NMR Structure of an Archaeal Homologue of Ribonuclease P Protein Rpp29[†]

David J. Sidote and David W. Hoffman*

Department of Chemistry and Biochemistry, Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas 78712

Received July 14, 2003

ABSTRACT: A protein component of the *Archaeoglobus fulgidus* RNase P was expressed in *Escherichia coli*, purified, and structurally characterized using multidimensional NMR methods. The dominant structural feature of this 11 kDa protein is a sheet of six antiparallel β-strands, wrapped around a core of conserved hydrophobic amino acids. Amide proton exchange and ^{15}N relaxation rate data provide evidence that the first 16 residues of the protein, located before the start of the first β-strand, and the last 24 residues, located past the end of the last β-strand, are relatively flexible; this contrasts with the relatively rigid and well-defined structure of the β-sheet. Amino acid sequence comparisons among a diverse set of species indicate that the *A. fulgidus* protein is homologous to the human RNase P protein Rpp29, yeast RNase P protein Pop4, and a known archaeal RNase P protein from *Methanobacter thermoautotrophicus*; conserved hydrophobic residues indicate that the homologous protein in each of these species contains a similar β-sheet structure. Conserved surface residues located in the loop connecting strands β2 and β3, the loop connecting strands β4 and β5, and in the flexible N- and C-terminal tails are most likely to have specific interactions with the RNA and other proteins of RNase P. The structural model of an RNase P protein component provided by the present work provides an essential step toward eventually understanding the overall architecture of this complex enzyme and the mechanism by which it performs its functions.

Ribonuclease P (RNase P)[1] is an RNA−protein complex that cleaves the 5′-termini of precursor tRNAs to yield mature tRNAs (for reviews see refs *1*−*4*). RNase P is both ubiquitous and essential, having been identified in all kingdoms of life. Bacteria, eukaryotes, and archaea contain distinctive forms of the enzyme, and RNase P activity has been identified in mitochondria and chloroplasts as well. Bacterial RNase P contains one RNA component [typically about 375 nucleotides (nt)], which is the catalytically active unit (*5*), and a single protein component (typically 14 kDa) that functions in substrate recognition and release (*6*). Although the RNA component of RNase P in eukaryotes is similar in size to that of the bacterial form, the protein component is significantly more complex. For example, RNase P in humans and yeast each contain at least nine proteins (*7*, *8*), so that the total mass of the protein component exceeds that of the RNA. Archaeal RNase P contains at least four protein components, each of which exhibits significant primary sequence homology to a protein of eukaryotic RNase P (*9*). The eukaryotic and archaeal RNase P proteins are not significantly homologous to the bacterial protein. As is the case in bacteria, the RNA component of archaeal RNase P is by itself catalytically active at high ionic strength (*10*), although the protein components are required for activity under physiological conditions. Comparative sequence analyses indicate that the RNA component of archaeal RNase P is structurally similar to that of the bacterial form (*11*).

In terms of structure, bacterial RNase P is the best characterized form of the enzyme. Although the structure of the complete bacterial RNase P has not yet been determined, a substantial amount of information is available regarding its overall architecture. Models of the tertiary structure of the RNA component of *Escherichia coli* and *Bacillus subtilis* RNase P have been constructed on the basis of comparative phylogenetic analyses (*12*). A crystal structure of the 154 nucleotide specificity domain of *B. subtilis* RNase P has recently been reported (*13*), and a 27 nt RNA hairpin from the P4 region of *E. coli* RNase P has been studied by NMR methods (*14*). Structures of the protein subunits from *B. subtilis* and *Staphylococcus aureus* have been determined by X-ray crystallography (*15*) and NMR (*16*), respectively. Footprinting results have provided information regarding the RNA−protein contact sites within the tertiary structure of the RNA component (*12*). In the absence of the protein component, the RNA component of bacterial RNase P is a monomer; however, small-angle X-ray scattering data have provided evidence of a dimeric form of the RNase P RNA−protein complex, containing two of the RNA and two of the protein subunits (*17*). Solutions of RNase P from *B. subtilis* can contain both monomeric and dimeric forms, with the relative amounts of each form being sensitive to ionic conditions (*18*).

RNase P is closely related to another ribonucleoprotein complex, RNase MRP, which processes preribosomal RNA as part of the pathway toward appropriately post-transcriptionally modified ribosomal RNA in eukaryotes and also plays an essential role in yeast cell cycle progression

[†] This work was supported by Grant F-1353 from the Welch Foundation.

* Corresponding author. E-mail: dhoffman@mail.utexas.edu. Phone: 512-471-7859. Fax: 512-471-8696.

[1] Abbreviations: RNase P, ribonuclease P; aRpp29, archaeal RNase P protein 29; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; PCR, polymerase chain reaction; HMQC, heteronuclear single-quantum coherence.

10.1021/bi030170z CCC: $25.00
Published on Web 10/29/2003

(*19*−*22*). The RNA components of RNase MRP and RNase P are similar in terms of their primary and secondary structure, and each RNase contains some of the same protein components (*23*, *24*), although there is at least one protein component that is specific to each (*25*, *26*).

A thorough understanding of the RNase P mechanism and function will require that detailed structural information be available, such as that which can be provided by NMR or X-ray crystallographic methods. Currently, there are no NMR or X-ray structures available for any of the RNA or protein components of either eukaryotic or archaeal RNase P.

Archaeoglobus fulgidus was selected as the preferred species for our recently initiated biophysical studies of RNase P. This strictly anaerobic sulfur-metabolizing organism is found in hydrothermal environments and has an optimum growth temperature of 83 °C. The *A. fulgidus* RNase P has several features that make it attractive for structural studies. Its RNA component is relatively small, containing only 227 nucleotides. The *A. fulgidus* genome has been sequenced (*27*), which greatly simplifies the identification of the likely protein components of its RNase P. In addition, it has been noted that hyperthermophiles have an excellent record of being the source of materials for successful biophysical studies. Finally, *A. fulgidus* can be cultured in the laboratory, which may ultimately be important if in the course of our studies it becomes necessary or desirable to isolate the natural form of the RNase P enzyme.

As a step toward furthering our understanding of RNase P structure, we have recombinantly expressed, purified, and structurally characterized the protein from *A. fulgidus* that is homologous to the human RNase P protein Rpp29 and the yeast RNase P protein Pop4. The *A. fulgidus* homologue of the Rpp29 protein (henceforth referred to as aRpp29) is a particularly attractive target for structural analysis for several reasons: (1) Its primary sequence is not significantly homologous to any protein of known structure. (2) Its sequence is homologous to known archaeal RNase P proteins from *Methanobacter thermoautotrophicus* (*9*) and *P. horikoshii* (*29*), providing strong evidence that the *A. fulgidus* aRpp29 protein is actually an RNase P component. (3) The homology of archaeal aRpp29 with Rpp29 and Pop4 implies that any structural information obtained will be relevant for these human and yeast RNase P proteins as well. (4) Interaction maps based on data from two- and three-hybrid assays indicate that the yeast and human homologues of aRpp29 directly contact the RNA component of RNase P and are involved in a relatively large number of protein−protein interactions (*8*, *28*), suggesting that aRpp29 and its homologues are central to, rather than a peripheral component of, the overall RNase P structure. (5) The human homologue of aRpp29 is present in both RNase P and RNase MRP (*24*), so that the structural results provided by the present study will be relevant to a protein component of RNase MRP as well. (6) It has recently been demonstrated that a functional archaeal RNase P can be reconstituted from four recombinant *P. horikoshii* RNase P proteins and in vitro transcribed RNA (*29*); one of these four proteins is the homologue of *A. fulgidus* aRpp29. (*7*) The yeast homologue of aRpp29, the Pop4 protein, is encoded by a gene that is essential for viability (*26*), providing further evidence of the importance of this conserved gene product.

EXPERIMENTAL PROCEDURES

Protein Cloning, Expression, and Purification. The aRpp29 gene was obtained by PCR from the genomic DNA of *A. fulgidus* cells (American Type Culture Collection, clone number 49558). PCR primers were designed to include unique *Eco*RI and *Bam*HI restriction sites to allow the PCR product to be cloned into a plasmid encoding maltose binding protein (MBP) (pMAL-c2T, derived from pMAL-c2x; New England Biolabs) followed by a TEV protease site; this plasmid was transformed into BL21(DE3) cells (Novagen) supplemented with ampicillin. The recombinant MBP-aRpp29 fusion protein was produced by growing typically 4 L of *E. coli* cells in Luria broth at 37 °C until the cells reached an OD_{600} of 0.5. The cells were then induced with 0.6 mM isopropyl β-D-thiogalactopyranoside (IPTG) and allowed to grow for an additional 6 h, harvested by centrifugation, and stored at −80 °C. Thawed cells were lysed by sonication, and the nucleic acids were precipitated by the addition of 0.5% poly(ethylenimine) (v/v). The cell lysate was centrifuged at 12000*g* for 20 min to remove cellular debris and nucleic acids. Cellular proteins were precipitated by addition of ammonium sulfate up to 70% saturation and separated by centrifugation at 12000*g*. The protein pellet was dissolved in 10 mM potassium phosphate buffer at pH 5.8, loaded onto an SP-Sepharose column (Sigma), and eluted using a 0−1 M NaCl gradient. The fractions containing the MBP-aRpp29 fusion protein were identified using SDS−PAGE and pooled. The typical yield was 75 mg of fusion protein/L of cell culture. The aRpp29 was cleaved from the MBP fusion protein by incubation with TEV protease at room temperature for 24 h. The cleavage products were then loaded to an SP-Sepharose column, and the aRpp29 protein was eluted using a 0−1 M NaCl gradient. Fractions containing the aRpp29 protein were identified by SDS−PAGE, pooled, and concentrated. The typical yield of purified aRpp29 was 5 mg/L of cell culture. N-Terminal sequencing and mass spectrometry were used to confirm the identity of the purified protein.

Samples of aRpp29 enriched in ^{15}N and/or ^{13}C were prepared as above but with M9 minimal media containing 0.5 g/L [^{15}N]ammonium chloride and/or 3 g/L [^{13}C]glucose (Cambridge Isotope Laboratories) as the source of nitrogen and/or carbon. A protein sample selectively labeled with [^{15}N]lysine was prepared by growing the cells in M9 minimal media supplemented with 50 mg/L ^{15}N-labeled lysine and 100 mg/L of the other 19 unlabeled amino acids. A protein sample selectively labeled with [^{15}N]glycine, [^{15}N]serine, and [^{15}N]tryptophan was prepared using M9 minimal media supplemented with 100 mg/L ^{15}N-labeled glycine and 100 mg/L each of 16 unlabeled amino acid types (all types except glycine, serine, cysteine, and tryptophan). Samples that were simultaneously enriched with ^{15}N (uniform) and [1-^{13}C]-leucine (or [1-^{13}C]isoleucine or [1-^{13}C]phenylalanine or [1-^{13}C]valine) were prepared by growing the cells in M9 minimal media containing 1 g/L [^{15}N]ammonium chloride and 100 mg/L 1-^{13}C -labeled amino acid.

NMR Spectroscopy. NMR spectra were recorded at 20 or 30 °C using a 500 MHz Varian Inova spectrometer equipped with a triple-resonance probe and *z*-axis pulsed-field gradient. NMR samples typically contained 1.8−2.0 mM aRpp29 protein in 90% H_2O/10% D_2O plus 100 mM NaCl and a

buffer of either 10 mM sodium phosphate at pH 5.8 or 10 mM deuterated acetic acid at pH 3. Pulse sequences were obtained from Lewis Kay's group at the Toronto NMR center and were optimized before use on the local NMR instrumentation. Backbone resonance assignments were obtained using three-dimensional HNCA, HNCACB, and HN(CO)-CACB spectra (*30*), HNCO spectra (*31*), and HACACBCO spectra (*32*); together, these spectra correlate the backbone protons to the N, C^{α}, C^{o}, and C^{β} signals of the same and adjacent amino acid residues. Side chain resonance assignments were obtained by analyzing three-dimensional ^{15}N-edited HMQC-TOCSY and ^{13}C-edited HCCH-TOCSY (33) spectra and two-dimensional homonuclear 2QF-COSY and TOCSY spectra. NOE cross-peaks were detected using two-dimensional ^{1}H$-^{1}$H NOESY, three-dimensional ^{15}N$-^{1}$H$-^{1}$H HSQC-NOESY, and three-dimensional ^{13}C$-^{1}$H$-^{1}$H HSQC-NOESY (*34*) spectra; the NOE mixing time was 60 ms for spectra used to derive distance constraints. A ^{13}C-edited HSQC-NOESY spectrum (mixing time 80 ms) was acquired in 90% H_2O/10% D_2O solvent so that NOE cross-peaks between amide and side chain protons could be resolved by the chemical shift of the ^{13}C nucleus coupled to the side chain proton. In several instances, chemical shift assignment ambiguities were removed with the help of the amino acid type-specifically labeled samples. For example, amide resonances of amino acids that follow isoleucine were identified in an HNCO spectrum of protein that was uniformly enriched in ^{15}N and specifically enriched with [1-^{13}C]isoleucine; an analogous procedure was used to assign resonances of amino acids that follow valine, leucine, and phenylalanine. Similarly, assignments of glycine, serine, tryptophan, and lysine resonances were confirmed using HSQC spectra of protein samples selectively enriched with the specific amino acids types labeled with ^{15}N. Data were processed using either NMR-Pipe (*35*) or Felix (Hare Research). ^{1}H, ^{15}N, and ^{13}C chemical shifts were referenced as recommended (*36*), with proton chemical shifts referenced to internal 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) at 0 ppm. The 0 ppm ^{13}C and ^{15}N reference frequencies were determined by multiplying the 0 ppm ^{1}H reference frequency by 0.251449530 and 0.101329118, respectively. Chemical shift assignments for the aRpp29 protein have been deposited in the BioMagResBank (accession number BMRB-5805).

^{15}N$-^{1}$H heteronuclear NOE and ^{15}N T_1 and T_2 relaxation times were measured using pulse sequences that feature gradient selection and sensitivity enhancement and pulses for minimizing saturation of the solvent water (*37*). The ^{15}N$-^{1}$H heteronuclear NOE was measured by comparing spectra acquired with either a 5 s delay between each free induction decay or a 2 s delay followed by a 3 s series of 120° nonselective ^{1}H pulses. For T_1 relaxation measurements, 2-D spectra with relaxation delays of 10, 260, 510, 760, and 1010 ms were obtained; for T_2 relaxation measurements, 2-D spectra with relaxation delays of 29, 58, 87, 116, and 145 ms were acquired; in each case the relaxation delay between the acquisition of each free induction decay was 3 s. T_1 and T_2 relaxation times were determined from the slope of plots of the logarithm of peak height versus relaxation delay.

Structure Calculation. Structure calculations were performed using the restrained simulated annealing protocol in the program CNS version 1.1 (*38*), with the goal of identifying the full range of structures that are consistent with the distance and angle constraints derived from the NMR data while having reasonable molecular geometry, consistent with a minimum value of the CNS energy function. Distance restraints were derived from the intensities of cross-peaks within NOE spectra obtained with relatively short mixing times of 60 ms (in the case of homonuclear 2-D and ^{15}N-resolved 3-D spectra) to 80 ms (in the case of the ^{13}C-resolved 3-D NOE spectrum) to minimize the effects of spin diffusion. On the basis of the cross-peak intensity in the homonuclear 60 ms NOE spectra, distance restraints were classified as strong (<2.8 Å), medium (<3.2 Å), weak (<3.8 Å), and very weak (<4.2 Å); these distance bounds were calibrated by using interproton distances in regions of regular secondary structure as internal distance standards. Additional cross-peaks were observed in the three-dimensional ^{15}N- and ^{13}C-resolved NOE spectra (60 and 80 ms mixing times, respectively) and were assigned to distance restraints as strong (<5.0 Å), medium (<5.5 Å), weak (<6.0 Å), and very weak (<6.5 Å). Due to the possibility of observing the effects of spin diffusion in long mixing time (200 ms) experiments, NOE cross-peaks that were only observed in these spectra were assigned to a distance restraint of 6.9 Å. Pseudoatom corrections were applied to the distance constraints as follows: NOEs from valine or leucine methyl

Table 1: Summary of Refinement and Structural Statistics for the *A. fulgidus* aRpp29 Protein (Residues 17−77)[a]

intraresidue NOEs	215
sequential NOEs (residue *i* to *i* + 1)	178
medium-range NOEs (residue *i* to *i* + 2, 3, 4)	18
long-range NOEs	143
dihederal angle restraints	70
hydrogen bond restraints	27
total structural restraints	651
no. of unique starting structures for simulated annealing	10
no. of simulated annealing runs, differing in initial trajectories	200
rmsd for backbone atoms (residues 17−77)	0.87 Å
rmsd for side chain atoms (residues 17−77)	1.78 Å
av no. of NOE violations >0.2 Å (per structure)	3.2 ± 1.0
av no. of NOE violations >0.5 Å (per structure)	0
residues in most favored regions of the Ramachandran plot	71.2 ± 2.6%
residues in additionally allowed regions of the Ramachandran plot	21.2 ± 4.6%
residues in generously allowed regions of the Ramachandran plot	5.8 ± 2.7%
residues in disallowed regions of the Ramachandran plot	1.9 ± 0.9%
rmsd for covalent bonds	0.0034 ± 0.0001
rmsd for covalent angles	0.511 ± 0.015
rmsd for improper angles	0.581 ± 0.016

[a] Statistics are derived from a set of 12 low- energy structures that is representative of the range of structures that are consistent with the NMR-derived constraints. Intra-side chain NOEs with no meaningful structural information were not included in the NOE list. The number of observed NOEs is fairly modest due to a substantial fraction of the structure being flexible; however, the final structural models are well defined. Over 90% of the residues in the refined models occur in the most favored or additionally allowed regions of a Ramachandran plot. Two residues (S25 and E30) in the loop connecting strands β1 and β2 have ϕ and ψ angles that extend slightly into the disallowed Ramachandran regions; however, this may not be significant since there is evidence of flexibility in this loop, and the rmsd for the coordinates of the backbone atoms of these residues is approximately 1.3.

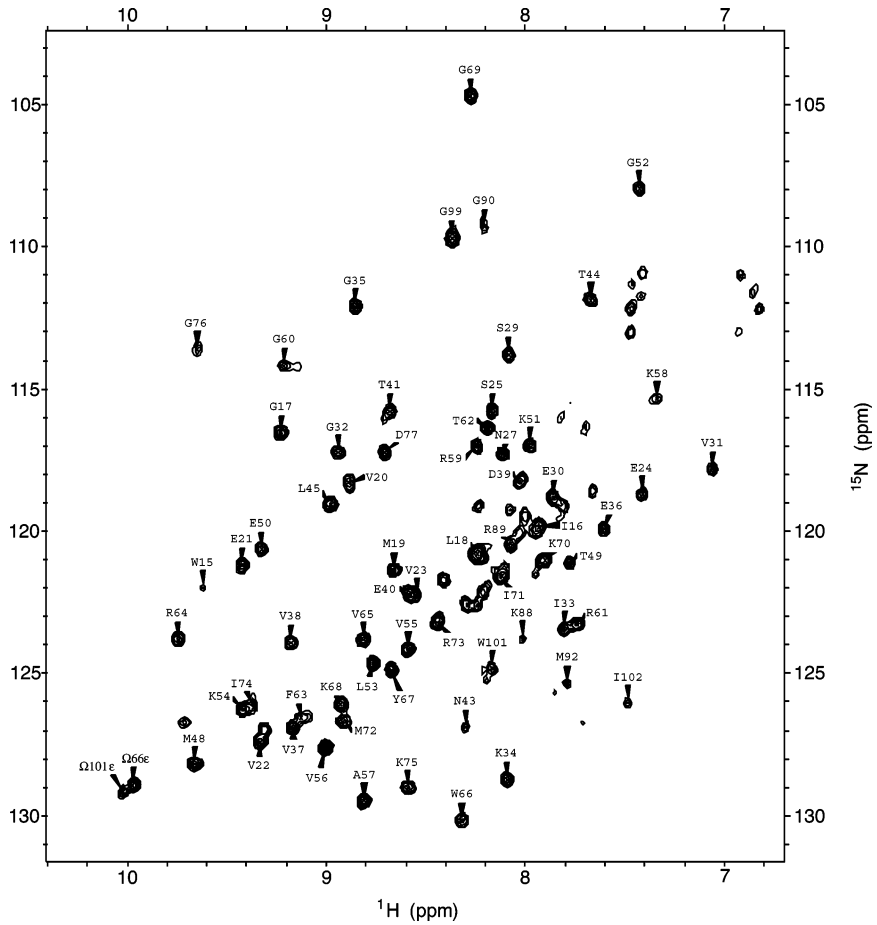

FIGURE 1: ^{15}N−^{1}H correlated HSMQC spectrum (*40*) of the aRpp29 protein from *A. fulgidus* obtained at 30 °C and pH 5.8, using presaturation for H_2O suppression, so that the amide protons that exchange relatively rapidly with the solvent are attenuated. Assignments for the best resolved cross-peaks are labeled in the figure. Additional cross-peaks are observable in ^{15}N−^{1}H correlated spectra obtained without using presaturation for solvent suppression; these are mostly located near the chemical shifts typical of random coil (not shown). ^{1}H, ^{13}C, and ^{15}N chemical shift assignments have been deposited in the BioMagResBank and assigned the accession number BMRB-5805.

groups that were not stereospecifically assigned were measured from the center of the two methyl groups, and 2.5 Å was added to the interproton distance. For NOEs involving other methyl protons, distances were measured from the center of the methyl group, and an additional 1.0 Å was added to the interproton distance. For NOEs involving methylene protons with no stereospecific assignment, distances were measured from the center of the methylene group, and 0.7 Å was added to the interproton distance. For NOEs involving δ and ϵ protons on tyrosine rings, distances were measured from the center of the two δ protons (or ϵ protons), and 2.4 Å was added to the interproton distance. Backbone dihedral angle restraints of $\psi = 150 \pm 25°$ and $\phi = -120 \pm 25°$ were included for residues within the regions of regular β-strand secondary structure, identified by characteristic NOE cross-peaks and patterns of protection of the amide protons from exchange with the solvent. Hydrogen bonds were defined using distance bounds for amide protons that were clearly located within the regions of regular β-sheet structure. Experimental restraints used for the structure calculations and structural statistics are summarized in Table 1.

An initial set of 10 structures was generated from an extended peptide conformation using a simulated annealing protocol with dihedral angle restraints only. These structures with low overall energies were selected for further refinement. The selected structures were then used as starting points to generate 20 structures each, via restrained simulated annealing using different initial trajectories. A set of refined conformers having the lowest energy was retained for final analysis and evaluation using Procheck-NMR (*39*), with statistics reported in Table 1. These final structures are a fair representation of the full range of structures that are consistent with the experimental data, while having reasonable molecular geometry, and have no NOE-derived distance constraint violations greater than 0.5 Å. Searches for similar structures within the Protein Data Bank were carried out using the Vector Alignment Search Tool (VAST), located at the National Center for Biotechnology Information (NCBI) web site, and the DALI search tool (*40*). The coordinates for aRpp29 have been deposited in the Protein Data Bank and assigned PDB code 1PC0.

RESULTS

The purified *A. fulgidus* aRpp29 protein was found to be soluble, stable, and suitable for structural analysis by NMR methods. About 68 of 100 possible correlation peaks are observed in a ^{15}N−^{1}H HSMQC spectrum (*41*) obtained at pH 5.8 and 30 °C using presaturation for solvent suppression (Figure 1); these peaks are well dispersed as is typical for a folded protein. When a nonsaturating method of solvent suppression is used, additional ^{15}N−^{1}H correlation peaks

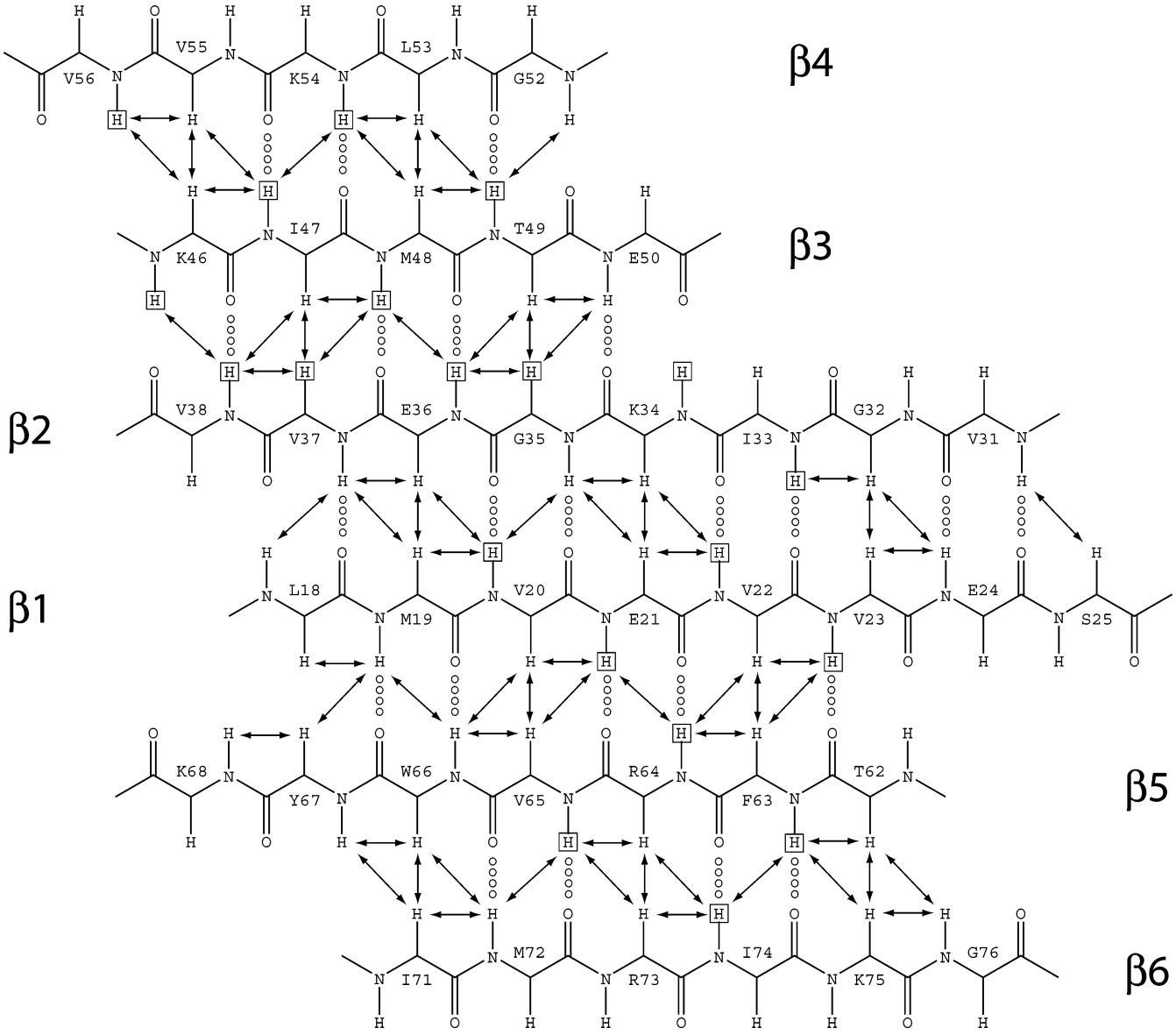

FIGURE 2: Diagram of the six-stranded β-sheet structure within the aRpp29 protein from *A. fulgidus*. Pairs of protons for which unambiguous NOEs are observed are connected by arrows, and interstrand hydrogen bonds are indicated by dotted lines. Amide protons that require 12 h or more to exchange with deuterated solvent are boxed.

corresponding to amide protons with relatively rapid solvent exchange rates are observed in or near the region of the spectrum typical of random coil structure. These observations provided an initial indication that about two-thirds of the protein structure forms a folded and stable domain, while the remaining one-third of the structure is significantly more flexible. Triple-resonance NMR methods were used to obtain unambiguous spectrum assignments for nuclei of all residues except 4−7, 79−84, and 93. Mass spectrometry and N-terminal sequencing indicated that residues 1−3 are absent from the structure of the purified protein, presumably due to proteolytic cleavage or degradation.

Structural results derived from the NMR data show that the dominant feature of the protein is an antiparallel sheet of six β-strands (Figure 2), formed by residues 18−76. Within the β-sheet, strands β1 and β2 are connected by a short loop, as are strands β2 and β3, while strands β3 and β4 are connected by a compact turn, as are strands β5 and β6. An interesting feature of the β-sheet is a loop consisting of residues 57−61; this five-residue loop connects strands β4 and β5, which are on opposite sides of the sheet as it is represented in two dimensions (Figure 2) resulting in strands β5 and β6 being folded onto the first four strands to form a sandwich-like structure (Figure 3). The six β-strands are wrapped around a hydrophobic core containing residues L18, V20, V22, I33, V37, L45, I47, V56, F63, V65, M72, and I74. The hydrophobic nature of these residues is conserved among a wide range of species (Figure 4A), providing strong evidence that the RNase P proteins homologous to aRpp29 in the archaea, Rpp29 in humans, and Pop4 in yeast all contain a similarly structured β-sheet.

The six strands of the β-sheet are the best defined region of the protein structure, due to the relatively high density of long-range NOE-derived distance constraints. Within a family of structures that satisfies the NMR-derived restraints equally well, the rmsd is less than 1 Å for the backbone atoms within the β-strands (Figure 5). The loops that connect strands β1−β2 and β2−β3 are the least well defined of the connecting structures; these loops are primarily defined by sequential and intraresidue NOE-derived restraints.

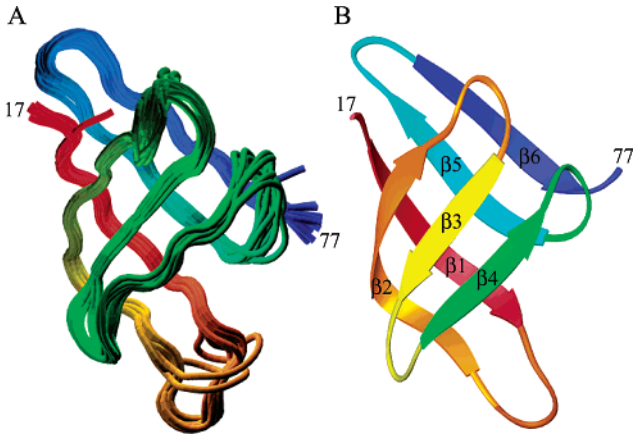

FIGURE 3: (A) Superposition of the backbones of 12 low-energy structures of aRpp29 that are equally consistent with the NMR data and have the minimum value of the CNS energy function, color-ramped from red at the N-terminus to blue at the C-terminus of the protein. The 12 models are a fair representation of the full range of structures that are consistent with the NMR-derived constraints and reasonable molecular geometry. The superposition of the 12 models was performed by minimizing the differences in the coordinates of the backbone heavy atoms of residues 17−77; the backbone is shown for residues 17−77 of the protein. The figure was created using the program VMD (*54*). (B) Ribbon diagram of residues 17−77 of the aRpp29 protein created using the program RIBBONS (*55*).

Most of the hydrogen-bonded amide protons within the β-sheet remain unexchanged with deuterated solvent after 12 h or more, providing evidence that the sheet is a relatively stable and rigid structure. This conclusion is further supported by ^{15}N relaxation data. The T_1 and T_2 relaxation times for the β-strand residues were found to be strikingly uniform, averaging 515 and 125 ms, respectively, and the ^{15}N−^{1}H heteronuclear NOE was found to be consistently positive with an average value of 0.72. These data are consistent with a well-ordered structure undergoing isotropic rotation with a correlation time of 6.7 ns; this rotational time is appropriate for a monomeric protein with a molecular mass of 11 kDa. The amide nitrogens of residues 30, 41, 58, and 69 (all located in the turns or loops connecting the β-strands) have ^{15}N relaxation times that deviate from these average values, consistent with their having less restricted motion. Amide proton exchange rates are significantly faster for the loop residues than for the β-strands, which is also consistent with the loops having the greater flexibility.

In contrast to the relatively rigid β-sheet structure, the positions of the residues prior to 17 and past 78 are poorly defined, with the majority of the observable NOEs being between protons of the same or sequential residues. The ^{15}N−^{1}H heteronuclear NOE was found to be negative or near zero for the amide groups of residues 88, 90, 92, 99, 101, and 102, consistent with their having short rotational correlation times and a relatively large amount of flexibility. ^{15}N−^{1}H heteronuclear NOE data were not obtainable for most other residues in the N- and C-terminal regions due to their rapid amide proton exchange rates, an observation which by itself is consistent with the N- and C-termini of the structure being solvent exposed and relatively flexible. Chemical shift index (*42*) values for residues 89−98 hint at the presence of helical structure, with an average backbone carbonyl ^{13}C shift of 176.3 ppm. However, the proton−proton NOEs typical of a helix are not observed, suggesting that the existence of this helix is only transient, if it exists at all. The ^{13}C chemical shifts may suggest a rapid equilibrium between helical and unfolded structure in the C-terminal region of the protein; a more stable helix may form in the context of intact RNase P.

Triple-resonance and ^{15}N-resolved 3-D NOE and 3-D TOCSY spectra were collected at reduced pH (3.0) and a slightly reduced temperature (20 °C), in an effort to slow the solvent-exchange rates of the amide protons within the terminal regions of the protein and perhaps obtain additional structural information for these flexible regions of the molecule. However, only modest changes in the chemical shifts were observed, and little significant additional NOE information was obtained, indicating that the protein structure is not significantly changed at these alternate conditions.

The substantial flexibility of the N- and C-terminal regions of the protein might suggest that the protein is a poor candidate for crystallization. Surprisingly, however, it was found that the protein crystallizes readily, in orthorhombic space group $P2_12_12_1$ with one molecule per asymmetric unit, and diffracts X-rays to 1.6 Å resolution on a local (copper anode) source. A selenomethionine derivative of the protein has also been prepared and crystallized, and the analysis of the structure using X-ray crystallographic methods is in progress.

DISCUSSION

Antiparallel β-sheet and β-barrel structures have previously been identified in RNA-associated proteins with functions related to tRNA and translation. For example, bacterial translation initiation factor-1 (IF-1) (*43*), ribosomal protein S17 (*44*), and the β-barrel domain of ribosomal protein S1 (*45*) each contain five antiparallel β-strands and are representatives of the oligomer-binding (OB) protein fold (*46*). RNA-associated proteins containing structures of six antiparallel β-strands have also been identified; these include domain III of EF-Tu (*47*) and domain VI within translation initiation factor IF2 (*48*, *49*). However, in each case the connectivity of the β-strands within the antiparallel β-sheet differs from that found in aRpp29, providing evidence against a close evolutionary or structural relationship between aRpp29 and these other proteins. The aRpp29 is not structurally similar to the bacterial RNase P protein, which has been structurally characterized by both X-ray and NMR methods (*15*, *16*) and found to contain a mixed $\alpha-\beta$ structure.

The Structural Classification of Proteins (SCOP) database of protein folds (*50*) contains a substantial number of small proteins which like aRpp29 consist of a six-stranded antiparallel β-sheet. However, none of these other six-stranded proteins are connected in such a way that their topology is the same as that of aRpp29. Searches carried out using the programs DALI (*40*) and VAST (available at the NCBI web site) did not identify any structures within the Protein Data Bank that are significantly similar to aRpp29 in terms of backbone atom coordinates. Interestingly, the first five strands of aRpp29 are connected in the same way as is observed in the Sm family of RNA-associated proteins, a representative of which is the translational regulator Hfq (*51*).

When the amino acid sequence of aRpp29 is aligned with that of homologous proteins from several species (Figure

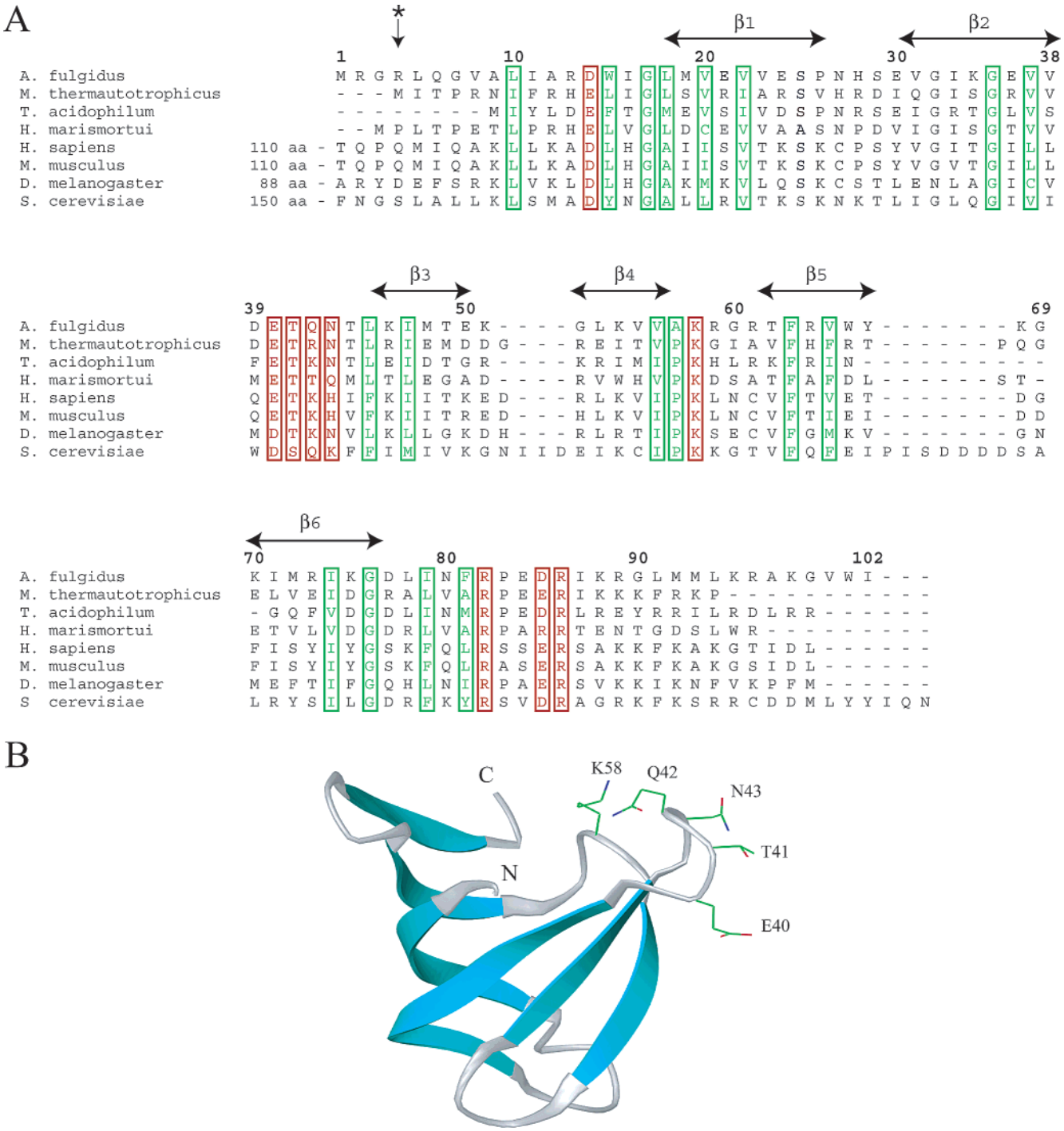

FIGURE 4: (A) Alignment of the amino acid sequence of the *A. fulgidus* aRpp29 protein with homologous proteins from three other species of archaea and four species of eukaryotes. Archaeal sequences are shown from *M. thermoautotrophicus*, *Thermoplasma acidophilum*, and *Haloarcula marismortui*. Representative species of eukaryotes shown are from *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, and *Saccharomyces cerevisiae*. The most conserved hydrophobic residues are boxed in green. Conserved residues that are accessible on the protein surface and are likely candidates for directly interacting with the RNase P RNA or other conserved protein components are boxed in red. Although only eight sequences are shown in the figure, a larger set of sequences was compared in making the decision as to which residues are relatively well conserved. (B) Ribbon diagram of residues 17−77 of the *A. fulgidus* aRpp29 protein, created using RIBBONS (*55*), with the relative positions of some of the most conserved surface residues indicated. Conserved residues E40, T41, Q42, and N43, located in the loop connecting strands β2 and β3, are near conserved residue K58 located within the loop that connects strands β4 and β5. The other conserved surface residues indicated in panel A are located within the flexible tails in the N- and C-terminal regions of the protein. It should be noted that the rmsd for the side chain coordinates of these residues is in the range of 1−3 Å. Hence, the NMR data do not define the precise conformation of each surface side chain but do serve to indicate the relative positions of the residues that are on the protein surface.

4A) and this alignment is compared with the structural results, it is possible to speculate as to the functional importance of each conserved amino acid. The majority of the conserved hydrophobic amino acids are located within the core of the protein that is enclosed by the six β-strands; these positions are likely to be critical for protein folding and stability. Several glycines located in turns are also well conserved. A few hydrophobic amino acids (at positions 10,

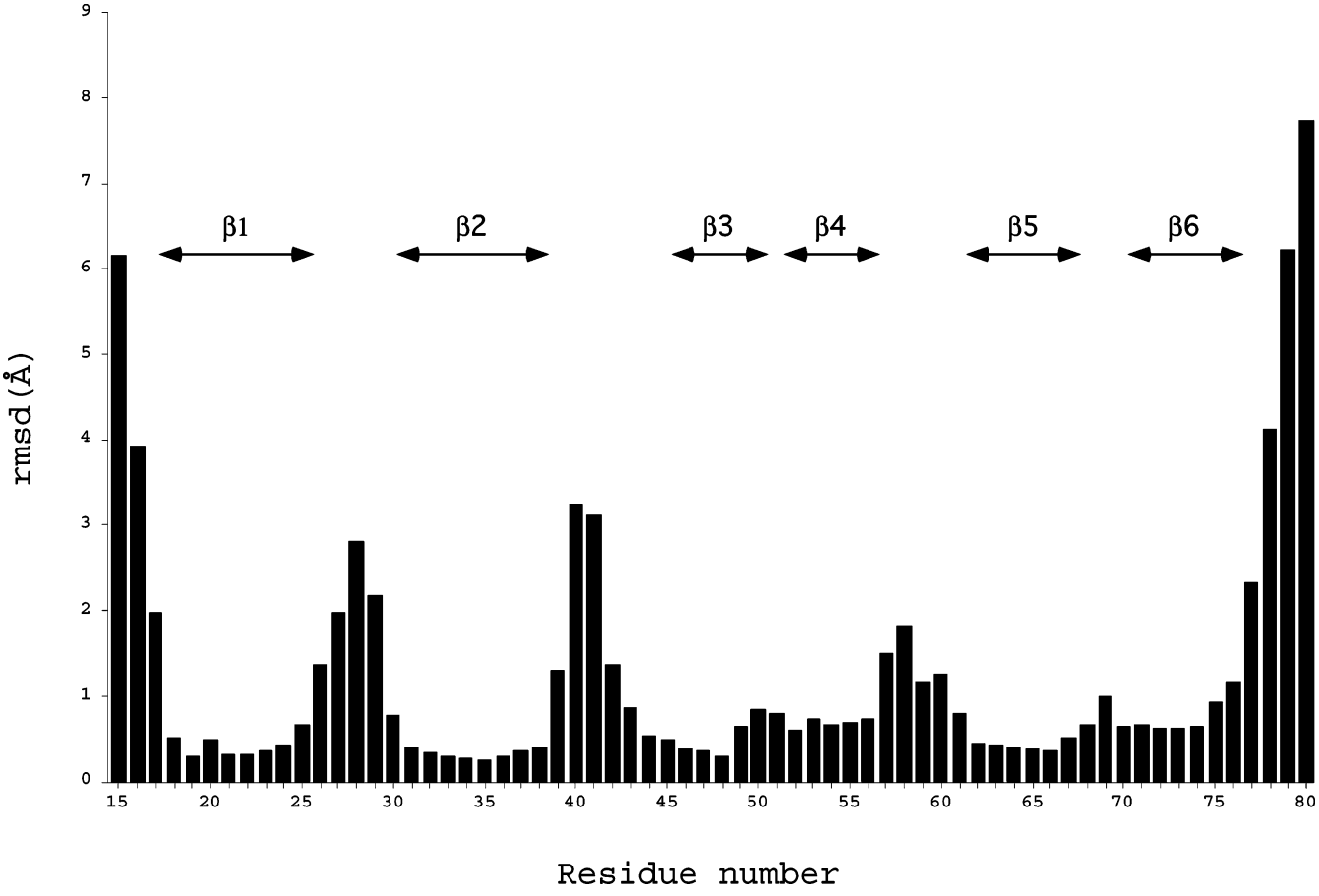

FIGURE 5: Plot of the root mean square deviation for the coordinates of the backbone heavy atoms for the aRpp29 protein, versus residue number, calculated using 12 structures that satisfy the NMR-derived structural constraints and are a fair representation of the full range of structures that are consistent with the NMR data. The rmsd values were calculated using a set of structures that were superimposed by minimizing the differences in the coordinates of the backbone atoms of residues 17−77. The figure shows that the β-strands are quite well defined by the NMR data (rmsd <1 Å) and the loops connecting the strands are moderately well defined (rmsd 1−3.5 Å), while the positions of the residues nearer to the N- and C-termini are not well determined (rmsd >3 Å).

15, 79, and 81) are located outside the well-ordered β-sheet structure and are also fairly well conserved; these residues may become more ordered and protected from the solvent when in the context of the fully assembled RNase P ribonucleoprotein particle.

Conserved hydrophilic amino acids on the surface of the protein are the most likely to be involved in essential intermolecular contacts with the RNase P RNA, other RNase P protein components, or other RNase P interaction partners. The protein contains several surface positions where positive charges (K58, R82, and R86), negative charges (D14, E40, and D85), and ability to serve as a hydrogen bond donor (T41, Q42, and N43) are conserved (Figure 4A). The conserved lysine and arginines are excellent candidates for residues likely to make direct contact with the phosphate backbone of the RNase P RNA, while the conserved negative charges and hydrogen bond donors have the potential to interact with specific RNA bases or other protein components within RNase P. The eukaryotic homologues of aRpp29 contain an additional ~100 amino acids at their N-terminus; however, these additional N-terminal residues are not particularly well conserved and are thus more likely to be involved in species-specific interactions. The observation that aRpp29 contains surface amino acids that are conserved among the eukaryotic as well as the archaeal homologues suggests that there are specific intermolecular interactions within RNase P that are similar among members of these kingdoms.

The presence of "tails" of flexible structure at the N- and C-termini of aRpp29 is reminiscent of several of the ribosomal proteins, such as L15, L21e, and L37e (*49*), which contain tails of extended structure at the N- and/or C-terminus of an otherwise globular domain. Within the context of the structure of the complete ribosomal subunit, these tails were found to penetrate to the interior of the ribosome and make specific contacts with the ribosomal RNA. By analogy, it is possible that the N- and C-terminal residues of aRpp29 may become structured and have specific interactions within the context of the complete and functional RNase P ribonucleoprotein complex, perhaps penetrating into the interior of the RNase P RNA. Lack of rigid structure in the N- and C-terminal residues of aRpp29 does therefore not imply that they are functionally unimportant.

Given that the catalytic activity associated with RNase P is carried out by the RNA component, what are the likely roles of the RNase P proteins? It can be speculated that one function of the proteins may be to stabilize the RNA component in a conformation that is optimal for catalytic activity. Studies of bacterial RNase P suggest that although the RNA is the catalytic unit, the protein component may have roles in substrate recognition (*6*, *52*, *53*) and dimerization (*17*); this suggests the possibility for more extensive functional

roles for the archaeal and eukaryotic RNase P proteins as well. A more detailed description as to how the protein and RNA components of RNase P work together in the functioning enzyme will likely be provided by additional structural and biochemical studies.

REFERENCES

1. Frank, D. N., and Pace, N. R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme, *Annu. Rev. Biochem. 67*, 153−180.
2. Altman, S., and Kirsebom, L. A. (1999) in *The RNA World* (Gesteland, R. F., Cech, T., and Atkins, J., Eds.) 2nd ed., pp 351−380, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3. Hall, T. A., and Brown, J. W. (2001) The ribonuclease P family, *Methods Enzymol. 341*, 56−77.
4. Xiao, S., Scott, F., Fierke, C. A., and Engelke, D. R. (2002) Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes, *Annu. Rev. Biochem. 71*, 165−189.
5. Guerrier-Takada, C., Gardner, K., Marsh, T., Pace, N. R., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme, *Cell 35*, 849−857.
6. Kurz, J. C., Niranjanakumari, S., and Fierke, C. A. (1998) Protein component of *Bacillus subtilis* RNase P specifically enhances the affinity for precursor-tRNA$_{Asp}$, *Biochemistry 37*, 2393−2400.
7. Jarrous, N., Reiner, R., Wesolowski, D., Mann, H., Guerrier-Takada, C., and Altman, S. (2001) Function and subnuclear distribution of Rpp21, a protein subunit of the human ribonucleoprotein ribonuclease P, *RNA 7*, 1153−1164.
8. Houser-Scott, F., Xiao, S., Millikin, C. E., Zengel, J. M., Lindahl, L., and Engelke, D. R. (2002) Interactions among the protein and RNA subunits of *Saccharomyces cerevisiae* nuclear RNase P, *Proc. Natl. Acad. Sci. U.S.A. 99*, 2684−2689.
9. Hall, T. A., and Brown, J. W. (2002) Archaeal RNase P has multiple protein subunits homologous to eukaryotic nuclear RNase P proteins, *RNA 8*, 296−306.
10. Pannucci, J. A., Hass, E. S., Hall, T. A., Harris, J. K., and Brown, J. W. (1999) RNase P RNAs from some Archaea are catalytically active, *Proc. Natl. Acad. Sci. U.S.A. 96*, 7803−7808.
11. Harris, J. K., Haas, E. S., Williams, D., Frank, D. N., and Brown, J. W. (2001) New insight into RNase P RNA structure from comparative analysis of the archaeal RNA, *RNA 7*, 220−232.
12. Tsai, H. Y., Masquida, B., Biswas, R., Westhof, E., and Gopalan, V. (2003) Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme, *J. Mol. Biol. 325*, 661−675.
13. Krasilnikov, A. S., Yang, X., Pan, T., and Mondragon, A. (2003) Crystal structure of the specificity domain of ribonuclease P, *Nature 421*, 760−764.
14. Schmitz, M., and Tinoco, I., Jr. (2000) Solution structure and metal-ion binding of the P4 element from bacterial RNase P RNA, *RNA 6*, 1212−1225.
15. Stams, T., Niranjanakumari, S., Fierke, C. A., and Christianson, D. W. (1998) Ribonuclease P protein structure: evolutionary origins in the translational apparatus, *Science 280*, 752−755.
16. Spitzfaden, C., Nicholson, N., Jones, J. J., Guth, S., Lehr, R., Prescott, C. D., Hegg, L. A., and Eggleston, D. S. (2000) The structure of ribonuclease P protein from *Staphylococcus aureus* reveals a unique binding site for single-stranded RNA, *J. Mol. Biol. 295*, 105−115.
17. Fang, X. W., Yang, X. J., Littrell, K., Niranjanakumari, S., Thiyagarajan, P., Fierke, C. A., Sosnick, T. R., and Pan, T. (2001) The *Bacillus subtilis* RNase P holoenzyme contains two RNase P RNA and two RNase P protein subunits, *RNA 7*, 233−241.
18. Barrera, A., Fang, X., Jacob, J., Casey, E., Thiyagarajan, P., and Pan, T. (2002) Dimeric and monomeric *Bacillus subtilis* RNase P holoenzyme in the absence and presence of pre-tRNA substrates, *Biochemistry 41*, 12986−12994.
19. Schmitt, M. E., and Clayton, D. A. (1993) Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in *Saccharomyces cerevisiae*, *Mol. Cell. Biol. 13*, 7935−7941.
20. Chu, S., Archer, R. H., Zengel, J. M., and Lindahl, L. (1994) The RNA of RNase MRP is required for normal processing of ribosomal RNA, *Proc. Natl. Acad. Sci. U.S.A. 91*, 659−963.
21. Lygerou, Z., Allmang, C., Tollervey, D., and Seraphin, B. (1996) hPop1: an autoantigenic protein subunit shared by the human RNase P and RNase MRP ribonucleoproteins, *Science 272*, 268−270.
22. Cai, T., Aulds, J., Gill, T., Cerio, M., and Schmitt, M. E. (2002) The *Saccharomyces cerevisiae* RNase mitochondrial RNA processing is critical for cell cycle progression at the end of mitosis, *Genetics 161*, 1029−1042.
23. Lygerou, Z., Mitchell, P., Petfalski, E., Seraphin, B., and Tollervey, D. (1994) The POP1 gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins, *Genes Dev. 8*, 1423−1433.
24. van Eenennaam, H., Pruijn, G. J., and van Venrooij, W. J. (1999) hPop4: a new protein subunit of the human RNase MRP and RNase P ribonucleoprotein complexes, *Nucleic Acids Res. 27*, 2465−2472.
25. Schmitt, M. E., and Clayton, D. A. (1994) Characterization of a unique protein component of yeast RNase MRP: an RNA-binding protein with a zinc-cluster domain, *Genes Dev. 8*, 2617−2628.
26. Chamberlain, J. R., Lee, Y., Lane, W. S., and Engelke, D. R. (1998) Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP, *Genes Dev. 12*, 1678−1690.
27. Klenk, H. P., et al. (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*, *Nature 390*, 364−370.
28. Jiang, T., Guerrier-Takada, C., and Altman, S. (2001) Protein-RNA interactions in the subunits of human nuclear RNase P, *RNA 7*, 937−941.
29. Kouzuma, Y., Mizoguchi, M., Takagi, H., Fukuhara, H., Tsukamoto, M., Numata, T., and Kimura, M. (2003) Reconstitution of archaeal ribonuclease P from RNA and four protein components, *Biochem. Biophys. Res. Commun. 306*, 666−673.
30. Muhandiram, D. R., and Kay, L. E. (1994) Gradient-enhanced triple resonance three-dimensional NMR experiments with improved sensitivity, *J. Magn. Reson., Ser. B 103*, 203−216.
31. Grzesiek, S., and Bax, A. (1992) ^{1}H, ^{13}C, and ^{15}N NMR backbone assignments and secondary structure of human interferon-gamma, *J. Magn. Reson. 96*, 432−440.
32. Kay, L. E. (1993) Pulsed-field gradient-enhanced three-dimensional NMR experiment for correlating ^{13}C.alpha./.beta., ^{13}C′, and ^{1}H.alpha. chemical shifts in uniformly carbon-13-labeled proteins dissolved in water, *J. Am. Chem. Soc. 115*, 2055−2057.
33. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., and Forman-Kay, J. D. (1993) A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain ^{1}H and ^{13}C Correlations in H_2O Samples of Proteins, *J. Magn. Reson., Ser. B 101*, 333−337.
34. Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1994) Simultaneous Acquisition of ^{15}N- and ^{13}C-Edited NOE Spectra of Proteins Dissolved in H_2O, *J. Magn. Reson., Ser. B 103*, 197−201.
35. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR 6*, 277−293.
36. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) ^{1}H, ^{13}C, and ^{15}N chemical shift referencing in biomolecular NMR, *J. Biomol. NMR 6*, 135−140.
37. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation, *Biochemistry 33*, 5984−6003.
38. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54* (Part 5), 905−921.
39. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR 8*, 477−486.
40. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123−138.

41. Zuiderweg, E. R. P. (1990) A proton-detected heteronuclear chemical-shift correlation experiment with improved resolution and sensitivity, *J. Magn. Reson. 86*, 346−352.
42. Wishart, D. S., and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination, *Methods Enzymol. 239*, 363−392.
43. Sette, M., van Tilborg, P., Spurio, R., Kaptein, R., Paci, M., Gualerzi, C. O., and Boelens, R. (1997) The structure of the translational initiation factor IF1 from E. coli contains an oligomer-binding motif, *EMBO J. 16*, 1436−1443.
44. Golden, B. L., Hoffman, D. W., Ramakrishnan, V., and White, S. W. (1993) Ribosomal protein S17: characterization of the three-dimensional structure by ^{1}H and ^{15}N NMR, *Biochemistry 32*, 12812−12820.
45. Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M., and Murzin, A. G. (1997) The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold, *Cell 88*, 235−242.
46. Murzin, A. G. (1993) OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for nonhomologous sequences, *EMBO J. 12*, 861−867.
47. Nissen, P., Kjeldgaard, M., Thirup, S., Polekhina, G., Reshetnikova, L., Clark, B. F., and Nyborg, J. (1995) Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog, *Science 270*, 1464−1472.
48. Meunier, S., Spurio, R., Czisch, M., Wechselberger, R., Guenneugues, M., Gualerzi, C. O., and Boelens, R. (2000) Structure of the fMet-tRNA(fMet)-binding domain of B. stearothermophilus initiation factor IF2, *EMBO J. 19*, 1918−1926.
49. Roll-Mecak, A., Cao, C., Dever, T. E., and Burley, S. K. (2000) X-ray structures of the universal translation initiation factor IF2/eIF5B: conformational changes on GDP and GTP binding, *Cell 103*, 781−792; Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution, *Science 289*, 905−920.
50. Murzin, A. G., Brenner, S. E., Hubbard T., and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol. 247*, 536−540.
51. Schumacher, M. A., Pearson, R. F., Møller, T., Valentin-Hansen, P., and Brennan, R. G. (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein, *EMBO J. 13*, 3546−3556.
52. Crary, S. M., Niranjanakumari, S., and Fierke, C. A. (1998) The protein component of *Bacillus subtilis* ribonuclease P increases catalytic efficiency by enhancing interactions with the 5′ leader sequence of pre-tRNAAsp, *Biochemistry 37*, 9409−9416.
53. Niranjanakumari, S., Stams, T., Crary, S. M., Christianson, D. W., and Fierke, C. A. (1998) Protein component of the ribozyme ribonuclease P alters substrate recognition by directly contacting precursor tRNA, *Proc. Natl. Acad. Sci. U.S.A. 95*, 15212−15217.
54. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J. Mol. Graphics 14*, 33−38.
55. Carson, M. (1991) Ribbons 2.0, *J. Appl. Crystallogr. 24*, 958.

BI030170Z