting.

A PCR product spanning the human *MRE11* open reading frame was generated and differentially end-labeled to determine polarity of the exonuclease activity. DNA was 5′-end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase (NEB, Beverly, MA) or 3′-labeled by digestion with *Bam*HI and 'filling-in' with [α-^{32}P]dATP and Klenow fragment (NEB). Exonuclease assays contained 0.1 pmol DNA ends and 0.09 U nuclease activity of peak nickel-purified fractions. Reactions were carried out at 37°C and aliquots stopped by addition of EDTA to 40 mM. All reactions were TCA precipitated and processed as above.

Northern blot analysis

Northern blots were prehybridized for 1 h at 65°C in ExpressHyb Solution (Clontech, Palo Alto, CA). *HEX1* PCR product (see above) was labeled using the Megaprime DNA Labeling System (Amersham) and [α-^{32}P]dCTP (Amersham) and hybridized in ExpressHyb Solution at 65°C for 1–2 h. Blots were washed once at room temperature for 20 min and twice at 50°C for 20 min with 50 mM $NaPO_4$, pH 7.4, 0.5% SDS, 1 mM EDTA. Images were obtained using a Molecular Dynamics PhosphorImager Storm 860 or autoradiography. Blots were normalized with β-actin (Clontech).

HEX1 genomic DNA isolation and chromosomal mapping

An arrayed human genomic BAC library (Research Genetics, Huntsville, AL) was screened by hybridizing with ^{32}P-labeled *HEX1* PCR product. First round hybridization identified six positive human BAC clones (76330, 91679, 243587, 247473, 266482 and 288592). BAC 266482 hybridized with both 5′- and 3′-end *HEX1* cDNA fragments and was chosen for sequencing. The entire BAC (98.6 kb) was sequenced using a shotgun approach as previously described (26). The sequence of the BAC insert was reconstructed using PHRAP and edited using CONSED. The *HEX1* cDNA sequence was aligned with the genomic sequence to define intron–exon boundaries using TBLASTX and BLASTN (25). Additional genomic elements within BAC 266482 were identified using XGRAIL 1.3 (27) and putative transcription factor binding sites were identified with TRANSFAC, TRRD, COMPEL and MatInspector (28,29). The sequence from BAC 266482 has been assigned Genbank accession no. AC004783.

Chromosomal mapping by fluorescence *in situ* hybridization (FISH) with *HEX1* BAC clones 266482 and 243587 was performed as previously described (30). Briefly, biotinylated BAC probes were generated by nick translation and hybridized to prometaphase chromosome spreads obtained from synchronized human lymphocyte cultures. Probes were detected by two layers of avidin–FITC and localized using concurrent DAPI–actinomycin bands.

RESULTS

Identification of a human homolog of *S.cerevisiae* Exo1

To identify mammalian nucleases that may participate in recombinational repair, the EST database was searched for cDNAs encoding proteins with homology to *S.cerevisiae* Exo1. The strongest identity was seen with IMAGE clone 843301. Dideoxy nucleotide sequencing of the 843301 insert exposed an open reading frame of 803 amino acids with the ATG start codon in a near consensus Kozak sequence of GGCACCATGG (31). This protein, termed Hex1 (for human exonuclease 1), has a predicted molecular weight of 89.1 kDa and a theoretical pI of 8.8 (ExPASy, ProtParam tool). Two potential nuclear localization signals, KRPR (at amino acids 418–421) and KRKH (at amino acids 775–778), are also present. When queried against the NCBI non-redundant database, Hex1 was found to display significant homology to *D.melanogaster* Tosca (~33% identity across the entire protein) and the yeast Exo1 proteins (~26% identity) and showed some, but less, homology to XPG (~17% identity across the entire protein) and Fen1 (~26% identity across the length of Fen1). The amino acid alignment of Hex1 with Exo1 of *S.pombe*, Exo1 of *S.cerevisiae* and Tosca revealed several highly conserved amino acid blocks in the first ~300 residues of the N-terminus (Fig. 1), with most conservation occurring in the N and I nuclease domains of the Rad2 family (17). Furthermore, many of the active site residues of T4 RNase H (22) shown to be required for the nuclease or binding activity of Fen1 (24) are conserved in Hex1. Using the *HEX1* cDNA as a probe, Southern blot analysis revealed that gene equivalents are present in monkey, rat, mouse, dog, cow, rabbit and chicken (data not shown).

Exonuclease activity of Hex1

Data obtained with the yeast Exo1 proteins, plus conservation of the N and I domains and the presumed catalytic residues, indicated that Hex1 may act as an exonuclease. To test for such an enzymatic function, crude cell extracts were prepared from IPTG-treated bacteria possessing either the pET28a vector or a Hex1 expression construct and assayed for general exonuclease activity. Extracts prepared from bacteria harboring independent pET28Hex1-His plasmid isolates exhibited on average a 13-fold higher specific activity than extracts prepared from bacteria carrying pET28a (Table 1). Similarly, crude extracts prepared from cells expressing non-fusion Hex1 protein (from pET11Hex1) displayed a higher exonuclease activity than the vector controls (Table 1). Western blot analysis confirmed that full-length His-tagged Hex1 protein was expressed in bacteria containing a recombinant expression construct (Fig. 2A).

Table 1. Exonuclease activity of crude cell extracts

Clone	Specific activity (U/mg)
BL21/pET28a	1.0 (0.2)
BL21/pET28a-Hex1-His	12.8 (3.6)
BL21/pET11-Hex1	6.1 (1.1)

Crude cell extracts were prepared from *E.coli* expressing Hex1 as a 6× His-tagged fusion protein (pET28a-Hex1-His) or an untagged protein (pET11-Hex1) or from *E.coli* carrying pET28a (mock). Extracts were assayed for exonuclease activity, indicated as specific activity (U/mg), as described in Materials and Methods. Standard errors from at least three individual determinations and at least two independent recombinant plasmid isolates are indicated in parentheses. One unit is equivalent to the release of 10 nmol TCA-soluble nucleotide monophosphate in 30 min at 37°C.

C-Terminal-tagged Hex1-His protein was partially purified from IPTG-induced *E.coli* extracts by anion exchange and nickel chromatography. Fractions eluted from the nickel column were assayed for general exonuclease activity and the peak of activity, which corresponded to the peak of Hex1-His protein as determined by western blotting, was centered at fraction 11 (Fig. 2B). Fractions from an equivalent mock purification using an extract prepared from *E.coli* carrying the pET28 vector showed only background levels of exonuclease activity (Fig. 2B). One point worth noting is the amount of degradation that was observed during fractionation of the Hex1-His protein. This instability was most prominent following dilution or freeze–thaw cycles and was not remedied by the addition of protease inhibitors. Similar degradation has been observed with the yeast Exo1 proteins (11,14,15).

In order to determine the polarity of the Hex1 exonuclease activity, fraction 11 from the nickel column was used in assays containing DNA labeled specifically at either the 5′- or 3′-end. Affinity-purified Hex1-His was found to release an increasing amount of radioactivity from 5′-^{32}P-end-labeled DNA substrates over time, with ~50% of the total counts released in 30 min

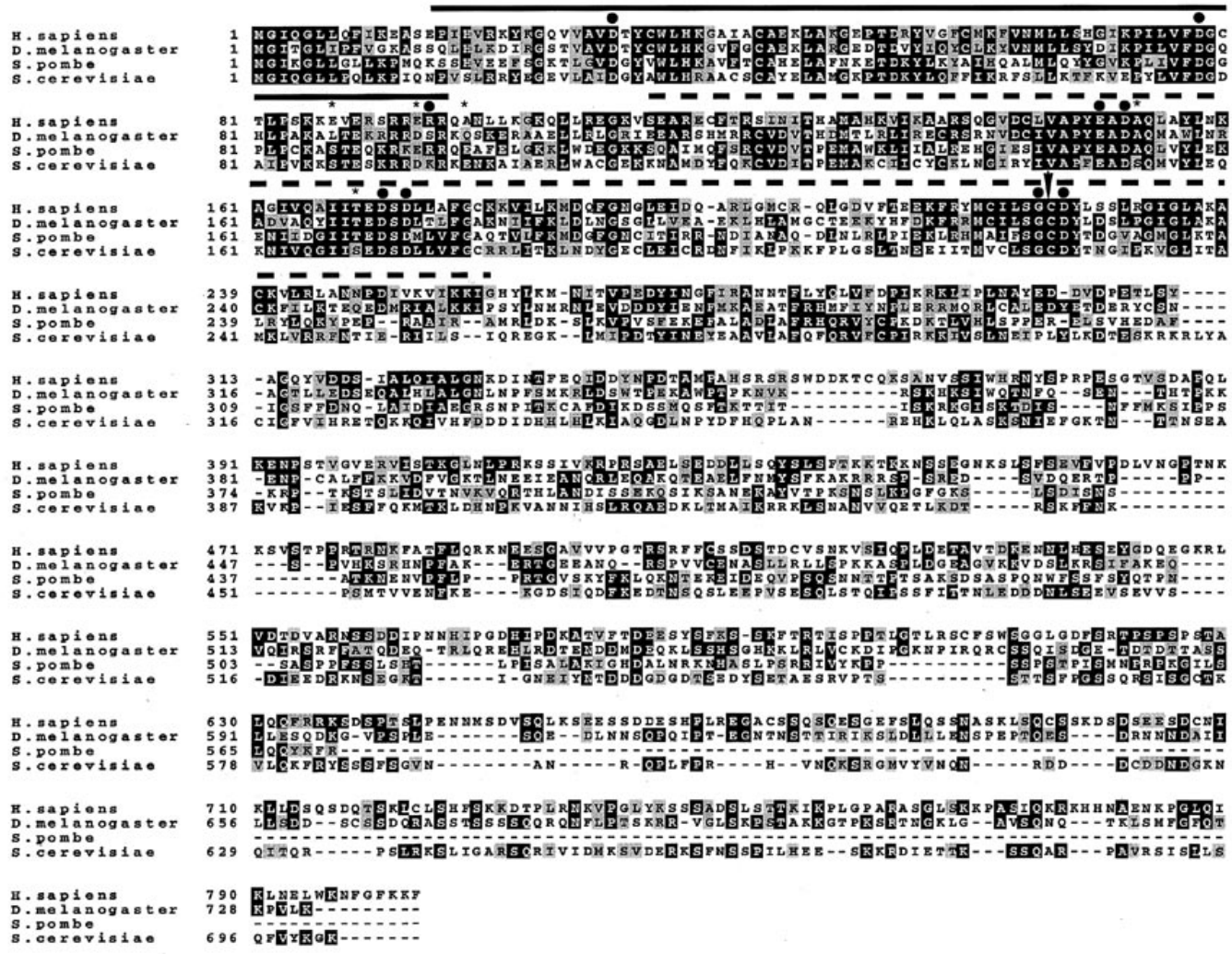

Figure 1. Alignment of the Rad2 class III proteins: *H.sapiens* Hex1, *D.melanogaster* Tosca (13), *S.pombe* Exo1 (12) and *S.cerevisiae* Exo1 (14). Identical residues are depicted in dark blocks. Gray blocks indicate conserved amino acid differences, while white indicates no sequence conservation. Dashes represent gaps or no residue in the sequence alignment. Solid line shows the N domain, whereas the dashed line indicates the I domain. The solid circles represent amino acid residues of Hex1 that are identical to residues in the active site of T4 RNase H (22), the asterisks represent T4 RNase H active site residues that are not identical in Hex1 and the arrow depicts the position of an active site residue that is absent from the Exo1-like proteins. Protein alignment was performed using CLUSTAL (Baylor College of Medicine) and the figure was produced with Boxshade (http://ulrec3.unil.ch/software/BOX_form.html).

(Fig. 2C). Simultaneous assays with the same Hex1-His-containing fractions showed only background counts being released from 3′-^{32}P-labeled DNA substrates over 30 min (Fig. 2C). These results indicate an exonuclease activity with a 5′→3′ polarity, consistent with the activity observed for the yeast Exo1 proteins (11,15).

Expression pattern of *HEX1* RNA

To determine the tissue specificity of *HEX1* mRNA expression, human multi-tissue northern (MTN) blots were probed. A single ~4 kb *HEX1* transcript was detected prominently in fetal liver and adult bone marrow, whereas only low levels of mRNA were observed in the other fetal tissues or in adult stomach, thyroid, spinal cord, lymph node, trachea and adrenal gland (Fig. 3). We saw similar low level expression in adult heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small intestine, colon (no mucosa) and peripheral blood leukocytes (Clontech MTN I and II; data not shown). Irradiation of HeLa cells with 2 or 10 Gy revealed no obvious induction of the normally low levels of *HEX1* mRNA (data not shown).

Genomic organization and chromosomal mapping of *HEX1*

The entire coding region, including the 5′- and 3′-untranslated regions (UTRs) of the *HEX1* cDNA, were contained in a single BAC clone of 98 574 bp; the cDNA and genomic sequences were identical. Alignment of the genomic and cDNA nucleotide sequences revealed that *HEX1* consists of 14 exons and 13 introns spanning ~42 kb of DNA (Fig. 4). The intron–exon boundaries all fit the consensus splice donor–acceptor rule (32). The sizes of the exons ranged from 96 to 595 bp, with exon 2 possessing the translational start. The first exon was comprised of only untranslated sequence. It should be noted that XGRAIL was

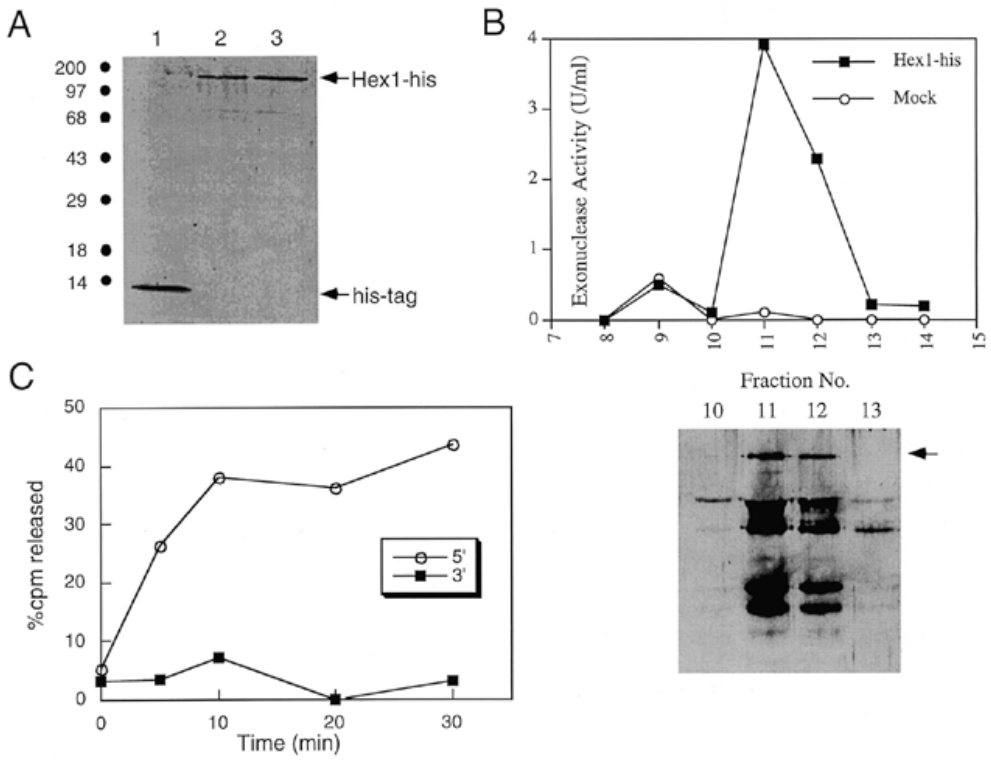

Figure 2. Analysis of Hex1 exonuclease activity. (**A**) Western blot analysis of crude cell extracts. Extracts were prepared from IPTG-induced *E.coli* carrying either pET28a (lane 1) or pET28a/Hex1-His (lane 2 and 3). Twenty five micrograms of total protein was fractionated on an SDS–polyacrylamide gel, transferred to nitrocellulose and probed with His-specific antibodies as described in Materials and Methods. Position of full-length Hex1-His, the His tag portion alone and the molecular weight protein standards (in kDa) are indicated. (**B**) Chromatogram of Hex1-His activity from a nickel affinity column. Fractions eluted from the nickel affinity column were assayed for exonuclease activity as described in Materials and Methods. Activity (in U/ml) is indicated for each fraction from the Hex1-His purification and the equivalent mock purification (top). Additionally, the fractions from the Hex1-His purification were subjected to western blot using a His-specific antibody (bottom). The arrow indicates the position of the full-length Hex1-His protein. (**C**) Polarity of Hex1-His exonuclease activity. Assays containing end-labeled PCR product (0.1 pmol) as indicated and an aliquot of a Hex1-His-containing fraction (0.09 U) were sampled and processed at the indicated times as described in Materials and Methods. The percent released of the initial radioactivity is shown. Results are representative of three independent experiments.

unable to identify exons 3, 9 and 12 and predicted a different exon from those found in the *HEX1* cDNA. Finally, within the entire 98 574 bp, XGRAIL identified only one other convincing open reading frame (located upstream of *HEX1*), which shares significant identity with the ribosomal protein RL6.

Several consensus trascription factor binding sequences were found near the 5′-end of the sequenced *HEX1* IMAGE cDNA (Fig. 4B). This upstream region, which lacks a clear TATA box-like promoter, is preceded by a CpG island at position –1250 relative to the cDNA 5′-end (Fig. 4A). Two additional CpG Islands are located within the *HEX1* gene (Fig 4A). FISH analysis, performed with two different BAC DNA isolates as well as the *HEX1* cDNA PCR product, revealed that the gene mapped to chromosome position 1q42–43 (Fig. 5).

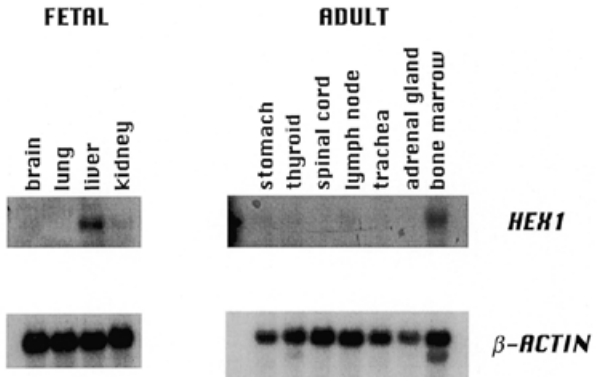

Figure 3. Expression pattern of *HEX1* mRNA in fetal and adult human tissues. Northern blot analysis was performed (see Materials and Methods) and the hybridization signal detected by phosphorimaging or autoradiography. Images were produced by scanning the autoradiograph using Adobe Photoshop 3.0. Developmental stage, tissue type and transcript are indicated.

DISCUSSION

The biological role of Hex1

We report here the identification of the first mammalian *RAD2* class III family member, *HEX1* (17). The 5′→3′ exonuclease activity of Hex1 described within, in combination with the studies performed in yeast and *Drosophila* (11–15), suggests that the human protein could function in aspects of recombination and/or mismatch repair. At DNA double-strand breaks, Hex1 could produce the single-stranded 3′-overhangs needed for strand invasion and the exchange of genetic information during homologous recombination or for microidentity alignment during DNA end-joining or single strand annealing. While yeast genetics suggests that Mre11 is the primary nuclease in non-homologous and homologous recombination (33), studies performed with the yeast and *Drosophila* Hex1 equivalents indicate that these proteins contribute to both meiotic and mitotic recombination. Furthermore,

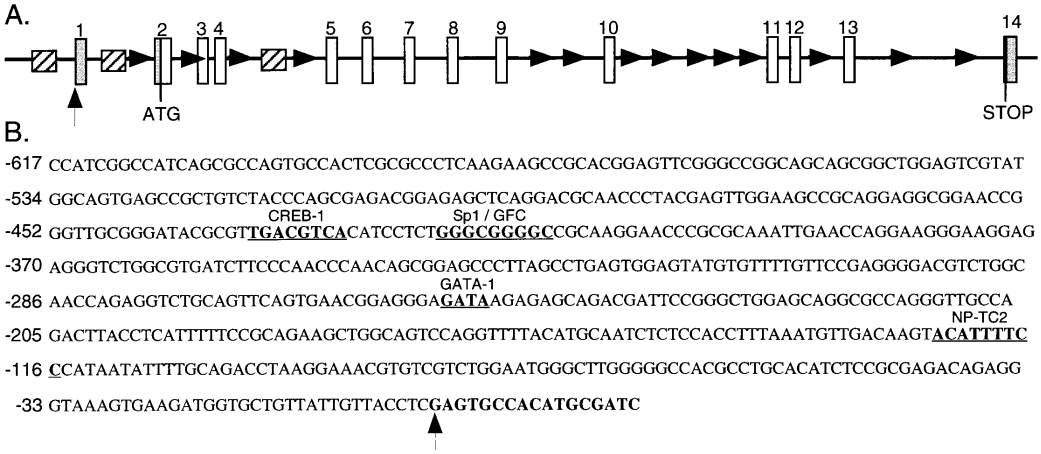

Figure 4. Genomic organization of *HEX1*. (**A**) A schematic depicting the relative positions of several *HEX1* genomic elements (drawing not to scale). Coding regions are represented by open boxes, with the exon number indicated above. The 5′- and 3′-UTRs are shown as shaded boxes. CpG islands are marked by boxes containing diagonal lines. Arrows display the location and direction of interspersed *Alu* repetitive sequences. The up arrow marks the 5′-end of the IMAGE *HEX1* cDNA clone 843301 in both (A) and (B). (**B**) The proximal promoter of *HEX1*. Consensus binding sequences for several transcription factors are indicated.

Mre11 was recently shown to operate as a 3′→5′ exonuclease (34), a polarity that is contrary to what has been predicted for this enzyme or for the facilitation of most recombination events.

Seeing that *HEX1* is expressed at low levels in all tissues examined, the encoded protein may in fact contribute globally to aspects of recombination and repair. However, the observation that *HEX1* mRNA expression is highest in fetal liver and adult bone marrow, areas of active hemopoiesis, suggests that Hex1 could act more prominently in DNA metabolic processes that occur during differentiation of the various stem cell lineages. It is tempting to speculate that the nuclease activity of Hex1 participates in V(D)J rearrangement and the generation of antibody diversity or also, potentially, in antibody class switching (35–37). Given the broad spectrum of DNA structures recognized by the Rad2 nuclease family, it will be of interest to see if Hex1 processess specific DNA conformations (such as hairpins) that are uniquely formed during V(D)J recombination, although incision of some hairpin loops has recently been appointed to Mre11 (34).

Of the putative promoter elements identified near the 5′-end of the sequenced cDNA clone (Fig. 5B), the NP-TCII consensus element is most intriguing. While this sequence motif is found upstream of several genes, the TC-II transcription factor is constitutively expressed only in lymphocytes and not in non-hemopoietic cell types (38) and, thus, it may be an important factor in determining the specificity of *HEX1* expression. The CpG islands located near exon 1 may also influence transcriptional regulation via their methylation status (39). The promoter elements that act coordinately to regulate *HEX1* transcription obviously need to be determined.

As for mismatch repair, Hex1 may operate with Msh2 to excise mismatched nucleotides that arise from misincorporations during DNA replication or from the inexact complementarity of DNA strands annealed during homologous recombination. Although it was reported that the yeast Exo1 protein interacts with human Msh2 (14), it is presently unknown whether there is a direct interaction between Hex1 and hMsh2.

Whereas the N-terminal region of Hex1 maintains several of the active site residues and the conserved N and I nuclease domains found in the Exo1 proteins (as well as other Rad2 proteins), the C-terminal region of Hex1, which is distinguished by its high serine content, displays no obvious homology to any protein in the database. This portion of the protein may determine the biological

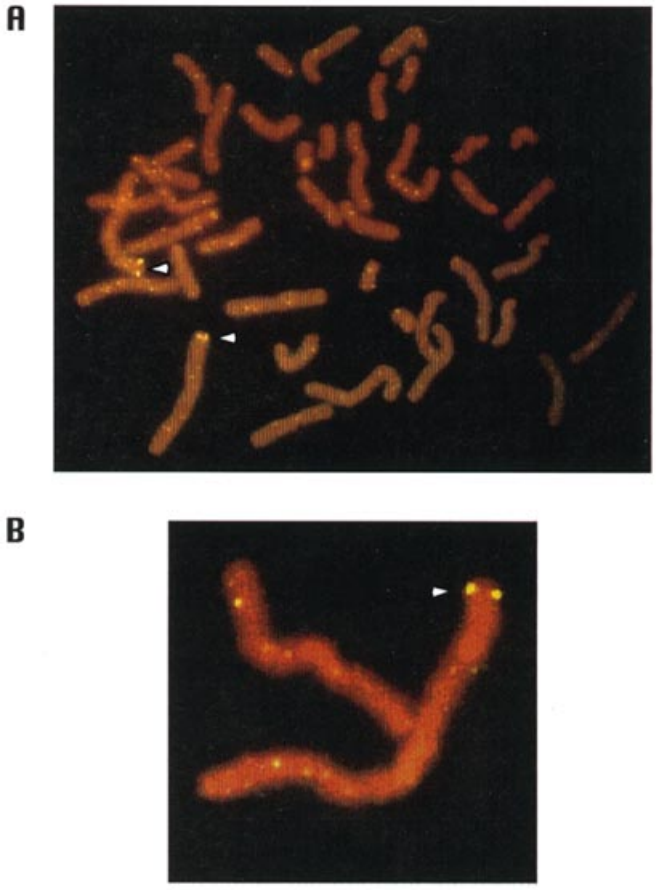

Figure 5. Chromosomal mapping of the *HEX1* gene. FISH was performed with BAC DNAs spanning the *HEX1* genomic region (see Materials and Methods). The arrow indicates the position of *HEX1* hybridization signal as detected by FITC staining. (**A**) FISH signal obtained on a complete mitotic chromosome spread. (**B**) Expanded view of the hybridization signal on chromosome 1, indicating localization near the telomere of the long arm. Images were processed without the use of digital enhancement.

specificity of Hex1 by directing interactions with components of the different DNA metabolic pathways. Furthermore, since *S.cerevisiae* possesses two additional *EXO1*-like genes, *DIN7* (40) and *YEN1* (41), it seems plausible that other human Exo1-like proteins may exist that, while maintaining the Rad2 nuclease domains, possess divergent C-terminal regions which impart distinct biological functions. However, searching the EST database has revealed no additional mammalian Rad2-like proteins. Moreover, Southern blot experiments performed at low stringency suggest that only a single *HEX1*-like gene exists in humans.

Hex1 and human cancers

Several acute leukemias and solid tumors have been found to undergo chromosomal rearrangements in the region of the *HEX1* gene. Specifically, balanced translocations involving 1q42 have been detected in two cases of leiomyoma of the uterus and four cases of non-Hodgkin's lymphoma (NHL) (42). Deletion of 1q42 has also been documented in multiple cases of acute lymphocytic leukemia, acute megakaryoblastic leukemia and NHL and loss of 1q42 has been observed alone or as part of more complex karyotypes in some breast and ovarian cancers (43) and other solid tumors (42). The association of genetic changes within 1q42 with multiple tumor types suggests that the loss of some gene(s) in this region is involved in tumor progression or selection, but is not required for tumor initiation, since the observed deletions generally occur as secondary aberrations (44).

Mutations in the *MSH2* mismatch repair gene have been causally associated with cases of HNPCC (45,46) and, more recently, precursor T cell lymphoblastoid malignancies (47). Given the hypothesis that Hex1 functions in mismatch repair or recombination with Msh2, it seems plausible that loss of Hex1 function contributes to tumorigenesis of hemopoietic lineages. The strong expression of *HEX1* in primary lymphoid tissues and the association of NHL with both deletions and balanced translocations in the 1q42 region suggest that *HEX1* may function as a tumor suppressor. It is noteworthy that the genetic composition of this region has not been extensively characterized at the molecular level and no specific genetic lesions have been identified in the reported neoplasms exhibiting loss of 1q42 (42). Molecular characterization of *HEX1* expression levels or the genomic content in neoplasms or tumor cell lines exhibiting loss of 1q42 will be important in establishing a relationship between the *HEX1* gene product and tumor progression. Additionally, the creation of mammalian cell lines with a deletion of this gene will contribute substantially to our understanding of the involvement of Hex1 in repair, recombination and protection against disease.

ACKNOWLEDGEMENTS

We thank C.Prange, J.Erzberger and Drs A.Feeney, B.Morgan, A.Islas and J.King for assistance and helpful comments and Dr J.Petrini for providing the hMRE11 cDNA. This work was done under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory under contract no. W-7405-ENG-48 and was supported by an LDRD (97-ERD-002) to D.M.W. J.P.C. was supported by an NIH Training grant CA09215-12.

REFERENCES

1 Lindahl,T., Karran,P. and Wood,R.D. (1997) *Curr. Opin. Genet. Dev.*, **7**, 158–169.
2 Fishel,R. and Wilson,T. (1997) *Curr. Opin. Genet. Dev.*, **7**, 105–113.
3 Arnheim,N. and Shibata,D. (1997) *Curr. Opin. Genet. Dev.*, **7**, 364–370.
4 Modrich,P. (1997) *J. Biol. Chem.*, **272**, 24727–24730.
5 Wood,R.D. (1997) *J. Biol. Chem.*, **272**, 23465–23468.
6 Wilson III,D.M. and Thompson,L.H. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 12754–12757.
7 Sun,H., Treco,D. and Szostak,J.W. (1991) *Cell*, **64**, 1155–1161.
8 Alani,E., Padmore,R. and Kleckner,N. (1990) *Cell*, **61**, 419–436.
9 Cao,L., Alani,E. and Kleckner,N. (1990) *Cell*, **61**, 1089–1101.
10 Henderson,G. and Simons,J.P. (1997) *Mol. Cell. Biol.*, **17**, 3779–3785.
11 Szankasi,P. and Smith,G.R. (1992) *J. Biol. Chem.*, **267**, 3014–3023.
12 Szankasi,P. and Smith,G.R. (1995) *Science*, **267**, 1166–1169.
13 Digilio,F.A., Pannuti,A., Lucchesi,J.C., Furia,M. and Polito,L.C. (1996) *Dev. Biol.*, **178**, 90–100.
14 Tishkoff,D.X., Boerger,A.L., Bertrand,P., Filosi,N., Gaida,G.M., Kane,M.F. and Kolodner,R.D. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 7487–7492.
15 Fiorentini,P., Huang,K.N., Tishkoff,D.X., Kolodner,R.D. and Symington,L.S. (1997) *Mol. Cell. Biol.*, **17**, 2764–2773.
16 Huang,K.N. and Symington,L.S. (1993) *Mol. Cell. Biol.*, **13**, 3125–3134.
17 Lieber,M.R. (1997) *BioEssays*, **19**, 233–240.
18 O'Donovan,A., Davies,A.A., Moggs,J.G., West,S.C. and Wood,R.D. (1994) *Nature*, **371**, 431–435.
19 Harrington,J.J. and Lieber,M.R. (1994) *EMBO J.*, **13**, 1235–1246.
20 Siegal,G., Turchi,J.J., Myers,T.W. and Bambara,R.A. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 9377–9381.
21 Waga,S. and Stillman,B. (1994) *Nature*, **369**, 207–212.
22 Mueser,T.C., Nossal,N.G. and Hyde,C.C. (1996) *Cell*, **85**, 1101–1112.
23 Ceska,T.A., Sayers,J.R., Stier,G. and Suck,D. (1996) *Nature*, **382**, 90–93.
24 Shen,B., Nolan,J.P., Sklar,L.A. and Park,M.S. (1996) *J. Biol. Chem.*, **271**, 9173–9176.
25 Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
26 Brookman,K.W., Lamerdin,J.E., Thelen,M.P., Hwang,M., Reardon,J.T., Sancar,A., Zhou,Z.Q., Walter,C.A., Parris,C.N. and Thompson,L.H. (1996) *Mol. Cell. Biol.*, **16**, 6553–6562.
27 Uberbacher,E.C. and Mural,R.J. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 11261–11265.
28 Heinemeyer,T., Wingender,E., Reuter,I., Hermjakob,H., Kel,A.E., Kel,O.V., Ignatieva,E.V., Ananko,E.A., Podkolodnaya,O.A., Kolpakov,F.A. *et al.* (1988) *Nucleic Acids Res.*, **26**, 364–370.
29 Quandt,K., Frech,K., Karas,H., Wingender,E. and Werner,T. (1995) *Nucleic Acids Res.*, **23**, 4878–4884.
30 Trask,B., Fertitta,A., Christensen,M., Youngblom,J., Bergmann,A., Copeland,A., de Jong,P., Mohrenweiser,H., Olsen,A., Carrano,A. *et al.* (1993) *Genomics*, **15**, 133–145.
31 Kozak,M. (1991) *J. Biol. Chem.*, **266**, 19867–19870.
32 Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.
33 Tsubouchi,H. and Ogawa,H. (1998) *Mol. Cell. Biol.*, **18**, 260–268.
34 Paull,T.T. and Gellert,M. (1998) *Mol. Cell*, **1**, 969–979.
35 Gellert,M. (1997) *Adv. Immunol.*, **64**, 39–64.
36 Stavnezer,J. (1996) *Curr. Opin. Immunol.*, **8**, 199–205.
37 Chu,G. (1997) *J. Biol. Chem.*, **272**, 24097–24100.
38 Lattion,A.L., Espel,E., Reichenbach,P., Fromental,C., Bucher,P., Israel,A., Baeuerle,P., Rice,N.R. and Nabholz,M. (1992) *Mol. Cell. Biol.*, **12**, 5217–5227.
39 Zingg,J.M. and Jones,P.A. (1997) *Carcinogenesis*, **18**, 869–882.
40 Mieczkowski,P.A., Fikus,M.U. and Ciesla,Z. (1997) *Mol. Gen. Genet.*, **253**, 655–665.
41 Smith,V., Chou,K.N., Lashkari,D., Botstein,D. and Brown,P.O. (1996) *Science*, **274**, 2069–2074.
42 Mitelman,F., Mertens,F. and Johansson,B. (1997) *Nature Genet.*, **15**, 417–474.
43 Heim,S., Teixeira,M.R., Dietrich,C.U. and Pandis,N. (1997) *Cancer Genet. Cytogenet.*, **95**, 16–19.
44 Johansson,B., Mertens,F. and Mitelman,F. (1995) *Blood*, **86**, 3905–3914.
45 Fishel,R., Lescoe,M.K., Rao,M.R., Copeland,N.G., Jenkins,N.A., Garber,J., Kane,M. and Kolodner,R. (1993) *Cell*, **75**, 1027–1038.
46 Liu,B., Parsons,R.E., Hamilton,S.R., Petersen,G.M., Lynch,H.T., Watson,P., Markowitz,S., Willson,J.K., Green,J., de la Chapelle,A. *et al.* (1994) *Cancer Res.*, **54**, 4590–4594.
47 Lowsky,R., DeCoteau,J.F., Reitmair,A.H., Ichinohasama,R., Dong,W.F., Xu,Y., Mak,T.W., Kadin,M.E. and Minden,M.D. (1997) *Blood*, **89**, 2276–2282.