GENOMICS **53,** 300–307 (1998)
ARTICLE NO. GE985497

Cloning, Expression, and Mapping of Ribonucleases H of Human and Mouse Related to Bacterial RNase HI

Susana M. Cerritelli and Robert J. Crouch[1]

Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892

Received June 8, 1998; accepted July 21, 1998

We identified two human sequences and one mouse sequence in the database of expressed sequence tags that are highly homologous to the N-terminal sequence of eukaryotic RNases H1. The cDNAs for human *RNASEH1* and mouse *Rnaseh1* were obtained, their nucleotide sequences determined, and the proteins expressed in *Escherichia coli* and partially purified. Both proteins have RNase H activity *in vitro* and they bind to dsRNA and RNA–DNA hybrids through the N-terminal conserved motif present in eukaryotic RNases H1. The *RNASEH1* gene is expressed in all human tissues at similar levels, indicating that RNase H1 may be a housekeeping protein. The human *RNASEH1* and mouse *Rnaseh1* cDNAs were used to isolate BAC genomic clones that were used as probes for fluorescence *in situ* hybridization. The human gene was localized to chromosome 17p11.2 and the mouse gene to a nonsyntenic region on chromosome 12A3. The chromosomal location and possible disease association of the human *RNASEH1* gene are discussed.

INTRODUCTION

Ribonucleases H (RNases H) are ubiquitously occurring enzymes that recognize an RNA–DNA duplex and cleave only the RNA strand. Even retroviruses, such as HIV, contain RNase H as a part of their reverse transcriptase to carry out synthesis of double-stranded DNA from the RNA genome (Crouch, 1990). It is clear that at least one type of cellular RNase H is similar in sequence and structure to the retroviral enzyme (Katayanagi *et al.*, 1992; Yang *et al.*, 1990; Davies *et al.*, 1991), and if the RNase H of HIV is to be a useful target for drug treatment, the properties of the human and related RNases H need to be determined. Ribonuclease H directs degradation of specific mRNAs, and this process is mediated by introduction of DNA oligonucleotides complementary to the target mRNA (Nellen and Lichtenstein, 1993). Not all cells respond equally to this antisense approach, and understanding the enzyme, its regulation, and its general properties could provide strategies to enhance the efficacy of drug design, delivery, and potency.

Although RNases H were first described in extracts from eukaryotic cells (Stein and Hausen, 1969), and have been studied extensively in a wide variety of organisms (Karwan and Wintersberger, 1988; Frank *et al.*, 1994; Cazenave *et al.*, 1994), only recently have the genes for several eukaryotic RNases H1 been cloned. In the absence of cloned genes, and sequence comparisons, eukaryotic RNases H were classified by their biochemical purification characteristics and properties into Class I and Class II (Büsen and Hausen, 1975). Two different RNases H have been cloned and characterized in *Escherichia coli:* RNase HI (Kanaya and Crouch, 1983) and RNase HII (Itaya, 1990). Retroviral RNases H are homologous in sequence and structure to *E. coli* RNase HI (Davies *et al.*, 1991). When the first RNase H clone was obtained from a eukaryote (Itaya *et al.*, 1991), it was found to be related to RNase HI of *E. coli;* hence the gene was called *RNH1.* Other eukaryotic homologs of the *Saccharomyces cerevisiae* RNase H1 have been found in *Schizosaccharomyces pombe* (AF048992), *Crithidia fasciculata* (Campbell and Ray, 1993), *Trypanosoma brucei* (Hesslein and Campbell, 1997), and *Drosophila melanogaster* (Filippov *et al.*, 1997), and because of sequence similarity they have been called RNase H1. Based on certain biochemical properties Campbell and Ray suggested that the enzyme from *C. fasciculata* is biochemically related to Class II RNase H (Campbell and Ray, 1993). A recent report (Frank *et al.*, 1998) suggests that the enzyme from *S. cerevisiae* with sequence homology to *E. coli* RNase HII is immunologically similar to the Class I RNase H. Thus, there is some confusion as to how to correlate the proteins that have been identified by sequence similarity to the proteins that have been identified by biochemical properties. Actually, limiting the number of RNases H to two groups may be incorrect. As the number of cloned RNase H genes increases, the

Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. AF048993 (mouse *Rnaseh1*) and AF048994 and AF048995 (human *RNASEH1*).

[1] To whom correspondence should be addressed. Fax: (301) 496-0243. E-mail: robert__crouch@nih.gov.

0888-7543/98 $25.00

classification into Class I and Class II may need to be expanded and redefined.

E. coli RNase HI is involved in DNA replication, preventing replication from occurring at sites other than *oriC*, the normal origin of replication (Crouch, 1990). The role of retroviral RNase H in viral replication is also well characterized, acting at several crucial stages of copying the single-stranded RNA into double-stranded DNA (Powell *et al.*, 1997). However, the function of the eukaryotic analogs is completely unknown. Understanding the function of eukaryotic RNases H1 is complicated by the fact that they contain an additional domain, not present in prokaryotes or retroviral RNases H, that *in vitro* binds to duplex RNAs, either double-stranded RNA or RNA–DNA hybrids (Cerritelli and Crouch, 1995; Cerritelli *et al.*, 1998). All known eukaryotic RNases H1 have an N-terminal region with one or two (only *S. cerevisiae*) copies of a conserved sequence of about 40 amino acids (Cerritelli *et al.*, 1998). A very similar amino acid sequence is present in caulimovirus ORF VI proteins (Mushegian *et al.*, 1994; Cerritelli *et al.*, 1998), where it plays a role in translation reinitiation (De Tapia *et al.*, 1993). The conserved motif from cauliflower mosaic virus, when in place of the *S. cerevisiae* sequence, binds well to dsRNA and RNA–DNA hybrids (Cerritelli *et al.*, 1998), indicating that both types of proteins may interact with similar nucleic acids *in vivo*.

We identified in the database of expressed sequence tags (dbEST) (Boguski *et al.*, 1993) two ESTs from human and one EST from mouse that were highly homologous to the N-terminal sequence of eukaryotic RNases H1. In this study, we obtained the cDNAs encoding human and mouse RNases H1, determined their nucleotide sequence, and expressed the proteins in *E. coli*. We characterized the RNase H and nucleic acid-binding activity of the recombinant proteins and determined the chromosomal location of the genes in mouse and human.[2]

MATERIALS AND METHODS

Cloning of Human and Mouse RNase H1 cDNAs

Similarity searches employing the BLASTN algorithm (Altschul *et al.*, 1990), using *S. cerevisiae* RNase H1 as the query sequence, detected two ESTs from Jurkat T-cells VI *Homo sapiens* cDNA (GenBank Accession Nos. AA312554 and AA356084) and one EST from mouse embryonic carcinoma (GenBank Accession No. AA163387). Clones for cDNAs of the three ESTs were obtained from ATCC. DNA sequence determination showed each clone to contain a complete cDNA encoding mouse or human RNase H1. The sequences have been deposited with GenBank (Accession Nos. for human AF048994 and AF048995 and for mouse AF048993).

For overexpression of the proteins in *E. coli*, the mouse and human cDNAs were cloned in the pET-15b expression vector (Novagen) by PCR amplification of the cDNA with primers that joined the vector cloning site *Nde*I with the second ATG of the cDNA coding sequence at the 5′ end of the gene and the TGA termination sequence with the *Bam*HI cloning site. After digestion with *Nde*I–*Bam*HI, the vector and the PCR products were ligated and transformed into *E. coli* competent cells. The constructs were confirmed by DNA sequence determination. The human cDNA gene was also cloned in the same way in the vector pET-3a (Novagen) using PCR amplification and *Nde*I–*Bam*HI cloning sites to produce a protein starting at the second ATG of the cDNA coding sequence, without an N-terminal His tag. The construct was confirmed by DNA sequence determination. Attempts to clone the cDNA for mouse and human starting at the first ATG were unsuccessful.

Expression and Purification of Recombinant Human and Mouse RNases H1

The human *RNASEH1* gene cloned in the pET-3a and pET-15b vectors (both from Novagen) was expressed in the *E. coli* strain BL21(DE3) harboring the plasmid pLysS (Studier *et al.*, 1990). The mouse *Rnaseh1* gene cloned in the pET-15b vector was expressed in the *E. coli* strain BL21(DE3). *E. coli* transformants were grown and induced to express recombinant RNases H1 as previously described (Cerritelli and Crouch, 1995). Crude cell extracts were prepared from *E. coli* cells expressing human RNase H1 in the pET-3a vector. After growth and induction of a 3-ml culture, 1 ml was pelleted and resuspended in 100 μl SDS cracking buffer (Cerritelli *et al.*, 1993). Human and mouse RNases H1 expressed in the pET-15b vector were purified using His-binding resin (Novagen). For purification, a 50-ml culture was grown, induced, and lysed, as previously described (Cerritelli *et al.*, 1998). After lysis, 1 μl each of soluble and insoluble fractions were separated in SDS–PAGE. Both proteins were partially insoluble. Purification proceeded from the soluble and the insoluble fraction, following the same protocol, with the only difference that purification from the insoluble fraction was performed in the presence of 6 M urea (Cerritelli *et al.*, 1998).

RNase H Activity after Electrophoresis in SDS–PAGE

Human and mouse His-tagged RNases H1 affinity purified by Ni^{2+}–NTA resin chromatography and crude protein extracts of cells induced to express untagged human RNase H1 were electrophoresed in 15% SDS–polyacrylamide gel containing 2–10 nmol (5×10^7 cpm) of [^{32}P]poly(rA)–poly(dT). The bands of RNase H activity were detected as previously described (Carl *et al.*, 1980) with the exception that the concentration of β-mercaptoethanol was increased to 10 mM.

Nucleic Acid Binding Assays

Northwestern assays using dsRNA and RNA–DNA were performed in the absence of $MgCl_2$ as previously described (Cerritelli and Crouch, 1995). After electrophoresis in SDS–PAGE, the proteins were transferred to Immobilon P membranes (Millipore) and allowed to renature. Renatured membrane-bound proteins were incubated with radiolabeled dsRNA and RNA–DNA hybrids of same size and sequence as previously described (Cerritelli *et al.*, 1998). Labeled dsRNA and RNA–DNA hybrids were generated as previously described (Cerritelli *et al.*, 1998) and purified in NucTrap push columns (Stratagene). [^{32}P]Poly(rA)–poly(dT) was prepared as described (Cerritelli and Crouch, 1995). After 4-h incubations, the unbound nucleic acid was washed off the membranes, and the filters were exposed to X-ray film overnight.

Hybridization Screening of Human and Mouse Bacterial Artificial Chromosome (BAC) Library

An *Eco*RI fragment containing the entire cDNA of human *RNASEH1* was used to screen a complete human BAC library (Genome Systems). The screening was performed by Genome Systems, Inc. Two clones were isolated, which were confirmed by Southern analysis to hybridize with human *RNASEH1* cDNA. A 530-bp *Sac*I fragment containing part of the coding region of mouse *Rnaseh1* cDNA was used to screen a complete mouse BAC library (Genome

[2] After submission of the manuscript, the paper by Wu *et al.* (1998) reporting the nucleotide sequence of a human RNase H cDNA, identical in sequence to that reported here, and some properties of the protein expressed in *E. coli* was brought to our attention.

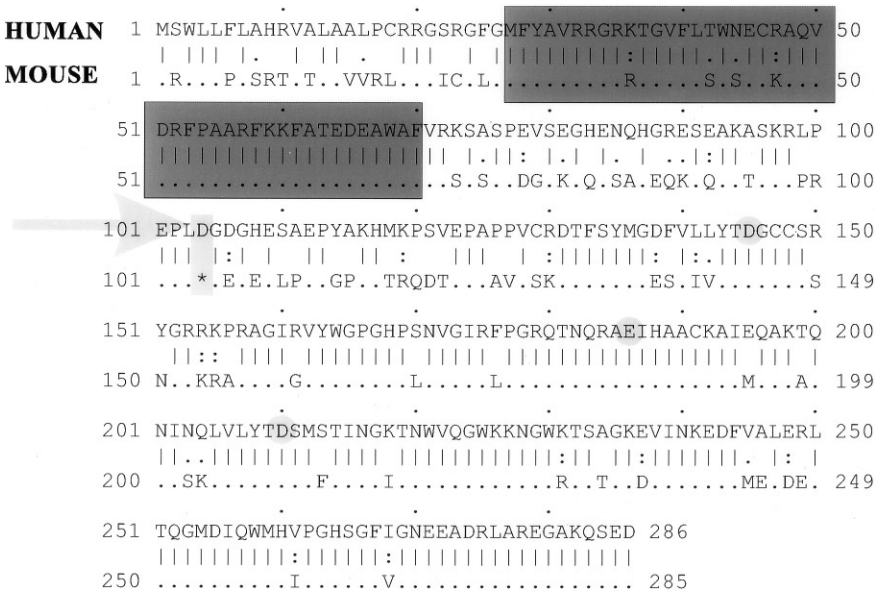


FIG. 1. Alignment of the amino acid sequences of human and mouse RNases H1 reveals 78% identity. The complete human sequence is shown, for mouse only amino acids that are different from the human sequence are shown, and identical residues are represented by dots. The nucleic acid-binding motif is included in the shaded box. The arrow points to the extra amino acid present in the human sequence. The conserved residues of the RNase H domain involved in catalysis are covered by a shaded circle. The alignment was generated using the BestFit program of the Wisconsin Sequence Analysis Package (Genetics Computer Group).

Systems). The screening was also performed by Genome Systems, Inc., and three clones were isolated, which were confirmed by Southern analysis to hybridize with mouse *Rnaseh1* cDNA.

Fluorescence in Situ Hybridization (FISH)

Human. DNA from a BAC clone obtained from the hybridization screen was labeled with digoxigenin–dUTP by nick-translation. The labeled probe was combined with sheared human DNA and hybridized to normal metaphase chromosomes derived from PHA-stimulated peripheral blood lymphocytes in a solution containing 50% formamide, 10% dextran sulfate, and 2× SSC. Specific hybridization signals were detected by incubating the hybridized slides in fluoresceinated anti-digoxigenin antibodies followed by counterstaining with DAPI.

Mouse. DNA from a BAC clone obtained from the hybridization screen was labeled with digoxigenin–dUTP by nick- translation. The labeled probe was combined with sheared mouse DNA and hybridized to normal metaphase chromosomes derived from mouse embryo fibroblast cells. FISH was performed as above.

RESULTS

Cloning of Human and Mouse cDNAs Encoding RNases H1

Two ESTs (AA312554 and AA356084) derived from Jurkat T-cells VI of *H. sapiens* cDNA were found to contain a sequence similar to the N-terminal region of eukaryotic RNases H1 (Cerritelli *et al.*, 1998). The complete DNA sequence of both cDNAs revealed the presence of the 3′ end of the human *RNASEH1* cDNA, including a poly(A) tail. Both cDNAs were identical with the exception of an extra 34 nucleotides in the 5′ end of AA312554. The human *RNASEH1* cDNA sequence was determined to be 1088 nucleotides in length and contained an open reading frame of 858 nucleotides with the first AUG at nucleotides 28–30. A second AUG in the same frame at nucleotide 79–81 potentially may serve as a translation initiation codon (Kozak, 1987). The protein, starting at the first AUG, was predicted to consist of 286 amino acids, with a calculated molecular mass of 32,245 Da and an isoelectric point of 9.98. The protein starting at the second AUG would contain 260 residues, with a calculated molecular mass of 29,195 Da and an isoelectric point of 8.85 (Wisconsin Sequence Analysis Package, Genetics Computer Group).

An EST (AA163387) from mouse embryonic carcinoma was also found to have sequence homology to the N-terminal region of eukaryotic RNases H1. The complete cDNA sequence includes a poly(A) tail, indicating that it contains the 3′ end of the mouse RNase H1 gene. The mouse *Rnaseh1* cDNA was determined to be 1407 nucleotides in length with an open reading frame of 855 nucleotides. There is a 3′ untranslated sequence of 510 nucleotides, 302 nucleotides longer than the human 3′UTR. As with the human *RNASEH1* cDNA, two in-frame AUGs, the first at nucleotide 40–42 and the second at nucleotide 119–121, could act as a translation start site. The protein starting at the first AUG would contain 285 amino acids, with a calculated molecular mass of 31,804 Da and an isoelectric point of 10.01. The protein starting at the second AUG would have 259 residues, with a calculated molecular mass of 28,885 and an isoelectric point of 8.85 (Wisconsin Sequence Analysis Package, Genetics Computer Group).

RNase H1 amino acid sequences from mouse and human were 78% identical (Fig. 1). Both proteins have an RNase H motif at the C-terminal end of the protein, with all the conserved amino acids that constitute the hallmark of the RNase H1 family of proteins (Crouch, 1990), and, like all the other eukaryotic RNases H1 described so far (Cerritelli *et al.*, 1998; Filippov *et al.*,

1997; Hesslein and Campbell, 1997; Campbell and Ray, 1993; Wu *et al.*, 1998), mouse and human RNases H1 have a conserved motif at the N-terminal region of the protein that in other proteins has been implicated in dsRNA and RNA–DNA hybrid binding (Cerritelli *et al.*, 1998). Human and mouse RNase H1 proteins are remarkably similar; perhaps the main difference between the two proteins is the presence of an extra residue (D104) in the connecting region between the N-terminal conserved domain and the RNase H domain of the human RNase H1.

Expression of Recombinant Mouse and Human RNases H1 in E. coli

To study the functional properties of mouse and human RNases H1, we cloned them into the pET-15b expression plasmid (Novagen) to express a fusion protein with a poly-His tag at the N-terminus followed by the mouse RNase H1. We were able to obtain only clones in which the His tag was fused to the gene at the second AUG codon, generating a protein not containing the first 26 amino acids of the open reading frame shown in Fig. 1. The cDNA clone of human *RNASEH1* was also subcloned into the pET-15b, and as for the mouse, we obtained clones expressing a fusion protein with a poly-His tag attached at the N-terminus of human RNase H1 starting at the second Met (M-27 in Fig. 1). However, the human protein could not be expressed in the host *E. coli* strain BL21(DE3), the strain we used for expression of the mouse and other eukaryotic RNases H1. The presence of the pLysS plasmid, which increases the tolerance of BL21(DE3) for plasmids with toxic inserts (Studier *et al.*, 1990), allowed the cloning and expression of human RNase H1, suggesting that the protein may be deleterious for growth in *E. coli.*

Following induction by IPTG, the fused proteins were partially insoluble in *E. coli* cell extracts. We recovered the insoluble fraction in 6 M urea and purified the soluble and the insoluble fractions using Ni^{2+}–NTA resin (Fig. 2A). The purified proteins were used to characterize the functional properties of human and mouse RNases H1. For eliminating the possible influence of the His tag in our assays, we also cloned the human RNase H1 without the His tag in the vector pET-3a (Novagen) starting at the second Met, and as before the protein was expressed in BL21(DE3) cells containing the pLysS plasmid.

A renaturation gel assay was performed with partially purified human and mouse His-tagged RNase H1 and with total crude extract of cells induced to express human RNase H1 without a His tag. This permits specific detection of RNase H proteins in an SDS–polyacrylamide gel that contains a radiolabeled RNA–DNA hybrid. Bands of RNase H1 activity were observed for mouse and human RNase H1, with and without His tag (data not shown), indicating that the proteins expressed in *E. coli* are fully active, without requiring modifications or additional proteins. The same has been observed for *D. melanogaster* RNase H1 (Filippov *et al.*, 1997).

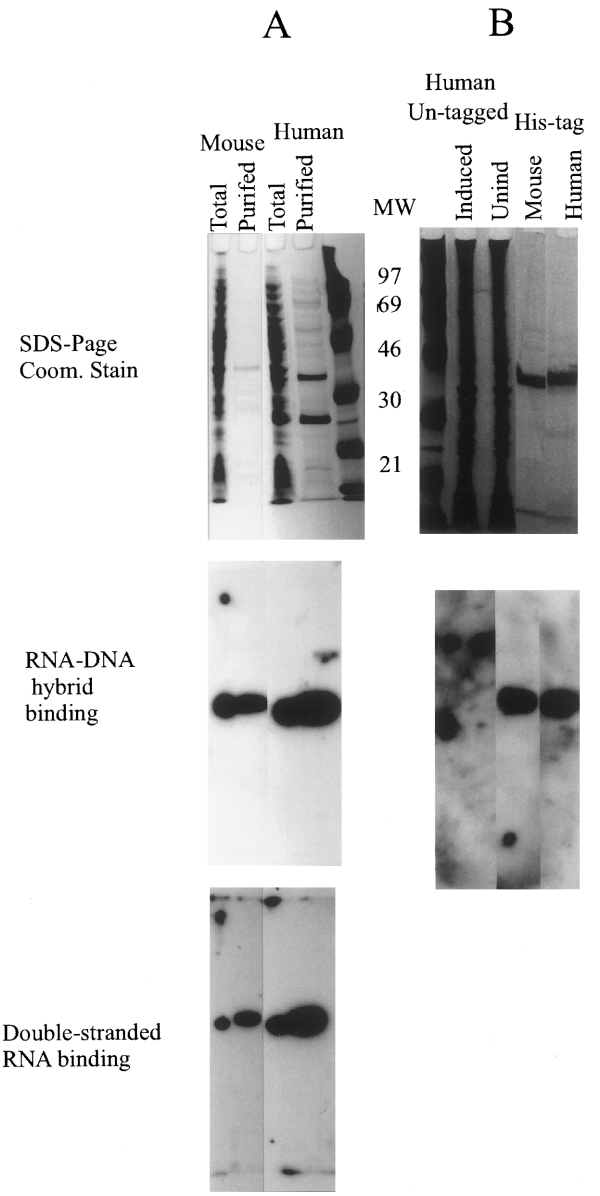


FIG. 2. Nucleic acid binding activity of human and mouse RNase H1. (**A**) Proteins from total crude extract of *E. coli* cells expressing human and mouse His-tagged RNase H1 and partially purified proteins that were electrophoresed in a polyacrylamide gel and stained with Coomassie blue (Coom. Stain) or transferred to an Immobilon P membrane, renatured, and incubated with a radiolabeled RNA–DNA hybrid (middle) or dsRNA (bottom) of identical size (100 bp) and sequence. (**B**) Proteins from total crude extract of *E. coli* cells expressing human untagged RNase H1 or partially purified His-tagged mouse and human RNase H1 that were electrophoresed in a polyacrylamide gel and stained with Coomassie blue (Coom. Stain) or transferred to an Immobilon P membrane, renatured, and incubated with a radiolabeled poly(rA)–poly(dT). Prestained molecular weight markers (Amersham) are shown (MW), and their masses in kDa are indicated.

Yeast RNase H1 has been shown to possess nucleic acid binding activity (predominantly dsRNA and RNA–DNA hybrids) through an N-terminal region of the protein, a conserved motif present in all known eukaryotic RNases H and also in caulimovirus ORF VI

proteins (Cerritelli *et al.*, 1998). Because mouse and human RNases H1 also have the N-terminal conserved motif, we tested the binding activity of proteins partially purified or present in total *E. coli* extracts expressing mouse and human RNase H1 with and without an N-terminal His tag. We used Northwestern assays, in which membrane-bound proteins were incubated with radiolabaled dsRNA and RNA–DNA hybrids of equal size and composition. The fusion proteins, partially purified or in crude extracts, bound both substrates very well (Fig. 2A), indicating that nucleic acid binding is also a property of higher eukaryotic RNases H1. *E. coli* cell lysates from cultures induced to express the untagged human RNase H1 were tested for binding to RNA–DNA hybrid. Human RNase H1, and another *E. coli* protein previously described (Cerritelli and Crouch, 1995), bound well to the substrate (Fig. 2B). The untagged protein started at the second Met of the coding sequence, immediately prior to the first amino acid of the conserved nucleic acid binding domain, indicating that the first 26 amino acids are not required for nucleic acid binding.

Hybridization Screening of a Human BAC Library

Two BAC clones were obtained by hybridization with a fragment of the coding region of the cDNA clone. The BAC plasmids were confirmed to contain the genomic sequence of the *RNASEH1* gene by partial DNA sequence determination and Southern and PCR analyses (data not shown).

Hybridization Screening of a Mouse BAC Library

Three BAC clones were obtained by hybridization with an internal fragment of the cDNA clone. All three BAC plasmids contained mouse *Rnaseh1* genomic sequences as confirmed by partial DNA sequence determination and Southern and PCR analyses (data not shown).

Fluorescence in Situ Hybridization

Human. Among 80 metaphase chromosomes analyzed, 71 exhibited specific signals on one pair of chromosomes. DAPI banding (Heng and Tsui, 1993) resulted in specific labeling of the short arm of a group E chromosome that was believed to be chromosome 17 on the basis of size, morphology, and banding pattern (Fig. 3A). A genomic probe that has been previously mapped to 17q24 was cohybridized with the *RNASEH1* probe and resulted in the specific labeling of the long and the short arms of chromosome 17 (Fig. 3B). Measurements of 10 specifically labeled chromosomes 17 demonstrated that the *RNASEH1* gene is located at a position that is 25% of the distance from the centromere to the telomere of chromosome 17p, an area that corresponds to band 17p11.2 (Fig. 5).

Mouse. A total of 80 metaphase chromosomes were analyzed with 72 exhibiting specific labeling. The initial experiment resulted in specific labeling of the proximal portion of a medium-sized chromosome that was believed to be chromosome 12 on the basis of banding pattern (Fig. 4A). A second experiment was conducted using a probe that is specific for the telomeric region of chromosome 12 and that was cohybridized with the *Rnaseh1* probe. This experiment resulted in the specific labeling of the telomeric region and the proximal portion of chromosome 12 (Fig. 4B). Measurement of 10 specifically hybridized chromosomes 12 demonstrated that the *Rnaseh1* gene is located at a position that is 17% of the distance from the heterochromatic–euchromatic boundary to the telomere of chromosome 12, an area that corresponds to band 12A3 (Fig. 5).

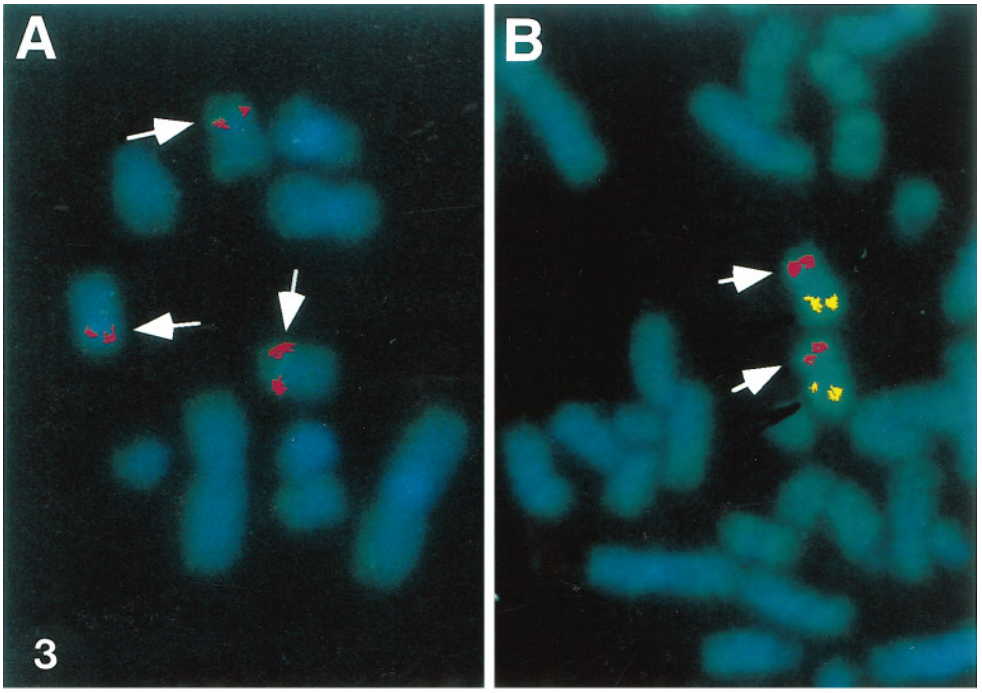
Abundance and Tissue Specificity of Human RNASEH1 mRNA

The abundance and tissue specificity of human *RNASEH1* mRNA was determined by PCR using Multiple Tissue cDNA Panels (Clontech) and primers specific for the *RNASEH1* gene. We observed PCR bands in all tissues analyzed with 30 cycles being required to visualize the PCR product of the *RNASEH1*, indicating a medium to low-abundance mRNA (data not shown). We quantified the intensity of each DNA band and repeated the experiment several times, obtaining band intensities with no more than a threefold difference, within the experimental error reported in the MTC Panels manual (Clontech). This indicated that the mRNA of the *RNASEH1* gene was expressed at a similar level in all tissues, suggesting that RNase H1 is a housekeeping protein. The results are consistent with those of Wu *et al.* (1998).

DISCUSSION

Human and mouse RNases H1 share a high degree of homology and are also very similar to other known eukaryotic RNases H1 (Fig. 1). They have, in addition to the RNase H domain, an N-terminal region with a conserved motif that *in vitro* binds to dsRNA and RNA–DNA hybrids (Fig. 2). The nucleic acid-binding domain is approximately 40 amino acids in length and is highly similar to a region of caulimovirus ORF VI family of proteins. This family of proteins is very different and is involved in translation reinitation (De Tapia *et al.*, 1993), possibly by interaction with ribosomes or mRNA through the dsRNA-binding motif. By analogy, the nucleic acid-binding domain of RNase H1 may also interact with ribosomal RNA, perhaps in the nucleus during the synthesis/processing of ribosomal RNA.

There is small, but significant sequence similarity between the nucleic acid-binding motif of eukaryotic RNases H1 (Cerritelli and Crouch, 1995; Cerritelli *et al.*, 1998) and a larger, less conserved dsRNA-binding domain (Green and Mathews, 1992; St Johnston *et al.*, 1992). The dsRNA-binding motif is about 70 amino acids long, while the RNase H1 conserved domain is only around 40 residues. The extra 30 amino acids at



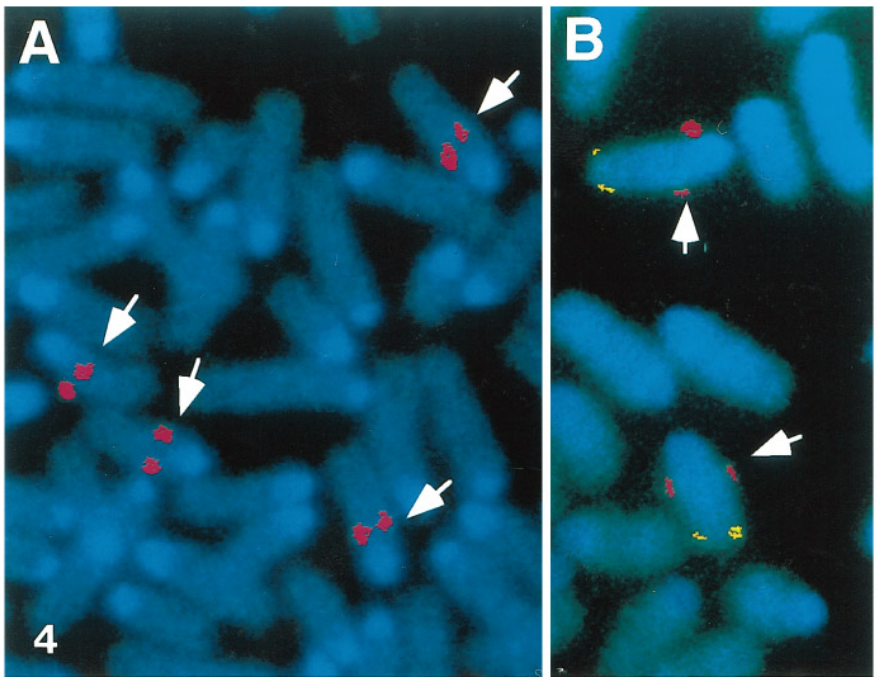


FIG. 3. FISH of the human *RNASEH1* BAC to metaphase chromosomes. (**A**) The location of the hybridization signal from the BAC *RNASEH1* labeled probe. (**B**) The cohybridization of the BAC *RNASEH1* labeled probe (in red and marked by arrows) and a genomic probe that has been previously mapped to 17q24 (in yellow). Measurement of 10 specifically labeled chromosomes 17 demonstrates that *RNASEH1* is located in band 17p11.2. Photographs represent computer-enhanced images.

FIG. 4. FISH of the mouse *Rnaseh1* BAC to metaphase chromosomes. (**A**) The location of the hybridization signal from the BAC *Rnaseh1* labeled probe. (**B**) The cohybridization of the BAC *Rnaseh1* labeled probe (in red and marked by arrows) and a genomic probe specific for the telomeric region of chromosome 12 (in yellow). Measurement of 10 specifically labeled chromosomes 12 demonstrates that *Rnaseh1* is located in an area that corresponds to band 12A3. Photographs represent computer-enhanced images.

the N-terminus of the dsRNA-binding domain constitute α-1, a part of the structure that is directly involved in contacting the dsRNA substrate (S. Schultz, Boulder, CO, pers. comm., Nov. 1997). Human *RNASEH1* and mouse *Rnaseh1* mRNAs have two in-frame AUGs that could be used as translation initiation signals. The nucleic acid-binding motif starts immediately after the second AUG codon. Proteins starting at the first AUG would have an additional 26-amino acid sequence that is absent from some eukaryotic RNases H1. Interestingly, we were unable to obtain clones for which the proteins would start at the

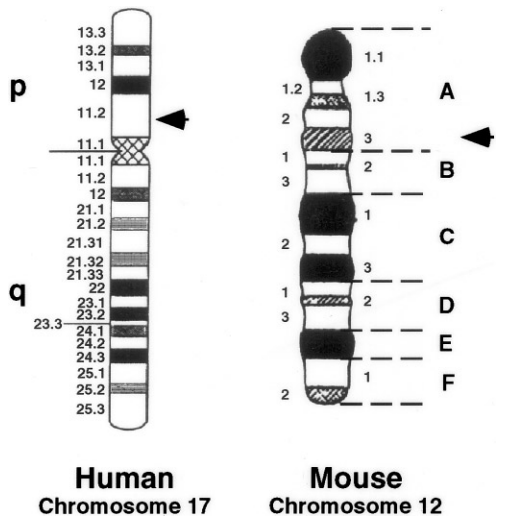


FIG. 5. Idiograms of the results showed in Figs. 3 and 4. Based on 10 metaphase spreads, the human *RNASEH1* gene was shown to localize to chromosome 17p11.2, indicated by an arrow. Based on 10 metaphase spreads, the mouse *Rnaseh1* gene was shown to localize to a nonsyntenic region on chromosome 12A3, indicated by an arrow. The images are not to scale.

first AUG in the *E. coli* expression system used to overproduce RNase H1, whereas Wu *et al.* (1998) report no such problems. The plasmid pET15b used as expression vector in most of our experiments includes an N-terminal His tag absent in pET17 employed by Wu *et al.* (1998). We made numerous attempts to use other pET vectors without obtaining clones starting at the first ATG. The *D. melanogaster* proteins also start at the second ATG of the cDNA (Filippov *et al.*, 1997). The 26 amino acids at the N-termini of human and mouse RNase H1 may be playing a role similar to that of the α-1 region of the dsRNA-binding motif, interacting with the substrate and enhancing the affinity of the protein for nucleic acids, making it a stronger binder than proteins lacking this region and potentially more deleterious for *E. coli* growth.

The connecting region between the N-terminal nucleic acid-binding domain and the C-terminal RNase H region is variable in length and sequence among eukaryotic RNase H1. The only common feature is the presence of a large number of S, T, and P residues. The connecting region is important for binding to nucleic acids (Cerritelli *et al.*, 1998) and may play a role in switching between the two activities of the protein. The human and mouse RNase H1 amino acid sequences are 78% identical (Fig. 1), with the connecting domain being the most dissimilar region. One significant difference is the presence of an extra residue (D104) in the human protein (Fig. 1). Although both proteins are remarkably similar, in our experiments, the human RNase H1 seems to be more harmful for *E. coli*, requiring pLysS, a plasmid that increases the tolerance for toxic inserts (Studier *et al.*, 1990). The presence of the extra D residue may affect the binding ability of the human protein, switching the equilibrium between the two activities toward a strong binding and a lower RNase H activity that could be potentially deleterious for *E. coli*. We have been unable to obtain catalytic-site mutants in the RNase H domain of *S. cerevisiae* RNase H1 (unpublished), suggesting that altering the equilibrium between nucleic acid binding and RNase H activity is harmful for *E. coli*.

The mouse *Rnaseh1* cDNA has a long 3′ untranslated region. The presence of this sequence is not a cloning artifact, because PCR amplification of the BAC plasmids containing the mouse genomic DNA with primers specific for the 3′UTR region present in the cDNA demonstrated that the 3′UTR sequence is also present in the chromosomal gene (data not shown). Also, a second EST sequence of mouse (AA217317) has the same 3′ sequence as the clone we report. Perhaps, the long mouse 3′UTR contains regulatory elements that modulate the expression or localization of the RNase H1 gene. In contrast, the human *RNASEH1* cDNA we describe has a much shorter 3′UTR. Interestingly, Wu *et al.* (1998) report one human *RNASEH1* cDNA clone that is 530 nucleotides longer than another; most of the size difference being attributed to the 3′UTR. At present we are in the process of determining the sequence of and characterizing human and mouse genomic DNAs corresponding to the cDNAs.

The *RNASEH1* gene is located on human chromosome 17p11.2 and mouse chromosome 12A3 (Fig. 5), defining a new region of synteny. Another gene that is found on mouse chromosome 12 and human chromosome 17p is the Ryk-2/RYK-2 receptor-like tyrosine kinase-like 1 (Gough *et al.*, 1995). The juxtacentromeric region of the human chromosome 17 short arm (17p11.2–p12), the site of the *RNASEH1* gene, is a very interesting and active chromosomal location. It contains genes involved in the Charcot–Marie–Tooth type 1A disease (CMT1A) and the Smith–Magenis syndrome (SMS) (Chevillard *et al.*, 1993). CMT1A is associated with a duplication of a short segment, whereas SMS is linked to microdeletions, extending toward the centromere. A repeated-gene cluster has been mapped to the ends and the middle of the common deletion interval found in SMS patients, suggesting that recombination between flanking repeated-gene clusters is a frequent cause of the SMS common deletion (Chen *et al.*, 1997).

There are some other DNA markers with several low-copy-number repeats localized in this region, including the human genes encoding *U3* snRNA (Chen *et al.*, 1997). The human *U3* genes are also clustered on chromosome 17p11.2, with evidence for large inverted duplications within the cluster (Gao *et al.*, 1997). The gene duplications within the *RNU3* locus appear to be of recent evolutionary origin, indicating that this region of chromosome 17 is highly recombinagenic, as is also evident in the SMS and CMT1A syndromes. The presence of tandem repetition in the SMS region and the CMT1A segment and inverted repeat in the *RNU3* locus can facilitate homologous recombination and DNA rearrangement on this region of the human chromosome 17p, which could have led to the insertion of the *RNASEH1* gene at this location. Current experi-

ments are under way to determine if the human *RNASEH1* gene falls within the SMS deletion and if the deletion of *RNASEH1* contributes to the SMS phenotype.

ACKNOWLEDGMENTS

We thank Dr. Steve Schultz for communicating results prior to publication. This work was supported in part by the National Institutes of Health Intramural AIDS Targeted Antiviral Program.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Boguski, M. S., Lowe, T. M., and Tolstoshev, C. M. (1993). dbEST—database for "expressed sequence tags." *Nat. Genet.* **4:** 332–333.

Büsen, W., and Hausen, P. (1975). Distinct ribonuclease H activities in calf thymus. *Eur. J. Biochem.* **52:** 179–190.

Campbell, A. G., and Ray, D. S. (1993). Functional complementation of an *Escherichia coli* ribonuclease H mutation by a cloned genomic fragment from the trypanosomatid *Crithidia fasciculata*. *Proc. Natl. Acad. Sci. USA* **90:** 9350–9354.

Carl, P. L., Bloom, L., and Crouch, R. J. (1980). Isolation and mapping of a mutation in *Escherichia coli* with altered levels of ribonuclease H. *J. Bacteriol.* **144:** 28–35.

Cazenave, C., Frank, P., Toulme, J. J., and Büsen, W. (1994). Characterization and subcellular localization of ribonuclease H activities from *Xenopus laevis* oocytes. *J. Biol. Chem.* **269:** 25185–25192.

Cerritelli, S. M., and Crouch, R. J. (1995). The non-RNase H domain of *Saccharomyces cerevisiae* binds double-stranded RNA: Magnesium modulates the switch between double-stranded RNA binding and RNase H activity. *RNA* **1:** 246–259.

Cerritelli, S. M., Fedoroff, O. Y., Reid, B. R., and Crouch, R. J. (1998). A common 40 amino acid motif in eukaryotic RNases H1 and caulimovirus ORF VI proteins binds to duplex RNAs. *Nucleic Acids Res.* **26:** 1834–1840.

Cerritelli, S. M., Shin, D. Y., Chen, H. C., Gonzales, M., and Crouch, R. J. (1993). Proteolysis of *Saccharomyces cerevisiae* RNase H1 in *E. coli*. *Biochimie* **75:** 107–111.

Chen, K. S., Manian, P., Koeuth, T., Potocki, L., Zhao, Q., Chinault, A. C., Lee, C. C., and Lupski, J. R. (1997). Homologous recombination of a flanking repeat gene cluster is a mechanism for a common contiguous gene deletion syndrome. *Nat. Genet.* **17:** 154–163.

Chevillard, C., Le Paslier, D., Passage, E., Ougen, P., Billaut, A., Boyer, S., Mazan, S., Bachellerie, J. P., Vignal, A., Cohen, D., *et al.* (1993). Relationship between Charcot–Marie–Tooth 1A and Smith–Magenis regions. snU3 may be a candidate gene for the Smith–Magenis syndrome. *Hum. Mol. Genet.* **2:** 1235–1243.

Crouch, R. J. (1990). Ribonuclease H: From discovery to 3D structure. *New Biol.* **2:** 771–777.

Davies, J. F., II, Hostomska, Z., Hostomsky, Z., Jordan, S. R., and Matthews, D. (1991). Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science* **252:** 88–95.

De Tapia, M., Himmelbach, A., and Hohn, T. (1993). Molecular dissection of the cauliflower mosaic virus translation transactivator. *EMBO J.* **12:** 3305–3314.

Filippov, V., Filippova, M., and Gill, S. S. (1997). Functional characterization of RNase H1 from *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* **240:** 844–849.

Frank, P., Albert, S., Cazenave, C., and Toulme, J. J. (1994). Purification and characterization of human ribonuclease HII. *Nucleic Acid Res.* **22:** 5247–5254.

Frank, P., Braunshofer-Reiter, C., and Wintersberger, U. (1998). Yeast RNase H(35) is the counterpart of the mammalian RNase H1, and is evolutionarily related to prokaryotic RNase HII. *FEBS Lett.* **421:** 23–26.

Gao, L., Frey, M. R., and Matera, A. G. (1997). Human genes encoding U3 snRNA associate with coiled bodies in interphase cells and are clustered on chromosome 17p11.2 in a complex inverted repeat structure. *Nucleic Acids Res.* **25:** 4740–4747.

Gough, N. M., Rakar, S., Hovens, C. M., and Wilks, A. (1995). Localization of two mouse genes encoding the protein tyrosine kinase receptor-related protein RYK. *Mamm. Genome* **6:** 255–256.

Green, S. R., and Mathews, M. B. (1992). Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI. *Genes Dev.* **6:** 2478–2490.

Heng, H. H., and Tsui, L. C. (1993). Modes of DAPI banding and simultaneous *in-situ* hybridization. *Chromosoma* **102:** 325–332.

Hesslein, D. G., and Campbell, A. G. (1997). Molecular cloning and expression of a ribonuclease H from the kinetoplastid, *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* **86:** 121–126.

Itaya, M. (1990). Isolation and characterization of a second RNase H (RNase HII) of *Escherichia coli* K-12 encoded by the *rnhB* gene. *Proc. Natl. Acad. Sci. USA* **87:** 8587–8591.

Itaya, M., McKelvin, D., Chatterjie, S. K., and Crouch, R. J. (1991). Selective cloning of genes encoding RNase H from *Salmonella typhimurium*, *Saccharomyces cerevisiae* and *Escherichia coli rnh* mutant. *Mol. Gen. Genet.* **227:** 438–445.

Kanaya, S., and Crouch, R. J. (1983). DNA sequence of the gene coding for *Escherichia coli* RNase H. *J. Biol. Chem.* **258:** 1276–1281.

Karwan, R., and Wintersberger, U. (1988). In addition to RNase H(70) two other proteins of *Saccharomyces cerevisiae* exhibit ribonuclease H activity. *J. Biol. Chem.* **263:** 14970–14977.

Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T., and Morikawa, K. (1992). Structural details of ribonuclease H from *Escherichia coli* as refined to an atomic resolution. *J. Mol. Biol.* **223:** 1029–1052.

Kozak, M. (1987). An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15:** 8125–8148.

Mushegian, A. M., Edskes, H. K., and Koonin, E. V. (1994). Eukaryotic RNase H shares a conserved domain with caulimovirus proteins that facilitate translation of polycistronic RNA. *Nucleic Acids Res.* **22:** 4163–4166.

Nellen, W., and Lichtenstein, C. (1993). What makes an mRNA anti-sense-itive? *Trends Biochem. Sci.* **18:** 419–423.

Powell, M. D., Ghosh, M., Jacques, P. S., Howard, K. J., Le Grice, S. F., and Levin, J. G. (1997). Alanine-scanning mutations in the "primer grip" of p66 HIV-1 reverse transcriptase resultin selective loss of RNA priming activity. *J. Biol. Chem.* **272:** 13262–13269.

Stein, H., and Hausen, P. (1969). Enzyme from calf thymus degrading the RNA moiety of DNA–RNA hybrids: Effect on DNA dependent RNA polymerase. *Science* **166:** 393–395.

St Johnston, D., Brown, N. H., Gall, J. G., and Jantsch, M. (1992). A conserved double-stranded RNA-binding domain. *Proc. Natl. Acad. Sci. USA* **89:** 10979–10983.

Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct the expression of cloned genes. *Methods Enzymol.* **185:** 60–89.

Wu, H. J., Lima, W. F., and Crooke, S. T. (1998). Molecular cloning and expression of cDNA for human RNase H. *Antisense Nucleic Acid Drug Dev.* **8:** 53–61.

Yang, W., Hendrickson, W. A., Crouch, R. J., and Satow, Y. (1990). Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. *Science* **249:** 1398–1405.