JOURNAL OF BACTERIOLOGY, June 1997, p. 3828–3836 Vol. 179, No. 12
0021-9193/97/$04.00+0

The *rnhB* Gene Encoding RNase HII of *Streptococcus pneumoniae* and Evidence of Conserved Motifs in Eucaryotic Genes

YIAN-BIAO ZHANG,† SAHLU AYALEW, AND SANFORD A. LACKS*

Biology Department, Brookhaven National Laboratory, Upton, New York 11973

Received 23 January 1997/Accepted 7 April 1997

A single RNase H enzyme was detected in extracts of *Streptococcus pneumoniae*. The gene encoding this enzyme was cloned and expressed in *Escherichia coli*, as demonstrated by its ability to complement a double-mutant *rnhA recC* strain. Sequence analysis of the cloned DNA revealed an open reading frame of 290 codons that encodes a polypeptide of 31.9 kDa. The predicted protein exhibits a low level of homology (19% identity of amino acid residues) to RNase HII encoded by *rnhB* of *E. coli*. Identification of the *S. pneumoniae* RNase HII translation start site by amino-terminal sequencing of the protein and of mRNA start sites by primer extension with reverse transcriptase showed that the major transcript encoding *rnhB* begins at the protein start site. Comparison of the *S. pneumoniae* and *E. coli* RNase HII sequences and sequences of other, putative bacterial *rnhB* gene products surmised from sequencing data revealed three conserved motifs. Use of these motifs to search for homologous genes in eucaryotes demonstrated the presence of *rnhB* genes in a yeast and a roundworm. Partial *rnhB* gene sequences were detected among expressed sequences of mouse and human cells. From these data, it appears that RNase HII is universally present in living cells.

RNase H cleaves the RNA strand in hybrid molecules containing paired RNA and DNA strands (39). This enzymatic activity has been found in a variety of sources ranging from bacteria to mammalian tissues (7). In *Escherichia coli*, two different genes were found to encode such enzymes; *rnhA* encodes RNase HI (16), and *rnhB* encodes RNase HII (12). The two enzymes contain 155 and 209 amino acid residues, respectively, and their sequences show no significant similarity. With respect to eucaryotes, a gene encoding an enzyme of 166 amino acid residues that is homologous to RNase HI was cloned from the yeast *Saccharomyces cerevisiae* (13). Several retroviruses have been shown to produce a polypeptide with RNA-directed DNA polymerase activity, of which a carboxyl-terminal segment of ~150 residues is homologous to *E. coli* RNase HI (15). Recently, a gene from the trypanosome *Crithidia fasciculata* was cloned and shown to encode a 494-amino-acid protein, in which the last 200 residues contained four sequence motifs common to RNase HI enzymes (2), and a homologous gene was identified in a cDNA library prepared from chicken lens RNA (24). Although genes apparently homologous to *E. coli rnhB* occur in several sequenced bacterial genomes (GenBank accession no. D32253, D90899, S76857, U32786, U67470, and Z74024), so far no other RNase HII enzymes have been directly identified.

RNase H enzymes purified from vertebrate sources have been placed into two classes, depending on their response to activators and inhibitors and the size of the active polypeptides (4). Distinction on the basis of size may be deceptive, however, in that smaller polypeptides that become prominent during purification could result from proteolytic action (30). In the absence of sequence information, the relationship of these eucaryotic enzyme classes to *E. coli* RNase HI and II is unknown. However, Walder and coworkers found an important distinction between a human RNase H from erythroleukemic cells and *E. coli* RNase HI. The latter enzyme required a minimum stretch of four ribonucleotides in the RNA segment of a DNA-RNA-DNA/DNA hybrid substrate for cleavage to occur (11). In contrast, the human enzyme was able to cleave at a single ribonucleotide in one strand of the hybrid substrate (8).

Despite their ubiquity, little is known of the functions of RNases H. The retrovirus enzyme degrades the RNA strand in the hybrid product, thereby allowing synthesis of a DNA duplex that can be integrated into the cellular genome (15). It has been proposed that *E. coli* RNase HI participates in DNA replication by preventing aberrant chromosomal initiation (17) and by removing RNA primers from Okazaki fragments of the discontinuously synthesized strand (25). Although it is presumably not a normal function, when antisense oligonucleotides are introduced into a cell, RNase H can specifically degrade the complementary mRNA (4).

An RNase H like the one from erythroleukemic cells could have a repair function in removing ribonucleotides accidentally incorporated into DNA (8). We wondered whether RNase H action at such adventitious ribonucleotides in newly synthesized DNA could provide strand breaks for targeting mismatch repair after DNA replication. DNA base mismatch repair is carried out by a homologous system in virtually all cells to eliminate potential mutations due to mistakes in DNA replication (6, 9, 26). In humans, defects in the mismatch repair system result in a propensity for cancer (9). The repair system recognizes the base mismatch and removes a lengthy strand segment from the newly synthesized strand (6, 9). In *Streptococcus pneumoniae*, the same system acts on heteroduplex DNA formed during DNA-mediated genetic transformation to remove the donor DNA strand segment prior to its ligation (6). Although in enterobacteria the system has an added fillip that recognizes unmethylated sites in the nascent DNA (9), in most cells, including *S. pneumoniae*, yeast, fruit fly and human cells, the presence of strand breaks targets the DNA strand to be corrected (9). In DNA replication, strand breaks in the lagging strand result from its discontinuous synthesis, but the leading strand is presumably continuous. Adventitious incorporation of ribonucleotides and cleavage by an RNase H could provide strand breaks and allow mismatch correction in the leading

* Corresponding author. Phone: (516) 344-3369. Fax: (516) 344-3407. E-mail: lacks@sun2.bnl.gov.

† Present address: Department of Biochemistry and Cell Biology, State University of New York, Stony Brook, NY 11794.

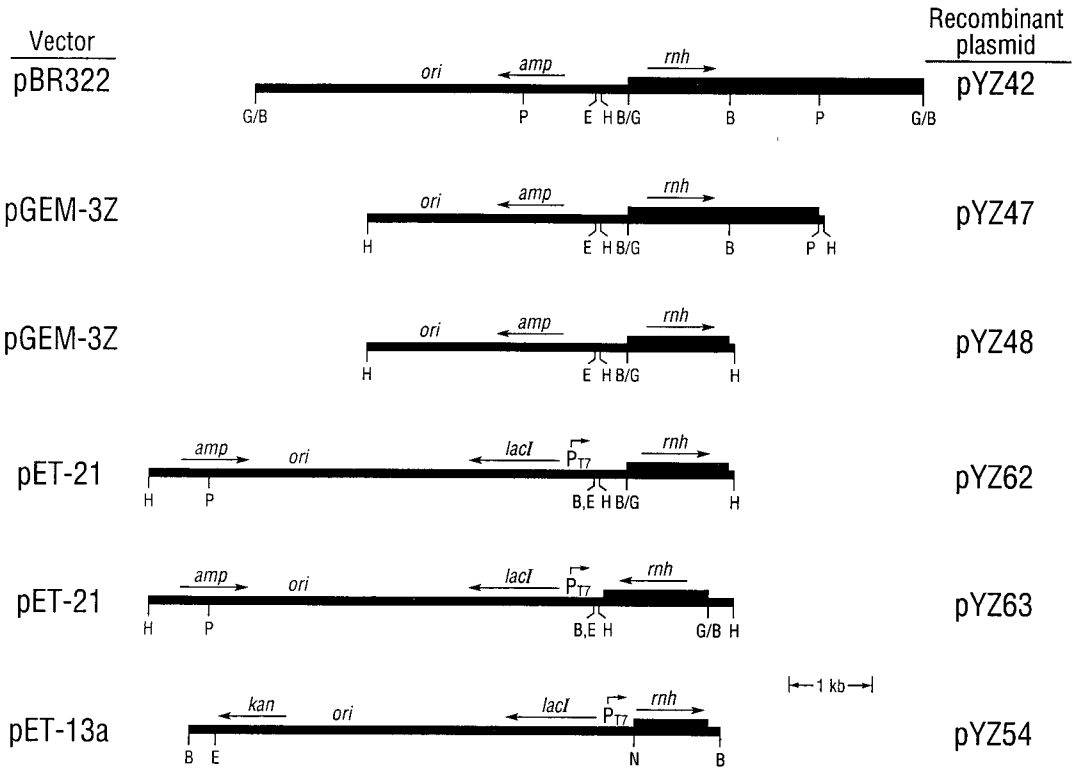


FIG. 1. Cloning and subcloning of the *rnhB* gene of *S. pneumoniae*. Thin bar, vector; thick bar, chromosomal segment. Restriction sites indicated below bar: B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; H, *Hin*dIII; N, *Nde*I; P, *Pst*I. Gene designations are above bar, with arrows showing direction of transcription and translation; *ori*, plasmid origin of replication; P_{T7}, phage T7 RNA polymerase promoter.

strand. To explore this possibility, we set out to characterize enzymes in *S. pneumoniae* with RNase H activity. A single RNase H enzyme was found, the gene encoding it was cloned and sequenced, and expression of the protein was investigated.

MATERIALS AND METHODS

Bacterial strains and plasmids. Strains of *S. pneumoniae* used, derivatives of R6, were 216 (*malM597*), 533 (*sul-a sul-d str bry nov ery*), and 1181 (*malM567 trt-1 endA1 exoA2 ung-1*) (5). Strains of *E. coli* used and their relevant genotypes were DH5α, BL21(DE3) (40), BW9091 (*xthA1*) (23), and MIC3001 (*rnhA339*::*cat recC271*) (13). Plasmids used, some of which are indicated in Fig. 1, were pBR322 (41), pGEM-3Z (Promega), pET-9c (40), pET-13a (40), and pET-21 (Novagen).

Growth and transformation of bacteria. Cultures of *S. pneumoniae* were grown as previously described (19). *E. coli* was grown in LB broth (33) and transformed by electroporation in a BTX *E. coli* TransPorator according to the protocol supplied by the manufacturer. Transformants were selected with ampicillin at 50 μg/ml or kanamycin at 25 μg/ml.

DNA manipulation and plasmid construction. DNA preparation, manipulation, and analysis were carried out by standard methods (33). DNA fragments were separated by electrophoresis on agarose gels and purified by binding to silica (Geneclean II kit; Bio 101, Inc.). To form pYZ47, the 2.6-kb *Eco*RI-*Pst*I fragment from pYZ42 was inserted between the *Eco*RI and *Pst*I sites in pGEM-3Z. To form pYZ48, pYZ47 was cut with *Bam*HI and *Pst*I, treated with Klenow fragment, and ligated to join the blunt ends of the vector-containing fragment. Plasmids pYZ62 and pYZ63 were formed by inserting the 1.6-kb *Hin*dIII fragment from pYZ48 into the *Hin*dIII site of pET21 in opposite orientations. To form pYZ54, an *Nde*I site was introduced at the start of the *rnh* gene (details to be reported elsewhere [44]), and an *Nde*I-*Bam*HI fragment containing the gene was inserted into pET-13a.

DNA sequence determination. Sequences were determined by the method of Sanger et al. (34), using the Sequenase 2.0 derivative of T7 DNA polymerase (U.S. Biochemical). Four deletions of increasing size, extending leftward (Fig. 1) from the *Pst*I site to different extents past the *Bam*HI site, were produced in pYZ47. This was accomplished by cutting pYZ47 with *Bam*HI and *Pst*I, treating for various lengths of time with exonuclease III, briefly with nuclease S1, and then with Klenow fragment, and ligating to circularize the deleted plasmid (10). Using these deleted plasmids and pYZ48 as templates, we determined sequences from oligonucleotide primers corresponding to the pBR322 sequence just to the left of its *Bam*HI site and the M13 sequence just to the right of the multicloning segment in pGEM-3Z (orientations as shown in Fig. 1). Five additional primers were synthesized to complete the sequence determination on both strands of the cloned segment extending from the left end of the insert to the *Bam*HI site.

Preparation of cell extracts. To prepare crude extracts for enzymatic assay after gel electrophoresis, cultures of *S. pneumoniae* were grown to an optical density at 650 nm of 0.5. Cells from 15 ml were sedimented by centrifugation and suspended in 180 μl of buffer containing 10 mM Tris-HCl (pH 7.6), 50 mM NaCl, and 1 mM dithiothreitol (DTT). The cells were lysed by addition of 20 μl of 5% Triton X-100 and 15 min of incubation at 30°C. The viscosity of the extracts was reduced by passage through a 27-gauge needle. Extracts of *E. coli* were similarly prepared except that after suspension in buffer, 10 μl of lysozyme at 4 mg/ml and 10 μl of Triton X-100 were added, and the cells were held at 0°C for 20 min and then frozen and thawed four times. To induce cultures of BL21(DE3), isopropylthiogalactoside (IPTG) was added to 1 mM and the cultures were incubated for 90 min before harvesting.

RNase H assay after gel electrophoresis. The RNA-DNA hybrid substrate was prepared by synthesizing a poly(A) RNA strand labeled with ^{32}P on a poly(dT) template (35). The reaction mixture contained 50 mM Tris-HCl (pH 7.9), 5 mM $MgCl_2$, 1 mM $MnCl_2$, 1 mM DTT, 5% glycerol, 17 μg of poly(dT) (Sigma), 0.1 mM ATP, 50 μCi of [α-^{32}P]ATP (800 Ci/mmol), and 20 U of *E. coli* RNA polymerase (Sigma) in a total volume of 0.2 ml. After incubation at 37°C for 30 min, the mixture was shaken with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1). One-tenth volume of 3 M sodium acetate and 2 volumes of ethanol were added to the aqueous supernatant to precipitate the hybrid RNA-DNA, which was dissolved in 0.1 ml of 10 mM Tris-HCl (pH 8.0)–0.1 mM EDTA and passed through a Sephadex G-50 spin column (Princeton Separations) to eliminate residual [α-^{32}P]ATP.

Samples of cell extracts or purified proteins were subjected to polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS) as previously described (31), with the following changes: EDTA was not added to any of the solutions; the running and separating gel buffers were at pH 8.8; and bovine serum albumin at 10 μg/ml was added to the separating gel. SDS from Matheson, Coleman and Bell (lot 27) was used at 0.1%; use of pure SDS requires higher detergent concentrations (20). Approximately 1.5 μCi of the ^{32}P-labeled RNA-DNA hybrid substrate was added to the separating gel. In some cases, herring sperm DNA at 10 μg/ml was also added. After electrophoresis, the gel was washed three times by gently shaking for 30 min with 200 ml of 50 mM Tris-HCl (pH 7.5), 4 mM $MgCl_2$–0.1 mM DTT. Gels were incubated at 37°C in the same

buffer for several days. At intervals, the wet gels were covered with polyethylene film and exposed to a Molecular Dynamics PhosphorImager screen; clear bands indicate the absence of radioactivity and degradation of the RNA-DNA substrate. Addition of excess DNA prevents the early appearance of clear bands due to nonspecific nucleases. Identification of DNases after prolonged incubation was accomplished by staining the gel with ethidium bromide and looking for clear bands in the stained DNA background (31).

Determination of amino-terminal polypeptide sequence. The *S. pneumoniae* RNase H protein was produced in the BL21(DE3)[pYZ54] overexpressing strain and purified to homogeneity (44). Amino-terminal sequencing of the protein was carried out by the Bioscience Core Facility of Notre Dame University.

Mapping of transcription start sites. RNA was prepared as previously described (32) from *E. coli* DH5α, DH5α[pYZ62], and DH5α[pYZ63] and *S. pneumoniae* 216. The 5′ ends of the *rnhB* mRNAs were determined by primer extension using oligonucleotide A (Oligo A; 5′-CCTTCACCCTGAAGC-3′), which is complementary to nucleotides (nt) 286 to 272 in the *rnhB* sequence, and Oligo B (5′-GATTCTTGCTAGGAGCCAGACTGG-3′), which is complementary to nt 204 to 181 in the *rnhB* sequence. These oligonucleotides were labeled at their 5′ ends with polynucleotide kinase and [γ-^{32}P]ATP (3,000 Ci/mmol), and after purification by ethanol precipitation, 1 pmol of the labeled oligonucleotide was annealed with 15 μg of total RNA in 26 μl of 50 mM Tris-HCl (pH 8.3)–40 mM KCl–1 mM $MgCl_2$–33 U of RNase inhibitor (RNasin; Promega) for 5 min at 56°C for Oligo A or 74°C for Oligo B. The mixture was quickly cooled in a dry ice-ethanol bath. After thawing at 0°C and subsequent addition of the four deoxynucleoside triphosphates, each to 100 μM, 1 mM DTT, and 200 U of Moloney murine leukemia virus reverse transcriptase (volume now 35 μl), the mixture was incubated at 37°C for 30 min. Products were examined on a sequencing gel and compared to a dideoxynucleotide-terminated DNA sequencing ladder obtained with Oligo A. Radioactive bands in the gel were detected by autoradiography with Amersham Hyperfilm MP and quantified with a Molecular Dynamics PhosphorImager.

Nucleotide sequence accession number. The DNA sequence data described in this report have been deposited in GenBank with accession no. U93576.

RESULTS

Cloning of a gene from *S. pneumoniae* encoding RNase H. A library of *Bgl*II-cut fragments of chromosomal DNA from *S. pneumoniae* 533 ligated into the *Bam*HI site of *E. coli* plasmid pBR322 was used to transform *E. coli* MIC3001. This strain contains a null RNase HI mutation, *rnhA339*::*cat*, and a temperature-sensitive *recC271* mutation, which results in a temperature-sensitive growth phenotype that can be complemented by clones producing either RNase HI (13) or RNase HII (12). A single transformant able to grow at 42°C was found. It contained a plasmid, pYZ42, with a 3.5-kb insert (Fig. 1).

Location and expression of the RNase H gene. The gene responsible for the RNase H function in the complementation assay was localized to a 1.2-kb fragment at one end of the original insert. This was done by subcloning the fragment first into pGEM-3Z and then into pET-21 (Fig. 1). Removal of the *Hin*dIII-*Bam*HI fragment from pYZ42 separates the pneumococcal insert from the *tet* promoter in the pBR322 vector (41). When the fragment was cloned into pET-21 in either orientation, it was able to complement the temperature sensitivity of strain MIC3001, as shown by growth at 42°C resulting from transformation with either pYZ62 or pYZ63 (Fig. 2). The complementation in both orientations indicated that the fragment carries its own promoter for transcription of the *rnh* gene encoding the RNase H.

To determine the size and number of RNase H enzymes in *S. pneumoniae*, an in situ gel assay for RNase H (3) was used. This assay was based on the finding by Rosenthal and Lacks (31) that enzyme proteins could be renatured after gel electrophoresis in the presence of SDS, thereby allowing the power of this technique to separate proteins on the basis of size to be coupled to an activity assay. Enzymes consisting of monomers or identical dimers are readily renatured (20). As shown in Fig. 3, lane 3, *S. pneumoniae* contains only a single RNase H activity, corresponding to a polypeptide of ~30 kDa. The strain tested is defective in *endA* and *exoA*. These genes encode an endonuclease that cleaves RNA as well as DNA (29, 31) and

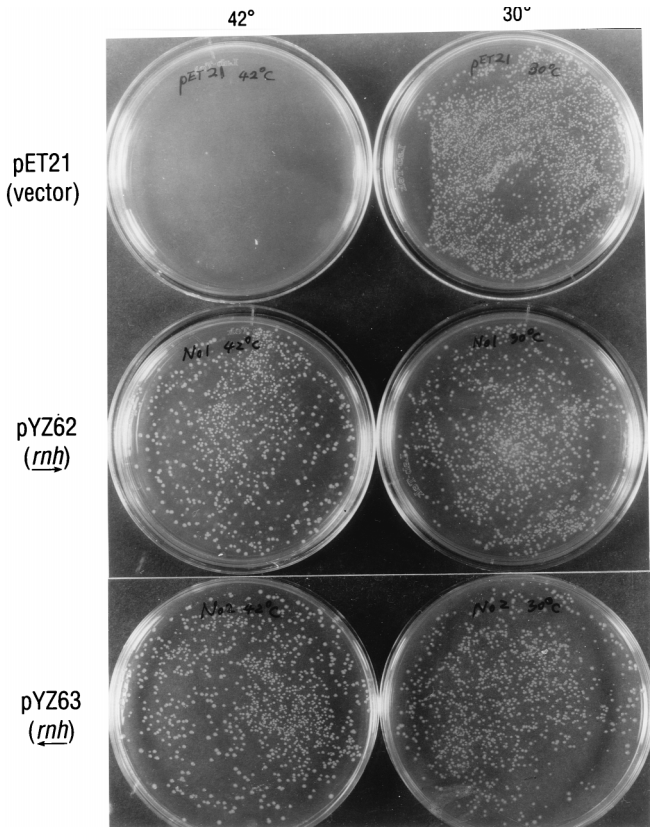


FIG. 2. Complementation by a segment of *S. pneumoniae* DNA of an *E. coli* strain deficient in RNase H. Samples of cultures of *E. coli* MIC3001 containing plasmid pET-21, pYZ62, or pYZ63 were spread on duplicate plates of LB agar and incubated for 16 h in parallel at 42 and 30°C. Plasmid pET-21 is the vector, in which the *S. pneumoniae* segment is inserted in opposite directions in pYZ62 and pYZ63. Strain MIC3001 cannot grow at 42°C due to a deficiency of RNase H.

an exonuclease homologous to *E. coli* exonuclease III (28), respectively. Both enzymes act on hybrid RNA-DNA; therefore, with a wild-type extract, two additional bands are observed, but these enzyme bands can be distinguished from RNase H by their ability to degrade DNA as well as the hybrid substrate (data not shown). Additional evidence for the existence of a single RNase H in *S. pneumoniae* corresponding to the cloned gene comes from preliminary results showing that an insertion mutation of the gene in *S. pneumoniae* eliminates the activity band in the gel assay (45).

In contrast to the single RNase H in *S. pneumoniae*, *E. coli* contains two such enzymes, the 18-kDa RNase HI (16) and the 23-kDa RNase HII (12), both of which are evident in Fig. 3, lane 1. The weak band at 30 kDa given by the *E. coli* extract corresponds to exonuclease III, as previously reported (3) and confirmed by us; that is, a band at that position acts also on DNA and no band appears there with extracts from the *xth* strain BW9091 (data not shown). The band at 19 kDa, just above RNase HI, corresponds to an unidentified nuclease that acts on DNA as well (data not shown). From the amounts of protein needed to produce bands of equal intensity, it appears that the level of RNase H activity in *S. pneumoniae* is 10-fold greater than the activity of either RNase H enzyme in *E. coli*. However, renaturation rates and ionic requirements might vary for the several enzymes, and so this result may not reflect native activities in vivo.

Although pET-21 is an expression vector based on the phage

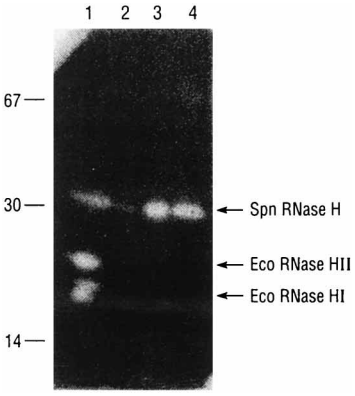


FIG. 3. Expression of RNase H after electrophoresis of extracts of *S. pneumoniae* (Spn) and *E. coli* (Eco). Lanes contained bacterial extracts as follows: 1, *E. coli* BL21DE3[pET-9c], induced with IPTG, 40 μg of protein; 2, same extract as in lane 1, 4 μg of protein; 3, *S. pneumoniae* 1181, 3 μg of protein; 4, BL21(DE3) [pYZ54], induced with IPTG, 0.04 μg of protein. Plasmid pET-9c is a vector similar to pET-13a used to construct pYZ54. Extracts were subjected to polyacrylamide gel electrophoresis in the presence of SDS. The gel contained poly ([^{32}P]rA)-poly(dT) substrate. After renaturation of the enzymes and incubation at 37°C as described in Materials and Methods, the gel was exposed to a PhosphorImager screen, and the image was obtained with a laser printer. White bands result from enzymatic degradation of the radioactive substrate. Positions of protein markers (bovine serum albumin, bovine carbonic anhydrase, and chicken egg lysozyme) were determined by staining a separate portion of the gel.

T7 RNA polymerase (40), relatively little expression of the *rnh* product was detected in strain BL21(DE3) carrying pYZ62 after induction of the polymerase (data not shown). However, after introduction of a ribosome binding site just upstream from a lengthy open reading frame in the subcloned fragment of pYZ54, overexpression of RNase H was obtained (Fig. 3, lane 4). Judging from the renatured activity, we estimate the level of overexpressed pneumococcal RNase H to be 1,000-fold higher than the basal level of RNase H in *E. coli* and 100-fold greater than the chromosomally produced enzyme level in *S. pneumoniae*. Furthermore, expression of the full-sized RNase H protein by pYZ54 localizes the *rnh* gene to a 1.2-kb segment at the left end of the original insert as depicted in Fig. 1.

Sequence of the *S. pneumoniae rnh* gene. The DNA sequence of a 990-bp segment of the cloned chromosomal insert from *S. pneumoniae* is shown in Fig. 4. A single large open reading frame is evident. From the putative ATG start codon at bp 117, it encodes a protein of 290 amino acid residues with a predicted mass of 31.9 kDa, which agrees reasonably well with the size shown by the enzymatic activity. The next possible start site, TTG at bp 222, would encode a protein of mass 28.5 kDa. Neither possible start site has a Shine-Dalgarno sequence for ribosome binding of mRNA associated with it. Direct sequencing of the amino terminus of the *S. pneumoniae* RNase H protein expressed in *E. coli* gave, for 16 cycles of the Edman reaction, the sequence ASITLTPSEKDIQAFL. Except for the absence of the initial methionine residue, which apparently had been removed, this result corresponds exactly to a translation start codon at bp 117. Inasmuch as the protein produced in *S. pneumoniae* and the overexpressed protein migrate identically in SDS-gel electrophoresis (Fig. 3, lanes 3 and 4), this start site is evidently used in *S. pneumoniae* as well.

Immediately upstream from the translation start site, a consensus procaryotic promoter sequence with a −35 box and an extended −10 box characteristic of *S. pneumoniae* promoters (32) is present (Fig. 4). The −35 box agrees with the consensus at five of six positions and the extended −10 box at eight of nine positions. The predicted transcription start site for this promoter would be very close to the translation start site, but this may be irrelevant for gene expression inasmuch as no ribosome binding sequence is present upstream of the coding region, and this situation has been observed for two other cases in *S. pneumoniae* (22, 32). In addition to this promoter at bp 79 to 109, another possible promoter without an extended −10 box, but otherwise with good agreement to the consensus, five of six for both −35 and −10 boxes, occurs at bp 24 to 52 (Fig. 4).

Comparison of the predicted protein sequence of the *S. pneumoniae* RNase H to that of *E. coli* RNase HI showed no evidence of homology, but the pneumococcal protein did show similarity to RNase HII. When the 213-amino-acid *E. coli*

```
TGTAAATTTG AATCGATTTA GATTTGCCAT AAAATTCACC TCACGATATT ATACCAAAAT TCGCTAATTT TGTCAGTTTT TACAAATTTT  90
                          -35                          -10                          -35

ACTGCTTTTG TGGTACAATA GAAACTATGG CAAGTATCAC ACTCACACCA AGCGAAAAGG ATATTCAGGC TTTTCTTGAA CACTATCAAA 180
              -10          M  A  S  I  T  L  T  P  S  E  K  D  I  Q  A  F  L  E  H  Y  Q

CCAGTCTGGC TCCTAGCAAG AATCCCTATA TCCGCTACTT TTTGAAACTA CCTCAAGCAA CGGTTTCTAT CTATACTTCT GGAAAAATCT 270
T  S  L  A  P  S  K  N  P  Y  I  R  Y  F  L  K  L  P  Q  A  T  V  S  I  Y  T  S  G  K  I

TGCTTCAGGG TGAAGGGGCT GAAAAATACG CCAGTTTCTT TGGCTATCAA GCTGTAGAGC AAACCAGCGG ACAAAATCTT CCTTTAATTG 360
L  L  Q  G  E  G  A  E  K  Y  A  S  F  F  G  Y  Q  A  V  E  Q  T  S  G  Q  N  L  P  L  I

GGACAGATGA GGTGGGAAAT GGTTCCTACT TTGGTGGGCT TGCAGTTGTG GCTGCCTTTG TCACACCTGA CCAGCACGAC TTTTTACGAA 480
G  T  D  E  V  G  N  G  S  Y  F  G  G  L  A  V  V  A  A  F  V  T  P  D  Q  H  D  F  L  R

AACTCGGTGT GGGGGATTCT AAGACTCTGA CCGACCAAAA GATCCGTCAG ATTGCTCCTA TTCTCAAGGA AAAAATCCAG CACCAGGCAC 540
K  L  G  V  G  D  S  K  T  L  T  D  Q  K  I  R  Q  I  A  P  I  L  K  E  K  I  Q  H  Q  A

TCCTTCTCTC ACCCAGCAAG TACAACGAGG TCATCGGAGA CCGCTACAAT GCTGTTTCGG TTAAGGTTGC CCTCCATAAT CAGGCTATCT 630
L  L  L  S  P  S  K  Y  N  E  V  I  G  D  R  Y  N  A  V  S  V  K  V  A  L  H  N  Q  A  I

ATCTCCTCCT TCAAAAAGGT GTTCAGCCTG AGAAAATTGT GATTGATGCC TTTACCAGTG CTAAAAATTA TGACAAGTAC TTGGCACAAG 720
Y  L  L  L  Q  K  G  V  Q  P  E  K  I  V  I  D  A  F  T  S  A  K  N  Y  D  K  Y  L  A  Q

AGACCAATCG TTTCAGCAAT CCTATCAGCT TAGAAGAAAA GGCTGAGGGC AAATACTTGG CTGTCGCAGT TTCTTCTGTC ATTGCGCGTG 810
E  T  N  R  F  S  N  P  I  S  L  E  E  K  A  E  G  K  Y  L  A  V  A  V  S  S  V  I  A  R

ATCTCTTTCT GGAAAATCTT GAAAATCTGG GACGAGAACT GGGTTATCAG CTTCCAAGTG GAGCTGGAAC GGCTTCTGAC AAGGTGGCTA 900
D  L  F  L  E  N  L  E  N  L  G  R  E  L  G  Y  Q  L  P  S  G  A  G  T  A  S  D  K  V  A

GCCAGATTTT GCAAGCCTAT GGTATGCAGG GACTCAACTT GTGCGCTAAG CTACACTTTA AAAATACTGA AAAAGCGAAA AACGCTTAGA 990
S  Q  I  L  Q  A  Y  G  M  Q  G  L  N  F  C  A  K  L  H  F  K  N  T  E  K  A  K  N  A  *
```

FIG. 4. Nucleotide sequence of a 990-bp segment of cloned DNA from the *S. pneumoniae* chromosome containing *rnhB*. The sequence of one DNA strand is shown with its 5′ end at nt 1 and its 3′ end at nt 990. Promoters for observed transcripts are underlined. The predicted amino acid sequence of the RNase HII polypeptide is shown. Determination of the amino-terminal sequence of the purified RNase HII protein revealed the underlined residues.

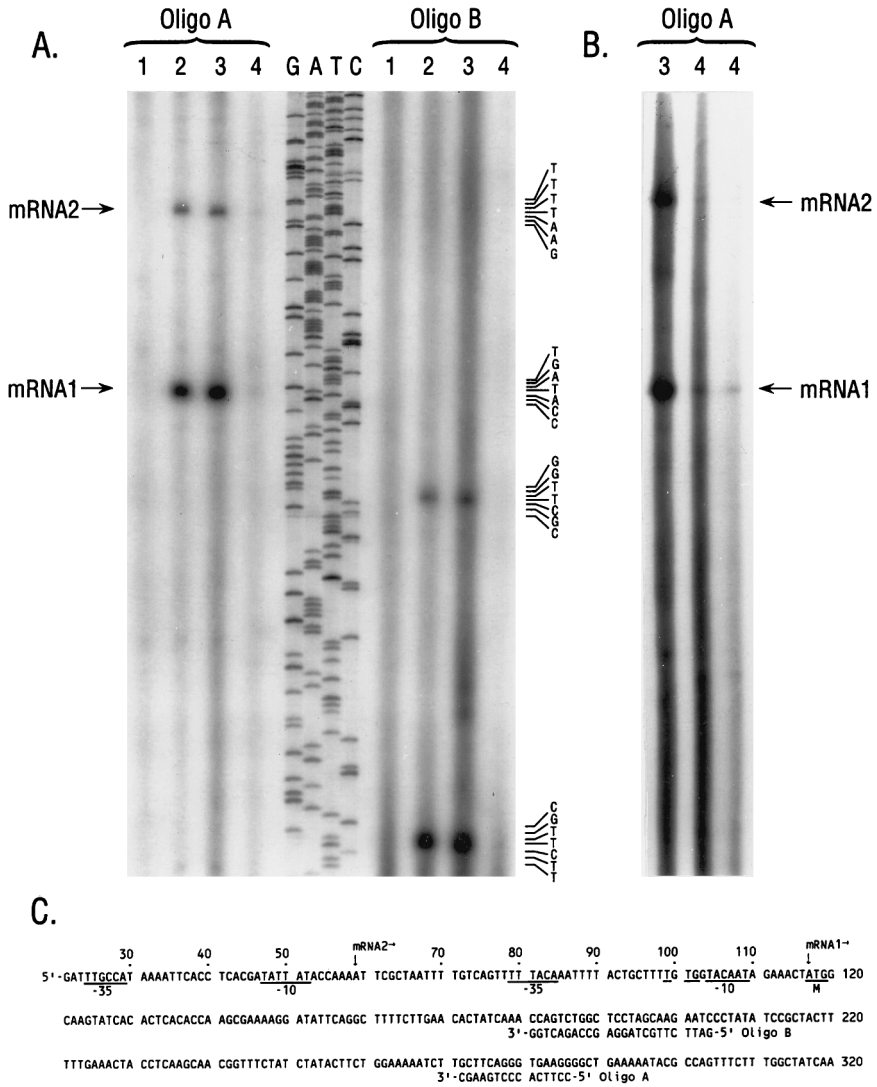
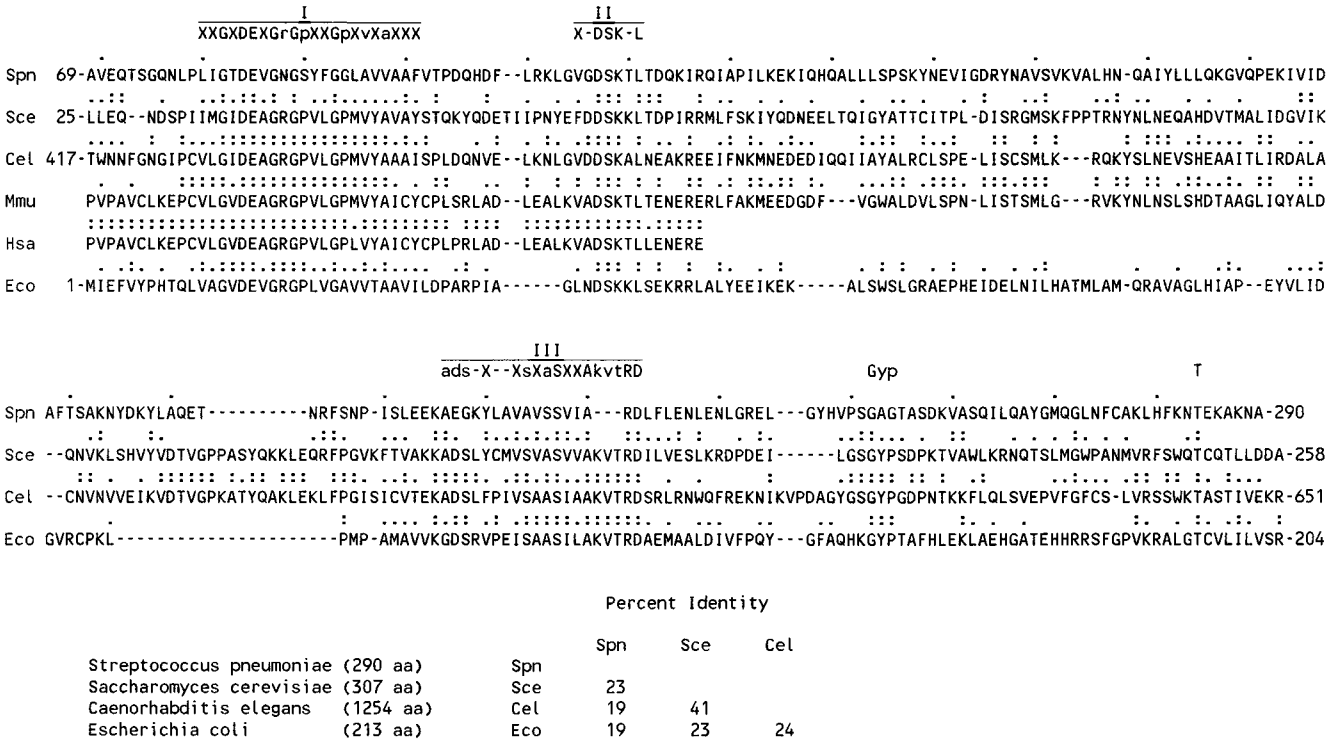


FIG. 5. Mapping of transcription start sites for mRNAs expressing *S. pneumoniae rnhB*. (A and B) Mapping of the 5′ endpoint by primer extension. Lanes G, A, T, and C, DNA sequencing ladder primed by Oligo A, one of the two oligonucleotides used for primer extension. RNA samples: lane 1, *E. coli* DH5α; lane 2, DH5α[pYZ62]; lane 3, DH5α[pYZ63]; lane 4, *S. pneumoniae* 216. Arrows mark primer extension products with Oligo A corresponding to mRNA1 and mRNA2. Base positions in the sequencing ladder are indicated for extension products of both Oligo A and Oligo B. (A) Experimental conditions as described in Materials and Methods. (B) Like panel A except that 40 μg of RNA was used for primer extension, and two different RNA preparations from *S. pneumoniae* were used. (C) Sequence upstream from *rnhB* gene showing transcription and translation start sites and the two promoters with bases underlined. Positions of the 5′-labeled oligonucleotides used for primer extension are shown downstream from the translation start codon (M). Vertical arrows indicate transcription start sites; horizontal arrows show the direction of transcription.

RNase HII protein was aligned with its larger pneumococcal counterpart, 19% of the residues were identical. Furthermore, as indicated below, the two proteins shared several amino acid sequence motifs that were also present in other putative RNase HII-encoding genes. Consequently, we shall refer to the pneumococcal gene as *rnhB* and to its product as RNase HII.

Transcription of the *S. pneumoniae rnhB* gene. Mapping of the 5′ end of the *rnhB* mRNA by primer extension revealed two transcription start sites (Fig. 5). In the absence of the *S. pneumoniae rnhB* gene (Fig. 5A, lane 1) no product is observed. With the cloned gene in either orientation in the *E. coli* plasmid, a strong band is seen at the position marked mRNA1; a weaker band (mRNA2) extends 58 nt further (Fig. 5A, lanes 2 and 3). The presence of both bands with either orientation of the pneumococcal insert indicates that both of the promoters are located within the *S. pneumoniae* DNA insert of pYZ62 and pYZ63, as was deduced above from the functional complementation. In *S. pneumoniae*, mRNA1 is produced from *rnhB* in its native location on the chromosome, albeit in lesser amount (approximately 1/15 of the amount from the multicopy plasmid in *E. coli*) from the single copy of the gene (Fig. 5A, lane 4). The minor band corresponding to mRNA2 is also evident. Both mRNAs from *S. pneumoniae* cells are more clearly seen in Fig. 5B, where larger amounts of two different preparations of RNA from *S. pneumoniae* were analyzed.

Although the primer extension bands in the heavily exposed autoradiograms of Fig. 5 are broad, lesser exposures and, particularly, PhosphorImager results gave sharper bands (data not shown) and allowed unambiguous assignment of lengths to base positions in the sequencing ladder of Fig. 5A. Start sites determined by extensions of Oligo A were read directly from the sequencing ladder made with the same oligonucleotide.



Percent Identity

		Spn	Sce	Cel
Streptococcus pneumoniae	(290 aa)	Spn		
Saccharomyces cerevisiae	(307 aa)	Sce	23	
Caenorhabditis elegans	(1254 aa)	Cel	19	41
Escherichia coli	(213 aa)	Eco	19	23	24

FIG. 6. Sequence comparison of *S. pneumoniae* RNase HII and homologous proteins. Numbers indicate positions in protein sequence with 1 corresponding to the N terminus. Dashes within the sequences indicate gaps giving optimal alignment; dashes at the ends indicate that the sequence continues after the numbered residue. Demonstrated RNase H enzymes: Spn, *S. pneumoniae*; Eco, *E. coli* (12). Putative RNase H enzymes: Sce, *S. cerevisiae* (GenBank accession no. Z71348, open reading frame YNL072w); Cel, *C. elegans* (GenBank accession no. Z66524, product T13H5.2). Putative RNase H fragments from expressed sequence tags: Mmu, *Mus musculus* (GenBank accession no. W11783, W71720, and AA050283 showed good agreement; W11783 was used for protein prediction); Hsa, *Homo sapiens* (GenBank accession no. W05602 and H43540; these sequences, from fetal lung and breast, respectively, showed numerous differences possibly due to errors in reverse transcription and sequence determination; the predicted protein was obtained after insertion of T residues after nt 124, 169, and 174 and removal of an A residue at nt 280 of sequence no. W05602). Symbols: :, identical residues; ., similar residues based on these groups: A, G, P; D, E, N, Q; A, C, F, I, L, M, T, V, W, Y; H, K, R; S, T. Dots above the Spn sequence indicate every 10th residue in Spn. Motifs of conserved residues are indicated by overlines with roman numerals. Residues conserved in all examples are shown by uppercase or, in all but one, by lowercase; X, represents hydrophobic residue A, C, F, I, L, M, T, V, W, or Y. Percent identity is shown for pairwise comparisons of the full proteins and refers to the smaller protein compared. aa, amino acids.

For extensions of Oligo B, start sites were calculated by determining the corresponding length from the sequencing ladder and measuring that length from the start of Oligo B. Both primers gave identical start sites for each of the two mRNAs.

The positions of the transcription start sites relative to the oligonucleotides used to determine them, their promoters, and the translation start site are shown in Fig. 5C. The major start site at A_{117} corresponds to the promoter immediately upstream from the coding region of the gene. Therefore, the major transcript, mRNA1, begins right at the translation start site. The other transcript, mRNA2, is produced in one-fourth the amount of mRNA1, as determined from radioactivity measurements of the bands shown in Fig. 5A. Its start site at A_{59} corresponds to the promoter at bp 24 to 52.

Common sequence motifs in RNase HII homologs. The protein sequence of *S. pneumoniae* RNase HII is compared in Fig. 6 with its *E. coli* homolog and with the products of two eucaryotic gene sequences from the GenBank database in which we found homologous sequence motifs and which we present as putative RNase HII-encoding genes. In addition to these sequences, genomic sequencing of various bacterial species has revealed homologs of *E. coli* RNase HII, to which RNase H function has been putatively ascribed (GenBank accession numbers listed above). These homologs also carry the conserved motifs (results not shown). One of the eucaryotic sequences shown in Fig. 6 (Sce) is from a yeast, *S. cerevisiae* (GenBank accession no. Z71348). This predicted protein contains 307 amino acid residues and is similar in size to the pneumococcal enzyme. The other sequence is from a roundworm, *Caenorhabditis elegans* (GenBank accession no. Z66524). The predicted protein in this case is large, 1,254 amino acid residues, but the region with RNase HII motifs corresponds to a central portion containing 235 residues, and it is quite similar to the yeast sequence, with 41% identity.

Three motifs are common to the RNase HII protein sequences. Motif I, the longest, contains 13 hydrophobic, 1 aspartyl, 1 glutamyl, and 4 glycyl residues. Motif II contains an invariant DSK sequence. Motif III contains five hydrophobic, two dicarboxylic, two basic, and six small (A, G, or S) residues. Some of the positively and negatively charged residues may be important for the hydrolytic mechanism; the conserved small and hydrophobic residues may reflect requirements for structural integrity in the polypeptide segments corresponding to the motifs.

By using the sequences in motifs I and II to probe the GenBank database, we found mouse- and human-expressed sequence tags that encode proteins containing these motifs. They are compared to the demonstrated RNase HII proteins from *S. pneumoniae* and *E. coli*, as well as to the putative RNase HII protein sequences predicted from *S. cerevisiae* and *C. elegans*, in Fig. 6. Only fragments of the genes appear to be present, and so we do not know the sizes of the proteins. Despite their fragmentary nature, the mammalian sequences clearly show homology to the other eucaryotic sequences even

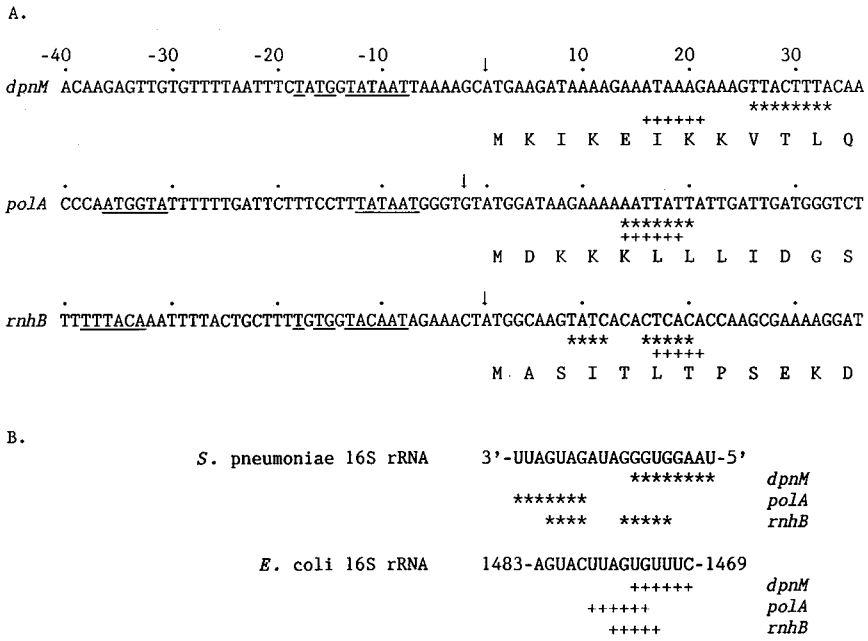


FIG. 7. Gene transcripts of *S. pneumoniae* beginning at or near translation start sites. (A) DNA segments containing the start sites for the *dpnM* (32), *polA* (22), and *rnhB* genes are shown with the first 12 amino acid residues of the proteins. Numbering is relative to the translation start site, designated +1. Promoters are underlined, and transcription start sites are indicated by arrows. Putative downstream ribosome binding sequences complementary to the *E. coli* 16S rRNA from nt 1469 to 1483 are indicated by plus signs; sequences complementary to a similarly located region in *S. pneumoniae* 16S rRNA are indicated by asterisks. Complementing base pairs include A:U, G:C, and G:U. (B) Segments of the 16S rRNAs and the portions complementary to the putative downstream binding sites. Symbols are as in panel A.

outside the motifs. The widespread occurrence of RNase HII suggests that it may be universally present in living cells.

DISCUSSION

The major transcript encoding *rnhB* in *S. pneumoniae* begins at a promoter with an extended −10 box, with the extension consensus T-TG- preceding TAcAAT at the −10 site. It is another example of a promoter with such an extended −10 site, which is prevalent in *S. pneumoniae*. Over 60% of promoters in this species exhibit this extension (32). A minor transcript encoding *rnhB* begins further upstream at a promoter with an ordinary −10 site. This promoter overlaps a putative promoter in the opposite direction, and the opposing promoters may be subject to coordinate regulation. Although the function of the opposing operon is unknown, it could conceivably be related to the *rnhB* function, and this may explain the existence of the minor promoter. The major transcript has the unusual property of beginning right at the translation start site for *rnhB*.

The coincidence of transcription and translation start sites is the third such case found in *S. pneumoniae*. The other two instances, in the *dpnM* (32) and *polA* (22) genes, are compared to *rnhB* in Fig. 7. In all three cases, there is no Shine-Dalgarno sequence upstream from the coding region. Normally, *S. pneumoniae* genes are associated with upstream ribosome binding sites complementary to the 3′ end of 16S rRNA. Although rare in *E. coli*, the transcription of genes without a 5′ leader in the mRNA has been observed a number of times in other bacteria, particularly in the streptomycetes (14). One well-studied instance is that of the repressor gene of phage lambda, in which one of its two promoters initiates transcription at the translation start site (27). In this case, evidence has been presented that a sequence downstream from the translation start plays a role in ribosome binding (36). More generally, it has been proposed that part of the region in *E. coli* 16S rRNA from nt 1483 to 1469 can bind to mRNA between nt 10 and 40 within the coding region (38).

Potential downstream binding sequences can be found within the first 9 to 33 nt of the mRNA of all three cases at hand (Fig. 7). The observation for *dpnM* was previously reported (32). Binding sites in this region are present for both *E. coli* (38) and *S. pneumoniae* (1) 16S rRNAs, which differ in 6 of 15 bases in their corresponding downstream binding boxes (Fig. 7). In the case of *S. pneumoniae*, the putative downstream binding box was extended to 19 nt. Complementing sequences include G:U as well as A:U and G:C base pairs. Unbroken complementing sequences range from 5 to 8 bp in length (Fig. 7). Five of the six matches occur between nt 16 and 22 of the mRNA. Similarly, five of six matches occurred within the 10 nt of the rRNA box at its 5′ end. Presumably, these optimal positions are determined by the distance on the ribosome between the junction of the 16S rRNA binding box and the downstream mRNA binding sequence and the point of juxtaposition of the 5′ end of the mRNA and the initiation site on the ribosome. The presence of binding sites in the mRNAs for both *S. pneumoniae* and *E. coli* ribosomes is consistent with the observation that all three proteins are expressed by the same signals in both species (references 22 and 32 and this work).

Three main motifs are evident in RNase HII (Fig. 6). It should be of interest to test particular residues in these motifs for enzymatic function. There may be separate sites for binding to DNA and for recognition of RNA in the duplex nucleic acid. The hydrolytic mechanism remains to be determined. In the case of RNase HI, there is a hydrolytic triad composed of one glutamyl and two aspartyl residues (43), two of which are believed to bind the Mg^{2+} cofactor (17). Only one motif of RNase HII resembles any of the four motifs recognized in RNase HI (2). That is motif II, which shares the *E. coli* RNase HI Asp_{70}-Ser_{71} sequence that is invariant in RNases HI. A

hydrolytic mechanism in which two carboxyl residues bind to the required Mg^{2+} ion is also plausible for RNase HII, inasmuch as the conserved motifs contain one glutamyl residue and three aspartyl residues.

It is not known why many species produce both RNase HI and RNase HII. Although it has not been established that the two types differ in substrate specificity, it is possible that RNase HII can cleave at single ribonucleotides, whereas RNase HI cleaves only at runs of four or more. The case of *S. pneumoniae* appears to be the first in which only RNase HII is present and it may be that RNase HI is the more dispensable of the two enzymes. It is possible that RNase HII is universally present in cells. Thus, evidence of the expression of such an enzyme in human and other mammalian cells was presented above (Fig. 6).

Either RNase HI or RNase HII may carry out the important DNA replication step of primer removal in Okazaki fragments. In bacteria, the 5′-to-3′ exonuclease domain of PolA (22), or separately encoded bacteriophage proteins corresponding to this domain (22), can also catalyze the reaction, thereby providing triple redundancy for this important function. In the introduction, we hypothesized that RNase HII might function in repair of newly synthesized DNA by removal of ribonucleotides inserted adventitiously into the nascent strand. We have not yet tested the newly identified *S. pneumoniae* enzyme for the appropriate specificity, that is, cleavage at a single ribonucleotide in DNA, and the possibility of such function remains open. If indeed, the RNase HII shows this specificity, it could well play a role in strand targeting in DNA mismatch repair. In this case, its role would be to produce single-strand breaks in the continuously synthesized strand of nascent DNA. These possibilities remain to be tested.

Whatever may be the natural functions of RNase HII, its universal presence in cells can be exploited in the use of antisense oligodeoxynucleotide drugs. On complexing with a target mRNA, such agents render the mRNA susceptible to cleavage in a specific manner that can be therapeutically beneficial (4). Greater understanding of the molecular structure of the RNase HII protein might lead to better use of this approach. The *S. pneumoniae* protein, 31.9 kDa, is of a reasonable size for structural determination, as are the proteins in other bacteria and, apparently, in yeast. Although its counterpart in *C. elegans* appears to be a domain in a much larger protein, this may be an anomaly of that invertebrate species. A similar case was observed in the folate biosynthetic system, present in *Pneumocystis carinii*, where four domains present as single proteins in other species (21) are joined together to make a single polyprotein (42). Accordingly, the polypeptide components of RNase HII in human cells may be similar in size to that in *S. pneumoniae*. This is suggested by the ~35-kDa subunit reported for at least one type of human RNase H (4, 8).

ACKNOWLEDGMENTS

We thank R. Crouch and S. Cerritelli for providing *E. coli* MIC3001.

This research was carried out under the auspices of the U.S. Department of Energy Office of Health and Environmental Research and supported by Public Health Service grant AI14885.

REFERENCES

1. **Bacot, C. M., and R. H. Reeves.** 1991. Novel tRNA gene organization in the 16S-23S intergenic spacer of the *Streptococcus pneumoniae* rRNA gene cluster. J. Bacteriol. **173:**4234–4236.
2. **Campbell, A. G., and D. S. Ray.** 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.
3. **Carl, P. L., L. Bloom, and R. J. Crouch.** 1980. Isolation and mapping of a mutation in *Escherichia coli* with altered levels of ribonuclease H. J. Bacteriol. **144:**28–35.
4. **Cazenave, C., P. Frank, J.-J. Toulmé, and W. Büsen.** 1994. Characterization and subcellular localization of ribonuclease H activities from *Xenopus laevis* oocytes. J. Biol. Chem. **269:**25185–25192.
5. **Chen, J.-D., and S. A. Lacks.** 1991. Role of uracil-DNA glycosylase in mutation avoidance by *Streptococcus pneumoniae*. J. Bacteriol. **173:**283–290.
6. **Claverys, J. P., and S. A. Lacks.** 1986. Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. Microbiol. Rev. **50:**133–165.
7. **Crouch, R. J., and M. L. Dirksen.** 1982. Ribonucleases H, p. 211–241. *In* S. M. Linn and R. J. Roberts (ed.), Nucleases. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. **Eder, P. S., and J. A. Walder.** 1991. Ribonuclease H from K562 human erythroleukemia cells: purification, characterization, and substrate specificity. J. Biol. Chem. **266:**6472–6479.
9. **Friedberg, E. C., G. C. Walker, and W. Siede.** 1995. DNA repair and mutagenesis. American Society for Microbiology, Washington, D.C.
10. **Henikoff, S.** 1987. Unidirectional digestion with exonuclease III in DNA sequence analysis. Methods Enzymol. **155:**156–165.
11. **Hogrefe, H. H., R. I. Hogrefe, R. Y. Walder, and J. A. Walder.** 1990. Kinetic analysis of *Escherichia coli* RNase H using DNA-RNA-DNA/DNA substrates. J. Biol. Chem. **265:**5561–5566.
12. **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.
13. **Itaya, M., D. McKelvin, S. K. Chatterjie, and R. J. Crouch.** 1991. Selective cloning of genes encoding RNase H from *Salmonella typhimurium*, *Saccharomyces cerevisiae*, and *Escherichia coli rnh* mutant. Mol. Gen. Genet. **227:** 438–445.
14. **Jannsen, G. R.** 1993. Eubacterial, archaebacterial, and eucaryotic genes that encode leaderless mRNA, p. 59–67. *In* R. H. Baltz, G. D. Hegeman, and P. L. Skatrud (ed.), Industrial microorganisms: basic and applied molecular genetics. American Society for Microbiology, Washington, D.C.
15. **Johnson, M. S., M. A. McClure, D.-F. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. Proc. Natl. Acad. Sci. USA **83:**7648–7652.
16. **Kanaya, S., and R. J. Crouch.** 1983. DNA sequence of the gene coding for *Escherichia coli* RNase H. J. Biol. Chem. **258:**1276–1281.
17. **Katayanagi, K., M. Okumura, and K. Morikawa.** 1993. Crystal structure of *Escherichia coli* RNase HI in complex with Mg^{2+} at 2.8 Å resolution: proof for a single Mg^{2+} binding site. Proteins Struct. Funct. Genet. **17:**337–346.
18. **Kogoma, T., N. L. Subia, and K. von Meyenburg.** 1985. Function of ribonuclease H in initiation of DNA replication in *Escherichia coli* K-12. Mol. Gen. Genet. **200:**103–109.
19. **Lacks, S.** 1966. Integration efficiency and genetic recombination in pneumococcal transformation. Genetics **53:**207–235.
20. **Lacks, S. A., and S. S. Springhorn.** 1980. Renaturation of enzymes after polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate. J. Biol. Chem. **255:**7467–7473.
21. **Lacks, S. A., B. Greenberg, and P. Lopez.** 1995. A cluster of four genes encoding enzymes for five steps in the folate biosynthetic pathway of *Streptococcus pneumoniae*. J. Bacteriol. **177:**66–74.
22. **Lopez, P., S. Martinez, A. Diaz, M. Espinosa, and S. A. Lacks.** 1989. Characterization of the *polA* gene of *Streptococcus pneumoniae* and comparison of the DNA polymerase I it encodes to homologous enzymes from *Escherichia coli* and phage T7. J. Biol. Chem. **264:**4255–4263.
23. **Milcarek, C., and B. Weiss.** 1972. Mutants of *Escherichia coli* with altered deoxyribonucleases. I. Isolation and characterization of mutants for exonuclease III. J. Mol. Biol. **68:**303–318.
24. **Mushegian, A. R., H. K. Edskes, and E. V. Koonin.** 1994. Eukaryotic RNAse H shares a conserved domain with caulimovirus proteins that facilitate translation of polycistronic RNA. Nucleic Acids Res. **22:**4163–4166.
25. **Ogawa, T., and T. Okazaki.** 1984. Function of RNase H in DNA replication revealed by RNase H defective mutants of *Escherichia coli*. Mol. Gen. Genet. **193:**231–237.
26. **Priebe, S. D., S. M. Hadi, B. Greenberg, and S. A. Lacks.** 1988. Nucleotide sequence of the *hexA* gene for DNA mismatch repair in *Streptococcus pneumoniae* and homology of *hexA* to *mutS* of *Escherichia coli* and *Salmonella typhimurium*. J. Bacteriol. **170:**190–196.
27. **Ptashne, M., K. Backman, M. Z. Humayun, A. Jeffrey, R. Maurer, B. Meyer, and R. T. Sauer.** 1976. Autoregulation and function of a repressor in bacteriophage lambda. Science **194:**156–161.
28. **Puyet, A., B. Greenberg, and S. A. Lacks.** 1989. The *exoA* gene of *Streptococcus pneumoniae* and its product, a DNA exonuclease with apurinic endonuclease activity. J. Bacteriol. **171:**2278–2286.
29. **Puyet, A., B. Greenberg, and S. A. Lacks.** 1990. Genetic and structural characterization of EndA, a membrane-bound nuclease required for transformation of *Streptococcus pneumoniae*. J. Mol. Biol. **171:**727–738.
30. **Rong, Y. W., and P. L. Carl.** 1990. On the molecular weight and subunit composition of calf thymus ribonuclease H1. Biochemistry **29:**383–389.
31. **Rosenthal, A. L., and S. A. Lacks.** 1977. Nuclease detection in SDS-poly-

acrylamide gel electrophoresis. Anal. Biochem. **80:**76–90.
32. **Sabelnikov, A. G., B. Greenberg, and S. A. Lacks.** 1995. An extended 10 promoter alone directs transcription of the *Dpn*II operon of *Streptococcus pneumoniae*. J. Mol. Biol. **250:**144–155.
33. **Sambrook, J., T. Maniatis, and E. F. Fritsch.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
34. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
35. **Sarngadharan, M. G., J. P. Leis, and R. C. Gallo.** 1975. Isolation and characterization of a ribonuclease from human leukemic blood cells specific for ribonucleic acid of ribonucleic acid-deoxyribonucleic acid hybrid molecules. J. Biol. Chem. **250:**365–373.
36. **Shean, C. S., and M. E. Gottesman.** 1992. Translation of the prophage λ *cI* transcript. Cell **70:**513–522.
37. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254:**34–38.
38. **Sprengart, M. L., H. P. Fatscher, and E. Fuchs.** 1990. The initiation of translation in *E. coli*: apparent base pairing between the 16s rRNA and downstream sequences of the mRNA. Nucleic Acids Res. **18:**1719–1723.
39. **Stein, H., and P. Hausen.** 1969. Enzyme from calf thymus degrading the RNA moiety of DNA-RNA hybrids: effect on DNA dependent RNA polymerase. Science **166:**393–395.
40. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185:**60–89.
41. **Sutcliffe, J. G.** 1978. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **43:**77–90.
42. **Volpe, F., M. Dyer, J. G. Scaife, G. Darby, D. K. Stammers, and C. J. Delves.** 1992. The multifunctional folic acid synthesis *fas* gene of *Pneumocystis carinii* appears to encode dihydropteroate synthase and hydroxymethyldihydropterin pyrophosphokinase. Gene **112:**213–218.
43. **Yang, W., W. A. Hendrickson, R. J. Crouch, and Y. Satow.** 1990. Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. Science **249:**1398–1405.
44. **Zhang, Y.-B., and S. A. Lacks.** Unpublished data.
45. **Zhang, Y.-B., B. Greenberg, and S. A. Lacks.** Unpublished data.