

Copyright
by
David Joseph Sidote
2004

**The Dissertation Committee for David Joseph Sidote Certifies that this is
the approved version of the following dissertation:**

**Structural Characterization of the Ribonuclease P Protein
aRpp29 from the Hyperthermophilic Sulfate-reducing Archaeon
*Archaeoglobus fulgidus***

Committee:

David W. Hoffman, Supervisor

Karen Browning

Arlen Johnson

Jon Robertus

John Tesmer

**Structural Characterization of the Ribonuclease P Protein
aRpp29 from the Hyperthermophilic Sulfate-reducing Archaeon
*Archaeoglobus fulgidus***

by

David Joseph Sidote, B.A.

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

**The University of Texas at Austin
December, 2004**

To my family

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. David Hoffman, for his supervision, guidance, and of course his seemingly inexhaustible patience (Thanks for putting up with me Dave!). I would like to thank my committee members Dr. Karen Browning, Dr. John Tesmer, Dr. Jon Robertus, and Dr. Arlen Johnson for their suggestions and for serving on my dissertation committee.

I would like to take this opportunity to thank members of the Hoffman Lab, past and present, for many helpful discussions and suggestions. Especially, Simrit Dhaliwal, whose cheerful and positive disposition is an inspiration. Special thanks go to Johanna Heideker for her contribution to my project and for the spirited discussions about keeping my cats indoors (where they will remain). I would also like to thank Romana Kristelly for providing the pMAL and TEV vectors that enabled this project to move forward at a time when I was stuck, Dave Lodowski for his help with crystallography related things, and Dr. Tesmer for taking the time to show me how to collect and process crystallography data.

I would like to thank my parents and my sister for their unwavering support and patience. They have always believed in me, even when I did not. I would like to thank Fang En Lee for her friendship and support. Lastly, I would like to thank my fiancé, Afreen Khan, for the encouragement, support, and understanding she has given me throughout my graduate career.

**Structural Characterization of the Ribonuclease P Protein
aRpp29 from the Hyperthermophilic Sulfate-reducing Archaeon
*Archaeoglobus fulgidus***

Publication No. _____

David Joseph Sidote, PhD.

The University of Texas at Austin, 2004

Supervisor: David W. Hoffman

The process of tRNA maturation requires extensive posttranscriptional modification. These modifications include 5' leader removal, 3' CCA addition, intron splicing, and extensive base modification. The enzyme responsible for the removal of the 5' leader is known as Ribonuclease P (RNase P). This ribonucleoprotein complex is present in all cells and cellular compartments that perform translation. In this dissertation, the archaeal Ribonuclease P protein aRpp29 from *Archaeoglobus fulgidus* was structurally characterized using nuclear magnetic resonance (NMR) and X-ray crystallography techniques.

The structure of aRpp29 consists of an amino terminal α -helix followed by a six-stranded, anti-parallel β -sheet and then an α -helix at the carboxy terminus. The three dimensional structure forms a semi-closed barrel, wrapped around a well-conserved hydrophobic core. The α -helices align in an anti-parallel orientation, capping the open end of the structure. There are well-conserved charged residues that may present a surface for interactions with either the RNase P RNA or the substrate tRNA. An interesting feature of this structure is an internal salt bridge formed by a triad of conserved residues. This feature may confer the unusual stability observed over a wide range of pH and temperatures.

The investigation of the structure of aRpp29 using NMR revealed distinct differences when compared to the structure solved using X-ray crystallography. The solution structure forms the same six stranded anti-parallel β -sheet but lacks stable amino and carboxy terminal helices, indicating that ~40% of the protein is in an equilibrium between a folded and unfolded state. This finding was further investigated by measuring circular dichroism and amide proton exchange rates.

The structure of aRpp29 reveals that it is a variant of the Sm-fold (or like-SM) family of proteins. These proteins are involved in a variety of processes involving RNA, including splicing and transcriptional regulation. Sm proteins and

their homologs form heptameric rings in solution, although untracentrifugation studies show that aRpp29 forms a monomer in solution.

The structural studies of archaeal Ribonuclease P protein Rpp29 presented in this dissertation provide an essential step toward understanding the overall architecture of ribonuclease P.

Table of Contents

List of Tables	xiii
List of Figures	xiv
Chapter 1 Introduction	1
1.1 The tRNA Maturation Process	1
1.2 Bacterial RNase P	2
1.3 Archaeal RNase P	7
1.4 Eukaryal RNase P	15
1.4.1 Nuclear	15
1.4.2 Mitochondrial	24
1.4.3 Chloroplast.....	25
1.5 The Relationship Between RNase P and RNase MRP	26
1.6 Understanding RNase P Structure and Function Using Archaea as a Model System	28
Chapter 2 NMR Structure of RNase P Protein aRpp29 from <i>A. fulgidus</i>	31
2.1 Introduction.....	31
2.1.1 Sample Preparation.....	32
2.1.2 Spectrum Assignment.....	33
2.1.3 Deriving Structural Constraints from Spectra	38
2.1.4 Structure Refinement	39
2.2 Materials and Methods.....	40
2.2.1 Cloning, Expression, and Purification of aRpp29.....	40
2.2.2 Preparation of Enriched Samples	42
2.2.3 Expression and Purification of TEV Protease	44
2.2.4 NMR Spectroscopy	45
2.2.5 NMR Dynamics.....	50

2.2.6 Structure Calculation	51
2.3 Results and Discussion	55
2.3.1 Initial Sample Characterization	55
2.3.2 Description of the aRpp29 Structure	58
2.3.3 Dynamics of aRpp29 in Solution	65
2.3.4 The termini of aRpp29 are Flexible	66
2.3.5 Data Collection at pH 3 and 20°C	69
Chapter 3 X-ray Crystal Structure of the RNase P Protein aRpp29 from A. fulgidus	70
3.1 Introduction.....	70
3.1.1 Crystallization of Proteins.....	71
3.1.2 Symmetry and Space Groups	72
3.1.3 Crystal Characteristics.....	72
3.1.4 Data Collection.....	73
3.1.5 Data Processing.....	74
3.1.6 The Phase Problem	74
3.1.7 Refinement and Model Building	82
3.1.8 The Crystal Structure of aRpp29.....	84
3.2 Materials and Methods.....	85
3.2.1 Crystallization and X-ray Diffraction	85
3.2.2 Crystal Structure Determination and Refinement	87
3.3 Results and Discussion	92
3.3.1 Overview of aRpp29 Crystallography	92
3.3.2 Description of the Crystal Structure	95
3.3.3 Conserved Surface Residues.....	106
3.3.4 What About Molecular Replacement?	110
Chapter 4 Biophysical Analysis of the Termini of aRpp29	113
4.1 Introduction.....	113
4.1.1 Amide Proton Exchange	114

4.1.2 Circular Dichroism Spectroscopy	115
4.2 Materials and Methods	119
4.2.1 NMR Spectroscopy	119
4.2.2 Amide Proton Exchange Rate Measurements	119
4.2.3 Circular Dichroism Measurements	121
4.2.4 Analytical Centrifugation	122
4.3 Results and Discussion	122
4.3.1 Comparison of the Crystal and Solution Structures of aRpp29.....	122
4.3.2 Reconciling the Differences Between the Crystal and Solution Structures of aRpp29	124
4.3.3 The Effect of Elevated NaCl Concentration on the Stability of the Termini of aRpp29.....	127
4.3.4 Conclusion	129
Chapter 5 Summary and Conclusions	134
Appendix 1: Chemical Shift Assignments for aRpp29 From <i>A. fulgidus</i>	136
Appendix 2: NOE Assignments for aRpp29 From <i>A. fulgidus</i>	155
Appendix 3: Dihedral Angle Restraints for aRpp29 From <i>A. fulgidus</i>	173
Appendix 4: Hydrogen Bond Restraints for aRpp29 From <i>A. fulgidus</i>	178
Glossary	182
Bibliography.....	184
Vita	205

List of Tables

Table 1.1: Protein subunit composition of nuclear RNase P from eukarya and RNase P from archaea.....	11
Table 2.1: Reagents used for preparation of M9 minimal media for isotope labeling of <i>A. fulgidus</i> aRpp29 (per liter)	43
Table 2.2: Summary of NMR spectra collected for resonance assignment and structure calculation	47
Table 2.3: Summary of the refinement and structural statistics for the NMR structure of <i>A. fulgidus</i> aRpp29.....	53
Table 3.1: Structure determination and refinement statistics for the X-ray crystal structure of <i>A. fulgidus</i> aRpp29.....	93

List of Figures

Figure 1.1: X-ray crystal structure of the specificity domain of the bacterial type B RNase P RNA.....	5
Figure 1.2: Cartoon representation of the X-ray crystal structure of the P protein from <i>B. subtilis</i>	6
Figure 1.3: X-ray crystal structure of the archaeal homolog of Rpp1 from <i>P. horokoshii</i>	14
Figure 1.4: Yeast RNase P interaction map determined by yeast two-hybrid and yeast three-hybrid assays	22
Figure 1.5: Human RNase P interaction map determined by yeast two-hybrid and yeast three-hybrid assays	23
Figure 2.1: Two-dimensional TOCSY spectrum of aRpp29 at pH 5.8 and 30°C.....	35
Figure 2.2: Sequential assignments of amino acids using homonuclear NMR.....	36
Figure 2.3: Backbone assignments using heteronuclear NMR.....	37
Figure 2.4: ^{15}N - ^1H HSQC spectrum of aRpp29 selectively labeled with ^{15}N -gly/ser/trp	48
Figure 2.5: ^{15}N - ^1H HSQC spectrum of aRpp29 selectively labeled with ^{15}N -lys.....	49
Figure 2.6: One dimensional spectrum of RNase P protein aRpp29 from <i>A. fulgidus</i>	56

Figure 2.7: ^{15}N - ^1H HSMQC spectrum of the aRpp29 protein from <i>A. fulgidus</i>	57
Figure 2.8: Diagram of the six-stranded β -sheet structure of aRpp29 derived from NOE crosspeaks.....	59
Figure 2.9: Stereo view of a cartoon representation of aRpp29.....	60
Figure 2.10: Amino acid sequence alignment of <i>A. fulgidus</i> aRpp29 and RNase P proteins from other archaea and eukarya.....	61
Figure 2.11: Superposition of the backbones of 12 low energy structures of aRpp29.....	63
Figure 2.12: Plot of the root mean square deviation for the coordinates of the backbone heavy atoms for the aRpp29.....	64
Figure 2.13: Backbone structures of protein subunits containing regions lacking regular secondary structure from the archaeal large ribosomal subunit.....	68
Figure 3.1: Anomalous contribution to scattering and its use in SAD phasing.....	80
Figure 3.2: Crystals of <i>A. fulgidus</i> RNase P protein aRpp29.....	86
Figure 3.3: X-ray diffraction image of aRpp29 to 1.7 Å.....	88
Figure 3.4: Fluorescence scan of an aRpp29 crystal containing 5 selenomethionine residues.....	89
Figure 3.5: Anomalous difference map showing the positions of the selenium atoms in the aRpp29 structure.....	91
Figure 3.6: Cartoon representation of the aRpp29 native and SeMet crystal structures.....	96

Figure 3.7: Amino acid sequence alignment of <i>A. fulgidus</i> aRpp29 and RNase P proteins from other archaea and eukarya.....	97
Figure 3.8: Stereo view of the X-ray crystal structure of aRpp29 from <i>A. fulgidus</i>	98
Figure 3.9: Lattice packing of the native and SeMet forms of aRpp29	100
Figure 3.10: Two-dimensional diagram of the aRpp29 variation of the SM-fold.....	101
Figure 3.11: Structural homologs of aRpp29 exhibiting the SM-fold	104
Figure 3.12: Electrostatic surface representation demonstrating the overall positive character of aRpp29	105
Figure 3.13: Conserved surface residues provide a potential site of interaction	107
Figure 3.14: A region of the aRpp29 2IfoI-IFCl electron density map showing the three-way internal salt bridge formed by residues E40, K58, and R86	108
Figure 3.15: Stereo diagram showing two orthogonal views of the <i>A. fulgidus</i> aRpp29 average NMR structure superimposed on the crystal structure	112
Figure 4.1: Circular dichroism spectra of the Fausman standards in the helical, β -sheet, and random coil conformations	117
Figure 4.2: Circular dichroism spectra used to determine the helical content of aRpp29 in solution.....	126

Figure 4.3: Analysis of rates of saturation-transfer between solvent water and amide protons can be used to gain insight into the motions within the aRpp29 protein 128

Figure 4.4: ^{15}N - ^1H HSQC spectra of aRpp29 collected at 250 mM NaCl overlaid on a spectrum collected with 100 mM NaCl.. 130

Figure 4.5: ^{15}N - ^1H HSQC spectra of aRpp29 collected at 500 mM NaCl overlaid on a spectrum collected with 100 mM NaCl.. 131

Figure 4.6: ^{15}N - ^1H HSQC spectra of aRpp29 collected at 1 M NaCl overlaid on a spectrum collected with 100 mM NaCl.. 132

Chapter 1

Introduction

1.1 The tRNA Maturation Process

Transfer RNA (tRNA) plays a central role in biology, serving as a link between the genetic code and protein. Despite their relatively small size, tRNA molecules are extensively processed. The maturation process occurs in 5 steps: (1) The removal of the 5' leader sequence by the ribonucleoprotein complex RNase P, (2) the removal of the 3' trailer sequence by a combination of endo- and exonucleases, (3) the addition of the 3'CCA in the tRNA of eukarya, some archaea, and many bacterial species, (4) eukarya and some archaeal tRNA genes contain introns which must be precisely removed. The intron is excised by an endonuclease and a ligase that joins the exons, finally, (5) numerous base modifications are made at multiple positions. More than 80 base modifications from tRNAs have been identified in various organisms (Björk, 1995).

The first step in the tRNA maturation process is performed by the ubiquitous ribonucleoprotein, RNase P. The removal of the 5' leader is a uniform process across all kingdoms of life as RNase P is found in all organisms and organelles that synthesize tRNA and in almost all cases contains an RNA moiety

crucial for cleavage of the 5' leader from the pre-tRNA. In order to provide a context for the results presented in this dissertation, the data that makes up our current understanding of RNase P, its function and structure, will be discussed below.

1.2 Bacterial RNase P

Bacterial RNase P is the best-characterized example of the RNase P family. The holoenzyme contains two components, a large RNA subunit ranging in size from 350 - 450 nucleotides, and a protein subunit ranging in size from 120 – 140 amino acids (Brown 1998). The RNA is encoded by the *mpB* gene and can be classified as one of three types (Haas *et al.*, 1996). Type A, or ancestral, is found in most bacteria, that is, Gram-negative and the high Gram-positives. The *E. coli* RNase P RNA is the prototype for this category. The Type B is found in the low Gram-positives with *B. subtilis* being the prototype. The type C is found in green nonsulfur bacteria (Haas *et al.*, 1998). The differences between the three types are the positions of the P6 and P10.1 stem loops (Siegel *et al.*, 1996). Covariation analysis has led to the determination of a minimal consensus sequence (Haas *et al.*, 1994).

Three-dimensional models of the core region have been constructed for both the *E. coli* and *B. subtilis* RNase P RNAs using covariation data, solvent accessibility studies, and cross-linking (Harris *et al.*, 1994, Westhof *et al.*, 1994,

Westhof *et al.*, 1996, Harris *et al.*, 1997, Chen *et al.*, 1998). The central element in the model is the P4 helix formed by 2 distantly located conserved sequences. The domains J16-15 contain a binding site for the pre-tRNA 3' CCA end and may also be a binding site for catalytic Mg²⁺ ions (Oh *et al.*, 1998). Specific phosphates have been identified that bind the acceptor stem of the substrate, Mg²⁺, as well as other metal ions that are possibly important for substrate binding and catalysis (Hardt *et al.*, 1995, Kleineidum *et al.*, 1993, Kufel *et al.*, 1996, Warnecke *et al.*, 1996). The pre-tRNA 3' CCA has been shown to be important for recognition and binding by RNase P in *E. coli* (Oh *et al.*, 1994, Sarvard *et al.*, 1996, Tallsjo *et al.*, 1996. Hardt *et al.*, 1995).

In both *E. coli* and *B. subtilis*, the T stem loop of the pre-tRNA binds to the P8 region of the RNase P RNA (Pan *et al.*, 1995, Loria and Pan 1997, Chen *et al.*, 1998). In *E. coli*, the exact location of the cleavage site on the pre-tRNA substrate is determined by the 3' end and the acceptor and T-stem helices (Kirsebom, 1995, Kirsebom and Vioque, 1996). A conserved adenosine, A248, in *E. coli* RNase P RNA interacts with the substrate nucleotide immediately 5' of the cleavage site (Zahler *et al.*, 2003). The preferred residue for this position is a uracil (Zahler *et al.*, 2003). The biochemical data has led to the identification of two domains within the P RNA. The P RNA can be roughly divided into two regions, one consisting of the elements responsible for specificity and binding of

the pre-tRNA substrate, and the other that contains that contains the elements necessary for catalysis of the substrate pre-tRNA.

Although no three-dimensional data exists for the holoenzyme or the complete P RNA, the crystal structure of the 154 nucleotide specificity domain from *B. subtilis* (type B) and the crystal structure of the 161 nucleotide specificity domain from *T. thermophilis* (type A) have been determined by X-ray crystallography. The domains from both types of RNase P RNA perform the same function and the functionally important domains are conserved. The differences arise from the tertiary fold of the RNA, resulting from structural changes due to variability in the secondary elements. These elements function to stabilize the core region of the RNA. These data provide a structural basis for the biochemical analysis (Krasilnikov et al. 2003, Krasilnikov, 2004) (Figure 1.1).

The three-dimensional structure of the P protein from *S. aureus* has been solved by NMR methods and from *T. maritima* and *B. subtilis* by X-ray crystallography (Spitzfaden *et al.*, 2000, Kazantsev *et al.*, 2003, Stams *et al.*, 1998). The structure reveals an arrangement of 4 α -helices surrounding a 4 stranded β -sheet (Figure 1.2). Cross-linking experiments indicate that the central cleft formed by one α -helix and the β -sheet directly interacts with the single stranded 5' leader sequence of the pre-tRNA, 4-8 bases from the site of cleavage (Niranjanakumari *et al.*, 1998). Binding studies with the pre-tRNA and the P protein show that the

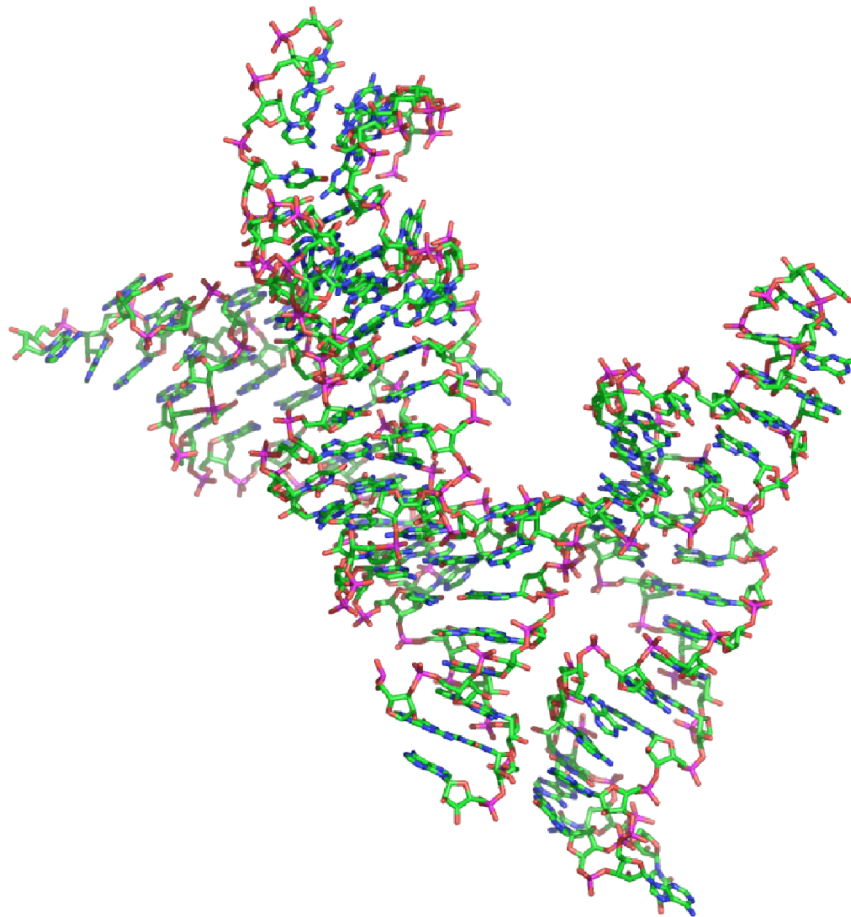


Figure 1.1. X-ray crystal structure of the specificity domain of the bacterial type B RNase P RNA. Stick representation of the 154 nucleotide specificity domain of the *B. subtilis* RNase P RNA (PDB code 1NBS, Krasilnikov *et al.*, 2003). The structure reveals the interactions required to maintain the overall architecture of this domain as well as regions that have been shown to be important in substrate recognition.

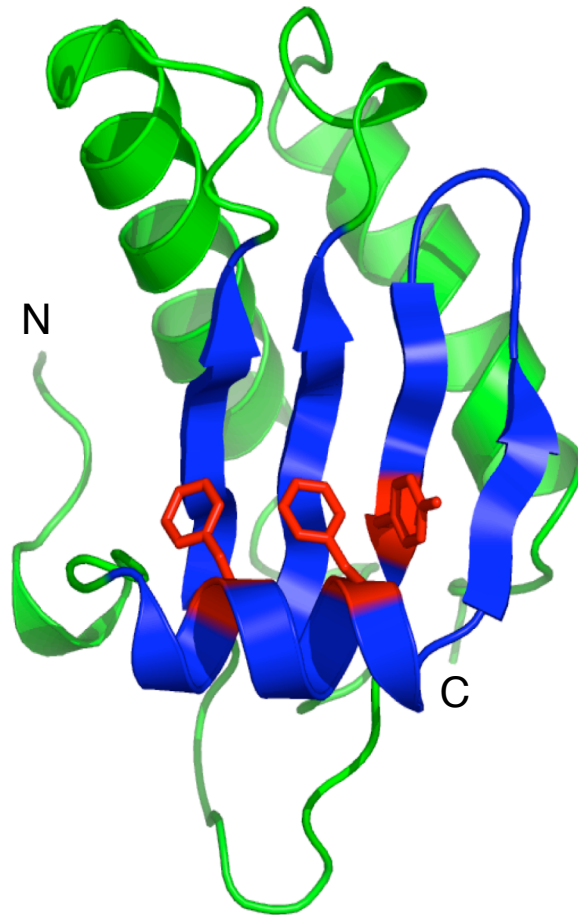


Figure 1.2. Cartoon representation of the crystal structure of the P protein from *B. subtilis*. The X-ray crystal structure of *B. subtilis* P protein has been solved at 2.6 Å (PDB code 1A6F, Stams *et al.*, 1998). The fold of the protein consists of a 4 stranded β -sheet with 3 α -helices oriented around the sheet. The P protein is required for efficient catalysis under physiological conditions as it enhances substrate binding through interaction with the pre-tRNA substrate. The pre-tRNA binds to the central cleft of the protein (colored blue). Aromatic residues Phe16, Phe20, and Tyr 34 are directly involved in mediating this interaction (colored in red)

5' leader sequence must be 4 -5 nucleotides in length for optimal affinity by the P protein (Crary *et al.*, 1998). Moreover, aromatic residues Phe16, Phe20, and Tyr 34 located in the central cleft of the *E. coli* protein, identified by mutagenesis, are involved with substrate binding (Gopalan *et al.*, 1997). Two other sites of possible RNA interaction are a negatively charged metal binding loop and the conserved RNR motif that is situated in an unusual β - α - β crossover connection.

Using small angle X-ray scattering, the *B. subtilis* RNase P enzyme was found to exist as a dimer containing two P RNA molecules and two P protein molecules in the absence of the pre-tRNA substrate (Fang *et al.*, 2001). This dimerization was found to be sensitive to the environment and is mediated by the protein subunit (Fang *et al.*, 2001, Barrera *et al.*, 2002). In substrates that contain two tRNA molecules, the enzyme-substrate molecule contains the dimeric form of the holoenzyme indicating that this form may exist in vivo in order to process tRNA molecules that are organized in large operons (Barrera *et al.*, 2002).

1.3 Archaeal RNase P

Archaeal RNase P has only recently been the subject of closer investigation. The archaeal RNase P RNA subunit is similar in both primary sequence and secondary structure to the phylogenetically conserved core of the bacterial P RNA (Haas *et al.*, 1996). The archaeal RNase P RNA can be

classified into two groups, type A, which is similar to the type A found in most bacteria, and type M, which is a small class that lacks elements known to be involved in substrate recognition and binding in the bacterial P RNA (Andrews *et al.*, 2001, Harris *et al.*, 2001). These distinct differences consist of alterations in the P7, P9, and P10/P11 elements, as well as deletions of the P8, P15/16/17/18, and P6 elements (Massire *et al.*, 1998, Harris *et al.*, 2001). There are no additional elements found in the type M RNase P RNA that may compensate for these differences, but it is possible that the protein component may be involved in restoring the necessary functions played by the missing elements (Harris *et al.*, 2001).

Very low levels of catalytic activity have been detected in the type A RNase P RNA, in the absence of the protein component, from *methanobacteria*, *thermococci*, and *halobacteria* in conditions of very high ionic strength (300mM MgCl₂, 3.0-4.0 M ammonium acetate), 45-50°C, and pH 8.0 (Pannucci *et al.*, 1999). No activity could be detected for the type M RNase P RNA (Pannucci *et al.*, 1999). The low levels of activity may be due to poor substrate affinity, as substrate cleavage rates increase linearly with substrate concentration up to 10 uM (Pannucci *et al.*, 1999).

Chimeric RNase P, composed of the archaeal P RNA and the *B. subtilis* protein subunit elicited activity in conditions of low ionic strength, consistent with data using only the conserved core of the bacterial P RNA and the protein

subunit. These data suggest that RNA contains the necessary elements needed for substrate recognition and cleavage, but lacks the phylogenetically variable stability elements and the RNase P RNA may be structurally unstable in the absence of the protein component (Hall *et al.*, 2002). Chimeric enzymes were produced from type M RNase P RNA but activity was not detected, although an active enzyme consisting of *H. volcanii* RNase P RNA and the *B. subtilis* P protein has been reported (Neiulandt *et al.*, 1991).

The characteristics of RNase P from representatives of both type A and type M RNase P have been reported. The RNase P holoenzyme from *Methanothermobacter thermoautotrophicus* was found to have a buoyant density of 1.42 g/ml indicating a ribonucleoprotein complex with a substantial protein component (Andrews *et al.*, 2001). Enzymatic activity for *M. thermoautotrophicus* was determined to have a K_m of 34.5 nM and a k_{cat} of 52.6 min⁻¹ at 65°C in 800 mM ammonium acetate with 5 mM MgCl₂ (Andrews *et al.*, 2001).

The RNase P holoenzyme from *Methanococcus jannashii*, a representative of type M RNase P, was isolated and found to have a buoyant density in Cs₂SO₄ of 1.39 g/ml consistent with that from the type A RNase P found in *M. thermoautotrophicus* (Andrews *et al.*, 2001). This is consistent with an enzyme complex consisting of roughly an equal ratio of protein to RNA (Andrews *et al.*, 2001). The enzyme was found to be active over a wide range of temperatures, with the optimal activity being recorded at 80°C (*M. jannashii's*

optimal growth temperature is 83°C), magnesium concentration, and ionic strength. The kinetic properties were examined by reaction velocity as a function of pre-tRNA concentration. The K_m was found to be 68.4 nM, and the k_{cat} was determined to be 34.1 min⁻¹ at 50°C in 50 mM ammonium acetate, and 30 mM MgCl₂ (Andrews *et al.*, 2001).

Comprehensive searches of archaeal genomes have not produced any open reading frames homologous to the bacterial RNase P protein subunit. The density of the *M. thermoautotrophicus* RNase P holoenzyme, which contains a type A RNA, in Cs₂SO₄ (1.42 g/ml) is consistent with a holoenzyme ~200 kDa, the RNA has a molecular weight ~96 kDa and the protein component ~98 kDa (Hall *et al.*, 2002). Interestingly, four open reading frames have been identified in archaeal genomes that have sequence similarity to four of the nine proteins from *S. cerevisiae* RNase P (Hall *et al.*, 2002, Kouzuma *et al.*, 2003). The four protein subunits are homologous to the yeast RNase P proteins Pop4p, Pop5p, Rpr2p, and Rpp1p, with molecular weights of 10.7, 14.6, 17.0 and 27.7 kDa respectively (Hall *et al.*, 2002) (Table 1.1). Each of these proteins has been shown to physically interact with RNase P by immunoprecipitating RNase P activity from partially purified extracts using antisera generated against each of the proteins (Hall *et al.*, 2002). The P RNA from *M. thermoautotrophicus* is active at very low levels in the absence of the protein in high ionic conditions, but the presence of the protein component enhances the substrate affinity over 1000 fold (Pannucci

Table 1.1 Protein subunit composition of nuclear RNase P from eukarya and RNase P from archaea.

Yeast	Human	Archaea
Pop1p	hPop1	-----
Pop3p	-----	-----
Pop4p	Rpp29	aRpp29
Pop5p	hPop5	aPop5
Pop6p	-----	-----
Pop7p	Rpp20	-----
Pop8p	-----	-----
Rpr2p	Rpp21	aRpp21
Rpp1p	Rpp30	aRpp30
-----	Rpp38	-----
-----	Rpp40	-----
-----	Rpp25	-----
-----	Rpp14	-----

et al., 1999). The K_m of the *M. thermiautotrophicus* P RNA alone is 40 μ M, while the K_m of the holoenzyme was determined to be 34.5 nM (Pannucci *et al.*, 1999, Andrews *et al.*, 2001).

The calculated molecular weight of the protein component, based on the CsSO_4 data, was determined to be 98 kDa. The expected molecular weights of the four proteins shown to be associated with the archaeal RNase P holoenzyme total \sim 70 kDa, suggesting that there are yet to be identified factor(s) or the stoichiometry of the protein subunits is not one per holoenzyme (Hall *et al.*, 2002).

An investigation by Kouzuma *et al.*, (2003) demonstrated that the RNase P enzyme from *Pyrococcus horikoshii* could be reconstituted from proteins expressed in *E. coli* and *in vitro* transcribed RNA. Each of the protein subunits homologous to those from yeast and identified by Hall *et al.*, (2001) was able to bind the RNase P RNA subunit and together could restore activity in conditions of low ionic strength. These reconstitution experiments showed that the yeast homologues of Pop4p, Pop5p and Rpr2p are sufficient for RNase P activity, and the addition of the homolog of Rpp1p greatly increases the ability of the enzyme to bind its substrate pre-tRNA (Kouzuma *et al.*, 2003). The optimal temperature for the *in vitro* reconstituted RNase P was observed at 55°C, while optimal activity for the native RNase P was observed at 70°C, suggesting that the RNA was in a slightly different conformation or other protein factors contribute some

sort of temperature stability (Kouzuma *et al.*, 2003) These data demonstrate that the four proteins constitute the minimal protein component necessary for RNase P activity (Kouzuma *et al.*, 2003).

Structural characterization of the archaeal RNase P has recently provided valuable insight into its overall architecture. While structural data for the holoenzyme or the RNA component of archaeal RNase P has not been reported, the proteins have become the focus of intense structural characterization. The archaeal Rpp29/Pop4p homologue has been studied by both NMR and X-ray crystallography (Sidote and Hoffman, 2003, Boomershine *et al.*, 2003, Numata *et al.*, 2004, and Sidote *et al.*, 2004). The structure of the archaeal homolog of the Rpp29/Pop4p protein is presented in this dissertation and will be discussed in detail in subsequent chapters. The structure of the Rpp30/Rpp1p homologue from *Pyrococcus horikoshii* has been reported (Takagi *et al.*, 2004). It contains a TIM barrel structure, consisting of ten α -helices and seven β -strands and can interact with the RNase P RNA (Kouzuma *et al.*, 2003) (Figure 1.3). Surface residues thought to be important for RNA binding were investigated by site-directed mutagenesis. Residues Arg90, Arg107, and Lys123 were mutated to Ala and found to cause decreases in RNase P activity by 42-48%. Mutations made to Arg176 and Lys196 caused a reduction in activity of reconstituted RNase P by 32%. These data suggest that these residues are involved interacting with the substrate pre-tRNA or the RNase P RNA.

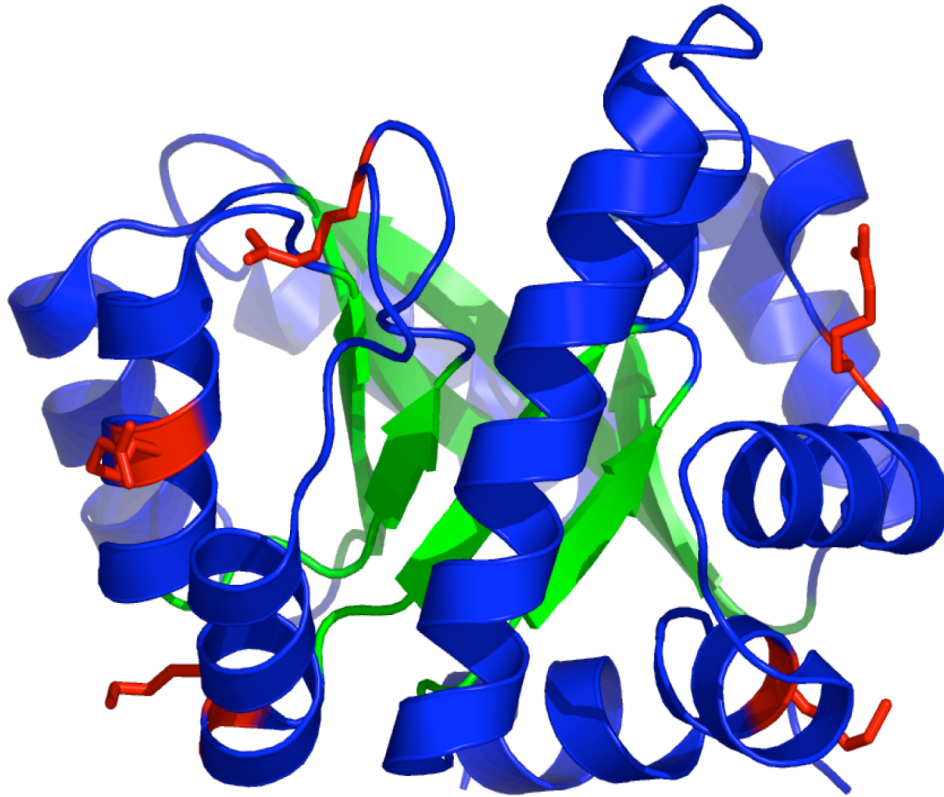


Figure 1.3. X-ray crystal structure of the archaeal homolog of Rpp30 from *P. horikoshii*. Cartoon representation of the X-ray crystal structure of aRpp30 with helices and loops colored blue and β -strands colored green (PDB code 1V77, Takagi *et al.*, 2004). The 24 kDa protein forms a TIM barrel structure, consisting of 10 α -helices and 7 β -strands. Residues important for RNase P function archaea are colored red. Mutation of these residues causes 32-48% reduction in activity.

1.4 Eukaryotic RNase P

1.4.1 Nuclear RNase P

Nuclear RNase P has been purified to homogeneity and the components of the complex determined (Lee *et al.*, 1991, Lee *et al.*, 1991, Lygerou *et al.*, 1994, Dicht *et al.*, 1997, Chu *et al.*, 1997, Stolc and Altman 1997, Stolc *et al.*, 1998). The yeast RNase P RNA subunit was first described fifteen years ago but the purification of the holoenzyme has only recently been reported. Biochemical purification has confirmed the presence of the RNA, termed RPR1, and the existence of nine protein subunits: Pop1p, Pop3p, Pop4p, Pop5p, Pop6p, Pop7p, Pop8p, Rpr2p, and Rpp1p (Chamberlain *et al.*, 1998). The protein subunits are essential for RNase P activity and yeast viability (Chamberlain *et al.*, 1998, Lee *et al.*, 1991, Lygerou *et al.*, 1994, Dicht *et al.*, 1997, Chu *et al.*, 1997, Stolc and Altman 1997). The other well studied representative of eukaryotic RNase P, human RNase P has been purified and the RNA subunit, termed H1, and 10 protein subunits have been shown to be part of the complex, termed Rpp14, Rpp21, Rpp25, Rpp29, Rpp30, Rpp38, Rpp40, hPop1, and hPop5 (Bartkiewicz *et al.*, 1989, Lygerou *et al.*, 1996, van Eenennaam *et al.*, 1999, Jarrous *et al.*, 1999, van Eennannaam *et al.*, 2001, Jarrous *et al.*, 2001). The human and yeast RNase P share some similarity as 6 of the 10 human protein subunits are homologous to the yeast protein subunits.

The RNA Component of Nuclear RNase P

The RNA component of RNase P has been isolated from several eukaryotic organisms but the majority of the data comes from analysis of the yeast RNase P RNA, RPR1 (Frank and Pace, 1998, Brown. 1999, Frank *et al.*, 2000). In *S. cerevisiae*, the RPR1 RNA is transcribed by RNA polymerase III as a 486-nucleotide precursor (Lee *et al.*, 1994). The 84-nucleotide leader sequence and the 33-nucleotide trailer sequence are removed, although the enzymes that perform the cleavage are unknown (Lee *et al.*, 1991).

In yeast cells growing at a normal rate, the ratio of pre-RPR1 to mature RPR1 is ~1:9 and both can be immunoprecipitated by affinity tagged protein subunits (Chamberlain *et al.*, 1998, Lygerou *et al.*, 1994, Dichtl *et al.*, 1997, Chu *et al.*, 1997, Stolc and Altman, 1997). Both the pre-RPR1 and mature RPR1 can be purified using similar protocols, suggesting that the RNA processing occurs after the protein subunits are present on the RNA (Chamberlain *et al.*, 1996). It has been shown that the pre-RPR1 enzyme is as active as the enzyme containing mature RPR1 (Srisawat *et al.*, 2002). Interestingly, the protein subunits Pop3p and Rpr2p are not part of the pre-RPR1 complex and seem to bind to the complex after RPR1 has been processed (Srisawat *et al.*, 2002).

The pre-RPR1 and mature RPR1 forms are localized to the nucleolus, suggesting that the assembly of the RNase P holoenzyme takes place there as well (Bertrant *et al.*, 1998, Srisawat *et al.*, 2002). Studies of the human H1 RNA

have also suggested localization occurs in the nucleolus (Jacobson *et al.*, 1997). A complex of seven Sm-like proteins (lsm2-8) has been shown to interact with the pre-RPR1 RNA. Sm/Lsm proteins have been shown to be involved in mRNA processing (Toro *et al.*, 2001). However, they are not part of the mature RPR1 complex and therefore may be involved in assembly or recycling of RNase P (Salgado-Garrido *et al.*, 1999, Pannone *et al.*, 2001).

The secondary structure of the RPR1 RNA has been predicted based on phylogenetic analysis, and structure sensitive RNA footprinting (Tranguch *et al.*, 1993, Zimmerly *et al.*, 1990, Frank *et al.*, 2000, Tranguch *et al.*, 1994). A conserved core structure can be identified from bacteria to eukarya (Chen *et al.*, 1997, Frank *et al.*, 2000). There are five conserved regions that contain conserved residues and stems at similar positions, these regions are called CR1-CRIV (Xiao *et al.*, 2002). The crystal structure of the RPR1 RNA has not yet been reported. Data regarding the structure and function of the RPR1 RNA comes mainly from mutagenesis, crosslinking, chemical probing and computer generated models.

The P4 helix, formed by base pairing of the CR1 and CRV regions, has been predicted to be the catalytic core of the bacterial RNase P RNA (Altman 1993, Kazakov *et al.*, 1991, Darr *et al.*, 1992, Guerrier-takada *et al.*, 1993, Haas *et al.*, 1994, Harris *et al.*, 1995, Pagan-Ramos *et al.*, 1996). It has also been shown to be a binding site for the P protein in Bacterial RNase P (Biswas *et al.*,

2000). This same region was studied in the RPR1 RNA using mutagenesis and was found to be important for substrate binding, catalysis, and RPR1 maturation (Pagan-Ramos *et al.*, 1996). Several of the changes in and around the P4 helix caused a 10-fold reduction in catalytic efficiency, as well as decreases in levels of the mature RPR1 RNA (Pagan-Ramos *et al.*, 1996). RNA footprinting results demonstrate that the P4 helix is not accessible to nucleases, suggesting that is buried (Tranguch *et al.*, 1994).

As is the case with most RNA enzymes, magnesium concentration plays an important role. The CRII, CRIII, P10/P11 and P12 regions form a domain whose structure and function depend on magnesium concentration (Pagan-Ramos *et al.*, 1996, Ziehler *et al.*, 1998). Deletion experiments have shown that this region is vital for RNase P activity and viability in yeast (Pagan-Ramos *et al.*, 1994). Mutagenesis of region IV produces large decreases in K_{cat} but negligible changes in K_m , and do not alter the ratios of pre-RPR1 to mature RPR1 (Pagan-Ramos *et al.*, 1996).

The RPR1 RNA contains some unique features not observed in the bacterial RNase P RNA. The P3 element can be found in both the bacterial and yeast RNase P RNA, but the yeast element contains a helix-loop-helix structure not found in the bacterial version (Tranguch *et al.*, 1993, Pittule *et al.*, 1998, Chen *et al.*, 1997, Frank *et al.*, 2000). The P3 region of RPR1 has shown to be crucial for RNase P activity, mutational analysis results in defects in pre-tRNA

processing, RPR1 maturation, and disruption of the binding between RPR1 and Pop1p (Ziehler *et al.*, 2001). The function of the P3 element has been studied in the human RNase P RNA and found to be important for localization to the nucleolus as well as acting as a protein binding site (Yuan *et al.*, 1991, Liu *et al.*, 1994, Jacobson *et al.*, 1997). The function of the P15 region is unknown. The P15 region in the bacterial RNA is larger than the simple structure found in yeast and altogether absent in the human RNase P RNA (Xiao *et al.*, 2002).

The Protein components of Nuclear RNase P

The protein subunits of nuclear RNase P from both yeast and humans have been characterized. RNA foot-printing shows that the majority of the yeast RNase P RNA subunit is covered by protein (Tranguch *et al.*, 1994). The yeast proteins, designated Pop proteins, for processor of precursor, range in size from 15.5 to 100.5 kDa. All of the proteins are very basic ($pI > 9$), with the exception of Pop5p, which has a pI of 7.8, and Pop8p, which is acidic with a pI of 4.6. It is interesting to note that none of the proteins contain a defined RNA binding motif, rather they contain only stretches of basic residues. Some of the proteins contain putative nuclear localization signals.

Currently, there is no data on the assembly order of RNase P. It is possible that the proteins bind the RNA as a preformed complex, or individually. Depletion studies of most of the RNase P proteins from yeast show that they are

necessary to maintain the proper ratio of mature RPR1 RNA (Chamberlain *et al.*, 1998, Chu *et al.*, 1997, Stolc *et al.*, 1997, Stolc *et al.*, 1998). Specific activities and functions of the yeast protein subunits in relation to RNase P activity are unknown at this time. Some human RNase P protein subunits have been found to possess specific activities. It has been reported that Rpp14 binds to OIP2 and forms a 3' → 5' exoribonuclease with phosphorolytic activity. This has only been demonstrated for crude extracts, as further purification removes this activity (Jiang and Altman. 2002). Rpp14, Rpp21, and Rpp29 have been shown to bind pre-tRNA in gel mobility assays but the specificity for such interaction remains to be verified (Jarrous *et al.*, 2001, Jarrous 2002). It has been reported that Rpp20 contains ATPase activity and contains a signature motif found in a subunit of the ABC transporter, a regulator of membrane traffic, and a variation of the DEAD box motif found in many ATPases (Li *et al.*, 2001). This DEAD box like motif is not present in the yeast homolog, Pop7p, indicating that this protein may fulfill a different role in the yeast RNase enzyme. Rpp20 has been shown to interact with the small heat shock protein, Hsp27. The nature of this interaction may be in a regulatory capacity as the addition of Hsp27 stimulates RNase P activity in a concentration dependent manner (Jiang *et al.*, 2001). Although the specific functions of the nuclear RNase P protein subunits remains to be established, it may be that they serve multiple roles, not just in substrate recognition and cleavage, but also in discerning which RNAs are to be processed.

The crystal structure of the eukaryotic RNase P holoenzyme has not yet been determined, although interaction maps for both yeast and human RNase P have been generated using cross-linking and yeast two-hybrid and yeast three-hybrid assays (Fields *et al.*, 1989, Houser-Scott *et al.*, 2002, Jiang and Altman, 2001). The protein-protein interaction between the yeast RNase P protein subunits was mapped using yeast two-hybrid assay (Houser-scott *et al.*, 2002). It was determined that Pop1p, Pop4p, Pop5p, Pop6, Rpp1p and Rpr2p interact with at least one other subunit and Pop4p and Pop1p make multiple interactions (Houser-Scott *et al.*, 2002). Pop4p interacts with 7 of the 8 protein subunits (Houser-Scott *et al.*, 2002). Yeast three-hybrid assays were used to map the RNA-protein interactions and revealed that Pop4p and Pop1p interact with the RPR1 RNA (Houser-Scott *et al.*, 2002). The Pop1p-RNA interaction has been mapped to the P3 region of the RPR1 RNA, although the Pop4p region remains to be determined (Zeihler *et al.*, 2001) (Figure 1.4).

The human RNase P interaction map has also been constructed and shows some variation to the yeast RNase P (Jiang and Altman 2001) (Figure 1.5). Yeast two-hybrid assays determined that Rpp21, Rpp29, Rpp30, Rpp38, Rpp40, and hPop1 make extensive but weak interactions with the other subunits (Jiang and Altman 2001). Using UV crosslinking and yeast three-hybrid assays, the H1 RNA was shown to interact with Rpp21, Rpp29, Rpp30, and Rpp38 (Jiang *et al.*,

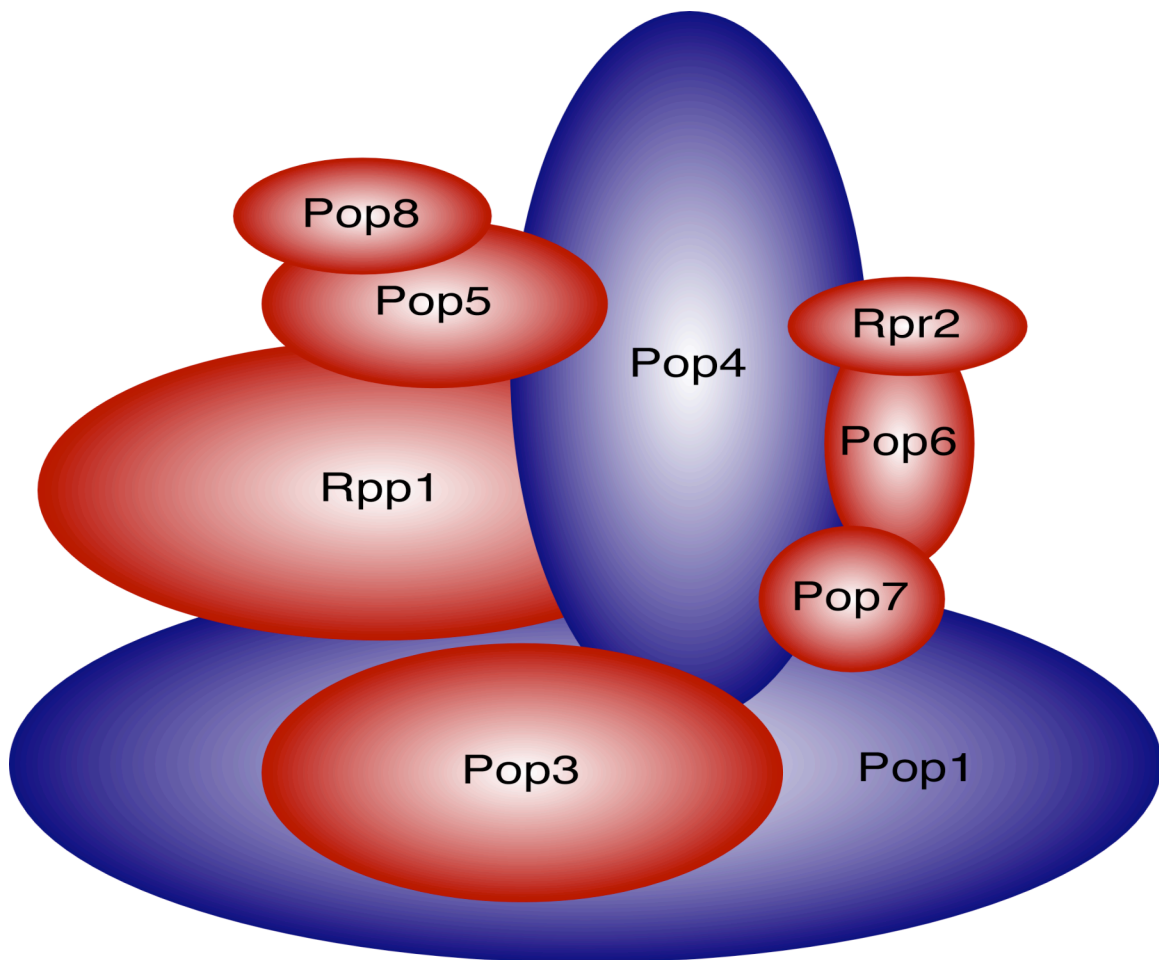


Figure 1.4. Yeast RNase P interaction map determined by yeast two-hybrid and yeast three-hybrid assays. The individual protein subunits are represented as ellipsoids with sizes proportional to the molecular weight of each protein. Protein-protein interactions were detected by yeast two-hybrid assays. The shaded ellipsoids have been shown to interact with the yeast RNA subunit (shown in blue) as determined by yeast three-hybrid assay. Interactions detected between protein subunits are depicted by overlapping ellipsoids. Pop4p forms the center of the enzyme, interacting with 7 of the 8 protein subunits as well as the RNA. Interactions between Pop1 and the RNase P RNA have been demonstrated but it is not known where Pop4p interacts with the RNase P RNA. Figure was adapted from Houser-Scott *et al.*, 2002.

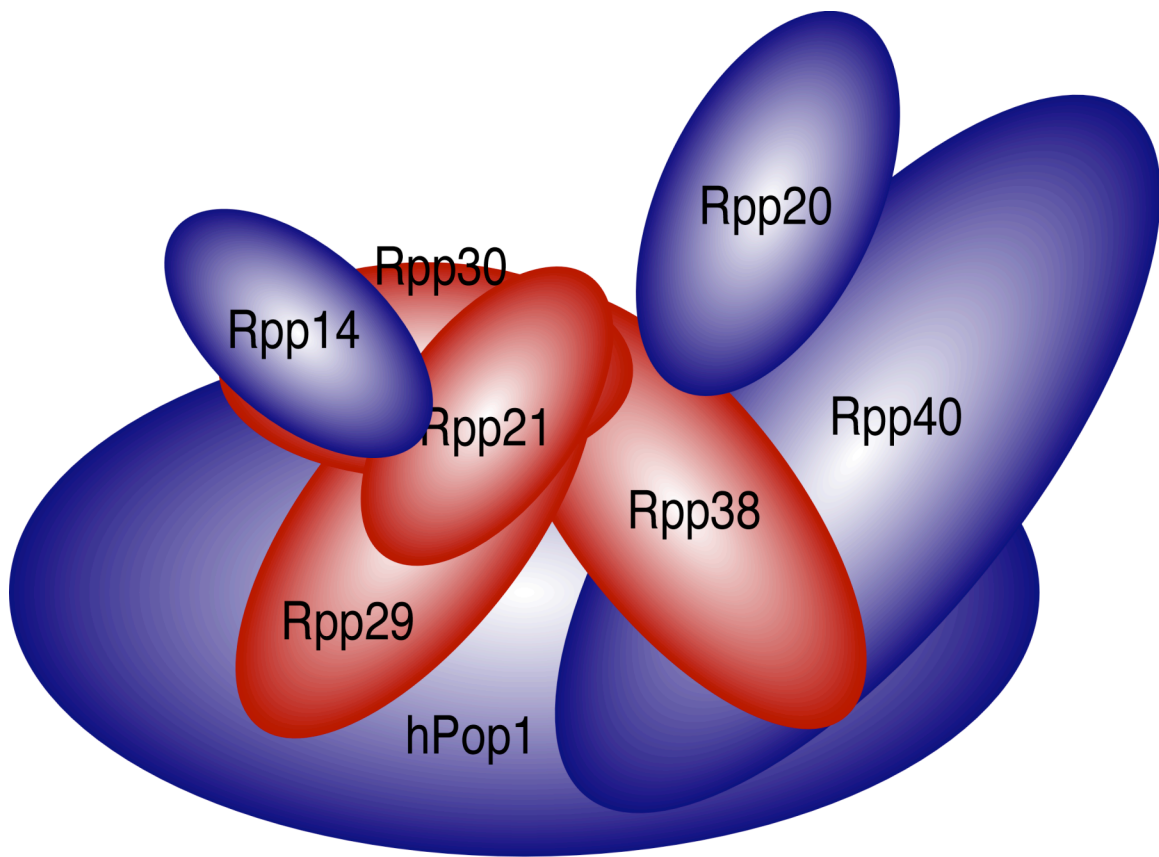


Figure 1.5 Human RNase P interaction map determined by yeast two-hybrid and yeast three-hybrid assays. Each protein subunit is represented as an ellipsoid and the size is roughly proportional to the molecular weight of the protein. Overlapping ellipsoids indicate that protein-protein interactions were detected. The red ellipsoids indicate that the protein subunit was shown to interact with the RNA subunit in yeast three-hybrid assays. As in the yeast interaction map, the p29 (Rpp29) protein interacts with many of the protein subunits and the PNase P RNA. Interactions are represented as overlapping ellipsoids. Figure was adapted from Jiang et al., 2001 and Jiang and Altman, 2001). This is in contrast to the yeast interaction map, as the only subunit that interacts in both organisms is Rpp29.

1.4.2 Mitochondrial RNase P

The activity of RNase P from mitochondria has been characterized mainly in yeast. In *S. cerevisiae*, mitochondrial RNase P consists of a mitochondrial genome encoded RNA subunit (Rpm1r) that is 490 nucleotides, and a 105 kDa protein subunit (Rpm2p) that is encoded in the nuclear genome (Miller and Martin 1983, Underbrink-Lyon *et al.*, 1983, Morales *et al.*, 1992, Dang *et al.*, 1993). Phylogenetic analysis has led to the proposal of a conserved RNA core structure that is similar to the bacterial RNA core structure (Wise 1991). The protein subunit is required for RNase P activity as mutations of the Rpm2p gene causes accumulation of pre-tRNA lacking processing of the 5' end (Morales *et al.*, 1992). Moreover, Rpm2p is needed for proper processing of the mitochondrial RNase P RNA (Stribinskis *et al.*, 1996, Stribinskis *et al.*, 2001). Deletion of the gene altogether, prevents the fermentative growth of yeast cells (Kassenbrock *et al.*, 1995).

Human mitochondrial RNase P composition is still a matter of debate. It was first reported that mitochondrial RNase P from HeLa cells consists entirely of protein and exhibits mitochondrial specific substrate specificity (Rossmanith and Karwan 1998, Roassmanith *et al.*, 1995). Recent data shows that the mitochondrial RNase P may contain an RNA identical to the HI RNA found in the nuclear RNase P (Puranam and Attardi, 2001). It should be noted that there is a

possibility that the presence of H1 RNA may be due to a contaminant from a non-mitochondrial cellular compartment.

1.4.3 Chloroplast RNase P

Chloroplast RNase P activity has been characterized in tobacco and spinach (Frank and Pace 1998). The presence of an RNA component has not been established for this RNase P and presents the only exception. Several lines of evidence indicate that chloroplast RNase P may consist entirely of protein: 1) spinach chloroplast is insensitive to micrococcal nuclease digestion, 2) No RNA has been detected in purified enzyme preparations, and 3) the buoyant density of the enzyme in CsCl_2 is 1.28 g/ml, which is the same buoyant density as the protein (reviewed in Gegenheimer, 1995, Pace and Brown, 1995).

It appears that the cleavage mechanism used by the chloroplast RNase P differs from that of the bacterial enzyme. In the bacterial RNase P, replacement of the pro-Rp non-bridging oxygen with a sulfur at the scissile bond of the pre-tRNA greatly decreases RNase P cleavage, while the same change has little effect on the chloroplast RNase P (Thomas *et al.*, 2000) The Chloroplast RNase P can still efficiently bind to the substrate pre-tRNA with a K_m of 16 nM (Thomas *et al.*, 2000).

RNase P from cyanobacteria of the primitive alga *Cyanophora paradoxa* has been characterized (Cordier and Schön, 1999). The cyanobacterium is a photosynthetic

organelle from cyanobacteria and belongs to a different phylogenetic branch than the chloroplasts of higher plants and green alga (Cordier and Schön, 1999). The RNA component has been shown to be required for activity as micrococcal nuclease treatment removes RNase P activity (Schön, 1999). The protein components constitute 80% of the mass of the holoenzyme and are required for RNase P activity (Baum *et al.*, 1996, Pascual and Vioque, 1999, Schön, 1999). RNA footprinting data shows that the proteins cover the RNA component extensively (Cordier and Schön, 1999).

1.5 The relationship between RNase P and RNase MRP

RNase P is related to another ribonucleoprotein enzyme found only in eukarya, called RNase MRP. RNase MRP plays an important role in the processing of ribosomal RNA (Tollervey *et al.*, 1995, Lindahl and Zenger 1995). RNase MRP RNA has been found to be localized mainly in the nucleolus (Riemer *et al.*, 1988, Kiss *et al.*, 1992, Li *et al.*, 1994, Jacobsen *et al.*, 1995). It has been determined that RNase MRP cleaves pre-rRNA at the A3 site in the first internal transcribed spacer region, which is an essential step for generating mature 5.8S rRNA (Schmitt and Clayton 1993, Lygerou *et al.*, 1996, Chu *et al.*, 1994). RNase MRP has been shown to be required for yeast viability, even though its established functions have not yet to be determined (Tollervey *et al.* 1995).

Characterization of RNase MRP comes mainly from yeast and HeLa human cells. The yeast RNase MRP contains an RNA component called NME1 RNA (Schmitt and Clayton 1992, Gold *et al.*, 1989). Secondary structures of NME1 RNA have been predicted based on phylogenetic data (Forster and Altman, 1990, Karwan, 1993, Schmitt *et al.*, 1993, Schmitt, 1999). The NME1 RNA folds into a structure similar to that of RNase P, containing the regions CR1, CRIV, CRV, and the P3 loop (Lindahl *et al.*, 2000). In human RNase P and RNase MRP, the P3 domain is required for several proteins to bind near or at the P3 domain (Pluk *et al.*, 1999, Liu *et al.*, 1994, Yuan *et al.*, 1991).

The protein component of yeast RNase MRP has been found to contain eight of the nine protein subunits found in yeast nuclear RNase P (Chamberlain *et al.*, 1998). In addition to the overlapping eight proteins, each enzyme contains one distinct protein subunit. The distinct subunit found in RNase MRP is Snm1p, and in RNase P it is the Rpr2p protein. The Snm1p protein has a putative zinc binding domain and can bind the NME1 RNA in gel shift assays (Schmitt 1994). Both Rpr2p and Snm1p have been shown to bind pop4p, although Rpr2p does not interact with the RNase P RNA (Houser-Scott *et al.*, 2002).

1.5 Understanding RNase P structure and function using archaea as a model system

The archaeon *Archaeoglobus fulgidus* was selected as the preferred species for the biophysical study of RNase P presented in this dissertation. 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 (*Klenk et al., 1997*), which greatly simplifies the identification of the likely protein components of its RNase P. Moreover, it has been noted that hyperthermophiles have an excellent record of being the source of materials for successful biophysical studies.

As a step toward furthering our understanding of RNase P structure, the protein from *A. fulgidus* that is homologous to the human RNase P protein Rpp29 and the yeast RNase P protein Pop4 has been recombinantly expressed, purified and structurally characterized.

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) It's primary sequence is not significantly homologous to any protein of known structure; 2) It's sequence is homologous to known archaeal RNase P proteins from *M. thermoautotrophicus* (*Hall and Brown, 2002*) and *P. horikoshii*

(Kouzuma *et al.*, 2003), 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 (Houser-Scott *et al.*, 2002, Jiang *et al.*, 2001), 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 (Chamberlain *et al.*, 1998), 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 (Kouzuma *et al.*, 2003); 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 (Chamberlain *et al.*, 1998), providing further evidence of the importance of this conserved gene product.

This dissertation is divided into 5 chapters. In chapter 1, a detailed analysis of the current understanding of RNase P has been discussed. In chapters

2 and 3, the NMR and crystal structures of the archaeal RNase P protein aRpp29 will be presented. The results of biophysical analysis will be presented in chapter 4 and chapter 5 will summarize the results and provide prospective.

Chapter 2

NMR Structure of an Archaeal Homologue of RNase P Protein

*aRpp29 from *Archaeoglobus fulgidus**

2.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for characterizing the three-dimensional structure of proteins in solution. It exploits the intrinsic magnetic spin properties of atomic nuclei to detect atoms that are near each other in space, either by being bonded to each other or through the folded protein's spatial arrangement. This information is then used to derive the three-dimensional structure of the macromolecule. The most common nuclei having a magnetic spin used for structure determination are ^1H , ^{15}N , and ^{13}C . In order to simplify the analysis of NMR data, protein samples can be enriched with ^{15}N and ^{13}C isotopes to allow the identification of specific isotopes and chemical groups. An overview of the procedure used for determining a protein structure using NMR methods is discussed below. A more detailed explanation can be found in texts such as Wuthrich, 1986 and Cavanaugh, 1996.

The structure determination of a protein from NMR data proceeds through 3 distinct stages: 1) spectrum assignment; 2) derivation of the structural constraints from the spectra; and 3) structure refinement and refinement.

2.1.1 Sample Preparation

A typical NMR experiment will start with a concentrated, homogeneous protein sample in a suitable buffer. Buffers such as potassium and sodium phosphate are most common. Due to the insensitivity of certain NMR experiments, sample concentration should be about 1 mM or above. Normally, protein samples are prepared in 90% H₂O/10% D₂O at an acidic pH (between 3.0 and 5.0), so that the peptide backbone amide protons are visible and the exchange with the solvent is kept to a minimum. The D₂O is required to keep the magnetic field constant during the course of the experiment. The temperature at which the NMR data is collected can also have an impact on the quality of the data. Elevated temperatures increase the overall and local mobility of the protein, causing decreased resonance linewidth and increased resolution, therefore the thermal stability of the protein should be examined prior to NMR data collection. Most proteins require low concentrations of salt to maintain solubility. The ionic strength of the sample buffer should be kept to the minimum that the protein will tolerate, as high ionic strength solutions can have an effect on the sensitivity of some NMR experiments. Other reagents necessary for protein stability such as

DTT or certain buffers can be used but must be obtained in a deuterated form to prevent the introduction of artifacts in the NMR spectra. Sodium azide to a concentration of a few millimolar is added to the protein sample to inhibit bacterial and fungal growth.

2.1.2 Spectrum Assignment

In order to define a structure it is first necessary to assign as many NMR resonances as possible to specific nuclei within individual amino acids. The first step in assignment strategy is to determine the resonances associated with specific amino acid types. This can be accomplished by analyzing NMR spectra that detect through bond interactions such as TOtal COrelated Spectroscopy (TOCSY), Double-quantum filtered COrelated Spectroscopy (DFQ-COSY), and ^{15}N -edited-HSQC-TOCSY. The determination of the specific type of amino acid depends on identifying H_N , H_α , and side chain resonances that belong to the same amino acid.

Typically, the backbone amide protons are well resolved and can easily be assigned. The H_α and sidechain resonances are determined by the observation of direct peaks to the backbone amide protons. The specific amino acid type can be deduced by examining the resonances corresponding to the side chain. The sidechain assignment is crucial for identifying the specific amino acid type. The number of protons, as well as the approximate random coil chemical shift value

for each proton on a sidechain can be used as a guide. In the next step, the adjacent residues in the sequence are determined by observing sequential NOE crosspeaks between the amide proton on one amino acid and the backbone protons on the preceding amino acid. The H_{α} resonance can also provide information on the local secondary structure, as H_{α} resonances upfield of the water indicate β -sheet and resonances downfield of the water indicate a helix. Generally, when the residue types of 3 – 5 sequential amino acids are determined, their unique position in relation to the primary sequence can be made (Figure 2.1).

There are a number of factors, such as chemical shift degeneracy and incompletely observed sidechain resonances, which prevent the unambiguous amino acid type determination from homonuclear data. To circumvent these problems three-dimensional heteronuclear NMR experiments are used. Triple resonance NMR spectra, using uniformly labeled ^{15}N - ^{13}C protein, can be used to sequentially assign amino acids without prior knowledge of their type by detecting through bond interaction across the peptide backbone. The ^{13}C chemical shifts of C_{α} , C_{β} , and C_{ω} can be used to determine the sequential orientation of residues using experiments such as HNCA, HNCOC, HNCACB, and HN(CO)CACB. For example, the HNCACB and HN(CO)CACB experiments can be used in conjunction to provide connectivities between sequential C_{α} and C_{β} resonances. Even using multi-dimensional NMR methods, it is not usually possible to work

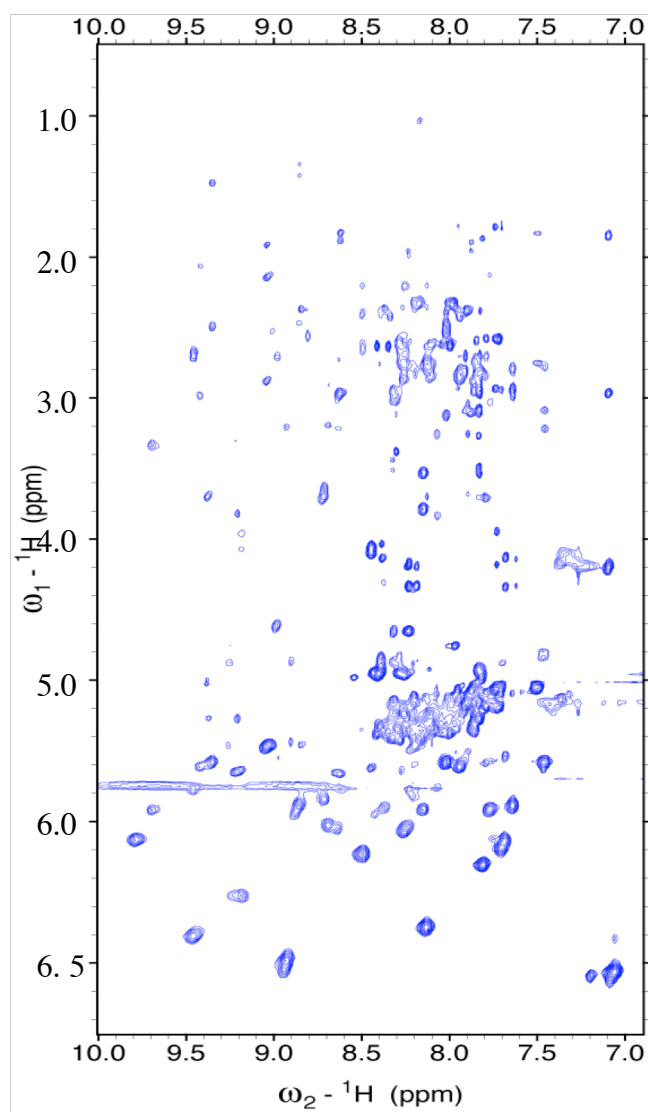


Figure 2.1 Two-dimensional TOCSY spectrum of aRpp29 at pH 5.8 and 30°C. Each axis represents a ^1H dimension. Vertical peaks having the same chemical shift belong to the same amino acid. This information can be used to determine the amino acid type and when used in conjunction with the corresponding NOESY spectrum, sequential amino acid assignments can be made.

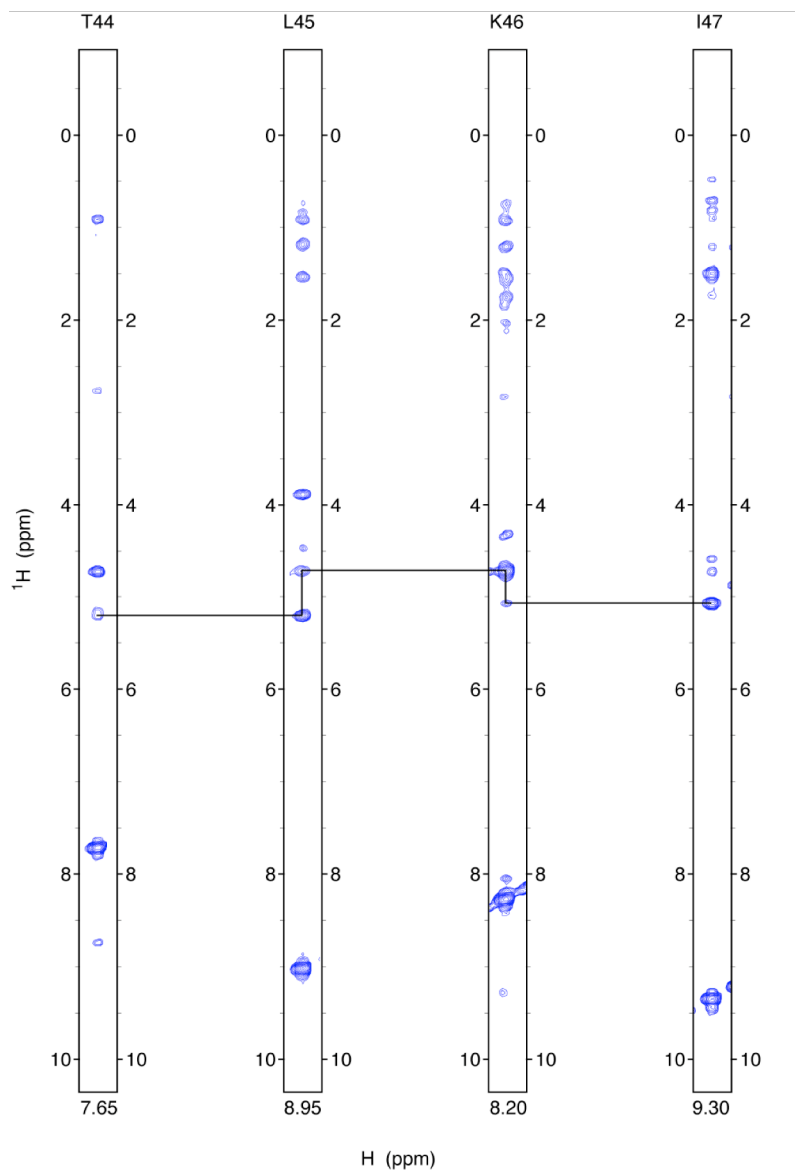


Figure 2.2. Sequential assignments of amino acids using homonuclear NMR. Slices of a three-dimensional NOESY experiment are shown. In this view the ^1H - ^1H dimensions can be viewed by ^{15}N chemical shift. Horizontal lines represent NOE interactions between the amide proton on one amino acid and the $\text{H}\alpha$ protons of the preceding amino acid.

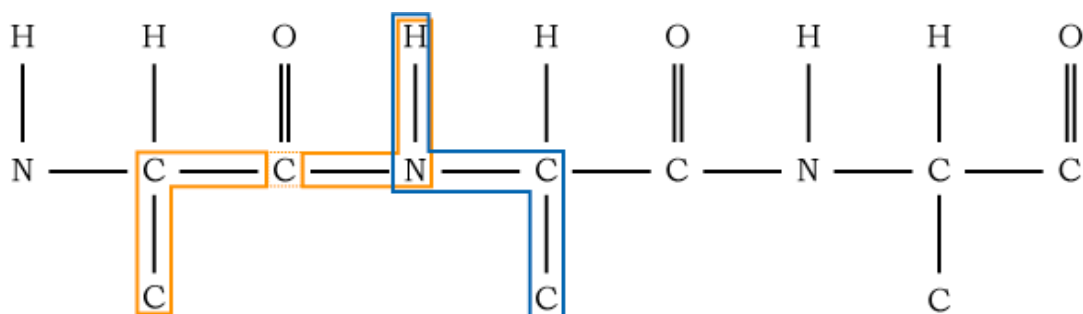


Figure 2.3. Backbone assignments using heteronuclear NMR. Schematic diagram of a peptide backbone demonstrating how a HNCACB (blue) 3D NMR experiment can be used in conjunction with a HN(CO)CACB (orange) to assign the peptide backbone. The amide proton and nitrogen chemical shifts can be used to link together sequential amino acids.

from one end of the protein to the other using triple resonance spectra as overlap in the peaks may lead to ambiguity in the assignments. Proline residues present another challenge due to the fact that they do not contain an amide proton.

Triple resonance NMR data is also useful for unambiguously determining the specific type of amino acids. Amino acids such as alanine, glycine, serine and threonine have distinct C_{α} and C_{β} chemical shifts. Specific ^{13}C type labeled samples can be prepared for amino acids such as leucine, isoleucine, valine, and phenylalanine and HNCO spectra collected to resolve assignment ambiguities. Resonance assignments for the sidechains can also be determined by analyzing HCCH-TOCSY experiments.

2.1.3 Deriving structural constraints from spectra

The three-dimensional structure of a protein is derived mainly from the through-space NMR spectra that involve the Nuclear Overhauser effect (NOE). This is a relaxation effect in which the volume of the peak is used as a measure of the distance between the two nuclei, at distances of approximately 6 Å or less. The NMR spectra designed to observe the NOE effect are known as NOESY spectra. The intensity of the NOESY peaks are classified as strong, medium, and weak and define the NOE distance constraints that are used in structure calculation. However, NOE data can be distorted by spin-diffusion and internal motions within the molecule. Spin-diffusion causes an exaggeration of the nuclear relaxation effect resulting in observation of NOE cross peaks with abnormally strong intensity. The contribution of spin-diffusion can be measured by altering the NMR parameters of the NOESY experiments.

Short and medium range NOEs are dependent on the secondary structure. To define the tertiary structure it is necessary to observe and assign long range NOEs between nuclei on amino acids that are distant from each other in the primary sequence, but are brought into proximity by the fold of the protein.

In addition to NOE data, NMR experiments can be acquired to derive data for dihedral angles and hydrogen bond donors. These data are used as additional constraints to increase the precision of the final structure. Dihedral angles are calculated from J-coupling constants, which are a description of the

very fine splitting of NMR peaks caused by the effect of the local environment on non-equivalent nuclei that are 2, 3, or 4 bonds away. Nuclei that are hydrogen bond donors can be detected by deuterium exchange by following the exchange of labile protons over time.

2.1.4 Structure Refinement

The three-dimensional protein structure is calculated by analyzing the violations of distance constraints and the energy functions for each structure generated using all the structural constraints. The constraints are then modified to satisfy the violations in subsequent structure calculations. This is an iterative process requiring multiple rounds of structure calculation.

The final NMR structures should satisfy the energy conditions and have no violations outside the experimental error for all of the constraints used in their calculations. NMR structures are presented as ensembles of structures that represent the full range of structures that satisfy the NMR derived constraints and have reasonable molecular geometry. Ensemble representations are useful for determining regions of the protein that are well defined and those that are not well defined. These poorly defined areas of the structure are often the result of regions of the proteins that have very few NMR derived structural restraints to define them; typically these are in loops and the termini of the protein. The quality of the structure can be accessed by examining the root mean square deviation

(r.m.s.d.) of the ensemble for the backbone and sidechain atoms and by examining the Ramachandran plot to indicate whether the secondary structure adheres to the rules regarding the distance and orientation of atoms in proteins.

In this chapter, the NMR structure of the Ribonuclease P protein aRpp29 from *Archaeoglobus fulgidus* is presented. This chapter was adapted from Sidote and Hoffman, 2003. The results of the structure calculation, a description of the structure, and interesting features of the protein will be discussed.

2.2 Materials and Methods

2.2.1 Cloning, expression, and purification of aRpp29

The 102 amino acid archaeal homologue of the human RNase P protein Rpp29 (aRpp29) was identified by a homology search using BLAST. The aRpp29 gene was obtained by PCR from the genomic DNA of *Archaeoglobus fulgidus* cells (obtained from American Type Culture Collection, clone number 49558). PCR primers were designed to include unique *EcoRI* and *BamHI* restriction sites to allow the PCR product to be cloned into a plasmid encoding maltose binding protein (MBP) (pMAL-c2T (Kapust and Waugh, 1999), derived from pMAL-c2x, New England Biolabs) followed by a TEV protease site; this plasmid was transformed into BL21(DE3) cells (Novagen) supplemented with 100 mg/L ampicillin.

Six independent PCR reactions were performed, and representatives from each reaction were cloned into the plasmid pMAL-c2t for sequence verification by automated DNA sequencing (ICMB Core Facility). All clones contain a base substitution that results in a Glu at position 9 in the protein being mutated to an Ala. The lack of conservation of this amino acid in the primary sequence alignment of the protein suggests that this mutation is either unique to the species of *A. fulgidus* used in this study or a sequencing mistake exists in the genomic database.

The recombinant MPB-aRpp29 fusion protein was produced by growing typically 4 liters of *E. coli* cells in Luria broth at 37 °C until the cells reached an OD₆₀₀ of 0.5. The cells were then induced with 0.6 mM isopropyl-B-D-thiogalactopyranoside (IPTG) and allowed to grow for an additional 6 hours, 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% polyethylenimine (v/v). The cell lysate was centrifuged at 12,000 g for 20 minutes to remove cellular debris and nucleic acids. Cellular proteins were precipitated by adding ammonium sulfate up to 70% saturation, and separated by centrifugation at 12,000 g. The protein pellet was dissolved in 10 mM potassium phosphate buffer at pH 5.8, loaded to an SP sepharose column (Sigma), and eluted using a 0 to 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 per liter of cell culture.

The aRpp29 was cleaved from the MBP fusion protein by incubating with TEV protease at room temperature for 24 hours. The cleavage products were then loaded to an SP sepharose column, and the aRpp29 protein was eluted using a 0 to 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 per liter of cell culture. N-terminal sequencing and mass spectrometry were used to confirm the identity of the purified protein.

2.2.2 Preparation of Enriched Samples

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 (Table 2.1). A protein sample selectively labeled with ^{15}N lysine was prepared by growing the cells in M9 minimal media supplemented with 50 mg/l of ^{15}N labeled lysine and 100 mg/l of the other 19 unlabeled amino acids. A protein sample selectively labeled with ^{15}N glycine, serine and tryptophan was prepared using M9 minimal media supplemented with 100 mg/l of ^{15}N labeled glycine and 100 mg/l of each of 16 unlabeled amino acid types (all types except glycine,

Table 2.1 Reagents used for preparation of M9 minimal media for isotope labeling of *A. fulgidus* aRpp29 (per liter).

Na ₂ HPO ₄	8.0 g
KH ₂ PO ₄	3.0 g
NaCl	0.5 g
NH ₄ Cl	0.5 g
Glucose	3.0 g
MgSO ₄	2 mM
FeCl ₃	1 uM
CaCl ₂	0.1 mM
Ampicillin	100 mg
Trace Metal Solution	1 ml
Vitamin	1 ml

Trace metal solution (per liter) (Battiste *et al.*, 2000)

CoCl ₂	0.8 mg
CuSO ₄	0.7 mg
MoNa ₂ O ₄	2.0 mg
MnSO ₄	4.0 mg
ZnSO ₄	5.0 mg

Vitamin (per ml)

Thiamine	5.0 mg
Biotin	1.0 mg
Folic acid	1.0 mg
Niacinamide	1.0 mg
Panthothenate	1.0 mg
Riboflavin	1.0 mg

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 of the 1- ^{13}C -labeled amino acid.

2.2.3 Expression and purification of TEV protease

Tobacco etch virus protease was chosen for its ability to be purified 'in-house' and for its sequence specificity as it recognizes the seven amino acid consensus sequence, Glu-X-X-Tyr-X-Gln/Ser (Kapust *et al.*, 2002). Recombinant tobacco etch virus (TEV) protease was purified from *E. coli* cells. The expression plasmid containing the TEV protease gene was transformed in BL21 (DE3) expression cells (Novagen) supplemented with ampicillin.

Typically, 2 liters of cells were grown at 37°C until the OD_{600} reached ~ 0.5 . The cells were induced with IPTG to a final concentration of 1 mM and allowed to grow for an additional 4 hours at 30°C. The cells were harvested by centrifugation, resuspended in lysis buffer (50 mM potassium phosphate pH 8, 100 mM NaCl, 10% glycerol, and 25 mM imidazole.), and lysed by the addition of 10 mg/L of lysozyme, followed by incubation on ice for 30 minutes. The nucleic acids were precipitated by the addition of 0.5% polyethylenimine (v/v) and spun at 12,000 rpm for 15 minutes. The supernatant was applied to a Ni-NTA column pre-

equilibrated with lysis buffer and further washed with 5 volumes of lysis buffer. The TEV protease was eluted from the column using a 10 column volume gradient to 50 mM potassium phosphate pH 8, 100 mM NaCl, 10% glycerol, and 500 mM imidazole. The fractions were analyzed by SDS-PAGE and pooled. Following the addition of EDTA and DTT to 1 mM, the enzyme was concentrated to 1mg/ml and stored at -20°C.

2.2.4 NMR Spectroscopy

NMR spectra were recorded at 20 °C 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% H₂O/10% D₂O 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 3-dimensional HNCA, HNCACB and HN(CO)CACB spectra (Muhandiram and Kay, 1994), HNCO spectra (Grzesiek and Bax, 1992) and HACACBCO spectra (Kay, 1993); together these spectra correlate the backbone protons to the N, C α , C ω and C β signals of the same and adjacent amino acid residues. Side chain resonance assignments were obtained by analyzing 3-dimensional ¹⁵N-edited HMQC-

TOCSY and ^{13}C -edited HCCH-TOCSY spectra (Wang *et al.*, 1995), and 2-dimensional homonuclear 2QF-COSY and TOCSY spectra (Table 2.2).

NOE cross peaks were detected using 2-dimensional ^1H - ^1H NOESY, 3-dimensional ^{15}N - ^1H - ^1H HSQC-NOESY, and 3-dimensional ^{13}C - ^1H - ^1H HSQC-NOESY (Pascal *et al.*, 1994) spectra. The NOE mixing time was 60 msec for spectra used to derive distance constraints. A ^{13}C -edited HSQC-NOESY spectrum (mixing time 80 msec) 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 . (Figure 2.4, 2.5)

Data were processed using either NMR-Pipe (Delaglio *et al.*, 1995) or Felix (Hare Research). ^1H , ^{15}N and ^{13}C chemical shifts were referenced as

Table 2.2. Summary of NMR spectra collected for resonance assignment and structure calculation.

<u>Spectrum</u>	<u>Sample (pH 5.8, 30°C)</u>
2D TOCSY	Unlabeled, 90% H ₂ O, 10% D ₂ O
2D TOCSY	Unlabeled, 90% H ₂ O, 10% D ₂ O, pH 3.0
2D TOCSY	Unlabeled, 99% D ₂ O
2D 2QF-COSY	Unlabeled, 90% H ₂ O, 10% D ₂ O
2D 2QF-COSY	Unlabeled, 99% D ₂ O
2D NOESY	Unlabeled, 90% H ₂ O, 10% D ₂ O
2D NOESY	Unlabeled, 90% H ₂ O, 10% D ₂ O, pH 3.0
2D NOESY	Unlabeled, 99% D ₂ O
2D ¹⁵ N- ¹ H HSQC	Uniform ¹⁵ N labeled
2D ¹⁵ N- ¹ H HSQC	¹⁵ N lysine labeled
2D ¹⁵ N- ¹ H HSQC	¹⁵ N glycine, serine, tryptophan labeled
3D ¹⁵ N-edited HSQC-TOCSY*	Uniform ¹⁵ N labeled
3D ¹⁵ N-edited HSQC-NOESY*	Uniform ¹⁵ N labeled, mix=60ms
3D ¹³ C-edited HCCH-TOCSY	Uniform ¹³ C and ¹⁵ N labeled
3D ¹³ C-edited HSQC-NOESY	Uniform ¹³ C and ¹⁵ N labeled, mix=80ms
3D HNCA	Uniform ¹³ C and ¹⁵ N labeled
3D HNCA	Uniform ¹³ C and ¹⁵ N labeled, pH 3.0
3D HNCACB	Uniform ¹³ C and ¹⁵ N labeled
3D HNCACB	Uniform ¹³ C and ¹⁵ N labeled, pH 3.0
3D HNCO	Uniform ¹³ C and ¹⁵ N labeled
3D HN(CO)CACB	Uniform ¹³ C and ¹⁵ N labeled
3D HN(CO)CACB	Uniform ¹³ C and ¹⁵ N labeled, pH 3.0
3D HACACBCO	Uniform ¹³ C and ¹⁵ N labeled
3D HNCO	Uniform ¹⁵ N labeled, 1- ¹³ C leucine
3D HNCO	Uniform ¹⁵ N labeled, 1- ¹³ C isoleucine
3D HNCO	Uniform ¹⁵ N labeled, 1- ¹³ C valine
3D HNCO	Uniform ¹⁵ N labeled, 1- ¹³ C phenylalanine

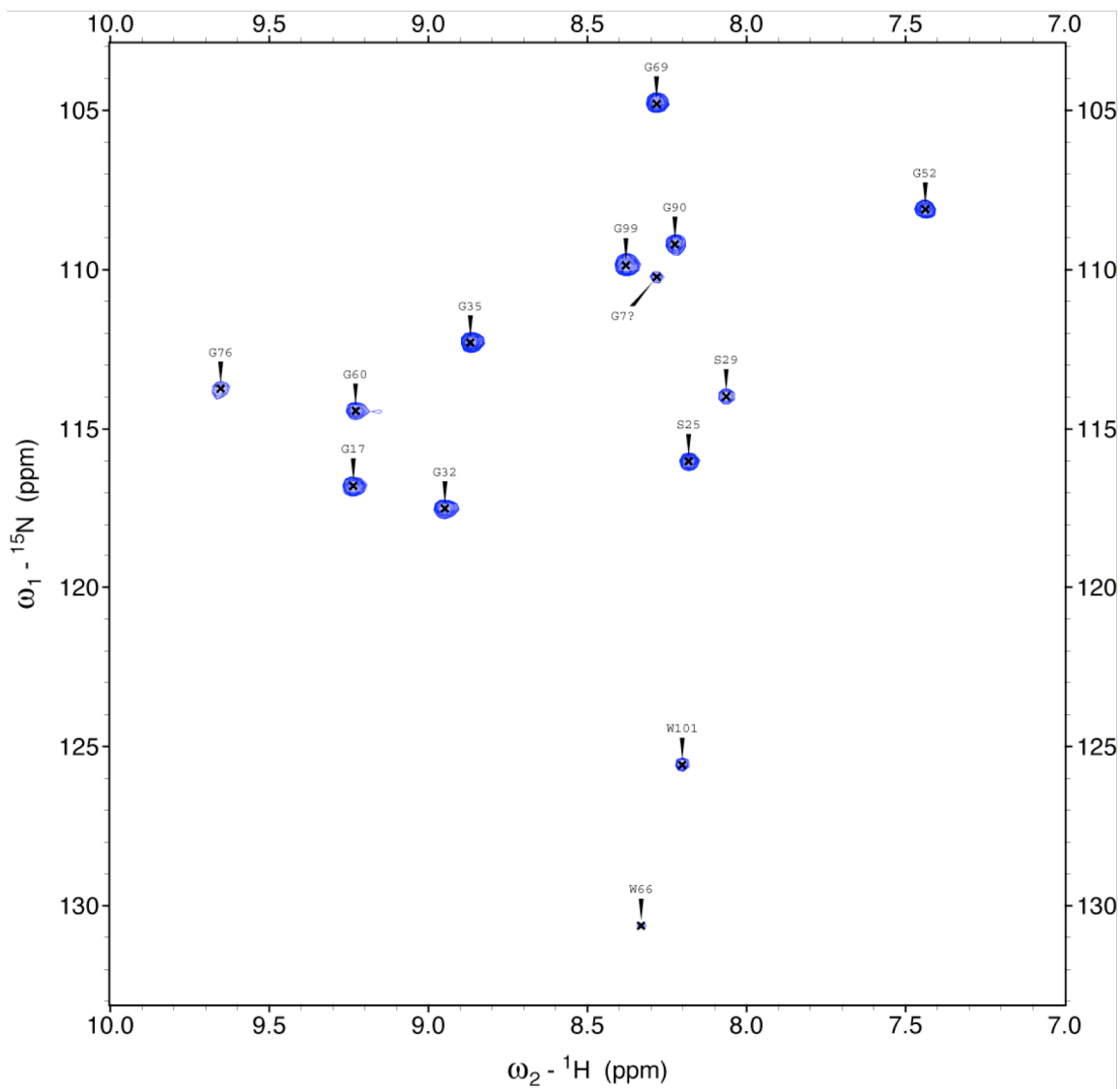


Figure 2.4. ${}^{15}\text{N}$ - ${}^1\text{H}$ - HSQC spectrum of aRpp29 selectively labeled with ${}^{15}\text{N}$ – gly/trp/ser. The spectrum was obtained using a sample (1.8 mM) in 10 mM potassium phosphate pH 5.8 with 100 mM NaCl at 30°C. Peaks are observed only from ${}^{15}\text{N}$ labeled glycine, tryptophan, and serine.

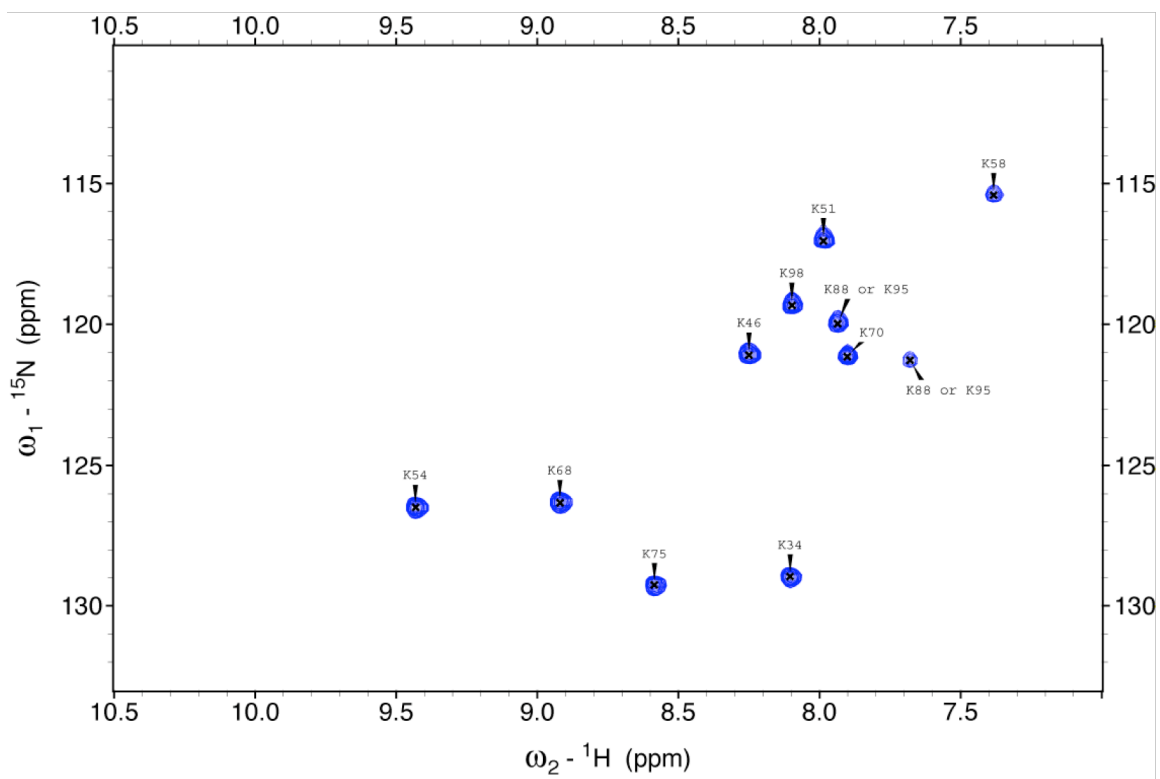


Figure 2.5. ^{15}N - ^1H - HSQC spectrum of aRpp29 selectively labeled with ^{15}N -lys. The spectrum was collected on a sample (1.8 mM) in 10 mM potassium phosphate pH 5.8 with 100 mM NaCl at 30°C. Peaks are observed only from labeled lysines. Two of the peaks could not be unambiguously assigned.

recommended (Wishart *et al.*, 1995), 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 ^1H 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).

2.2.5 NMR Dynamics

^{15}N - ^1H 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 (Farrow *et al.*, 1994). The ^{15}N - ^1H heteronuclear NOE was measured by comparing spectra acquired with either a 5 sec delay between each free induction decay or a 2 sec delay followed by a 3 sec series of 120° nonselective ^1H pulses. For T_1 relaxation measurements, 2-D spectra with relaxation delays of 10, 260, 510, 760 and 1010 msec were obtained; for T_2 relaxation measurements 2-D spectra with relaxation delays of 29, 58, 87, 116 and 145 msec were acquired; in each case the relaxation delay between the acquisition of each free induction decay was 3 sec. T_1 and T_2 relaxation times were determined from the slope of plots of the logarithm of peak height versus relaxation delay.

2.2.6 Structure Calculation

Structure calculations were performed using the restrained simulated annealing protocol in the program CNS version 1.1 (Brünger *et al.*, 1998), 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 msec (in the case of homonuclear 2-D and ^{15}N -resolved 3-D spectra) to 80 msec (in the case of the ^{13}C -resolved 3-D NOE spectrum) to minimize the effects of spin diffusion.

Based on the cross peak intensity in the homonuclear 60 msec NOE spectra, distance restraints were classified as strong ($< 2.8 \text{ \AA}$), medium ($< 3.2 \text{ \AA}$), weak ($< 3.8 \text{ \AA}$) and very weak ($< 4.2 \text{ \AA}$); these distance bounds were calibrated by using inter-proton 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 msec mixing time, respectively) and were assigned to distance restraints as strong ($< 5.0 \text{ \AA}$), medium ($< 5.5 \text{ \AA}$), weak ($< 6.0 \text{ \AA}$), and very weak ($< 6.5 \text{ \AA}$). Due to the possibility of observing the effects of spin diffusion in long mixing time (200 msec)

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 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 delta and epsilon protons on tyrosine rings, distances were measured from the center of the two delta protons (or epsilon protons), and 2.4 Å was added to the interproton distance.

Backbone dihedral angle restraints of $\psi = 150 \pm 25$ and $\phi = -120 \pm 25$ degrees 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 2.3.

Table 2.3. Summary of the refinement and structural statistics for the NMR structure of *A. fulgidus* aRpp29.

Intraresidue NOEs	215
Sequential NOEs (residue <i>i</i> to <i>i</i> + 1)	178
Medium-range NOEs (residue <i>i</i> to <i>i</i> + 2, 3, 4)	18
Long-range NOEs	143
Dihedral angle restraints	70
Hydrogen bond restraints	27
Total structural restraints	651
Number of unique starting structures for simulated annealing	10
Number of simulated annealing runs, differing in initial trajectories	200
R. m. s. d. for backbone atoms (residues 17-77)	0.87 Å
R. m. s. d. for side chain atoms (residues 17-77)	1.78 Å
Average number of NOE violations > 0.2 Å (per structure)	3.17
Average number of NOE violations > 0.5 Å (per structure)	0
Residues in the most favored regions of the Ramachandran plot	71.2 %
Residues in the additional allowed regions of the Ramachandran plot	21.2 %
Residues in generously allowed regions of the Ramachandran plot	5.8 %
Residues in the disallowed regions of the Ramachandran plot	1.9 %
R. m. s. d. for covalent bonds	0.0034 ± 0.0001
R. m. s. d. for covalent angles	0.511 ± 0.015
R. m. s. d. for improper angles	0.581 ± 0.016

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 (Laskowski *et al.*, 1996), with statistics reported in Table 2.2. 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 (Holm and Sander, 1993). The coordinates for aRpp29 have been submitted to the Protein Data Bank and assigned PDB code 1PC0. The resonance assignments, NOE assignments, dihedral restraints, and hydrogen bond restraints are listed in appendices A, B, C, D, respectively.

2.3 Results and Discussion

2.3.1 Initial Sample Characterization

The purified *Archaeoglobus fulgidus* aRpp29 protein was found to be soluble, stable, and suitable for structural analysis by NMR methods. (Figure 2.6). About 68 of 100 possible correlation peaks are observed in a ^{15}N - ^1H HSMQC spectrum (Zuiderweg, 1990) obtained at pH 5.8 and 30 °C using presaturation for solvent suppression (Figure 2.7); these peaks are well dispersed as is typical for a folded protein. When a non-saturating method of solvent suppression is used, additional ^{15}N - ^1H correlation peaks 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 indicated that residues 1-3 are absent from the structure of the purified protein.

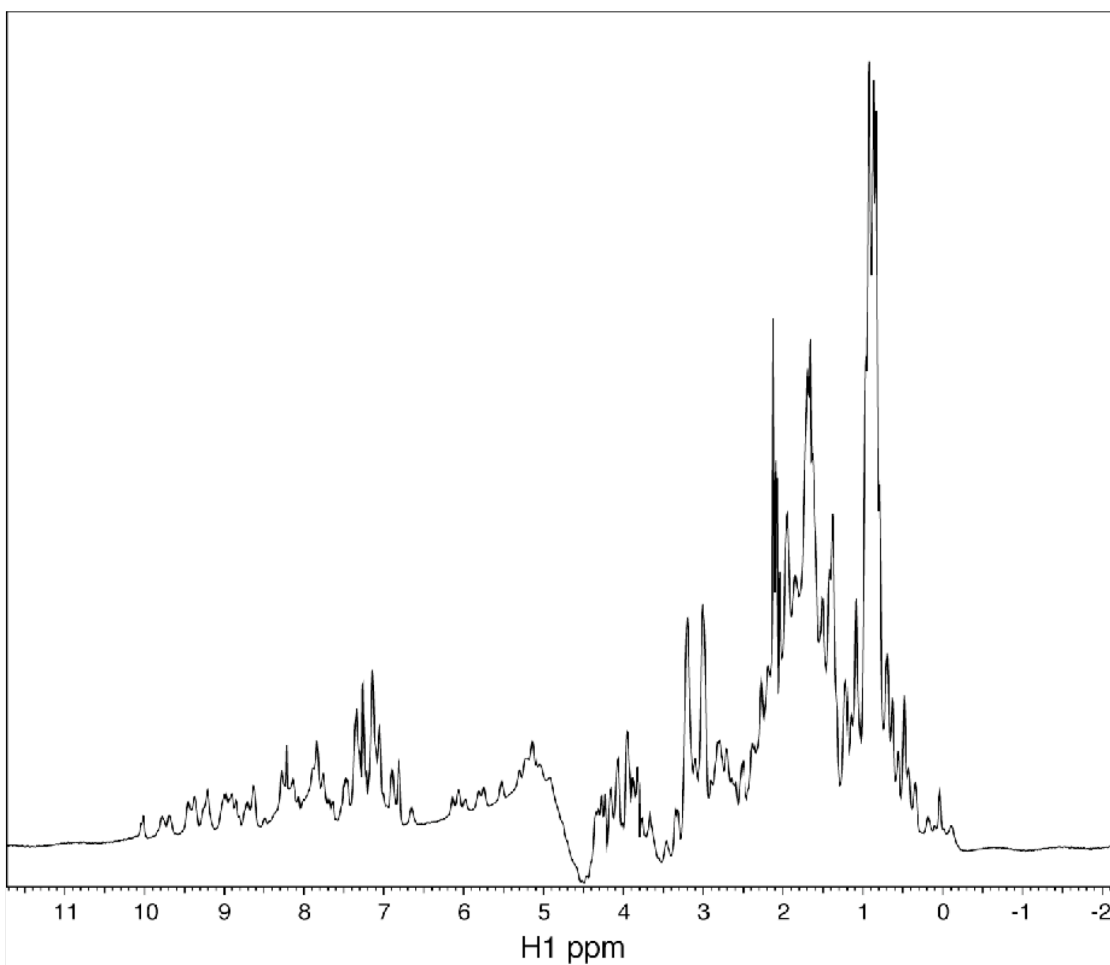


Figure 2.6. One dimensional spectrum of RNase P protein aRpp29 from *A. fulgidus*. The sample contains 1.8 mM protein in potassium phosphate buffffer consisting of 90% H₂O/10% D₂O at pH 5.8. The spectrum was acquired at 30°C. The wide dispersed peaks far from random coil values indicate the protein is folded and suitable for structure determination by NMR. Peaks between 5.0 and 6.0 ppm indicate the presence of b strands in the structure.

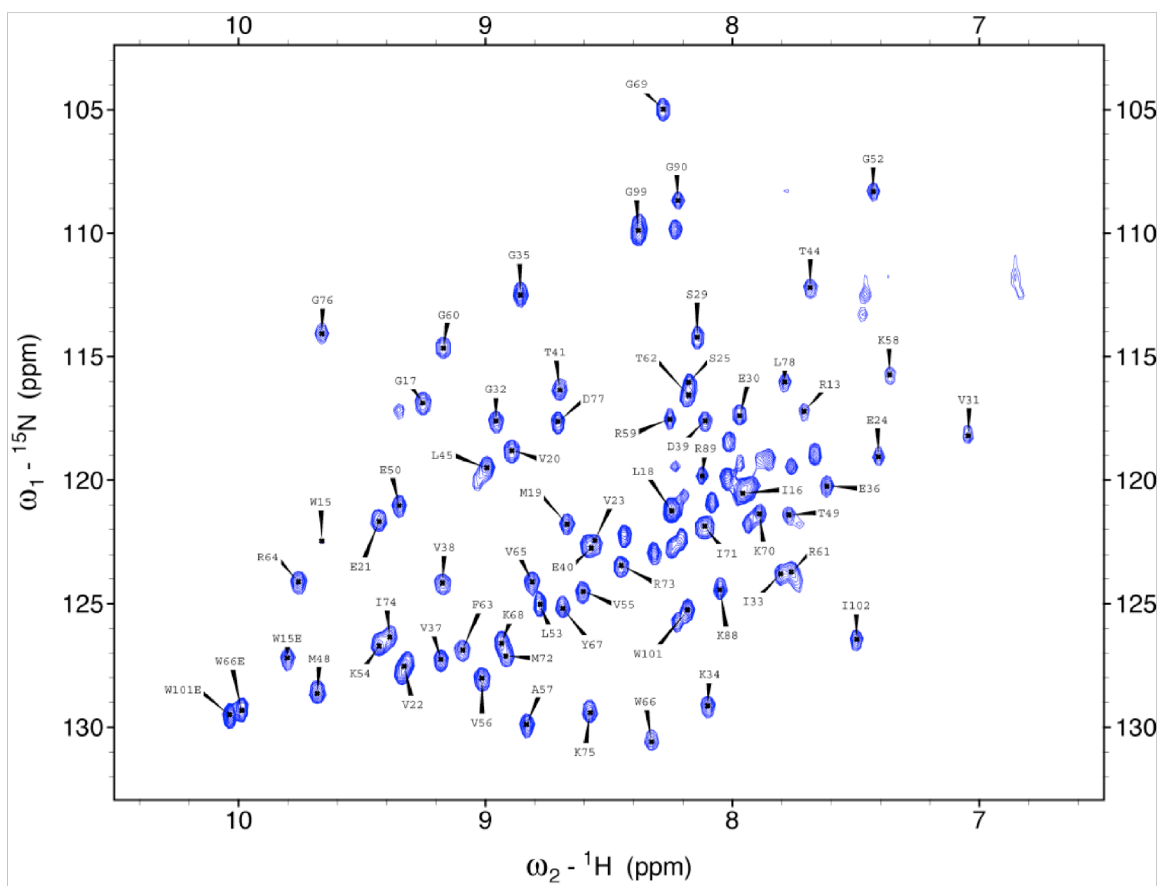


Figure 2.7. ^{15}N - ^1H correlated HSMQC spectrum of the aRpp29 protein from *Archaeoglobus fulgidus*. The spectrum was collected at 30 °C and pH 5.8, using presaturation for H_2O suppression, so that the amide protons that exchange relatively rapidly with the solvent are attenuated. Assignments for the best resolved cross peaks are labeled. Resonance assignments for 66 of the clearly resolved peaks are labeled. The remaining unlabeled peaks could not be unambiguously assigned due to fast exchange or spectral overlap.

2.3.2 Description of the aRpp29 structure

Structural results derived from the NMR data show that the dominant feature of the protein is an antiparallel sheet of six β -strands (Figure 2.8), formed by residues 18 through 76. Within the β -sheet, strands β 1 and β 2 are connected by a short loop, as are strands β 2 and β 3, while strands β 3 and β 4 are connected by a compact turn, as are strands β 5 and β 6. An interesting feature of the β -sheet is a loop consisting of residues 57-61; this five residue loop connects strands β 4 and β 5, which are on opposite sides of the sheet as it is represented in two dimensions resulting in strands β 5 and β 6 being folded onto the first four strands to form a sandwich-like structure (Figure 2.8, 2.9).

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 2.10), 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

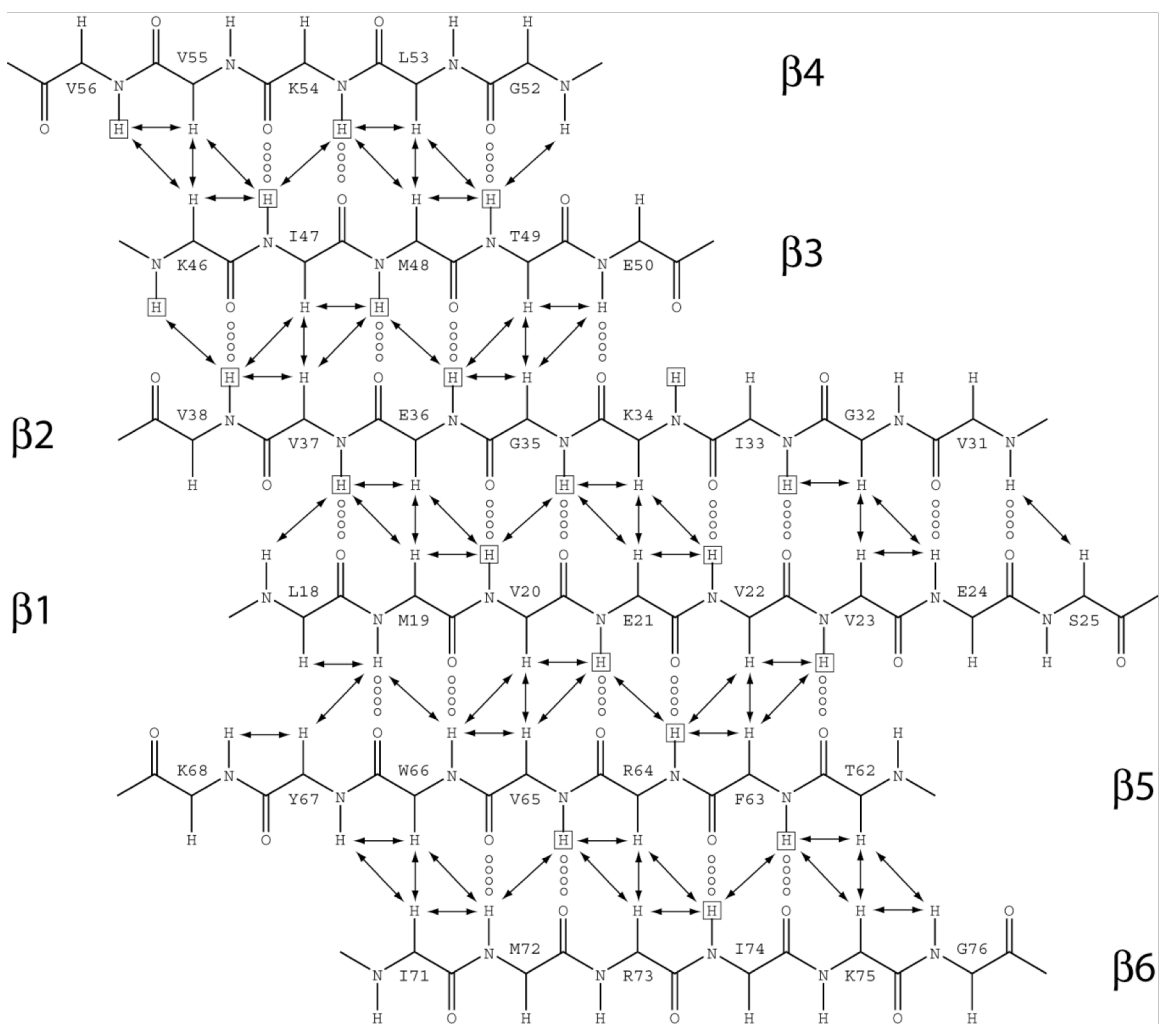


Figure 2.8. Diagram of the six-stranded β -sheet structure of aRpp29 protein derived from NOE crosspeaks. Pairs of protons for which unambiguous NOEs are observed are connected by arrows, and inter-strand hydrogen bonds are indicated by dotted lines. Amide protons that require 12 hours or more to exchange with deuterated solvent are boxed.

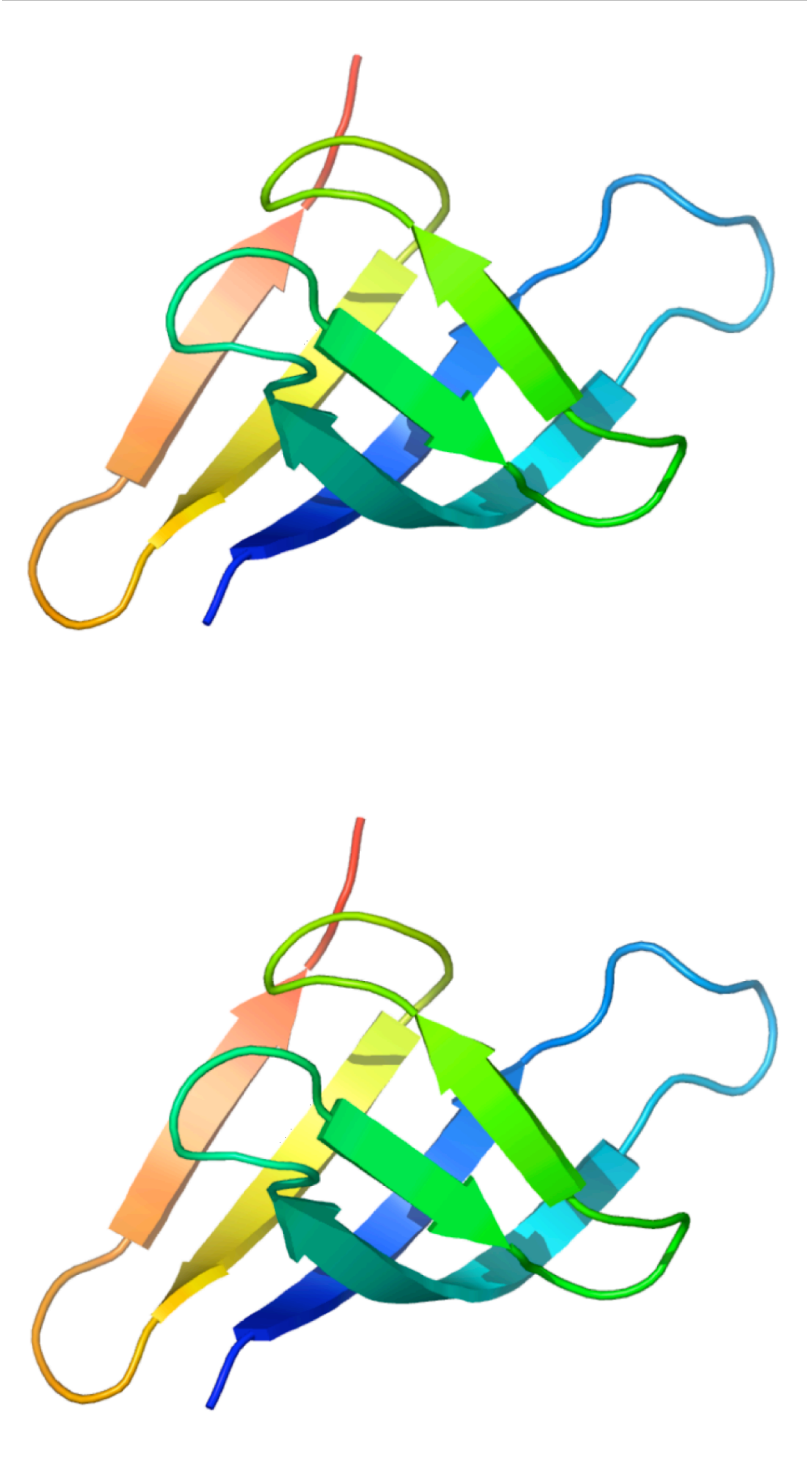
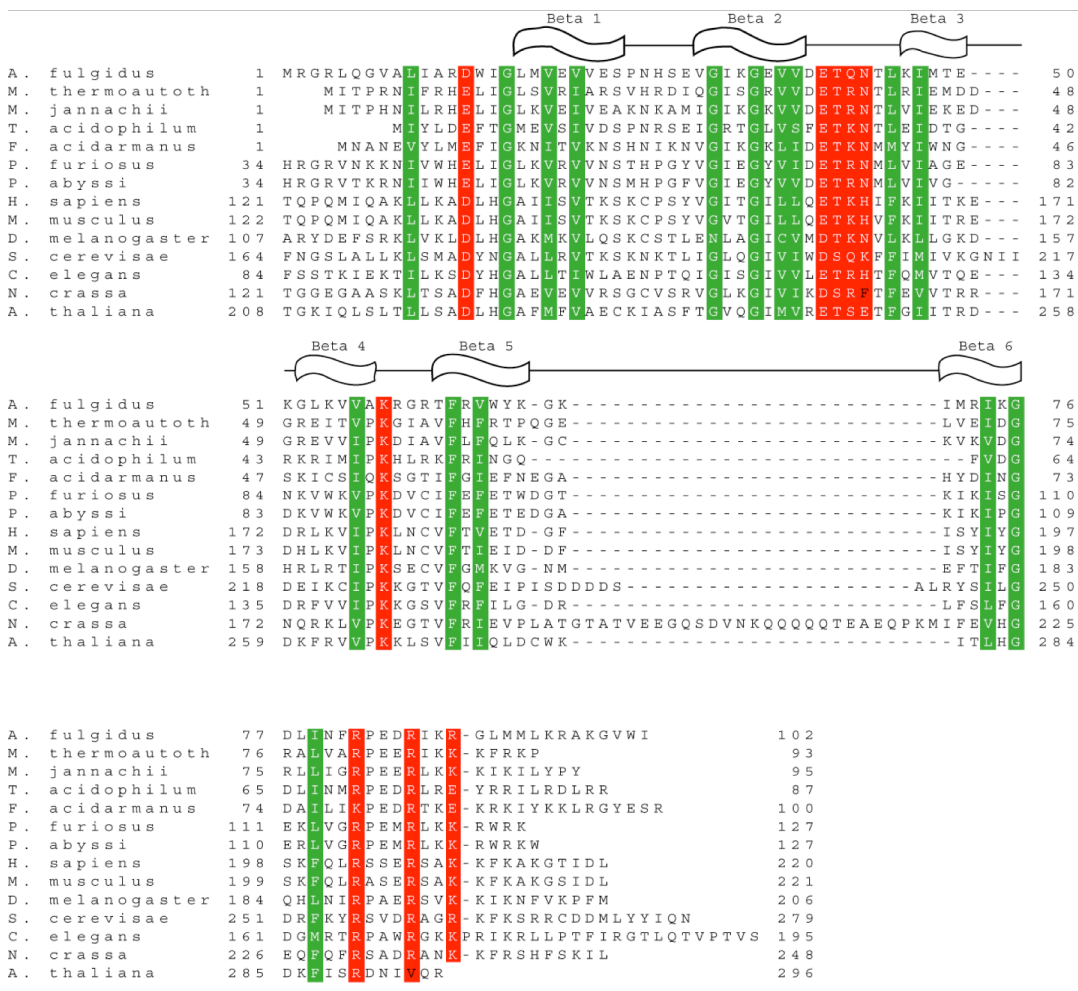


Figure 2.9. Stereo view of a cartoon representation of aRpp29. Shown are residues 17 – 77 color ramped from blue at the N-terminus to red at the C-terminus.



2.10. Amino acid sequence alignment of *A. fulgidus* aRpp29 and RNase P proteins from other archaea and eukarya. Sequences from seven archaeal species are shown: *M. thermoautotrophicus*, *M. jannaschii*, *T. acidophilum*, *T. acidarmanus*, *P. furiosus*, and *P. abyssi*. The eukaryotic sequences are from *H. sapiens*, *M. musculus*, *D. melanogaster*, *S. cerevisiae*, *C. elegans*, *N. crassa*, and *A. thaliana*. Archaeal aRpp29 is homologous to the C-terminal half of eukaryotic Rpp29. The most conserved hydrophobic residues, which form the hydrophobic core of aRpp29, are boxed in green. Conserved surface residues that are most likely to be involved in mediating protein-RNA or protein-protein interactions within RNase P are boxed in red.

equally well, the r.m.s.d. is less than 1 Å for the backbone atoms within the β -strands (Figure 2.11, 2.12). The loops that connect strands b1-b2 and b2-b3 are the least well defined of the connecting structures; these loops are primarily defined by sequential and intraresidue NOE-derived restraints (Figure 2.12)

When the amino acid sequence of aRpp29 is aligned with that of homologous proteins from several species (Figure 2.10) 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, 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

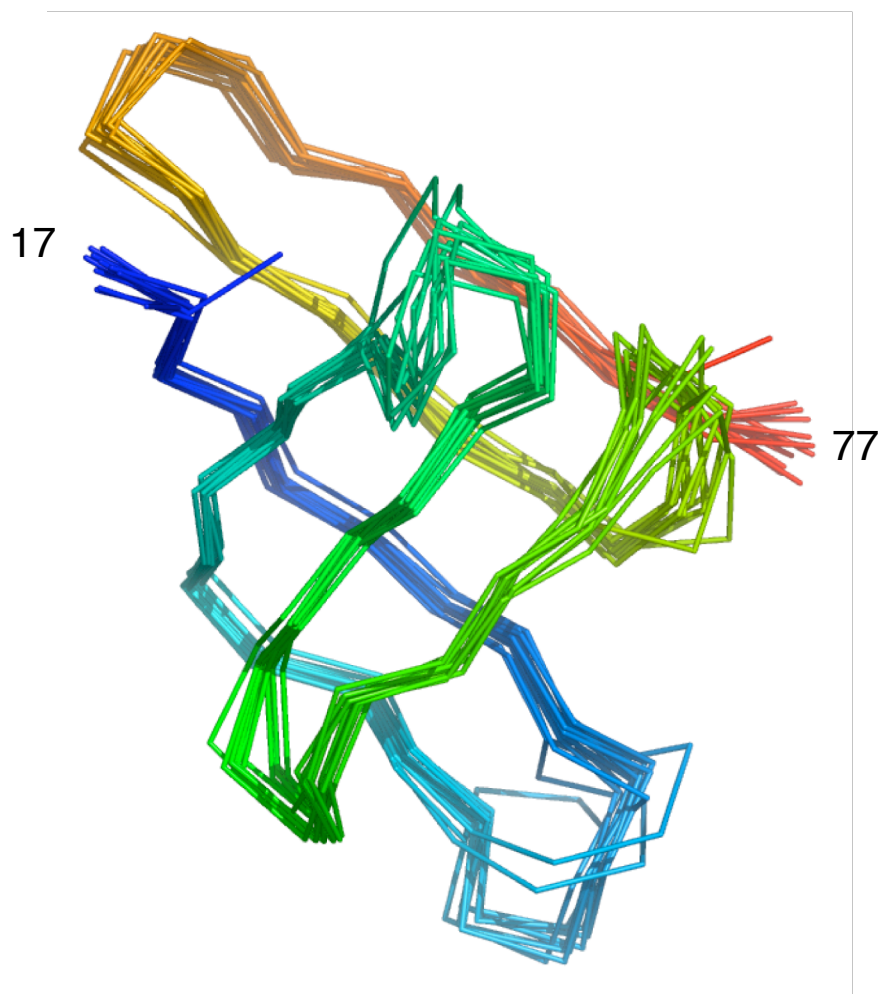


Figure 2.11. Superposition of the backbones of 12 low energy structures of aRpp29. The structures are equally consistent with the NMR data and have the minimum value of the CNS energy function, color-ramped from red at the N-terminal to blue at the C-terminal of the protein. The 12 models are a fair 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 through 77; the backbone is shown for residues 17 through 77 of the protein.

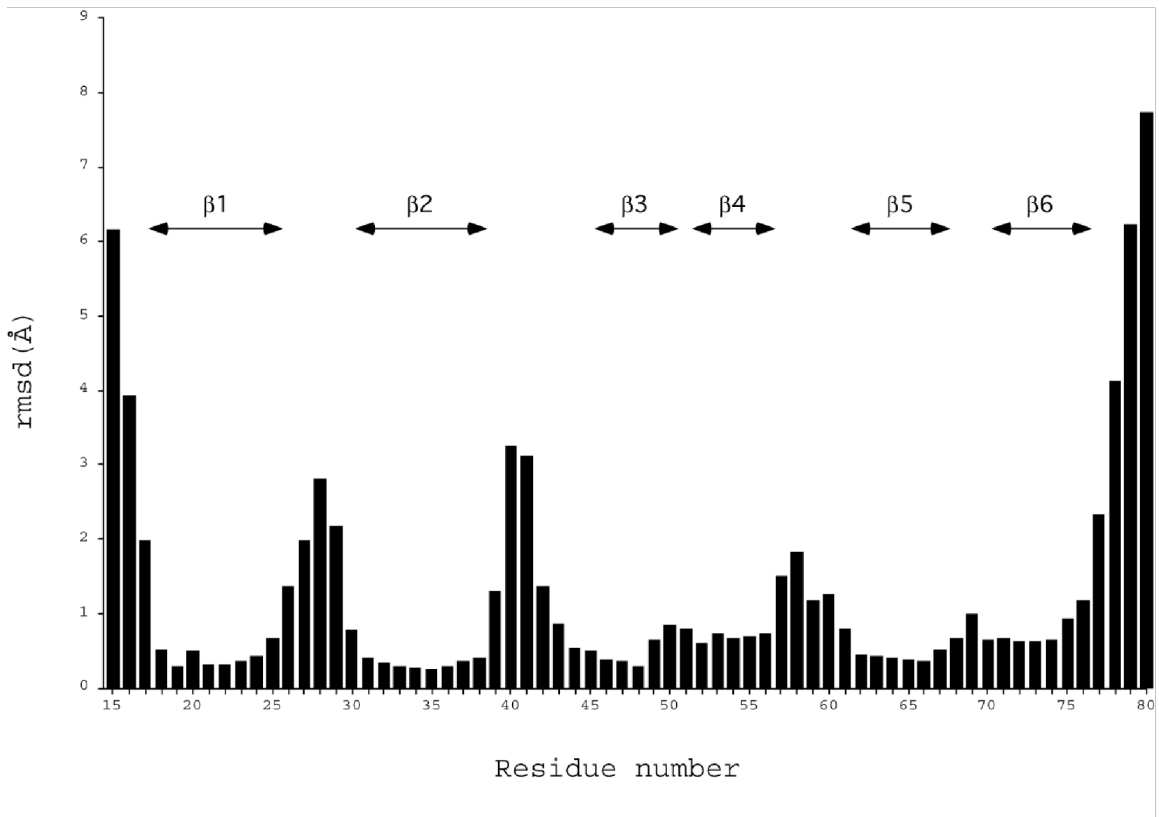


Figure 2.12. Plot of the root mean square deviation for the coordinates of the backbone heavy atoms for the aRpp29. R.m.s.d. is plotted versus residue number, calculated using 12 structures that satisfy the NMR-derived structural constraints, and are a fair representation of the full range of structures that are consistent with the NMR data. The r.m.s.d. 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 through 77. The figure shows that the b-strands are quite well-defined by the NMR data (r.m.s.d. < 1 Å), the loops connecting the strands are moderately well-defined (r.m.s.d. 1 to 3.5 Å), while the positions of the residues nearer to the N- and C-termini are not well determined (r.m.s.d. > 3 Å).

hydrogen bond donor (T41, Q42, N43) are conserved (Figure 2.10). 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.

2.3.3 Dynamics of aRpp29 in solution

Most of the hydrogen bonded amide protons within the β -sheet remain unexchanged with deuterated solvent after 12 hours 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 msec, respectively, and the ^{15}N - ^1H 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 nsec; this rotational time is appropriate for a monomeric protein with a molecular weight 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.

2.3.4 The termini of aRpp29 are flexible

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 - ^1H 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 - ^1H 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 (Wishart and Sykes, 1994) values for residues 89 through 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. 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.

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 L2, L3, L15, L21e and L37e (Roll-Mecak *et al.*, 2000, Ban *et al.*, 2000), which contain tails of extended structure at the N- and/or C-terminus of an otherwise globular domain (Figure 2.13). 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.

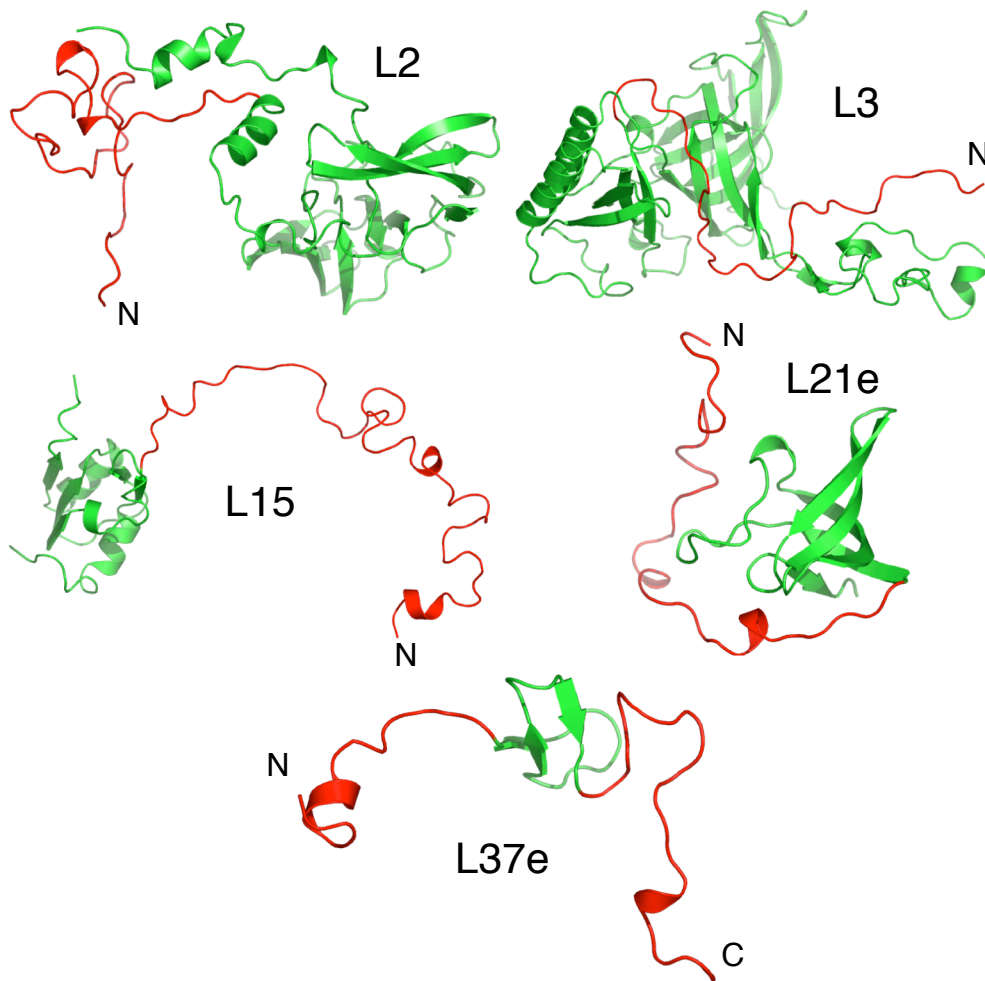


Figure 2.13. Backbone structures of protein subunits containing regions lacking regular secondary structure from the archaeal large ribosomal subunit. Shown are the structures of ribosomal proteins L2, L3, L15, L21e, and L37e from the X-ray crystal structure of the *H. marismortui* large ribosomal subunit (PDB code 1S72, Ban *et al.*, 2000). Globular domains are colored in green and regions without regular secondary structure are colored in red. Regions colored red are involved in anchoring the proteins to the ribosomal RNA and by analogy may serve a similar function in the case of aRPP29 and the RNase P RNA.

2.3.5 Data Collection at pH 3 and 20°C

Triple resonance and ^{15}N -resolved 3-D NOE and 3-D TOCSY, as well as homonuclear 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.

Chapter 3

X-ray crystal structure of the RNase P protein aRpp29 from *A. fulgidus*

3.1 Introduction

X-ray crystallography and NMR are complementary methods of investigating the three-dimensional structures of biological macromolecules. X-ray crystallography was the first method available to provide detailed atomic resolution structures of proteins. Like NMR, it is not an imaging technique, but rather exploits the diffraction of X-rays by the distributions of electrons within a crystal lattice.

The process of determining the atomic structure of a biological macromolecule by X-ray crystallography is a multistep process. Outlined below are the steps necessary to solve the structure of a biological molecule using X-ray crystallography from initial crystal growth to final model validation. This is presented merely to give the reader a sense of the process and is not meant to provide a comprehensive explanation. For a detailed treatment of protein crystallography refer to texts by Drenth, 1999 and Blundell, 1976.

3.1.1 Crystallization of Proteins

Crystallizing proteins for structure analysis typically requires large amounts of pure, homogeneous, soluble material. Obtaining crystals that are suitable for structure determination can be an exhausting and time consuming process. The principle of crystallization is to gently induce the protein to come out of solution and arrange in a well-ordered lattice, forming crystals as large as possible. If the process happens too fast, amorphous precipitation will likely occur.

The procedure for determining the correct conditions for crystallization is mainly by trial and error. Solvent conditions that are typically varied include, pH buffer, salt, precipitant, and temperature. Additives such as organic solvents and detergents may also be used. The idea is to screen as many different conditions as possible in order to grow large, well ordered crystals. The method used to grow the crystal, such as vapor diffusion, sitting drop, or dialysis can also affect crystal growth. Crystals can take anywhere from hours to months to grow to full size and the time may be dependent on changes the protein must undergo, such as degradation, before the conditions are appropriate for crystal formation. Crystals vary considerably in morphology. They can assume geometric shapes, needles, or clusters. The crystals should at least > 0.2 mm in the longest dimension for adequate diffraction on a home source X-ray generator/detector.

The availability of high-energy synchrotron X-rays allows the use of smaller crystals to obtain adequate diffraction.

3.1.2 Symmetry and Space groups

When proteins crystallize, they arrange in specific three-dimensional orientations caused by regular packing and lowest free energy, forming special relationships of symmetry with each other. There are 3 vectors, a , b , and c and the angles between them, α , β , γ that define the three dimensions of the unit cell. The unit cell is the regular repeating unit that defines the crystal. If the crystal lattice has symmetry higher than triclinic, the unit cell will contain internal symmetry, and in the case of biological macromolecules, the molecular structure will be repeated a number of times due to symmetry operations. This is known as the asymmetric unit. The number of molecules in the unit cells does not have to be the same as the number of asymmetric units, as there can be more than one molecule per asymmetric unit. There are 65 possible ways of combining symmetry operations in a crystal of a chiral molecule such as a protein.

3.1.3 Crystal Characteristics

The ability of a crystal to diffract X-rays to a high resolution depends on the ordering of the molecules in the unit cell. However, the positions of the molecules are not fixed due to static disorder and thermal vibrations. Most

crystals contain defects in the repetition of the unit cells, known as mosaicity. This causes each unit cell to contribute a slightly different diffraction pattern to the overall diffraction of the crystal. Crystals of good quality will have slight mosaicity, which can be accounted for during data processing.

3.1.4 Data Collection

Crystals are placed in the X-ray beam between the X-ray source and the detector. Generally, they are flash cooled in liquid nitrogen before being placed in the X-ray beam. This step is useful to reduce the effects of radiation damage and increase resolution by reducing thermal vibrations and disorder. The crystal is placed in a nylon loop or capillary tube (if cryo-cooling is not used) and positioned in the beam in a known orientation, preferably aligned along one axis of the crystal. Experimental conditions, such as exposure time, distance from the crystal to the detector, and degrees of oscillation are determined empirically. The crystal is rotated a fixed number of degrees to collect a complete set of data, meaning that each unique reflection is collected at least once. The number of degrees the crystals needs to be rotated depends on the space group of the crystal, as crystals of higher symmetry require fewer degrees of oscillation to collect a complete dataset.

3.1.5 Data Processing

Protein crystals diffract X-rays weakly because they are composed of mainly light atoms and have large unit cells. The intensity of the diffracted spots can vary due to the amplitude of the wave and their phase. The accuracy of the intensity measurements is important for successful structure determination, especially in cases where small anomalous differences are being measured, as in the case of single anomalous diffraction (SAD) (Borek et al., 2003, Gonzalez, 2003). Once the diffraction data has been collected, the intensities of the reflections (“spots”) are measured by integration and then scaled against the rest of the reflections in the dataset so they can be related. Careful monitoring of the scaling statistics can be useful for rejecting bad spots or complete diffraction images and can be used to determine the overall quality of the dataset. At this point a number of useful parameters can be determined such as unit cell dimensions, space group, solvent content of the crystal, and number of molecules per asymmetric unit.

3.1.6 The Phase problem

Diffraction data provides information pertaining to the intensities of the waves reflected from planes in the crystal (referred to as hkl). The amplitude of the wave $|F_{hkl}|$ is proportional to the square root of the intensity (I_{hkl}). In order to

calculate the electron density of a point (xyz) in the crystal requires the following summation over all the hkl planes:

$$\rho(xyz) = 1/V \sum |F_{hkl}| \exp(i\alpha_{hkl}) \exp(2\pi i(hx + ky + lz))$$

where V is the volume of the unit cell and α_{hkl} is the phase corresponding to the structure factor amplitude F_{hkl} . At this point we can measure the intensities, but the phases are lost. Determining the value for α_{hkl} is referred to as the “phase problem”. The only relationship between the intensities and the phases is through the electron density (Taylor, 2003).

There are a number of methods that can be used to derive the phase. A brief description of each is provided below.

Isomorphous replacement

This technique requires crystals to be soaked in a heavy atom salt solution to form isomorphous heavy atom derivatives. The diffraction patterns of the native and the derivative crystal are then compared and the measurable differences in intensity due to the heavy atoms are used to calculate their position. Once the positions of the heavy atoms are known they can be used to phase the protein. The main problem that arises with this method is nonisomorphism. The heavy atom salts can cause rearrangements in the crystal lattice, leading to changes in unit cell dimensions, poor diffraction, or no

diffraction. Heavy atom soaking can also change the orientation of the protein as well as cause conformational changes.

In most cases more than one heavy atom derivative (multiple isomorphous replacement, MIR) is needed to derive the exact phase, but a single derivative (single isomorphous replacement, SIR) can also be used if it provides a sufficiently strong difference in intensity. The amplitudes of a reflection are measured for the native crystal, $|F_p|$, and for the derivative crystal, $|F_{PH}|$. The isomorphous difference can be calculated by $|F_H| = |F_{PH}| - |F_p|$ and can be used to determine the heavy atom positions (Drenth, 1999). The heavy atom positions can be refined and used to calculate a more accurate $|F_H|$ and its corresponding phase, α_{hkl} .

The phase calculation leads to two possible solutions of the heavy atom phase. Examining the phase probability distributions can break the phase ambiguity. The phase distribution probabilities are calculated by taking into account errors in the measurement of F_{PH} and applying maximum-likelihood methods to arrive at the correct solution. Alternatively, electron density maps can be calculated for each possible phase following density modification and the correct phase determined by visual inspection of the map. The use of multiple heavy atom derivatives or the presence of anomalous scattering from the single derivative can be used to break the phase ambiguity (Drenth, 1999).

The phases can be further improved using solvent flattening, histogram matching and non-crystallographic symmetry averaging. Solvent flattening alters the electron density by setting regions of solvent to a lower density than that of the protein. The principle of this method is to recognize interpretable features in the electron density map and use them to generate a solvent mask around the protein molecule. The bulk solvent region is recognized as low density and the protein as regions of higher density. This is then used to generate a new set of phases that are then combined with the experimental phases and repeated. This technique is extremely useful as long as the solvent boundary is correctly identified and the errors in the phases are accounted for. Histogram matching changes values of the electron density points to be consistent with the expected distribution of electron density values. Noncrystallographic symmetry averaging forces the electron density values to be equal for each molecule in an asymmetric unit when more than one molecule is present.

Molecular Replacement

Molecular replacement is the most rapid way to solve an unknown structure. This technique is useful when the structure of a homologous molecule has been solved. Generally, the sequence identity needs to be at least 25 % and the r.m.s.d. should be less than 2 Å between the C α atoms of the model and the unknown structure. Molecular replacement can be thought of as crystallography

in reverse: the structure factors and phases from the model are used to calculate new structure factors and phases for the unknown molecule. Before the phases can be applied the search model must be positioned in the same way as the unknown molecule. The placement of the model is performed using rotational and translational searches to determine the position accurately enough to provide an interpretable electron density map.

Direct Methods

Direct methods are based on the atomicity and positivity of electron density that leads to phase relationships between the structure factors. This is the least useful technique for biological macromolecular crystallography as it requires very high resolution data ($< 1.2\text{\AA}$) and has successfully been used on small molecules (< 1000 atoms). In the case of larger molecules, this method has been implemented in many phasing software packages to determine the heavy atom substructure, which can then be used in phase calculation.

Anomalous scattering

In most cases, the electrons scatter as if they are “free”. For atoms with high atomic numbers, the electrons do not behave this way. These atoms absorb X-rays at specific wavelengths at and near their absorption edges, causing excited electrons to jump from a lower energy level to a higher energy level,

which leads to anomalous scattering. The absorption edges for lighter atoms such as C, N, and O are not near the X-ray wavelengths used in protein crystallography, so these atoms do not contribute to the anomalous scattering.

When anomalous scattering is present, the atomic scattering factor becomes:

$$f = f_0 + f' + if''$$

where f_0 is the normal scattering factor, and f' and f'' are the real and imaginary components of the anomalous scattering factors (Blundell, 1976)(Figure 3.1A). The anomalous scattering factors depend on the wavelength but do not change with resolution. The value of f' can be positive or negative. The imaginary component, f'' , is 90° ahead in phase in relation to f_0 and is proportional to the absorption of the X-rays and to their fluorescence.

The presence of anomalous components in the atomic scattering factors leads to the breakdown of Friedel's law. This law states that symmetry related reflections will be equal in intensity and their phases are complementary, that is, $F_{\alpha|hk|l} = -F_{\alpha|-h|-k|-l}$. This is no longer the case when the when anomalous scattering is present. The anomalous or Bijvoet difference can be used the same way isomorphous differences are used to determine the positions of the anomalous scattering atoms (Blundell, 1976). The phases for the structure factors can then be calculated as in the SIR/MIR case.

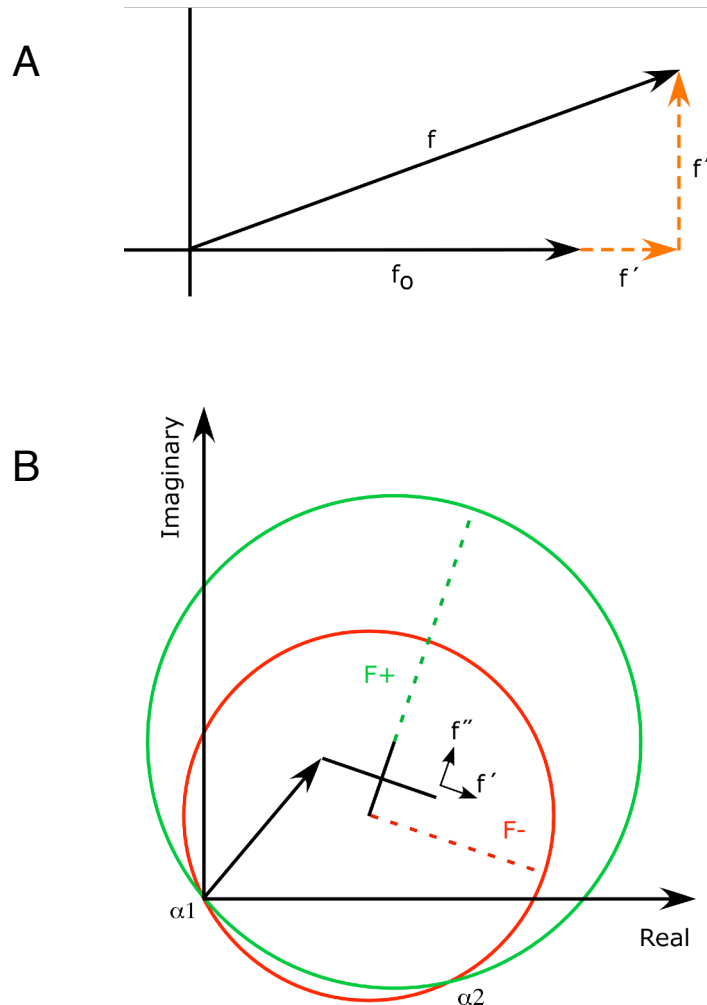


Figure 3.1. Anomalous contribution to scattering and its use in SAD phasing. (A) Atomic scattering factor when an anomalous contribution is present. The anomalous contribution is composed of a real component, f' , and an imaginary component, f'' . (B) Harker diagram for determining the protein phase angle by single-wavelength anomalous scattering. $F+$ and $F-$ are the structure factor amplitudes of the Friedel mates of the anomalous scatterers. The two

points of intersections represent the two possible values for the protein phase angles.

Mutiple-wavelength anomalous diffraction (MAD)

If the protein contains atoms that contribute to anomalous scattering, this can be exploited to solve the phase. The anomalous scatterer can be soaked into the crystal, as in the case of isomorphous replacement, occur naturally, such as in metalloproteins, or be incorporated during protein preparation, such as replacing methionine with selenomethionine in the growth medium. For selenomethionine, the general rule requires at least one Se per 100 amino acids to detect a useful anomalous signal.

Using MAD to solve the “phase problem” involves collecting diffraction data at three or more different wavelengths in order to maximize the anomalous contribution to the scattering (Ealick, 2000). This requires access to a tunable synchrotron X-ray source. The wavelengths collected correspond to the wavelength where the peak anomalous signal is detected (determined for each crystal by fluorescence), the wavelength where the dispersive difference is smallest, and at a remote wavelength. Typically, all of the data are collected on a single frozen crystal with as high a completeness as possible and a high degree of redundancy to make the measurements as statistically significant as possible. The changes in structure factor amplitudes due to anomalous scattering are

small and require very accurate measurement to be useful. Once diffraction data has been collected the MAD case can be essentially be treated as MIR.

Single-wavelength diffraction (SAD)

It is possible to calculate the phase using the anomalous difference from just the X-ray wavelength corresponding to peak wavelength if the anomalous difference is sufficient (Rice *et al.*, 2000, Dauter *et al.*, 2002, Dodson, 2003). Phasing by SAD is analogous to the SIR method, were the phase calculation leads to two possibilities, corresponding to the two possible hands of the positions of the anomalous scatterers (Figure 3.1B). To resolve this problem, phase calculations are performed in parallel on both hands followed by density modification. The phase ambiguity can be broken by examining the phase probability distribution or by visually inspecting the density modified electron density calculated for both phase possibilities (Broderson *et al.*, 2000). The correct solution should have better contrast between the solvent and protein areas of the map as well as contain more stretches of connected density.

3.1.7 Refinement and model building

Once the phases have been derived and refined, an electron density map can be calculated by Fourier transformation. If the map quality is sufficient to

reliably place the amino acids or nucleotides, model building can be used to reconstruct the protein. The model is refined to the lowest energy and through iterative rounds of manual model building and refinement. Refinement is the adjustment of the calculated and observed structure factors until they are as close to agreement as possible. The two main methods of refinement are to use a maximum-likelihood, or molecular dynamics. The maximum-likelihood method minimizes the coordinate parameters to satisfy a maximum-likelihood function. The molecular dynamics approach uses a process called simulated annealing. Simulated annealing allows a molecule to sample all possible conformations by raising the temperature to, for example, 2500°C and constraining the model by energy terms derived from the input model and the experimental diffraction data, the structure can fold and refold until it reaches an energetically favorable conformation.

The accuracy of the model can be determined by comparing the observed diffraction data to the diffraction data calculated from the model. This statistic is referred to as the crystallographic R factor, and it is defined as:

$$R = \frac{\sum_{hkl} ||F_{OBS}| - k|F_{CALC}||}{\sum_{hkl} |F_{OBS}|} \times 100$$

The crystallographic R factor is intended to provide a good approximation of how well the model reflects the experimental data. It is necessary to point out that the R factor alone is not a sufficient indicator of the structures accuracy. Other indicators of the quality of the structure are provided by the Ramachandran plot, which shows if phi and psi angles have allowed values for each residue in the protein molecule, how much the bond lengths and angles deviate from idealized values, rotamer side chain fit values, and hydrogen-bond analysis (Kleywegt, 2000).

3.1.6 The X-ray crystal structure of aRpp29

In chapter 2, NMR studies of *A. fulgidus* aRpp29 show that the central structural feature of this 102 residue protein is a sheet of six antiparallel β -strands, wrapped around a core of conserved hydrophobic amino acids. In contrast to the relatively rigid and well-defined β -sheet, about 40 percent of the protein was poorly defined in solution; this includes the first 16 residues, located before the start of the first β -strand, and the last 24 residues, located past the end of the last β -strand. In this chapter, the X-ray crystal structure of the aRpp29 from *A. fulgidus* will be presented. The X-ray structure of aRpp29 differs from the solution structure in that two well ordered helices are present at the amino and carboxy termini. Residues 6 – 12 and 83 – 96 form α -helices arranged in an anti-

parallel orientation. This chapter of the dissertation was adapted from Sidote *et al.*, 2004.

3.2 Methods and Materials

3.2.1 Crystallization and X-ray diffraction

The 102 amino acid archaeal homologue of the human RNase P protein Rpp29 (aRpp29) was cloned from *Archaeoglobus fulgidus* genomic DNA and expressed as a fusion with maltose binding protein. Following cleavage of the fusion with tobacco etch virus (TEV) protease, the protein was purified to homogeneity as described chapter 2. The protein was crystallized using the hanging drop vapor diffusion method with samples at concentrations of 15 mg/ml. Native crystals were grown in 0.1 M sodium citrate (pH 5.5) and 2.2 M ammonium sulfate at room temperature, with crystal formation occurring after 2 months. Protein containing selenomethionine was produced as described by VanDuyne *et al.* (1993), and purified exactly as the native protein. Crystals of the selenomethionine form of aRpp29 were grown from a solution of 0.1 M HEPES (pH 7.5) and 1.2 M ammonium sulfate, with crystal formation occurring after 4 to 5 days at room temperature (Figure 3.2). Crystals were transferred to a 25% glycerol cryoprotectant solution for 1 minute prior to freezing, removed from the cryoprotectant using a nylon loop (Hampton Research), and frozen in the

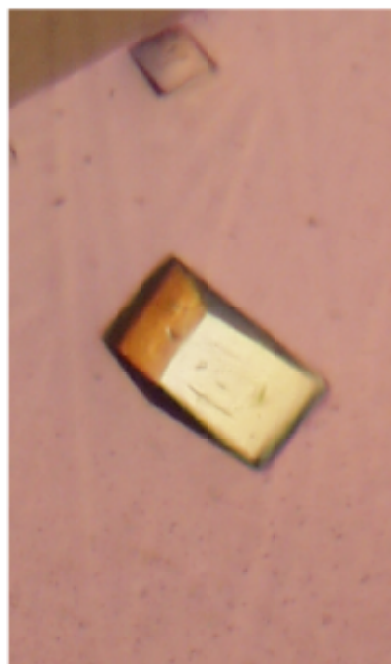
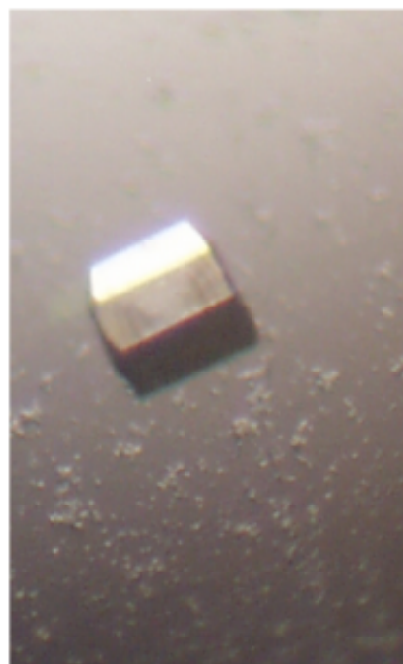
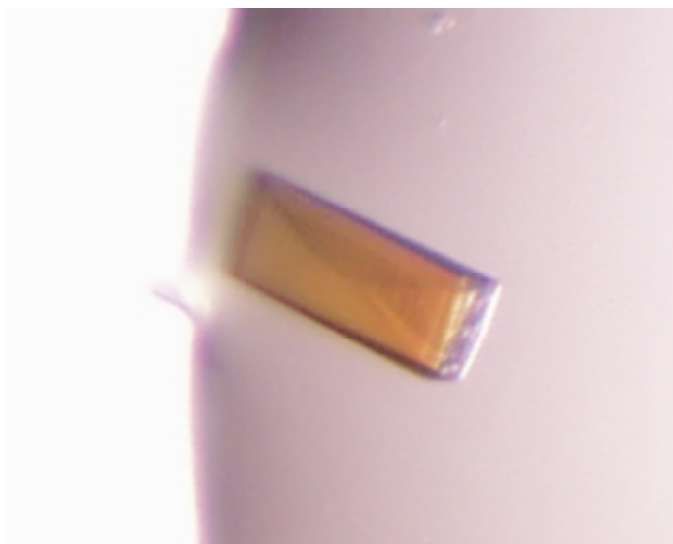


Figure 3.2. Crystals of *A. fulgidus* RNase P protein aRpp29. Different crystal forms of the aRpp29 protein were grown by vapor diffusion using the hanging drop technique. The native form grew in the space group $P2_12_12_1$ and the SeMet form in the space group $P2_12_12$. In both cases, crystals grew to an approximate diameter of 0.2-0.3 mm.

nitrogen stream at 100K. X-ray diffraction data at 1.54 Å were collected for the native and selenomethionine forms using a rotating copper anode X-ray source equipped with a MAR3450 detector (MAR-USA, Inc.) (Figure 3.3). Synchrotron data were collected for the selenomethionine form at the PX beamline at the Center for Advanced Microstructures and Devices (CAMD) in Baton Rouge, LA, equipped with a MAR CCD detector (MAR-USA, Inc.); data were collected at 0.97945 Å, corresponding to the peak selenium anomalous signal according to a fluorescence scan (Figure 3.4). Diffraction data were integrated and scaled using DENZO and SCALEPACK (Otwinowski and Minor, 1997) to a resolution of 1.7 Å.

3.2.2 Crystal structure determination and refinement

The crystal structure of the selenomethionine form of aRpp29 was solved using the single wavelength anomalous diffraction (SAD) method. Alternative methods for solving the phase problem, such as isomorphous replacement and molecular replacement were attempted without success. The unit cell constants of the selenomethionine crystals, which grew faster than the native and therefore were more abundant for heavy atom soaking, varied considerably from crystal to crystal after cryoprotection. In most cases, as is the case with soaking crystals in heavy atom salts, the quality of diffraction drastically decreased or caused the crystal to deteriorate while soaking. Molecular replacement was attempted with

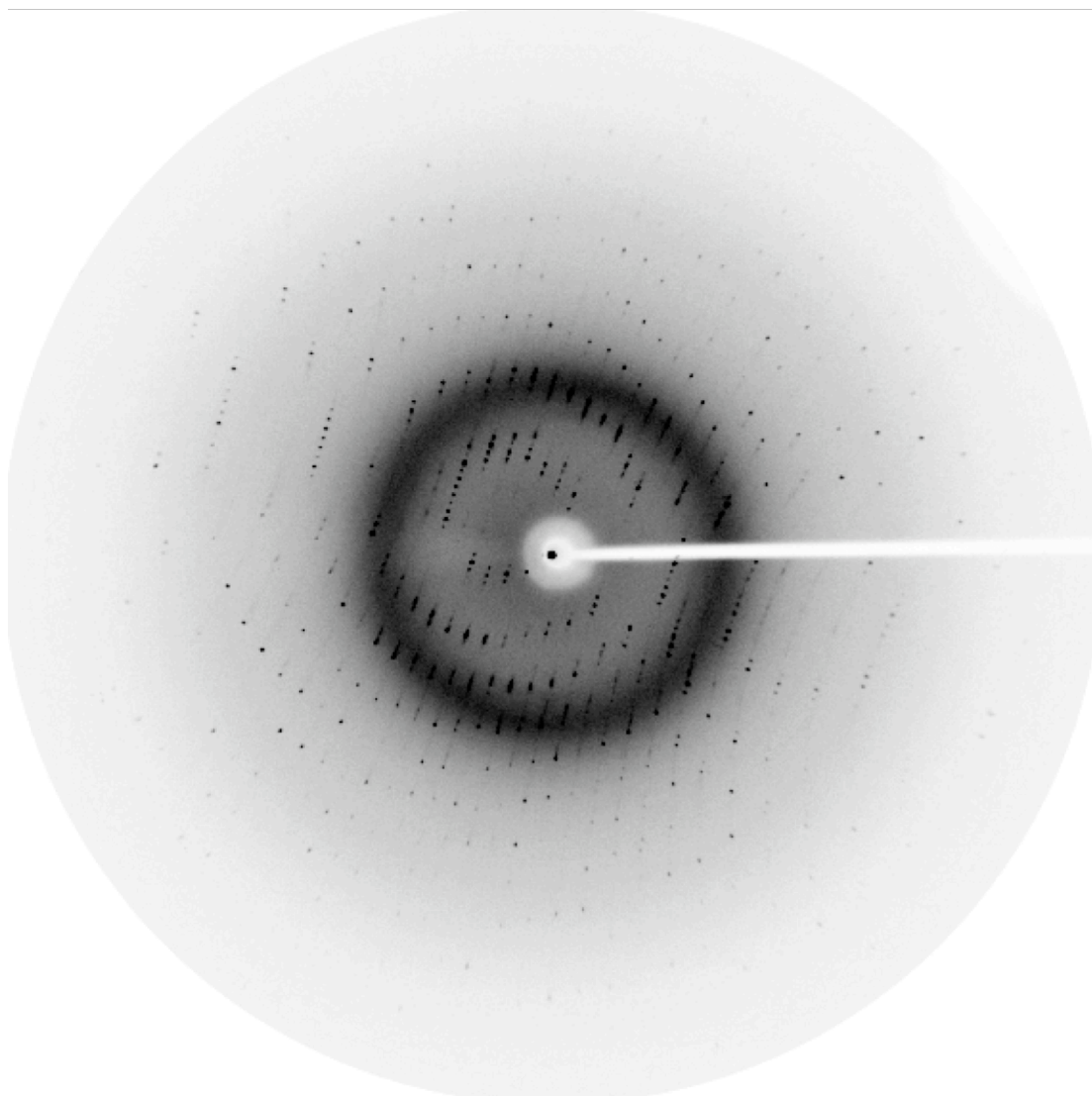


Figure 3.3. X-ray diffraction image of aRpp29 to 1.7Å resolution. These data were collected on a “home source” rotating copper anode X-ray generator equipped with a MAR 3450 detector. The crystal was flash cooled to 100K with 25% glycerol used a cryoprotectant prior to data collection.

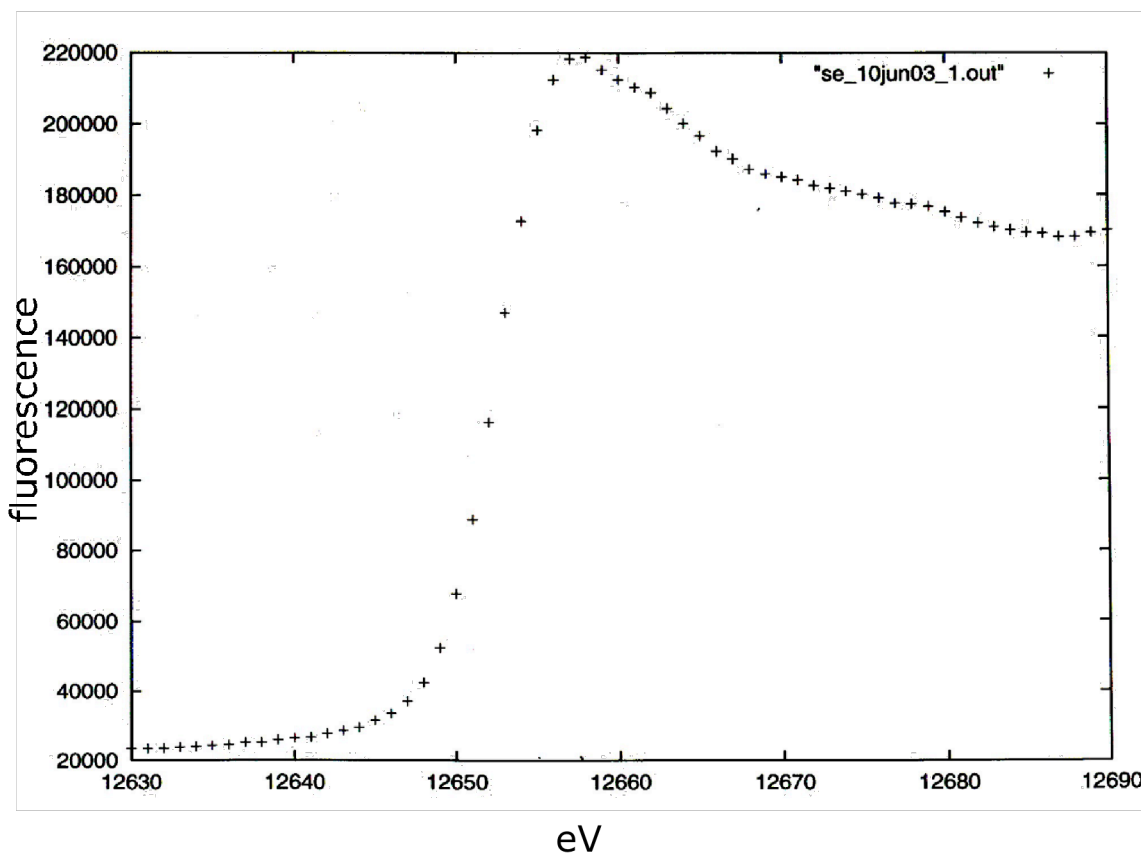


Figure 3.4. Fluorescence scan of an aRpp29 crystal containing 5 selenomethionine residues. In order to determine the wavelength of X-rays that produces the maximum anomalous signal, a fluorescence scan is performed. The fluorescence is proportional to the imaginary component of the anomalous scattering factor. The peak wavelength derived from this scan for this crystal was 0.979495Å.

many search models and with different software suites but a correct solution was not found. The possible reasons for this are addressed later.

The locations of the five selenium atoms and initial phases were calculated at 3.5 Å using SHARP (Bricogne *et al.*, 2003) (Figure 3.5). After solvent flattening using ARP/warp (Perrakis *et al.*, 2001), a model of the structure was built into the electron density using XtalView (McRee, 1999). The phase ambiguity that is inherent in the SAD was broken by examining the phase distribution probabilities calculated by SHARP and was confirmed by visually inspecting the electron density maps to access which phase calculation is correct. After positional refinement, density modification was repeated at 1.8 Å to improve the phases, followed by multiple cycles of re-building (using omit maps) and positional refinement using the 1.54 Å diffraction data to a resolution of 1.7 Å. Atomic B-factors were refined using CNS version 1.1 (Brünger *et al.*, 1998). The structure of the native form of aRpp29 (which crystallized in a different space group) was solved by the molecular replacement method using CNS, with the structure of the selenomethionine form of the protein as a search model.

After multiple cycles of re-building using omit maps, and positional and B-factor refinement, the structure of the native form of the protein was completed using data to a resolution of 1.7 Å. Water molecules were added to each structure where the difference electron density was greater than 3 σ and the

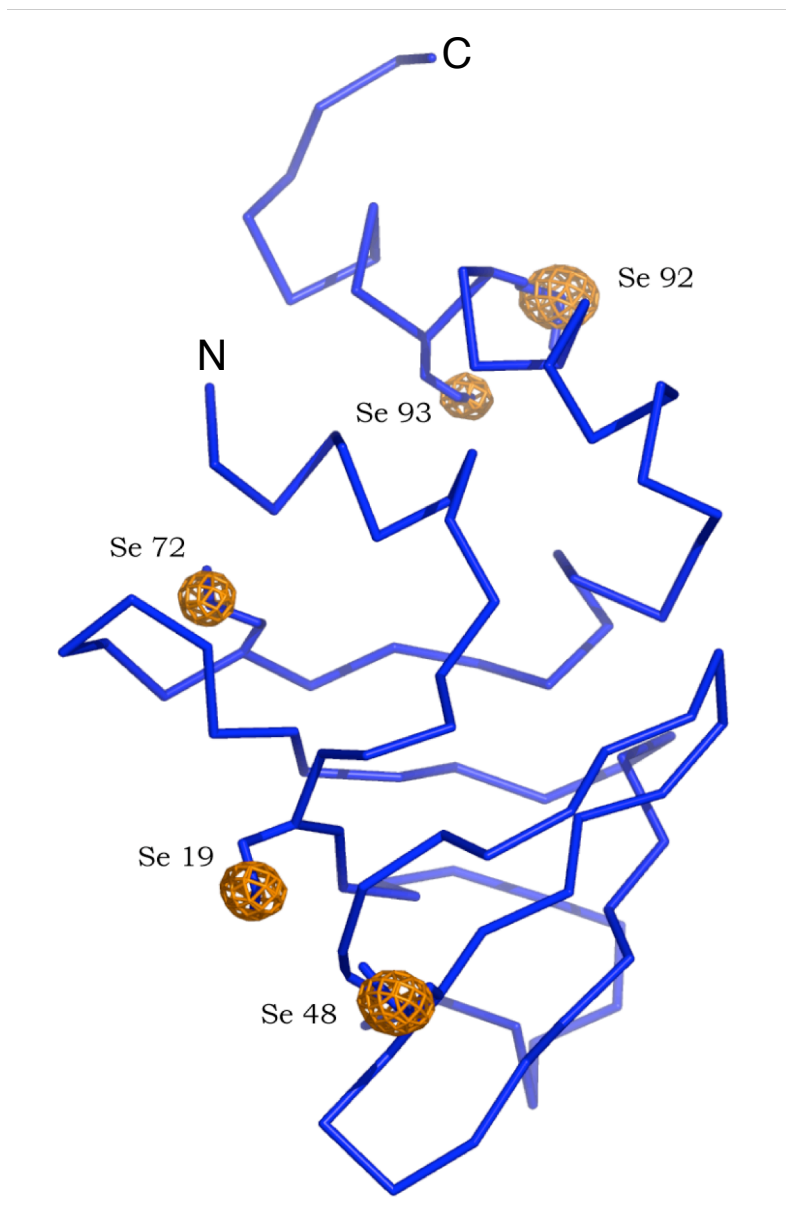


Figure 3.5. Anomalous difference map showing the positions of the selenium atoms in the aRpp29 structure. The backbone of the aRpp29 crystal structure is shown in blue. The sidechains of SeMet19, -48, -72, -92, -93 are also drawn in blue. The orange mesh represents the electron density corresponding to the anomalous difference in scattering from the selenium atoms contoured at 10σ .

density in the $2|F_o| - |F_c|$ map was greater than 1 σ . Summaries of the crystallographic and refinement statistics for the native and selenomethionine forms of the aRpp29 protein are given in Table 3.1. The quality and stereochemistry of the final structures were evaluated using PROCHECK (Laskowski *et al.*, 1996). The programs DALI (Holm and Sander, 1993) and VAST (Gilbrat *et al.*, 1996) were used to search the RCSB database for proteins that are structurally similar to aRpp29. Ribbon and stick diagrams were generated using Pymol (Delano, 2002).

3.3 Results and Discussion

3.3.1 Overview of the aRpp29 crystal structure

Crystals of the *A. fulgidus* aRpp29 protein were grown from a solution of sodium citrate with ammonium sulfate as the precipitant; the crystals formed in the space group $P2_12_12_1$ and diffract X-rays to 1.7 Å resolution. A derivative of aRpp29 containing selenomethionine (SeMet) in place of methionine was prepared as a tool for solving the phase problem by anomalous scattering methods. Interestingly, the SeMet-containing aRpp29 did not crystallize under the same conditions as the native form of the protein. However, crystals with a slightly larger unit cell diffracting to 1.7 Å were obtained in space group $P2_12_12$ from a solution of HEPES and ammonium sulfate.

Table 3.1. Structure determination and refinement statistics for the X-ray crystal structure of *A. fulgidus* aRpp29.

Data set	Native (Cu K)	SeMet (Cu K)	SeMet (CAMD)
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2
Cell constants (Å, °)	a = 39.6 b = 43.8 c = 49.2	a = 35.9 b = 71.8 c = 32.4	a = 38.1 b = 73.0 c = 32.5
	$\alpha, \beta, \gamma = 90.0$	$\alpha, \beta, \gamma = 90.0$	$\alpha, \beta, \gamma = 90.0$
Number of crystals	1	1	1
Wavelength (Å)	1.542	1.542	0.97954
Dmin (Å)	1.7	1.7	2.3
Rsym (%)	4.3 (45.6)	4.0 (24.8)	8.0 (24.8)
Completeness (%)	90.1 (98.5)	99.1 (88.1)	99.1 (93.1)
I/ σ	25.6 (3.0)	138.7 (21.9)	23.2 (5.3)
Phasing power			1.52
No. Observed Reflections	34,081	124,791	27,573
No. Unique reflections	8955	10,241	7688
Average redundancy	3.8	12.2	3.6
Residues in structural model	6 – 86	5 - 102	
Overall figure of merit	0.81	0.87	
Resolution range for refinement (Å)	15.0 to 1.7	15.0 to 1.7	
Rwork (%)	25.0	20.3	
Rfree (%)	28.4	23.9	
Average B factor (Å ²)	18.9	22.2	
Number of protein atoms	650	781	
Number of water molecules	99	144	
r.m.s.d. bond lengths (Å)	0.00500	0.00446	
r.m.s.d. bond angles (°)	1.360	1.350	
r.m.s.d. bonded B factors (Å ²)	1.379	1.297	
Residues in most favored regions (%)	95.7	91.7	
Residues in additional allowed regions (%)	4.3	8.3	
Residues in generously allowed regions (%)	0.0	0.0	
Residues in disallowed regions (%)	0.0	0.0	

All x-ray data were collected at 100 K. The values in parentheses correspond to the highest resolution shell of the x-ray data. The highest resolution shells for the native, SeMet, and CAMD datasets are 1.76-1.70, 1.72-1.66, and 2.38-2.30 respectively.

Obtaining two different crystal forms of aRpp29 was fortuitous in that it provided for two independent views of the protein structure. Several of the methionine residues were found to be located on the inter-molecular contacts within the crystal, thus providing an explanation for the change in molecular packing upon substituting selenomethionine for methionine.

The electron density maps for each crystal form of aRpp29 were of excellent quality. For the SeMet form of the protein, residues 5 to 102 were built into the maps and refined. For the native protein, only residues 6 to 86 were clearly observed in the electron density maps; inspection of the molecular packing suggests that the C-terminal residues were absent in the native form of the protein, presumably due to degradation prior to crystallization (Figure 3.6). This observation is consistent with the significant difference in crystallization time for the two crystal forms: The SeMet protein crystallizes in several days, while the native protein requires two months to form crystals.

The native and SeMet structures were refined to 1.7 Å with $R_{\text{working}}/R_{\text{free}}$ of 25.0%/28.4% and 20.3%/23.9%, respectively. The solvent contents of 31.7% and 30.2% (v/v) for the native and SeMet crystal forms, respectively, are at the lower end of the range typical for protein crystals; this is consistent with relatively tight packing of the protein molecules and excellent diffraction. The two structures can be superimposed with an r.m.s.d. of 0.8 Å for the C α atoms, with most of the variation coming from the last few residues at the C-terminus. Both structures

have been deposited to the Protein Data Bank, and assigned accession numbers 1TS9 and 1TSF.

3.3.2 Description of the structure of aRpp29

The central feature of the aRpp29 protein is a sheet of six antiparallel β -strands formed by residues 17 through 75; this β -sheet is significantly twisted and wrapped around a core of conserved hydrophobic amino acids Val20, Val22, Val37, Val38, Leu45, Ile47, Val56, Phe63, Val65, and Ile74 (Figure 3.6). The high degree of conservation of these hydrophobic residues in the archaeal aRpp29, yeast Pop4, and human Rpp29 indicates that each of these proteins contains a similar β -sheet structure (Figure 3.7). Within the β -sheet, strand β 1 is connected to β 2 by a 7-residue loop containing one turn of helix centered at residue 28, and strand β 2 is connected to β 3 by a short loop, as is β 3 to β 4. Strand β 4 is connected to β 5, located at the opposite side of the protein, by a 5-residue loop that spans the open part of the twisted β -sheet. Strand β 5 is connected to β 6 by a compact turn, followed by a 7-residue loop containing one turn of helix centered at residue 76.

The N-terminal region of the aRpp29 protein contains a two-turn α -helix (residues 6 through 12), followed by a short loop leading to the first β -strand. Side chains of residues Leu10, Ile11 and Trp15 are oriented toward the hydrophobic face of the β -sheet; residues at these positions are hydrophobic

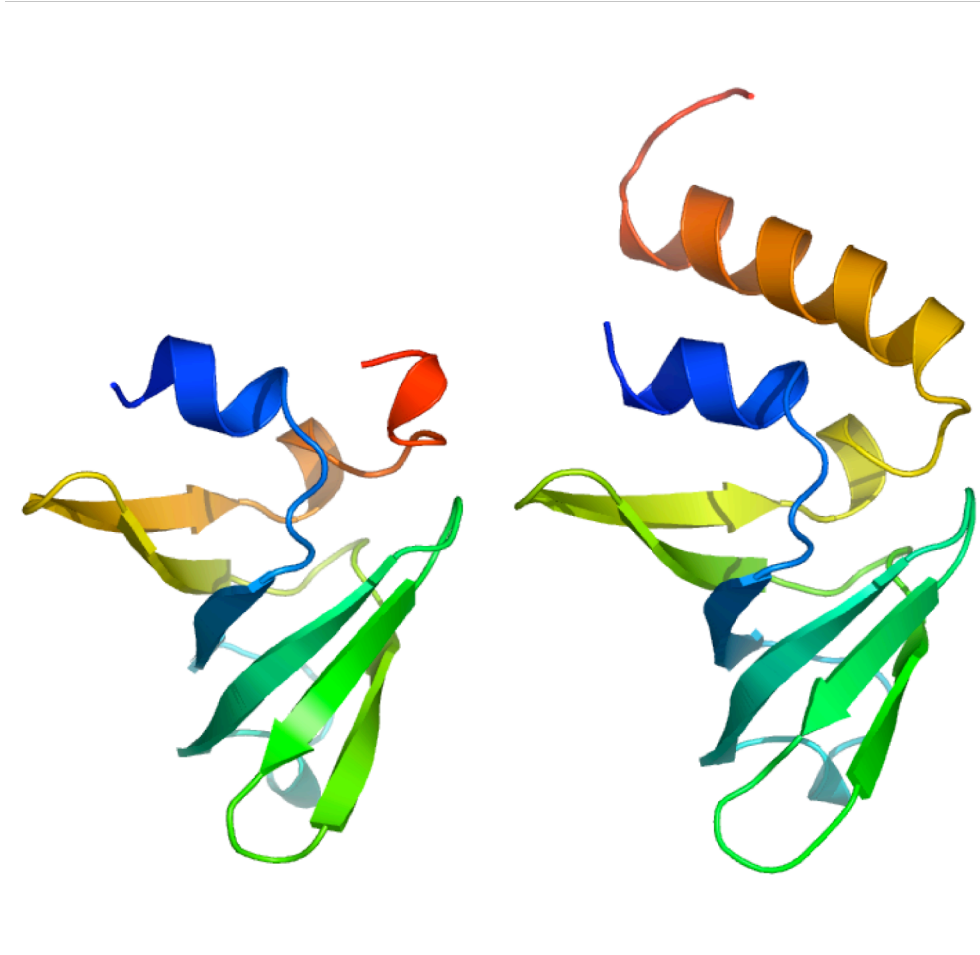


Figure 3.6. Cartoon representation of the aRpp29 native and SeMet crystal structures. The cartoons are color ramped from blue at the N-terminus to red at the C-terminus. (A) The native form of aRpp29 crystallized in the space group $P2_12_12_1$. The structure consists of residues 6 - 86. Inspection of the crystal packing suggests that residues 87 - 102 have degraded and are not missing due to disorder. (B) The SeMet form of aRpp29 crystallized in the space group $P2_12_12$. The space group difference can be attributed to the positions of the SeMet residues, as they are located on the surface and form crystal contacts with neighboring molecules. The SeMet structure consists of residues 5 - 102.

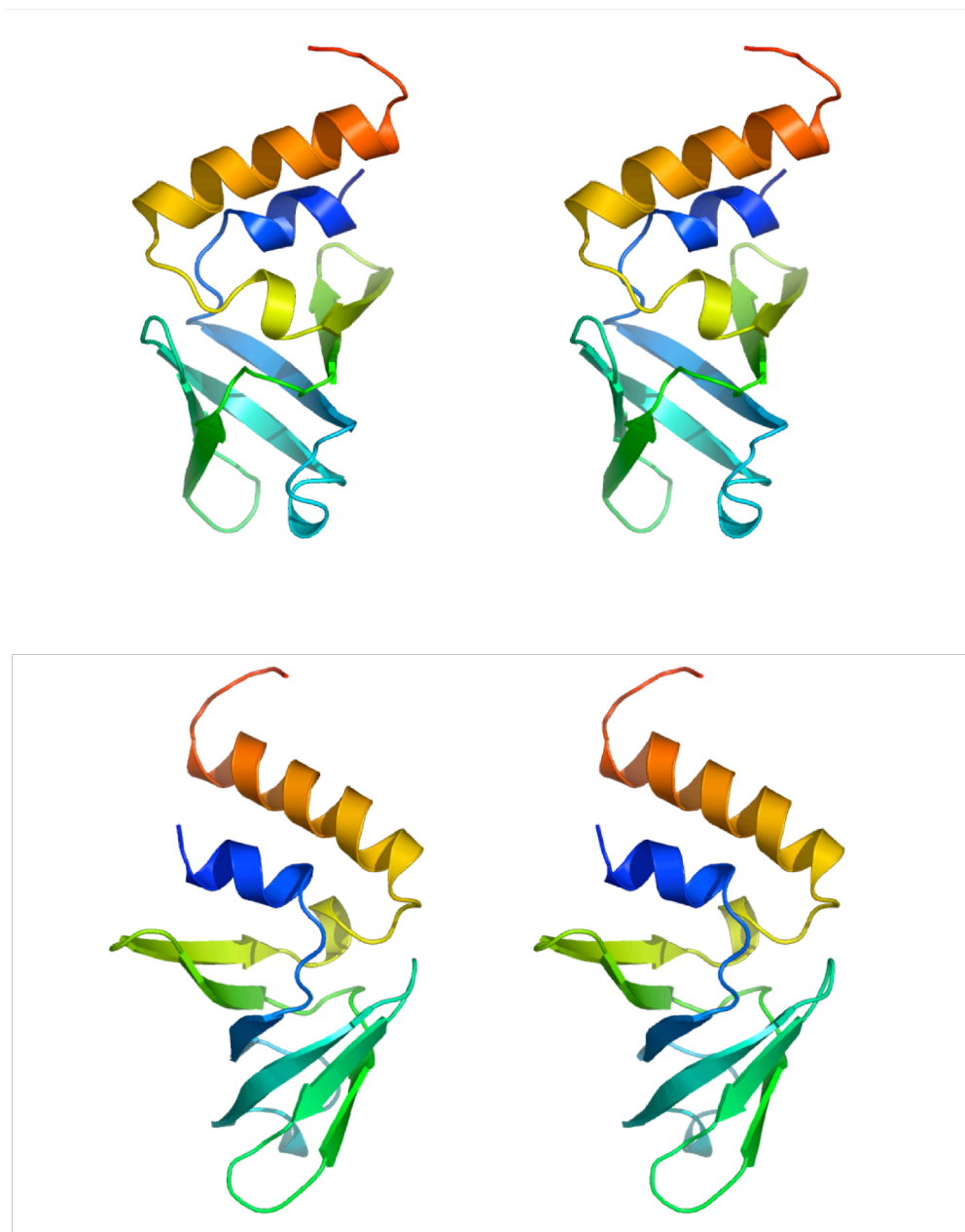


Figure 3.8. Stereo view of the crystal structure of aRpp29 from *A. fulgidus*. Cartoon diagram showing two orthogonal views of the SeMet form of *A. fulgidus* aRpp29, color ramped from blue at the N-terminus to red at the C-terminus. The figure was generated using the backbone atoms of residues 5 to 102.

in aRpp29 from most species, suggesting that the N-terminal helix is a conserved feature of the protein structure. The C-terminal region of the SeMet form of the protein contains a four-turn α -helix formed by residues 83 to 96, followed by an irregular loop to residue Trp101. In the native form of the protein, only a single turn of this helix is observed in the electron density map.

The N-terminal and C-terminal helices are aligned antiparallel to each other, which may be a stabilizing feature since it provides for a favorable alignment of the helix dipoles. In the SeMet form of the protein, the N- and C-terminal helices make inter-molecular contacts near a two-fold axis within the crystal. The selenomethionine residues are all exposed on the protein surface, with SeMet 19, 48, and 92 being directly involved in forming crystal contacts. In contrast, the packing of the native protein is mediated by the N-terminal helix and the C-terminal region from Asp77 to Arg86 packing against the surface of the β -sheet formed by strands β 1, β 2, β 5, and β 6 (Figure 3.9).

The crystal structure of aRpp29 is structurally similar to the Sm and Sm-like (Lsm) proteins which are characterized by a motif consisting of two distinct sequences. The Sm-fold is defined by an N-terminal helix followed by a five-stranded antiparallel β -sheet that forms an open barrel (Toro *et al.*, 2002) (Figure 3.10). The structure of aRpp29 is therefore a variation of this fold, as it contains the N-terminal helix and five β -strands (β 1- β 5) exhibiting the same connectivity, with strands β 6 and the C-terminal helix being extra elements. The Sm-like

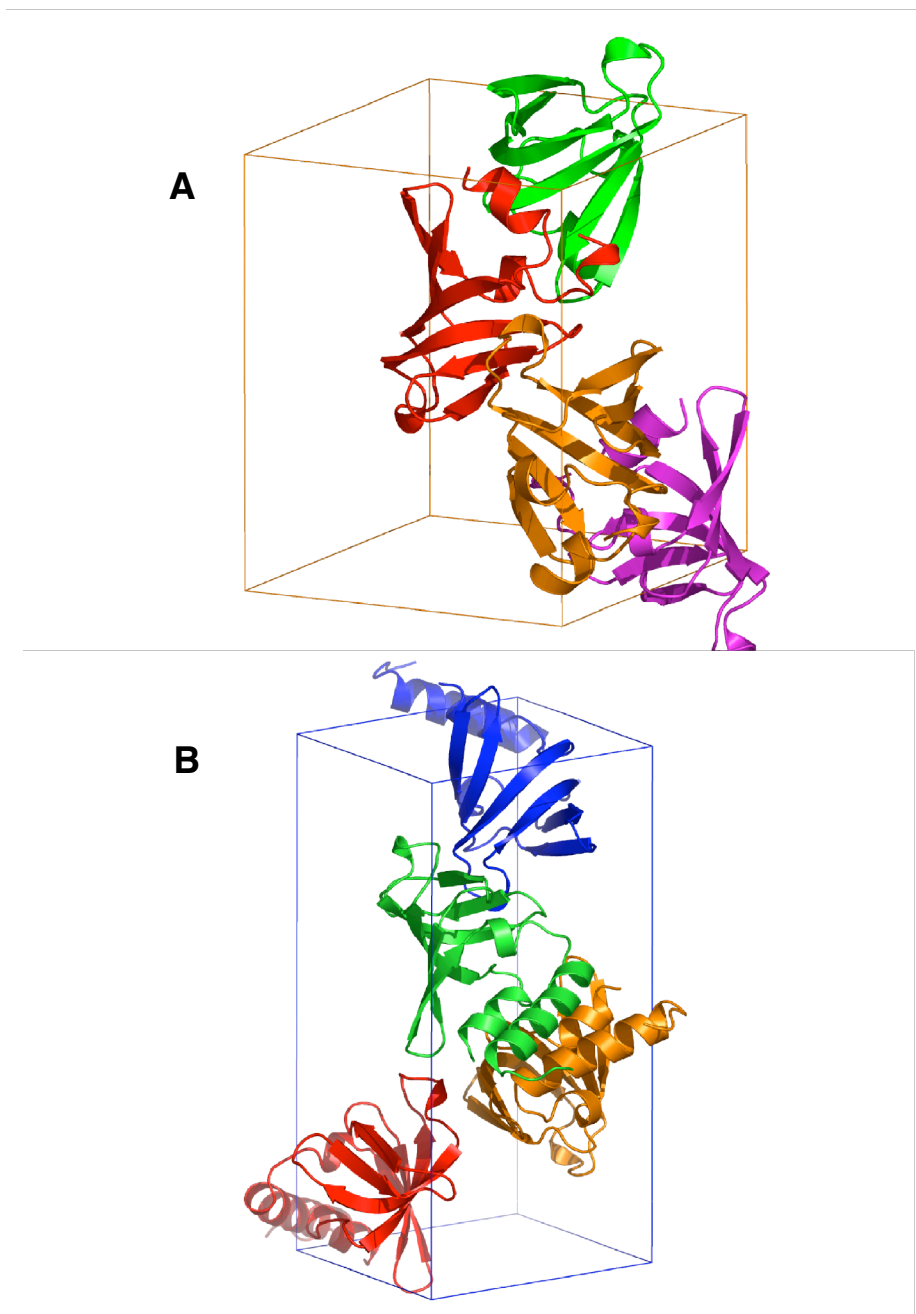


Figure 3.9. Lattice packing of the native and SeMet forms of aRpp29. (A) Native aRpp29 packing in space group $P2_12_12_1$. a, b, c denote the unit cell axes. (B) SeMet aRpp29 packing in space group $P2_12_12$. Selenomethionine residues at positions 19, 48, and 92 form crystal contacts. The SeMet residue at position 92 is involved in making crystal contacts along the 2-fold axis (green and orange molecules in panel B).

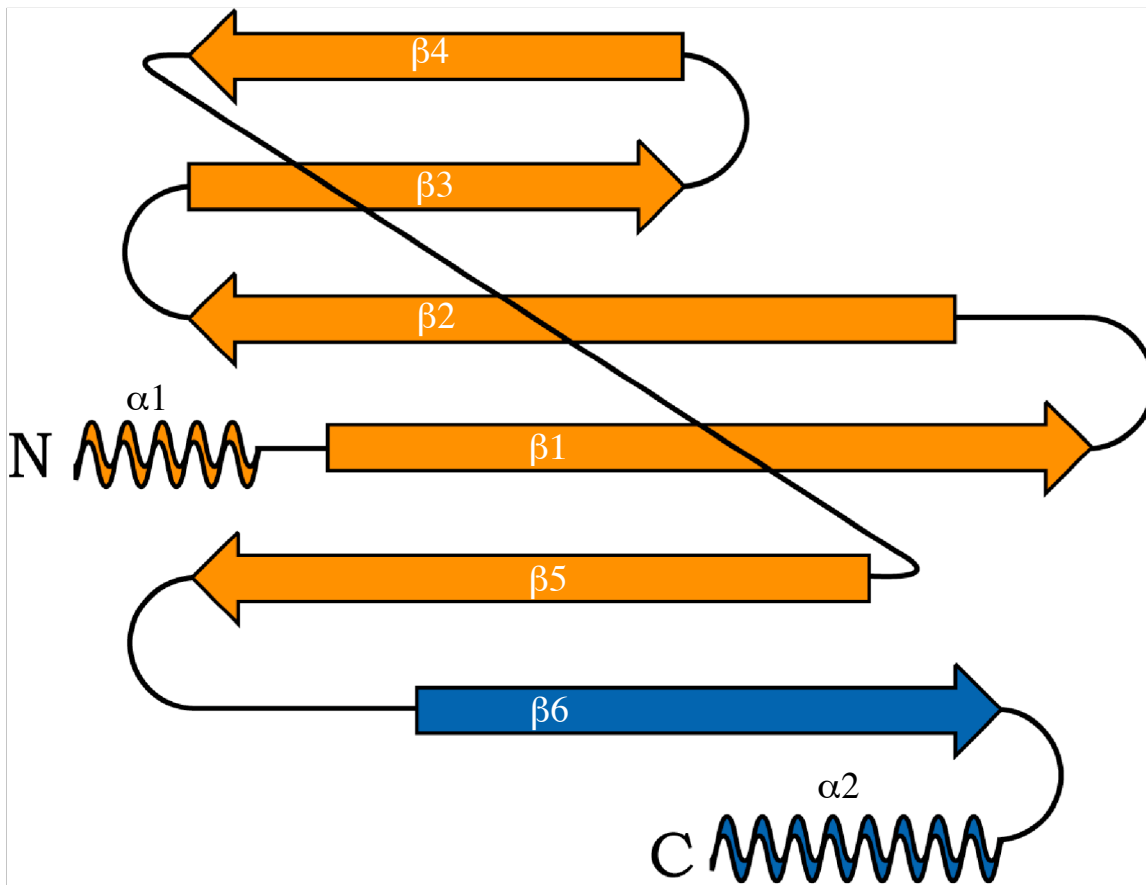


Figure 3.10. Two-dimensional diagram of the aRpp29 variation of the SM-fold. The SM and Lsm (SM-like) fold consists of an N-terminal α -helix followed by 5 β -strands forming an anti-parallel β -sheet (Shown in orange). The structure of aRpp29 is a variation of the SM fold, as it contains the distinguishing features of the fold with an additional anti-parallel β -strand followed by a C-terminal α -helix (colored in blue).

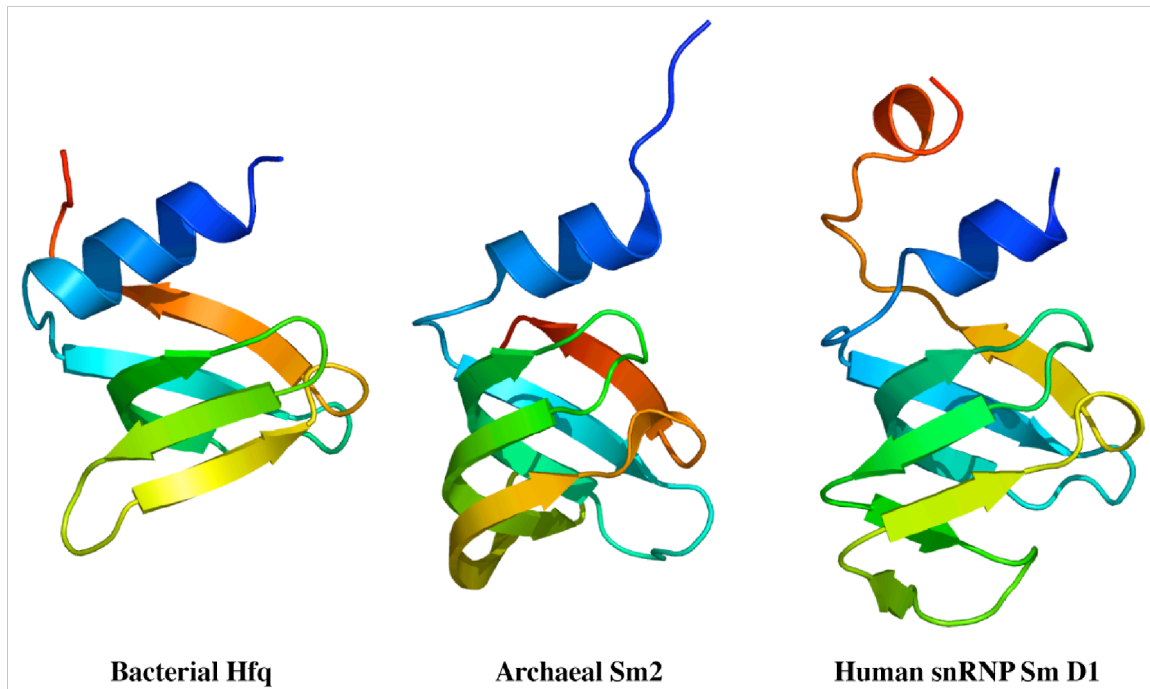
proteins occur in all three domains of life (bacteria, archaea and eukarya), and are known to associate with U-rich RNA sequences. Sm and Sm-like proteins are known to be involved in a variety of RNA processing events, including pre-mRNA splicing (Pannone *et al.*, 2001) and the biogenesis of the small nuclear ribonucleoprotein particles (snRNPs), as well as interaction with pre-RNase P RNA (Toro *et al.*, 2001, reviewed by Valentin-Hansen *et al.*, 2004).

The structural similarity between the β -sheet regions of aRpp29 and the Sm-like protein and bacterial transcription factor Hfq becomes even more apparent now that the crystal structure of aRpp29 is known, since Hfq contains a helix that corresponds to the N-terminal helix of aRpp29. In addition to their structural similarity, there is a slight but detectable amino acid sequence similarity between aRpp29 and Hfq. The bacterial Hfq protein contains the glycine at the bend in the central region of strand β 2 (corresponding to the well-conserved Gly35 in *A. fulgidus* aRpp29) that appears to be essential for forming the characteristically bent β -sheet. The similarity between aRpp29 and other Sm-family members is not always noticeable at the level of primary sequence, although the structural similarity is fairly obvious. Unlike the apparently monomeric aRpp29, many of the Sm-like proteins form hexameric or heptameric complexes (Toto *et al.*, 2002, Schumacher *et al.*, 2002, Thore *et al.*, 2003; reviewed by Valentin-Hansen *et al.*, 2004). The structural similarity to aRpp29

can be seen in the ribbon diagrams of the monomer units of the Sm and Sm-like proteins Hfq, Sm2 and Human snRNP D1 shown in Figure 3.11.

The structure of a complex of Hfq with a single-stranded oligoribonucleotide (Schumacher *et al.*, 2002) shows that the RNA binds in a basic cleft, however the residues of Hfq that contact the RNA do not correspond to conserved residues in aRpp29, suggesting that the RNA-binding modes of these proteins are different. An NMR chemical shift perturbation study of the *M. thermoautotrophicus* homologue of aRpp29 by Boomershine *et al.* (2003) suggested that a broad surface involving a number of the β -sheet residues may contact the RNA, which differs from the conserved surface that we suggest in Figure 1B.

In the recent crystal structure of the *P. horikoshii* homologue of aRpp29, Numata *et al.* (2004) propose a concave surface near the β 2- β 3 connecting loop as a possible site of interaction with RNA. The *A. fulgidus* aRpp29 as a whole is a basic protein, and conserved basic residues such as Arg82 and Arg89 on the protein surface are likely candidates for RNA contacts. An electrostatic surface plot based on the present crystallographic work indicates that the surface is overwhelmingly positive in charge (Figure 3.12). It appears clear that structural results for an aRpp29 protein-RNA complex will be required before the details of the protein-RNA interaction are understood.



3.11. Structural homologs of aRpp29 exhibiting the SM-fold. Cartoon representations of proteins that are similar in structure to aRpp29 and exhibit the Sm-fold, demonstrating its conservation across all three domains of life. The diagrams are color ramped from blue at the N-terminus to red at the C-terminus. The proteins shown are: Host-factor q (Hfq) from *E. coli* (PDB entry 1HK9), an RNA-binding protein involved in regulating translation; Sm-2 from *A. fulgidus* (PDB entry 1LJO); Sm-D1 from *H. sapiens* (PDB entry 1B34), which forms a dimer with Sm-D2 and interacts with the 3'-terminal U-tract of U6 snRNA. The r.m.s.d. between the backbone atoms of *A. fulgidus* aRpp29 (residues 17-77) and Hfq, Sm2, and Sm-D1 are 2.2 Å, 2.1Å, and 3.5 Å, respectively.

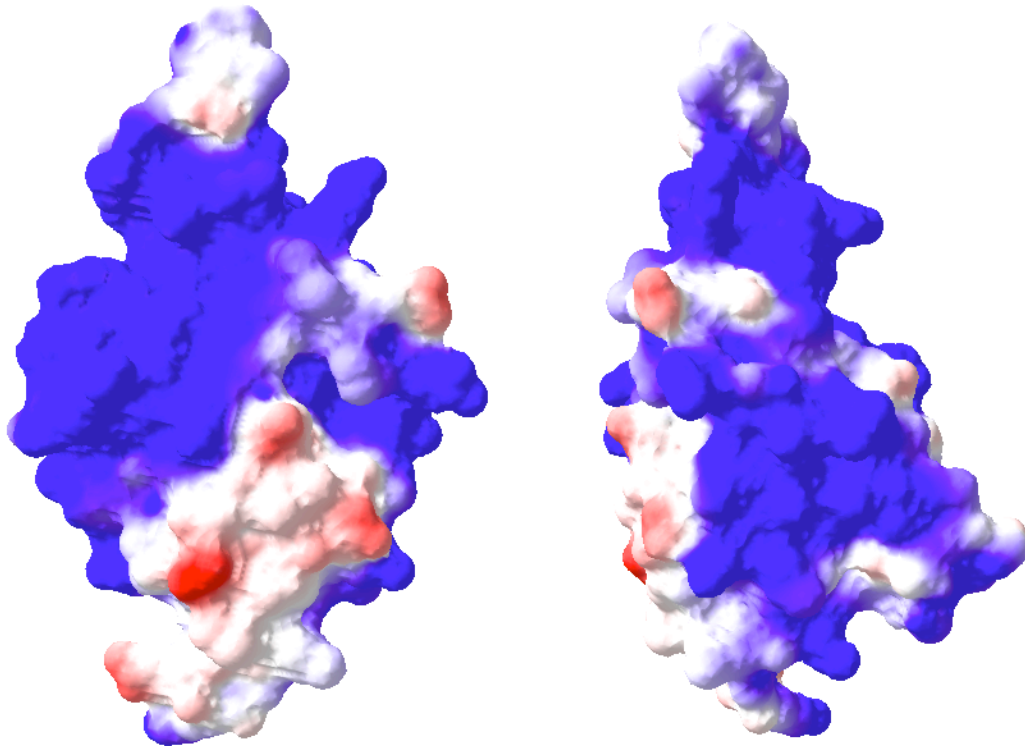


Figure 3.12. Electrostatic surface representation demonstrating the overall positive character of aRpp29. aRpp29 is an RNA binding protein with an isoelectric point of approximately 10.5. The surface contains an overall positive charge which may be the sites of interaction with RNase P RNA or the pre-tRNA substrate. The electrostatic calculation was performed using Swiss PDB viewer.

3.3.3 Conserved surface residues of aRpp29

Amino acids are generally conserved among homologous proteins from a diverse set of species when they are important for purposes of structure, function or for inter-molecular interactions. In the case of aRpp29 and its homologues (Figure 3.7), conserved hydrophobic residues in the protein core are likely to be essential for stabilizing the overall protein fold. The conserved residues on the protein surface are the most likely to be important for inter-molecular interactions, such as contacting the RNase P RNA or other protein components.

The most-conserved residues on the surface of aRpp29 are spread throughout the primary sequence of the protein, but are spatially clustered into a single region of the protein surface (figure 3.13). Six surface residues (Asp14, Thr41, Gln42, Asn43, Arg82 and Arg88) that are well-conserved in archaeal aRpp29, as well as the RNase P proteins Rpp29 and Pop4 in eukarya, are located either within or near the loop connecting strands $\beta 2$ and $\beta 3$, and the beginning of helix $\alpha 2$. Each of these six conserved surface residues may act as either a hydrogen bond donor or acceptor, and is potentially a site of interaction with the RNase P RNA, the substrate pre-tRNA, or both.

Another interesting feature of the protein is a salt bridge formed by the triad of residues Glu40, Lys58, and Arg86 (Figure 3.14). These hydrophilic amino acid types are most commonly found on the surface of proteins, however in aRpp29, the charged ends of these residues are buried so that the salt bridge

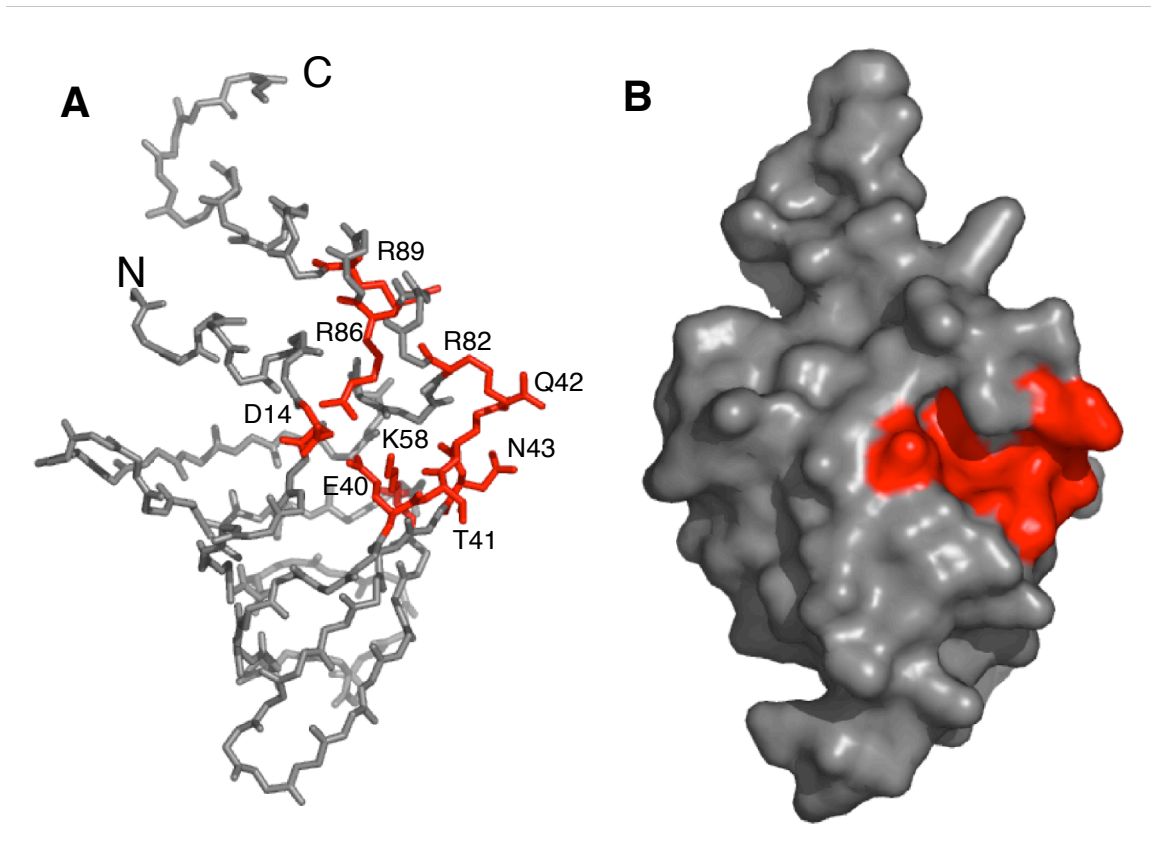


Figure 3.13. Conserved surface residues provide a potential site of interaction. (A) Surface residues that are conserved in a wide range of species and considered likely to interact with the RNase P RNA or other protein subunits are shown in red; these conserved residues are clustered into a relatively small region of the protein surface. (B) Surface representation of aRpp29 with conserved surface residues colored red.

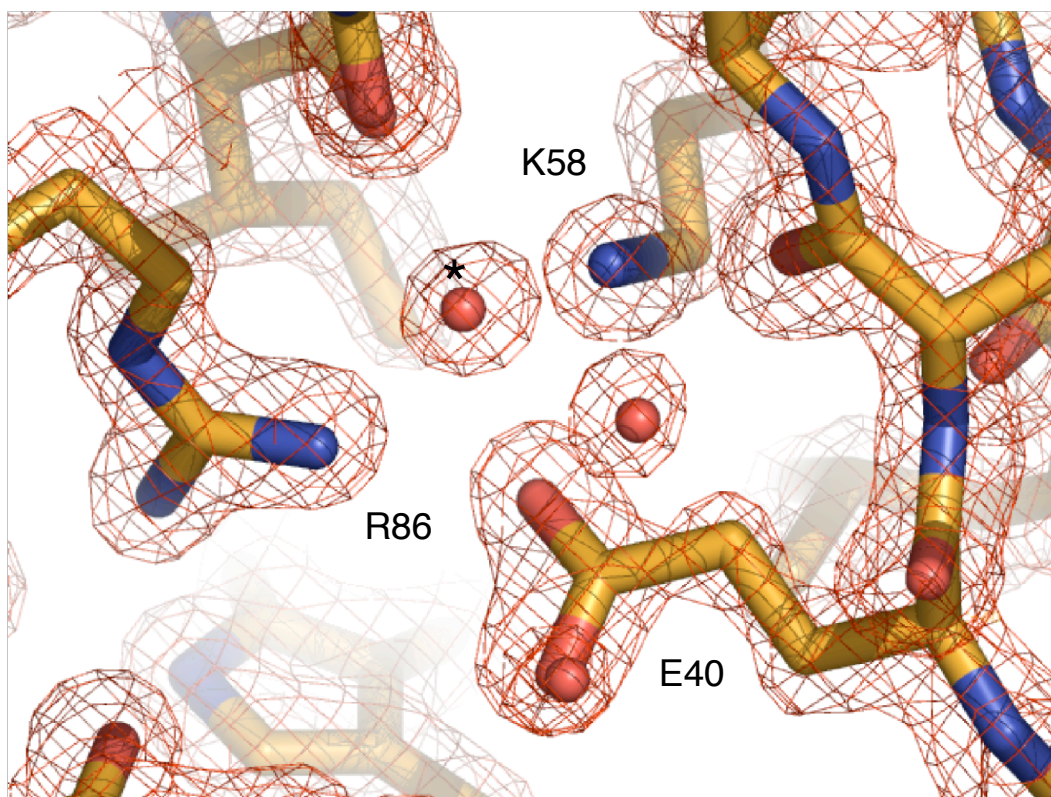


Figure 3.14. A region of aRpp29 2IF_o-|f_c| electron density map showing the three-way internal salt bridge formed by residues E40, K58, and R86. These residues are nearly completely inaccessible to the solvent. The distances from Glu40 O ϵ to Lys58 N ζ and Glu40 O ϵ to Arg86 N η are 2.71Å and 2.63Å, respectively. The water molecule marked with an asterisk is involved in mediating the interaction between Lys58 and Arg86; the oxygen of this water molecule is 3.10Å from Lys58 N ζ and 2.94Å from Arg86 N η . The electron density is contoured at 1.5 σ .

spans an interior region of the protein. This triad of residues is conserved from aRpp29 in *A. fulgidus* to Rpp29 in humans and Pop4 in yeast. A study of salt-bridges within protein structures indicates that such a three-way internal structure is rather unusual (Kumar and Nussinov, 1999), and the degree to which the identity of each of these residues is conserved is also quite striking (Figure 3.7). It seems reasonable to speculate that the residues of the salt-bridge are conserved for a reason. One possibility is that the salt-bridge is required simply for protein stability. However, the present study provides evidence that the N- and C-terminal helices are far from rigidly stable, suggesting that the salt-bridge is often not as buried as it appears in the crystal. Also, the observation that these residues are conserved not just in thermophilic archaea, but also in higher eukarya, suggests the possibility that their function may be more complex. A relatively complex recognition mechanism could explain why the involved amino acids are so well conserved from archaea to humans.

The alignment of the aRpp29 sequence with sequences of homologous proteins (Figure 3.7) indicates that one of the most conserved residues is a glycine (Gly35 in *A. fulgidus* aRpp29) located at a bend near the center of strand β 2. A Ramachandran plot for aRpp29 shows that Gly35 has phi and psi angles that would be unfavorable for any other amino acid type (phi and psi are 163 and -169 degrees, respectively). This glycine therefore appears to be essential to the

fold of the protein, in that it allows the structure to form its strongly bent open barrel.

3.3.4 What about molecular replacement?

There are many examples of NMR structures being used successfully as molecular replacement search models for phasing X-ray crystal structures (reviewed in Chen *et al.*, 2000). However, the circumstances for each problem are different. In the case of aRpp29, the molecular replacement method was attempted using all variations of the NMR structure as a search model. Specifically, the search models included using the complete β -strand region, with and without the loop regions, generating models with ideal geometry, the *M. thermoautotrophicus* NMR structure, and the crystal structure of Hfq. It is worth noting that our NMR model was only 60 % of the total structure so using smaller models might have required the r.m.s.d. be even smaller than for the complete structure. Molecular replacement was attempted with each search model using the computer programs Molrep, Amore, CNS, and EPMR without success.

The failure to find a molecular replacement solution can be understood now that the crystal structure has been solved using SAD. The r.m.s.d. between the backbone atoms of the crystal and the average of all the calculated NMR structures is calculated to be 3.7Å. Molecular replacement was attempted using individual as well as the average NMR models. This difference is beyond the

positioning capabilities used in molecular replacement. When the NMR structure is superimposed onto the crystal structure the differences between the two structures become apparent. The β -sheet consisting of $\beta 1$, $\beta 2$, and $\beta 5$ is twisted in relation to the crystal structure (Figure 3.15). It is possible that, because the β -sheet does not form a closed barrel, the dynamic nature of the termini of aRpp29 exhibit a force on the β -sheet part of the structure causing slight changes in the orientation of one part of the β -sheet toward another.

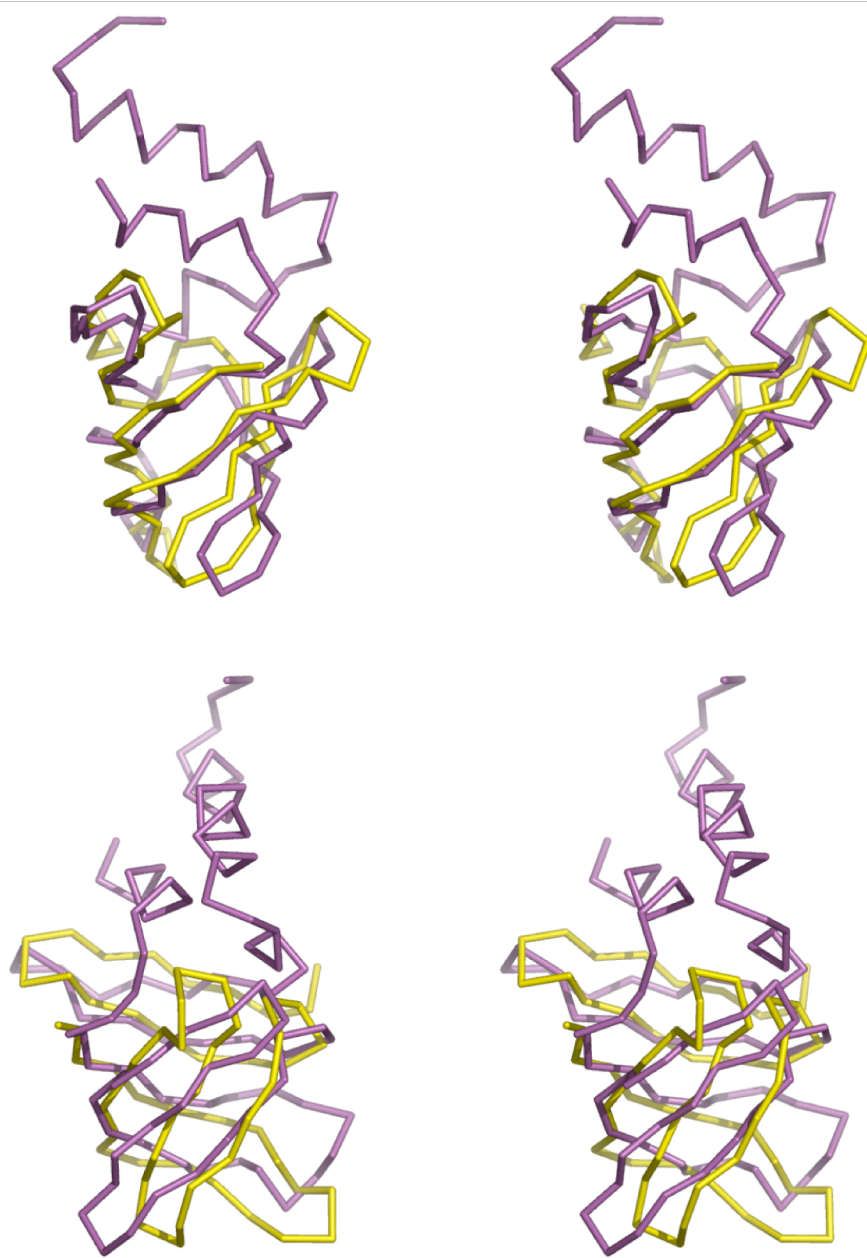


Figure 3.15. Stereo diagram showing two orthogonal views of the *A. fulgidus* aRpp29 average NMR structure superimposed on the crystal structure. The backbone atoms of NMR model (yellow), consisting of residues 17-77, were superimposed on the backbone atoms of the crystal structure (pink) with an r.m.s.d. of 3.7 Å.

Chapter 4 Biophysical analysis of the termini of aRpp29

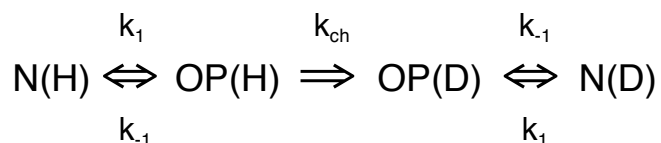
4.1 Introduction

The NMR and X-ray structures of aRpp29 provide two distinct views of the molecule. The NMR structure consists of a relatively rigid six-stranded antiparallel β -sheet with flexible termini. The X-ray crystal structure contains the same six-stranded antiparallel β -sheet, but in contrast to the NMR structure, the amino and carboxy termini form well ordered α -helices at the amino and carboxy termini. This raises some interesting questions; 1) If the helices exist in solution, why were they not detectable by NMR? 2) Are the helices in the crystal structure artifacts of the crystallization?

It can be hypothesized that the helices are in equilibrium between an unstructured state, consistent with the NMR, and an ordered state, as is observed in the crystal structure. To test this hypothesis, amide proton exchange, HSQC experiments, and circular dichroism were used to examine the dynamic characteristics of the termini of aRpp29. This chapter of the dissertation was adapted from Sidote *et al.*, 20004.

4.1.1 Amide Protein Exchange

Amide proton exchange can be used to measure the exchange rate of amide protons with the solvent, and thus can provide information on the conformational stability of flexible parts of a protein. Unprotected amide protons can readily exchange with the solvent during conformational changes. An amide that is hydrogen bonded can exchange only a very small fraction of the time, and is usually an indication that those amides are involved in making interactions in stable regions of the protein. Typically, the proton exchange rate can be explained by the “EX2” mechanism proposed by Hvidt and Nielson (1966) which is derived from:



where, N(H) is the protonated form of the protein, OP(H) is the exchange capable locally unfolded form of the protein, OP(D) represents the deuterated locally unfolded form, N(D) represents the deuterated folded form of the protein, and k_{ch} is the rate of amide proton exchange for the exchange capable form of the protein (Lillemoen *et al.*, 1997). Generally, the native state is considered to be the

most abundant form so it can be assumed that $k_{-1} \gg k_1$ and $k_{-1} \gg k_{ch}$. From these assumptions, the following formula can be proposed:

$$k_{HX} = k_{OP}k_{CH}$$

where k_{HX} is the observed rate of exchange, k_{OP} is equilibrium constant for the unfolded protein (Lillemoen et al. 1997). The “EX2” model implies that the rate of exchange increases 10 fold with each increasing pH unit and has been confirmed in previous studies (Lillemoen *et al.*, 1997).

The “EX2” mechanism can be used to study the conformational stability of the helices in aRpp29 by determining K_{OP} from measurements of K_{HX} . The ratio of k_{CH}/k_{HX} is referred to as the protection factor. Using this model, the rates at which the helices are in the open, or unfolded form can be quantitatively calculated by measuring k_{OP} .

4.1.2 Circular dichroism spectroscopy

Circular dichroism (CD) is a type of absorption spectroscopy that can provide information on the structure of biological macromolecules. In addition to determining secondary structure content, CD can be used to monitor protein folding/unfolding, structural changes induced by pH and temperature, and

changes induced by ligand binding. CD experiments require low concentrations of material, require no extensive preparation, are insensitive to molecular weight, and allow measurements in solution.

CD measures the difference between the absorption of right and left handed circularly polarized light by chiral molecules. The difference is always very small and is measured as a function of wavelength, typically in the UV range, from 190 to 300 nm. After passing through the sample, the right and left hand circularly polarized light have different amplitudes and the combination of the two unequal beams become elliptically polarized light. CD measures the ellipticity of the transmitted light.

The peptide backbone is intrinsically asymmetric and always optically active. There are characteristic wavelengths in which the secondary structures of proteins elicit positive and negative ellipticity peaks based on Fasman standard curves for polylysine (Figure 4.1). For example, proteins with helical content will produce positive peaks at 208 and 222 nm and a negative peak at 192 nm, β -sheet will produce a positive peak at 216 nm and a negative peak at 195nm, and random coil will produce only a positive peak at 200nm.

There are a number parameters pertaining to the sample that are important for accurate data. In order to get accurate estimates of the structural content of the protein, it is essential that the concentration of the sample be accurately calculated. The sample should be as homogenous as possible to

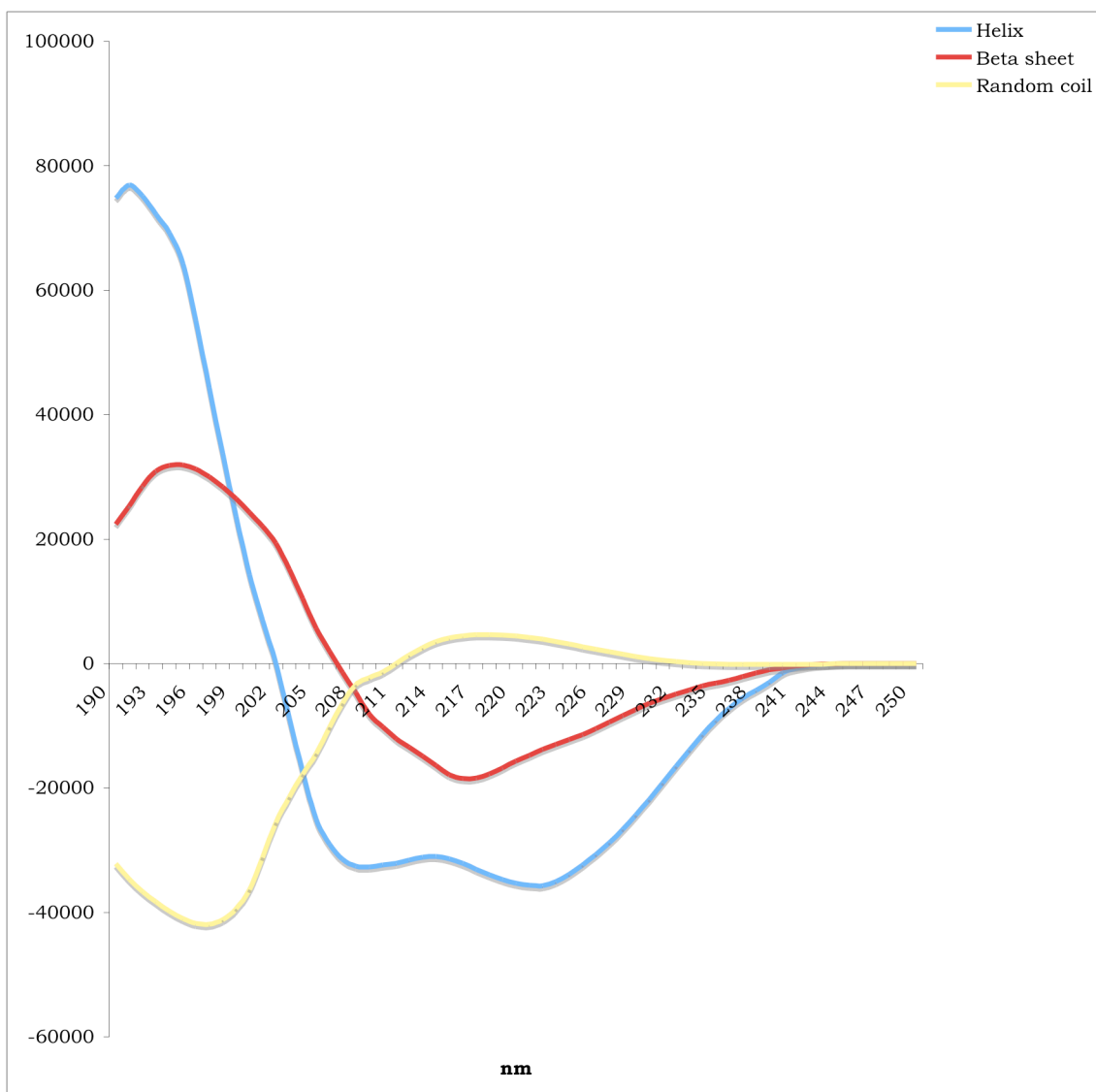


Figure 4.1. Circular dichroism spectra of the Fausman standards in the helical, β -sheet, and random coil conformations. The Fausman standards are consist of poly(Lys) in 3 different conformations. The 3 types of secondary structure each have a unique CD curve. The blue line represents the spectra of in the helical conformation; the red line represents the β -sheet conformation; and the yellow line represents the random coil conformation.

prevent artifacts. The amount of sample required for CD is very low, so to avoid excessive spectral noise, the total absorbance should not exceed 1 unit. The buffer concentration should be as low as possible while still maintaining the proper pH and should be limited to phosphate, borate or low concentration Tris (Martin, 1996). It should also be noted that Cl⁻ ions absorb strongly below 195 nm and should be substituted with fluoride or sulfate salts.

CD data is analyzed by fitting the experimental data curve to a set of standard curves using least squares minimization. The standards are proteins with very high-resolution crystal structures containing different ratios of helix and β -sheet. The best fitting models use data from many different proteins (≥ 20). The accuracy of structure determination depends on the type of secondary structure. The accuracy for helices, β -sheets, and turns is 0.97, 0.75, and 0.50 respectively (Manavalan and Johnson, 1987).

Aside from the inaccuracy of measuring β -sheet and turns, there are other limitations to structure analysis using CD. The shape of the absorption curve can be influenced by aromatic side chains and by the tertiary structure. Another problem arises from the use of reference spectra corresponding to 100 % helix, β -sheet and turns, as they are not directly applicable to proteins that contain short sections of helix or β -sheet. The CD of a helix is known to increase as the length of the helix increases and the CD of β -sheets is sensitive to environment and geometry.

4.2 Methods and Materials

4.2.1 NMR spectroscopy

NMR data were collected using a 500 MHz Varian Inova spectrometer. Spectra were collected at 30°C unless stated otherwise. The ^1H , ^{15}N , and ^{13}C chemical shifts for the nuclei of *A. fulgidus* aRpp29 were previously assigned using two and three-dimensional NMR spectra as described in chapter 2. Two-dimension ^1H - ^{15}N correlated spectra for the present work were collected using sweep widths of 8000 Hz and 5000 Hz in the ^1H and ^{15}N dimensions, respectively. Data were processed using nmrPipe (Delaglio *et al.*, 1995) and analyzed using Sparky (Goddard and Kneller, 2002).

4.2.2 Amide Proton Exchange Rate Measurements

The rates with which the backbone amide protons of the aRpp29 protein exchange with protons from the solvent were measured using a saturation-transfer method, essentially as described by Lillemoen *et al.* (1997).

Uniformly ^{15}N -labeled samples of aRpp29 were equilibrated with buffers of 10mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4/100\text{mM NaCl}$, at pH 5.8, 6.8, 7.8 and 8.8. A pair of ^1H - ^{15}N correlated 2-D NMR spectra was obtained at each pH. In all cases a selective excitation method was used for water suppression. For one spectrum of each pair, the water resonance was pre-saturated for 3 seconds, resulting in

attenuation of the backbone amide ^1H resonances due saturation transfer from the solvent, which depends on the proton exchange rate, as well as a nuclear Overhauser effect (NOE). The NOE was assumed to be constant at each pH, while the proton exchange rate was assumed to increase tenfold with each unit increase in pH (Spera *et al.*, 1991).

By comparing the spectra obtained with and without presaturation at each pH, the effects of solvent exchange and NOE could be determined (Lillemoen *et al.*, 1997). Amide proton exchange rates were normalized to pH 7 for comparison. Peak heights were measured for each assigned residue and fit to the equation: $I_{\text{presat}}/I_o = (1-N)/\{10^{(7-\text{pH})}T_1k+1\}$ where k is the amide proton exchange rate. I_{presat} and I_o are the peak heights in the ^1H - ^{15}N correlated spectra with and without water presaturation, respectively. N represents the NOE contribution. The longitudinal relaxation time T_1 was estimated to be 0.7 s using an inversion recovery experiment.

Alternatively, solvent exchange rates for relatively slowly exchanging amide protons were determined simply by transferring the protein to deuterated solvent and monitoring the disappearance of resonances in 2-D ^1H - ^{15}N correlated spectra, as ^1H is replaced with ^2H . A solution of uniformly ^{15}N labeled Rpp29 was passed through a column of Sephadex G-25 beads (Sigma-Aldrich) that had been pre-equilibrated in buffer (10 mM $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$ pH 5.8/100 mM NaCl) in 99.9% D_2O overnight. The protein was immediately placed in the NMR

spectrometer, and 2-D ^1H - ^{15}N correlated spectra were collected at time intervals of 6 hours over a total of 24 hours, and after 48, 72, and 144 hours. All spectra were processed identically using nmrPipe (Delaglio *et al.*, 1995) and peak heights were measured using Sparky (Goddard and Kneller, 2002). The decrease of the peak height over time was fit to a first-order exponential decay equation. Solvent exchange rates were measured for 58 backbone amide protons; overlapping or partially overlapping peaks were not included in the analysis.

4.2.3 Circular Dichroism Measurements

Circular dichroism data were collected for the aRpp29 protein using an AVIV Circular Dichroism Spectrometer Model 62DS. The concentration of the protein sample was 14.2 μM in 10 mM potassium phosphate pH 7.0 and 100 mM NaF. Data were collected at 25°C using a cell with a path length of 1 mm. Nine scans were averaged to achieve a high degree of signal-to-noise. The parameters of each scan were as follows: wavelength range 260-185 nm; bandwidth 0.5 nm; scan time 2.0 s. The data were analyzed using Selcon3 (Sreerama *et al.*, 2000).

4.2.4 Analytical Ultracentrifugation

Sedimentation data were collected for the *A. fulgidus* aRpp29 protein for the purpose of determining its aggregation state in solution. Data were collected using a Beckman Coulter Optima XL-1 Analytical Centrifuge, operated at 20 °C with a sample concentration of 156 μM. A total of 60 UV absorbance scans were obtained at a rotor speed of 42,000 rpm, with an interval of 10 minutes between scans. The experiment was performed in triplicate, and data were analyzed using Ultrascan (Demeler *et al.*, 1997). The aRpp29 protein was found to have a sedimentation coefficient of 1.25 ± 0.06 S at 20 °C in the same buffer as was used for the NMR experiments (100 mM NaCl, 10 mM K₂HPO₄/KH₂PO₄ pH 5.8); for comparison, hen egg white lysozyme, with 129 amino acids, was found to have a sedimentation coefficient that is slightly greater (1.60 ± 0.04 S) under these same conditions. These sedimentation data are most consistent with aRpp29 being a monomer in solution.

4.3 Results and Discussion

4.3.1 Comparison of the crystal and solution structures of aRpp29

The crystal and NMR structures are similar in that each method describes essentially the same six-stranded antiparallel twisted β-sheet (Figure 2.9, 3.8), and the large majority of NMR-derived inter-proton distances are consistent with

the crystal structure. Indeed, the NMR structure of *A. fulgidus* aRpp29 identified exactly the same set of hydrogen bonds connecting the antiparallel β -strands as is observed in the crystal.

On the other hand, there are immediately obvious differences between the crystal and solution structures (Figure 3.14). Most notably, the well-ordered helices at the N- and C-termini of the protein in the crystal are not observed in solution, even upon re-examination of the NMR data in view of the crystallography results. NMR chemical shift index values for the residues of the C-terminal helix provided a hint of helical structure in solution, however the ^1H - ^1H NOE cross-peaks typical of helical structure were not observed for either the N-terminal or C-terminal helix. The r.m.s.d. between the backbone atoms of residues 18-76 of the crystal and NMR structures of the *A. fulgidus* protein exceeds 3 Å, which is larger than the likely uncertainty in the structures, suggesting that there likely are some real differences in the overall twist of the β -sheet.

Several observed NOE cross-peaks involving loop residues (particularly Ser25) are inconsistent with distances in the crystal structure, indicating there are some differences in the loop structures of the solution and crystalline forms of the protein. In addition, the internal three-way salt bridge between residues Glu40, Lys58, and Arg86 in the crystal (Figure 3.13) was not observed in the solution structures of either archaeal protein, although the salt-bridge between two

residues of this triad (Glu40-Lys58) were detected in the *M. thermoautotrophicus* solution structure (Boomershine *et al.*, 2003).

The apparent differences between the crystal and NMR structures suggest that the crystallization process may lead to a "freezing-out" of a single ordered form of the molecule, and that in solution the helices at the N- and C-termini of the protein may be in equilibrium between folded and unfolded forms. The biophysical tools of NMR and circular dichroism provide additional means to test this hypothesis, and reconcile the apparent differences between the crystal and solution structures of aRpp29. Understanding these differences is important since a simple static model of the structure may not provide an accurate description of the protein surface; a detailed understanding of the protein surface may be required to identify which residues make the essential inter-subunit and enzyme-substrate interactions within RNase P.

4.3.2 Reconciling differences the between the crystal and solution structures of aRpp29.

The large majority of the helical content within the crystal structure of aRpp29 resides in the N- and C-terminal helices; the helical content of the crystal structure is approximately 28%. Circular dichroism (CD) measurements provide information from which the helical content of aRpp29 can be estimated in

solution. CD data from *A. fulgidus* aRpp29 collected at 25 °C were analyzed using the program Selcon3 (Sreerama *et al.*, 2000), indicating a helical content of 7.4% in solution; other methods of CD data analysis indicated a helical content of 5 to 12% (Figure 4.2). The CD data therefore indicate that the helical content of aRpp29 in solution is approximately one-third of that in the crystal. The CD data are consistent with either 1) approximately two-thirds of the helical structure of aRpp29 being unfolded in solution; or 2) the helices of aRpp29 existing in equilibrium between folded and unfolded states, with the equilibrium slightly favoring the unfolded form.

While circular dichroism measurements provide information regarding the average helical content of aRpp29 in solution, measurements of the rates at which backbone amide protons exchange with the solvent can provide insight into how frequently the helices unfold. Amide protons along the protein backbone are only competent to exchange with the solvent when they are not involved in stable hydrogen bonds. Thus, the frequency with which a proton exchanges with solvent sets an upper limit on how frequently a hydrogen bond opens.

The solvent exchange rates for the backbone amide protons of aRpp29 were measured using NMR methods: Slowly exchanging amide protons were monitored by simply observing the decay of ¹H NMR signals after transferring the protein to deuterated solvent, while rates of rapidly exchanging protons were

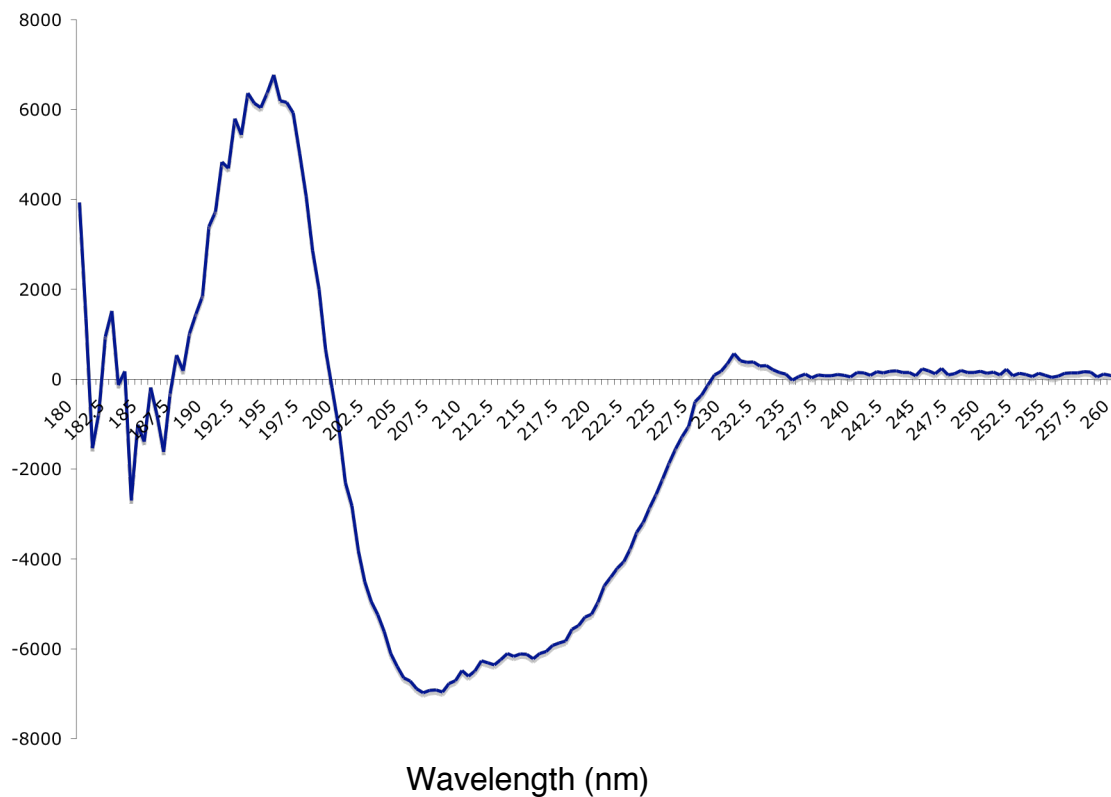


Figure 4.2. Circular dichroism spectra used to determine the helical content of aRpp29 in solution. Data were collected at 25°C in 10 mM potassium phosphate with 100 NaF. Analysis of the data shows that the helical content of aRpp29 is approximately 5 – 12%.

measured using a saturation transfer method (Figure 4.3). Amide proton exchange rates in aRpp29 were found to range over several orders of magnitude. Amide protons within the central regions of the antiparallel β -sheet were found to have exchange rates of less than 10^{-4} sec^{-1} at pH 7, consistent with being in a stable hydrogen bonded structure that opens infrequently. In contrast, the exchange rates for amide protons of the residues of the N- and C-terminal helices are several orders of magnitude more rapid. For example, solvent exchange rates of amide protons of Lys88 and Gly90, located in the middle of the C-terminal helix α_2 in the crystal structure, are approximately 1.0 sec^{-1} , indicating that these hydrogen bonds open much more frequently than those in the β -sheet. For the residues of the N-terminal helix α_1 , the backbone amide protons exchanged with the solvent too rapidly to measure by the saturation transfer method, or were altogether unobservable due to rapid exchange.

4.3.3 The effect of elevated NaCl concentrations on the termini of aRpp29

It is possible that the helices observed in the crystal structure are a result of increased stability due to the elevated salt concentration used in the crystallization of aRpp29. The sample preparation of aRpp29 for the NMR and crystallographic studies were identical. In the case of NMR, a decrease in sensitivity can be observed in some experiments with elevated salt concentration so samples are typically prepared with the minimum salt concentration necessary

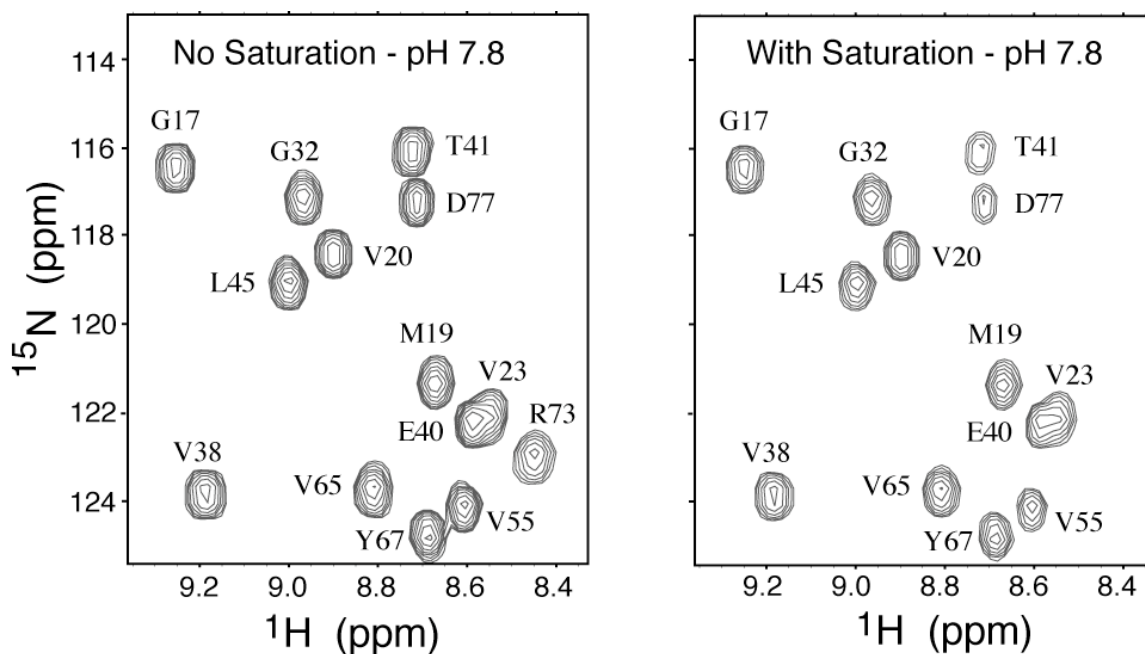


Figure 4.3. Analysis of rates of saturation-transfer between solvent water and amide protons can be used to gain insight into the motions within the Rpp29 protein. Left: A section of the ^{15}N - ^1H correlated NMR spectrum of Rpp29 obtained using a selective excitation method for solvent suppression. Right: A spectrum obtained under the same conditions, but in this case saturating the solvent water resonance. NMR signals from amide protons involved in stable hydrogen bonds are relatively uninfluenced by saturating the solvent resonance. In contrast, signals from amide protons that exchange with protons from the solvent during the saturation period are significantly reduced in intensity; examples are Thr41 and Asp77 which are partially attenuated, and Arg73, in the lower right of the figure, which is completely attenuated when the water resonance is saturated. The influence of the nuclear Overhauser effect is separated from that of amide exchange by performing the experiment at several different values of pH.

to keep the protein folded (100 mM NaCl in the case of aRpp29). The crystallization of aRpp29 was successful only with salt concentrations ranging from 1.1 – 2.2 M ammonium sulfate.

To address this issue, ^1H - ^{15}N HSQC NMR experiments were collected under conditions of increasing concentrations of NaCl. If the helices are stabilized by elevated NaCl concentrations, then additional peaks should appear corresponding to more ordered structures with slower exchange times. NMR spectra were collected with NaCl concentrations of 0.1, 0.25, 0.5, and 1.0 M (Figures 4.4, 4.5, 4.6). Due to the decreased sensitivity of the NMR experiments, the highest NaCl concentration measurable was 1 M. By comparing the spectra of aRpp29 collected at 100 mM to those collected at the elevated NaCl concentrations, it can be concluded that absences of any additional peaks indicate that the elevated NaCl concentrations had little effect on stabilizing the helices.

4.4.4 Conclusion

In summary, CD and NMR data support a model where the N- and C-terminal helices that are clearly observed in the crystal structure of *A. fulgidus*

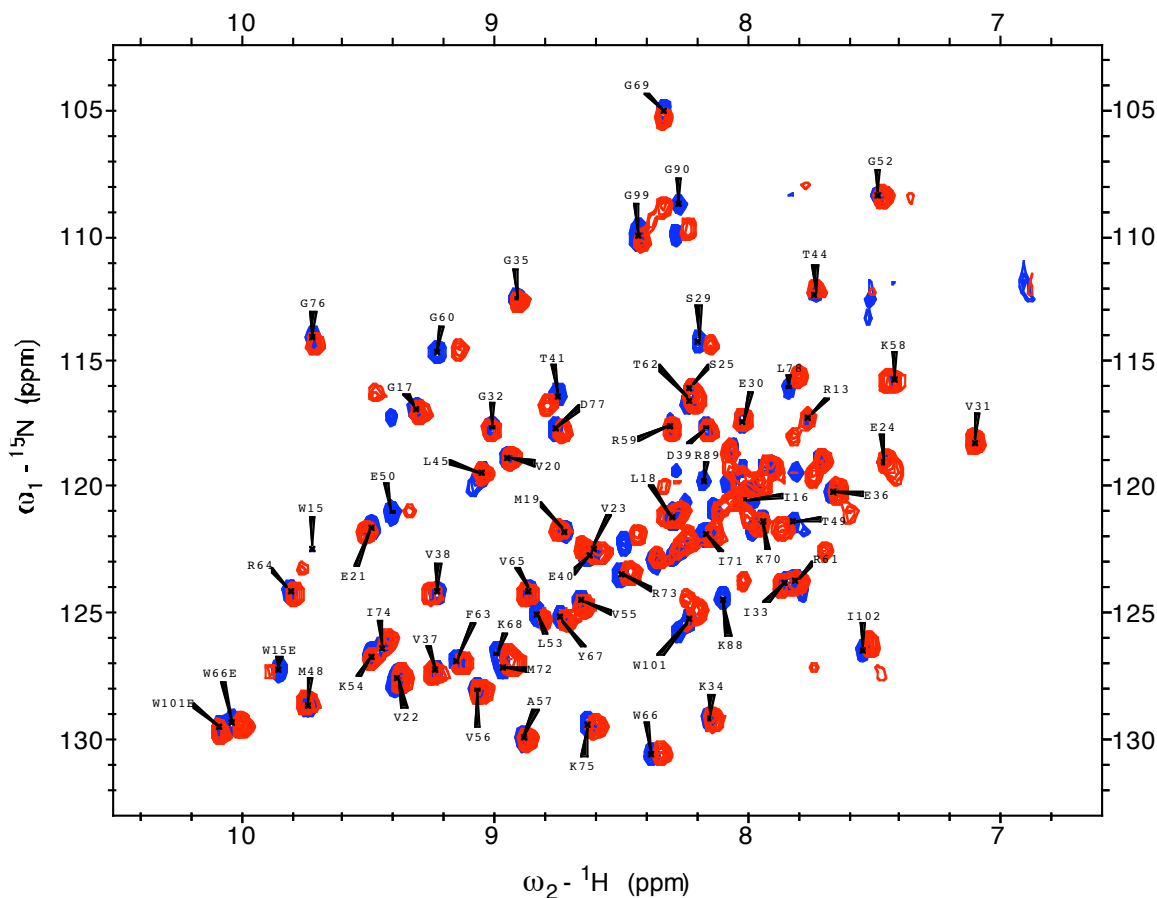


Figure 4.4. ^1H - ^{15}N HSQC spectra of aRpp29 collected at 250 mM NaCl (red) overlaid on a spectrum collected at 100 mM NaCl (blue). Both spectra were collected at 30° and pH 5.8. The samples used to determine the NMR structure of aRpp29 contained 100 mM NaCl. The crystals of aRpp29 using 1.1 – 2.2 M salt as the precipitant and it can be hypothesized that this increased NaCl concentration helps stabilize the N- and C-terminal helices visible in the X-ray crystal structure. To address this issue, spectra were collected with increasing NaCl concentrations. If the NaCl concentration was responsible for stabilizing the helices, additional peaks corresponding to the residues in the helices should become detectable. At 250 mM NaCl, peaks shift due to the change in salt concentration but the number of peaks does not increase.

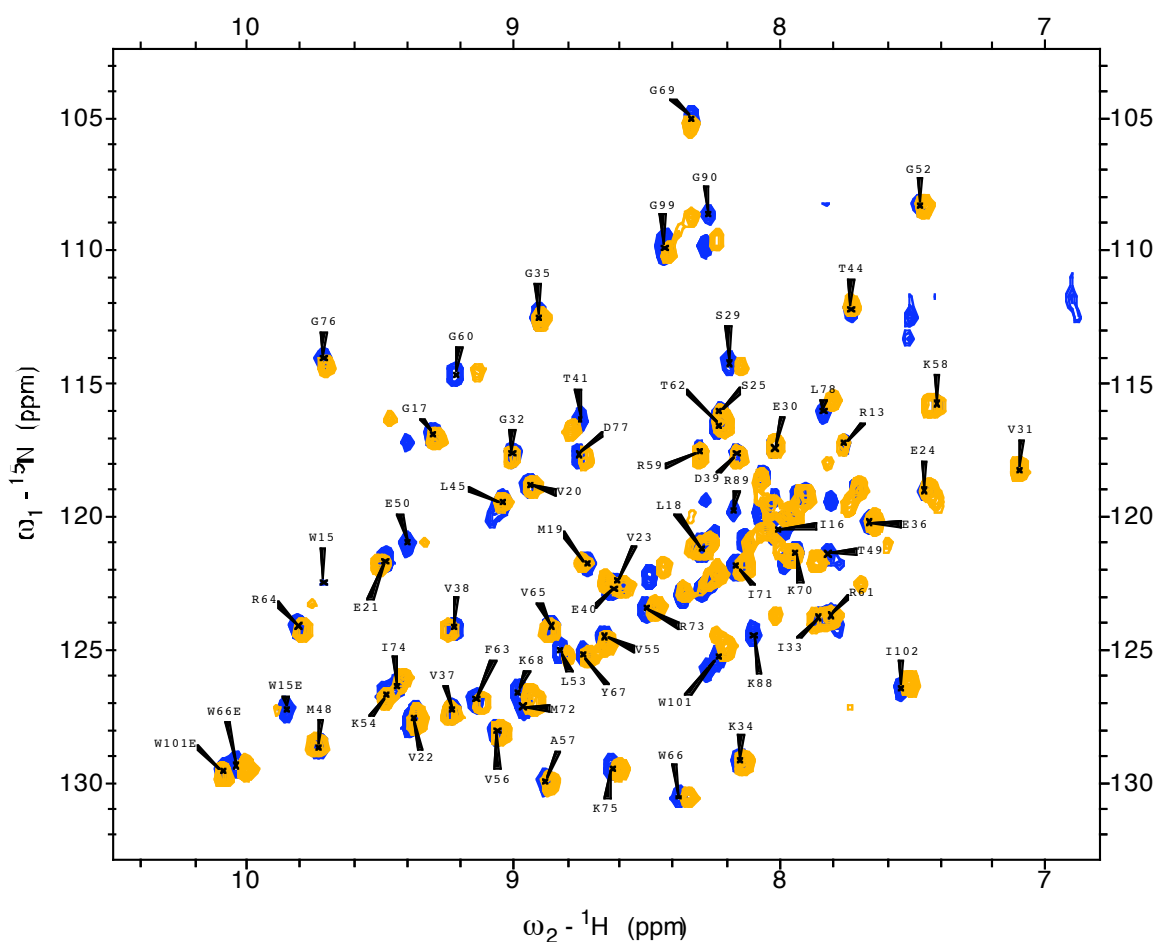


Figure 4.4. ^1H - ^{15}N HSQC spectra of aRpp29 collected at 1 M NaCl (orange) overlaid on a spectrum collected at 100 mM NaCl (blue). Both spectra were collected at 30° and pH 5.8. Increasing the NaCl concentration from 100 mM to 1M again did not affect the number of peaks detectable by NMR. A concentration 1 M is about the limit that can be used in an NMR sample as the sensitivity decreases dramatically when a higher concentration is used. Based on the data collected up to 1 M NaCl, the NaCl concentration does not have an effect on the stability of the N- and C-terminal helices of aRpp29.

aRpp29 are either unfolded in solution, or are in rapid equilibrium between folded and unfolded states. These observations are consistent with the results reported in chapter 2, using measurements of ^{15}N NMR relaxation rates (an alternative method of identifying flexible regions within a protein structure), which indicated that the residues of the β -sheet for *A. fulgidus* aRpp29 are significantly less mobile in solution than those of the termini.

Chapter 5

Summary and conclusions

With the exception of the ribosome, RNase P is the only RNA containing enzyme found in all organisms. RNase P is an integral part of a complicated tRNA maturation process, whereby it cleaves the 5' leader from the pre-tRNA. The composition and complexity of RNase P varies across bacterial, archaeal, and eukaryal organisms, with the common element being a catalytic RNA subunit.

Structural characterization of RNase P has been limited to the bacterial enzyme, which consists of a catalytic RNA and a protein subunit. The structure of the holoenzyme has not been reported, however the structure of the protein subunit and the specificity domain of the RNA component has been reported. The eukaryotic RNase P is a complex ribonucleoprotein, consisting of an RNA subunit and up to 10 protein subunits. To date, no structural data exists for the eukaryal RNase P enzyme. The archaeal RNase P is intermediate in terms of complexity, containing an RNA subunit and at least four protein subunits. For this reason, using archaea is an attractive model system for studying RNase structure and function.

The goal of the structural studies presented in this dissertation was to investigate the structure of the archaeal RNase P protein aRpp29. The NMR and crystal structures presented provide a unique view of aRpp29, indicating that it is dynamic structure and perhaps this characteristic plays an important role in the function or assembly of the RNase P holoenzyme in archaea. This dissertation presents the first atomic structure of a non-bacterial RNase P protein subunit.

Future work on archaeal RNase P will continue towards the structure of the holoenzyme. Because RNase P is a multisubunit enzyme containing an RNA moiety, it is likely that this goal will be very difficult. The relatively low abundance of RNase P makes isolation directly from the source, as was the case for structural studies of the ribosome, very unlikely. Investigations of the individual subunits are underway and so far, in addition to the structures of *A. fulgidus* Rpp29 described in this thesis, there are structures of Rpp29 from *P. horikoshii* and *M. methanohalophilus*. In addition, the structure of the archaeal homolog of Rpp30 has also been solved. The structural analysis of the archaeal homologs of Pop5, Rpr2/Rpp21, and the RNA subunit are in progress.

Appendix 1. Resonance assignments for *A. fulgidus* aRpp29

Atom no.	Residue no.	Residue type	Atom name	Atom type	Chemical shift
1	7	GLY	CA	C	43.2
2	8	VAL	H	H	8.33
3	8	VAL	HA	H	4.03
4	8	VAL	HB	H	1.9
5	8	VAL	HG1	H	0.75
6	8	VAL	HG2	H	0.8
7	8	VAL	C	C	175.4
8	8	VAL	CA	C	62.1
9	8	VAL	CB	C	32.9
10	8	VAL	CG1	C	20.8
11	8	VAL	CG2	C	20.8
12	8	VAL	N	N	119.2
13	9	ALA	H	H	8.35
14	9	ALA	HA	H	4.31
15	9	ALA	HB	H	1.43
16	9	ALA	C	C	177.1
17	9	ALA	CA	C	52.2
18	9	ALA	CB	C	19.3
19	9	ALA	N	N	127.6
20	10	LEU	H	H	8.24
21	10	LEU	HA	H	4.34
22	10	LEU	HB2	H	1.95
23	10	LEU	HB3	H	1.95
24	10	LEU	HG	H	1.71
25	10	LEU	HD1	H	1.32
26	10	LEU	HD2	H	0.76
27	10	LEU	C	C	175.8
28	10	LEU	CA	C	56.5
29	10	LEU	CB	C	42.1
30	10	LEU	N	N	8.24
31	11	ILE	H	H	7.66
32	11	ILE	HA	H	4.08
33	11	ILE	HB	H	2.01
34	11	ILE	C	C	175.2
35	11	ILE	CA	C	62.3
36	11	ILE	CB	C	38.4
37	11	ILE	N	N	114.3
38	12	ALA	H	H	7.66
39	12	ALA	HA	H	4.49
40	12	ALA	HB	H	1.36
41	12	ALA	C	C	175.4
42	12	ALA	CA	C	51.8
43	12	ALA	CB	C	19.6
44	12	ALA	N	N	124.4
45	13	ARG	H	H	7.63
46	13	ARG	HA	H	4.44

47	13	ARG	HB2	H	1.83
48	13	ARG	HB3	H	1.83
49	13	ARG	HG2	H	1.99
50	13	ARG	HG3	H	1.99
51	13	ARG	HD2	H	1.92
52	13	ARG	HD3	H	1.92
53	13	ARG	C	C	174.8
54	13	ARG	CA	C	55.2
55	13	ARG	CB	C	31.9
56	13	ARG	N	N	117.4
57	14	ASP	H	H	8.17
58	14	ASP	HA	H	4.58
59	14	ASP	HB2	H	3.02
60	14	ASP	HB3	H	3.02
61	14	ASP	C	C	176
62	14	ASP	CA	C	54.9
63	14	ASP	CB	C	42.1
64	14	ASP	N	N	120
65	15	TRP	H	H	9.65
66	15	TRP	HA	H	4.62
67	15	TRP	HB2	H	2.86
68	15	TRP	HB3	H	2.86
69	15	TRP	C	C	178.3
70	15	TRP	CA	C	56.4
71	15	TRP	CB	C	29.6
72	16	ILE	H	H	7.93
73	16	ILE	HA	H	3.46
74	16	ILE	HB	H	1.95
75	16	ILE	HG12	H	1.73
76	16	ILE	HG13	H	1.73
77	16	ILE	HG2	H	0.88
78	16	ILE	HD1	H	0.97
79	16	ILE	C	C	177.5
80	16	ILE	CA	C	64.1
81	16	ILE	CB	C	36
82	16	ILE	CG1	C	27.5
83	16	ILE	CG2	C	17.6
84	16	ILE	CD1	C	12.8
85	16	ILE	N	N	119.1
86	17	GLY	H	H	9.27
87	17	GLY	HA2	H	4.46
88	17	GLY	HA3	H	3.86
89	17	GLY	C	C	176.6
90	17	GLY	CA	C	44.8
91	17	GLY	N	N	116.6
92	18	LEU	H	H	8.27
93	18	LEU	HA	H	4.68
94	18	LEU	HB2	H	1.34
95	18	LEU	C	C	175.4
96	18	LEU	CA	C	54.3

97	18	LEU	CB	C	41.8
98	18	LEU	N	N	8.27
99	19	MET	H	H	8.7
100	19	MET	HA	H	5.04
101	19	MET	HB2	H	1.96
102	19	MET	HB3	H	1.92
103	19	MET	HG2	H	2.19
104	19	MET	HG3	H	2.21
105	19	MET	C	C	174.8
106	19	MET	CA	C	54.2
107	19	MET	CB	C	32.2
108	19	MET	CG	C	31.7
109	19	MET	N	N	121.4
110	20	VAL	H	H	8.94
111	20	VAL	HA	H	5.99
112	20	VAL	HB	H	2.22
113	20	VAL	HG1	H	0.94
114	20	VAL	HG2	H	0.99
115	20	VAL	C	C	173.4
116	20	VAL	CA	C	58.8
117	20	VAL	CB	C	37
118	20	VAL	CG1	C	21.2
119	20	VAL	CG2	C	21.2
120	20	VAL	N	N	118.3
121	21	GLU	H	H	9.45
122	21	GLU	HA	H	5.81
123	21	GLU	HB2	H	2.26
124	21	GLU	HB3	H	2.26
125	21	GLU	HG2	H	1.98
126	21	GLU	HG3	H	2.21
127	21	GLU	C	C	176
128	21	GLU	CA	C	53.6
129	21	GLU	CB	C	37
130	21	GLU	CG	C	36.3
131	21	GLU	N	N	121.2
132	22	VAL	H	H	9.36
133	22	VAL	HA	H	4
134	22	VAL	HB	H	2.69
135	22	VAL	HG1	H	0.6
136	22	VAL	HG2	H	1.06
137	22	VAL	C	C	175.7
138	22	VAL	CA	C	63.3
139	22	VAL	CB	C	30.9
140	22	VAL	CG1	C	21.1
141	22	VAL	CG2	C	21.1
142	22	VAL	N	N	127.5
143	23	VAL	H	H	8.58
144	23	VAL	HA	H	3.66
145	23	VAL	HB	H	2.19
146	23	VAL	HG1	H	0.67

147	23	VAL	HG2	H	0.87
148	23	VAL	C	C	174.8
149	23	VAL	CA	C	61.8
150	23	VAL	CB	C	33
151	23	VAL	CG1	C	21.5
152	23	VAL	CG2	C	21.5
153	23	VAL	N	N	122.6
154	24	GLU	H	H	7.43
155	24	GLU	HA	H	4.58
156	24	GLU	HB2	H	2.06
157	24	GLU	HB3	H	2.22
158	24	GLU	HG2	H	1.78
159	24	GLU	HG3	H	1.96
160	24	GLU	C	C	174
161	24	GLU	CA	C	56.1
162	24	GLU	CB	C	34
163	24	GLU	CG	C	36.1
164	24	GLU	N	N	118.8
165	25	SER	H	H	8.2
166	25	SER	HA	H	4.82
167	25	SER	HB2	H	3.34
168	25	SER	HB3	H	3.88
169	25	SER	HG	H	6.13
170	25	SER	C	C	171.5
171	25	SER	CA	C	54.8
172	25	SER	CB	C	64.9
173	25	SER	N	N	115.8
174	26	PRO	HA	H	4.53
175	26	PRO	HB2	H	1.95
176	26	PRO	HB3	H	1.84
177	26	PRO	HG2	H	2.03
178	26	PRO	HG3	H	2.33
179	26	PRO	HD2	H	3.71
180	26	PRO	HD3	H	3.71
181	26	PRO	C	C	175.8
182	26	PRO	CA	C	63.8
183	26	PRO	CB	C	31.8
184	26	PRO	CG	C	27.2
185	26	PRO	CD	C	50.8
186	27	ASN	H	H	8.12
187	27	ASN	HA	H	4.87
188	27	ASN	HB2	H	2.48
189	27	ASN	HB3	H	2.74
190	27	ASN	C	C	174.9
191	27	ASN	CA	C	51
192	27	ASN	CB	C	38.1
193	27	ASN	N	N	117.3
194	28	HIS	H	H	9.08
195	28	HIS	HA	H	4.23
196	28	HIS	HB2	H	3.3

197	28	HIS	HB3	H	3.36
198	28	HIS	HD1	H	7.35
199	28	HIS	HD2	H	7.35
200	28	HIS	HE1	H	8.52
201	28	HIS	HE2	H	8.52
202	28	HIS	C	C	176.5
203	28	HIS	CA	C	59.5
204	28	HIS	CB	C	29.2
205	28	HIS	N	N	123.2
206	29	SER	H	H	8.1
207	29	SER	HA	H	4.23
208	29	SER	HB2	H	3.92
209	29	SER	HB3	H	3.92
210	29	SER	C	C	177
211	29	SER	CA	C	60.8
212	29	SER	CB	C	62.9
213	29	SER	N	N	113.8
214	30	GLU	H	H	7.85
215	30	GLU	HA	H	4.14
216	30	GLU	HB2	H	2.07
217	30	GLU	HB3	H	2.07
218	30	GLU	HG2	H	2.25
219	30	GLU	HG3	H	2.25
220	30	GLU	C	C	176.4
221	30	GLU	CA	C	56.1
222	30	GLU	CB	C	31.2
223	30	GLU	CG	C	36.4
224	30	GLU	N	N	118.9
225	31	VAL	H	H	7.06
226	31	VAL	HA	H	3.2
227	31	VAL	HB	H	1.96
228	31	VAL	HG1	H	0.62
229	31	VAL	HG2	H	0.85
230	31	VAL	C	C	176.7
231	31	VAL	CA	C	65.7
232	31	VAL	CB	C	30.5
233	31	VAL	CG1	C	21.3
234	31	VAL	CG2	C	21.3
235	31	VAL	N	N	117.9
236	32	GLY	H	H	8.93
237	32	GLY	HA2	H	4.54
238	32	GLY	HA3	H	3.78
239	32	GLY	C	C	175.1
240	32	GLY	CA	C	44.5
241	32	GLY	N	N	117.2
242	33	ILE	H	H	7.8
243	33	ILE	HA	H	4.02
244	33	ILE	HB	H	1.87
245	33	ILE	HG12	H	1.1
246	33	ILE	HG13	H	0.95

247	33	ILE	HG2	H	0.78
248	33	ILE	C	C	174.2
249	33	ILE	CA	C	64.1
250	33	ILE	CB	C	37.9
251	33	ILE	CG1	C	27.3
252	33	ILE	CG2	C	17.1
253	33	ILE	N	N	123.4
254	34	LYS	H	H	8.13
255	34	LYS	HA	H	5.75
256	34	LYS	HB2	H	1.75
257	34	LYS	HB3	H	1.75
258	34	LYS	HG2	H	1.21
259	34	LYS	HG3	H	1.36
260	34	LYS	HD2	H	1.56
261	34	LYS	HD3	H	1.56
262	34	LYS	C	C	174.2
263	34	LYS	CA	C	54.5
264	34	LYS	CB	C	36.7
265	34	LYS	CG	C	27.4
266	34	LYS	CD	C	30
267	34	LYS	N	N	128.7
268	35	GLY	H	H	8.88
269	35	GLY	HA2	H	4.43
270	35	GLY	HA3	H	3.86
271	35	GLY	C	C	170
272	35	GLY	CA	C	45.6
273	35	GLY	N	N	112.2
274	36	GLU	H	H	7.6
275	36	GLU	HA	H	4.89
276	36	GLU	HB2	H	1.8
277	36	GLU	HB3	H	2.78
278	36	GLU	HG2	H	1.95
279	36	GLU	HG3	H	2.35
280	36	GLU	C	C	171.8
281	36	GLU	CA	C	54.8
282	36	GLU	CB	C	32.6
283	36	GLU	CG	C	33.9
284	36	GLU	N	N	120
285	37	VAL	H	H	9.16
286	37	VAL	HA	H	4.27
287	37	VAL	HB	H	2.81
288	37	VAL	HG1	H	24.2
289	37	VAL	HG2	H	24.2
290	37	VAL	C	C	176.9
291	37	VAL	CA	C	64.2
292	37	VAL	CB	C	31.2
293	37	VAL	CG1	C	24.2
294	37	VAL	CG2	C	24.2
295	37	VAL	N	N	126.9
296	38	VAL	H	H	9.17

297	38	VAL	HA	H	4.65
298	38	VAL	HB	H	2.31
299	38	VAL	HG1	H	0.81
300	38	VAL	HG2	H	1
301	38	VAL	C	C	174.7
302	38	VAL	CA	C	61.4
303	38	VAL	CB	C	33.1
304	38	VAL	CG1	C	21.2
305	38	VAL	CG2	C	21.2
306	38	VAL	N	N	124.1
307	39	ASP	H	H	8.04
308	39	ASP	HA	H	4.77
309	39	ASP	HB2	H	2.25
310	39	ASP	HB3	H	2.84
311	39	ASP	C	C	175.7
312	39	ASP	CA	C	53.8
313	39	ASP	CB	C	43.8
314	39	ASP	N	N	118.4
315	40	GLU	H	H	8.58
316	40	GLU	HA	H	4.67
317	40	GLU	HB2	H	1.73
318	40	GLU	HB3	H	1.9
319	40	GLU	HG2	H	2.05
320	40	GLU	HG3	H	2.19
321	40	GLU	C	C	173.4
322	40	GLU	CA	C	55.9
323	40	GLU	CB	C	33.1
324	40	GLU	CG	C	35.3
325	40	GLU	N	N	122.4
326	41	THR	H	H	8.67
327	41	THR	HA	H	4.59
328	41	THR	HB	H	4.44
329	41	THR	HG2	H	1.1
330	41	THR	C	C	174.7
331	41	THR	CA	C	59.5
332	41	THR	CB	C	71.4
333	41	THR	CG2	C	21.1
334	41	THR	N	N	115.6
335	42	GLN	H	H	8.91
336	42	GLN	HA	H	4.65
337	42	GLN	HB2	H	2.27
338	42	GLN	HB3	H	2.27
339	42	GLN	HG2	H	2.19
340	42	GLN	HG3	H	2.19
341	42	GLN	C	C	174.9
342	42	GLN	CA	C	62.6
343	42	GLN	CB	C	32.7
344	42	GLN	CG	C	33.9
345	42	GLN	N	N	119
346	43	ASN	H	H	8.32

347	43	ASN	HA	H	5.15
348	43	ASN	HB2	H	2.79
349	43	ASN	HB3	H	2.91
350	43	ASN	C	C	176.1
351	43	ASN	CA	C	53
352	43	ASN	CB	C	41.5
353	43	ASN	N	N	126.8
354	44	THR	H	H	7.66
355	44	THR	HA	H	5.17
356	44	THR	HB	H	3.87
357	44	THR	HG2	H	0.94
358	44	THR	C	C	171.7
359	44	THR	CA	C	60.6
360	44	THR	CB	C	73.6
361	44	THR	CG2	C	24.7
362	44	THR	N	N	112
363	45	LEU	H	H	8.99
364	45	LEU	HA	H	4.68
365	45	LEU	HB2	H	1.16
366	45	LEU	HB3	H	1.53
367	45	LEU	HG	H	0.18
368	45	LEU	HD1	H	0.69
369	45	LEU	HD2	H	0.69
370	45	LEU	C	C	172.5
371	45	LEU	CA	C	54
372	45	LEU	CB	C	46.1
373	45	LEU	CG	C	26.2
374	45	LEU	CD1	C	23.5
375	45	LEU	CD2	C	23.5
376	45	LEU	N	N	119.2
377	46	LYS	H	H	8.21
378	46	LYS	HA	H	5.06
379	46	LYS	HB2	H	1.73
380	46	LYS	HB3	H	1.73
381	46	LYS	HG2	H	1.21
382	46	LYS	HG3	H	1.21
383	46	LYS	HD2	H	1.55
384	46	LYS	HD3	H	1.55
385	46	LYS	C	C	174.1
386	46	LYS	CA	C	55.6
387	46	LYS	CB	C	32.9
388	46	LYS	CG	C	24.2
389	46	LYS	CD	C	30
390	46	LYS	N	N	121
391	47	ILE	H	H	9.31
392	47	ILE	HA	H	4.58
393	47	ILE	HB	H	1.49
394	47	ILE	HG12	H	0.73
395	47	ILE	HG2	H	0.48
396	47	ILE	C	C	174.1

397	47	ILE	CA	C	60.1
398	47	ILE	CB	C	41.7
399	47	ILE	CG1	C	28.8
400	47	ILE	CG2	C	20.5
401	47	ILE	N	N	126.9
402	48	MET	H	H	9.65
403	48	MET	HA	H	4.92
404	48	MET	HB2	H	1.95
405	48	MET	HB3	H	2.78
406	48	MET	HG2	H	1.82
407	48	MET	HG3	H	2.34
408	48	MET	C	C	174.4
409	48	MET	CA	C	54.8
410	48	MET	CB	C	31.4
411	48	MET	CG	C	33.5
412	48	MET	N	N	128.2
413	49	THR	H	H	7.78
414	49	THR	HA	H	5.3
415	49	THR	HB	H	4.57
416	49	THR	HG2	H	1.24
417	49	THR	C	C	175.2
418	49	THR	CA	C	60.3
419	49	THR	CB	C	72.8
420	49	THR	CG2	C	22.2
421	49	THR	N	N	121.1
422	50	GLU	H	H	9.32
423	50	GLU	HA	H	4.27
424	50	GLU	HB2	H	2.17
425	50	GLU	HB3	H	2.17
426	50	GLU	HG2	H	2.41
427	50	GLU	HG3	H	2.41
428	50	GLU	C	C	176.2
429	50	GLU	CA	C	28.8
430	50	GLU	CB	C	29.3
431	50	GLU	CG	C	36.4
432	50	GLU	N	N	120.6
433	51	LYS	H	H	7.98
434	51	LYS	HA	H	4.58
435	51	LYS	HB2	H	2.12
436	51	LYS	HB3	H	2.12
437	51	LYS	HG2	H	1.52
438	51	LYS	HG3	H	1.52
439	51	LYS	HE2	H	3.76
440	51	LYS	HE3	H	3.76
441	51	LYS	C	C	175.1
442	51	LYS	CA	C	54.7
443	51	LYS	CB	C	32.7
444	51	LYS	CG	C	29.2
445	51	LYS	CE	C	42.3
446	51	LYS	N	N	117

447	52	GLY	H	H	7.42
448	52	GLY	HA2	H	4.61
449	52	GLY	HA3	H	3.83
450	52	GLY	C	C	174.4
451	52	GLY	CA	C	44
452	52	GLY	N	N	108
453	53	LEU	H	H	8.76
454	53	LEU	HA	H	4.75
455	53	LEU	HB2	H	1.56
456	53	LEU	HB3	H	1.56
457	53	LEU	HG	H	2.1
458	53	LEU	HD1	H	0.8
459	53	LEU	HD2	H	0.8
460	53	LEU	C	C	176.6
461	53	LEU	CA	C	56
462	53	LEU	CB	C	43.2
463	53	LEU	CG	C	27
464	53	LEU	CD1	C	24.5
465	53	LEU	CD2	C	24.5
466	53	LEU	N	N	124.6
467	54	LYS	H	H	9.43
468	54	LYS	HA	H	4.78
469	54	LYS	HB2	H	2.28
470	54	LYS	HB3	H	2.28
471	54	LYS	HG2	H	1.44
472	54	LYS	HG3	H	1.44
473	54	LYS	HD2	H	1.7
474	54	LYS	HD3	H	1.7
475	54	LYS	C	C	173.7
476	54	LYS	CA	C	53.6
477	54	LYS	CB	C	35.3
478	54	LYS	CG	C	24.9
479	54	LYS	CD	C	29.2
480	54	LYS	N	N	126.3
481	55	VAL	H	H	8.58
482	55	VAL	HA	H	4.75
483	55	VAL	HB	H	1.97
484	55	VAL	HG1	H	0.85
485	55	VAL	HG2	H	0.98
486	55	VAL	C	C	175.2
487	55	VAL	CA	C	62
488	55	VAL	CB	C	32.6
489	55	VAL	CG1	C	21.2
490	55	VAL	CG2	C	21.2
491	55	VAL	N	N	124.1
492	56	VAL	H	H	9
493	56	VAL	HA	H	4.47
494	56	VAL	HB	H	1.88
495	56	VAL	HG1	H	0.91
496	56	VAL	HG2	H	1.15

497	56	VAL	C	C	173.4
498	56	VAL	CA	C	60.3
499	56	VAL	CB	C	35.8
500	56	VAL	CG1	C	23.5
501	56	VAL	CG2	C	23.5
502	56	VAL	N	N	127.6
503	57	ALA	H	H	8.8
504	57	ALA	HA	H	4.45
505	57	ALA	HB	H	1.37
506	57	ALA	C	C	175
507	57	ALA	CA	C	51.9
508	57	ALA	CB	C	18.7
509	57	ALA	N	N	129.5
510	58	LYS	H	H	7.33
511	58	LYS	HA	H	4.14
512	58	LYS	HB2	H	1.94
513	58	LYS	HB3	H	1.94
514	58	LYS	HD2	H	2.01
515	58	LYS	HD3	H	2.01
516	58	LYS	HE2	H	2.3
517	58	LYS	HE3	H	2.3
518	58	LYS	C	C	177.5
519	58	LYS	CA	C	59.2
520	58	LYS	CB	C	34.7
521	58	LYS	CG	C	29.6
522	58	LYS	CE	C	43.9
523	58	LYS	N	N	115.3
524	59	ARG	H	H	8.25
525	59	ARG	HA	H	3.82
526	59	ARG	HB2	H	1.77
527	59	ARG	HB3	H	1.77
528	59	ARG	HG2	H	1.52
529	59	ARG	HG3	H	1.52
530	59	ARG	HD2	H	3.2
531	59	ARG	HD3	H	3.2
532	59	ARG	C	C	177.1
533	59	ARG	CA	C	58.6
534	59	ARG	CB	C	29.8
535	59	ARG	CG	C	27.4
536	59	ARG	CD	C	43.6
537	59	ARG	N	N	116.9
538	60	GLY	H	H	9.24
539	60	GLY	HA2	H	4.1
540	60	GLY	HA3	H	3.75
541	60	GLY	C	C	173.5
542	60	GLY	CA	C	45.6
543	60	GLY	N	N	114.3
544	61	ARG	H	H	7.73
545	61	ARG	HA	H	4.91
546	61	ARG	HB2	H	1.71

547	61	ARG	HB3	H	1.98
548	61	ARG	HG2	H	1.14
549	61	ARG	HG3	H	1.14
550	61	ARG	HD2	H	2.75
551	61	ARG	HD3	H	2.91
552	61	ARG	C	C	175.8
553	61	ARG	CA	C	56.5
554	61	ARG	CB	C	31
555	61	ARG	CG	C	26.9
556	61	ARG	CD	C	44.2
557	61	ARG	N	N	123.3
558	62	THR	H	H	8.19
559	62	THR	HA	H	4.77
560	62	THR	HB	H	3.65
561	62	THR	HG2	H	0.97
562	62	THR	C	C	171.8
563	62	THR	CA	C	62.5
564	62	THR	CB	C	69.5
565	62	THR	CG2	C	21.2
566	62	THR	N	N	116.5
567	63	PHE	H	H	9.15
568	63	PHE	HA	H	5.52
569	63	PHE	HB2	H	2.97
570	63	PHE	HB3	H	3.05
571	63	PHE	HD1	H	7.12
572	63	PHE	HD2	H	7.12
573	63	PHE	HE1	H	7.01
574	63	PHE	HE2	H	7.01
575	63	PHE	C	C	174.4
576	63	PHE	CA	C	56.7
577	63	PHE	CB	C	42.3
578	63	PHE	N	N	126.5
579	64	ARG	H	H	9.75
580	64	ARG	HA	H	5.12
581	64	ARG	HB2	H	1.63
582	64	ARG	HB3	H	1.63
583	64	ARG	HG2	H	1.49
584	64	ARG	HG3	H	1.49
585	64	ARG	HD2	H	2.9
586	64	ARG	HD3	H	2.9
587	64	ARG	C	C	174.1
588	64	ARG	CA	C	55.3
589	64	ARG	CB	C	33.1
590	64	ARG	CG	C	27.4
591	64	ARG	CD	C	44.2
592	64	ARG	N	N	123.8
593	65	VAL	H	H	8.82
594	65	VAL	HA	H	4.88
595	65	VAL	HB	H	1.48
596	65	VAL	HG1	H	0.38

597	65	VAL	HG2	H	0.44
598	65	VAL	C	C	173.6
599	65	VAL	CA	C	60.8
600	65	VAL	CB	C	35.4
601	65	VAL	CG1	C	21.7
602	65	VAL	CG2	C	21.7
603	65	VAL	N	N	123.8
604	66	TRP	H	H	8.34
605	66	TRP	HA	H	4.9
606	66	TRP	HB2	H	3.13
607	66	TRP	HB3	H	3.3
608	66	TRP	HD1	H	7.13
609	66	TRP	HE1	H	10.01
610	66	TRP	HE3	H	7.75
611	66	TRP	HZ2	H	7.36
612	66	TRP	HZ3	H	6.88
613	66	TRP	HH2	H	7.32
614	66	TRP	C	C	174.8
615	66	TRP	CA	C	58.5
616	66	TRP	CB	C	31.2
617	66	TRP	N	N	130.2
618	67	TYR	H	H	8.68
619	67	TYR	HA	H	4.84
620	67	TYR	HB2	H	2.6
621	67	TYR	HB3	H	2.68
622	67	TYR	HD1	H	7.04
623	67	TYR	HD2	H	7.04
624	67	TYR	HE1	H	6.8
625	67	TYR	HE2	H	6.8
626	67	TYR	C	C	173.5
627	67	TYR	CA	C	57.3
628	67	TYR	CB	C	41.2
629	67	TYR	N	N	125
630	68	LYS	H	H	8.94
631	68	LYS	HA	H	3.62
632	68	LYS	HB2	H	1.39
633	68	LYS	HB3	H	1.69
634	68	LYS	HG2	H	0.44
635	68	LYS	HG3	H	0.67
636	68	LYS	HE2	H	2.85
637	68	LYS	HE3	H	2.85
638	68	LYS	C	C	175.6
639	68	LYS	CA	C	57.1
640	68	LYS	CB	C	30.1
641	68	LYS	CG	C	23.9
642	68	LYS	CE	C	42.2
643	68	LYS	N	N	126
644	69	GLY	H	H	8.28
645	69	GLY	HA2	H	4.15
646	69	GLY	HA3	H	3.66

647	69	GLY	C	C	173.2
648	69	GLY	CA	C	45.6
649	69	GLY	N	N	104.9
650	70	LYS	H	H	7.91
651	70	LYS	HA	H	4.61
652	70	LYS	HB2	H	1.84
653	70	LYS	HB3	H	1.84
654	70	LYS	HG2	H	1.41
655	70	LYS	HG3	H	1.41
656	70	LYS	HD2	H	1.65
657	70	LYS	HD3	H	1.65
658	70	LYS	C	C	174.4
659	70	LYS	CA	C	54.3
660	70	LYS	CB	C	35.3
661	70	LYS	CG	C	24.9
662	70	LYS	CD	C	29.2
663	70	LYS	N	N	121
664	71	ILE	H	H	8.13
665	71	ILE	HA	H	4.41
666	71	ILE	HB	H	1.31
667	71	ILE	HG12	H	1.11
668	71	ILE	HG13	H	-0.17
669	71	ILE	HG2	H	0.01
670	71	ILE	HD1	H	0.64
671	71	ILE	C	C	175.6
672	71	ILE	CA	C	60.8
673	71	ILE	CB	C	38.8
674	71	ILE	CG1	C	26.9
675	71	ILE	CG2	C	17.9
676	71	ILE	CD1	C	13.9
677	71	ILE	N	N	121.7
678	72	MET	H	H	8.94
679	72	MET	HA	H	4.69
680	72	MET	HB2	H	1.94
681	72	MET	HB3	H	1.94
682	72	MET	HG2	H	1.72
683	72	MET	HG3	H	1.72
684	72	MET	C	C	173.4
685	72	MET	CA	C	54.2
686	72	MET	CB	C	36
687	72	MET	CG	C	33.3
688	72	MET	N	N	126.7
689	73	ARG	H	H	8.46
690	73	ARG	HA	H	5.22
691	73	ARG	HB2	H	1.2
692	73	ARG	HB3	H	1.22
693	73	ARG	HG2	H	1.37
694	73	ARG	HG3	H	1.64
695	73	ARG	HD2	H	3.07
696	73	ARG	HD3	H	3.21

697	73	ARG	C	C	174.9
698	73	ARG	CA	C	55.2
699	73	ARG	CB	C	32.1
700	73	ARG	CG	C	27.4
701	73	ARG	CD	C	43.5
702	73	ARG	N	N	123.3
703	74	ILE	H	H	9.4
704	74	ILE	HA	H	4.63
705	74	ILE	HB	H	1.98
706	74	ILE	HG12	H	1.42
707	74	ILE	HG13	H	1.42
708	74	ILE	HG2	H	1.08
709	74	ILE	C	C	174.5
710	74	ILE	CA	C	59.1
711	74	ILE	CB	C	41.9
712	74	ILE	CG1	C	24.9
713	74	ILE	CG2	C	24.9
714	74	ILE	N	N	126
715	75	LYS	H	H	8.6
716	75	LYS	HA	H	4.66
717	75	LYS	HB2	H	1.75
718	75	LYS	HB3	H	1.75
719	75	LYS	HG2	H	2.04
720	75	LYS	HG3	H	2.04
721	75	LYS	HD2	H	1.63
722	75	LYS	HD3	H	1.63
723	75	LYS	C	C	176.8
724	75	LYS	CA	C	56.9
725	75	LYS	CB	C	32.9
726	75	LYS	CG	C	24.6
727	75	LYS	CD	C	29.2
728	75	LYS	N	N	129
729	76	GLY	H	H	9.63
730	76	GLY	HA2	H	4.7
731	76	GLY	HA3	H	3.7
732	76	GLY	C	C	175.3
733	76	GLY	CA	C	47.2
734	76	GLY	N	N	113.4
735	77	ASP	H	H	8.71
736	77	ASP	HA	H	4.26
737	77	ASP	HB2	H	2.43
738	77	ASP	HB3	H	2.7
739	77	ASP	C	C	176.7
740	77	ASP	CA	C	56.6
741	77	ASP	CB	C	40.8
742	77	ASP	N	N	117.1
743	78	LEU	H	H	7.83
744	78	LEU	HA	H	4.33
745	78	LEU	HB2	H	1.94
746	78	LEU	HB3	H	1.94

747	78	LEU	HG	H	1.74
748	78	LEU	HD1	H	1.1
749	78	LEU	HD2	H	1.1
750	78	LEU	C	C	175.8
751	78	LEU	CA	C	56.3
752	78	LEU	CB	C	42.5
753	78	LEU	CG	C	27.2
754	78	LEU	CD1	C	25.2
755	78	LEU	CD2	C	25.2
756	78	LEU	N	N	116.2
757	85	ASP	H	H	8.29
758	85	ASP	HA	H	4.34
759	85	ASP	HB2	H	2.06
760	85	ASP	HB3	H	2.24
761	85	ASP	C	C	175.8
762	85	ASP	CA	C	55.3
763	85	ASP	CB	C	42.5
764	85	ASP	N	N	118.7
765	86	ARG	H	H	7.81
766	86	ARG	HA	H	4.27
767	86	ARG	HB2	H	1.94
768	86	ARG	HB3	H	1.94
769	86	ARG	HG2	H	2.1
770	86	ARG	HG3	H	2.1
771	86	ARG	C	C	176.3
772	86	ARG	CA	C	57.3
773	86	ARG	CB	C	30.8
774	86	ARG	CG	C	27.3
775	86	ARG	N	N	125.1
776	87	ILE	H	H	7.8
777	87	ILE	HA	H	4.07
778	87	ILE	HB	H	1.76
779	87	ILE	HG12	H	1.94
780	87	ILE	HG13	H	1.94
781	87	ILE	HG2	H	1.36
782	87	ILE	HD1	H	0.81
783	87	ILE	C	C	175.9
784	87	ILE	CA	C	62.2
785	87	ILE	CB	C	38.4
786	87	ILE	CG1	C	26.4
787	87	ILE	CG2	C	26.4
788	87	ILE	CD1	C	13.4
789	87	ILE	N	N	119.8
790	88	LYS	H	H	8.09
791	88	LYS	HA	H	4.28
792	88	LYS	HB2	H	1.85
793	88	LYS	HB3	H	1.85
794	88	LYS	HG2	H	1.44
795	88	LYS	HG3	H	1.44
796	88	LYS	HD2	H	1.67

797	88	LYS	HD3	H	1.67
798	88	LYS	HE2	H	2.99
799	88	LYS	HE3	H	2.99
800	88	LYS	C	C	174.5
801	88	LYS	CA	C	56.7
802	88	LYS	CB	C	32.9
803	88	LYS	CG	C	25.4
804	88	LYS	CD	C	29.6
805	88	LYS	CE	C	42.3
806	88	LYS	N	N	123.8
807	89	ARG	H	H	8.13
808	89	ARG	HA	H	4.33
809	89	ARG	HB2	H	1.8
810	89	ARG	HB3	H	1.8
811	89	ARG	HG2	H	1.53
812	89	ARG	HG3	H	1.53
813	89	ARG	C	C	177.5
814	89	ARG	CA	C	56.5
815	89	ARG	CB	C	30.9
816	89	ARG	CG	C	27.9
817	89	ARG	N	N	120.6
818	90	GLY	H	H	8.24
819	90	GLY	HA2	H	3.96
820	90	GLY	HA3	H	3.96
821	90	GLY	C	C	176.8
822	90	GLY	CA	C	45.5
823	90	GLY	N	N	108.8
824	91	LEU	H	H	7.99
825	91	LEU	HA	H	4.6
826	91	LEU	HB2	H	1.84
827	91	LEU	HB3	H	1.84
828	91	LEU	HG	H	1.41
829	91	LEU	HD1	H	1.66
830	91	LEU	HD2	H	1.66
831	91	LEU	C	C	176.3
832	91	LEU	CA	C	55.3
833	91	LEU	CB	C	42.6
834	91	LEU	CG	C	27.2
835	91	LEU	CD1	C	24.2
836	91	LEU	CD2	C	24.2
837	91	LEU	N	N	121
838	92	MET	H	H	7.82
839	92	MET	HA	H	4.27
840	92	MET	HB2	H	1.95
841	92	MET	HB3	H	1.95
842	92	MET	HG2	H	2.11
843	92	MET	HG3	H	2.53
844	92	MET	C	C	180.1
845	92	MET	CA	C	57.4
846	92	MET	CB	C	34.3

847	92	MET	CG	C	33.4
848	92	MET	N	N	124.8
849	93	MET	C	C	176.2
850	93	MET	CA	C	56
851	93	MET	CB	C	31.1
852	94	LEU	H	H	7.9
853	94	LEU	HA	H	4.77
854	94	LEU	C	C	172.2
855	94	LEU	CA	C	53.1
856	94	LEU	CB	C	42.7
857	94	LEU	N	N	116.9
858	95	LYS	H	H	8.61
859	95	LYS	HA	H	4.78
860	95	LYS	HB2	H	1.73
861	95	LYS	HB3	H	1.73
862	95	LYS	C	C	175.2
863	95	LYS	CA	C	55.6
864	95	LYS	CB	C	32.7
865	95	LYS	N	N	122.4
866	96	ARG	H	H	8.21
867	96	ARG	HA	H	4.24
868	96	ARG	HB2	H	1.78
869	96	ARG	HB3	H	1.78
870	96	ARG	C	C	175.9
871	96	ARG	CA	C	56.7
872	96	ARG	CB	C	30.7
873	96	ARG	N	N	122
874	97	ALA	H	H	8.05
875	97	ALA	HA	H	4.34
876	97	ALA	HB	H	1.44
877	97	ALA	C	C	177
878	97	ALA	CA	C	52.9
879	97	ALA	CB	C	19.2
880	97	ALA	N	N	123.8
881	98	LYS	H	H	8.11
882	98	LYS	HA	H	4.22
883	98	LYS	C	C	176.2
884	98	LYS	CA	C	56.6
885	98	LYS	CB	C	33.1
886	98	LYS	N	N	119.3
887	99	GLY	H	H	8.38
888	99	GLY	HA2	H	3.96
889	99	GLY	HA3	H	3.95
890	99	GLY	C	C	172.9
891	99	GLY	CA	C	45.2
892	99	GLY	N	N	110.8
893	100	VAL	H	H	7.72
894	100	VAL	HA	H	4.07
895	100	VAL	HB	H	1.94
896	100	VAL	HG1	H	0.81

897	100	VAL	HG2	H	0.81
898	100	VAL	C	C	174.8
899	100	VAL	CA	C	62.2
900	100	VAL	CB	C	32.9
901	100	VAL	CG1	C	21.4
902	100	VAL	CG2	C	21.4
903	100	VAL	N	N	118.9
904	101	TRP	H	H	8.2
905	101	TRP	HA	H	4.78
906	101	TRP	HB2	H	3.19
907	101	TRP	HB3	H	3.35
908	101	TRP	HD1	H	7.23
909	101	TRP	HE1	H	10.06
910	101	TRP	HE3	H	7.68
911	101	TRP	HZ2	H	7.47
912	101	TRP	HZ3	H	7.15
913	101	TRP	HH2	H	7.21
914	101	TRP	C	C	174.2
915	101	TRP	CA	C	56.9
916	101	TRP	CB	C	30
917	101	TRP	N	N	124.8
918	102	ILE	H	H	7.51
919	102	ILE	HA	H	4.22
920	102	ILE	HB	H	1.8
921	102	ILE	HG12	H	1.39
922	102	ILE	HG13	H	1.39
923	102	ILE	HG2	H	1.09
924	102	ILE	HD1	H	0.85
925	102	ILE	C	C	174.6
926	102	ILE	CA	C	62.8
927	102	ILE	CB	C	40
928	102	ILE	CG1	C	28.1
929	102	ILE	CG2	C	15.8
930	102	ILE	CD1	C	12.4
931	102	ILE	N	N	126.1

Appendix 2. NOE assignment table for *A. fulgidus* aRpp29

NOE cross peaks observed in 2 dimensional NOESY experiments with 60 ms mixing times. Peak intensities converted to distance restraints using peak height as determined using the program sparky

distance restraint	

strong peak	2.8 Å {1}
medium peak	3.2 Å {2}
weak peak	3.8 Å {3}
very weak peak	4.2 Å {4}

NOE cross peaks observed in 3 dimensional NOESY experiments with 60ms mixing time. Peak intensities converted to distance restraints using peak height as determined using the program Sparky

distance restraint	

strong peak	5.0 Å {5}
medium peak	5.5 Å {6}
weak peak	6.0 Å {7}
very weak	6.5 Å {8}

NOE cross peaks observed in 2 dimensional NOESY experiments with 200 ms mixing time have the following distance restraints:

6.9 Å {9}

i = intraresidue
s = sequential
l = long range

{A9}	
assign (resid 9 and name hn) (resid 9 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 9 and name hn) (resid 9 and name hb*)	6.0 6.0 1.0 {7i}
{L10}	
assign (resid 10 and name hn) (resid 10 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 10 and name hn) (resid 9 and name ha)	6.0 6.0 0.0 {7s}

{I11}	
assign (resid 11 and name hn) (resid 11 and name ha)	5.5 5.5 0.0 {6i}
assign (resid 11 and name hn) (resid 11 and name hb)	5.5 5.5 0.0 {6i}
assign (resid 11 and name hn) (resid 14 and name hn)	6.5 6.5 0.0 {8l}
{A12}	
assign (resid 12 and name hn) (resid 12 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 12 and name hn) (resid 12 and name hb*)	5.5 5.5 1.0 {6i}
assign (resid 12 and name hn) (resid 11 and name ha)	5.5 5.5 0.0 {6s}
assign (resid 12 and name hn) (resid 11 and name hb)	6.5 6.5 0.0 {8s}
{R13}	
assign (resid 13 and name hn) (resid 13 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 13 and name hn) (resid 13 and name hb*)	6.0 6.0 0.7 {7i}
assign (resid 13 and name hn) (resid 12 and name ha)	5.5 5.5 0.0 {6s}
assign (resid 13 and name hn) (resid 12 and name hb*)	5.5 5.5 1.0 {6s}
{D14}	
assign (resid 14 and name hn) (resid 14 and name ha)	6.9 6.9 0.0 {9i}
assign (resid 14 and name hn) (resid 14 and name hb*)	6.9 6.9 0.7 {9i}
{W15}	
assign (resid 15 and name hn) (resid 15 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 15 and name hn) (resid 15 and name hb*)	6.5 6.5 0.7 {8i}
{I16}	
assign (resid 16 and name hn) (resid 16 and name ha)	4.2 4.2 0.0 {4i}
assign (resid 16 and name hn) (resid 16 and name hb)	3.8 3.8 0.0 {4i}
assign (resid 16 and name hn) (resid 16 and name hd1)	4.2 4.2 0.0 {4i}
assign (resid 16 and name hn) (resid 16 and name hg12)	6.5 6.5 1.0 {8i}
assign (resid 16 and name hn) (resid 16 and name hg13)	6.5 6.5 1.0 {8i}
assign (resid 16 and name ha) (resid 15 and name hb*)	6.5 6.5 0.7 {8l}
{G17}	
assign (resid 17 and name hn) (resid 17 and name ha*)	3.8 3.8 0.7 {3i}
assign (resid 17 and name hn) (resid 16 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 17 and name hn) (resid 16 and name hb)	6.0 6.0 0.0 {7s}
assign (resid 17 and name hn) (resid 16 and name hg2)	3.8 3.8 0.0 {3s}
assign (resid 17 and name hn) (resid 16 and name hd1)	5.5 5.5 0.0 {6s}
assign (resid 17 and name hn) (resid 18 and name hn)	3.8 3.8 0.0 {3l}
{L18}	
assign (resid 18 and name hn) (resid 18 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 18 and name hn) (resid 18 and name hb*)	6.0 6.0 0.0 {7i}

assign (resid 18 and name hn) (resid 17 and name ha*)	6.0 6.0 0.0 {7i}
assign (resid 18 and name hn) (resid 37 and name hg**)	6.0 6.0 2.5 {7j}
assign (resid 18 and name hn) (resid 37 and name hn)	6.5 6.5 0.0 {8j}
{M19}	
assign (resid 19 and name hn) (resid 19 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 19 and name hn) (resid 19 and name hb*)	3.8 3.8 0.7 {3i}
assign (resid 19 and name hn) (resid 19 and name hg*)	4.2 4.2 0.7 {4i}
assign (resid 19 and name hn) (resid 18 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 19 and name hn) (resid 18 and name hb*)	4.2 4.2 0.7 {4s}
assign (resid 19 and name ha) (resid 36 and name ha)	6.9 6.9 0.0 {9j}
assign (resid 19 and name ha) (resid 20 and name hg**)	6.5 6.5 2.5 {8s}
assign (resid 19 and name hb*)(resid 66 and name he3)	6.5 6.5 0.0 {8j}
assign (resid 19 and name hn) (resid 67 and name ha)	3.8 3.8 0.0 {3j}
assign (resid 19 and name hn) (resid 66 and name hb*)	6.5 6.5 0.7 {8j}
assign (resid 19 and name hn) (resid 66 and name hn)	6.5 6.5 0.0 {8j}
assign (resid 19 and name hn) (resid 20 and name hg**)	4.2 4.2 2.5 {4j}
{V20}	
assign (resid 20 and name hn) (resid 20 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 20 and name hn) (resid 20 and name hb)	3.8 3.8 0.0 {3i}
assign (resid 20 and name hn) (resid 20 and name hg**)	4.2 4.2 0.0 {4i}
assign (resid 20 and name hn) (resid 19 and name ha)	2.8 2.8 0.0 {1s}
assign (resid 20 and name hn) (resid 19 and name hb*)	6.0 6.0 0.7 {7s}
assign (resid 20 and name ha) (resid 65 and name ha)	6.9 6.9 0.0 {9j}
assign (resid 20 and name ha) (resid 66 and name he3)	6.5 6.5 0.0 {8j}
assign (resid 20 and name hb) (resid 63 and name hd*)	6.5 6.5 2.4 {8j}
assign (resid 20 and name hg**) (resid 63 and name hd*)	6.5 6.5 2.4 {8j}
assign (resid 20 and name hn) (resid 35 and name hn)	6.5 6.5 0.0 {8j}
assign (resid 20 and name hn) (resid 36 and name ha)	6.0 6.0 0.0 {7j}
assign (resid 20 and name hn) (resid 66 and name he3)	3.8 3.8 0.0 {3j}
assign (resid 20 and name hn) (resid 66 and name hz3)	3.8 3.8 0.0 {3j}
{E21}	
assign (resid 21 and name hn) (resid 21 and name ha)	4.2 4.2 0.0 {4i}
assign (resid 21 and name hn) (resid 21 and name hb*)	3.8 3.8 0.7 {3i}
assign (resid 21 and name hn) (resid 21 and name hg*)	3.8 3.8 0.7 {3i}
assign (resid 21 and name hn) (resid 20 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 21 and name hn) (resid 20 and name hb)	5.0 5.0 0.0 {5s}
assign (resid 21 and name hn) (resid 20 and name hg**)	6.0 6.0 2.5 {7s}

assign (resid 21 and name ha) (resid 34 and name ha)	6.9 6.9 0.0 {9l}
assign (resid 21 and name ha) (resid 34 and name hb*)	6.0 6.0 0.7 {7l}
assign (resid 21 and name ha) (resid 22 and name hg**)	6.5 6.5 2.5 {8l}
assign (resid 21 and name ha) (resid 23 and name hg**)	6.5 6.5 2.5 {8l}
assign (resid 21 and name hb*)(resid 23 and name hg**)	6.5 6.5 2.5 {8l}
assign (resid 21 and name hn) (resid 64 and name hn)	5.5 5.5 0.0 {6l}
assign (resid 21 and name hn) (resid 65 and name ha)	3.8 3.8 0.0 {3l}
assign (resid 21 and name hn) (resid 66 and name he3)	4.2 4.2 0.0 {4l}
assign (resid 21 and name hn) (resid 66 and name hz3)	3.8 3.8 0.0 {3l}
{V22}	
assign (resid 22 and name hn) (resid 22 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 22 and name hn) (resid 22 and name hb)	3.8 3.8 0.0 {3i}
assign (resid 22 and name hn) (resid 22 and name hg**)	3.8 3.8 2.5 {3i}
assign (resid 22 and name hn) (resid 21 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 22 and name hn) (resid 21 and name hb*)	6.0 6.0 0.7 {7s}
assign (resid 22 and name hn) (resid 21 and name hg*)	6.0 6.0 0.7 {7s}
assign (resid 22 and name ha) (resid 63 and name ha)	6.5 6.5 0.0 {8l}
assign (resid 22 and name hg**)(resid 33 and name ha)	6.5 6.5 2.5 {8l}
assign (resid 22 and name hn) (resid 33 and name hn)	6.5 6.5 0.0 {8l}
assign (resid 22 and name hn) (resid 34 and name ha)	6.5 6.5 0.0 {8l}
assign (resid 22 and name hn) (resid 34 and name hb*)	6.5 6.5 0.7 {8l}
assign (resid 22 and name hn) (resid 63 and name hd*)	6.5 6.5 2.4 {8l}
assign (resid 22 and name ha) (resid 64 and name hn)	6.5 6.5 0.0 {8l}
{v23}	
assign (resid 23 and name hn) (resid 23 and name ha)	6.5 6.5 0.0 {8i}
assign (resid 23 and name hn) (resid 23 and name hb)	6.0 6.0 0.0 {7i}
assign (resid 23 and name hn) (resid 23 and name hg**)	3.8 3.8 2.5 {3i}
assign (resid 23 and name hn) (resid 22 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 23 and name hn) (resid 22 and name hg**)	3.8 3.8 2.5 {3s}
assign (resid 23 and name hn) (resid 63 and name ha)	3.8 3.8 0.0 {3l}
assign (resid 23 and name hn) (resid 24 and name hn)	6.5 6.5 0.0 {8l}
assign (resid 23 and name hb) (resid 24 and name hn)	6.5 6.5 0.0 {8l}
{E24}	
assign (resid 24 and name hn) (resid 24 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 24 and name hn) (resid 24 and name hb*)	4.2 4.2 0.7 {4i}
assign (resid 24 and name hn) (resid 24 and name hg*)	4.2 4.2 0.7 {4i}
assign (resid 24 and name hn) (resid 23 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 24 and name hn) (resid 23 and name hg**)	3.8 3.8 2.5 {3s}
assign (resid 24 and name hb*)(resid 23 and name ha)	6.5 6.5 0.7 {8s}
assign (resid 24 and name hg*)(resid 23 and name ha)	6.0 6.0 0.4 {7s}

assign (resid 24 and name hn) (resid 27 and name hb*)	6.5 6.5 0.7 {8l}
assign (resid 24 and name hn) (resid 22 and name hg**)	4.2 4.2 2.5 {4l}
assign (resid 24 and name hn) (resid 32 and name ha*)	6.5 6.5 0.7 {8l}
assign (resid 24 and name hn) (resid 63 and name ha)	6.5 6.5 0.0 {8l}
assign (resid 24 and name hn) (resid 61 and name ha)	6.5 6.5 0.0 {8l}
{S25}	
assign (resid 25 and name hn) (resid 25 and name ha)	3.2 3.2 0.0 {2i}
assign (resid 25 and name hn) (resid 25 and name hb*)	4.2 4.2 0.7 {4i}
assign (resid 25 and name hn) (resid 24 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 25 and name hn) (resid 24 and name hb*)	6.0 6.0 0.7 {7s}
assign (resid 25 and name hn) (resid 24 and name hg*)	3.8 3.8 0.7 {3s}
assign (resid 25 and name hb*)(resid 25 and name hg)	6.0 6.0 0.0 {7i}
assign (resid 25 and name hb*)(resid 24 and name hg*)	6.5 6.5 0.0 {8s}
assign (resid 25 and name hb*)(resid 31 and name hg**)	6.0 6.0 2.5 {7l}
assign (resid 25 and name hn) (resid 26 and name hd*)	5.5 5.5 0.7 {6l}
assign (resid 25 and name hn) (resid 22 and name hg**)	6.0 6.0 2.5 {7l}
assign (resid 25 and name hn) (resid 23 and name hg**)	6.0 6.0 2.5 {7l}
{N27}	
assign (resid 27 and name hn) (resid 27 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 27 and name hn) (resid 27 and name hb*)	3.8 3.8 0.0 {3i}
assign (resid 27 and name hn) (resid 26 and name ha)	4.2 4.2 0.0 {4s}
assign (resid 27 and name hn) (resid 26 and name hb*)	6.0 6.0 0.7 {7s}
assign (resid 27 and name hn) (resid 26 and name hg*)	6.5 6.5 0.7 {8s}
assign (resid 27 and name hn) (resid 26 and name hd*)	5.5 5.5 0.7 {6s}
assign (resid 27 and name hn) (resid 25 and name hb*)	6.5 6.5 0.7 {4l}
assign (resid 27 and name hn) (resid 25 and name hg)	6.5 6.5 0.0 {3l}
assign (resid 27 and name hn) (resid 28 and name hd*)	6.5 6.5 0.7 {8l}
{H28}	
assign (resid 28 and name hn) (resid 28 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 28 and name hn) (resid 28 and name hb*)	6.0 6.0 0.0 {7i}
assign (resid 28 and name hn) (resid 27 and name ha)	6.5 6.5 0.0 {8s}
assign (resid 28 and name hn) (resid 27 and name hb*)	6.5 6.5 0.7 {8s}
{S29}	
assign (resid 29 and name hn) (resid 29 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 29 and name hn) (resid 29 and name hb*)	3.2 3.2 0.7 {2i}
assign (resid 29 and name hn) (resid 30 and name hn)	6.0 6.0 0.0 {7l}
{E30}	
assign (resid 30 and name hn) (resid 30 and name ha)	3.8 3.8 0.0 {3i}

assign (resid 30 and name hn) (resid 30 and name hb*)	3.8 3.8 0.7 {3i}
assign (resid 30 and name hn) (resid 30 and name hg*)	3.8 3.8 0.7 {3i}
assign (resid 30 and name hn) (resid 29 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 30 and name hn) (resid 29 and name hb*)	4.2 4.2 0.7 {4s}
assign (resid 30 and name hn) (resid 28 and name hb*)	6.5 6.5 0.7 {8l}
assign (resid 30 and name hn) (resid 25 and name hg)	6.5 6.5 0.0 {8l}
assign (resid 30 and name hn) (resid 31 and name hn)	6.5 6.5 0.0 {8l}
assign (resid 30 and name hn) (resid 31 and name hg**)	4.2 4.2 2.5 {4l}
assign (resid 30 and name ha) (resid 25 and name hg)	6.5 6.5 0.0 {8l}
{V31}	
assign (resid 31 and name hn) (resid 31 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 31 and name hn) (resid 31 and name hb)	3.2 3.2 0.0 {2i}
assign (resid 31 and name hn) (resid 31 and name hg**)	3.2 3.2 2.5 {2i}
assign (resid 31 and name hn) (resid 30 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 31 and name hn) (resid 30 and name hb*)	4.2 4.2 0.7 {4s}
assign (resid 31 and name hn) (resid 30 and name hg*)	6.0 6.0 0.7 {7s}
assign (resid 31 and name ha) (resid 30 and name hn)	6.5 6.5 0.0 {8l}
assign (resid 31 and name hg**)(resid 30 and name ha)	6.0 6.0 2.5 {7l}
{G32}	
assign (resid 32 and name hn) (resid 32 and name ha*)	3.8 3.8 0.7 {3i}
assign (resid 32 and name hn) (resid 31 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 32 and name hn) (resid 31 and name hb)	4.2 4.2 0.0 {4s}
assign (resid 32 and name hn) (resid 31 and name hg**)	3.8 3.8 2.5 {3s}
assign (resid 32 and name ha) (resid 23 and name ha)	6.9 6.9 0.0 {9l}
{I33}	
assign (resid 33 and name hn) (resid 33 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 33 and name hn) (resid 33 and name hb)	3.8 3.8 0.0 {3i}
assign (resid 33 and name hn) (resid 33 and name hg12)	3.8 3.8 0.0 {3i}
assign (resid 33 and name hn) (resid 33 and name hg13)	3.8 3.8 0.0 {3i}
assign (resid 33 and name hn) (resid 33 and name hg2)	4.2 4.2 0.0 {4i}
assign (resid 33 and name hn) (resid 32 and name ha*)	4.2 4.2 0.7 {4i}
assign (resid 33 and name hn) (resid 22 and name hb)	3.8 3.8 0.0 {3l}
assign (resid 33 and name hn) (resid 22 and name hn)	6.5 6.5 0.0 {8l}
assign (resid 33 and name hn) (resid 23 and name hg**)	6.5 6.5 2.5 {8l}
{K34}	
assign (resid 34 and name hn) (resid 34 and name ha)	3.8 3.8 0.7 {3i}
assign (resid 34 and name hn) (resid 34 and name hb*)	3.8 3.8 0.7 {3i}

assign (resid 34 and name hn) (resid 34 and name hg*)	3.8 3.8 0.7 {3i}
assign (resid 34 and name hn) (resid 34 and name hd*)	6.9 6.9 0.7 {9i}
assign (resid 34 and name ha) (resid 22 and name hn)	5.5 5.5 0.0 {6l}
assign (resid 34 and name ha) (resid 21 and name ha)	6.0 6.0 0.0 {7l}
assign (resid 34 and name ha) (resid 33 and name hg12)	6.5 6.5 0.0 {8l}
assign (resid 34 and name hn) (resid 33 and name ha)	2.8 2.8 0.0 {1s}
assign (resid 34 and name hn) (resid 33 and name hb)	3.8 3.8 0.0 {3s}
assign (resid 34 and name hn) (resid 33 and name hg12)	3.8 3.8 0.0 {3s}
assign (resid 34 and name hn) (resid 33 and name hg13)	3.8 3.8 0.0 {3s}
assign (resid 34 and name hn) (resid 33 and name hg2)	3.8 3.8 0.0 {3s}
assign (resid 34 and name hn) (resid 32 and name hn)	6.5 6.5 0.0 {8l}
assign (resid 34 and name hn) (resid 50 and name ha)	6.5 6.5 0.0 {8l}
{G35}	
assign (resid 35 and name hn) (resid 35 and name ha*)	3.8 3.8 0.0 {3i}
assign (resid 35 and name ha*)(resid 34 and name hg*)	6.0 6.0 0.7 {7l}
assign (resid 35 and name hn) (resid 34 and name ha)	5.0 5.0 0.0 {5s}
assign (resid 35 and name hn) (resid 34 and name hb*)	3.8 3.8 0.7 {3s}
assign (resid 35 and name hn) (resid 34 and name hg*)	6.0 6.0 0.7 {7s}
assign (resid 35 and name hn) (resid 34 and name hd*)	6.5 6.5 0.7 {8s}
assign (resid 35 and name hn) (resid 20 and name hn)	6.5 6.5 0.0 {8l}
assign (resid 35 and name hn) (resid 21 and name ha)	6.5 6.5 0.0 {8l}
assign (resid 35 and name ha) (resid 49 and name ha)	6.9 6.9 0.0 {9l}
assign (resid 47 and name hn) (resid 47 and name hg2)	5.5 5.5 0.0 {6l}
{E36}	
assign (resid 36 and name hn) (resid 36 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 36 and name hn) (resid 36 and name hb*)	3.8 3.8 0.7 {5i}
assign (resid 36 and name hn) (resid 36 and name hg*)	3.8 3.8 0.7 {7i}
assign (resid 36 and name hn) (resid 35 and name ha*)	3.8 3.8 0.7 {3s}
assign (resid 36 and name hn) (resid 49 and name ha)	6.5 6.5 0.0 {8l}
assign (resid 36 and name hn) (resid 48 and name hn)	4.2 4.2 0.0 {4l}
assign (resid 36 and name hn) (resid 47 and name hg2)	3.8 3.8 0.0 {4l}
assign (resid 36 and name hn) (resid 37 and name ha)	6.5 6.5 0.0 {8l}
{V37}	
assign (resid 37 and name hn) (resid 37 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 37 and name hn) (resid 37 and name hb)	5.0 5.0 0.0 {5i}
assign (resid 37 and name hn) (resid 37 and name hg**)	5.5 5.5 2.5 {6i}
assign (resid 37 and name hn) (resid 36 and name ha)	5.0 5.0 0.0 {5s}
assign (resid 37 and name hn) (resid 36 and name hb*)	6.0 6.0 0.7 {7s}

assign (resid 37 and name hn) (resid 36 and name hg*)	5.5 5.5 0.7 {6s}
assign (resid 37 and name ha) (resid 48 and name hn)	6.0 6.0 0.0 {7i}
assign (resid 37 and name ha) (resid 39 and name hn)	6.0 6.0 0.0 {7i}
{V38}	
assign (resid 38 and name hn) (resid 38 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 38 and name hn) (resid 38 and name hb)	4.2 4.2 0.0 {4i}
assign (resid 38 and name hn) (resid 38 and name hg**)	3.8 3.8 2.5 {3i}
assign (resid 38 and name hg**)(resid 40 and name hn)	6.5 6.5 0.0 {8i}
assign (resid 38 and name hg**)(resid 46 and name hn)	6.5 6.5 0.0 {8i}
assign (resid 38 and name hn) (resid 37 and name ha)	2.8 2.8 0.0 {1s}
assign (resid 38 and name hn) (resid 37 and name hb)	3.2 3.2 0.0 {2s}
assign (resid 38 and name hn) (resid 37 and name hg**)	3.2 3.2 2.5 {2s}
assign (resid 38 and name hn) (resid 39 and name hn)	6.5 6.5 0.0 {8i}
assign (resid 38 and name hn) (resid 46 and name hn)	6.5 6.5 0.0 {8i}
{D39}	
assign (resid 39 and name hn) (resid 39 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 39 and name hn) (resid 39 and name hb*)	4.2 4.2 0.7 {4i}
assign (resid 39 and name hb*)(resid 40 and name hn)	6.5 6.5 0.7 {8i}
assign (resid 39 and name hb*)(resid 38 and name hg**)	6.5 6.5 2.5 {8i}
assign (resid 39 and name hn) (resid 38 and name ha)	4.2 4.2 0.0 {4s}
assign (resid 39 and name hn) (resid 38 and name hb)	4.2 4.2 0.0 {4s}
assign (resid 39 and name hn) (resid 38 and name hg**)	4.2 4.2 0.0 {3s}
assign (resid 39 and name hn) (resid 37 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 39 and name hn) (resid 43 and name hn)	6.0 6.0 0.0 {7i}
assign (resid 39 and name hn) (resid 40 and name hn)	6.5 6.5 0.0 {8i}
{E40}	
assign (resid 40 and name hn) (resid 40 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 40 and name hn) (resid 40 and name hb*)	6.0 6.0 0.7 {7i}
assign (resid 40 and name hn) (resid 40 and name hg*)	6.5 6.5 0.7 {8i}
assign (resid 40 and name hn) (resid 39 and name ha)	2.8 2.8 0.0 {1s}
assign (resid 40 and name hn) (resid 39 and name hb*)	6.0 6.0 0.7 {7s}
{T41}	
assign (resid 41 and name hn) (resid 41 and name hg*)	6.0 6.0 1.0 {7i}
assign (resid 41 and name hn) (resid 40 and name hb*)	6.5 6.5 0.7 {8s}
assign (resid 41 and name hn) (resid 40 and name hg*)	6.5 6.5 0.7 {8s}
{Q43}	
assign (resid 42 and name hn) (resid 42 and name ha)	6.0 6.0 0.0 {7i}

assign (resid 42 and name hn) (resid 42 and name hb*)	5.5 5.5 0.7 {6i}
assign (resid 42 and name hn) (resid 42 and name hg*)	5.5 5.5 0.7 {6i}
assign (resid 42 and name hn) (resid 41 and name hg)	6.0 6.0 0.0 {7s}
assign (resid 42 and name hn) (resid 39 and name hn)	6.0 6.0 0.0 {7i}
{T44}	
assign (resid 44 and name hn) (resid 44 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 44 and name hn) (resid 44 and name hb)	6.5 6.5 0.0 {7i}
assign (resid 44 and name hn) (resid 44 and name hg*)	5.5 5.5 1.0 {6i}
assign (resid 44 and name hn) (resid 43 and name ha)	6.0 6.0 0.0 {7s}
assign (resid 44 and name hn) (resid 43 and name hb*)	6.5 6.5 0.7 {8s}
{L45}	
assign (resid 45 and name hn) (resid 45 and name ha)	3.2 3.2 0.0 {2i}
assign (resid 45 and name hn) (resid 45 and name hb*)	5.5 5.5 0.7 {6i}
assign (resid 45 and name hn) (resid 45 and name hg)	6.5 6.5 0.0 {8i}
assign (resid 45 and name hn) (resid 45 and name hd**)	6.5 6.5 2.5 {8i}
assign (resid 45 and name hn) (resid 44 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 45 and name hn) (resid 44 and name hb)	5.0 5.0 0.0 {5s}
assign (resid 45 and name hn) (resid 44 and name hg*)	6.0 6.0 1.0 {7s}
{K46}	
assign (resid 46 and name hn) (resid 46 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 46 and name hn) (resid 46 and name hb*)	5.5 5.5 0.7 {6i}
assign (resid 46 and name hn) (resid 46 and name hg*)	5.0 5.0 0.7 {5i}
assign (resid 46 and name hn) (resid 46 and name hd*)	5.5 5.5 0.7 {6i}
assign (resid 46 and name ha) (resid 47 and name hn)	6.5 6.5 0.0 {8i}
assign (resid 46 and name ha) (resid 55 and name ha)	6.5 6.5 0.0 {8i}
assign (resid 46 and name ha) (resid 55 and name hg**)	6.5 6.5 2.5 {8i}
assign (resid 45 and name ha) (resid 45 and name hb*)	6.5 6.5 0.7 {8i}
assign (resid 45 and name ha) (resid 45 and name hd**)	6.0 6.0 2.5 {7i}
assign (resid 46 and name hn) (resid 45 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 46 and name hn) (resid 45 and name hb*)	3.8 3.8 0.7 {3s}
assign (resid 46 and name hn) (resid 45 and name hg)	6.5 6.5 0.0 {8s}
assign (resid 46 and name hn) (resid 45 and name hd**)	4.2 4.2 2.5 {4s}
assign (resid 46 and name hn) (resid 40 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 46 and name hn) (resid 38 and name hg**)	6.0 6.0 2.5 {7i}
{I47}	
assign (resid 47 and name hn) (resid 47 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 47 and name hn) (resid 47 and name hb)	3.2 3.2 0.0 {2i}
assign (resid 47 and name hn) (resid 47 and name hg12)	4.2 4.2 2.5 {4i}
assign (resid 47 and name hn) (resid 47 and name hg13)	4.2 4.2 2.5 {4i}

assign (resid 47 and name hn) (resid 47 and name hg2)	4.2 4.2 0.0 {4i}
assign (resid 47 and name hn) (resid 46 and name ha)	2.8 2.8 0.0 {1s}
assign (resid 47 and name hn) (resid 46 and name hb*)	6.5 6.5 0.7 {8s}
assign (resid 47 and name hn) (resid 46 and name hg*)	6.0 6.0 0.7 {7s}
assign (resid 47 and name hn) (resid 46 and name hd*)	6.5 6.5 0.7 {8s}
assign (resid 47 and name hn) (resid 54 and name hn)	6.5 6.5 0.0 {8l}
assign (resid 47 and name ha) (resid 37 and name ha)	6.5 6.5 0.0 {8l}
assign (resid 47 and name hb) (resid 54 and name hn)	6.0 6.0 0.0 {7l}
assign (resid 47 and name hb) (resid 45 and name hd**)	6.0 6.0 2.5 {7l}
{M48}	
assign (resid 48 and name hn) (resid 48 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 48 and name hn) (resid 48 and name hb*)	3.2 3.2 0.7 {2i}
assign (resid 48 and name hn) (resid 48 and name hg*)	3.8 3.8 0.7 {3i}
assign (resid 48 and name hn) (resid 47 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 48 and name hn) (resid 47 and name hb)	4.2 4.2 0.0 {3s}
assign (resid 48 and name hn) (resid 47 and name hg12)	3.8 3.8 2.5 {3s}
assign (resid 48 and name hn) (resid 47 and name hg13)	3.8 3.8 2.5 {3s}
assign (resid 48 and name hn) (resid 47 and name hg2)	3.8 3.8 0.0 {3s}
assign (resid 48 and name hn) (resid 36 and name hn)	4.2 4.2 0.0 {4l}
assign (resid 48 and name hn) (resid 37 and name ha)	6.0 6.0 0.0 {7l}
assign (resid 48 and name ha) (resid 47 and name hg12)	5.5 5.5 2.5 {6l}
assign (resid 48 and name ha) (resid 47 and name hg13)	5.5 5.5 2.5 {6l}
{T49}	
assign (resid 49 and name hn) (resid 49 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 49 and name hn) (resid 49 and name hg*)	3.8 3.8 1.0 {3i}
assign (resid 49 and name hn) (resid 48 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 49 and name hn) (resid 48 and name hb*)	6.5 6.5 0.7 {8s}
assign (resid 49 and name hn) (resid 48 and name hg*)	4.2 4.2 0.7 {4s}
assign (resid 49 and name hn) (resid 32 and name hn)	6.0 6.0 0.0 {7l}
assign (resid 49 and name hn) (resid 53 and name ha)	6.5 6.5 0.0 {8l}
assign (resid 49 and name ha) (resid 35 and name ha)	6.5 6.5 0.0 {8l}
assign (resid 49 and name hn) (resid 25 and name hg)	6.5 6.5 0.0 {8l}
{E50}	
assign (resid 50 and name hn) (resid 50 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 50 and name hn) (resid 50 and name hb*)	3.8 3.8 0.7 {3i}
assign (resid 50 and name hn) (resid 50 and name hg*)	3.8 3.8 0.7 {3i}
assign (resid 50 and name hn) (resid 49 and name ha)	3.8 3.8 0.0 {3s}

assign (resid 50 and name hn) (resid 49 and name hg*)	3.8 3.8 0.7 {3s}
assign (resid 50 and name hn) (resid 35 and name ha)	6.9 6.9 0.0 {9l}
{K51}	
assign (resid 51 and name hn) (resid 51 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 51 and name hn) (resid 51 and name hb*)	3.8 3.8 0.7 {3i}
assign (resid 51 and name hn) (resid 51 and name hg*)	3.8 3.8 0.7 {3i}
assign (resid 51 and name hn) (resid 50 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 51 and name hn) (resid 50 and name hb*)	6.5 6.5 0.7 {8s}
assign (resid 51 and name hn) (resid 50 and name hg*)	6.5 6.5 0.7 {8s}
assign (resid 51 and name hn) (resid 25 and name hg)	6.5 6.5 0.0 {4l}
assign (resid 51 and name hn) (resid 32 and name hn)	3.8 3.8 0.0 {3l}
{G52}	
assign (resid 52 and name hn) (resid 52 and name ha*)	3.8 3.8 0.7 {3i}
assign (resid 52 and name hn) (resid 51 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 52 and name hn) (resid 51 and name hb*)	4.2 4.2 0.7 {4s}
assign (resid 52 and name hn) (resid 51 and name hg*)	6.0 6.0 0.7 {7s}
assign (resid 52 and name hn) (resid 25 and name hg)	6.5 6.5 0.0 {4l}
{L53}	
assign (resid 53 and name hn) (resid 53 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 53 and name hn) (resid 53 and name hb*)	3.2 3.2 0.7 {2i}
assign (resid 53 and name hn) (resid 53 and name hg)	4.2 4.2 0.0 {4i}
assign (resid 53 and name hn) (resid 53 and name hd**)	4.2 4.2 2.5 {4i}
assign (resid 53 and name ha) (resid 48 and name ha)	6.5 6.5 0.0 {8l}
assign (resid 53 and name hn) (resid 52 and name ha*)	3.2 3.2 0.7 {2s}
{K54}	
assign (resid 54 and name hn) (resid 54 and name ha)	3.2 3.2 0.0 {2i}
assign (resid 54 and name hn) (resid 54 and name hb*)	3.8 3.8 0.7 {3i}
assign (resid 54 and name hn) (resid 54 and name hg*)	3.8 3.8 0.7 {3i}
assign (resid 54 and name hn) (resid 54 and name hd*)	3.8 3.8 0.7 {3i}
assign (resid 54 and name hn) (resid 54 and name he*)	6.5 6.5 0.7 {8i}
assign (resid 54 and name hn) (resid 53 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 54 and name hn) (resid 53 and name hb*)	3.8 3.8 0.7 {3s}
assign (resid 54 and name hn) (resid 53 and name hg)	3.8 3.8 0.0 {3s}
assign (resid 54 and name hn) (resid 53 and name hd**)	3.8 3.8 2.5 {4s}
assign (resid 54 and name hn) (resid 47 and name hn)	6.5 6.5 0.0 {8l}
assign (resid 54 and name ha) (resid 48 and name ha)	6.0 6.0 0.0 {7l}
{V55}	
assign (resid 55 and name hn) (resid 55 and name ha)	3.8 3.8 0.0 {3i}

assign (resid 55 and name hn) (resid 55 and name hb)	3.8 3.8 0.0 {3i}
assign (resid 55 and name hn) (resid 55 and name hg**)	3.8 3.8 2.5 {3i}
assign (resid 55 and name hn) (resid 54 and name ha)	2.8 2.8 0.0 {1s}
assign (resid 55 and name hn) (resid 54 and name hb*)	4.2 4.2 0.7 {4s}
assign (resid 55 and name hn) (resid 54 and name hg*)	4.2 4.2 0.7 {4s}
assign (resid 55 and name hn) (resid 54 and name hd*)	3.8 3.8 0.7 {3s}
assign (resid 55 and name ha) (resid 46 and name ha)	6.5 6.5 0.0 {8l}
{V56}	
assign (resid 56 and name hn) (resid 56 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 56 and name hn) (resid 56 and name hb)	3.8 3.8 0.0 {3i}
assign (resid 56 and name hn) (resid 56 and name hg**)	3.8 3.8 2.5 {3i}
assign (resid 56 and name hn) (resid 55 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 56 and name hn) (resid 55 and name hb)	3.2 3.2 0.0 {2s}
assign (resid 56 and name hn) (resid 55 and name hg**)	3.8 3.8 2.5 {3s}
assign (resid 56 and name hn) (resid 46 and name ha)	4.2 4.2 0.0 {4l}
assign (resid 56 and name hg**)(resid 58 and name hn)	6.5 6.5 2.5 {8l}
{A57}	
assign (resid 57 and name hn) (resid 57 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 57 and name hn) (resid 57 and name hb*)	3.2 3.2 1.0 {2i}
assign (resid 57 and name hn) (resid 56 and name ha)	2.8 2.8 0.0 {1s}
assign (resid 57 and name hn) (resid 56 and name hg**)	2.8 2.8 2.5 {1s}
assign (resid 57 and name hb*)(resid 59 and name hn)	6.5 6.5 1.0 {8l}
{K58}	
assign (resid 58 and name hn) (resid 58 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 58 and name hn) (resid 57 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 58 and name hn) (resid 57 and name hb*)	4.2 4.2 1.0 {4s}
assign (resid 58 and name hn) (resid 45 and name hg)	4.2 4.2 0.0 {4l}
assign (resid 58 and name hn) (resid 59 and name hn)	5.5 5.5 0.0 {6l}
{R59}	
assign (resid 59 and name hn) (resid 59 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 59 and name hn) (resid 59 and name hb*)	2.8 2.8 0.7 {1i}
assign (resid 59 and name hn) (resid 59 and name hg*)	3.2 3.2 0.7 {2i}
assign (resid 59 and name hn) (resid 59 and name hd*)	6.5 6.5 0.7 {3i}
assign (resid 59 and name hn) (resid 58 and name hn)	3.8 3.8 0.0 {3l}
assign (resid 59 and name hn) (resid 58 and name ha)	3.8 3.8 0.0 {3l}

{G60}		
assign (resid 60 and name hn) (resid 60 and name ha*)	4.2 4.2 0.7 {4i}	
assign (resid 60 and name hn) (resid 59 and name ha)	3.8 3.8 0.0 {3s}	
assign (resid 60 and name hn) (resid 59 and name hb*)	6.5 6.5 0.7 {8s}	
{R61}		
assign (resid 61 and name hn) (resid 61 and name ha)	3.8 3.8 0.0 {3i}	
assign (resid 61 and name hn) (resid 61 and name hb*)	6.0 6.0 0.7 {7i}	
assign (resid 61 and name hn) (resid 61 and name hg*)	6.0 6.0 0.7 {7i}	
assign (resid 61 and name hn) (resid 61 and name hd*)	6.0 6.0 0.7 {7i}	
assign (resid 61 and name hn) (resid 60 and name ha*)	3.8 3.8 0.7 {3s}	
assign (resid 61 and name hn) (resid 60 and name hn)	5.5 5.5 0.0 {6i}	
{T62}		
assign (resid 62 and name hn) (resid 62 and name ha)	5.5 5.5 0.0 {6i}	
assign (resid 62 and name hn) (resid 62 and name hb)	5.5 5.5 0.0 {6i}	
assign (resid 62 and name hn) (resid 62 and name hg*)	6.0 6.0 1.0 {7i}	
assign (resid 62 and name hn) (resid 61 and name ha)	5.0 5.0 0.0 {5s}	
assign (resid 62 and name hn) (resid 61 and name hb*)	5.5 5.5 0.7 {6s}	
assign (resid 62 and name hn) (resid 61 and name hg*)	6.0 6.0 0.7 {7s}	
assign (resid 62 and name hn) (resid 24 and name hn)	5.5 5.5 0.0 {6i}	
assign (resid 62 and name hb) (resid 24 and name hn)	6.5 6.5 0.0 {8i}	
assign (resid 62 and name hb) (resid 23 and name hg**)	6.0 6.0 2.5 {7i}	
{F63}		
assign (resid 63 and name hn) (resid 63 and name ha)	4.2 4.2 0.0 {4i}	
assign (resid 63 and name hn) (resid 63 and name hb*)	4.2 4.2 0.0 {4i}	
assign (resid 63 and name hn) (resid 62 and name ha)	3.8 3.8 0.0 {3s}	
assign (resid 63 and name hn) (resid 62 and name hg*)	3.8 3.8 1.0 {3s}	
assign (resid 63 and name hn) (resid 74 and name hn)	6.0 6.0 0.0 {7i}	
assign (resid 63 and name ha) (resid 22 and name ha)	6.5 6.5 0.0 {8i}	
assign (resid 63 and name hn) (resid 75 and name ha)	6.5 6.5 0.0 {8i}	
assign (resid 63 and name hd*)(resid 22 and name hg**)	6.5 6.5 2.5 {8i}	
assign (resid 63 and name hd*)(resid 65 and name hg**)	6.5 6.5 2.5 {8i}	
assign (resid 63 and name he*)(resid 65 and name hg**)	6.5 6.5 2.4 {8i}	
assign (resid 63 and name hb*)(resid 63 and name hd*)	3.8 3.8 2.4 {3i}	
{R64}		
assign (resid 64 and name hn) (resid 64 and name ha)	3.8 3.8 0.0 {3i}	
assign (resid 64 and name hn) (resid 64 and name hb*)	3.8 3.8 0.7 {3i}	
assign (resid 64 and name hn) (resid 64 and name hg*)	3.8 3.8 0.7 {3i}	

assign (resid 64 and name hn) (resid 63 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 64 and name hn) (resid 63 and name hb*)	3.8 3.8 0.7 {3s}
assign (resid 64 and name hn) (resid 63 and name hd*)	3.8 3.8 2.4 {3s}

assign (resid 64 and name hn) (resid 22 and name ha)	4.2 4.2 0.0 {4l}
assign (resid 64 and name hn) (resid 21 and name hn)	3.8 3.8 0.0 {3l}

{V65}

assign (resid 65 and name hn) (resid 65 and name ha)	4.2 4.2 0.0 {4i}
assign (resid 65 and name hn) (resid 65 and name hb)	3.8 3.8 0.0 {3i}
assign (resid 65 and name hn) (resid 65 and name hg**)	4.2 4.2 2.5 {4i}

assign (resid 65 and name hn) (resid 64 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 65 and name hn) (resid 64 and name hb*)	6.0 6.0 0.7 {7s}
assign (resid 65 and name hn) (resid 64 and name hg*)	6.5 6.5 0.7 {8s}

assign (resid 65 and name hn) (resid 73 and name Ha)	6.0 6.0 0.0 {7l}
------------------------------------------------------	------------------

{W66}

assign (resid 66 and name hn) (resid 66 and name ha)	3.2 3.2 0.0 {2i}
assign (resid 66 and name hn) (resid 66 and name hb*)	4.2 4.2 0.7 {4i}

assign (resid 66 and name hn) (resid 65 and name ha)	4.2 4.2 0.0 {4s}
assign (resid 66 and name hn) (resid 65 and name hg**)	3.8 3.8 2.5 {3s}
assign (resid 66 and name ne1)(resid 71 and name hd1)	5.5 5.5 0.0 {6l}
assign (resid 66 and name he1)(resid 71 and name hd12)	6.5 6.5 2.5 {8l}
assign (resid 66 and name he1)(resid 71 and name hd13)	6.5 6.5 2.5 {8l}
assign (resid 66 and name hn) (resid 20 and name ha)	4.2 4.2 0.0 {4l}
assign (resid 66 and name ha) (resid 71 and name ha)	6.5 6.5 0.0 {8l}
assign (resid 66 and name hb*)(resid 66 and name hd1)	6.5 6.5 0.7 {8i}
assign (resid 66 and name hn) (resid 19 and name hn)	6.5 6.5 0.0 {8l}

{Y67}

assign (resid 67 and name hn) (resid 67 and name ha)	4.2 4.2 0.0 {4i}
assign (resid 67 and name hn) (resid 67 and name hb*)	3.8 3.8 0.7 {3i}
assign (resid 67 and name ha) (resid 67 and name hd*)	4.2 4.2 2.4 {4i}
assign (resid 67 and name hb*)(resid 67 and name hd*)	4.2 4.2 2.4 {4i}

assign (resid 67 and name hn) (resid 66 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 67 and name hn) (resid 66 and name hb*)	4.2 4.2 0.7 {4s}

assign (resid 67 and name hn) (resid 71 and name ha)	6.0 6.0 0.0 {7l}
assign (resid 67 and name hb*)(resid 70 and name hn)	6.5 6.5 0.0 {8l}
assign (resid 67 and name hb*)(resid 68 and name hg*)	6.5 6.5 2.5 {8l}
assign (resid 67 and name hd*)(resid 37 and name hg**)	6.5 6.5 5.0 {8l}

{K68}

assign (resid 68 and name hn) (resid 68 and name ha)	3.2 3.2 0.0 {2i}
------------------------------------------------------	------------------

assign (resid 68 and name hn) (resid 68 and name hb*)	3.8 3.8 0.7 {3i}
assign (resid 68 and name hn) (resid 68 and name hg*)	4.2 4.2 0.7 {4i}
assign (resid 68 and name hn) (resid 67 and name ha)	6.0 6.0 0.0 {5s}
assign (resid 68 and name hn) (resid 67 and name hb*)	3.8 3.8 0.7 {3s}
assign (resid 68 and name hn) (resid 69 and name hn)	6.0 6.0 0.7 {7i}
{G69}	
assign (resid 69 and name hn) (resid 69 and name ha*)	3.8 3.8 0.7 {3i}
assign (resid 69 and name hn) (resid 68 and name ha)	3.2 3.2 0.0 {2i}
assign (resid 69 and name hn) (resid 68 and name hb*)	6.5 6.5 0.7 {8s}
assign (resid 69 and name hn) (resid 70 and name hn)	6.5 6.5 0.0 {8i}
{K70}	
assign (resid 70 and name hn) (resid 70 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 70 and name hn) (resid 70 and name hb*)	5.5 5.5 0.7 {6i}
assign (resid 70 and name hn) (resid 70 and name hg*)	5.5 5.5 0.7 {6i}
assign (resid 70 and name hn) (resid 69 and name ha*)	6.0 6.0 0.7 {7s}
{I71}	
assign (resid 71 and name hn) (resid 71 and name ha)	4.2 4.2 0.0 {4i}
assign (resid 71 and name hn) (resid 71 and name hb)	3.8 3.8 0.0 {3i}
assign (resid 71 and name hn) (resid 71 and name hg12)	3.8 3.8 0.0 {3i}
assign (resid 71 and name hn) (resid 71 and name hg13)	3.8 3.8 0.0 {3i}
assign (resid 71 and name hn) (resid 71 and name hd1)	6.0 6.0 0.0 {7i}
assign (resid 71 and name hn) (resid 71 and name hg2)	4.2 4.2 0.0 {4i}
assign (resid 71 and name hn) (resid 70 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 71 and name hn) (resid 70 and name hb*)	3.8 3.8 0.7 {3s}
assign (resid 71 and name hn) (resid 70 and name hg*)	6.0 6.0 0.4 {7s}
assign (resid 71 and name ha) (resid 66 and name ha)	6.5 6.5 0.0 {8i}
assign (resid 71 and name hd1)(resid 58 and name hn)	6.5 6.5 0.0 {8i}
{M72}	
assign (resid 72 and name hn) (resid 72 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 72 and name hn) (resid 72 and name hb*)	4.2 4.2 0.7 {4i}
assign (resid 72 and name hn) (resid 72 and name hg*)	6.0 6.0 0.7 {7i}
assign (resid 72 and name hn) (resid 71 and name ha)	3.2 3.2 0.0 {3s}
assign (resid 72 and name hn) (resid 71 and name hb)	6.5 6.5 0.0 {8s}
assign (resid 72 and name hn) (resid 71 and name hg12)	6.5 6.5 1.0 {8s}
assign (resid 72 and name hn) (resid 71 and name hg13)	6.5 6.5 1.0 {8s}
assign (resid 72 and name hn) (resid 71 and name hg2)	5.5 5.5 0.0 {6s}
assign (resid 72 and name hn) (resid 66 and name ha)	5.5 5.5 0.0 {6i}

assign (resid 72 and name hn) (resid 65 and name hn)	6.5 6.5 0.0 {8l}
{R73}	
assign (resid 73 and name hn) (resid 73 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 73 and name hn) (resid 73 and name hb*)	6.0 6.0 0.7 {7i}
assign (resid 73 and name hn) (resid 73 and name hg*)	3.8 3.8 0.7 {3i}
assign (resid 73 and name hn) (resid 72 and name ha)	3.8 3.8 0.0 {3s}
assign (resid 73 and name hn) (resid 72 and name hb*)	4.2 4.2 0.7 {4s}
assign (resid 73 and name hn) (resid 72 and name hg*)	3.8 3.8 0.7 {3s}
assign (resid 73 and name ha) (resid 64 and name ha)	6.5 6.5 0.0 {8l}
{I74}	
assign (resid 74 and name hn) (resid 74 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 74 and name hn) (resid 74 and name hb)	4.2 4.2 0.0 {4i}
assign (resid 74 and name hn) (resid 74 and name hg12)	6.0 6.0 1.0 {7i}
assign (resid 74 and name hn) (resid 74 and name hg13)	6.0 6.0 1.0 {7i}
assign (resid 74 and name hn) (resid 74 and name hd1)	6.5 6.5 0.0 {8i}
assign (resid 74 and name hn) (resid 74 and name hg2)	3.8 3.8 0.0 {3i}
assign (resid 74 and name hn) (resid 73 and name ha)	3.2 3.2 0.0 {2s}
assign (resid 74 and name hn) (resid 73 and name hb*)	4.2 4.2 0.7 {4s}
assign (resid 74 and name hn) (resid 73 and name hg*)	6.5 6.5 0.7 {8s}
assign (resid 74 and name hn) (resid 64 and name ha)	4.2 4.2 0.0 {4l}
assign (resid 74 and name hn) (resid 63 and name hn)	6.0 6.0 0.0 {7l}
assign (resid 74 and name hg2)(resid 63 and name hd*)	6.0 6.0 2.4 {7l}
assign (resid 74 and name hg2)(resid 63 and name he*)	6.5 6.5 2.4 {8l}
assign (resid 74 and name hd1)(resid 63 and name hd*)	6.5 6.5 2.4 {8l}
{K75}	
assign (resid 75 and name hn) (resid 75 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 75 and name hn) (resid 75 and name hb*)	6.0 6.0 0.7 {7i}
assign (resid 75 and name hn) (resid 75 and name hg*)	6.5 6.5 0.7 {8i}
assign (resid 75 and name hn) (resid 75 and name hd*)	6.0 6.0 0.7 {7i}
assign (resid 75 and name hn) (resid 74 and name ha)	3.2 3.2 0.0 {3s}
assign (resid 75 and name hn) (resid 74 and name hb)	5.5 5.5 0.0 {6s}
assign (resid 75 and name hn) (resid 74 and name hg12)	6.5 6.5 2.5 {8s}
assign (resid 75 and name hn) (resid 74 and name hg13)	6.5 6.5 2.5 {8s}
assign (resid 75 and name hn) (resid 74 and name hg2)	5.0 5.0 0.0 {5s}
assign (resid 75 and name hn) (resid 74 and name hd1)	6.5 6.5 0.0 {8s}
assign (resid 75 and name ha) (resid 62 and name ha)	6.0 6.0 0.0 {7s}
{G76}	
assign (resid 76 and name hn) (resid 76 and name ha*)	6.0 6.0 0.7 {7i}

assign (resid 76 and name hn) (resid 75 and name ha)	5.0 5.0 0.0 {5s}
assign (resid 76 and name hn) (resid 75 and name hb*)	6.5 6.5 0.7 {8s}
assign (resid 76 and name hn) (resid 75 and name hd*)	6.5 6.5 0.7 {8s}
assign (resid 76 and name hn) (resid 62 and name ha)	6.5 6.5 0.0 {8l}
{D77}	
assign (resid 77 and name hn) (resid 77 and name ha)	3.8 3.8 0.0 {3i}
assign (resid 77 and name hn) (resid 77 and name hb*)	3.8 3.8 0.7 {3i}
assign (resid 77 and name hn) (resid 76 and name ha*)	4.2 4.2 0.7 {4s}
assign (resid 77 and name hn) (resid 75 and name hg*)	6.0 6.0 0.7 {7l}
assign (resid 77 and name hn) (resid 75 and name hd*)	6.0 6.0 0.7 {7l}
assign (resid 77 and name hn) (resid 78 and name hd**)	6.0 6.0 2.5 {7l}
assign (resid 77 and name hn) (resid 78 and name hn)	5.5 5.5 0.0 {6l}
{L78}	
assign (resid 78 and name hn) (resid 78 and name ha)	4.2 4.2 0.0 {4i}
assign (resid 78 and name hn) (resid 78 and name hb*)	5.5 5.5 0.7 {6i}
assign (resid 78 and name hn) (resid 78 and name hg)	5.5 5.5 0.0 {6i}
assign (resid 78 and name hn) (resid 78 and name hd**)	6.0 6.0 2.5 {7i}
assign (resid 78 and name hn) (resid 77 and name hb*)	6.5 6.5 0.7 {8s}
{G90}	
assign (resid 90 and name hn) (resid 90 and name ha*)	6.0 6.0 0.7 {7i}
assign (resid 90 and name hn) (resid 91 and name hb*)	6.5 6.5 0.7 {8s}
assign (resid 90 and name hn) (resid 89 and name ha)	6.5 6.5 0.0 {8s}
{R96}	
assign (resid 96 and name hn) (resid 96 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 96 and name hn) (resid 96 and name hb*)	6.0 6.0 0.7 {7i}
assign (resid 96 and name hn) (resid 95 and name ha)	6.0 6.0 0.0 {7s}
assign (resid 96 and name hn) (resid 95 and name hb*)	6.5 6.5 0.7 {7s}
{A97}	
assign (resid 97 and name hn) (resid 97 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 97 and name hn) (resid 97 and name hb*)	6.5 6.5 0.7 {8i}
assign (resid 97 and name hn) (resid 96 and name ha)	6.0 6.0 0.0 {8s}
assign (resid 97 and name hn) (resid 96 and name hb*)	6.0 6.0 0.7 {7s}
{K98}	
assign (resid 98 and name hn) (resid 98 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 98 and name hn) (resid 97 and name ha)	6.5 6.5 0.0 {8s}
{V100}	
assign (resid 100 and name hn) (resid 100 and name ha)	5.5 5.5 0.0 {6i}

assign (resid 100 and name hn) (resid 100 and name hb)	5.5 5.5 0.0 {6i}
assign (resid 100 and name hn) (resid 100 and name hg**)	5.5 5.5 2.5 {6i}
assign (resid 100 and name hn) (resid 99 and name ha*)	5.5 5.5 0.7 {6s}
{W101}	
assign (resid 101 and name hn) (resid 101 and name ha)	6.0 6.0 0.0 {7i}
assign (resid 101 and name hn) (resid 101 and name hb*)	5.5 5.5 0.0 {6i}
assign (resid 101 and name hn) (resid 101 and name hd1)	6.0 6.0 0.0 {7i}
assign (resid 101 and name hn) (resid 101 and name he3)	6.0 6.0 0.0 {7i}
assign (resid 101 and name hn) (resid 100 and name ha)	6.0 6.0 0.0 {7s}
assign (resid 101 and name hn) (resid 100 and name hb)	5.5 5.5 0.0 {6s}
assign (resid 101 and name hn) (resid 100 and name hg**)	5.5 5.5 2.5 {6s}
{I102}	
assign (resid 102 and name hn) (resid 102 and name ha)	5.5 5.5 0.0 {6i}
assign (resid 102 and name hn) (resid 102 and name hb)	5.5 5.5 0.0 {6i}
assign (resid 102 and name hn) (resid 102 and name hg12)	6.5 6.5 1.0 {8i}
assign (resid 102 and name hn) (resid 102 and name hg13)	6.5 6.5 1.0 {8i}
assign (resid 102 and name hn) (resid 102 and name hg2)	6.0 6.0 0.0 {7i}
assign (resid 102 and name hn) (resid 102 and name hd1)	6.0 6.0 0.0 {7i}
assign (resid 102 and name hn) (resid 101 and name ha)	5.0 5.0 0.0 {5s}
assign (resid 102 and name hn) (resid 101 and name hb*)	6.0 6.0 0.7 {7s}

APPENDIX 3. Dihedral angle restraint table for the NMR structure of *A. fulgidus* aRpp29

Dihedral angle constraints are used as follows:

Torsion angle restraints are used to restrict dihedral angles to values within the range typical of beta sheets and alpha helices, when these regions of regular secondary structure are clearly identified by characteristic NOE cross peaks and chemical shift values.

{*** beta strand 19 to 24 ***}

assign (resid 18 and name o) (resid 19 and name n)
(resid 19 and name ca) (resid 19 and name o) 1.0 -120.0 25.0 2

assign (resid 19 and name n) (resid 19 and name ca)
(resid 19 and name c) (resid 20 and name n) 1.0 150.0 25.0 2

assign (resid 19 and name o) (resid 20 and name n)
(resid 20 and name ca) (resid 20 and name o) 1.0 -120.0 25.0 2

assign (resid 20 and name n) (resid 20 and name ca)
(resid 20 and name c) (resid 21 and name n) 1.0 150.0 25.0 2

assign (resid 20 and name o) (resid 21 and name n)
(resid 21 and name ca) (resid 21 and name o) 1.0 -120.0 25.0 2

assign (resid 21 and name n) (resid 21 and name ca)
(resid 21 and name c) (resid 22 and name n) 1.0 150.0 25.0 2

assign (resid 21 and name o) (resid 22 and name n)
(resid 22 and name ca) (resid 22 and name o) 1.0 -120.0 25.0 2

assign (resid 22 and name n) (resid 22 and name ca)
(resid 22 and name c) (resid 23 and name n) 1.0 150.0 25.0 2

assign (resid 22 and name o) (resid 23 and name n)
(resid 23 and name ca) (resid 23 and name o) 1.0 -120.0 25.0 2

assign (resid 23 and name n) (resid 23 and name ca)
(resid 23 and name c) (resid 24 and name n) 1.0 150.0 25.0 2

assign (resid 23 and name o) (resid 24 and name n)
(resid 24 and name ca) (resid 24 and name o) 1.0 -120.0 25.0 2

assign (resid 24 and name n) (resid 24 and name ca)
(resid 24 and name c) (resid 25 and name n) 1.0 150.0 25.0 2

{*** beta strand 32 to 38 ***}

assign (resid 31 and name o) (resid 32 and name n)
(resid 32 and name ca) (resid 32 and name o) 1.0 -120.0 25.0 2

assign (resid 32 and name n) (resid 32 and name ca)
(resid 32 and name c) (resid 33 and name n) 1.0 150.0 25.0 2

assign (resid 32 and name o) (resid 33 and name n)
(resid 33 and name ca) (resid 33 and name o) 1.0 -120.0 25.0 2

assign (resid 33 and name n) (resid 33 and name ca)
(resid 33 and name c) (resid 34 and name n) 1.0 150.0 25.0 2

assign (resid 33 and name o) (resid 34 and name n)
(resid 34 and name ca) (resid 34 and name o) 1.0 -120.0 25.0 2

assign (resid 34 and name n) (resid 34 and name ca)
(resid 34 and name c) (resid 35 and name n) 1.0 150.0 25.0 2

assign (resid 34 and name o) (resid 35 and name n)
(resid 35 and name ca) (resid 35 and name o) 1.0 -120.0 25.0 2

assign (resid 35 and name n) (resid 35 and name ca)
(resid 35 and name c) (resid 36 and name n) 1.0 150.0 25.0 2

assign (resid 35 and name o) (resid 36 and name n)
(resid 36 and name ca) (resid 36 and name o) 1.0 -120.0 25.0 2

assign (resid 36 and name n) (resid 36 and name ca)
(resid 36 and name c) (resid 37 and name n) 1.0 150.0 25.0 2

assign (resid 36 and name o) (resid 37 and name n)
(resid 37 and name ca) (resid 37 and name o) 1.0 -120.0 25.0 2

assign (resid 37 and name n) (resid 37 and name ca)
(resid 37 and name c) (resid 38 and name n) 1.0 150.0 25.0 2

assign (resid 37 and name o) (resid 38 and name n)
(resid 38 and name ca) (resid 38 and name o) 1.0 -120.0 25.0 2

assign (resid 38 and name n) (resid 38 and name ca)
(resid 38 and name c) (resid 39 and name n) 1.0 150.0 25.0 2

{*** beta strand 44 to 49 ***}

assign (resid 43 and name o) (resid 44 and name n)
(resid 44 and name ca) (resid 44 and name o) 1.0 -120.0 25.0 2

assign (resid 44 and name n) (resid 44 and name ca)
(resid 44 and name c) (resid 45 and name n) 1.0 150.0 25.0 2

assign (resid 44 and name o) (resid 45 and name n)
(resid 45 and name ca) (resid 45 and name o) 1.0 -120.0 25.0 2

assign (resid 45 and name n) (resid 45 and name ca)
(resid 45 and name c) (resid 46 and name n) 1.0 150.0 25.0 2

assign (resid 45 and name o) (resid 46 and name n)
(resid 46 and name ca) (resid 46 and name o) 1.0 -120.0 25.0 2

assign (resid 46 and name n) (resid 46 and name ca)
(resid 46 and name c) (resid 47 and name n) 1.0 150.0 25.0 2

assign (resid 46 and name o) (resid 47 and name n)
(resid 47 and name ca) (resid 47 and name o) 1.0 -120.0 25.0 2

assign (resid 47 and name n) (resid 47 and name ca)
(resid 47 and name c) (resid 48 and name n) 1.0 150.0 25.0 2

assign (resid 47 and name o) (resid 48 and name n)
(resid 48 and name ca) (resid 48 and name o) 1.0 -120.0 25.0 2

assign (resid 48 and name n) (resid 48 and name ca)
(resid 48 and name c) (resid 49 and name n) 1.0 150.0 25.0 2

{*** beta strand 53 to 57 ***}

assign (resid 52 and name o) (resid 53 and name n)
(resid 53 and name ca) (resid 53 and name o) 1.0 -120.0 25.0 2

assign (resid 53 and name n) (resid 53 and name ca)
(resid 53 and name c) (resid 54 and name n) 1.0 150.0 25.0 2

assign (resid 53 and name o) (resid 54 and name n)
(resid 54 and name ca) (resid 54 and name o) 1.0 -120.0 25.0 2

assign (resid 54 and name n) (resid 54 and name ca)
(resid 54 and name c) (resid 55 and name n) 1.0 150.0 25.0 2

assign (resid 54 and name o) (resid 55 and name n)

(resid 55 and name ca) (resid 55 and name o) 1.0 -120.0 25.0 2
 assign (resid 55 and name n) (resid 55 and name ca)
 (resid 55 and name c) (resid 56 and name n) 1.0 150.0 25.0 2
 assign (resid 55 and name o) (resid 56 and name n)
 (resid 56 and name ca) (resid 56 and name o) 1.0 -120.0 25.0 2
 assign (resid 56 and name n) (resid 56 and name ca)
 (resid 56 and name c) (resid 57 and name n) 1.0 150.0 25.0 2
 assign (resid 56 and name o) (resid 57 and name n)
 (resid 57 and name ca) (resid 57 and name o) 1.0 -120.0 25.0 2
 assign (resid 57 and name n) (resid 57 and name ca)
 (resid 57 and name c) (resid 58 and name n) 1.0 150.0 25.0 2

 {*** beta strand 62 to 67 ***}

 assign (resid 61 and name o) (resid 62 and name n)
 (resid 62 and name ca) (resid 62 and name o) 1.0 -120.0 25.0 2
 assign (resid 62 and name n) (resid 62 and name ca)
 (resid 62 and name c) (resid 63 and name n) 1.0 150.0 25.0 2
 assign (resid 62 and name o) (resid 63 and name n)
 (resid 63 and name ca) (resid 63 and name o) 1.0 -120.0 25.0 2
 assign (resid 63 and name n) (resid 63 and name ca)
 (resid 63 and name c) (resid 64 and name n) 1.0 150.0 25.0 2
 assign (resid 63 and name o) (resid 64 and name n)
 (resid 64 and name ca) (resid 64 and name o) 1.0 -120.0 25.0 2
 assign (resid 64 and name n) (resid 64 and name ca)
 (resid 64 and name c) (resid 65 and name n) 1.0 150.0 25.0 2
 assign (resid 64 and name o) (resid 65 and name n)
 (resid 65 and name ca) (resid 65 and name o) 1.0 -120.0 25.0 2
 assign (resid 65 and name n) (resid 65 and name ca)
 (resid 65 and name c) (resid 66 and name n) 1.0 150.0 25.0 2
 assign (resid 65 and name o) (resid 66 and name n)
 (resid 66 and name ca) (resid 66 and name o) 1.0 -120.0 25.0 2
 assign (resid 66 and name n) (resid 66 and name ca)
 (resid 66 and name c) (resid 67 and name n) 1.0 150.0 25.0 2

assign (resid 66 and name o) (resid 67 and name n)
(resid 67 and name ca) (resid 67 and name o) 1.0 -120.0 25.0 2

assign (resid 67 and name n) (resid 67 and name ca)
(resid 67 and name c) (resid 68 and name n) 1.0 150.0 25.0 2

{*** beta strand 71 to 76 ***}

assign (resid 70 and name o) (resid 71 and name n)
(resid 71 and name ca) (resid 71 and name o) 1.0 -120.0 25.0 2

assign (resid 71 and name n) (resid 71 and name ca)
(resid 71 and name c) (resid 72 and name n) 1.0 150.0 25.0 2

assign (resid 71 and name o) (resid 72 and name n)
(resid 72 and name ca) (resid 72 and name o) 1.0 -120.0 25.0 2

assign (resid 72 and name n) (resid 72 and name ca)
(resid 72 and name c) (resid 73 and name n) 1.0 150.0 25.0 2

assign (resid 72 and name o) (resid 73 and name n)
(resid 73 and name ca) (resid 73 and name o) 1.0 -120.0 25.0 2

assign (resid 73 and name n) (resid 73 and name ca)
(resid 73 and name c) (resid 74 and name n) 1.0 150.0 25.0 2

assign (resid 73 and name o) (resid 74 and name n)
(resid 74 and name ca) (resid 74 and name o) 1.0 -120.0 25.0 2

assign (resid 74 and name n) (resid 74 and name ca)
(resid 74 and name c) (resid 75 and name n) 1.0 150.0 25.0 2

assign (resid 74 and name o) (resid 75 and name n)
(resid 75 and name ca) (resid 75 and name o) 1.0 -120.0 25.0 2

assign (resid 75 and name n) (resid 75 and name ca)
(resid 75 and name c) (resid 76 and name n) 1.0 150.0 25.0 2

assign (resid 75 and name o) (resid 76 and name n)
(resid 76 and name ca) (resid 76 and name o) 1.0 -120.0 25.0 2

assign (resid 76 and name n) (resid 76 and name ca)
(resid 76 and name c) (resid 77 and name n) 1.0 150.0 25.0 2

Appendix 4. Hydrogen bond restraint table for NMR structure of *A. fulgidus* aRpp29.

Hydrogen bonds in regular alpha helices and beta sheets are defined using distance restraints.

{ ***** link strand 19-24 to 31-37 antiparallel ***** }

assign (resid 18 and name o) (resid 37 and name hn)	2.15	0.0039	0.0039
assign (resid 18 and name c) (resid 37 and name n)	4.35	0.0039	0.0039
assign (resid 18 and name o) (resid 37 and name n)	3.14	0.0039	0.0039
assign (resid 18 and name c) (resid 37 and name hn)	3.38	0.0039	0.0039

assign (resid 35 and name o) (resid 20 and name hn)	2.15	0.0039	0.0039
assign (resid 35 and name c) (resid 20 and name n)	4.35	0.0039	0.0039
assign (resid 35 and name o) (resid 20 and name n)	3.14	0.0039	0.0039
assign (resid 35 and name c) (resid 20 and name hn)	3.38	0.0039	0.0039

assign (resid 20 and name o) (resid 35 and name hn)	2.15	0.0039	0.0039
assign (resid 20 and name c) (resid 35 and name n)	4.35	0.0039	0.0039
assign (resid 20 and name o) (resid 35 and name n)	3.14	0.0039	0.0039
assign (resid 20 and name c) (resid 35 and name hn)	3.38	0.0039	0.0039

assign (resid 33 and name o) (resid 22 and name hn)	2.15	0.0039	0.0039
assign (resid 33 and name c) (resid 22 and name n)	4.35	0.0039	0.0039
assign (resid 33 and name o) (resid 22 and name n)	3.14	0.0039	0.0039
assign (resid 33 and name c) (resid 22 and name hn)	3.38	0.0039	0.0039

assign (resid 22 and name o) (resid 33 and name hn)	2.15	0.0039	0.0039
assign (resid 22 and name c) (resid 33 and name n)	4.35	0.0039	0.0039
assign (resid 22 and name o) (resid 33 and name n)	3.14	0.0039	0.0039
assign (resid 22 and name c) (resid 33 and name hn)	3.38	0.0039	0.0039

assign (resid 31 and name o) (resid 24 and name hn)	2.15	0.0039	0.0039
assign (resid 31 and name c) (resid 24 and name n)	4.35	0.0039	0.0039
assign (resid 31 and name o) (resid 24 and name n)	3.14	0.0039	0.0039
assign (resid 31 and name c) (resid 24 and name hn)	3.38	0.0039	0.0039

assign (resid 24 and name o) (resid 31 and name hn)	2.15	0.0039	0.0039
assign (resid 24 and name c) (resid 31 and name n)	4.35	0.0039	0.0039
assign (resid 24 and name o) (resid 31 and name n)	3.14	0.0039	0.0039
assign (resid 24 and name c) (resid 31 and name hn)	3.38	0.0039	0.0039

{ ***** link strand 18-24 to 62-67 antiparallel ***** }

assign (resid 66 and name o) (resid 19 and name hn)	2.15	0.0039	0.0039
assign (resid 66 and name c) (resid 19 and name n)	4.35	0.0039	0.0039
assign (resid 66 and name o) (resid 19 and name n)	3.14	0.0039	0.0039
assign (resid 66 and name c) (resid 19 and name hn)	3.38	0.0039	0.0039

assign (resid 19 and name o) (resid 66 and name hn)	2.15	0.0039	0.0039
------------------------------------------------------	------	--------	--------

assign (resid 19 and name c) (resid 66 and name n)	4.35	0.0039	0.0039
assign (resid 19 and name o) (resid 66 and name n)	3.14	0.0039	0.0039
assign (resid 19 and name c) (resid 66 and name hn)	3.38	0.0039	0.0039
assign (resid 64 and name o) (resid 21 and name hn)	2.15	0.0039	0.0039
assign (resid 64 and name c) (resid 21 and name n)	4.35	0.0039	0.0039
assign (resid 64 and name o) (resid 21 and name n)	3.14	0.0039	0.0039
assign (resid 64 and name c) (resid 21 and name hn)	3.38	0.0039	0.0039
assign (resid 21 and name o) (resid 64 and name hn)	2.15	0.0039	0.0039
assign (resid 21 and name c) (resid 64 and name n)	4.35	0.0039	0.0039
assign (resid 21 and name o) (resid 64 and name n)	3.14	0.0039	0.0039
assign (resid 21 and name c) (resid 64 and name hn)	3.38	0.0039	0.0039
assign (resid 62 and name o) (resid 23 and name hn)	2.15	0.0039	0.0039
assign (resid 62 and name c) (resid 23 and name n)	4.35	0.0039	0.0039
assign (resid 62 and name o) (resid 23 and name n)	3.14	0.0039	0.0039
assign (resid 62 and name c) (resid 23 and name hn)	3.38	0.0039	0.0039
assign (resid 23 and name o) (resid 62 and name hn)	2.15	0.0039	0.0039
assign (resid 23 and name c) (resid 62 and name n)	4.35	0.0039	0.0039
assign (resid 23 and name o) (resid 62 and name n)	3.14	0.0039	0.0039
assign (resid 23 and name c) (resid 62 and name hn)	3.38	0.0039	0.0039
{ ***** link strand 62-67 to 71-76 antiparallel ***** }			
assign (resid 61 and name o) (resid 76 and name hn)	2.15	0.0039	0.0039
assign (resid 61 and name c) (resid 76 and name n)	4.35	0.0039	0.0039
assign (resid 61 and name o) (resid 76 and name n)	3.14	0.0039	0.0039
assign (resid 61 and name c) (resid 76 and name hn)	3.38	0.0039	0.0039
assign (resid 74 and name o) (resid 63 and name hn)	2.15	0.0039	0.0039
assign (resid 74 and name c) (resid 63 and name n)	4.35	0.0039	0.0039
assign (resid 74 and name o) (resid 63 and name n)	3.14	0.0039	0.0039
assign (resid 74 and name c) (resid 63 and name hn)	3.38	0.0039	0.0039
assign (resid 63 and name o) (resid 74 and name hn)	2.15	0.0039	0.0039
assign (resid 63 and name c) (resid 74 and name n)	4.35	0.0039	0.0039
assign (resid 63 and name o) (resid 74 and name n)	3.14	0.0039	0.0039
assign (resid 63 and name c) (resid 74 and name hn)	3.38	0.0039	0.0039
assign (resid 72 and name o) (resid 65 and name hn)	2.15	0.0039	0.0039
assign (resid 72 and name c) (resid 65 and name n)	4.35	0.0039	0.0039
assign (resid 72 and name o) (resid 65 and name n)	3.14	0.0039	0.0039
assign (resid 72 and name c) (resid 65 and name hn)	3.38	0.0039	0.0039
assign (resid 65 and name o) (resid 72 and name hn)	2.15	0.0039	0.0039
assign (resid 65 and name c) (resid 72 and name n)	4.35	0.0039	0.0039
assign (resid 65 and name o) (resid 72 and name n)	3.14	0.0039	0.0039
assign (resid 65 and name c) (resid 72 and name hn)	3.38	0.0039	0.0039

assign (resid 70 and name o) (resid 67 and name hn)	2.15	0.0039	0.0039
assign (resid 70 and name c) (resid 67 and name n)	4.35	0.0039	0.0039
assign (resid 70 and name o) (resid 67 and name n)	3.14	0.0039	0.0039
assign (resid 70 and name c) (resid 67 and name hn)	3.38	0.0039	0.0039

{ ***** link strand 44-49 to 53-57 antiparallel ***** }

assign (resid 56 and name o) (resid 45 and name hn)	2.15	0.0039	0.0039
assign (resid 56 and name c) (resid 45 and name n)	4.35	0.0039	0.0039
assign (resid 56 and name o) (resid 45 and name n)	3.14	0.0039	0.0039
assign (resid 56 and name c) (resid 45 and name hn)	3.38	0.0039	0.0039

assign (resid 45 and name o) (resid 56 and name hn)	2.15	0.0039	0.0039
assign (resid 45 and name c) (resid 56 and name n)	4.35	0.0039	0.0039
assign (resid 45 and name o) (resid 56 and name n)	3.14	0.0039	0.0039
assign (resid 45 and name c) (resid 56 and name hn)	3.38	0.0039	0.0039

assign (resid 54 and name o) (resid 47 and name hn)	2.15	0.0039	0.0039
assign (resid 54 and name c) (resid 47 and name n)	4.35	0.0039	0.0039
assign (resid 54 and name o) (resid 47 and name n)	3.14	0.0039	0.0039
assign (resid 54 and name c) (resid 47 and name hn)	3.38	0.0039	0.0039

assign (resid 47 and name o) (resid 54 and name hn)	2.15	0.0039	0.0039
assign (resid 47 and name c) (resid 54 and name n)	4.35	0.0039	0.0039
assign (resid 47 and name o) (resid 54 and name n)	3.14	0.0039	0.0039
assign (resid 47 and name c) (resid 54 and name hn)	3.38	0.0039	0.0039

assign (resid 52 and name o) (resid 49 and name hn)	2.15	0.0039	0.0039
assign (resid 52 and name c) (resid 49 and name n)	4.35	0.0039	0.0039
assign (resid 52 and name o) (resid 49 and name n)	3.14	0.0039	0.0039
assign (resid 52 and name c) (resid 49 and name hn)	3.38	0.0039	0.0039

{ ***** link strand 32-37 to 44-49 antiparallel ***** }

assign (resid 38 and name o) (resid 46 and name hn)	2.15	0.0039	0.0039
assign (resid 38 and name c) (resid 46 and name n)	4.35	0.0039	0.0039
assign (resid 38 and name o) (resid 46 and name n)	3.14	0.0039	0.0039
assign (resid 38 and name c) (resid 46 and name hn)	3.38	0.0039	0.0039

assign (resid 46 and name o) (resid 38 and name hn)	2.15	0.0039	0.0039
assign (resid 46 and name c) (resid 38 and name n)	4.35	0.0039	0.0039
assign (resid 46 and name o) (resid 38 and name n)	3.14	0.0039	0.0039
assign (resid 46 and name c) (resid 38 and name hn)	3.38	0.0039	0.0039

assign (resid 36 and name o) (resid 48 and name hn)	2.15	0.0039	0.0039
assign (resid 36 and name c) (resid 48 and name n)	4.35	0.0039	0.0039
assign (resid 36 and name o) (resid 48 and name n)	3.14	0.0039	0.0039
assign (resid 36 and name c) (resid 48 and name hn)	3.38	0.0039	0.0039

assign (resid 48 and name o) (resid 36 and name hn)	2.15	0.0039	0.0039
------------------------------------------------------	------	--------	--------

assign (resid 48 and name c) (resid 36 and name n)	4.35	0.0039	0.0039
assign (resid 48 and name o) (resid 36 and name n)	3.14	0.0039	0.0039
assign (resid 48 and name c) (resid 36 and name hn)	3.38	0.0039	0.0039

List of Abbreviations

CD	circular dichroism
CNS	Crystallography and NMR System
COSY	correlation spectroscopy
DNA	deoxyribonucleic acid
D ₂ O	deuterium (² H ₂ O)
DFQ	double quantum filter
DTT	dithiothreitol
HEPES	N-[2-hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
HMQC	heteronuclear multiple-quantum correlation
HSMQC	heteronuclear single- and multiple-quantum correlation
HSQC	heteronuclear single-quantum correlation
kDa	kilodalton
MAD	multiple-wavelength anomalous diffraction
MIR	multiple isomorphous replacement
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NMR	nuclear magnetic resonance

OD	optical density
PAGE	polyacryamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
pI	isoelectric point
ppm	parts per million
r.m.s.d.	root mean square deviation
RNA	ribonucleic acid
SAD	single-wavelength anomalous diffraction
SeMet	selenomethionine
SIR	single isomorphous replacement
TEV	tobacco etch virus
TOCSY	total correlation spectroscopy
tRNA	transfer RNA
UV	ultraviolet

Bibliography

- Altman, S., D. Wesolowski, and R. S. Puranam. 1993. "Nucleotide sequences of the RNA subunit of RNase P from several mammals." *Genomics* 18:418-22.
- Andrews, A. J., T. A. Hall, and J. W. Brown. 2001. "Characterization of RNase P holoenzymes from *Methanococcus jannaschii* and *Methanothermobacter thermoautotrophicus*." *Biol Chem* 382:1171-7.
- Ban, N., P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz. 2000. "The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution." *Science* 289:905-20.
- Barrera, A., X. Fang, J. Jacob, E. Casey, P. Thiyagarajan, and T. Pan. 2002. "Dimeric and monomeric *Bacillus subtilis* RNase P holoenzyme in the absence and presence of pre-tRNA substrates." *Biochemistry* 41:12986-94.
- Bartkiewicz, M., H. Gold, and S. Altman. 1989. "Identification and characterization of an RNA molecule that copurifies with RNase P activity from HeLa cells." *Genes Dev* 3:488-99.
- Battiste, J. L., T. V. Pestova, C. U. Hellen, and G. Wagner. 2000. "The eIF1A solution structure reveals a large RNA-binding surface important for scanning function." *Mol Cell* 5:109-19.
- Baum, M. and A. Schön. 1996. "Localization and expression of the closely linked cyanelle genes for RNase P RNA and two transfer RNAs." *FEBS Lett* 382:60-4.
- Beebe, J. A. and C. A. Fierke. 1994. "A kinetic mechanism for cleavage of precursor tRNA(Asp) catalyzed by the RNA component of *Bacillus subtilis* ribonuclease P." *Biochemistry* 33:10294-304.
- Beebe, J. A., J. C. Kurz, and C. A. Fierke. 1996. "Magnesium ions are required by *Bacillus subtilis* ribonuclease P RNA for both binding and cleaving precursor tRNA^{Asp}." *Biochemistry* 35:10493-505.

- Bertrand, E., F. Houser-Scott, A. Kendall, R. H. Singer, and D. R. Engelke. 1998. "Nucleolar localization of early tRNA processing." *Genes Dev* 12:2463-8.
- Biswas, R., D. W. Ledman, R. O. Fox, S. Altman, and V. Gopalan. 2000. "Mapping RNA-protein interactions in ribonuclease P from *Escherichia coli* using disulfide-linked EDTA-Fe." *J Mol Biol* 296:19-31.
- Bjork, G. R. 1995. "Genetic dissection of synthesis and function of modified nucleosides in bacterial transfer RNA." *Prog Nucleic Acid Res Mol Biol* 50:263-338.
- Blundell, T. L. and Johnson, L.N. 1976. "Protein crystallography." Academic Press, New York, NY.
- Boomershine, W. P., C. A. McElroy, H. Y. Tsai, R. C. Wilson, V. Gopalan, and M. P. Foster. 2003. "Structure of Mth11/Mth Rpp29, an essential protein subunit of archaeal and eukaryotic RNase P." *Proc Natl Acad Sci U S A* 100:15398-403.
- Borek, D., Minor, W., and Otwinoski, Z. 2003. "Measurement errors and their consequences in protein crystallography." *Acta Crystallogr D Biol Crystallogr* 59:2031-8.
- Bricogne, G., C. Vonrhein, C. Flensburg, M. Schiltz, and W. Paciorek. 2003. "Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0." *Acta Crystallogr D Biol Crystallogr* 59:2023-30.
- Brodersen, D.E., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J., and Kjeldgaard, M. 2000. "Applications of single-wavelength anomalous dispersion at high and atomic resolution." *Acta Crystallogr D Biol Crystallogr* 56:431-41.
- Brown, J. W. 1998. "The ribonuclease P database." *Nucleic Acids Res* 26:351-2.
- Brown, J. W. 1999. "The Ribonuclease P Database." *Nucleic Acids Res* 27:314.
- Brown, J. W., E. S. Haas, and N. R. Pace. 1993. "Characterization of ribonuclease P RNAs from thermophilic bacteria." *Nucleic Acids Res* 21:671-9.

- Brünger , A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. 1998. "Crystallography & NMR system: A new software suite for macromolecular structure determination." *Acta Crystallogr D Biol Crystallogr* 54 (Pt 5):905-21.
- Brünger , A. T., P. D. Adams, and L. M. Rice. 1998. "Recent developments for the efficient crystallographic refinement of macromolecular structures." *Curr Opin Struct Biol* 8:606-11.
- Cate, J. H., A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, and J. A. Doudna. 1996. "Crystal structure of a group I ribozyme domain: principles of RNA packing." *Science* 273:1678-85.
- Cavanaugh, J., Fairbrother, W., Palmer, A.G., and Skelton, N.J. 1996. "Protein NMR spectroscopy: principles and practice." Academic Press, San Diego, CA.
- Chamberlain, J. R., Y. Lee, W. S. Lane, and D. R. Engelke. 1998. "Purification and characterization of the nuclear RNase P holoenzyme complex reveals extensive subunit overlap with RNase MRP." *Genes Dev* 12:1678-90.
- Chamberlain, J. R., Ramos Pagan, D. W. Kindelberger, and D. R. Engelke. 1996. "An RNase P RNA subunit mutation affects ribosomal RNA processing." *Nucleic Acids Res* 24:3158-66.
- Chen, J. L., J. M. Nolan, M. E. Harris, and N. R. Pace. 1998. "Comparative photocross-linking analysis of the tertiary structures of Escherichia coli and Bacillus subtilis RNase P RNAs." *Embo J* 17:1515-25.
- Chen, J. L. and N. R. Pace. 1997. "Identification of the universally conserved core of ribonuclease P RNA." *RNA* 3:557-60.
- Chen, Y. W. and G. M. Clore. 2000. "A systematic case study on using NMR models for molecular replacement: p53 tetramerization domain revisited." *Acta Crystallogr D Biol Crystallogr* 56 Pt 12:1535-40.
- Chen, Y. W., E. J. Dodson, and G. J. Kleywegt. 2000. "Does NMR mean "not for molecular replacement"? Using NMR-based search models to solve protein crystal structures." *Structure Fold Des* 8:R213-20.

- Chu, S., R. H. Archer, J. M. Zengel, and L. Lindahl. 1994. "The RNA of RNase MRP is required for normal processing of ribosomal RNA." *Proc Natl Acad Sci U S A* 91:659-63.
- Chu, S., J. M. Zengel, and L. Lindahl. 1997. "A novel protein shared by RNase MRP and RNase P." *RNA* 3:382-91.
- Cordier, A. and A. Schön. 1999. "Cyanelle RNase P: RNA structure analysis and holoenzyme properties of an organellar ribonucleoprotein enzyme." *J Mol Biol* 289:9-20.
- Crary, S. M., S. Niranjana Kumari, and C. A. Fierke. 1998. "The protein component of *Bacillus subtilis* ribonuclease P increases catalytic efficiency by enhancing interactions with the 5' leader sequence of pre-tRNA^{Asp}." *Biochemistry* 37:9409-16.
- Dang, Y. L. and N. C. Martin. 1993. "Yeast mitochondrial RNase P. Sequence of the RPM2 gene and demonstration that its product is a protein subunit of the enzyme." *J Biol Chem* 268:19791-6.
- Darr, S. C., K. Zito, D. Smith, and N. R. Pace. 1992. "Contributions of phylogenetically variable structural elements to the function of the ribozyme ribonuclease P." *Biochemistry* 31:328-33.
- Dauter, Z., Dauter, M., and Dodson, E. 2002. "Jolly SAD." *Acta Crystallogr D Biol Crystallogr* 58:494-506.
- Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." *J Biomol NMR* 6:277-93.
- DeLano, W. L. 2002. "The PyMol Molecular Graphics System." *DeLano Scientific, San Carlos, CA USA*.
- Demeler, B., H. Saber, and J.C. Hansen. 1997. "Identification and Interpretation of Complexity in Sedimentation Velocity Boundaries." *Biophysical Journal* 72:397-407.
- DePristo, M. A., de Bakker, P.I., and Blundell, T.L. 2004. "Heterogeneity and inaccuracy in protein structures solved by x-ray crystallography." *Structure* 12:831-8.

- Dichtl, B. and D. Tollervey. 1997. "Pop3p is essential for the activity of the RNase MRP and RNase P ribonucleoproteins in vivo." *Embo J* 16:417-29.
- Dodson. E. 2003. "Is it jolly SAD?" *Acta Crystallogr D Biol Crystallogr* 59:1958-65.
- Drenth, J. 1999. "Principles of protein crystallography." Springer-Verlag Inc., New York, NY.
- Ealick, S.E. 2000. "Advances in multiple wavelength anomalous diffraction crystallography." *Curr Opin Chem Biol* 4:495-9.
- Fang, X. W., X. J. Yang, K. Littrell, S. Niranjankumari, P. Thiyagarajan, C. A. Fierke, T. R. Sosnick, and T. Pan. 2001. "The *Bacillus subtilis* RNase P holoenzyme contains two RNase P RNA and two RNase P protein subunits." *RNA* 7:233-41.
- Farrow, N. A., R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay. 1994. "Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation." *Biochemistry* 33:5984-6003.
- Ferre-D'Amare, A. R., K. Zhou, and J. A. Doudna. 1998. "Crystal structure of a hepatitis delta virus ribozyme." *Nature* 395:567-74.
- Fields, S. and O. Song. 1989. "A novel genetic system to detect protein-protein interactions." *Nature* 340:245-6.
- Frank, D. N., C. Adamidi, M. A. Ehringer, C. Pitulle, and N. R. Pace. 2000. "Phylogenetic-comparative analysis of the eukaryal ribonuclease P RNA." *RNA* 6:1895-904.
- Frank, D. N. and N. R. Pace. 1998. "Ribonuclease P: unity and diversity in a tRNA processing ribozyme." *Annu Rev Biochem* 67:153-80.
- Gegenheimer, P. 1995. "Structure, mechanism and evolution of chloroplast transfer RNA processing systems." *Mol Biol Rep* 22:147-50.

- Gibrat, J. F., T. Madej, and S. H. Bryant. 1996. "Surprising similarities in structure comparison." *Curr Opin Struct Biol* 6:377-85.
- Goddard, T.D. and D.G. Kneller. 2002. "Sparky 3." *University of California, San Francisco*.
- Gold, H. A., J. N. Topper, D. A. Clayton, and J. Craft. 1989. "The RNA processing enzyme RNase MRP is identical to the Th RNP and related to RNase P." *Science* 245:1377-80.
- Gonzalez, A. 2003. "Optimizing data collection for structure determination." *Acta Crystallogr D Biol Crystallogr* 59:1935-42.
- Gopalan, V., A. D. Baxevanis, D. Landsman, and S. Altman. 1997. "Analysis of the functional role of conserved residues in the protein subunit of ribonuclease P from *Escherichia coli*." *J Mol Biol* 267:818-29.
- Grzesiek, S., H. Dobeli, R. Gentz, G. Garotta, A. M. Labhardt, and A. Bax. 1992. "¹H, ¹³C, and ¹⁵N NMR backbone assignments and secondary structure of human interferon-gamma." *Biochemistry* 31:8180-90.
- Guerrier-Takada, C. and S. Altman. 1993. "A physical assay for and kinetic analysis of the interactions between M1 RNA and tRNA precursor substrates." *Biochemistry* 32:7152-61.
- Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace, and S. Altman. 1983. "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." *Cell* 35:849-57.
- Haas, E. S., D. W. Armbruster, B. M. Vucson, C. J. Daniels, and J. W. Brown. 1996. "Comparative analysis of ribonuclease P RNA structure in Archaea." *Nucleic Acids Res* 24:1252-9.
- Haas, E. S. and J. W. Brown. 1998. "Evolutionary variation in bacterial RNase P RNAs." *Nucleic Acids Res* 26:4093-9.
- Haas, E. S., J. W. Brown, C. Pitulle, and N. R. Pace. 1994. "Further perspective on the catalytic core and secondary structure of ribonuclease P RNA." *Proc Natl Acad Sci U S A* 91:2527-31.

- Hall, T. A. and J. W. Brown. 2002. "Archaeal RNase P has multiple protein subunits homologous to eukaryotic nuclear RNase P proteins." *RNA* 8:296-306.
- Hardt, W. D., J. Schlegl, V. A. Erdmann, and R. K. Hartmann. 1995. "Kinetics and thermodynamics of the RNase P RNA cleavage reaction: analysis of tRNA 3'-end variants." *J Mol Biol* 247:161-72.
- Harris, J. K., E. S. Haas, D. Williams, D. N. Frank, and J. W. Brown. 2001. "New insight into RNase P RNA structure from comparative analysis of the archaeal RNA." *RNA* 7:220-32.
- Harris, M. E., A. V. Kazantsev, J. L. Chen, and N. R. Pace. 1997. "Analysis of the tertiary structure of the ribonuclease P ribozyme-substrate complex by site-specific photoaffinity crosslinking." *RNA* 3:561-76.
- Harris, M. E., J. M. Nolan, A. Malhotra, J. W. Brown, S. C. Harvey, and N. R. Pace. 1994. "Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA." *Embo J* 13:3953-63.
- Harris, M. E. and N. R. Pace. 1995. "Identification of phosphates involved in catalysis by the ribozyme RNase P RNA." *RNA* 1:210-8.
- Hartmann, E. and R. K. Hartmann. 2003. "The enigma of ribonuclease P evolution." *Trends Genet* 19:561-9.
- Holm, L., C. Ouzounis, C. Sander, G. Tuparev, and G. Vriend. 1992. "A database of protein structure families with common folding motifs." *Protein Sci* 1:1691-8.
- Holm, L. and C. Sander. 1993. "Protein structure comparison by alignment of distance matrices." *J Mol Biol* 233:123-38.
- Houser-Scott, F., S. Xiao, C. E. Millikin, J. M. Zengel, L. Lindahl, and D. R. Engelke. 2002. "Interactions among the protein and RNA subunits of *Saccharomyces cerevisiae* nuclear RNase P." *Proc Natl Acad Sci U S A* 99:2684-9.

- Hsieh, J., A. J. Andrews, and C. A. Fierke. 2004. "Roles of protein subunits in RNA-protein complexes: lessons from ribonuclease P." *Biopolymers* 73:79-89.
- Jacobson, M. R., L. G. Cao, K. Taneja, R. H. Singer, Y. L. Wang, and T. Pederson. 1997. "Nuclear domains of the RNA subunit of RNase P." *J Cell Sci* 110 (Pt 7):829-37.
- Jacobson, M. R., L. G. Cao, Y. L. Wang, and T. Pederson. 1995. "Dynamic localization of RNase MRP RNA in the nucleolus observed by fluorescent RNA cytochemistry in living cells." *J Cell Biol* 131:1649-58.
- Jarrous, N. 2002. "Human ribonuclease P: subunits, function, and intranuclear localization." *RNA* 8:1-7.
- Jarrous, N. and S. Altman. 2001. "Human ribonuclease P." *Methods Enzymol* 342:93-100.
- Jarrous, N., P. S. Eder, D. Wesolowski, and S. Altman. 1999. "Rpp14 and Rpp29, two protein subunits of human ribonuclease P." *RNA* 5:153-7.
- Jarrous, N., R. Reiner, D. Wesolowski, H. Mann, C. Guerrier-Takada, and S. Altman. 2001. "Function and subnuclear distribution of Rpp21, a protein subunit of the human ribonucleoprotein ribonuclease P." *RNA* 7:1153-64.
- Jiang, T. and S. Altman. 2001. "Protein-protein interactions with subunits of human nuclear RNase P." *Proc Natl Acad Sci U S A* 98:920-5.
- Jiang, T. and S. Altman. 2002. "A protein subunit of human RNase P, Rpp14, and its interacting partner, OIP2, have 3'→5' exoribonuclease activity." *Proc Natl Acad Sci U S A* 99:5295-300.
- Jiang, T., C. Guerrier-Takada, and S. Altman. 2001. "Protein-RNA interactions in the subunits of human nuclear RNase P." *RNA* 7:937-41.
- Kapust, R. B. and D. S. Waugh. 1999. "Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused." *Protein Sci* 8:1668-74.

- Kapust, R. B., Tozser, J., Copeland, T. D., and Waugh, D. S. 2002. "The P1' specificity of tobacco etch virus protease." *Biochem Biophys Res Commun* 294:949-55.
- Karwan, R. 1993. "RNase MRP/RNase P: a structure-function relation conserved in evolution?" *FEBS Lett* 319:1-4.
- Kassenbrock, C. K., G. J. Gao, K. R. Groom, P. Sulo, M. G. Douglas, and N. C. Martin. 1995. "RPM2, independently of its mitochondrial RNase P function, suppresses an ISP42 mutant defective in mitochondrial import and is essential for normal growth." *Mol Cell Biol* 15:4763-70.
- Kay, L. E. 1993. "Pulse-field gradient-enhanced three-dimensional NMR experiment for correlating ^{13}C alpha/beta, ^{13}C , and ^1H alpha chemical shifts in uniformly carbon-13-labeled proteins dissolved in water." *J. Am. Chem Soc.* 115:2055-2057.
- Kazakov, S. and S. Altman. 1991. "Site-specific cleavage by metal ion cofactors and inhibitors of M1 RNA, the catalytic subunit of RNase P from Escherichia coli." *Proc Natl Acad Sci U S A* 88:9193-7.
- Ke, A., K. Zhou, F. Ding, J. H. Cate, and J. A. Doudna. 2004. "A conformational switch controls hepatitis delta virus ribozyme catalysis." *Nature* 429:201-5.
- Kirsebom, L. A. 1995. "RNase P--a 'Scarlet Pimpernel'." *Mol Microbiol* 17:411-20.
- Kirsebom, L. A. and A. Vioque. 1995. "RNase P from bacteria. Substrate recognition and function of the protein subunit." *Mol Biol Rep* 22:99-109.
- Kiss, T., C. Marshallsay, and W. Filipowicz. 1992. "7-2/MRP RNAs in plant and mammalian cells: association with higher order structures in the nucleolus." *Embo J* 11:3737-46.
- Kleineidam, R. G., C. Pitulle, B. Sproat, and G. Krupp. 1993. "Efficient cleavage of pre-tRNAs by E. coli RNase P RNA requires the 2'-hydroxyl of the ribose at the cleavage site." *Nucleic Acids Res* 21:1097-101.
- Klenk, H. P., R. A. Clayton, J. F. Tomb, O. White, K. E. Nelson, K. A. Ketchum, R. J. Dodson, M. Gwinn, E. K. Hickey, J. D. Peterson, D. L. Richardson, A. R. Kerlavage, D. E. Graham, N. C. Kyrpides, R. D. Fleischmann, J. Quackenbush, N. H. Lee, G. G. Sutton, S. Gill, E. F. Kirkness, B. A.

- Dougherty, K. McKenney, M. D. Adams, B. Loftus, J. C. Venter, and et al. 1997. "The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*." *Nature* 390:364-70.
- Kleywegt, G. J. 2000. "Validation of protein crystals structures." *Acta Crystallogr D Biol Crystallogr* 56:249-65.
- Kouzuma, Y., M. Mizoguchi, H. Takagi, H. Fukuhara, M. Tsukamoto, T. Numata, and M. Kimura. 2003. "Reconstitution of archaeal ribonuclease P from RNA and four protein components." *Biochem Biophys Res Commun* 306:666-73.
- Krasilnikov, A. S., Y. Xiao, T. Pan, and A. Mondragon. 2004. "Basis for structural diversity in homologous RNAs." *Science* 306:104-7.
- Krasilnikov, A. S., X. Yang, T. Pan, and A. Mondragon. 2003. "Crystal structure of the specificity domain of ribonuclease P." *Nature* 421:760-4.
- Kufel, J. and L. A. Kirsebom. 1996. "Residues in *Escherichia coli* RNase P RNA important for cleavage site selection and divalent metal ion binding." *J Mol Biol* 263:685-98.
- Kumar, S. and R. Nussinov. 1999. "Salt bridge stability in monomeric proteins." *J Mol Biol* 293:1241-55.
- Kurz, J. C., S. Niranjanakumari, and C. A. Fierke. 1998. "Protein component of *Bacillus subtilis* RNase P specifically enhances the affinity for precursor-tRNA^{Asp}." *Biochemistry* 37:2393-400.
- Laskowski, R. A., J. A. Rullmann, M. W. MacArthur, R. Kaptein, and J. M. Thornton. 1996. "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." *J Biomol NMR* 8:477-86.
- Lee, J. Y., C. F. Evans, and D. R. Engelke. 1991. "Expression of RNase P RNA in *Saccharomyces cerevisiae* is controlled by an unusual RNA polymerase III promoter." *Proc Natl Acad Sci U S A* 88:6986-90.
- Lee, J. Y., C. E. Rohlman, L. A. Molony, and D. R. Engelke. 1991. "Characterization of RPR1, an essential gene encoding the RNA component of *Saccharomyces cerevisiae* nuclear RNase P." *Mol Cell Biol* 11:721-30.

- Li, K., C. S. Smagula, W. J. Parsons, J. A. Richardson, M. Gonzalez, H. K. Hagler, and R. S. Williams. 1994. "Subcellular partitioning of MRP RNA assessed by ultrastructural and biochemical analysis." *J Cell Biol* 124:871-82.
- Li, Y. and S. Altman. 2001. "A subunit of human nuclear RNase P has ATPase activity." *Proc Natl Acad Sci U S A* 98:441-4.
- Lillemoen, J., C. S. Cameron, and D. W. Hoffman. 1997. "The stability and dynamics of ribosomal protein L9: investigations of a molecular strut by amide proton exchange and circular dichroism." *J Mol Biol* 268:482-93.
- Lindahl, L., S. Fretz, N. Epps, and J. M. Zengel. 2000. "Functional equivalence of hairpins in the RNA subunits of RNase MRP and RNase P in *Saccharomyces cerevisiae*." *RNA* 6:653-8.
- Lindahl, L. and J. M. Zengel. 1995. "RNase MRP and rRNA processing." *Mol Biol Rep* 22:69-73.
- Liu, F. and S. Altman. 1994. "Differential evolution of substrates for an RNA enzyme in the presence and absence of its protein cofactor." *Cell* 77:1093-100.
- Loria, A. and T. Pan. 1997. "Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from *Bacillus subtilis* ribonuclease P." *Biochemistry* 36:6317-25.
- Loria, A. and T. Pan. 1998. "Recognition of the 5' leader and the acceptor stem of a pre-tRNA substrate by the ribozyme from *Bacillus subtilis* RNase P." *Biochemistry* 37:10126-33.
- Lygerou, Z., P. Mitchell, E. Petfalski, B. Seraphin, and D. Tollervey. 1994. "The POP1 gene encodes a protein component common to the RNase MRP and RNase P ribonucleoproteins." *Genes Dev* 8:1423-33.
- Lygerou, Z., H. Pluk, W. J. van Venrooij, and B. Seraphin. 1996. "hPop1: an autoantigenic protein subunit shared by the human RNase P and RNase MRP ribonucleoproteins." *Embo J* 15:5936-48.

- Mann, H., Y. Ben-Asouli, A. Schein, S. Moussa, and N. Jarrous. 2003. "Eukaryotic RNase P: role of RNA and protein subunits of a primordial catalytic ribonucleoprotein in RNA-based catalysis." *Mol Cell* 12:925-35.
- Manavalan, P. and Johnson, W. C., Jr.. 1987. Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal Biochem* 167:76-85.
- Massire, C., L. Jaeger, and E. Westhof. 1998. "Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis." *J Mol Biol* 279:773-93.
- Mattaj, I. W. 1993. "RNA recognition: a family matter?" *Cell* 73:837-40.
- McRee, D. E. 1999. "XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density." *J Struct Biol* 125:156-65.
- Mildvan, A. S. 1997. "Mechanisms of signaling and related enzymes." *Proteins* 29:401-16.
- Miller, D. L. and N. C. Martin. 1983. "Characterization of the yeast mitochondrial locus necessary for tRNA biosynthesis: DNA sequence analysis and identification of a new transcript." *Cell* 34:911-7.
- Morales, M. J., Y. L. Dang, Y. C. Lou, P. Sulo, and N. C. Martin. 1992. "A 105-kDa protein is required for yeast mitochondrial RNase P activity." *Proc Natl Acad Sci U S A* 89:9875-9.
- Muhandiram, D.R. and L.E. Kay. 1994. "Gradient enhanced triple resonance three-dimensional NMR experiments with improved sensitivity." *J. Magn. Reson., Series B* 103:203-216.
- Nieuwlandt, D. T., E. S. Haas, and C. J. Daniels. 1991. "The RNA component of RNase P from the archaebacterium *Haloferax volcanii*." *J Biol Chem* 266:5689-95.
- Niranjanakumari, S., J. C. Kurz, and C. A. Fierke. 1998. "Expression, purification and characterization of the recombinant ribonuclease P protein component from *Bacillus subtilis*." *Nucleic Acids Res* 26:3090-6.

- Niranjanakumari, S., T. Stams, S. M. Crary, D. W. Christianson, and C. A. Fierke. 1998. "Protein component of the ribozyme ribonuclease P alters substrate recognition by directly contacting precursor tRNA." *Proc Natl Acad Sci U S A* 95:15212-7.
- Numata, T., I. Ishimatsu, Y. Kakuta, I. Tanaka, and M. Kimura. 2004. "Crystal structure of archaeal ribonuclease P protein Ph1771p from *Pyrococcus horikoshii* OT3: an archaeal homolog of eukaryotic ribonuclease P protein Rpp29." *RNA* 10:1423-32.
- Oh, B. K., D. N. Frank, and N. R. Pace. 1998. "Participation of the 3'-CCA of tRNA in the binding of catalytic Mg²⁺ ions by ribonuclease P." *Biochemistry* 37:7277-83.
- Oh, B. K. and N. R. Pace. 1994. "Interaction of the 3'-end of tRNA with ribonuclease P RNA." *Nucleic Acids Res* 22:4087-94.
- Otwinowski, Z. and W. Minor. 1997. "Processing of X-ray Diffraction Data Collected in Oscillation Mode." *Methods in Enzymology* 276:307-326.
- Pace, N. R. and J. W. Brown. 1995. "Evolutionary perspective on the structure and function of ribonuclease P, a ribozyme." *J Bacteriol* 177:1919-28.
- Pagan-Ramos, E., Y. Lee, and D. R. Engelke. 1996. "Mutational analysis of *Saccharomyces cerevisiae* nuclear RNase P: randomization of universally conserved positions in the RNA subunit." *RNA* 2:441-51.
- Pagan-Ramos, E., A. J. Tranguch, D. W. Kindelberger, and D. R. Engelke. 1994. "Replacement of the *Saccharomyces cerevisiae* RPR1 gene with heterologous RNase P RNA genes." *Nucleic Acids Res* 22:200-7.
- Pan, T. 1995. "Higher order folding and domain analysis of the ribozyme from *Bacillus subtilis* ribonuclease P." *Biochemistry* 34:902-9.
- Pannone, B. K., S. D. Kim, D. A. Noe, and S. L. Wolin. 2001. "Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein." *Genetics* 158:187-96.
- Pannucci, J. A., E. S. Haas, T. A. Hall, J. K. Harris, and J. W. Brown. 1999. "RNase P RNAs from some Archaea are catalytically active." *Proc Natl Acad Sci U S A* 96:7803-8.

- Pascal, S. M., D.R. Muhandiram, T. Yamazaki, J. D. Forman-Kay, and L. E. Kay. 1994. "Simultaneous acquisition of ¹⁵N and ¹³C-Edited NOE spectra of proteins dissolved in water." *J. Magn. Reson., Series B* 103:197-201.
- Pascual, A. and A. Vioque. 1994. "Sequence and structure of the RNA subunit of RNase P from the cyanobacterium *Pseudoanabaena* sp. PCC6903." *Biochim Biophys Acta* 1218:463-5.
- Pascual, A. and A. Vioque. 1999. "Functional reconstitution of RNase P activity from a plastid RNA subunit and a cyanobacterial protein subunit." *FEBS Lett* 442:7-10.
- Perrakis, A., M. Harkiolaki, K. S. Wilson, and V. S. Lamzin. 2001. "ARP/wARP and molecular replacement." *Acta Crystallogr D Biol Crystallogr* 57:1445-50.
- Pitulle, C., M. Garcia-Paris, K. R. Zamudio, and N. R. Pace. 1998. "Comparative structure analysis of vertebrate ribonuclease P RNA." *Nucleic Acids Res* 26:3333-9.
- Pley, H. W., K. M. Flaherty, and D. B. McKay. 1994. "Three-dimensional structure of a hammerhead ribozyme." *Nature* 372:68-74.
- Pluk, H., H. van Eenennaam, S. A. Rutjes, G. J. Pruijn, and W. J. van Venrooij. 1999. "RNA-protein interactions in the human RNase MRP ribonucleoprotein complex." *RNA* 5:512-24.
- Puranam, R. S. and G. Attardi. 2001. "The RNase P associated with HeLa cell mitochondria contains an essential RNA component identical in sequence to that of the nuclear RNase P." *Mol Cell Biol* 21:548-61.
- Reich, C., G. J. Olsen, B. Pace, and N. R. Pace. 1988. "Role of the protein moiety of ribonuclease P, a ribonucleoprotein enzyme." *Science* 239:178-81.
- Reimer, G., I. Raska, U. Scheer, and E. M. Tan. 1988. "Immunolocalization of 7-2-ribonucleoprotein in the granular component of the nucleolus." *Exp Cell Res* 176:117-28.

- Rice, L.M., Earnst, T. N., and Brunger, A. T. 2000. "Single-wavelength anomalous diffraction phasing revisited." *Acta Crystallogr D Biol Crystallogr* 56:1413-20.
- Robertson, H. D., S. Altman, and J. D. Smith. 1972. "Purification and properties of a specific *Escherichia coli* ribonuclease which cleaves a tyrosine transfer ribonucleic acid precursor." *J Biol Chem* 247:5243-51.
- Roll-Mecak, A., C. Cao, T. E. Dever, and S. K. Burley. 2000. "X-Ray structures of the universal translation initiation factor IF2/eIF5B: conformational changes on GDP and GTP binding." *Cell* 103:781-92.
- Rossmannith, W. and R. M. Karwan. 1998. "Characterization of human mitochondrial RNase P: novel aspects in tRNA processing." *Biochem Biophys Res Commun* 247:234-41.
- Rossmannith, W., A. Tullo, T. Potuschak, R. Karwan, and E. Sbisà. 1995. "Human mitochondrial tRNA processing." *J Biol Chem* 270:12885-91.
- Salgado-Garrido, J., E. Bragado-Nilsson, S. Kandels-Lewis, and B. Seraphin. 1999. "Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin." *Embo J* 18:3451-62.
- Schmitt, M. E. 1999. "Molecular modeling of the three-dimensional architecture of the RNA component of yeast RNase MRP." *J Mol Biol* 292:827-36.
- Schmitt, M. E. and D. A. Clayton. 1992. "Yeast site-specific ribonucleoprotein endoribonuclease MRP contains an RNA component homologous to mammalian RNase MRP RNA and essential for cell viability." *Genes Dev* 6:1975-85.
- Schmitt, M. E. and D. A. Clayton. 1993. "Nuclear RNase MRP is required for correct processing of pre-5.8S rRNA in *Saccharomyces cerevisiae*." *Mol Cell Biol* 13:7935-41.
- Schmitt, M. E. and D. A. Clayton. 1994. "Characterization of a unique protein component of yeast RNase MRP: an RNA-binding protein with a zinc-cluster domain." *Genes Dev* 8:2617-28.
- Schön, A. 1999. "Ribonuclease P: the diversity of a ubiquitous RNA processing enzyme." *FEMS Microbiol Rev* 23:391-406.

- Schumacher, M. A., R. F. Pearson, T. Moller, P. Valentin-Hansen, and R. G. Brennan. 2002. "Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein." *Embo J* 21:3546-56.
- Scott, W. G., J. B. Murray, J. R. Arnold, B. L. Stoddard, and A. Klug. 1996. "Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme." *Science* 274:2065-9.
- Sidote, D. J., Heideker, J., and Hoffman, D. W. 2004. "Crystal Structure of Archaeal Ribonuclease P Protein aRpp29 from *Archaeoglobus fulgidus*." *Biochemistry* In Press.
- Sidote, D. J. and D. W. Hoffman. 2003. "NMR structure of an archaeal homologue of ribonuclease P protein Rpp29." *Biochemistry* 42:13541-50.
- Siegel, R. W., A. B. Banta, E. S. Haas, J. W. Brown, and N. R. Pace. 1996. "Mycoplasma fermentans simplifies our view of the catalytic core of ribonuclease P RNA." *RNA* 2:452-62.
- Smith, D. and N. R. Pace. 1993. "Multiple magnesium ions in the ribonuclease P reaction mechanism." *Biochemistry* 32:5273-81.
- Smith, J. S. and M. J. Roth. 1993. "Purification and characterization of an active human immunodeficiency virus type 1 RNase H domain." *J Virol* 67:4037-49.
- Spera, S., M. Ikura, and A. Bax. 1991. "Measurement of the exchange rates of rapidly exchanging amide protons: application of calmodulin and its complex with a myosin light chain kinase fragment." *J Biomol NMR* 1:155-65.
- Spitzfaden, C., N. Nicholson, J. J. Jones, S. Guth, R. Lehr, C. D. Prescott, L. A. Hegg, and D. S. Eggleston. 2000. "The structure of ribonuclease P protein from *Staphylococcus aureus* reveals a unique binding site for single-stranded RNA." *J Mol Biol* 295:105-15.
- Sreerama, N., S. Y. Venyaminov, and R. W. Woody. 2000. "Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis." *Anal Biochem* 287:243-51.

- Srisawat, C., F. Houser-Scott, E. Bertrand, S. Xiao, R. H. Singer, and D. R. Engelke. 2002. "An active precursor in assembly of yeast nuclear ribonuclease P." *RNA* 8:1348-60.
- Stams, T., S. Niranjankumari, C. A. Fierke, and D. W. Christianson. 1998. "Ribonuclease P protein structure: evolutionary origins in the translational apparatus." *Science* 280:752-5.
- Steitz, T. A. and J. A. Steitz. 1993. "A general two-metal-ion mechanism for catalytic RNA." *Proc Natl Acad Sci U S A* 90:6498-502.
- Stolc, V. and S. Altman. 1997. "Rpp1, an essential protein subunit of nuclear RNase P required for processing of precursor tRNA and 35S precursor rRNA in *Saccharomyces cerevisiae*." *Genes Dev* 11:2926-37.
- Stolc, V., A. Katz, and S. Altman. 1998. "Rpp2, an essential protein subunit of nuclear RNase P, is required for processing of precursor tRNAs and 35S precursor rRNA in *Saccharomyces cerevisiae*." *Proc Natl Acad Sci U S A* 95:6716-21.
- Stribinskis, V., G. J. Gao, S. R. Ellis, and N. C. Martin. 2001. "Rpm2, the protein subunit of mitochondrial RNase P in *Saccharomyces cerevisiae*, also has a role in the translation of mitochondrially encoded subunits of cytochrome c oxidase." *Genetics* 158:573-85.
- Stribinskis, V., G. J. Gao, P. Sulo, Y. L. Dang, and N. C. Martin. 1996. "Yeast mitochondrial RNase P RNA synthesis is altered in an RNase P protein subunit mutant: insights into the biogenesis of a mitochondrial RNA-processing enzyme." *Mol Cell Biol* 16:3429-36.
- Svard, S. G., U. Kagardt, and L. A. Kirsebom. 1996. "Phylogenetic comparative mutational analysis of the base-pairing between RNase P RNA and its substrate." *RNA* 2:463-72.
- Takagi, H., M. Watanabe, Y. Kakuta, R. Kamachi, T. Numata, I. Tanaka, and M. Kimura. 2004. "Crystal structure of the ribonuclease P protein Ph1877p from hyperthermophilic archaeon *Pyrococcus horikoshii* OT3." *Biochem Biophys Res Commun* 319:787-94.

- Talbot, S. J. and S. Altman. 1994a. "Gel retardation analysis of the interaction between C5 protein and M1 RNA in the formation of the ribonuclease P holoenzyme from *Escherichia coli*." *Biochemistry* 33:1399-405.
- Talbot, S. J. and S. Altman. 1994b. "Kinetic and thermodynamic analysis of RNA-protein interactions in the RNase P holoenzyme from *Escherichia coli*." *Biochemistry* 33:1406-11.
- Tallsjo, A. and L. A. Kirsebom. 1993. "Product release is a rate-limiting step during cleavage by the catalytic RNA subunit of *Escherichia coli* RNase P." *Nucleic Acids Res* 21:51-7.
- Tallsjo, A., J. Kufel, and L. A. Kirsebom. 1996. "Interaction between *Escherichia coli* RNase P RNA and the discriminator base results in slow product release." *RNA* 2:299-307.
- Taylor, G. 2003. "The phase problem." *Acta Crystallogr D Biol Crystallogr* 59:1881-90.
- Thomas, B. C., X. Li, and P. Gegenheimer. 2000. "Chloroplast ribonuclease P does not utilize the ribozyme-type pre-tRNA cleavage mechanism." *RNA* 6:545-53.
- Thore, S., C. Mayer, C. Sauter, S. Weeks, and D. Suck. 2003. "Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya." *J Biol Chem* 278:1239-47.
- Tollervey, D. 1995. "Genetic and biochemical analyses of yeast RNase MRP." *Mol Biol Rep* 22:75-9.
- Toro, I., J. Basquin, H. Teo-Dreher, and D. Suck. 2002. "Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile *Archaeoglobus fulgidus*." *J Mol Biol* 320:129-42.
- Toro, I., S. Thore, C. Mayer, J. Basquin, B. Seraphin, and D. Suck. 2001. "RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex." *Embo J* 20:2293-303.

- Tranguch, A. J. and D. R. Engelke. 1993. "Comparative structural analysis of nuclear RNase P RNAs from yeast." *J Biol Chem* 268:14045-55.
- Tranguch, A. J., D. W. Kindelberger, C. E. Rohlman, J. Y. Lee, and D. R. Engelke. 1994. "Structure-sensitive RNA footprinting of yeast nuclear ribonuclease P." *Biochemistry* 33:1778-87.
- Tsai, H. Y., B. Masquida, R. Biswas, E. Westhof, and V. Gopalan. 2003. "Molecular modeling of the three-dimensional structure of the bacterial RNase P holoenzyme." *J Mol Biol* 325:661-75.
- Underbrink-Lyon, K., D. L. Miller, N. A. Ross, H. Fukuhara, and N. C. Martin. 1983. "Characterization of a yeast mitochondrial locus necessary for tRNA biosynthesis. Deletion mapping and restriction mapping studies." *Mol Gen Genet* 191:512-8.
- Valentin-Hansen, P., M. Eriksen, and C. Udesen. 2004. "The bacterial Sm-like protein Hfq: a key player in RNA transactions." *Mol Microbiol* 51:1525-33.
- Van Duyne, G.D., R.F. Standaert, P.A. Karplus, S.L. Schreiber, and J. Clardy. 1993. "Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin." *J Mol Biol* 229:105-24.
- van Eenennaam, H., D. Lugtenberg, J. H. Vogelzangs, W. J. van Venrooij, and G. J. Pruijn. 2001. "hPop5, a protein subunit of the human RNase MRP and RNase P endoribonucleases." *J Biol Chem* 276:31635-41.
- van Eenennaam, H., G. J. Pruijn, and W. J. van Venrooij. 1999. "hPop4: a new protein subunit of the human RNase MRP and RNase P ribonucleoprotein complexes." *Nucleic Acids Res* 27:2465-72.
- Wang, H. and E. R. Zuiderweg. 1995. "HCCH-TOCSY spectroscopy of ¹³C-labeled proteins in H₂O using heteronuclear cross-polarization and pulsed-field gradients." *J Biomol NMR* 5:207-11.
- Warnecke, J. M., J. P. Furste, W. D. Hardt, V. A. Erdmann, and R. K. Hartmann. 1996. "Ribonuclease P (RNase P) RNA is converted to a Cd(2+)-ribozyme by a single Rp-phosphorothioate modification in the precursor tRNA at the RNase P cleavage site." *Proc Natl Acad Sci U S A* 93:8924-8.

- Wedekind, J. E. and D. B. McKay. 2003. "Crystal structure of the leadzyme at 1.8 Å resolution: metal ion binding and the implications for catalytic mechanism and allo site ion regulation." *Biochemistry* 42:9554-63.
- Westhof, E. and S. Altman. 1994. "Three-dimensional working model of M1 RNA, the catalytic RNA subunit of ribonuclease P from *Escherichia coli*." *Proc Natl Acad Sci U S A* 91:5133-7.
- Westhof, E., D. Wesolowski, and S. Altman. 1996. "Mapping in three dimensions of regions in a catalytic RNA protected from attack by an Fe(II)-EDTA reagent." *J Mol Biol* 258:600-13.
- Wimberly, B. T., D. E. Brodersen, W. M. Clemons, Jr., R. J. Morgan-Warren, A. P. Carter, C. Vornrhein, T. Hartsch, and V. Ramakrishnan. 2000. "Structure of the 30S ribosomal subunit." *Nature* 407:327-39.
- Wise, C. A. and N. C. Martin. 1991. "Dramatic size variation of yeast mitochondrial RNAs suggests that RNase P RNAs can be quite small." *J Biol Chem* 266:19154-7.
- Wishart, D. S., C. G. Bigam, A. Holm, R. S. Hodges, and B. D. Sykes. 1995. "¹H, ¹³C and ¹⁵N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects." *J Biomol NMR* 5:67-81.
- Wishart, D. S. and B. D. Sykes. 1994. "The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data." *J Biomol NMR* 4:171-80.
- Wuthrich, K. 1986. "NMR of proteins and nucleic acids." John Wiley & Sons Inc.
- Xiao, S., F. Scott, C. A. Fierke, and D. R. Engelke. 2002. "Eukaryotic ribonuclease P: a plurality of ribonucleoprotein enzymes." *Annu Rev Biochem* 71:165-89.
- Yuan, Y., E. Tan, and R. Reddy. 1991. "The 40-kilodalton to autoantigen associates with nucleotides 21 to 64 of human mitochondrial RNA processing/7-2 RNA in vitro." *Mol Cell Biol* 11:5266-74.

- Zahler, N. H., E. L. Christian, and M. E. Harris. 2003. "Recognition of the 5' leader of pre-tRNA substrates by the active site of ribonuclease P." *RNA* 9:734-45.
- Ziehler, W. A., J. Morris, F. H. Scott, C. Millikin, and D. R. Engelke. 2001. "An essential protein-binding domain of nuclear RNase P RNA." *RNA* 7:565-75.
- Ziehler, W. A., J. Yang, A. V. Kurochkin, P. O. Sandusky, E. R. Zuiderweg, and D. R. Engelke. 1998. "Structural analysis of the P10/11-P12 RNA domain of yeast RNase P RNA and its interaction with magnesium." *Biochemistry* 37:3549-57.
- Zimmerly, S., V. Gamulin, U. Burkard, and D. Soll. 1990. "The RNA component of RNase P in *Schizosaccharomyces* species." *FEBS Lett* 271:189-93.
- Zuiderweg, E. R. P. 1990. "A proton-detected heteronuclear chemical-shift correlation experiment with improved resolution and sensitivity." *J. Magn. Reson.* 86:346-352.

VITA

David Joseph Sidote was born in New Brunswick, NJ on July 28, 1971, the son of Ruth and William Sidote. After graduating from St. Joseph's High School, Metuchen, in 1989, he decided to work before entering college. In 1994, he entered Rutgers University in New Brunswick, NJ and received the degree of Bachelor of Arts in May 1998. During the following years he was employed as an associate scientist at the R.W. Johnson Pharmaceutical Research Institute and then as a technician at the Center for Advanced Biotechnology and Medicine in Piscataway, NJ. After his supervisor convinced him that he would do well to obtain a PhD., he entered the Graduate School of the University of Texas in September of 2000.

Permanent Address: 31 Plennert Road, Flemington, NJ 08822

This dissertation was typed by the author.