Bacillus subtilis RNA deprotection enzyme RppH recognizes guanosine in the second position of its substrates

Jérémie Piton^a, Valéry Larue^b, Yann Thillier^c, Audrey Dorléans^a, Olivier Pellegrini^a, Inés Li de la Sierra-Gallay^{a,1}, Jean-Jacques Vasseur^c, Françoise Debart^c, Carine Tisné^b, and Ciarán Condon^{a,2}

^aCentre National de la Recherche Scientifique (CNRS), Unité Propre de Recherche 9073 (affiliated with Université Paris Diderot, Sorbonne Paris Cité) Institut de Biologie Physico-Chimique, 75005 Paris, France; ^bUnité Mixte de Recherche (UMR) 8015, CNRS, Université Paris Descartes, Sorbonne Paris Cité, 75006 Paris, France; and ^cInstitut des Biomolécules Max Mousseron, UMR 5247, CNRS–Université Montpellier 1–Université Montpellier 2, 34095 Montpellier Cedex 05, France

Edited by Ben F. Luisi, University Cambridge, Cambridge, United Kingdom, and accepted by the Editorial Board March 15, 2013 (received for review January 7, 2013)

The initiation of mRNA degradation often requires deprotection of its 5' end. In eukaryotes, the 5'-methylguanosine (cap) structure is principally removed by the Nudix family decapping enzyme Dcp2, yielding a 5'-monophosphorylated RNA that is a substrate for 5' exoribonucleases. In bacteria, the 5'-triphosphate group of primary transcripts is also converted to a 5' monophosphate by a Nudix protein called RNA pyrophosphohydrolase (RppH), allowing access to both endo- and 5' exoribonucleases. Here we present the crystal structures of *Bacillus subtilis* RppH (BsRppH) bound to GTP and to a triphosphorylated dinucleotide RNA. In contrast to *Bdellovibrio bacteriovorus* RppH, which recognizes the first nucleotide of its RNA targets, the *B. subtilis* enzyme has a binding pocket that prefers guanosine residues in the second position of its substrates. The identification of sequence specificity for RppH in an internal position was a highly unexpected result. NMR chemical shift mapping in solution shows that at least three nucleotides are required for unambiguous binding of RNA. Biochemical assays of BsRppH on RNA substrates with single-base-mutation changes in the first four nucleotides confirm the importance of guanosine in position two for optimal enzyme activity. Our experiments highlight important structural and functional differences between BsRppH and the RNA deprotection enzymes of distantly related bacteria.

RNA stability | 5'-processing | RNA decapping

RNA turnover is a major target for the control of gene expression. Although the RNA maturation and degradation machineries of two of the best studied model bacteria, *Escherichia coli* and *Bacillus subtilis*, differ significantly (1, 2), they share an enzyme that deprotects the 5' ends of primary transcripts by converting the 5'-triphosphate group to a 5' monophosphate (3, 4). The 5' monophosphorylated RNA is a much better substrate for the major endoribonuclease RNase E in *E. coli* (5) and for the 5'-3' exoribonuclease RNase J1 in *B. subtilis* (6). The structural basis for this preference is known in both cases. Binding of a 5' monophosphate to a specific pocket stimulates the efficiency of RNase E cleavage at downstream sites, through a conformational change in the enzyme (7), whereas in the case of RNase J1, a 5' triphosphate is thought to reduce enzyme activity because the distance between the 5'-phosphate binding pocket and the active site is optimized for a nucleoside 5' monophosphate (8). The primary endoribonuclease of *B. subtilis*, RNase Y, has also been shown to prefer the 5'-monophosphorylated version of at least one RNA, the *yitJ* riboswitch, but the molecular basis for this preference is not yet known (9).

The enzyme responsible for RNA deprotection in both *E. coli* and *B. subtilis* is RNA pyrophosphohydrolase (RppH). This enzyme is a member of the very ancient family of nucleoside diphosphate linked to X (Nudix) hydrolases, involved in a wide range of important biological reactions, including the hydrolysis of ADP ribose, 8-oxoguanosine, and AppppA (10). Removal of the methylguanosine "cap" structure of eukaryotic mRNAs is also catalyzed

by a Nudix protein, decapping enzyme 2 (Dcp2) (11), suggesting that Nudix-mediated deprotection of RNA predates the evolutionary separation of bacteria and eukaryotes.

Nudix proteins have a characteristic signature motif GX₅EX₇REUXEEXGU (U is isoleucine, leucine, or valine and X is any residue) that forms a short α helix and contains the residues involved in metal (usually magnesium) ion binding (10). This family of enzymes has a characteristic fold consisting of two β sheets flanked by three α helices. Whereas *E. coli* RppH (EcRppH) and *Bdellovibrio bacteriovorus* RppH (BdRppH) catalyze the conversion of RNA 5' triphosphate to 5' monophosphate in a single step (i.e., with the release of pyrophosphate) (3, 12), the *B. subtilis* enzyme performs this reaction in two steps, releasing two phosphate ions (4). Intrigued by this difference and conscious of the importance of *B. subtilis* as an alternative biological model for bacterial mRNA decay, we resolved the crystal structure of the *B. subtilis* RppH (BsRppH) enzyme in its free form, as well as that bound to GTP and a 5'-triphosphorylated dinucleotide RNA. To corroborate these data, we compared the 2D NMR spectra (¹H-¹⁵N) of RppH bound to 1, 2, or 3 nucleotides of 5'-triphosphorylated RNA and we performed enzyme assays on longer RNAs bearing different nucleotides in the first four positions. We propose a model whereby the RNA is bound not only at the γ position of the 5'-triphosphate group, but also at the position of the second base, via a nucleotide binding pocket with a preference for guanosine residues.

Results

Structure of the BsRppH Nudix Domain. We first solved the crystal structure of a Glu68 to Ala catalytic mutant (E68A) of *B. subtilis* RppH bearing an N-terminal His-tag, at 2.2-Å resolution. BsRppH was not sufficiently similar to other Nudix proteins in the Protein Data Bank (PDB), including the *B. bacteriovorus* and *E. coli* RppH enzymes, to be able to solve the structure by molecular replacement. We thus used sodium-iodide-soaked crystals and phasing by single isomorphous replacement with anomalous scattering (SIRAS)

Author contributions: J.P., A.D., O.P., I.L.d.I.S.-G., J.-J.V., F.D., C.T., and C.C. designed research; J.P., V.L., Y.T., A.D., and O.P. performed research; Y.T., O.P., J.-J.V., and F.D. contributed new reagents/analytic tools; J.P., V.L., A.D., I.L.d.I.S.-G., C.T., and C.C. analyzed data; and J.P., C.T., and C.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. B.F.L. is a guest editor invited by the Editorial Board.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4JZ5, 4JZT, 4JZU, 4JZV).

See Commentary on page 8765.

¹Present address: Institut de Biochimie et Biophysique Moléculaire et Cellulaire, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8619, Bâtiment 430 Université Paris-Sud, 91405 Orsay, France.

²To whom correspondence should be addressed. E-mail: condon@ibpc.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221510110/-DCSupplemental.

to determine the structure. The structural data are presented in Table S1. Although the asymmetric unit contains a dimer, BsRppH is a monomer in solution (Fig. S14). This is in contrast to BdRppH, which forms a dimer in solution (12) and whose dimerization interface is very different from that seen in crystals of the *B. subtilis* enzyme (Fig. S1B).

Each of the monomers is composed of a typical Nudix fold (Fig. 1A), consisting of a central four-stranded mixed β sheet (β strands 1, 3, 4, and 5) and an antiparallel β sheet (β strands 2 and 6) sandwiched between three α helices ($\alpha 1$ and $\alpha 2$, $\alpha 3$). To facilitate comparisons, we have kept the same nomenclature as that of the previously published BdRppH enzyme (12). The topology of the secondary structures is shown in Fig. 1B. The ~ 20 N-terminal amino acids, which were not visible in the BdRppH structure (and are essentially absent from the *E. coli* sequence), form a long loop and two additional antiparallel β strands that we have called $\beta(-1)$ and $\beta(-2)$. These extend the central β sheet to six strands (shown in red in Fig. 1A and B) through six hydrogen bond interactions between residues in $\beta 4$ and $\beta(-1)$. The N-terminal domain is further stabilized by four potential salt bridges (Fig. 1C) and by hydrophobic interactions between residues in helix $\alpha 3$ and strands $\beta(-1)$ and $\beta(-2)$ (Fig. 1D). The Nudix motif is contained principally in helix $\alpha 1$ and contains all of the key residues for an active member of this family (Fig. S2). Despite the overall conservation of the Nudix topology, the structures of the RppH enzymes from *B. subtilis*, *E. coli*, and *B. bacteriovorus* are substantially different, independent of differences in the N-terminal domain (Fig. S3).

Structure of BsRppH Bound to GTP. We soaked crystals of BsRppH (E68A) in a solution of GTP, anticipating that the nucleotide would occupy a similar position to that seen with the BdRppH enzyme. Crystals of the resulting complex with GTP were resolved at 2.9-Å resolution by molecular replacement. In BdRppH, GTP binds close to the active site, with the α and β phosphates

coordinated by three Mg ions and the purine ring recognized by residues Pro52, Phe53, and Asn136 (12). In the *B. subtilis* enzyme, GTP clearly binds to a different pocket, about 8–10 Å distant from this site (Figs. 2 and 3A). The guanine ring is held in place by hydrogen bonds with Asp6, Tyr86, Lys97, Asp141, and a water molecule (Fig. S4) and the α phosphate by His27. The phosphate groups of GTP bound to this pocket are clearly too far from the active site to be a substrate for catalysis. The BdRppH guanosine binding pocket does not exist in the *B. subtilis* enzyme and is, in fact, largely occupied by a tryptophan residue (Trp29). Furthermore, Asn136, which plays a key role in guanosine recognition in BdRppH, has no functional equivalent in the *B. subtilis* enzyme (Table S2). No metal ions were visible in the E68A mutant enzyme, which was not unexpected as the equivalent residue in BdRppH (E70) coordinates two of the three metal ions.

Crystal Structure(s) of BsRppH Bound to a Dinucleotide. Wary of the possibility that the absence of Mg ions in the mutant enzyme might have an effect on the position of GTP and curious to know how this enzyme dephosphorylates RNA, we attempted to generate complexes between the nontagged wild-type enzyme and mono-(G), di-(GG), or trinucleotides (GGA), bearing 5'-triphosphate groups. To prevent hydrolysis of these substrates, we used 5' nucleotides bearing methylene groups between the γ and β phosphates (pcp). We successfully obtained crystals of wild-type BsRppH bound to the pcp-pGpG dinucleotide in a different space group (Table S1). The asymmetric unit was also a dimer, but different from both that of the first space group and that of BdRppH (Fig. S1B), further evidence that BsRppH functions as a monomer in solution.

Two different complexes were obtained (in the same drop) that contained either the first or second guanosine residue in the nucleotide binding pocket. These structures were resolved by molecular replacement at 1.7-Å and 2.2-Å resolution, respectively. Although only one G residue was visible in each structure, in each case bound to the nucleotide pocket, the positions of the 5' and 3' phosphates give important clues as to how BsRppH binds and dephosphorylates RNA. The complex with the first G residue (G1) in the nucleotide pocket shows the 5' nucleotide in an identical conformation to that of GTP bound to the E68A mutant, with the γ and β phosphates too far from the active site for hydrolysis (Fig. 3B). Furthermore, no metal ions were visible in the active site of this structure, despite the fact that the enzyme was wild type, providing an additional indication that this is a nonproductive complex. One important difference with the GTP-bound complex, however, is that we can see the position of the 5'-phosphate group of the second residue, which indicates that the path of the RNA molecule is toward the N-terminal domain of RppH. This phosphate group is principally in interaction with Asn10 in strand $\beta(-1)$.

The structure containing the second guanosine residue (G2) in the nucleotide binding pocket is likely to represent the productive enzyme/substrate complex (Fig. 3C). In this structure, Gly54 and Glu72 coordinate one metal ion (Mg1), whereas Glu68, Glu72, and Glu115 coordinate a second (Mg2) (Fig. 3D). These metal ions are in similar positions to two of the three Mg ions in BdRppH, but play different roles. In BdRppH, the metal ions share coordination of the α and β phosphates (Fig. 3E), whereas in BsRppH, only the γ phosphate is coordinated by the magnesium ions (Fig. 3D). Although the α , β , and γ phosphates of G1 are clearly visible in this structure (Fig. 3C), the ribose and base moieties are not, suggesting that they are flexible. We can also clearly see the 5' monophosphate of G2 in the nucleotide pocket (Fig. 3C). These experiments clearly show that the pppGG dinucleotide can have at least two different configurations on BsRppH and that longer molecules are likely required for efficient enzyme function.

NMR Chemical Shift Mapping upon Binding of Oligonucleotides to ^{15}N -Labeled BsRppH. Given the ambiguous binding of the pppGG dinucleotide in crystals, we turned to NMR footprinting experiments

Fig. 1. Structure of BsRppH. (A) Cartoon view of BsRppH showing the classical Nudix domain in dark blue (β strands) and yellow (α helices). The N-terminal domain, not visible in the BdRppH structure and absent from the EcRppH sequence, is shown in red. Strands and helices are labeled as for the BdRppH structure, with additional N-terminal strands labeled as $\beta(-1)$ and $\beta(-2)$. (B) Topology of BsRppH. Strands of the classical Nudix domain are shown in dark blue, helices in yellow and the N-terminal extension in red. (C) Salt bridges (green dotted lines) and hydrogen bonds (gray dotted lines) stabilizing N-terminal domain (red). (D) Hydrophobic interactions (blue arc) between helix $\alpha 3$ and N-terminal domain (red).

Fig. 2. The nucleotide binding pocket of BsRppH is in a different location to the *B. bacteriovorus* enzyme. Superposition of BsRppH (green) and BdRppH (gray) structures. GTP residues bound to each enzyme are labeled. The three Mg ions of the BdRppH enzyme are shown as orange spheres.

to determine how RppH binds RNAs in solution. We performed NMR chemical shift mapping using heteronuclear (^1H - ^{15}N) transverse relaxation optimized spectroscopy (TROSY) on wild-type ^{15}N -labeled BsRppH bound to pcp-GMP, pcp-pGpG, and pcp-pGpGpA. Using ^{15}N - ^{13}C -labeled BsRppH, we were able to assign TROSY peaks to the backbone amide groups of about two-thirds of BsRppH residues (104 of 158). Binding of pcp-GMP to BsRppH caused a major perturbation, i.e., disappearance, of seven peaks corresponding to residues near the guanosine binding pocket, consistent with the crystallography data (Fig. 4A and Table 1). A further 15 peaks showed significant chemical shift variations (≥ 0.05 ppm), including four residues (8–11) in the N-terminal domain near where GTP binds. The 2D spectra are shown in Fig. S5.

Binding of the pcp-pGpG dinucleotide caused six additional major chemical shift perturbations and 10 supplementary chemical shift variations in the TROSY spectrum of BsRppH (Table 1). Consistent with the observation that the dinucleotide can bind

BsRppH in two different ways, most of these peaks also corresponded to residues clustered around the guanosine binding site (Fig. 4B). Indeed many of the chemical shifts were a reinforcement of displacements already seen with pcp-GMP, notably residues 8–10 of the N terminus, Val55 and Ile138. Only a strong displacement of peaks corresponding to Glu44 and Glu115 gave an indication of an interaction with residues near the catalytic site. Glu115 is one of two potential catalytic bases in BsRppH (Discussion).

More dramatic chemical shift changes were seen when the trinucleotide pcp-pGpGpA was added to BsRppH; a further 13 peaks “disappeared” compared with the dinucleotide (Table 1). The majority of the “new” shifts correspond to residues close to the catalytic site, notably residues 45–47, 72–74, and many between 112 and 118 (Fig. 4C), suggesting that the RNA molecule has found its correct niche and confirming the idea that at least three residues are required to prevent ambiguous binding. The middle G residue of the GGA trinucleotide is the most likely residue to be in the nucleotide binding pocket in this complex, because the crystal structure suggests there is only enough space to accommodate one nucleotide between the binding pocket and the active site. Three peaks corresponding to residues Tyr8, Gln9, and Ala102 showed additional chemical shifts of intermediate proportions upon binding pppGGA. Indeed Tyr8 and Gln9 in the N terminus showed increased peak displacements with each added nucleotide (Fig. 4) consistent with these residues interacting optimally with the third residue of the trinucleotide RNA.

Guanosine Is Preferred in the Second Position of BsRppH Substrates.

We were unsuccessful in attempts to obtain crystals of BsRppH with ATP in the nucleotide binding pocket under conditions identical to those that were effective for GTP. Guanine can also form more hydrogen bonds (six) with the key residues of the pocket than any of the other three bases (Fig. S4). These observations, along with the crystal structure of the productive enzyme–substrate complex, suggested a general preference for guanosine in the second position of BsRppH substrates. To test this idea, we made variants of a 280-nt RNA molecule, synthesized by T7 RNA polymerase, that we had previously identified as a substrate of BsRppH in vitro. The substrate begins with a GGGA sequence and first 9 nucleotides are predicted to be in a single stranded conformation. We changed positions 2 or 3 of this RNA to each of the other three possible bases A, C, or U. We also made variants in the first (AGGA) and fourth positions (GGGU) as controls. The different substrates were γ - ^{32}P -labeled at the 5' end and subjected to hydrolysis by increasing concentrations of BsRppH. The products were then run on a 20% polyacrylamide gel to resolve the liberated γ phosphate (inorganic phosphate, Pi). In these experiments, BsRppH showed about a 5- to 10-fold preference for guanine in position 2 compared with the other three bases, whereas the identity of the base in positions 1, 3, and 4 had much smaller effects on enzyme activity (Fig. 5). Although we cannot rule out the possibility that some of these single-base mutations have an effect on the secondary structure of the 5' end of the RNA and thus the amount of Pi released, the fact that three different mutations in position 2 all significantly reduce activity is strong evidence in support of a key role for guanosine in the second position of BsRppH substrates. A similar conclusion was reached in a parallel study in the companion article by Hsieh et al. (13). Our data are also consistent with the possibility that the enzyme has a modest preference for A in position 3, as observed in the accompanying paper (13).

Discussion

We have solved the crystal structures of the *B. subtilis* RNA pyrophosphohydrolase BsRppH alone, and in complex with GTP and a triphosphorylated RNA dinucleotide (pppGG). The latter is the first complex of a Nudix protein with an RNA substrate bound in the catalytic site. Although BsRppH has a classical Nudix fold, the *B. subtilis* enzyme has important characteristics that distinguish it from the previously solved structures of *B. bacteriovorus* and *E. coli* RppH, the main difference being the

Fig. 3. Nucleotide and phosphate binding by BsRppH. Surface plots of BsRppH showing omit maps (Fo-Fc) of electron density corresponding to (A) GTP, (B) pcp-pGpG, with G1 in the nucleotide binding pocket, and (C) pcp-pGpG, with G2 in the nucleotide binding pocket, at 2.0σ above the mean. Metal ions are shown as orange spheres. (D) Metal ion coordination of phosphate residues in the active site of BsRppH. (E) Metal ion coordination of phosphate residues in the active site of BdRppH.

Fig. 4. Chemical shift mapping of BsRppH bound to one, two, or three nucleotides. Residues showing chemical shift variations upon binding to (A) pcp-GMP (B) pcp-pGpG, and (C) pcp-pGpGpA. Key residues are labeled, and the locations of the metal ions and GTP binding pocket are indicated. Assigned residues are in green and unassigned residues in gray. For each mixture, the average chemical shift variation (δ) was calculated (0.025 ppm for A, 0.07 ppm for B, and 0.08 ppm for C). Residues showing chemical shift variations are shown in colors ranging from beige ($\geq 2\delta$) to orange ($\geq 4\delta$) to violet ($\geq 6\delta$) to red (disappearance), with the strength of the shift indicated both by the color (see gradient) and thickness of the backbone ribbon.

primary site of nucleotide recognition. Whereas the *B. bacteriovorus* enzyme has a binding pocket for the first residue of its RNA substrate, the *B. subtilis* enzyme primarily recognizes the second base. Although no structure of the *E. coli* enzyme bound to a nucleotide is currently available, the key residues of the *B. subtilis* nucleotide binding pocket are absent in *E. coli*, suggesting that its RNA recognition mechanism is also different from BsRppH. Further differences were seen in the N-terminal domain, which was not visible in the BdRppH structure and is absent from EcRppH, and in the loop structures of the three enzymes. Thus, although they are obviously derived from a common Nudix ancestor, and share overall topology, they have clearly had time to evolve into recognizably different structures with interesting divergent features. One such divergent property is that BdRppH forms dimers in solution, whereas BsRppH is a monomer. Indeed, the *B. subtilis* enzyme cannot form the same dimer as BdRppH because of a steric clash between the N-terminal domains. The crystallography data suggested that the RNA path away from the catalytic site is likely to be via the N-terminal domain in BsRppH; this RNA exit pathway is also likely to be available in the *Bdellovibrio* enzyme despite its dimerization.

A second major contrasting feature is that both the Bd and EcRppH enzymes release pyrophosphate from the 5' ends of primary transcripts (3, 12), whereas the *B. subtilis* enzyme catalyzes

this reaction in two steps, releasing two phosphates (4). The Nudix nucleoside triphosphatase YmdB of *E. coli* has also been shown to produce Pi instead of pyrophosphate (PPI), but the basis for this mechanism is unknown (14). The difference in the catalytic mechanisms of the Bd and BsRppH enzymes is likely explained by the difference in phosphate coordination by the metal ions of each protein. The α and β phosphates of GTP are coordinated by Mg in BdRppH, favoring cleavage between them and liberation of PPI (Fig. 3E). In contrast, only the γ phosphate is coordinated by metal ions in BsRppH (Fig. 3D) and thus the enzyme must proceed in a sequential manner. The flexibility of the base and sugar moiety of the first nucleotide may facilitate the repositioning of the β phosphate adjacent to the metal ions, after the γ phosphate has left, for the second round of catalysis.

Catalytic mechanisms have been proposed for several Nudix hydrolases, among them the mutator phosphohydrolase MutT, ADP ribose pyrophosphatase (ADPRP), diadenosine tetraphosphate pyrophosphatase (Ap4AP), and GDP-mannose mannosyl hydrolase (GDPMH) (10). These can be divided into two major classes based on the location of the catalytic base. In MutT, Ap4AP, and *Thermus thermophilus* (Tt) ADPRP (10, 15, 16), the catalytic base is thought to be the second glutamate residue of the Nudix signature sequence (Glu68 in BsRppH), whereas in EcADPRP and GDPMH the catalytic base is either a glutamate

Table 1. Residues corresponding to displaced TROSY peaks upon substrate binding to BsRppH

Substrate	Residue
pcp-pG	Trp29 (NE1), Val94, Ile95, Val96, Tyr100, Phe137, Leu144, <i>Tyr8, Gln9, Asn10, Thr11, Gly53, Val55, Leu83, Gln85, Val88, Asn98, Ile99, Ile138, Lys140, Asp141, Lys152</i>
pcp-pGpG	Trp29 (NE1), Val94, Ile95, Val96, Tyr100, Phe137, Leu144, Glu44, Val55, Leu89, Asn98, Glu115, Ile138, <i>Tyr8, Gln9, Asn10, Gly53, Leu83, Gln85, Ile99, Asp112, Thr116, Lys140</i>
pcp-pGpGpA	Trp29 (NE1), Val94, Ile95, Val96, Tyr100, Phe137, Leu144, Glu44, Val55, Leu89, Asn98, Glu115, Ile138, Asp45, Arg46, Gly47, Gly52, Gly53, Glu72, Thr73, Gly74, Asp112, Phe114, Thr116, Lys117, Gly118, <i>Asn10, Leu83, Gln85, Ile99,</i> <i>Tyr8, Gln9, Ala102</i>

Boldface represents peaks corresponding to residues that disappear upon substrate binding whereas non-boldface residues show intermediate chemical shifts (two or more times average measurable shift). Residues showing a significant chemical shift relative to the previous substrate ($n - 1$ nt) are in italics.

Fig. 5. BsRppH has a preference for guanosine in position 2. (A) Representative autoradiograms of BsRppH reactions in vitro. The portions of the gel corresponding to the full-length (FL) γ - ^{32}P -labeled RNA substrate and inorganic phosphate (Pi) product are shown. The sequences of the first four nucleotides of the 280-nt substrates are shown above the autoradiogram, with mutations underlined. Right angled triangles indicate direction of increasing enzyme concentration (0.1, 0.3, and 1 μM). Specific activities were calculated from the lowest concentration of enzyme giving a visible product (Pi). Disappearance of the substrate is not a good indicator of enzyme activity; at higher enzyme concentrations, RppH forms a visible complex with the substrate that has difficulty entering the gel. (B) Histogram showing quantification of three independent experiments similar to that shown in A with SE as shown.

or a histidine residue located in the loop between strands $\beta 5$ and $\beta 6$ (17, 18). His116 in this loop has been proposed as a potential catalytic base for BdRppH (12) and the equivalent residue in BsRppH is Glu115 (Table S2). Both Glu68 and Glu115 are hydrogen bonded to the same water molecule, which also serves as a ligand for Mg2, in BsRppH (Fig. 3D). This water molecule is a potential candidate to perform the nucleophilic attack on the γ phosphate of BsRppH upon activation by Glu68 or Glu115 (or both).

Based on the two different crystal structures of BsRppH bound to the pppGG dinucleotide and taking into account the NMR chemical shift mapping data of BsRppH bound to trinucleotide, we have built a working model of the enzyme bound to a pppGGA sequence (Fig. 6). The position of the ribose moiety of the first guanosine residue was determined by the positions of the α phosphates of the first and second residues, which were clearly defined in the two crystal structures of BsRppH bound to pppGG. Only weak and incomplete density is visible for the G1 base (not included in the deposited PDB file). Nevertheless, this position of the G1 base is also supported by the NMR chemical shift data, which shows a major perturbation of the peaks corresponding to the nearby residues Asp45, Arg46, and Gly47 in the loop between

strands $\beta 2$ and $\beta 3$. The 5' phosphate of the G2 residue was clearly visible in the crystal structure with G1 in the nucleotide binding pocket. This defines the position of the α phosphate of the adenosine residue (A3) in the pppGGA trinucleotide. The NMR chemical shift data show a progressive increase in the chemical shift variation of residues in the loop between strands $\beta(-1)$ and $\beta(-2)$ in the N-terminal domain, notably Tyr8 and Gln9, with each added nucleotide. In the model of pppGGA bound to BsRppH, we have stacked the A3 base against the tyrosine ring to take into account the NMR data; other orientations of the sugar and base are possible, however.

There are currently over 70 distinct Nudix proteins in the PDB, with about 15 different known or predicted functions. It is remarkable that the Bs, Ec, and BdRppH structures were sufficiently different that it was not possible to solve the BsRppH structure by molecular replacement. Indeed a BLAST search for homologs to BsRppH in *E. coli* only identifies EcRppH in fifth position (24% identity) among 13 Nudix proteins encoded by its genome; it is more closely related to *E. coli* MutT at the sequence level (32% identity). Similarly, a Dali search (http://ekhidna.biocenter.helsinki.fi/dali_server/) of the PDB using the 3D structure of BsRppH ranks seven *E. coli* Nudix proteins ahead of EcRppH by z score. It is thus impossible to distinguish RppH from other Nudix proteins in distantly related bacteria by sequence or structure alignment; experimental validation is required.

Our data and that of our colleagues (13) showed that BsRppH has a preference for guanosine in the second position of its substrates. Indeed, two previously identified substrates of BsRppH, the *yhxA-glpP* (regulation of glycerol metabolism) and *ermC* (erythromycin resistance) mRNAs have a G in position 2. Preference for a nucleotide in an internal position of the RNA substrate has not been seen previously among Nudix family decapping proteins and suggests that RppH has a preference for a subset of *B. subtilis* RNAs. The start points of about 600 *B. subtilis* primary transcripts have recently been identified by differential RNA sequencing (dRNA-seq) at single nucleotide resolution (19). We examined nucleotide distribution over the first 10 positions of these 600 transcripts to see whether there was a preference for particular nucleotides in position 2. The first 10 nucleotides are particularly rich in A (41%) and poor in C (10%) residues (Fig. S6). Only the first two positions were dramatically different from the others; from positions 3–10, the distribution was relatively homogenous, but still not quite the same as the genome average.

Fig. 6. Hypothetical model of BsRppH bound to a trinucleotide RNA. Electrostatic surface map of BsRppH bound to pppGGA. Positively charged surfaces are shown in blue, negatively charged surfaces in red. Metal ions are shown as orange spheres. Nucleotides are labeled according to their position relative to the 5' end.

As expected, A (56%) and G (33%) were strongly preferred as the starting nucleotide. Position 2, on the other hand, showed a strong preference for A (37%) or T (40%) and there were relatively few G residues (15%) compared with positions 3–10 (21%). A similar distribution has been seen in *E. coli*, where G is even less frequent in position 2 (20). This suggests a counter selection for guanosine residues in position 2 among both *B. subtilis* and *E. coli* primary transcripts. It is interesting to speculate that this may have been influenced by the presence of BsRppH in *B. subtilis* over the course of evolution. We have previously suggested that *B. subtilis* contains at least one additional RNA pyrophosphohydrolase (4). The fact that BsRppH preferentially dephosphorylates only a subset of RNA substrates suggests that this second enzyme may be of major importance.

Materials and Methods

Protein Production and Purification. The *B. subtilis* *rppH* gene was amplified by PCR using oligonucleotide pairs CC546/547 and CC579/580, digested with NdeI/BamHI and BamHI/Sall, respectively, and cloned in pET28a cut with the same enzymes. The resulting plasmids expressed N-terminal and C-terminal His-tagged derivatives of BsRppH. A plasmid expressing an E68A mutant derivative of N-terminal His-tagged BsRppH was made by site-directed mutagenesis (QuikChange) using oligos CC566 and CC567. A plasmid expressing a nontagged wild-type version of BsRppH was made by introducing a stop codon before the C-terminal His-tagged version by site-directed mutagenesis (QuikChange) using oligos CC979 and CC980. BsRppH was overproduced in BL21 CodonPlus cells by induction with 0.5 mM isopropylthio- β -galactoside for 4 h. Full details of purification are provided in *SI Materials and Methods*.

RNA Synthesis. Oligonucleotides pcp-pGpG and pcp-pGpGpA were chemically synthesized on solid supports, triphosphorylated, deprotected, and purified by HPLC. Full details are provided in *SI Materials and Methods*. GMP-pcp was purchased from Sigma.

Crystallization of BsRppH. Preliminary crystallization trials were performed at 293K by sitting drop vapor diffusion and manually optimized by vapor diffusion in hanging drops containing 1 μ L of reservoir solution and 1 μ L of protein. N-terminal His-tagged BsRppH (E68A) crystallized at 15 mg/mL in 1.9 M ammonium sulfate, 0.1 M Tris pH 8.3, with 10 mM Tris(2-carboxyethyl)phosphine (TCEP) hydrochloride as an additive in the mixed drop. Complexes of BsRppH (E68A) with GTP were obtained by soaking the crystals in a solution containing 1.9 M ammonium sulfate, 0.1 M Tris pH 8.3, 10 mM TCEP hydrochloride, and 50 mM GTP for 24 h. Cocrystals of the complex between wild-type BsRppH and pcp-pGpG were obtained in 100 mM Hepes pH 7.5, 25% (wt/vol) PEG 1000 (condition C3 from the JBScreen Classic 1 kit) with a protein concentration of 18 mg/mL and a 1:1 ratio of pcp-pGpG. Details of data collection, structure determination, and refinement are provided in *SI Materials and Methods*. Images were produced using PyMol software (DeLano Scientific). The coordinates and structure factors have been deposited in the Brookhaven Protein Data Bank (PDB ID codes 4JZS, 4JZT, 4JZU, 4JZV).

NMR Chemical Shift Mapping. NMR spectra were recorded at 20 °C on a Bruker Avance 600 MHz spectrometer equipped with a TCI 5-mm cryoprobe. The $^{15}\text{N}/^{13}\text{C}$ BsRppH sample was prepared at a concentration of 1 mM in a 20 mM Na_2HPO_4 buffer pH 7.0 containing 2.5 mM MgCl_2 , 0.5 mM DTT, and 10% (vol/vol) $^2\text{H}_2\text{O}$. Backbone assignments were obtained using a set of standard 3D NMR experiments [HNCA, HNCACB, CBCA(CO)NH, HNCO, HNCACO (where H is hydrogen, N is nitrogen, CA is alpha-carbon, CB is beta-carbon and CO is main-chain carbonyl), NOESY-HSQC (nuclear Overhauser effect spectroscopy–heteronuclear single quantum correlation), and TOCSY (total correlation spectroscopy)-HSQC]. Chemical shift mappings of the interaction between RNA and BsRppH were obtained using ^{15}N -labeled BsRppH (0.25 mM) mixed with one equivalent of pcp-GMP, pcp-pGpG, or pcp-pGpGpA. For each mixture, one TROSY experiment (21) was recorded. NMRpipe (22) and Sparky (23) software were used to process and analyze NMR data. Chemical shift differences $\Delta(\text{H},\text{N})$ were derived from ^1H and ^{15}N chemical shifts: $\Delta(\text{H},\text{N}) = \sqrt{[(\Delta^{15}\text{N} W_{\text{N}})^2 + (\Delta^1\text{H} W_{\text{H}})^2]}$, where $\Delta = \delta_{\text{complex}} - \delta_{\text{free}}$ and $W_{\text{H}} = 1$ and $W_{\text{N}} = 1/6$.

BsRppH Assays in Vitro. Templates for in vitro transcription using T7 RNA polymerase were made by PCR, with the upper oligonucleotide containing the T7 promoter sequence. Wild-type (CC169) and mutagenic (CC1124–1134) oligonucleotides were paired with CC170 (Table S3). The sequence amplified extended from 38 nt upstream to 240 nt into the 16S rRNA gene of *rrnW* (ribosomal RNA operon) and has been described previously (6). Labeled (γ - ^{32}P -GTP) substrate RNAs were prepared by in vitro transcription using T7 RNA polymerase (Promega). In a first incubation step (15 min at 37 °C), only γ - ^{32}P -GTP and ATP, CTP, and UTP (0.5 mM final concentration) were added. Cold GTP (0.5 mM final concentration) was then added and incubation continued for another 90 min.

RNA phosphohydrolase assays were performed using wild-type BsRppH enzyme at 1, 0.3, and 0.1 μ M final concentration in 5- μ L reactions in 20 mM Mes pH 6, 100 mM NH_4Cl , 5 mM MgCl_2 , and 0.1 mM DTT. Reactions were incubated at 37 °C for 30 min and stopped by addition of 5 μ L 95% (vol/vol) formamide, 20 mM EDTA, 0.05% (wt/vol) bromophenol blue, 0.05% (wt/vol) xylene cyanol, and run on 20% (wt/vol) polyacrylamide/7 M urea gels. The amount of radiolabeled Pi released was measured by phosphor imaging using a Typhoon apparatus (GE Healthcare).

ACKNOWLEDGMENTS. We thank European Synchrotron Radiation Facility ID14eh1 and ID23eh1 beamline staff for assistance in data collection; V. Normand for technical help; D. Picot, W. Winkler, and members of our laboratories for helpful discussions; and P. Weber and A. Haouz (Platform 6, Institut Pasteur) for performing robot-driven crystallization trials. This work was supported by funds from the Centre National de la Recherche Scientifique (Unité Propre de Recherche 9073 and Unité Mixte de Recherche 8015), Université Paris VII-Denis Diderot, Université Paris Descartes, and the Agence Nationale de la Recherche (subtilRNA2). This paper is dedicated to the memory of NAS member Marianne Grunberg-Manago, who made important contributions to sciences, and in particular, the solving of the genetic code.

- Condon C (2007) Maturation and degradation of RNA in bacteria. *Curr Opin Microbiol* 10(3):271–278.
- Condon C (2009) RNA processing in bacteria. *Encyclopedia of Microbiology*, ed Schaechter M (Elsevier, Oxford), 3rd Ed, Vol 5, pp 395–408.
- Deana A, Celesnik H, Belasco JG (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* 451(7176):355–358.
- Richards J, et al. (2011) An RNA pyrophosphohydrolase triggers 5'-exonucleolytic degradation of mRNA in *Bacillus subtilis*. *Mol Cell* 43(6):940–949.
- Mackie GA (1998) Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* 395(6703):720–723.
- Mathy N, et al. (2007) 5'-to-3' exoribonuclease activity in bacteria: Role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* 129(4):681–692.
- Callaghan AJ, et al. (2005) Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature* 437(7062):1187–1191.
- Dorléans A, et al. (2011) Molecular basis for the recognition and cleavage of RNA by the bifunctional 5'-3' exo/endoribonuclease RNase J. *Structure* 19(9):1252–1261.
- Shahbadian K, Jamali A, Zig L, Putzer H (2009) RNase Y, a novel endoribonuclease, initiates riboswitch turnover in *Bacillus subtilis*. *EMBO J* 28(22):3523–3533.
- Mildvan AS, et al. (2005) Structures and mechanisms of Nudix hydrolases. *Arch Biochem Biophys* 433(1):129–143.
- Dunckley T, Parker R (1999) The DCP2 protein is required for mRNA decapping in *Saccharomyces cerevisiae* and contains a functional MutT motif. *EMBO J* 18(19):5411–5422.
- Messing SA, et al. (2009) Structure and biological function of the RNA pyrophosphohydrolase BdRppH from *Bdellovibrio bacteriovorus*. *Structure* 17(3):472–481.
- Hsieh P-K, Richards J, Liu Q, Belasco JG (2013) Specificity of RppH-dependent RNA degradation in *Bacillus subtilis*. *Proc Natl Acad Sci USA* 110:8864–8869.
- Xu W, Dunn CA, O'handley SF, Smith DL, Bessman MJ (2006) Three new Nudix hydrolases from *Escherichia coli*. *J Biol Chem* 281(32):22794–22798.
- Harris TK, Wu G, Massiah MA, Mildvan AS (2000) Mutational, kinetic, and NMR studies of the roles of conserved glutamate residues and of lysine-39 in the mechanism of the MutT pyrophosphohydrolase. *Biochemistry* 39(7):1655–1674.
- Ooga T, Yoshida S, Nakagawa N, Kuramitsu S, Masui R (2005) Molecular mechanism of the *Thermus thermophilus* ADP-ribose pyrophosphatase from mutational and kinetic studies. *Biochemistry* 44(26):9320–9329.
- Gabelli SB, et al. (2002) Mechanism of the *Escherichia coli* ADP-ribose pyrophosphatase, a Nudix hydrolase. *Biochemistry* 41(30):9279–9285.
- Legner PM, Massiah MA, Mildvan AS (2002) Mutational, kinetic, and NMR studies of the mechanism of *E. coli* GDP-mannose mannosyl hydrolase, an unusual Nudix enzyme. *Biochemistry* 41(35):10834–10848.
- Irnov I, Sharma CM, Vogel J, Winkler WC (2010) Identification of regulatory RNAs in *Bacillus subtilis*. *Nucleic Acids Res* 38(19):6637–6651.
- Hawley DK, McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res* 11(8):2237–2255.
- Pervushin KV, Wider G, Wüthrich K (1998) Single transition-to-single transition polarization transfer (ST2-PT) in [^{15}N , ^1H]-TROSY. *J Biomol NMR* 12(2):345–348.
- Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293.
- Goddard TD, Kneller DG (2008) SPARKY 3. (Univ of California, San Francisco).