Multiple Ribonuclease H–Encoding Genes in the Caenorhabditis elegans Genome Contrasts with the Two Typical Ribonuclease H–Encoding Genes in the Human Genome

Arulvathani Arudchandran,* Susana M. Cerritelli,* Nathan J. Bowen,† Xiongfong Chen,‡ Michael W. Krause,§ and Robert J. Crouch*

*Laboratory of Molecular Genetics, †Laboratory of Gene Regulation and Development, ‡Unit on Biologic Computation, National Institute of Child Health and Human Development, and §Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland

Database searches of the Caenorhabditis elegans and human genomic DNA sequences revealed genes encoding ribonuclease H1 (RNase H1) and RNase H2 in each genome. The human genome contains a single copy of each gene, whereas C. elegans has four genes encoding RNase H1–related proteins and one gene for RNase H2. By analyzing the mRNAs produced from the C. elegans genes, examining the amino acid sequence of the predicted protein, and expressing the proteins in Escherichia coli we have identified two active RNase H1–like proteins. One is similar to other eukaryotic RNases H1, whereas the second RNase H (rnh-1.1) is unique. The rnh-1.0 gene is transcribed as a dicistronic message with three dsRNA-binding domains; the mature mRNA is transspliced with SL2 splice leader and contains only one dsRNA-binding domain. Formation of RNase H1 is further regulated by differential cis-splicing events. A single rnh-2 gene, encoding a protein similar to several other eukaryotic RNase H2's, also has been examined. The diversity and enzymatic properties of RNase H homologues are other examples of expansion of protein families in C. elegans. The presence of two RNases H1 in C. elegans suggests that two enzymes are required in this rather simple organism to perform the functions that are accomplished by a single enzyme in more complex organisms. Phylogenetic analysis indicates that the active C. elegans RNases H1 are distantly related to one another and that the C. elegans RNase H1 is more closely related to the human RNase H1. The database searches also suggest that RNase H domains of human endogenous retroviruses are more closely related to cellular RNases H1, but numerous RNase H domains of human endogenous retroviruses are more closely related to cellular RNases H.

Introduction

Data generated from the various genome-sequencing projects provide useful information on diversity in structure and functions of genes and proteins, providing us with a foothold to gain an understanding about related proteins from various organisms. Completion of the genomic sequences of Caenorhabditis elegans (The C. elegans Sequencing Consortium 1998) and humans (Lander et al. 2001; Venter et al. 2001) directed our interest toward the characterization of ribonuclease H (RNase H) homologues of these multicellular animals.

The RNases H in a variety of organisms have been studied extensively with respect to structure, function, and enzymic properties on the basis of their specific degradation of the RNA in RNA-DNA hybrids (Crouch and Toulmé 1998, pp. 1–265). The RNases H participate in cellular processes, such as DNA replication, repair, and transcription, as well as in the replication of retroviral genomes. Escherichia coli has two RNases H (HI and HII), each having its own characteristic amino acid sequence (Ohtani et al. 1999a, 1999b). Bacillus subtilis also has two active RNases H, both related by sequence to E. coli RNase HII (Itaya et al. 1999; Ohtani et al. 1999a, 1999b). Despite their similarity at the amino acid sequence level, these two B. subtilis proteins have very different specific activities, specificities of cleavage sites, and strikingly different divalent metal ion preferences and, therefore, have been classified as RNases HII and HIII. Bacillus subtilis has a gene encoding a protein with strong sequence similarity to E. coli RNase H1 but lacks a portion of the basic protrusion and has other changes that render the protein nonfunctional as an RNase H (Itaya et al. 1999; Ohtani et al. 1999a). The presence of a gene encoding an inactive RNase H1 in B. subtilis and the disparate activities of RNases H1 and HIII point out the difficulties in assigning a function merely on the basis of amino acid sequence similarity.

Thus far, at least one gene encoding an RNase H–like protein is present in all prokaryotic and archael genomes (Ohtani et al. 1999b). Most often there are two genes, either a combination of HI and HII or HII and HIII. Little is known about the number and types of RNases H in eukaryotes. Two proteins from mammalian sources and Saccharomyces cerevisiae (RNase H1 and RNase H2L) (Crouch and Cerritelli 1998; Frank, Braunshofer-Reiter, and Wintersberger 1998) have been shown to be related by amino acid sequence to E. coli RNases HI and HII. A third RNase H, RNase H70 (Frank et al. 1999), also is present in S. cerevisiae that has sequence similarity to several other proteins in S. cerevisiae, including Rex3p (RNA exonuclease), Rex4p, and Pan2p, the last being a subunit of the polyA ribonuclease. All these proteins are related to exonuclease III, an enzyme known for many years to degrade RNA of RNA-DNA hybrids (Keller and Crouch 1972). At present, there is no clear consensus amino acid sequence that will permit defining a protein related to RNase H70 as an RNase H.

Key words: ribonuclease H, splicing, multiple genes, Caenorhabditis elegans, human, genome, double-stranded RNA, RNA-DNA hybrids.

Address for correspondence and reprints: Robert J. Crouch, Building 6B Room 2B-231, 6 Center Drive MSC 2790, National Institutes of Health, Bethesda, MD 20892. E-mail: robert.crouch@nih.gov.

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038
An analysis of the genome sequence and EST data from *C. elegans* reveals four RNase HI–like genes. This large number of potential RNases H in a single organism is unprecedented, and there are features of the predicted proteins that are unique. In contrast, the human genome has two genes encoding RNase H1 and RNase H2. We characterized cDNAs obtained from the mRNA of *C. elegans* for different RNases H and identified unique structural features of some proteins. The splicing events for the regulation of the *rnh-1.0* gene, their phylogenetic relationship, and the enzymatic properties of different RNases H expressed in *E. coli* are also described.

Materials and Methods

Bacterial Strains

The *E. coli* strains used in this study were TOP10F* (Invitrogen), DH5α (Life Technologies), BL21(DE3)pLysS (Novagen), MIC1066 [rnhA-339::cat recB270(Ts)] (Cazenave, Mizrahi, and Crouch 1998).

Computer Analysis

*Caenorhabditis elegans* BLAST searches (Altschul et al. 1990) were performed at either the Sanger (http://www.sanger.ac.uk/Projects/C_elegans/blast_server.html) or National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/BLAST/) web sites. Query RNase H protein sequences were *E. coli* RNase H1 and RNases H1 of *S. pombe*, *S. cerevisiae*, and human. The human genome sequences at NCBI or National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/) were examined separately. Query RNase H protein sequences were submitted. Duplicate accession numbers were eliminated, creating a file used for Batch Entrez (http://www.ncbi.nlm.nih.gov/entrez/batchentrez.cgi?db=Nucleotide). The output from the Batch Entrez search is the Unmasked database. The Unmasked database was searched by the hurep.ref and hurep.sub file (http://www.girinst.org/server/RepBase/) (Jurka 2000), removing repeat sequences and generating the Masked database. The Masked database was searched using local tBLASTn with each of the RNase H queries to determine which sequences were selected by all or only some of the protein queries. The Celera database (http://publication.celera.com) was searched using the human RNase H1 protein and mRNA (cDNA) sequence as queries using BLASTp and BLASTn, respectively. Analyses of DNA and protein sequences were done using Wisconsin Package Version 10.0 (Genetics Computer Group [GCG], Madison, Wisc.).

Phylogenetic Analysis

The phylogenetic analysis was performed on the multiple sequence alignment shown in figure 1A using the maximum likelihood method implemented in version 3.6a2.1 of the PHYLIP package (Felsenstein 2001). The alignment was generated using PILEUP of the GCG package, and adjustments were made, taking into account the structural data for *E. coli* RNase HI. The sequence alignment was converted to the PHYLIP format, and SEQBOOT was used to generate 100 data replicates. The data were subsequently analyzed with PROML (constant rate of change), followed by NEIGHBOR, and finally with CONSENSE to generate the bootstrapped tree. The phylogram was rooted with *E. coli* RNase HI, and the tree was visualized with TreeViewPPC version 1.5.3 (Page 1996).

cDNA Cloning

The mRNA isolated from the total *C. elegans* RNA (Krause 1995) was reverse transcribed (Life Technologies) to synthesize the first strand of cDNA by using the oligo dT primer. Primers specific for corresponding *rnh* genes were designed on the basis of the sequence information obtained from the splicing predicted by the GENEFINDER program (table 1). More than one set of primers was used in PCR reactions to characterize each type of cDNA clone. Total PCR product or the gel purified (Gene clean II kit, Bio 101) product was cloned into the pCRII-TOPO TA cloning vector and transformed into TOP10F* competent cells (Invitrogen). More than two different isolates were sequenced in each case from two different RNA preparations.

Northern Blotting

Five micrograms of mRNA isolated from a mixed-stage population of *C. elegans* was run on a 1.0% agarose-LE gel and blotted to Bright Star Plus nylon membrane (Ambion). Northern hybridization was carried out by using the NorthernMax-Gly protocol, as described by the manufacturer (Ambion). 32P-Labeled antisense transcripts were generated by in vitro transcription (Promega) using DNA for the R1-R2 dsRNA-binding domains or the RNase H domain.

Transcript Analysis

The 5′- and 3′-rapid amplification of cDNA end (RACE) reactions were carried out to analyze transcripts of different RNase H genes. The 3′, 5′-RACE primers (Life Technologies) with gene-specific primers were used in PCR reactions to characterize the 3′- and 5′-ends of the messages. In some cases, primers specific for splice leaders (Huang and Hirsh 1989) were used with the gene-specific primers to identify the 5′-end of messages. To confirm that the PCR reaction products were indeed derived from the target mRNA when splice leader and gene-specific primers were used in PCR reactions, Southern analysis of the PCR products was carried out using the appropriate probes. Two different mRNA preparations were used.

Expression in *E. coli*

Once the complete cDNA sequence was obtained and the coding region determined, primers were synthe-
sized to amplify the coding regions such that an NdeI or NcoI site was at the first Met codon and the downstream primer included a BamHI or XhoI restriction enzyme site. The PCR products were cloned into pCRII-TOPO and digested with the appropriate enzymes, and the fragment was cloned into the pET15b expression vector (Novagen) and transformed into the BL21(DE3) pLysS E. coli strain (Novagen). Cells were grown at 32°C to mid-log phase, and expression was induced by the addition of IPTG (final concentration 1 mM), followed by incubation for 3 h. HIS-tagged proteins were purified from HIS-bind columns, as described by the manufacturer (Novagen, Clontech).

RNAse H Activity

In Gel Assay

Renaturation gel assays were carried out with the partially purified HIS-tagged proteins (Han, Ma, and Crouch 1997; Cazenave, Mizrahi, and Crouch 1998). RNase H activity was measured by renaturation gel assays carried out with the partially purified HIS-tagged proteins (Han, Ma, and Crouch 1997; Cazenave, Mizrahi, and Crouch 1998). Renaturation gel assays were carried out with the partially purified HIS-tagged proteins (Han, Ma, and Crouch 1997; Cazenave, Mizrahi, and Crouch 1998).
The pET15b vectors harboring different rnh-like cDNAs derived from mRNAs of *C. elegans* were transformed into the MIC1066 *E. coli* strain (Cazenave, Mizrahi, and Crouch 1998). Transformants were plated on LB-amp plates at 32 and 42°C. Growth at 42°C indicates that a functional RNase H is present in MIC1066. The pET15b vector was used as a control in this study.

**Results**

Database Searches for RNase H–Related Proteins in the Genomes of *C. elegans* and Humans

Using tBLASTn with RNases HI or H1 of *E. coli*, *S. pombe*, *S. cerevisiae*, and human, four *C. elegans* genes related to RNases HI or H1 were found, each having several of the hallmarks of RNases HI or H1. Another gene related by sequence to RNase HII of *S. cerevisiae*, *Escherichia coli*, and *C. elegans* was also observed. Table 2 shows the results of tBLASTn searches against the Sanger Centre and NCBI Databases. F59A6.6, ZK938.7, and CO4F12.9 give very significant “expect” values (see table 2) when any of the four RNases H is used as the query sequence. In contrast, ZK1290.6 was detected with low probability scores by only two of the RNases H queries. The alignments indicate that most but not all of the canonical RNase H residues are present in the predicted amino acid sequences (Ohtani et al. 1999b). In addition to the four genes shown in table 2, a few genes having low probability scores were detected. Upon examining the sequences carefully, we found these additional genes encode proteins that had none or only some of the highly conserved amino acid residues of RNases H.

Despite the similarity of *E. coli* RNase HI and eukaryotic RNases H1 to the sequence of RNase H found in reverse transcriptases and retrotransposons, none of the proteins in table 2 is from the retroviral elements of *Caenorhabditis elegans*.

### Table 1

**Oligonucleotides**

<table>
<thead>
<tr>
<th>Query</th>
<th>F59A6 (RNase H1)</th>
<th>ZK1290 (RNase H1A)</th>
<th>ZK938 (RNase H1B)</th>
<th>CO4F12 (RNase H1C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Escherichia coli H1</td>
<td>7 × 10^{-11}</td>
<td>9.0 × 10^{-10}</td>
<td>&gt;10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Saccharomyces pombe H1</td>
<td>3.3 × 10^{-14}</td>
<td>1.0 × 10^{-15}</td>
<td>0.69</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Human H1</td>
<td>9.5 × 10^{-23}</td>
<td>1.0 × 10^{-14}</td>
<td>0.0095</td>
<td>0.001</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae H1</td>
<td>3.5 × 10^{-10}</td>
<td>1.0 × 10^{-10}</td>
<td>&gt;10</td>
<td>8.0</td>
</tr>
</tbody>
</table>

**Note.** Query RNase H sequences were used to search the Sanger Centre or NCBI databases using tBLASTn. Five RNase H–related sequences were selected and the expect values of the different searches are shown. Values from the Sanger Centre are shown first, followed by those from NCBI. “The Expect value is a parameter that describes the number of hits one can ‘expect’ to see just by chance when searching a database of a particular size. It decreases exponentially with the Score that is assigned to a match between two sequences. Essentially, the E value describes the random background noise that exists for matches between sequences.” For details and source of quotation see: http://www.ncbi.nlm.nih.gov/BLAST/blast_FAQs.html#Expect.
C. elegans (Britten 1995; Bowen and McDonald 1999). Unlike the reverse transcriptase of HIV-1, the RNases H of the C. elegans retrotransposons are so distantly related to the cellular RNases H as used as query sequences that BLAST does not detect them. We believe that we have uncovered several interesting genes that are useful for studying RNases H in C. elegans, but because of our failure to find any C. elegans retroviral sequence, there may be more cellular rnh genes to be found. Only one candidate clone is found when using any of several RNase H1- or RNase H2L-like sequences as the query amino acid sequence (data not shown).

**Human Genome**

Having found several RNase H1 proteins in C. elegans, we searched the human genomic sequence for genes encoding RNase H1-like proteins. Although gaps are still present in the genomic sequence of the human DNA, we were unable to uncover any RNase H1-encoding genes except for the RNASEH1 gene on chromosome 2 (our modification of AC018488—see supplementary material at MBE web site: http://www.molbiolevol.org) and two pseudogenes; one located on chromosome 17p11.2 (AC022596.9) and one on chromosome 2q32.1-4 (AL035414.30) (supplementary material at MBE web site: http://www.molbiolevol.org) and two pseudogenes; one located on chromosome 17p11.2 (AC022596.9) and one on chromosome 2q32.1-4 (AL035414.30) (supplementary material). Unlike the search of C. elegans, the human DNA has numerous retroviral sequences that yield very low (significant) expect scores (supplementary material). One class of these is shown in figure 2 and is a member of the human endogenous retrovirus L (HERVL) family. Elimination of these sequences for further examination using RepMask sequences reduced the total number of sequences to about 47 (supplementary material). These sequences score well in the BLAST searches because of the three highly conserved Trp residues because the substitution matrix Blosum62 (Henikoff S. and Henikoff J. G. 1992) credits 11 points to Trp residues compared with 4 points for Leu-Leu matches. Interestingly, the conserved Trp residues are in the α10-αc-α17-region (fig. 1A), where HIV-1 RT has a deletion when compared with cellular RNases H1. Thus, the RNase H domain of HERVL elements are more similar to cellular RNase H1 than to the RNase H domain of HIV-1.

**Four C. elegans Genes Produce RNase H1-like Proteins**

The four RNase H1-related genes detected by the database searches were examined by cloning cDNAs generated by reverse transcription, followed by PCR amplification. We examined several independently derived clones for each gene. Amino acid sequences of the proteins expressed from these cDNAs are shown in figure 1A. Each of the cDNAs was cloned into pET15b for expression as a His-RNase H fusion protein in E. coli. Determination of RNase H activity of these four proteins expressed in E. coli was accomplished by two different assays. A typical, assay for RNase H activity based on measuring the degradation of RNA-DNA hybrids in a solution-based procedure was of limited usefulness because of the limited solubility of the proteins expressed in E. coli.

**In Gel Assay**

Results obtained from the gel renaturation assays with labeled RNA-DNA hybrids are presented in figure 3. RNase H1 (F59A6.6) and RNase H1A (ZK1290.6) exhibited enzymatic activity in this assay with activity detected in two bands for both samples. The major activity of RNase H1 (marked with an arrow) is coincident with the stained protein with some minor activity migrating at a position of a dimer of the 33-kDa protein. We occasionally see dimer bands in this gel assay (Han, Ma, and Crouch 1997). RNase H1A activity marked by the arrow corresponds to the full-length protein, whereas the band at about 35 kDa most likely represents activity of RNase H1 (marked with an arrow) is coincident with the stained protein with some minor activity migrating at a position of a dimer of the 33-kDa protein. We occasionally see dimer bands in this gel assay (Han, Ma, and Crouch 1997).RNase H1A activity marked by the arrow corresponds to the full-length protein, whereas the band at about 35 kDa most likely represents activity derived from a proteolytic product of RNase H1A. Proteins containing only the RNase H domain of RNase H1 or RNase H1A have significant RNase H activity in the gel assay (data not shown). Retention of the substrate in the region in the Activity Gel (fig. 3B—H1A—arrow at about 66 kDa) indicates the binding of the enzyme to the substrate without substantial degradation. This phenomenon is related to the N-terminal portion of RNase H1A to bind to some types of nucleic acids independent
Multiple mRNAs are Produced from the \textit{rnh-1.0} Gene

The mRNA encoding the \textit{C. elegans} RNase H1 protein is derived by the transcription of a 1.25-kb di-
FIG. 5.—Transcription and splicing of rnh-1.0. (A) Northern analysis of rnh-1.0. Northern analysis probing for transcripts derived from the rnh-1.0 gene. The blot was probed with antisense transcripts from the R1-R2 region (blue-R1, red-R2, and lane A) and from the RNase H region (black and lane B). The two transspliced products are detected as 0.5 kb (lane A, R1-R2) and 0.8 kb (lane B, R3-H). Purple denotes R3. (B) Splice patterns of C. elegans rnh-1.0. Splice patterns for multiple splice products are shown with boxed areas noting exons connected by thin carat lines marking the intronic regions. (i) is the full-length transcript (1.25 kb) detected by PCR using primers BC673 and BC675 and is labeled as “Dicistronic cDNA.” (ii) is the dicistronic precursor to the mature mRNAs shown in (iii) and (iv). The upstream R1-R2-encoding transcript (iii) results from cleavage and polyadenylation of (ii). The mRNA encoding the active RNase H1 protein is shown in (iv), having a purple duplex RNA-binding domain with black boxes designating the RNase H domain—noted as “Processed cDNA.” The duplex RNA-binding regions are shown as blue (R1), red (R2), and purple (R3). “polyA” denotes the position of polyA addition to the first mRNA, and SL2 indicates the site to which the SL2 sequence has been transspliced. Numbers above each indicate the exon number starting with the first exon in each case. The lower part of the figure represents an enlargement of the region in (i) and (iv) sometimes having an extra 19 nt (green) at the position marked by *.  

identified a locus in C. elegans cosmid T13H5, with the reported product (T13H5.2) being related to the squid retinal-binding protein. The large T13H5.2 protein of 1,264 amino acids has an RNase H2L-like 298-amino acid central domain with a 455-amino acid N-terminal and a 561-amino acid C-terminal domain. We made a series of oligonucleotides both upstream and downstream of the cDNA sequence encoding the RNase H2L protein, all of which failed to produce PCR products when paired with oligonucleotides internal to the rnh-2 gene (data not shown). By using primers both to the N-terminal and C-terminal regions and by analyzing the 5’- and 3’-ends of transcripts (RACE reactions), we have cloned and characterized the cDNA for RNase H2L (accession number AF181619). We conclude that the RNase H2L protein is not part of a larger polypeptide derived from either the N-terminal or the C-terminal domains, as predicted in the database.

Discussion
Multiple RNases H

Now, the embarkation point for examining proteins of cells can be from information provided by genomic DNA sequence determinations. Several surprises have already been uncovered when examining complete genome sequences for genes encoding RNase H-like proteins. In addition, proteins with RNase HI-related sequences do not always exhibit RNase H activity (Ohtani et al. 1999a). Because of the poor conservation of RNase HI sequences, the failure to find more than one RNase H in some organisms may be a reflection of the search engines used to find homologous protein sequences.

We have taken advantage of the wealth of information about the DNA sequences of C. elegans and humans to examine potential genes coding for RNases H. To know what the number and types of RNases H might be in animals on the basis of the genomic DNA sequence, we used several RNases H as query sequences in a BLAST search (tBLASTn). Use of multiple proteins helps to overcome the poor conservation of sequences of RNases H1, including the spacing between those regions that are most indicative of RNases H. Even when employing these searches, no retrotransposon element of C. elegans was detected. The alignment shown in figure 1A helps explain why we failed to detect C. elegans
retrotransposons in our searches. Only 17 of 155 amino acids are identical in the alignment, and three important conserved regions are missing (the αC region is missing as well as the region between β4 and α5, including the important His124-E. coli numbering). Even though Cer1-1 contains a Glu48 residue (fig. 1A), the context in which it resides bears little resemblance to the other RNases H1. In contrast, numerous human endogenous retrovirus (HERV) elements are readily detected in the human genome by using almost any of the RNase H proteins as query sequences. This indicates that many HERV RNase H domains are more recently derived from cellular RNases H or vice versa, whereas the separation between the origin of retroelements in 

C. elegans and cellular enzymes is much greater. This conclusion supports that of Malik and Eickbush (2001).

Five different genes yielding proteins related to RNases H were found in the C. elegans genome. In organisms such as E. coli and S. cerevisiae, cells deleted for both RNase H-encoding genes are viable (Frank et al. 1999; Itaya et al. 1999; Arudchandran et al. 2000; A. Arudchandran and R. J. Crouch, unpublished data). We found that RNAi (Fire et al. 1998) inactivation of any or several of the RNase H mRNAs also produced no easily detectable phenotype (data not shown). Thus far, only in Drosophila melanogaster is there a serious defect related to an RNase H mutation (Filippov, Filipova, and Gill 2001).

Of the five RNase H–related genes, one is similar to RNase H2 or RNase HIII. When expressed in E. coli or when assayed in a gel renaturation assay, no enzymatic activity is detected (fig. 3). We have expressed RNase H2 from S. cerevisiae, human, and mouse and uniformly find no enzymatic activity (Crouch and Cerritelli 1998). A similar observation has been reported for the human protein (Lima, Wu, and Crooke 2001), but others have observed very weak activity after re-folding of the S. cerevisiae RNase H2 in E. coli (Qiu et al. 1999). The RNase H2 may be composed of two sub-units (Frank et al. 1998) or may require modification to exhibit RNase H activity.

Two C. elegans proteins, similar to RNase HI of E. coli in amino acid sequence, exhibit RNase H activity when expressed in E. coli. Caenorhabditis elegans RNase H1 is similar to most RNases H1 of eukaryotes having a duplex RNA–binding domain at its N-terminus and the RNase H domain at the C-terminus. RNase H1A is unique. The N-terminal region is not found in any of the other RNase H sequences and contains a large number of the Arg-Ser repeats (fig. 1B), typical of SR proteins involved in splicing (Graveley 2000). The Arg-Ser repeats are important for protein-protein interactions and may direct these proteins to the spliceosome (Yuryev et al. 1996). There seems to be no obvious direct role for RNase H in splicing, and we are unaware of any report indicating a requirement for RNase H in splicing. The importance of the amino terminal region is unclear, particularly in light of the fact that it is not required for enzymatic activity (data not shown). The RNase H domain does differ from those of other active RNases H1. In particular, the αB-C-D-helices of RNase H1A are more similar in size and content to HIV-1 RNase H than to E. coli RNase HI, suggesting that additional amino acids are important for the binding of the protein to nucleic acid substrates. It may be that the C-terminal extension seen in RNase H1A supplies the binding function through the many basic amino acid residues present there. It should be pointed out that several of the Arg residues are followed by Ser, similar to what is found near the N-terminus of the protein. We are currently examining RNase H1A for determinants of RNase H activity and inquiring into the role of the non–RNase H domain.

The RNases H1B and H1C are inactive, as expressed from the cDNAs we have cloned. The genes encoding RNases H1B and H1C contain introns and, therefore, are probably not pseudogenes. RNase H1C does not have the C-terminal αC-helix whose presence is necessary for enzymatic activity (Haruki et al. 1994; Goedken, Raschke, and Marqusee 1997). If a splice were to occur near the end of the gene, an α-helix could possibly be attached. We have examined four independently derived cDNAs and have found no example of a transcript encoding the putative αC-helix. The defect in RNase H1B is most likely due to the unusual nature of the αβ±αC±αD±region. In RNase H1B, there are numerous Thr and Ser residues rather than the typical basic and Trp residues at conserved locations (fig. 1A). We have obtained three independent clones of rnh-1.2, all of which have the same sequence. The RNases H1 and H1C contain introns between the coding sequences for the αA and β1 (fig. 1A—reverse letters indicate splice site). RNase H1B has no equivalent splice site. Translation of the mRNA in the αA±αC±αD±region of the rnh-1.2 in all three reading frames reveals the presence of an out-of-frame coding sequence that yields a very good αβ±αC±αD±region, thereby suggesting that the formation of an active RNase H1B is possible (data not shown). The RNases H1B and H1C may have a function(s) other than providing RNase H activity, but it is also possible that their expression in an RNase H active form may be limited to specialized situations.

The abundance and diversity of alternatively spliced mRNAs of C. elegans RNase H1 is striking (fig. 5B and supplementary material) and makes it clear that synthesis of RNase H1 is regulated by splicing. The primary transcript is differentially spliced to produce two types of dicistronic messages, one of which is processed by the usual pathway for the generation of two mRNAs encoding two different proteins. Because the intercistronic region having the poly(A) addition signal is deleted when splicing occurs in the alternatively spliced mRNA (fig. 5B exon 3 to exon 4), the two coding regions remain on a single message and would require internal initiation of translation to produce RNase H1. The alternative splice joining exon 5 to exon 6 (fig. 5B) does not permit the synthesis of an active RNase H1.

Generality of Multiple Genes in C. elegans

In contrast to the C. elegans genome, we have been unable to find evidence for multiple RNase H1–like pro-
teins in the human genome. This disparity in the numbers of proteins of a particular type between \textit{C. elegans} and other genomes is not unique to RNases H (Combes et al. 2000; Keiper et al. 2000; Robertson 2000; Hodgkin 2001). Sternberg (2001) has suggested that each of the small number of cells comprising \textit{C. elegans} may be more complex or may respond in more complex ways due to increased molecular diversity within each cell. One prime example is in olfactory neuronal cells where one cell possesses multiple receptors and yet can sense different odors (Bargmann 1998). In simpler cell types, a single protein may perform many functions but would be limited in its role in any cell by the presence of one or only a few substrates. For example, RNase H1 in human cells may have multiple functions, but within a given cell type these functions may be limited by environmental factors such as substrates. In \textit{C. elegans}, RNases H of several types may be present within a single cell type but may of necessity be limited to one function or one cell organelle, and the apparent regulation of RNase H1 levels by splicing may indicate that this protein may be able to recognize all the cellular substrates and, therefore, needs to be kept under tight control so that it does not subsume another enzyme’s role. Alternatively, RNase H1A may have a requirement uniquely present in \textit{C. elegans} for splicing or some splice-related event, as indicated by the N-terminal SR character (fig. 1B).

Acknowledgments

We thank Vivian Tsai for initial characterization of the yk177f5 clone and Dr. Yuji Kohara for several cDNA clones.

LITERATURE CITED


David Irwin, reviewing editor

Accepted July 15, 2002