Identification of the Genes Encoding Mn^{2+}-Dependent RNase HII and Mg^{2+}-Dependent RNase HIII from *Bacillus subtilis*: Classification of RNases H into Three Families†

Naoto Ohtani,‡ Mitsuru Haruki,‡ Masaaki Morikawa,‡ Robert J. Crouch,§ Mitsuhiro Itaya,|| and Shigenori Kanaya*,‡

Department of Material and Life Science, Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka 565-0871, Japan, Laboratory of Molecular Genetics, NICHD, National Institutes of Health, Bethesda, Maryland 20892, and Mitsubishi-Kasei Institute of Life Sciences, 11 Minamiooya, Machida, Tokyo 194-8511, Japan

Received September 14, 1998; Revised Manuscript Received November 5, 1998

ABSTRACT: Database searches indicated that the genome of *Bacillus subtilis* contains three different genes encoding RNase H homologues. The *ypdQ* gene encodes an RNase HI homologue with 132 amino acid residues, whereas the *rnh* and *ysgB* genes encode RNase HII homologues with 255 and 313 amino acid residues, respectively. RNases HI and HII show no significant sequence similarity. These genes were individually expressed in *Escherichia coli*; the recombinant proteins were purified, and their enzymatic properties were compared with those of *E. coli* RNases HI and HII. We found that the *ypdQ* gene product showed no RNase H activity. The 2.2 kb pair genomic DNA containing this gene did not suppress the RNase H deficiency of an *E. coli rnhA* mutant, indicating that this gene product shows no RNase H activity in vivo as well. In contrast, the *rnh* (*rnhB*) gene product (RNase HII) showed a preference for Mn^{2+}, as did *E. coli* RNase HII, whereas the *ysgB* (*rnhC*) gene product (RNase HIII) exhibited a Mg^{2+}-dependent RNase H activity. Oligomeric substrates digested with these enzymes indicate similar recognition of these substrates by *B. subtilis* and *E. coli* RNases HII. Likewise, *B. subtilis* RNase HIII and *E. coli* RNase HI have generated similar products. These results suggest that *B. subtilis* RNases HII and HIII may be functionally similar to *E. coli* RNases HII and HI, respectively. We propose that Mn^{2+}-dependent RNase HII is universally present in various organisms and Mg^{2+}-dependent RNase HIII, which may have evolved from RNase HII, functions as a substitute for RNase HI.

As is the case for many proteins, the ribonucleases H (RNases H)[1] of *Escherichia coli* provide a starting point for comparison of these enzymes, which specifically degrade the RNA of RNA−DNA hybrids (*1*), with similar proteins from other sources. *E. coli* RNase HI is a protein comprising 155 amino acids (*2*) that has been extensively studied (*3*−*5*). Its structure has been determined, and amino acids involved in catalysis are known; two models for the mechanism of action have been promulgated, extensive kinetic analyses performed, and its sites of cleavage of RNA−DNA hybrids determined. In many cases, proteins and genes related to *E. coli* RNase HI have also been described in widely divergent organisms. For example, an *rnhA* homologue or the gene encoding a multidomain protein, in which the *rnhA*-like segment encodes one of these domains, has been found in ∼20 organisms (Table 1). Nevertheless, the physiological function(s) of these enzymes is not yet completely understood, although mutations in some genes related to recombination, DNA replication, and repair in combination with *rnhA* mutant genes suggest that RNase HI may be involved in these processes. RNase H activity of retroviral reverse transcriptases (RTs) is required for the conversion of genomic RNA to double-stranded DNA, and these enzymes are, therefore, essential for the proliferation of retroviruses.

The second RNase H (RNase HII) of *E. coli* was discovered on the basis of its ability to support growth of *E. coli* strains defective in RNase HI (*6*). Other than the fact that the protein has a molecular mass of 23 kDa and can degrade RNA of RNA−DNA hybrids, little is known about any distinctive enzymatic properties or the secondary and tertiary structures or types of amino acid residues involved in catalytic function and substrate binding. Likewise, no physiological function has been assigned to RNase HII. However, an *rnhB* homologue or the gene encoding a multidomain protein, in which the *rnhB*-like segment encodes one of these domains, has been found in ∼25 organisms (Table 1).

Database searches, using programs such as BLAST, rely upon local sequence similarities to establish a relationship between two or more proteins or nucleic acids. RNases H vary in the degree of conservation, with the RNase H of rous sarcoma virus having but eight amino acids in common with *E. coli* RNase HI (*7*). Not only is the RSV RNase H

† This work was supported in part by a Grant-in-Aid for Scientific Research (08455382) from the Ministry of Education, Science and Culture of Japan.

* To whom correspondence and reprint requests should be addressed. Telephone or fax: 81-6-879-7938. E-mail: kanaya@chem.eng.osaka-u.ac.jp.

‡ Osaka University.

§ National Institutes of Health.

|| Mitsubishi-Kasei Institute of Life Sciences.

[1] Abbreviation: RNase H, ribonuclease H.

10.1021/bi982207z CCC: $18.00
Published on Web 12/16/1998

Table 1: List of the RNase H Homologues So Far Identified

accession number	size (no. of amino acid residues)	origin
RNase HI family		
Swiss-Prot P00647	155	*E. coli*
EMBL X57159	155	*Salmonella typhimurium*
Swiss-Prot P43807	154	*Haemophilus influenzae*
Swiss-Prot Q08885	161	*Buchnera aphidicola*
Swiss-Prot P56120	143	*Helicobacter pylori*
GenBank AF060218	156	*Zymomonas mobilis*
Swiss-Prot P29253	166	*Thermus thermophilus*
Swiss-Prot Q07705	159	*M. smegmatis*
DDBJ D64004	160	*Synechpcystis* sp. PCC6803
Swiss-Prot Q04740	348	*Sa. cerevisiae*
Swiss-Prot Q07762	494	*Crithidia fasciculata*
GenBank AF032921	333	*D. melanogaster*
Swiss-Prot D26340	293	*Gallus gallus*
GenBank AF048993	285	*Mus musculus*
GenBank AF048994	286	*H. sapiens*
GenBank AF039652	286	*H. sapiens*
Swiss-Prot U53871	560	HIV-1[a]
RNase HII family		
Swiss-Prot P10442	213	*E. coli*
Swiss-Prot P43808	197	*H. influenzae*
Swiss-Prot P56121	209	*He. pylori*
DDBJ D32253	201	*Magnetospirillum* sp.
GenBank AF054610	220	*Brucella melitensis*
Swiss-Prot P52975	192	*Caulobacter crescentus*
GenBank AE001118	181	*Borrelia burgdorferi*
Swiss-Prot Q10793	264	*M. tuberculosis*
EMBL Z97369	240	*Mycobacterium leprae*
EMBL AL022374	233	*Streptomyces coelicolor*
ORF B. subtilis rnh	255	*B. subtilis*
DDBJ D90899	190	*Synechocystis* sp. PCC6803
GenBank U93576	290	*S. pneumoniae*
ORF A. aeolicus rnh	196	*Aquifex aeolicus*
GenBank U67470	230	*Methanococcus jannaschii*
ORF MTH1023	206	*Methanobacterium thermoautotrophicum*
ORF PH1650	220	*Pyrococcus horikoshii*
ORF AF0621	205	*Archaeoglobus fulgidus*
EMBL Z71348	306	*Sa. cerevisiae*
Swiss-Prot Q10236	326	*Schizosaccharomyces pombe*
EMBL Z66524	1254	*Caenorhabditis elegans*
EMBL Z97029	299	*H. sapiens*

[a] The RNase H domain of HIV-1 RT is listed as a representative of the RNase H domains of various retroviral RTs, because multiple alignment has already been carried out for these sequences by Doolittle et al. (*43*).

undetected in BLAST searches but the RNases H of TY elements of *Saccharomyces cerevisiae* are also so distantly related to most RNases HI that they too are not selected by the BLAST search. Nevertheless, such searches can prove to be useful for finding RNase H-related genes and proteins. In the three kingdoms, bacteria, eukarya, and archaea (*8*, *9*), database searches detect RNase HII but not RNase HI in the genomes of archaea, whereas *rnhB* genes are found in the genomes of the organisms of representatives of all three kingdoms. In addition, it has been suggested that Gram-positive bacterium *Streptococcus pneumoniae* contains only the *rnhB* gene (*10*). Likewise, determination of the complete genome sequences of the Gram-positive bacterium *Bacillus subtilis* (*11*) and Gram-negative bacterium *Aquifex aeolicus* (*12*) annotate a single *rnh* gene, which is homologous to the *rnhB* gene.

Because of the presence of both *rnhA* and *rnhB* in bacterial and eukaryotic organisms, it is curious that the *rnhA* genes are not recognizable or absent from some genomes of bacteria and archaea. Perhaps, a gene exhibiting weak sequence similarity to the *rnhA* gene may be present in these genomes. Therefore, we have decided to examine whether an *rnh* gene of *B. subtilis* encodes such a functional enzyme and whether the *B. subtilis* genome contains additional *rnh* genes.

Here, we report that the *B. subtilis* genome contains three *rnh* genes, the *ypdQ* gene of *B. subtilis*, an *rnhA* homologue, an *rnhB* gene, encoding RNase HII, and *rnhC*, encoding RNase HIII. The amino acid sequences of RNase HII and RNase HIII are similar to that of the *E. coli* RNase HII and to each other, but are not significantly similar to that of *E. coli* RNase HI. We also show that the protein expressed in *E. coli* from the *rnhA* homologue has no RNase H activity. *B. subtilis* RNases HII and HIII are similar to *E. coli* RNases HII and HI, respectively, in isoelectric points, divalent cation preference, and sequence specificity for the hydrolysis of oligomeric substrates. We propose that *B. subtilis* RNases HII and HIII function as substitutes for *E. coli* RNases HII and HI, respectively.

EXPERIMENTAL PROCEDURES

Database Search and Sequence Analysis. E. coli RNases HI and HII and *S. pneumoniae* RNase HII were used as query sequences in the *B. subtilis* genome database searches performed with a program called BLASTP. The *B. subtilis* genome database is the BSORF site (http://bacillus.tokyo-center.genome.ad.jp:8008/BSORF-DB.html). Amino acid sequence analyses, including determination of molecular masses and multiple alignment, were performed using DNASIS software (Hitachi Software). For the phylogenetic analyses of the RNase HII and HIII sequences, the multiple alignment was made by using a program called CLUSTAL W (*13*) with default parameters. The phylogenetic tree was constructed by using a program called TreeView (*14*) based on the distance matrix produced by CLUSTAL W.

The accession numbers of the RNase H sequences examined in this study in Database are listed in Table 1.

Cells and Plasmids. The *B. subtilis* strain MI113 was previously isolated (*15*). The *rnhA* mutant *E. coli* strains MIC3009 [F^-, *sup*E44, *sup*F58, *lac*Y1 or Δ(*lac*IZY)6, *trp*R55, *gal*K2, *gal*T22, *met*B1, *hsd*R14(r_K^-, m_K^+), *rnh*-339:: cat], MIC3001 [MIC3009 plus *recB270*] (*16*), and MIC1066 [MIC1007 (*16*) and λ*DE3*] (*17*) have been described. *E. coli* HMS174(DE3)pLysS [F^-, *recA1*, *hsd*R(r_{K12}^-, m_{K12}^+), *Rif*^{R}-(DE3), pLysS] and the plasmid pET-3a were from Novagen. Plasmid pMIC2721 for overproduction of *E. coli* RNase HII (*6*) and the plasmid pBR860 for overproduction of *E. coli* RNase HI (*18*) were previously constructed. Plasmid pJLA503 was from Medac Gentechnologie. *E. coli* MIC3009 and MIC1066 transformants were grown in Luria-Bertani medium (*19*) containing 50 mg/L ampicillin. *E. coli* HMS174-(DE3)pLysS transformants were grown in NZCYM medium (Novagen) containing 50 mg/L ampicillin.

Materials. [γ-^{32}P]ATP (>5000 Ci/mmol) and [α-^{32}P]-ddATP (>5000 Ci/mmol) were obtained from Amersham. *Crotalus durissus* phosphodiesterase was from Boehringer Mannheim. *E. coli* RNase HI was previously purified (*20*). Restriction and modifying enzymes were from either Toyobo Co., Ltd., or Takara Shuzo Co., Ltd.

General DNA Manipulations. Genomic DNA was prepared from a Sarkosyl lysate of *B. subtilis* cells (*15*) and used as a template to amplify the genes encoding RNase H homologues by PCR. Digestion of DNA with restriction enzymes,

Table 2: PCR Primers[a]

plasmid	primer	sequence	restriction site
pBR2.2	5′-primer	5′-TTCGCTAAAAGCATGCAATGAGACGGATCCCTGTCC-3′	*Sph*I
	3′-primer	5′-AGCGGGGTGCTCGCGACGACGAACGGCGTCAACGCC-3′	*Nru*I
pBR800es	5′-primer	5′-TTCAAGAATTCTCATGTTTGAC-3′	*Eco*RI
	3′-primer	5′-GCGGTCGACCCTCGCCATTAGGATGAACC-3′	*Sal*I
	5′-fusion primer	5′-TCTACCAGAGATG*CCTACAGAAATATATGT*-3′	
	3′-fusion primer	5′-*CTGTAGG*CATCTCTGGTAGACTTCCTGTAA-3′	
pJAL500su	5′-primer	5′-GGAGTTCATACATATGCCTACAGAAATATATGTAGA-3′	*Nde*I
	3′-primer	5′-GCGGTCGACCACCGCCGGTCAGCCGGCCT-3′	*Sal*I
pJAL900su	5′-primer	5′-GGGAGAAGAGCATATGAATACATTAACCGTAAAGGAC-3′	*Nde*I
	3′-primer	5′-GGATCCGAATTCTGCGGGACGGCTGTACGCCCG-3′	*Eco*RI
pET1000su	5′-primer	5′-AGGAGATTATCATATGTCCCATTCAGTGATAAAAG-3′	*Nde*I
	3′-primer	5′-CAAGGATCCGAATTCGTTAGAAGACACATATCAAGC-3′	*Bam*HI

[a] For 5′- and 3′-primers, underlined bases show the positions of the restriction sites. For the 5′- and 3′-fusion primers, the underlined bases show the position of the codon for the initial methionine residue, and the italic residues represent those of the *ypdQ* gene.

analysis of DNA fragments by agarose gel electrophoresis, transformation with calcium chloride procedure, and small-scale preparation of plasmid DNA with the alkaline lysis method were performed as described previously (*21*). Large-scale plasmid DNA preparation was performed by using a Qiagen plasmid kit (Qiagen). PCR was performed in 25 cycles with GeneAmp PCR system 2400 (Perkin-Elmer) using a Gene Amp kit (Takara Shuzo Co., Ltd.) according to the procedure recommended by the supplier. All oligodeoxyribonucleotides were synthesized by Sawady Technology Co., Ltd.

Construction of Plasmids for Complementation Experiments. Plasmid pBR2.2 was constructed by ligating the 2.2 kb pair DNA fragment containing the *ypdQ* gene, which was amplified by PCR with the primers listed in Table 2, to the *Sph*I−*Nru*I site of pBR322. This 2.2 kb pair DNA fragment contains the 800 and 1000 bp DNA fragments, which are located upstream and downstream of the *ypdQ* gene as well. In plasmid pBR2.2, transcription and translation of the *ypdQ* gene are probably controlled by its own promoter and Shine-Dalgarno (SD) sequence. Plasmid pBR800es, in which the transcription and translation of the *ypdQ* gene are controlled by the promoter and the SD sequence of the *E. coli rnhA* gene, was constructed by performing PCR twice with the primers listed in Table 2. In the first PCR, the 400 bp DNA fragment containing the promoter and the SD sequence of the *E. coli rnhA* gene was amplified with the 5′-primer and 3′-fusion primer by using pBR860 as a template. Likewise, the 400 bp DNA fragment containing the entire *ypdQ* gene was amplified with the 5′-fusion primer and 3′-primer by using the *B. subtilis* genome as a template. These two 400 bp DNA fragments were mixed and amplified with 5′- and 3′-primers. The resultant 800 bp DNA fragment was ligated to the *Eco*RI−*Sal*I site of pBR322.

Overproduction and Purification of a B. subtilis RNase HI Homologue. Plasmid pJAL500su was constructed by ligating the 500 bp DNA fragment containing the *ypdQ* gene, which was amplified by PCR with the primers listed in Table 2, to the *Nde*I−*Sal*I site of pJLA503. The overproducing strain was constructed by transforming *E. coli* MIC3009 with pJAL500su. Overproduction was performed as described previously for *E. coli* RNase HI (*20*). Cells were harvested by centrifugation at 6000*g* for 10 min, and the enzyme was purified at 4 °C. Cells were suspended in 10 mM Tris-HCl (pH 7.5) containing 1 mM EDTA (TE buffer), disrupted by sonication with a model 450 sonifier from Branson Ultrasonic Corp., and centrifuged at 15000*g* for 30 min. The supernatant was applied to a column (1 mL) of DE-52 (Whatman) equilibrated with 10 mM Tris-HCl (pH 7.5). The flow-through fraction containing the protein was applied to a gel filtration column (1.6 cm × 60 cm) of HiLoad 16/60 Superdex 200 pg (Pharmacia LKB Biotechnology) equilibrated with 10 mM Tris-HCl (pH 7.5) containing 150 mM NaCl. Fractions containing the pure protein were combined and used for further characterizations.

Overproduction and Purification of B. subtilis RNase HII. Plasmid pJAL900su was constructed by ligating the 900 bp DNA fragment containing the *rnhB* gene, which was amplified by PCR with the primers listed in Table 2, to the *Nde*I−*Eco*RI site of pJLA503. The overproducing strain was generated by transforming *E. coli* MIC3009 with pJAL900su. Construction of an overproducing strain, overproduction, and sonication lysis of the cells were performed as described for MIC3009/pJAL500su. The following purification procedures were carried out at 4 °C. RNase HII was precipitated at a 70% saturation level with ammonium sulfate to remove nucleic acids from the supernatant obtained after sonication lysis. The precipitates were collected by centrifugation at 10000*g* for 20 min, dissolved in 10 mM Tris-HCl (pH 8.0), and applied to a column (1 mL) of DE-52 equilibrated with the same buffer. The enzyme was eluted from the column by linearly increasing the NaCl concentration from 0 to 0.5 M. Fractions containing the enzyme were combined and applied to a column of HiLoad 16/60 Superdex 200 pg equilibrated with 10 mM Tris-HCl (pH 7.5) containing 150 mM NaCl. Fractions containing the pure enzyme were combined and used for further analysis.

Overproduction and Purification of B. subtilis RNase HIII. Plasmid pET1000su was constructed by ligating the 1000 bp DNA fragment containing the *ysgB* (*rnhC*) gene, which was amplified by PCR with the primers listed in Table 2, to the *Nde*I−*Bam*HI site of pET-3a. A strain from which the enzyme could be overproduced was constructed by transforming *E. coli* HMS174(DE3)pLysS with pET1000su. For overproduction, this transformant was grown at 37 °C. When the absorbance at 660 nm of the culture reached around 0.6, isopropyl β-D-thiogalactopyranoside (IPTG) was added to the culture medium (final concentration of 1 mM) and induction was continued for an additional 6 h. Cells were harvested by centrifugation at 6000*g* for 10 min. Protein purification procedures were carried out at 4 °C. Sonication lysis was performed as described for MIC3009/pJAL500su.

The supernatant obtained after sonication lysis was applied to a column (1 mL) of DE-52 equilibrated with 10 mM Tris-HCl (pH 7.5). The flow-through fraction containing the enzyme was applied to a column (1 mL) of P-11 (Whatman) equilibrated with 10 mM Tris-HCl (pH 7.5). The flow-through fraction containing the enzyme was applied to the column of HiLoad 16/60 Superdex 200 pg equilibrated with 10 mM Tris-HCl (pH 7.5) containing 150 mM NaCl. Fractions containing the pure enzyme were combined, and the enzyme was precipitated at a 70% saturation level with ammonium sulfate to remove nucleic acids. The precipitates were collected by centrifugation at 10000*g* for 20 min, dissolved in 10 mM Tris-HCl (pH 7.5), and dialyzed against the same buffer. This solution containing the pure enzyme was used for further characterizations.

Overproduction and Purification of E. coli RNase HII. An overproducing strain was constructed by transforming *E. coli* MIC1066 with pMIC2721. The enzyme was overproduced and purified as described previously (*6*). Because the solubility of the enzyme in a buffer, which not containing a detergent, is very low, the purified enzyme was stored in a denatured form in 20 mM Tris-HCl (pH 8.5) containing 4 M urea at 4 °C, until just prior to enzymatic activity analysis.

Protein Concentrations. The protein concentrations were determined from the extent of UV absorption with $A_{280}^{0.1\%}$ values of 0.57 for the *B. subtilis* RNase HI homologue, 0.93 for *B. subtilis* RNase HII, 0.62 for *B. subtilis* RNase HIII, 2.02 for *E. coli* RNase HI, and 0.56 for *E. coli* RNase HII. The value for *E. coli* RNase HI was experimentally determined (*22*). Other values were calculated by using ϵ values of 1576 M^{-1} cm^{-1} for Tyr and 5225 M^{-1} cm^{-1} for Trp at 280 nm (*23*).

Biochemical Characterizations. Purities of the enzymes were determined by SDS−PAGE on a 12% polyacrylamide gel (*24*), followed by staining with Coomassie Brilliant Blue. For estimation of the molecular mass of the protein by gel filtration, bovine serum albumin, ovalbumin, chymotrypsinogen, and ribonuclease A were used as standard proteins. The CD spectra were measured on a J-720W spectropolarimeter of Japan Spectroscopic Co., Ltd. The far-UV (200−260 nm) CD spectrum was obtained at 25 °C by using a solution containing the enzyme at 0.39 mg/mL in 10 mM Tris-HCl buffer (pH 7.5) containing 150 mM NaCl in a cell with an optical path length of 2 mm. The mean residue ellipticity, which has units of degrees per square centimeter per decamole, was calculated by using an average amino acid molecular weight of 110.

Thermal Denaturation. The thermal denaturation curves and the temperature of the midpoint of the transition (T_m) were determined as described previously (*25*) by monitoring the change in the CD value at 220 nm. The protein was dissolved in 10 mM Tris-HCl (pH 7.5) containing 150 mM NaCl and 2.0 M urea. To analyze the effect of an additive (NaCl and $MgCl_2$) on protein stability, thermal denaturation curves were measured in the absence and presence of 2.5 M NaCl and/or 50 mM $MgCl_2$. The protein concentration was approximately 0.1 mg/mL, and a cell with an optical path length of 2 mm was used.

Enzymatic Activity. The RNase H activity of each enzyme was determined at 30 °C in an appropriate buffer in the presence of 1 mM 2-mercaptoethanol (2-Me) and 50 μg/mL bovine serum albumin (BSA) by measuring the amount of radioactivity of the acid-soluble digestion product from the substrate, ^{3}H-labeled M13 DNA/RNA hybrid, as previously described (*26*). The buffers used for assays were as follows: *E. coli* RNase HI, 10 mM Tris-HCl (pH 8.0) containing 10 mM $MgCl_2$ and 50 mM NaCl; *E. coli* RNase HII, 10 mM Tris-HCl (pH 8.0) containing 10 mM $MnCl_2$ and 50 mM NaCl; *B. subtilis* RNase HII, 10 mM Tris-HCl (pH 8.5) containing 10 mM $MnCl_2$ and 50 mM KCl; and *B. subtilis* RNase HIII, 10 mM Tris-HCl (pH 8.5) containing 50 mM $MgCl_2$ and 100 mM NaCl. One unit is defined as the amount of enzyme producing 1 μmol of acid-soluble material per minute at 30 °C. The specific activity was defined as the enzymatic activity per milligram of protein. When the effect of divalent cations on the enzymatic activity was analyzed, RNase H activity was determined in 10 mM Tris-HCl (pH 8.0) containing 50 mM NaCl, 1 mM 2-Me, and 50 μg/mL BSA in the presence of 10 mM $MgCl_2$, $MnCl_2$, $BaCl_2$, $CaCl_2$, $CoCl_2$, $CuCl_2$, $FeCl_2$, $NiCl_2$, $SrCl_2$, or $ZnCl_2$. For the analysis of pH dependence, 10 mM BisTris-HCl (pH 5.7−7.1), 10 mM Tris-HCl (pH 7.1−8.8), or 10 mM glycine-NaOH (pH 8.3−10.0) was used as a buffer for assay procedures.

Cleavage of Oligomeric RNA−DNA and DNA-RNA-DNA−DNA Substrates. The 29 b DNA-RNA-DNA (5′-AATAGAGAAAAAGaaaaAAGATGGCAAAG-3′), in which DNA and RNA are represented by uppercase and lowercase letters, respectively, and the 29 b DNA, which is complementary to this 29 b DNA-RNA-DNA, were kindly donated by ID Biomedical Corp. The 12 b RNA (5′-cggagaugacgg-3′) was chemically synthesized by Toray Research Center Co., Ltd. The 12 b DNA, complementary to the 12 b RNA, and the 13 b DNA, complementary to the 5′-terminal sequence of the 29 b DNA-RNA-DNA, were synthesized by Sawady Technology Co., Ltd. The 12 b RNA was 5′-end-labeled, and the 29 b DNA-RNA-DNA was either 5′- or 3′-end-labeled with standard procedures. The hybrid duplexes (1.0 μM) were prepared by hybridizing either the end-labeled 12 b RNA or DNA-RNA-DNA with 1.5 molar equiv of DNA oligomers. Hydrolysis of the substrate was carried out under the same condition used to determine the enzymatic activity for the hydrolysis of the M13 DNA−RNA hybrid. Products were analyzed on a 20% polyacrylamide gel containing 7 M urea and quantified using the Instant Imager (Packard). When the hybrid duplex, in which RNA or DNA-RNA-DNA is 5′-end-labeled, was used as a substrate, products were identified by comparing their patterns of migration on the gel with those of the oligonucleotides generated by partial digestion of RNA or DNA-RNA-DNA with snake venom phosphodiesterase (*27*). When the hybrid duplex, in which DNA-RNA-DNA is 3′-end-labeled, was used as a substrate, products were identified by comparing their patterns of migration with that of the end-labeled DNA, which is complementary to the 3′-terminal sequence of DNA-RNA-DNA.

For construction of the 12 bp RNA−RNA duplex, another RNA, complementary to the 12 b RNA, was synthesized by Takara Shuzo Co., Ltd. The RNA−RNA duplex was prepared as described for the construction of the 12 bp RNA−DNA hybrid.

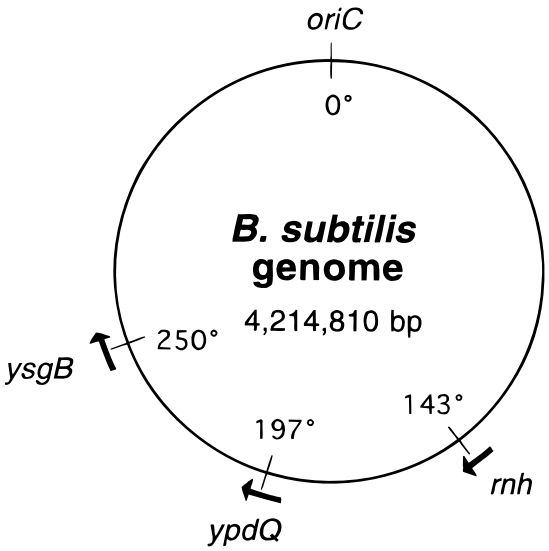

FIGURE 1: Location of the genes encoding the RNase HI and RNase HII homologues on the chromosomal map of *B. subtilis*. The *rnh* gene encodes RNase HII with 255 amino acid residues. The *ypdQ* and *ysgB* genes encode proteins with 132 and 313 amino acid residues, respectively. The latter two genes had previously been designated as having unknown functions. The locations of these genes on the *B. subtilis* chromosome are from 1 676 850 to 1 677 617 for *rnh*, from 2 309 661 to 2 310 059 for *ypdQ*, and from 2 925 133 to 2 926 074 for *ysgB*. An arrow shows the direction of the transcription for each gene.

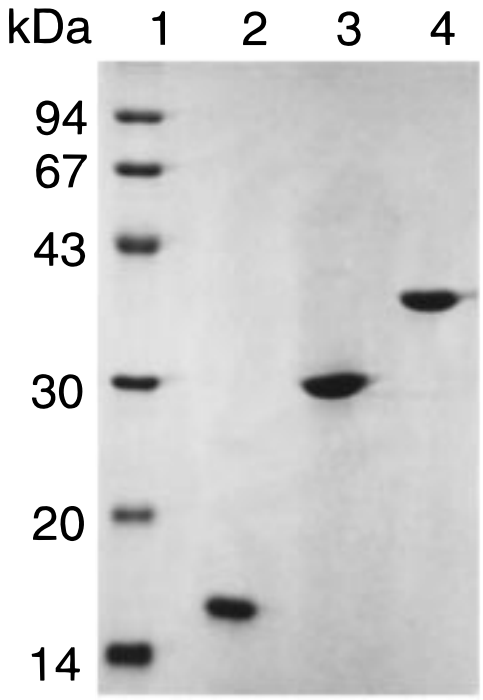

FIGURE 2: SDS−PAGE of the purified *B. subtilis* RNase H homologues. All recombinant proteins were purified as described in Experimental Procedures. Samples were subjected to 12% SDS−PAGE and stained with Coomassie Brilliant Blue: lane 1, a low-molecular weight marker kit (Pharmacia LKB Biotechnology) containing phosphorylase b, albumin, ovalbumin, carbonic anhydrase, trypsin inhibitor, and α-lactalbumin; lane 2, *B. subtilis* RNase HI homologue; lane 3, *B. subtilis* RNase HII; and lane 4, *B. subtilis* RNase HIII. Numbers along the gel represent the molecular weights of individual standard proteins.

RESULTS

Genes Encoding RNase HI and RNase HII Homologues from B. subtilis

When the *B. subtilis* genome was searched for genes encoding RNase H homologues by using *E. coli* RNases HI and HII as query sequences, only the *ypdQ* and *rnh* genes were found to encode RNase HI and RNase HII homologues, respectively. However, when *S. pneumoniae* RNase HII was used as a query sequence, the *ysgB* gene was found to encode another RNase HII homologue. The *ypdQ* and *ysgB* genes have been reported to encode proteins with unknown functions (*11*). The localization of these three genes in the *B. subtilis* genome is shown in Figure 1. An RNase HI homologue encoded by the *ypdQ* gene is composed of 132 amino acid residues with a calculated molecular weight of 14 670 and an isoelectric point (p*I*) of 5.85. An RNase HII homologue encoded by the *rnh* gene is composed of 255 amino acid residues with a calculated molecular weight of 28 322 and a p*I* of 5.62. Another RNase HII homologue encoded by the *ysgB* gene is composed of 313 amino acid residues with a calculated molecular weight of 34 037 and a p*I* of 10.47. As shown later, the protein encoded by the *ysgB* gene is an RNase H protein. Therefore, we will refer to it as RNase HIII and the gene encoding it as *rnhC*.

Overexpression and Purification. To facilitate the purification of the RNase HI homologue, RNase HII, and RNase HIII from *B. subtilis* in an amount sufficient for enzymatic and biochemical characterizations, overproducing strains were constructed as described in Experimental Procedures. Because the initiation codons of the genes encoding RNases HII and HIII in the *B. subtilis* genome are GTG, this codon was altered to ATG, when the plasmids for the overproduction of these proteins were constructed. Induction of the gene encoding either one of these proteins caused accumulation of the protein in a soluble form as the most abundant protein in cells. The production level of the protein was estimated to be roughly 30, 25, and 40 mg/L of culture for the RNase HI homologue, RNase HII, and RNase HIII, respectively. The amount of the protein purified from 1 L of culture was approximately 10, 5, and 6 mg for the RNase HI homologue, RNase HII, and RNase HIII, respectively. The molecular weight of the protein was estimated to be 15 000, 30 000, and 35 000 from SDS−PAGE (Figure 2) and 22 000, 36 000, and 40 000 from gel filtration column chromatography (data not shown) for the RNase HI homologue, RNase HII, and RNase HIII, respectively. Thus, all of these proteins exist in a monomeric form. Because of the marked differences in protein properties uncovered during this investigation, we have chosen to present the following data in separate sections for the RNase HI homologue and RNases HII and HIII of *B. subtilis*.

RNase HI Homologue

Amino Acid Sequence Comparison. An RNase HI homologue from *B. subtilis* shows significant amino acid sequence similarity to all of the RNase HI enzymes. For example, it shows sequence identities of 29% to *E. coli* RNase HI, 24% to *Thermus thermophilus* RNase HI, 27% to *Mycobacterium smegmatis* RNase HI, 26% to *Sa. cerevisiae* RNase HI, 22% to *Drosophila melanogaster* RNase HI, 21% to *Homo sapiens* RNase HI, and 19% to RNase H domain of HIV-1 RT. However, among five catalytically important residues, a histidine residue (His^{124} for *E. coli* RNase HI) is

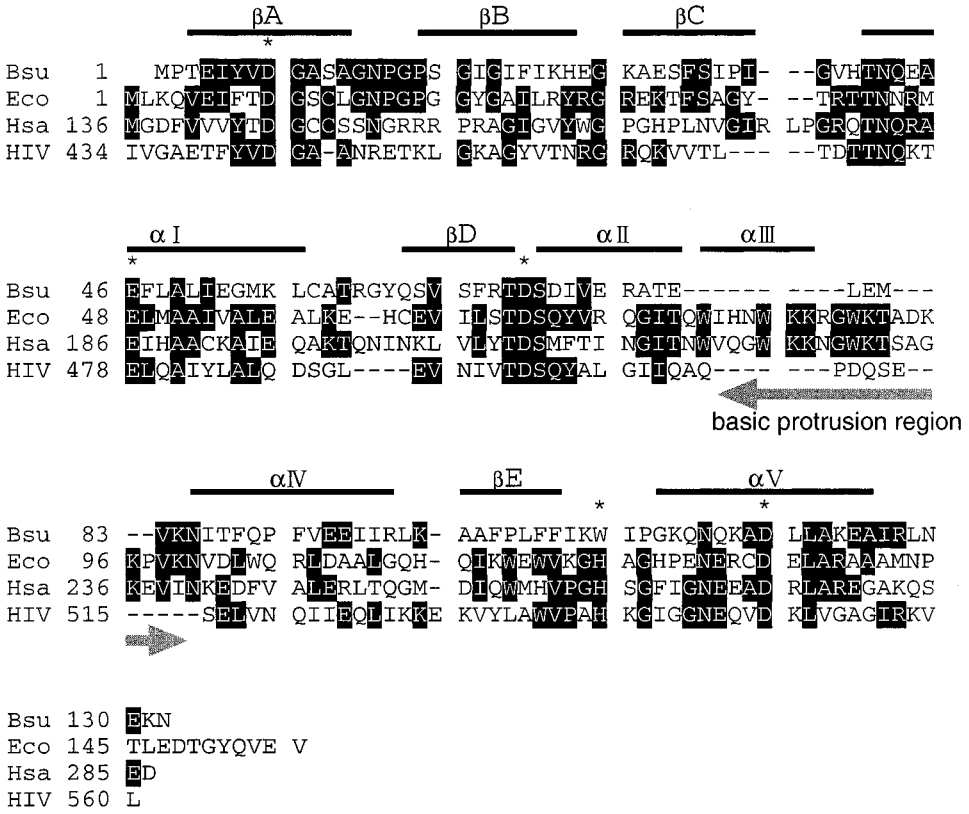

FIGURE 3: Alignment of the sequences of RNase HI and its homologue from *E. coli* (Eco), human (Hsa), HIV-1 (RNase H domain of RT) (HIV), and *B. subtilis* (Bsu). Amino acid residues which are conserved in at least two different proteins are highlighted in black. When two types of amino acid residues were conserved at a given position, the residue conserved in the *B. subtilis* sequence was selected. Gaps are denoted by dashes. The conserved residues in various RNase HI sequences, which are involved in metal ion binding and catalytic function, are denoted by asterisks (*). They are Asp^{10}, Glu^{48}, Asp^{70}, His^{124}, and Asp^{134} for *E. coli* RNase HI. The region which forms a basic protrusion in the *E. coli* RNase HI structure is indicated. Numbers represent the positions of the amino acid residues that start from the initiator methionine for each enzyme. Lines above the sequences indicate the extent of each of the five α-helices and the five *β*-strands of *E. coli* RNase HI (*44*). The accession numbers for these proteins, except for *B. subtilis* protein, are listed in Table 1.

replaced by Trp (Trp^{110}) in this protein sequence (Figure 3). In addition, a region constituting the basic protrusion, which is important for substrate binding (*28*), is missing in this protein sequence.

Enzymatic Activity. The enzymatic activities of the RNase HI from *B. subtilis* were determined under various conditions by using an M13 DNA−RNA hybrid as a substrate. RNase H requires divalent cations for activity, but the preference of the divalent cations varies for different RNases H. Therefore, we have analyzed the dependence of any RNase H activity using various divalent cations, as described in Experimental Procedures. We found that the *B. subtilis* RNase HI homologue exhibited no activity above background levels (specific activity of <0.0005 unit/mg) in the presence of any divalent cation examined. We also examined two other RNase H substrates (see Cleavage of Oligomeric Substrates below). Neither substrate was cleaved by the RNase HI homologue from *B. subtilis*.

Complementation Assay. The *rnh* mutant strain *E. coli* MIC3001 shows an RNase H-dependent temperature-sensitive growth phenotype (*16*). To examine whether the gene encoding the *B. subtilis* RNase HI homologue complements the temperature-sensitive growth phenotype, *E. coli* MIC3001 cells were transformed with pBR2.2 and pBR800es. Plasmid pBR2.2 contains a 2.2 kb pair DNA fragment of the *B. subtilis* genome, in which the *ypdQ* gene is located in the middle of this DNA fragment. In plasmid pBR800es, the transcription and translation of the *ypdQ* gene are under the control of the promoter and SD for *E. coli* RNase HI. Both of these transformants failed to grow at the nonpermissive temperature, indicating that the *B. subtilis* RNase HI homologue does not have RNase H activity in vivo.

Conformation of the B. subtilis RNase HI Homologue. Four invariant acidic amino acid residues, which form the metal ion-binding site in RNases HI, are also conserved in the amino acid sequence of the *B. subtilis* RNase HI homologue. The CD spectrum of this protein gave a broad trough with a shape that is similar to that of *E. coli* RNase HI (Figure 4), suggesting that this protein may fold into a structure somewhat related to that of *E. coli* RNase HI. The depth of this trough was shallower than that of *E. coli* RNase HI by 2-fold, probably because the αIII-helix is completely missing in this protein structure. In fact, this CD spectrum is quite similar to that of the RNase H domain of HIV-1 RT (*29*), in which the αIII-helix is absent.

We have previously shown that binding of a Mg^{2+} ion to the active site of *E. coli* RNase HI increased the thermal stability of this protein (*30*). To examine whether the *B. subtilis* RNase HI homologue assumes a conformation, in which the Mg^{2+} ion binds, thermal denaturation of this protein was analyzed in the presence or absence of 50 mM $MgCl_2$ and/or 2.5 M NaCl as described in Experimental Procedures. Thermal unfolding of this protein is fully reversible under these conditions. Thermal denaturation curves of the *B. subtilis* RNase HI homologue in the absence of these salts, in the presence of 50 mM $MgCl_2$, in the

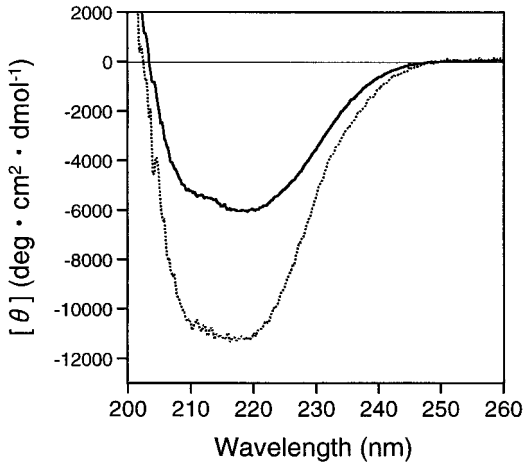

FIGURE 4: Far-UV CD spectra of *E. coli* RNase HI and the *B. subtilis* RNase HI homologue. The far-UV CD spectra of the *B. subtilis* RNase HI homologue (solid line) and *E. coli* RNase HI (dotted line) are shown. The spectra were recorded at 25 °C by using a solution containing enzyme at 0.39 mg/mL in 10 mM Tris-HCl buffer (pH 7.5) and 150 mM NaCl. The optical path length of the cell was 2 mm.

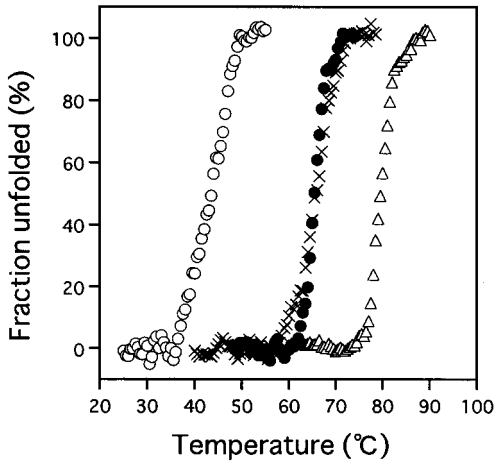

FIGURE 5: Thermal denaturation curves of the *B. subtilis* RNase HI homologue. The apparent fraction of the unfolded protein is shown as a function of temperature. Thermal denaturation curves were determined by monitoring the change in the CD value at 220 nm, as described in Experimental Procedures. The proteins were dissolved in a buffer containing 10 mM Tris-HCl (pH 7.5), 2 M, urea and 150 mM NaCl with no salt (○), 2.5 M NaCl (●), 50 mM $MgCl_2$ (×), or both 2.5 M NaCl and 50 mM $MgCl_2$ (△).

presence of 2.5 M NaCl, and in the presence of both 50 mM $MgCl_2$ and 2.5 M NaCl are shown in Figure 5. We included only these data because the plots of ΔT_m versus concentration of either one of these salts showed a saturation curve and protein stability increased nearly to the maximum level (~23 °C in ΔT_m) in the presence of 50 mM $MgCl_2$ or 2.5 M NaCl (data not shown). The concentrations of $MgCl_2$ and NaCl which increase protein stability by half of the maximum value (~11.5 °C in T_m) were 1.5 mM and 1 M, respectively, suggesting that the dissociation constants of the Mg^{2+} ion and the Na^+ or Cl^- ion are close to these values. The thermal denaturation curve of the protein was measured in the presence of both 50 mM $MgCl_2$ and 2.5 M NaCl as well, to examine whether the stabilization effects of these salts are additive. In the absence of 2.5 M NaCl, the protein was stabilized by 50 mM $MgCl_2$ by 21.7 °C in T_m. In the presence of 2.5 M NaCl, it was stabilized by 50 mM $MgCl_2$ by 14.1 °C in T_m. Because these values are comparable to each other, it is likely that the four invariant acidic residues form a Mg^{2+} ion binding site in the *B. subtilis* RNase HI homologue. Moreover, stabilization of the protein in the presence of the high concentration of NaCl is not primarily due to the binding of the Na^+ ion to this site.

RNase HII Homologues (RNases HII and HIII)

Enzymatic Activities. The enzymatic activities of RNase HII and RNase HIII from *B. subtilis* were determined under various conditions using the M13 DNA−RNA hybrid as a substrate. Because optimum conditions for the activity of *E. coli* RNase HII had not been determined, we also included this protein in our analysis. We have analyzed the effects on RNase H activities of these proteins of various divalent cations as described in Experimental Procedures. *B. subtilis* and *E. coli* RNases HII exhibited Mn^{2+}-dependent RNase H activity, while *B. subtilis* RNase HIII demonstrated a Mg^{2+}-dependent RNase H activity. Both *B. subtilis* and *E. coli* RNases HII showed very weak activities in the presence of the Mg^{2+} ion, which were 0.3−1% of that found using the Mn^{2+} ion. However, these enzymes showed slightly different responses with various divalent cations. For example, *B. subtilis* RNase HII was approximately 20 times more active in the presence of Mn^{2+} ion than in the presence of the Co^{2+} ion. In contrast, *E. coli* RNase HII had no activity in the presence of the Co^{2+} ion. Both enzymes exhibited little RNase H activity in the presence of any other divalent cation (data not shown). Like *E. coli* RNase HI, which prefers Mg^{2+} to Mn^{2+} for activity and exhibits 2−3% of the maximal activity in the presence of the Mn^{2+} ion (*31*), *B. subtilis* RNase HIII exhibited weak enzymatic activity in the presence of the Mn^{2+} ion. In addition, only weak RNase H activity was observed when assays were performed in the presence of the Co^{2+}, Ni^{2+}, and Sr^{2+} ions. The levels of the enzymatic activities determined in the presence of these divalent cations relative to that determined in the presence of the Mg^{2+} ion were 5.0% for Mn^{2+}, 4.3% for Co^{2+}, 0.7% for Ni^{2+}, and 0.5% for Sr^{2+}. *E. coli* RNase HI exhibited little activity in the presence of these divalent cations. Both *E. coli* RNase HI and *B. subtilis* RNase HIII exhibited little activity in the presence of any other divalent cation. As shown in Figure 6, *B. subtilis* RNase HII exhibited maximal activity in the presence of 10 mM $MnCl_2$, whereas the *B. subtilis* RNase HIII optimum activity was at 50 mM $MgCl_2$.

The influence of pH on enzymatic activities of *B. subtilis* RNases HII and HIII was analyzed in the presence of 10 mM $MnCl_2$ and 50 mM $MgCl_2$, respectively. *B. subtilis* RNase HII required a pH of at least 7.0 for activity. Similarly, *B. subtilis* RNase HIII exhibited enzymatic activity only at pH >7.5. The enzymatic activity of *B. subtilis* RNase HII increased as the pH increased from 5 to 10, whereas the *B. subtilis* RNase HIII activity gave an optimum pH of 9.5 (data not shown). However, because the solubility of the divalent cations decreases as the pH increases, and both the RNA−DNA substrate and enzyme may be unstable at a highly alkaline pH, we decided to measure the *B. subtilis* RNases HII and HIII activities at a relatively mild pH of 8.5. The specific activities of *B. subtilis* RNases HII and HIII determined at this pH are ~50 and ~80% of the maximal ones, respectively.

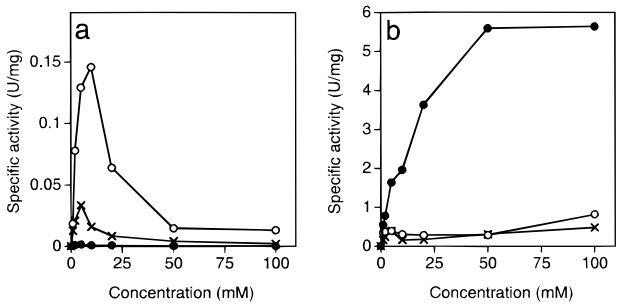

FIGURE 6: Effect of divalent metal ion concentrations on the *B. subtilis* RNase HII and HIII activities. Enzymatic activities of *B. subtilis* RNases HII (a) and HIII (b) were determined at 30 °C in 10 mM Tris-HCl (pH 8.5) containing 50 mM NaCl, 1 mM 2-mercaptoethanol, 10 μg/mL bovine serum albumin, and various concentrations of $MgCl_2$ (●), $MnCl_2$ (○), or $CoCl_2$ (×), by using an M13 DNA−RNA hybrid as a substrate. The concentration range was 1−100 mM for these metal ions. It is noted that *B. subtilis* RNase HIII showed little RNase H activity in the presence of 10−100 μM $MnCl_2$.

Table 3: Comparison of the Specific Activities of the Enzymes in the Presence of either Mg^{2+} or Mn^{2+} Ions[a]

enzyme	metal ion	specific activity (units/mg)	relative activity (%)
E. coli RNase HI	$MgCl_2$	9.5	100
	$MnCl_2$	0.29	3.1
E. coli RNase HII	$MgCl_2$	0.001	0.01
	$MnCl_2$	0.31	3.3
B. subtilis RNase HII	$MgCl_2$	0.005	0.05
	$MnCl_2$	0.50	5.3
B. subtilis RNase HIII	$MgCl_2$	10.0	105
	$MnCl_2$	0.50	5.3

[a] The hydrolysis of the M13 DNA−RNA hybrid with the enzyme was carried out at 30 °C for 15 min under the conditions described in Experimental Procedures. The concentrations of the metal ions were 10 mM, except for that of $MgCl_2$ (50 mM) for the assay of the *B. subtilis* RNase HIII activity. Errors, which represent the 67% confidence limits, are all at or below ±30% of the values reported.

We have also examined the dependence of enzymatic activity on the concentration of NaCl or KCl. *B. subtilis* RNase HII showed nearly twice as much activity in the presence of 50 mM KCl as when using 50 mM NaCl. This activity sharply decreased as the concentration of the salt increased beyond 100 mM. In contrast, *B. subtilis* RNase HIII activity responded equally in the presence of NaCl and KCl with the highest activity observed in the presence of 100−200 mM NaCl or KCl. In the absence of added salt, both enzymes showed only 25−33% of maximal activity. Therefore, we decided to measure the *B. subtilis* RNase HII and RNase HIII activities in the presence of 50 mM KCl and 100 mM NaCl, respectively.

The specific activities of *B. subtilis* RNases HII and HIII, which were determined in the presence of Mg^{2+} and Mn^{2+} ions, were compared with those of *E. coli* RNases HI and HII in Table 3. Interestingly, the specific activities of *B. subtilis* RNase HIII determined in the presence of the Mg^{2+} and Mn^{2+} ions were nearly identical with those of *E. coli* RNase HI. In addition, the specific activities of *B. subtilis* RNase HII determined in the presence of these divalent cations were nearly identical with those of *E. coli* RNase HII.

Cleavage of Oligomeric Substrates. Two different oligomeric substrates have been examined extensively for cleavage by *E. coli* RNase HI. One is a 12 bp RNA−DNA hybrid that is cleaved at multiple sites, but most preferably at a9-c10 (*32*). The second is a DNA-RNA-DNA−DNA substrate that is cleaved only in the middle of the tetraribonucleotide (*33*). We have used these two substrates, labeled at their 5′-ends, to examine whether and how *E. coli* RNase HII, *B. subtilis* RNase HII, and *B. subtilis* RNase HIII hydrolyze them. As a standard, we have also cleaved these substrates with *E. coli* RNase HI. The results are shown in Figure 7 and summarized in Figure 8. When the 12 bp RNA−DNA hybrid was used as a substrate, all of these enzymes cleaved it at multiple sites. Each enzyme generated products resulting in differences and relative frequencies of cleavage by the various enzymes. *E. coli* RNase HI cleaved this substrate preferentially at a9-c10 with minor products resulting from cleavages at a6-u7 and u7-g8. *B. subtilis* RNase HIII favored cleavage at a4-g5 over a second site, a9-c10, and a third site, c1-g2. In contrast to these enzymes, which show relatively high sequence specificities, *E. coli* and *B. subtilis* RNases HII, especially the latter, cleaved the substrate in an almost random manner. *B. subtilis* RNase HII cleaved the substrate at all sites, except for c1-g2 and g2-g3, with nearly equal frequencies. When the 29 bp DNA-RNA-DNA−DNA substrate was used as a substrate, both *E. coli* and *B. subtilis* RNases HII specifically cleaved it at the phosphodiester bond between the third and fourth adenosines (a16-a17). In contrast, *E. coli* RNase HI and *B. subtilis* RNase HIII specifically cleaved the substrate at the middle of the tetra-adenosine (a15-a16). *B. subtilis* RNase HIII cleaved the chimeric substrate at a14-a15 and a16-a17 as well, but with lower frequencies. These experiments were repeated using a substrate with a 3′-end ^{32}P-label, confirming both the limited site cleavages and relative frequencies of hydrolysis as determined for the 5′-labeled substrate (data not shown). In the absence of the complementary DNA, the 29 b DNA-RNA-DNA was not cleaved by any of the enzymes tested, further indicating their specificities for cleaving RNA−DNA hybrids (data not shown).

It should be pointed out that *B. subtilis* RNase HIII cleaved the DNA-RNA-DNA−DNA substrate much less effectively than *E. coli* RNase HI, because the amount of *B. subtilis* RNase HIII required for complete cleavage of the DNA-RNA-DNA−DNA substrate is roughly 300 times larger than that of *E. coli* RNase HI (Figure 7B). These enzymes cleave the M13 DNA−RNA hybrid with nearly equal efficiencies. Because *E. coli* RNase HI cleaves both oligomeric substrates and the M13 DNA−RNA hybrid with nearly equal efficiencies, *B. subtilis* RNase HII cleaves the DNA-RNA-DNA−DNA substrate much less effectively than the M13 DNA−RNA hybrid. Likewise, *E. coli* and *B. subtilis* RNases HII cleaved the 12 bp RNA−DNA hybrid less effectively than the M13 DNA−RNA hybrid, because the amounts of these enzymes required for complete hydrolysis of this oligomeric RNA−DNA hybrid are much larger than those expected from the difference in the specific activities between *E. coli* RNase HI and these RNase HII enzymes (Figure 7A). These results indicate that the size and sequence of an RNA−DNA substrate seriously affect its susceptibility to the cleavage with these RNase H enzymes.

Specificity for RNA−DNA Hybrids. When the 12 b RNA, the 29 b DNA-RNA-DNA, and the 12 bp RNA−RNA duplex, each of which was labeled at its 5′-end, were used

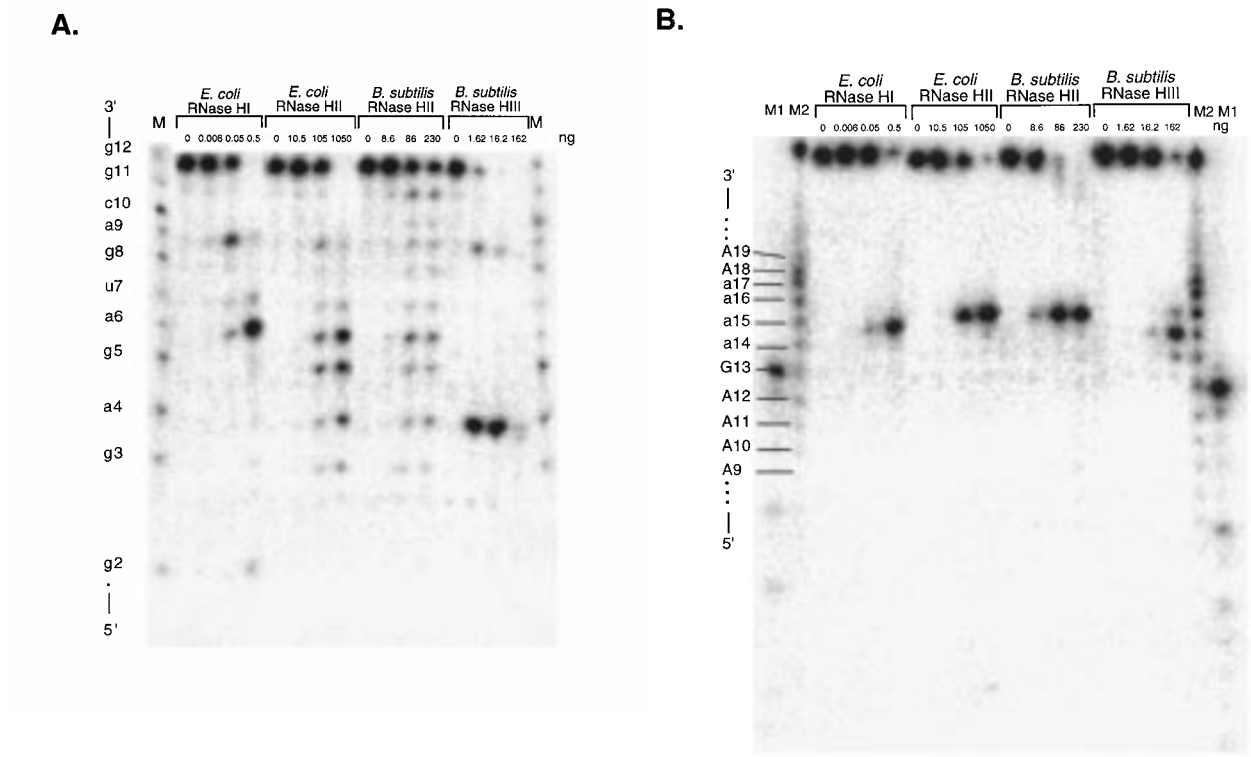

FIGURE 7: Autoradiograph of cleavage reaction products. Hydrolyses of the 5′-end-labeled 12 b RNA hybridized to the 12 b DNA (A) and the 5′-end-labeled 29 b DNA-RNA-DNA hybridized to the 29 b DNA (B) with *E. coli* RNase HI, *E. coli* RNase HII, *B. subtilis* RNase HII, and *B. subtilis* RNase HIII were carried out at 30 °C for 15 min. Products were separated on a 20% polyacrylamide gel containing 7 M urea as described in Experimental Procedures. The concentration of the substrate is 1.0 μM. M and M2 represent products resulting from partial digestion of the 12 b RNA and 29 b DNA-RNA-DNA with snake venom phosphodiesterase, respectively. M1 represents the 5′-end-labeled 13 b DNA with the sequence, which is identical with residues 1−13 of the 29 b DNA-RNA-DNA. Numbers indicate the amount (nanograms) of the enzyme used for hydrolysis of the substrate. The 3′-terminal residue of each oligonucleotide generated by the partial digestion with snake venom phosphodiesterase is shown along the gel.

as a substrate, no cleavage by the RNase HI homologue, RNase HII, or RNase HIII from *B. subtilis* was observed (data not shown). These results indicate that these proteins cannot cleave this single-stranded RNA, single-stranded DNA, or this RNA−RNA duplex. Because none of these proteins cleaved the 29 bp DNA-RNA-DNA−DNA substrate at the double-stranded DNA region, it is clear that they do not cleave a DNA−DNA duplex. *B. subtilis* RNases HII and HIII do not have any nuclease activity other than the RNase H activity. Examination for the cleavage of these substrates by *E. coli* RNase HII supports the previous observation that *E. coli* RNase HII specifically cleaves an RNA−DNA hybrid (*6*) (data not shown).

Amino Acid Sequence Comparison Reveals Multiple Genes Encoding RNase HII. To gain more insight into the rather dramatic differences between *E. coli* and *B. subtilis rnh* genes and the enzymes described above, we examined the sequence relatedness of a number of RNases H whose sequences are presently in the databases. By searching for *rnh* genes using complementation of an *rnhA recBCD* temperature-sensitive strain of *E. coli*, only a single gene was obtained from *S. pneumoniae* DNA (*10*). However, when we performed database searches using the site http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html, a second gene related to *rnhB* of *E. coli* was detected. Nearly the entire region of the gene was present in the database (residues 1−193), and the predicted protein shows 47.9% identity to the *E. coli* RNase HII. We also found the *A. aeolicus* genome contains two genes encoding proteins related to *E. coli* RNase HII. In contrast, only a single RNase HII-like encoding gene could be detected in *M. genitalium* and *M. pneumoniae* genomes. Thus, it appears that different organisms contain one or more copies of genes encoding proteins related to *E. coli* RNase HII.

The *B. subtilis* RNase HII-like sequences were compared with a number of other RNase HII sequences (Figure 9). One problem in this comparison was obvious from a comparison of *B. subtilis* RNases HII and HIII, namely, their relationship at the amino acid sequence level but their quite different biochemical properties. To aid in making these alignments and to understand the evolutionary relationships of these various RNases H, phylogenetic analyses were performed on 23 different RNase HII-like sequences. As shown in Figure 10, these enzymes fell into roughly four subgroups: (i) bacterial proteins related to *B. subtilis* RNase HII, (ii) bacterial proteins related to *B. subtilis* RNase HIII, (iii) archaeal RNases HII, and (iv) eukaryotic RNases HII. Apparently, eukaryotic RNases HII are evolutionarily more closely related to archaeal RNases HII than to bacterial RNases HII and HIII. Because it has not been previously recognized that more than one RNase HII-like protein is present in some organisms, most of the genes for this group of enzymes have been related solely to *E. coli* RNase HII and are generally noted in the database as *rnhB* or RNase HII. One particular instance for which the protein has actually been studied is that of *S. pneumoniae* RNase HII (*10*). Clearly by alignment of sequences, *S. pneumoniae* RNase HII is more closely related to *B. subtilis* RNase HIII, and the gene we identified from the partial genomic DNA sequence currently in the database is more closely related

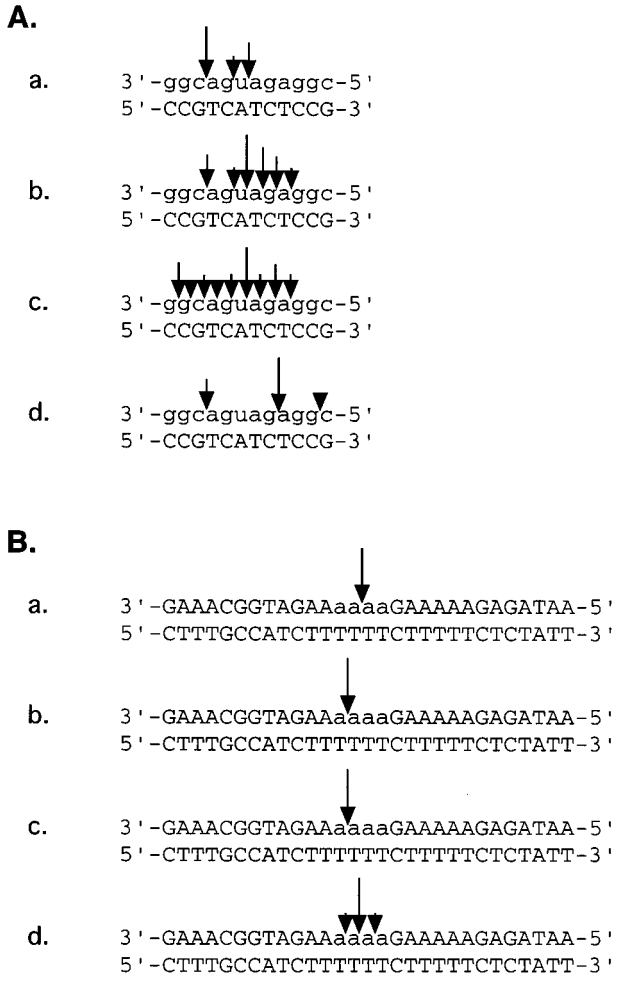

FIGURE 8: Graphical representation of sites and extents of cleavages by RNases H. Cleavage sites of the 12 bp RNA−DNA hybrid (A) and those of the 29 bp DNA-RNA-DNA−DNA substrate (B) with *E. coli* RNase HI (a), *E. coli* RNase HII (b), *B. subtilis* RNase HII (c), and *B. subtilis* RNase HIII (d) are denoted with arrows. Differences in the size of the arrows reflect the relative cleavage intensities at the indicated position. Deoxyribonucleotides are denoted with uppercase letters, and ribonucleotides are denoted with lowercase letters.

to *B. subtilis* RNase HII (Figure 9). Data presented by Zhang et al. (*10*) do not permit us to make any statement concerning specific activity, divalent metal ion preference, or most of the other parameters tested in this paper.

In Figure 9, *E. coli*, *M. jannaschii*, and *Sa. cerevisiae* RNases HII represent bacterial, archaeal, and eukaryotic RNases HII, respectively. The *A. aeolicus* RNase HII sequence and newly identified *S. pneumoniae* RNase HII sequence were also shown, because *A. aeolicus* and *S. pneumoniae* contain both enzymes which are homologous to *B. subtilis* RNases HII and HIII. Likewise, the *B. subtilis* RNase HIII sequence was compared with the sequences of *A. aeolicus* RNase HIII and *S. pneumoniae* RNase HII of Zhang et al. in the same figure. *B. subtilis* RNase HII showed the amino acid sequence identities of 55% to newly identified *S. pneumoniae* RNases HII, 47% to *E. coli* RNase HII, 47% to *A. aeolicus* RNase HII, 25% to *M. jannaschii* RNase HII, 24% to *Sa. cerevisiae* RNase HII, and less than 20% to *B. subtilis*, *A. aeolicus*, and *S. pneumoniae* RNases HII of Zhang et al. On the other hand, *B. subtilis* RNase HIII showed the amino acid sequence identities of 38% to *S. pneumoniae* RNase HII of Zhang et al., 23% to *A. aeolicus* RNase HIII,

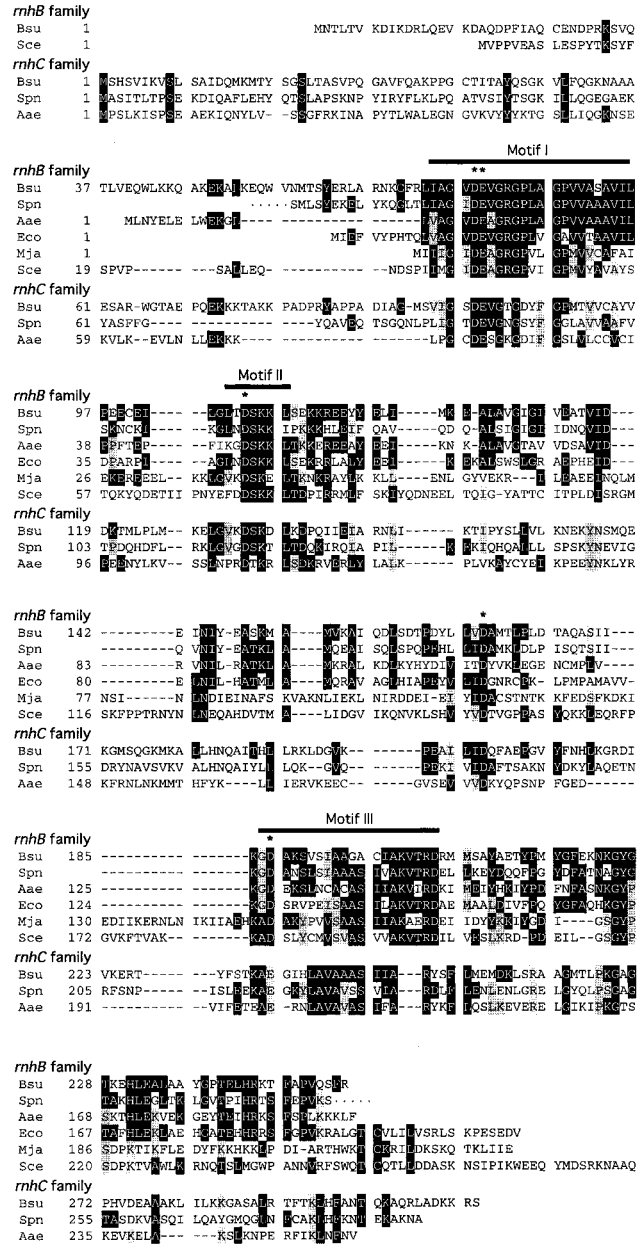

FIGURE 9: Alignment of RNase HII and RNase HIII sequences. The accession numbers for the proteins involved in an *rnhB* family are B. subtilis rnh (ORF) for *B. subtilis* RNase HII (Bsu), A. aeolicus rnh (ORF) for *A. aeolicus* RNase HII (Aae), P10442 (Swiss-Prot) for *E. coli* RNase HII (Eco), U67470 (GenBank) for *M. jannaschii* RNase HII (Mja), and Z71348 (EMBL) for *Sa. cerevisiae* RNase HII (Sce). Those for the proteins involved in an *rnhC* family are *ysgB* (ORF) for *B. subtilis* RNase HIII (Bsu), U93576 (GenBank) for *S. pneumoniae* RNase HII of Zhang et al. (Spn), and aq-1768 (ORF) for *A. aeolicus* RNase HIII (Aae). The accession number for another *S. pneumoniae* protein (Spn) involved in an *rnhB* family is not available. This sequence information is available at the site http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html. Gaps are denoted with dashes. Amino acid residues, which are conserved in at least three different sequences, are denoted with inverse letters. If two types of amino acid residues were conserved at a given position, the most conserved one was denoted with inverse letters; the other one was denoted with a gray-shaded background. If they were conserved equally, the residue conserved in the *B. subtilis* RNase HII sequence is denoted with inverse letters. The conserved amino acid residues with carboxyl groups are denoted with asterisks (*). Numbers represent the positions of the amino acid residues which start from the initiator methionine for each enzyme. The three sequence motifs, which are well conserved in the RNase HII sequences (*10*), are indicated by their identities and lines above the sequences.

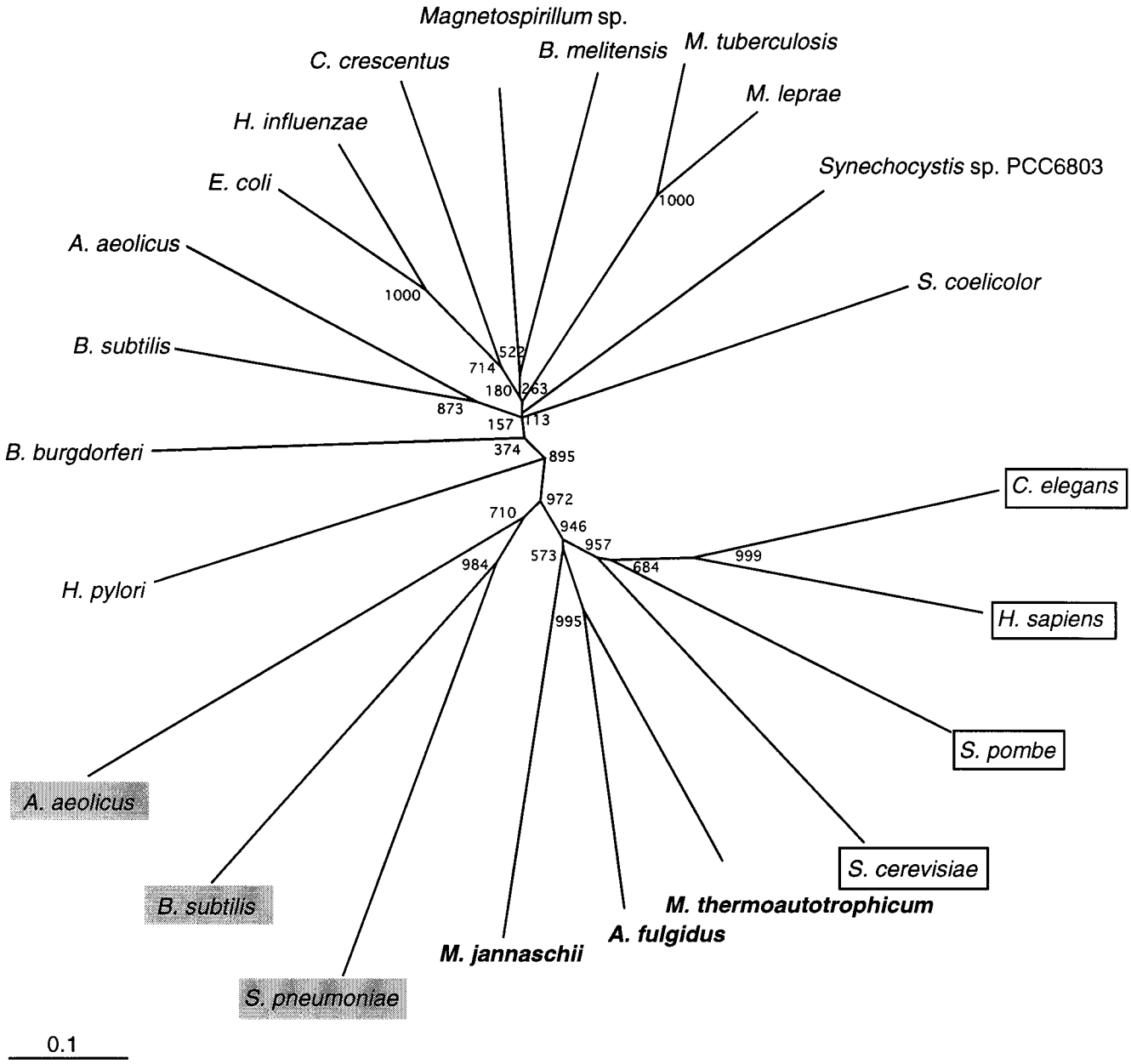

FIGURE 10: Phylogenetic tree derived from RNase HII and RNase HIII sequences of 23 organisms. The phylogenetic tree was constructed by comparing the 23 RNase HII and RNase HIII sequences by using the TreeView program (*14*) on the basis of the distance matrix produced by the CLUSTAL W program (*13*). Of 23 organisms, eukaryotes are boxed, archaea are denoted with boldface, and bacteria, which contain the genes homologous to the *B. subtilis rnhC* gene, are denoted with a gray-shaded background. The accession numbers of the proteins analyzed are presented in Table 1 or the legend of Figure 9. The numbers show how often the target sequence is located at the same position of the tree when the analyses were repetitively performed 1000 times. The distance of the branch from a diverging point is proportional to the frequency of the amino acid substitution at each position in average. The scale bar corresponds to this frequency of 0.1.

23% to *E. coli* RNase HII, and less than 20% to *B. subtilis*, *A. aeolicus*, *M. jannaschii*, and newly identified *S. pneumoniae* RNases HII. RNases HIII are evolutionarily related to RNases HII, particularly through three motifs (motifs I−III), which are highly conserved in the various RNase HII sequences (*10*) (Figure 9).

DISCUSSION

Three B. subtilis rnh Genes. Three genes similar in sequence to *E. coli* RNases HI and HII were studied in this report by cloning and expressing their products in *E. coli*. The gene most closely resembling *E. coli rnhA* yielded a protein with no detectable RNase H activity, employing assays that measured the activity in vitro and in vivo. In contrast, the two genes related to one another and to *E. coli* RNase HII did possess RNase H activity, but surprisingly, the two proteins had remarkably different biochemical properties.

Enzymatic Properties of B. subtilis RNases HII and HIII. The enzymatic activities of *B. subtilis* RNases HII and HIII were determined by using the M13 DNA−RNA hybrid, the 12 bp RNA−DNA hybrid, and the 29 bp DNA-RNA-DNA−DNA duplex as substrates. The enzymatic activities of *E. coli* RNases HI and HII were also determined for comparative purposes. The results indicated that these four enzymes could be placed in two groups on the basis of the differences in the divalent cation specificities, specific activities, and cleavage products generated by hydrolysis of oligomeric

substrates. One group includes *B. subtilis* RNase HIII and *E. coli* RNase HI (group I), and the other group includes *B. subtilis* and *E. coli* RNases HII (group II). Group II enzymes exhibited a Mn^{2+}-dependent RNase H activity with a specific activity of 0.3−0.5 unit/mg (Table 3) and cleaved the 12 bp RNA−DNA hybrid almost nonspecifically and the 29 bp DNA-RNA-DNA−DNA substrate at the phosphodiester bond between the third and fourth adenosines (a16-a17) in the tetra-adenosine region (a14-a17) (Figure 8). In contrast, group I enzymes exhibited a Mg^{2+}-dependent RNase H activity with a specific activity of ∼10 units/mg and cleaved the 12 bp RNA−DNA hybrid sequence primarily at a single site and the 29 bp DNA-RNA-DNA−DNA substrate in the middle of the tetra-adenosine (a15-a16). The large difference in the specific activities of these enzymes suggests that the major RNase H activities detected in the *E. coli* and *B. subtilis* cells may be ascribed to RNase HI and RNase HIII, respectively.

By amino acid sequence analysis, *B. subtilis* RNase HIII is more closely related to *B. subtilis* and *E. coli* RNases HII than to *E. coli* RNase HI which is likely to also be reflected in the tertiary structure. Nevertheless, the enzymatic properties of *B. subtilis* RNase HIII resemble those of *E. coli* RNase HI. In addition, the electrostatic natures of these enzymes are similar. The calculated isoelectric points of *B. subtilis* RNase HIII and *E. coli* RNase HI are 9.0−10.1, whereas those of *B. subtilis* and *E. coli* RNases HII are 5.6−6.9. Does RNase HIII acquire RNase HI-like activity by assuming a very different structure than does RNase HII? To answer this question, structural analyses for these proteins will need to be performed.

Functions of RNases HII and HIII. The physiological roles of RNases HII and HIII remained to be determined. One of the major phenotypes of *E. coli rnhA* mutants is the suppression of oriC-defective initiation. Under conditions of limiting RNase HI activity in *E. coli*, new origins of replication are used permitting bypass of the use of oriC, the normal origin of replication. It will be interesting to know if mutations in either *rnhB* or *rnhC* of *B. subtilis* result in replication of DNA from sites not used for normal DNA replication. Both RNase HII and RNase HIII of *B. subtilis* are able to function in *E. coli* to complement the temperature-sensitive phenotype of an *rnhA recC* (Ts) strain (M. Itaya, unpublished). The requirement for RNase H activity in strains defective for RecBCD is not understood. However, these results suggest that removal of RNA−DNA hybrids is sufficient to permit growth of strains mutated in *recB*, *recC*, or *recD*, and that the biochemical differences reported here for the *B. subtilis* RNase HII and HIII do not affect their ability to remove these "undesirable" RNA−DNA hybrids.

RNA-primed DNA synthesis yields one potential substrate for RNase H and misincorporation of ribonucleotides during DNA synthesis. Unlike the RNase H purified from human erythroleukemic cells which cleaves both of these substrates (*34*), *E. coli* RNase HI is unable to cleave at the RNA−DNA junction of Okazaki fragments and requires at least four contiguous ribonucleotides in a duplex DNA for cleavage (*33*).

Similar to *E. coli* RNase HI, none of the RNases H examined in this study cleaved the 29 bp DNA-RNA-DNA−DNA substrate at either the DNA−RNA or RNA−DNA junction (Figure 8). Thus, neither RNase HII nor RNase HIII may have a repair function, by which single ribonucleotides misincorporated into DNA are removed, and they cannot eliminate single ribonucleotides attached to the 5′-end of the DNA strand in Okazaki fragments. Whatever their functions may be, it seems possible, given the problematic nature of the *rnhA* gene in *B. subtilis*, that an enzyme such as the highly active RNase HIII might substitute for RNase HI in the organisms where they are present. It is even reasonable to suggest that *B. subtilis* RNases HII and HIII are functionally related to *E. coli* RNases HII and HI, respectively.

Catalytic Mechanism. Enzymes requiring divalent cations for activity and cleaving their substrates at alkaline pH typically cleave the P−O3′ bond of the RNA in an endonucleolytic manner. From the migration of the products produced by the *B. subtilis* RNases HII and HIII (Figure 7), we deduce that these enzymes also cleave endonucleolytically at the P−O3′ bond. Three or four carboxylates provide ligands for the binding of the metal ion(s). Comparison of the various RNases HII and HIII sequences indicated that five acidic amino acid residues are fully conserved as noted (Figure 9). Interestingly, no histidine residue is conserved in the RNase HII sequences or in the RNase HIII sequences, suggesting a slightly different source for activation of an attacking H_2O molecule from that suggested for *E. coli* RNase HI (*30*, *35*).

Is RNase HI Dispensable? RNase HI is widely present in various organisms from *E. coli* to humans (*36*, *37*). However, an *rnhA* gene, encoding such an enzyme, has not been recognized by computer-based similarity searches of the genomic DNA sequences of those mycoplasma and archaea whose complete DNA sequence has been determined. In addition, such searches failed to find a similar gene in certain Gram-negative and Gram-positive bacteria, such as *A. aeolicus* and *S. pneumoniae*. These results raise the following question. Is RNase HI dispensable in these microorganisms?

In this report, we showed the RNase HI homologue, which is encoded by the *ypdQ* gene, does not have RNase H activity either in vivo or in vitro, despite its ability to fold into an RNase H-like structure (Figure 4). The two major differences between the protein produced by the *B. subtilis rnhA* gene and that for other active RNases HI are (i) replacement of Trp for His at the position, in which the active-site histidine residue is fully conserved, and (ii) the absence of amino acids corresponding to the basic protrusion in the structure of *E. coli* RNase HI. Perhaps this protein requires an additional subunit to supply the binding function provided by the basic protrusion in other RNases HI. Preliminary experiments for testing this possibility have been so far unsuccessful (N. Ohtani, unpublished). A second possibility is that the catalytic portion of RNase HI has been co-opted by another system to direct this nuclease to a completely different target molecule. Yet a third possibility is that the protein has a function completely unrelated to that of an RNase H. *B. subtilis* is not unique in having an *rnhA*-like gene missing the basic protrusion and having Trp substituted for His. A similar RNase HI homologue has been found in *Enterococcus faecalis* (ebsB) and *Mycobacterium tuberculosis* (MTCY427.09c).

Multiple rnh Genes. While using computer-based searches for *rnh*-like genes may be limiting, useful information has been obtained from these searches. Together with the cloning and characterization of the *B. subtilis* RNases H reported in this paper, these permit us to describe the current status of

such genes in all kingdoms. Bacteria, with the exception of Mycoplasma, all have two recognizable and, presumably like for *B. subtilis*, active RNases H. For the Mycoplasma, whose complete genomic DNA sequences have been determined, only a single *rnhC*-like gene was detected. We were even able to find a second RNase HII-like gene in *S. pneumoniae*, an organism previously thought to have but one *rnh* gene (*10*). The amino acid sequence of this newly discovered sequence is more closely related to *B. subtilis* RNase HII than is the original RNase HII described by Zhang et al. (*10*). The combination of *rnh*-like genes in bacteria varies in a nonobvious manner. Some organisms contain *rnhA* and *rnhB* genes (AB type), while others have *rnhB* and *rnhC* genes (BC type). Apparently, these types of genomes are not correlated with the bacterial species. For example, both the Gram-negative bacteria, such as *A. aeolicus*, and Gram-positive bacteria, such as *B. subtilis* and *S. pneumoniae*, have BC type genes. Likewise, the Gram-negative bacteria, such as *E. coli*, have the AB type gene. It is also likely that the Gram-positive bacterium *M. smegmatis* has the AB type genes as well, because it has been shown that this genome contains the *rnhA* gene (*38*).

Archaea, on the other hand, seem to have only the *rnhB*-like gene. We have recently found that the enzymatic properties of RNase H from a hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 are more similar to those of *B. subtilis* RNase HII than to those of *B. subtilis* RNase HIII (*39*). Because this enzyme is clearly a member of archaeal RNase H enzymes, it is relevant to classify them as RNase HII.

In contrast to bacteria, eukaryotes have so far been shown to contain only the AB type genes. We classified eukaryotic enzymes as RNases HII, because of their close evolutionary relationships with archaeal RNases HII. Only a single gene can be found in the *Sa. cerevisiae* genomic DNA sequence database with sufficient sequence similarity to any of several widely divergent members of the *rnhB* or *rnhC* type of genes to be detected by computer-assisted searches. The *rnhA* homologue of *Sa. cerevisiae* has been extensively studied (*40*, *41*). Also, searching for RNase HII-related protein sequences in human and mouse EST databases, one finds many different cDNA clones. When these amino acid sequences are compared, it is clear that they are encoded by very closely related genes that are much more closely related than *B. subtilis rnhB* and *rnhC*. In fact, they are likely to have been encoded by the same gene. This singularity of RNase HII sequences does not preclude the presence of other RNase HII-like proteins because these EST sequences represent frequencies of RNA transcripts and not gene frequencies. Frank et al. (*42*) have suggested that the *Sa. cerevisiae* RNase HII-encoding gene (*rnh35*) is actually responsible for RNase H activity in *Sa. cerevisiae*. Clones of the *Sa. cerevisiae*, mouse, and human cDNAs encoding these RNase HII-like proteins produce proteins with no RNase H activity when expressed in *E. coli* (A. Arudchandran and R. J. Crouch, unpublished). Further progress in these studies, as well as in the genome sequence analyses, may allow us to identify the BC type genome, or even a novel type of the genome containing a novel *rnh* gene, in eukaryotes.

ACKNOWLEDGMENT

We thank ID Biomedical Corp. for providing the DNA-RNA-DNA probe for the RNase H assay and A. Muroya, T. Kochi, and N. Hirano for helpful discussions.

REFERENCES

1. Crouch, R. J., and Dirksen, M.-L. (1982) in *Nuclease* (Linn, S. M., and Roberts, R. J., Eds.) pp 211−241, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Kanaya, S., and Crouch, R. J. (1983) *J. Biol. Chem. 258*, 1276−1281.
3. Crouch, R. J. (1990) *New Biol. 2*, 771−777.
4. Hostomsky, Z., Hostomska, Z., and Matthews, D. A. (1993) in *Nucleases* (Linn, S. M., and Roberts, R. J., Eds.) 2nd ed., pp 341−376, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
5. Kanaya, S., and Ikehara, M. (1995) in *Subcellular Biochemistry, Vol. 24, Proteins: Structure, Function, and Engineering* (Biswas, B. B., and Roy, S., Eds.) pp 377−422, Plenum Press, New York.
6. Itaya, M. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8587−8591.
7. Yang, W., Hendrickson, W. A., Crouch, R. J., and Satow, Y. (1990) *Science 249*, 1398−1405.
8. Woese, C. R., and Fox, G. E. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5088−5099.
9. Woese, C. R., Kandler, O., and Wheelis, M. L. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4576−4579.
10. Zhang, Y.-B., Ayalew, S., and Lackes, S. A. (1997) *J. Bacteriol. 179*, 3828−3836.
11. Kunst, F., et al. (1997) *Nature 390*, 249−256.
12. Deckert, G., et al. (1998) *Nature 392*, 353−358.
13. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res. 22*, 4673−4680.
14. Page, R. D. M. (1996) *Comput. Appl. Biosci. 12*, 357−358.
15. Imanaka, T., Fujii, M., and Aiba, S. (1981) *J. Bacteriol. 146*, 1091−1097.
16. Itaya, M., and Crouch, R. J. (1991) *Mol. Gen. Genet. 227*, 424−432.
17. Zhan, X., and Crouch, R. J. (1997) *J. Biol. Chem. 272*, 22023−22029.
18. Haruki, M., Noguchi, E., Akasako, A., Oobatake, M., Itaya, M., and Kanaya, S. (1994) *J. Biol. Chem. 269*, 26904−26911.
19. Miller, J. H. (1972) in *Experiments in Molecular Genetics* (Miller, J. H., Ed.) p 433, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
20. Kanaya, S., Kohara, A., Miyagawa, M., Matsuzaki, T., Morikawa, K., and Ikehara, M. (1989) *J. Biol. Chem. 264*, 11546−11549.
21. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22. Kanaya, S., Kimura, S., Katsuda, C., and Ikehara, M. (1990) *Biochem. J. 271*, 59−66.
23. Goodwin, T. W., and Morton, R. A. (1946) *Biochem. J. 40*, 628−632.
24. Laemmli, U. K. (1970) *Nature 227*, 680−685.
25. Kimura, S., Nakamura, H., Hashimoto, T., Oobatake, M., and Kanaya, S. (1992) *J. Biol. Chem. 267*, 21535−21542.
26. Kanaya, S., Katsuda, C., Kimura, S., Nakai, T., Kitakuni, E., Nakamura, H., Katayanagi, K., Morikawa, K., and Ikehara, M. (1991) *J. Biol. Chem. 266*, 6038−6044.
27. Jay, E., Bambara, R., Padmanabham, P., and Wu, R. (1974) *Nucleic Acids Res. 1*, 331−353.
28. Kanaya, S., Katsuda-Nakai, C., and Ikehara, M. (1991) *J. Biol. Chem. 266*, 11621−11627.
29. Becerra, S. P., Clore, G. M., Gronenborn, A. M., Karlstrom, A. R., Stahl, S. J., Wilson, S. H., and Wingfield, P. T. (1990) *FEBS Lett. 270*, 76−80.
30. Kanaya, S., Oobatake, M., and Liu, Y.-Y. (1996) *J. Biol. Chem. 271*, 32729−32736.
31. Berkower, I., Leis, J., and Hurwitz, J. (1973) *J. Biol. Chem. 248*, 5914−5921.

32. Kanaya, E., and Kanaya, S. (1995) *Eur. J. Biochem. 231*, 557−562.
33. Hogrefe, H. H., Hogrefe, R. I., Walder, R. Y., and Walder, J. A. (1990) *J. Biol. Chem. 265*, 5561−5566.
34. Eder, P. S., and Walder, J. A. (1991) *J. Biol. Chem. 266*, 6472−6479.
35. Kashiwagi, T., Jeanteur, D., Haruki, M., Katayanagi, K., Kanaya, S., and Morikawa, K. (1996) *Protein Eng. 9*, 857−867.
36. Cerritelli, S. M., and Crouch, R. J. (1998) *Genomics 53*, 300−307.
37. Wu, H., Lima, W. F., and Crooke, S. T. (1998) *Antisense Nucleic Acid Drug Dev. 8*, 53−61.
38. Dawes, S. S., Crouch, R. J., Morris, S. L., and Mizrahi, V. (1995) *Gene 165*, 71−75.
39. Haruki, M., Hayashi, K., Kochi, T., Muroya, A., Koga, Y., Morikawa, M., Imanaka, T., and Kanaya, S. (1998) *J. Bacteriol. 180*, 6207−6214.
40. Cerritelli, S. M., and Crouch, R. J. (1995) *RNA 1*, 246−259.
41. Cerritelli, S. M., Fedoroff, O. Y., Reid, B. R., and Crouch, R. J. (1998) *Nucleic Acids Res. 26*, 1834−1840.
42. Frank, P., Braunshofer-Reiter, C., and Wintersberger, U. (1998) *FEBS Lett. 421*, 23−26.
43. Doolittle, R. F., Feng, D.-F., Johnson, M. S., and McClure, M. A. (1989) *Q. Rev. Biol. 64*, 1−30.
44. Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Ikehara, M., Matsuzaki, T., and Morikawa, K. (1990) *Nature 347*, 306−309.

BI982207Z