Specificity and Evolutionary Conservation of the Escherichia coli RNA Pyrophosphohydrolase RppH*5

Received for publication, December 27, 2014, and in revised form, January 30, 2015 Published, JBC Papers in Press, February 5, 2015, DOI 10.1074/jbc.M114.634659

Patricia L. Foley1,2, Ping-kun Hsieh1, Daniel J. Luciano2, and Joel G. Belasco3
From the Kimmel Center for Biology and Medicine at the Skirball Institute and the Department of Microbiology, New York University School of Medicine, New York, New York 10016

Background: In bacteria, RppH triggers mRNA degradation by converting the 5′-terminal triphosphate to a monophosphate.

Results: The substrate specificity of E. coli RppH has been determined, along with residues important for this property and species in which they are conserved.

Conclusion: The specificity of RppH is species-dependent.

Significance: These findings help to explain the selective influence of RppH on rates of mRNA decay.

Bacterial RNA degradation often begins with conversion of the 5′-terminal triphosphate to a monophosphate by the RNA pyrophosphohydrolase RppH, an event that triggers rapid ribonucleolytic attack. Besides its role as the master regulator of 5′-end-dependent mRNA decay, RppH is important for the ability of pathogenic bacteria to invade host cells, yet little is known about how it chooses its targets. Here, we show that Escherichia coli RppH (EcRppH) requires at least two unpaired nucleotides at the RNA 5′ end and prefers three or more such nucleotides. It can tolerate any nucleotide at the first three positions but has a modest preference for A at the 5′ terminus and either a G or A at the second position. Mutational analysis has identified EcRppH residues crucial for substrate recognition or catalysis. The promiscuity of EcRppH differentiates it from its Bacillus subtilis counterpart, which has a strict RNA sequence requirement. EcRppH orthologs likely to share its relaxed sequence specificity are widespread in all classes of Proteobacteria, except Deltaproteobacteria, and in flowering plants. By contrast, the phylogenetic range of recognizable B. subtilis RppH orthologs appears to be restricted to the order Bacillales. These findings help to explain the selective influence of RppH on bacterial mRNA decay and show that RppH-dependent degradation has diversified significantly during the course of evolution.

In bacteria, the ends of primary transcripts are protected by a 5′-terminal triphosphate and a 3′-terminal stem-loop, both of which are nuclease-resistant. As a consequence, bacterial mRNA degradation generally begins by either of two mechanisms in which those barriers are bypassed or removed. The bypass mechanism involves direct access by an endonuclease to an internal cleavage site (1, 2). In Escherichia coli, RNase J, an enzyme that is absent from other Gram-positive bacteria, cleaves the 5′ end of the monophosphorylated (3), whereas in Bacillus subtilis, the endonuclease Rnase Y plays a similarly important role (3). Less frequently, internal cleavage by another endonuclease such as Rnase III triggers mRNA decay. The other mechanism for initiating mRNA degradation involves conversion of the 5′-terminal triphosphate to a monophosphate by the RNA pyrophosphohydrolase RppH, a member of the Nudix hydrolase family (4–6). In E. coli, this modification creates better substrates for RNase E, whose endonucleolytic activity is greatly enhanced when the RNA 5′ end is monophosphorylated (7), whereas in B. subtilis it exposes the 5′ end to attack by the 5′-exonuclease Rnase J, an enzyme that is absent in E. coli (6). mRNA degradation by the 5′-end-dependent pathway is important for bacterial pathogenesis, as evidenced by the impaired invasiveness of mutant strains that lack RppH (8–10).

The specificity of the two enzymes involved in RppH-dependent degradation in B. subtilis has been investigated. Purified B. subtilis RppH (BsRppH) requires two unpaired nucleotides at the RNA 5′ end and prefers three or more such nucleotides (11). Furthermore, BsRppH is sequence-dependent, requiring G as the second nucleotide of its substrates, preferring a purine at the third position, and slightly favoring A over G at position 1 (11, 12). In contrast, Rnase J appears to require at least 9–10 unpaired nucleotides at the 5′ end of the monophosphorylated intermediate for maximum 5′-exonuclease activity and 4–5 unpaired nucleotides for exonuclease processivity, but it exhibits little if any sequence specificity (13). Consistent with these observations, RppH-dependent mRNA degradation in B. subtilis requires an unpaired 5′ end and a G at position 2 (6, 11).

* This work was supported, in whole or in part, by National Institutes of Health Research Grant R01GM035769 (to J. G. B.)
2 Supported by National Institutes of Health Fellowship T32AI007180.
3 To whom correspondence should be addressed: Skirball Institute, New York University School of Medicine, 540 First Ave., New York, NY 10016 Tel.: 212-263-5409; Fax: 212-263-2150; E-mail: joel.belasco@med.nyu.edu.

The abbreviations used are: BsRppH, B. subtilis RppH; EcRppH, E. coli RppH; oligo, oligodeoxynucleotide; Ap4A, diadenosine tetraphosphate.
Interestingly, *B. subtilis* appears to contain a second, as yet unidentified, RNA pyrophosphohydrolase whose action is sequence-independent (6, 11).

In *E. coli*, hundreds of messages are targeted for degradation by the RppH-dependent pathway, but even more are not (5, 14). However, little is understood about the basis for the specificity of this pathway in *E. coli* other than its requirement for an unpaired 5′-end (1, 5, 15, 16). Neither the number of unpaired 5′-terminal nucleotides required for pyrophosphate removal by *E. coli* RppH (EcRppH) or for 5′-monophosphate-assisted cleavage by RNase E nor the influence of the sequence of those nucleotides has previously been determined.

We have now examined the specificity of EcRppH. Our findings show that, although its 5′-terminal length requirements resemble those of BsRppH, its sequence preferences are much more permissive. Phylogenetic comparison indicates which other bacterial species contain an EcRppH-like ortholog and which contain a BsRppH-like ortholog. Finally, mutational analysis of EcRppH has identified amino acid residues important for its catalytic activity and relaxed substrate specificity.

**EXPERIMENTAL PROCEDURES**

**Purification of RppH—** Wild-type and mutant forms of *E. coli* RppH bearing an amino-terminal hexahistidine tag were produced in *E. coli* cells that contained plasmid pPac-RppH6 (5) or a derivative thereof and lacked the chromosomal genes encoding RppH and RNase I. Protein synthesis was induced by adding isopropyl-1-thio-D-galactopyranoside (1 mM) to a log phase culture (A460 = 0.5) growing in LB medium (1000 ml) at 37 °C. After 2 h, the cells were pelleted, resuspended in 10 ml of buffer E (10 mM HEPES, pH 7.5, 0.3 M sodium chloride) containing 0.25% Genapol, 0.1 mM PMSF, and Complete EDTA-free protease inhibitor (Roche Applied Science), and disrupted in an EmulsiFlex-C3 high pressure homogenizer (Avestin). The cell lysates were cleared twice by centrifugation at 14,500 × *g* for 15 min at 4 °C, and the supernatants were incubated with 2 ml of BD TALON metal affinity resin (Clontech) for 1 h at 4 °C with gentle agitation. The resin was washed five times with buffer E containing 0–20 mM imidazole, and RppH was then eluted with buffer E containing 400 mM imidazole. The imidazole was then removed by gel filtration through Sephadex G-25 (1 ml). Peak fractions were stored at −80 °C in buffer E containing 50% (v/v) glycerol. Protein concentrations were measured colorimetrically by using the Bio-Rad Protein Assay and verified by SDS-PAGE and staining with Coomassie Blue.

**Monitoring Pyrophosphate Removal by RppH—** RppH substrates bearing a 5′-terminal γ-32P label and a single internal fluorescein label were prepared by *in vitro* transcription and purified by PAGE, as described previously (11).

To examine the specificity of wild-type EcRppH, equal amounts of the 5′-γ-32P-labeled RNA to be tested and 5′-γ-32P-labeled A8XL (0.4 pmol each) were prewarmed to 37 °C for 3 min in EcRppH reaction buffer (20 mM HEPES, pH 7.5, 10 mM magnesium chloride, 1 mM dithiothreitol, 1% glycerol; total volume of 99 μl). A 0-min sample (9 μl) was removed, mixed with 8 μl of loading buffer (90 mM EDTA, 82% formamide), and cooled on ice. RppH (10 μl of an 80 nM solution in EcRppH reaction buffer) was added to the remainder (final RppH concentration of 8 nM), and 10-μl samples were withdrawn at time intervals (1, 2, 4, and 8 min) and quenched on ice with 8 μl of a loading buffer containing excess EDTA. (Under these conditions, the reaction rate is limited by the concentration of RppH.) All of the samples were then heated to 95 °C for 3 min and subjected to electrophoresis on a 13.5% polyacrylamide, 8 M urea gel. After visualizing the fluorescence of the RNA bands with a Typhoon Trio scanner (GE Healthcare), the gel was dried, and the radioactivity of the bands was visualized with a Storm 820 PhosphorImager (GE Healthcare). Band intensities were quantified by using ImageQuant software, and radioactivity was normalized to fluorescence.

To examine the effect of amino acid substitutions on the activity and specificity of EcRppH, equal amounts of 5′-γ-32P-labeled A8XL and either A4_AGAA, or A4_ACAA (final concentration of 5 nM each) were prewarmed to 37 °C for 5 min in EcRppH reaction buffer containing unlabeled total *E. coli* RNA (1 μg/ml) as a competitor. EcRppH or a mutant thereof was added to a final concentration of either 15 nM (for A4_AGAA) or 60 nM (for A4_ACAA), and samples (10 μl) were withdrawn at time intervals (0, 1, 2, 4, and 8 min) and quenched on ice with a loading buffer containing excess EDTA. Sample analysis was then performed as described above.

**Examining mRNA Degradation in E. coli—** To determine the effect of sequence on the 5′ phosphorylation state of *yeiP* mRNA in *E. coli*, either strain BW25113 containing plasmid pYeip1 (17) or strain BW25113ΔyeipP (17) containing plasmid pYeip1-U2G was grown to exponential phase at 37 °C in MOPS/glucose medium. Total cellular RNA was then extracted and analyzed by PABLO as described previously (17), except that 10–23 deoxyribozyme Yeip-DZM7 (GTAATTCAGTGGCTAGCTACAACGACATATCCTTTT) was used instead of Yeip-DZM3, oligo X22 (GAAA-CATAATGAATGATAACTTG) was used instead of YeiP-DZM3, and oligo Yyeip-U2G (AGTCGAAAATGTCAAAAATATCAAGTTATCATTCA) was used instead of oligo Yyeip-DZM3. (AGTCGAAAATGTCAAAAATATCAAGTTATCATTCA) was used instead of oligo Yyeip-U2G (AGTCGAAAATGTCAAAAATATCAAGTTATCATTCA) to determine the phosphorylation state of the *yeiP*-U2G transcript, and purified EcRppH was used instead of tobacco acid pyrophosphatase to generate fully monophosphorylated RNA *in vitro*.

Rates of *yeiP* mRNA degradation in these strains or in congenic strains lacking the rppH gene were measured by Northern blot analysis of total RNA extracted at time intervals after inhibiting transcription, as described previously (17).

**Identification of EcRppH and BsRppH Orthologs—** Bacterial proteins with a high degree of sequence homology to EcRppH and BsRppH were identified by using Blastp. True orthologs of these two enzymes were distinguished from other Nudix hydrolases on the basis of their *p* value (≤3 × 10−12 for proteobacterial EcRppH orthologs or ≤4 × 10−15 for BsRppH orthologs) and confirmed by the presence of conserved amino acid residues characteristic of each protein clade. Nudix hydrolases with higher *p* values had clear deficiencies in sequence homology outside the vicinity of the Nudix motif. Subclades of EcRppH orthologs and their evolutionary relatedness were determined by using ClustalW2 to analyze protein sequences. The degree to which individual amino acid residues are conserved in orthologs of EcRppH and BsRppH was examined by
using the Protein versus All Alignment tool on the website of the Comprehensive Microbial Resource.

RESULTS

Requirement for Unpaired Nucleotides at the 5′Terminus—To determine the minimum number of unpaired 5′-terminal nucleotides required by EcRppH, we used a set of structurally unambiguous substrates previously employed to examine the specificity of BsRppH. A8, the prototype of these RNA substrates, comprised an 8-nucleotide single-stranded segment followed by two stem-loop structures, the first of which contained the only U nucleotide in the entire molecule (Fig. 1). It was synthesized by in vitro transcription in the presence of [γ-32P]ATP and fluorescein-12-UTP so as to incorporate a γ radiolabel within the 5′-terminal triphosphate and a single fluorescein label at the top of the first stem-loop. For use as an internal standard, we also prepared doubly labeled A8XL RNA, which was identical to A8 except for an additional stem-loop at the 3′ end.

To monitor pyrophosphate removal from these RNAs, we combined equal amounts of each with EcRppH in a Mg2+-containing buffer, quenched reaction samples periodically with EDTA, and subjected the reaction products to gel electrophoresis. The extent of reaction at the various time points was then quantified for each substrate by normalizing the radioactivity of the corresponding gel band to its fluorescence intensity. As expected, the rate of pyrophosphate removal was very similar for A8 and A8XL (Fig. 2A and B, graphs). RppH-catalyzed phosphate removal from A8, A4, A3, A2, A1, and A1 + 3 or from G8, G4, G3, G2, G1, and G0 was monitored as in A and quantified by normalizing the radioactivity remaining in each RNA to the corresponding fluorescence intensity. Each time point is the average of two or more independent measurements. Error bars have been omitted to improve the legibility of the graph; instead, the standard error of each measurement is reported in supplemental Table 1.

![FIGURE 1. EcRppH substrates. The sequence and expected secondary structure of A8, A4, A3, A2, A1, A1+3, G8, and A8XL RNA are shown. Each bore a 5′-terminal triphosphate (ppp), a γ-32P radiolabel (* at the 5′ end, and a fluorescein label at the top of the first stem-loop. In each RNA name, the letter indicates the identity of the 5′-terminal nucleotide, and the numeral indicates the number of unpaired nucleotides at the 5′ end. Truncated derivatives of A8 (A4, A3, A2, A1) lacked 4–7 nucleotides from the 3′ boundary of the 5′-terminal single-stranded segment. G8, G4, G3, G2, G1, and G0 were identical to A1 except for the presence of G instead of A at the 5′ end. A1 + 3 was the same as A1 except for three additional nucleotides at the 3′ end.

![FIGURE 2. Effect of the length of the 5′-terminal single-stranded segment on reactivity with EcRppH in vitro. A, representative gel images. In vitro transcribed A8 and A1 bearing a γ-32P radiolabel and an internal fluorescein label were mixed with labeled A8XL and treated with purified EcRppH (8 nM), and the radioactivity (P-32) and fluorescence (Fluor) of each RNA were monitored as a function of time by gel electrophoresis. B, graphs. RppH-catalyzed phosphate removal from A8, A4, A3, A2, A1, and A1 + 3 or from G8, G4, G3, G2, G1, and G0 was monitored as in A and quantified by normalizing the radioactivity remaining in each RNA to the corresponding fluorescence intensity. Each time point is the average of two or more independent measurements. Error bars have been omitted to improve the legibility of the graph; instead, the standard error of each measurement is reported in supplemental Table 1.

We then shortened the single-stranded segment at the 5′ end of A8 from 8 to 4, 3, 2, or 1 nucleotides by removing nucleotides from its 3′ boundary so as to generate A4, A3, A2, and A1 (Fig.
The reactivity of each toward EcRppH was compared by monitoring pyrophosphate removal in the presence of A8XL. A4 and A3 were nearly as reactive as A8, whereas A2 was significantly less reactive, and A1 was completely unreactive (Fig. 2A and B). Adding three unpaired nucleotides to the 3'-end of A1 (A1+3) did not increase its reactivity (Fig. 2B), demonstrating that its resistance to pyrophosphate removal by EcRppH was due to an insufficient number of unpaired nucleotides at the 5'-end and not merely to its shorter overall length. A similar length dependence was observed for a related set of RNA substrates in which the 5'-terminal nucleotide was changed to G (Fig. 2C). These results show that substrates of EcRppH, like those of BsRppH, must have at least two unpaired nucleotides at the 5'-end and preferably three or more.

Effect of 5'-Terminal RNA Sequence—To ascertain whether EcRppH has sequence preferences at the 5'-end of its substrates, we introduced single nucleotide substitutions into A4 (hereafter referred to as A4AGAA to indicate both the identity of the 5'-terminal nucleotide and the sequence of unpaired nucleotides at the 5'-end) and tested their effect on the rate of pyrophosphate removal. The reaction rate of G4GGAA, in which the first nucleotide was changed from A to G (the other one commonly present at the 5'-end of primary transcripts), was somewhat slower than that of A4AGAA (Fig. 3A). Changing the identity of the third nucleotide had only a small effect, as A4AGGA and A4AGCA were as reactive as A4AGAA, and A4AGUA was only slightly less reactive (Fig. 3B). By contrast, substitutions at the second position had a clear effect on reactivity. Although replacing the G at position 2 of G4GGAA with A (to create G4GAAA) impeded the reaction only slightly, substituting a pyrimidine at that position in the context of either G4GGAA or A4AGAA (to create G4GCAA, G4GUAA, A4ACAA, or A4AUAA) allowed the reaction to proceed but slowed it significantly (Fig. 3C). Thus, the 5'-terminal sequence specificity of EcRppH differs from that of BsRppH in one crucial respect: whereas BsRppH has an absolute requirement for G at position 2, EcRppH prefers either G or A at that position but can also tolerate either C or U.

Influence of 5'--Terminal RNA Sequence in E. coli—To determine whether the sequence preference of EcRppH is also evident in vivo, we examined the influence of the second nucleotide on the steady-state ratio of monophosphorylated to triphosphorylated yeiP mRNA, a transcript known to be degraded by an RppH-dependent mechanism in E. coli (5, 17).
Specificity and Conservation of E. coli RppH

FIGURE 4. Effect of the identity of the second RNA nucleotide on monophosphorylation in vivo. Total RNA was extracted from log-phase E. coli cells containing either wild-type yeiP mRNA (5' AU) or a variant (yeiP-U2G) in which the second nucleotide was changed from U to G (5' AG), and the phosphorylation state of each transcript was determined by PABLO. RNA samples that had first been treated in vitro with an excess of purified EcRppH (PPase) were analyzed in parallel so that the ligation yields of the fully monophosphorylated transcripts could be used as correction factors for calculating the percentage of yeiP and yeiP-U2G that was monophosphorylated. The steady-state ratios of monophosphorylated to triphosphorylated mRNA reported in the text (mean values and standard deviations) were each calculated from measurements made on three independent RNA preparations.

This ratio, which reflects the relative values of the rate constants for RppH-dependent pyrophosphate removal and subsequent ribonucleolytic degradation (14), was measured by PABLO, a splinted ligation assay that takes advantage of the ability of T4 DNA ligase to covalently join a DNA oligonucleotide to the 5’ end of monophosphorylated RNA, but not triphosphorylated RNA, when the two are juxtaposed by simultaneous base pairing to a bridging oligonucleotide (4, 18). The ligation product and any unligated RNA that remains are then separated and quantified by gel electrophoresis and Northern blotting, and the percentage of 5’ ends that are monophosphorylated is calculated from the ratio of the ligation yields before and after enzymatic conversion of all 5’ ends to ligatable monophosphates.

In accordance with a previous report (17), the ratio of monophosphorylated to triphosphorylated yeiP 5’ ends was found to be 0.78 ± 0.02 in E. coli (Fig. 4), a sign that the rate constants for pyrophosphate removal and subsequent degradation are in close balance for this mRNA. Replacing the U ordinarily present at the second position of yeiP with G increased this ratio 3-fold to 2.4 ± 0.2, a finding consistent with more rapid pyrophosphate removal as a consequence of substituting the nucleotide that EcRppH prefers there. Nevertheless, although RppH is crucial for the rapid decay of yeiP mRNA (17), the U-to-G substitution had little effect on the half-life of this transcript in E. coli (data not shown), suggesting that formation of the monophosphorylated intermediate is not the rate-limiting event in the multistep process of yeiP degradation. The latter conclusion is consistent with the high relative abundance of this decay intermediate at steady state.

Phylogenetic Distribution of EcRppH and BsRppH Homologs—Alignment of the sequences of EcRppH and BsRppH shows them to be quite different, with only 23% identity (Fig. 5A). Much of what little similarity exists is localized in and around the sequence motif (GXE,EXE,RXE,EXE,G) characteristic of all Nudix hydrolases, irrespective of their function. As a result, many of the BsRppH residues implicated by x-ray crystallography in substrate binding (12) appear to have no counterpart in the sequence of EcRppH. However, superposition of the three-dimensional structures of BsRppH and EcRppH (12, 19) reveals a pair of nonhomologous residues, BsRppH Lys-97 and EcRppH Lys-140, that, despite their disparate sequence contexts, appear to be similarly situated in the active site of each enzyme.

Screening for orthologs by phylogenetic comparison reveals that recognizable homologs of EcRppH are widespread among 1,367 sequenced species representing all classes of Proteobacteria except Deltaproteobacteria, where they are very rare (Fig. 5, B and C). By contrast, the phylogenetic range of BsRppH homologs is much narrower, as they are present in 186 sequenced species of the order Bacillales but almost entirely absent from the closely related order Lactobacillales and from other classes of Firmicutes (Fig. 5C). In bacterial species that lack an identifiable EcRppH or BsRppH ortholog, sequence divergence makes it difficult to predict which of their Nudix hydrolases functions as an RNA pyrophosphohydrolase.

Sequence alignment of proteobacterial EcRppH homologs shows a high degree of conservation of the amino acid residues that line the active site, as defined by the structure of the unliganded enzyme determined by NMR and corroborated by x-ray crystallography (19, 20). Not only most of the residues characteristic of the Nudix motif but also 14 other residues are strictly conserved in virtually all of those orthologs (Fig. 5, A and B). According to the structure of ligand-associated EcRppH reported in the accompanying article by Vasilyev and Serganov (20), all but a couple of these residues are situated in close proximity to the bound substrate. Conserved residues Glu-53, Glu-56, Glu-57, and Glu-120 coordinate magnesium ions near the 5’ triphosphate of the trinucleotide ligand; Arg-8 and Gly-39 directly contact the triphosphate; Gln-37, Tyr-77, and Gln-95 interact with the phosphate between nucleotides 1 and 2; and Arg-27, Phe-139, and Lys-140 contact the base of the second nucleotide.

Alignment of the BsRppH orthologs in various Bacillales species reveals a greater number of strictly conserved residues (Fig. 5A), as expected from the closer evolutionary relationship of those species. According to the structure of ligand-associated BsRppH (12), most of these conserved residues line the substrate-binding pocket. In particular, Gly-52, Glu-68, Glu-72, and Glu-115 coordinate magnesium ions in the vicinity of the 5’ triphosphate; His-43 and Arg-46 make direct contact with the 5’ triphosphate; Lys-54 interacts with the phosphate between nucleotides 1 and 2; and Asp-6, Tyr-86, Lys-97, and Asp-141 surround the base of the second nucleotide.

Effect of Active Site Mutations in EcRppH—To assess the catalytic importance of the conserved EcRppH glutamate residues that coordinate Mg2+, we examined the effect of alanine substitutions at these positions on the ability of the purified enzyme to remove pyrophosphate from A4AGAA and A8XL (Fig. 6A). These experiments showed that mutating Glu-53, Glu-56, or Glu-57 abolished enzyme activity. By contrast, replacing Glu-120 with alanine had little effect.

We next investigated the importance of the five EcRppH residues that surround the base of the second nucleotide of substrate RNAs, as initially suggested by modeling and later substantiated by x-ray crystallography. These include the conserved residues Arg-27, Phe-139, and Lys-140 as well as Ser-32 and Val-137 (20).
FIGURE 5. RppH sequence alignment. A, alignment of EcRppH and BsRppH. The sequences were aligned by analysis with ClustalW. Asterisks mark amino acid residues that are identical. Residues that are conserved in virtually all bacterial orthologs of EcRppH or BsRppH are depicted as red or blue letters, respectively.

B, alignment of proteobacterial RppH orthologs. Sequences are shown for RppH from *E. coli*, *Legionella pneumophila*, *Burkholderia mallei*, *Azoarcus* sp. strain BH72, *Rhodobacter sphaeroides*, *Agrobacterium tumefaciens*, *Campylobacter jejuni*, and *Nautilia profundicola*, except that the nonhomologous carboxyl-terminal tails of the enzymes from *E. coli* (ENTPKPQNASAYRRKRG), *L. pneumophila* (RRTPYGLKRKRGNQRA), *B. mallei* (QRTDKSRGPRAPRYPRVANGHAASEAPAAIDTSVCSEVPGANALTRPRL), and *Azoarcus* sp. strain BH72 (KPAELPEGYRQGTASQA) have been omitted. Asterisks or colons mark amino acid residues that are identical or chemically similar, respectively. Residues that were mutated in the EcRppH variants characterized in Figs. 6 and 7 are depicted on a gray background. Numbers correspond to the sequence of EcRppH. C, evolutionary conservation of EcRppH and BsRppH orthologs. The dendrogram includes only bacterial phyla in which recognizable sequence homologs of EcRppH (p value 3.10^{-13} / H11349 3.10^{-11} / H11003 10^{-12} / H11002 10^{-14}) or BsRppH (p value 3.10^{-13} / H11349 3.10^{-11} / H11003 10^{-12} / H11002 10^{-14}) are common. The number of species within a class or order whose genomes were represented in the Blastp database and found to encode an ortholog of EcRppH (red) or BsRppH (blue) is indicated. Phylogenetic branching reflects RppH clades, as calculated by ClustalW2. Line lengths are arbitrary.
Specificity and Conservation of E. coli RppH

Although the S32A substitution mutant retained full activity toward A4AGAA and A8XL, the R27A, V137A, and K140M mutants were somewhat less active, and the F139A mutant lacked detectable activity (Fig. 6B). To determine whether these five amino acid residues influence the preference of EcRppH for a purine at position 2, the rate at which the same substitution mutants remove pyrophosphate from A4AGAA and A4ACAA was compared (Fig. 7). Interestingly, the ability of the enzyme to discriminate between G and C at that position was considerably diminished in the K140M mutant. By contrast, the S32A mutation increased the preference for G over C at position 2, as did an S32D or S32E substitution. Mutating Arg-27 or Val-137 had a more modest effect on the sequence bias of EcRppH, whereas the inactivity of the F139A mutant precluded a meaningful assessment of its specificity. Unexpectedly, the preference of the enzyme for G at position 2 was also enhanced by replacing the Mg2+-coordinating residue Glu-120 with alanine.

DISCUSSION

Both E. coli and B. subtilis can degrade some but not all mRNAs by a 5′-end-dependent mechanism in which the triphosphorylated 5′ end of primary transcripts is converted by RppH to a monophosphate amenable to ribonuclease binding. The specificity of this pathway for a subset of messages has been a subject of considerable interest. However, apart from the requirement for an unpaired 5′ terminus, nothing was known about the features of E. coli transcripts that govern their susceptibility to degradation by this mechanism. Here, we have shown that pyrophosphate removal by purified EcRppH requires at least two unpaired nucleotides at the 5′ end and prefers three or more, exactly as has been reported for BsRppH (11). However, unlike BsRppH, which has a strict requirement for G at the second position of its substrates (11, 12), EcRppH prefers three or more, exactly as has been reported for BsRppH (11). However, unlike BsRppH, which has a strict requirement for G at the second position of its substrates (11, 12), EcRppH prefers three or more, exactly as has been reported for BsRppH (11). However, unlike BsRppH, which has a strict requirement for G at the second position of its substrates (11, 12), EcRppH can act on unpaired 5′ ends with any nucleotide sequence. Nevertheless, it has a modest preference for A over G at position 1 and a greater preference for a purine at position 2.

The greater promiscuity of EcRppH in comparison with BsRppH requires at least two unpaired nucleotides at the 5′ end and prefers three or more, exactly as has been reported for BsRppH (11). However, unlike BsRppH, which has a strict requirement for G at the second position of its substrates (11, 12), EcRppH can act on unpaired 5′ ends with any nucleotide sequence. Nevertheless, it has a modest preference for A over G at position 1 and a greater preference for a purine at position 2.

The greater promiscuity of EcRppH in comparison with BsRppH is consistent with the smaller number and more flexible hydrogen bonding potential of the EcRppH side chains that interact with the edges of the base at position 2 (20). When bound to either enzyme, the planar surfaces of this base are sandwiched between hydrophobic or cationic residues that are conserved or conservatively substituted in their respective
The importance of the four glutamate residues that coordinate the three Mg\(^{2+}\) ions in the catalytic center of EcRppH has also been investigated. Although the three Mg\(^{2+}\)-coordinating glutamates that are part of the sequence motif characteristic of every Nudix hydrolase (Glu-53, Glu-56, and Glu-57) are all essential for activity, a fourth Mg\(^{2+}\)-coordinating glutamate residue (Glu-120) is not, even though a fourth Mg\(^{2+}\)-coordinating glutamate (Glu-115) is similarly situated in the catalytic center of BsRppH, and a homologous glutamate residue is reportedly important for the hydrolysis of Ap4A by an EcRppH homolog present in *Lupinus angustifolius* (12, 21). In view of crystallographic evidence that the two Mg\(^{2+}\) ions coordinated by Glu-120 are critical for enzyme function (20), it seems likely that their productive binding does not require this glutamate residue at the Mg\(^{2+}\) concentration used in these experiments, a postulate consistent with the absence of an additional Mg\(^{2+}\)-coordinating glutamate residue in some Nudix hydrolases (22, 23). Interestingly, an E120A substitution enhances the preference of EcRppH for G over C at position 2, presumably by an indirect mechanism because of the large distance between them.

Nearly all bacterial species contain multiple Nudix hydrolases with diverse functions. Although a Nudix hydrolase with RNA pyrophosphohydrolase activity has so far been authenticated only in the Gammaproteobacterium *E. coli*, the Deltaproteobacterium *Bdellovibrio bacteriovorus*, and the Firmicute *B. subtilis* (5, 6, 23), there is reason to believe that they are widespread in bacteria, as 5'-end-dependent RNA degradation has been observed in other species (24–26). However, the divergent evolution of these enzymes often makes them difficult to recognize solely on the basis of sequence homology, a complication that has hindered their discovery and characterization. Nevertheless, sequence homologs of EcRppH are readily identifiable in all classes of Proteobacteria except Deltaproteobacteria, including many important pathogens such as *Yersinia pestis*, *Vibrio cholerae*, *Neisseria meningitidis*, *Bordetella pertussis*, and *Helicobacter pylori*. Interestingly, they are also present in flowering plants, where they presumably originated by horizontal gene transfer from colonizing or infecting bacteria (most likely Alphaproteobacteria, judging by homology) and where their biological role is unknown.

Because four of the five EcRppH residues that surround the base of nucleotide 2 are generally conserved (Arg-27, Phe-139, and Lys-140) or conservatively replaced (Val-137) in its orthologs, these enzymes might be expected to share the relaxed sequence specificity of EcRppH. However, the identity of the fifth such residue, corresponding to Ser-32, is highly variable in proteobacterial homologs of EcRppH; although it is usually serine, it is often replaced by alanine and sometimes by a residue with a large hydrophobic side chain or a side chain that is negatively or positively charged. Therefore, our evidence that S32A, S32D, and S32E substitutions reduce the promiscuity of EcRppH raises the possibility that the sequence specificity of many of its orthologs may be more pronounced. Selectivity for the nucleotide at position 2 may also be affected by the protein loop in which Ser-32 is embedded. Anchored by the conserved residues Arg-27 and Trp-33, this loop is highly variable in length and sequence, which may influence the interaction of

---

**FIGURE 8. Interaction of EcRppH with the second base of RNA substrates.**

A, structure of the nucleotide-binding pocket of EcRppH with a guanine ligand. The illustration is based on the EcRppH:2MgpppAGU structure reported in the accompanying article by Vasilyev and Serganov (20). The base of the second RNA nucleotide is blue; EcRppH side chains that contact the edge of the base (Ser-32 and Lys-140) are red; EcRppH side chains that contact the planar surfaces of the base (Arg-27, Val-137, and Phe-139) are pink; and water is green. Dashed lines represent hydrogen bonds. B, hypothetical structure of the nucleotide-binding pocket of EcRppH with a cytosine ligand.

orthologs (Arg-27, Val-137, and Phe-139 in EcRppH; Val-88, Ile-95, Phe-137, and Ile-138 in BsRppH), and a lysine residue (Lys-140 in EcRppH; Lys-97 in BsRppH) forms hydrogen bonds with O6 and N7 of guanine (Fig. 8) (12, 20). Our data show that all four of these EcRppH residues contribute to enzyme activity, especially Phe-139. Furthermore, Lys-140 is important for the ability of EcRppH to discriminate between G and C at position 2, a finding consistent with the more favorable interaction of the positively charged side chain of this lysine residue with the carbonyl oxygen (O6) and imino nitrogen (N7) of guanine than with the amino group (N4) of cytosine. Besides Lys-140, the only other EcRppH residue that interacts with the edge of this base is Ser-32, which does so by accepting a hydrogen bond from the N2 amino group of guanine (Fig. 8). Because the hydroxyl-bearing side chain of this serine residue can either accept or donate hydrogen bonds, it has the potential to accommodate any of the four possible RNA bases. Changing Ser-32 either to alanine (whose side chain cannot hydrogen bond) or to aspartate or glutamate (both obligate hydrogen bond acceptors) increases the preference of EcRppH for G over C at position 2, possibly by preventing the protein from donating a hydrogen bond to either the carbonyl oxygen (O2) or imino nitrogen (N3) of cytosine and thereby eliminating all productive interactions with the edges of that base. In contrast, mutating EcRppH residues that engage the planar surfaces of the second base in cation-π or hydrophobic interactions (Arg-27 or Val-137, respectively) has only a small effect on substrate specificity. Unlike EcRppH, the nucleotide-binding pocket of BsRppH contacts the edge of guanine via hydrogen bond interactions involving not only Lys-97 but also three other residues (Asp-6, Tyr-86, and Asp-141) that are strictly conserved, two of which (Asp-6 and Asp-141) have negatively charged side chains that can accept but not donate hydrogen bonds (12). The number and specificity of these interactions likely explain the inability of BsRppH to tolerate other nucleotides at this position.
**Specificity and Conservation of E. coli RppH**

EcRppH homologs with the second nucleotide of potential substrates.

Orthologs of BsRppH are readily identifiable throughout the order Bacillales (e.g. *Bacillus anthracis, Staphylococcus aureus, and Listeria monocytogenes*) but not in Lactobacillales or other Firmicutes. Among the Bacillales homologs, the strict conservation of the four BsRppH residues that contact the guanine at position 2 of substrate RNAs (Asp-6, Tyr-86, Lys-97, and Asp-141) suggests that the stringent sequence specificity of BsRppH is also conserved. Although an ortholog of BsRppH is not recognizable in Lactobacillales, there nevertheless is evidence for 5′-end-dependent mRNA degradation in one such species, *Streptococcus pyogenes*, whose transcript can be stabilized by 5′-terminal base pairing (26). Therefore, Lactobacillales, which contain the 5′-monophosphate-dependent exonuclease RNase I, presumably also contain an RNA pyrophosphohydrolase, most likely a Nudix hydrolase with a divergent sequence.

In addition to BsRppH, *B. subtilis* appears to contain a second as yet unidentified RNA pyrophosphohydrolase that is indifferent to the identity of the second RNA nucleotide (6, 11). Its presence is thought to broaden the subset of transcripts that are capable of being degraded by a 5′-end-dependent mechanism, which would otherwise be limited to those in which the second nucleotide is G. By contrast, there is as yet no evidence for an alternative RNA pyrophosphohydrolase in *E. coli*, where the greater promiscuity of EcRppH may be adequate to allow targeting of a sufficient variety of cellular transcripts. Whether the presence of an additional RNA pyrophosphohydrolase is generally correlated with the sequence specificity of RppH in other bacterial species remains to be determined.

**Acknowledgments**—We are grateful to Nikita Vasilyev and Alexander Serganov for sharing information about the structure of EcRppH prior to publication; to Quansheng Liu for developing the methods that we used to examine the specificity of RppH; to Xiaoshan Shi, Atilio Deana, and Erik Sakowski for assistance with plasmid construction; and to Kevin Belasco and Monica Hui for help with phylogenetic analysis.

**REFERENCES**

1. Arnold, T. E., Yu, I., and Belasco, J. G. (1998) mRNA stabilization by the omega 5′ untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *RNA* 4, 319–330.

2. Baker, K. E., and Mackie, G. A. (2003) Ectopic RNase E sites promote bypass of 5′-end-dependent mRNA decay in *Escherichia coli*. *Mol. Microbiol.* 47, 75–88.

3. Laalami, S., Zig, L., and Putzer, H. (2014) Initiation of mRNA decay in bacteria. *Cell. Mol. Life Sci.* 71, 1799–1828.

4. Césenik, H., Deana, A., and Belasco, J. G. (2007) Initiation of RNA decay in *Escherichia coli* by 5′ pyrophosphate removal. *Mol. Cell* 27, 79–90.

5. Deana, A., Césenik, H., and Belasco, J. G. (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5′ pyrophosphate removal. *Nature* 451, 355–358.

6. Richards, J., Liu, Q., Pellegrini, O., Césenik, H., Yao, S., Bechhofer, D. H., Condon, C., and Belasco, J. G. (2011) An RNA pyrophosphohydrolase triggers 5′-exonucleolytic degradation of mRNA in *Bacillus subtilis*. *Mol. Cell* 43, 940–949.

7. Mackie, G. A. (1998) Ribonuclease E is a 5′-end-dependent endonuclease. *Nature* 395, 720–723.

8. Mitchell, S. J., and Minnick, M. F. (1995) Characterization of a two-gene locus from *Bartonella bacilliformis* associated with the ability to invade human erythrocytes. *Infect. Immun.* 63, 1552–1562.

9. Badger, J. L., Wiss, C. A., and Kim, K. S. (2000) Identification of *Escherichia coli* K1 genes contributing to human brain microvascular endothelial cell invasion by differential fluorescence induction. *Mol. Microbiol.* 36, 174–182.

10. Liu, H., Semino-Mora, C., and Dubois, A. (2012) Mechanism of *H. pylori* intracellular entry: an in vitro study. *Front. Cell. Infect. Microbiol.* 2, 13.

11. Hsieh, P. K., Richards, J., Liu, Q., and Belasco, J. G. (2013) Specificity of RppH-dependent RNA degradation in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8864–8869.

12. Piton, J., Larue, V., Thillier, Y., Dorlêans, A., Pellegrini, O., Li de la Sierra-Gallay, I., Vasseur, J. J., Debart, F., Tisné, C., and Condon, C. (2013) *Bacillus subtilis* RNA deprotection enzyme RppH recognizes guanosine in the second position of its substrates. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8858–8863.

13. Dorlêans, A., Li de la Sierra-Gallay, I., Piton, J., Zig, L., Gilet, L., Putzer, H., and Condon, C. (2011) Molecular basis for the recognition and cleavage of RNA by the bifunctional 5′-3′ exo/endonuclease RNase I. *J. Structure* 19, 1252–1261.

14. Luciano, D. J., Hui, M. P., Deana, A., Foley, P. L., Belasco, K. J., and Belasco, J. G. (2012) Differential control of the rate of 5′-end-dependent mRNA degradation in *Escherichia coli*. *J. Bacteriol.* 194, 6233–6239.

15. Emory, S. A., Bouvet, P., and Belasco, J. G. (1992) A 5′-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. *Genes Dev.* 6, 135–148.

16. Bouvet, P., and Belasco, J. G. (1992) Control of RNase E-mediated RNA degradation by 5′-terminal base pairing in *E. coli*. *Nature* 360, 488–491.

17. Richards, J., Luciano, D. J., and Belasco, J. G. (2012) Influence of translation on RppH-dependent mRNA degradation in *Escherichia coli*. *Mol. Microbiol.* 86, 1063–1072.

18. Césenik, H., Deana, A., and Belasco, J. G. (2008) PARLO analysis of RNA: 5′-phosphorylation state and 5′-end mapping. *Methods Enzymol.* 447, 83–98.

19. Bi, Y., Li, H., Fan, S., Xia, B., and Jin, C. (2009) 1H, 13C and 15N resonance assignments of RNA pyrophosphohydrolase RppH from *Escherichia coli*. *Biomol. NMR Assign.* 3, 149–151.

20. Vasilyev, N., and Serganov, A. (February 5, 2015) Structures of *E. coli* RppH-RNA complexes unveil the basis for specific 5′-end-dependent mRNA decay. *J. Biol. Chem.* 10.1074/jbc.M114.634824.

21. Maksel, D., Gooley, P. R., Swarbrick, J. D., Guranowski, A., Gange, C., Mildvan, A. S., and Amzel, L. M. (2004) Structure and mechanism of GDP-mannose glycosyl hydrolase, a Nudix enzyme that cleaves at carbon instead of phosphorus. *Biochem. J.* 379, 395–405.

22. Gabelli, S. B., Bianchet, M. A., Azurmendi, H. F., Xia, Z., Saravat, V., Mildvan, A. S., and Amzel, L. M. (2004) Structure and mechanism of GDP-mannose glycosylase, a Nudix enzyme that cleaves at carbon instead of phosphorus. *Structure* 12, 927–935.

23. Messing, S. A., Gabelli, S. B., Liu, Q., Césenik, H., Belasco, J. G., Pêrêno, S. A., and Amzel, L. M. (2009) Structure and biological function of the RNA pyrophosphohydrolase BdRppH from *Bdellovibrio bacteriovorus*. *Structure* 17, 472–481.

24. Heck, C., Rothfuchs, R., Jäger, A., Rauhut, R., and Klug, G. (1996) Effect of the pufQ-pufB intercistronic region on puf mRNA stability in *Rhodobacter capsulatus*. *Mol. Microbiol.* 20, 1165–1178.

25. Unniran, S., Chatterji, M., and Nagaraja, V. (2002) A hairpin near the 5′ end stabilizes the DNA gyrase mRNA in *Mycobacterium smegmatis*. *Nucleic Acids Res.* 30, 5376–5381.

26. Ramirez-Peña, E., Treviño, J., Liu, Z., Perez, N., and Sumby, P. (2010) The group A *Streptococcus* small regulatory RNA FdxS enhances streptokinese activity by increasing the stability of the ska mRNA transcript. *Mol. Microbiol.* 78, 1332–1347.